Yellow Taxi Fare Prediction

Rahul Parmar[1], Murli Alva[2], Pathik Patel[3], Vatsal Suthar[4]
[1]rahul.p1@ahduni.edu.in
[2]murli.a@ahduni.edu.in
[3]pathik.p@ahduni.edu.in
[4]vatsal.s3@ahduni.edu.in

***Abstract*—Taxis are preferred as one of the favorite ride in late 19th century in new york. Mostly preople used to prefer taxis over their personal vehicle. There are two types of taxis yellow and green. They are selected according to the location. fares vary according to location and time. In this project we are going to predict taxi fare and classify the areas where most taxis are booked.**

***Keywords*—Yellow taxi, New york, Machine learning, analysis, Regression, classification**

I. Introduction

In NYC, taxicabs come in two varieties: yellow and green; they are widely recognizable symbols of the city. Taxis painted yellow (medallion taxis) are able to pick up passengers anywhere in the five boroughs. in Upper Manhattan, the Bronx, Brooklyn, Queens. The yellow taxi cab was first introduced in 1915 by a car salesman named John Hertz. Hertz decided to paint his taxis yellow because of a study by a Chicago university to establish what colour would grab the attention of passers-by more easily. The results proved that yellow with a touch of red was most noticeable. As a result, Hertz started to paint all his taxi cabs yellow and went on to start the Chicago-based Yellow Cab Company in 1915.

II. Literature Survey

Yellow taxi cabs are most famous in New York city. There are 2 types of taxi green and yellow. Over 200,000 TLC licensees complete approximately 1,000,000 trips each day. To operate for hire, drivers must first undergo a background check, have a safe driving record, and complete 24 hours of driver training. TLC-licensed vehicles are inspected for safety and emissions at TLC's Woodside Inspection Facility.

III. Implementation

We have collected data set from nyc taxi website. First of all we understood all the columns and its meaning. We researched about yellow and green taxis as there are 2 taxi service in New York. we planned to go with yellow taxi fare prediction which are used in Upper Manhattan, the Bronx, Brooklyn, Queens and stalen island.

IV. Results

We have explored the October month dataset and understood the meaning of different columns. We have merged extra amount into a single column named final amount. We have performed correlation of all the columns and got the output as a matrix and visualized different columns. We have also split date and time into different columns.

V. Conclusion

During pre processing of data there were many outliers such as there was 100 dollars fare for 0 mile trip. Than there were few outliers in ratecode id. We pre processed and removed them all and cleaned the data. After cleaning the data we visualized data in which we got different insights people like to travel single in the taxi. Area 236 has most taxi bookings. Also we observed that at mid night(1 to 6 am) people doesn't like to travel much often.

References

[1] https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page

[2] https://www.youtube.com/watch?v=lJitdOke2bsabchannel=CodingNest

[3] https://towardsdatascience.com/new-york-taxi-data-set-analysis-7f3a9ad84850

[4] https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page