E-ABOVE

NYC-TAXI

SmartInvest

E-Above Report

Mohammed Arebi IBS5

Agenda

Part 1

- Data Collection and Sampling
- Problem/Questions
- Overview of Results

Part 2

- Data Cleaning
- Analysis Discussion
- Pickup Prediction Model

Data Collection

E-ABOVE
NYC-TAXI

- Data collected from this link: http://www.andresmh.com/nyctaxitrips/
- 2013 data was used for the analysis
- Data contains information regarding taxi trips
- The information includes trip duration, distance, amount, payment type, time and location data

Data Sampling

E-ABOVE
NYC-TAXI

- Reservoir random sampling was used to sample 150K records/trips per month of 2013

- Fare data (Amount information) and trip data (time, duration, etc. information) were merged together for complete trip information

- Sampling resulted in around 2 million records

Problem and Questions

E-ABOVE
NYC-TAXI

- Given the data we tried to analyse the patterns and trends inherent in the dataset.
- The insights generated from the analysis will assist the business and decision makers in decision making and it will also enable the optimisation of operations
- Patterns analysed include:
  - Effect and relationship of trip distance on total amount
  - Effect of location-specific and time-specific data on the number of pickups (pickup density) and traffic volume
  - Effect of location-specific and time-specific data on the average speed
  - Effect of location-specific and time-specific data on the tipping behaviour
  - Others...

Overview of Results

E-ABOVE
NYC-TAXI

- Total Amount is highly positively correlated with the trip distance (0.94) and trip duration (0.83)
- Conservative restrictions on trip distance do not significantly affect the total amount more severe restrictions do significantly affect the amount
- More taxi trips were starting in the South of New York compared to the North
- Average overall taxi speed in New York in 2013 was 13.66 miles per hour or 21.99 km per hour
- Highest average speeds are between 12AM and 6AM while the lowest are between 8AM and 4PM
- Highest average speeds are on Monday and the weekend, with Friday being the lowest
- Trips from the Queens to Staten Island have the fastest trips while trips from Staten Island to Manhattan have the slowest
- 8AM to 10AM and 6PM to 10PM are the busiest hours in terms of taxi traffic i..e most trips are in the morning and evening. With exceptions discussed later
- Manhattan is the busiest with a large margin followed by Queens and Brooklyn

Overview of Results

E-ABOVE
NYC-TAXI

- Fridays are the busiest day followed closely by the weekend
- Traffic volume per month? Uniformly Distributed due to sampling, so not significant
- Tip Amount is slightly positively correlated with trip distance and duration
- Staten Island is the highest tipping borough (Queens Without Outliers)
- 5AM is the highest tipping hour (Late Night and Night as well). and 11AM to 1PM being the lowest tipping hours (not a significant difference, discussed later)

Data Cleaning

Location

Data was restricted to trips only within New York and outliers removed as data contained erronous values

Trip Duration

Outlier and invalid trip durations were corrected and removed using the recorded trips time. Slide

Trip Distance

Outlier and invalid trip distances were corrected and removed

Trip Speed

After computing the trip speed from distance and duration outliers were removed (found speed of 7000 miles/hr !!)

Trip Total Amount

Outlier and invalid trip amounts were corrected and removed

§58-20 Operations – Responsibilities with Respect to Drivers

(a) *Hours of Operation. Maximum Driving Hours.* An Owner must not require a Driver to operate one or more Taxicabs for more than 12 consecutive hours.

§58-20(a)	Fine: $50	Appearance NOT REQUIRED

Data Cleaning (Continued)

Passsenger Count

Trips with no passengers were replaced with the mean passenger count (Only 7 Records)

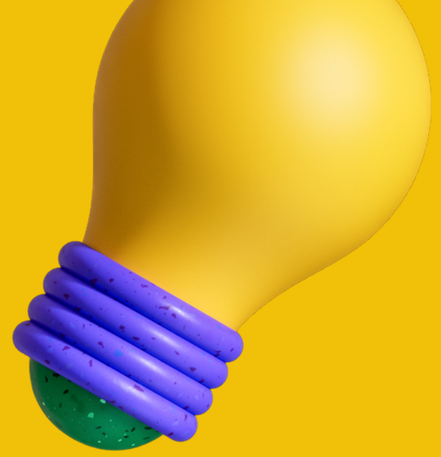
Rate Code

Rate code is between 1-6, found records with rate ocode of 0 (33 to be exact). They were replaced with Standard Rate Code

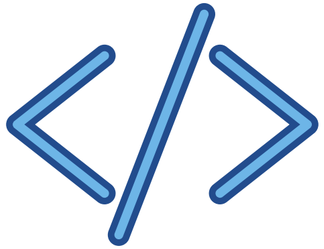
Trips in Water

Some trips were recorded in water, these were removed

Analysis Discussion



E-ABOVE
NYC-TAXI



Let's move to the other presentation for visuals and discussion in Part 2

Thanks for listening

Any Questions?

Address : Habsburgerstr 10, Freiburg

Phone : +49 15775654385

Email Address : mohammed.fateh.arebi@hs-furtwangen.de