

题目五：2006_2010 年美国五个著名城市空气中一氧化碳污染情况分析

【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）的空气中一氧化碳（NO）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) CO 污染物有 5 个专栏：

CO 单位（CO Unites）：测量 CO 的单位

CO 平均值（CO Mean）：给定日内 CO 浓度的算术平均值

CO 第一最大值（CO 1st Max Value）：给定日期的 CO 浓度的最大值

CO 第一最大值时间：（CO 1st Max Hour）：给定日期的 CO 浓度的最大值所处的时间。

CO AQI（CO2 AQI）：一天内 CO 计算的空气质量指数

【任务】

- 1、用 pandas 库读取“5. pollution_us_5city_2006_2010_CO.csv”文件，查看前五、后两行，并将缺失值全部丢弃处理。
- 2、选择 City==“New York”、Date Local、CO Mean、CO 1st Max Hour 四列导出为文本文件“pollution_us_NewYork_2006_2010_COMean.txt”，要求数据之间用空格分隔，每行末尾包含换行符。
- 3、读取文本文件“pollution_us_NewYork_2006_2010_COMean.txt”，选择 CO 1st Max Hour=“20”的所有行，以 Date Local 为横轴，以 CO Mean 为纵轴，画折线图。包括图例、图标题，x 轴刻度以年显示，y 轴显示刻度值，曲线颜色分别为红色。
- 4、选择 City==“New York”、Date Local、CO AQI 三列，按照列 CO AQI 降序排序，并将排序后结果导出为 Excel 文件“pollution_us_NewYork_2006_2010_COAQI.xlsx”。

- 5、读取 Excel 文件 “pollution_us_NewYork_2006_2010_COAQI.xlsx”，利用 `category = [0, 5, 10, 15, 20, 25, 30]`和 `labels = ['Good', 'Moderate', 'SubUnhealthy', 'Unhealthy', 'VeryUnhealthy', 'Hazardous']`将 CO AQI 进行离散化，并根据离散化结果画出饼状图，保存为 “CO_AQI_pie.png”，要求分辨率不低于 300dpi。

【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。