**模块B：数据分析（40分）**

## 背景简介

校服被形象地称为中小学生的“第二层皮肤”，其质量关系学生的健康成长，式样影响学生的形象和气质养成。优质、合体、美观、舒适的校服是培育校园文化的重要载体，是培养团队意识、传播平等精神的有益方式，是传承中华优秀传统文化的积极探索。中国作为人口大国、教育强国，对于校服的需求规模巨大，根据2021年教育事业统计数据结果显示，全国共有各级各类学校52.93万所，在校生2.91亿人，中小学生1.97亿人。

面对巨大的校服市场规模，如何分配资源，例如：不同月份的供应数量、各校服颜色受欢迎的程度等。通过数据分析，能让我们以数据的方式去看待校服各个方面的需求，在满足用户需求的情况下，提高自己的利润，实现双赢。

## 数据介绍

校服销售订单表：

| **NO.** | **字段名** | **备注** |
| --- | --- | --- |
| 1 | 店铺名称 |  |
| 2 | 订单号 | 客户实际购买生成的单号 |
| 3 | 系统单号 |  |
| 4 | 仓库编码 |  |
| 5 | 仓库名称 |  |
| 6 | 系统订单状态 |  |
| 7 | 线上订单状态 |  |
| 8 | 明细状态 | 订单完成的情况 |
| 9 | 原订单号 | 客户下单生成的单号（下单未付款、下单付款） |
| 10 | 商品编码 |  |
| 11 | 商品名称 |  |
| 12 | 规格名称 |  |
| 13 | 品牌 |  |
| 14 | 参考进价 |  |
| 15 | 单价 |  |
| 16 | 折后单价 | 商品打折后的价格 |
| 17 | 数量 | 购买的数量 |
| 18 | 单位 | 件 |
| 19 | 应收 | 这行数据中所有折后商品的价格与数量的乘积 |
| 20 | 销售金额 |  |
| 21 | 发货时间 | 订单发货时间 |
| 22 | 完成时间 | 客户确认收货的时间 |
| 23 | 付款时间 |  |
| 24 | 买家ID(客户昵称) |  |

校服小类信息表

| **NO.** | **字段名** | **备注** |
| --- | --- | --- |
| 1 | 款号 |  |
| 2 | SKU | 库存保有单位 |
| 3 | 学校 |  |
| 4 | 吊牌价 |  |
| 5 | 销售价 |  |
| 6 | 季节 | 商品所属季节 |
| 7 | 小类 | 商品类别 |

**任务一：数据探索与处理**

1.1读取数据

读取校服销售订单表、校服小类信息表两个表的数据，将校服销售订单表，字段名为数量这一列设为整形。

校服销售订单表路径：'./校服销售订单.xlsx'

校服小类信息表路径：'./校服小类信息.xlsx'

将读取校服销售订单表的所有数据保存到变量clothes\_initial中，读取校服小类信息表的所有数据保存到变量news\_initial中，并运行给出的答案保存代码保存答案

例如：变量名 = pd.read\_excel('表名.xlsx',dtype = {'字段名1':str, '字段名2':int})

1.2 查看校服小类信息表中学校是卡帕足球的所有数据，并保存查看的结果到变量news\_soccer中，并运行给出的答案保存代码保存答案

例如：表名.query('字段名==“元素名”')

1.3 对校服小类信息表进行处理，删除任务1.2查询出的所有数据,以及存在空值的所有行，将处理结果保存到变量news\_clean中，并运行给出的答案保存代码保存答案

1.4 查找news\_clean中'SKU'为'K0AY2CQ92-134'的所有数据，并将这些数据所对应的吊牌价替换为399,将处理后的news\_clean保存到变量news中，并运行给出的答案保存代码保存答案

1.5 对变量clothes\_initial进行处理,通过付款时间字段生成月份、季节两列，变量名分别为month、season，将处理结果保存到变量clothes\_all中，并运行给出的答案保存代码保存答案

季节划分规则：3、4、5为春，6、7、8为夏，9、10、11为秋，12、1、2为冬

变量clothes\_initial中字段season为用户购买时所属季度，变量news中字段季节为商品所属季度

**任务二：订单分析**

2.1分析变量clothes\_all，哪个季节完成付款的订单数量最多？结果保存到变量task2\_1中，并运行给出的答案保存代码保存答案

2.2分析变量clothes\_all，在线上订单状态为线下订单的购买记录中，计算其中打折商品的件数占比，结果保存到变量task2\_2中，并运行给出的答案保存代码保存答案（保留两位小数）

打折商品：折后单价小于单价的商品

2.3分析变量clothes\_all，计算订单发货后快递运输的天数（完成时间等于0的数据和线上订单状态的线下订单数据不考虑），然后返回运输天数的中位数（相同原订单号只计算一次），并保存到变量task2\_3中，并运行给出的答案保存代码保存答案，并运行给出的答案保存代码保存答案

例如9月8号发货，9月9号到达，快递运输天数就为一天，9月8号发货，9月10号到达，快递运输天数就为2天

2.4处理变量clothes\_all，对完成时间字段等于0的异常值数据按照以下公式处理，并保存处理完成后的数据到原变量clothes\_all，并运行给出的答案保存代码保存答案

2.5分析变量clothes\_all，预售商品中最受欢迎的款式是什么？结果保存到变量task2\_5中，并运行给出的答案保存代码保存答案

预售商品:字段商品编码前两位为'YS'的商品

款式在商品名称字段里’KAPPA校服 ‘与’YS‘之间获取，例如：'预售 小学款 十四中 KAPPA校服 针织开衫外套中性款YS0AY2WK92'，款式就是'针织开衫外套中性款'

**任务三：用户行为分析**

3.1 按照以下要求处理变量clothes\_all和news:

* 对clothes\_all变量进行处理，通过商品编码字段构建新的SKU字段。例如：商品编码为'K0AY2AK93-554-175',SKU就是'K0AY2AK93-554'
* 对clothes\_all变量进行处理，以'YS'开头的SKU值需将'YS'替换为'K'。例如：'YS0AY2CQ92-134',替换后为'K0AY2CQ92-134'
* clothes\_all变量通过SKU字段左连接news变量（news变量只保留学校,SKU,小类,款号,季节五个字段）。
* 保存处理完成后的数据到变量clothes,并运行给出的答案保存代码保存答案。

3.2分析变量clothes，求复购用户数量，将结果保存到变量task3\_2中，并运行给出的答案保存代码保存答案

复购用户：进行两次或两次以上购物的用户（相同原订单号只算一次购物）

3.3分析变量clothes，求新用户占比最低的月份，结果保存为变量task3\_3中，并运行给出的答案保存代码保存答案

3.4分析变量clothes，客单价最高的店铺是？结果保存为变量task3\_4中，并运行给出的答案保存代码保存答案

3.5 分析变量clothes，根据付款时间字段，按以下要求进行分段，求用户每天最喜欢在什么时间段进行购物，结果保存到变量task3\_5中，并运行给出的答案保存代码保存答案

3.6 分析变量clothes，在10月份参与购物的老用户中，进行两次或两次以上购物的用户占比是多少（保留两位小数）？结果保存到变量task3\_6中，并运行给出的答案保存代码保存答案（相同订单只算一次购物）

**任务四：商品分析**

4.1 分析变量clothes，卡帕校服徐水仓共售卖了多少件罗马蓝颜色的商品？结果保存到变量task4\_1中，并运行给出的答案保存代码保存答案

4.2 分析变量clothes，在夏季购物中，不同商品款号最多销售多少件，结果保存到变量task4\_2中，并运行给出的答案保存代码保存答案

4.3 分析变量clothes，按以下价格区间进行分类，求什么价格区间的商品最受欢迎？结果保存到变量task4\_3中，并运行给出的答案保存代码保存答案

**任务五：成本利润分析**

5.1分析变量clothes，9月份打折商品的总销售额，结果保存到变量task5\_1中，并运行给出的答案保存代码保存答案

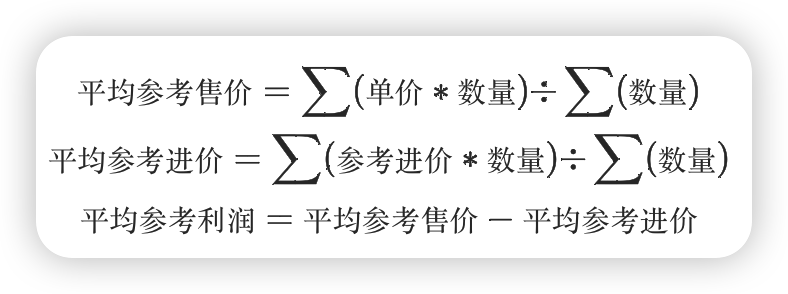
打折商品：折后单价小于单价的商品

5.2 分析变量clothes，查询出利润最大的小类。结果保存到变量task5\_2中，并运行给出的答案保存代码保存答案

5.3 分析变量clothes，计算各自不同商品小类的平均成本，最贵成本需要多少钱？结果保存到变量task5\_3中，并运行给出的答案保存代码保存答案（成本价格，保留两位小数）

5.4 分析变量clothes，'卡帕校服徐水仓'9月的利润率为多少？结果保存到变量task5\_4中，并运行给出的答案保存代码保存答案（保留两位小数）

5.5 分析变量clothes，按小类字段分类别计算平均参考售价、平均参考进价、平均参考利润，结果保存到变量price\_cost中，并运行给出的答案保存代码保存答案(格式保存为pandas DataFrame表，字段名为平均参考售价，平均参考进价，平均参考利润)



5.6 分析变量clothes，构建出'冲锋衣','防晒衣','休闲裤','运动裤','短袖POLO','短袖图案衫'这六类别搭配的所有组合（不考虑顺序，组合中要求外套、短袖、裤子各一件）结果保存到变量result中，并运行给出的答案保存代码保存答案（2分）

外套：'冲锋衣','防晒衣'

裤子：'休闲裤','运动裤'

短袖：'短袖POLO','短袖图案衫'

结果保存格式,例： [('冲锋衣', '短袖POLO', '休闲裤'), ..., ('防晒衣', '短袖图案衫', '运动裤')]