

Strategy to Increase New York Taxi Drivers' Income

Yue Ren, Tianyu Wang

Problem Statement:

The taxi is one of most important transportation in New York City and the tip can decide the income of the taxi driver. There are several factors that can affect the amount of the tip and we want to find the pattern between these factors and tips to know how the taxi driver can increase his income in our project. There are so many existing datasets, such as New York City Taxi Trips [1] indicating the history record of each trip, and trip records under hourly weather data [2].

In this project, we will explore some datasets, such as New York City Taxi Trip - Distance Matrix [4], Weather data in New York City [5], New York City Bus Data [6], and Uber Pickups in New York City [3]. Our aim is to try to find a better way for taxi drivers to gain more income during their regular working time, or to suggest a better strategy. For example, a long-distance trip means much more income per trip. There might be a pattern, location and time, to indicate where and when might be easier to pick up a customer with long-distance trip demand. Also, there are so many people in New York City taking public transit and Uber. How to avoid the competition from these, and find out where and when the customer might be more willing to choose taxi as their first option.

First, we will explore the relation between location, time and long-distance trip, how to be more possible have a long-distance-trip customer. Second, we will explore the relation between trip and weather, how normal and extreme might affect the trip. Third, we will explore the competition between taxi, uber, and public transit, how to avoid the competition and when taxi might be the first option for customers. Finally, we will try to find a strategy to increase income of taxi drivers, under the assumption -- working 8 hours per day.

Data Acquisition:

The data will be downloaded, loaded, processed into python. The datasets we could download are from 2016 to 2019. In New York City Taxi Trips [1], we can obtain details about taxis such as, pickup datetime, drop off datetime, passenger count, trip distance and tip amount and the data is in a CSV format that was collected from driver input and from the GPS. This data will be the main dataset that we use in our project. Moreover, the most important factor that we want to explore is weather and weather information of New York can be obtained from New York City Taxi Trip - Hourly Weather Data [2]. The dataset of weather contains details of datetime and temperature. Furthermore, the weather information in 2016 can be obtained from Weather data in New York City - 2016 [5]. We can get the information about Uber in New York City from Uber Pickups in New York City[3]. We can

obtain the location, time, schedule and more information of the bus for New York City from New York City Bus Data [6].

Feature Choices

For this program, we will use python to build. In the code, we need a comment for each function to show the purpose of this function, the mean of parameters and return variable of the function.

Algorithm choices/ideas

In the first three exploration, such as the relation between weather and trip, between location/time and trip, we might use regression to see which condition, like weather, time interval, will affect the trip more. Later, we will use a decision tree to split the work hour to find a better strategy to increase the income. Also, a linear regression of our model will be used to compute how much a taxi driver might earn under our strategy, which will be used to compare with the current taxi driver census.

Performance Evaluation

We will evaluate our performance by test dataset and we will separate our data to training dataset and test dataset. Training dataset will be used to construct a model to find the relationship between several factors and income of taxi. Test dataset will be used to check the accuracy of our result. Moreover, we also evaluate our algorithm.

Main References:

1. <https://www.kaggle.com/kentonnlp/2014-new-york-city-taxi-trips>
2. <https://www.kaggle.com/meinertsen/new-york-city-taxi-trip-hourly-weather-data>
3. <https://www.kaggle.com/fivethirtyeight/uber-pickups-in-new-york-city>
4. <https://www.kaggle.com/debanjanpaul/new-york-city-taxi-trip-distance-matrix>
5. <https://www.kaggle.com/mathijs/weather-data-in-new-york-city-2016>
6. <https://www.kaggle.com/stoney71/new-york-city-transport-statistics>