数据预处理和可视化作业-2

处理 HUMI , PRES , TEMP 数据

1. 读入 csv 文件
2. 提取出 HUMI , PRES , TEMP 这三列数据, 并计算均值和标准差
3. 对这三列的数据进行处理:
    i. 首先判断值是否大于平均值加上三倍标准差, 如果大于就替换为平均值加上三倍标准差
    ii. 如果数据为 NaN, 则使用线性插值方法 np.interp 将缺失值替换为前一个有效值和后一个有效值的线性插值结果
4. 将结果写入到 humi_pres_temp.csv 文件中

修改 PM 值

1. 读入 csv 文件
2. 提取出 PM_Dongsi , PM_Dongsihuan , PM_Nongzhanguan , PM_US Post 这几列 (因为后面还需要对 PM 值进行操作, 所以把 PM_US Post 的数据也处理一下)
3. 判断如果大于 500 则设为 500, 如果为 NaN 则进行线性插值处理
4. 将结果写入到 pm.csv 文件中

修改 cbwd 列

1. 读取 csv 文件
2. 将 cbwd 列中值为 cv 的数据用它后面一行的数据进行替换
3. 将结果写入到 cbwd.csv 文件中

对 DEWP 和 TEMP 进行归一化处理

1. 读取 csv 文件
2. 选取 DEWP 和 TEMP 这两列数据, 分别进行 0-1 归一化 和 Z-Score 归一化
3. 使用归一化后的数据和原始数据绘制散点图
4. 将结果写入到 scatter.png 中

将空气质量进行离散化, 并分级计算天数

1. 读入第二题处理后的数据 `pm.csv`
2. 根据日期分组, 使用 `PM_Dongsi`， `PM_Dongsihuan`， `PM_Nongzhanguan`， `PM_US Post` 这四列数据计算每天 PM 平均值, 并新建一个 `AirQuality` 列用于存放数据
3. 将第 2 步得到的平均值按照 `[0, 50, 100, 150, 200, 300, float('inf')]` 分为 `['优', '良', '轻度污染', '中度污染', '重度污染', '严重污染']` 六个等级, 并存到新的 `AQI` 列中
4. 将目前得到的数据写入到 `daily_pm.csv` 文件中
5. 使用第 4 步得到的数据计算每个污染级别的天数
6. 将最终得到的数据写入到 `counts.csv` 文件中