

## 2020 秋双学士《商务统计与数据分析》期末课程作业

### 1、课程作业要求：

- (1) 按规定完成相应的数据分析任务，装订成册（加封面，加注名字和学号）后提交；
- (2) 字体使用宋体小四号，英文使用 Times New Roman；单倍行距；
- (3) 独立完成作业，作业内容需包括任务相关的 python 代码和结果；
- (4) 请在规定期限内提交作业，否则按照未提交作业计算成绩；
- (5) 如果文献引用参考，请加注参考文献来源；
- (6) 注意整个作业手册和绘制的图表的美观性以及专业性，这里会有 10 分评价成绩，每一页右上角加注页码；
- (7) 整个课业报告页数不少于 20 页；
- (8) 请将本作业要求作为封面之后的第一页装订在报告里面。

### 2、数据分析任务

2.1 餐饮团购活动.xlsx 这个数据集里记录了某团购网站上某段时间内餐饮团购活动内容的具体信息。餐饮团购店铺.xlsx 这个数据集中记录了团购店铺的相关数据，可以从这个数据集了解到每家餐饮店铺的特征数据。基于这两个数据，进行相应的数据整理与清洗，试着回答以下问题：（合计 20 分）

- (1) 有多少家店铺参加了团购活动？
- (2) 最受消费者喜爱的 Top5 餐饮类型是哪些？
- (3) 销售最多的店铺 Top10 是哪些？
- (4) 团购优惠的额度是否和购买数量存在关系？什么关系？
- (5) 不同行政区域，口碑良好（选择评分大于 4 分）的商家数量？

2.2 使用 python 对京东超市销售数据.csv 数据集中的合适变量进行基本的描述性统计分析，主要包括数据的频数分析、数据的集中趋势分析、数据离散程度分析、数据的分布、变量间的相关性以及一些基本的统计图形比如直方图和盒状图，并对结果进行管理学意义上的讨论。（小组为单位完成；合计 20 分）；

2.3 成都市青羊区锦江区是公认的小学初中教育资源最好的行政区，因此二手房价相对较高。课程作业文件夹中包含青羊区二手房.csv 格式数据文档文件一份。数据为青羊区挂牌二手房的相关数据。要求：整理与描述青羊区挂牌二手房的相关信息，建立二手房房价的预测模型，揭示影响二手房房价的主要因素并对结果进行管理学意义上的阐述（小组为单位完成；合计 20 分）；

2.4 基于纽约地铁站人流量数据（Newyork.csv），使用 FineBI 或 Python 进行数据的可视化描述天气状况如何影响地铁客流？这种影响效应是否存在差异性（在不同的条件下影响效应具有差异）？输出图表结果并对结果进行较为详细的阐述（小组为单位完成；合计 20 分）

2.5 51job.csv 存放了 51job 网站成都地区的关于数据分析相关的职位的信息，试使用 python 的 numpy 和 pandas 包对数据进行整理清洗与分析，回答如下问题：（个人完成；合计 20 分）

- (1) 哪个区对数据分析相关的岗位需求最高？
- (2) 对数据分析相关有需求的企业具有什么特征？
- (3) 薪资最高的 10 家企业？
- (4) 数据分析相关的职位有哪些并进行排序？

(5) 对岗位职位进行可视化词云分析，并对结果进行阐述；