

## 题目九：2008\_2013 年世界各国 PM2.5 数据分析

### 【数据说明】

2008\_2013 年世界各国家和地区的年平均 PM2.5、PM10 数据。

字段说明：

- 1) Region: 地区
- 2) Subregion: 次级地区
- 3) Country: 国家
- 4) City/station: 城市、站
- 5) PM10: PM10 年平均浓度 ( $\mu\text{g}/\text{m}^3$ )
- 6) PM10 Year: PM 10 记录年份
- 7) pm2.5: PM2.5 年平均浓度 ( $\mu\text{g}/\text{m}^3$ )
- 8) PM 2.5 Year: PM 2.5 记录年份

### 【任务】

- 1、用 pandas 库读取 “9-10. world\_pm25\_pm10.csv” 文件，查看前五、后两行，缺失值丢弃处理。
- 2、选择 Region、Country、City/station、PM 2.5、PM2.5 Year 共 5 列数据导出到文本文件 “world\_pm25.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 “world\_pm25.txt”，按照列 PM 2.5 降序排序，并将排序后结果导出为 CSV 文件 “world\_pm25\_descending.csv”。
- 4、读取新的数据集 “world\_pm25\_descending.csv”，利用 `category = [0, 50, 100, 150, 200]` 和 `labels = ['One', 'Two', 'Three', 'Four']` 将 PM 2.5 进行离散化，并根据离散化结果进行直方图统计，分别画出统计直方图和饼状图，并分别将直方图和饼状图保存为 “world\_pm25\_hist.png”、“world\_pm25\_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。