

## 第 12 期 Datawhale 组队学习

### Pandas 教程（上）综合练习

#### 一、2002 年-2018 年上海机动车拍照拍卖

##### 问题

- (1) 哪一次拍卖的中标率首次小于 5%?
- (2) 按年统计拍卖最低价的下列统计量：最大值、均值、0.75 分位数，要求显示在同一张表上。
- (3) 将第一列时间列拆分成两个列，一列为年份（格式为 20××），另一列为月份（英语缩写），添加到列表作为第一第二列，并将原表第一列删除，其他列依次向后顺延。
- (4) 现在将表格行索引设为多级索引，外层为年份，内层为原表格第二至第五列的变量名，列索引为月份。
- (5) 一般而言某个月最低价与上月最低价的差额，会与该月均值与上月均值的差额具有相同的正负号，哪些拍卖时间不具有这个特点？
- (6) 将某一个月牌照发行量与其前两个月发行量均值的差额定义为发行增益，最初的两个用 0 填充，求发行增益极值出现的时间。

#### 二、2007 年-2019 年俄罗斯机场货运航班运载量

##### 问题

- (1) 求每年货运航班总运量。
- (2) 每年记录的机场都是相同的吗？

- (3) 按年计算 2010 年-2015 年全年货运量记录为 0 的机场航班比例。
- (4) 若某机场至少存在 5 年或以上满足所有月运量记录都为 0，则将其所有年份的记录信息从表中删除，并返回处理后的表格
- (5) 采用一种合理的方式将所有机场划分为东南西北四个分区，并给出 2017 年-2019 年货运总量最大的区域。
- (6) 在统计学中常常用秩代表排名，现在规定某个机场某年某个月的秩为该机场该月在当年所有月份中货运量的排名（例如 \*\*\* 机场 19 年 1 月运量在整个 19 年 12 个月中排名第一，则秩为 1），那么判断某月运量情况的相对大小的秩方法为将所有机场在该月的秩排名相加，并将这个量定义为每一个月的秩综合指数，请根据上述定义计算 2016 年 12 个月的秩综合指数。

### 三、新冠肺炎在美国的传播

#### 问题

- (1) 用 `corr()` 函数计算县（每行都是一个县）人口与表中最后一天记录日期死亡数的相关系数。
- (2) 截止到 4 月 1 日，统计每个州零感染县的比例。
- (3) 请找出最早出确证病例的三个县。
- (4) 按州统计单日死亡增加数，并给出哪个州在哪一天确诊数增加最大（这里指的是在所有州和所有天两个指标一起算，不是分别算）。
- (5) 现需对每个州编制确证与死亡表，第一列为时间，并且起始时间为该州开始出现死亡比例的那一天，第二列和第三列分别为确证数和死亡数，每个州需要保存为一个单独的 csv 文件，文件名为“州名.csv”。
- (6) 现需对 4 月 1 日至 4 月 10 日编制新增确证数与新增死亡数表，第一列为州名，第二列和第三列分别为新增确证数和新增死亡数，分别保存为十个单独的 csv 文件，文件名为“日期.csv”。