Modelling Tipping Behaviour in Urban Transportation

Progress Report

Group Members

- Archanakumaree Dineshbhai Bhusara - abhusara -50560481
- Nihar Madasu - niharmad – 50559501
- Navaneethakannan Arumugam - arumuga3 – 50560498
- Yiyuan Wang – yiyuanwa – 50311193

Introduction:

Our project aims to build a comprehensive model to analyse tipping behaviour in urban transportation systems, more specifically the New York City Taxi system. By bringing in external data points like weather, demographic information and socioeconomic factors, we hope to discover patterns that influence the amount of tips an average customer pays, which could help us gain an insight into consumer behaviour in general.

Literature Review

1. Tang (2019) - Big Data Analytics of Taxi Operations in New York City:

Tang employs big data analytics to investigate taxi operations in New York City, leveraging vast datasets to uncover insights into various aspects of service provision and customer satisfaction. Methodologically, Tang's study involves the collection and analysis of large-scale transactional data from taxi services, encompassing variables such as trip duration, route optimization, and passenger feedback. This study helps our research by pulling us in the right direction in terms of variable selection and EDA.

2. Elliot et al. (2017) - Tippers and Stiffers: An Analysis of Tipping Behavior in Taxi Trips:

Elliot and colleagues conducted a nuanced analysis of tipping behaviour in taxi trips, distinguishing between different types of passengers based on their propensity to tip. The study employs a mixed-method approach, combining surveys, interviews, and observational data to capture the multifaceted nature of tipping decisions.

3. Correa (2017) - Exploring the Taxi and Uber Demands in New York City: An Empirical Analysis and Spatial Modeling:

Correa's research adopts an empirical analysis and spatial modelling approach to examine the demands for both traditional taxi services and ride-hailing platforms in New York City. It provides us with significant insight into using spatial analysis in our research as well.

4. Judd Cramer and Alan B. Krueger - Disruptive Change in the Taxi Business: The Case of Uber:

Cramer and Krueger conduct a comparative study to analyse the disruptive impact of ride-hailing platforms like Uber on the traditional taxi business. The quantitative analysis of market data to examine shifts in consumer behaviour provides us with guidance towards using another predictor in our model.

5. Riascos and Mateos (2020) - Networks and Long-range Mobility in Cities: A Study of More Than One Billion Taxi Trips in New York City:

Riascos and Mateos employ network analysis techniques to study long-range mobility patterns in urban areas using taxi trip data. This information provides us with an analysis of mobility which provides us with another possible predictor of customer tipping behaviour

6. Xie et al. (2021) - Revealing Spatiotemporal Travel Demand and Community Structure Characteristics with Taxi Trip Data: A Case Study of New York City:

Xie and colleagues conduct a case study utilising taxi trip data to reveal spatiotemporal travel demand patterns and community structure characteristics in New York City. Methodologically, the study involves the collection and analysis of detailed trip data from taxi services, encompassing variables such as trip origin-destination pairs, travel times, and passenger demographics.

7. Hu and Du (2024) - Passenger Group Size and Tipping: An Empirical Study of 50 Million NYC Yellow Taxi Rides:

Hu and Du undertake an empirical study to investigate the relationship between passenger group size and tipping behaviour using a dataset of 50 million NYC yellow taxi rides. It focuses on one of the predictors we plan to add to our model, namely the passenger count.

8. Aydin and Acun - An Investigation of Tipping Behavior as a Major Component in Service Economy: The Case of Taxi Tipping:

Aydin and Acun conduct an in-depth investigation of tipping behaviour as a major component of the service economy, focusing specifically on taxi tipping. The study employs a mixed-method approach, combining surveys, experiments, and observational data to explore the psychological, social, and economic factors shaping tipping decisions.

9. Tan and Zhang (2021) - Good Days, Bad Days: Stock Market Fluctuation and Taxi Tipping Decisions:

Tan and Zhang examine the impact of stock market fluctuations on taxi tipping decisions, employing econometric techniques to analyze transactional data from taxi services. This research helps us in identifying another possible key predictor to add to our model

10. Zhan et al. (2016) - A Graph-Based Approach to Measuring the Efficiency of an Urban Taxi Service System:

Zhan and colleagues propose a graph-based approach to measuring the efficiency of an urban taxi service system, utilising network analysis techniques to assess service quality and performance. Using

efficiency as a factor to add to our prediction model is also an approach we considered after looking at this paper.

11. Lee and Sohn - A Large-scale Data-based Investigation on the Relationship between Bad Weather and Taxi Tipping:

Lee and Sohn undertake a large-scale data-based investigation to examine the relationship between adverse weather conditions and taxi tipping behaviour. The study involves the collection and analysis of transactional data from taxi services, along with meteorological data on weather conditions. This research provided us guidance in using weather data in conjunction with the dataset, in order to add it as a predictor in our model.

12. Ferreira Neto et al. - Do Tourists Tip More Than Local Consumers? Evidence from Taxi Rides in New York City:

Ferreira Neto and colleagues investigate differences in tipping behaviour between tourists and local consumers using transactional data from taxi rides in New York City. Methodologically, the study involves the segmentation of passengers based on residency status and the analysis of tipping amounts and frequencies.

Novelty:
1. We referred to multiple research studies, all of which focused on different predictors or points of analysis for customer tipping behaviour
2. Using all of these predictors, we plan to build a comprehensive model to analyse tipping behaviour
3. We hope to uncover previously unknown patterns, and also understand the relationships between the predictors themselves, if any, and how they affect the result of the overall model

Progress:
1. Collection and preprocessing of data is completed.
2. Preliminary EDA in progress. Identifying top predictors and drawing plots to understand relationships between predictors and response and within predictors.
3. Explored the use of geopandas for spatial analysis, specifically to draw a mapping between taxi zones and zip codes in order to make it easier to bring weather data and demographic information into the dataset.

Upcoming Experiments:
1. Finalising the top predictors to use for the model.
2. Combining all of our datasets into one appropriate dataset, including all of the data coming from external sources i.e. weather, stock market, etc.
3. Employing model selection techniques to pick out the best algorithm for our model.
4. Evaluating the performance of our model on appropriate test data.

Group Effort Statement:

Everyone in the group has contributed equally, with tasks ranging from data collection, data processing, exploration of outside datasets, EDA and documentation.

References:

1. Tang, Y. (2019) Big Data Analytics of Taxi Operations in New York City. *American Journal of Operations Research*, **9**, 192-199
2. Elliott, David & Tomasini, Marcello & Oliveira, Marcos & Menezes, Ronaldo. (2017). Tippers and Stiffers: an Analysis of Tipping Behavior in Taxi Trips.
3. Correa, Diego, Exploring the Taxi and Uber Demands in New York City: An Empirical Analysis and Spatial Modeling (August 1, 2017).
4. Judd Cramer, Alan B. Krueger. Disruptive Change in the Taxi Business: The Case of Uber.
5. Riascos, A.P., Mateos, J.L. Networks and long-range mobility in cities: A study of more than one billion taxi trips in New York City. *Sci Rep* 10, 4022 (2020).
6. Xie C, Yu D, Zheng X, Wang Z, Jiang Z (2021) Revealing spatiotemporal travel demand and community structure characteristics with taxi trip data: A case study of New York City. PLoS ONE 16(11): e0259694.
7. Hu, Ye and Du, Rex Yuxing, Passenger Group Size and Tipping: An Empirical Study of 50 Million NYC Yellow Taxi Rides (January 25, 2024).
8. Asli Elif Aydin, Yüksel Acun,An investigation of tipping behavior as a major component in service economy: The case of taxi tipping,Journal of Behavioral and Experimental Economics
9. Weiqiang Tan and Jian Zhang, Good Days, Bad Days: Stock Market Fluctuation and Taxi Tipping Decisions, Management Science 2021
10. X. Zhan, X. Qian and S. V. Ukkusuri, "A Graph-Based Approach to Measuring the Efficiency of an Urban Taxi Service System," in IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 9, pp. 2479-2489, Sept. 2016
11. Won Kyung Lee, So Young Sohn,A large-scale data-based investigation on the relationship between bad weather and taxi tipping,Journal of Environmental Psychology
12. Amir B. Ferreira Neto, Adam Nowak, and Amanda Ross, Do Tourists Tip More Than Local Consumers? Evidence from Taxi Rides in New York City, International Regional Science Review 2019