1. 数值型与数值型 分布依赖图

类别型和数值型 箱图(看的是中间那条众线)

类别型和类别型 条形图

要注意分开来考虑

1. 共线变量的问题，要注意多个变量的这个问题
2. 一定要想清楚这个项目的架构，这个很重要
3. # 注意X[:, i]和X[:, i:i+1]的区别
4. 要预留出基线缓存，不要一下子就替换所有，为了避免不要重头再来
5. 把训练集和测试集合在一起操作，这样会好很多的
6. 在最初一定要把变量的类型弄好，弄成字符串或者时间类型，不要让他搞一些object出来
7. 准备一个基础的模型用来分析特征添加是否有效，然后用一个基础的特征来分析模型的是否有效
8. 必须要一个个模型去做一遍，再stacking才有意义，直接跑stacking也没用
9. 记录 每天完成的工作是什么，吾日三省吾身这很重要

预处理：

1. 一个变量一个变量拿过来看相关性分析
2. 然后对多类别进行筛选，尽量得缩减类别数量，尽量的使稀疏数据变密集，进行统计分组
3. 一直忽略了归一化哎。。。
4. 画图不如看数据，因为要很明显图才可以看出来