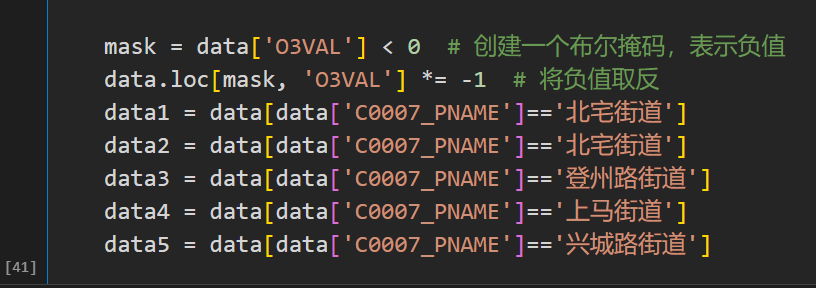
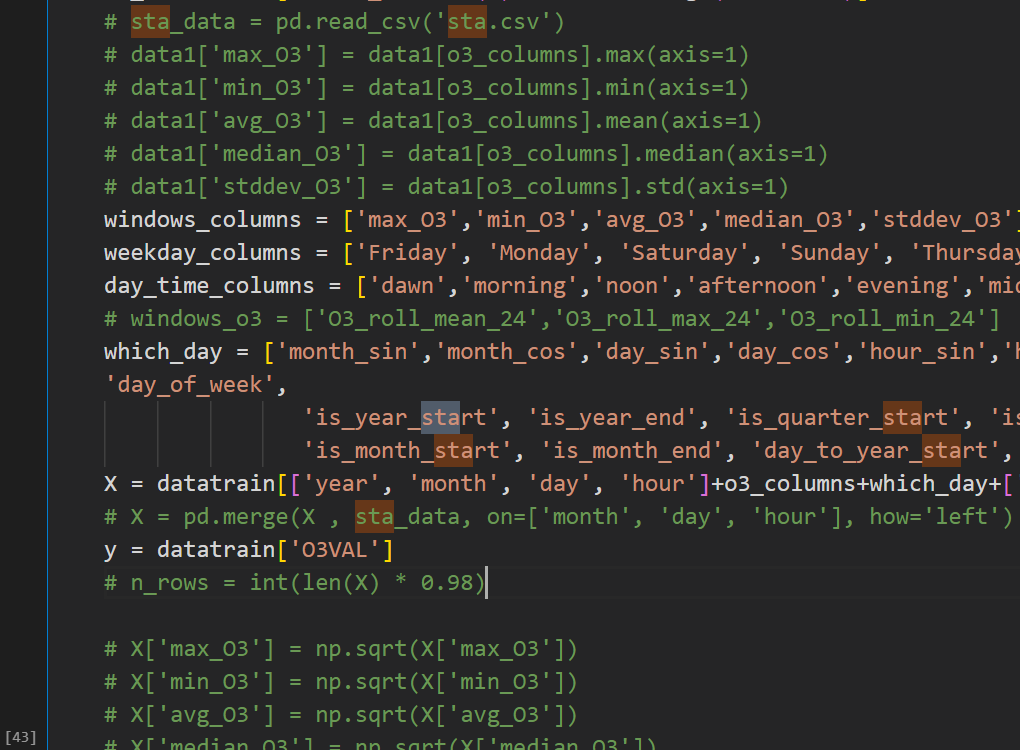
# 训练：

**1. 要换不同街道训练的话，把data1改成对应街道名称即可，后面的data1保持不变。因为只给了八大关的训练代码，所以从八大关训练的基础上更改。**

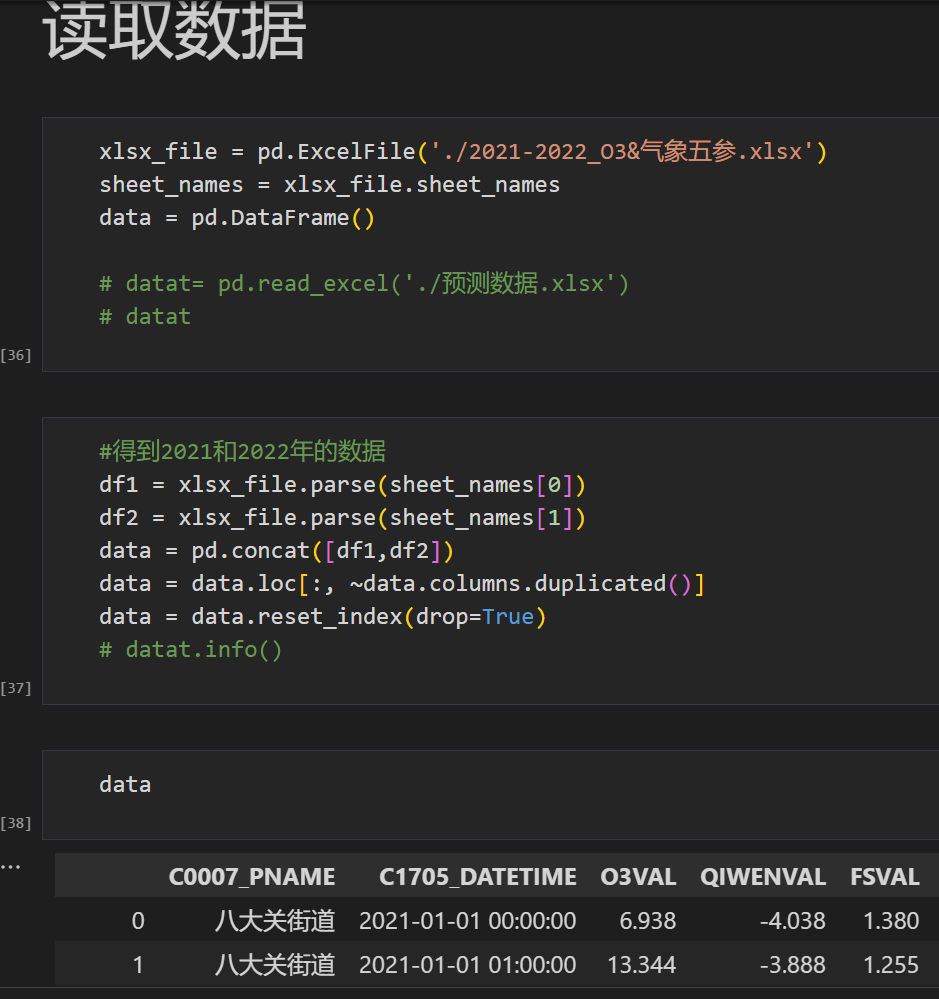


**2. 在数据集划分中把sta\_data相关的两行代码去掉。（学长说这部分没用）**



**3. 最后一个block的代码会出现重复特征的报错，做出两处更改：**

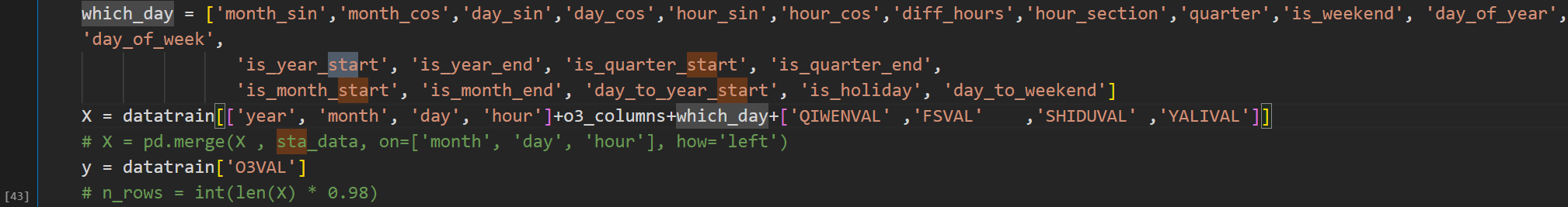
a)在第三个block中两个data之间加入 data = data.loc[:, ~data.columns.duplicated()]（来自GPT）

****

b)在数据集划分的代码中，把：

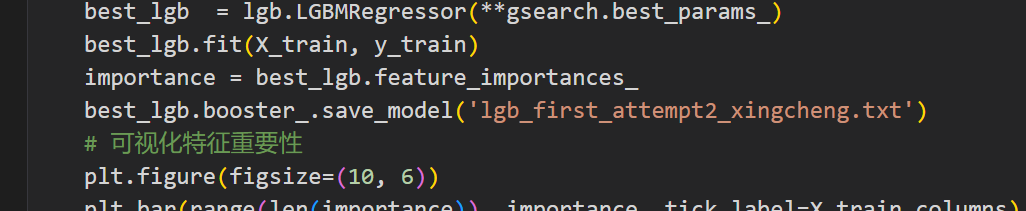
X = datatrain[['year', 'month', 'day', 'hour']+o3\_columns+which\_day+['QIWENVAL' ,'FSVAL' ,'SHIDUVAL' ,'YALIVAL']]

中的which\_day去掉。



八大关、北宅、登州路在进行a)后就没再报错，但上马和兴城路还要进行b)。

**4. 为了区分不同训练出的模型，在训练部分把输出名改一改。**



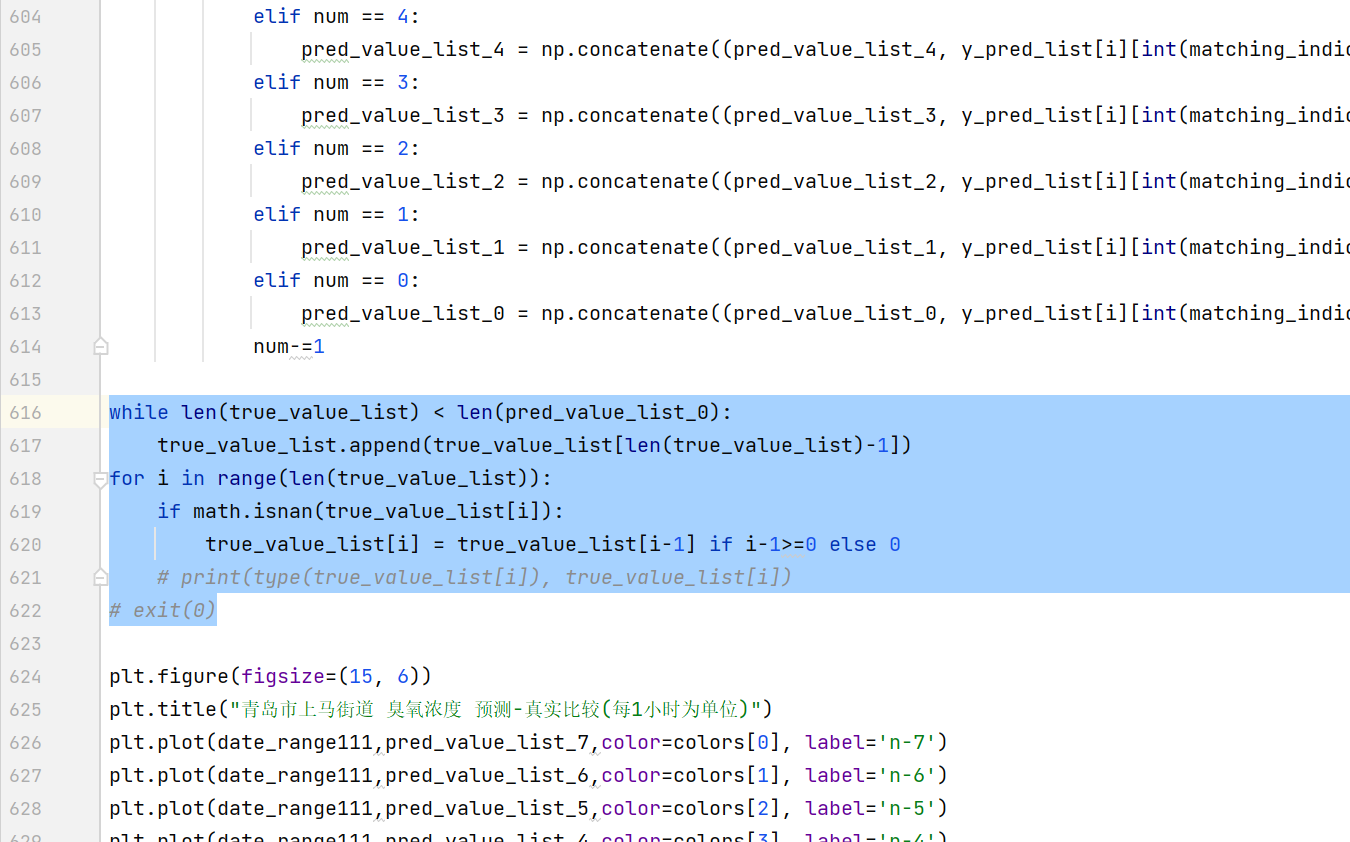
# 预测：

预测部分给了五个单独的代码，分别在所给代码基础上修改。

**1. 删了sta\_data相关的两行代码。**

**2. true\_value\_list和pred\_value\_listX的长度不匹配，加入以下代码调整true\_value\_list的长度：**

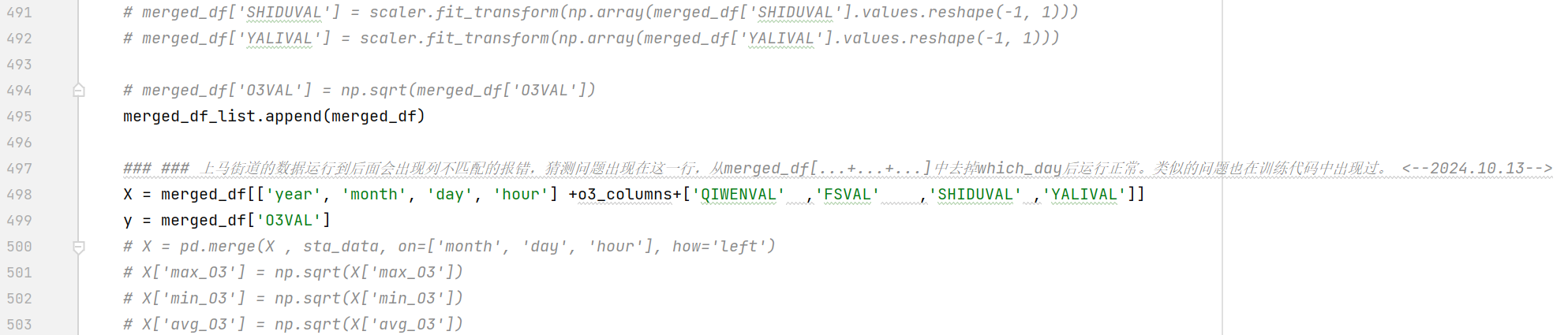
while len(true\_value\_list) < len(pred\_value\_list\_0):  
 true\_value\_list.append(true\_value\_list[len(true\_value\_list)-1])  
for i in range(len(true\_value\_list)):  
 if math.isnan(true\_value\_list[i]):  
 true\_value\_list[i] = true\_value\_list[i-1] if i-1>=0 else 0  
 *# print(type(true\_value\_list[i]), true\_value\_list[i])  
# exit(0)*

****

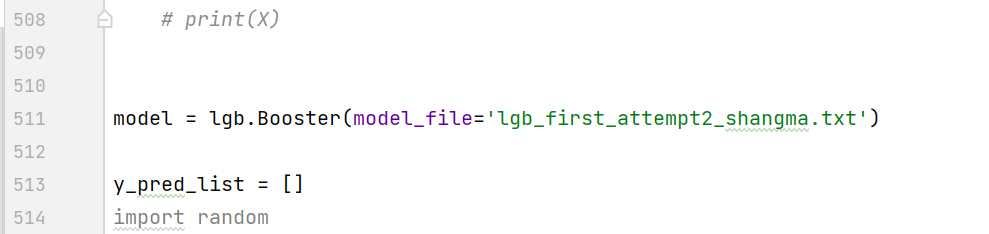
**3. 上马和兴城路会报错列数（特征数）不匹配，和训练部分类似**，把

X = merged\_df[['year', 'month', 'day', 'hour'] +o3\_columns+['QIWENVAL' ,'FSVAL' ,'SHIDUVAL' ,'YALIVAL']]

中的which\_day去掉。



**4. 使用与数据匹配的模型。把模型改成对应街道的。**



**\*\*\*输出预测数据**

arr = [true\_value\_list, pred\_value\_list\_0, pred\_value\_list\_1, pred\_value\_list\_2, pred\_value\_list\_3, pred\_value\_list\_4, pred\_value\_list\_5, pred\_value\_list\_6, pred\_value\_list\_7]  
for cnt,lst in zip(range(len(arr)), arr):  
 f = open(f"tmp{cnt-1}" + ".txt", "w", encoding="utf-8")  
 for i in lst:  
 *# f.write("\""+str(i)+"\""+",\n")* f.write(str(i) + ",\n")  
 f.close()

