

职业技能等级认定指导手册

商务数据分析师 (三级)

上海市职业技能鉴定中心 指导编写

目 录

商务数据分析师职业简介·····	(1)
第 1 部分 商务数据分析师(三级)认定方案·····	(2)
第 2 部分 认定要素细目表·····	(4)
第 3 部分 理论知识复习题·····	(20)
第 4 部分 操作技能复习题·····	(72)
第 5 部分 理论知识考试模拟试卷及答案·····	(232)
第 6 部分 操作技能考核模拟试卷·····	(244)

商务数据分析师职业简介

一、职业名称

商务数据分析师

二、职业定义

从事商务行为相关数据采集、清洗、挖掘、分析，发现问题、研判规律，形成数据分析报告并指导他人应用的人员。

三、主要工作内容

商务数据分析师的工作内容围绕数据采集与处理、数据挖掘与分析、数据可视化、商务数据分析、贸易数据申报与报告撰写五大核心模块展开，覆盖从基础操作到高级决策支持的全流程。

四、职业等级

本职业共设四个等级，分别为：四级/中级工、三级/高级工、二级/技师、一级/高级技师。

职业技能等级认定评价项目评价方案

商务数据分析师（三级）

一、评价方式

商务数据分析师（三级）的评价方式分为理论知识考试和操作技能考核。理论知识考试采用闭卷机考方式，操作技能考核采用计算机模拟操作考核方式。理论知识考试和操作技能考核均实行百分制，成绩皆达 60 分及以上者为合格。理论知识或操作技能不及格者可按规定分别补考。

二、理论知识考试方案（考试时间 90 分钟）

题型	题库参数	考试方式	题库量	考试题量	分值	配分
判断题		闭卷机考	300	40	0.5 分/题	20
单选题			600	140	0.5 分/题	70
多选题			300	10	1 分/题	10
合计		—	1200	190	—	100

三、操作技能考核方案（考试时间 120 分钟）

考核项目表

职业（工种）		商务数据分析师			等级	三级			
职业代码		4-07-02-05				考核方式	考核时间（分钟）	配分	题库题量
1	数据采集、清洗和处理	1	采集、清洗数据	操作	抽一	20	20	5	1
		2	查询数据	操作		20	20	5	
2	数据挖掘和分析	1	开展数据预处理、数据建模	操作	必考	15	10	5	1
		2	开展数据计算、开展数据挖掘	操作	抽一	15	10	5	
		3	开展数据分析	操作		15	10	5	

3	数据可视化	1	设计数据可视化方案	操作	必考	15	10	5	1
		2	分析数据可视化结果	操作	必考	10	10	5	1
4	商务分类数据分析、数据申报和报告缮制	1	分析行业和竞争数据并缮制商务数据分析报告	操作	必考	25	20	5	1
		2	采集申报贸易数据及开展波特五力模型的应用分析	操作	必考	20	20	5	1
合 计						120	100	45	7
备注									

商务数据分析师（三级） 理论知识考试要素细目表

职业（工种）名称					商务数据分析师	等级	三级
序号	细目点代码				名称·内容	分数系数	备注
	章	节	目	点			
	0				基本要求	20	
	0	1			职业道德	5	
	0	1	1		职业道德基本知识	3	
1	0	1	1	1	广义职业道德概念		
2	0	1	1	2	狭义职业道德概念		
3	0	1	1	3	职业道德与企业文化的关系		
4	0	1	1	4	商务数据分析师职业的定义		
5	0	1	1	5	商务数据分析师职业的特点		
6	0	1	1	6	数字伦理与数据责任		
7	0	1	1	7	社会责任的内涵及关切点		
8	0	1	1	8	商务数据分析师的社会责任		
9	0	1	1	9	职业道德在商务数据分析中的功能		
	0	1	2		职业守则	2	
10	0	1	2	1	尊重与保护用户数据		
11	0	1	2	2	知识产权保护与合规使用		
12	0	1	2	3	遵守行业规范与自律		
13	0	1	2	4	保守商业机密		
14	0	1	2	5	精益求精的方法		
15	0	1	2	6	忠于职守的关键点		
	0	2			基础知识	15	
	0	2	1		理论基础知识	3	
16	0	2	1	1	操作系统的概念		
17	0	2	1	2	操作系统的目标		

18	0	2	1	3	操作系统的主要功能		
19	0	2	1	4	计算机网络的概念		
20	0	2	1	5	计算机网络的分类		
21	0	2	1	6	计算机网络的功能		
22	0	2	1	7	数据库及数据库系统的概念		
23	0	2	1	8	数据库原理及技术特点		
24	0	2	1	9	数据库系统的构成		
	0	2	2		技术基础知识	4	
25	0	2	2	1	统计学的研究对象		
26	0	2	2	2	统计学的基本特点		
27	0	2	2	3	统计学的基本方法		
28	0	2	2	4	数据分析的功能		
29	0	2	2	5	数据分析的原则		
30	0	2	2	6	数据分析方法		
31	0	2	2	7	常见离散和连续分布的特性		
32	0	2	2	8	数据管理的概念		
33	0	2	2	9	数据治理的意义与功能		
34	0	2	2	10	数据治理的三个阶段		
35	0	2	2	11	数据治理框架		
36	0	2	2	12	数据流程优化方法		
	0	2	3		安全知识	2	
37	0	2	3	1	数据安全的定义		
38	0	2	3	2	数据安全的关键要素		
39	0	2	3	3	数据安全的功能		
40	0	2	3	4	数据服务安全架构		
41	0	2	3	5	数据泄露防护（DLP）策略		
42	0	2	3	6	数据安全管理的原则		
	0	2	4		商务基础知识	4	

43	0	2	4	1	商品的定义与基本属性		
44	0	2	4	2	商品的四大要素		
45	0	2	4	3	商品生命周期管理方法		
46	0	2	4	4	营销组合策略		
47	0	2	4	5	营销策划的特点与原则		
48	0	2	4	6	市场营销策划的基本功能		
49	0	2	4	7	商贸流通的特点与原则		
50	0	2	4	8	商品流通的主要渠道和模式		
51	0	2	4	9	企业管理的原则		
52	0	2	4	10	企业管理的职能		
53	0	2	4	11	企业组织结构的设计原则		
54	0	2	4	12	成本控制的原则		
	0	2	5		相关法律、法规知识	2	
55	0	2	5	1	《中华人民共和国电子商务法》亮点及特点		
56	0	2	5	2	《中华人民共和国保守国家秘密法》亮点及特点		
57	0	2	5	3	《中华人民共和国数据安全法》亮点及特点		
58	0	2	5	4	《中华人民共和国个人信息保护法》亮点及特点		
59	0	2	5	5	《中华人民共和国网络安全法》亮点及特点		
60	0	2	5	6	《全国人民代表大会常务委员会关于加强网络信息保护的決定》亮点及特点		
	1				数据采集、清洗和处理	25	
	1	1			采集、清洗数据	18	
	1	1	1		信息系统配置知识	2	
61	1	1	1	1	信息系统的定义		
62	1	1	1	2	信息系统的功能		
63	1	1	1	3	数据采集策略		
64	1	1	1	4	数据清洗规则		

65	1	1	1	5	数据清洗流程		
66	1	1	1	6	数据质量控制与监控规则		
	1	1	2		数据监测与迁移知识	4	
67	1	1	2	1	数据监测的定义		
68	1	1	2	2	数据监测的目的		
69	1	1	2	3	数据监测指标定义及设计原则		
70	1	1	2	4	数据异常监测与预警机制		
71	1	1	2	5	实时数据监测与处理机制		
72	1	1	2	6	数据迁移的基本概念		
73	1	1	2	7	数据迁移的基本类型		
74	1	1	2	8	数据迁移的策略		
75	1	1	2	9	数据迁移工具使用注意事项		
76	1	1	2	10	数据迁移中的数据一致性保障方法		
77	1	1	2	11	数据迁移的性能优化方法		
78	1	1	2	12	数据安全性与合规注意事项		
	1	1	3		数据存储策略知识	4	
79	1	1	3	1	数据存储类型		
80	1	1	3	2	数据备份与恢复方法		
81	1	1	3	3	数据冗余与去重方法		
82	1	1	3	4	数据生命周期管理策略		
83	1	1	3	5	数据安全性与加密使用注意事项		
84	1	1	3	6	数据压缩与解压缩使用注意事项		
85	1	1	3	7	分布式存储技术特点		
86	1	1	3	8	数据访问控制技术特点		
87	1	1	3	9	数据存储的目标		
88	1	1	3	10	数据存储的原则		
89	1	1	3	11	数据存储的流程		
90	1	1	3	12	数据存储性能评估方法		

	1	1	4		数据清洗需求分析方法	4	
91	1	1	4	1	数据清洗需求分析的定义		
92	1	1	4	2	数据质量问题识别方法		
93	1	1	4	3	清洗数据价值评估策略		
94	1	1	4	4	数据清洗预期目标设定方法		
95	1	1	4	5	数据清洗关联业务指标		
96	1	1	4	6	数据源与数据结构的分析方法		
97	1	1	4	7	数据清洗工具使用注意事项		
98	1	1	4	8	数据完整性分析方法		
99	1	1	4	9	数据准确性分析方法		
100	1	1	4	10	数据一致性分析方法		
101	1	1	4	11	数据缺失值识别方法		
102	1	1	4	12	数据异常值识别方法		
	1	1	5		数据清洗方案设置知识	4	
103	1	1	5	1	数据清洗方案的概念		
104	1	1	5	2	数据清洗需求的分析		
105	1	1	5	3	数据清洗工具的选择依据		
106	1	1	5	4	数据清洗规则的设定规则		
107	1	1	5	5	数据清洗流程的设计规则		
108	1	1	5	6	数据清洗中的质量控制方法		
109	1	1	5	7	数据清洗方案的评估方法		
110	1	1	5	8	数据清洗中的异常处理方法		
111	1	1	5	9	数据清洗与数据治理方法		
112	1	1	5	10	数据清洗的版本控制与回溯方法		
113	1	1	5	11	数据归一化处理方法		
114	1	1	5	12	数据合并处理方法		
	1	2			查询数据	7	
	1	2	1		查询引擎知识	4	

115	1	2	1	1	查询引擎基础概念		
116	1	2	1	2	SQL: BI 查询引擎优缺点		
117	1	2	1	3	MDX: 多维数据查询引擎优缺点		
118	1	2	1	4	DAX: 数据模型查询引擎优缺点		
119	1	2	1	5	Spark SQL: 大数据处理查询引擎优缺点		
120	1	2	1	6	Presto: 交互式查询的大数据查询引擎优缺点		
121	1	2	1	7	FineReport 和 FineVis: 多种查询引擎的 BI 工具优缺点		
122	1	2	1	8	BI 工具选择查询引擎的考虑因素		
123	1	2	1	9	查询引擎优化和性能调优方法		
124	1	2	1	10	查询引擎的安全性和数据保护方法		
125	1	2	1	11	查询性能评估指标项		
126	1	2	1	12	查询优化原理		
	1	2	2		数据分词知识	3	
127	1	2	2	1	数据分词基础概念		
128	1	2	2	2	分词工具与库使用注意事项		
129	1	2	2	3	基于字符串匹配的分词方法		
130	1	2	2	4	基于统计的分词方法		
131	1	2	2	5	基于理解的分词方法		
132	1	2	2	6	基于深度学习的分词方法		
133	1	2	2	7	自定义词典与规则的基本概念		
134	1	2	2	8	分词粒度控制测量		
135	1	2	2	9	分词效果评估策略		
	2				数据挖掘和分析	20	
	2	1			开展数据预处理	4	
	2	1	1		数据库基础操作知识	2	
136	2	1	1	1	数据库的创增查选改删注意事项		
137	2	1	1	2	数据表的创增查改名删注意事项		

138	2	1	1	3	基本查询语句注意事项		
139	2	1	1	4	数据库单表查询注意事项		
140	2	1	1	5	数据库聚合函数查询注意事项		
141	2	1	1	6	数据库连接查询和复合条件查询注意事项		
	2	1	2		数据结构化与非结构化知识	2	
142	2	1	2	1	结构化数据的定义		
143	2	1	2	2	结构化数据的特点		
144	2	1	2	3	非结构化数据的定义		
145	2	1	2	4	非结构化数据的特点		
146	2	1	2	5	结构化数据的应用场景		
147	2	1	2	6	非结构化数据的应用场景		
	2	2			开展数据建模	5	
	2	2	1		层次建模知识	2	
148	2	2	1	1	层次数据模型的概念		
149	2	2	1	2	层次数据模型的特征		
150	2	2	1	3	层次建模的优缺点		
151	2	2	1	4	层次建模的基本要素		
152	2	2	1	5	层次建模的方法		
153	2	2	1	6	层次建模的存储结构		
	2	2	2		维度建模知识	2	
154	2	2	2	1	维度建模的概念		
155	2	2	2	2	维度建模的特征		
156	2	2	2	3	维度建模的优缺点		
157	2	2	2	4	维度模型的基本要素		
158	2	2	2	5	维度模型的方法		
159	2	2	2	6	维度建模的存储结构		
	2	2	3		读写性能知识	1	
160	2	2	3	1	数据存储的原理		

161	2	2	3	2	数据存储的三类方式		
162	2	2	3	3	数据读写性能的指标		
	2	3			开展数据计算	3	
	2	3	1		常用算法与数据结构知识	2	
163	2	3	1	1	数据的常用算法优缺点		
164	2	3	1	2	数据的常用结构优缺点		
165	2	3	1	3	决策树算法及其应用方法		
166	2	3	1	4	支持向量机（SVM）与核函数方法		
167	2	3	1	5	Python 内置的数据结构		
168	2	3	1	6	排序算法知识		
	2	3	2		数据画像知识	1	
169	2	3	2	1	数据画像的概念		
170	2	3	2	2	数据画像的构建流程		
171	2	3	2	3	数据画像中的标签体系		
	2	4			开展数据挖掘	3	
	2	4	1		特征工程知识	2	
172	2	4	1	1	特征工程的概念		
173	2	4	1	2	特征工程的意义		
174	2	4	1	3	特征选择方法		
175	2	4	1	4	特征提取方法		
176	2	4	1	5	缺失值处理策略方法		
177	2	4	1	6	特征构造与变换方法		
	2	4	2		数据挖掘类库应用知识	1	
178	2	4	2	1	数据挖掘类库的基本概念		
179	2	4	2	2	python 数据挖掘类库的作用		
180	2	4	2	3	python 数据挖掘类库的分类		
	2	5			开展数据分析	5	
	2	5	1		概要、描述性统计分析知识	1	

181	2	5	1	1	概要统计的基本概念		
182	2	5	1	2	概要统计目的		
183	2	5	1	3	描述性统计分析的指标体系		
	2	5	2		统计工具使用知识	3	
184	2	5	2	1	统计工具的基本概念		
185	2	5	2	2	统计工具的分类		
186	2	5	2	3	Excel 统计工具使用知识注意点		
187	2	5	2	4	Python 统计使用知识注意点		
188	2	5	2	5	SPSS 统计使用知识注意点		
189	2	5	2	6	Power BI 统计使用知识注意点		
190	2	5	2	7	Stata 统计使用知识注意点		
191	2	5	2	8	SAS 统计使用知识注意点		
192	2	5	2	9	MATLAB 统计使用知识注意点		
	2	5	3		线性相关及回归相关知识	1	
193	2	5	3	1	线性相关的概念、特点与内容		
194	2	5	3	2	一元线性回归的概念与方程		
195	2	5	3	3	多元线性回归的概念与模型		
	3				数据可视化	10	
	3	1			设计数据可视化方案	7	
	3	1	1		数据可视化设计知识	4	
196	3	1	1	1	数据可视化的基本概念		
197	3	1	1	2	视觉编码原则		
198	3	1	1	3	数据可视化设计原则		
199	3	1	1	4	常见数据可视化类型		
200	3	1	1	5	数据可视化工具的使用注意事项		
201	3	1	1	6	数据可视化设计的色彩理论		
202	3	1	1	7	图表布局与排版注意事项		
203	3	1	1	8	交互性设计规则		

204	3	1	1	9	数据可视化质量评估标准		
205	3	1	1	10	数据可视化中的视觉通道特点		
206	3	1	1	11	图标类型与数据匹配原则		
207	3	1	1	12	数据可视化设计的准确性原则		
	3	1	2		可视化组件库知识	3	
208	3	1	2	1	可视化组件库的基本概念		
209	3	1	2	2	组件库的分类与结构		
210	3	1	2	3	基础图表组件的使用注意事项		
211	3	1	2	4	高级分析组件的使用注意事项		
212	3	1	2	5	交互组件的使用注意事项		
213	3	1	2	6	组件的属性与参数设置注意事项		
214	3	1	2	7	组件库的集成与使用注意事项		
215	3	1	2	8	组件的性能优化方法		
216	3	1	2	9	组件的兼容性与适配方法		
	3	2			分析数据可视化结果	3	
	3	2	1		数据可视化结果业务分析知识	3	
217	3	2	1	1	业务指标规则		
218	3	2	1	2	对比能力分析方法		
219	3	2	1	3	趋势预测能力方法		
220	3	2	1	4	因果关系推断方法		
221	3	2	1	5	异常值与离群点分析方法		
222	3	2	1	6	数据分布与统计特征分析方法		
223	3	2	1	7	数据可视化结果在业务决策中的应用注意事项		
224	3	2	1	8	数据可视化结果与业务 KPI 的关联分析注意事项		
225	3	2	1	9	数据可视化结果的质量评估方法		
	4				商务分类数据分析、数据申报和报告编制	25	
	4	1			分析行业数据	4	

	4	1	1		市场分析知识	2	
226	4	1	1	1	市场行情分析的常用指标		
227	4	1	1	2	市场行情数据的获取方式		
228	4	1	1	3	市场供给与需求分析方法		
229	4	1	1	4	市场容量与趋势分析方法		
230	4	1	1	5	市场潜力分析方法		
231	4	1	1	6	市场的划分分类		
	4	1	2		行业分析知识	2	
232	4	1	2	1	行业状况分析的常用指标		
233	4	1	2	2	行业状况数据的获取方式		
234	4	1	2	3	行业状况分析的常用方法		
235	4	1	2	4	行业的一般特征分析方法		
236	4	1	2	5	行业结构及发展分析方法		
237	4	1	2	6	行业生命周期分析概念		
	4	2			分析竞争数据	6	
	4	2	1		竞争对手分析知识	3	
238	4	2	1	1	竞争对手的识别方法		
239	4	2	1	2	竞争对手的分类		
240	4	2	1	3	竞争对手分析的主要指标		
241	4	2	1	4	竞争对手分析的意义		
242	4	2	1	5	竞争对手分析的常用方法		
243	4	2	1	6	竞争对手下单转化率分析方法		
244	4	2	1	7	竞争对手客户拥有量分析方法		
245	4	2	1	8	竞争对手 SWOT 分析方法		
246	4	2	1	9	竞争对手合作与联盟策略		
	4	2	2		竞争产品分析知识	3	
247	4	2	2	1	竞争产品识别与分类方法		
248	4	2	2	2	竞争产品特性分析方法		

249	4	2	2	3	竞争产品定价策略分析方法		
250	4	2	2	4	竞争产品分销渠道分析方法		
251	4	2	2	5	竞争产品促销策略分析方法		
252	4	2	2	6	竞争产品市场份额与增长率分析方法		
253	4	2	2	7	竞争产品客户反馈与评价分析方法		
254	4	2	2	8	竞争产品生命周期分析方法		
255	4	2	2	9	竞争产品供应链分析方法		
	4	3			开展波特五力模型的应用分析，波特五力模型知识	3	
	4	3	1		波特五力模型知识	1	
256	4	3	1	1	波特五力模型概念		
257	4	3	1	2	波特五力模型的使用步骤注意事项		
258	4	3	1	3	波特五力模型的动态性特征		
	4	3	2		开展波特五力模型的应用分析	2	
259	4	3	2	1	新进入者威胁分析方法		
260	4	3	2	2	替代品竞争评估与应对策略		
261	4	3	2	3	买方议价能力的多维分析方法		
262	4	3	2	4	卖方议价力量分析方法		
263	4	3	2	5	现有竞争者分析与监测方法		
264	4	3	2	6	波特五力模型的功能		
	4	4			采集申报贸易数据	10	
	4	4	1		中国国际贸易单一窗口知识	5	
265	4	4	1	1	单一窗口的概念		
266	4	4	1	2	单一窗口的功能		
267	4	4	1	3	单一窗口作用和意义		
268	4	4	1	4	跨境电子商务数据申报基础概念		
269	4	4	1	5	跨境电子商务数据申报流程		
270	4	4	1	6	跨境电子商务数据申报注意事项		
271	4	4	1	7	出口退税数据申报流程		

272	4	4	1	8	外贸企业出口退税申报流程		
273	4	4	1	9	生产企业出口退税申报流程		
274	4	4	1	10	口岸物流的内涵		
275	4	4	1	11	口岸物流的特点		
276	4	4	1	12	口岸物流的功能		
277	4	4	1	13	口岸物流数据申报流程		
278	4	4	1	14	电子口岸的概念		
279	4	4	1	15	电子口岸物流数据申报的注意事项		
	4	4	2		跨境电子商务知识	2	
280	4	4	2	1	跨境电子商务的定义		
281	4	4	2	2	跨境电子商务的特点		
282	4	4	2	3	跨境电子商务的交易模式		
283	4	4	2	4	跨境电子商务的物流方式		
284	4	4	2	5	跨境电子商务的支付与结算的注意事项		
285	4	4	2	6	跨境电子商务的风险与应对的注意事项		
	4	4	3		出口退税知识	3	
286	4	4	3	1	出口退税的概念		
287	4	4	3	2	出口退税的税种		
288	4	4	3	3	出口退税的类型		
289	4	4	3	4	出口退税的适用范围		
290	4	4	3	5	出口退税的基本政策		
291	4	4	3	6	出口退税的操作流程		
292	4	4	3	7	出口退税的计算方法		
293	4	4	3	8	出口退税的申报方式		
294	4	4	3	9	出口退税风险防范方法		
	4	5			缮制商务数据分析报告	2	
	4	5	1		商务数据分析报告撰写要求与规范、 分析报告框架构建知识	2	
295	4	5	1	1	商务数据分析报告撰写的原则		

296	4	5	1	2	商务数据分析报告撰写的常见技巧		
297	4	5	1	3	缮制商务数据分析报告的常见错误		
298	4	5	1	4	商务数据分析报告框架的基本构成		
299	4	5	1	5	商务数据分析报告框架的设计原则		
300	4	5	1	6	商务数据分析报告框架设计的注意事项		

商务数据分析师（三级） 操作技能考核要素细目表

职业（工种）名称				商务数据分析师	等级	三级
序号	细目点代码			名称·内容	备注	
	项目	单元	细目			
	1			数据采集、清洗和处理		
	1	1		采集、清洗数据		
1	1	1	1	根据业务需求进行在线、离线数据采集		
2	1	1	2	根据业务需求参与制定数据迁移方案		
3	1	1	3	根据业务特性制订数据存储策略		
4	1	1	4	根据数据质量要求制定数据清洗策略及评估方案		
	1	2		查询数据		
5	1	2	1	使用计算引擎优化数据查询效率		
6	1	2	2	通过计算平台构建检索分析系统		
7	1	2	3	同构计算平台显示检索分析结果		
	2			数据挖掘和分析		
	2	1		开展数据预处理、数据建模		
8	2	1	1	根据业务需求对多源数据进行整合		
9	2	1	2	根据业务规则对数据格式进行转换		
10	2	1	3	根据数据归一性原则对数据进行单位、数值规约		
11	2	1	4	根据数据建模规范设计数据模型		
12	2	1	5	根据业务需求对数据模型进行优化		
	2	2		开展数据计算、开展数据挖掘		
13	2	2	1	根据业务需求编写批量数据计算作业并优化作业参数		
14	2	2	2	根据业务规则设计相应数据指标计算算法		
15	2	2	3	评估挖掘需求并使用工具对数据进行特征工程处理		
16	2	2	4	调用常规模型进行模型训练		
17	2	2	5	基于评价指标进行模型测试		

	2	3		开展数据分析	
18	2	3	1	结合业务场景使用工具对数据集进行概要、描述性统计分析	
19	2	3	2	在描述结果基础上对数据进行特征和规律的分析与推测	
20	2	3	3	结合业务场景编写数据统计报告	
	3			数据可视化	
	3	1		设计数据可视化方案	
21	3	1	1	设计数据可视化实现方式	
22	3	1	2	设计可视化展板	
23	3	1	3	根据业务需求及分析结果制定数据展示方案	
	3	2		分析数据可视化结果	
24	3	2	1	对数据可视化结果进行业务分析	
25	3	2	2	输出分析报告	
26	3	2	3	撰写数据驱动决策	
	4			商务分类数据分析、数据申报和报告缮制	
	4	1		分析行业和竞争数据并缮制商务数据分析报告	
27	4	1	1	进行市场行情分析	
28	4	1	2	进行行业状况分析	
29	4	1	3	进行竞争对手分析	
30	4	1	4	进行竞争产品分析	
31	4	1	5	撰写商务数据分析报告	
	4	2		采集申报贸易数据及开展波特五力模型的应用分析	
32	4	2	1	进行跨境电子商务数据申报	
33	4	2	2	进行出口退税数据申报	
34	4	2	3	进行口岸物流数据申报	
35	4	2	4	画出波特五力模型图	
36	4	2	5	进行波特五力分析	

第3部分

理论知识复习题

一、判断题（将判断结果填入括号中。正确的填“√”，错误的填“×”）

1. 计算机网络中的负载均衡功能是指将网络流量均匀分配到不同的网络设备上，以提高网络性能。（ ）
2. 数据库系统就是存储数据的仓库，不包括其他组成部分。（ ）
3. 数据库系统只能存储结构化数据，无法处理非结构化或半结构化数据。（ ）
4. 数据库系统的核心组成部分是数据库和数据库管理系统。（ ）
5. 统计学的研究对象仅限于数值型数据，不包括非数值型数据。（ ）
6. 统计学的基本特点之一是只关注数据的收集，而不关注数据的解释和应用。（ ）
7. 相关性分析可以确定两个变量之间的因果关系。（ ）
8. 数据分析的功能之一是帮助企业识别新的市场机会和潜在的增长点。（ ）
9. 数据分析的透明度意味着分析过程、方法和结果都应该清晰、可验证。（ ）
10. 相关性分析可以揭示两个变量之间的因果关系。（ ）
11. 离散分布只能描述离散变量，连续分布只能描述连续变量。（ ）
12. 数据管理仅仅是对数据进行存储和备份的过程。（ ）
13. 数据治理只是对数据进行管理和规范，与企业的业务战略无关。（ ）
14. 数据治理的第一个阶段主要是建立数据架构和组织，不涉及具体的数据管理活动。（ ）
15. 数据治理框架的实施只需要技术层面的支持，不需要高层管理的参与。（ ）
16. 在数据流程优化过程中，对流程进行定期评估和监控的目的是识别新的改进机会。（ ）
17. 数据安全仅关注数据的机密性，而不涉及数据的完整性和可用性。（ ）
18. 数据安全策略中的访问控制仅涉及对用户身份的验证，而不包括对用户权限的管理。（ ）
19. 数据审计功能可以记录和追踪数据的操作历史，以便在发生安全事件时进行追溯和调查。（ ）
20. 在设计数据服务安全架构时，不需要考虑数据的生命周期管理。（ ）
21. DLP 策略只能防止内部员工造成的数据泄露，无法防御外部攻击。（ ）
22. 数据安全管理的可审计性原则要求对数据的操作进行记录和追踪，以便在发生安全事件时能够追溯和调查。（ ）
23. 商品的使用价值是指其能够满足人们某种需要的属性。（ ）
24. 商品的核心利益是消费者购买商品时最直接和主要的需求满足点。（ ）
25. 在商品生命周期的衰退期应该积极收集客户反馈，对商品进行持续改进和优化。（ ）
26. 营销组合策略（ ）包括产品、价格、促销和渠道。（ ）
27. 营销策划一旦确定，就不需要再进行调整和修改，以保持策略的稳定性。（ ）

- 28 . 分销策略是市场营销策划中的关键环节，它决定了如何将产品有效地送达目标消费者手中。()
- 29 . 商贸流通的广泛性意味着它只涉及国内市场的商品交换。()
- 30 . 分销模式总是包括多个层级的分销商。()
- 31 . 系统性原则要求企业管理中的各个部分都要相互协调，形成一个有机整体。()
- 32 . 计划职能只是为企业设定长期目标，不涉及日常运营。()
- 33 . 企业组织结构的设计应具有灵活性，能够随企业战略和市场环境的变化而调整。()
- 34 . 成本控制就是简单地降低生产成本。()
- 35 . 《电子商务法》要求平等地对待线上线下的商务活动，促进线上线下融合发展。()
- 36 . 《中华人民共和国保守国家秘密法》规定，涉密人员在脱密期内，可以以任何方式泄露国家秘密。()
- 37 . 《中华人民共和国数据安全法》要求数据处理者未经授权处理个人信息，或处理超过个人信息范围的其他信息，均要承担法律责任。()
- 38 . 《中华人民共和国个人信息保护法》对人脸识别的使用没有任何限制。()
- 39 . 《中华人民共和国网络安全法》不适用于个人用户在网络上的行为。()
- 40 . 即使未经电子信息接收者同意或请求，也可以偶尔向其固定电话、移动电话或个人电子邮箱发送商业性电子信息。()
- 41 . 信息系统只是技术工具，不涉及人与人之间的直接交流过程。()
- 42 . 信息系统只能处理结构化数据，无法处理非结构化数据。()
- 43 . 数据采集的起点是明确数据采集的目标和需求。()
- 44 . 数据编码转换是数据清洗过程中的一个必要步骤，以确保数据在不同系统或平台间的兼容性和一致性。()
- 45 . 数据清洗是一次性过程，不需要在数据生命周期中重复进行。()
- 46 . 数据质量控制与监控是数据管理中的一次性活动，不需要持续进行。()
- 47 . 数据监测不仅关注数据的收集，还涉及数据的后续处理和分析。()
- 48 . 数据监测的唯一目的是为了企业的利润率。()
- 49 . 设计数据监测指标时，不必考虑指标之间的逻辑关系。()
- 50 . 实施数据异常监测与预警，简单设置阈值即可，无需复杂算法支持。()
- 51 . 实时数据监测与处理不可以完全替代传统的批处理数据处理方式。()
- 52 . 在数据迁移过程中，保障数据的完整性和一致性并非必须。()
- 53 . 有类型的数据迁移都涉及到数据的结构变更。()
- 54 . 所有的数据迁移都应该选择一次性大规模停机的方式进行，以确保数据的完整性。()
- 55 . 数据迁移工具的使用无需考虑其对生产环境的影响，只需关注迁移效果即可。()
- 56 . 数据迁移时，只要迁移了所有数据，就可以保证数据的一致性。()
- 57 . 数据迁移的性能优化仅与迁移工具的选择有关，与其他因素无关。()
- 58 . 数据迁移过程中，只要使用了加密技术，就可以完全保证数据的安全与合规。()
- 59 . 不是所有的数据存储类型在访问速度、存储容量和成本方面都是相同的。()
- 60 . 数据备份是可选的，因为它占用了宝贵的存储资源。()
- 61 . 数据去重只应在数据备份场景中使用，以提高备份效率。()

- 62 . 数据生命周期管理策略不是一成不变的, 需要随着业务需求和技术环境的变化进行调整。()
- 63 . 加密是确保数据安全的万无一失的方法, 不需要其他安全措施作为补充。()
- 64 . 数据压缩总是能显著减少存储空间需求, 且不会对数据访问性能产生负面影响。()
- 65 . 分布式存储系统由于将数据分散存储在多个节点上, 因此比传统存储系统更容易受到网络延迟和带宽限制的影响。()
- 66 . 数据访问控制技术不可以完全消除数据泄露的风险。()
- 67 . 数据存储的唯一目标是确保数据的完整性, 其他考虑因素如成本、性能和安全性都不重要。()
- 68 . 数据存储的原则要求所有数据都必须存储在同一类型的介质上, 以确保一致性。()
- 69 . 数据存储流程的第一步总是数据采集。()
- 70 . 评估数据存储性能时, 仅考虑读写速度就足够了。()
- 71 . 数据清洗需求分析是数据预处理阶段不可或缺的一部分。()
- 72 . 识别数据质量问题时, 仅依靠技术手段就足够了, 无需考虑业务背景和实际需求。()
- 73 . 清洗数据的价值只能通过其直接提升的数据质量来评估, 无需考虑其对业务的具体贡献。()
- 74 . 设定数据清洗预期目标时, 只需关注技术层面的要求, 无需考虑业务层面的需求。()
- 75 . 关联数据清洗与业务指标是确保数据清洗工作能够为业务带来实际价值的关键步骤。()
- 76 . 数据源分析只需要关注数据的量, 而无需考虑数据的质量。()
- 77 . 数据清洗工具可以自动发现并解决所有数据质量问题。()
- 78 . 数据完整性分析只需要关注数据是否完整, 无需考虑数据的准确性和一致性。()
- 79 . 数据准确性分析可以完全依赖自动化工具进行, 无需人工干预。()
- 80 . 数据一致性分析只需要关注数据在单个系统内的表现, 无需考虑跨系统的一致性。()
- 81 . 数据缺失值总是可以通过简单的统计函数()完全识别出来。()
- 82 . 所有超出平均值一定范围的数据点都应被视为异常值。()
- 83 . 数据清洗方案仅关注于纠正数据中的错误, 不涉及数据质量的整体提升。()
- 84 . 数据清洗需求分析是制定数据清洗方案的第一步, 也是最重要的一步。()
- 85 . 数据清洗工具的选择应完全基于其价格, 价格越低越好。()
- 86 . 数据清洗规则的设定应尽可能复杂, 以覆盖所有可能的数据质量问题。()
- 87 . 数据清洗流程的设计是一个迭代的过程, 随着数据质量问题的不断变化和业务需求的更新, 流程也需要进行相应的修改和优化。()
- 88 . 数据清洗的质量控制完全依赖于数据清洗工具的高级功能, 而不需要人工参与。()
- 89 . 数据清洗方案的评估应仅在清洗过程完成后进行, 以确保评估结果的准确性。()
- 90 . 异常值的处理应基于数据特性和业务需求, 有时保留异常值并分析其原因也是必要的。()

- 91 . 数据治理可以完全替代数据清洗，因为数据治理涵盖了数据生命周期的所有方面。()
- 92 . 版本控制不仅有助于确保数据的一致性，还能帮助追踪数据清洗过程中的变化，快速定位并解决问题，从而支持异常处理和数据质量的持续提升。()
- 93 . 归一化主要调整数据的尺度，但不一定改变其分布形态。()
- 94 . 在数据合并过程中，只要字段名相同，就可以直接合并两个数据集。()
- 95 . 索引构建虽然是一个初始步骤，但索引本身需要定期更新以反映数据库中数据的变化。()
- 96 . SQL 查询引擎在处理大量数据时总是比 NoSQL 数据库快。()
- 97 . MDX 仅支持基于 OLAP 的多维数据模型，无法与关系型数据库集成。()
- 98 . DAX 是专为大数据处理设计的查询语言。()
- 99 . Spark SQL 通过其 DataFrame API 和 Catalyst 优化器，提供了对 Hive 数据仓库的广泛支持和优化查询性能的能力。()
- 100 . Presto 的查询性能总是优于其他所有类型的大数据处理框架。()
- 101 . 许多 BI 工具，包括 FineReport 和 FineVis，都支持多种查询引擎。()
- 102 . 在选择 BI 工具时，查询引擎的易用性总是比其性能更重要。()
- 103 . 查询引擎优化不仅包括 SQL 查询的优化，还涉及数据库的物理设计、配置调整等。()
- 104 . 加密查询结果可以保证查询引擎在处理过程中的所有数据都是安全的。()
- 105 . 查询性能评估时，只需关注查询的执行时间，其他系统资源消耗()的评估不重要。()
- 106 . 数据库查询优化总是能显著提升查询性能，不受数据规模、查询复杂度等因素的影响。()
- 107 . 中文分词是自然语言处理中的一个基础任务，其核心在于准确地识别出文本中的词汇边界。()
- 108 . 分词工具的准确率受多个因素影响，包括词典的完整性、算法的优化程度、对特定领域的适应性等。()
- 109 . 基于字符串匹配的分词方法完全依赖于预定义的词典，不需要任何统计信息或机器学习算法的支持。()
- 110 . 虽然统计分词方法主要依赖于统计规律，但预定义的词典可以作为辅助信息，帮助提升分词效果。()
- 111 . 基于理解的分词方法能够完全准确地理解任何复杂句子的语义并进行精确分词。()
- 112 . 基于深度学习的分词方法一定比基于统计或基于理解的分词方法效果更好。()
- 113 . 自定义词典和规则是提高分词系统性能的唯一途径。()
- 114 . 分词粒度越大，分词结果中的词汇数量就越少，但这并不意味着分词效果就一定更好。()
- 115 . 评估分词效果时，需要综合考虑多个方面的因素，包括准确率、召回率、分词速度、语义一致性以及对未登录词的处理能力等。()
- 116 . 创建一个数据库后，既可以修改数据库的相关参数，也可以修改数据库名。()
- 117 . 在更新数据的时候，只要你不写 where 语句，就可以更新全表。()
- 118 . 在 select 后面列的前使用 delete 可以消除重复的行。()
- 119 . 若在数据库中允许有重复的元组，改用 ALL 来 替换 DISTINCT，或直接去掉 DISTINCT 即可。()

- 120 . 集合函数最大的特点是它们根据一组数据求出一个值。()
- 121 . 利用复杂的条件来筛选数据库时, 必须使用“高级筛选”功能。()
- 122 . 结构化数据可以通过非关系型数据库进行存储和管理。()
- 123 . 半结构化数据是介于结构化数据和非结构化数据之间的数据。()
- 124 . 非结构化数据是指信息没有一个预先定义好的数据模型或者没有以一个预先定义的方式来组织。()
- 125 . 非结构化数据无法用固定的结构来逻辑表达, 例如用户在网络平台上发表的评价。()
- 126 . 结构化数据标记能让网站以更好的姿态展示在搜索结果当中的方式。()
- 127 . 非结构化数据被广泛应用于全文检索和各种多媒体信息处理领域。()
- 128 . 层次模型是一种用于描述事物、组织结构等具有网状结构的数据模型。()
- 129 . 层次模型能够直观地展示数据的层次关系, 便于理解和分析。()
- 130 . 层次模型查询节点的时候必须知道其双亲节点的, 因此限制了对数据库存取路径的控制。()
- 131 . 层次模型采用树形结构来表示数据和数据之间的联系, 每个节点表示一个实体, 实体之间的联系通过节点之间的连线表示。()
- 132 . 层次分析法()是一种定性与定量相结合的分析方法, 通过系统化的层次结构来处理复杂问题。()
- 133 . 层次建模的链接法按照层次树前序穿越的顺序把所有记录值依次邻接存放, 即通过物理空间的位置相邻来体现层次顺序。()
- 134 . 维度建模是数据仓库设计中的一种核心方法, 旨在以业务角度组织数据, 使其更易于理解、查询和分析。()
- 135 . 维度建模需要不断在用户需求和数据源事实之间进行平衡, 以提交可执行性好的设计。()
- 136 . 如果只是依靠单纯的维度建模, 不能保证数据来源的一致性和准确性, 而且在数据仓库的底层, 不是特别适用于维度建模的方法。()
- 137 . 事实表是树型模式中的中心表格, 用于存储业务事实或事件的数据, 通常与多个维度表通过外键关联。()
- 138 . 雪花模式是最常用的维度建模方式, 它以事实表为中心, 所有的维度表直接连接在事实表上, 形成一个雪花结构。()
- 139 . 维度表描述了事实表中度量数据的上下文或背景信息, 例如时间、地点、产品、客户等。()
- 140 . 数据存储对象包括数据流在加工过程中产生的临时文件或加工过程中需要查找的信息。()
- 141 . 数据存储可以按照不同的技术进行分类, 包括文件系统存储、关系型数据库存储、非关系型数据库存储、对象存储、块存储、归档存储等。()
- 142 . 对于需要大量数据传输的应用, 低延迟率能够确保数据传输的效率。()
- 143 . 二分查找算法仅适用于有序数组, 不适用于动态数据。()
- 144 . 数据结构中的队列结构复杂, 但实现简单。()
- 145 . 决策树模型结构清晰, 易于理解和解释, 尤其是浅层的决策树, 视觉上非常直观, 容易为非技术用户所理解。()
- 146 . 支持向量机()的目标是找到一个最大化类间间隔的超平面, 只能用于线性分类问题。()
- 147 . Python 字典()中的键可以是任何数据类型。()

- 148 . 堆排序是一种稳定的排序算法。()
- 149 . 用户数据画像是指通过收集和分析大量数据来创建个人或群体的详细描述。()
- 150 . 在数据画像构建中, 不需要进行数据清洗和预处理。()

二、单项选择题（选择一个正确的答案，将相应的字母填入题内的括号中）

- 1 . 职业道德的诚实守信是指为人做事实事求是, 讲信用、守诺言, 这是职业道德的()。
- (A) 根本准则 (B) 基本准则 (C) 根本要求 (D) 本质内容
- 2 . ()是职业道德的最基本的要求, 是职业道德的基础与核心。
- (A) 诚实守信 (B) 办事公道 (C) 服务群众 (D) 爱岗敬业
- 3 . 狭义的职业道德是指在一定职业活动中应遵循的、体现一定职业特征的、调整一定职业关系的()准则和规范。
- (A) 职业法律 (B) 职业意识 (C) 职业态度 (D) 职业行为
- 4 . 狭义的职业道德强调的是()职业的特殊要求和规范, 这些规范通常是根据该职业的特点和需求制定的。
- (A) 特定 (B) 某种 (C) 特殊 (D) 所有
- 5 . 职业道德在企业文化中的作用主要体现在()。
- (A) 塑造企业形象 (B) 提高企业利润率 (C) 决定企业战略方向 (D) 优化办公环境
- 6 . 企业的核心价值观通常与()紧密相连。
- (A) 员工个人兴趣 (B) 市场流行趋势 (C) 职业道德规范 (D) 政治经济环境
- 7 . 商务数据分析师的()技能要求其熟悉常用的数据分析语言和工具, 如Python、R等。
- (A) 编程技能 (B) 业务知识 (C) 沟通能力 (D) 数据处理能力
- 8 . 商务数据分析师的()技能要求其了解企业的业务流程和行业知识, 能够从业务角度解读数据。
- (A) 数据处理能力 (B) 业务指导能力 (C) 业务知识 (D) 沟通能力
- 9 . 商务数据分析师职业()的特点凸显其除了具备数据分析技能, 还需要具备一定的商业和市场知识, 以及良好的沟通能力,
- (A) 就业前景广泛 (B) 专业性较强 (C) 资格认证方便 (D) 具备跨领域知识
- 10 . 在()领域, 商务数据分析师主要负责风险控制、资产管理和投资决策等方面的工作。
- (A) 电商 (B) 金融 (C) 制造业 (D) 媒体行业
- 11 . 随着大数据技术的发展, 个人数据面临着被滥用、泄露和篡改的风险属于数据伦理面临的挑战的()。
- (A) 数据隐私与安全 (B) 数据偏见与歧视 (C) 数据所有权与控制 (D) 公众参与增多
- 12 . 数据伦理准则不包括()。
- (A) 尊重他人 (B) 行善原则 (C) 公平公正 (D) 可视化误导
- 13 . 企业社会责任()的核心关切点通常不包括()。
- (A) 环境保护 (B) 员工福利 (C) 股东利益最大化 (D) 社区贡献
- 14 . 企业在实施社会责任时, ()最能体现其长期承诺。

- (A) 一次性捐赠大额资金给慈善机构 (B) 定期发布社会责任报告 (C) 仅为响应公关危机而采取临时措施 (D) 通过社交媒体宣传其社会责任活动
- 15 . 商务数据分析师确保其分析结果对社会负责体现在()。
- (A) 只关注企业内部需求 (B) 忽略数据隐私保护 (C) 促进数据透明性和可解释性 (D) 避免公开任何分析结果
- 16 . 商务数据分析师在推动企业社会责任方面可以()。
- (A) 增加销售额 (B) 提高生产效率 (C) 减少员工培训成本 (D) 通过数据洞察支持可持续发展策略减少员工培训成本
- 17 . 在商务数据分析中, 职业道德的核心作用是()。
- (A) 确保数据真实性和准确性 (B) 提高数据分析效率 (C) 优化数据分析工具 (D) 增强数据可视化效果
- 18 . 商务数据分析师在处理敏感数据时, 应遵循的职业道德原则是()。
- (A) 最大化数据利用价值 (B) 保护数据安全和隐私 (C) 追求数据分析结果的精确性 (D) 快速响应数据分析需求
- 19 . 当用户要求删除其个人数据时, 商务数据分析师应()。
- (A) 删除数据, 保留合同中授权使用的部分 (B) 立即删除所有相关数据 (C) 仅删除部分数据, 保留备份 (D) 通知用户数据无法删除
- 20 . 商务数据分析师在使用用户数据进行分析时, 应遵循的原则是()。
- (A) 最大化数据利用价值 (B) 确保数据的匿名性和安全性 (C) 随意分享数据以提高透明度 (D) 忽视数据保护法规
- 21 . 商务数据分析师在处理客户数据时, ()是不恰当的。
- (A) 遵守所有相关的隐私法规 (B) 未经客户许可, 将数据用于个人项目 (C) 对敏感数据进行加密处理 (D) 保留数据的备份以防丢失
- 22 . 商务数据分析师在使用开源软件时, 首先应该()。
- (A) 随意修改代码以满足需求 (B) 忽视版权信息, 直接使用 (C) 检查软件的许可协议, 确保合规使用 (D) 仅在内部使用, 不公开发布
- 23 . 商务数据分析师在使用设备时, 应注意(), 合理安排工作时间, 避免长时间连续使用设备导致资源浪费。
- (A) 节约资源 (B) 铺张浪费 (C) 分享数据 (D) 团队合作
- 24 . ()不符合行业规范和自律要求。
- (A) 定期更新数据分析工具和技术 (B) 使用未经授权的数据集进行分析 (C) 确保分析结果的准确性和可靠性 (D) 寻求法律咨询或伦理指导
- 25 . ()是指加强数据在传输和存储过程中的安全性, 避免数据泄露和窃取。
- (A) 安全防护措施 (B) 加强员工的安全意识培养 (C) 数据加密 (D) 控制员工的上网行为
- 26 . 商务数据分析师在进行数据共享时应()。
- (A) 无需告知数据提供方 (B) 仅在内部团队中分享 (C) 遵守数据使用协议并确保数据安全 (D) 在公共平台分享
- 27 . 精益求精的方法强调在工作中()。
- (A) 不断寻求改进和优化 (B) 只关注效率提升 (C) 减少工作量以减轻压力 (D) 保持现状, 不做出改变
- 28 . 精益求精的方法中, 持续改进的关键在于()。
- (A) 快速完成任务 (B) 建立反馈机制并持续评估 (C) 减少员工培训 (D) 忽视小问题, 只关注大问题

- 29 . 在商务数据分析过程中, 忠于职守的核心要求是()。
- (A) 确保数据准确性, 维护分析公正性 (B) 追求数据分析技术的先进性 (C) 提高数据分析效率 (D) 美化数据分析结果
- 30 . 商务数据分析过程中, 忠于职守要求分析师()。
- (A) 只报告客户想听的结果 (B) 为了效率可以牺牲部分准确性 (C) 避免利益冲突, 保持分析独立性 (D) 追求分析结果的最大化
- 31 . 操作系统作为计算机系统的核心软件, 其主要功能是管理计算机的()资源。
- (A) 硬件和软件 (B) 网络和数据 (C) 用户和进程 (D) 输入和输出
- 32 . 操作系统的内核是操作系统的核心部分, 它负责管理()等底层硬件资源。
- (A) 文件和目录 (B) 用户界面和应用程序 (C) CPU、内存和 I/O 设备 (D) 网络通信和安全
- 33 . 操作系统的核心目标是提供一个()的计算机环境。
- (A) 用户友好且资源有效管理 (B) 仅供程序员使用的底层接口 (C) 单一任务执行平台 (D) 硬件与软件完全隔离的抽象层
- 34 . ()不是操作系统追求的主要目标。
- (A) 系统安全性 (B) 资源利用率 (C) 用户便利性 (D) 硬件成本最小化
- 35 . 操作系统的主要功能不包括()。
- (A) 处理器管理 (B) 存储器管理 (C) 办公软件提供 (D) 设备管理
- 36 . ()操作系统结构允许多个用户同时运行各自的应用程序。
- (A) 单用户单任务 (B) 单用户多任务 (C) 多用户多任务 (D) 实时操作系统
- 37 . 计算机网络中的“资源”主要指的是()。
- (A) 硬件设备 (B) 软件和数据 (C) 网络协议 (D) 用户账户
- 38 . 计算机网络是指将地理位置分散的()按照一定的通信协议相互连接, 以实现资源共享和信息交换的系统。
- (A) 多台计算机设备 (B) 单个计算机 (C) 打印机 (D) 路由器
- 39 . 在计算机网络分类中, 通常用于连接校园内各建筑物内设备的网络是()。
- (A) 广域网 (B) 局域网 (C) 城域网 (D) 互联网
- 40 . ()不属于计算机网络的分类依据。
- (A) 覆盖范围 (B) 传输介质 (C) 拓扑结构 (D) 使用目的
- 41 . 计算机网络最基本的功能是()。
- (A) 数据传输 (B) 资源共享 (C) 分布式处理 (D) 信息交换
- 42 . 计算机网络能够协调多个计算机系统的工作, 这体现了其()功能。
- (A) 信息传输 (B) 数据交换 (C) 分布式处理 (D) 安全保护
- 43 . 数据库系统是由数据库、数据库管理系统、应用程序、用户和()构成的。
- (A) 硬件及操作系统 (B) 网络通信设备 (C) 数据库设计工具 (D) 数据库备份设备
- 44 . 数据库是按照一定数据结构来组织、存储和管理数据的仓库, 其英文缩写通常为()。
- (A) DBMS (B) DB (C) DBA (D) SQL
- 45 . 数据库技术的基础是()。
- (A) 数据管理技术 (B) 数据安全技术 (C) 数据传输技术 (D) 计算机技术
- 46 . 数据库中的索引机制主要用于()。

- (A) 加快数据写入速度 (B) 减慢数据检索速度 (C) 提高数据检索速度 (D) 降低磁盘 I/O 操作
- 47 . 数据库系统中, 负责数据定义、数据操纵、数据控制功能的是()。
(A) 数据库管理员 (B) 数据库管理系统 (C) 数据库 (D) 操作系统
- 48 . 在数据库系统的三层模式中, 最内层是()。
(A) 外模式 (B) 概念模式 (C) 内模式 (D) 子模式
- 49 . 统计学的研究对象主要是(), 用以揭示数据的内在规律和特征。
(A) 数据 (B) 理论模型 (C) 个体现象 (D) 经济指标
- 50 . 在统计学中, 通过对()的研究来推断总体的性质。
(A) 个别案例 (B) 局部现象 (C) 特定事件 (D) 样本数据
- 51 . 统计学作为一门科学, 其核心特点是基于()来进行研究和分析。
(A) 数据 (B) 理论 (C) 假设 (D) 模型
- 52 . 统计学的基本特点之一是能够处理和分析()的数据, 揭示其内在规律。
(A) 少量 (B) 单一来源 (C) 静态 (D) 大量且多样
- 53 . 描述性统计的核心目的是(), 以揭示数据的基本特征。
(A) 总结和概括数据 (B) 推断总体参数 (C) 检验假设 (D) 预测未来趋势
- 54 . 方差分析()主要用于(), 以判断不同来源的变异是否显著。
(A) 两个独立样本的比较 (B) 多个样本均值的比较 (C) 两个相关样本的比较 (D) 时间序列数据的分析
- 55 . 数据分析的主要功能之一是帮助企业(), 从而做出更明智的决策。
(A) 洞察数据中的模式和趋势 (B) 设计数据可视化图表 (C) 收集更多数据 (D) 提升数据处理速度
- 56 . 企业利用数据分析进行风险管理时, 主要关注的是()的预测和预防。
(A) 市场波动 (B) 竞争对手动态 (C) 技术革新 (D) 潜在威胁和危机
- 57 . 数据分析的首要原则是确保数据的(), 以支持准确的结论和决策。
(A) 准确性 (B) 速度 (C) 数量 (D) 来源
- 58 . 数据分析应遵循()原则, 以确保分析结果能够应用于实际业务场景。
(A) 理论优先 (B) 技术导向 (C) 数据驱动 (D) 实用主义
- 59 . ()这一数据分析方法常用于预测未来趋势或结果。
(A) 描述性统计 (B) 推断性统计 (C) 预测性分析 (D) 探索性分析
- 60 . ()方法是用于探索数据集中各变量间关系的。
(A) 聚类分析 (B) 回归分析 (C) 关联规则挖掘 (D) 主成分分析
- 61 . ()是离散分布。
(A) 正态分布 (B) 二项分布 (C) 均匀分布 (D) 指数分布
- 62 . ()是一种连续分布, 其图形呈钟形曲线。
(A) 正态分布 (B) t 分布 (C) 均匀分布 (D) 指数分布
- 63 . 数据管理是指对数据的()进行规划、组织、控制和利用的过程。
(A) 全生命周期 (B) 单一阶段 (C) 存储环节 (D) 分析环节
- 64 . 数据管理的一个重要环节是数据的(), 它有助于企业更好地理解 and 利用数据资产。
(A) 元数据管理 (B) 数据可视化 (C) 数据分析 (D) 数据备份
- 65 . 数据治理对于提升企业的决策效率具有关键作用, 这主要体现了数据治理的()

意义。

- (A) 保障数据安全性 (B) 优化决策过程 (C) 提高数据质量 (D) 促进数据共享
- 66 . 数据治理有助于企业及时发现并解决数据问题，避免数据错误对企业运营造成负面影响，这体现了数据治理的()重要性。
- (A) 提升数据可视化能力 (B) 增强数据安全性 (C) 风险管理与预防 (D) 促进数据文化形成
- 67 . 数据治理的第三阶段侧重于()。
- (A) 数据架构的规划 (B) 数据质量的监控 (C) 数据价值的挖掘和利用 (D) 数据安全的保障
- 68 . 在数据治理的过程中，确保数据的一致性和准确性的阶段被称为()。
- (A) 数据规划阶段 (B) 数据质量管理阶段 (C) 数据应用阶段 (D) 数据监控阶段
- 69 . 数据治理框架的核心目的是()。
- (A) 提高数据查询速度 (B) 保障数据质量和一致性 (C) 减少数据存储成本 (D) 优化数据库性能
- 70 . 数据治理框架的核心组件不包括()。
- (A) 数据质量管理 (B) 数据安全治理 (C) 数据存储管理 (D) 数据生命周期管理
- 71 . 在数据流程优化中，()通过减少不必要的步骤和简化流程来提高效率。
- (A) 数据标准化 (B) 数据流程再造 (C) 数据清洗 (D) 数据备份
- 72 . 数据流程优化中，用于识别瓶颈和潜在改进点的技术是()。
- (A) 流程图分析 (B) 数据挖掘 (C) 数据仓库 (D) 数据可视化
- 73 . 数据安全主要是指保护数据的()。
- (A) 机密性、完整性和可用性 (B) 美观性、完整性和可读性 (C) 高效性、完整性和可扩展性 (D) 经济性、完整性和安全性
- 74 . 在数据安全领域，访问控制是指()。
- (A) 对数据进行加密处理 (B) 对数据进行备份和恢复 (C) 对数据进行定期清理 (D) 限制对数据的访问权限
- 75 . 数据安全中，确保授权用户能够在需要时访问和使用数据的要素是()。
- (A) 保密性 (B) 真实性 (C) 可用性 (D) 可追踪性
- 76 . 在数据安全领域，通过日志记录和监控来追踪数据的使用和修改历史的是()。
- (A) 可审计性 (B) 可恢复性 (C) 完整性 (D) 机密性
- 77 . 数据安全的核心功能是确保数据的()。
- (A) 保密性 (B) 完整性 (C) 可用性 (D) 可恢复性
- 78 . 数据安全功能中，防止数据被非法访问或泄露的措施属于()。
- (A) 访问控制 (B) 数据备份 (C) 数据加密 (D) 数据审计
- 79 . 数据服务安全架构中，()主要负责监控、检测和响应安全事件。
- (A) 数据加密模块 (B) 访问控制模块 (C) 安全事件管理模块 (D) 数据备份与恢复模块
- 80 . 数据安全主要关注的是()。
- (A) 数据传输的速度 (B) 数据存储的容量 (C) 数据的保密性、完整性和可用性 (D) 数据处理的效率
- 81 . 数据泄露防护()策略的核心目的是()。
- (A) 提高数据可用性 (B) 防止敏感数据非法泄露 (C) 优化数据存储性能 (D) 确保数据完整性

- 82 . ()不属于 DLP 策略的实施手段。
(A) 数据加密 (B) 访问控制 (C) 数据脱敏 (D) 数据备份
- 83 . 数据安全管理的首要原则是()。
(A) 保密性 (B) 完整性 (C) 可用性 (D) 可审计性
- 84 . 数据安全管理中, 确保授权用户能够根据需要访问和使用数据的是()。
(A) 最小权限原则 (B) 责任明确原则 (C) 可用性原则 (D) 数据备份原则
- 85 . 商品的基本属性中, 决定其使用价值的因素是()。
(A) 价格 (B) 质量 (C) 品牌 (D) 包装
- 86 . 从经济学的角度看, 商品的价值是由()决定的。
(A) 生产成本 (B) 市场需求 (C) 社会必要劳动时间 (D) 销售策略
- 87 . 商品的四大要素中, 体现商品基本功能和用途的是()。
(A) 核心利益 (B) 有形产品 (C) 附加产品 (D) 期望产品
- 88 . 商品四大要素的核心是(), 它决定了商品的基本价值。
(A) 核心利益 (B) 附加产品 (C) 有形产品 (D) 期望产品
- 89 . 当产品进入成长和成熟阶段时, 企业应该()应对以满足市场需求和消费者偏好的变化。
(A) 减少研发投入, 降低成本 (B) 停止产品创新, 维持现状 (C) 不断改进和升级产品 (D) 放弃该产品, 开发新产品
- 90 . 商品生命周期管理方法中, ()阶段主要关注产品的开发、测试和引入市场。
(A) 引入期 (B) 成长期 (C) 成熟期 (D) 衰退期
- 91 . 营销组合策略通常被称为()。
()4Ps()5Cs()STP()SWOT
- 92 . Promotion 在营销组合中的主要目的是()。
(A) 提高产品质量 (B) 降低生产成本 (C) 增加产品知名度 (D) 扩大生产规模
- 93 . 营销策划中, 强调以消费者需求为导向的原则, 体现了()。
(A) 成本领先策略 (B) 差异化策略 (C) 市场导向原则 (D) 集中化策略
- 94 . 营销策划的可行性原则要求()。
(A) 营销策略必须是最先进的 (B) 营销策略应基于企业资源和市场环境 (C) 营销策略必须追求最大利润 (D) 营销策略必须包括所有可能的营销手段
- 95 . 市场营销策划的首要基本功能是()。
(A) 确定市场定位 (B) 提升产品质量 (C) 增加员工福利 (D) 优化生产流程
- 96 . 营销策划的核心目的是()。
(A) 提升市场竞争力 (B) 降低生产成本 (C) 增加员工福利 (D) 扩大生产规模
- 97 . 商贸流通的广泛性体现在其涉及的()。
(A) 单一行业 (B) 多个行业与领域 (C) 少数消费者 (D) 固定地域范围
- 98 . 商贸流通的高效性要求()。
(A) 快速响应市场需求 (B) 延长商品流通时间 (C) 增加流通环节 (D) 提高生产成本
- 99 . ()不是商品流通的主要渠道。
(A) 直销渠道 (B) 零售渠道 (C) 生产渠道 (D) 分销渠道
- 100 . 商品流通的主要模式不包括()。
(A) 直销模式 (B) 零售模式 (C) 多级分销模式 (D) 电子商务模式

- 101 . 企业管理中强调的“以人为本”原则主要体现在()。
- (A) 重视员工发展与福利 (B) 追求最大利润 (C) 严格成本控制 (D) 强化技术革新
- 102 . 企业管理的权变原则指的是()。
- (A) 固定不变的管理模式 (B) 忽视市场变化 (C) 不接受任何创新 (D) 根据内外部环境灵活调整管理策略
- 103 . 企业管理的核心职能之一是()。
- (A) 计划职能 (B) 销售职能 (C) 生产职能 (D) 财务职能
- 104 . 企业内部结构的构建和优化属于()。
- (A) 领导职能 (B) 组织职能 (C) 激励职能 (D) 创新职能
- 105 . 企业组织结构的设计应遵循()这一核心原则, 以确保组织能够高效、灵活地应对市场变化。
- (A) 最大化利润 (B) 最小化成本 (C) 精简人员 (D) 战略导向与灵活性
- 106 . 企业组织结构设计中, 强调各部门和岗位之间既要分工明确, 又要相互协作的原则是()。
- (A) 目标导向原则 (B) 精干高效原则 (C) 分工协作原则 (D) 集权分权结合原则
- 107 . 成本控制的原则中, ()强调在不影响产品质量的前提下降低成本。
- (A) 经济性原则 (B) 全面性原则 (C) 责任性原则 (D) 效益性原则
- 108 . 成本控制的首要原则是()。
- (A) 最大化利润 (B) 最小化成本 (C) 提高生产效率 (D) 保持产品质量
- 109 . 《电子商务法》把调整范围严格限定在()这一区域内。
- (A) 中华人民共和国境内 (B) 中华人民共和国境外 (C) 亚洲地区 (D) 欧美地区
- 110 . 《中华人民共和国电子商务法》主要适用于()。
- (A) 实体零售 (B) 互联网信息服务 (C) 通过互联网等信息网络销售商品或提供服务 (D) 制造业
- 111 . 《中华人民共和国保守国家秘密法》规定, 国家秘密的保密期限已满的, 应该()。
- (A) 进行调整 (B) 上报上级单位 (C) 正式公开 (D) 自行解密
- 112 . 根据《中华人民共和国保守国家秘密法》的规定, 涉密人员离岗离职时, 应实行()。
- (A) 脱密期管理 (B) 保密培训管理 (C) 公开透明管理 (D) 无需特别管理
- 113 . 《中华人民共和国数据安全法》在制度设计上强调了()。
- (A) 只需关注数据安全的具体制度 (B) 只需构建整体治理框架 (C) 加强具体制度与整体治理框架的衔接 (D) 无需具体制度与整体治理框架的协调
- 114 . 《中华人民共和国数据安全法》的核心思想是()。
- (A) 安全第一, 发展次之 (B) 发展优先, 安全为辅 (C) 安全与发展并重 (D) 只关注数据安全, 不考虑发展
- 115 . 根据《中华人民共和国个人信息保护法》的规定, ()不属于个人信息处理活动。
- (A) 收集 (B) 存储 (C) 公开 (D) 售卖
- 116 . 《中华人民共和国个人信息保护法》法在空间适用上主要遵循()。
- (A) 属地管辖 (B) 属人管辖 (C) 保护性管辖 (D) 普遍管辖
- 117 . 根据《中华人民共和国网络安全法》, 网络运营者在为用户提供服务时, 应要求用户提供()。
- (A) 虚拟身份信息 (B) 无需提供任何信息 (C) 真实身份信息 (D) 网络昵称

- 118 . 《中华人民共和国网络安全法》明确的核心原则是()。
- (A) 网络安全优先于网络发展 (B) 网络发展优先于网络安全 (C) 网络空间主权原则 (D) 网络自由原则
- 119 . 《全国人民代表大会常务委员会关于加强网络信息保护的決定》主要聚焦于()的问题。
- (A) 网络安全基础设施建设 (B) 网络技术创新与发展 (C) 公民个人电子信息和隐私保护 (D) 网络犯罪打击与预防
- 120 . ()不是《全国人民代表大会常务委员会关于加强网络信息保护的決定》所强调的内容。
- (A) 禁止非法获取、出售或提供公民个人电子信息 (B) 网络信息服务提供者应对用户信息保密, 不得泄露 (C) 鼓励网络信息服务提供者进行大规模数据收集与分析 (D) 采取技术措施和其他必要措施确保信息安全
- 121 . 信息系统主要依赖()来收集、存储、处理、传输和展示信息。
- (A) 手工 (B) 物理 (C) 电子 (D) 机械
- 122 . 信息系统的核心组成部分不包括()。
- (A) 数据库 (B) 用户界面 (C) 原材料 (D) 网络通信
- 123 . 信息系统通过()将原始数据转化为有用的信息和知识。
- (A) 数据采集 (B) 数据分析 (C) 数据存储 (D) 数据展示
- 124 . ()不是信息系统的基本功能。
- (A) 自动化业务流程 (B) 社交娱乐 (C) 决策支持 (D) 信息集成
- 125 . 数据采集的首要步骤是()。
- (A) 数据清洗 (B) 确定数据源 (C) 数据分析 (D) 数据存储
- 126 . ()适合捕捉社交媒体上的实时数据
- (A) 批量导入 (B) API 接口 (C) 手工录入 (D) 日志文件分析
- 127 . 数据清洗的主要目的是()。
- (A) 增加数据量 (B) 纠正数据错误 (C) 数据加密 (D) 数据压缩
- 128 . 数据清洗的常见步骤不包括()。
- (A) 缺失值填充 (B) 噪声数据过滤 (C) 数据编码转换 (D) 数据建模
- 129 . 在数据清洗流程中, 首先应当进行()。
- (A) 数据分析 (B) 识别数据质量问题 (C) 数据转换 (D) 数据备份
- 130 . 数据清洗中的“去重”步骤主要解决()。
- (A) 数据类型不一致 (B) 数据量过大 (C) 数据中存在重复记录 (D) 数据缺失
- 131 . 数据质量控制的核心目的是()。
- (A) 增加数据量 (B) 加快数据处理速度 (C) 确保数据的准确性和完整性 (D) 降低存储成本
- 132 . ()不是数据质量监控的常用方法。
- (A) 设定数据质量标准 (B) 定期执行数据审计 (C) 实时数据流监控 (D) 数据备份与恢复
- 133 . 数据监测是对()进行持续的观察和记录。
- (A) 个人隐私数据 (B) 系统运行状态及业务数据 (C) 社交媒体趋势 (D) 自然环境变化
- 134 . ()不属于数据监测的范畴。
- (A) 实时数据收集 (B) 数据分析与预测 (C) 数据可视化展示 (D) 数据销毁处理
- 135 . 数据监测的主要目的是()。

- (A) 娱乐休闲 (B) 为企业决策提供数据支持 (C) 监测员工工作效率 (D) 展示公司技术实力
- 136 . 数据监测不能帮助()。
- (A) 优化业务流程 (B) 预测市场趋势 (C) 提高员工士气 (D) 评估系统性能
- 137 . 在设计数据监测指标时, 首要需要考虑()。
- (A) 多样性 (B) 可度量性 (C) 创新性 (D) 美观性
- 138 . ()不是设计数据监测指标时应遵循的原则。
- (A) 相关性原则 (B) 唯一性原则 (C) 实时性原则 (D) 主观性原则
- 139 . 数据异常监测旨在()。
- (A) 仅仅为了收集更多数据 (B) 及时发现并处理数据中的异常情况 (C) 增加数据存储的复杂性 (D) 改善用户体验
- 140 . 在构建数据异常预警机制时, 最核心的要素是()。
- (A) 数据存储介质 (B) 高效的异常识别算法 (C) 绚丽的可视化界面 (D) 定期的数据备份
- 141 . 实时数据监测的核心价值体现在()。
- (A) 提升数据收集速度 (B) 保证数据的时效性, 以支持即时决策 (C) 节省数据存储成本 (D) 美化数据呈现
- 142 . ()不属于实时数据处理的范畴。
- (A) 数据流的实时接收 (B) 数据的实时清洗与预处理 (C) 数据的永久归档存储 (D) 数据的即时分析
- 143 . 数据迁移的定义是()。
- (A) 数据在不同存储设备间的简单复制 (B) 数据环境或系统间的迁移 (C) 数据的销毁与重建 (D) 数据安全加密技术的应用
- 144 . 数据迁移过程中, 首要关注的是()。
- (A) 迁移工具的成本 (B) 迁移操作的简便性 (C) 数据的完整性与一致性 (D) 网络延迟对迁移的影响
- 145 . ()不是数据迁移的常见类型。
- (A) 结构化数据迁移 (B) 非结构化数据迁移 (C) 半结构化数据迁移 (D) 无效数据迁移
- 146 . 当需要将数据库中的表及其关系从一个数据库系统迁移到另一个时, 这属于()。
- (A) 跨平台迁移 (B) 异构迁移 (C) 同构迁移 (D) 云迁移
- 147 . 数据迁移前, 首要需要()。
- (A) 选择迁移工具 (B) 制定详细的迁移计划 (C) 执行数据备份 (D) 直接开始迁移
- 148 . 在数据迁移策略中, ()强调在最小化对生产系统影响的同时完成迁移。
- (A) 大规模停机迁移 (B) 滚动迁移 (C) 镜像迁移 (D) 实时同步迁移
- 149 . 使用数据迁移工具时, 不需要考虑()。
- (A) 工具的兼容性 (B) 工具的易用性 (C) 工具的免费性 (D) 工具的性能与稳定性
- 150 . 在选择数据迁移工具时, 首先需要评估的是()。
- (A) 工具的知名度 (B) 工具的价格 (C) 工具对目标系统的支持程度 (D) 工具的界面美观度
- 151 . 在数据迁移过程中, 为了保障数据一致性, 最常用的方法是()。
- (A) 暂停所有业务操作 (B) 使用事务管理确保操作的原子性 (C) 忽略数据不一致, 迁移后再处理 (D) 仅迁移最新数据

- 152 . ()常用于在数据库迁移中保证数据的一致性,特别是在分布式系统中。
(A)两阶段提交(B)消息队列(C)数据分片(D)数据冗余存储
- 153 . 为了提高数据迁移的性能,不推荐()。
(A)并行处理迁移任务(B)在业务高峰期进行迁移(C)选择高效的数据传输方式
(D)对数据进行合理分块处理
- 154 . 数据迁移过程中,()可以优化数据传输效率。
(A)增加网络带宽(B)压缩传输数据(C)使用低速存储设备(D)频繁进行数据校验
- 155 . 在进行数据迁移时,确保数据安全的首要步骤是()。
(A)选择高性能的迁移工具(B)评估数据敏感性并制定加密策略(C)忽略数据备份
(D)仅迁移部分非敏感数据
- 156 . 数据迁移过程中,符合 GDPR()要求的关键措施是()。
(A)不收集用户数据(B)匿名化处理所有个人数据(C)加密传输和存储敏感数据
(D)不跨境传输数据
- 157 . ()主要用于随机访问大量数据,如数据库系统。
(A)固态硬盘(B)机械硬盘(C)磁带存储(D)闪存卡
- 158 . 在云环境中,()提供了最大的灵活性和可扩展性,但可能伴随着更高的访问延迟。
(A)对象存储(B)块存储(C)文件存储(D)网络附加存储
- 159 . 在进行数据备份时,()通过记录自上次备份以来数据的所有更改来减少备份时间和存储需求。
(A)完全备份(B)增量备份(C)差异备份(D)镜像备份
- 160 . ()用于灾难恢复,允许在发生严重故障时快速恢复业务运。
(A)数据压缩(B)数据加密(C)热备份(D)远程镜像
- 161 . 数据冗余的主要目的是()。
(A)减少数据存储成本(B)提高数据访问速度(C)增加数据的可靠性和容错性(D)
便于数据迁移
- 162 . ()可以在存储数据之前检测并消除重复数据块,从而减少存储空间需求。
(A)数据压缩(B)数据去重(C)数据加密(D)数据索引
- 163 . 数据生命周期管理的核心目标是()。
(A)最小化数据存储空间(B)确保数据永远可用(C)优化数据在整个生命周期中的
成本、风险和价值(D)加速数据处理速度
- 164 . 在数据生命周期管理中,()涉及数据的归档和长期存储。
(A)数据创建(B)数据使用(C)数据保留(D)数据销毁
- 165 . 在数据安全中,加密的主要目的是()。
(A)提高数据传输速度(B)防止数据在未经授权的情况下被访问或泄露(C)减少数
据存储成本(D)改进数据压缩效率
- 166 . ()使用单个密钥进行加密和解密。
(A)对称加密(B)非对称加密(C)哈希函数(D)散列函数
- 167 . 数据压缩的主要目的是()。
(A)提高数据传输速度(B)降低数据存储空间需求(C)加密数据以保护隐私(D)提
升数据处理的准确性
- 168 . ()通常用于无损压缩,即在压缩和解压缩过程中不会丢失原始数据。
()JPEG ()ZIP ()MP3 (()PNG

- 169 . ()是分布式存储系统的核心优势。
(A)数据易于丢失 (B)高可扩展性 (C)单点故障导致整体服务中断 (D)数据访问速度受限
- 170 . 在分布式存储中,数据通常被分割成()以便跨多个节点存储。
(A)完整文件 (B)数据块 (C)数据流 (D)数据库记录
- 171 . 数据访问控制的主要目的是()。
(A)确保数据的完整性 (B)提高数据传输效率 (C)防止未经授权的数据访问 (D)减少数据存储成本
- 172 . ()是基于角色的权限分配。
(A)自主访问控制(B)强制访问控制(C)基于角色的访问控制(D)访问控制列表
- 173 . 数据存储的主要目标不包括()。
(A)确保数据的安全性 (B)提供高效的数据访问 (C)最小化存储成本 (D)阻止所有用户访问数据
- 174 . ()强调了在需要时能够快速检索和恢复数据的重要性。
(A)数据完整性 (B)数据备份 (C)数据可用性 (D)数据冗余
- 175 . ()不是数据存储应遵循的基本原则。
(A)数据应备份以防止丢失 (B)存储系统应支持高并发访问 (C)所有数据都应无条件地公开访问 (D)数据的访问应受到适当的安全控制
- 176 . 数据存储设计时应考虑(),以确保未来数据量的增长不会导致系统瓶颈。
(A)数据保密性 (B)数据完整性 (C)可扩展性 (D)实时性
- 177 . ()涉及将数据从临时存储位置转移到长期存储介质。
(A)数据采集(B)数据转换(C)数据迁移(D)数据备份
- 178 . 在数据存储流程中,()是确保数据在存储前符合特定格式和标准。
(A)数据清洗(B)数据索引(C)数据压缩(D)数据加密
- 179 . 评估数据存储系统性能时,()用于衡量数据访问速度。
(A)吞吐量(B)延迟(C)可靠性(D)冗余度
- 180 . 在评估数据存储系统的可扩展性时,主要关注的是()。
(A)初始投资成本(B)存储空间利用率(C)系统能否轻松增加存储容量或处理能力(D)数据备份策略
- 181 . 数据清洗需求分析的主要目的是()。
(A)选择合适的数据存储技术 (B)确定数据清洗的具体步骤 (C)识别并解决当前数据中的质量问题 (D)评估数据分析结果的准确性
- 182 . ()不属于数据清洗需求分析的内容。
(A)确定数据清洗的必要性和紧迫性 (B)设计数据清洗的算法细节 (C)识别数据中存在的缺失、错误和异常 (D)设定清洗后数据应达到的质量标准
- 183 . 在数据质量问题识别中,()常用于发现数据中的异常值。
(A)数据可视化 (B)数据排序 (C)相关性分析 (D)完整性检查
- 184 . ()不是直接用于识别数据缺失。
(A)非空值验证 (B)逻辑规则检查 (C)聚类分析 (D)数据比对
- 185 . 评估清洗数据价值时,最重要的考虑因素是()。
(A)清洗数据的成本 (B)清洗后数据对业务决策的改进 (C)清洗工具的先进性 (D)清洗数据的数量
- 186 . ()最有效地量化清洗数据的价值。

- (A) 对比清洗前后业务指标的变化 (B) 单纯计算清洗数据所占的比例 (C) 评估数据清洗的复杂度 (D) 检查清洗后数据的完整性
- 187 . 设定数据清洗预期目标时, () 最关键。
(A) 选择合适的清洗工具 (B) 确定具体的数据质量标准 (C) 分析清洗后数据的潜在应用 (D) 立即启动清洗操作
- 188 . () 不是设定数据清洗预期目标时应考虑的因素。
(A) 业务对数据质量的实际需求 (B) 清洗工具的最新功能 (C) 现有数据质量的评估结果 (D) 清洗工作的时间限制
- 189 . 在评估数据清洗对业务的影响时, 应首先关注()。
(A) 数据清洗的成本 (B) 清洗后数据的存储效率 (C) 直接支持业务决策的关键指标 (D) 清洗工具的市场口碑
- 190 . () 不是用于评估数据清洗对业务指标影响的常用手段。
(A) 对比清洗前后业务指标的变化 (B) 进行 A/B 测试, 比较清洗前后数据对业务的影响 (C) 单纯查看清洗数据的数量 (D) 使用统计方法评估数据清洗对业务指标的贡献度
- 191 . 在分析数据源时, 首先需要考虑()。
(A) 数据清洗的难易程度 (B) 数据源的可信度与权威性 (C) 数据存储的硬件要求 (D) 数据可视化工具的兼容性
- 192 . 数据结构分析的主要目的是()。
(A) 确定数据是否需要清洗 (B) 优化数据存储和检索效率 (C) 验证数据的准确性 (D) 转换数据格式以兼容不同系统
- 193 . 使用数据清洗工具时, 最重要的是()。
(A) 追求最新最先进的技术 (B) 深入了解工具的功能和限制 (C) 立即开始清洗而不做计划 (D) 依赖工具自动完成所有清洗工作
- 194 . () 不是选择数据清洗工具时应考虑的因素。
(A) 工具的易用性和学习曲线 (B) 工具的扩展性和与其他系统的集成能力 (C) 工具的价格是否最低 (D) 工具的维护和支持服务
- 195 . 数据完整性分析主要关注的是()。
(A) 数据中是否存在缺失值 (B) 数据是否准确无误 (C) 数据在不同系统间是否一致 (D) 数据是否满足特定业务规则
- 196 . () 不是用于分析数据完整性。
(A) 检查数据字段是否为空 (B) 验证数据类型的正确性 (C) 比较数据记录之间的逻辑关系 (D) 使用哈希函数检查数据一致性
- 197 . 数据准确性分析的核心是()。
(A) 检查数据是否完整 (B) 验证数据是否符合业务规则 (C) 比较数据与其他可靠来源的一致性 (D) 评估数据清洗工具的性能
- 198 . () 常用于数据准确性验证。
(A) 数据压缩 (B) 聚类分析 (C) 交叉验证 (D) 数据可视化
- 199 . 数据一致性分析主要关注的是()。
(A) 数据在不同系统间的同步状态 (B) 数据字段是否完整 (C) 数据是否准确无误 (D) 数据是否满足特定格式要求
- 200 . () 常用于检测数据不一致性。
(A) 数据排序 (B) 哈希校验 (C) 数据比对 (D) 数据分析可视化
- 201 . () 不是用于识别数据缺失值。

- (A)非空值检查 (B)逻辑规则判断 (C)数据可视化分析 (D)数据排序
- 202 . 在 Excel 中, () 函数常用于统计缺失值的数量。
(A)COUNT (B)COUNTIF (C)COUNTA (D)COUNTBLANK
- 203 . () 常用于识别数据中的异常值。
(A)数据排序 (B)箱线图分析 (C)数据类型转换 (D)数据规范化
- 204 . 在统计学中, () 用于检测单变量数据中的异常值。
(A)T 检验 (B)方差分析 (C)Z 分数 (D)卡方检验
- 205 . 数据清洗方案的主要目的是 ()。
(A)增加数据量 (B)提高数据质量 (C)简化数据分析过程 (D)节省存储空间
- 206 . () 不是数据清洗方案通常包含的内容。
(A)识别数据中的错误和异常 (B)填充缺失值 (C)数据备份策略 (D)转换数据格式
- 207 . 在进行数据清洗需求分析时, 首先应考虑的是 ()。
(A)数据的来源和类型 (B)清洗后数据的用途 (C)清洗工具的可用性 (D)清洗过程的时间限制
- 208 . () 不是数据清洗需求分析的一部分。
(A)确定数据中存在的具体问题 (B)评估数据清洗的优先级 (C)设定数据清洗的预算 (D)选择具体的清洗技术
- 209 . 选择数据清洗工具时, () 不是主要考虑点。
(A)工具的易用性和用户友好性 (B)工具支持的数据类型和格式 (C)工具的价格 (D)工具的自动化程度
- 210 . 在处理大规模数据集时, 选择数据清洗工具时应优先考虑 ()。
(A)工具的图形界面 (B)工具的批处理能力 (C)工具的自定义脚本支持 (D)工具的实时数据处理能力
- 211 . 设定数据清洗规则时, 最重要的是 ()。
(A)规则的复杂性 (B)规则的一致性 (C)规则的数量 (D)规则的执行速度
- 212 . 在处理缺失值时, 设定清洗规则时通常不包括 ()。
(A)识别缺失值的模式 (B)选择填充缺失值的方法 (C)设定缺失值填充的阈值 (D)评估填充后数据的质量
- 213 . 设计数据清洗流程时, () 不是必需的。
(A)数据收集与导入 (B)清洗规则的定义与实现 (C)清洗过程的自动化 (D)数据分析与解释
- 214 . 在设计数据清洗流程时, 应优先考虑 ()。
(A)流程的灵活性 (B)流程的复杂性 (C)流程的效率 (D)流程的可追溯性
- 215 . 数据清洗过程中, 实施质量控制的关键一步是 ()。
(A)选择适合的清洗工具 (B)设定明确的清洗规则 (C)定期进行清洗结果的验证 (D)尽快完成清洗任务以节省时间
- 216 . 在数据清洗质量控制中, () 有助于发现潜在的数据错误和异常。
(A)简单的数据排序 (B)依赖数据可视化进行主观判断 (C)统计分析和假设检验 (D)随意选择部分数据进行检查
- 217 . 评估数据清洗方案效果时, 首要关注的是 ()。
(A)清洗工具的性能 (B)清洗后数据的质量提升 (C)清洗过程的时间消耗 (D)清洗成本的节约
- 218 . () 适用于评估数据清洗方案对业务决策的影响。

- (A) 对比清洗前后数据的统计特征 (B) 询问数据使用者的反馈 (C) 单纯分析数据清洗的技术细节 (D) 检查数据清洗的代码实现
- 219 . 在处理数据中的异常值时, 通常不会采取()。
- (A) 删除异常值 (B) 填充异常值为某个固定值 (C) 使用统计方法 (替换异常值 (D) 忽略异常值, 不进行任何处理
- 220 . () 有助于识别数据中的异常值。
- (A) 数据排序 (B) 箱型图分析 (C) 数据压缩 (D) 数据加密
- 221 . 数据清洗与数据治理的主要区别在于()。
- (A) 数据清洗是数据治理的一个子集 (B) 数据治理关注数据的整个生命周期管理, 而数据清洗只是其中一个环节 (C) 数据治理侧重于技术实现, 而数据清洗侧重于业务逻辑 (D) 数据治理是可选的, 而数据清洗是必须的
- 222 . 数据治理中, () 与数据清洗直接相关。
- (A) 数据架构设计 (B) 数据质量管理 (C) 数据安全治理 (D) 数据使用政策制定
- 223 . 在数据清洗过程中实施版本控制的主要目的是()。
- (A) 确保数据清洗工具始终保持最新版本 (B) 跟踪和记录数据清洗的各个阶段和变化 (C) 加速数据清洗过程 (D) 减少数据清洗过程中的人工干预
- 224 . () 用于实现数据清洗的版本控制。
- (A) 数据备份 (B) 版本控制系统 (C) 数据加密 (D) 数据压缩
- 225 . 在数据预处理中, 将特征缩放到[0, 1]区间的方法称为()。
- (A) 标准化 (B) 归一化 (C) 离散化 (D) 编码
- 226 . 有一组数据, 其最大值为 100, 最小值为 10, 要将其归一化到[0, 1]区间, 应使用的公式是()。
- () () () () () () () ()
- 227 . 将两个或多个数据集中的数据合并到一个数据集中的过程称为()。
- (A) 数据聚合 (B) 数据整合 (C) 数据合并 (D) 数据迁移
- 228 . 在合并两个数据集时, 需要特别注意的是()。
- (A) 数据集的大小 (B) 数据集的缺失值 (C) 数据集的来源 (D) 数据集之间的兼容性
- 229 . 负责理解用户输入的查询语句并将其转换为计算机可理解形式的组件是()。
- (A) 索引构建器 (B) 查询解析器 (C) 排序器 (D) 结果合并器
- 230 . () 通常用于提高数据库查询的速度, 通过预先构建的数据结构来加速数据检索过程。
- (A) 查询优化器 (B) 索引 (C) 分区 (D) 缓存
- 231 . SQL 作为 BI 查询引擎, 其最显著的特点是()。
- (A) 适用于非结构化数据 (B) 提供丰富的数据操作能力 (C) 仅支持内存计算 (D) 专为图形界面设计
- 232 . () 不是 SQL 查询引擎常用于 BI 分析的优点。
- (A) 成熟且广泛支持 (B) 灵活的数据查询和转换能力 (C) 实时分析能力强 (D) 强大的数据聚合与统计功能
- 233 . MDX 主要用于()。
- (A) 文本挖掘 (B) 实时数据流处理 (C) 多维数据分析 (D) 图形数据分析
- 234 . 在使用 MDX 时, () 是理解和构建多维数据集的关键。
- (A) 数据行 (B) 数据列 (C) 维度 (D) 数据表
- 235 . DAX 主要用在()平台中。

- (A)SQL Server (B)Power BI (C)Hadoop (D)MongoDB
- 236 . DAX 提供了()的数据操作能力。
(A)低级数据位操作 (B)内存数据库管理 (C)数据模型的查询、计算和定义 (D)实时数据流处理
- 237 . Spark SQL 提供的主要优势是()。
(A)实时数据流处理 (B)基于内存的分布式数据处理 (C)专为结构化数据设计 (D)低延迟的图形数据处理
- 238 . 在 Spark SQL 中, ()用于将 SQL 查询转换为 Spark 作业。
(A)DataFrame API (B)SQLContext/SparkSession (C)RDD API (D)Catalyst Optimizer
- 239 . Presto 作为交互式查询引擎, 其主要优势是()。
(A)低延迟查询 (B)高效的数据转换 (C)强大的实时数据处理 (D)复杂的数据挖掘
- 240 . Presto 特别适用于()。
(A)批处理大数据作业 (B)流式数据处理 (C)低延迟的交互式分析查询 (D)图形和图像数据处理
- 241 . FineReport 和 FineVis 相比, ()更侧重于数据可视化。
(A)FineReport (B)FineVis (C)两者相同 (D)两者都不侧重
- 242 . FineReport 支持(), 使其成为一个全面的 BI 工具。
(A)实时数据流处理 (B)复杂的数据挖掘算法 (C)多种数据源的集成与查询 (D)高级的图形渲染技术
- 243 . 在选择 BI 工具的查询引擎时, ()通常不需要首要考虑。
(A)查询性能 (B)支持的数据源类型 (C)工具的易用性 (D)工具的价格是否足够低
- 244 . ()最能体现查询引擎的扩展性。
(A)并发用户数 (B)支持的查询类型 (C)分布式处理能力 (D)图形界面的美观程度
- 245 . ()常用于查询引擎的性能优化。
(A)加密技术 (B)索引优化 (C)实时数据压缩 (D)数据清洗
- 246 . 在进行查询引擎优化时, ()不是必须的。
(A)分析查询计划 (B)增加更多的硬件资源 (C)调整查询逻辑以减少数据扫描量 (D)监控并优化数据库连接
- 247 . ()主要用于保护查询引擎免受 SQL 注入攻击。
(A)数据加密 (B)访问控制列表 (C)使用参数化查询 (D)定期备份数据
- 248 . 在保证查询引擎安全性时, ()不是直接相关。
(A)防火墙和入侵检测系统 (B)强制使用 HTTPS (C)定期对数据库进行碎片整理 (D)加密敏感数据字段
- 249 . ()直接反映了查询操作执行时间的长短。
(A)CPU 使用率 (B)响应时间 (C)磁盘 I/O 次数 (D)网络带宽
- 250 . 在数据库查询性能评估中, QPS()指的是()。
(A)每秒完成的查询数量 (B)每秒插入的记录数 (C)每秒更新的行数 (D)每秒删除的记录数
- 251 . 查询优化器的主要功能是()。
(A)自动修正 SQL 语句的语法错误 (B)分析并优化 SQL 查询计划, 以提高执行效率 (C)预测查询结果 (D)管理数据库索引
- 252 . ()不属于数据库查询优化的基本策略。

- (A) 使用索引减少数据扫描范围 (B) 避免使用 SELECT *, 指定具体列 (C) 频繁对数据库进行全表扫描 (D) 利用查询缓存减少重复计算
- 253 . 数据分词主要用于处理()。
- (A) 数值型数据 (B) 图像数据 (C) 文本数据 (D) 音频数据
- 254 . 中文分词的主要目的是()。
- (A) 将句子拆分为单词, 以便进行语言翻译 (B) 识别文本中的关键词 (C) 将连续的中文文本切分成具有完整意义的词汇 (D) 压缩文本数据
- 255 . 在选择分词工具或库时, 首要考虑的因素是()。
- (A) 工具的易用性 (B) 工具的准确率 (C) 工具的更新频率 (D) 工具的开源程度
- 256 . 使用分词库时, 常见的性能瓶颈是()。
- (A) 词汇量不足 (B) 单词识别错误率高 (C) 处理大数据量时的内存消耗 (D) 不支持多语言分词
- 257 . 基于字符串匹配的分词方法主要依赖于()。
- (A) 语义分析 (B) 词典匹配 (C) 机器学习算法 (D) 文本统计特征
- 258 . ()属于基于字符串匹配的分词方法。
- (A) 最大正向匹配法 (B) 基于隐马尔可夫模型的分词 (C) 基于条件随机场的分词 (D) 基于深度学习的分词
- 259 . 基于统计的分词方法主要依据()来判断词边界。
- (A) 预定义词典 (B) 词汇出现的统计规律 (C) 语义信息 (D) 文本格式
- 260 . ()常用于基于统计的分词方法中。
- (A) 最大正向匹配 (B) 隐马尔可夫模型 (C) 条件随机场 (D) 神经网络
- 261 . 基于理解的分词方法的核心思想是()。
- (A) 依赖预定义词典 (B) 利用统计规律 (C) 深入理解句子的语义和结构 (D) 依赖文本格式
- 262 . ()常用于基于理解的分词方法中, 以分析句子的语义和句法结构。
- (A) 隐马尔可夫模型 (B) 自然语言处理技术 (C) 最大熵模型 (D) 最大正向匹配
- 263 . 基于深度学习的分词方法主要依赖于()。
- (A) 手工制定的规则 (B) 预定义的词典 (C) 神经网络模型 (D) 统计特征
- 264 . ()深度学习模型常用于分词任务。
- (A) 卷积神经网络 (B) 递归神经网络 (C) 深度置信网络 (D) 生成对抗网络
- 265 . 自定义词典在分词中的作用是()。
- (A) 替代预定义词典 (B) 作为分词结果的最终判断依据 (C) 辅助分词系统处理特定领域或特殊词汇 (D) 决定分词方法的类型
- 266 . 在分词系统中添加自定义规则的主要目的是()。
- (A) 提高分词速度 (B) 简化分词算法 (C) 处理词典中未包含的特定词汇或结构 (D) 减少内存消耗
- 267 . 分词粒度指的是()。
- (A) 分词结果中词汇的数量 (B) 分词结果中词汇的平均长度 (C) 分词系统处理文本的速度 (D) 分词结果中词汇的详细程度或精度
- 268 . 在分词过程中, ()量化分词粒度的大小。
- (A) 通过比较分词结果中词汇的总数 (B) 通过评估分词结果对原文语义的保留程度 (C) 通过计算分词结果中词汇的平均字符数 (D) 上述都是可能的
- 269 . 评估分词效果时, 最常用的指标之一是()。

- ()分词速度 ()准确率()召回率 ()F1 值
- 270 . ()不是直接用于评估分词效果的技术。
(A)人工评估 (B)交叉验证 (C)使用外部词典进行比对 (D)文本相似度计算
- 271 . 在()数据库时, 需要选择合适的字符集和排序规则, 以便支持不同语言和文本的存储和检索。
(A) 创建 (B) 检索 (C) 删除 (D) 查询
- 272 . 在查看数据库时, 可以通过()或客户端工具来查看数据库的结构和内容。
(A) 输入框 (B) 视图窗 (C) 命令行 (D) 数据列
- 273 . insert 和 select 搭配使用的要求是在删除表中的重复记录时, 重复的数据只能有()。
(A) 一份 (B) 两份 (C) 三份 (D) 四份
- 274 . 使用 select into 语句将现有表中的数据添加到新建表中 , 要注意新表是在执行查询语句的时候创建的, 不能够()。
(A) 永久删除 (B) 持续添加 (C) 手动创建 (D) 预先存在
- 275 . 使用 SELECT 语句不可以查询()。
(A) 一个数据表 (B) 表中的数据 (C) 表中的一列或多列 (D) 表中字的总和
- 276 . 使用查询语句“select * from 表名”时, from 关键字后面写(), 表示数据来源于是这张表。
(A) 数据列名 (B) 数据表名 (C) 数据行名 (D) 数据集名
- 277 . 如果要从很多记录中查询出指定的记录, 那么就需要一个查询的()。
(A) 函数 (B) 数据 (C) 条件 (D) 记录
- 278 . 为了方便用户提高工作效率, SQL 语言允许用户将 SELECT 后的<值表达式>省略为()。
(A) * (B) % (C) # (D) @
- 279 . 聚合函数要搭配()一起使用, 它才能显示出结果。
(A) count (B) having (C) insert (D) select
- 280 . group by 子句进行分组以后, 需要对分组结果再进行条件过滤时, 不能使用 where 语句, 而需要用()。
(A) count (B) having (C) insert (D) select
- 281 . 交叉连接是不带()的连接, 即两个表的笛卡尔积是量表中元组的交叉乘积。
(A) 逻辑运算符 (B) 谓词 (C) 连接词 (D) 名词
- 282 . 自连接就是一个(), 同本身进行连接。
(A) 表 (B) 数据 (C) 数据集 (D) 数据列
- 283 . 结构化数据是指表现为()的数据。
(A) 一维形式 (B) 二维形式 (C) 三维形式 (D) 平面形式
- 284 . 结构化数据是指具有()格式和规则的数据, 通常以表格、树状结构或关系模型的形式呈现。
(A) 固定 (B) 非固定 (C) 多变 (D) 相同
- 285 . 结构化数据特点是数据以行为单位, 一行数据表示一个实体的信息, 每一行数据的()是相同的。
(A) 数量 (B) 内容 (C) 属性 (D) 间距
- 286 . 结构化数据的特点是()。
(A) 没有预定义的模式 (B) 难以通过数据库进行管理 (C) 可以用关系型数据库表

- 来表示 (D) 只能用于文本分析
- 287 . 非结构化数据是指数据结构不规则或不完整, 没有预定义的数据模型, 不方便用数据库()来表现的数据。
(A) 一维文件夹 (B) 二维文件夹 (C) 一维逻辑表 (D) 二维逻辑表
- 288 . 非结构化数据的格式和标准具有(), 因此, 在技术上比结构化信息更难标准化和理解。
(A) 单一性 (B) 多样性 (C) 复杂性 (D) 规则性
- 289 . 非结构化数据一般指()数据, 但是数据中有很多诸如时间, 数字等的信息。
(A) 文字型 (B) 函数型 (C) 数值型 (D) 货币型
- 290 . 非结构化数据可以从论坛、微博、微信等渠道获得属于其()的特征。
(A) 格式多样 (B) 容量大 (C) 来源多样 (D) 数据类型多样
- 291 . 结构化数据是指那些以固定格式或模式存储的信息, 如()。
(A) 文本文件 (B) 客户信息 (C) 图片 (D) 社交媒体帖子
- 292 . 不属于结构化数据的应用场景的是()。
(A) 微博评论 (B) 企业 ERP (C) 财务系统 (D) 教育一卡通
- 293 . 非结构化数据()通过集成 OCR 等第三方生态应用, 实现内容赋能业务运营, 包括智能识别提取业务系统内容, 实现业务流程的自动化, 提升数字化服务能力和用户体验等。
(A) 前台 (B) 中台 (C) 后台 (D) 终端
- 294 . 通过非结构化数据中台, 可以实现数据(), 推动学习型组织建设, 赋能业务运营。
(A) 清晰化 (B) 结构化 (C) 知识化 (D) 格式化
- 295 . 层次数据模型就是的图形表示就是一个()的树, 其基本数据结构是树或二叉树。
(A) 倒立生长 (B) 直立生长 (C) 交叉生长 (D) 网状生长
- 296 . 层次模型的基本思想是将一个大的系统或结构分解为若干个相对()的子系统或部分, 这些子系统或部分之间以特定的层次关系相互联系。
(A) 无关 (B) 联结 (C) 交叉 (D) 独立
- 297 . 层次数据模型中()节点没有父节点, 其余的节点必须有且仅有一个父节点, 但是所有的节点都可以不存在子节点。
(A) 至少有一个 (B) 至少有两个 (C) 有且仅有两个 (D) 有且仅有一个
- 298 . 层次模型通过限制节点之间的关系来保证数据的(), 避免数据的冗余和不一致。
(A) 多元性 (B) 一致性 (C) 耦合性 (D) 独立性
- 299 . 层次数据模型的结构特征不包括()。
(A) 简单 (B) 清晰 (C) 复杂 (D) 呆板
- 300 . 层次数据模型只能表示实体之间的()的关系, 不能表示 m:n 的复杂关系, 因此现实世界中的很多模型不能通过该模型方便的表示。
() 1:n () 2:n () n:1 () n:m
- 301 . 层次建模的要素中, ()用于唯一标识一个实体的属性集合, 确保数据的一致性和完整性。
(A) 实体 (B) 键 (C) 属性 (D) 关系
- 302 . 在层次建模中, 需要构造判断矩阵来比较不同元素之间的相对重要性, 判断矩阵的元素使用 1 到 9 及其()来表示不同因素之间的相对重要性。
(A) 中位数 (B) 方差 (C) 倒数 (D) 众数

- 303 . 层次建模过程中, 通过线性和非线性的联系来表示()之间的关系。
(A) 数据 (B) 属性 (C) 键 (D) 实体
- 304 . 一致性检验的步骤不包括()。
(A) 计算一致性指标 CI (B) 查找对应的平均随机一致性指标 RI (C) 计算一致性比例 CR (D) 查询一致性数据 IR
- 305 . 层次建模的存储结构主要采用(), 即层次模型。
(A) 网状结构 (B) 拓扑结构 (C) 树状结构 (D) 星型结构
- 306 . 在层次数据库中不仅要存储数据本身, 还要存储数据之间的()。
(A) 层次联系 (B) 平行关系 (C) 联结方式 (D) 内部参数
- 307 . ()指数据仓库的数据中保存数据的细化程度或综合程度的级别。
(A) 数据宽度 (B) 数据维度 (C) 数据粒度 (D) 数据粘性
- 308 . 数据仓库是企业所决策制定过程, 提供所有支持类型的()。
(A) 数据模型 (B) 数据库 (C) 数据集合 (D) 数据表
- 309 . 维度建模是分析数据的首选技术, 可以以商业用户可理解的方式发布数据以及提供高效的()性能。
(A) 繁殖 (B) 分析 (C) 查询 (D) 处理
- 310 . 为了提高查询性能, 维度建模允许增加数据(), 采用反规范化的设计技术。
(A) 冗余 (B) 集 (C) 表 (D) 列
- 311 . 维度建模的缺点包括()。
(A) 模型简单易理解 (B) 可扩展性好 (C) 支持高效的数据分析 (D) 数据冗余
- 312 . 维度建模是()标准框架, 允许数据库系统和最终用户查询工具在数据方面生成强大的假设条件, 这些数据主要在表现和性能方面起作用。
(A) 规则的 (B) 复杂的 (C) 不可预测的 (D) 可预测的
- 313 . ()是维度模型的核心, 包含了业务过程的度量信息, 每条记录代表了一个度量事件。
(A) 事实表 (B) 维度表 (C) 退化维度 (D) 度量
- 314 . 维度表通常比较(), 是扁平型非规范表, 包含大量的低粒度的文本属性。
(A) 长 (B) 短 (C) 窄 (D) 宽
- 315 . ()是指确定事实表中一条事实记录所代表的业务事件的详细程度。
(A) 声明粒度 (B) 识别维度 (C) 识别事实 (D) 选择业务过程
- 316 . ()是指明确业务过程和维度所属的主题域、明确维度与业务过程的关系, 最终形成一个总线矩阵图表。
(A) 构建明细模型 (B) 构建总线矩阵 (C) 构建汇总模型 (D) 明确统计指标
- 317 . 事实表中包含了业务过程的(), 可以通过与维度表关联来进行分析和查询。
(A) 事实 (B) 维度 (C) 度量 (D) 量表
- 318 . 每个维度表都包含单一的(), 维度表的主键可以作为与之关联的任何事实表的外键, 维度表行的描述环境应与事实表行完全对应。
(A) 主键列 (B) 主键行 (C) 主键 (D) 数据列
- 319 . 定点存储的方式指的是将数据的()转化为反码, 再将反码转化为补码, 最后以补码的形式存储于内存的存储方式。
(A) 二进制原码 (B) 八进制原码 (C) 十进制原码 (D) 十六进制原码
- 320 . 在计算机进行计算等操作时, 如果得到的结果太大以至于原有的存储空间不足以存

- 储操作结果时就会发生()现象。
- (A) 迭代 (B) 另存 (C) 溢出 (D) 自毁
- 321 . ()通过电缆直接与服务器相连接, 存储设备作为服务器的附加硬件, 不带操作系统, 直接接收所连服务器的 I/O 请求, 完全依托服务器, 通过服务器上的网卡向用户提供数据。
- (A) 网络附加存储 (B) 存储区域网络 (C) 直接附加存储 (D) 归档存储
- 322 . ()采用集中式数据存储模式, 将存储设备与服务器彻底分离, 其优点是易于管理和扩展, 适用于需要集中存储和共享数据的场景。
- (A) 网络附加存储 (B) 存储区域网络 (C) 直接附加存储 (D) 对象分离存储
- 323 . IOPS 是指()能处理的读写()请求数, 这是衡量存储性能的主要指标之一。
- (A) 每天 (B) 每小时 (C) 每分钟 (D) 每秒钟
- 324 . 对于需要顺序处理大量数据的应用, 如视频编辑、大数据分析, ()是关键指标, 能够确保数据处理的连续性和效率。
- (A) 顺序读写速度 (B) 随机读写速度 (C) 延迟率 (D) 吞吐率
- 325 . 递归算法解决问题的优势包括()。
- (A) 递归就是在过程或函数里调用自身 (B) 在使用递归策略时, 必须有一个明确的递归结束条件 (C) 递归算法解题通常显得很简洁 (D) 递归算法解题的运行效率较低
- 326 . 哈希算法的优点不包括()。
- (A) 高效性 (B) 哈希冲突 (C) 安全性 (D) 固定长度输出
- 327 . 由于链表在(), 它适用于需要频繁进行这些操作的应用场景。
- (A) 插入和删除操作上的高效性 (B) 长度上的动态性 (C) 内存上的利用率高 (D) 访问上的元素效率低
- 328 . 栈的操作基于()的原则, 入栈和出栈操作都只涉及栈顶元素, 因此操作的时间复杂度为 $O(1)$, 执行速度较快。
- (A) 只进不出 (B) 同时进出 (C) 先进后出 (D) 后进先出
- 329 . 根据选择的特征评估标准, 从上至下递归地生成子节点, 直到数据集不可分则停止生长属于决策树算法构造过程的()。
- (A) 特征选择 (B) 决策树生成 (C) 剪枝 (D) 数据分析
- 330 . 在决策树中, 如果一个节点的所有样本都属于同一个类, 那么该节点被称为()。
- (A) 根节点 (B) 内部节点 (C) 叶节点 (D) 分支节点
- 331 . 支持向量机()是一种常用的机器学习算法, 主要用于线性和非线性分类问题, 回归以及()检测。
- (A) 回归值 (B) 常量值 (C) 向量值 (D) 异常值
- 332 . SVM 中的“支持向量”是()。
- (A) 离分类边界最近的训练样本 (B) 所有的训练样本 (C) 离分类边界最远的训练样本 (D) 错误分类的训练样本
- 333 . 在 Python 中, 函数()用来检查一个元素是否在集合中。
- (A) `isin()` (B) `has_key()` (C) `contains()` (D) `in`
- 334 . ()不是 Python 中的基本数据结构。
- (A) 列表 (list) (B) 字典 (dictionary) (C) 数组 (array) (D) 集合 (set)
- 335 . ()可以通过构建最大堆或最小堆来实现排序。
- (A) 冒泡排序 (B) 快速排序 (C) 堆排序 (D) 归并排序

- 336 . 在排序过程中需要额外的存储空间来存储中间结果的是()。
- (A) 冒泡排序 (B) 直接插入排序 (C) 归并排序 (D) 选择排序
- 337 . 数据画像是通过挖掘和分析海量的数据, 将数据转化为()的展示形式, 从而更加直观地揭示个体和群体的行为特征。
- (A) 图像化 (B) 形象化 (C) 具象化 (D) 抽象化
- 338 . 数据画像的主要目的是()。
- (A) 提高销售额 (B) 减少成本 (C) 理解目标受众 (D) 增加市场份额
- 339 . 在创建数据画像时, 最常用的数据分析方法是()。
- (A) 描述性统计分析 (B) 回归分析 (C) 聚类分析 (D) 假设检验
- 340 . 在数据画像构建流程中, ()涉及到选择相关的特征。
- (A) 数据收集 (B) 数据清洗 (C) 特征工程 (D) 模型评估
- 341 . 婚姻、有无女孩、有无男孩、家里是否有老人、性取向等属于用户标签的()。
- (A) 基础属性 (B) 社会关系 (C) 行为特征 (D) 业务相关
- 342 . 在标签体系中, ()可以确保标签的有效性和准确性。
- (A) 忽略用户反馈 (B) 避免使用动态标签 (C) 定期审查和更新标签 (D) 只使用一种标签类型
- 343 . 特征工程就是一个把原始()转变成特征的过程, 这些特征可以很好的描述这些数据, 并且利用它们建立的模型在未知数据上的表现性能可以达到最优。
- (A) 数据 (B) 特征 (C) 代码 (D) 函数
- 344 . 特征工程就是寻找那些对()具有显著影响的特征, 我们通常称这些影响目标变量的因素为自变量 x , 即特征。
- (A) 自变量 x (B) 自变量 y (C) 目标变量 x (D) 目标变量 y
- 345 . 特征工程的主要目标是()。
- (A) 提高模型的准确性 (B) 减少数据存储空间 (C) 增加数据收集的难度 (D) 降低计算成本
- 346 . ()不属于特征工程的一部分。
- (A) 归一化 (B) 主成分分析 (C) 线性回归 (D) 独热编码
- 347 . ()通过评估每个特征的重要性来选择特征。
- (A) 包裹法 (B) 过滤法 (C) 嵌入法 (D) 验证法
- 348 . ()通过构建一个目标函数来选择特征。
- (A) 包裹法 (B) 过滤法 (C) 嵌入法 (D) 验证法
- 349 . 主成分分析方法()旨在找到数据中的主成分来表征原始数据, 从而达到()的目的。
- (A) 升级 (B) 升维 (C) 降级 (D) 降维
- 350 . ()通过自动训练出区分图像的特征分类器来提取特征。
- (A) 聚类方法 (B) 线或边缘提取方法 (C) 深度学习的方法 (D) 映射方法
- 351 . 在处理缺失值时, 如果数据是正态分布的, 常用的填充方法是()。
- (A) 众数填充 (B) 中位数填充 (C) 均值填充 (D) 随机填充
- 352 . 当数据集中的缺失值比例很高时, 最不推荐的方法是()。
- (A) 删除记录 (B) 使用算法预测填充 (C) 多重插补 (D) 增加缺失值
- 353 . 聚合特征构造通常基于 ID 值或()的某个类别计算数值特征的一些统计量。
- (A) 存储特征 (B) 文本特征 (C) 类别特征 (D) 数值特征

- 354 . 在特征变换中, 标准化的目的是()。
- (A) 使所有特征具有相同的单位 (B) 将特征转换为分类变量 (C) 减少特征的数量
(D) 增加数据的复杂性
- 355 . ()是 Python 数值计算的基础包, 提供了高效的数据结构和计算方法。
- (A) NumPy (B) Pandas (C) scikit-learn (D) Seaborn
- 356 . ()是一个开源的深度学习框架, 也可以用于构建神经网络模型。
- (A) NumPy (B) Pandas (C) scikit-learn (D) TensorFlow
- 357 . 在 scikit-learn 库中, K-Means 算法属于()。
- (A) 分类算法 (B) 回归算法 (C) 聚类算法 (D) 降维算法
- 358 . TensorFlow 中的计算图的主要作用是()。
- (A) 可视化数据 (B) 优化模型性能 (C) 存储训练数据 (D) 定义和运行计算模型
- 359 . ()是一个用于数据分析和处理的库, 提供了 DataFrame 和 Series 等数据结构。
- (A) NumPy (B) Pandas (C) scikit-learn (D) Seaborn
- 360 . matplotlib 属于 Python()类库。
- (A) 机器学习库 (B) 数据处理和分析库 (C) 可视化库 (D) 文本处理库
- 361 . 概要统计以()的形式描述数据的分布特征、集中趋势和离散程度等, 从而帮助人们快速了解数据的整体状况。
- (A) 简明 (B) 详细 (C) 复杂 (D) 抽象
- 362 . ()是衡量数据离散程度的一个重要指标, 它表示数据点相对于平均数的偏离程度。
- (A) 方差 (B) 平均数 (C) 标准差 (D) 众数
- 363 . 概要统计的目的不包括()。
- (A) 描述数据状况 (B) 总结分析结论 (C) 揭示数据规律 (D) 支持决策制定
- 364 . ()不会影响概要统计的准确性。
- (A) 样本大小 (B) 样本的代表性 (C) 数据的质量 (D) 样本的存放位置
- 365 . 统计量()用来衡量数据集中趋势。
- (A) 方差 (B) 众数 (C) 标准差 (D) 极差
- 366 . 在描述一组数据的离散程度时, 统计量()考虑了每个数据与平均数的偏差的平方。
- (A) 众数 (B) 标准差 (C) 方差 (D) 极差
- 367 . 统计工具的运用旨在从数据中提取有用的(), 以便更好地理解现象、预测趋势并做出明智的决策。
- (A) 信息 (B) 代码 (C) 文章 (D) 函数
- 368 . 描述统计的主要目的不包括()。
- (A) 总结数据特征 (B) 进行预测 (C) 描述数据分布 (D) 提供决策依据
- 369 . 按统计工具的功能分类, ()用于将数据转化为图表、图形等形式, 使数据更加直观易懂。
- (A) 数据分析工具 (B) 数据类库 (C) 数据可视化工具 (D) 数据搜集工具
- 370 . 按统计工具的应用领域分类, ()用于市场调研、销售预测、客户关系管理等。
- (A) 商业统计工具 (B) 科研统计工具 (C) 政府统计工具 (D) 数据类库
- 371 . excel 最长数字精度是()而身份证是 18 位, 后面多出的三位会变成 0。
()13 位()14 位()15 位()16 位

- 372 . Excel 的 COUNTIF 函数用于()。
- (A) 计算满足特定条件的单元格数量 (B) 计算数据集中的唯一值数量 (C) 计算数据集的平均值 (D) 计算数据集的总和
- 373 . 有关线性表操作的描述中, 不正确的是()。
- (A) 插入元素 (append, insert) (B) 删除元素 (remove, pop) (C) 查找元素 (index) (D) 遍历元素 (len)
- 374 . 在进行 Python 数据分析时, ()不能提高结果的可靠性。
- (A) 清洗和准备数据 (B) 使用适当的统计方法 (C) 验证分析结果 (D) 仅依赖默认设置进行分析
- 375 . 在使用 SPSS 进行数据分析时, 为保证数据的可靠性, 要检验数据的()。
- (A) 来源 (B) 存放位置 (C) 信效度 (D) 大小
- 376 . 在进行数据分析之前, 要明确研究问题和研究(), 选择适当的统计方法和分析技术, 避免分析过程中的迷茫和盲目。
- (A) 数据 (B) 目的 (C) 方法 (D) 结果
- 377 . 在 Power BI 中, 每次可视化都要求用户在将可视化拖动到画布后, 从“字段”窗格中选择一个或多个字段, 用户必须选中(), 才能在“字段”窗格中包含该字段以进行可视化。
- (A) 复选框 (B) 文本框 (C) 无线框 (D) 输入框
- 378 . Power BI 中数据模型主要包括的三种对象是()。
- (A) 表格、列和行 (B) 度量值、维度和事实 (C) 数据源、模型和报表 (D) 查询、映射和计算
- 379 . 在 Stata 中, ()命令用于导入 Excel 数据。
- (A) use excel (B) import excel (C) read excel (D) load excel
- 380 . 在 Stata 中, 用于生成频数分布表的命令是()。
- (A) tabulate (B) summarize (C) regress (D) list
- 381 . 在 SAS 程序编写时, 有一类输入语句称为(), 它不仅作用于当前编写的那段程序, 而且作用于此次关闭 SAS 前所有的程序过程, 除非接下来另行修改。
- (A) 部分声明 (B) 要点声明 (C) 补充声明 (D) 全局声明
- 382 . 在 SAS 程序编写时, 可以加入一些()来增进理解, 也便于下一次的阅读和修改。
- (A) 概括性语言 (B) 总结性语言 (C) 解释性语言 (D) 补充性语言
- 383 . 在 MATLAB 中, 使用 plot、scatter、bar 等函数绘制()。
- (A) 三维图标 (B) 二维图标 (C) 交互式图表 (D) 自定义图表
- 384 . 在编写 MATLAB 代码时, 需要注意代码的优化和调试, 可以使用 MATLAB 提供的调试工具()等来调试代码, 并使用性能分析工具来优化代码的执行效率。
- (A) 断点、单步执行 (B) 插件 Process (C) 内置 Debug 工具 (D) IDE
- 385 . 线性相关是指一组向量中存在线性关系, 即这些向量可以通过其他向量的线性()来表示。
- (A) 方程 (B) 组合 (C) 关系 (D) 变量
- 386 . 当两个变量的散点图呈现出明显的直线趋势时, 我们可以说这两个变量之间存在()。
- (A) 弱相关 (B) 强相关 (C) 逻辑相关 (D) 线性相关
- 387 . 在一元线性回归中, 斜率()代表()。
- (A) 自变量的平均值 (B) 因变量的平均值 (C) 自变量每增加一个单位时, 因变量

- 的平均变化量 (D) 数据的离散程度
- 388 . 在一元线性回归分析中, R^2 值越接近于 1, 说明模型的拟合效果()。
- (A) 越差 (B) 不变 (C) 越好 (D) 无法确定
- 389 . 多元线性回归分析预测法是指通过对()的自变量与一个因变量的相关性分析, 建立预测模型进行预测和控制的方法。
- (A) 一个 (B) 大于等于二个 (C) 大于等于三个 (D) 大于等于四个
- 390 . 在多元线性回归中, 斜率()代表在其他自变量保持不变的情况下, 某个自变量每增加一个单位时()。
- (A) 自变量的平均值 (B) 因变量的平均值 (C) 数据的离散程度 (D) 因变量的平均变化量
- 391 . 数据可视化是指通过()将数据信息以图形化形式展示出来。
- (A) 文本描述 (B) 语音播放 (C) 图形、图表和动画 (D) 数据库存储
- 392 . 数据可视化的主要目的是()。
- (A) 隐藏数据中的信息 (B) 增加数据处理的复杂性 (C) 帮助人们更快地理解和分析数据 (D) 仅仅为了美观
- 393 . 在视觉编码中, 位置()通常用于表示()。
- (A) 数量大小 (B) 类别 (C) 时间顺序 (D) 层级关系
- 394 . ()有助于提高图表的可读性。
- (A) 相近色调的对比 (B) 高饱和度的颜色对比 (C) 互补色的对比 (D) 灰色调对比
- 395 . ()强调在数据可视化中应突出显示最重要的数据点。
- (A) 简洁性 (B) 层次性 (C) 焦点与上下文 (D) 一致性
- 396 . 数据可视化设计时, 应()平衡信息的复杂性与清晰度。
- (A) 使用尽可能多的细节 (B) 简化到只显示最重要的数据 (C) 牺牲清晰度以显示更多数据 (D) 两者之间的适度折中
- 397 . ()最适合表示时间序列数据。
- (A) 散点图 (B) 条形图 (C) 折线图 (D) 热力图
- 398 . 当你需要比较不同类别的数量时, ()最为直观。
- (A) 饼图 (B) 雷达图 (C) 箱线图 (D) 散点图
- 399 . 使用数据可视化工具时, 首要考虑的是()。
- (A) 工具的知名度 (B) 工具的功能是否齐全 (C) 工具的易用性和学习曲线 (D) 工具的价格
- 400 . ()不是使用数据可视化工具时的良好实践。
- (A) 确保数据的准确性和完整性 (B) 盲目追求图表的美观而忽略数据的实际意义 (C) 根据目标受众调整可视化风格 (D) 定期检查并更新图表
- 401 . 在数据可视化中, 使用色彩对比的主要目的是()。
- (A) 使图表更美观 (B) 强调数据间的差异性 (C) 增加图表的复杂性 (D) 隐藏不重要的数据
- 402 . ()最适合用于表示数据的变化趋势。
- (A) 邻近色搭配 (B) 对比色搭配 (C) 渐变色搭配 (D) 灰度色搭配
- 403 . 图表布局中, ()通常用于引导观众的视线流动。
- (A) 标题 (B) 网格线 (C) 视觉层次 (D) 图例
- 404 . 在设计图表时, ()是排版中应避免的。
- (A) 保持文字清晰可读 (B) 使用过多字体样式 (C) 合理对齐元素 (D) 确保信息层次

清晰

- 405 . 在数据可视化中，交互性设计的主要目的是()。
(A)增加图表的复杂度 (B)使观众能够更深入地探索数据 (C)仅仅为了美观 (D)减少图表的信息量
- 406 . ()常用于筛选和突出显示图表中的数据点。
(A)鼠标悬停提示 (B)动态过滤 (C)缩放和平移 (D)数据排序
- 407 . 评估数据可视化质量时，()不是主要考虑的因素。
(A)数据的准确性 (B)视觉的吸引力 (C)信息的完整性 (D)图表的复杂度
- 408 . ()直接关联到数据可视化是否能够帮助观众快速理解信息。
(A)美观度 (B)信息层次清晰 (C)交互性 (D)数据源的可靠性
- 409 . 在数据可视化中，()最适合用于表示数量的多少。
(A)颜色 (B)大小 (C)形状 (D)位置
- 410 . ()最容易被人眼捕捉和感知。
(A)颜色的细微变化 (B)位置的移动 (C)形状的轻微变形 (D)大小的明显变化
- 411 . 当需要展示数据的占比关系时，最适合使用的图标类型是()。
(A)折线图(B)柱状图(C)饼图(D)散点图
- 412 . 最适合用于展示时间序列数据的变化趋势的是()。
(A)柱状图(B)饼图 (C)折线图 (D)热力图
- 413 . 在数据可视化设计中，确保准确性的首要步骤是()。
(A)选择合适的颜色方案(B)验证数据源的可靠性 (C)设计吸引人的布局(D)应用交互功能
- 414 . 当在数据可视化中展示百分比数据时，()最符合准确性原则。
(A)使用整数表示百分比(B)使用小数表示百分比，并保留到小数点后两位(C)使用百分号符号(%)明确标识(D)将百分比转换为分数形式表示
- 415 . 可视化组件库主要用于()。
(A)存储和管理图片 (B)提供预定义的界面元素和图表，简化开发流程 (C)设计网站的整体布局 (D)进行数据分析
- 416 . ()不是可视化组件库通常包含的元素。
(A)图表组件 (B)按钮和输入框 (C)完整的后台管理系统 (D)布局容器
- 417 . ()不属于常见的组件库分类方式。
(A)按行业领域 (B)按开发平台(C)按功能(D)按颜色主题
- 418 . 组件库中的“原子设计”理念强调了()。
(A)组件应该尽可能大，包含更多功能 (B)组件应该是不可拆分的最小单元 (C)组件应该根据用户习惯进行定制 (D)组件之间应该相互独立
- 419 . 在使用基础图表组件时，首先需要考虑()。
(A)图表的外观是否美观 (B)数据源的正确性和完整性 (C)图表类型是否流行 (D)图表的颜色搭配
- 420 . ()不是在使用基础图表组件时推荐的。
(A)根据数据特点选择合适的图表类型 (B)确保图表的标签和标题清晰可读 (C)忽略图表的交互性，只关注静态展示 (D)对图表进行必要的注释和说明
- 421 . 高级分析组件主要服务于()需求。
(A)简单的数据展示 (B)深入的数据分析和挖掘 (C)用户界面的美化 (D)网站的 SEO 优化

- 422 . 在使用高级分析组件时, ()不是必要的准备工作。
(A)了解组件的算法原理 (B)准备充足且准确的数据集 (C)设计美观的组件外观
(D)评估组件的性能和资源消耗
- 423 . 交互组件的主要目的是()。
(A)装饰用户界面 (B)提供用户与数据之间的动态交互 (C)增加页面的动画效果
(D)隐藏不必要的信息
- 424 . 设计交互组件时, 最重要的是()。
(A)组件的外观设计 (B)组件的响应速度和流畅性 (C)组件的复杂性 (D)组件的品牌一致性
- 425 . 在设置组件属性时, ()是不推荐的。
(A)仔细阅读组件文档, 了解每个属性的作用和限制 (B)随意设置属性值, 不考虑其对组件外观和功能的影响 (C)根据实际需求调整属性, 确保组件符合预期效果 (D)使用默认值作为起点, 逐步调整以达到最佳效果
- 426 . 当组件的某个属性有多种类型可选时, 应该()。
(A)总是选择最简单的类型, 以降低出错率 (B)选择最复杂的类型, 以支持更多功能
(C)根据组件的具体用途和数据类型, 选择最合适的类型 (D)随意选择,
- 427 . 在集成一个新的组件库到项目中时, 首先需要()。
(A)直接将组件库文件拖入项目中 (B)查阅组件库的集成文档, 了解集成步骤和依赖关系 (C)跳过集成文档, 直接参考示例代码 (D)假设组件库与项目当前使用的技术栈完全兼容
- 428 . 使用组件库时, 为了保持代码的可维护性和可读性, 应该()。
(A)将所有组件的引用和配置集中在一个文件中 (B)随意放置组件的使用代码, 不考虑模块化和封装 (C)根据组件的功能和用途, 将相关代码组织在不同的模块或组件中 (D)使用全局变量来管理组件的状态和配置
- 429 . 为了优化组件的性能, ()是有效的。
(A)增加组件的复杂度, 以提供更多功能 (B)减少不必要的 DOM 操作, 优化渲染过程
(C)忽略组件的懒加载和代码分割 (D)使用大量的动画和过渡效果, 以增强用户体验
- 430 . 在 Web 应用中, ()可以优化长列表组件的性能。
(A)一次性加载和渲染所有列表项 (B)使用虚拟滚动技术, 只渲染可视区域内的列表项
(C)增加列表项的复杂度, 以提高用户满意度 (D)禁用浏览器的滚动平滑效果
- 431 . 在处理组件的兼容性问题时, ()不是必需的。
(A)确定目标浏览器和设备的范围 (B)使用 shim 来填补旧浏览器中的功能缺失 (C)忽略非目标浏览器的兼容性问题 (D)进行跨浏览器测试, 确保组件在所有目标环境中正常工作
- 432 . 响应式设计主要解决的是()。
(A)组件在不同屏幕尺寸下的布局问题 (B)组件的动画效果在不同浏览器中的一致性问题
(C)组件的兼容性问题 (D)组件的性能优化问题
- 433 . 在电商业务中, 衡量用户粘性的关键业务指标是()。
(A)订单量 (B)用户复购率 (C)平均客单价 (D)页面访问量
- 434 . ()不属于衡量网站运营效率的常见业务指标。
(A)页面加载时间 (B)转化率 (C)用户满意度调查 (D)跳出率
- 435 . 在进行竞品分析时, 采用对比能力分析方法主要是为了()。
(A)确定自身的市场定位 (B)预测未来市场趋势 (C)评估不同竞品之间的优势和劣势
(D)制定营销策略

- 436 . ()不常用于产品功能对比。
(A)雷达图 (B)柱状图 (C)散点图 (D)折线图
- 437 . 在进行销售趋势预测时,最常用的时间序列分析方法是()。
(A)回归分析 (B)聚类分析 (C)移动平均法 (D)决策树分析
- 438 . ()不是影响趋势预测准确性的关键因素。
(A)数据的质量与完整性 (B)预测模型的复杂度 (C)预测时间范围的长短 (D)预测者的主观判断
- 439 . 在社会科学研究中,为了推断因果关系,最常用的方法是()。
(A)相关性分析 (B)回归分析 (C)聚类分析 (D)假设检验
- 440 . ()不是进行因果关系推断所必需的。
(A)时间顺序性 (B)相关性 (C)排除其他合理解释 (D)数据的随机性
- 441 . 在数据分析中,异常值()通常是指()。
(A)数据中最大或最小的值 (B)与数据集中其他观测值明显不同的值 (C)任何不在平均值周围的价值 (D)由错误测量或数据录入错误产生的值
- 442 . 在处理数据集中的异常值时,通常不会采用的方法是()。
(A)删除异常值 (B)将异常值替换为平均值 (C)转换数据以减少异常值的影响 (D)使用专门的统计技术来识别和分析异常值
- 443 . 在数据分析中,描述数据集中趋势的统计量通常是()。
(A)标准差 (B)中位数 (C)众数 (D)方差
- 444 . 偏度用于描述数据分布的()。
(A)对称性 (B)集中趋势 (C)离散程度 (D)尖峰或扁平程度
- 445 . 在业务决策中使用数据可视化结果时,最重要的是()。
(A)确保数据的准确性 (B)选择最美观的可视化方式 (C)快速完成可视化过程 (D)依赖单一的可视化结果
- 446 . ()不是数据可视化在业务决策中的优势。
(A)提高数据理解速度 (B)隐藏数据的复杂性 (C)辅助识别趋势和模式 (D)便于跨部门沟通
- 447 . 在进行数据可视化结果与业务 KPI 关联分析时,首先要确保()。
(A)可视化结果的美观性 (B)KPI 定义明确性 (C)数据更新的实时性 (D)可视化工具的功能强大
- 448 . ()不是关联分析时常见的挑战。
(A)数据源的不一致性 (B)可视化结果难以理解 (C)KPI 之间的独立性 (D)关联关系的复杂性
- 449 . 评估数据可视化结果质量时,()最重要。
(A)图表的美观性 (B)数据的准确性和完整性 (C)视觉元素的多样性 (D)可视化技术的先进性
- 450 . ()不是评估数据可视化结果质量的有效方法。
(A)清晰度测试:评估非专家用户能否轻松理解可视化所传达的信息。(B)一致性检查:确保数据可视化中的信息与原始数据源保持一致。(C)美感评分:仅基于图表的美观程度进行评分,不考虑信息的传达效果。(D)目标受众反馈
- 451 . 市场行情分析的常用指标中,主要用()衡量市场的超买超卖状态。
(A)移动平均线 (B)相对强弱指数 (RSI) (C)成交量比例 (D)中间意愿指标 (CR)
- 452 . 市场行情分析中,()指标是通过计算一定时期内的平均价格来平滑价格数据的。

- (A) 移动平均线 (B) 相对强弱指数 (C) 成交量比例 (D) 中间意愿指标
- 453 . ()不是市场行情数据的主要来源。
(A) 财经新闻网站 (B) 社交媒体上的谣言 (C) 金融数据服务提供商 (D) 政府机构发布的报告
- 454 . 实时市场行情数据通常通过()从交易所获取。
(A) 交易所 API (B) 移动平均线 (C) 中间意愿指标 (D) 成交量比例
- 455 . 在市场供给与需求分析中, ()是描述生产者在不同价格水平下愿意提供的商品数量的曲线。
(A) 需求曲线 (B) 供给曲线 (C) 成本曲线 (D) 收益曲线
- 456 . 在市场供给与需求分析方法中, 当供给量小于需求量时, 市场价格通常会()上升。
(A) 上升 (B) 下降 (C) 保持不变 (D) 无法确定
- 457 . 关于市场容量与趋势分析的说法中, ()是不正确的。
(A) 市场容量是指市场的最大承载能力 (B) 市场容量分析有助于了解市场的规模和潜力 (C) 市场趋势分析只能基于历史数据 (D) 市场趋势分析包括识别市场的周期性行为
- 458 . 市场容量分析主要关注()。
(A) 市场的短期波动 (B) 市场的长期发展趋势 (C) 单个投资者的行为 (D) 产品的生产成本
- 459 . 在分析新兴经济体的市场潜力时, 企业最应关注的是()。
(A) 该经济体的政治制度稳定性 (B) 该经济体的 GDP 增长率 (C) 该经济体的自然资源储备 (D) 该经济体的文化习俗差异
- 460 . 市场潜力分析中, 用于衡量产品或服务在目标市场中的潜在需求量的方法是()。
(A) 移动平均线 (B) 市场调研 (C) 成交量比例 (D) 中间意愿指标
- 461 . 市场分类从区分目的来说有()这两大种类。
(A) 地理型和投资型 (B) 管理型和地理型 (C) 管理型和投资型 (D) 管理型和消费者导向型
- 462 . GICS 行业分类系统属于()的行业分类。
(A) 地理型 (B) 投资型 (C) 管理型 (D) 消费者导向型
- 463 . 行业数据分析不包括()。
(A) 行业数据采集 (B) 产业链分析 (C) 市场生命周期分析 (D) 竞争产品分析
- 464 . 在进行行业盈利能力分析时, ()指标可以反映行业的平均利润水平。
(A) 行业市盈率 (B) 行业毛利率 (C) 行业杠杆率 (D) 行业流动比率
- 465 . 企业获取行业状况数据的最常见和直接的方式是()。
(A) 通过政府发布的官方统计数据 (B) 依赖个人经验和直觉判断 (C) 雇佣专业市场研究机构进行定制调研 (D) 利用公开的互联网资源和行业报告
- 466 . 在获取行业状况数据时, 最权威且官方的数据来源通常是()。
(A) 政府统计局网站 (B) 证券交易所 (C) 行业媒体 (D) 市场调研公司
- 467 . ()不属于市场分析法。
(A) 波特五力模型 (B) SWOT 分析法 (C) 角色扮演法 (D) PEST 分析法
- 468 . 用于评估行业内部企业间竞争态势的分析方法是()。
(A) 成交量比例分析 (B) 波特五力模型 (C) 移动平均线预测 (D) 相对强弱指数对比

- 469 . 当行业进入成熟期时, () 的特征最为显著。
 (A) 市场份额高度集中, 少数几家企业主导市场 (B) 技术创新频繁, 产品更新换代快 (C) 市场需求增长放缓, 竞争趋于稳定 (D) 企业数量急剧减少, 市场逐渐萎缩
- 470 . 在行业的一般特征分析中, 用于反映行业技术水平和创新能力的是()。
 (A) 市场份额 (B) 行业利润率 (C) 行业集中度 (D) 技术发展水平
- 471 . 市场集中度指标主要包括()、成交销量和流量。
 (A) 成交笔数 (B) 成交金额 (C) 成交时间 (D) 成交区域
- 472 . 市场集中度是指产值、产量、销售额、()、职工人数、资产总额等的总和。
 (A) 销售量 (B) 销售度 (C) 销售数 (D) 销售值
- 473 . 行业生命周期分析中, 描述行业处于高速增长阶段, 企业数量迅速增加, 竞争激烈化的时期是()。
 (A) 初创期 (B) 成长期 (C) 成熟期 (D) 衰退期
- 474 . 行业生命周期分析中, 行业增长率开始放缓, 企业利润趋于稳定的时期是()。
 (A) 初创期 (B) 引入期 (C) 成熟期 (D) 衰退前期
- 475 . 竞争对手的识别方法中最常用且有效的是()。
 (A) 通过市场调研了解行业内的主要企业和品牌 (B) 依赖个人经验和直觉判断 (C) 仅仅关注直接竞争对手, 忽略潜在和间接竞争对手 (D) 通过社交媒体上的用户评论来识别
- 476 . 通过分析企业自身的资源和能力, 以及潜在竞争对手进入市场的难易程度来识别竞争对手的方法是()。
 (A) 市场分析法 (B) 进入壁垒分析 (C) 价值链分析 (D) PEST 分析
- 477 . 企业把同一行业中以相似的价格向相同的顾客提供类似产品或服务的企业称为()竞争者。
 (A) 行业 (B) 需要 (C) 品牌 (D) 消费
- 478 . 企业把提供同种或同类产品, 但规格、型号、款式不同的企业称为()竞争者。
 (A) 需要 (B) 行业 (C) 品牌 (D) 消费
- 479 . () 属于分析竞争对手商品信息的角度。
 (A) 客单价 (B) 商品细节 (C) 店铺收藏 (D) 网店地址
- 480 . 从商品价格、物流方式、商品细节等方面展示商品品质属于分析竞争对手的() 角度。
 (A) 详情页 (B) 推广活动 (C) 商品评论 (D) 商品信息
- 481 . 进行竞争对手分析的主要意义是()。
 (A) 了解竞争对手的财务状况, 以便进行投资决策 (B) 单纯为了了解竞争对手的优劣势, 不用于指导企业战略 (C) 识别竞争对手的战略意图和动向, 为企业制定竞争策略提供依据 (D) 仅仅为了收集竞争对手的信息, 不进行深入分析
- 482 . 企业在进行竞争对手实力分析时, 对于实力相当的竞争对手, () 的描述是正确的。
 (A) 是企业学习和追赶的目标 (B) 竞争力较小, 不构成威胁 (C) 行业中强劲的对手, 相互竞争大 (D) 有可能是黑马, 能迅速进入行业前列
- 483 . 竞争对手分析中, 用于识别竞争对手优势和劣势的分析工具是()。
 (A) 市场份额分析 (B) SWOT 分析 (C) 五力模型 (D) 成本效益分析
- 484 . 在竞争对手分析中, 通过分析竞争对手的成本结构来评估其盈利能力的方法是

- ()。
- (A) 市场份额分析 (B) 成本结构分析 (C) 价值链分析 (D) 市场渗透率分析
- 485 . 下单转化率分析是分析()。
- (A) 访问量与注册用户数的比例 (B) 访问量与下单用户数的比例 (C) 下单用户数与支付成功用户数的比例 (D) 支付成功用户数与复购用户数的比例
- 486 . 在分析竞争对手下单转化率时, 首先要收集的数据是()。
- (A) 竞争对手的订单量与访客数 (B) 竞争对手的库存量 (C) 竞争对手的营销策略 (D) 竞争对手的财务状况
- 487 . 评估竞争对手客户拥有量的关键指标不包括()。
- (A) 客户满意度 (B) 客户增长率 (C) 市场份额 (D) 客户满意度调查问卷的数量
- 488 . 在竞争对手客户拥有量分析中, 通过对比竞争对手与自身在特定市场区域的客户数量, 可以评估的方面是()。
- (A) 市场份额 (B) 客户满意度 (C) 产品定价 (D) 品牌知名度
- 489 . 竞争对手 SWOT 分析中的“S”代表()。
- (A) 优势 (B) 劣势 (C) 机会 (D) 威胁
- 490 . 竞争对手 SWOT 分析中的“W”代表劣势, ()不是竞争对手的劣势。
- (A) 产品创新能力不足 (B) 资金链紧张, 融资困难 (C) 市场营销策略不够灵活 (D) 市场份额持续增长, 盈利能力强
- 491 . 在构建与竞争对手的合作与联盟时, 企业通常采取()来确保合作的顺利进行并最大化双方的利益。
- (A) 严格保密, 避免与竞争对手分享任何关键信息 (B) 强调竞争, 通过合作中的小范围竞争来激发创新 (C) 签订详细合同, 明确双方的权利、义务和利益分配机制 (D) 单纯依赖口头协议和信任, 不诉诸法律文件
- 492 . 竞争对手合作与联盟策略中, 通过共享资源和技术以降低成本的合作方式被称为()。
- (A) 战略联盟 (B) 成本共享联盟 (C) 营销联盟 (D) 研发联盟
- 493 . 产品功能形态和解决方案不同, 但用户群高度重合, 解决的用户需求相同的产品是()。
- (A) 直接竞品 (B) 间接竞品 (C) 替代品 (D) 一般竞品
- 494 . 在竞争产品识别过程中, 通过市场调研来收集并分析竞争对手产品信息的方法被称为()。
- (A) 市场情报法 (B) 技术分析法 (C) 成本效益法 (D) 用户反馈法
- 495 . 评估竞争产品耐用性时, 常采用的方法是()。
- (A) 用户满意度调查 (B) 品牌声誉分析 (C) 使用寿命测试 (D) 价格对比分析
- 496 . 竞争产品特性分析中, 关注产品是否符合特定行业标准或法规的方法是()。
- (A) 功能创新评估 (B) 合规性检查 (C) 市场趋势分析 (D) 用户体验测试
- 497 . ()主要用于分析竞争产品的定价策略。
- (A) 成本加成法 (B) 竞争对比法 (C) 价值定价法 (D) 市场渗透定价法
- 498 . ()不是分析竞争产品定价策略时常用的工具或方法。
- (A) 波特五力模型 (B) 价格敏感度分析 (C) 价值曲线分析 (D) 成本结构对比
- 499 . 竞争产品分销渠道的宽度, 通常指的是()。
- (A) 分销渠道的深度 (B) 分销渠道中中间商的数量 (C) 分销渠道的覆盖范围 (D) 分销渠道的利润率

- 500 . 竞争产品分销渠道的效率评估, 主要依据的是()。
- (A) 分销渠道的知名度 (B) 分销渠道的多样性 (C) 产品从生产到消费者手中的时间 (D) 分销渠道的广告投入
- 501 . 在竞争产品的促销策略分析中, 评估促销活动的直接经济效果通常使用的方法是()。
- (A) 销售数据分析 (B) 消费者心理分析 (C) 竞争对手动态监测 (D) 市场趋势预测
- 502 . 竞争产品的促销策略分析中, 分析促销活动对消费者购买决策的影响程度时, 常用的方法是()。
- (A) 市场调研问卷 (B) 竞争对手财务分析 (C) 消费者行为研究 (D) 产品生命周期分析
- 503 . ()可以用来衡量竞争产品在市场中的表现。
- (A) 毛利率 (B) 市场占有率 (C) 客户满意度 (D) 研发投入
- 504 . 在分析竞争产品的市场份额与增长率时, 企业通常会使用()图表来直观展示。
- (A) 柱状图 (B) 折线图 (C) 饼图 (D) 散点图
- 505 . ()不能用于收集和分析竞争产品的客户反馈与评价。
- (A) 在线调查 (B) 社交媒体监听 (C) 客户访谈 (D) 竞争对手分析报告
- 506 . 在分析竞争产品的客户反馈与评价时, 企业为了确保信息的准确性和客观性, 不可以()。
- (A) 采用多种数据来源进行交叉验证 (B) 避免仅依赖单一渠道或个体的反馈 (C) 定期对收集到的反馈进行汇总和分析 (D) 建立客户反馈的激励机制以提高反馈数量
- 507 . 产品生命周期分析中, 通常认为产品进入衰退期的标志是()。
- (A) 销售额持续增长 (B) 市场份额稳步提高 (C) 销售额增长放缓或下降 (D) 产品成本大幅降低
- 508 . 竞争产品生命周期分析中, 评估产品市场潜力的关键指标是()。
- (A) 产品生产成本 (B) 市场需求规模 (C) 产品研发周期 (D) 产品更新换代速度
- 509 . 在分析竞争产品的供应链时, 企业可以通过()的方法来评估其供应链的竞争力。
- (A) 仅考虑供应链的成本优势 (B) 仅关注供应链的响应速度 (C) 综合评估供应链的可靠性、成本效率、灵活性和创新能力 (D) 忽视供应链的长期效益, 仅关注短期利润
- 510 . 在分析竞争产品供应链时, 识别潜在供应链风险的关键步骤是()。
- (A) 评估产品市场需求 (B) 分析产品生产成本 (C) 监测竞争对手市场份额 (D) 评估供应商与物流合作伙伴的稳定性
- 511 . 在波特五力模型中, ()这一力量关注行业内现有企业之间的竞争程度。
- (A) 供应商的议价能力 (B) 购买者的议价能力 (C) 行业内竞争者的竞争程度 (D) 替代品的威胁
- 512 . 波特五力模型由()管理学家提出。
- (A) 迈克尔·波特 (B) 彼得·德鲁克 (C) 菲利普·科特勒 (D) 弗雷德里克·温斯洛·泰勒
- 513 . 在应用波特五力模型之前, 首先需要明确的是()。
- (A) 竞争对手的财务状况 (B) 行业的定义和范围 (C) 消费者的个人喜好 (D) 政府的最新政策
- 514 . 在使用波特五力模型时, 首先需要做的是()。
- (A) 明确分析目的和范围 (B) 收集行业数据 (C) 评估各力量的强弱 (D) 制定应

对策略

- 515 . 替代品威胁的动态特征主要体现在()的变化上。
(A) 行业内企业的成本结构 (B) 替代品的功能、价格与市场接受度 (C) 供应商的议价能力 (D) 潜在进入者的数量
- 516 . 波特五力模型中, 供应商议价能力的动态性主要体现在()上。
(A) 行业内企业的数量 (B) 替代品的市场占有率 (C) 消费者需求的变化 (D) 供应商集中度与产品差异化程度
- 517 . ()不是影响新进入者威胁的主要因素。
(A) 规模经济 (B) 转换成本 (C) 市场需求增长 (D) 品牌忠诚度
- 518 . 在分析新进入者威胁时, 最常用的方法是()。
(A) SWOT 分析 (B) 进入壁垒分析 (C) 五力模型分析 (D) 价值链分析
- 519 . 替代品威胁的大小主要取决于()。
(A) 替代品的价格 (B) 替代品的性能 (C) 消费者的偏好 (D) 上述都是
- 520 . 在应对策略中, 为了降低替代品威胁, 企业可以采取的措施是()。
(A) 提高生产成本 (B) 减少产品差异化 (C) 提升产品附加值 (D) 降低产品质量
- 521 . ()的情况最可能增强买方的议价能力。
(A) 买方数量少且集中 (B) 产品差异化程度高 (C) 买方购买量小且频繁 (D) 替代品少且性能不佳
- 522 . 买方议价能力的高低主要取决于()。
(A) 产品的生产成本 (B) 产品的销售渠道 (C) 买方的购买实力与谈判技巧 (D) 产品的品牌形象
- 523 . 卖方议价力量分析中, 最关键的是评估()。
(A) 产品的生产成本 (B) 供应商的集中度与议价能力 (C) 产品的销售渠道 (D) 产品的品牌形象
- 524 . 卖方议价力量分析中, 若市场上存在大量替代品, 则卖方()议价能力。
A) 增强 (B) 不变 (C) 减弱 (D) 无法判断
- 525 . ()不是分析现有竞争者时需要考虑的因素。
(A) 竞争者的市场份额 (B) 竞争者的财务状况 (C) 竞争者的营销策略 (D) 政府的税收政策
- 526 . 在监测现有竞争者时, ()不是主要关注的方面。
(A) 竞争对手的新产品开发能力 (B) 竞争对手的市场份额变化 (C) 竞争对手的营销策略 (D) 产品的差异化程度
- 527 . 波特五力模型主要用于分析企业的()。
(A) 行业竞争态势 (B) 内部运营效率 (C) 市场营销策略 (D) 供应链管理
- 528 . 波特五力模型主要用于分析企业的()环境。
(A) 竞争 (B) 内部 (C) 市场 (D) 财务
- 529 . 单一窗口是指()。
(A) 国际贸易中, 为简化手续、提高效率而建立的信息共享平台 (B) 国内贸易中, 为降低成本而建立的交易平台 (C) 海关部门内部的信息管理系统 (D) 企业内部的业务流程优化系统
- 530 . ()不是单一窗口的直接作用对象。
(A) 进出口企业 (B) 海关、税务等政府部门 (C) 国内非贸易企业 (D) 物流、金融等配套服务企业

- 531 . 通过登录单一窗口平台，可一次提交申报数据和随附单证，满足口岸管理相关部门及运输工具、监管场所的监管要求，这属于()功能。
(A) 贸易企业 (B) 运输企业 (C) 监管场所 (D) 口岸管理相关部门
- 532 . ()不是单一窗口的功能。
(A) 运输企业递交航行数据 (B) 贸易企业提交进口货物单证 (C) 海关发布税收政策 (D) 口岸管理部门获取申报数据
- 533 . 单一窗口的主要目的是()。
(A) 简化通关流程 (B) 增加企业成本 (C) 降低政府监管 (D) 阻碍国际贸易
- 534 . 单一窗口的建设对于提升()具有重要意义。
(A) 政府行政效率 (B) 国际贸易竞争力 (C) 企业内部管理水平 (D) 个人生活质量
- 535 . 跨境电子商务数据申报主要通过()平台进行。
(A) 阿里巴巴 (B) 亚马逊 (C) 国际贸易“单一窗口” (D) 海关官方网站
- 536 . 跨境电子商务数据申报的内容通常不包括()。
(A) 交易双方信息 (B) 商品信息 (C) 物流信息 (D) 消费者个人偏好
- 537 . 在跨境电子商务数据申报中，()是必须提供的。
(A) 商品的尺寸和颜色 (B) 交易双方的银行卡账户信息 (C) 商品的详细描述和 HS 编码 (D) 物流公司的内部编号
- 538 . 跨境电子商务数据申报流程中，()是确保商品能够顺利清关的关键。
(A) 商品发货 (B) 数据提交与申报 (C) 数据审核与反馈 (D) 物流运输
- 539 . 跨境电子商务数据申报时，若商品涉及特殊监管要求，应()进行额外申报。
(A) 按照相关规定 (B) 忽略该要求 (C) 仅向海关口头说明 (D) 向电商平台报告即可
- 540 . 跨境电子商务数据申报时，若商品涉及禁运或限制进口，应特别注意()的合规性。
(A) 商品质量 (B) 进口许可和限制 (C) 商品价格 (D) 商品包装
- 541 . 在出口退税数据申报流程中，首先需要完成的是()环节。
(A) 数据采集与整理 (B) 提交退税申请 (C) 审核与反馈 (D) 退税款到账确认
- 542 . 企业在提交出口退税数据申报前，应确保所有()均已正确无误。
(A) 银行账户信息 (B) 生产记录 (C) 出口发票及报关单信息 (D) 市场推广计划
- 543 . 外贸企业在完成出口退税申报前，首先需要()。
(A) 提交财务报表 (B) 取得出口产品经营权批准文件 (C) 签订出口合同 (D) 支付关税
- 544 . 出口退税申报系统中，外贸企业需要录入的核心信息包括()。
(A) 商品生产成本 (B) 国内销售价格 (C) 出口发票号和报关单号 (D) 市场推广费用
- 545 . 生产企业在进行出口退税申报时，()环节是确保数据准确性的关键
(A) 数据录入 (B) 关联检查 (C) 税务咨询 (D) 物流安排
- 546 . 生产企业在进行出口退税申报前，首先需要完成的是()工作。
(A) 生产及出口数据的整理与核对 (B) 向海关提交报关单 (C) 向银行申请外汇结算 (D) 向税务局提交财务报表
- 547 . 口岸物流的核心是()。
(A) 货物运输 (B) 港口服务 (C) 国际贸易支持 (D) 金融服务

- 548 . 在口岸物流中, 实现货物快速通关的关键在于()的协同作业。
 (A) 海关、检验检疫、物流等多部门 (B) 物流公司内部各部门 (C) 金融机构与物流公司 (D) 政府与旅游机构
- 549 . 口岸物流的最大特点是()。
 (A) 货物存储 (B) 换载、接驳货物 (C) 跨境支付 (D) 货物加工
- 550 . 口岸物流在国际贸易中扮演着()的角色, 是连接国内外市场的桥梁。
 (A) 辅助性 (B) 从属性 (C) 关键性 (D) 边缘性
- 551 . ()不是口岸物流的主要功能。
 (A) 装卸搬运 (B) 加工制造 (C) 仓储 (D) 信息处理
- 552 . 口岸物流的核心功能之一是(), 它确保了货物在口岸的高效流通。
 (A) 货物的集散与转运 (B) 商品的生产与加工 (C) 市场的开发与营销 (D) 金融服务的提供
- 553 . 口岸物流数据申报流程的第一步是()。
 (A) 缴纳税费 (B) 提交单证 (C) 查验与放行 (D) 货物装船
- 554 . 口岸物流数据申报时, ()环节涉及将货物信息、运输信息及报关单信息整合。
 (A) 海关查验 (B) 数据录入系统 (C) 缴纳税费 (D) 物流跟踪
- 555 . 电子口岸平台通过(), 实现信息的实时传递与共享。
 (A) 互联网与专用网络 (B) 企业内部网络 (C) 公共电话网络 (D) 移动通信网络
- 556 . 电子口岸是指利用现代信息技术, 将()等相关部门的业务数据进行电子交换和共享的平台。
 (A) 海关、税务、外汇管理 (B) 银行、保险、物流 (C) 电商、支付、物流 (D) 制造、销售、服务
- 557 . 在电子口岸进行物流数据申报时, 确保数据的()是首要原则。
 (A) 准确性 (B) 及时性 (C) 完整性 (D) 保密性
- 558 . 电子口岸物流数据申报中, 若货物涉及特殊监管条件, 需特别注意()的填写。
 (A) 货物尺寸 (B) 货物包装 (C) 许可证编号 (D) 运输路线
- 559 . 跨境电子商务是指()通过互联网及其相关信息技术实现商品或服务的交易, 并通过跨境支付完成结算的商业模式。
 (A) 同一国家内的交易双方 (B) 不同国家或地区的交易双方 (C) 政府机构之间的交易 (D) 消费者与金融机构之间的交易
- 560 . 在跨境电子商务中, 电子商务平台主要扮演()的角色。
 (A) 商品生产者 (B) 物流服务商 (C) 交易中介 (D) 金融监管机构
- 561 . 跨境电子商务的核心特征是(), 即交易双方分属不同国家或地区。
 (A) 线上交易 (B) 跨境性 (C) 物流快速 (D) 支付便捷
- 562 . 跨境电子商务相比传统国际贸易, 不具有()特点。
 (A) 交易环节简化 (B) 交易成本低 (C) 交易效率高 (D) 交易风险更低
- 563 . 企业与企业之间的电子商务交易属于()。
 ()C2C 模式()B2C 模式()B2B 模式()跨境直运
- 564 . 根据交易主体类型, 跨境电子商务主要分为 B2B 和()两种模式。
 ()B2C()C2C()B2C 与 C2C()G2B
- 565 . 快速、高效和广泛覆盖, 适用于中小件商品和需要较快交货的情况下, 可以选择()方式。

- (A) 国际快递 (B) 海运运输 (C) 航空运输 (D) 铁路运输
- 566 . 跨境电子商务物流方式中, () 通常最经济实惠但速度较慢的方式。
(A) 国际快递 (B) 航空运输 (C) 海运运输 (D) 铁路运输
- 567 . () 方式不属于跨境电子商务常见的支付方式。
(A) 信用卡支付 (B) 现金交易 (C) 第三方支付 (D) 银行转账
- 568 . 在跨境电子商务中, 为了保障支付安全, 消费者应优先选择() 进行支付。
(A) 未经认证的第三方支付平台 (B) 知名且经过安全认证的支付网关 (C) 直接银行转账 (D) 邮件汇款
- 569 . 在跨境电子商务中, 为了降低物流风险, 商家应优先选择() 的物流公司。
(A) 信誉良好且经验丰富的 (B) 价格最低的 (C) 新成立的 (D) 未经过认证的
- 570 . 跨境电子商务中, 面对汇率风险, 商家可以采取的应对措施是()。
(A) 不关注汇率变动 (B) 使用金融工具进行风险管理 (C) 随意调整商品价格 (D) 仅依赖支付平台处理
- 571 . 出口退税的监管主要由() 负责。
(A) 海关 (B) 税务局 (C) 商务部 (D) 外汇管理局
- 572 . 出口退税的主要是为了()。
(A) 鼓励出口 (B) 限制进口 (C) 增加财政收入 (D) 降低企业成本
- 573 . 出口退税主要涉及() 两种税种。
(A) 增值税和所得税 (B) 增值税和关税 (C) 增值税和消费税 (D) 营业税和增值税
- 574 . 出口退税政策主要针对的是() 类型的税。
(A) 流转税 (如增值税、消费税) (B) 所得税 (C) 财产税 (D) 行为税
- 575 . () 不属于出口退税的类型。
(A) 出口免税并退税 (B) 出口征税 (C) 出口退税但不免税 (D) 出口免税但不退税
- 576 . “免、抵、退” 税办法中的“免” 指的是()。
(A) 免征进口环节税 (B) 免征出口环节税 (C) 免征企业所得税 (D) 免征增值税和消费税
- 577 . () 不属于出口退税的适用范围。
(A) 生产企业自营出口货物 (B) 外贸企业代理出口货物 (C) 进口货物再出口 (D) 生产企业视同自产货物出口
- 578 . 出口退税政策主要适用于() 类型的企业。
(A) 具有出口经营权的生产企业 (B) 所有国内企业 (C) 外资企业 (D) 个体工商户
- 579 . 出口退税的基本政策是为了() 我国出口货物在国际市场上的竞争力。
(A) 增强 (B) 削弱 (C) 保持 (D) 减少
- 580 . 根据出口退税政策, 出口货物的退税税率是由() 确定的。
(A) 出口企业 (B) 海关 (C) 国家税务部门 (D) 商务部
- 581 . 生产企业在提交出口退税申报材料后, 税务局的初步审核期限通常为() 个工作日。
() 5 () 10 () 15 () 20
- 582 . 出口退税操作流程中, 企业首先需要完成的是() 环节。
(A) 出口货物报关 (B) 提交退税申请 (C) 审核退税资料 (D) 收到退税款

- 583 . 出口退税的计算主要基于()。
- (A) 出口货物的到岸价格 (B) 出口货物的离岸价格 (C) 进口货物的成本 (D) 国内市场销售价格
- 584 . 出口退税的计算方法中, 关键要素之一是出口货物的()。
- (A) 离岸价格 (B) 到岸价格 (C) 生产成本 (D) 销售价格
- 585 . 当企业选择在线申报系统进行出口退税申报时, 通常需要登录的是()平台。
- (A) 海关总署官网 (B) 商务部网站 (C) 国家税务总局网站 (D) 外汇管理局网站
- 586 . ()方式不属于出口退税的申报方式。
- (A) 纸质申报 (B) 在线申报 (C) 代理申报 (D) 电话申报
- 587 . 出口退税风险防范中, 企业需特别注意退税政策的()变化, 及时调整经营策略。
- (A) 长期 (B) 短期 (C) 国际 (D) 国内
- 588 . 在出口退税风险防范中, 企业应加强与()的沟通, 及时了解退税政策动态。
- (A) 税务机关 (B) 海关 (C) 银行 (D) 外汇管理局
- 589 . 商务数据分析报告撰写时, 首要考虑的是报告的()以确保读者能够明确理解分析目的。
- (A) 清晰性 (B) 详细性 (C) 创新性 (D) 艺术性
- 590 . 商务数据分析报告应包含足够的()以支持分析结论的可靠性。
- (A) 个人观点 (B) 数据证据 (C) 市场传闻 (D) 行业常识
- 591 . ()是提升商务数据分析报告可读性的有效技巧。
- (A) 使用大量文字描述 (B) 简洁明了的语言 (C) 复杂的图表设计 (D) 忽视数据来源
- 592 . 撰写商务数据分析报告时, 逻辑结构的重要性在于它能帮助读者()。
- (A) 快速找到感兴趣的部分 (B) 欣赏报告的文采 (C) 了解数据分析的技术细节 (D) 跟随分析思路, 理解结论
- 593 . 在缮制商务数据分析报告时, 未对数据进行充分清洗和校验, 可能导致报告中的()问题。
- (A) 数据准确性不足 (B) 报告格式不美观 (C) 结论过于乐观 (D) 分析方法不先进
- 594 . 报告中未对关键数据点进行突出展示, 可能会让读者()。
- (A) 对分析方法产生质疑 (B) 对报告格式不满意 (C) 对结论持保留意见 (D) 难以抓住重点信息
- 595 . 商务数据分析报告框架中, 通常位于报告开头部分的是()。
- (A) 摘要或概述 (B) 数据分析方法 (C) 详细数据展示 (D) 结论与建议
- 596 . 报告框架中的“数据收集与处理”部分通常紧随()之后。
- (A) 背景介绍 (B) 数据分析方法 (C) 数据结果分析 (D) 结论与建议
- 597 . 设计商务数据分析报告框架时应遵循的首要原则是()。
- (A) 美观性 (B) 图标丰富 (C) 逻辑清晰 (D) 报告简短
- 598 . 在设计商务数据分析报告框架时, 首要考虑的原则是(), 以确保报告能够满足读者的需求。
- (A) 针对性 (B) 创新性 (C) 艺术性 (D) 详细性
- 599 . 报告框架设计时, 为了增强可读性, ()部分应突出显示关键数据和洞察。
- (A) 引言 (B) 方法论 (C) 附录 (D) 关键发现与分析
- 600 . 商务数据分析报告框架设计时, 确保报告内容()是提高报告可读性的关键。

(A) 详细冗长 (B) 条理清晰 (C) 视觉华丽 (D) 语言深奥

三、多项选择题(选择至少有两个正确的答案,将相应的字母填入题内的括号中)

- 1 . 设计数据监测指标时,应遵循()。
(A) SMART 原则 (B) 反映业务目标原则 (C) 数据易于获取和处理原则 (D) 指标之间应相互独立原则 (E) 指标设计应追求复杂和全面原则
- 2 . 数据异常监测与预警系统通常由()组成。
(A) 明确的数据源 (B) 精确的异常定义规则 (C) 强大的实时数据处理引擎 (D) 有效的预警通知渠道 (E) 全面的数据恢复计划
- 3 . ()是实时数据监测与处理系统的关键技术支撑。
(A) 流式处理平台 (B) 分布式存储架构 (C) 实时数据分析引擎 (D) 离线批处理框架 (E) 低延迟网络传输技术
- 4 . 数据迁移过程涵盖()。
(A) 数据源与目标环境的详细分析 (B) 详尽的迁移计划与策略制定 (C) 数据的有效清洗与转换 (D) 迁移过程的严格监控与日志记录 (E) 数据安全与隐私保护措施
- 5 . 数据迁移的基本类型包括()。
(A) 本地到云迁移 (B) 云到云迁移 (C) 跨数据库系统迁移 (D) 离线批量迁移 (E) 实时流式迁移
- 6 . 设计数据迁移策略时,需要考虑()。
(A) 迁移的时间窗口 (B) 数据的敏感性与合规性要求 (C) 业务连续性影响 (D) 迁移后的验证与测试 (E) 迁移工具的成本与效益
- 7 . 使用数据迁移工具时,应该()。
(A) 确保工具与目标系统的兼容性 (B) 详细了解工具的迁移能力与限制 (C) 进行充分的测试与验证 (D) 制定应对迁移过程中可能出现问题的应急预案 (E) 忽略工具的技术支持与更新情况
- 8 . 保障数据迁移中数据一致性的方法包括()。
(A) 设计并执行详细的数据验证计划 (B) 使用事务日志来记录迁移过程中的变化 (C) 实施数据一致性检查点 (D) 迁移前后进行数据的完整性比对 (E) 忽略小量的数据不一致
- 9 . 优化数据迁移性能的策略包括()。
(A) 评估并优化网络性能 (B) 使用高效的数据压缩和解压算法 (C) 合理安排迁移时间窗口 (D) 对迁移过程进行实时监控与调整 (E) 忽视迁移过程中的错误与警告
- 10 . 保障数据安全与合规的注意事项包括()。
(A) 评估并遵守相关法律法规 (B) 对敏感数据进行加密处理 (C) 实施数据访问权限控制 (D) 在迁移前后进行数据完整性验证 (E) 忽略数据迁移过程中的安全漏洞
- 11 . 现代数据存储类型包括()。
(A) 直接附加存储 (B) 存储区域网络 (C) 网络附加存储 (D) 对象存储 (E) 云端存储服务
- 12 . ()有效的数据备份与恢复策略。
(A) 定期全备份 (B) 增量或差异备份 (C) 异地备份 (D) 备份数据加密 (E) 恢复计划的定期测试
- 13 . ()可以实现数据冗余与去重。
(A) 镜像备份 (B) RAID 技术 (C) 数据去重软件 (D) 数据分片 (E) 数据压缩算法
- 14 . 数据生命周期管理策略包括()。
(A) 数据分类与策略制定 (B) 数据存储位置的选择 (C) 数据访问权限的控制 (D) 数

- 据的定期评估与审查 (E)数据的合规性处理
- 15 . 在数据安全与加密使用中, 需要注意()。
(A)选择合适的加密算法以满足安全需求 (B)确保加密密钥的安全存储与分发 (C)定期更换加密密钥以防止潜在的安全威胁 (D)对敏感数据进行加密处理, 无论是在存储还是传输过程中 (E)忽视加密带来的性能开销, 因为它不是关键考虑因素
- 16 . 使用数据压缩与解压缩时, 需要注意()。
(A)评估压缩比与解压缩时间之间的平衡 (B)选择适合数据类型的压缩算法 (C)确保压缩过程不会破坏数据的完整性 (D)压缩和解压缩过程应易于自动化和集成 (E)忽略压缩对系统性能可能产生的影响
- 17 . 分布式存储技术的特点包括()。
(A)数据冗余和容错机制 (B)高可用性和数据持久性 (C)可扩展性和灵活性 (D)集中管理和统一访问接口 (E)所有数据始终存储在单一节点上
- 18 . 数据访问控制技术的特点包括()。
(A)细粒度的权限控制 (B)支持动态权限变更 (C)确保数据的机密性和完整性 (D)总是允许所有用户访问所有数据 (E)灵活适应不同的安全需求
- 19 . 数据存储的目标包括()。
(A)数据的长期保存 (B)提供高效的数据检索和访问 (C)确保数据的完整性和一致性 (D)保护数据的机密性和隐私 (E)最小化存储成本和优化资源利用
- 20 . 数据存储的原则包括()。
(A)数据的冗余和备份, 以防止数据丢失 (B)数据的可访问性和易用性 (C)数据的安全性, 包括加密和访问控制 (D)可扩展性, 以应对未来数据量的增长 (E)数据的质量保证, 包括数据完整性和准确性
- 21 . 数据存储流程包括()。
(A)数据采集 (B)数据验证 (C)数据存储介质选择 (D)数据备份与恢复 (E)数据销毁
- 22 . 评估数据存储系统性能时, 应考虑()。
(A)读写速度 (B)并发处理能力 (C)存储空间成本 (D)数据恢复时间 (E)安全性与合规性
- 23 . 数据清洗需求分析的定义中, 包含()。
(A)评估当前数据质量状况 (B)设定清洗后数据应满足的业务需求 (C)确定数据清洗的优先级和顺序 (D)立即实施数据清洗操作 (E)分析数据清洗对业务决策的潜在影响
- 24 . 在数据质量问题识别过程中, 可以采用()。
(A)统计分析 (B)数据可视化 (C)逻辑检查 (D)聚类分析 (E)数据分布分析
- 25 . 在评估清洗数据价值时, 应考虑()。
(A)数据质量改进的直接效果 (B)对业务决策的潜在影响 (C)数据清洗的成本效益 (D)清洗后数据的市场竞争力 (E)清洗工具的技术规格
- 26 . 在设定数据清洗预期目标时, 需要考虑()。
(A)业务对数据质量的期望 (B)当前数据质量的问题点 (C)清洗后数据应满足的具体标准 (D)清洗工作的优先级和资源分配 (E)清洗工具的易用性
- 27 . 在关联数据清洗与业务指标时, 应考虑()。
(A)识别哪些业务指标受数据质量直接影响 (B)分析数据清洗如何改善这些指标 (C)设定清洗后业务指标的具体提升目标 (D)监测数据清洗实施后业务指标的变化 (E)仅关注数据清洗的技术实现, 忽略业务指标
- 28 . 数据源与数据结构分析时, 需要考虑()。

- (A)数据源的可靠性和更新频率 (B)数据结构是否满足业务需求 (C)数据格式的一致性和兼容性 (D)数据的安全性和隐私保护 (E)数据存储的成本和可扩展性
- 29 . 使用数据清洗工具时, 需要注意()。
- (A)备份原始数据以防万一 (B)详细了解工具的清洗规则和算法 (C)验证清洗后的数据质量 (D)忽视工具的性能对清洗效率的影响 (E)考虑工具的长期维护和支持
- 30 . 分析数据完整性时, 可以采取()的策略。
- (A)设定非空值规则 (B)验证数据字段的长度和格式 (C)检查数据记录之间的依赖关系 (D)使用统计方法识别异常值 (E)比较数据与其他可靠来源的一致性
- 31 . 在数据准确性分析中, 应该()。
- (A)设定数据范围和阈值 (B)使用业务规则进行校验 (C)依赖专家进行人工审核 (D)比较历史数据与当前数据的变化 (E)依赖数据清洗工具自动完成
- 32 . 在数据一致性分析中, 需要考虑()。
- (A)不同数据源间数据同步的实时性 (B)数据字段在不同系统中的命名一致性 (C)数据记录在不同系统间的唯一性 (D)数据类型在不同系统中的兼容性 (E)数据可视化工具的准确性
- 33 . 识别数据缺失值时, 可以考虑()。
- (A)使用统计函数检查每个字段的缺失值数量 (B)设定数据完整性规则, 如某些字段不允许为空 (C)编写代码或脚本来自动检测并报告缺失值 (D)依靠数据可视化工具来直观展示缺失值分布情况 (E)忽略缺失值, 因为它们不影响数据分析结果
- 34 . 识别数据异常值时, 可以采用()。
- (A)箱线图 (B)Z分数或标准分数方法 (C)聚类分析 (D)散点图或直方图进行视觉检查 (E)忽略异常值
- 35 . 数据清洗方案应包含()。
- (A)数据质量问题的定义和识别 (B)清洗策略和技术的选择 (C)清洗后的数据验证和质量控制 (D)清洗过程的时间表和里程碑 (E)数据存储和备份策略
- 36 . 数据清洗需求分析涉及()。
- (A)数据质量的现状评估 (B)清洗后数据的预期质量标准 (C)利益相关者的需求和期望 (D)清洗过程的技术和资源要求 (E)数据的法律合规性要求
- 37 . 选择数据清洗工具时, 应该依据()这些因素。
- (A)工具的兼容性 (B)工具的性能 (C)工具的灵活性 (D)工具的社区支持和文档质量 (E)工具的价格和成本效益
- 38 . 设定数据清洗规则时, 需要考虑()。
- (A)数据的业务含义和上下文 (B)数据的统计特性和分布 (C)清洗规则的优先级和顺序 (D)清洗规则的灵活性和可扩展性 (E)清洗规则对后续数据分析的影响
- 39 . 设计数据清洗流程时, 需要遵循()这些原则。
- (A)流程应清晰、可理解, 便于团队沟通和协作 (B)流程应具有可扩展性, 以适应未来可能的数据质量问题 (C)流程应支持数据的增量清洗, 以提高处理效率 (D)流程的每个步骤都应有明确的输入和输出, 确保数据的可追溯性 (E)流程的设计应完全依赖于特定的清洗工具或技术
- 40 . 数据清洗中实施质量控制时, 可以采用()这些策略。
- (A)制定详细的质量控制计划 (B)设定数据质量标准和指标 (C)使用自动化工具进行实时数据监控 (D)定期对清洗后的数据进行抽样检查 (E)依赖数据拥有者的直觉和经验
- 41 . 评估数据清洗方案时, 应考虑()。

- (A)数据清洗的效率和准确性 (B)清洗后数据对业务目标的贡献 (C)清洗过程中可能引入的风险 (D)清洗方案的成本效益分析 (E)清洗工具的技术先进性和市场声誉
- 42 . 处理数据清洗中的异常值时,可以采取()。
(A)直接删除异常值 (B)使用统计方法修正异常值 (C)将异常值视为缺失值处理 (D)保留异常值,但在后续分析中单独考虑 (E)忽略异常值
- 43 . 数据治理中涉及数据清洗的方面包括()。
(A)制定数据清洗的标准和流程 (B)监控数据清洗的执行情况 (C)评估数据清洗的效果和影响 (D)培训数据清洗团队的技术能力 (E)制定数据使用政策以确保数据合规性
- 44 . 在数据清洗中实施版本控制与回溯时,需要()。
(A)使用版本控制软件来管理数据清洗脚本和结果的变更历史 (B)定期备份数据清洗的中间结果和最终输出 (C)为数据清洗的每个阶段创建明确的文档说明 (D)设置自动化的数据清洗任务,但不进行版本控制 (E)实施严格的访问控制,确保只有授权人员可以修改数据清洗过程
- 45 . 数据归一化处理的目的包括()。
(A)加快模型的收敛速度 (B)提高模型的预测准确性 (C)消除不同量纲的影响 (D)减少计算资源的消耗 (E)改变数据的分布形态
- 46 . 数据合并处理前,应考虑()以确保合并的有效性。
(A)数据集的时间范围是否一致 (B)数据集是否包含重复的记录 (C)数据集的字段类型和含义是否相同 (D)数据集的存储格式是否兼容 (E)数据集的收集方法是否相同
- 47 . ()对于提高查询效率至关重要。
(A) 查询优化器 (B) 索引 (C) 缓存 (D) 分布式处理 (E) 实时数据更新
- 48 . SQL 在 BI 查询引擎中的优势可能包括()。
(A)高度标准化的查询语言 (B)支持复杂的数据分析和报表生成 (C)易于与其他 BI 工具集成 (D)自动优化查询性能 (E)仅适用于小型数据集
- 49 . MDX 作为多维数据查询引擎的优势包括()。
(A)支持复杂的切片和切块操作 (B)直观理解业务数据 (C)高效的聚合计算 (D)主要用于关系型数据库 (E)良好的可视化支持
- 50 . DAX 在数据模型查询中的特点包括()。
(A)支持复杂的计算表达式 (B)可用于创建度量值和计算列 (C)主要面向实时数据处理 (D)易于与其他 BI 工具集成 (E)专门用于大型分布式系统
- 51 . Spark SQL 在大数据处理中的优势包括()。
(A)支持 SQL 语法,降低学习曲线 (B)高度可扩展和容错 (C)基于内存的数据处理,提高速度 (D)强大的数据聚合和排序能力 (E)仅适用于小型数据集
- 52 . Presto 在大数据环境下的优点包括()。
(A)分布式架构,支持大规模并行处理 (B)优化的 SQL 查询引擎,低延迟响应 (C)支持多种数据源,包括 Hadoop、关系数据库等 (D)专为实时分析设计 (E)内置机器学习库,用于预测分析
- 53 . FineReport 和 FineVis 作为 BI 工具,其优点包括()。
(A)易于使用的拖放界面 (B)支持多种查询引擎,如 SQL、MDX 等 (C)丰富的图表和可视化组件 (D)强大的数据报表生成能力 (E)内置 AI 助手,自动分析数据
- 54 . 在选择 BI 工具的查询引擎时,应考虑()。

- (A)性能和响应时间 (B)支持的数据源和格式 (C)易用性和学习曲线 (D)可扩展性和弹性 (E)安全性和合规性
- 55 . 查询引擎优化和性能调优的方法包括()。
- (A)索引和分区策略优化 (B)SQL 语句重写以提高性能 (C)缓存频繁访问的数据 (D)分布式并行处理 (E)频繁进行数据备份和恢复
- 56 . 查询引擎的安全性和数据保护方法包括()。
- (A)访问控制和角色基安全模型 (B)加密存储和传输中的数据 (C)SQL 注入防护和参数化查询 (D)审计和日志记录功能 (E)禁用所有外部网络访问
- 57 . 查询性能评估时,通常会关注()。
- (A)响应时间 (B)吞吐量 (C)CPU和内存资源消耗 (D)磁盘 I/O 操作次数 (E)网络延迟
- 58 . 数据库查询优化涉及()方面的技术或策略。
- (A)索引优化 (B)SQL 语句重写 (C)查询缓存利用 (D)数据库架构调整 (E)减少网络延迟
- 59 . 数据分词在()有重要应用。
- (A)自然语言处理 (B)搜索引擎优化 (C)文本挖掘与分析 (D)机器学习中的文本预处理 (E)数据库管理
- 60 . 在使用分词工具或库时,需要注意()。
- (A)准确率与召回率的平衡 (B)处理速度和效率 (C)对特殊词汇和领域的支持 (D)工具的易用性和文档完善程度 (E)工具的开源协议和许可
- 61 . 基于字符串匹配的分词方法包括()。
- (A)最大正向匹配法(MM) (B)最大逆向匹配法(RMM) (C)双向最大匹配法(Bi-MM) (D)最小匹配法 (E)基于统计的分词方法
- 62 . 基于统计的分词方法涉及()技术或概念。
- (A)词频统计 (B)互信息 (C)熵值计算 (D)神经网络模型 (E)语义分析
- 63 . 基于理解的分词方法包括()。
- (A)句法分析 (B)语义角色标注 (C)实体识别 (D)字符串匹配 (E)深度学习模型应用
- 64 . 基于深度学习的分词方法相比传统方法优势在于()。
- (A)无需手动制定大量规则 (B)能够自动学习复杂的语言特征 (C)分词效果受预定义词典限制较小 (D)计算复杂度低 (E)适用于多种语言和领域
- 65 . 关于自定义词典与规则,()是正确的。
- (A)自定义词典可以扩展分词系统的词汇覆盖范围。(B)自定义规则总是能够比统计方法更准确地处理所有分词任务。(C)自定义词典和规则通常需要领域专家根据具体需求进行定制。(D)自定义词典和规则一旦确定,就不需要再进行调整或更新。(E)自定义词典和规则可以显著提高分词系统在某些特定领域的性能。
- 66 . 分词粒度控制涉及()。
- (A)词汇的细分程度 (B)分词结果中词汇的数量 (C)词汇的合并策略 (D)分词算法的选择 (E)分词结果的语义准确性
- 67 . 评估分词效果时,需要考虑()。
- (A)分词结果的准确率 (B)分词结果的召回率 (C)分词速度 (D)分词结果的语义一致性 (E)分词系统对未登录词的处理能力
- 68 . 在修改数据库时,需要注意()。
- (A)先备份重要的数据 (B)谨慎操作,避免失误 (C)做好权限控制 (D)检查依

- 赖关系 (E) 打开所有连接
- 69 . truncate 语句用来删除数据, 它和 delete 的不同点在于()。
(A) 只能对整表操作, 不能针对部分数据操作 (B) 在删除数据时, 不经过真正事物, 无法回滚 (C) 会重置 AUTO_INCREMENT 项 (D) 只要你不写 where 语句, 就可以更新全表 (E) 可以搭配 order by 来显示倒数两行的数据
- 70 . 数据库查询语句中的逻辑运算符包括()。
(A) and (B) where (C) or (D) in (E) not
- 71 . 数据库查询语句中的谓词包括()。
(A) between (B) in (C) like (D) null (E) exists
- 72 . 简单的聚合函数包括()。
(A) count (B) sum (C) avg (D) max (E) min
- 73 . 内连接是根据指定的连接条件进行连接查询, 包括()。
(A) 相等内连接 (B) 非等内连接 (C) 自连接 (D) 长连接 (E) 短连接
- 74 . 结构化数据的数据形式包括()。
(A) 电子表格中的数据 (B) JSON 格式的数据 (C) 数据库中的表格数据 (D) 自由形式的文本数据 (E) 图形中的数据
- 75 . 结构化数据的特点包括()。
(A) 格式固定 (B) 规则明确 (C) 数据相同 (D) 应用广泛 (E) 易于管理
- 76 . 非结构化数据包括()。
(A) 文本 (B) 图形 (C) 图像 (D) 音频 (E) 视频
- 77 . 非结构化数据则没有固定的结构和模式, 处理起来更为复杂, 通常需要()等过程。
(A) 净化 (B) 挖掘 (C) 提取 (D) 分析 (E) 存储
- 78 . 结构化数据使用的应用场景包括()。
(A) 电子商务网站的用户交易记录 (B) 企业的财务报表 (C) 社交媒体的用户生成内容 (D) 科学研究实验数据 (E) 图文工具制作的图片
- 79 . 在数字化转型过程中, 文档数据种类多样且持续积累, 传统的文档存储和管理模式已不再适应当前需求, 通过非结构化数据中台, 可以实现文档的()。
(A) 统一管理 (B) 安全合规 (C) 数字资产沉淀 (D) 集中分析 (E) 高效挖掘
- 80 . 关于层次结构的概念, 正确的说法包括()。
(A) 层次结构是一种将复杂系统或信息组织成多个层级或级别的表示方法 (B) 在层次结构中, 每一层都仅与其上一层或下一层有直接联系, 与其他层无直接联系 (C) 层次结构有助于简化和理解复杂系统或信息, 通过分解和归类来实现 (D) 层次结构中的每一层都包含相同数量和类型的信息 (E) 层次结构广泛应用于计算机科学、生物学、管理学等多个领域
- 81 . 层次模型的特征包括()。
(A) 层次结构清晰 (B) 单一父节点 (C) 多个子节点 (D) 数据的一致性 (E) 灵活性和扩展性
- 82 . 层次建模的优势包括()。
(A) 层次数据模型的结构简单、清晰、明朗 (B) 操作层次数据类型的数据库语句比较简单 (C) 查询效率较高 (D) 层次数据模型提供了较好的数据完整性支持 (E) 查询节点的时候必须知道其双亲节点的
- 83 . 层次建模的核心是将问题分解为不同的层次, 每个层次包含不同的元素, 通常包括()三个层次。

- (A) 最高层 (B) 中上层 (C) 中间层 (D) 中下层 (E) 最底层
- 84 . 运用层次分析法建模, 其步骤包括()。
(A) 建立递阶层次结构模型 (B) 构造出各层次中的所有判断矩阵 (C) 一致性检验 (D) 求权重后进行评价 (E) 删除多余的节点
- 85 . 层次数据模型的存储结构包括()。
(A) 实接法 (B) 短接法 (C) 虚接法 (D) 邻接法 (E) 链接法
- 86 . 数仓是一个()的数据集合, 随着大数据技术的发展, 其作用不再局限于决策分析、还可以为业务应用、审计、追踪溯源等多方面提供数据支撑, 帮助企业完成数字化转型。
(A) 面向主题 (B) 集成的 (C) 相对稳定 (D) 反应历史变化 (E) 数据离散
- 87 . 维度建模的特点包括()。
()可以没有属性, 只包含实体和关系的描述()需要对实体的属性进行补全
()不要求必须满足 3NF()包含的信息与 3NF 相同()以一种用户可理解的, 利于查询、灵活多变的方式对 3NF 进行包装
- 88 . 维度建模的优点包括()。
(A) 后期数据产品性能好 (B) 切换使用不同的维度查询很方便 (C) 可扩展性好
(D) 数据预处理开销和数据冗余 (E) 数据分层处理, 可能会导致数据处理丢失或者异常
- 89 . 事实表的类型包括()。
(A) 事务事实表 (B) 周期快照事实表 (C) 累积快照事实表 (D) 数据存储事实表
(E) 信息快照事实表
- 90 . 维度建模的四个步骤包括()。
(A) 选择业务过程 (B) 设计调查问卷 (C) 声明粒度 (D) 确认维度 (E) 确认事实
- 91 . 维度建模的存储结构包括()。
(A) 事实表 (B) 维度表 (C) 树状表 (D) 结构表 (E) 虚拟表
- 92 . 数据存储的根本原理是利用可区分的物理状态来表示和识别数字或字符信息, 以便于信息的()。
(A) 调研 (B) 生成 (C) 保持 (D) 检索 (E) 处理
- 93 . 数据的储存方式包括()。
(A) NDS (B) SIM (C) DAS (D) NAS (E) SAN
- 94 . 数据读写性能的主要指标包括()。
(A) IOPS (B) 吞吐率 (C) 延迟 (D) 顺序读写速度 (E) 随机读写速度
- 95 . 排序算法的种类包括()等。
(A) 冒泡排序 (B) 归并排序 (C) 基数排序 (D) 快速排序 (E) 希尔排序
- 96 . 数组的缺点包括()。
(A) 大小固定 (B) 插入和删除效率低 (C) 不支持动态关联数据 (D) 浪费空间 (E) 内存效率高
- 97 . 决策树算法在实际应用中可能面临的问题包括()。
(A) 过拟合 (B) 欠拟合 (C) 无法处理非线性关系 (D) 对异常值敏感 (E) 缺失值较多
- 98 . 在使用 SVM 进行分类时, ()可以提高模型的泛化能力。
()增加更多的特征()使用交叉验证选择最佳参数()应用正则化技术, 如 L1 或 L2 惩罚()使用更大的数据集进行训练()选择不敏感的参数

- 99 . python 内置的数据结构包括()。
(A) 列表 (B) 元组 (C) 字典 (D) 集合 (E) 数组
- 100 . ()可以在最坏情况下可以达到O()。
(A) 冒泡排序 (B) 快速排序 (C) 堆排序 (D) 直接插入排序 (E) 归并排序
- 101 . 数据画像可以帮助企业实现()目标。
(A) 个性化营销 (B) 改善客户服务 (C) 开发新产品 (D) 优化运营流程 (E) 降低生产成本
- 102 . 用户数据画像的构建流程主要包括数据收集、数据清洗、数据分析和数据可视化()。
(A) 用户数据收集 (B) 用户数据清洗 (C) 用户数据分析 (D) 用户数据可视化 (E) 用户数据上传
- 103 . 在构建标签体系时, 需要考虑的因素包括()。
(A) 业务目标 (B) 数据可用性 (C) 技术限制 (D) 用户隐私 (E) 数据价格
- 104 . 特征处理作为特征工程的核心组成部分, 包含了()等多个环节。
(A) 数据预处理 (B) 特征提取 (C) 特征构造 (D) 特征选择 (E) 降维
- 105 . 特征工程的作用包括()。
(A) 提高模型性能 (B) 减少模型训练时间 (C) 增加模型可解释性 (D) 防止过拟合 (E) 提高模型的泛化能力
- 106 . 具体特征选择方法包括()。
(A) 相关性系数法 (B) 信息增益法 (C) 互信息法 (D) 特征向量法 (E) 信息相切法
- 107 . 特征提取在多个领域有广泛应用, 包括()。
(A) 数字信号处理 (B) 信息成本降低 (C) 信息来源多样化 (D) 智能风控 (E) 机器学习
- 108 . 在处理缺失值之前, 应该考虑()等因素。
(A) 缺失值的比例 (B) 缺失值的类型 (C) 缺失值对分析结果的影响 (D) 缺失值的来源 (E) 缺失值的位置
- 109 . 在特征构造过程中, ()有助于提高模型的性能。
(A) 特征选择 (B) 特征缩放 (C) 特征组合 (D) 特征转换 (E) 特征放大
- 110 . Pandas 库的主要功能包括()。
(A) 数据清洗和准备 (B) 数据分析和探索 (C) 机器学习模型训练 (D) 数据可视化 (E) 数据生成和处理
- 111 . scikit-learn 中常用的监督学习算法有()。
(A) 线性回归 (B) 逻辑回归 (C) K 最近邻 (D) 随机森林 (E) 矩阵运算
- 112 . python 数据处理和分析库包括()。
(A) pandas (B) NumPy (C) scipy (D) Statsmodels (E) SQLAlchemy
- 113 . 概要统计通过计算诸如()等统计量, 来揭示数据的内在规律和特征。
(A) 平均数 (B) 中位数 (C) 众数 (D) 方差 (E) 标准差
- 114 . 概要统计的目的包括()。
(A) 发现数据的特点与规律 (B) 分析总体结构 (C) 提供决策依据 (D) 明确调查的目的 (E) 确定调查对象和调查表
- 115 . 描述性统计分析的指标体系主要包括()。
(A) 集中趋势分析 (B) 离散程度分析 (C) 分布分析 (D) 频数分析 (E) 标准差分析

- 116 . 统计工具是指用于()数据的一系列方法和技术。
(A) 收集 (B) 整理 (C) 分析 (D) 解释 (E) 处理
- 117 . 数据分析工具包括()。
(A) spss (B) stata (C) R 语言 (D) Tableau (E) Excel
- 118 . 在使用 Excel 进行数据分析时, 应注意()。
(A) 确保数据的准确性和完整性 (B) 正确选择和使用统计函数 (C) 忽略数据的可视化展示 (D) 定期保存工作簿以防数据丢失 (E) 使用适当的统计方法
- 119 . 在使用 Python 进行数据分析时, 应注意()。
(A) 确保数据的准确性和完整性 (B) 正确选择和使用统计函数 (C) 忽略数据的可视化展示 (D) 定期保存工作以防数据丢失 (E) 使用适当的统计方法
- 120 . SPSS 软件的基本使用方式包括()。
(A) 窗口菜单方式 (B) 程序运行方式 (C) 混合运行方式 (D) 函数运行方式 (E) 网络端口运行方式
- 121 . Power BI 报表分享的方式有()。
(A) 发布到 Power BI Service (B) 导出为 PDF 或 PPT 文件 (C) 嵌入网页或应用 (D) 通过邮件发送链接 (E) 通过网盘传输
- 122 . 在 Stata 中处理缺失值时, ()是可行的。
(A) 使用 drop if missing() 删除缺失值 (B) 使用 mvdecode 命令将缺失值转换为特定值 (C) 使用 mi 系列命令进行多重插补 (D) 忽略缺失值进行分析 (E) 手动补充缺失值
- 123 . 在 SAS 程序编写时, 变量和数据集的名字都不能超过 32 个字符, 且名称中只允许出现()。
(A) 空格 (B) 破折号 (C) 下划线 (D) 字母 (E) 数字
- 124 . 使用 MATLAB 基本统计指标进行计算数据, 正确的是()。
(A) 使用 mean() 函数计算数据的平均值 (B) 使用 median() 函数计算数据的中位数 (C) 使用 std() 函数计算数据的标准差 (D) 使用 var() 函数计算数据的方差 (E) 使用 drop() 函数删除数据
- 125 . 在线性相关分析中, ()等因素可能会影响相关系数的大小。
(A) 样本量的大小 (B) 数据的分布形状 (C) 异常值的存在 (D) 变量的单位 (E) 不同的相关系数计算方法
- 126 . 一元线性回归模型的优点包括()。
(A) 简单易懂 (B) 计算速度快 (C) 适用于非线性关系 (D) 容易解释 (E) 不能捕捉复杂的非线性模式
- 127 . 多元线性回归模型的局限性包括()。
(A) 假设自变量和因变量之间存在线性关系 (B) 不能捕捉复杂的非线性模式 (C) 对异常值敏感 (D) 可以处理多个自变量 (E) 提供有关每个自变量对因变量影响的量化信息
- 128 . 数据可视化在()具有广泛的应用。
(A) 商业分析 (B) 科学研究 (C) 教育培训 (D) 社交媒体 (E) 医疗健康
- 129 . 视觉编码原则中, ()是设计有效可视化时需要考虑的。
(A) 颜色 (B) 大小 (C) 形状 (D) 纹理 (E) 位置
- 130 . 数据可视化设计时应遵循()的基本原则。
(A) 简洁性 (B) 准确性 (C) 层次性 (D) 互动性 (E) 美观性
- 131 . ()是可视化不同类型数据的常用图表类型。

- (A) 条形图 (B) 折线图 (C) 散点图 (D) 树状图 (E) 热力图
- 132 . 使用数据可视化工具时, () 是需要特别注意的方面。
(A) 数据源的选择与验证 (B) 图表类型的适当选择 (C) 颜色的合理运用与对比 (D) 图表的交互性与动态性 (E) 图表的美观度与装饰性
- 133 . 在设计数据可视化时, 选择色彩时需要考虑()。
(A) 数据的性质 (B) 目标受众的视觉偏好 (C) 色彩的心理效应 (D) 色彩的可用性和可访问性 (E) 图表的复杂程度
- 134 . 图表布局与排版时, 需要考虑()。
(A) 图表的整体平衡 (B) 元素间的间距 (C) 文字的字体和大小 (D) 颜色的协调 (E) 数据的准确性
- 135 . 在设计具有交互性的数据可视化时, 需要考虑()。
(A) 用户的操作习惯 (B) 交互的响应速度 (C) 数据的安全性和隐私 (D) 交互的直观性和易用性 (E) 交互的视觉反馈
- 136 . 评估数据可视化质量时, () 需要考虑。
(A) 数据的准确性和可靠性 (B) 视觉的清晰度和吸引力 (C) 信息的完整性和深度 (D) 图表的交互性和易用性 (E) 设计的原创性和创新性
- 137 . 在数据可视化中, 视觉通道的选择和应用需要考虑()。
(A) 视觉通道的信息传达效率 (B) 观众对视觉通道的感知能力 (C) 数据类型和属性的匹配度 (D) 视觉通道的可用性和可访问性 (E) 图表设计的整体风格和美观性
- 138 . 在选择图表类型以匹配数据时, 需要综合考虑()。
(A) 数据的性质 (B) 需要传达的信息类型 (C) 观众的背景和偏好 (D) 图表的复杂性和易读性 (E) 数据的可用性和可访问性
- 139 . 在数据可视化设计中, () 有助于确保准确性原则。
(A) 使用官方或可靠的数据源 (B) 在可视化中明确标注数据的来源和采集方法 (C) 对数据进行清洗和预处理, 去除错误和异常值 (D) 在可视化中使用明确的比例尺和单位 (E) 在必要时提供数据的原始值和计算过程
- 140 . 可视化组件库可以带来() 这些好处。
(A) 加快开发速度 (B) 保持设计一致性 (C) 减少编码错误 (D) 提高网站或应用的安全性 (E) 便于组件的重用和维护
- 141 . 组件库通常包含()。
(A) 原子组件 (如按钮、图标) (B) 分子组件 (由原子组件组合而成) (C) 组织组件 (由分子组件进一步组合) (D) 页面模板 (完整页面的布局和设计) (E) 开发文档和 API 参考
- 142 . 在使用基础图表组件时, 需要考虑()。
(A) 数据量大小对图表性能的影响 (B) 图表的可访问性 (如屏幕阅读器支持) (C) 图表在不同设备和浏览器上的兼容性 (D) 图表的动画效果是否过于复杂 (E) 图表颜色对观众情绪的影响
- 143 . 高级分析组件的使用涉及()。
(A) 数据聚合与汇总 (B) 预测分析与机器学习 (C) 实时数据流处理 (D) 复杂的图表联动与交互 (E) 简单的数据排序和过滤
- 144 . 在设计和交互组件时, () 是需要考虑的。
(A) 用户的需求和行为模式 (B) 组件的可用性和易用性 (C) 组件在不同设备和浏览器上的兼容性 (D) 组件的反馈机制 (如加载状态、错误提示) (E) 组件的动画效果是否分散用户注意力

- 145 . 在设置组件属性与参数时，应考虑()。
- (A)属性的必要性和默认值 (B)属性的类型和可接受的值范围 (C)属性间的相互依赖和约束关系 (D)属性对组件性能的影响 (E)属性是否支持响应式设计
- 146 . 在集成和使用组件库时，()是有益的。
- (A)仔细阅读并遵循组件库的集成指南 (B)在项目中统一命名规范和代码风格 (C)定期检查并更新组件库到最新版本 (D)避免过度依赖组件库，保持项目的独立性和可迁移性 (E)忽略组件库的文档和示例，直接开始使用
- 147 . 优化组件性能的方法包括()。
- (A)懒加载和按需加载 (B)减少 DOM 操作和重绘/重排 (C)使用 Web Workers 进行非 UI 线程的计算 (D)优化组件的算法和数据结构 (E)增加组件的视觉复杂度以提高吸引力
- 148 . 为了确保组件的广泛兼容性和良好适配，()是有效的。
- (A)使用 CSS 媒体查询来根据屏幕尺寸和分辨率调整样式 (B)编写适应不同屏幕尺寸和分辨率的响应式布局 (C)使用现代浏览器特性时，提供 polyfills 或降级方案以支持旧浏览器 (D)进行彻底的跨浏览器测试，包括移动设备 (E)忽略少数非主流浏览器和设备，专注于主流平台的优化
- 149 . 在制定业务指标规则时，应考虑()。
- (A)指标的可衡量性 (B)指标与业务目标的关联性 (C)数据的可获得性和准确性 (D)指标的时效性和动态性 (E)员工的个人偏好
- 150 . 在运用对比能力分析方法时，()是关键。
- (A)明确对比目标和维度 (B)收集相关数据并进行标准化处理 (C)选择合适的对比图表展示结果 (D)分析差异原因并提出改进建议 (E)忽略异常值以确保结果的准确性

第4部分

操作技能复习题

数据采集、清洗和处理

一、电商平台用户行为数据处理方案设计（考核时间：20 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装 BI 类商务数据分析软件、压缩软件、Python 3.8+（安装 pandas、mysql-connector-python 库）、MySQL 8.0+及办公软件；
- 3) 商务数据分析师考试平台。

（2）工作任务

- 1) 根据业务需求，设计在线（JSON）和离线（CSV）用户行为数据采集方案，存入 MySQL，确保实时性和完整性；
- 2) 根据业务需求，制定数据迁移方案，将用户行为数据从源 MySQL 迁移至目标 MySQL（模拟云端），保证一致性和效率；
- 3) 根据业务特性，优化用户行为数据存储结构，添加索引，提升查询性能；
- 4) 根据数据质量要求，制定清洗策略处理重复、缺失、格式问题，并评估清洗后数据质量（准确率和填充率）。

（3）技能要求

- 1) 根据业务需求进行在线、离线数据采集；
- 2) 根据业务需求参与制定数据迁移方案；
- 3) 根据业务特性制订数据存储策略；
- 4) 根据数据质量要求制定数据清洗策略及评估方案。

（4）质量指标

- 1) 确保数据采集覆盖率高，实时数据快速入库，离线数据无丢失；
- 2) 保证迁移后数据一致，无记录丢失，迁移过程高效；
- 3) 优化存储结构，支持快速查询并降低存储成本；
- 4) 清洗数据确保行为类型准确，缺失值有效填充。

背景资料

某电商平台致力于优化用户行为数据管理，以提升个性化推荐系统的精准性和营销策略的有效性。业务场景涵盖用户在平台上的在线浏览、点击、购买行为，以及离线生成的订单和库存数据。数据来源多样，包括网站日志（JSON 格式，记录实时操作）、移动端 API（实时传输交互数据）、第三方支付平台（CSV 格式，记录交易信息）以及本地 MySQL 数据库（存储历史订单和库存）。这些数据存在诸多问题，如重复记录（因多设备登录导致）、缺失值（行为类型未正确记录）、格式不一致（时间戳格式多样）等，严重影响数据分析的准确性。为实现数据驱动决策，需设计一套高效的数据处理方案，涵盖以下核心任务：首先，根据业务需求进行在线和离线数据采集，确保实时性和完整性，满足推荐系统对低延迟的要求；其次，制定数据迁移方案，将历史数据从本地 MySQL 数据库迁移至云端数据仓库（模拟为目标 MySQL 表），保证数据一致性和迁移效率；再次，根据业务特性制定数据存储策略，通过表结构优化和索引设计，提升查询性能并降低存储成本；最后，制定数据清洗策略及评估方案，处理重复、缺失、格式问题，提升数据准确性和一致性。目标是构建一个高效、可靠的数据处理流程，支持实时推荐、高并发查询和动态营销策略，同时确保数据安全和系统扩展性。

操作要求：

- （1）编写代码导入必要的 Python 库。使用 `try...except ImportError` 结构对 `pandas` 和 `mysql.connector` 的导入进行异常处理，以验证库的可用性。
- （2）实现数据库连接，使用 `mysql.connector.connect(db_config)` 的字典解包方式。连接成功后，执行 SQL 查询以获取 `user_behavior` 表的前 5 条记录，验证数据预览功能。
- （3）使用 `pandas` 从 JSON 文件读取在线数据，在验证其记录数、字段数及数据完整性后，使用 `cursor.executemany()` 的批量插入方法将其存入数据库。
- （4）使用 `pandas` 从 CSV 文件读取离线数据，在验证记录数和字段数满足要求后，使用 `cursor.executemany()` 的批量插入方法将其追加到数据库的相应表中。
- （5）编写数据迁移逻辑，从源表查询数据后，在 Python 中使用 `executemany()` 批量插入到目标表。同时，使用参数化查询 (`VALUES (%s, ...)`) 将迁移日志写入 `migration_log` 表。
- （6）为指定表创建基于 `user_id` 和 `action_type` 的复合索引。代码中必须包含检查索引是否已存在的逻辑，以优化查询性能并确保脚本可重复执行。

(7) 使用 pandas 的向量化操作完成数据清洗：使用 `df.drop_duplicates()`去重，使用 `pd.to_datetime()`格式化时间戳，并使用 `df.fillna()`填充 `action_type` 的缺失值，确保数据一致性和完整性。

(8) 使用 `isin().mean()`的向量化方法高效计算 `action_type` 字段的有效值比例，以评估数据质量并返回准确率指标。

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M1 - 库验证成功”；无 <code>ModuleNotFoundError</code> 。	结果正确，且使用了推荐的 <code>try...except ImportError</code> 结构进行验证，得 2 分。 结果正确，但仅使用了 <code>import</code> 语句，未进行异常处理，得 1 分。 结果错误，得 0 分。		
M2	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M2 - 数据库连接成功”“M2 - SQL 查询执行成功”“M2 - 数据预览（表格格式）”“M2 - 返回记录数: [1-5]”。	结果正确，且使用了推荐的 <code>mysql.connector.connect(**db_config)</code> 解包字典方式连接，得 2 分。 结果正确，但手动逐一列出了连接参数，得 1 分。 结果错误，得 0 分。		
M3	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M3 - 数据验证通过”“M3 - 记录数: ≥ 1000 ”“M3 - 字段数: 8”“M3 - 数据验证状态: success”； <code>ecommerce.sql</code> 中 <code>user_behavior</code> 表记录数 $\geq 1,000$ ，字段数=8，无空值。	结果正确，且使用了推荐的 <code>pd.read_json()</code> 和 <code>cursor.executemany()</code> 批量插入方法，得 2 分。 结果正确，但使用了任何形式的逐条循环插入（如 <code>iterrows</code> , <code>apply</code> 等），得 1 分。 结果错误，得 0 分。		
M4	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M4 - 数据验证通过”“M4 - 记录数: ≥ 10000 ”“M4 - 字段数: 8”“M4 - 数据验证状态: success”； <code>ecommerce.sql</code> 中 <code>user_behavior</code> 表记录数 $\geq 10,000$ ，字段数=8。	结果正确，且使用了推荐的 <code>pd.read_csv()</code> 和 <code>cursor.executemany()</code> 批量插入方法，得 2 分。 结果正确，但使用了任何形式的逐条循环插入（如 <code>iterrows</code> , <code>apply</code> 等），得 1 分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
			结果错误，得 0 分。		
M5	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M5 - 数据迁移成功”“M5 - 迁移记录数: [≈11000]”“M5 - 迁移状态: success”；ecommerce.sql 中 user_behavior_target 表记录数 ≈ 11,000。	结果正确，且使用了推荐的 Python SELECT 后 executemany() 批量 INSERT 的方法，得 2 分。 结果正确，但使用了 INSERT INTO ... SELECT FROM ... SQL 语句或逐条迁移，得 1 分。 结果错误，得 0 分。		
M6	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M6 - 迁移日志记录成功”“M6 - 日志状态: success”；ecommerce.sql 中 migration_log 表记录存在。	结果正确，且使用了推荐的参数化查询 (VALUES (%s, ...)) 以防止 SQL 注入，得 2 分。 结果正确，但使用了 f-string 或字符串拼接来构造 SQL，得 1 分。 结果错误，得 0 分。		
M7	2	日志显示“M7 - 索引创建成功”或“M7 - 索引 idx_user_action 已存在”；ecommerce.sql 中 user_behavior_target 表存在索引 idx_user_action。	结果正确，且代码中包含了推荐的检查索引是否已存在的逻辑，得 2 分。 结果正确，但仅执行 CREATE INDEX，未处理索引已存在的情况，得 1 分。 结果错误，得 0 分。		
M8	2	日志显示“M8 - 数据去重成功”“M8 - 移除重复记录: [≈200] 条”“M8 - 去重状态: success”；user_behavior_cleaned.csv 记录数 ≈ 9,800，字段数=8，无乱码。	结果正确，且使用了推荐的 df.drop_duplicates() 向量化方法，得 2 分。 结果正确，但使用了循环等其他逻辑上可行但低效的方法，得 1 分。 结果错误，得 0 分。		
M9	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M9 - 数据清洗成功”“M9 - 填充 action_type 缺失值: [≈200] 个”“M9 - 格式化 timestamp 成功，填充空时间戳: [0-50] 个”“M9 - 数据保存成功”“M9 - 保存状态: success”；user_behavior_cleaned.csv 记录数 ≈ 9,800，字段数=8，无缺失值，timestamp 格式为 YYYY-MM-DD HH:MM:SS，无乱码。	结果正确，且使用了推荐的 df.fillna() 和 pd.to_datetime() 等向量化操作，得 2 分。 结果正确，但使用了 .apply() 配合自定义函数等低效方法，得 1 分。 结果错误，得 0 分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M10	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M10 - 数据质量评估成功”“M10 - 行为类型准确率： $[\geq 0.99]$ ”“M10 - 评估状态：success”； user_behavior_cleaned.csv 的 action_type 均为合法值（浏览数据、点击数据、购买数据、数据）。	结果正确，且使用了推荐的 isin().mean() 向量化方法，得 2 分。 结果正确，但使用了循环计数的方法，得 1 分。 结果错误，得 0 分。		
合计配分	20	合计得分			

二、生鲜配送订单数据处理方案设计（考核时间：20 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装 BI 类商务数据分析软件、压缩软件、Python 3.8+（安装 pandas、mysql-connector-python 库）、MySQL 8.0+及办公软件；
- 3) 商务数据分析师考试平台。

（2）工作任务

- 1) 设计实时（JSON）和离线（CSV）订单数据采集方案，存入 MySQL，确保实时性和完整性；
- 2) 制定数据迁移方案，将订单数据从源 MySQL 迁移至目标 MySQL，保证一致性和效率；
- 3) 优化订单数据存储结构，添加索引，提升查询性能；
- 4) 制定清洗策略处理重复、缺失、格式问题，评估清洗后数据质量（准确率和填充率）。

（3）技能要求

- 1) 根据业务需求进行在线、离线数据采集；
- 2) 根据业务需求参与制定数据迁移方案；
- 3) 根据业务特性制订数据存储策略；

4) 根据数据质量要求制定数据清洗策略及评估方案。

(4) 质量指标

- 1) 确保数据采集覆盖率高，实时订单快速入库，离线数据无丢失；
- 2) 保证迁移后数据一致，无记录丢失，迁移过程高效；
- 3) 优化存储结构，支持快速查询并降低存储成本；
- 4) 清洗数据确保订单状态准确，缺失值有效填充。

背景资料

某新零售生鲜配送平台致力于优化订单数据管理，以提升配送效率和客户满意度。业务场景涵盖用户通过移动端或网页端下单购买生鲜商品，涉及实时订单生成、库存更新和配送调度。数据来源多样，包括移动端 API（JSON 格式，记录实时订单）、第三方物流系统（CSV 格式，记录配送信息）、本地 MySQL 数据库（存储历史订单和库存）以及供应商接口（提供商品库存更新）。这些数据存在问题，包括重复订单（因网络延迟导致）、缺失值（订单状态未记录）、格式不一致（配送时间格式多样）等，影响配送调度和库存管理的准确性。为实现高效运营，需设计一套订单数据处理方案，涵盖以下核心任务：首先，采集实时和离线订单数据，确保数据的实时性和完整性，满足配送系统对低延迟的要求；其次，制定数据迁移方案，将历史订单数据从本地 MySQL 迁移至云端数据仓库（模拟为目标 MySQL 表），保证数据一致性和迁移效率；再次，优化订单数据存储结构，通过索引设计提升查询性能，降低存储成本；最后，制定数据清洗策略及评估方案，处理重复、缺失、格式问题，提升订单数据的准确性和一致性。目标是构建一个高效、可靠的订单数据处理流程，支持实时配送调度、高并发查询和库存优化，同时确保数据安全和系统扩展性。

操作要求：

- (1) 导入 `pandas` 和 `mysql.connector` 库，以确保支持数据处理和数据库操作。
- (2) 连接到 MySQL 数据库并查询 `orders` 表的前 5 条记录，以验证连接和预览数据。
- (3) 读取 `online_orders.json` 文件，验证其记录数不少于 1,000 且字段数为 6，然后将数据插入到 `orders` 表中。
- (4) 读取 `offline_orders.csv` 文件，验证其记录数不少于 10,000 且字段数为 6，

格式化 delivery_time 字段后，将数据插入到 orders 表中。

(5) 将 orders 表中的所有数据完整迁移到 orders_target 表中。

(6) 使用参数化查询，将数据迁移的源表名、记录数、状态和时间等信息记录到 migration_log 表中。

(7) 为 orders_target 表的 user_id 和 order_status 字段创建一个名为 idx_user_order 的复合索引。

(8) 对数据进行清洗，包括基于 user_id, item_id, delivery_time 移除重复记录，将 order_status 的缺失值填充为 'pending'，并统一 delivery_time 的格式，最后将清洗结果保存至 output/orders_cleaned.csv。

(9) 读取 output/orders_cleaned.csv 文件，评估 order_status 字段中合法值 ('pending', 'confirmed', 'delivered') 的比例，确保其准确率不低于 0.99。

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M1 - 库验证成功”；无 ModuleNotFoundError。	结果正确，且使用了 try...except ImportError 结构进行验证，得 2 分。 结果正确，但仅使用了 import 语句，未进行异常处理，得 1 分。 结果错误，得 0 分。		
M2	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M2 - 数据库查询成功”“M2 - 查询记录数: 5”“M2 - 查询状态: success”；包含 5 条记录的字段值。	结果正确，且使用了 mysql.connector.connect(**db config) 解包字典方式连接，得 2 分。 结果正确，但手动逐一列出了连接参数，得 1 分。 结果错误，得 0 分。		
M3	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M3 - 数据验证通过”“M3 - 记录数: [≥1000]”“M3 - 字段数: 6”“M3 - 数据验证状态: success”；fresh_delivery_db.sql 中 orders 表记录数 ≥ 1,000，字段数=6，无空值。	结果正确，且使用了逐条循环 INSERT 方法，得 2 分。 结果正确，但使用了其他方法（如 cursor.executemany() 批量插入），得 1 分。 结果错误，得 0 分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M4	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M4 - 数据验证通过”“M4 - 记录数: [≥10000]”“M4 - 字段数: 6”“M4 - 数据验证状态: success”；fresh_delivery_db.sql 中 orders 表记录数 ≥ 10,000，字段数=6。	结果正确，且使用了逐条循环 INSERT 方法，得 2 分。 结果正确，但使用了其他方法（如 cursor.executemany() 批量插入），得 1 分。 结果错误，得 0 分。		
M5	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M5 - 数据迁移成功”“M5 - 迁移记录数: [≈11200]”“M5 - 迁移状态: success”；fresh_delivery_db.sql 中 orders_target 表记录数 ≈ 11,200。	结果正确，且使用了逐条循环迁移方法，得 2 分。 结果正确，但使用了其他方法（如 executemany() 或 INSERT INTO ... SELECT），得 1 分。 结果错误，得 0 分。		
M6	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M6 - 迁移日志记录成功”“M6 - 日志状态: success”；fresh_delivery_db.sql 中 migration_log 表记录存在。	结果正确，且使用了参数化查询（VALUES (%s, ...)），得 2 分。 结果正确，但使用了 f-string 或字符串拼接来构造 SQL，得 1 分。 结果错误，得 0 分。		
M7	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M7 - 索引创建成功”或“M7 - 索引 idx_user_order 已存在”；fresh_delivery_db.sql 中 orders_target 表存在复合索引 idx_user_order。	结果正确，且使用了直接 CREATE INDEX 方法，得 2 分。 结果正确，但增加了检查索引是否已存在的逻辑，得 1 分。 结果错误，得 0 分。		
M8	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M8 - 数据去重成功”“M8 - 移除重复记录: [≈200] 条”“M8 - 填充 order_status 缺失值: [≈200] 个”“M8 - 格式化 delivery_time 成功，填充空时间戳: [0-50] 个”“M8 - 数据保存成功”“M8 - 保存状态: success”；orders_cleaned.csv 记录数 ≈ 9,800，字段数=6，无缺失值，delivery_time 格式为 YYYY-MM-DD HH:MM:SS，无乱码。	结果正确，且四项操作均使用了特定向量化方法（drop_duplicates, fillna, pd.to_datetime, to_csv），得 3 分。 结果正确，但部分操作使用了其他逻辑正确的方法（如用 .apply() 格式化时间），得 2 分。 结果正确，但主要或全部使用了低效方法（如多个 .apply() 或循环），得 1 分。 结果错误，得 0 分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M9	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M9 - 数据质量评估成功”“M9 - 订单状态准确率: ≥ 0.99 ”“M9 - 评估状态: success”；orders_cleaned.csv 的 order_status 均为合法值（pending、confirmed、delivered）。	结果正确，且使用了 isin().mean() 向量化方法，得 3 分。 结果正确，且使用了其它正确的向量化方法（如 isin().sum() / len(df)），得 2 分。 结果正确，但使用了循环计数的低效方法，得 1 分。 结果错误，得 0 分。		
合计配分	20	合计得分			

三、社交平台内容互动数据处理方案设计（考核时间：20 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装 BI 类商务数据分析软件、压缩软件、Python 3.8+（安装 pandas、mysql-connector-python 库）、MySQL 8.0+及办公软件；
- 3) 商务数据分析师考试平台。

（2）工作任务

- 1) 设计实时（JSON）和离线（CSV）内容互动数据采集方案，存入 MySQL；
- 2) 将内容数据从源 MySQL 迁移至目标 MySQL，保证一致性和效率；
- 3) 优化存储结构，添加索引，提升查询性能；
- 4) 处理重复、缺失、格式问题，评估数据质量。

（3）技能要求

- 1) 根据业务需求进行在线、离线数据采集；
- 2) 根据业务需求参与制定数据迁移方案；
- 3) 根据业务特性制订数据存储策略；
- 4) 根据数据质量要求制定数据清洗策略及评估方案。

（4）质量指标

- 1) 确保覆盖率高，实时数据快速入库，离线数据无丢失；
- 2) 保证数据一致，无记录丢失，迁移高效；

- 3) 优化结构，支持快速查询；
- 4) 确保互动类型准确，缺失值有效填充。

背景资料

某社交平台专注于优化内容互动数据管理，以提升用户互动和内容推荐效果。业务场景涵盖用户在平台上发布帖子、点赞、评论等互动行为，涉及实时内容生成和历史数据分析。数据来源包括移动端 API(JSON 格式，记录实时互动)、第三方分析工具(CSV 格式，记录历史互动)、本地 MySQL 数据库(存储帖子和评论)以及外部爬虫(抓取热门话题)。数据存在问题，如重复互动(因多设备操作)、缺失值(互动类型未记录)、格式不一致(发布时间格式多样)等，影响推荐算法和用户分析的准确性。为提升平台体验，需设计一套内容数据处理方案，涵盖以下任务：首先，采集实时和离线内容互动数据，确保实时性和完整性；其次，制定迁移方案，将历史互动数据从本地 MySQL 迁移至云端(模拟为目标 MySQL 表)，保证一致性和效率；再次，优化数据存储结构，通过索引提升查询性能；最后，制定清洗策略及评估方案，处理重复、缺失、格式问题，提升数据质量。目标是构建高效的内容数据处理流程，支持实时推荐、高并发查询和用户分析。

操作要求：

- (1) 导入 pandas 和 mysql.connector 库，以确保支持后续的数据处理和数据库操作。
- (2) 连接到 MySQL 数据库并查询 content_data 表的前 5 条记录，以验证连接并预览数据。
- (3) 从 data/online_content.json 文件读取在线数据，验证记录数不少于 1,000 且字段数为 6 且无空值后，将其插入到 content_data 表中。
- (4) 从 data/content_data.csv 文件读取离线数据，验证记录数不少于 10,000 且字段数为 6 后，将其插入到 content_data 表中。
- (5) 将 content_data 源表中的所有数据完整迁移到 content_data_target 目标表中。
- (6) 使用参数化查询，将数据迁移的关键信息(如源表名、迁移记录数、状态和时间)记录到 migration_log 表中。
- (7) 为 content_data_target 表中基于 user_id 字段创建一个名为 idx_user_id 的索引，并应处理索引已存在的情况。
- (8) 对数据进行清洗，包括移除重复记录、将 action_type 字段的缺失值填充

为 'post'、并将 post_time 字段统一格式化为 YYYY-MM-DD HH:MM:SS，然后将清洗后的数据保存至 output/content_data_cleaned.csv 文件，并最终评估 action_type 的准确性。

2. 评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	2	考评员运行考生提交的PY文件，检查运行过程日志显示“M1 - 库验证成功”；无ModuleNotFoundError。	结果正确，且正确填写 import pandas as pd 和 import mysql.connector，得2分。 结果正确，但未正确填写 import pandas as pd 和 import mysql.connector，得1分。 结果错误，得0分。		
M2	2	考评员运行考生提交的PY文件，检查运行过程日志显示“M2 - 数据库连接成功”“M2 - SQL 查询执行成功”“M2 - 数据预览（表格格式）”“M2 - 返回记录数：[1-5]”。	结果正确，且使用了mysql.connector.connect(db_config)解包字典方式连接，得2分。 结果正确，但手动逐一列出了连接参数，得1分。 结果错误，得0分。		
M3	3	考评员运行考生提交的PY文件，检查运行过程日志显示“M3 - 数据验证通过”“M3 - 记录数：[≥1000]”“M3 - 字段数：6”“M3 - 数据验证状态：success”；social_platform.sql 中 content_data 表记录数 ≥ 1,000，字段数=6，无空值。	结果正确，且使用了逐条循环INSERT方法，得3分。 结果正确，但使用了其他逻辑正确的循环方法（如 itertuples 或 apply），得2分。 结果正确，但使用了与模板要求不符的批量插入方法（如 executemany），得1分。 结果错误，得0分。		
M4	3	考评员运行考生提交的PY文件，检查运行过程日志显示“M4 - 数据验证通过”“M4 - 记录数：[≥10000]”“M4 - 字段数：6”“M4 - 数据验证状态：success”；social_platform.sql 中 content_data 表记录数 ≥ 10,000，字段数=6。	结果正确，且使用了逐条循环INSERT方法，得3分。 结果正确，但使用了其他逻辑正确的循环方法（如 itertuples 或 apply），得2分。 结果正确，但使用了与模板要求不符的批量插入方法（如 executemany），得1分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
			结果错误，得 0 分。		
M5	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M5 - 数据迁移成功”“M5 - 迁移记录数: [≈11000]”“M5 - 迁移状态: success”；social_platform.sql 中 content_data_target 表记录数 ≈ 11,000。	结果正确，且使用了逐条循环迁移方法，得 3 分。 结果正确，但使用了 INSERT INTO ... SELECT FROM ... SQL 语句，得 2 分。 结果正确，但使用了与模板要求不符的批量插入方法（如 executemany），得 1 分。 结果错误，得 0 分。		
M6	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M6 - 迁移日志记录成功”“M6 - 日志状态: success”；social_platform.sql 中 migration_log 表记录存在。	结果正确，且使用了参数化查询（VALUES (%s, ...)），得 2 分。 结果正确，但使用了 f-string 或字符串拼接来构造 SQL，得 1 分。 结果错误，得 0 分。		
M7	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M7 - 索引创建成功”或“M7 - 索引 idx_user_id 已存在”；social_platform.sql 中 content_data_target 表存在索引 idx_user_id。	结果正确，且使用了先检查后创建（SHOW INDEX ...）的完整逻辑，得 2 分。 结果正确，但仅执行 CREATE INDEX 语句，未处理索引已存在的情况，得 1 分。 结果错误，得 0 分。		
M8	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M8 - 数据清洗成功”“M8 - 移除重复记录: [≈200] 条”“M8 - 填充 action_type 缺失值: [≈200] 个”“M8 - 格式化 post_time 成功”“M8 - 数据质量评估成功”“M8 - 行为类型准确率: [≥0.99]”；content_data_cleaned.csv 记录数 ≈ 9,800，字段数=6，无缺失值，post_time 格式为 YYYY-MM-DD HH:MM:SS，无乱码。	结果正确，且所有清洗和评估操作均使用了特定向量化方法（`drop_duplicates(keep='first')`，`fillna('post')`，`pd.to_datetime`，`isin().mean()`），得 3 分。 结果正确，但部分操作使用了其他逻辑正确的方法（如用 `.apply()` 格式化时间），得 2 分。 结果正确，但主要或全部使用了低效方法（如多个 `.apply()` 或循环），得 1 分。 结果错误，得 0 分。		
合计配分	20	合计得分			

四、电商平台商品评论数据处理方案设计（考核时间：20 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装 BI 类商务数据分析软件、压缩软件、Python 3.8+（安装 pandas、mysql-connector-python 库）、MySQL 8.0+及办公软件；
- 3) 商务数据分析师考试平台。

（2）工作任务

- 1) 设计实时（JSON）和离线（CSV）评论数据采集方案，存入 MySQL；
- 2) 将评论数据从源 MySQL 迁移至目标 MySQL，保证一致性和效率；
- 3) 优化存储结构，添加索引，提升查询性能；
- 4) 处理重复、缺失、格式问题，评估数据质量。

（3）技能要求

- 1) 根据业务需求进行在线、离线数据采集；
- 2) 根据业务需求参与制定数据迁移方案；
- 3) 根据业务特性制订数据存储策略；
- 4) 根据数据质量要求制定数据清洗策略及评估方案。

（4）质量指标

- 1) 确保覆盖率高，实时数据快速入库，离线数据无丢失；
- 2) 保证数据一致，无记录丢失，迁移高效；
- 3) 优化结构，支持快速查询；
- 4) 确保评分准确，缺失值有效填充。

背景资料

某电商平台致力于优化商品评论数据管理，以提升客户信任和产品推荐效果。业务场景涵盖用户对购买的商品发布评论、评分和反馈，涉及实时评论提交和历史评论分析。数据来源多样，包括网页和移动端 API（JSON 格式，记录实时评论）、第三方分析工具（CSV 格式，存储历史评论）、本地 MySQL 数据库（存储评论和商品数据）以及外部情感分析服务（提供评论情感评分）。数据存在问题，包括重复评论（因多次提交导致）、缺失值（评分未提供）、格式不一致（评论时间格式多样）等，影响情感分析和推荐系统的准确性。为实现数据驱动决策，需设计一套评论数据处理方案，涵盖以下任务：首先，采集实时和离线评论数据，确保实时性和完整性；其次，制定迁移方案，将历史评论数据从本地 MySQL 迁移至云端（模拟为目标 MySQL 表），保证一致性和效率；再次，

优化数据存储结构，通过索引提升查询性能；最后，制定清洗策略及评估方案，处理重复、缺失、格式问题，提升数据质量。目标是构建高效的评论数据处理流程，支持实时情感分析、高并发查询和推荐优化。

操作要求：

(1) 编写代码以导入 `pandas` 和 `mysql.connector` 库，以支持后续的数据处理和数据库操作。

(2) 实现数据库连接并执行查询，从 `review_data` 表获取前 5 条记录，以验证数据库连接和数据预览功能。

(3) 分别从 `online_reviews.json` 和 `review_data.csv` 文件读取在线与离线数据，在验证其记录数、字段数及数据完整性后，将所有数据统一插入到 `review_data` 表中。

(4) 将源表 `review_data` 中的所有数据迁移到目标表 `review_data_target`，并使用参数化查询将本次迁移的日志（包括表名、记录数、状态和时间戳）记录到 `migration_log` 表中。

(5) 为 `review_data_target` 表的 `user_id` 字段创建一个名为 `idx_user_id` 的索引，以优化数据库查询性能。

(6) 对数据进行清洗，包括移除重复记录，将 `rating` 字段的缺失值统一填充为 3，并格式化 `review_time` 字段，最终将清洗结果保存至 `output/review_data_cleaned.csv` 文件。

(7) 评估已清洗数据的质量，计算并报告 `rating` 字段中评分值在 1 到 5 之间的合法比例。

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	2	考评员运行考生提交的 PY 文件，检查运行过程日志“M1 - 库验证成功：mysql-connector-python 和 pandas 已导入”。	结果正确，且正确填写 <code>import pandas as pd</code> 和 <code>import mysql.connector</code> ，得 2 分。结果正确，但未正确填写 <code>import pandas as pd</code> 和 <code>import mysql.connector</code> ，得 1 分。结果错误，得 0 分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M2	2	考评员运行考生提交的 PY 文件，显示“M2 - 数据库连接成功”“M2 - SQL 查询执行成功”“M2 - 返回记录数: [1-5]”；ecommerce.sql 中 review_data 表记录数 ≥ 11,200，字段数=6。	结果正确，且使用了 mysql.connector.connect(**db_config) 解包字典方式连接，得 2 分。 结果正确，但手动逐一列出了连接参数，得 1 分。 结果错误，得 0 分。		
M3	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M3 - 数据验证通过”“M3 - 记录数: 1000”“M3 - 数据验证状态: success”；ecommerce.sql 中 review_data 表包含 1,000 条 JSON 数据，字段数=6。	结果正确，且使用了逐条循环 INSERT 方法，得 2 分。 结果正确，但使用了其他方法（如 cursor.executemany() 批量插入），得 1 分。 结果错误，得 0 分。		
M4	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M4 - 数据验证通过”“M4 - 记录数: [≥10200]”“M4 - 数据验证状态: success”；ecommerce.sql 中 review_data 表记录数 ≥ 11,200，字段数=6。	结果正确，且使用了逐条循环 INSERT 方法，得 2 分。 结果正确，但使用了其他方法（如 cursor.executemany() 批量插入），得 1 分。 结果错误，得 0 分。		
M5	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示包含“M5 - 数据迁移成功”“M5 - 迁移状态: success”；数据库 review_data_target 表记录数在 10,800-11,400 之间。	结果正确，且使用了逐条循环迁移方法，得 2 分。 结果正确，但使用了其他方法（如 executemany() 或 INSERT INTO ... SELECT），得 1 分。 结果错误，得 0 分。		
M6	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M6 - 迁移日志记录成功”“M6 - 日志状态: success”；ecommerce.sql 中 migration_log 表包含记录，status 为 "success"。是=	结果正确，且使用了参数化查询 (VALUES (%s, ...))，得 2 分。 结果正确，但使用了 f-string 或字符串拼接来构造 SQL，得 1 分。 结果错误，得 0 分。		
M7	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M7 - 索引创建成功”或“M7 - 索引 idx_user_id 已存在”；ecommerce.sql 中	结果正确，且使用了直接 CREATE INDEX 方法，得 3 分。 结果正确，但使用了 try...except 结构来捕获重复创建的错误，得 2 分。 结果正确，但增加了检查索引是否已		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
		review_data_target 表包含索引 idx_user_id。	存在的逻辑（如 SHOW INDEX），得 1 分。 结果错误，得 0 分。		
M8	3	考评员运行考生提交的 PY 文件，检查运行过程日志包含“M8 - 数据清洗成功”“M8 - 移除重复记录”“M8 - 填充 rating 缺失值”“M8 - 格式化 review_time 成功”；review_data_cleaned.csv 记录数在 9,500–10,200 之间，字段数 =6（user_id, rating, product_id, review_time, data_source, data_status），无缺失值，review_time 格式为 YYYY-MM-DD HH:MM:SS，移除重复记录数在 150–250 之间，填充 rating 缺失值数在 0–250 之间。	结果正确，且四项操作均使用了特定向量化方法（drop_duplicates(keep='first'), fillna(3), pd.to_datetime, to_csv)，得 3 分。 结果正确，但部分操作使用了其他逻辑正确的方法（如用 .apply() 格式化时间），得 2 分。 结果正确，但主要或全部使用了低效方法（如多个 .apply() 或循环），得 1 分。 结果错误，得 0 分。		
M9	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M9 - 数据质量评估成功”“M9 - 评分合法比例：[≥0.99]”，review_data_cleaned.csv 中 rating 合法值（1-5）比例 ≥0.99。	结果正确，且使用了 isin([...]).mean() 向量化方法，得 2 分。 结果正确，但使用了循环计数等其他方法，得 1 分。 结果错误，得 0 分。		
合计配分	20	合计得分			

五、金融平台交易记录数据处理方案设计（考核时间：20 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装 BI 类商务数据分析软件、压缩软件、Python 3.8+（安装 pandas、mysql-connector-python 库）、MySQL 8.0+ 及办公软件；
- 3) 商务数据分析师考试平台。

（2）工作任务

- 1) 设计实时（JSON）和离线（CSV）交易数据采集方案，存入 MySQL；

- 2) 将交易数据从源 MySQL 迁移至目标 MySQL，保证一致性和效率；
- 3) 优化存储结构，添加索引，提升查询性能；
- 4) 处理重复、缺失、格式问题，评估数据质量。

(3) 技能要求

- 1) 根据业务需求进行在线、离线数据采集；
- 2) 根据业务需求参与制定数据迁移方案；
- 3) 根据业务特性制订数据存储策略；
- 4) 根据数据质量要求制定数据清洗策略及评估方案。

(4) 质量指标

- 1) 确保覆盖率高，实时数据快速入库，离线数据无丢失；
- 2) 保证数据一致，无记录丢失，迁移高效；
- 3) 优化结构，支持快速查询；
- 4) 确保金额准确，缺失值有效填充。

背景资料

某金融平台需要高效管理交易记录数据，以支持风险监控、合规审计和用户分析。业务场景包括用户通过网页或移动端发起的实时交易以及批量导入的历史交易记录。数据来源多样，包括实时交易接口（JSON 格式，记录实时交易）、银行对账文件（CSV 格式，存储历史交易）、本地 MySQL 数据库（存储交易和用户信息）以及外部风控服务（提供交易风险评分）。数据面临问题，如重复交易（因网络重试导致）、缺失值（交易金额未记录）、格式不一致（交易时间格式多样）等，影响风控分析和审计准确性。为实现数据驱动的金融服务，需设计一套交易记录处理方案，涵盖以下任务：首先，采集实时和离线交易数据，确保高效性和完整性；其次，制定迁移方案，将历史交易数据从本地 MySQL 迁移至云端（模拟为目标 MySQL 表），保证一致性和合规性；再次，优化数据存储结构，通过索引提升查询性能；最后，制定清洗策略及评估方案，处理重复、缺失、格式问题，提升数据质量。目标是构建高效的交易数据处理流程，支持实时风控、高并发查询和审计需求。

操作要求：

- (1) 编写代码以导入 pandas 和 mysql.connector 库，以支持后续的数据处理和数据库操作。
- (2) 实现数据库连接并执行查询，从 transaction_data 表获取前 5 条记录，以验证数据库连接和数据预览功能。

(3) 分别从 `online_transactions.json` 和 `transaction_data.csv` 文件读取在线与离线数据,在验证其记录数和字段数后,将所有数据统一插入到 `transaction_data` 表中。

(4) 将源表 `transaction_data` 中的所有数据迁移到目标表 `transaction_data_target`, 并使用参数化查询将本次迁移的日志(包括表名、记录数、状态和时间戳)记录到 `migration_log` 表中。

(5) 为 `transaction_data_target` 表的 `user_id` 字段创建一个名为 `idx_user_id` 的索引,以优化数据库查询性能。

(6) 对数据进行清洗,包括移除重复记录,将 `amount` 字段的缺失值统一填充为 `100.0`,并格式化 `transaction_time` 字段,最终将清洗结果保存至 `output/transaction_data_cleaned.csv` 文件。

(7) 评估已清洗数据的质量,计算并报告 `transaction_type` 字段中值为 `'transfer'`, `'payment'`, 或 `'investment'` 的合法比例。

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	2	考评员运行考生提交的 PY 文件,检查运行过程日志显示“M1 - 库验证成功”;无 <code>ModuleNotFoundError</code> 。	结果正确,且正确填写 <code>`import pandas as pd`</code> 和 <code>`import mysql.connector`</code> ,得 2 分。 结果正确,但未正确填写 <code>`import pandas as pd`</code> 和 <code>`import mysql.connector`</code> ,得 1 分。 结果错误,得 0 分。		
M2	2	考评员运行考生提交的 PY 文件,检查运行过程日志显示“M2 - 数据库连接成功”“M2 - SQL 查询执行成功”“M2 - 数据预览(表格格式)”“M2 - 返回记录数: [1-5]”。	结果正确,且使用了 <code>mysql.connector.connect(**db_config)</code> 解包字典方式连接,得 2 分。 结果正确,但手动逐一列出了连接参数,得 1 分。 结果错误,得 0 分。		
M3	2	考评员运行考生提交的 PY 文件,检查运行过程日志显示“M3 - 数据验证通过”“M3 - 记录数:	结果正确,且使用了逐条循环 <code>INSERT</code> 方法,得 2 分。 结果正确,但使用了其他方法(如		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
		[≥1000]”“M3 - 字段数: 6”“M3 - 数据验证状态: success”; financial_platform.sql 中 transaction_data 表记录数 ≥ 1,000, 字段数=6, 无空值。	cursor.executemany() 批量插入), 得 1 分。 结果错误, 得 0 分。		
M4	2	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M4 - 数据验证通过”“M4 - 记录数: [≥10000]”“M4 - 字段数: 6”“M4 - 数据验证状态: success”; financial_platform.sql 中 transaction_data 表记录数 ≥ 10,000, 字段数=6。	结果正确, 且使用了逐条循环 INSERT 方法, 得 2 分。 结果正确, 但使用了其他方法 (如 cursor.executemany() 批量插入), 得 1 分。 结果错误, 得 0 分。		
M5	2	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M5 - 数据迁移成功”“M5 - 迁移记录数: [≈11000]”“M5 - 迁移状态: success”; financial_platform.sql 中 transaction_data_target 表记录数 ≈ 11,000。	结果正确, 且使用了逐条循环迁移方法, 得 2 分。 结果正确, 但使用了其他方法 (如 executemany() 或 INSERT INTO ... SELECT), 得 1 分。 结果错误, 得 0 分。		
M6	2	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M6 - 迁移日志记录成功”“M6 - 日志状态: success”; financial_platform.sql 中 migration_log 表记录存在。	结果正确, 且使用了参数化查询 (VALUES (%s, ...), 得 2 分。 结果正确, 但使用了 f-string 或字符串拼接来构造 SQL, 得 1 分。 结果错误, 得 0 分。		
M7	2	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M7 - 索引创建成功”或“M7 - 索引 idx_user_id 已存在”; financial_platform.sql 中 transaction_data_target 表存在索引 idx_user_id。	结果正确, 且使用了直接 CREATE INDEX 方法, 得 2 分。 结果正确, 但增加了检查索引是否已存在的逻辑 (如 SHOW INDEX 或 try...except), 得 1 分。 结果错误, 得 0 分。		
M8	3	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M8 - 数据清洗成功”“M8 - 移除重复记录: [≈200] 条”“M8 - 填充 amount 缺失值: [≈200] 个”“M8 - 格式化 transaction_time 成	结果正确, 且三项操作均使用了特定向量化方法 (drop_duplicates(keep='first'), fillna(100.0), pd.to_datetime), 得 3 分。 结果正确, 但部分操作使用了其他逻辑		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
		功 ” ; transaction_data_cleaned.csv 记录数 ≈ 9,800, 字段数=6, 无缺失值, transaction_time 格式为 YYYY-MM-DD HH:MM:SS, 无乱码。	辑正确的方法(如用 .apply() 格式化时间), 得 2 分。 结果正确, 但主要或全部使用了低效方法(如多个 .apply() 或循环), 得 1 分。 结果错误, 得 0 分。		
M9	3	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M9 - 数据质量评估成功”“M9 - 交易类型准确率: [≥0.99]”。	结果正确, 且使用了 isin([...]).mean() 向量化方法, 得 3 分。 结果正确, 且使用了其它正确的向量化方法(如 isin().sum() / len(df)), 得 2 分。 结果正确, 但使用了循环计数的低效方法, 得 1 分。 结果错误, 得 0 分。		
合计配分	20	合计得分			

六、教育平台学习记录数据查询与分析系统设计（考核时间：20 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装 BI 类商务数据分析软件、压缩软件、Python 3.8+（`pandas`、`mysql-connector-pytho` 库）、MySQL 8.0+及办公软件；
- 3) 数据库及商务数据分析师考试平台。

（2）工作任务

- 1) 实现按用户标识或时间范围查询学习记录；
- 2) 使用 MySQL 索引提升查询性能；
- 3) 通过 MySQL 和 Python 分析用户学习活跃度或评分分布；
- 4) 在 Python 环境中展示分析结果。

（3）技能要求

- 1) 使用计算引擎优化数据查询效率；
- 2) 通过计算平台构建检索分析系统；
- 3) 通过计算平台显示检索分析结果。

(4) 质量指标

- 1) 查询学习记录功能正确实现；
- 2) 查询效率通过索引优化得到提升；
- 3) 检索分析系统准确分析用户学习活跃度和评分分布；
- 4) 分析结果以清晰格式展示。

背景材料

某教育平台需高效管理用户学习记录数据，以支持课程推荐、学习进度跟踪和教学优化。业务场景包括用户通过网页或移动端提交的实时学习记录（如课程评分、学习时长）以及批量导入的历史学习记录。数据来源包括实时学习接口（JSON 格式，记录用户实时学习行为）、历史学习归档（CSV 格式，存储历史记录）和本地 MySQL 数据库（存储学习记录和用户信息）。数据存在问题，包括重复记录、缺失值、时间格式不一致（需统一为 YYYY-MM-DD HH:MM:SS），影响查询效率和分析准确性。为实现数据驱动的教育服务，需在局域网环境中构建数据查询与分析系统，涵盖以下功能：查询数据，支持按用户标识或时间范围检索学习记录；使用 MySQL 索引优化查询效率；通过 MySQL 和 Python 构建检索分析系统，分析用户学习活跃度或评分分布；在 Python 环境中展示分析结果。

操作要求：

(1) 从 `learning_records` 表中查询学习记录，要求能分别实现：

按用户标识检索：通过参数化查询筛选特定 `user_id` 的记录。

按时间范围检索：使用 `BETWEEN` 语法查询特定 `learning_time` 区间内的记录。

(2) 为 `learning_records` 表的 `user_id` 和 `learning_time` 字段创建 `idx_user_id` 和 `idx_learning_time` 索引，并且必须通过先使用 `SHOW INDEX` 检查索引是否存在的方式，来确保索引创建操作的幂等性。

(3) 执行以下两种独立的 SQL 聚合分析：

用户学习活跃度分析：统计 `learning_records` 表中每个用户的记录总数，并按记录数 (`learning_count`) 降序排列，最终返回前 10 名的用户。

评分分布分析：统计 `learning_records` 表中所有非空评分 (`rating`) 的分布情况，并按评分值升序排列结果。

(4) 将上述两种数据分析得到的 `Pandas DataFrame` 结果，均使用 `to_string(index=False)` 方法转换为不包含索引的表格字符串格式进行输出展示。

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示 M1 - 查询用户标识数据状态: <code>success`</code> 且数据库返回至少 1 条记录。	<p>结果正确，且 SQL 查询使用了基础语句 (<code>WHERE 1=1</code>) 并结合了标准的参数化条件，得 3 分。</p> <p>结果正确，但 SQL 查询语句或参数化方式与标准实现不完全一致（如未使用 <code>WHERE 1=1</code>），得 2 分。</p> <p>结果正确，但使用了有安全风险的字符串拼接方式传递参数，得 1 分。</p> <p>结果错误，得 0 分。</p>		
M2	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示 M2 - 查询时间范围数据状态: <code>success</code> 且数据库返回至少 1 条记录。	<p>结果正确，且使用了 <code>BETWEEN %s AND %s</code> 语法进行时间范围查询，得 3 分。</p> <p>结果正确，但使用了其他逻辑正确的 SQL 语法（如 <code>>= %s AND <= %s</code>），得 2 分。</p> <p>结果正确，但使用了有安全风险的字符串拼接方式传递参数，得 1 分。</p> <p>结果错误，得 0 分。</p>		
M3	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示 M3 - 索引创建状态: <code>success</code> 且索引存在。	<p>结果正确，且使用了先检查索引是否存在 (<code>SHOW INDEX</code>) 再执行创建的完整逻辑，确保了代码的幂等性，得 3 分。</p> <p>结果正确，但仅使用 <code>try...except</code> 捕获错误来处理索引已存在的情况，得 2 分。</p> <p>结果正确，但仅执行</p>		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
			CREATE INDEX 语句，未处理索引已存在的情况，得 1 分。 结果错误，得 0 分。		
M4	3	考评员运行考生提交的 PY 文件，通过 SQL 查询统计每个用户的学习记录数，并取前 10 名，查询执行成功后，检查运行过程日志显示 M4 - 用户学习活跃度分析状态: success 。	结果正确，且 SQL 查询语句完整运用了 GROUP BY, ORDER BY DESC, 和 LIMIT 10, 得 3 分。 结果正确，但 SQL 查询缺少部分排序或限制条件（如仅 GROUP BY），得 2 分。 结果正确，但使用了低效的 Python 循环处理代替 SQL 聚合，得 1 分。 结果错误，得 0 分。		
M5	3	考评员运行考生提交的 PY 文件，通过 SQL 查询统计每个评分的记录数，排除空值。查询执行成功后，明确记录日志 M5 - 评分分布分析状态: success。	结果正确，且 SQL 查询语句完整运用了 WHERE rating IS NOT NULL, GROUP BY, 和 ORDER BY, 得 3 分。 结果正确，但 SQL 查询缺少部分条件（如未排除 NULL 值或未排序），得 2 分。 结果正确，但使用了低效的 Python 循环处理代替 SQL 聚合，得 1 分。 结果错误，得 0 分。		
M6	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示 M6 - 用户学习活跃度展示状态: success 且输出包含 user_id learning_count。	结果正确，且使用了标准的 to_string(index=False) 方法将 DataFrame 转换为无索引的表格字符串，得 3 分。 结果正确，但使用了其他 Pandas 输出方法（如 to_markdown,		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
			to_html), 得 2 分。 结果正确, 但使用了简单的 print(dataframe), 未指定格式, 得 1 分。 结果错误, 得 0 分。		
M7	2	考评员运行考生提交的 PY 文件, 检查运行过程日志显示 M7 - 评分分布展示状态: success 且输出包含 rating rating_count。	结果正确, 且使用了标准的 to_string(index=False) 方法将 DataFrame 转换为无索引的表格字符串, 得 2 分。 结果正确, 但使用了其他格式化输出方法, 得 1 分。 结果错误, 得 0 分。		
合计配分	20	合计得分			

七、社交平台内容分享数据查询与分析系统设计（考核时间：20 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装 BI 类商务数据分析软件、压缩软件、Python 3.8+（pandas、mysql-connector-pytho 库）、MySQL 8.0+及办公软件；
- 3) 数据库及商务数据分析师考试平台。

（2）工作任务

- 1) 查询数据：实现按用户标识或时间范围查询内容数据；
- 2) 优化查询效率：使用 MySQL 索引提升查询性能；
- 3) 构建检索分析系统：通过 MySQL 和 Python 分析用户活跃度或评分分布；
- 4) 展示分析结果：在 Python 环境中展示分析结果。

（3）技能要求

- 1) 使用计算引擎优化数据查询效率；

- 2) 通过计算平台构建检索分析系统;
 - 3) 通过计算平台显示检索分析结果。
- (4) 质量指标
- 1) 查询内容数据功能正确实现;
 - 2) 正确查询效率通过索引优化得到提升;
 - 3) 检索分析系统准确分析用户活跃度和评分分布;
 - 4) 分析结果以清晰格式展示。

背景材料

某社交平台需高效管理用户生成的内容数据，以支持内容推荐、用户行为分析和广告精准投放。业务场景包括用户通过网页或移动端发布的实时内容（如帖子、评论）以及批量导入的历史分享记录。数据来源包括实时内容接口（JSON 格式，记录用户实时发布的内容）、历史归档文件（CSV 格式，存储历史分享记录）和本地 MySQL 数据库（存储内容和用户信息）。数据存在问题，如重复内容、缺失值、时间格式不一致（需统一为 YYYY-MM-DD HH:MM:SS），影响查询效率和分析准确性。为实现数据驱动的社交服务，需在局域网环境中构建数据查询与分析系统，涵盖以下功能：查询数据，支持按用户标识或时间范围检索内容；使用 MySQL 索引优化查询效率；通过 MySQL 和 Python 构建检索分析系统，分析用户活跃度或内容评分分布；在 Python 环境中展示分析结果。

操作要求：

- (1) 从 `content_data` 表中查询内容数据，要求能分别实现：
按用户标识检索：通过参数化查询筛选特定 `user_id` 的记录。
按时间范围检索：使用 `BETWEEN` 语法查询特定 `created_at` 区间内的记录。
- (2) 为 `content_data` 表的 `user_id` 和 `created_at` 字段创建 `idx_user_id` 和 `idx_created_at` 索引，并且必须通过先使用 `SHOW INDEX` 检查索引是否存在的方式，来确保索引创建操作的幂等性。
- (3) 执行以下两种独立的 SQL 聚合分析：
用户活跃度分析：统计 `content_data` 表中每个用户的发布内容总数，并按内容数 (`post_count`) 降序排列，最终返回前 10 名的用户。
评分分布分析：统计 `content_data` 表中所有非空评分 (`rating`) 的分布情况，并按评分值升序排列结果。

(4) 将上述两种数据分析得到的 Pandas DataFrame 结果，均使用 to_string(index=False) 方法转换为不包含索引的表格字符串格式进行输出展示。

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示包含“M1 - 查询用户标识数据状态: success”且数据库返回至少 1 条记录。	<p>结果正确，且 SQL 查询使用了基础语句（WHERE 1=1）并结合了标准的参数化条件，得 3 分。</p> <p>结果正确，但 SQL 查询语句或参数化方式与标准实现不完全一致（如未使用 WHERE 1=1），得 2 分。</p> <p>结果正确，但使用了有安全风险的字符串拼接方式传递参数，得 1 分。</p> <p>结果错误，得 0 分。</p>		
M2	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示包含“M2 - 查询时间范围数据状态: success”且数据库返回至少 1 条记录。	<p>结果正确，且使用了 BETWEEN %s AND %s 语法进行时间范围查询，得 3 分。</p> <p>结果正确，但使用了其他逻辑正确的 SQL 语法（如 >= %s AND <= %s），得 2 分。</p> <p>结果正确，但使用了有安全风险的字符串拼接方式传递参数，得 1 分。</p> <p>结果错误，得 0 分。</p>		
M3	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示包含“M3 - 索引创建状态:	结果正确，且使用了先检查索引是否存在		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
		success” 且索引存在。	<p>(SHOW INDEX) 再执行创建的完整逻辑，确保了代码的幂等性，得 3 分。</p> <p>结果正确，但仅使用 try...except 捕获错误来处理索引已存在的情况，得 2 分。</p> <p>结果正确，但仅执行 CREATE INDEX 语句，未处理索引已存在的情况，得 1 分。</p> <p>结果错误，得 0 分。</p>		
M4	3	<p>考评员运行考生提交的 PY 文件，检查运行过程日志显示包含 “M4 - 用户活跃度分析状态: success” 且输出包含表格 (user_id post_count, 10 行) 与数据库查询 SELECT user_id, COUNT(*) FROM content_data GROUP BY user_id ORDER BY COUNT(*) DESC LIMIT 10 一致，结果正确。</p>	<p>结果正确，且 SQL 查询语句完整运用了 GROUP BY, ORDER BY DESC, 和 LIMIT 10, 得 3 分。</p> <p>结果正确，但 SQL 查询缺少部分排序或限制条件 (如仅 GROUP BY)，得 2 分。</p> <p>结果正确，但使用了低效的 Python 循环处理代替 SQL 聚合，得 1 分。</p> <p>结果错误，得 0 分。</p>		
M5	3	<p>考评员运行考生提交的 PY 文件，检查运行过程日志显示包含 “M5 - 评分分布分析状态: success” 且表格 (rating rating_count, rating 在 1-5 范围内) 与数据库查询 SELECT rating, COUNT(*) FROM content_data WHERE rating IS NOT NULL GROUP BY rating ORDER BY rating 一致，结果正确。</p>	<p>结果正确，且 SQL 查询语句完整运用了 WHERE rating IS NOT NULL, GROUP BY, 和 ORDER BY, 得 3 分。</p> <p>结果正确，但 SQL 查询缺少部分条件 (如未排除 NULL 值或未</p>		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
			排序), 得 2 分。 结果正确, 但使用了低效的 Python 循环处理代替 SQL 聚合, 得 1 分。 结果错误, 得 0 分。		
M6	3	考评员运行考生提交的 PY 文件, 检查运行过程日志显示包含 “M6 - 用户活跃度展示状态: success” 且输出包含 user_id post_count(前 10 行) 格式正确与 M4 结果一致。	结果正确, 且使用了标准的 to_string(index=False) 方法将 DataFrame 转换为无索引的表格字符串, 得 3 分。 结果正确, 但使用了其他 Pandas 输出方法 (如 to_markdown, to_html), 得 2 分。 结果正确, 但使用了简单的 print(dataframe), 未指定格式, 得 1 分。 结果错误, 得 0 分。		
M7	2	考评员运行考生提交的 PY 文件, 检查运行过程日志显示包含 “M7 - 评分分布展示状态: success” 且输出包含 rating rating_count(rating 在 1-5 范围内), 格式正确并与 M5 结果一致。	结果正确, 且使用了标准的 to_string(index=False) 方法将 DataFrame 转换为无索引的表格字符串, 得 2 分。 结果正确, 但使用了其他格式化输出方法, 得 1 分。 结果错误, 得 0 分。		
合计配分	20	合计得分			

八、电商平台商品评论数据查询与分析系统设计（考核时间：20 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装 BI 类商务数据分析软件、压缩软件、Python 3.8+（`pandas`、`mysql-connector-pytho` 库）、MySQL 8.0+及办公软件；
- 3) 数据库及商务数据分析师考试平台。

（2）工作任务

- 1) 实现按用户标识或时间范围查询评论数据；
- 2) 使用 MySQL 索引提升查询性能；
- 3) 通过 MySQL 和 Python 分析用户评论活跃度或评分分布；
- 4) 在 Python 环境中展示分析结果。

（3）技能要求

- 1) 使用计算引擎优化数据查询效率；
- 2) 通过计算平台构建检索分析系统；
- 3) 同构计算平台显示检索分析结果。

（4）质量指标

- 1) 查询评论数据功能正确实现；
- 2) 查询效率通过索引优化得到提升；
- 3) 检索分析系统准确分析用户评论活跃度和评分分布；
- 4) 分析结果以清晰格式展示。

背景材料

某电商平台需高效管理用户提交的商品评论数据，以支持商品推荐、商家管理和用户体验优化。业务场景包括用户通过网页或移动端提交的实时评论（如评分、文字评价）以及批量导入的历史评论记录。数据来源包括实时评论接口（JSON 格式，记录用户实时提交的评论）、历史评论归档（CSV 格式，存储历史评论）和本地 MySQL 数据库（存储评论和商品信息）。数据存在问题，如重复评论、缺失值、时间格式不一致（需统一为 YYYY-MM-DD HH:MM:SS），影响查询效率和分析准确性。为实现数据驱动的电商服务，需在局域网环境中构建数据查询与分析系统，涵盖以下功能：查询数据，支持按用户标识或时间范围检索评论；使用 MySQL 索引优化查询效率；通过 MySQL 和 Python 构建检索分析系统，分析用户评论活跃度或评分分布；在 Python 环境中展示分析结果。

操作要求:

(1) 从 learning_records 表中查询学习记录, 要求能分别实现:

按用户标识检索: 通过参数化查询筛选特定 user_id 的记录。

按时间范围检索: 使用 BETWEEN 语法查询特定 learning_time 区间内的记录。

(2) 为 learning_records 表的 user_id 和 learning_time 字段创建 idx_user_id 和 idx_learning_time 索引, 并且必须通过先使用 SHOW INDEX 检查索引是否存在的方式, 来确保索引创建操作的幂等性。

(3) 执行以下两种独立的 SQL 聚合分析:

用户学习活跃度分析: 统计 learning_records 表中每个用户的记录总数, 并按记录数 (learning_count) 降序排列, 最终返回前 10 名的用户。

评分分布分析: 统计 learning_records 表中所有非空评分 (rating) 的分布情况, 并按评分值升序排列结果。

(4) 将上述两种数据分析得到的 Pandas DataFrame 结果, 均使用 to_string(index=False) 方法转换为不包含索引的表格字符串格式进行输出展示。

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	3	考评员运行考生提交的 PY 文件, 检查运行过程日志显示包含 “M1 - 查询用户标识数据状态: success” 且数据库返回至少 1 条记录。	结果正确, 且 SQL 查询使用了基础语句 (WHERE 1=1) 并结合了标准的参数化条件, 得 3 分。 结果正确, 但 SQL 查询语句或参数化方式与标准实现不完全一致 (如未使用 WHERE 1=1), 得 2 分。 结果正确, 但使用了		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
			有安全风险的字符串拼接方式传递参数，得 1 分。 结果错误，得 0 分。		
M2	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示包含“M2 - 查询时间范围数据状态: success”且数据库返回至少 1 条记录。	结果正确，且使用了 BETWEEN %s AND %s 语法进行时间范围查询，得 3 分。 结果正确，但使用了其他逻辑正确的 SQL 语法(如 >= %s AND <= %s)，得 2 分。 结果正确，但使用了有安全风险的字符串拼接方式传递参数，得 1 分。 结果错误，得 0 分。		
M3	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示包含“M3 - 索引创建状态: success”且索引存在。	结果正确，且使用了先检查索引是否存在 (SHOW INDEX) 再执行创建的完整逻辑，确保了代码的幂等性，得 3 分。 结果正确，但仅使用 try...except 捕获错误来处理索引已存在的情况，得 2 分。 结果正确，但仅执行 CREATE INDEX 语句，未处理索引已存在的情况，得 1 分。 结果错误，得 0 分。		
M4	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示包含“M4 - 用户评论活跃度分析状态: success”且结果正确，程序输出的前 10 条活跃用户数据结果正确，即输出包含字段	结果正确，且 SQL 查询语句完整运用了 GROUP BY, ORDER BY DESC, 和		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
		user_id 和 comment_count。	LIMIT 10, 得 3 分。 结果正确, 但 SQL 查询缺少部分排序或限制条件 (如仅 GROUP BY), 得 2 分。 结果正确, 但使用了低效的 Python 循环处理代替 SQL 聚合, 得 1 分。 结果错误, 得 0 分。		
M5	3	考评员运行考生提交的 PY 文件, 检查运行过程日志显示包含 “M5 - 评分分布分析状态: success” 且结果正确, 程序输出的评分分布数据结果正确, 即输出包含字段 rating 和 rating_count, 且统计值与数据库真实数据一致。	结果正确, 且 SQL 查询语句完整运用了 WHERE rating IS NOT NULL, GROUP BY, 和 ORDER BY, 得 3 分。 结果正确, 但 SQL 查询缺少部分条件 (如未排除 NULL 值或未排序), 得 2 分。 结果正确, 但使用了低效的 Python 循环处理代替 SQL 聚合, 得 1 分。 结果错误, 得 0 分。		
M6	3	考评员运行考生提交的 PY 文件, 检查运行过程日志显示包含 “M6 - 用户评论活跃度展示状态: success” 且输出包含 user_id comment_count。	结果正确, 且使用了标准的 to_string(index=False) 方法将 DataFrame 转换为无索引的表格字符串, 得 3 分。 结果正确, 但使用了其他 Pandas 输出方法 (如 to_markdown, to_html), 得 2 分。 结果正确, 但使用了		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
			简单的 <code>print(dataframe)</code> , 未指定格式, 得 1 分。 结果错误, 得 0 分。		
M7	2	考评员运行考生提交的 PY 文件, 检查运行过程日志显示包含 “M7 - 评分分布展示状态: success” 且输出包含 <code>rating rating_count</code> 。	结果正确, 且使用了标准的 <code>to_string(index=False)</code> 方法将 DataFrame 转换为无索引的表格字符串, 得 2 分。 结果正确, 但使用了其他格式化输出方法, 得 1 分。 结果错误, 得 0 分。		
合计配分	20	合计得分			

九、金融平台交易记录数据查询与分析系统设计（考核时间：20 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装 BI 类商务数据分析软件、压缩软件、Python 3.8+（`pandas`、`mysql-connector-pytho` 库）、MySQL 8.0+及办公软件；
- 3) 数据库及商务数据分析师考试平台。

（2）工作任务

- 1) 实现按用户标识或时间范围查询交易记录；
- 2) 使用 MySQL 索引提升查询性能；
- 3) 通过 MySQL 和 Python 分析用户交易活跃度或评分分布；
- 4) 在 Python 环境中展示分析结果。

（3）技能要求

- 1) 使用计算引擎优化数据查询效率；

- 2) 通过计算平台构建检索分析系统;
 - 3) 通过计算平台显示检索分析结果。
- (4) 质量指标
- 1) 查询交易记录功能正确实现;
 - 2) 查询效率通过索引优化得到提升;
 - 3) 检索分析系统准确分析用户交易活跃度和评分分布;
 - 4) 分析结果以清晰格式展示。

背景材料

某金融平台需高效管理用户交易记录数据,以支持风险管理、用户行为分析和市场预测。业务场景包括用户通过网页或移动端完成的实时交易(如转账、投资)以及批量导入的历史交易记录。数据来源包括实时交易接口(JSON格式,记录用户实时交易)、历史交易归档(CSV格式,存储历史交易)和本地MySQL数据库(存储交易和用户信息)。数据存在问题,如重复交易、缺失值、时间格式不一致(需统一为YYYY-MM-DD HH:MM:SS),影响查询效率和分析准确性。为实现数据驱动的金融服务,需在局域网环境中构建数据查询与分析系统,涵盖以下功能:查询数据,支持按用户标识或时间范围检索交易记录;使用MySQL索引优化查询效率;通过MySQL和Python构建检索分析系统,分析用户交易活跃度或评分分布;在Python环境中展示分析结果。

操作要求:

- (1) 从 `transaction_data` 表中查询交易记录,要求能分别实现:
按用户标识检索: 通过参数化查询筛选特定 `user_id` 的记录。
按时间范围检索: 使用 `BETWEEN` 语法查询特定 `transaction_time` 区间内的记录。
- (2) 为 `transaction_data` 表的 `user_id` 和 `transaction_time` 字段创建 `idx_user_id` 和 `idx_transaction_time` 索引,并且必须通过先使用 `SHOW INDEX` 检查索引是否存在的方式,来确保索引创建操作的幂等性。
- (3) 执行以下两种独立的SQL聚合分析:
用户交易活跃度分析: 统计 `transaction_data` 表中每个用户的交易总数,并按交易数(`transaction_count`)降序排列,最终返回前10名的用户。
评分分布分析: 统计 `transaction_data` 表中所有非空评分(`rating`)的分布情况,并按评分值升序排列结果。

(4) 将上述两种数据分析得到的 Pandas DataFrame 结果，均使用 to_string(index=False) 方法转换为不包含索引的表格字符串格式进行输出展示。

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示包含“M1 - 查询用户标识数据状态: success”且数据库返回至少 1 条记录。	<p>结果正确，且 SQL 查询使用了基础语句 (WHERE 1=1) 并结合了标准的参数化条件，得 3 分。</p> <p>结果正确，但 SQL 查询语句或参数化方式与标准实现不完全一致 (如未使用 WHERE 1=1)，得 2 分。</p> <p>结果正确，但使用了有安全风险的字符串拼接方式传递参数，得 1 分。</p> <p>结果错误，得 0 分。</p>		
M2	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示包含“M2 - 查询时间范围数据状态: success”且数据库返回至少 1 条记录。	<p>结果正确，且使用了 BETWEEN %s AND %s 语法进行时间范围查询，得 3 分。</p> <p>结果正确，但使用了其他逻辑正确的 SQL 语法 (如 >= %s AND <= %s)，得 2 分。</p> <p>结果正确，但使用了有安全风险的字符串拼接方式传递参数，得 1 分。</p> <p>结果错误，得 0 分。</p>		
M3	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示包含“M3 - 索引创建状态: success”且索引存在。	<p>结果正确，且使用了先检查索引是否存在 (SHOW INDEX) 再执行创建的完整逻辑，得 3 分。</p> <p>结果正确，但未使用 SHOW INDEX 检查索引是否存在，得 2 分。</p> <p>结果正确，但使用了有安全风险的字符串拼接方式传递参数，得 1 分。</p> <p>结果错误，得 0 分。</p>		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
			<p>辑，确保了代码的幂等性，得 3 分。</p> <p>结果正确，但仅使用 <code>try...except</code> 捕获错误来处理索引已存在的情况，得 2 分。</p> <p>结果正确，但仅执行 <code>CREATE INDEX</code> 语句，未处理索引已存在的情况，得 1 分。</p> <p>结果错误，得 0 分。</p>		
M4	3	<p>考评员运行考生提交的 PY 文件，检查运行过程日志显示包含“M4 - 用户交易活跃度分析状态: success”且结果正确，程序输出的前 10 条活跃用户数据结果正确，即输出包含字段 <code>user_id</code> 和 <code>transaction_count</code>。</p>	<p>结果正确，且 SQL 查询语句完整运用了 <code>GROUP BY, ORDER BY DESC, 和 LIMIT 10</code>，得 3 分。</p> <p>结果正确，但 SQL 查询缺少部分排序或限制条件（如仅 <code>GROUP BY</code>），得 2 分。</p> <p>结果正确，但使用了低效的 Python 循环处理代替 SQL 聚合，得 1 分。</p> <p>结果错误，得 0 分。</p>		
M5	3	<p>考评员运行考生提交的 PY 文件，检查运行过程日志显示包含“M5 - 评分分布分析状态: success”且结果正确，程序输出的评分分布数据结果正确，即输出包含字段 <code>rating</code> 和 <code>rating_count</code>，且统计值与数据库真实数据一致。</p>	<p>结果正确，且 SQL 查询语句完整运用了 <code>WHERE rating IS NOT NULL, GROUP BY, 和 ORDER BY</code>，得 3 分。</p> <p>结果正确，但 SQL 查询缺少部分条件（如未排除 <code>NULL</code> 值或未排序），得 2 分。</p> <p>结果正确，但使用了低效的 Python 循环处理代替 SQL 聚合，得 1 分。</p> <p>结果错误，得 0 分。</p>		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M6	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示包含“M6 - 用户交易活跃度展示状态：success”且输出包含 user_id transaction_count。	<p>结果正确，且使用了标准的 to_string(index=False) 方法将 DataFrame 转换为无索引的表格字符串，得 3 分。</p> <p>结果正确，但使用了其他 Pandas 输出方法（如 to_markdown, to_html），得 2 分。</p> <p>结果正确，但使用了简单的 print(dataframe)，未指定格式，得 1 分。</p> <p>结果错误，得 0 分。</p>		
M7	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示包含“M7 - 评分分布展示状态：success”且输出包含 rating rating_count。	<p>结果正确，且使用了标准的 to_string(index=False) 方法将 DataFrame 转换为无索引的表格字符串，得 2 分。</p> <p>结果正确，但使用了其他格式化输出方法，得 1 分。</p> <p>结果错误，得 0 分。</p>		
合计配分	20	合计得分			

十、医疗平台患者记录数据查询与分析系统设计（考核时间：20 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装 BI 类商务数据分析软件、压缩软件、Python 3.8+（`pandas`、`mysql-connector-pytho` 库）、MySQL 8.0+ 及办公软件；
- 3) 数据库及商务数据分析师考试平台。

(2) 工作任务

- 1) 实现按用户标识或时间范围查询患者记录；
- 2) 使用 MySQL 索引提升查询性能；
- 3) 通过 MySQL 和 Python 分析患者活跃度或评分分布；
- 4) 在 Python 环境中展示分析结果。

(3) 技能要求

- 1) 使用计算引擎优化数据查询效率；
- 2) 通过计算平台构建检索分析系统；
- 3) 通过计算平台显示检索分析结果。

(4) 质量指标

- 1) 查询患者记录功能正确实现；
- 2) 查询效率通过索引优化得到提升；
- 3) 检索分析系统准确分析患者活跃度和评分分布；
- 4) 分析结果以清晰格式展示。

背景材料

某医疗平台需高效管理患者记录数据，以支持诊疗优化、患者管理和健康分析。业务场景包括患者通过线上或线下方式生成的实时记录（如就诊、化验）以及批量导入的历史患者记录。数据来源包括实时记录接口（JSON 格式，记录患者实时数据）、历史归档文件（CSV 格式，存储历史记录）和本地 MySQL 数据库（存储患者记录和用户信息）。数据存在问题，如重复记录、缺失值、时间格式不一致（需统一为 YYYY-MM-DD HH:MM:SS），影响查询效率和分析准确性。为实现数据驱动的医疗服务，需在局域网环境中构建数据查询与分析系统，涵盖以下功能：查询数据，支持按用户标识或时间范围检索患者记录；使用 MySQL 索引优化查询效率；通过 MySQL 和 Python 构建检索分析系统，分析患者活跃度或评分分布；在 Python 环境中展示分析结果。

操作要求：

(1) 从 `patient_data` 表中查询患者记录，要求能分别实现：

按用户标识检索：通过参数化查询筛选特定 `user_id` 的记录。

按时间范围检索：使用 `BETWEEN` 语法查询特定 `record_time` 区间内的记录。

(2) 为 `patient_data` 表的 `user_id` 和 `record_time` 字段创建 `idx_user_id` 和 `idx_record_time` 索引，并且必须通过先使用 `SHOW INDEX` 检查索引是否存在

的方式，来确保索引创建操作的幂等性。

(3) 执行以下两种独立的 SQL 聚合分析：

患者活跃度分析：统计 `patient_data` 表中每个用户的记录总数，并按记录数 (`record_count`) 降序排列，最终返回前 10 名的用户。

评分分布分析：统计 `patient_data` 表中所有非空评分 (`rating`) 的分布情况，并按评分值升序排列结果。

(4) 将上述两种数据分析得到的 `Pandas DataFrame` 结果，均使用 `to_string(index=False)` 方法转换为不包含索引的表格字符串格式进行输出展示。

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示包含“M2 - 查询时间范围数据状态: success”且数据库返回至少 1 条记录。	<p>结果正确，且 SQL 查询使用了基础语句 (<code>WHERE 1=1</code>) 并结合了标准的参数化条件，得 3 分。</p> <p>结果正确，但 SQL 查询语句或参数化方式与标准实现不完全一致（如未使用 <code>WHERE 1=1</code>），得 2 分。</p> <p>结果正确，但使用了有安全风险的字符串拼接方式传递参数，得 1 分。</p> <p>结果错误，得 0 分。</p>		
M2	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示包含“M2 - 查询时间范围数据状态: success”且数据库返回至少 1 条记录。	<p>结果正确，且使用了 <code>BETWEEN %s AND %s</code> 语法进行时间范围查询，得 3 分。</p> <p>结果正确，但使用了其他逻辑正确的 SQL</p>		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
			语法(如 <code>>= %s AND <= %s</code>), 得 2 分。 结果正确, 但使用了有安全风险的字符串拼接方式传递参数, 得 1 分。 结果错误, 得 0 分。		
M3	3	考评员运行考生提交的 PY 文件, 检查运行过程日志显示包含 “M3 - 索引创建状态: success” 且索引存在。	结果正确, 且使用了先检查索引是否存在 (SHOW INDEX) 再执行创建的完整逻辑, 确保了代码的幂等性, 得 3 分。 结果正确, 但仅使用 <code>try...except</code> 捕获错误来处理索引已存在的情况, 得 2 分。 结果正确, 但仅执行 CREATE INDEX 语句, 未处理索引已存在的情况, 得 1 分。 结果错误, 得 0 分。		
M4	3	考评员运行考生提交的 PY 文件, 检查运行过程日志显示包含 “M4 - 患者活跃度分析状态: success” 且结果正确, 程序输出的前 10 条活跃用户数据结果正确, 即输出包含字段 <code>user_id</code> 和 <code>record_count</code> 。	结果正确, 且 SQL 查询语句完整运用了 GROUP BY, ORDER BY DESC, 和 LIMIT 10, 得 3 分。 结果正确, 但 SQL 查询缺少部分排序或限制条件 (如仅 GROUP BY), 得 2 分。 结果正确, 但使用了低效的 Python 循环处理代替 SQL 聚合, 得 1 分。 结果错误, 得 0 分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M5	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示包含“M5 - 评分分布分析状态: success”且结果正确，程序输出的评分分布数据结果正确，即输出包含字段 rating 和 rating_count，且统计值与数据库真实数据一致。	<p>结果正确，且 SQL 查询语句完整运用了 WHERE rating IS NOT NULL, GROUP BY, 和 ORDER BY, 得 3 分。</p> <p>结果正确，但 SQL 查询缺少部分条件（如未排除 NULL 值或未排序），得 2 分。</p> <p>结果正确，但使用了低效的 Python 循环处理代替 SQL 聚合，得 1 分。</p> <p>结果错误，得 0 分。</p>		
M6	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示包含“M6 - 患者活跃度展示状态: success”且输出包含 user_id record_count。	<p>结果正确，且使用了标准的 to_string(index=False) 方法将 DataFrame 转换为无索引的表格字符串，得 3 分。</p> <p>结果正确，但使用了其他 Pandas 输出方法（如 to_markdown, to_html），得 2 分。</p> <p>结果正确，但使用了简单的 print(dataframe), 未指定格式，得 1 分。</p> <p>结果错误，得 0 分。</p>		
M7	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示包含“M7 - 评分分布展示状态: success”且输出包含 rating rating_count。	<p>结果正确，且使用了标准的 to_string(index=False) 方法将 DataFrame 转换为无索引的表格字符串，得 2 分。</p> <p>结果正确，但使用了</p>		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
			其他格式化输出方法，得1分。 结果错误，得0分。		
合计配分	20	合计得分			

数据挖掘和分析

一、执行费用数据标准化处理与建模分析（考核时间：15 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装数据分析软件、压缩软件及办公软件；
- 3) 数据库及商务数据分析师考试平台。

（2）工作任务

- 1) 根据企业多源异构数据，完成多源数据的整合工作；
- 2) 根据某公司数据，完成金额数据的规范化格式处理任务；
- 3) 根据业务要求，完成预算费用数据的规约处理任务；
- 4) 根据维度建模理论，完成数据表关系的分析与建立；
- 5) 根据业务需求，完成查询重构及模型的优化工作。

（3）技能要求

- 1) 根据相关数据表特点，进行多源数据的整合；
- 2) 根据业务要求，对金额数据格式进行精准转换；
- 3) 根据数据归一性原则，开展预算费用数据的规约；
- 4) 参照数据建模标准规范，设计星型分析模型；
- 5) 根据业务需求，对数据进行重构并优化数据模型。

（4）质量指标

- 1) 准确识别多源数据表的关键字段，正确完成数据匹配与整合，确保信息完整

无误；

- 2) 正确执行金额数据格式转换，精确实现小数点后三位标准化处理；
- 3) 准确筛选关键数据特征，正确完成数据规约处理；
- 4) 正确分析数据表间逻辑关系，准确建立符合规范的星型模型；
- 5) 准确理解业务需求，正确优化查询逻辑，精确提升模型性能。

背景材料

商飞集团作为综合性企业，在日常运营中产生多源异构数据，包括 CSV 和 Excel 等格式的费用明细记录和成本中心数据等。其中，成本中心数据表详细记录了各分店、城市及对应成本中心的关联信息；费用科目体系采用分级编码管理，涵盖包括咖啡豆、牛奶等原材料在内的材料费、设备维护费、市场营销费、物业费及薪酬福利等类别。基于北京、上海和杭州三地近两年的季度数据，通过数据预处理和建模分析技术，对各区域分店的预算执行情况进行多维度评估，重点分析材料成本波动、费用管控效率等关键指标，为集团精细化管理和资源优化配置提供数据支持。

操作要求：

- (1) 利用 SQL、MySQL、ETL、BI 等数据分析工具,实现“费用数据”多级科目分类属性与以及“费用明细”中全部数据的精准关联匹配，完成多源异构数据的深度融合与统一管理；
- (2) 将费用明细表中的各项费用支出统一转换为人民币计量单位，且确保数据精度达到千分位，为财务数据分析提供可靠的数据基础；
- (3) 在预算表中通过 Min-Max 归一化算法，构建“总费用_归一化”衍生指标，对预算数据实施特征缩放，以量化费用波动的相对幅度；
- (4) 运用维度建模理论，针对“费用数据”、“费用明细”以及“成本中心数据”这三个数据源中的所有数据表，构建星型模型数据架构；
- (5) 实施基于业务语义的水平分片策略，按区域属性对预算表进行物理分区，生成北京/上海专属子集，并通过优化模型重构维度关联逻辑。（请提交操作完成后的数据文件至考试平台。）

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	2	正确对多源异构数据进行横向整合	正确加载并整合=2分； 仅正确加载=1分； 正确加载任意1个数据源或仅正确进行整合分析=0.5分； 错误=0分。		
M2	2	正确转换数据格式	正确转换2处=2分； 正确转换1处=1分； 错误=0分。		
M3	2	正确对数据进行 Min-Max 归一化处理	是=2分；否=0分。		
M4	2	正确建立关系模型	正确建立3处关系=2分； 建立2处关系=1分； 建立1处或错误=0分。		
M5	2	正确处理表并完善表间关系，以优化模型	正确编辑表并完善表间关系=2分； 仅正确编辑表=1分； 错误=0分。		
合计配分	10	合计得分			

二、开展客户数据预处理并构建关系模型（考核时间：15 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装数据分析软件、压缩软件及办公软件；
- 3) 数据库及商务数据分析师考试平台。

（2）工作任务

- 1) 根据某公司多源异构客户数据，完成多源数据的整合工作；
- 2) 根据某公司客户信息数据表，完成数据格式转换的任务；
- 3) 根据业务要求，完成客户信息表的数据规约处理任务；
- 4) 根据公司需求，完成数据表关系的分析与管理，并建立数据模型；
- 5) 根据相关数据，完成新表创建及模型优化工作。

（3）技能要求

- 1) 能根据相关数据特点，进行多源数据的整合；

- 2) 能根据业务规则, 对数据格式进行精准转换;
- 3) 能根据数据归一性原则, 开展性别和等级数据的规约;
- 4) 能参照数据建模标准规范设计数据雪花模型;
- 5) 能根据业务需求, 新建订单总额表并优化原有数据模型。

(4) 质量指标

- 1) 准确分析并筛选相关数据, 正确整合多源配送信息表;
- 2) 准确分析数据特点, 正确转换客户信息表的数据格式;
- 3) 正确分析年龄数据, 正确对数据进行规约处理;
- 4) 准确分析数据表间关系, 正确建立符合企业需求的雪花模型;
- 5) 正确新建订单总额表, 准确优化原有数据模型。

背景材料

海东集团作为零售行业头部企业, 当前面临客户数据资产价值挖掘的迫切需求。由于业务系统历经多年迭代, 客户信息分散在订单管理、会员服务、等级信息等多个异构系统中, 导致数据维度定义不统一, 例如性别字段存在“男/女”“M/F”等多种表示方式。为解决这些数据孤岛与质量问题, 该集团开展了客户数据治理任务, 通过整合多源数据构建标准化客户信息表, 将原始等级编码、性别等标识映射为统一标准, 并基于年龄数据建立分层分析维度, 为后续业务场景提供高质量数据支持。

操作要求:

- (1) 利用 SQL、MySQL、ETL、BI 等数据分析工具, 将“信息”和“客户等级”异构数据源聚合成一个包含基础信息与相应会员等级的“客户信息表”, 实施多维度的数据治理, 形成客户全景信息;
- (2) 将客户信息表中的数值型等级标识 1/2/3 级映射为商业语义等级青铜/白银/黄金; 同步转换性别标识为国际通用符号 F、M, 提升数据规范性与可解析性;
- (3) 对客户信息表中连续型年龄变量执行智能分箱操作, 生成 50 岁以上、41 至 49 岁等五个分析适配区间, 建立适配精准营销的离散特征;
- (4) 基于维度建模理论, 针对“客户信息表”、“信息”、“客户等级”和“订单数据”构建支持弹性扩展的雪花分析模型, 实现多维度多粒度的数据洞察需求;
- (5) 新增扩展表用于展示各订单的总金额。为优化原有数据模型, 通过外键关联机制构建该表与关联表间关系, 构建逻辑清晰的关系模型。(请提交操作完成后的数据文件至考试平台。)

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	1.5	正确对多源异构数据进行整合	正确聚合数据源并整合=1.5分；仅正确聚合=1分；正确聚合任意1个数据源或仅正确整合=0.5分；错误=0分。		
M2	0.5	正确重命名整合数据	是=0.5分；否=0分。		
M3	2	正确转换数据格式	正确转换2处=2分；正确转换1处=1分；错误=0分。		
M4	2	正确对数据进行规约	是=2分；否=0分。		
M5	1	正确建立关系模型	正确建立全部4处关系=1分；正确建立2-3处关系=0.5分；建立1处关系或错误=0分。		
M6	2	正确建立新表	是=2分；否=0分。		
M7	1	正确建立新建表与关联表间关系，以优化原有模型	正确新建表并完善表间关系=2分；正确新建表但关系不全或错误=1分；未建新表或错误=0分。		
合计配分	10	合计得分			

三、预处理培训数据并优化模型（考核时间：15 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装数据分析软件、压缩软件及办公软件；
- 3) 数据库及商务数据分析师考试平台。

（2）工作任务

- 1) 根据某公司数据，完成多源数据的整合工作；

- 2) 根据某公司持证情况表，完成性别数据格式转换处理任务；
- 3) 根据业务要求，完成档案表的数据规约处理任务；
- 4) 根据数据特点，完成数据表关系的分析与管理；
- 5) 根据业务需求，完成对所建立模型的针对性优化工作。

(3) 技能要求

- 1) 根据相关数据特点，进行多源数据的整合；
- 2) 根据业务要求，对性别数据格式进行精准转换；
- 3) 根据数据归一性原则，开展相关数据的规约；
- 4) 参照数据建模标准规范，设计星型模型；
- 5) 根据业务需求，对数据模型进行针对性优化。

(4) 质量指标

- 1) 准确分析并筛选相关数据，正确整合多源配送信息表；
- 2) 准确分析数据特点，正确转换性别的数据格式；
- 3) 准确把握相关数据，正确对数据进行规约处理；
- 4) 正确分析数据表间关系，建立符合数据建模规范的星型模型；
- 5) 准确分析业务要求，优化数据模型，正确建立完整的关系。

背景材料

当下制造业对安全生产与员工技能要求趋严，华锐公司生产涉及叉车、电工、高处作业、天车等特种作业场景，各环节紧密关联且安全风险不容小觑。员工是否持有有效特种作业证件，直接影响生产安全与效率。目前，公司需摸清员工叉车、电工、高处作业、天车等证件的持有情况，包括是否持证、证件类型等，避免无证上岗。同时，要梳理员工培训时间安排是否合理，防止因时间冲突降低参与度；分析培训项目内容是否贴合实际工作，有无涵盖最新行业规范；明确培训处于进行中、未达标完成还是已达标等状态。为此，员工张风将对收集上来的数据进行处理，以便公司及时安排复审或考证工作，杜绝因证件过期导致关键岗位人手“空缺”，打乱既定生产计划与进度。

操作要求：

(1) 利用 SQL、MySQL、ETL、BI 等数据分析工具，构建四类特种作业证书叉车、电工、高处作业、天车的标准化“持证情况”表，实现多源证书信息的语义级融合与结构化聚合；

(2) 针对持证情况表中的性别标识 0 与 1，实施业务语义驱动的离散值映射，将数值编码自动化转换为“男”和“女”可读性标签，增强数据可读性；

(3) 对“培训项目情况”中档案表进行多条件过滤，当员工同时满足在职状态、应在培训名单中、应在体检名单中三重条件时，自动生成“体检、培训”备注标签，其余记录保持默认无标注状态；

(4) 对异构数据源“叉车证件”、“证书情况”、“培训项目情况”以及持证情况表进行建模分析，并剔除冗余关联关系，提升模型轻量化水平；

(5) 通过数据特征关联分析，挖掘培训时间与培训项目表、培训状态表的潜在联系，构建两表之间的外键关联关系，完善及优化数据模型。（请提交操作完成后的数据文件至考试平台。）

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	1.5	正确对多源异构数据进行整合	正确聚合并整合=1.5分；仅正确聚合=1分；正确聚合任意1个数据源或仅正确整合=0.5分；错误=0分。		
M2	0.5	正确重命名整合数据	是=0.5分；否=0分。		
M3	2	正确转换数据格式	正确转换2处=2分；正确转换1处=1分；错误=0分。		
M4	2	正确对数据进行规约	是=2分；否=0分。		
M5	2	正确建立关系模型	正确建立7处关系=2分；建立4-6处关系=1分；建立3处及以下关系或错误=0分。		
M6	2	正确处理表并完善表间关系，以优化模型	正确处理表并完善表间关系=2分；仅正确处理表=1分；错误=0分。		
合计配分	10	合计得分			

四、开展销售数据预处理并构建雪花模型（考核时间：15 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装数据分析软件、压缩软件及办公软件；
- 3) 数据库及商务数据分析师考试平台。

（2）工作任务

- 1) 根据某公司多源数据，完成多源数据的整合工作；
- 2) 根据某公司销售数据表，完成金额数据格式转换的任务；
- 3) 根据业务要求，完成新建产品信息表的数据规约处理任务；
- 4) 根据公司数据，完成数据模型的建立及优化任务。

（3）技能要求

- 1) 根据相关数据特点，进行多源数据的整合；
- 2) 根据业务规则，对数据格式进行精准转换；
- 3) 根据数据归一性原则，开展季度和价格数据的规约；
- 4) 参照数据建模标准规范和业务要求，设计雪花模型及优化数据模型。

（4）质量指标

- 1) 准确分析并筛选相关数据，正确整合多源数据表；
- 2) 准确分析数据特点，正确转换金额的数据格式；
- 3) 精准把握业务需求，正确对数据进行规约处理；
- 4) 准确分析数据表间关系，正确建立符合建模标准规范的雪花模型，准确优化数据模型。

背景材料

在当下市场环境日趋复杂、消费者需求碎片化特征加剧的竞争态势下，飞跃公司亟需借助数据驱动模式实现业务增长突破。公司积累的多源数据中，结构化表格数据以货号、年份、上市时间、类别、系列、产品名称、价格等字段为维度，完整呈现产品从研发到退市的全生命周期轨迹，销售与收款明细则精准记录交易动态。非结构化文本数据潜藏着客户编码、客户名称、所在城市等极具价值的信息。目前，公司正积极推动数据驱动战略深度落地，以多源数据融合为引擎驱动业务革新，通过将结构化数据中的产品信息与非结构化文本里的消费需求洞察精准对接，开展深度数据分析，为产品迭代升级、营销策略优化提供切实可行的数

据方案，助力公司在碎片化需求市场中突围。

操作要求：

(1) 利用 SQL、MySQL、ETL、BI 等数据分析工具，整合多源系统中的近三年销售数据，构建覆盖产品全生命周期的标准化“产品信息”表，实现多维度基础数据的聚合与统一管理；

(2) 对销售明细的货币型字段实施货币归一化处理，统一转换为精确到元级的人民币基准单位，以增强数据的规范性与可解析性；

(3) 采用时序数据离散化技术，将产品信息表中的产品上市时间进行结构化转换，生成“第 X 季度”形式的标准化时间维度标签，以支持产品生命周期阶段性的趋势分析与季节性特征挖掘；

(4) 基于产品信息表的价格分布特征，采用 Jenks 自然断点法构建包含 500 元以下、500-1000 元等四级价格分层体系，实现价格区间的科学划分；

(5) 对“产品信息”、“客户信息表”和近三年销售数据进行建模分析与优化，采用雪花分析模型架构并剔除冗余关联关系，满足多维度、多粒度数据洞察需求。
(请提交操作完成后的数据文件至考试平台。)

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	1.5	正确对多源异构数据进行整合	正确加载并整合=1.5分；仅正确加载数据=1分；正确加载任意1个数据源或仅正确整合=0.5分；错误=0分。		
M2	0.5	正确重命名整合数据	是=0.5分；否=0分。		
M3	1	正确转换数据格式	正确完成4处格式转换=1分；正确转换2-3处=0.5分；正确转换1处或错误=0分。		
M4	2	正确对数据进行规约	正确生成数据列=2分；仅正确生成数据列但表达形式错误=1分；错误=0分。		
M5	2	正确对数据进行处理	是=2分；否=0分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M6	3	建立关系模型，以优化数据模型	正确建立 7 处关系=3 分；建立 7 处关系但仍存在非活动状态关联=2.5 分；建立 4-5 处关联且无非活动状态关联=2 分；建立 4-5 处关联但仍存在非活动状态关联=1.5 分；建立 2-3 处关联且无非活动状态关联=1 分；建立 2-3 处关联但仍存在非活动关联=0.5 分；建立 1 处关联或错误=0 分。		
合计配分	10	合计得分			

五、开展数据预处理与数据建模（考核时间：15 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装数据分析软件、压缩软件及办公软件；
- 3) 数据库及商务数据分析师考试平台。

（2）工作任务

- 1) 根据某公司多源配送信息表，完成多源数据的整合工作；
- 2) 根据某公司配送信息数据表，完成格式转换处理的任务；
- 3) 根据业务要求，完成新配送信息表的数据规约处理任务；
- 4) 根据公司业务需求，完成数据表关系的分析与管理，并建立数据模型；
- 5) 根据分析模型中呈现的问题，完成对所建立模型的针对性优化工作。

（3）技能要求

- 1) 能根据相关数据特点，进行多源数据的整合；
- 2) 能根据业务规则，对数据格式进行精准转换；
- 3) 能根据数据归一性原则，开展满意度数据的规约；
- 4) 能参照数据建模标准规范设计数据雪花模型；
- 5) 能根据业务需求，对数据模型进行针对性优化。

（4）质量指标

- 1) 准确分析并筛选相关数据，正确整合多源配送信息表；
- 2) 准确分析数据特点，正确转换配送信息表的数据格式；
- 3) 正确分析各满意度，正确对数据进行规约处理；
- 4) 正确分析数据表间关系，建立符合企业需求的数据模型；
- 5) 精准捕捉并处理数据模型间的问题，优化数据模型，正确建立完整的关系。

背景材料

上海天英商贸集团，作为物流配送行业的标杆企业，始终致力于通过数据驱动决策以提升服务品质与运营效率。为更精准地洞察配送业务动态、优化服务策略体系，集团系统化地收集并整合了近阶段的配送业务详情数据。该数据集以高度结构化的形式，详尽记录了配送员核心信息、配送站点分布、配送区域覆盖范围、服务客户数量统计，以及客户针对配送服务质量、货物完好度及沟通效能所给出的多维度评分反馈。评分体系采用国际通行的 5 分制标准，其中 5 分代表着客户对服务体验的高度认可与“非常满意”的极致评价，是衡量集团服务质量的关键指标。现阶段，基于该数据集开展全面的数据预处理与建模工作，旨在通过科学的分析方法挖掘数据价值，为后续的精准确策、服务优化及战略规划提供坚实的数据支撑。

操作要求：

- (1) 利用 SQL、MySQL、ETL、BI 等数据分析工具,整合所有月份配送业务产生的结构化与非结构化数据，统一归集至“各月配送信息表”，形成时序化的配送业务数据；
- (2) 针对各月配送信息表中的服务态度、货物完好度、及时沟通三项指标的 1-5 分量化评分，实施语义映射转换，将分值映射为非常不满意、较不满意、一般、较满意、非常满意五级标签，提升数据的业务可读性与分析直观性；
- (3) 基于三项满意度语义化标签的组合特征，构建多条件判断逻辑，当三项指标同时为“非常满意度”时，自动生成“继续保持”的正面业务备注；其余情况统一标注为“需要提升”的改进建议标签；
- (4) 对“各月配送信息表”、“2025 年 1 月”、“2025 年 2 月”和“2024 年配送信息”异构数据源进行建模分析，构建层次分明、逻辑清晰的雪花型分析模型架构；
- (5) 通过数据表结构优化和关键路径筛选，构建低冗余、高性能的轻量化数据关联体系，实现所有数据表间关联的全面优化。（请提交操作完成后的数据文件至考试平台。）

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	1.5	正确对多源异构数据进行整合	正确聚合数据源并整合=1.5分；仅正确聚合=1分；正确聚合任意1个数据源或仅正确整合=0.5分；错误=0分。		
M2	0.5	正确重命名整合数据	是=0.5分；否=0分。		
M3	2	正确转换数据格式	正确转换5处=2分；正确转换3-4处=1分；正确转换2处=0.5分；正确转换1处或错误=0分。		
M4	2	正确对满意度进行规约	是=2分；否=0分。		
M5	2	正确建立关系模型	正确建立5个关系=2分；正确建立3-4个关系=1分；正确建立2个关系=0.5分；正确建立1个或错误=0分。		
M6	2	正确编辑表并完善和优化数据模型	正确编辑表并完善模型和剔除非活动关系=2分；仅编辑表或仅未剔除非活动关系=1分；错误=0分。		
合计配分	10	合计得分			

六、电信客户流失预测考核（考核时间：15 min）

1. 试题单

（1）场地设备要求

1) 装有 Windows 操作系统的计算机 (Windows 7 及以上), CPU 4 核, 内存 8GB, 硬盘 50GB 可用空间;

2) 系统中已安装 BI 类商务数据分析软件、压缩软件、Python 3.8+ (安装 pandas、numpy、scikit-learn、matplotlib、seaborn 库)、MySQL 8.0+ 及办公软件;

3) 商务数据分析师考试平台。

(2) 工作任务

- 1) 计算月均账单金额和高账单低通话标志；
- 2) 处理年龄缺失值，标准化通话分钟数和账单金额；
- 3) 训练逻辑回归模型，计算准确率，生成通话分钟数分布图和混淆矩阵；
- 4) 生成 HTML 报告，包含特征计算结果、分布图、模型评估结果、混淆矩阵和业务洞察。

(3) 技能要求

- 1) 根据业务需求编写批量数据计算作业并优化作业参数；
- 2) 根据业务规则设计相应数据指标计算算法；
- 3) 评估挖掘需求并使用工具对数据进行特征工程处理；
- 4) 调用常规模型进行模型训练；
- 5) 基于评价指标进行模型测试。

(4) 质量指标

- 1) 高效准确提取月均账单金额和高账单低通话标志；
- 2) 完整规范处理年龄缺失值并标准化特征；
- 3) 逻辑回归模型精准分类客户流失风险；
- 4) HTML 报告清晰呈现特征计算结果、模型性能和图表。

背景材料

电信行业客户流失是企业面临的重要挑战，流失客户会导致收入下降和市场份额减少。通过分析客户行为数据（如账单金额、通话分钟数等），企业可以识别高风险客户并采取挽留措施。本任务基于 10,000 条电信客户数据，包含客户 ID（customer_id）、账单金额（bill_amount，年总金额，元）、通话分钟数（call_minutes，年总分钟数）、年龄（age，部分缺失）、是否流失（churn，0=未流失，1=流失）等字段。目标是构建预测模型，识别可能流失的客户，并生成分析报告。

操作要求：

(1) 在 `compute_features` 函数中，计算月均账单金额，确保 `月均账单金额` 列被正确创建，其值为 `账单金额` 列除以 12。

(2) 在 `compute_features` 函数中，使用向量化操作计算高账单低通话标志，确保 `高账单低通话` 列被正确创建，其逻辑为当 `账单金额` 大于均值且 `通话分钟数` 小于均值时为 1，否则为 0。

(3) 在 `feature_engineering` 函数中，填补 `年龄` 列的缺失值，确保该列最终无

任何缺失数据，并优先使用中位数或均值进行填补。

(4) 在 `feature_engineering` 函数中，对 通话分钟数 和 账单金额 两列进行标准化处理，确保处理后的新列均值约等于 0 且标准差约等于 1。

(5) 在 `train_and_evaluate` 函数中，计算并返回逻辑回归模型在测试集上的准确率。

(6) 在 `train_and_evaluate` 函数中，生成并保存混淆矩阵热图至 `confusion_matrix.png`，该图必须为蓝色调 (`cmap='Blues'`)，在每个单元格中标注具体数值，并确保中文标题 (“混淆矩阵”) 与坐标轴标签 (“预测值”、“真实值”) 显示正常无乱码。

(7) 在 `train_and_evaluate` 函数中，生成并保存通话分钟数分布的直方图至 `call_minutes_distribution.png`，该图必须包含 30 个条柱 (`bins`)，条柱颜色为 `skyblue` 且带有黑色边框，并确保中文标题 (“通话分钟数分布”) 与坐标轴标签 (“通话分钟数”、“频次”) 显示正常无乱码。

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M1-月均账单金额计算成功”“M1-示例值:[值]”；验证 <code>processed_data.csv</code> 包含正确月均账单金额列 (账单金额/12)。	结果正确，且使用了 <code>df['bill_amount'] / 12</code> 的标准计算方法，得 2 分。 结果正确，但未使用 <code>df['bill_amount'] / 12</code> 的标准计算方法，得 1 分。 结果错误，得 0 分。		
M2	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M2-高账单低通话标志计算成功”“M2-示例值:[0 或 1]”；验证 <code>processed_data.csv</code> 包含正确高账单低通话列 (账单金额 > 均值且通话分钟数 < 均	结果正确，且使用了布尔掩码 <code>astype(int)</code> 或 <code>np.where</code> 的向量化方法，得 2 分。 结果正确，但使用了循环判断的低效方法，得 1 分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
		值)。	结果错误, 得 0 分。		
M3	2	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M3-年龄缺失值填补成功”“M3-填补记录数:[值]” (记录填补前缺失值数量, 约 500 条); 验证 processed_data.csv 的年龄列无缺失值。	结果正确, 且使用了中位数或均值进行填补, 得 2 分。 结果正确, 但使用了众数或固定值进行填补, 得 1 分。 结果错误, 得 0 分。		
M4	1	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M4-特征标准化成功”“M4-通话分钟数均值:[≈0]”“M4-账单金额标准差:[≈1]”; 验证 processed_data.csv 的通话分钟数和账单金额列均值 ≈0、标准差 ≈1 (接受 StandardScaler 或手动标准化等方法)。	结果正确, 且使用了 StandardScaler 或等效的手动标准化方法, 得 1 分。 结果正确, 但使用了其他缩放方法 (如 MinMaxScaler), 得 0.5 分。 结果错误, 得 0 分。		
M5	1	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M5-模型准确率计算成功”“M5-准确率:[值]”; 验证准确率值在合理范围内 (0.0-1.0)。	结果正确, 且使用了 accuracy_score 函数, 得 1 分。 结果正确, 但使用了手动计算 (如 np.sum(y_test == y_pred) / len(y_test)) 的方法, 得 0.5 分。 结果错误, 得 0 分。		
M6	1	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M6-混淆矩阵热图生成成功”png 文件存在, 为蓝色调, 带数字标注, 中文标题	结果正确, 且完全满足所有格式要求 (蓝色调、带数字标注), 得 1 分。 结果正确, 但未完全		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
		（“混淆矩阵”）和轴标签（“预测值”“真实值”）无乱码。	满足所有格式要求（如未使用蓝色调、或缺少数字标注），得 0.5 分。 结果错误（如图未生成或中文乱码），得 0 分。		
M7	1	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M7-通话分钟数分布图生成成功”；png 文件存在，为直方图（30 个 bins,skyblue 颜色，黑色边框），中文标题（“通话分钟数分布”）和轴标签（“通话分钟数”“频次”）无乱码。	结果正确，且完全满足所有格式要求（30 个 bins、skyblue 颜色、黑色边框），得 1 分。 结果正确，但未完全满足所有格式要求（如 bins 数量不对、或颜色/边框不符），得 0.5 分。 结果错误（如图未生成或中文乱码），得 0 分。		
合计配分	10	合计得分			

七、电商平台销量预测考核（考核时间：15 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装 BI 类商务数据分析软件、压缩软件、Python 3.8+（安装 pandas、numpy、scikit-learn、matplotlib、seaborn 库）、MySQL 8.0+ 及办公软件；
- 3) 数据库及商务数据分析师考试平台。

（2）工作任务

- 1) 计算月均销量和促销期间销量；
- 2) 处理价格缺失值，转换促销活动为整数；
- 3) 训练 XGBoost 模型，计算 RMSE，生成销量分布图和预测散点图；
- 4) 生成 HTML 报告，包含特征计算结果、分布图、模型评估结果、散点图和业务洞察。

(3) 技能要求

- 1) 根据业务需求编写批量数据计算作业并优化作业参数；
- 2) 根据业务规则设计相应数据指标计算算法；
- 3) 评估挖掘需求并使用工具对数据进行特征工程处理；
- 4) 调用常规模型进行模型训练；
- 5) 基于评价指标进行模型测试。

(4) 质量指标

- 1) 高效准确提取月均销量和促销期间销量；
- 2) 完整规范处理价格缺失值并转换数据类型；
- 3) XGBoost 模型精准预测商品销量；
- 4) HTML 报告清晰呈现特征计算结果、模型性能和图表。

背景材料

电商平台通过分析销售数据优化定价和促销策略，以提升销量和市场竞争力。继电信客户流失预测聚焦分类任务后，本任务转向回归分析，基于 10,000 条电商销售数据，包含商品 ID (`product_id`)、价格 (`price`, 单位: 元, 部分缺失)、促销活动 (`promotion`, 0=无促销, 1=有促销) 和销量 (`sales`, 月度销量) 等字段。目标是构建销量预测模型，分析促销和价格对销量的影响，并生成分析报告，为平台决策提供支持。

操作要求:

(1) 在 `compute_features` 函数中，计算月均销量，确保 `月均销量` 列被正确创建，其值为 `sales` 列除以 12。

(2) 在 `compute_features` 函数中，使用向量化操作 (如 `apply` 或 `np.where`) 计算促销期间销量，确保 `促销期间销量` 列被正确创建，其逻辑为当 `promotion` 值为 1 时，取促销商品的平均销量，否则取所有商品的平均销量。

(3) 在 `feature_engineering` 函数中，填补 `价格` 列的缺失值，确保该列最终无任何缺失数据，并必须使用均值进行填补。

(4) 在 `feature_engineering` 函数中，使用 `astype` 方法将 `促销活动` 列的数据类型转换为整数类型 (`int32` 或 `int64`)。

(5) 在 `train_and_evaluate` 函数中, 使用 `np.sqrt(mean_squared_error(...))` 标准方法计算并返回 XGBoost 模型在测试集上的均方根误差 (RMSE)。

(6) 在 `train_and_evaluate` 函数中, 生成并保存预测值与真实值的散点图至 `prediction_scatter.png`, 该图必须包含蓝色的数据点和一条红色的拟合曲线, 并确保中文标题 (“预测值 vs 真实值”) 与坐标轴标签 (“真实销量”、“预测销量”) 显示正常无乱码。

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	2	考评员运行考生提交的 PY 文件, 检查运行过程日志显示 “M1-月均销量计算成功”“M1-示例值:[值]”; 验证 <code>processed_data.csv</code> 包含正确月均销量列 (<code>sales/12</code>)。	结果正确, 且使用了 <code>data['sales']/12</code> 的标准计算方法, 得 2 分。 结果正确, 但未使用 <code>data['sales']/12</code> 的标准计算方法, 得 1 分。 结果错误, 得 0 分。		
M2	2	考评员运行考生提交的 PY 文件, 检查运行过程日志显示 “M2-促销期间销量计算成功”“M2-示例值:[值]”; 验证 <code>processed_data.csv</code> 包含正确促销期间销量列 (<code>promotion=1</code> 时的 <code>sales</code> 均值)。	结果正确, 且使用了 <code>apply</code> 或 <code>np.where</code> 的向量化方法, 得 2 分。 结果正确, 但使用了循环判断的低效方法, 得 1 分。 结果错误, 得 0 分。		
M3	2	考评员运行考生提交的 PY 文件, 检查运行过程日志显示 “M3-价格缺失值填补成功”“M3-填补记录数:[值]”(记录填补前缺失值数量, 约 500 条); 验证 <code>processed_data.csv</code> 的价格列无缺失值 (接受均值填补)。	结果正确, 且使用了均值进行填补, 得 2 分。 结果正确, 但使用了中位数、众数或固定值进行填补, 得 1 分。 结果错误, 得 0 分。		
M4	1	考评员运行考生提交的 PY 文件, 检查运行过程日志显示 “M4-促销活动类型转换成功”“M4-数据类型:[int]” 验证 <code>processed_data.csv</code> 的促销活动	结果正确, 且使用了 <code>astype('int32')</code> 或 <code>astype('int64')</code> , 得 1 分。 结果正确, 但使用了其他方法 (如 <code>pd.to_numeric</code>), 得 0.5		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
		列为整数类型 (int32 或 int64)。	分。 结果错误, 得 0 分。		
M5	2	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M5-RMSE 计算成功”“M5-RMSE:[值]”; 验证 RMSE 值在合理范围内 (>0)。	结果正确, 且使用了 np.sqrt(mean_squared_error(...)) 的标准方法, 得 2 分。 结果正确, 但使用了手动计算的方法, 得 1 分。 结果错误, 得 0 分。		
M6	1	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M6-预测散点图生成成功”; prediction_scatter.png 存在, 为蓝色点, 红色拟合曲线 (接受二次多项式或其他平滑曲线), 中文标题 (“预测值 vs 真实值”) 和轴标签 (“真实销量”“预测销量”) 无乱码。	结果正确, 且完全满足所有格式要求 (蓝色点、红色拟合曲线), 得 1 分。 结果正确, 但未完全满足所有格式要求 (如缺少拟合曲线、或颜色不符), 得 0.5 分。 结果错误 (如图未生成或中文乱码), 得 0 分。		
合计配分	10	合计得分			

八、银行信用风险评估考核 (考核时间: 15 min)

1. 试题单

(1) 场地设备要求

- 1) 装有 Windows 操作系统的计算机 (Windows 7 及以上);
- 2) 系统中已安装 BI 类商务数据分析软件、压缩软件、Python 3.8+ (安装 pandas、numpy、scikit-learn、matplotlib、seaborn 库)、MySQL 8.0+ 及办公软件;
- 3) 数据库及商务数据分析师考试平台。

(2) 工作任务

- 1) 计算贷款收入比和信用评分等级;
- 2) 处理收入缺失值, 标准化贷款金额和信用评分;
- 3) 训练随机森林模型, 计算准确率, 生成信用评分分布图和特征重要性图;
- 4) 生成 HTML 报告, 包含特征计算结果、分布图、模型评估结果、特征重要性图和业务洞察。

(3) 技能要求

- 1) 根据业务需求编写批量数据计算作业并优化作业参数；
- 2) 根据业务规则设计相应数据指标计算算法；
- 3) 评估挖掘需求并使用工具对数据进行特征工程处理；
- 4) 调用常规模型进行模型训练；
- 5) 基于评价指标进行模型测试。

(4) 质量指标

- 1) 高效准确提取贷款收入比和信用评分等级；
- 2) 完整规范处理收入缺失值并标准化特征；
- 3) 随机森林模型精准分类违约风险；
- 4) HTML 报告清晰呈现特征计算结果、模型性能和图表。

背景材料

银行通过分析客户信用数据评估贷款违约风险，以降低不良贷款率。继电信和电商任务后，本任务回归分类问题，基于 10,000 条银行信用数据，包含客户 ID、收入（年收入，单位：元，部分缺失）、贷款金额（单位：元）、信用评分（0-100）和是否违约（0=未违约，1=违约）等字段。目标是构建违约预测模型，识别高风险客户，并生成分析报告，为信贷决策提供依据。

操作要求：

- (1) 在 `compute_features` 函数中，计算贷款收入比，确保 `贷款收入比` 列被正确创建，其值为 `loan_amount` 列除以 `income` 列。
- (2) 在 `compute_features` 函数中，使用向量化方法（如 `pd.cut` 或 `np.select`）创建 `信用评分等级` 列，其逻辑为：信用评分低于 60 为'低'，60 至 80 之间为'中'，高于 80 为'高'。
- (3) 在 `feature_engineering` 函数中，填补 `income` 列的缺失值，确保该列最终无任何缺失数据，并且必须使用均值进行填补。
- (4) 在 `feature_engineering` 函数中，使用 `StandardScaler` 对 `loan_amount` 和 `credit_score` 两列进行标准化处理，确保处理后的新列均值约等于 0 且标准差约等于 1。
- (5) 在 `train_and_evaluate` 函数中，使用 `accuracy_score` 函数计算并返回随机森林模型在测试集上的准确率。
- (6) 在 `train_and_evaluate` 函数中，生成并保存特征重要性条形图至 `特征重要性.png`，该图必须为绿色的水平条形图（`kind='barh', color='green'`），并确保中文

标题（“特征重要性”）与坐标轴标签（“重要性”、“特征”）显示正常无乱码。

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	2	考评员运行考生提交的PY文件，检查运行过程日志显示“M1-贷款收入比计算成功”“M1-示例值:[值]”；验证处理后数据.csv包含正确贷款收入比列（贷款金额/收入）。	结果正确，且使用了 <code>data['loan_amount'] / data['income']</code> 的标准计算方法，得2分。 结果正确，但未使用 <code>data['loan_amount'] / data['income']</code> 的标准计算方法，得1分。 结果错误，得0分。		
M2	2	考评员运行考生提交的PY文件，检查运行过程日志显示“M2-信用评级等级计算成功”“M2-示例值:[值]”；验证处理后数据.csv包含正确信用评级等级列（<60=低，60-80=中，>80=高）。	结果正确，且使用了 <code>pd.cut</code> 或 <code>np.select</code> 的向量化方法，得2分。 结果正确，但使用了循环判断的低效方法，得1分。 结果错误，得0分。		
M3	2	考评员运行考生提交的PY文件，检查运行过程日志显示“M3-收入缺失值填补成功”“M3-填补记录数:[值]”（记录填补前缺失值数量，约500条）；验证处理后数据.csv的收入列无缺失值（接受均值填补）。	结果正确，且使用了均值进行填补，得2分。 结果正确，但使用了中位数、众数或固定值进行填补，得1分。 结果错误，得0分。		
M4	1	考评员运行考生提交的PY文件，检查运行过程日志显示“M4-标准化成功”“M4-均值和标准差:[值]”；验证处理后数据.csv的贷款金额和信用评分列均值≈0、标准差≈1。	结果正确，且使用了 <code>StandardScaler</code> 或等效的手动标准化方法，得1分。 结果正确，但使用了其他缩放方法（如 <code>MinMaxScaler</code> ），得0.5分。 结果错误，得0分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M5	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M5-准确率计算成功”“M5-准确率:[值]”；验证准确率值在合理范围内（0-1）。	结果正确，且使用了 accuracy_score 函数，得 2 分。 结果正确，但使用了手动计算（如 np.sum(y_test == y_pred) / len(y_test)）的方法，得 1 分。 结果错误，得 0 分。		
M6	1	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M6-特征重要性图生成成功”；特征重要性.png 存在，为绿色条形，中文标题（“特征重要性”）和轴标签（“重要性”“特征”）无乱码。	结果正确，且完全满足所有格式要求（绿色条形），得 1 分。 结果正确，但未完全满足所有格式要求（如使用了其他颜色），得 0.5 分。 结果错误（如图未生成或中文乱码），得 0 分。		
合计配分	10	合计得分			

九、医疗影像数据分类考核（考核时间：15 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装 BI 类商务数据分析软件、压缩软件、Python 3.8+（安装 pandas、numpy、scikit-learn、matplotlib、seaborn 库）、MySQL 8.0+及办公软件；
- 3) 数据库及商务数据分析师考试平台。

（2）工作任务

- 1) 计算像素均值标准差和对比度等级；
- 2) 处理像素均值缺失值，标准化像素均值和对比度；
- 3) 训练支持向量机（SVM）模型，计算准确率，生成像素均值分布图和混淆矩阵图；

4) 生成 HTML 报告，包含特征计算结果、分布图、模型评估结果、混淆矩阵图和业务洞察。

(3) 技能要求

- 1) 根据业务需求编写批量数据计算作业并优化作业参数；
- 2) 根据业务规则设计相应数据指标计算算法；
- 3) 评估挖掘需求并使用工具对数据进行特征工程处理；
- 4) 调用常规模型进行模型训练；
- 5) 基于评价指标进行模型测试。

(4) 质量指标

- 1) 高效准确提取像素均值标准差和对比度等级；
- 2) 完整规范处理像素均值缺失值并标准化特征；
- 3) 支持向量机模型精准分类影像异常；
- 4) HTML 报告清晰呈现特征计算结果、模型性能和图表。

背景材料

医疗影像数据分析是辅助诊断的关键技术，通过分析影像特征（如像素均值、对比度）识别异常（如肿瘤或病变），提升诊断准确性。本任务为分类问题，基于 10,000 条医疗影像数据，包含患者 ID、像素均值（影像灰度均值，单位：无，10%缺失）、对比度（影像对比度强度，单位：无，范围 30-90）和诊断结果（0=正常，1=异常，异常占比 15%）等字段。目标是构建影像异常分类模型，识别高风险影像，并生成分析报告，为临床诊断提供依据。

操作要求：

(1) 在 `compute_features` 函数中，使用 `data['pixel_mean'].std()` 方法计算像素均值标准差，并确保 像素均值标准差 列被正确创建。

(2) 在 `compute_features` 函数中，使用向量化方法（如 `pd.cut` 或 `np.select`）创建 对比度等级 列，其逻辑为：对比度低于 50 为'低'，50 至 75 之间为'中'，高于 75 为'高'。

(3) 在 `feature_engineering` 函数中，填补 `pixel_mean` 列的缺失值，确保该列最终无任何缺失数据，并且必须使用均值进行填补。

(4) 在 `feature_engineering` 函数中，使用 `StandardScaler` 对 `pixel_mean` 和 `contrast` 两列进行标准化处理，确保处理后的新列均值约等于 0 且标准差约等于 1。

(5) 在 `train_and_evaluate` 函数中，使用 `accuracy_score` 函数计算并返回 SVM 模型在测试集上的准确率。

(6) 在 `train_and_evaluate` 函数中, 生成并保存混淆矩阵图至 `混淆矩阵.png`, 该图必须为蓝色热图 (`cmap='Blues'`), 在每个单元格中标注具体数值, 并确保中文标题 (“混淆矩阵”)、坐标轴标签 (“预测标签”、“真实标签”) 以及分类标签 (“正常”、“异常”) 显示正常无乱码。

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	2	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M1-像素均值标准差计算成功”“M1-示例值:[值]”; 验证处理后数据.csv 包含正确像素均值标准差列。	结果正确, 且使用了 <code>data['pixel_mean'].std()</code> 的标准计算方法, 得 2 分。结果正确, 但使用了其他方法 (如 <code>np.std</code> 或手动计算), 得 1 分。结果错误, 得 0 分。		
M2	2	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M2-对比度等级计算成功”“M2-示例值:[值]”; 验证处理后数据.csv 包含正确对比度等级列 (<50=低, 50-75=中, >75=高)。	结果正确, 且使用了 <code>pd.cut</code> 或 <code>np.select</code> 的向量化方法, 得 2 分。结果正确, 但使用了循环判断的低效方法, 得 1 分。结果错误,		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
			得 0 分。		
M3	2	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M3-像素均值缺失值填补成功”“M3-填补记录数:[值]” (记录填补前缺失值数量, 约 1000 条); 验证处理后数据.csv 的 pixel_mean 列无缺失值 (接受均值填补)。	结果正确, 且使用了均值进行填补, 得 2 分。 结果正确, 但使用了中位数、众数或固定值进行填补, 得 1 分。 结果错误, 得 0 分。		
M4	1	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M4-标准化成功”“M4-均值和标准差:[值]”; 验证处理后数据.csv 的 pixel_mean 和 contrast 列均值 ≈ 0 、标准差 ≈ 1 。	结果正确, 且使用了 StandardScaler 或等效的手动标准化方法, 得 1 分。 结果正确, 但使用了其他缩放方法 (如 MinMaxScaler), 得 0.5 分。 结果错误, 得 0 分。		
M5	2	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M5-准确率计算成功”“M5-准确率:[值]”; 验证准确率值在合理范围内 (0.5-1.0)。	结果正确, 且使用了 accuracy_score 函数, 得 2 分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
			结果正确,但使用了手动计算(如 np.sum(y_test == y_pred)/len(y_test))的方法,得 1 分。结果错误,得 0 分。		
M6	1	考评员运行考生提交的 PY 文件,检查运行过程日志显示“M6-混淆矩阵图生成成功”;混淆矩阵.png 存在,为蓝色热图,中文标题(“混淆矩阵”)和轴标签(“预测标签”“真实标签”)无乱码,分类标签为“正常”和“异常”。	结果正确,且完全满足所有格式要求(蓝色热图、带数字标注、分类标签),得 1 分。结果正确,但未完全满足所有格式要求(如未使用蓝色调、或缺少数字/分类标签),得 0.5 分。结果错误(如图未生成或中文乱码),得 0 分。		
合计配分	10	合计得分			

十、交通流量预测考核（考核时间：15 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装 BI 类商务数据分析软件、压缩软件、Python 3.8+（安装 pandas、numpy、scikit-learn、matplotlib、seaborn 库）、MySQL 8.0+及办公软件；
- 3) 数据库及商务数据分析师考试平台。

（2）工作任务

- 1) 计算小时均车流量和高峰时段标志；
- 2) 处理车流量缺失值，转换天气为整数；
- 3) 训练梯度提升树模型，计算 MAE，生成车流量分布图和预测散点图；
- 4) 生成 HTML 报告，包含特征计算结果、分布图、模型评估结果、散点图和业务洞察。

（3）技能要求

- 1) 根据业务需求编写批量数据计算作业并优化作业参数；
- 2) 根据业务规则设计相应数据指标计算算法；
- 3) 评估挖掘需求并使用工具对数据进行特征工程处理；
- 4) 调用常规模型进行模型训练；
- 5) 基于评价指标进行模型测试。

（4）质量指标

- 1) 高效准确提取小时均车流量和高峰时段标志；
- 2) 完整规范处理车流量缺失值并转换数据类型；
- 3) 梯度提升树模型精准预测车流量；
- 4) HTML 报告清晰呈现特征计算结果、模型性能和图表。

背景材料

交通流量预测是智慧城市管理的重要组成部分，通过分析车流量数据，城市管理者可以优化交通信号、缓解拥堵并提高道路效率。继电信、电商、银行和医疗任务后，本任务聚焦回归分析，与电商销量预测类似，基于 10,000 条交通流量数据，包含路段 ID（segment_id）、车流量（traffic_flow（每小时车辆数，部分缺失））、时间段（hour（0-23 小时））、天气（weather（0=晴，1=雨，2=雾））和流量预测值（predicted_flow（目标值））等字段。目标是构建预测模型，分析时间段和天气对车流量的影响，并生成分析报告，为交通管理提供支持。

操作要求:

(1) 在 `compute_features` 函数中, 先对 `traffic_flow` 列的缺失值进行临时均值填充, 然后计算小时均车流量 (填充后的值/24), 确保 `小时均车流量` 列被正确创建。

(2) 在 `compute_features` 函数中, 使用向量化方法 (如 `apply` 或 `np.where`) 创建 `高峰时段` 列, 其逻辑为: 当 `hour` 值为 7, 8, 9, 17, 18, 或 19 时为 1, 否则为 0。

(3) 在 `feature_engineering` 函数中, 填补 `traffic_flow` 列的缺失值, 确保该列最终无任何缺失数据, 并且必须使用均值进行填补。

(4) 在 `feature_engineering` 函数中, 使用 `map` 方法将 `weather` 列的文本值 ('晴', '雨', '雾') 精确映射为对应的整数 (0, 1, 2)。

(5) 在 `train_and_evaluate` 函数中, 使用 `mean_absolute_error` 函数计算梯度提升树模型在测试集上的平均绝对误差 (MAE), 并确保结果正确。

(6) 在 `train_and_evaluate` 函数中, 生成并保存预测值与真实值的散点图至 `预测散点图.png`, 该图尺寸必须为 576x432 像素, 包含蓝色半透明的数据点、一条红色对角线, 并确保中文标题 (“预测值 vs 真实值”) 与坐标轴标签 (“真实车流量”、“预测车流量”) 显示正常无乱码。

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	2	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M1-小时均车流量计算成功”“M1-示例值:[值]”; 验证处理后数据.csv 包含正确小时均车流量列 (<code>traffic_flow/24</code> , 缺失值已填补)。	结果正确, 且使用了先临时填补缺失值再计算的标准方法, 得 2 分。 结果正确, 但处理缺失值的方法不规范 (如先全局填补再计算), 得 1 分。 结果错误, 得 0 分。		
M2	2	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M2-高峰时段标志计算成功”“M2-示例值:[值]”; 验证处理后数据.csv 包	结果正确, 且使用了 <code>apply</code> 或 <code>np.where</code> 的向量化方法, 得 2 分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
		含正确高峰时段列（7-9 或 17-19=1，其余=0）。	结果正确，但使用了循环判断的低效方法，得 1 分。 结果错误，得 0 分。		
M3	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M3-车流量缺失值填补成功”“M3-填补记录数:[约 1000]”；验证处理后数据.csv 的 traffic_flow 列无缺失值（接受均值填补）。	结果正确，且使用了均值进行填补，得 2 分。 结果正确，但使用了中位数、众数或固定值进行填补，得 1 分。 结果错误，得 0 分。		
M4	1	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M4-天气转换成功”“M4-示例值:[值]”；验证处理后数据.csv 的 weather 列为整数类型（0=晴，1=雨，2=雾）。	结果正确，且使用了 map 方法进行映射，得 1 分。 结果正确，但使用了其他方法（如 pd.Categorical 或 replace），得 0.5 分。 结果错误，得 0 分。		
M5	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M5-MAE 计算成功”“M5-MAE:[值]”；验证 MAE 值在 0-1000 范围内（示例：221.91）。	结果正确，且使用了 mean_absolute_error 函数，得 2 分。 结果正确，但使用了手动计算（如 np.abs(y_test - y_pred).mean()）的方法，得 1 分。 结果错误，得 0 分。		
M6	1	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M6-预测散点图生成成功”“M6-图像尺寸:576x432”；预测散点图.png 存在，尺寸为 576x432 像素（figsize=(8,6),dpi=72），包含标题（“预测值 vs 真实值”）、轴标签（“真实车流量”“预测车流量”）、蓝色点（推荐 alpha=0.5）和红色对角线，中文无乱码。	结果正确，且完全满足所有格式要求（蓝色点、红色对角线、正确尺寸），得 1 分。 结果正确，但未完全满足所有格式要求（如缺少对角线、颜色不符或尺寸不对），得 0.5 分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
			结果错误（如图未生成或中文乱码），得0分。		
合计配分	10	合计得分			

十一、开展销售数据分析、预测及编写统计报告（考核时间：15 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装数据分析软件、压缩软件及办公软件；
- 3) 数据库及商务数据分析师考试平台。

（2）工作任务

- 1) 根据某公司提供的销售数据，完成数据汇总与描述性统计分析；
- 2) 根据公司业务数据，绘制线性回归模型，并对某个时间销售额进行预测；
- 3) 根据数据分析的结果，描述数据关系，完成一份简要数据统计报告。

（3）技能要求

- 1) 结合业务场景使用工具对数据集进行概要、描述性统计分析；
- 2) 在描述结果基础上对数据进行特征和规律的分析与推测；
- 3) 结合业务场景编写数据统计报告。

（4）质量指标

- 1) 正确计算均值、中位数、最大值、最小值和标准差；
- 2) 正确预测 2026 年 1 月份销售额；
- 3) 准确编写简要的数据统计报告，语句通顺、逻辑合理。

背景材料

国通公司是一家经营多种食品和饮品的企业，拥有多个项目组 and 众多员工。目前公司积累了大量的销售数据、员工数据、产品数据以及顾客数据等，存储在不同工作表中。公司管理层希望通过对这些数据的分析，了解销售业务的整体情况，以便制定更合理的销售策略和资源分配方案。

操作要求:

(1) 根据国通公司提供的数据, 计算 2023 年月销售额的均值、中位数、最大值、最小值和标准差, 计算 2024 年月成本的均值、中位数、最大值、最小值和标准差;

(2) 根据这两年的月销售额数据, 利用最小二乘法生成回归模型并预测 2026 年 1 月份该公司的销售额;

(3) 基于上述分析, 撰写一份数据统计报告。

(注: 所有计算结果都保留两位小数)

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	0.5	正确计算 2023 年月销售额均值。	是=0.5 分; 否=0 分。		
M2	0.5	正确计算 2023 年月销售额中位数。	是=0.5 分; 否=0 分。		
M3	0.5	正确计算 2023 年月销售额最大值。	是=0.5 分; 否=0 分。		
M4	0.5	正确计算 2023 年月销售额最小值。	是=0.5 分; 否=0 分。		
M5	0.5	正确计算 2023 年月销售额标准差。	是=0.5 分; 否=0 分。		
M6	0.5	正确计算 2024 年月成本均值。	是=0.5 分; 否=0 分。		
M7	0.5	正确计算 2024 年月成本中位数。	是=0.5 分; 否=0 分。		
M8	0.5	正确计算 2024 年月成本最大值。	是=0.5 分; 否=0 分。		
M9	0.5	正确计算 2024 年月成本最小值。	是=0.5 分; 否=0 分。		
M10	0.5	正确计算 2024 年月成本标准差。	是=0.5 分; 否=0 分。		
M11	2	正确写出一元线性回归模型。	是=2 分; 否=0 分。		
M12	1	正确预测加班时间 5 小时, 新增保单的数量。	是=1 分; 否=0 分。		
合计配分	8	合计得分			

评价分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	2	答案与数据高度相关，解释清晰、逻辑严谨，结论合理，且分析深入或有独到见解=2分。				
		答案与数据相关，解释基本合理，但分析较浅显或存在少量逻辑不严谨之处=1分。				
		答案与数据无关，解释不合理，或未提供任何有效分析=0分。				
合计配分	2	合计得分				

十二、开展物流数据分析、预测及编写统计报告（考核时间：15 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装数据分析软件、压缩软件及办公软件；
- 3) 数据库及商务数据分析师考试平台。

（2）工作任务

- 1) 根据某公司提供的物流数据，完成数据汇总与描述性统计分析；
- 2) 根据公司业务数据，绘制二元线性回归模型，并进行预测；
- 3) 根据数据分析的结果，完成一份简要数据统计报告。

（3）技能要求

- 1) 结合业务场景使用工具对数据集进行概要、描述性统计分析；
- 2) 在描述结果基础上对数据进行特征和规律的分析与推测；
- 3) 结合业务场景编写数据统计报告。

（4）质量指标

- 1) 正确计算均值、中位数和标准差；
- 2) 正确预测重量为 100000、提单量为 200000 时的金额；
- 3) 准确编写简要的数据统计报告，语句通顺、逻辑合理。

背景材料

东方港通物流有限公司是国内领先的港口物流综合服务商，专注于大宗散货、集装箱及特种货物的国内运输及合同销售业务，服务网络覆盖全国主要港口枢纽。公司建立了完善的智能化数据采集系统，系统性地记录了提单号、作业日期、货物重量、运输金额、提单量、作业港口、货物类型、客户等级等 30 余项核心业务指标，构建了贯穿业务全流程的完整数据库体系。

在物流行业数字化转型的浪潮下，公司高度重视数据驱动的科学决策。本次专项分析聚焦运输金额、货物重量与提单量三大关键指标的关联性研究，由数据分析主管张伟领衔的专业团队负责实施。研究将运用先进的数据分析方法，深入挖掘指标间的内在联系和变化规律，旨在为公司的运力资源配置优化和动态定价策略制定提供精准的数据支撑。通过建立科学的分析模型和预测机制，该项目将有效提升公司的业务预判能力，优化经营管理决策，助力企业实现降本增效的高质量发展目标。

操作要求：

(1) 根据该公司提供的数据，分别计算重量、提单量和金额的均值、中位数和标准差；

(2) 根据金额与重量、提单量之间的关系，利用线性最小二乘法生成二元回归模型并预测重量为 100000、提单量为 200000 时的金额；

(3) 基于上述分析，撰写一份数据统计报告。

(注：所有计算结果都保留两位小数)

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	0.5	正确计算重量均值。	是=0.5 分； 否=0 分。		
M2	0.5	正确计算重量中位数。	是=0.5 分； 否=0 分。		
M3	0.5	正确计算重量标准差。	是=0.5 分； 否=0 分。		
M4	0.5	正确计算提单量均值。	是=0.5 分； 否=0 分。		
M5	0.5	正确计算提单量中位数。	是=0.5 分； 否=0 分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M6	0.5	正确计算提单量标准差。	是=0.5分； 否=0分。		
M7	0.5	正确计算金额均值。	是=0.5分； 否=0分。		
M8	0.5	正确计算金额中位数。	是=0.5分； 否=0分。		
M9	0.5	正确计算金额标准差。	是=0.5分； 否=0分。		
M10	2	正确写出二元线性回归模型。	是=2分； 否=0分。		
M11	1.5	正确预测金额。	是=1.5分； 否=0分。		
合计配分	8	合计得分			

评价分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	2	答案与数据高度相关，解释清晰、逻辑严谨，结论合理，且分析深入或有独到见解=2分。				
		答案与数据相关，解释基本合理，但分析较浅显或存在少量逻辑不严谨之处=1分。				
		答案与数据无关，解释不合理，或未提供任何有效分析=0分。				
合计配分	2	合计得分				

十三、开展保险数据分析、预测及编写统计报告（考核时间：15 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装数据分析软件、压缩软件及办公软件；
- 3) 数据库及商务数据分析师考试平台。

(2) 工作任务

- 1) 根据某公司提供的保险数据，完成数据汇总与描述性统计分析；
- 2) 根据公司业务数据，绘制线性回归模型，并对加班 5 小时的保单数量进行预测；
- 3) 根据数据分析的结果，完成一份简要数据统计报告。

(3) 技能要求

- 1) 结合业务场景使用工具对数据集进行概要、描述性统计分析；
- 2) 在描述结果基础上对数据进行特征和规律的分析与推测；
- 3) 结合业务场景编写数据统计报告。

(4) 质量指标

- 1) 正确计算均值、中位数、最大值、最小值和标准差；
- 2) 正确预测加班 5 小时的新增保单数量；
- 3) 准确编写简要的数据统计报告，语句通顺、逻辑合理。

背景材料

为了更好地优化员工的工作安排并提升工作效率，上海海东保险公司决定深入分析其营业部每周加班时间与签发的新保单数目之间的潜在关系。公司希望通过这一分析，了解加班时间是否对新保单的签发数量有显著影响，从而为制定合理的工作时间政策提供数据支持。为此，公司在连续 60 周的时间内，系统地收集了每周的加班时间（以小时为单位）以及对应签发的新保单数目（以件为单位）的详细数据。通过对这些数据进行统计分析，公司希望能够揭示两者之间的相关性，并为未来的资源分配和工作计划提供科学依据。

操作要求：

(1) 根据上海海东保险公司提供的数据，计算每周加班时间均值、中位数、最大值、最小值和标准差，计算新保单数目（件）的均值、中位数、最大值、最小值和标准差；

(2) 根据提供的数据，利用线性最小二乘法生成回归模型并预测加班时间 5 小时新增保单的数量；

(3) 基于上述分析，撰写一份数据统计报告。

(注：所有计算结果都保留两位小数)

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	0.5	正确计算加班时间均值。	是=0.5分； 否=0分。		
M2	0.5	正确计算加班时间中位。	是=0.5分； 否=0分。		
M3	0.5	正确计算加班时间最大值。	是=0.5分； 否=0分。		
M4	0.5	正确计算加班时间额最小值。	是=0.5分； 否=0分。		
M5	0.5	正确计算加班时间标准差。	是=0.5分； 否=0分。		
M6	0.5	正确计算新保单数目均值。	是=0.5分； 否=0分。		
M7	0.5	正确计算新保单数目中位数。	是=0.5分； 否=0分。		
M8	0.5	正确计算新保单数目最大值。	是=0.5分； 否=0分。		
M9	0.5	正确计算新保单数目最小值。	是=0.5分； 否=0分。		
M10	0.5	正确计算新保单数目标标准差。	是=0.5分； 否=0分。		
M11	2	正确写出一元线性回归模型。	是=2分； 否=0分。		
M12	1	正确预测新增保单的数量。	是=1分； 否=0分。		
合计配分	8	合计得分			

评价分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	2	正确揭示“签发新保单的数目与每周加班时间”的相关性=2分。				
		提及“签发新保单的数目与每周加班时间”的相关性,其中之一=1分。				

		未提及“签发新保单的数目与每周加班时间”的相关性=0分。				
合计配分	2	合计得分				

十四、开展车辆数据分析、预测及编写统计报告（考核时间：15 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装数据分析软件、压缩软件及办公软件；
- 3) 数据库及商务数据分析师考试平台。

（2）工作任务

- 1) 根据某公司提供的用车数据，完成数据汇总与描述性统计分析；
- 2) 根据公司用车数据，绘制二元线性回归模型，并进行预测；
- 3) 根据数据分析的结果，完成一份简要数据统计报告。

（3）技能要求

- 1) 结合业务场景使用工具对数据集进行概要、描述性统计分析；
- 2) 在描述结果基础上对数据进行特征和规律的分析与推测；
- 3) 结合业务场景编写数据统计报告。

（4）质量指标

- 1) 正确计算均值、中位数和标准差；
- 2) 正确预测晚上 8 点，里程为 50 时的订单金额；
- 3) 准确编写简要的数据统计报告，语句通顺、逻辑合理。

背景材料

新锐科技股份有限公司作为一家快速发展的科技型企业，业务范围涵盖研发、生产、销售等多个领域。随着业务规模不断扩大，员工因工作需要频繁使用打车服务。为加强差旅费用精细化管理，公司建立了完善的用车审批制度，将用车类型细分为生产用车、加班用车、因公用车、出差用车和业务用车。财务部门通过数字化管理系统详细记录了每笔打车订单的核心数据，包括订单编号、用车时间、行驶里程、使用人员工号、所属部门代码、用车事由说明及费用明细等，形成了完整的打车业务数据库。

鉴于近期打车费用持续攀升，公司决定启动专项数据分析工作。本次分析主要围绕三个重点展开，首先研究不同用车类型特别是生产用车和业务用车的费用与里程之间的相关性，其次分析 24 小时内打车时段的分布特征，最后探究打车费用与用车时间行驶里程之间的关系。数据分析主管带领团队开展此项工作，其中员工小张负责部分专项分析任务。通过此次系统性的数据分析，旨在识别潜在的优化空间，为优化用车管理和加强费用管控提供数据支撑和决策依据。

操作要求：

(1) 根据该公司提供的数据，分别计算用车时间（小时时段）、订单总里程和订单金额的均值、中位数和标准差；

(2) 根据订单金额与用车时间、订单总里程之间的关系，利用线性最小二乘法生成二元回归模型并预测晚上 8 点，里程为 50 时的订单金额；

(3) 基于上述分析，撰写一份数据统计报告。

(注：所有计算结果都保留两位小数)

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	0.5	正确计算用车时间均值。	是=0.5 分； 否=0 分。		
M2	0.5	正确计算用车时间中位数。	是=0.5 分； 否=0 分。		
M3	0.5	正确计算用车时间标准差。	是=0.5 分； 否=0 分。		
M4	0.5	正确计算总里程均值。	是=0.5 分； 否=0 分。		
M5	0.5	正确计算总里程中位数。	是=0.5 分； 否=0 分。		
M6	0.5	正确计算总里程标准差。	是=0.5 分； 否=0 分。		
M7	0.5	正确计算订单金额均值。	是=0.5 分； 否=0 分。		
M8	0.5	正确计算订单金额中位数。	是=0.5 分； 否=0 分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M9	0.5	正确计算订单金额标准差。	是=0.5分； 否=0分。		
M10	2	正确写出二元线性回归模型。	是=2分； 否=0分。		
M11	1.5	正确预测订单金额。	是=1.5分； 否=0分。		
合计配分	8	合计得分			

评价分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	2	答案与数据高度相关，解释清晰、逻辑严谨，结论合理，且分析深入或有独到见解=2分。				
		答案与数据相关，解释基本合理，但分析较浅显或存在少量逻辑不严谨之处=1分。				
		答案与数据无关，解释不合理，或未提供任何有效分析，得0分。				
合计配分	2	合计得分				

十五、开展配送数据分析、预测及编写统计报告（考核时间：15 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装数据分析软件、压缩软件及办公软件；
- 3) 数据库及商务数据分析师考试平台。

（2）工作任务

- 1) 根据某公司提供的配送数据，完成数据汇总与描述性统计分析；
- 2) 根据公司配送数据，绘制二元线性回归模型，并进行预测；
- 3) 根据数据分析的结果，完成一份简要数据统计报告。

（3）技能要求

- 1) 结合业务场景使用工具对数据集进行概要、描述性统计分析；

- 2) 在描述结果基础上对数据进行特征和规律的分析与推测;
- 3) 结合业务场景编写数据统计报告。

(4) 质量指标

- 1) 正确计算均值、中位数和标准差;
- 2) 正确预测总单量为 500, 准时送达量为 400 时的金额;
- 3) 准确编写简要的数据统计报告, 语句通顺、逻辑合理。

背景材料

顺达物流作为国内领先的第三方物流服务商, 已在全国 20 多个主要城市建立了完善的物流服务网络, 日均订单处理量突破 1 万单。公司目前在哈尔滨、大连、沈阳、北京、成都等核心城市布局了 30 多个专业化运营站点, 形成了覆盖东北、华北、华东、西南等重点区域的物流配送体系。随着电商物流需求的快速增长和公司业务规模的持续扩张, 现有运力资源配置面临新的挑战, 特别是在大件货物配送方面, 对持有 B1 驾照的专业货车司机需求激增。

为科学优化运力配置, 公司决定启动“运力优化提升计划”, 由数据分析团队负责人张明牵头开展专项研究。项目组将对各站点运营数据进行系统性分析, 包括订单量、配送时效、司机配置等关键指标。张明将带领团队运用描述性统计、相关性分析等方法, 深入挖掘数据规律, 识别运力需求特征, 并建立预测模型评估未来司机缺口。通过数据驱动的决策支持, 公司计划制定精准的招聘与调配方案, 在保障 95% 以上配送准时率的同时, 实现运力资源的最优配置, 为业务持续扩张提供坚实支撑。

操作要求:

- (1) 根据顺达物流公司提供的数据, 分别计算总单量、准时送达量和 B1 所需就业人数的均值、中位数和标准差;
- (2) 根据 B1 需就业人数与总单量、准时送达量之间的关系, 利用线性最小二乘法生成二元回归模型并预测总单量为 500, 准时送达量为 400 时的就业人数;
- (3) 基于上述分析, 撰写一份数据统计报告。

(注: 所有计算结果都保留两位小数)

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	0.5	正确计算总单量均值。	是=0.5分； 否=0分。		
M2	0.5	正确计算总单量中位数。	是=0.5分； 否=0分。		
M3	0.5	正确计算总单量标准差。	是=0.5分； 否=0分。		
M4	0.5	正确计算准时送达量均值。	是=0.5分； 否=0分。		
M5	0.5	正确计算准时送达量中位数。	是=0.5分； 否=0分。		
M6	0.5	正确计算准时送达量标准差。	是=0.5分； 否=0分。		
M7	0.5	正确计算就业人数均值。	是=0.5分； 否=0分。		
M8	0.5	正确计算就业人数中位数。	是=0.5分； 否=0分。		
M9	0.5	正确计算就业人数标准差。	是=0.5分； 否=0分。		
M10	2	正确写出二元线性回归模型。	是=2分； 否=0分。		
M11	1.5	正确预测就业人数。	是=1.5分； 否=0分。		
合计配分	8	合计得分			

评价分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	2	答案与数据高度相关，解释清晰、逻辑严谨，结论合理，且分析深入或有独到见解=2分。				
		答案与数据相关，解释基本合理，但分析较浅显或存在少量逻辑不严谨之处=1分。				

		答案与数据无关，解释不合理，或未提供任何有效分析=0分。				
合计配分	2	合计得分				

数据可视化

一、设计商铺数据可视化方案（考核时间：15 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装数据处理软件、压缩软件及办公软件；
- 3) 商务数据分析师考试平台。

（2）工作任务

- 1) 根据业务需求及公司数据状况，完成数据可视化实现方式的设计任务；
- 2) 根据业务场景及视觉对象特点，开展可视化展板的设计与布局规划工作；
- 3) 根据业务要求及分析结果，完成数据展示方案的制定与呈现。

（3）技能要求

- 1) 设计数据可视化实现方式；
- 2) 设计与布局可视化展板；
- 3) 结合业务要求和分析成果，制定并呈现数据展示方案。

（4）质量指标

- 1) 精准锚定公司业务需求，正确梳理数据类型与特性及筛选关联数据，合理设计可视化视觉载体；
- 2) 准确分析业务场景和视觉对象特征，合理规划可视化展板架构，科学布局画布页面；
- 3) 深度解析业务需求内核，分析结果要点提炼无误，合理构建数据展示方案，语句通顺，逻辑清晰。

背景材料

近年来中国百货行业处于变革浪潮，市场规模随城镇化与居民收入提升稳步扩张，三线及以下城市成为新增长极。消费场景加速融合，百货店转型为集购物、餐饮、娱乐、体验服务于一体的综合空间，美妆等高附加值品类销售占比提升，

拉动单店营收。竞争格局中，外资品牌，如老佛爷、伊势丹依托国际资源与高端服务深耕一线，本土连锁，如美罗、银泰借数字化与供应链下沉向低线渗透。消费者行为变迁显著，即时性消费需求凸显，对购物便捷时效要求提高，同时数字化深度影响决策，线上渠道成为信息获取、比价与分享的关键途径。

在此行业背景下，海天百货公司收集了 2023-2025 年上海分店的月度商铺明细，涵盖店铺数、店铺排名，以及包括餐饮、普通服饰、儿童、珠宝等在内的业态和楼层等详细信息。公司员工张海东承担起分析这些数据的重任，旨在帮助企业解决三大需求。市场规模与潜力，挖潜下沉市场，评估高附加值品类以优化货架；竞争格局应对，拆解外资与本土连锁策略，评估业态竞争力，寻差异化路径，强化本土服务与数字化能力；消费者行为适配，速应即时消费需求，优化供应链响应，借数字化渠道与社交营销破局，以体验式业态创新提升客流与消费。

操作要求:

- (1) 结合该公司的业务战略目标与现有数据维度，构建多维度、交互式的数据可视化实现方式，呈现各店铺与业态经营差异及核心特征，为决策提供依据；
- (2) 紧密贴合业务场景的实际需求与可视化视觉对象的独特性，运用前沿的数据可视化设计方法论，打造具备高度交互性、灵活扩展性的展板布局；
- (3) 作为本次数据分析任务主责人，基于业务战略研判与数据深度梳理，运用 PPT 定制呈现契合业务场景、赋能决策的数据展示方案，助力精准把握市场、优化运营。（请提交操作完成后的数据文件至考试平台。）

2.评分表

测量分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	3	所选属性合理、形成 3 个正确的可视化视觉对象=3 分。				
		所选属性基本合理、形成至少 2 个正确的可视化视觉对象=2 分。				
		所选属性基本合理、形成 1 个正确的可视化视觉对象=1 分。				
		所选特征不合理并形成错误的可视化视觉对象=0 分。				

J2	2	设计展板色彩体系完整规范、图表类型丰富多样、布局科学合理=2分。				
		设计展板色彩体系基本规范、图表类型较为单一、布局基本合理=1分。				
		设计展板色彩混乱，图表类型单一且不匹配数据特征，布局杂乱无章=0分。				
J3	1	设计方案框架中封面和目录完整、设计规范=1分。				
		设计方案框架中封面或目录缺失、设计基本规范=0.5分。				
		设计方案框架中缺失封面或目录、不规范=0分。				
J4	1	业务需求提及市场规模与潜力洞察需求、竞争格局与业态分化应对需求、消费者行为变迁适应需求等关键词中的2个及以上，且描述准确=1分。				
		业务需求提及市场规模与潜力洞察需求、竞争格局与业态分化应对需求、消费者行为变迁适应需求等关键词中的1个，或描述基本准确=0.5分。				
		业务需求未提及市场规模与潜力洞察需求、竞争格局与业态分化应对需求、消费者行为变迁适应需求等关键词，或语句不通顺、毫无逻辑=0分。				
J5	3	展示方案中呈现3个视觉对象，数据分析准确合理=3分。				
		展示方案中呈现2个视觉对象，或数据分析基本合理=2分。				
		展示方案中呈现1个视觉对象，或数据分析存在不足=1分。				
		展示方案中视觉对象缺失，或数据分析存在重大问题=0分。				
合计配分	10	合计得分				

二、设计销售数据可视化方案（考核时间：15 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装数据处理软件、压缩软件及办公软件；
- 3) 商务数据分析师考试平台。

（2）工作任务

- 1) 根据公司现有销售数据，完成数据可视化实现方式的设计任务；
- 2) 根据业务场景及视觉对象特点，开展可视化展板的设计与布局规划工作；
- 3) 根据业务要求及分析结果，完成数据展示方案的制定与呈现。

（3）技能要求

- 1) 设计数据可视化实现方式；
- 2) 设计与布局可视化展板；
- 3) 结合业务要求和分析成果，制定并呈现数据展示方案。

（4）质量指标

- 1) 精准锚定公司业务需求，精准识别关键数据维度与属性特征，科学构建可视化呈现形式；
- 2) 正确分析业务场景和视觉对象特征，合理设计可视化展板架构，页面布局科学高效；
- 3) 准确解析业务需求内核，准确提炼分析结果要点，合理构建数据展示方案，逻辑清晰。

背景材料

华海公司是一家专注于服装、配饰及鞋子的企业，旗下产品涵盖休闲系列和运动系列，包括短袖 T 恤、夹克、帽子、棉服、短裤、运动鞋等，产品上市时间为 2022 年至 2024 年。随着业务规模的扩大，企业需要全面分析近三年的销售数据，包括零售价、货号、客户信息，如上海未来商贸有限公司、北京理想科技有限公司等、收款明细、销售金额、回款金额以及销售明细，以优化库存管理、制定精准营销策略并提升整体营收效率。

此外，公司要求数据分析团队对销售数据进行深度挖掘，通过多维度的分析揭示潜在的销售规律和市场机会，深入了解各产品线的市场表现，识别高贡献客户群体，并评估不同区域的销售趋势。

公司员工李海东需运用数据分析工具进行可视化分析，最终形成结构化报告

或动态看板，清晰展示销售增长点、客户偏好及潜在业务机会，助力公司制定更具针对性的经营策略，提升市场竞争力。

操作要求:

- (1) 基于现有销售数据体系，构建支持多维度、实时交互的可视化分析，通过动态图表联动展示销售额的情况，为管理层提供直观的决策支持；
- (2) 紧密贴合业务场景与可视化视觉对象的特点，运用数据叙事和视觉感知理论，构建兼具业务洞察力与视觉表现力的智能分析看板；
- (3) 作为项目负责人，基于业务需求与数据洞察，通过定制化 PPT 报告呈现数据展示方案，助力精准把握市场、优化运营。(请提交操作完成后的数据文件至考试平台。)

2.评分表

测量分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	3	所选属性合理、形成 3 个正确的可视化视觉对象=3 分。				
		所选属性基本合理、形成至少 2 个正确的可视化视觉对象=2 分。				
		所选属性基本合理、形成 1 个正确的可视化视觉对象=1 分。				
		所选特征不合理并形成错误的可视化视觉对象=0 分。				
J2	2	设计展板色彩体系完整规范、图表类型丰富多样、布局科学合理=2 分。				
		设计展板色彩体系基本规范、图表类型较为单一、布局基本合理=1 分。				
		设计展板色彩混乱，图表类型单一且不匹配数据特征，布局杂乱无章=0 分。				
J3	1	设计方案框架中封面和目录完整、设计规范=1 分。				
		设计方案框架中封面或目录缺失、设计基本规范=0.5 分。				

		设计方案框架中缺失封面或目录、不规范=0分。				
J4	1	业务需求提及挖掘数据、销售规律和市场机会、产品线的市场表现、客户、销售趋势、可视化分析、经营策略等关键词中的3个及以上，且描述准确=1分。				
		业务需求提及挖掘数据、销售规律和市场机会、产品线的市场表现、客户、销售趋势、可视化分析、经营策略等关键词中的2个，或描述基本准确=0.5分。				
		未提及挖掘数据、销售规律和市场机会、产品线的市场表现、客户、销售趋势、可视化分析、经营策略等关键词，或语句不通顺、毫无逻辑=0分。				
J5	3	展示方案中呈现3个视觉对象，数据分析准确合理=3分。				
		展示方案中呈现2个视觉对象，或数据分析基本合理=2分。				
		展示方案中呈现1个视觉对象，或数据分析存在不足=1分。				
		展示方案中视觉对象缺失，或数据分析存在重大问题=0分。				
合计配分	10	合计得分				

三、设计车辆数据可视化方案（考核时间：15 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装数据处理软件、压缩软件及办公软件；
- 3) 商务数据分析师考试平台。

（2）工作任务

- 1) 根据公司用车数据，完成数据可视化实现方式的设计任务；
- 2) 根据业务场景及视觉对象特点，开展可视化展板的设计与布局规划工作；
- 3) 根据业务要求及分析结果，完成数据展示方案的制定与呈现。

（3）技能要求

- 1) 设计数据可视化实现方式；

- 2) 设计与布局可视化展板；
- 3) 结合业务要求和分析成果，制定并呈现数据展示方案。

(4) 质量指标

- 1) 精准锚定公司业务需求，精准识别关键数据维度与属性特征，科学构建可视化呈现形式；
- 2) 正确分析业务场景和视觉对象特征，合理设计可视化展板架构，页面布局科学高效；
- 3) 准确解析业务需求内核，准确提炼分析结果要点，合理构建数据展示方案，逻辑清晰。

背景材料

深圳市智慧城市科技发展集团有限公司下设财务管理部、纪检监察部、技术质量部等 12 个部门，2025 年 3 月各部门加班用车、出差用车等订单数据已汇总完毕，涵盖订单号、下单时间、车辆类型、用车类型、订单总金额、总里程、企业支付金额及员工姓名等信息。为优化车辆资源调配、控制用车成本并提升管理效率，公司需对用车数据进行深度分析，识别各部门用车规律及费用分布情况，为后续管理决策提供数据支持。

员工赵海东将负责本次数据分析工作，运用分析工具对各部门用车情况进行多维度统计与可视化呈现。以此来评估各部门用车需求合理性；识别异常用车行为及高成本订单，提出优化建议；同时结合员工加班数据，挖掘用车与加班时长的关联性，为人力资源调配及车辆管理策略提供依据。

操作要求:

- (1) 基于企业现有用车数据，设计支持多维度分析的交互式可视化视觉对象，通过动态图表联动展示用车订单数量的情况；
- (2) 聚焦员工加班打车场景，运用数据可视化美学设计原则，打造界面简洁、重点突出、色彩协调的交互式看板；
- (3) 基于业务需求与数据分析结果，运用 PPT 定制呈现数据展示方案，为企业合理管控出行成本、提升管理效能提供决策支持。（请提交操作完成后的数据文件至考试平台。）

2.评分表

测量分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	3	所选属性合理、形成 3 个正确的可视化视觉对象=3 分。				
		所选属性基本合理、形成至少 2 个正确的可视化视觉对象=2 分。				
		所选属性基本合理、形成 1 个正确的可视化视觉对象=1 分。				
		所选特征不合理并形成错误的可视化视觉对象=0 分。				
J2	2	设计展板色彩体系完整规范、图表类型丰富多样、布局科学合理=2 分。				
		设计展板色彩体系基本规范、图表类型较为单一、布局基本合理=1 分。				
		设计展板色彩混乱，图表类型单一且不匹配数据特征，布局杂乱无章=0 分。				
J3	1	设计方案框架中封面和目录完整、设计规范=1 分。				
		设计方案框架中封面或目录缺失、设计基本规范=0.5 分。				
		设计方案框架中缺失封面或目录、不规范=0 分。				
J4	1	业务需求提及车辆资源、用车需求、用车成本、用车规律、用车与加班、提升管理效率等关键词中的 3 个及以上，且描述准确=1 分。				
		业务需求提及车辆资源、用车需求、用车成本、用车规律、用车与加班、提升管理效率等关键词中的 2 个，或描述基本准确=0.5 分。				
		业务需求未提及车辆资源、用车需求、用车成本、用车规律、用车与加班、提升管理效率等关键词，或语句不通顺、毫无逻辑=0 分。				
J5	3	展示方案中呈现 3 个视觉对象，数据分析准确合理=3 分。				

		展示方案中呈现 2 个视觉对象，或数据分析基本合理=2 分。				
		展示方案中呈现 1 个视觉对象，或数据分析存在不足=1 分。				
		展示方案中视觉对象缺失，或数据分析存在重大问题=0 分。				
合计 配分	10	合计得分				

四、设计商品数据可视化方案（考核时间：15 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装数据处理软件、压缩软件及办公软件；
- 3) 商务数据分析师考试平台。

（2）工作任务

- 1) 根据公司数据，完成数据可视化实现方式的设计任务；
- 2) 根据业务场景及视觉对象特点，完成可视化展板的设计工作；
- 3) 根据业务要求及分析结果，完成数据展示方案的制定与呈现。

（3）技能要求

- 1) 设计数据可视化实现方式；
- 2) 设计与布局可视化展板；
- 3) 结合业务要求和分析成果，制定并呈现数据展示方案。

（4）质量指标

- 1) 精准锚定公司业务需求，精准识别关键数据维度与属性特征，科学构建可视化呈现形式；
- 2) 正确分析业务场景和视觉对象特征，合理设计可视化展板架构，页面布局科学高效；
- 3) 准确解析业务需求内核，准确提炼分析结果要点，合理构建数据展示方案，逻辑清晰，语句通顺。

背景材料

科技优品商城是一家专注于智能科技产品的线上零售平台，致力于为消费者提供高品质的智能生活解决方案。平台主营产品包括空气净化器、智能手机及配件、无人机、便携式投影仪等前沿科技商品，覆盖了智能家居、移动办公、数码娱乐等多个应用场景。目前平台已完成 2024 年全年运营数据的系统性采集工作，建立了完整的商品运营数据库。本次数据采集工作重点聚焦商品维度的关键运营指标，全面覆盖了各 SKU 的用户行为数据和交易转化数据。具体包括了反映商品吸引力的访客数、商品流量等曝光指标；衡量商品页面质量的详情页跳出率、平均停留时长等互动指标；以及体现最终转化效果的支付金额、支付买家数、客单价等交易指标。这些数据将为后续的商品运营策略优化提供坚实的数据支撑。

基于平台商品精细化运营的战略需求，数据分析师孙海东将牵头开展深入的商业数据分析工作。通过对商品数据的系统性挖掘和多维度可视化分析，将重点评估各商品的市场表现和运营效率，建立科学的商品评估体系。分析成果将应用于识别高潜力商品以重点推广，优化低效商品展示以提升转化，并制定差异化的商品定价策略和精准促销方案。最终目标是全面提升平台的商品运营效率，优化商品结构布局，推动整体销售业绩的持续增长，巩固平台在智能科技产品领域的市场竞争力。

操作要求:

- (1) 借助多维度数据可视化手段呈现商品分析情况，清晰展示关键指标，为经营决策提供有力数据支撑；
- (2) 依据业务场景特点及数据特性，运用先进的数据可视化设计理念，构建兼具美观性与商业洞察力的交互式智能看板；
- (3) 基于业务分析与数据分析结果，量身定制契合实际业务场景的 PPT 数据展示方案，推动运营效能不断提升。（请提交操作完成后的数据文件至考试平台。）

2.评分表

测量分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	3	所选属性合理、形成 3 个正确的可视化视觉对象=3 分。				
		所选属性基本合理、形成至少 2 个正确的可视化视觉对象=2 分。				

		所选属性基本合理、形成 1 个正确的可视化视觉对象=1 分。				
		所选特征不合理并形成错误的可视化视觉对象=0 分。				
J2	2	设计展板色彩体系完整规范、图表类型丰富多样、布局科学合理=2 分。				
		设计展板色彩体系基本规范、图表类型较为单一、布局基本合理=1 分。				
		设计展板色彩混乱，图表类型单一且不匹配数据特征，布局杂乱无章=0 分。				
J3	1	设计方案框架中封面和目录完整、设计规范=1 分。				
		设计方案框架中封面或目录缺失、设计基本规范=0.5 分。				
		设计方案框架中缺失封面或目录、不规范=0 分。				
J4	1	业务需求提及商品市场表现、挖掘商品潜力、运营问题、运营策略等关键词中的 2 个及以上，且描述准确=1 分。				
		业务需求提及商品市场表现、挖掘商品潜力、运营问题、运营策略等关键词中的 1 个，或描述基本准确=0.5 分。				
		业务需求未提及商品市场表现、挖掘商品潜力、运营问题、运营策略等关键词，或语句不通顺、毫无逻辑=0 分。				
J5	3	展示方案中呈现 3 个视觉对象，数据分析准确合理=3 分。				
		展示方案中呈现 2 个视觉对象，或数据分析基本合理=2 分。				
		展示方案中呈现 1 个视觉对象，或数据分析存在不足=1 分。				
		展示方案中视觉对象缺失，或数据分析存在重大问题=0 分。				
合计配分	10	合计得分				

五、设计网店数据可视化方案（考核时间：15 min）

1. 试题单

(1) 场地设备要求

- 1) 装有 Windows 操作系统的计算机 (Windows 7 及以上);
- 2) 系统中已安装数据处理软件、压缩软件及办公软件;
- 3) 商务数据分析师考试平台。

(2) 工作任务

- 1) 根据业务需求及公司数据, 完成数据可视化实现方式的设计任务;
- 2) 根据业务场景及视觉对象特点, 完成可视化展板的设计与布局规划工作;
- 3) 根据业务要求及分析结果, 完成数据展示方案的制定与呈现。

(3) 技能要求

- 1) 设计数据可视化实现方式;
- 2) 设计与布局可视化展板;
- 3) 结合业务要求和分析成果, 制定并呈现数据展示方案。

(4) 质量指标

- 1) 精准把握数据特点并识别关键数据维度与属性特征, 合理构建可视化呈现形式;
- 2) 正确分析业务场景和视觉对象特征, 合理设计可视化展板架构, 页面布局科学高效;
- 3) 准确解析业务需求内核, 准确提炼分析结果要点, 合理构建数据展示方案, 语句通顺, 逻辑清晰。

背景材料

潮搭衣橱是一家新成立的线上潮流服饰店铺, 主营年轻时尚风格的服装及配饰, 目标客群为 18-35 岁的都市青年。店铺开业初期需要重点关注访客数、浏览量、平均停留时长等流量指标, 以及加购人数、下单转化率、支付买家数等转化数据, 同时二次访问率和购物车放弃率等用户行为指标也能帮助评估用户粘性和购买体验。通过对这些核心运营数据的系统分析, 可以全面了解店铺当前的经营状况, 发现潜在问题并制定针对性的优化策略。

作为店铺运营负责人, 孙海东需要深入分析这些数据指标之间的关联性, 比如访客质量与转化率的关系、停留时长对加购行为的影响等。通过监测数据, 定期追踪关键指标的变化趋势, 结合用户反馈和市场动态, 不断调整商品展示和页面设计, 从而提升整体运营效率, 推动店铺业绩的稳步增长。

操作要求:

(1) 通过多维度数据可视化展示网店运营状况, 直观呈现关键指标, 为经营决策提供数据支持;

(2) 基于业务场景特性与数据特征, 运用前沿的数据可视化设计方法论, 打造兼具美学价值与商业洞察力的交互式智能看板;

(3) 基于业务分析与数据洞察, 定制开发符合实际业务场景的 PPT 数据展示方案, 驱动运营效能持续升级。(请提交操作完成后的数据文件至考试平台。)

2.评分表

测量分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	3	所选属性合理、形成 3 个正确的可视化视觉对象=3 分。				
		所选属性基本合理、形成至少 2 个正确的可视化视觉对象=2 分。				
		所选属性基本合理、形成 1 个正确的可视化视觉对象=1 分。				
		所选特征不合理并形成错误的可视化视觉对象=0 分。				
J2	2	设计展板色彩体系完整规范、图表类型丰富多样、布局科学合理=2 分。				
		设计展板色彩体系基本规范、图表类型较为单一、布局基本合理=1 分。				
		设计展板色彩混乱, 图表类型单一且不匹配数据特征, 布局杂乱无章=0 分。				
J3	1	设计方案框架中封面和目录完整、设计规范=1 分。				
		设计方案框架中封面或目录缺失、设计基本规范=0.5 分。				
		设计方案框架中缺失封面或目录、不规范=0 分。				
J4	1	业务需求提及数据关联性、追踪变化趋势、运营优化策略、决策支持等关键词				

		中的 2 个及以上，且描述准确=1 分。				
		业务需求提及数据关联性、追踪变化趋势、运营优化策略、决策支持等关键词中的 1 个，或描述基本准确=0.5 分。				
		业务需求未提及数据关联性、追踪变化趋势、运营优化策略、决策支持等关键词，或语句不通顺、毫无逻辑=0 分。				
J5	3	展示方案中呈现 3 个视觉对象，数据分析准确合理=3 分。				
		展示方案中呈现 2 个视觉对象，或数据分析基本合理=2 分。				
		展示方案中呈现 1 个视觉对象，或数据分析存在不足=1 分。				
		展示方案中视觉对象缺失，或数据分析存在重大问题=0 分。				
合计配分	10	合计得分				

六、开展家电营销数据可视化结果分析并实现客户留存（考核时间：10 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机 (Windows 7 及以上)，至少 8GB 内存，10GB 可用磁盘空间；
- 2) 系统中已安装数据处理软件、Python 3.8+、MySQL Server (8.0+ 版本)、压缩软件及办公软件；
- 3) 商务数据分析师考试平台。

（2）工作任务

- 1) 根据某公司业务需求，完成商务数据报告撰写任务；
- 2) 根据某公司运营数据，完成商务数据对比分析任务；
- 3) 根据某公司运营现状，完成商务数据驱动决策任务。

（3）技能要求

- 1) 能对数据可视化结果进行业务分析；
- 2) 能填写或阅读商务数据分析报表；
- 3) 能输出商务数据分析报告；

3) 能提出并撰写数据驱动决策。

(4) 质量指标

- 1) 根据相关数据，准确撰写商务数据分析报告或报表的内容，内容合理；
- 2) 准确解读工程文件中的商务数据，并对比分析相关数据，判断无误；
- 3) 根据相关数据，准确计算与描述分布率和占比等数据，数据精确；
- 4) 根据要求，针对忠实客户流失问题提出对策，对策合理、有效。

背景材料

上海海东商贸集团旗下拥有一家家电公司，经营 A、B、C、D、E 五种产品，开设了两家线下门店，并在天猫、京东和唯品会开设了三家网店，但由于本年度出现了经营决策失误，导致了部分忠诚客户的流失，集团对该问题非常重视，希望“海东家电”能够出具一份商务数据分析报告，你是“海东家电”的市场部经理，请撰写一份商务数据分析报告，报告的主体部分已经完成，但是，部分数据表和内容需要完善。

操作要求：

- (1) 请根据数据，简要描述 2016 年 11 月 15 日的基本客户情况，包括总客户数、忠实客户数、忠实客户新增数、忠实客户流失数；
 - (2) 已知公司忠实客户的流失率和流失女客户占比的“警戒线”分别为 20% 和 60%，请查找与对比相关数据，简要描述当前（即 2016 年 11 月）的忠实客户流失率和流失女客户占比，并判断是否超过“警戒线”。
 - (3) 请根据数据，分析会员类型分布情况，包括各种会员类型的人数、占比、最高占比和最低占比；
 - (4) 请根据流失客户数据及分析，提出 2 条忠诚客户流失的阻断策略。
- (注：所有比率保留两位小数)

上海海东家电有限责任公司 忠实客户流失数据分析报告

市场部 张三

一、基本客户情况

(1)

二、流失客户分析

(2)

在流失的忠实客户中，年龄段分布情况为：90后为941人，占比19.35%；80后为1258人，占比25.87%；70后为1159人，占比23.83%；60后为800人，占比16.45%；50后为705人，占比14.50%。其中，占比最高的是80后，最低的是50后。

在流失的忠实客户中，会员类型分布情况为：(3)

在流失的忠实客户中，最后消费品分布情况为：A商品为520件，占比10.69%；B商品为1422件，占比29.24%；C商品为1392件，占比28.62%；D商品为948件，占比19.49%；E商品为581件，占比11.95%。其中，占比最高的是B商品和C商品，最低的是A商品。

在流失的忠实客户中，最后消费渠道分布情况为：No1门店为671人，占比13.80%；No2门店为420人，占比8.64%；天猫为1594人，占比32.78%；京东为1713人，占比35.23%；唯品会为465人，占比9.56%。其中，占比最高的是天猫和京东，最低的是No2门店。

三、问题及原因分析

.....

四、提出对策

(4)

(请在考试系统相应位置填写答案。)

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	2	简要描述2016年11月15日的基本客户情况，包括总客户数、忠实客户数、忠实客户新增数、忠实客户流失数。	正确描述1项数据=0.5分，共4项；错误或未答题=0分。		
M2	2	在所提供的数据可视化结果文件中筛选与计算2016年11月的忠实客户流失率和流失女客户占比数据，并判断是否超过公司“警戒线”。	正确描述或判断1项数据=0.5分，共4项；错误或未答题=0分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M3	3	分析会员类型分布情况，包括各种会员类型的人数、占比、最高占比和最低占比。	正确分析一类会员数据=0.5分，共5类会员；正确分析最高占比和最低占比会员=0.5分；错误或未答题=0分。		
合计配分	7	合计得分			

评价分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	1.5	第一条建议合理有效、逻辑正确、具有创新性=1.5分。				
		第一条建议合理有效、逻辑正确=1分。				
		第一条建议基本合理=0.5分。				
		第一条建议不合理或未答题=0分。				
J2	1.5	第二条建议合理有效、逻辑正确、具有创新性=1.5分。				
		第二条建议合理有效、逻辑正确=1分。				
		第二条建议基本合理=0.5分。				
		第二条建议不合理或未答题=0分。				
合计配分	3	合计得分				

七、基于课程营销数据可视化图表提供销售决策建议（考核时间：10 min）

1. 试题单

（1）场地设备要求

1) 装有 Windows 操作系统的计算机 (Windows 7 及以上), 至少 8GB 内存, 10GB 可用磁盘空间;

2) 系统中已安装数据处理软件、Python 3.8+、MySQL Server (8.0+ 版本)、压缩软件及办公软件;

3) 商务数据分析师考试平台。

(2) 工作任务

1) 根据某企业业务需求, 完成商务数据分析报告或报表编写任务;

2) 根据某企业营销数据, 完成计算商务数据分析报告中的数据计算任务;

3) 根据某企业营销数据, 完成销售数据的直观分析和购物篮分析任务;

4) 根据某企业营销现状, 完成提出某产品营销问题的解决策略的任务。

(3) 技能要求

1) 能对数据可视化结果进行业务分析;

2) 能填写商务数据分析报表;

3) 能输出商务数据分析报告;

4) 能提出并撰写数据驱动决策。

(4) 质量指标

1) 准确撰写商务数据分析报告, 内容合理, 逻辑正确;

2) 准确计算销售额、平均销售额等数据, 数据精确;

3) 准确进行销售产品的购物篮分析, 逻辑合理, 内容详实;

4) 准确提出营销困境的解决对策, 逻辑合理, 对策实用。

背景材料

2023 年, 上海海东商贸集团旗下的“新海东”教育咨询有限公司新推出了三门数据分析类课程, 分别是“PP+DAX 数据建模课程”“PQ 数据查询课程”“可视化课程”, 由于内容丰富实用而备受学员青睐。因内部对账审计和销售情况分析需要, 现在需要每天对三门课程的销售额、销售量、销售人数等情况进行统计, 并形成商务数据分析课程销售数据报告, 报告的基本框架已经初步形成, 还有一些关键数据和营销决策内容需要填写。

操作要求:

(1) 已知同类公司商务数据分析课程的市场销售日平均总额为 15000 元, 请在销售数据概览中, 分析本公司三门商务数据分析课程的销售收入情况, 并判断销售总额的市场营销水平;

(2) 在销售数据概览中, 根据营业额数据, 填写平均订单金额的计算公式, 并计算平均每笔订单的金额 (结果保留两位小数);

- (3) 在销售数据概览中，根据数据，填写不同支付方式的支付金额分析；
- (4) 在数据分析中根据购物篮分析相关数据，填写可视化课程的购物篮分析；
- (5) 请针对当前课程销售的困境，提出 2 条有效的营销对策。

上海“新海东”教育咨询有限公司 商务数据分析课程销售数据报告

市场部 张三

一、引言

在数字化时代背景下，商务数据分析已成为企业决策的关键环节。本报告旨在通过分析商务数据分析课程的销售数据，揭示该领域的市场现状与未来趋势。商务数据分析涉及对大量商务数据的收集、处理和分析，以支撑企业的运营决策和战略规划。随着市场竞争的加剧，企业愈发重视数据驱动的决策模式，使得商务数据分析的需求持续增长。

二、市场需求分析

行业背景：随着互联网和电子商务的快速发展，企业面临的市场数据量呈爆炸式增长。这些数据来自多种来源，如社交媒体、客户反馈、销售记录等。有效地分析这些数据可以帮助企业更好地了解市场动态，优化产品和服务，提高市场竞争力。

需求概况：当前市场对商务数据分析的需求主要体现在以下几个方面：

数据驱动决策：企业需要通过数据分析来支持战略决策和日常运营，以提高决策的准确性和效率。

客户洞察：通过对客户行为数据的分析，企业可以更好地了解客户需求，进行精准营销。

运营优化：数据分析可以帮助企业发现运营中的问题，优化流程，降低成本。

风险管理：数据分析还可以帮助企业识别潜在风险，采取预防措施。

三、销售数据概览

数据来源：本次销售数据主要来源于内部数据库和电商平台后台的数据统计。所有数据均经过严格的筛选和验证，去除了退单数据，确保其真实性和可靠性。

数据概况：

总体销售额：当天，(1.填写三门商务数据分析课程的总销售额分析)，显示出市场对该课程的高度认可和需求旺盛。

订单数量：三门课程总体订单数量为 95 单，平均每笔订单金额计算方式及结果为 (2)，反映出消费者对课程价值的认同以及购买意愿的强烈。

支付方式： (3.填写不同支付方式下的支付金额分析)

用户画像：购买者主要以在职人员和大学生为主，年龄集中在 20-40 岁之间。他们大多

个性化需求增加：随着客户对商务数据分析认识的加深，他们对课程的个性化需求越来越高，期望课程能够针对自己的特定需求进行定制。然而，目前我们的课程体系还不够灵活，难以完全满足客户的个性化需求。

技术更新迅速：商务数据分析技术日新月异，客户对课程内容的时效性和前沿性提出了更高要求。如何及时更新课程内容以保持竞争力成为一大难题。

...

六、提出对策

(5.提出 2 条课程营销对策。)

七、结论

根据对商务数据分析课程销售数据的分析可以看出该课程在市场上具有较高的需求且需求持续增长。然而我们也面临着数据质量问题、激烈的市场竞争以及客户需求不断变化等挑战。通过对销售数据的深入分析我们了解到客户主要对数据驱动决策、客户洞察、运营优化和风险管理等方面感兴趣。这些信息为企业制定营销策略提供了有力支持。

(请在考试系统相应位置填写答案。)

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分	
M1	1	准确查找并定位相关数据，正确分析三门商务数据分析课程的销售情况，并判断销售总额的市场营销水平。	销售额分析正确=0.5分；市场水平判断准确=0.5分；错误或未答题=0分。			
M2	1	正确填写平均订单金额的计算公式，并计算平均每笔订单的金额。	公式正确=0.5分；计算金额正确=0.5分；错误或未答题=0分。			
M3	2	正确填写不同支付方式的支付金额分析。	正确分析一项=0.5分，共4处；错误或未答题=0分。			
M4	2	正确填写可视化课程的购物篮分析。	正确分析一处数据=0.5分；共4处；错误或未答题=0分。			
合计配分	6	合计得分				

评价分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	2	第一条对策合理有效、逻辑正确、具有创新性=2分。				
		第一条对策合理有效、逻辑正确=1分；				
		第一条对策基本合理=0.5分；				
		第一条对策不合理或未答题=0分。				
J2	2	第二条对策合理有效、逻辑正确、具有创新性=2分。				
		第二条对策合理有效、逻辑正确=1分；				
		第二条对策基本合理=0.5分；				
		第二条对策不合理或未答题=0分。				
合计配分	4	合计得分				

八、实施合同管理数据视图监控并开展业务流程优化咨询（考核时间：10 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机(Windows 7 及以上), 至少 8GB 内存, 10GB 可用磁盘空间;
- 2) 系统中已安装数据处理软件、Python 3.8+、MySQL Server (8.0+ 版本)、压缩软件及办公软件;
- 3) 商务数据分析师考试平台。

（2）工作任务

- 1) 根据某公司业务需求, 完成商务数据分析报告编写任务;

- 2) 根据某公司运营数据，完成商务数据统计表填写任务；
- 3) 根据某公司运营现状，完成提出运营问题对策的任务。

(3) 技能要求

- 1) 能对数据可视化结果进行业务分析；
- 2) 能填写商务数据分析报表；
- 3) 能输出商务数据分析报告；
- 4) 能提出并撰写数据驱动决策。

(4) 质量指标

- 1) 准确解读与描述商务数据工程文件中的基本业绩数据，数据精确；
- 2) 准确解读某公司业务需求，完善商务数据分析报告的内容；
- 3) 准确根据某公司合同进度问题，提出有效的对策，逻辑合理。

背景材料

“海东物流”是上海海东商贸集团旗下的一家主营进出口运输与内贸运输的物流公司。2021年2月，集团想要对海东物流的业务效率和收益率进行了解，并根据合作情况，调整合作规模。附件为2020年1月至2021年1月，海东物流与合作客户的合作业绩数据。你是海东物流的公司战略运营，请完成一份海东物流的合同进度数据分析报告。报告的主体内容已经拟定完成，还有一些关键数据和营销决策内容需要填写。

操作要求：

- (1) 请完善合同进度数据分析报告的基础业绩数据分析，需包括总合同数量(数据单位转换至万吨)、合同金额(数据单位转换至亿元)、合同计数和合作企业数量等数据；
- (2) 请在合同进度数据分析报告的基础业绩数据统计表中(表1)，并填写合同数量大于10个的合作企业、合同数量小于3个的合作企业；
- (3) 请在进度未完成的合同进度数据统计表中(表2)，填写“六安”提货中的任务编号；
- (4) 为了加快合同完成速度，请提出2条提升仓库出货效率的建议。

上海海东物流有限责任公司 合同进度数据分析报告

总经办 李四

	未开始	—
金鼎	提货中	KSDL2012338; KSDL2012330
	未开始	—
仙福	提货中	—
	未开始	—
恒威	提货中	—
	未开始	—
济钢	提货中	—
	未开始	KSDL2101013
建邦	提货中	—
	未开始	—
玉昆	提货中	—
	未开始	KSDL2101009;

三、问题分析

.....

四、对策分析

.....

五、小结

.....

(请在考试系统相应位置填写答案。)

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	2	正确完善合同进度数据分析报告的基础业绩数据分析。	正确填写一项数据=0.5分，共4处；错误或未答题=0分。		
M2	2	正确筛选与查找数据，填写合同数量大于10个的合作企业。	正确填写一项=0.5分，共4项；错误或未答题=0分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M3	1.5	正确筛选与查找数据，填写合同数量小于3个的合作企业。	正确填写一项=0.5分，共3项；错误或未答题=0分。		
M4	1.5	正确筛选与查找数据，填写“六安”提货中的任务编号。	正确填写一项数据=0.5分，共3处；错误或未答题=0分。		
合计配分	7	合计得分			

评价分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	1.5	第一条建议合理有效、逻辑正确、具有创新性=1.5分。				
		第一条建议合理有效、逻辑正确=1分；				
		第一条建议基本合理=0.5分；				
		第一条建议不合理或未答题=0分。				
J2	1.5	第二条建议合理有效、逻辑正确、具有创新性=1.5分。				
		第二条建议合理有效、逻辑正确=1分；				
		第二条建议基本合理=0.5分；				
		第二条建议不合理或未答题=0分。				
合计配分	3	合计得分				

九、解读库存数据可视化图表并优化商品仓储模式（考核时间：10 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机 (Windows 7 及以上), 至少 8GB 内存, 10GB 可用磁盘空间;
- 2) 系统中已安装数据处理软件、Python 3.8+、MySQL Server (8.0+ 版本)、压缩软件及办公软件;
- 3) 商务数据分析师考试平台。

（2）工作任务

- 1) 根据某公司业务需求, 完成商务数据分析报告编写任务;
- 2) 根据某公司库存数据, 完成商务数据统计表填写任务;
- 3) 根据某公司库存现状, 完成提出库存问题解决对策任务。

（3）技能要求

- 1) 能对数据可视化结果进行业务分析;
- 2) 能填写商务数据分析报表;
- 3) 能输出商务数据分析报告;
- 4) 能提出并撰写数据驱动决策。

（4）质量指标

- 1) 准确解读与填写商务数据工程文件中的库存数据, 数据精确;
- 2) 准确根据某企业需求, 完善库存数据分析报告, 内容合理;
- 3) 准确根据某企业库存问题, 提出调整库存问题的思路, 逻辑合理。

背景材料

“海东物流”是上海海东商贸集团旗下的一家主营进出口运输与内贸运输的物流公司。2021年2月, 集团想要对海东物流的业务效率和收益率进行了解, 并根据库存情况, 调整运输工作安排。附件为2020年1月至2021年1月, 海东物流与客户的合作业绩数据。你是海东物流的公司物流运营主管, 请完成一份海东物流的库存数据分析报告。报告的基本框架已经初步形成, 但还有一些关键数据和营销决策内容需要填写。

操作要求:

(1) 请根据库存分析数据, 在表 1 中填写在手库存 (吨)、应收款 (元)、抗跌率 (%)、低抗跌率批次号 (<10%) 和高在手库存量客户 (>80,000 吨) 数据。

(2) 请根据 BI 工程文件库存分析中的饼图数据, 简要分析在港库存、在途库存和未发货库存的吨数和占比情况。

(3) 请根据当前在港库存和未发货库存占比较大的情况提出 2 条调整思路, 以调整整个库存情况。

上海海东物流有限责任公司 库存数据分析报告

总经办 李四

一、基本情况

从总体业绩上看, 本年度总合同数量为 1043 万吨, 合同金额为 77.4 亿元, 合同计数为 154 个, 合作企业数量为 11 个。合作企业包括旋力、锦亿、六安、开滦、澳森、金鼎、仙福、恒威、济钢、建邦、玉昆。深度合作的企业为锦亿、六安、开滦、金鼎, 签订合同数量超过 10 个。

.....

二、库存分析

表 1. 总体库存数据统计		
1	在手库存 (吨)	(1)
2	应收款 (元)	(2)
3	抗跌率 (%)	(3)
4	低抗跌率批次号 (<10%)	(4)
5	高在手库存量客户 (>80,000 吨)	(5)

在港库存、在途库存和未发货库存的吨数和占比情况分析: _____ (6)

调整库存情况的建议: _____ (7)

三、合同进度分析

总整体合同进度来看,

表 2.合同进度数据统计（进度未完成）		
合作方	合同进度	任务编号
旋力	提货中	—
	未开始	KSDL2012341
锦亿	提货中	KSDL2010281; KSDL2010283; KSDL2012335; KSDL2012348; KSDL2012386
	未开始	KSDL2101031; KSDL21010051;KSDL2101008
六安	提货中	KSDL2011323; KSDL20113101; KSDL2012379
	未开始	KSDL2101001; KSDL2101002
开滦	提货中	KSDL2012385
	未开始	—
金鼎	提货中	KSDL2012338; KSDL2012330
	未开始	—
济钢	提货中	—
	未开始	KSDL2101013
玉昆	提货中	—
	未开始	KSDL2101009;

四、问题分析

.....

五、提出对策

.....

六、小结

.....

(请在考试系统相应位置填写答案。)

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	0.5	正确填写在手库存（吨）。	是=0.5分；否=0分。		
M2	0.5	正确填写应收款（元）。	是=0.5分；否=0分。		
M3	0.5	正确填写抗跌率（%）。	是=0.5分；否=0分。		
M4	1.5	填写低抗跌率批次号（<10%）。	正确填写一项=0.5分，共3处；错误或未答题=0分。		
M5	2	正确填写高在手库存量客户（>80,000吨）。	正确填写一项=0.5分，共4处；错误或未答题=0分。		
M6	3	简要分析在港库存、在途库存和未发货库存的吨数和占比情况。	正确分析一处数据=0.5分，共6处；错误或未答题=0分。		
合计配分	8	合计得分			

评价分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	1	提出在港库存占比较大的调整思路，合理有效、逻辑正确、具有创新性=1分。				
		提出在港库存占比较大的调整思路，合理有效、逻辑正确=0.5分				
		提出在港库存占比较大的调整思路，不合理或未答题=0分。				
J2	1	提出未发货库存占比较大的调整思路，合理有效、逻辑正确、具有创新性=1分。				
		提出未发货库存占比较大的调整思路，合理有效、逻辑正确=0.5分				
		提出未发货库存占比较大的调整思路，不合理或未答题=0分。				

合计配分	2	合计得分	
------	---	------	--

十、旅游数据可视化结果追踪及解决业务开拓难点（考核时间：10 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机(Windows 7 及以上), 至少 8GB 内存, 10GB 可用磁盘空间;
- 2) 系统中已安装数据处理软件、Python 3.8+、MySQL Server (8.0+ 版本)、压缩软件及办公软件;
- 3) 商务数据分析师考试平台。

（2）工作任务

- 1) 根据某公司市场调研数据, 完成商务数据分析报表撰写;
- 2) 根据某公司业务需求, 完成商务数据分析报告内容撰写;
- 3) 根据某公司市场调研数据的解读, 提出未来业务发展策略。

（3）技能要求

- 1) 能对数据可视化结果进行业务分析;
- 2) 能填写商务数据分析报表;
- 3) 能输出商务数据分析报告;
- 4) 能提出并撰写数据驱动决策。

（4）质量指标

- 1) 准确解读商务数据, 撰写商务数据分析报告和报表, 数据精确无误;
- 2) 准确解读某公司业务需求, 完善商务数据分析报告, 内容合理、语句通顺;
- 3) 准确分析某一产业发展趋势, 提出合理、有效的企业发展规划路径。

背景材料

上海海东商贸集团旗下新开设了一家旅行社“上海海东文旅有限公司”, 由于公司旅游业务广泛、服务优质而备受消费者青睐。同时, 该企业善于运用各类数据为未来发展战略提供决策依据。为了开拓业务, 需要对旅行社的国内周边游业务进行规划。你作为该公司的市场部战略规划顾问, 收集了 2010 年至 2017 年的国内周边游旅游数据大盘, 请根据该数据大盘制订商务数据分析报告框架, 并撰写报告内容, 以提交集团高层完成最终业务决策。

操作要求:

- (1) 请根据调研数据,在调研数据分析部分,分析近八年间全国游客周边游的出行现状,需包括出游总人数、平均预定周期和平均游玩天数等数据;
- (2) 分析国内八年间周边游热门景点运营现状,需包括 3-4 个最热门的景区景点、景点的类型、吸引的人群等数据;
- (3) 分析山东省近八年周边游人数的增减趋势,并描述顶峰时的收入数据(万元)和最热门景点;
- (4) 请对上海海东文旅有限公司国内周边游业务的开拓提供 2 条建议。

上海海东文旅有限公司 国内周边游业务数据分析报告

市场部 张三

一、引言

近年来,周边游业务在市场上发展迅猛。上海海东文旅有限公司立足上海这一区位优势,深度整合各地文化资源与现代化旅游服务,在国内周边游市场占据了独特地位。随着文旅产业向高质量、高效益方向转型,消费者需求从传统观光向深度文化体验转变,公司面临着产品创新与数字化转型的双重挑战。本报告通过系统分析市场动态、客群特征及业务表现,旨在为公司在产品优化、精准营销及全国化扩张方面提供数据支撑与策略指导。

二、调研背景及意义

(一) 行业发展趋势

中国观光休闲行业正经历结构性变革,市场规模预计将以 8% 的年复合增长率持续扩张,到 2030 年有望达到万亿元级规模。这一增长主要源于三大驱动力:消费升级背景下体验经济的崛起、年轻客群对文化沉浸式旅游的偏好深化,以及数字化转型对服务链路的全面重构。同时,行业竞争格局呈现高度集中态势,2024 年文旅融合行业 CR3(前三家企业市场份额合计)已超过 46.73%,头部企业通过资源整合与品牌输出加速全国布局。在此背景下,区域文旅企业需依托独特文化 IP 与精细化运营构建差异化壁垒,避免同质化竞争。

(二) 区域资源与政策机遇

海东公司坐拥周边文化旅游资产,近年来,地方政府通过政策与资金双向发力助推产业升级:一方面争取省级文旅引导资金 2 千余万元,实施旅行社纾困、景区优惠等政策;另一方面推动智慧文旅平台建设,整合物联网、云计算技术提升行业监管效能与公共服务水平。2023 年一季度数据印证了政策效果,海东公司接待游客超百万人次,旅游收入达 7 亿余元,增速位居全市首位。

三、调研目标及内容

(一) 目标分析

本调研围绕四大核心目标展开：一是量化评估周边游市场份额与营收贡献，定位主力产品线与区域；二是解构客源画像，包括人口属性、消费偏好及行为轨迹；三是诊断产品竞争力，识别爆款特质与运营短板；四是评估数字化工具的渗透率与效能缺口。

(二) 调研内容框架

市场规模分析.....

客群特征调研.....

景点发展调研.....

四、调研数据分析

(1.分析近八年间全国游客周边游的出行现状)

(2.分析国内周边游热门景点运营现状)

(3.分析山东省近八年周边游人数的增减趋势及顶峰时收入数据和最热门景点。)

三、问题分析

.....

四、对策分析

.....

五、结论

因此，未来海东公司的发展方向在于...预期目标为...

(请将答案填写在考试系统相应位置，并点击提交。)

2.评分表

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
------	----	--------	--------	--------	----

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	1.5	正确分析近八年间全国游客周边游的出行现状。	正确分析一项数据=0.5分，共3项；错误或未答题=0分。		
M2	2	正确分析国内周边游热门景点运营现状。	正确分析最热门景点=1分；正确分析景点类型0.5分；正确分析吸引的人群=0.5分；错误或未答题=0分。		
M3	3.5	正确分析山东省近八年周边游人数的增减趋势，并描述顶峰时的收入数据和最热门景点。	正确分析增减趋势=2分；正确分析峰值年份与收入数据=1分；正确分析最热门景点=0.5分；错误或未答题=0分。		
合计配分	7	合计得分			

评价分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	1.5	第一条建议合理有效、逻辑正确、具有创新性=1.5分。				
		第一条建议合理有效、逻辑正确=1分。				
		第一条建议基本合理=0.5分。				
		第一条建议不合理或未答题=0分。				
J2	1.5	第二条建议合理有效、逻辑正确、具有创新性=1.5分。				

		第二条建议合理有效、逻辑正确=1分。				
		第二条建议基本合理=0.5分。				
		第二条建议不合理或未答题=0分。				
合计配分	3	合计得分				

商务分类数据分析、数据申报和报告缮制

一、挖掘精密制造数据并研究竞争态势，缮制商务数据分析报告（考核时间：25 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装数据处理软件、压缩软件及办公软件；
- 3) 数据库及商务数据分析师考试平台。

（2）工作任务

- 1) 根据相关数据，完成市场行情分析的工作任务；
- 2) 根据相关数据，完成行业状况分析的工作任务；
- 3) 根据相关数据，完成竞争对手分析的工作任务；
- 4) 根据相关数据，完成竞争产品分析的工作任务；
- 5) 根据公司业务需求，完成撰写商务数据分析报告的工作任务。

（3）技能要求

- 1) 进行市场行情分析；
- 2) 进行行业状况分析；
- 3) 进行竞争对手分析；
- 4) 进行竞争产品分析；
- 5) 撰写商务数据分析报告。

（4）质量指标

- 1) 基于行业数据特征深度解析，精准匹配可视化分析范式，系统性分析市场行情；
- 2) 准确理解相关给定数据，正确选用恰当的可视化视觉对象，对行业情况进行

分析；

3) 基于行业数据特征深度解读，精准匹配可视化分析范式，精准分析竞争对手情况；

4) 基于行业数据特征深度解读，精准匹配可视化分析范式，精准分析竞争产品情况；

5) 基于多维度可视化视觉对象分析结果，撰写具备战略穿透力的行业商务分析报告，数据-结论-策略三级逻辑闭环，语句通顺，逻辑合理。

背景资料

精密制造业正经历技术迭代加速+跨界竞争渗透的双重挤压，行业生态呈现攻防易位的动态博弈：一方面，技术扩散周期缩短至 3-5 年，催生近 300 家初创企业（2021-2023 年全球融资额超 120 亿美元）以颠覆性创新切入细分赛道，对传统巨头形成技术奇袭；另一方面，替代品威胁从材料端、工艺端到模式端全链条渗透，叠加下游客户对“性能-成本-交付”三角的极致追求，推动行业年均价格战强度达 12%-15%；中国作为全球最大增量市场，在政策（“强基工程”专项补贴）、资本与产业协同（长三角精密制造集群）三重驱动下，正成为新老玩家混战的主战场，亟需通过数据化工具穿透技术替代风险、客户忠诚度衰减与成本结构异化，为竞争策略制定提供动态校准依据。

操作要求：

（1）请根据所提供的数据，选用一个恰当的可视化视觉对象，分析 2014 年至 2023 年间精密制造业的市场行情，可以基于供应商的数量、消费者偏好变化、市场份额等维度；

（2）请根据所提供的数据，选用一个恰当的可视化视觉对象，分析 2014 年至 2023 年间精密制造业的行业状况，可以基于新进入者数量与现有主要竞争者数量关系、替代品数量、供应商数量与现有竞争者数量等维度；

（3）请根据所提供的数据，选用一个恰当的可视化视觉对象，分析 2014 年至 2023 年间精密制造业的竞争对手情况，可以基于新进入者数量及市场份额、新进入企业等维度；

（4）请根据所提供的数据，选用一个恰当的可视化视觉对象，分析 2014 年至 2023 年间精密制造业的竞争产品情况，可以基于替代品市场份额、替代品年增长率等维度；

（5）基于提供的数据及以上数据分析，请为精密制造业书写一份商务数据分析

报告，报告需包含基本模块：标题、分析目的与意义、市场行情、行业情况、竞争对手、竞争产品分析、困境分析、策略分析、结论。要求数据-结论-策略三级逻辑闭环、量化结论支撑业务决策、可视化结论复用设计。（请提交操作完成后的数据文件至考试平台。）

2.评分表

测量分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	2	所选市场行情分析维度合理、适合并形成正确的可视化视觉对象=2分。				
		所选市场行情分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选市场行情分析维度不合理或形成错误的可视化视觉对象=0分。				
J2	2	所选行业状况分析维度合理、适合并形成正确的可视化视觉对象=2分。				
		所选行业状况分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选行业状况分析维度不合理或形成错误的可视化视觉对象=0分。				
J3	2	所选竞争对手分析维度合理、适合并形成正确的可视化视觉对象=2分。				
		所选竞争对手分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选竞争对手分析维度不合理或形成错误的可视化视觉对象=0分。				
J4	2	所选竞争产品分析维度合理、适合并形成正确的可视化视觉对象=2分。				
		所选竞争产品分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选竞争产品分析维度不合理或形成错误的可视化视觉对象=0分。				
J5	1	标题合理正确=1分。				

		标题不合理、错误=0分。				
J6	2	报告的目的与意义分析全面、逻辑合理=2分。				
		报告的目的与意义分析基本合理=1分。				
		报告的目的与意义分析不合理=0分。				
J7	1	分析市场行情全面、逻辑合理=1分。				
		分析市场行情基本合理=0.5分。				
		分析市场行情不合理=0分。				
J8	1	分析行业状况全面、逻辑合理=1分。				
		分析行业状况基本合理=0.5分。				
		分析行业状况不合理=0分。				
J9	1	分析竞争对手全面、逻辑合理=1分。				
		分析竞争对手基本合理=0.5分。				
		分析竞争对手不合理=0分。				
J10	1	分析竞争产品全面、逻辑合理=1分。				
		分析竞争产品基本合理=0.5分。				
		分析竞争产品不合理=0分。				
J11	2	分析行业困境全面、逻辑合理=2分。				
		分析行业困境基本合理=1分。				

		分析行业困境不合理=0分。				
J12	2	提出两条合理、有效、实用策略=2分。				
		提出一条合理、有效、实用策略=1分				
		提出一条基本合理策略=0.5分。				
		所提策略不合理=0分。				
J13	1	结论归纳全面=1分。				
		部分归纳=0.5分。				
		未归纳=0分。				
合计 配分	20	合计得分				

二、绘制旅游业图谱并破译关键数据, 撰写商务数据分析报告(考核时间: 25 min)

1. 试题单

(1) 场地设备要求

- 1) 装有 Windows 操作系统的计算机 (Windows 7 及以上);
- 2) 系统中已安装数据处理软件、压缩软件及办公软件;
- 3) 数据库及商务数据分析师考试平台。

(2) 工作任务

- 1) 根据相关数据, 完成市场行情分析的工作任务;
- 2) 根据相关数据, 完成行业状况分析的工作任务;
- 3) 根据相关数据, 完成竞争对手分析的工作任务;
- 4) 根据相关数据, 完成竞争产品分析的工作任务;
- 5) 根据公司业务需求, 完成撰写商务数据分析报告的工作任务。

(3) 技能要求

- 1) 进行市场行情分析;
- 2) 进行行业状况分析;

- 3) 进行竞争对手分析;
- 4) 进行竞争产品分析;
- 5) 撰写商务数据分析报告。

(4) 质量指标

- 1) 基于行业数据特征深度解析, 精准匹配可视化分析范式, 系统性分析市场行情;
- 2) 准确理解相关给定数据, 正确选用恰当的可视化视觉对象, 对行业情况进行分析;
- 3) 基于行业数据特征深度解读, 精准匹配可视化分析范式, 精准分析竞争对手情况;
- 4) 基于行业数据特征深度解读, 精准匹配可视化分析范式, 精准分析竞争产品情况;
- 5) 基于多维度可视化视觉对象分析结果, 撰写具备战略穿透力的行业商务分析报告, 数据-结论-策略三级逻辑闭环, 语句通顺, 逻辑合理。

背景资料

在当前全球经济格局不断演变的大背景下, 旅游行业却展现出了独特的韧性与活力, 持续蓬勃发展。随着人们生活水平的提高、交通基础设施的完善以及消费观念的转变, 旅游已经从过去的奢侈品逐渐转变为大众日常生活中的重要组成部分。

然而, 旅游行业在快速发展的过程中也面临着诸多挑战。市场竞争日益激烈, 新进入者和现有竞争者不断涌入, 价格战、营销战等现象时有发生, 企业的利润空间受到挤压。此外, 技术创新压力、成本上升等问题也给企业带来了不小的困扰。在此背景下, 对旅游市场进行全面且精准的行情分析显得尤为关键。通过把握消费趋势、游客行为以及国际市场动态, 企业能够精准捕捉市场动态, 深入挖掘旅游行业的潜在增长点和机遇, 从而为战略决策、产品优化及市场拓展提供强有力的数据支撑和洞察, 助力企业在激烈的市场竞争中脱颖而出, 实现可持续发展。

操作要求:

- (1) 请根据所提供的数据, 选用一个恰当的可视化视觉对象, 分析 2014 年至 2023 年间旅游业的市场行情, 可以基于入境游客的性别、年龄、六大洲游客数量、国际旅游收入等维度;

(2) 请根据所提供的数据，选用一个恰当的可视化视觉对象，分析 2014 年至 2023 年间旅游业的行业状况，可以基于国内城镇农村居民旅游花费、按国别细分入境游客、旅游业的游客数量发展等维度；

(3) 请根据所提供的数据，选用一个恰当的可视化视觉对象，分析 2014 年至 2023 年间旅游业的竞争对手情况，可以基于不同类别的交通工具的旅游、国内外游客花费等维度；

(4) 请根据所提供的数据，选用一个恰当的可视化视觉对象，分析 2014 年至 2023 年间旅游业的竞争产品情况，可以基于不同国别、不同交通工具旅游分类等维度；

(5) 基于提供的数据及以上数据分析，为旅游业书写一份商务数据分析报告，报告需包含基本模块：标题、分析目的与意义、市场行情、行业状况、竞争对手、竞争产品分析、困境分析、策略分析、结论。要求数据-结论-策略三级逻辑闭环、量化结论支撑业务决策、可视化结论复用设计。（请提交操作完成后的数据文件至考试平台）

2.评分表

测量分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	2	所选市场行情分析维度合理、适合并形成正确的可视化视觉对象=2分。				
		所选市场行情分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选市场行情分析维度不合理或形成错误的可视化视觉对象=0分。				
J2	2	所选行业状况分析维度合理、适合并形成正确的可视化视觉对象=2分。				
		所选行业状况分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选行业状况分析维度不合理或形成错误的可视化视觉对象=0分。				
J3	2	所选竞争对手分析维度合理、适合并形成正确的可视化视觉对象=2分。				

		所选竞争对手分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选竞争对手分析维度不合理或形成错误的可视化视觉对象=0分。				
J4	2	所选竞争产品分析维度合理、适合并形成正确的可视化视觉对象=2分。				
		所选竞争产品分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选竞争产品分析维度不合理或形成错误的可视化视觉对象=0分。				
J5	1	标题合理正确=1分。				
		标题不合理、错误=0分。				
J6	2	报告的目的与意义分析全面、逻辑合理=2分。				
		报告的目的与意义分析基本合理=1分。				
		报告的目的与意义分析不合理=0分。				
J7	1	分析市场行情全面、逻辑合理=1分。				
		分析市场行情基本合理=0.5分。				
		分析市场行情不合理=0分。				
J8	1	分析行业状况全面、逻辑合理=1分。				
		分析行业状况基本合理=0.5分。				
		分析行业状况不合理=0分。				
J9	1	分析竞争对手全面、逻辑合理=1分。				
		分析竞争对手基本合理=0.5分。				

		分析竞争对手不合理=0分。				
J10	1	分析竞争产品全面、逻辑合理=1分。				
		分析竞争产品基本合理=0.5分。				
		分析竞争产品不合理=0分。				
J11	2	分析行业困境全面、逻辑合理=2分。				
		分析行业困境基本合理=1分。				
		分析行业困境不合理=0分。				
J12	2	提出两条合理、有效、实用策略=2分。				
		提出一条合理、有效、实用策略=1分				
		提出一条基本合理策略=0.5分。				
		所提策略不合理=0分。				
J13	1	结论归纳全面=1分。				
		部分归纳=0.5分。				
		未归纳=0分。				
合计 配分	20	合计得分				

三、解构医疗战略脉络并深挖核心数据，输出商务数据分析报告（考核时间：25 min）

1. 试题单

(1) 场地设备要求

- 1) 装有 Windows 操作系统的计算机 (Windows 7 及以上);
- 2) 系统中已安装数据处理软件、压缩软件及办公软件;
- 3) 数据库及商务数据分析师考试平台。

(2) 工作任务

- 1) 根据相关数据, 完成市场行情分析的工作任务;
- 2) 根据相关数据, 完成行业状况分析的工作任务;
- 3) 根据相关数据, 完成竞争对手分析的工作任务;
- 4) 根据相关数据, 完成竞争产品分析的工作任务;
- 5) 根据公司业务需求, 完成撰写商务数据分析报告的工作任务。

(3) 技能要求

- 1) 进行市场行情分析;
- 2) 进行行业状况分析;
- 3) 进行竞争对手分析;
- 4) 进行竞争产品分析;
- 5) 撰写商务数据分析报告。

(4) 质量指标

- 1) 基于行业数据特征深度解析, 精准匹配可视化分析范式, 系统性分析市场行情;
- 2) 准确理解相关给定数据, 正确选用恰当的可视化视觉对象, 对行业情况进行分析;
- 3) 基于行业数据特征深度解读, 精准匹配可视化分析范式, 精准分析竞争对手情况;
- 4) 基于行业数据特征深度解读, 精准匹配可视化分析范式, 精准分析竞争产品情况;
- 5) 基于多维度可视化视觉对象分析结果, 撰写具备战略穿透力的行业商务分析报告, 数据-结论-策略三级逻辑闭环, 语句通顺, 逻辑合理。

背景资料

在当今社会, 随着人们健康意识的不断提升以及医疗技术的飞速发展, 医疗卫生行业正呈现出蓬勃发展的态势。然而, 与之相伴的是日益激烈的市场竞争环境。在这样的大背景下, 各大医院为了在市场中占据一席之地, 纷纷将提升医疗

技术水平、优化服务质量作为核心发展战略，不断加大投入、创新管理模式。与此同时，各医院也愈发关注自身在市场中的竞争力表现，力求精准把握自身的优势与不足，以便及时调整策略，增强市场竞争力。

近期，我们获取了一组极具价值的数据集，该数据集涵盖了2024年12月份的医院关键词查询情况。其中详细记录了关键词、医院名称、关键词所属科室以及竞争力分数等关键信息。这些数据犹如一面镜子，能够清晰地反映出各医院在不同科室领域的市场关注度和竞争力状况。通过对这十家医院竞争力进行深入分析，我们期望能够揭示各医院在不同科室的竞争力表现及其变化趋势，挖掘隐藏在数据背后的规律和潜在问题，为全面洞察医疗行业的竞争态势提供精准、可靠的依据，助力医院及相关机构在激烈的市场竞争中做出科学、合理的决策。

操作要求：

- (1) 请根据所提供的数据，选用一个恰当的可视化视觉对象，分析2024年12月各医院的市场行情，可以基于竞争力、竞争力分数、排名等维度；
- (2) 请根据所提供的数据，选用一个恰当的可视化视觉对象，分析2024年12月各医院的行业状况，可以基于竞争力排名、关键词、关键词所属科室等维度；
- (3) 请根据所提供的数据，选用一个恰当的可视化视觉对象，分析2024年12月各医院的竞争对手情况，可以基于对应医院与所属科室、关键词与对应医院等维度；
- (4) 请根据所提供的数据，选用一个恰当的可视化视觉对象，分析2024年12月各医院的竞争产品情况，可以基于所属科室、日期、小时等维度；
- (5) 基于提供的数据及以上数据分析，请写一份商务数据分析报告，报告需包含基本模块：标题、分析目的与意义、市场行情、行业状况、竞争对手、竞争产品分析、困境分析、策略分析、结论。要求数据-结论-策略三级逻辑闭环、量化结论支撑业务决策、可视化结论复用设计。（请提交操作完成后的数据文件至考试平台。）

2.评分表

测量分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	2	所选市场行情分析维度合理、适合并形成正确的可视化视觉对象=2分。				

		所选市场行情分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选市场行情分析维度不合理或形成错误的可视化视觉对象=0分。				
J2	2	所选行业状况分析维度合理、适合并形成正确的可视化视觉对象=2分。				
		所选行业状况分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选行业状况分析维度不合理或形成错误的可视化视觉对象=0分。				
J3	2	所选竞争对手分析维度合理、适合并形成正确的可视化视觉对象=2分。				
		所选竞争对手分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选竞争对手分析维度不合理或形成错误的可视化视觉对象=0分。				
J4	2	所选竞争产品分析维度合理、适合并形成正确的可视化视觉对象=2分。				
		所选竞争产品分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选竞争产品分析维度不合理或形成错误的可视化视觉对象=0分。				
J5	1	标题合理正确=1分。				
		标题不合理、错误=0分。				
J6	2	报告的目的与意义分析全面、逻辑合理=2分。				
		报告的目的与意义分析基本合理=1分。				
		报告的目的与意义分析不合理=0分。				
J7	1	分析市场行情全面、逻辑合理=1分。				
		分析市场行情基本合理=0.5分。				

		分析市场行情不合理=0分。				
J8	1	分析行业状况全面、逻辑合理=1分。				
		分析行业状况基本合理=0.5分。				
		分析行业状况不合理=0分。				
J9	1	分析竞争对手全面、逻辑合理=1分。				
		分析竞争对手基本合理=0.5分。				
		分析竞争对手不合理=0分。				
J10	1	分析竞争产品全面、逻辑合理=1分。				
		分析竞争产品基本合理=0.5分。				
		分析竞争产品不合理=0分。				
J11	2	分析行业困境全面、逻辑合理=2分。				
		分析行业困境基本合理=1分。				
		分析行业困境不合理=0分。				
J12	2	提出两条合理、有效、实用策略=2分。				
		提出一条合理、有效、实用策略=1分				
		提出一条基本合理策略=0.5分。				
		所提策略不合理=0分。				
J13	1	结论归纳全面=1分。				

		部分归纳=0.5 分。				
		未归纳=0 分。				
合计 配分	20	合计得分				

四、扫描地区竞争全景并透视产业数据，制作商务数据分析报告（考核时间：25 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装数据处理软件、压缩软件及办公软件；
- 3) 数据库及商务数据分析师考试平台。

（2）工作任务

- 1) 根据相关数据，完成市场行情分析的工作任务；
- 2) 根据相关数据，完成行业状况分析的工作任务；
- 3) 根据相关数据，完成竞争对手分析的工作任务；
- 4) 根据相关数据，完成竞争产品分析的工作任务；
- 5) 根据公司业务需求，完成撰写商务数据分析报告的工作任务。

（3）技能要求

- 1) 进行市场行情分析；
- 2) 进行行业状况分析；
- 3) 进行竞争对手分析；
- 4) 进行竞争产品分析；
- 5) 撰写商务数据分析报告。

（4）质量指标

- 1) 基于行业数据特征深度解析，精准匹配可视化分析范式，系统性分析市场行情；
- 2) 准确理解相关给定数据，正确选用恰当的可视化视觉对象，对行业情况进行分析；
- 3) 基于行业数据特征深度解读，精准匹配可视化分析范式，精准分析竞争对手情况；
- 4) 基于行业数据特征深度解读，精准匹配可视化分析范式，精准分析竞争产品

情况；

5) 基于多维度可视化视觉对象分析结果，撰写具备战略穿透力的行业商务分析报告，数据-结论-策略三级逻辑闭环，语句通顺，逻辑合理。

背景资料

在当今数字化与信息化浪潮席卷全球的时代背景下，数据已成为洞察经济态势、把握行业脉搏的关键要素。对于地区经济发展而言，精准剖析各行业发展趋势与特征，不仅能够为政府制定科学合理的产业政策提供有力依据，助力区域经济结构优化升级，还能为企业战略决策提供重要参考，使其在激烈的市场竞争中找准方向、抢占先机。

在此背景下，我们聚焦于某特定地区，致力于深入探索并分析该地区在 2019 年至 2023 年这一关键时间段内各行业的发展轨迹与内在特征。为了实现这一目标，我们获取了一份涵盖该时间段内该地区公开的重点企业明细数据文件。这份数据文件犹如一座蕴藏着丰富经济信息的宝藏，其中每一行数据都对应着一家企业的详细信息，这些信息不仅包含了企业的基础属性，如企业名称、成立时间、所属行业分类等，还涵盖了企业在经营过程中产生的各类关键指标，例如营业收入、利润总额、员工数量、研发投入等。通过对这些海量且复杂的企业数据进行系统性的处理与深度分析，我们期望能够揭开该地区各行业在这五年间发展演变的神秘面纱，挖掘出隐藏在数据背后的行业发展规律、潜在机遇与挑战，为推动该地区经济的高质量发展贡献数据驱动的智慧力量。

操作要求：

(1) 请根据所提供的数据，选用一个恰当的可视化视觉对象，分析 2019 年至 2023 年各行业的市场行情，可以基于行业类型、公司个数、行业细分等维度；

(2) 请根据所提供的数据，选用一个恰当的可视化视觉对象，分析 2019 年至 2023 年各行业的行业状况，可以基于公司个数、注册资本、行业细分等维度；

(3) 请根据所提供的数据，选用一个恰当的可视化视觉对象，分析 2019 年至 2023 年“批发和零售业”行业的竞争对手情况，可以基于行业的注册资本、公司个数、行业细分等维度；

(4) 请根据所提供的数据，选用一个恰当的可视化视觉对象，分析 2019 年至 2023 年“批发和零售业”行业的竞争产品情况，可以基于行业代码、行业细分、公司个数等维度；

(5) 基于提供的数据及以上数据分析，请写一份商务数据分析报告，报告需包

含基本模块：标题、分析目的与意义、市场行情、行业状况、竞争对手、竞争产品分析、困境分析、策略分析、结论。要求数据-结论-策略三级逻辑闭环、量化结论支撑业务决策、可视化结论复用设计。（请提交操作完成后的数据文件至考试平台）

2.评分表

测量分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	2	所选市场行情分析维度合理、适合并形成正确的可视化视觉对象=2分。				
		所选市场行情分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选市场行情分析维度不合理或形成错误的可视化视觉对象=0分。				
J2	2	所选行业状况分析维度合理、适合并形成正确的可视化视觉对象=2分。				
		所选行业状况分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选行业状况分析维度不合理或形成错误的可视化视觉对象=0分。				
J3	2	所选竞争对手分析维度合理、适合并形成正确的可视化视觉对象=2分。				
		所选竞争对手分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选竞争对手分析维度不合理或形成错误的可视化视觉对象=0分。				
J4	2	所选竞争产品分析维度合理、适合并形成正确的可视化视觉对象=2分。				
		所选竞争产品分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选竞争产品分析维度不合理或形成错误的可视化视觉对象=0分。				
J5	1	标题合理正确=1分。				

		标题不合理、错误=0分。				
J6	2	报告的目的与意义分析全面、逻辑合理=2分。				
		报告的目的与意义分析基本合理=1分。				
		报告的目的与意义分析不合理=0分。				
J7	1	分析市场行情全面、逻辑合理=1分。				
		分析市场行情基本合理=0.5分。				
		分析市场行情不合理=0分。				
J8	1	分析行业状况全面、逻辑合理=1分。				
		分析行业状况基本合理=0.5分。				
		分析行业状况不合理=0分。				
J9	1	分析竞争对手全面、逻辑合理=1分。				
		分析竞争对手基本合理=0.5分。				
		分析竞争对手不合理=0分。				
J10	1	分析竞争产品全面、逻辑合理=1分。				
		分析竞争产品基本合理=0.5分。				
		分析竞争产品不合理=0分。				
J11	2	分析行业困境全面、逻辑合理=2分。				
		分析行业困境基本合理=1分。				

		分析行业困境不合理=0分。				
J12	2	提出两条合理、有效、实用策略=2分。				
		提出一条合理、有效、实用策略=1分				
		提出一条基本合理策略=0.5分。				
		所提策略不合理=0分。				
J13	1	结论归纳全面=1分。				
		部分归纳=0.5分。				
		未归纳=0分。				
合计 配分	20	合计得分				

五、追踪瓷砖竞争裂变并研判市场数据，生成商务数据分析报告（考核时间：25 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装数据处理软件、压缩软件及办公软件；
- 3) 数据库及商务数据分析师考试平台。。

（2）工作任务

- 1) 根据相关数据，完成市场行情分析的工作任务；
- 2) 根据相关数据，完成行业状况分析的工作任务；
- 3) 根据相关数据，完成竞争对手分析的工作任务；
- 4) 根据相关数据，完成竞争产品分析的工作任务；
- 5) 根据公司业务需求，完成撰写商务数据分析报告的工作任务。

（3）技能要求

- 1) 进行市场行情分析；

- 2) 进行行业状况分析;
- 3) 进行竞争对手分析;
- 4) 进行竞争产品分析;
- 5) 撰写商务数据分析报告。

(4) 质量指标

- 1) 基于行业数据特征深度解析, 精准匹配可视化分析范式, 系统性分析市场行情;
- 2) 准确理解相关给定数据, 正确选用恰当的可视化视觉对象, 对行业情况进行分析;
- 3) 基于行业数据特征深度解读, 精准匹配可视化分析范式, 精准分析竞争对手情况;
- 4) 基于行业数据特征深度解读, 精准匹配可视化分析范式, 精准分析竞争产品情况;
- 5) 基于多维度可视化视觉对象分析结果, 撰写具备战略穿透力的行业商务分析报告, 数据-结论-策略三级逻辑闭环, 语句通顺, 逻辑合理。

背景资料

在当今竞争激烈且快速迭代的瓷砖市场环境中, 产品多样化与创新性已成为企业赢得消费者青睐、抢占市场份额以及构筑核心竞争优势的核心要素。瓷砖市场的风云变幻, 不仅体现在消费者需求的多元化与个性化上, 更反映在竞品层出不穷的新产品、新技术以及灵活多变的策略之中。

海东企业作为瓷砖行业的领军企业, 始终保持敏锐的市场洞察力, 深刻认识到精准把握市场动态、全面剖析竞品表现, 对于制定科学有效的市场策略具有至关重要的意义。这不仅关乎企业能否在激烈的市场竞争中站稳脚跟, 更决定了其能否持续引领行业发展潮流, 实现长期稳健增长。

基于此, 海东企业主动出击, 积极构建全面的市场数据收集体系, 广泛收集市场上各类瓷砖产品的详尽信息, 对海量竞品数据进行深度挖掘与系统分析, 力求从产品特性、市场定位、销售趋势、消费者反馈等多个维度, 全面、深入地洞察竞品优势与不足, 精准把握市场动态与消费者需求变化趋势, 从而为企业制定更具前瞻性、针对性和竞争力的市场策略提供坚实的数据支撑, 助力企业在复杂多变的 market 环境中脱颖而出, 进一步提升市场份额与品牌影响力。

操作要求:

(1) 请根据所提供的数据，选用一个恰当的可视化视觉对象，分析瓷砖行业的市场行情，可以基于店铺类别、客户来源、竞品市场分类等维度；

(2) 请根据所提供的数据，选用一个恰当的可视化视觉对象，分析瓷砖行业的行业状况，可以基于客户性别、学历、婚姻状况、职业等维度；

(3) 请根据提供的数据，选用一个恰当的可视化视觉对象，分析瓷砖行业的竞争对手情况，可以基于店铺类别、粉丝销售利润、店铺数量、交易次数等维度；

(4) 请根据所提供的数据，选用一个恰当的可视化视觉对象，分析瓷砖行业的竞争产品情况，可以基于竞品功能、竞品尺寸、竞品销售、竞品的客户等维度；

(5) 基于提供的数据及以上可视化视觉对象数据分析，请写一份商务数据分析报告，报告需包含基本模块：标题、分析目的与意义、市场行情、行业状况、竞争对手、竞争产品分析、困境分析、策略分析、结论。要求数据-结论-策略三级逻辑闭环、量化结论支撑业务决策、可视化结论复用设计。（请提交操作完成后的数据文件至考试平台。）

2.评分表

测量分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	2	所选市场行情分析维度合理、适合并形成正确的可视化视觉对象=2分。				
		所选市场行情分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选市场行情分析维度不合理或形成错误的可视化视觉对象=0分。				
J2	2	所选行业状况分析维度合理、适合并形成正确的可视化视觉对象=2分。				
		所选行业状况分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选行业状况分析维度不合理或形成错误的可视化视觉对象=0分。				
J3	2	所选竞争对手分析维度合理、适合并形成正确的可视化视觉对象=2分。				

		所选竞争对手分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选竞争对手分析维度不合理或形成错误的可视化视觉对象=0分。				
J4	2	所选竞争产品分析维度合理、适合并形成正确的可视化视觉对象=2分。				
		所选竞争产品分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选竞争产品分析维度不合理或形成错误的可视化视觉对象=0分。				
J5	1	标题合理正确=1分。				
		标题不合理、错误=0分。				
J6	2	报告的目的与意义分析全面、逻辑合理=2分。				
		报告的目的与意义分析基本合理=1分。				
		报告的目的与意义分析不合理=0分。				
J7	1	分析市场行情全面、逻辑合理=1分。				
		分析市场行情基本合理=0.5分。				
		分析市场行情不合理=0分。				
J8	1	分析行业状况全面、逻辑合理=1分。				
		分析行业状况基本合理=0.5分。				
		分析行业状况不合理=0分。				
J9	1	分析竞争对手全面、逻辑合理=1分。				
		分析竞争对手基本合理=0.5分。				

		分析竞争对手不合理=0分。				
J10	1	分析竞争产品全面、逻辑合理=1分。				
		分析竞争产品基本合理=0.5分。				
		分析竞争产品不合理=0分。				
J11	2	分析行业困境全面、逻辑合理=2分。				
		分析行业困境基本合理=1分。				
		分析行业困境不合理=0分。				
J12	2	提出两条合理、有效、实用策略=2分。				
		提出一条合理、有效、实用策略=1分				
		提出一条基本合理策略=0.5分。				
		所提策略不合理=0分。				
J13	1	结论归纳全面=1分。				
		部分归纳=0.5分。				
		未归纳=0分。				
合计 配分	20	合计得分				

六、跨境电商贸易数据申报与波特五力分析（考核时间：20 min）

1. 试题单

（1）场地设备要求

1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；

2) 系统中已安装 BI 类商务数据分析软件、压缩软件、安装 Python3.8+、Python 库 (pandas、mysql-connector-python、matplotlib、seaborn)、Mysql8.0+及办公软件;

3) 商务数据分析师考试平台。

(2) 工作任务

- 1) 统计各平台的交易总额, 生成用于跨境电商数据申报的汇总数据;
- 2) 筛选符合退税条件的交易, 计算退税金额, 生成出口退税申报数据;
- 3) 统计各口岸的货运总量, 生成口岸物流申报数据;
- 4) 绘制波特五力模型图, 清晰展示跨境电商行业的竞争格局;
- 5) 分析跨境电商行业的竞争者、供应商、买家、替代品和新进入者, 生成详细的市场竞争报告。

(3) 技能要求

- 1) 采集跨境电子商务申报数据;
- 2) 采集出口退税申报数据;
- 3) 采集口岸物流申报数据;
- 4) 画出波特五力模型图;
- 5) 进行波特五力分析。

(4) 质量指标

- 1) 跨境电商交易总额统计准确, 符合申报要求;
- 2) 退税金额计算精确, 数据筛选正确;
- 3) 口岸货运总量统计准确, 数据完整;
- 4) 波特五力模型图清晰, 包含所有竞争维度;
- 5) 波特五力分析全面, 结论合理且基于数据。

背景资料

某跨境电子商务企业通过多个平台 (如亚马逊、阿里巴巴、易贝) 向全球销售商品, 涉及跨境电商数据申报、出口退税申报和口岸物流数据申报。数据存储于 MySQL 数据库, 包含交易和物流记录。为支持市场竞争分析, 需基于数据应用波特五力模型, 评估行业竞争格局。

操作要求:

- (1) 使用 SQL 聚合查询 transactions 表, 按 platform 分组统计总交易额 (total_amount), 并将结果保存至 platform_amount.csv, 金额需保留两位小数。

(2) 使用 SQL WHERE 子句筛选 transactions 表中符合出口退税条件的交易记录 (amount ≥ 1000.00 且 country 为“美国”、“日本”或“德国”), 并将结果保存至 tax_rebate_eligible.csv, 金额需保留两位小数。

(3) 基于筛选出的退税交易记录, 使用 Pandas 向量化操作计算每笔交易的退税金额 (tax_rebate, 为 amount 的 10%), 并将包含新列的结果保存至 tax_rebate_amount.csv, 金额需保留两位小数。

(4) 使用 SQL 聚合查询 logistics 表, 按 port 分组统计总货运重量 (total_weight), 并将结果保存至 port_weight.csv, 重量需保留两位小数。

(5) 使用 SQL 聚合查询 transactions 表, 按 country 分组统计交易笔数 (transaction_count), 并将结果保存至 country_transaction_count.csv。

(6) 基于交易总额数据, 生成并保存交易总额柱状图至 platform_amount_bar.png, 图像尺寸必须为 576x432 像素, 并确保中文标题 (“各平台交易总额”) 与坐标轴标签 (“平台”、“交易总额 (元)”) 显示正常无乱码

(7) 生成并保存波特五力模型图至 porter_five_forces.png, 图像尺寸必须为 576x432 像素, 图中必须包含分别指向“供应商”、“买家”、“替代品”、“新进入者”的箭头, 并确保所有中文标注显示正常无乱码。

(8) 基于已处理的交易和物流数据, 撰写一份包含对“竞争者”、“供应商”、“买家”、“替代品”、“新进入者”五个维度分析的市场竞争报告, 报告内容必须明确引用相关统计数据 (如百分比、金额、总量等), 并保存至 porter_analysis.txt, 数值需保留两位小数。

2.评分表

评价分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	2.5	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M1 - 交易总额统计成功”“M1 - 示例值: [值]”; 验证 platform_amount.csv 包含正确 platform 和 total_amount 列, 金额无缺失且保	结果正确, 且使用了 SQL 聚合查询的方式, 得 2.5 分。 结果正确, 但使用了先查询后在 Pandas 中聚合的方法, 得 1.5		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
		留两位小数。	分。 结果错误，得 0 分。		
M2	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M2 - 退税条件筛选成功”“M2 - 示例值: [值]”; 验证 tax_rebate_eligible.csv 包含正确 transaction_id, amount, country 列，满足金额≥1000.00 且国家在“美国”“日本”“德国”，金额保留两位小数。	结果正确，且使用了 SQL WHERE 子句进行筛选，得 2.5 分。 结果正确，但使用了先查询后在 Pandas 中筛选的方法，得 1.5 分。 结果错误，得 0 分。		
M3	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M3 - 退税金额计算成功”“M3 - 示例值: [值]”; 验证 tax_rebate_amount.csv 包含正确 transaction_id, tax_rebate 列，退税金额为金额的 10%，保留两位小数。	结果正确，且使用了 Pandas 向量化 (df['amount'] * 0.1) 的方式计算，得 2.5 分。 结果正确，但使用了 .apply 或循环的低效方法，得 1.5 分。 结果错误，得 0 分。		
M4	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M4 - 货运总量统计成功”“M4 - 示例值: [值]”; 验证 port_weight.csv 包含正确 port 和 total_weight 列，重量无缺失且保留两位小数。	结果正确，且使用了 SQL 聚合查询的方式，得 2.5 分。 结果正确，但使用了先查询后在 Pandas 中聚合的方法，得 1.5 分。 结果错误，得 0 分。		
M5	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M5 - 交易笔数统计成功”“M5 - 示例值: [值]”; 验证 country_transaction_count.csv 包含正确 country 和 transaction_count 列，笔数无缺失	结果正确，且使用了 SQL 聚合查询 (COUNT(*)) 的方式，得 2.5 分。 结果正确，但使用了先查询后在 Pandas 中聚合（如		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
		且正确。	groupby().size())的方法, 得 1.5 分。 结果错误, 得 0 分。		
M6	2.5	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M6 - 交易总额柱状图生成成功”“M6 - 图像尺寸: 576x432”; platform_amount_bar.png 存在, 尺寸为 576x432 像素 (figsize=(8, 6), dpi=72), 包含标题和正确柱状图, 中文无乱码, 金额显示保留两位小数。	结果正确, 且完全满足所有格式要求 (seaborn 或 matplotlib 均可), 得 2.5 分。 结果正确, 但未完全满足所有格式要求 (如尺寸不对、或缺少标题/标签), 得 1.5 分。 结果错误 (如图未生成、图表类型错误或中文乱码), 得 0 分。		
M7	2.5	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M7 - 波特五力模型图生成成功”“M7 - 图像尺寸: 576x432”; porter_five_forces.png 存在, 尺寸为 576x432 像素 (figsize=(8, 6), dpi=72), 包含标题、箭头和五力标注, 中文无乱码。	结果正确, 且完全满足所有格式要求 (包含所有标注、颜色框和箭头), 得 2.5 分。 结果正确, 但未完全满足所有格式要求 (如缺少箭头、或缺标注不全), 得 1.5 分。 结果错误 (如图未生成、或中文乱码), 得 0 分。		
M8	2.5	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M8 - 波特五力分析成功”; porter_analysis.txt 存在, 包含竞争者、供应商、买家、替代品、新进入者分析, 每项至少 50 字, 基于数据且结论合理, 数值保留两位小	结果正确, 且分析内容明确引用了数据 (如百分比、金额、总量等), 且满足所有其他要求, 得 2.5 分。 结果正确, 但分析内容泛泛而谈, 未明确		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
		数。	引用数据，或字数不足，得 1.5 分。 结果错误（如文件未生成、缺少分析维度），得 0 分。		
合计配分	20				

七、国际文化产品出口数据分析与波特五力模型应用（考核时间：20 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装 BI 类商务数据分析软件、压缩软件、安装 Python3.8+、Python 库（pandas、mysql-connector-python、matplotlib、seaborn）、Mysql8.0+及办公软件；
- 3) 商务数据分析师考试平台。

（2）工作任务

- 1) 按平台统计交易总额，生成跨境贸易申报数据；
- 2) 按平台计算可退税金额，生成出口退税申报数据；
- 3) 按口岸统计货运总重量，生成口岸物流申报数据；
- 4) 绘制波特五力模型图，展示竞争格局；
- 5) 基于交易和物流数据，分析行业竞争者、供应商、买家、替代品和新进入者的影响力，生成结构化的市场竞争分析报告。

（3）技能要求

- 1) 采集跨境电子商务申报数据；
- 2) 采集出口退税申报数据；
- 3) 采集口岸物流申报数据；
- 4) 画出波特五力模型图；
- 5) 进行波特五力分析。

(4) 质量指标

- 1) 交易总额统计准确，符合申报要求；
- 2) 可退税金额计算精确，数据完整；
- 3) 货运重量统计准确，反映物流规模；
- 4) 波特五力模型图清晰，包含所有竞争维度；
- 5) 竞争分析全面，结论合理且基于数据。

背景资料

您是一家文化产品出口企业的市场专员，负责采集和申报交易及物流数据，支持市场拓展和品牌战略。企业通过多个平台（如国际展会、电商平台）出口文化产品（如手工艺品、影视作品），涉及多口岸物流（如厦门、大连）。为制定战略规划，您需基于数据应用波特五力模型，分析文化产品行业竞争格局。

操作要求：

- (1) 使用 SQL 聚合查询 transactions 表，按 platform 分组统计总交易金额 (total_amount)，并将结果保存至 platform_amount.csv，金额需保留两位小数。
- (2) 使用 SQL 聚合计算 transactions 表中各平台的总可退税金额 (tax_refunded，为总金额的 10%)，并将结果保存至 platform_tax_refunded.csv，金额需保留两位小数。
- (3) 使用 SQL 聚合查询 logistics 表，按 port 分组统计总货运重量 (total_weight)，并将结果保存至 port_weight.csv，重量需保留两位小数。
- (4) 使用 SQL 聚合查询 transactions 表，按 country 分组统计交易笔数 (transaction_count)，并将结果保存至 country_transaction_count.csv。
- (5) 使用 SQL 聚合查询 transactions 表，按 product_category 分组统计交易总额 (total_amount)，并将结果保存至 product_amount.csv，金额需保留两位小数。
- (6) 生成并保存波特五力模型图至 porter_five_forces.png，图像尺寸必须为 576x432 像素，图中必须包含分别指向“供应商”、“买家”、“替代品”、“新进入者”的箭头，并确保所有中文标注显示正常无乱码。
- (7) 基于已处理的交易和物流数据，撰写一份包含对“竞争者”、“供应商”、“买家”、“替代品”、“新进入者”五个维度分析的市场竞争报告，报告内容必须明确引用相关统计数据（如百分比、金额、总量等），并保存至 porter_analysis.txt，数值需保留两位小数。

2.评分表

评价分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M1 - 交易总额统计成功”“M1 - 示例值: [值]”；验证 platform_amount.csv 包含正确 platform 和 total_amount 列，金额无缺失且保留两位小数。	<p>结果正确，且使用了 SQL 聚合查询的方式，得 3 分。</p> <p>结果正确，但使用了先查询后在 Pandas 中聚合的方法，得 2 分。</p> <p>结果正确，但使用了低效的 Python 循环处理，得 1 分。</p> <p>结果错误，得 0 分。</p>		
M2	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M2 - 可退税金额计算成功”“M2 - 示例值: [值]”；验证 platform_tax_refunded.csv 包含正确 platform 和 tax_refunded 列，退税金额为总额的 10%，保留两位小数。	<p>结果正确，且使用了 SQL 聚合计算（SUM(amount) * 0.1）的方式，得 3 分。</p> <p>结果正确，但使用了先查询后在 Pandas 中计算的方法，得 2 分。</p> <p>结果正确，但使用了低效的 Python 循环处理，得 1 分。</p> <p>结果错误，得 0 分。</p>		
M3	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M3 - 货运总量统计成功”“M3 - 示例值: [值]”；验证 port_weight.csv 包含正确 port 和 total_weight 列，重量无缺失且保留两位小数。	<p>结果正确，且使用了 SQL 聚合查询的方式，得 3 分。</p> <p>结果正确，但使用了先查询后在 Pandas 中聚合的方法，得 2 分。</p> <p>结果正确，但使用了</p>		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
			低效的 Python 循环处理，得 1 分。 结果错误，得 0 分。		
M4	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M4 - 交易笔数统计成功”“M4 - 示例值: [值]”；验证 country_transaction_count.csv 包含正确 country 和 transaction_count 列，笔数无缺失且正确。	结果正确，且使用了 SQL 聚合查询 (COUNT(*)) 的方式，得 2.5 分。 结果正确，但使用了先查询后在 Pandas 中聚合 (如 groupby().size()) 的方法，得 1.5 分。 结果错误，得 0 分。		
M5	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M5 - 产品类别交易总额统计成功”“M5 - 示例值: [值]”；验证 product_amount.csv 包含正确 product_category 和 total_amount 列，金额无缺失且保留两位小数。	结果正确，且使用了 SQL 聚合查询的方式，得 2.5 分。 结果正确，但使用了先查询后在 Pandas 中聚合的方法，得 1.5 分。 结果错误，得 0 分。		
M6	3	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M6 - 波特五力模型图生成成功”“M6 - 图像尺寸: 576x432”；porter_five_forces.png 存在，尺寸为 576x432 像素 (figsize=(8, 6), dpi=72)，包含标题、箭头和五力标注，中文无乱码。	结果正确，且完全满足所有格式要求 (包含所有标注、颜色框和箭头)，得 3 分。 结果正确，但未完全满足所有格式要求 (如缺少箭头、或标注不全)，得 2 分。 结果正确，但仅生成了基础图形，缺少大部分格式化元素，得 1 分。 结果错误 (如图未生		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
			成、或中文乱码), 得 0 分。		
M7	3	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M7 - 波特五力分析成功”; porter_analysis.txt 存在, 包含竞争者、供应商、买家、替代品、新进入者分析, 每项至少 50 字, 基于数据且结论合理, 数值保留两位小数。	结果正确, 且分析内容明确引用了数据(如百分比、金额、总量等), 且满足所有其他要求, 得 3 分。 结果正确, 但分析内容泛泛而谈, 未明确引用数据, 或字数不足, 得 2 分。 结果正确, 但仅完成了部分维度的分析, 得 1 分。 结果错误(如文件未生成、缺少分析维度), 得 0 分。		
合计配分	20	合计得分			

八、跨境电商数据合规与波特五力分析（考核时间：20 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装 BI 类商务数据分析软件、压缩软件、安装 Python3.8+、Python 库（pandas、mysql-connector-python、matplotlib、seaborn）、Mysql8.0+及办公软件；
- 3) 商务数据分析师考试平台。

（2）工作任务

- 1) 按平台统计交易总额，生成跨境电子商务申报数据；

- 2) 按平台计算可退税金额，生成出口退税申报数据；
- 3) 按口岸统计货运总重量，生成口岸物流申报数据；
- 4) 绘制波特五力模型图，展示竞争格局；
- 5) 基于交易和物流数据，分析行业竞争者、供应商、买家、替代品和新进入者的影响力，生成结构化的市场竞争分析报告。

(3) 技能要求

- 1) 采集跨境电子商务申报数据；
- 2) 采集出口退税申报数据；
- 3) 采集口岸物流申报数据；
- 4) 画出波特五力模型图；
- 5) 进行波特五力分析。

(4) 质量指标

- 1) 交易总额统计准确，符合申报要求；
- 2) 可退税金额计算精确，数据完整；
- 3) 货运重量统计准确，反映物流规模；
- 4) 波特五力模型图清晰，包含所有竞争维度；
- 5) 竞争分析全面，结论合理且基于数据。

背景资料

您是一家跨境电子商务企业的合规分析师，负责采集和申报交易及物流数据，确保符合海关和税务要求。企业通过多个平台（如亚马逊、阿里巴巴）开展跨境销售，涉及多口岸物流（如上海、深圳）。为制定战略规划，您需基于数据应用波特五力模型，分析电商行业竞争格局。

操作要求：

- (1) 使用 SQL 聚合查询 transactions 表，按 platform 分组统计总交易金额 (total_amount)，并将结果保存至 platform_amount.csv，金额需保留两位小数。
- (2) 使用 SQL 聚合计算 transactions 表中各平台的总可退税金额 (tax_refunded，为总金额的 10%)，并将结果保存至 platform_tax_refunded.csv，金额需保留两位小数。
- (3) 使用 SQL 聚合查询 logistics 表，按 port 分组统计总货运重量 (total_weight)，并将结果保存至 port_weight.csv，重量需保留两位小数。

(4) 使用 SQL 聚合查询 transactions 表，按 country 分组统计交易笔数 (transaction_count)，并将结果保存至 country_transaction_count.csv。

(5) 使用 SQL 聚合查询 transactions 表，按 product_category 分组统计交易总额 (total_amount)，并将结果保存至 product_amount.csv，金额需保留两位小数。

(6) 生成并保存波特五力模型图至 porter_five_forces.png，图像尺寸必须为 576x432 像素，图中必须包含分别指向“供应商”、“买家”、“替代品”、“新进入者”的箭头，并确保所有中文标注显示正常无乱码。

(7) 基于交易总额数据，生成并保存平台交易总额柱状图至 platform_amount_bar.png，图像尺寸必须为 576x432 像素，并确保中文标题 (“各平台交易总额”) 与坐标轴标签显示正常无乱码。

(8) 基于已处理的交易和物流数据，撰写一份包含对“竞争者”、“供应商”、“买家”、“替代品”、“新进入者”五个维度分析的市场竞争报告，报告内容必须明确引用相关统计数据 (如百分比、金额、总量等)，并保存至 porter_analysis.txt，数值需保留两位小数。

2.评分表

评价分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M1 - 交易总额统计成功”“M1 - 示例值: [值]”；验证 platform_amount.csv 包含正确 platform 和 total_amount 列，金额无缺失且保留两位小数。	结果正确，且使用了 SQL 聚合查询的方式，得 2.5 分。 结果正确，但使用了先查询后在 Pandas 中聚合的方法，得 1.5 分。 结果错误，得 0 分。		
M2	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M2 - 可退税金额计算成功”“M2 - 示例值: [值]”；验证 platform_tax_refunded.csv 包含正确 platform 和 tax_refunded 列，退税金额为总额的 10%，保留两位小数。	结果正确，且使用了 SQL 聚合计算 (SUM(amount) * 0.1) 的方式，得 2.5 分。 结果正确，但使用了先查询后在 Pandas		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
			中计算的方法,得 1.5 分。 结果错误,得 0 分。		
M3	2.5	考评员运行考生提交的 PY 文件,检查运行过程日志显示“M3 - 货运总量统计成功”“M3 - 示例值:[值]”;验证 port_weight.csv 包含正确 port 和 total_weight 列,重量无缺失且保留两位小数。	结果正确,且使用了 SQL 聚合查询的方式,得 2.5 分。 结果正确,但使用了先查询后在 Pandas 中聚合的方法,得 1.5 分。 结果错误,得 0 分。		
M4	2.5	考评员运行考生提交的 PY 文件,检查运行过程日志显示“M4 - 交易笔数统计成功”“M4 - 示例值:[值]”;验证 country_transaction_count.csv 包含正确 country 和 transaction_count 列,笔数无缺失且正确。	结果正确,且使用了 SQL 聚合查询 (COUNT(*)) 的方式,得 2.5 分。 结果正确,但使用了先查询后在 Pandas 中聚合 (如 groupby().size()) 的方法,得 1.5 分。 结果错误,得 0 分。		
M5	2.5	考评员运行考生提交的 PY 文件,检查运行过程日志显示“M5 - 产品类别交易总额统计成功”“M5 - 示例值:[值]”;验证 product_amount.csv 包含正确 product_category 和 total_amount 列,金额无缺失且保留两位小数。	结果正确,且使用了 SQL 聚合查询的方式,得 2.5 分。 结果正确,但使用了先查询后在 Pandas 中聚合的方法,得 1.5 分。 结果错误,得 0 分。		
M6	2.5	考评员运行考生提交的 PY 文件,检查运行过程日志显示“M6 - 波特五力模型图生成成功”“M6 - 图像尺寸:576x432”;porter_five_forces.png 存在,尺寸为 576x432 像素 (figsize=(8, 6), dpi=72),包含标题、箭头和五力标注,中文无乱码。	结果正确,且完全满足所有格式要求 (包含所有标注、颜色框和箭头),得 2.5 分。 结果正确,但未完全满足所有格式要求 (如缺少箭头、或标注不全),得 1.5 分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
			结果错误（如图未生成、或中文乱码），得0分。		
M7	2.5	考评员运行考生提交的PY文件，检查运行过程日志显示“M7 - 交易总额柱状图生成成功”“M7 - 图像尺寸：576x432”；platform_amount_bar.png存在，尺寸为576x432像素，包含标题、正确柱状图，中文无乱码，金额显示保留两位小数。	结果正确，且完全满足所有格式要求（seaborn或matplotlib均可），得2.5分。 结果正确，但未完全满足所有格式要求（如尺寸不对、或缺少标题/标签），得1.5分。 结果错误（如图未生成、图表类型错误或中文乱码），得0分。		
M8	2.5	考评员运行考生提交的PY文件，检查运行过程日志显示“M8 - 波特五力分析成功”；porter_analysis.txt存在，包含竞争者、供应商、买家、替代品、新进入者分析，每项至少50字，基于数据且结论合理，数值保留两位小数。	结果正确，且分析内容明确引用了数据（如百分比、金额、总量等），且满足所有其他要求，得2.5分。 结果正确，但分析内容泛泛而谈，未明确引用数据，或字数不足，得1.5分。 结果错误（如文件未生成、缺少分析维度），得0分。		
合计配分	20	合计得分			

九、国际制造业出口数据分析与波特五力模型应用（考核时间：20 min）

1. 试题单

（1）场地设备要求

1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；

2) 系统中已安装 BI 类商务数据分析软件、压缩软件、安装 Python3.8+、Python 库 (pandas、mysql-connector-python、matplotlib、seaborn)、Mysql8.0+及办公软件;

3) 商务数据分析师考试平台。

(2) 工作任务

- 1) 按渠道统计交易总额, 生成跨境贸易申报数据;
- 2) 按渠道计算可退税金额, 生成出口退税申报数据;
- 3) 按口岸统计货运总重量, 生成口岸物流申报数据;
- 4) 绘制波特五力模型图, 展示竞争格局;
- 5) 基于交易和物流数据, 分析行业竞争者、供应商、买家、替代品和新进入者的影响力, 生成结构化的市场竞争分析报告。

(3) 技能要求

- 1) 采集跨境电子商务申报数据;
- 2) 采集出口退税申报数据;
- 3) 采集口岸物流申报数据;
- 4) 画出波特五力模型图;
- 5) 进行波特五力分析。

(4) 质量指标

- 1) 交易总额统计准确, 符合申报要求;
- 2) 可退税金额计算精确, 数据完整;
- 3) 货运重量统计准确, 反映物流规模;
- 4) 波特五力模型图清晰, 包含所有竞争维度;
- 5) 竞争分析全面, 结论合理。

背景材料

您是一家国际制造业企业的贸易专员, 负责采集和申报出口交易及物流数据, 支持出口申报和供应链优化。企业通过多个贸易渠道 (如直接出口、代理商) 出口机械设备, 涉及多口岸物流 (如青岛、广州)。为制定战略规划, 您需基于数据应用波特五力模型, 分析制造业竞争格局。

操作要求:

- (1) 使用 SQL 聚合查询 transactions 表, 按 channel 分组统计总交易金额 (total_amount), 并将结果保存至 platform_amount.csv, 金额需保留两位小数。

- (2) 使用 SQL 聚合计算 transactions 表中各渠道的总可退税金额 (tax_refunded, 为总金额的 10%), 并将结果保存至 platform_tax_refunded.csv, 金额需保留两位小数。
- (3) 使用 SQL 聚合查询 logistics 表, 按 port 分组统计总货运重量 (total_weight), 并将结果保存至 port_weight.csv, 重量需保留两位小数。
- (4) 使用 SQL 聚合查询 transactions 表, 按 country 分组统计交易笔数 (transaction_count), 并将结果保存至 country_transaction_count.csv。
- (5) 使用 SQL 聚合查询 transactions 表, 按 product_category 分组统计交易总额 (total_amount), 并将结果保存至 product_amount.csv, 金额需保留两位小数。
- (6) 生成并保存波特五力模型图至 porter_five_forces.png, 图像尺寸必须为 576x432 像素, 图中必须包含分别指向“供应商”、“买家”、“替代品”、“新进入者”的箭头, 并确保所有中文标注显示正常无乱码。
- (7) 基于渠道交易总额数据, 生成并保存渠道交易总额柱状图至 platform_amount_bar.png, 图像尺寸必须为 576x432 像素, 并确保中文标题 (“各渠道交易总额”) 与坐标轴标签显示正常无乱码。
- (8) 基于已处理的交易和物流数据, 撰写一份包含对制造业“竞争者”、“供应商”、“买家”、“替代品”、“新进入者”五个维度分析的市场竞争报告, 报告内容必须明确引用相关统计数据 (如百分比、金额、总量等), 并保存至 porter_analysis.txt, 数值需保留两位小数。

2.评分表

评价分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	2.5	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M1 - 交易总额统计成功”“M1 - 示例值: [值]”; 验证 platform_amount.csv 包含正确 channel 和 total_amount 列, 金额无缺失且保留两位小数。	结果正确, 且使用了 SQL 聚合查询的方式, 得 2.5 分。 结果正确, 但使用了先查询后在 Pandas 中聚合的方法, 得 1.5 分。 结果错误, 得 0 分。		
M2	2.5	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M2 - 可退税	结果正确, 且使用了 SQL 聚合计算		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
		金额计算成功”“M2 - 示例值: [值]” ; 验证 platform_tax_refunded.csv 包含正确 channel 和 tax_refunded 列, 退税金额为总额的 10%, 保留两位小数。	(SUM(amount) * 0.1) 的方式, 得 2.5 分。 结果正确, 但使用了先查询后在 Pandas 中计算的方法, 得 1.5 分。 结果错误, 得 0 分。		
M3	2.5	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M3 - 货运总量统计成功”“M3 - 示例值: [值]”; 验证 port_weight.csv 包含正确 port 和 total_weight 列, 重量无缺失且保留两位小数。	结果正确, 且使用了 SQL 聚合查询的方式, 得 2.5 分。 结果正确, 但使用了先查询后在 Pandas 中聚合的方法, 得 1.5 分。 结果错误, 得 0 分。		
M4	2.5	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M4 - 交易笔数统计成功”“M4 - 示例值: [值]”; 验证 country_transaction_count.csv 包含正确 country 和 transaction_count 列, 笔数无缺失且正确。	结果正确, 且使用了 SQL 聚合查询 (COUNT(*)) 的方式, 得 2.5 分。 结果正确, 但使用了先查询后在 Pandas 中聚合 (如 groupby().size()) 的方法, 得 1.5 分。 结果错误, 得 0 分。		
M5	2.5	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M5 - 产品类别交易总额统计成功”“M5 - 示例值: [值]”; 验证 product_amount.csv 包含正确 product_category 和 total_amount 列, 金额无缺失且保留两位小数。	结果正确, 且使用了 SQL 聚合查询的方式, 得 2.5 分。 结果正确, 但使用了先查询后在 Pandas 中聚合的方法, 得 1.5 分。 结果错误, 得 0 分。		
M6	2.5	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M6 - 波特五	结果正确, 且完全满足所有格式要求 (包		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
		力模型图生成成功”“M6 - 图像尺寸: 576x432”; porter_five_forces.png 存在, 尺寸为 576x432 像素 (figsize=(8, 6), dpi=72), 包含标题、箭头和五力标注, 中文无乱码。	含所有标注、颜色框和箭头), 得 2.5 分。 结果正确, 但未完全满足所有格式要求 (如缺少箭头、或标注不全), 得 1.5 分。 结果错误 (如图未生成、或中文乱码), 得 0 分。		
M7	2.5	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M7 - 交易总额柱状图生成成功”“M7 - 图像尺寸 : 576x432” ; platform_amount_bar.png 存在, 尺寸为 576x432 像素, 包含标题、正确柱状图, 中文无乱码, 金额显示保留两位小数。	结果正确, 且完全满足所有格式要求 (seaborn 或 matplotlib 均可), 得 2.5 分。 结果正确, 但未完全满足所有格式要求 (如尺寸不对、或缺少标题/标签), 得 1.5 分。 结果错误 (如图未生成、图表类型错误或中文乱码), 得 0 分。		
M8	2.5	考评员运行考生提交的 PY 文件, 检查运行过程日志显示“M8 - 波特五力分析成功”; porter_analysis.txt 存在, 包含竞争者、供应商、买家、替代品、新进入者分析, 每项至少 50 字, 基于数据且结论合理, 数值保留两位小数。	结果正确, 且分析内容明确引用了数据 (如百分比、金额、总量等), 且满足所有其他要求, 得 2.5 分。 结果正确, 但分析内容泛泛而谈, 未明确引用数据, 或字数不足, 得 1.5 分。 结果错误 (如文件未生成、缺少分析维度), 得 0 分。		
合	20	合计得分			

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
计配分					

十、跨境农产品贸易数据分析与波特五力模型应用（考核时间：20 min）

1. 试题单

（1）场地设备要求

- 1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- 2) 系统中已安装 BI 类商务数据分析软件、压缩软件、安装 Python3.8+、Python 库（pandas、mysql-connector-python、matplotlib、seaborn）、Mysql8.0+及办公软件；
- 3) 商务数据分析师考试平台。

（2）工作任务

- 1) 按渠道统计交易总额，生成跨境贸易申报数据；
- 2) 按渠道计算可退税金额，生成出口退税申报数据；
- 3) 按口岸统计货运总重量，生成口岸物流申报数据；
- 4) 绘制波特五力模型图，展示竞争格局；
- 5) 基于交易和物流数据，分析行业竞争者、供应商、买家、替代品和新进入者的影响力，生成结构化的市场竞争分析报告。

（3）技能要求

- 1) 采集跨境电子商务申报数据；
- 2) 采集出口退税申报数据；
- 3) 采集口岸物流申报数据；
- 4) 画出波特五力模型图；
- 5) 进行波特五力分析。

（4）质量指标

- 1) 交易总额统计准确，符合申报要求；
- 2) 可退税金额计算精确，数据完整；
- 3) 货运重量统计准确，反映物流规模；
- 4) 波特五力模型图清晰，包含所有竞争维度；

5) 竞争分析全面，结论合理且基于数据。

背景材料

您是一家农产品出口企业的物流分析师，负责采集和申报交易及物流数据，支持贸易申报和物流规划。企业通过多个渠道（如合作社、国际买家）出口农产品（如水果、谷物），涉及多口岸物流（如天津、宁波）。为制定战略规划，您需基于数据应用波特五力模型，分析农产品贸易竞争格局。

操作要求：

- (1) 使用 SQL 聚合查询 transactions 表，按 channel（贸易渠道）分组统计总交易金额（total_amount），并将结果保存至 platform_amount.csv，金额需保留两位小数。
- (2) 使用 SQL 聚合计算 transactions 表中各渠道的总可退税金额（tax_refunded，为总金额的 10%），并将结果保存至 platform_tax_refunded.csv，金额需保留两位小数。
- (3) 使用 SQL 聚合查询 logistics 表，按 port 分组统计总货运重量（total_weight），并将结果保存至 port_weight.csv，重量需保留两位小数。
- (4) 使用 SQL 聚合查询 transactions 表，按 country 分组统计交易笔数（transaction_count），并将结果保存至 country_transaction_count.csv。
- (5) 使用 SQL 聚合查询 transactions 表，按 product_category 分组统计交易总额（total_amount），并将结果保存至 product_amount.csv，金额需保留两位小数。
- (6) 生成并保存波特五力模型图至 porter_five_forces.png，图像尺寸必须为 576x432 像素，图中必须包含分别指向“供应商”、“买家”、“替代品”、“新进入者”的箭头，并确保所有中文标注显示正常无乱码。
- (7) 基于渠道交易总额数据，生成并保存渠道交易总额柱状图至 platform_amount_bar.png，图像尺寸必须为 576x432 像素，并确保中文标题（“各渠道交易总额”）与坐标轴标签显示正常无乱码。
- (8) 基于已处理的交易和物流数据，撰写一份包含对农产品贸易“竞争者”、“供应商”、“买家”、“替代品”、“新进入者”五个维度分析的市场竞争报告，报告内容必须明确引用相关统计数据（如百分比、金额、总量等），并保存至 porter_analysis.txt，数值需保留两位小数。

2.评分表

评价分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M1 - 交易总额统计成功”“M1 - 示例值:[值]”；验证 platform_amount.csv 包含正确 channel 和 total_amount 列，金额无缺失且保留两位小数。	结果正确，且使用了 SQL 聚合查询的方式，得 2.5 分。 结果正确，但使用了先查询后在 Pandas 中聚合的方法，得 1.5 分。 结果错误，得 0 分。		
M2	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M2 - 可退税金额计算成功”“M2 - 示例值:[值]”；验证 platform_tax_refunded.csv 包含正确 channel 和 tax_refunded 列，退税金额为总额的 10%，保留两位小数。	结果正确，且使用了 SQL 聚合计算（SUM(amount) * 0.1）的方式，得 2.5 分。 结果正确，但使用了先查询后在 Pandas 中计算的方法，得 1.5 分。 结果错误，得 0 分。		
M3	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M3 - 货运总量统计成功”“M3 - 示例值:[值]”；验证 port_weight.csv 包含正确 port 和 total_weight 列，重量无缺失且保留两位小数。	结果正确，且使用了 SQL 聚合查询的方式，得 2.5 分。 结果正确，但使用了先查询后在 Pandas 中聚合的方法，得 1.5 分。 结果错误，得 0 分。		
M4	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M4 - 交易笔数统计成功”“M4 - 示例值:[值]”；验证 country_transaction_count.csv 包含正确 country 和 transaction_count 列，笔	结果正确，且使用了 SQL 聚合查询（COUNT(*)）的方式，得 2.5 分。 结果正确，但使用了		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
		数无缺失且正确。	先查询后在 Pandas 中聚合（如 <code>groupby().size()</code> ）的方法，得 1.5 分。 结果错误，得 0 分。		
M5	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M5 - 产品类别交易总额统计成功”“M5 - 示例值: [值]”；验证 <code>product_amount.csv</code> 包含正确 <code>product_category</code> 和 <code>total_amount</code> 列，金额无缺失且保留两位小数。	结果正确，且使用了 SQL 聚合查询的方式，得 2.5 分。 结果正确，但使用了先查询后在 Pandas 中聚合的方法，得 1.5 分。 结果错误，得 0 分。		
M6	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M6 - 波特五力模型图生成成功”“M6 - 图像尺寸: 576x432”； <code>porter_five_forces.png</code> 存在，尺寸为 576x432 像素（ <code>figsize=(8, 6)</code> , <code>dpi=72</code> ），包含标题、箭头和五力标注，中文无乱码。	结果正确，且完全满足所有格式要求（包含所有标注、颜色框和箭头），得 2.5 分。 结果正确，但未完全满足所有格式要求（如缺少箭头、或标注不全），得 1.5 分。 结果错误（如图未生成、或中文乱码），得 0 分。		
M7	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M7 - 交易总额柱状图生成成功”“M7 - 图像尺寸: 576x432”； <code>platform_amount_bar.png</code> 存在，尺寸为 576x432 像素，包含标题、正确柱状图，中文无乱码，金额显示保留两位小数。	结果正确，且完全满足所有格式要求（ <code>seaborn</code> 或 <code>matplotlib</code> 均可），得 2.5 分。 结果正确，但未完全满足所有格式要求（如尺寸不对、或缺少标题/标签），得 1.5 分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
			结果错误（如图未生成、图表类型错误或中文乱码），得 0 分。		
M8	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M8 - 波特五力分析成功”；porter_analysis.txt 存在，包含竞争者、供应商、买家、替代品、新进入者分析，每项至少 50 字，基于数据且结论合理，数值保留两位小数。	结果正确，且分析内容明确引用了数据（如百分比、金额、总量等），且满足所有其他要求，得 2.5 分。 结果正确，但分析内容泛泛而谈，未明确引用数据，或字数不足，得 1.5 分。结果错误（如文件未生成、缺少分析维度），得 0 分。		
合计配分	20	合计得分			

第5部分

理论知识考试模拟试卷及答案

商务数据分析师（三级）理论知识试卷

注 意 事 项

1. 考试时间：90 min。
2. 请首先按要求在试卷的标封处填写您的姓名、准考证号和所在单位的名称。
3. 请仔细阅读各种题目的回答要求，在规定的位置填写您的答案。
4. 不要在试卷上乱写乱画，不要在标封区填写无关的内容。

	一	二	三	总 分
得 分				

得 分	
评分人	

一、判断题（第1题~第40题。将判断结果填入括号中。正确的填“√”，错误的填“×”。每题0.5分，满分20分）

1. 计算机网络中的负载均衡功能是指将网络流量均匀分配到不同的网络设备上，以提高网络性能。（ ）
2. 数据库系统就是存储数据的仓库，不包括其他组成部分。（ ）
3. 数据库系统只能存储结构化数据，无法处理非结构化或半结构化数据。（ ）
4. 数据库系统的核心组成部分是数据库和数据库管理系统。（ ）
5. 统计学的研究对象仅限于数值型数据，不包括非数值型数据。（ ）
6. 统计学的基本特点之一是只关注数据的收集，而不关注数据的解释和应用。（ ）
7. 相关性分析可以确定两个变量之间的因果关系。（ ）
8. 数据分析的功能之一是帮助企业识别新的市场机会和潜在的增长点。（ ）
9. 数据分析的透明度意味着分析过程、方法和结果都应该清晰、可验证。（ ）
10. 相关性分析可以揭示两个变量之间的因果关系。（ ）
11. 离散分布只能描述离散变量，连续分布只能描述连续变量。（ ）
12. 数据管理仅仅是对数据进行存储和备份的过程。（ ）
13. 数据治理只是对数据进行管理和规范，与企业的业务战略无关。（ ）
14. 数据治理的第一个阶段主要是建立数据架构和组织，不涉及具体的数据管理活动。（ ）
15. 数据治理框架的实施只需要技术层面的支持，不需要高层管理的参与。（ ）

- 16 . 在数据流程优化过程中, 对流程进行定期评估和监控的目的是识别新的改进机会。
()
- 17 . 数据安全仅关注数据的机密性, 而不涉及数据的完整性和可用性。()
- 18 . 数据安全策略中的访问控制仅涉及对用户身份的验证, 而不包括对用户权限的管理。
()
- 19 . 数据审计功能可以记录和追踪数据的操作历史, 以便在发生安全事件时进行追溯和调查。()
- 20 . 在设计数据服务安全架构时, 不需要考虑数据的生命周期管理。()
- 21 . DLP 策略只能防止内部员工造成的数据泄露, 无法防御外部攻击。()
- 22 . 数据安全管理的可审计性原则要求对数据的操作进行记录和追踪, 以便在发生安全事件时能够追溯和调查。()
- 23 . 商品的使用价值是指其能够满足人们某种需要的属性。()
- 24 . 商品的核心利益是消费者购买商品时最直接和主要的需求满足点。()
- 25 . 在商品生命周期的衰退期应该积极收集客户反馈, 对商品进行持续改进和优化。()
- 26 . 营销组合策略(4Ps)包括产品、价格、促销和渠道。()
- 27 . 营销策划一旦确定, 就不需要再进行调整和修改, 以保持策略的稳定性。()
- 28 . 分销策略是市场营销策划中的关键环节, 它决定了如何将产品有效地送达目标消费者手中。()
- 29 . 商贸流通的广泛性意味着它只涉及国内市场的商品交换。()
- 30 . 分销模式总是包括多个层级的分销商。()
- 31 . 系统性原则要求企业管理中的各个部分都要相互协调, 形成一个有机整体。()
- 32 . 计划职能只是为企业设定长期目标, 不涉及日常运营。()
- 33 . 企业组织结构的设计应具有灵活性, 能够随企业战略和市场环境的变化而调整。()
- 34 . 成本控制就是简单地降低生产成本。()
- 35 . 《电子商务法》要求平等地对待线上线下的商务活动, 促进线上线下融合发展。()
- 36 . 《中华人民共和国保守国家秘密法》规定, 涉密人员在脱密期内, 可以以任何方式泄露国家秘密。()
- 37 . 《中华人民共和国数据安全法》要求数据处理者未经授权处理个人信息, 或处理超过个人信息范围的其他信息, 均要承担法律责任。()
- 38 . 《中华人民共和国个人信息保护法》对人脸识别的使用没有任何限制。()
- 39 . 《中华人民共和国网络安全法》不适用于个人用户在网络上的行为。()
- 40 . 即使未经电子信息接收者同意或请求, 也可以偶尔向其固定电话、移动电话或个人电子邮箱发送商业性电子信息。()

得 分	
评分人	

二、单项选择题(第 1 题~第 140 题。选择一个正确的答案, 将相应的字母填入题内的括号中。每题 0.5 分, 满分 70 分)

- 1 . 商品流通的主要模式不包括()。
(A) 直销模式 (B) 零售模式 (C) 多级分销模式 (D) 电子商务模式
- 2 . 企业管理中强调的“以人为本”原则主要体现在()。
(A) 重视员工发展与福利 (B) 追求最大利润 (C) 严格成本控制 (D) 强化技术革新

- 3 . 成本控制的首要原则是 () 。
(A) 最大化利润 (B) 最小化成本 (C) 提高生产效率 (D) 保持产品质量
- 4 . 《电子商务法》把调整范围严格限定在 () 这一区域内。
(A) 中华人民共和国境内 (B) 中华人民共和国境外 (C) 亚洲地区 (D) 欧美地区
- 5 . 《中华人民共和国电子商务法》主要适用于 () 。
(A) 实体零售 (B) 互联网信息服务 (C) 通过互联网等信息网络销售商品或提供服务 (D) 制造业
- 6 . 信息系统主要依赖 () 来收集、存储、处理、传输和展示信息。
(A) 手工 (B) 物理 (C) 电子 (D) 机械
- 7 . () 不是信息系统的基本功能。
(A) 自动化业务流程 (B) 社交娱乐 (C) 决策支持 (D) 信息集成
- 8 . 数据采集的首要步骤是 () 。
(A) 数据清洗 (B) 确定数据源 (C) 数据分析 (D) 数据存储
- 9 . () 适合捕捉社交媒体上的实时数据
(A) 批量导入 (B) API 接口 (C) 手工录入 (D) 日志文件分析
- 10 . 数据清洗的主要目的是 () 。
(A) 增加数据量 (B) 纠正数据错误 (C) 数据加密 (D) 数据压缩
- 11 . 数据质量控制的核心目的是 () 。
(A) 增加数据量 (B) 加快数据处理速度 (C) 确保数据的准确性和完整性 (D) 降低存储成本
- 12 . () 不是数据质量监控的常用方法。
(A) 设定数据质量标准 (B) 定期执行数据审计 (C) 实时数据流监控 (D) 数据备份与恢复
- 13 . 数据监测的主要目的是 () 。
(A) 娱乐休闲 (B) 为企业决策提供数据支持 (C) 监测员工工作效率 (D) 展示公司技术实力
- 14 . 在构建数据异常预警机制时, 最核心的要素是 () 。
(A) 数据存储介质 (B) 高效的异常识别算法 (C) 绚丽的可视化界面 (D) 定期的数据备份
- 15 . 实时数据监测的核心价值体现在 () 。
(A) 提升数据收集速度 (B) 保证数据的时效性, 以支持即时决策 (C) 节省数据存储成本 (D) 美化数据呈现
- 16 . 数据迁移过程中, 首要关注的是 () 。
(A) 迁移工具的成本 (B) 迁移操作的简便性 (C) 数据的完整性与一致性 (D) 网络延迟对迁移的影响
- 17 . 在云环境中, () 提供了最大的灵活性和可扩展性, 但可能伴随着更高的访问延迟。
(A) 对象存储 (B) 块存储 (C) 文件存储 (D) 网络附加存储
- 18 . 在进行数据备份时, () 通过记录自上次备份以来数据的所有更改来减少备份时间和存储需求。
(A) 完全备份 (B) 增量备份 (C) 差异备份 (D) 镜像备份
- 19 . () 用于灾难恢复, 允许在发生严重故障时快速恢复业务运。
(A) 数据压缩 (B) 数据加密 (C) 热备份 (D) 远程镜像

- 20 . 数据冗余的主要目的是 () 。
(A)减少数据存储成本 (B)提高数据访问速度 (C)增加数据的可靠性和容错性
(D)便于数据迁移
- 21 . () 可以在存储数据之前检测并消除重复数据块, 从而减少存储空间需求。
(A)数据压缩 (B)数据去重 (C)数据加密 (D)数据索引
- 22 . 在数据生命周期管理中, () 涉及数据的归档和长期存储。
(A)数据创建 (B)数据使用 (C)数据保留 (D)数据销毁
- 23 . 在数据安全中, 加密的主要目的是 () 。
(A)提高数据传输速度 (B)防止数据在未经授权的情况下被访问或泄露 (C)减少数据存储成本 (D)改进数据压缩效率
- 24 . () 使用单个密钥进行加密和解密。
(A)对称加密 (B)非对称加密 (C)哈希函数 (D)散列函数
- 25 . 数据压缩的主要目的是 () 。
(A)提高数据传输速度 (B)降低数据存储空间需求 (C)加密数据以保护隐私 (D)提升数据处理的准确性
- 26 . 数据访问控制的主要目的是 () 。
(A)确保数据的完整性 (B)提高数据传输效率 (C)防止未经授权的数据访问 (D)减少数据存储成本
- 27 . () 涉及将数据从临时存储位置转移到长期存储介质。
(A)数据采集(B)数据转换(C)数据迁移(D)数据备份
- 28 . 在数据存储流程中, () 是确保数据在存储前符合特定格式和标准。
(A)数据清洗(B)数据索引(C)数据压缩(D)数据加密
- 29 . 数据清洗需求分析的主要目的是 () 。
(A)选择合适的数据存储技术 (B)确定数据清洗的具体步骤 (C)识别并解决当前数据中的质量问题 (D)评估数据分析结果的准确性
- 30 . () 不属于数据清洗需求分析的内容。
(A)确定数据清洗的必要性和紧迫性 (B)设计数据清洗的算法细节 (C)识别数据中存在的缺失、错误和异常 (D)设定清洗后数据应达到的质量标准
- 31 . 在数据质量问题识别中, () 常用于发现数据中的异常值。
(A)数据可视化 (B)数据排序 (C)相关性分析 (D)完整性检查
- 32 . () 不是直接用于识别数据缺失。
(A)非空值验证 (B)逻辑规则检查 (C)聚类分析 (D)数据比对
- 33 . 在分析数据源时, 首先需要考虑 () 。
(A)数据清洗的难易程度 (B)数据源的可信度与权威性 (C)数据存储的硬件要求 (D)数据可视化工具的兼容性
- 34 . 数据结构分析的主要目的是 () 。
(A)确定数据是否需要清洗 (B)优化数据存储和检索效率 (C)验证数据的准确性 (D)转换数据格式以兼容不同系统
- 35 . 使用数据清洗工具时, 最重要的是 () 。
(A)追求最新最先进的技术 (B)深入了解工具的功能和限制 (C)立即开始清洗而不做计划 (D)依赖工具自动完成所有清洗工作
- 36 . 数据完整性分析主要关注的是 () 。
(A)数据中是否存在缺失值 (B)数据是否准确无误 (C)数据在不同系统间是否一致 (D)数据是否满足特定业务规则

- 37 . () 不是用于分析数据完整性。
 (A)检查数据字段是否为空 (B)验证数据类型的正确性 (C)比较数据记录之间的逻辑关系 (D)使用哈希函数检查数据一致性
- 38 . 数据准确性分析的核心是 ()。
 (A)检查数据是否完整 (B)验证数据是否符合业务规则 (C)比较数据与其他可靠来源的一致性 (D)评估数据清洗工具的性能
- 39 . 在 Excel 中, () 函数常用于统计缺失值的数量。
 (A)COUNT (B)COUNTIF (C)COUNTA (D)COUNTBLANK
- 40 . () 常用于识别数据中的异常值。
 (A)数据排序 (B)箱线图分析 (C)数据类型转换 (D)数据规范化
- 41 . 在统计 B 学中, () 用于检测单变量数据中的异常值。
 (A)T 检验 (B)方差分析 (C)Z 分数 (D)卡方检验
- 42 . 选择数据清洗工具时, () 不是主要考虑点。
 (A)工具的易用性和用户友好性 (B)工具支持的数据类型和格式 (C)工具的价格 (D)工具的自动化程度
- 43 . 在处理大规模数据集时, 选择数据清洗工具时应优先考虑 ()。
 (A)工具的图形界面 (B)工具的批处理能力 (C)工具的自定义脚本支持 (D)工具的实时数据处理能力
- 44 . 设定数据清洗规则时, 最重要的是 ()。
 (A)规则的复杂性 (B)规则的一致性 (C)规则的数量 (D)规则的执行速度
- 45 . 在处理缺失值时, 设定清洗规则时通常不包括 ()。
 (A)识别缺失值的模式 (B)选择填充缺失值的方法 (C)设定缺失值填充的阈值 (D)评估填充后数据的质量
- 46 . () 适用于评估数据清洗方案对业务决策的影响。
 (A)对比清洗前后数据的统计特征 (B)询问数据使用者的反馈 (C)单纯分析数据清洗的技术细节 (D)检查数据清洗的代码实现
- 47 . 在处理数据中的异常值时, 通常不会采取 ()。
 (A)删除异常值 (B)填充异常值为某个固定值 (C)使用统计方法(替换异常值 (D)忽略异常值, 不进行任何处理
- 48 . 在数据预处理中, 将特征缩放到[0, 1]区间的方法称为 ()。
 (A)标准化 (B)归一化 (C)离散化 (D)编码
- 49 . 有一组数据, 其最大值为 100, 最小值为 10, 要将其归一化到[0, 1]区间, 应使用的公式是 ()。
 (A) $(\frac{x - \text{min}(x)}{\text{max}(x) - \text{min}(x)})$ (B) $(\frac{x - \mu}{\sigma})$ (C) $(\log(x))$ (D) $(\frac{1}{x})$
- 50 . 将两个或多个数据集中的数据合并到一个数据集中的过程称为 ()。
 (A)数据聚合 (B)数据整合 (C)数据合并 (D)数据迁移
- 51 . 在合并两个数据集时, 需要特别注意的是 ()。
 (A)数据集的大小 (B)数据集的缺失值 (C)数据集的来源 (D)数据集之间的兼容性
- 52 . SQL 作为 BI 查询引擎, 其最显著的特点是 ()。
 (A)适用于非结构化数据 (B)提供丰富的数据操作能力 (C)仅支持内存计算 (D)专为图形界面设计
- 53 . MDX 主要用于 ()。

- (A) 文本挖掘 (B) 实时数据流处理 (C) 多维数据分析 (D) 图形数据分析
- 54 . 在使用 MDX 时, () 是理解和构建多维数据集的关键。
(A) 数据行 (B) 数据列 (C) 维度 (D) 数据表
- 55 . DAX 主要用在 () 平台中。
(A) SQL Server (B) Power BI (C) Hadoop (D) MongoDB
- 56 . DAX 提供了 () 的数据操作能力。
(A) 低级数据位操作 (B) 内存数据库管理 (C) 数据模型的查询、计算和定义 (D) 实时数据流处理
- 57 . Spark SQL 提供的主要优势是 () 。
(A) 实时数据流处理 (B) 基于内存的分布式数据处理 (C) 专为结构化数据设计 (D) 低延迟的图形数据处理
- 58 . Presto 特别适用于 () 。
(A) 批处理大数据作业 (B) 流式数据处理 (C) 低延迟的交互式分析查询 (D) 图形和图像数据处理
- 59 . FineReport 和 FineVis 相比, () 更侧重于数据可视化。
(A) FineReport (B) FineVis (C) 两者相同 (D) 两者都不侧重
- 60 . 在选择 BI 工具的查询引擎时, () 通常不需要首要考虑。
(A) 查询性能 (B) 支持的数据源类型 (C) 工具的易用性 (D) 工具的价格是否足够低
- 61 . () 最能体现查询引擎的扩展性。
(A) 并发用户数 (B) 支持的查询类型 (C) 分布式处理能力 (D) 图形界面的美观程度
- 62 . () 常用于查询引擎的性能优化。
(A) 加密技术 (B) 索引优化 (C) 实时数据压缩 (D) 数据清洗
- 63 . 在进行查询引擎优化时, () 不是必须的。
(A) 分析查询计划 (B) 增加更多的硬件资源 (C) 调整查询逻辑以减少数据扫描量 (D) 监控并优化数据库连接
- 64 . () 直接反映了查询操作执行时间的长短。
(A) CPU 使用率 (B) 响应时间 (C) 磁盘 I/O 次数 (D) 网络带宽
- 65 . 在数据库查询性能评估中, QPS (Queries Per Second) 指的是 () 。
(A) 每秒完成的查询数量 (B) 每秒插入的记录数 (C) 每秒更新的行数 (D) 每秒删除的记录数
- 66 . 使用分词库时, 常见的性能瓶颈是 () 。
(A) 词汇量不足 (B) 单词识别错误率高 (C) 处理大数据量时的内存消耗 (D) 不支持多语言分词
- 67 . 在 () 数据库时, 需要选择合适的字符集和排序规则, 以便支持不同语言和文本的存储和检索。
(A) 创建 (B) 检索 (C) 删除 (D) 查询
- 68 . 在查看数据库时, 可以通过 () 或客户端工具来查看数据库的结构和内容。
(A) 输入框 (B) 视图窗 (C) 命令行 (D) 数据列
- 69 . group by 子句进行分组以后, 需要对分组结果再进行条件过滤时, 不能使用 where 语句, 而需要用 () 。
(A) count (B) having (C) insert (D) select
- 70 . 交叉连接是不带 () 的连接, 即两个表的笛卡尔积是量表中元组的交叉乘积。

- (A) 逻辑运算符 (B) 谓词 (C) 连接词 (D) 名词
- 71 . 自连接就是一个 () , 同本身进行连接。
(A) 表 (B) 数据 (C) 数据集 (D) 数据列
- 72 . 结构化数据是指表现为 () 的数据。
(A) 一维形式 (B) 二维形式 (C) 三维形式 (D) 平面形式
- 73 . 非结构化数据可以从论坛、微博、微信等渠道获得属于其 () 的特征。
(A) 格式多样 (B) 容量大 (C) 来源多样 (D) 数据类型多样
- 74 . 结构化数据是指那些以固定格式或模式存储的信息, 如 () 。
(A) 文本文件 (B) 客户信息 (C) 图片 (D) 社交媒体帖子
- 75 . 不属于结构化数据的应用场景的是 () 。
(A) 微博评论 (B) 企业 ERP (C) 财务系统 (D) 教育一卡通
- 76 . 非结构化数据 () 通过集成 OCR 等第三方生态应用, 实现内容赋能业务运营, 包括智能识别提取业务系统内容, 实现业务流程的自动化, 提升数字化服务能力和用户体验等。
(A) 前台 (B) 中台 (C) 后台 (D) 终端
- 77 . 层次建模的要素中, () 用于唯一标识一个实体的属性集合, 确保数据的一致性和完整性。
(A) 实体 (B) 键 (C) 属性 (D) 关系
- 78 . 为了提高查询性能, 维度建模允许增加数据 () , 采用反规范化的设计技术。
(A) 冗余 (B) 集 (C) 表 (D) 列
- 79 . 维度建模的缺点包括 () 。
(A) 模型简单易理解 (B) 可扩展性好 (C) 支持高效的数据分析 (D) 数据冗余
- 80 . 事实表中包含了业务过程的 () , 可以通过与维度表关联来进行分析和查询。
(A) 事实 (B) 维度 (C) 度量 (D) 量表
- 81 . 每个维度表都包含单一的 () , 维度表的主键可以作为与之关联的任何事实表的外键, 维度表行的描述环境应与事实表行完全对应。
(A) 主键列 (B) 主键行 (C) 主键 (D) 数据列
- 82 . 由于链表在 () , 它适用于需要频繁进行这些操作的应用场景。
(A) 插入和删除操作上的高效性 (B) 长度上的动态性 (C) 内存上的利用率高 (D) 访问上的元素效率低
- 83 . 栈的操作基于 () 的原则, 入栈和出栈操作都只涉及栈顶元素, 因此操作的时间复杂度为 $O(1)$, 执行速度较快。
(A) 只进不出 (B) 同时进出 (C) 先进后出 (D) 后进先出
- 84 . 根据选择的特征评估标准, 从上至下递归地生成子节点, 直到数据集不可分则停止生长属于决策树算法构造过程的 () 。
(A) 特征选择 (B) 决策树生成 (C) 剪枝 (D) 数据分析
- 85 . 在决策树中, 如果一个节点的所有样本都属于同一个类, 那么该节点被称为 () 。
(A) 根节点 (B) 内部节点 (C) 叶节点 (D) 分支节点
- 86 . 支持向量机 (SVM) 是一种常用的机器学习算法, 主要用于线性和非线性分类问题, 回归以及 () 检测。
(A) 回归值 (B) 常量值 (C) 向量值 (D) 异常值
- 87 . SVM 中的“支持向量”是 () 。
(A) 离分类边界最近的训练样本 (B) 所有的训练样本 (C) 离分类边界最远的训练样本 (D) 错误分类的训练样本

- 88 . 在创建数据画像时，最常用的数据分析方法是（ ）。
- (A) 描述性统计分析 (B) 回归分析 (C) 聚类分析 (D) 假设检验
- 89 . 在数据画像构建流程中，（ ）涉及到选择相关的特征。
- (A) 数据收集 (B) 数据清洗 (C) 特征工程 (D) 模型评估
- 90 . 婚姻、有无女孩、有无男孩、家里是否有老人、性取向等属于用户标签的（ ）。
- (A) 基础属性 (B) 社会关系 (C) 行为特征 (D) 业务相关
- 91 . 在标签体系中，（ ）可以确保标签的有效性和准确性。
- (A) 忽略用户反馈 (B) 避免使用动态标签 (C) 定期审查和更新标签 (D) 只使用一种标签类型
- 92 . （ ）通过自动训练出区分图像的特征分类器来提取特征。
- (A) 聚类方法 (B) 线或边缘提取方法 (C) 深度学习方法 (D) 映射方法
- 93 . 在处理缺失值时，如果数据是正态分布的，常用的填充方法是（ ）。
- (A) 众数填充 (B) 中位数填充 (C) 均值填充 (D) 随机填充
- 94 . 当数据集中的缺失值比例很高时，最不推荐的方法是（ ）。
- (A) 删除记录 (B) 使用算法预测填充 (C) 多重插补 (D) 增加缺失值
- 95 . 聚合特征构造通常基于 ID 值或（ ）的某个类别计算数值特征的一些统计量。
- (A) 存储特征 (B) 文本特征 (C) 类别特征 (D) 数值特征
- 96 . 在特征变换中，标准化的目的是（ ）。
- (A) 使所有特征具有相同的单位 (B) 将特征转换为分类变量 (C) 减少特征的数量 (D) 增加数据的复杂性
- 97 . （ ）是 Python 数值计算的基础包，提供了高效的数据结构和计算方法。
- (A) NumPy (B) Pandas (C) scikit-learn (D) Seaborn
- 98 . （ ）是一个开源的深度学习框架，也可以用于构建神经网络模型。
- (A) NumPy (B) Pandas (C) scikit-learn (D) TensorFlow
- 99 . 在 scikit-learn 库中，K-Means 算法属于（ ）。
- (A) 分类算法 (B) 回归算法 (C) 聚类算法 (D) 降维算法
- 100 . TensorFlow 中的计算图的主要作用是（ ）。
- (A) 可视化数据 (B) 优化模型性能 (C) 存储训练数据 (D) 定义和运行计算模型
- 101 . 在进行数据分析之前，要明确研究问题和研究（ ），选择适当的统计方法和分析技术，避免分析过程中的迷茫和盲目。
- (A) 数据 (B) 目的 (C) 方法 (D) 结果
- 102 . 在 Power BI 中，每次可视化都要求用户在将可视化拖动到画布后，从“字段”窗格中选择一个或多个字段，用户必须选中（ ），才能在“字段”窗格中包含该字段以进行可视化。
- (A) 复选框 (B) 文本框 (C) 无线框 (D) 输入框
- 103 . 在 Stata 中，用于生成频数分布表的命令是（ ）。
- (A) tabulate (B) summarize (C) regress (D) list
- 104 . 数据可视化的主要目的是（ ）。
- (A) 隐藏数据中的信息 (B) 增加数据处理的复杂性 (C) 帮助人们更快地理解和分析数据 (D) 仅仅为了美观
- 105 . 在视觉编码中，位置 (Position) 通常用于表示（ ）。
- (A) 数量大小 (B) 类别 (C) 时间顺序 (D) 层级关系
- 106 . （ ）有助于提高图表的可读性。

- (A)相近色调的对比 (B)高饱和度的颜色对比 (C)互补色的对比 (D)灰色调对比
- 107 . ()强调在数据可视化中应突出显示最重要的数据点。
(A)简洁性 (B)层次性 (C)焦点与上下文 (D)一致性
- 108 . 在数据可视化设计中, 确保准确性的首要步骤是()。
(A)选择合适的颜色方案(B)验证数据源的可靠性 (C)设计吸引人的布局(D)应用交互功能
- 109 . 当在数据可视化中展示百分比数据时, () 最符合准确性原则。
(A)使用整数表示百分比(B)使用小数表示百分比, 并保留到小数点后两位(C)使用百分号符号(%)明确标识(D)将百分比转换为分数形式表示
- 110 . 为了优化组件的性能, ()是有效的。
(A)增加组件的复杂度, 以提供更多功能 (B)减少不必要的 DOM 操作, 优化渲染过程 (C)忽略组件的懒加载和代码分割 (D)使用大量的动画和过渡效果, 以增强用户体验
- 111 . 在 Web 应用中, ()可以优化长列表组件的性能。
(A)一次性加载和渲染所有列表项 (B)使用虚拟滚动技术, 只渲染可视区域内的列表项 (C)增加列表项的复杂度, 以提高用户满意度 (D)禁用浏览器的滚动平滑效果
- 112 . 在进行销售趋势预测时, 最常用的时间序列分析方法是()。
(A)回归分析 (B)聚类分析 (C)移动平均法 (D)决策树分析
- 113 . ()不是影响趋势预测准确性的关键因素。
(A)数据的质量与完整性 (B)预测模型的复杂度 (C)预测时间范围的长短 (D)预测者的主观判断
- 114 . 在业务决策中使用数据可视化结果时, 最重要的是()。
(A)确保数据的准确性 (B)选择最美观的可视化方式 (C)快速完成可视化过程 (D)依赖单一的可视化结果
- 115 . 市场行情分析的常用指标中, 主要用() 衡量市场的超买超卖状态。
(A) 移动平均线 (B) 相对强弱指数 (RSI) (C) 成交量比例 (D) 中间意愿指标 (CR)
- 116 . 市场行情分析中, () 指标是通过计算一定时期内的平均价格来平滑价格数据的。
(A) 移动平均线 (B) 相对强弱指数 (C) 成交量比例 (D) 中间意愿指标
- 117 . () 不是市场行情数据的主要来源。
(A) 财经新闻网站 (B) 社交媒体上的谣言 (C) 金融数据服务提供商 (D) 政府机构发布的报告
- 118 . 实时市场行情数据通常通过() 从交易所获取。
(A) 交易所 API (B) 移动平均线 (C) 中间意愿指标 (D) 成交量比例
- 119 . 在市场供给与需求分析中, () 是描述生产者在不同价格水平下愿意提供的商品数量的曲线。
(A) 需求曲线 (B) 供给曲线 (C) 成本曲线 (D) 收益曲线
- 120 . 在市场供给与需求分析方法中, 当供给量小于需求量时, 市场价格通常会() 上升。
(A) 上升 (B) 下降 (C) 保持不变 (D) 无法确定
- 121 . 关于市场容量与趋势分析的说法中, () 是不正确的。
(A) 市场容量是指市场的最大承载能力 (B) 市场容量分析有助于了解市场的规模和潜力 (C) 市场趋势分析只能基于历史数据 (D) 市场趋势分析包括识别市场

的周期性行为

- 122 . 市场容量分析主要关注（ ）。
- (A) 市场的短期波动 (B) 市场的长期发展趋势 (C) 单个投资者的行为 (D) 产品的生产成本
- 123 . 在分析新兴经济体的市场潜力时，企业最应关注的是（ ）。
- (A) 该经济体的政治制度稳定性 (B) 该经济体的 GDP 增长率 (C) 该经济体的自然资源储备 (D) 该经济体的文化习俗差异
- 124 . 市场潜力分析中，用于衡量产品或服务在目标市场中的潜在需求量的方法是（ ）。
- (A) 移动平均线 (B) 市场调研 (C) 成交量比例 (D) 中间意愿指标
- 125 . 市场分类从区分目的来说有（ ）这两大种类。
- (A) 地理型和投资型 (B) 管理型和地理型 (C) 管理型和投资型 (D) 管理型和消费者导向型
- 126 . （ ）不属于市场分析法。
- (A) 波特五力模型 (B) SWOT 分析法 (C) 角色扮演法 (D) PEST 分析法
- 127 . 用于评估行业内部企业间竞争态势的分析方法是（ ）。
- (A) 成交量比例分析 (B) 波特五力模型 (C) 移动平均线预测 (D) 相对强弱指数对比
- 128 . 当行业进入成熟期时，（ ）的特征最为显著。
- (A) 市场份额高度集中，少数几家企业主导市场 (B) 技术创新频繁，产品更新换代快 (C) 市场需求增长放缓，竞争趋于稳定 (D) 企业数量急剧减少，市场逐渐萎缩
- 129 . 在行业的一般特征分析中，用于反映行业技术水平和创新能力的是（ ）。
- (A) 市场份额 (B) 行业利润率 (C) 行业集中度 (D) 技术发展水平
- 130 . 企业把提供同种或同类产品，但规格、型号、款式不同的企业称为（ ）竞争者。
- (A) 需要 (B) 行业 (C) 品牌 (D) 消费
- 131 . （ ）属于分析竞争对手商品信息的角度。
- (A) 客单价 (B) 商品细节 (C) 店铺收藏 (D) 网店地址
- 132 . 从商品价格、物流方式、商品细节等方面展示商品品质属于分析竞争对手的（ ）角度。
- (A) 详情页 (B) 推广活动 (C) 商品评论 (D) 商品信息
- 133 . 竞争对手 SWOT 分析中的“S”代表（ ）。
- (A) 优势 (B) 劣势 (C) 机会 (D) 威胁
- 134 . 在应用波特五力模型之前，首先需要明确的是（ ）。
- (A) 竞争对手的财务状况 (B) 行业的定义和范围 (C) 消费者的个人喜好 (D) 政府的最新政策
- 135 . 买方议价能力的高低主要取决于（ ）。
- (A) 产品的生产成本 (B) 产品的销售渠道 (C) 买方的购买实力与谈判技巧 (D) 产品的品牌形象
- 136 . 在跨境电子商务数据申报中，（ ）是必须提供的。
- (A) 商品的尺寸和颜色 (B) 交易双方的银行卡账户信息 (C) 商品的详细描述和 HS 编码 (D) 物流公司的内部编号
- 137 . “免、抵、退”税办法中的“免”指的是（ ）。
- (A) 免征进口环节税 (B) 免征出口环节税 (C) 免征企业所得税 (D) 免征增值税和消费税

- 138 商务数据分析报告撰写时，首要考虑的是报告的（ ）以确保读者能够明确理解分析目的。
(A) 清晰性 (B) 详细性 (C) 创新性 (D) 艺术性
- 139 在缮制商务数据分析报告时，未对数据进行充分清洗和校验，可能导致报告中的（ ）问题。
(A) 数据准确性不足 (B) 报告格式不美观 (C) 结论过于乐观 (D) 分析方法不先进
- 140 商务数据分析报告框架设计时，确保报告内容（ ）是提高报告可读性的关键。
(A) 详细冗长 (B) 条理清晰 (C) 视觉华丽 (D) 语言深奥

三、多项选择题（第 1 题~第 10 题。选择一个正确的答案，将相应的字母填入题内的括号中。每题 1 分，满分 10 分）

- 1 数据画像可以帮助企业实现（ ）目标。
(A) 个性化营销 (B) 改善客户服务 (C) 开发新产品 (D) 优化运营流程 (E) 降低生产成本
- 2 用户数据画像的构建流程主要包括数据收集、数据清洗、数据分析和数据可视化（ ）。
(A) 用户数据收集 (B) 用户数据清洗 (C) 用户数据分析 (D) 用户数据可视化 (E) 用户数据上传
- 3 在构建标签体系时，需要考虑的因素包括（ ）。
(A) 业务目标 (B) 数据可用性 (C) 技术限制 (D) 用户隐私 (E) 数据价格
- 4 特征处理作为特征工程的核心组成部分，包含了（ ）等多个环节。
(A) 数据预处理 (B) 特征提取 (C) 特征构造 (D) 特征选择 (E) 降维
- 5 特征工程的作用包括（ ）。
(A) 提高模型性能 (B) 减少模型训练时间 (C) 增加模型可解释性 (D) 防止过拟合 (E) 提高模型的泛化能力
- 6 具体特征选择方法包括（ ）。
(A) 相关性系数法 (B) 信息增益法 (C) 互信息法 (D) 特征向量法 (E) 信息相切法
- 7 特征提取在多个领域有广泛应用，包括（ ）。
(A) 数字信号处理 (B) 信息成本降低 (C) 信息来源多样化 (D) 智能风控 (E) 机器学习
- 8 在处理缺失值之前，应该考虑（ ）等因素。
(A) 缺失值的比例 (B) 缺失值的类型 (C) 缺失值对分析结果的影响 (D) 缺失值的来源 (E) 缺失值的位置
- 9 在特征构造过程中，（ ）有助于提高模型的性能。
(A) 特征选择 (B) 特征缩放 (C) 特征组合 (D) 特征转换 (E) 特征放大
- 10 Pandas 库的主要功能包括（ ）。
(A) 数据清洗和准备 (B) 数据分析和探索 (C) 机器学习模型训练 (D) 数据可视化 (E) 数据生成和处理

商务数据分析师（三级）理论知识试卷答案

一、判断题(第1题~第40题。将判断结果填入括号中。正确的填“√”，错误的填“×”。每题0.5分，满分20分)

- | | | | | | |
|-------|-------|-------|-------|-------|-------|
| 1. √ | 2. × | 3. × | 4. √ | 5. × | 6. × |
| 7. × | 8. √ | 9. √ | 10. × | 11. × | 12. × |
| 13. × | 14. √ | 15. × | 16. √ | 17. × | 18. × |
| 19. √ | 20. × | 21. × | 22. √ | 23. √ | 24. √ |
| 25. × | 26. √ | 27. × | 28. √ | 29. × | 30. × |
| 31. √ | 32. × | 33. √ | 34. × | 35. √ | 36. × |
| 37. √ | 38. × | 39. × | 40. × | | |

二、单项选择题(第1题~第140题。选择一个正确的答案，将相应的字母填入题内的括号中。每题0.5分，满分70分)

- | | | | | | |
|--------|--------|--------|--------|--------|--------|
| 1. C | 2. A | 3. B | 4. A | 5. C | 6. C |
| 7. B | 8. B | 9. B | 10. B | 11. C | 12. D |
| 13. B | 14. B | 15. B | 16. C | 17. A | 18. B |
| 19. D | 20. C | 21. B | 22. C | 23. B | 24. A |
| 25. B | 26. C | 27. C | 28. A | 29. C | 30. B |
| 31. A | 32. C | 33. B | 34. B | 35. B | 36. A |
| 37. D | 38. B | 39. D | 40. B | 41. C | 42. C |
| 43. B | 44. B | 45. C | 46. B | 47. D | 48. B |
| 49. A | 50. C | 51. D | 52. B | 53. C | 54. C |
| 55. B | 56. C | 57. B | 58. C | 59. B | 60. D |
| 61. C | 62. B | 63. B | 64. B | 65. A | 66. C |
| 67. A | 68. C | 69. B | 70. B | 71. A | 72. B |
| 73. C | 74. B | 75. A | 76. B | 77. B | 78. A |
| 79. D | 80. C | 81. A | 82. A | 83. D | 84. B |
| 85. C | 86. D | 87. A | 88. A | 89. C | 90. B |
| 91. C | 92. C | 93. C | 94. D | 95. C | 96. A |
| 97. A | 98. D | 99. C | 100. D | 101. B | 102. A |
| 103. A | 104. C | 105. B | 106. C | 107. C | 108. B |
| 109. C | 110. B | 111. B | 112. C | 113. D | 114. A |
| 115. B | 116. A | 117. B | 118. A | 119. B | 120. A |
| 121. C | 122. B | 123. B | 124. B | 125. C | 126. C |
| 127. B | 128. C | 129. D | 130. C | 131. B | 132. D |
| 133. A | 134. B | 135. C | 136. C | 137. B | 138. A |
| 139. A | 140. B | | | | |

三、多项选择题(第1题~第10题。选择一个正确的答案，将相应的字母填入题内的括号中。每题1分，满分10分)

- | | | | | | |
|---------|---------|---------|----------|----------|--------|
| 1. ABCD | 2. ABCD | 3. ABCD | 4. ABCDE | 5. ABCDE | 6. ABC |
| 7. ADE | 8. ABC | 9. ABCD | 10. ABD | | |

第6部分

操作技能考核模拟试卷

注 意 事 项

1. 考生根据操作技能考核通知单中所列的试题做好考核准备；
2. 请考生仔细阅读试题单中具体考核内容和要求，并按要求完成操作或进行笔答或口答，若有笔答请考生在答题卷上完成。
3. 操作技能考核时要遵守考场纪律，服从考场管理人员指挥，以保证考核安全顺利进行。

注：操作技能考核试题评分表及答案是考评员对考生考核过程及考核结果的评分记录表，也是评分依据。

商务数据分析师（三级）操作技能考核通知单

姓名：

准考证号：

考核日期：

试题 1

试题代码：1.1.1

试题名称：电商平台用户行为数据处理方案设计

考核时间：20 min

配分：20 分

试题 2

试题代码：2.1.1

试题名称：执行费用数据标准化处理与建模分析

考核时间：15 min

配分：10 分

试题 3

试题代码：2.2.1

试题名称：电信客户流失预测考核

考核时间：15min

配分：10 分

试题 4

试题代码：3.1.1

试题名称：设计商铺数据可视化方案

考核时间：15 min

配分：10 分

试题 5

试题代码：3.2.1

试题名称：开展家电营销数据可视化结果分析并实现客户留存

考核时间：10 min

配分：10 分

试题 6

试题代码：4.1.1

试题名称：挖掘精密制造数据并研究竞争态势，缮制商务数据分析报告

考核时间：25 min

配分：20 分

试题 7

试题代码：4.2.1

试题名称：跨境电商贸易数据申报与波特五力分析

考核时间：20 min

配分：20 分

商务数据分析师（三级）操作技能考核 试题单

试题代码：1.1.1

试题名称：电商平台用户行为数据处理方案设计

考核时间：20 min

1.场地设备要求

- (1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- (2) 系统中已安装 BI 类商务数据分析软件、压缩软件、Python 3.8+（安装 pandas、mysql-connector-python 库）、MySQL 8.0+及办公软件；
- (3) 商务数据分析师考试平台。

2.工作任务

- (1) 根据业务需求，设计在线（JSON）和离线（CSV）用户行为数据采集方案，存入 MySQL，确保实时性和完整性；
- (2) 根据业务需求，制定数据迁移方案，将用户行为数据从源 MySQL 迁移至目标 MySQL（模拟云端），保证一致性和效率；
- (3) 根据业务特性，优化用户行为数据存储结构，添加索引，提升查询性能；
- (4) 根据数据质量要求，制定清洗策略处理重复、缺失、格式问题，并评估清洗后数据质量（准确率和填充率）。

3.技能要求

- (1) 根据业务需求进行在线、离线数据采集；
- (2) 根据业务需求参与制定数据迁移方案；
- (3) 根据业务特性制订数据存储策略；
- (4) 根据数据质量要求制定数据清洗策略及评估方案。

4.质量指标

- (1) 确保数据采集覆盖率高，实时数据快速入库，离线数据无丢失；
- (2) 保证迁移后数据一致，无记录丢失，迁移过程高效；
- (3) 优化存储结构，支持快速查询并降低存储成本；
- (4) 清洗数据确保行为类型准确，缺失值有效填充。

背景材料

某电商平台致力于优化用户行为数据管理，以提升个性化推荐系统的精准性和营销策略的有效性。业务场景涵盖用户在平台上的在线浏览、点击、购买行为，以及离线生成的订单和库存数据。数据来源多样，包括网站日志（JSON 格式，记录实时操作）、移动端 API（实时传输交互数据）、第三方支付平台（CSV 格式，记录交易信息）以及本地 MySQL 数据库（存储历史订单和库存）。这些数据存在诸多问题，如重复记录（因多设备登录导致）、缺失值（行为类型未正确记录）、格式不一致（时间戳格式多样）等，严重影响数据分析的准确性。为实现数据驱动决策，需设计一套高效的数据处理方案，涵盖以下核心任务：首先，根据业务需求进行在线和离线数据采集，确保实时性和完整性，满足推荐系统对低延迟的要求；其次，制定数据迁移方案，将历史数据从本地 MySQL 数据库迁移至云端数据仓库（模拟为目标 MySQL 表），保证数据一致性和迁移效率；再次，根据业务特性制定数据存储策略，通过表结构优化和索引设计，提升查询性能并降低存储成本；最后，制定数据清洗策略及评估方案，处理重复、缺失、格式问题，提升数据准确性和一致性。目标是构建一个高效、可靠的数据处理流程，支持实时推荐、高并发查询和动态营销策略，同时确保数据安全和系统扩展性。

操作要求：

1. 编写代码导入必要的 Python 库。使用 `try...except ImportError` 结构对 `pandas` 和 `mysql.connector` 的导入进行异常处理，以验证库的可用性。
2. 实现数据库连接，使用 `mysql.connector.connect(db_config)` 的字典解包方式。连接成功后，执行 SQL 查询以获取 `user_behavior` 表的前 5 条记录，验证数据预览功能。
3. 使用 `pandas` 从 JSON 文件读取在线数据，在验证其记录数、字段数及数据完整性后，使用 `cursor.executemany()` 的批量插入方法将其存入数据库。
4. 使用 `pandas` 从 CSV 文件读取离线数据，在验证记录数和字段数满足要求后，使用 `cursor.executemany()` 的批量插入方法将其追加到数据库的相应表中。
5. 编写数据迁移逻辑，从源表查询数据后，在 Python 中使用 `executemany()` 批量插入到目标表。同时，使用参数化查询 (`VALUES (%s, ...)`) 将迁移日志写入 `migration_log` 表。
6. 为指定表创建基于 `user_id` 和 `action_type` 的复合索引。代码中必须包含检查索引是否已存在的逻辑，以优化查询性能并确保脚本可重复执行。
7. 使用 `pandas` 的向量化操作完成数据清洗：使用 `df.drop_duplicates()` 去重，使用 `pd.to_datetime()` 格式化时间戳，并使用 `df.fillna()` 填充 `action_type` 的缺失值，

确保数据一致性和完整性。

8. 使用 `isin().mean()` 的向量化方法高效计算 `action_type` 字段的有效值比例，以评估数据质量并返回准确率指标。

商务数据分析师（三级）操作技能考核 试题评分表

试题代码：1.1.1

试题名称：电商平台用户行为数据处理方案设计

考核时间：20 min

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M1 - 库验证成功”；无 ModuleNotFoundError。	结果正确，且使用了推荐的 try...except ImportError 结构进行验证，得 2 分。 结果正确，但仅使用了 import 语句，未进行异常处理，得 1 分。 结果错误，得 0 分。		
M2	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M2 - 数据库连接成功”“M2 - SQL 查询执行成功”“M2 - 数据预览（表格格式）”“M2 - 返回记录数: [1-5]”。	结果正确，且使用了推荐的 mysql.connector.connect(**db_config) 解包字典方式连接，得 2 分。 结果正确，但手动逐一列出了连接参数，得 1 分。 结果错误，得 0 分。		
M3	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M3 - 数据验证通过”“M3 - 记录数: [≥1000]”“M3 - 字段数: 8”“M3 - 数据验证状态: success”；ecommerce.sql 中 user_behavior 表记录数 ≥ 1,000，字段数=8，无空值。	结果正确，且使用了推荐的 pd.read_json() 和 cursor.executemany() 批量插入方法，得 2 分。 结果正确，但使用了任何形式的逐条循环插入（如 iterrows, apply 等），得 1 分。 结果错误，得 0 分。		
M4	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M4 - 数据验证通过”“M4 - 记录数: [≥10000]”“M4 - 字段数: 8”“M4 - 数据验证状态: success”；ecommerce.sql 中 user_behavior 表记录数 ≥ 10,000，字段数	结果正确，且使用了推荐的 pd.read_csv() 和 cursor.executemany() 批量插入方法，得 2 分。 结果正确，但使用了任何形式的逐条循环插入（如 iterrows, apply 等），得 1 分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
		=8。	结果错误，得 0 分。		
M5	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M5 - 数据迁移成功”“M5 - 迁移记录数: [≈11000]”“M5 - 迁移状态: success”；ecommerce.sql 中 user_behavior_target 表记录数 ≈ 11,000。	结果正确，且使用了推荐的 Python SELECT 后 executemany() 批量 INSERT 的方法，得 2 分。 结果正确，但使用了 INSERT INTO ... SELECT FROM ... SQL 语句或逐条迁移，得 1 分。 结果错误，得 0 分。		
M6	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M6 - 迁移日志记录成功”“M6 - 日志状态: success”；ecommerce.sql 中 migration_log 表记录存在。	结果正确，且使用了推荐的参数化查询 (VALUES (%s, ...)) 以防止 SQL 注入，得 2 分。 结果正确，但使用了 f-string 或字符串拼接来构造 SQL，得 1 分。 结果错误，得 0 分。		
M7	2	日志显示“M7 - 索引创建成功”或“M7 - 索引 idx_user_action 已存在”；ecommerce.sql 中 user_behavior_target 表存在索引 idx_user_action。	结果正确，且代码中包含了推荐的检查索引是否已存在的逻辑，得 2 分。 结果正确，但仅执行 CREATE INDEX，未处理索引已存在的情况，得 1 分。 结果错误，得 0 分。		
M8	2	日志显示“M8 - 数据去重成功”“M8 - 移除重复记录: [≈200] 条”“M8 - 去重状态: success”；user_behavior_cleaned.csv 记录数 ≈ 9,800，字段数=8，无乱码。	结果正确，且使用了推荐的 df.drop_duplicates() 向量化方法，得 2 分。 结果正确，但使用了循环等其他逻辑上可行但低效的方法，得 1 分。 结果错误，得 0 分。		
M9	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M9 - 数据清洗成功”“M9 - 填充 action_type 缺失值: [≈ 200] 个”“M9 - 格式化 timestamp 成功，填充空时间戳: [0-50] 个”“M9 - 数据保存成功”“M9 - 保存状态: success”；user_behavior_cleaned.csv 记录数 ≈ 9,800，字段数=8，无缺失值，timestamp 格式为 YYYY-MM-DD HH:MM:SS，无乱码。	结果正确，且使用了推荐的 df.fillna() 和 pd.to_datetime() 等向量化操作，得 2 分。 结果正确，但使用了 .apply() 配合自定义函数等低效方法，得 1 分。 结果错误，得 0 分。		
M10	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示	结果正确，且使用了推荐的 isin().mean() 向量化方法，得 2 分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
		<p>“M10 - 数据质量评估成功”</p> <p>“M10 - 行为类型准确率: [≥ 0.99]”</p> <p>“M10 - 评估状态: success”</p> <p>user_behavior_cleaned.csv 的 action_type 均为合法值(浏览数据、点击数据、购买数据、数据)。</p>	<p>结果正确, 但使用了循环计数的方法, 得 1 分。</p> <p>结果错误, 得 0 分。</p>		
合计配分	20	合计得分			

商务数据分析师（三级）操作技能考核 试题单

试题代码：2.1.1

试题名称：执行费用数据标准化处理与建模分析

考核时间：15 min

1.场地设备要求

- (1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- (2) 系统中已安装数据分析软件、压缩软件及办公软件；
- (3) 数据库及商务数据分析师考试平台。

2.工作任务

- (1) 根据企业多源异构数据，完成多源数据的整合工作；
- (2) 根据某公司数据，完成金额数据的规范化格式处理任务；
- (3) 根据业务要求，完成预算费用数据的规约处理任务；
- (4) 根据维度建模理论，完成数据表关系的分析与建立；
- (5) 根据业务需求，完成查询重构及模型的优化工作。

3.技能要求

- (1) 根据相关数据表特点，进行多源数据的整合；
- (2) 根据业务要求，对金额数据格式进行精准转换；
- (3) 根据数据归一性原则，开展预算费用数据的规约；
- (4) 参照数据建模标准规范，设计星型分析模型；
- (5) 根据业务需求，对数据进行重构并优化数据模型。

4.质量指标

- (1) 准确识别多源数据表的关键字段，正确完成数据匹配与整合，确保信息完整无误；
- (2) 正确执行金额数据格式转换，精确实现小数点后三位标准化处理；
- (3) 准确筛选关键数据特征，正确完成数据规约处理；
- (4) 正确分析数据表间逻辑关系，准确建立符合规范的星型模型；
- (5) 准确理解业务需求，正确优化查询逻辑，精确提升模型性能。

背景材料

商飞集团作为综合性企业，在日常运营中产生多源异构数据，包括 CSV 和 Excel 等格式的费用明细记录和成本中心数据等。其中，成本中心数据表详细记录了各分店、城市及对应成本中心的关联信息；费用科目体系采用分级编码管理，涵盖包括咖啡豆、牛奶等原材料在内的材料费、设备维护费、市场营销费、物业费及薪酬福利等类别。基于北京、上海和杭州三地近两年的季度数据，通过数据预处理和建模分析技术，对各区域分店的预算执行情况进行多维度评估，重点分析材料成本波动、费用管控效率等关键指标，为集团精细化管理和资源优化配置提供数据支持。

操作要求：

1. 利用 SQL、MySQL、ETL、BI 等数据分析工具，实现“费用数据”多级科目分类属性与以及“费用明细”中全部数据的精准关联匹配，完成多源异构数据的深度融合与统一管理；

2. 将费用明细表中的各项费用支出统一转换为人民币计量单位，且确保数据精度达到千分位，为财务数据分析提供可靠的数据基础；

3. 在预算表中通过 Min-Max 归一化算法，构建“总费用_归一化”衍生指标，对预算数据实施特征缩放，以量化费用波动的相对幅度；

4. 运用维度建模理论，针对“费用数据”、“费用明细”以及“成本中心数据”这三个数据源中的所有数据表，构建星型模型数据架构；

5. 实施基于业务语义的水平分片策略，按区域属性对预算表进行物理分区，生成北京/上海专属子集，并通过优化模型重构维度关联逻辑。（请提交操作完成后的数据文件至考试平台。）

商务数据分析师（三级）操作技能考核 试题评分表

试题代码：2.1.1

试题名称：执行费用数据标准化处理与建模分析

考核时间：15 min

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	2	正确对多源异构数据进行横向整合	正确加载并整合=2分； 仅正确加载=1分；正 确加载任意1个数据 源或仅正确进行整合 分析=0.5分；错误=0 分。		
M2	2	正确转换数据格式	正确转换2处=2分； 正确转换1处=1分； 错误=0分。		
M3	2	正确对数据进行 Min-Max 归一化处理	是=2分；否=0分。		
M4	2	正确建立关系模型	正确建立3处关系=2 分；建立2处关系=1 分；建立1处或错误=0 分。		
M5	2	正确处理表并完善表间关系，以优化模型	正确编辑表并完善表 间关系=2分；仅正确 编辑表=1分；错误=0 分。		
合计配分	10	合计得分			

商务数据分析师（三级）操作技能考核 试题单

试题代码：2.2.1

试题名称：电信客户流失预测考核

考核时间：15 min

1.场地设备要求

- (1) 装有 Windows 操作系统的计算机（Windows 7 及以上），CPU 4 核，内存 8GB，硬盘 50GB 可用空间；
- (2) 系统中已安装 BI 类商务数据分析软件、压缩软件、Python 3.8+（安装 pandas、numpy、scikit-learn、matplotlib、seaborn 库）、MySQL 8.0+及办公软件；
- (3) 商务数据分析师考试平台。

2.工作任务

- (1) 计算月均账单金额和高账单低通话标志；
- (2) 处理年龄缺失值，标准化通话分钟数和账单金额；
- (3) 训练逻辑回归模型，计算准确率，生成通话分钟数分布图和混淆矩阵；
- (4) 生成 HTML 报告，包含特征计算结果、分布图、模型评估结果、混淆矩阵和业务洞察。

3.技能要求

- (1) 根据业务需求编写批量数据计算作业并优化作业参数；
- (2) 根据业务规则设计相应数据指标计算算法；
- (3) 评估挖掘需求并使用工具对数据进行特征工程处理；
- (4) 调用常规模型进行模型训练；
- (5) 基于评价指标进行模型测试。

4.质量指标

- (1) 高效准确提取月均账单金额和高账单低通话标志；
- (2) 完整规范处理年龄缺失值并标准化特征；
- (3) 逻辑回归模型精准分类客户流失风险；
- (4) HTML 报告清晰呈现特征计算结果、模型性能和图表。

背景材料

电信行业客户流失是企业面临的重要挑战，流失客户会导致收入下降和市场份额减少。通过分析客户行为数据（如账单金额、通话分钟数等），企业可以识别高风险客户并采取挽留措施。本任务基于 10,000 条电信客户数据，包含客户 ID（`customer_id`）、账单金额（`bill_amount`，年总金额，元）、通话分钟数（`call_minutes`，年总分钟数）、年龄（`age`，部分缺失）、是否流失（`churn`，0=未流失，1=流失）等字段。目标是构建预测模型，识别可能流失的客户，并生成分析报告。

操作要求：

1. 在 `compute_features` 函数中，计算月均账单金额，确保 月均账单金额 列被正确创建，其值为 账单金额 列除以 12。
2. 在 `compute_features` 函数中，使用向量化操作计算高账单低通话标志，确保 高账单低通话 列被正确创建，其逻辑为当 账单金额 大于均值且 通话分钟数 小于均值时为 1，否则为 0。
3. 在 `feature_engineering` 函数中，填补 年龄 列的缺失值，确保该列最终无任何缺失数据，并优先使用中位数或均值进行填补。
4. 在 `feature_engineering` 函数中，对 通话分钟数 和 账单金额 两列进行标准化处理，确保处理后的新列均值约等于 0 且标准差约等于 1。
5. 在 `train_and_evaluate` 函数中，计算并返回逻辑回归模型在测试集上的准确率。
6. 在 `train_and_evaluate` 函数中，生成并保存混淆矩阵热图至 `confusion_matrix.png`，该图必须为蓝色调（`cmap='Blues'`），在每个单元格中标注具体数值，并确保中文标题（“混淆矩阵”）与坐标轴标签（“预测值”、“真实值”）显示正常无乱码。
7. 在 `train_and_evaluate` 函数中，生成并保存通话分钟数分布的直方图至 `call_minutes_distribution.png`，该图必须包含 30 个条柱（bins），条柱颜色为 `skyblue` 且带有黑色边框，并确保中文标题（“通话分钟数分布”）与坐标轴标签（“通话分钟数”、“频次”）显示正常无乱码。

商务数据分析师（三级）操作技能考核

试题评分表

试题代码：2.2.1

试题名称：电信客户流失预测考核

考核时间：15min

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M1-月均账单金额计算成功”“M1-示例值:[值]”；验证 processed_data.csv 包含正确月均账单金额列（账单金额/12）。	结果正确，且使用了 df['bill_amount'] / 12 的标准计算方法，得 2 分。 结果正确，但未使用 df['bill_amount'] / 12 的标准计算方法，得 1 分。 结果错误，得 0 分。		
M2	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M2-高账单低通话标志计算成功”“M2-示例值:[0 或 1]”；验证 processed_data.csv 包含正确高账单低通话列（账单金额>均值且通话分钟数<均值）。	结果正确，且使用了布尔掩码 astype(int) 或 np.where 的向量化方法，得 2 分。 结果正确，但使用了循环判断的低效方法，得 1 分。 结果错误，得 0 分。		
M3	2	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M3-年龄缺失值填补成功”“M3-填补记录数:[值]”（记录填补前缺失值数量，约 500 条）；验证 processed_data.csv 的年龄列无缺失值。	结果正确，且使用了中位数或均值进行填补，得 2 分。 结果正确，但使用了众数或固定值进行填补，得 1 分。 结果错误，得 0 分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M4	1	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M4-特征标准化成功”“M4-通话分钟数均值:[≈ 0]”“M4-账单金额标准差:[≈ 1]”；验证 processed_data.csv 的通话分钟数和账单金额列均值 ≈ 0 、标准差 ≈ 1 （接受 StandardScaler 或手动标准化等方法）。	结果正确，且使用了 StandardScaler 或等效的手动标准化方法，得 1 分。 结果正确，但使用了其他缩放方法（如 MinMaxScaler），得 0.5 分。 结果错误，得 0 分。		
M5	1	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M5-模型准确率计算成功”“M5-准确率:[值]”；验证准确率值在合理范围内（0.0-1.0）。	结果正确，且使用了 accuracy_score 函数，得 1 分。 结果正确，但使用了手动计算（如 <code>np.sum(y_test == y_pred) / len(y_test)</code> ）的方法，得 0.5 分。 结果错误，得 0 分。		
M6	1	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M6-混淆矩阵热图生成成功”png 文件存在，为蓝色调，带数字标注，中文标题（“混淆矩阵”）和轴标签（“预测值”“真实值”）无乱码。	结果正确，且完全满足所有格式要求（蓝色调、带数字标注），得 1 分。 结果正确，但未完全满足所有格式要求（如未使用蓝色调、或缺少数数字标注），得 0.5 分。 结果错误（如图未生成或中文乱码），得 0 分。		
M7	1	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M7-通话分钟数分布图生成成功”；png 文件存在，	结果正确，且完全满足所有格式要求（30 个 bins、skyblue 颜色、黑色边框），		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
		为直方图（30 个 bins, skyblue 颜色, 黑色边框），中文标题（“通话分钟数分布”）和轴标签（“通话分钟数”“频次”）无乱码。	得 1 分。 结果正确, 但未完全满足所有格式要求（如 bins 数量不对、或颜色/边框不符），得 0.5 分。 结果错误(如图未生成或中文乱码)，得 0 分。		
合计配分	10	合计得分			

商务数据分析师（三级）操作技能考核 试题单

试题代码：3.1.1

试题名称：设计商铺数据可视化方案

考核时间：15 min

1.场地设备要求

- (1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- (2) 系统中已安装数据分析软件、压缩软件及办公软件；
- (3) 数据库及商务数据分析师考试平台。

2.工作任务

- (1) 根据业务需求及公司数据状况，完成数据可视化实现方式的设计任务；
- (2) 根据业务场景及视觉对象特点，开展可视化展板的设计与布局规划工作；
- (3) 根据业务要求及分析结果，完成数据展示方案的制定与呈现。

3.技能要求

- (1) 设计数据可视化实现方式；
- (2) 设计与布局可视化展板；
- (3) 结合业务要求和分析成果，制定并呈现数据展示方案。

4.质量指标

- (1) 精准锚定公司业务需求，正确梳理数据类型与特性及筛选关联数据，合理设计可视化视觉载体；
- (2) 准确分析业务场景和视觉对象特征，合理规划可视化展板架构，科学布局画布页面；
- (3) 深度解析业务需求内核，分析结果要点提炼无误，合理构建数据展示方案，语句通顺，逻辑清晰。

背景材料

近年来中国百货行业处于变革浪潮，市场规模随城镇化与居民收入提升稳步扩张，三线及以下城市成为新增长极。消费场景加速融合，百货店转型为集购物、餐饮、娱乐、体验服务于一体的综合空间，美妆等高附加值品类销售占比提升，拉动单店营收。竞争格局中，外资品牌，如老佛爷、伊势丹依托国际资源与高端服务深耕一线，本土连锁，如美罗、银泰借数字化与供应链下沉向低线渗透。消费者行为变迁显著，即时性消费需求凸显，对购物便捷时效要求提高，同时数字化深度影响决策，线上渠道成为信息获取、比价与分享的关键途径。

在此行业背景下，海天百货公司收集了 2023-2025 年上海分店的月度商铺明细，涵盖店铺数、店铺排名，以及包括餐饮、普通服饰、儿童、珠宝等在内的业态和楼层等详细信息。公司员工张海东承担起分析这些数据的重任，旨在帮助企业解决三大需求。市场规模与潜力，挖潜下沉市场，评估高附加值品类以优化货架；竞争格局应对，拆解外资与本土连锁策略，评估业态竞争力，寻差异化路径，强化本土服务与数字化能力；消费者行为适配，速应即时消费需求，优化供应链响应，借数字化渠道与社交营销破局，以体验式业态创新提升客流与消费。

操作要求：

1. 结合该公司的业务战略目标与现有数据维度，构建多维度、交互式的数据可视化实现方式，呈现各店铺与业态经营差异及核心特征，为决策提供依据；
2. 紧密贴合业务场景的实际需求与可视化视觉对象的独特性，运用前沿的数据可视化设计方法论，打造具备高度交互性、灵活扩展性的展板布局；
3. 作为本次数据分析任务负责人，基于业务战略研判与数据深度梳理，运用 PPT 定制呈现契合业务场景、赋能决策的数据展示方案，助力精准把握市场、优化运营。（请提交操作完成后的数据文件至考试平台。）

商务数据分析师（三级）操作技能考核

试题评分表

试题代码：3.1.1

试题名称：设计商铺数据可视化方案

考核时间：15 min

评价分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	3	所选属性合理、形成3个正确的可视化视觉对象=3分。				
		所选属性基本合理、形成至少2个正确的可视化视觉对象=2分。				
		所选属性基本合理、形成1个正确的可视化视觉对象=1分。				
		所选特征不合理并形成错误的可视化视觉对象=0分。				
J2	2	设计展板色彩体系完整规范、图表类型丰富多样、布局科学合理=2分。				
		设计展板色彩体系基本规范、图表类型较为单一、布局基本合理=1分。				
		设计展板色彩混乱，图表类型单一且不匹配数据特征，布局杂乱无章=0分。				
J3	1	设计方案框架中封面和目录完整、设计规范=1分。				
		设计方案框架中封面或目录缺失、设计基本规范=0.5分。				
		设计方案框架中缺失封面或目录、不规范=0分。				
J4	1	业务需求提及市场规模与潜力洞察需求、竞争格局与业态分化应对需求、消费者行为变迁适应需求等关键词中的2个及以上，且描述准确=1分。				

		业务需求提及市场规模与潜力洞察需求、竞争格局与业态分化应对需求、消费者行为变迁适应需求等关键词中的 1 个，或描述基本准确=0.5 分。				
		业务需求未提及市场规模与潜力洞察需求、竞争格局与业态分化应对需求、消费者行为变迁适应需求等关键词，或语句不通顺、毫无逻辑=0 分。				
J5	3	展示方案中呈现 3 个视觉对象，数据分析准确合理=3 分。				
		展示方案中呈现 2 个视觉对象，或数据分析基本合理=2 分。				
		展示方案中呈现 1 个视觉对象，或数据分析存在不足=1 分。				
		展示方案中视觉对象缺失，或数据分析存在重大问题=0 分。				
合计配分	10	合计得分				

商务数据分析师（三级）操作技能考核 试题单

试题代码：3.2.1

试题名称：开展家电营销数据可视化结果分析并实现客户留存

考核时间：10 min

1. 场地设备要求

- (1) 装有 Windows 操作系统的计算机（Windows 7 及以上），至少 8GB 内存，10GB 可用磁盘空间；
- (2) 系统中已安装数据处理软件、Python 3.8+、MySQL Server（8.0+ 版本）、压缩软件及办公软件；
- (3) 商务数据分析师考试平台。

2. 工作任务

- (1) 根据某公司业务需求，完成商务数据报告撰写任务；
- (2) 根据某公司运营数据，完成商务数据对比分析任务；
- (3) 根据某公司运营现状，完成商务数据驱动决策任务。

3. 技能要求

- (1) 能对数据可视化结果进行业务分析；
- (2) 能填写或阅读商务数据分析报表；
- (3) 能输出商务数据分析报告；
- (3) 能提出并撰写数据驱动决策。

4. 质量指标

- (1) 根据相关数据，准确撰写商务数据分析报告或报表的内容，内容合理；
- (2) 准确解读工程文件中的商务数据，并对比分析相关数据，判断无误；
- (3) 根据相关数据，准确计算与描述分布率和占比等数据，数据精确；
- (4) 根据要求，针对忠实客户流失问题提出对策，对策合理、有效。

背景材料

上海海东商贸集团旗下拥有一家家电公司，经营 A、B、C、D、E 五种产品，开设了两家线下门店，并在天猫、京东和唯品会开设了三家网店，但由于本年度出现了经营决策失误，导致了部分忠诚客户的流失，集团对该问题非常重视，希望“海东家电”能够出具一份商务数据分析报告，你是“海东家电”的市场部经理，请撰写一份商务数据分析报告，报告的主体部分已经完成，但是，部分数据表和内容需要完善。

具体要求：

1. 请根据数据，简要描述 2016 年 11 月 15 日的基本客户情况，包括总客户数、忠实客户数、忠实客户新增数、忠实客户流失数；
 2. 已知公司忠实客户的流失率和流失女客户占比的“警戒线”分别为 20% 和 60%，请查找与对比相关数据，简要描述当前（即 2016 年 11 月）的忠实客户流失率和流失女客户占比，并判断是否超过“警戒线”。
 3. 请根据数据，分析会员类型分布情况，包括各种会员类型的人数、占比、最高占比和最低占比；
 4. 请根据流失客户数据及分析，提出 2 条忠诚客户流失的阻断策略。
- （注：所有比率保留两位小数）

上海海东家电有限责任公司 忠实客户流失数据分析报告

市场部 张三

一、基本客户情况

(1)

二、流失客户分析

(2)

在流失的忠实客户中，年龄段分布情况为：90 后为 941 人，占比 19.35%；80 后为 1258 人，占比 25.87%；70 后为 1159 人，占比 23.83%；60 后为 800 人，占比 16.45%；50 后为 705 人，占比 14.50%。其中，占比最高的是 80 后，最低

的是 50 后。

在流失的忠实客户中，会员类型分布情况为：_____ (3)

在流失的忠实客户中，最后消费品分布情况为：A 商品为 520 件，占比 10.69%；B 商品为 1422 件，占比 29.24%；C 商品为 1392 件，占比 28.62%；D 商品为 948 件，占比 19.49%；E 商品为 581 件，占比 11.95%。其中，占比最高的是 B 商品和 C 商品，最低的是 A 商品。

在流失的忠实客户中，最后消费渠道分布情况为：No1 门店为 671 人，占比 13.80%；No2 门店为 420 人，占比 8.64%；天猫为 1594 人，占比 32.78%；京东为 1713 人，占比 35.23%；唯品会为 465 人，占比 9.56%。其中，占比最高的是天猫和京东，最低的是 No2 门店。

三、问题及原因分析

.....

四、提出对策

_____ (4)

(请在考试系统相应位置填写答案。)

商务数据分析师（三级）操作技能考核 试题评分表

试题代码：3.2.1

试题名称：开展家电营销数据可视化结果分析并实现客户留存

考核时间：10 min

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	2	简要描述 2016 年 11 月 15 日的基本客户情况，包括总客户数、忠实客户数、忠实客户新增数、忠实客户流失数。	正确描述 1 项数据=0.5 分，共 4 项；错误或未答题=0 分。		
M2	2	在所提供的数据可视化结果文件中筛选与计算 2016 年 11 月的忠实客户流失率和流失女客户占比数据，并判断是否超过公司“警戒线”。	正确描述或判断 1 项数据=0.5 分，共 4 项；错误或未答题=0 分。		
M3	3	分析会员类型分布情况，包括各种会员类型的人数、占比、最高占比和最低占比。	正确分析一类会员数据=0.5 分，共 5 类会员；正确分析最高占比和最低占比会员=0.5 分；错误或未答题=0 分。		
合计配分	7	合计得分			

评价分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	1.5	第一条建议合理有效、逻辑正确、具有创新性=1.5 分。				
		第一条建议合理有效、逻辑正确=1 分。				

		第一条建议基本合理=0.5分。				
		第一条建议不合理或未答题=0分。				
J2	1.5	第二条建议合理有效、逻辑正确、具有创新性=1.5分。				
		第二条建议合理有效、逻辑正确=1分。				
		第二条建议基本合理=0.5分。				
		第二条建议不合理或未答题=0分。				
合计配 分	3	合计得分				

商务数据分析师（三级）操作技能考核 试题单

试题代码：4.1.1

试题名称：挖掘精密制造数据并研究竞争态势，缮制商务数据分析报告

考核时间：25 min

1.场地设备要求

- (1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- (2) 系统中已安装数据处理软件、压缩软件及办公软件；
- (3) 数据库及商务数据分析师考试平台。

2.工作任务

- (1) 根据相关数据，完成市场行情分析的工作任务；
- (2) 根据相关数据，完成行业状况分析的工作任务；
- (3) 根据相关数据，完成竞争对手分析的工作任务；
- (4) 根据相关数据，完成竞争产品分析的工作任务；
- (5) 根据公司业务需求，完成撰写商务数据分析报告的工作任务。

3.技能要求

- (1) 进行市场行情分析；
- (2) 进行行业状况分析；
- (3) 进行竞争对手分析；
- (4) 进行竞争产品分析；
- (5) 撰写商务数据分析报告。

4.质量指标

- (1) 基于行业数据特征深度解析，精准匹配可视化分析范式，系统性分析市场行情；
- (2) 准确理解相关给定数据，正确选用恰当的可视化视觉对象，对行业情况进行分析；
- (3) 基于行业数据特征深度解读，精准匹配可视化分析范式，精准分析竞争对手情况；
- (4) 基于行业数据特征深度解读，精准匹配可视化分析范式，精准分析竞争产

品情况；

(5) 基于多维度可视化视觉对象分析结果，撰写具备战略穿透力的行业商务分析报告，数据-结论-策略三级逻辑闭环，语句通顺，逻辑合理。

背景材料

精密制造业正经历技术迭代加速+跨界竞争渗透的双重挤压，行业生态呈现攻防易位的动态博弈：一方面，技术扩散周期缩短至 3-5 年，催生近 300 家初创企业（2021-2023 年全球融资额超 120 亿美元）以颠覆性创新切入细分赛道，对传统巨头形成技术奇袭；另一方面，替代品威胁从材料端、工艺端到模式端全链条渗透，叠加下游客户对“性能-成本-交付”三角的极致追求，推动行业年均价格战强度达 12%-15%；中国作为全球最大增量市场，在政策（“强基工程”专项补贴）、资本与产业协同（长三角精密制造集群）三重驱动下，正成为新老玩家混战的主战场，亟需通过数据化工具穿透技术替代风险、客户忠诚度衰减与成本结构化，为竞争策略制定提供动态校准依据。

操作要求：

1. 请根据所提供的数据，选用一个恰当的可视化视觉对象，分析 2014 年至 2023 年间精密制造业的市场行情，可以基于供应商的数量、消费者偏好变化、市场份额等维度；

2. 请根据所提供的数据，选用一个恰当的可视化视觉对象，分析 2014 年至 2023 年间精密制造业的行业状况，可以基于新进入者数量与现有主要竞争者数量关系、替代品数量、供应商数量与现有竞争者数量等维度；

3. 请根据所提供的数据，选用一个恰当的可视化视觉对象，分析 2014 年至 2023 年间精密制造业的竞争对手情况，可以基于新进入者数量及市场份额、新进入企业等维度；

4. 请根据所提供的数据，选用一个恰当的可视化视觉对象，分析 2014 年至 2023 年间精密制造业的竞争产品情况，可以基于替代品市场份额、替代品年增长率等维度；

5. 基于提供的数据及以上数据分析，请为精密制造业书写一份商务数据分析报告，报告需包含基本模块：标题、分析目的与意义、市场行情、行业情况、竞争对手、竞争产品分析、困境分析、策略分析、结论。要求数据-结论-策略三级逻辑闭环、量化结论支撑业务决策、可视化结论复用设计。（请提交操作完成后的数据文件至考试平台。）

商务数据分析师（三级）操作技能考核 试题评分表

试题代码：4.1.1

试题名称：挖掘精密制造数据并研究竞争态势，缮制商务数据分析报告

考核时间：25 min

评价分评分表

细则编号	配分	评分细则描述	考评员评分			得分
			1	2	3	
J1	2	所选市场行情分析维度合理、适合并形成正确的可视化视觉对象=2分。				
		所选市场行情分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选市场行情分析维度不合理或形成错误的可视化视觉对象=0分。				
J2	2	所选行业状况分析维度合理、适合并形成正确的可视化视觉对象=2分。				
		所选行业状况分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选行业状况分析维度不合理或形成错误的可视化视觉对象=0分。				
J3	2	所选竞争对手分析维度合理、适合并形成正确的可视化视觉对象=2分。				
		所选竞争对手分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选竞争对手分析维度不合理或形成错误的可视化视觉对象=0分。				
J4	2	所选竞争产品分析维度合理、适合并形成正确的可视化视觉对象=2分。				
		所选竞争产品分析维度基本合理并形成正确的可视化视觉对象=1分。				
		所选竞争产品分析维度不合理或形成错误的可视化视觉对象=0分。				

J5	1	标题合理正确=1分。				
		标题不合理、错误=0分。				
J6	2	报告的目的与意义分析全面、逻辑合理=2分。				
		报告的目的与意义分析基本合理=1分。				
		报告的目的与意义分析不合理=0分。				
J7	1	分析市场行情全面、逻辑合理=1分。				
		分析市场行情基本合理=0.5分。				
		分析市场行情不合理=0分。				
J8	1	分析行业状况全面、逻辑合理=1分。				
		分析行业状况基本合理=0.5分。				
		分析行业状况不合理=0分。				
J9	1	分析竞争对手全面、逻辑合理=1分。				
		分析竞争对手基本合理=0.5分。				
		分析竞争对手不合理=0分。				
J10	1	分析竞争产品全面、逻辑合理=1分。				
		分析竞争产品基本合理=0.5分。				
		分析竞争产品不合理=0分。				
J11	2	分析行业困境全面、逻辑合理=2分。				

		分析行业困境基本合理=1分。				
		分析行业困境不合理=0分。				
J12	2	提出两条合理、有效、实用策略=2分。				
		提出一条合理、有效、实用策略=1分				
		提出一条基本合理策略=0.5分。				
		所提策略不合理=0分。				
J13	1	结论归纳全面=1分。				
		部分归纳=0.5分。				
		未归纳=0分。				
合计 配分	20	合计得分				

商务数据分析师（三级）操作技能考核 试题单

试题代码：4.2.1

试题名称：跨境电商贸易数据申报与波特五力分析

考核时间：20 min

1.场地设备要求

- (1) 装有 Windows 操作系统的计算机（Windows 7 及以上）；
- (2) 系统中已安装 BI 类商务数据分析软件、压缩软件、安装 Python3.8+、Python 库（pandas、mysql-connector-python、matplotlib、seaborn）、Mysql8.0+ 及办公软件；
- (3) 商务数据分析师考试平台。

2.工作任务

- (1) 统计各平台的交易总额，生成用于跨境电商数据申报的汇总数据；
- (2) 筛选符合退税条件的交易，计算退税金额，生成出口退税申报数据；
- (3) 统计各口岸的货运总量，生成口岸物流申报数据；
- (4) 绘制波特五力模型图，清晰展示跨境电商行业的竞争格局；
- (5) 分析跨境电商行业的竞争者、供应商、买家、替代品和新进入者，生成详细的市场竞争报告。

3.技能要求

- (1) 采集跨境电子商务申报数据；
- (2) 采集出口退税申报数据；
- (3) 采集口岸物流申报数据；
- (4) 画出波特五力模型图；
- (5) 进行波特五力分析。

4.质量指标

- (1) 跨境电商交易总额统计准确，符合申报要求；
- (2) 退税金额计算精确，数据筛选正确；
- (3) 口岸货运总量统计准确，数据完整；
- (4) 波特五力模型图清晰，包含所有竞争维度；
- (5) 波特五力分析全面，结论合理且基于数据。

背景材料

某跨境电子商务企业通过多个平台（如亚马逊、阿里巴巴、易贝）向全球销售商品，涉及跨境电商数据申报、出口退税申报和口岸物流数据申报。数据存储在 MySQL 数据库，包含交易和物流记录。为支持市场竞争分析，需基于数据应用波特五力模型，评估行业竞争格局。

操作要求：

1. 使用 SQL 聚合查询 transactions 表，按 platform 分组统计总交易额（total_amount），并将结果保存至 platform_amount.csv，金额需保留两位小数。
2. 使用 SQL WHERE 子句筛选 transactions 表中符合出口退税条件的交易记录（amount ≥ 1000.00 且 country 为“美国”、“日本”或“德国”），并将结果保存至 tax_rebate_eligible.csv，金额需保留两位小数。
3. 基于筛选出的退税交易记录，使用 Pandas 向量化操作计算每笔交易的退税金额（tax_rebate，为 amount 的 10%），并将包含新列的结果保存至 tax_rebate_amount.csv，金额需保留两位小数。
4. 使用 SQL 聚合查询 logistics 表，按 port 分组统计总货运重量（total_weight），并将结果保存至 port_weight.csv，重量需保留两位小数。
5. 使用 SQL 聚合查询 transactions 表，按 country 分组统计交易笔数（transaction_count），并将结果保存至 country_transaction_count.csv。
6. 基于交易总额数据，生成并保存交易总额柱状图至 platform_amount_bar.png，图像尺寸必须为 576x432 像素，并确保中文标题（“各平台交易总额”）与坐标轴标签（“平台”、“交易总额（元）”）显示正常无乱码。
7. 生成并保存波特五力模型图至 porter_five_forces.png，图像尺寸必须为 576x432 像素，图中必须包含分别指向“供应商”、“买家”、“替代品”、“新进入者”的箭头，并确保所有中文标注显示正常无乱码。
8. 基于已处理的交易和物流数据，撰写一份包含对“竞争者”、“供应商”、“买家”、“替代品”、“新进入者”五个维度分析的市场竞争报告，报告内容必须明确引用相关统计数据（如百分比、金额、总量等），并保存至 porter_analysis.txt，数值需保留两位小数。

商务数据分析师（三级）操作技能考核

试题评分表

试题代码：4.2.1

试题名称：跨境电商贸易数据申报与波特五力分析

考核时间：20 min

测量分评分表

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
M1	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M1 - 交易总额统计成功”“M1 - 示例值：[值]”；验证 platform_amount.csv 包含正确 platform 和 total_amount 列，金额无缺失且保留两位小数。	结果正确，且使用了 SQL 聚合查询的方式，得 2.5 分。 结果正确，但使用了先查询后在 Pandas 中聚合的方法，得 1.5 分。 结果错误，得 0 分。		
M2	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M2 - 退税条件筛选成功”“M2 - 示例值：[值]”；验证 tax_rebate_eligible.csv 包含正确 transaction_id, amount, country 列，满足金额 ≥ 1000.00 且国家在“美国”“日本”“德国”，金额保留两位小数。	结果正确，且使用了 SQL WHERE 子句进行筛选，得 2.5 分。 结果正确，但使用了先查询后在 Pandas 中筛选的方法，得 1.5 分。 结果错误，得 0 分。		
M3	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M3 - 退税金额计算成功”“M3 - 示例值：[值]”；验证 tax_rebate_amount.csv 包含正确 transaction_id, tax_rebate 列，退税金额为金额的 10%，保留两位小数。	结果正确，且使用了 Pandas 向量化 (df['amount'] * 0.1) 的方式计算，得 2.5 分。 结果正确，但使用了 .apply 或循环的低效方法，得 1.5 分。		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
			结果错误，得 0 分。		
M4	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M4 - 货运总量统计成功”“M4 - 示例值: [值]”；验证 port_weight.csv 包含正确 port 和 total_weight 列,重量无缺失且保留两位小数。	结果正确，且使用了 SQL 聚合查询的方式，得 2.5 分。 结果正确，但使用了先查询后在 Pandas 中聚合的方法，得 1.5 分。 结果错误，得 0 分。		
M5	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M5 - 交易笔数统计成功”“M5 - 示例值: [值]”；验证 country_transaction_count.csv 包含正确 country 和 transaction_count 列,笔数无缺失且正确。	结果正确，且使用了 SQL 聚合查询 (COUNT(*))的方式，得 2.5 分。 结果正确，但使用了先查询后在 Pandas 中聚合 (如 groupby().size()) 的方法，得 1.5 分。 结果错误，得 0 分。		
M6	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M6 - 交易总额柱状图生成成功”“M6 - 图像尺寸: 576x432”；platform_amount_bar.png 存在，尺寸为 576x432 像素 (figsize=(8, 6), dpi=72)，包含标题和正确柱状图，中文无乱码，金额显示保留两位小数。	结果正确，且完全满足所有格式要求 (seaborn 或 matplotlib 均可)，得 2.5 分。 结果正确，但未完全满足所有格式要求 (如尺寸不对、或缺少标题/标签)，得 1.5 分。 结果错误 (如图未生成、图表类型错误或中文乱码)，得 0 分。		
M7	2.5	考评员运行考生提交的 PY 文件，检查运行过程日志显示“M7 - 波特五力模型图生成成功”“M7 -	结果正确，且完全满足所有格式要求 (包含所有标注、颜色框		

细则编号	配分	评分细则描述	规定或标称值	结果或实际值	得分
		图像尺寸: 576x432” ; porter_five_forces.png 存在, 尺寸为 576x432 像素 (figsize=(8, 6), dpi=72) , 包 含标题、箭头和五力标注, 中文无 乱码。	和箭头), 得 2.5 分。 结果正确, 但未完全 满足所有格式要求 (如缺少箭头、或标 注不全), 得 1.5 分。 结果错误 (如图未生 成、或中文乱码), 得 0 分。		
M8	2.5	考评员运行考生提交的 PY 文件, 检查运行过程日志显示 “M8 - 波 特五力分析成功” ; porter_analysis.txt 存在, 包含 竞争者、供应商、买家、替代品、 新进入者分析, 每项至少 50 字, 基于数据且结论合理, 数值保留两 位小数。	结果正确, 且分析内 容明确引用了数据 (如百分比、金额、 总量等), 且满足所 有其他要求, 得 2.5 分。 结果正确, 但分析内 容泛泛而谈, 未明确 引用数据, 或字数不 足, 得 1.5 分。 结果错误 (如文件未 生成、缺少分析维 度), 得 0 分。		
合计配 分	20				