

3.1.x知识点总结

数据统计

数据计数（偏好，频率之类统计）

**代码法**  
data['待统计列'].value\_counts().sort\_values(ascending=False)

**Excel法**  
1、插入数据透视表  
2、选择待统计字段作为行  
3、选择待统计字段作为值，值统计方式选择计数

时间段划分

代码法

**新建小时列**  
data['小时'] = pd.to\_datetime(data['时间列']).dt.hour 新建小时列

**划分时间段**  
data['时间段'] = pd.cut(data['小时'], bins=[0,6,12,18,24], labels=['凌晨', '上午', '下午', '晚上'], right=False)

**新建小时列**  
使用 =hour(时间列单元格) 公式新建一列存放小时

Excel法

划分时间段

**If函数方法**  
时间段列单元格 = IF(小时列单元格<6,"凌晨",IF(小时列单元格<12,"上午",IF(小时列单元格<18,"下午",IF(小时列单元格<24,"晚上","时间异常"))))

**透视表分组法**  
1、插入数据透视表  
2、选择小时列作为行  
3、右键透视表中的小时行选择分组，起始0，终止23，步长6

注意透视表分组法只能用于均匀划分

**注意事项**  
时间段的划分需要根据不同数据来选择不同的划分规则，因为划分规则不一致会导致数据统计的结果也不一致

建议  
1、3.1.3 按照'凌晨','早上','上午','中午','下午','傍晚','晚上' 划分，时间段为0-6，6-9，9-12，12-14，14-18，18-20，20-24

2、3.1.4 不用划分，提取出"时:分"之后直接以"时:分"为分组依据进行后续的数据统计

**代码法**  
data.groupby('时间段列')[['待计算列1', '待计算列2', ...]].mean()

**Excel法**  
1、创建数据透视表（如果使用透视表分组法进行时间段划分就算已经创建好了）  
2、添加时间段到行，待计算列到值，值统计方法选择平均值

**代码法**  
data.corr()

**Excel法**  
1、点击文件，选项，加载项，转到  
2、勾选“分析工具库”，确定  
3、回到表格页面，点击菜单栏的数据，分析栏里的数据分析，选择相关系数  
4、选择需要计算的列，勾选上“标志位于第一行”，点确定

根据时间段计算均值

计算相关系数矩阵

统计结果分析

**数据计数分析**  
对计算后的值排序后直接根据最大值和最小值得出偏好和频率或其他相关分析结论

**根据时间段计算均值分析**  
对计算后的值排序后直接根据最大值和最小值所在时间段得出相关分析结论

**相关系数矩阵**  
相关系数值再-1到1之间，为正则表示正相关，为负则表示负相关，越接近-1或1就表示两个值相关性越高。