

## **!** 纽约出租车简介





纽约出租车协会提供了两百万趟出租车行车记录

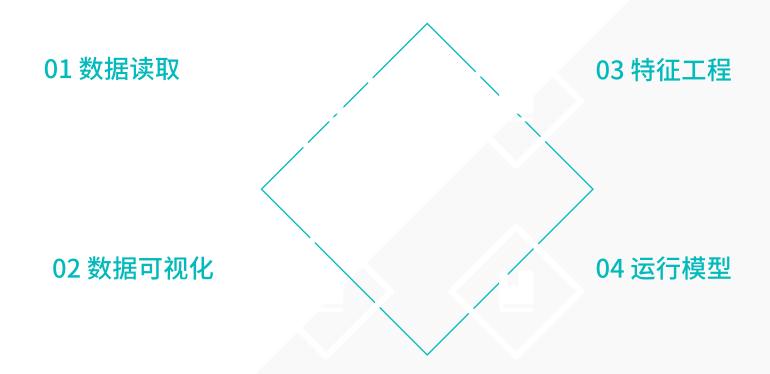


根据载客情况,准确回归模型预测每一单出车用时



预测纽约出租车用时,有助于在行程开始前对车费的预测

## 课程安排



#### 读取数据



#### **Pandas**

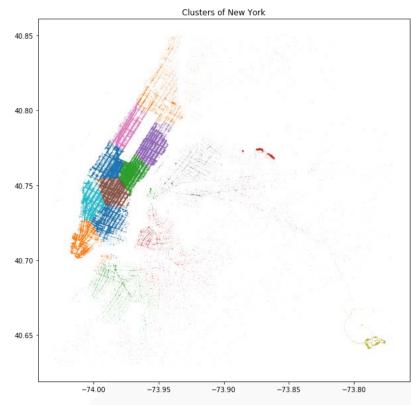
- 读取训练集和测试集
- 读取行程数据
- 读取节假日数据
- 读取天气数据



## 数据可视化



#### **Kmeans Clustering+ matplotlib**



## **特征工程**



#### **Scikit Learn**

- 时间特征
- 距离特征
- 区域拥挤特征



## ■运行模型



#### **XGB**

- One-Hot-Enconde 类别型数据
- 区分训练集,验证集和测试集
- 设置参数,评估得分
- 写出结果,上传提交



#### Reference

https://www.kaggle.com/c/nyc-taxi-trip-duration

https://www.kaggle.com/drgilermo/dynamics-of-new-york-city-animation

 $\underline{https://www.kaggle.com/karelrv/nyct-from-a-to-z-with-xgboost-tutorial}$ 

https://www.kaggle.com/gaborfodor/from-eda-to-the-top-lb-0-367

 $\underline{https://www.kaggle.com/maheshdadhich/strength-of-visualization-python-visuals-tutorial}\\$ 



# THANK YOU