项目

1. 数据清洗

数据预览：beijing\_17\_18\_aq.csv

|  |  |  |
| --- | --- | --- |
| 字段 | 数据类型 | 示例 |
| stationId | object | aotizhongxin\_aq |
| utc\_time | object | 2017-01-01 14:00:00 |
| PM2.5 | float64 | 453.0 |
| PM10 | float64 | 467.0 |
| NO2 | float64 | 156.0 |
| CO | float64 | 7.2 |
| O3 | float64 | 3.0 |
| SO2 | float64 | 9.0 |

缺失值统计：

缺失值数量 缺失值比例%

stationId 0 0.000000

utc\_time 0 0.000000

PM2.5 20389 6.555738

PM10 83263 26.771808

NO2 18651 5.996913

CO 42813 13.765795

O3 20421 6.566027

SO2 18548 5.963795

用时:0.02849102020263672

删除缺失值大于3的row：

缺失值数量 缺失值比例%

stationId 0 0.000000

utc\_time 0 0.000000

PM2.5 4399 1.491085

PM10 67273 22.802861

NO2 2661 0.901973

CO 26823 9.091926

O3 4431 1.501932

SO2 2558 0.867060

用时:0.06281137466430664

线性插值24小时以内的空值：

缺失值数量 缺失值比例%

stationId 0 0.000000

utc\_time 0 0.000000

PM2.5 457 0.154905

PM10 8342 2.827605

NO2 312 0.105756

CO 16957 5.747746

O3 441 0.149481

SO2 447 0.151515

用时:0.08076238632202148

根据最近的4个近邻加权填补 长缺失值：

填充长缺失

缺失值数量 缺失值比例%

stationId 0 0.000000

utc\_time 0 0.000000

PM2.5 0 0.000000

PM10 266 0.090163

NO2 0 0.000000

CO 16191 5.488103

O3 0 0.000000

SO2 0 0.000000

用时:1814.1346621513367

Final clean：

用时0.041471004486083984

1. 数据分析
2. 预测结果及展示