

### 题目三：2006\_2010 年美国五个著名城市空气中臭氧污染情况分析

#### 【数据说明】

2006\_2010 年美国五个著名城市纽约（New York）、华盛顿（Washington）、休斯顿（Houston）、洛杉矶(Los Angeles)、费城（Philadelphia）的空气中臭氧（O3）污染情况数据。

字段说明：

- 1) 序号（ID）：记录序号。
- 2) 州代码（State Code）：由美国环保局分配给每个州的代码
- 3) 县代码（County Code）：由美国环保署分配的特定州的代码
- 4) 地点编号（Site Num）：由美国环保局分配的特定县的地点编号
- 5) 地址（Address）：监测站点的地址
- 6) 状态（State）：监测点的状态
- 7) 县（Country）：县监测站点
- 8) 城市（City）：监测点的城市
- 9) 日期本地（Date Local）：监视日期
- 10) O3 污染物有 5 个专栏：
  - O3 单位（O3 Unites）：测量 O3 的单位
  - O3 平均值（O3 Mean）：给定日内 O3 浓度的算术平均值
  - O3 第一最大值（O3 1st Max Value）：给定日期的 O3 浓度的最大值
  - O3 第一最大值时间：（O3 1st Max Hour）：给定日期的 O3 浓度的最大值所处的时间。
  - O3 AQI（O3 AQI）：一天内 O3 计算的空气质量指数

#### 【任务】

- 1、用 pandas 库读取“pollution\_us\_5city\_2006\_2010\_O3.csv”文件，查看前五  
行、后两行。
- 2、选择 2007-2009 年数据集导出到 CSV 文件  
“pollution\_us\_5city\_2007\_2009\_O3.csv”。
- 3、读取新的数据集“pollution\_us\_5city\_2007\_2009\_O3.csv”，分别选择字段  
City==“Houston”、City==“New York”、City==“Washington”的所有数据集，  
分别导出为文本文件“pollution\_us\_Houston\_2007\_2009\_O3.txt”、  
“pollution\_us\_NewYork\_2007\_2009\_O3.txt”、  
“pollution\_us\_Washington\_2007\_2009\_O3.txt”，要求数据之间用空格分  
隔，每行末尾包含换行符。
- 4、分别读取文本文件“pollution\_us\_Houston\_2007\_2009\_O3.txt”、  
“pollution\_us\_NewYork\_2007\_2009\_O3.txt”、

“pollution\_us\_Washington\_2007\_2009\_O3.txt”，并转存到 Excel 文件  
“pollution\_us\_Houston\_2007\_2009\_O3.xlsx”、“pollution\_us\_NewYork\_2007\_2009\_O3.xlsx”、“pollution\_us\_Washington\_2007\_2009\_O3.xlsx”中。

- 5、读取 Excel 文件 “pollution\_us\_Houston\_2007\_2009\_O3.xlsx”、  
“pollution\_us\_NewYork\_2007\_2009\_O3.xlsx”、  
“pollution\_us\_Washington\_2007\_2009\_O3.xlsx”，利用 matplotlib 库，可视化对比显示三个城市 2007-2009 年的 O3 Mean、O3 AQI、O3 1st Max Hour 的值，要求三幅图的名称分别为  
“Houston\_NewYork\_Washington\_2007\_2009\_O3Mean”、  
“Houston\_NewYork\_Washington\_2007\_2009\_O3AQI”、  
“Houston\_NewYork\_Washington\_2007\_2009\_O31stMaxHour”，包括图例，x 轴刻度以年显示，y 轴显示刻度值，每幅图中三条曲线颜色分别为红色、绿色、蓝色。

**【要求】**

- 1、根据以上数据处理任务，设计并编程实现“数据分析与可视化系统”，要求
  - ① 各个任务选择用菜单实现（菜单可用字符串输出模拟，或者 Tkinter 形式实现）。
  - ② 各个任务名称自己定义，须由独立的函数实现，且每个任务执行成功与否须给出必要的文字提示。
  - ③ 数据输入和结果输出的文件名须由人工输入，且输出结果都要以文件形式保存。
  - ④ 为保持程序的健壮性，各个任务执行过程中需要进行必要的判断（如文件是否存在、输入是否合法等）、程序异常控制等。
- 2、根据以上统计结果，书写不少于 300 字的结果分析。