1. 数据准备

本实验使用的数据集来自[https://www.kaggle.com/c/predict-west-nile-virus/data]，其中包含train.csv,test.csv,weather.csv,我们先将train和weather结合在一起，然后利用一部分数据训练学习模型，另一部分数据用来检测。

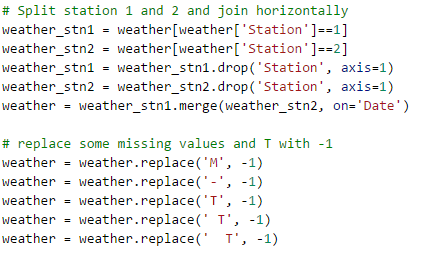
首先导入train.csv和weather.csv



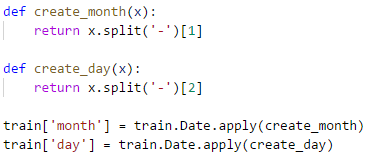
train数据集中WnvPresent列表示病毒是否出现，值为0或1，将这一列提取出来作为训练的标签。



整理weather数据集



整理train数据集，将Date分为month和day



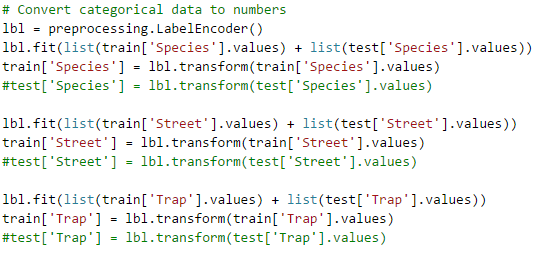
将latitude和longitude转换成整数



最后将train和weather合并



将标称数据转换成数值类型



去掉带有-1的列

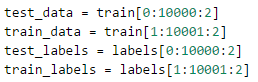


删除缺失值



1. 分类

数据集中的奇数项作为训练数据，偶数项作为测试数据



首先使用K近邻分类器分类



score为分类器爱测试数据集上的得分



然后使用SVC分类器



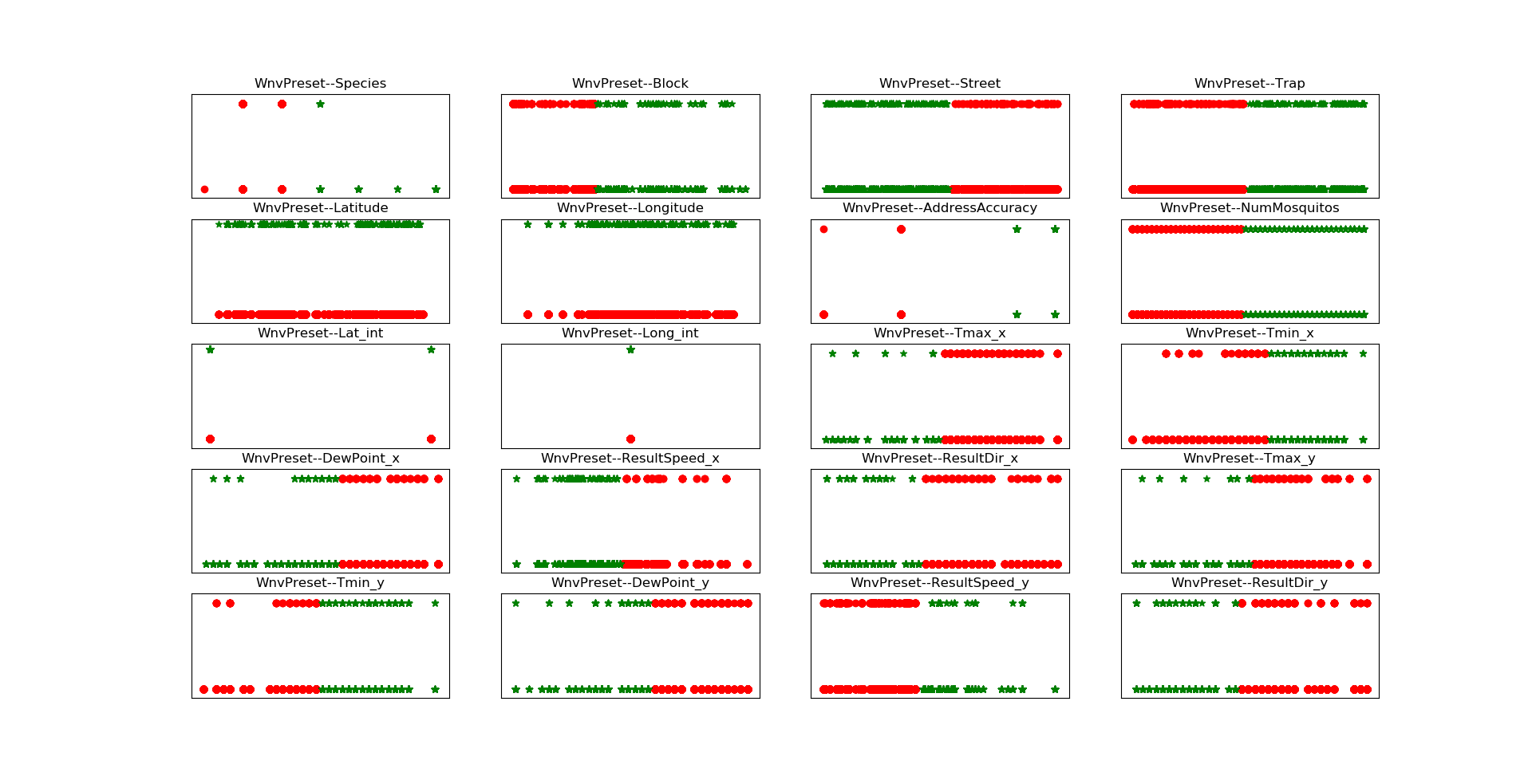


1. 聚类

为了便于可视化，聚类部分使用两列进行聚类，首先使用Kmeans算法进行聚类



聚类的结果如下：



使用MeansShift算法聚类



结果如下

