Report

1.

(a)

第k,p列数据完全一样

第5行数据完全一样

(b)

将这些列删除

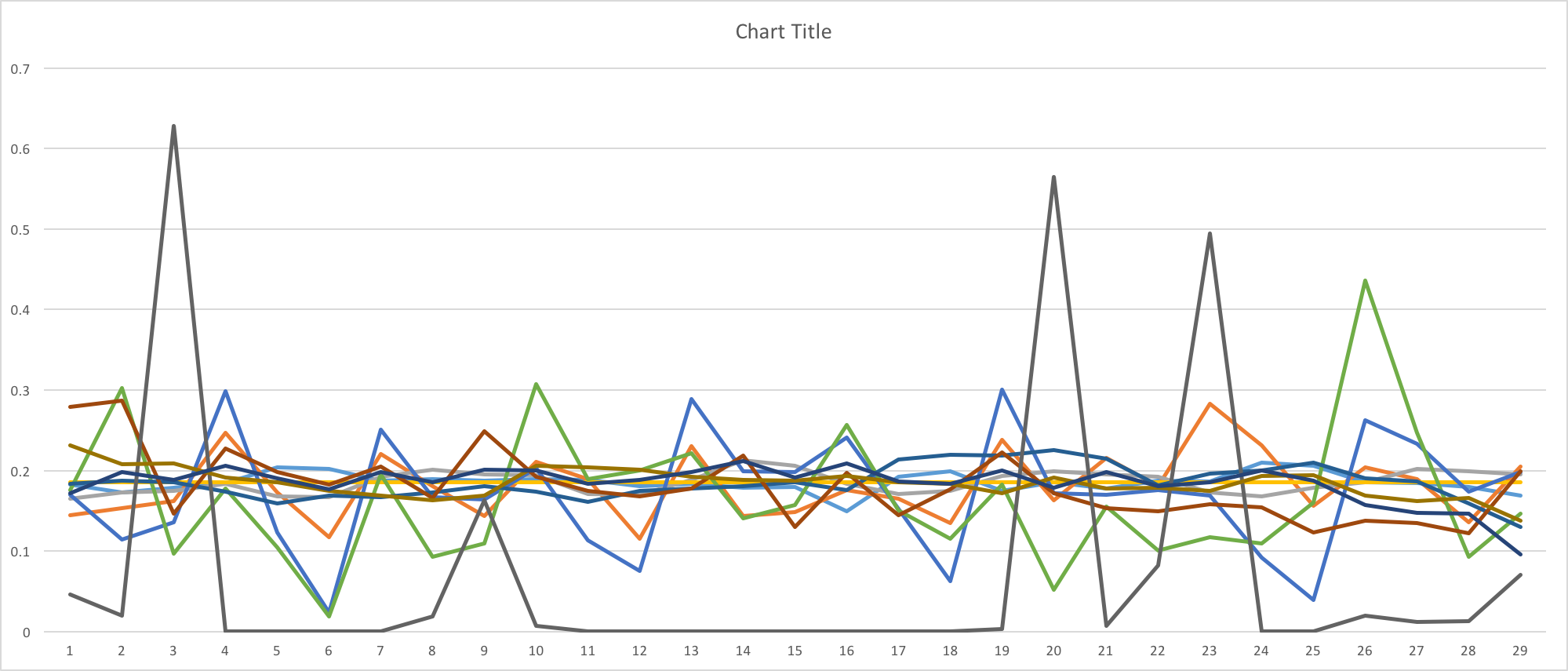
2.

观察折线图可以发现第9个人数据有突变，可能为异常

3.

(a)

i. 先观察折线图，根据第6天(周末)流量是否显著下降大致估计属于那种类型



ii. 编程先normalize所有数据，使它们scale一样，再使用KMeans 进行聚类

(b)

使用GBDT regression model, 用大小为5的滑动窗口依次遍历data, 用这5天数据predict第7天数据,后续可以使用其他model进行emsemble,滑动窗口大小也可条件，窗口内数据权重可以调节，RNN或LSTM模型因为数据量不够可能效果不好