

题目六：2010-2014 年北京市 PM2.5 数据分析

【数据说明】

2010-2014 年北京市每小时天气预报数据。

字段说明：

- 1) No: 行序号
- 2) year: 行数据记录年份
- 3) month: 行数据记录月份
- 4) day: 行数据记录日期
- 5) hour: 行数据记录时间
- 6) pm2.5: PM2.5 浓度 (ug/m³)
- 7) DEWP: Dew Point (单位: °C)
- 8) TEMP: 温度(单位: °C)
- 9) PRES: 压力(单位: hPa)
- 10) cbwd: 风向
- 11) lws: 风力(单位: m/s)
- 12) Is: 累计下雪时长
- 13) Ir: 累计下雨时长

【任务】

- 1、用 pandas 库读取 “PRSA_data_2010.1.1-2014.12.31.csv” 文件，查看前三行、后两行。
- 2、用 pandas 数据预处理模块将缺失值丢弃处理，删除列 DEWP、TEMP、PRES、cbwd、lws、Is、Ir，并将剩余列导出到新的 csv 文件 “pm25_data_2010.1.1-2014.12.31.csv”。
- 3、利用 pandas 库重新读取新的数据集 “pm25_data_2010.1.1-2014.12.31.csv”，并选择字段 pm2.5 大于 300 的所有数据集，导出为文本文件 “pm25_hazardous_data_2010.1.1-2014.12.31.txt”，要求数据之间用逗号分隔，每行末尾包含换行符。
- 4、读取文本文件 “pm25_hazardous_data_2010.1.1-2014.12.31.txt” 并转存到 Excel 文件 pm25_hazardous_data_2010.1.1-2014.12.31.xlsx 中。
- 5、重新读取文本文件读取文本文件 “pm25_hazardous_data_2010.1.1-2014.12.31.txt”，分别统计出现最多的 month、day、hour，并将 month、day、hour 的出现频次用直方图显示。要求包括图例、图标题，x y 轴均显示刻度值，直方图填充颜色分别为红色、绿色、蓝色；并将结果保存为 png

图片保存，分辨率为 400dpi，png 图片命名分别为

“pm25_hazardous_month_day_hour.png”。

【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。