**二、居民异常用电预测（30分）**

**1.题干**

用电量预测对于电网的供电调度和安全运行有着重要意义。针对某城镇居民用户的电量采集和分析挖掘，发现异常用户用电模式与一般用户整体用电模式存在一定的偏差，且具备了一定的时间性的规律和相关性。

针对某城镇用电居民2018年1年（365天）的日用电量数据city\_electricity\_data.csv (单位：度），请编写python程序（保存为city\_electricity\_analysis.py）对表中数据进行分析挖掘，完成以下题目：

1. 对于日用电量数据表city\_electricity\_data.csv， 截取2018/1/1日至2018/12/31 的365列数据，并剔除365天的日电量均为0的样本行，保存到city\_electricity\_0.csv 。求出该城镇每天的用电总量，保存到city\_electricity\_sum.csv 。（10分）

city\_electricity\_sum.csv格式为：

|  |  |  |
| --- | --- | --- |
|  | 日期 | 日用电总量 |
| 0 | 2018/1/1 |  |
| …… | …… |  |
| 364 | 2018/12/31 |  |

1. 利用（1）中计算得到的该城镇的日用电总量，绘制折线图，绘图要求如下：横轴标注date，刻度以30天为间隔显示，纵轴标注为electricity，标题为trend ，将图形保存为city\_electricity\_trend.jpg。（8分）
2. 为研究前六个月的电量是否呈线性下降，现取（1）中计算得到的该城镇的日用电总量的前181天的数据（2018/1/1 至2018/6/30），为缩放数据，以千度电为单位来进行分析，采用线性回归的方式，拟合时间和用电量的线性关系。

将日期转化为序号作为自变量（序号从1开始，181截止），将缩放后的日用电总量作为因变量，先研究两者的相关系数。再用train\_test\_split方法探索线性回归方程（将集合分为预测集和训练集，其中预测集占**30%**，**随机种子数为2**，构建一个线性模型，用训练集训练模型后，用预测集进行检验），求出均方差、均方差根。**相关系数、均方差、均方差根的值**写入city\_electricity\_result.csv ，**按顺序**每行对应一个值(**保留两位小数**）。（12分）

**2.知识点**

数据预处理、数据可视化（折线图）、数据挖掘（线性回归）、结构化文件输出

**3.评判方法**

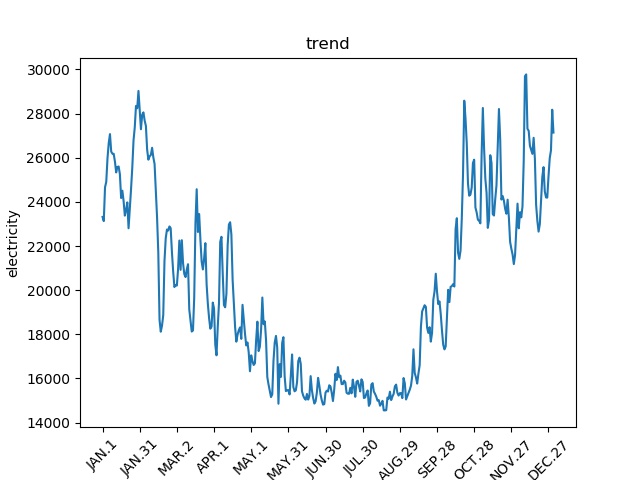
使用人工查看结果的方法，通过和标准答案比对进行判题，题（1）分值10分，结题过程有程序，逻辑完整清晰，代码能正常运行，得2分程序分。数据列截取正确得2分，全0行剔除正确得2分（city\_electricity\_0.csv 文件加表头共3277行），求和并保存文件正确得4分。

题（2）分值8分，结题过程有程序，逻辑完整清晰，代码能正常运行，得2分程序分。有X轴标注给2分，有Y轴标注给2分，有标题给2分，有保存的图片文件给2分。

题（3）分值12分。结题过程有程序，逻辑完整清晰，代码能正常运行，得6分程序分。相关性系数，均方差、均方差根结果正确每个2分。

1. **参考答案**
2. 

**2）**

****

**3）**

|  |  |  |
| --- | --- | --- |
| 相关性系数 | 均方差 | 均方差根 |
| -0.89 | 2.62 | 1.62 |

-0.89

2.62

1.62



参考程序

