

Traffic Analysis In New York City

Better city, better life.

Group 11

Yizhen Zhao

Yu Wang

Kai Zhang

Task background

1. NYC 10 years taxi rides records

From 2009 to 2019, records of all taxi rides in New York City were made public.

2. 20GB dataset per year

The 20 GB dataset (per year) includes time and place of passenger pickup and drop-off, trip time and distance, and information about how much was paid for the trip.



Task background

Find long-term trends

How NYC traffic and the habit of passengers using taxis has been evolving these years.

Find long-term changes

What factors have influenced the traffic over these years

Suggestion for NYC traffic

Give reasonable explanation of changes and suggestions for NYC traffic.



Early findings

Data Source	Data Characteristic	Data Quality & Data Noise	Supplement
Yellow taxi	<ol style="list-style-type: none">1. 2009 – 20192. Pick up in all places in NYC	<ol style="list-style-type: none">1. High2. Pick up/Drop off location may have noise	
Green taxi	<ol style="list-style-type: none">1. 2013 – 2019 (Four years less than Y taxi)2. Pick up in strict places in NYC	<ol style="list-style-type: none">1. High2. Pick up/Drop off location may have noise	
FHV	<ol style="list-style-type: none">1. 2015 – 2019 (Six years less than Y taxi)2. User get trip from internet service	<ol style="list-style-type: none">1. Medium2. Share-ride flag may have noise3. PUlocationID may have noise4. DOlocationID may have noise (But won't use)	
High Volume FHV	<ol style="list-style-type: none">1. 2019 (Only one year data)2. User get trip from internet service (Trip from company)	<ol style="list-style-type: none">1. Low2. Noise same as FHV	<ol style="list-style-type: none">1. Nearly same service as FHV so will combine together with FHV

Data Product

Data Product	Target	Additional Data Source
Taxi trend	1. How does green taxi influence NYC's traffic	
Holiday influence	1. Traffic situation may be abnormal during holidays	1. Public holidays of NYC
Weather influence	1. Bad weathers like rain, storm or snow may influence traffic	1. Historical weather by hours
Traffic circumstance	1. What is NYC traffic circumstance of different locations 2. What is the trend of NYC traffic situation	1. Population 2. Postcode-to-Taxi-Area list
Rush location	1. There may be some areas which taxi use time are much higher than other area. We will find out these areas' properties and try to relief it.	
Rush hours estimation	1. To find if real rush hour aligned to estimated rush hour	
FHV service trend (characteristic)	1. Trends of FHV service changes in 10 years	

Technologies plan to use

Spark

Deal with raw data
of taxi

Python crawler

Get the weather
information of past
10 years

Data Brick

Spark service &
data repository

AWS S3

A cloud server as a
platform to deal with
all tasks

Mapbox

1. Plan route for trips
2. Map data visualization

Tableau

Table data visualization

IBM SPSS

Deal with reliability, validity,
correlation, relevance, and
normal distribution

SQL

Store result data for
demonstration website

