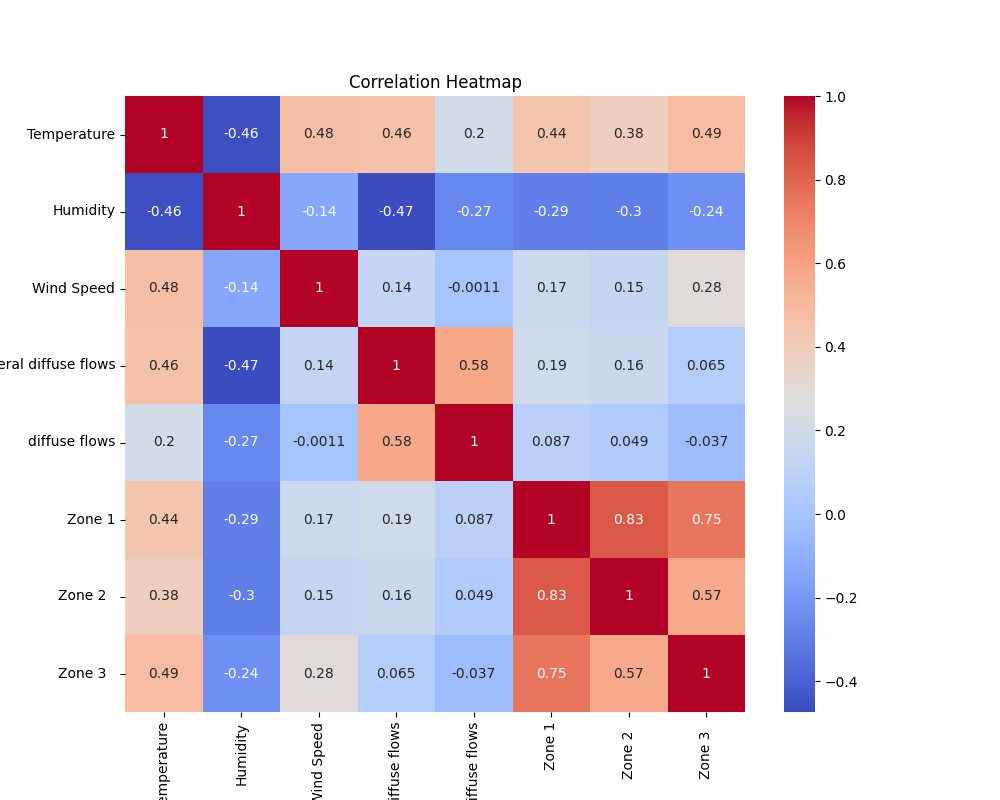
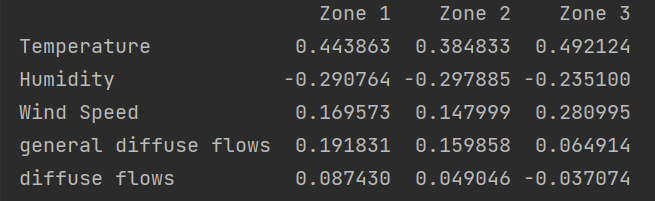
## 首先是图一：



这张代表的是，每个因素和各因素之间的相关性，你需要看的数据为：

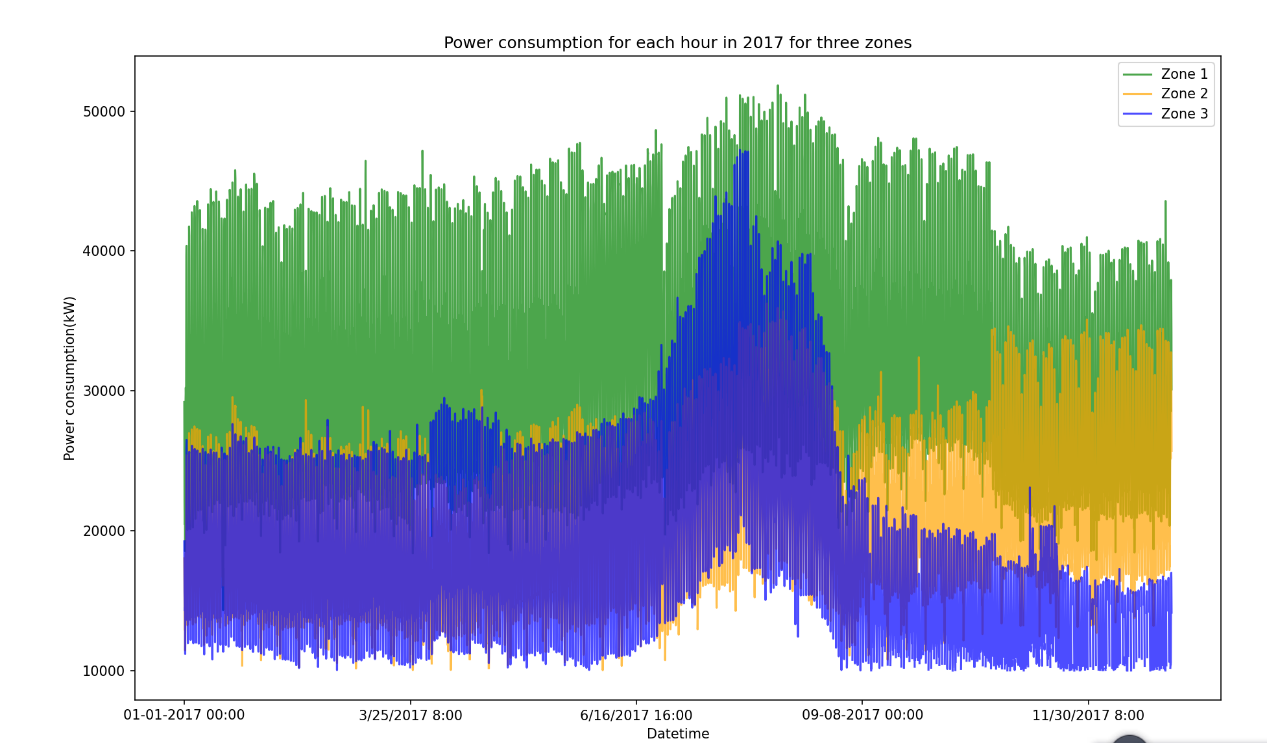


展示的是不同环境因素（Temperature、Humidity 等）与各区域（Zone 1、Zone 2、Zone 3 ）之间的相关性系数，代表的含义如下：

* **数值正负**：正值表明环境因素与区域数值呈正相关，即环境因素数值增加，对应区域数值倾向于增加；负值意味着负相关，即环境因素数值增加时，对应区域数值倾向于减少。
* **数值大小**：绝对值越接近 1，相关性越强；越接近 0，相关性越弱。

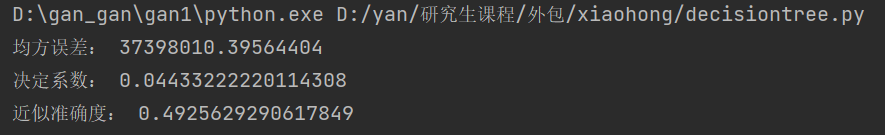
**结论：**Temperature 与 Zone 3 的相关性系数为 0.492124 ，有一定正相关；而 diffuse flows 与 Zone 3 的相关性系数为 -0.037074 ，相关性很弱。类似于这种

## 图二

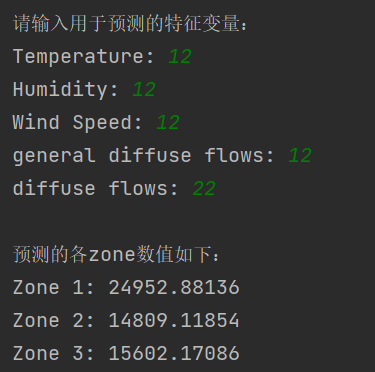


这张图是展示数据，将三个区域的电量进行可视化展示，底下横轴，是对应的时间，数轴是电量值，不同颜色就是不同的区域

## 准确率：



* **均方误差（37398010.39564404）**：均方误差是通过计算预测值与真实值之差的平方的平均值得到的。这个值越大，说明模型预测值与真实值之间的平均差异越大，模型的预测效果越差。这里均方误差数值这么高，表明该决策树回归模型在测试集上的预测值与真实值偏离程度较大 。
* **决定系数（0.04433222220114308）**：也叫*R*2 分数，取值范围在 0 到 1 之间。1 表示模型能够完美拟合数据，所有观测点都在回归线上；0 表示模型的预测效果和随机猜测差不多。此例中*R*2 分数接近 0，说明该模型对数据的拟合程度很低，大部分数据的变异不能被模型所解释，模型的预测能力较弱。
* **近似准确度（0.4925629290617849）**：因为这种预测，不是标签预测，所以没办法直接判断对错，我们是按照一定误差阈值判断预测正确样本数后计算得到的比例。该值接近 0.5，意味着大约有一半的预测结果在设定的误差范围内，说明模型的预测准确性处于中等偏下水平，有较大的改进空间。



这个单纯就是为了让你毕设答辩，有互动过程，加的现场演示代码，输入的参数就是数据集内容，输出就是我们预测的zone1，zone2，zone3的值，展示我们确实能做到预测功能。

