Take a Taxi in New York City

A study on the taxi travel pattern of NYC

Bolin Huang
Master Business Information Technology
University of Twente
Enschede, Netherlands
b.huang-2@student.utwente.nl

Yawen Li
Master Business Information Technology
University of Twente
Enschede, Netherlands
y.li-20@student.utwente.nl

Wan Barmawi Muhammad Razaq
Master Business Information Technology
University of Twente
Enschede, Netherlands
w.b.m.razaq@student.utwente.nl

Hakan Tirsi
Master Business Information Technology
University of Twente
Enschede, Netherlands
h.tirsi@student.utwente.nl

***Abstract*—In this study, we have investigated the taxi service in New York City, focussing on the difference between Green Taxi and HVFHV. Since 2013, the Green Taxis have been introduces to provide more service to the areas in New York outside Manhattan. Additionally, Uber and similar app-hail services have been active in the market for an extended period. Our research aimed to compare these two taxis, focusing on how the taxis are used for both long and short distance, and when they are preferred (morning, afternoon, evening or the night). Moreover, within our research we found significant changes to the usage of taxis after the pandemic. Our research revealed a stark reality regarding the number of rides for the Green taxis. Between 2019 and 2022, the Green Taxis experienced a decrease of more than 85% of their rides. This decline is noteworthy when compared to HVFHV, which has only saw 9.4% decrease. Despite the Green taxis have been introduces to provide more service outside Manhattan, customers continue to prefer the HVFHV. Lastly, our research indicates that customers prefer the Green Taxi for short-distance travel, while HVFHV is preferred for short to medium distances. You can access our code in this repositary.**

***Index Terms*—Green Taxi, HVFHV, Travel pattern, Spark**

I. Introduction

Taxi is one of the most important transportation alternatives in metropolises. Compared to public transportation, it provides a more customized service, more convenience and thus is more costly. Meanwhile, considering the high utilization rate of taxi in urban commuting, the analysis of taxi ride data can yield insightful deductions on the patterns of human mobility.[1] Previous research has identified that the human mobility in a street network should comply with the distribution pattern Levy Flight, which is widely observed in reality and applied in the simulation of random movements of animals and humans as a supplement method to Brownian motion.[2][3] Regarding the urban human mobility pattern, there were some illuminating studies on it using novel indicators before. For example, Brockmann et al.[4] have used the flow trajectory of dollar bills to study human mobility and mobile phone was also used in the research of Gonzalez et al.[5] to track mobility patterns.

Problem Statement

These above indicators both have their own limitations because on one hand, as the popularity of electronic payment grows, cash bills is not used as much as in the past and could be less representative nowadays. On the other hand, the data protection legislation varies in different regions so using mobile phone to track human mobility may sometimes end up in an ethically controversial situation.

Therefore, it is promising to delve into analyzing taxi ride data, because taxi is common in metropolises all around the world and the open data is more easily retrieved without sensitive issues. There is a lot of open taxi riding data on the internet, among which the data published by Taxi & Limousine Commission(TLC) of New York City(NYC) is of high quality to study. With the versatile dataset, researchers can examine many aspects of the mobility pattern in NYC. Moreover, it is noticed that since the entry of For-Hire-Vehicle(FHV, such as Uber and Lyft) into the NYC taxi market, it has brought a disruptive influence on the traditional taxi industry and caused some behavioral changes on taxi passengers[6]. For-Hire-Vehicle service vendors that dispatch more than 10000 trips per day, like Uber, Lyft and Via, are granted the category of High-Volume-For-Hire-Vehicle(HVFHV) independent from other For-Hire-Vehicle. They dominate the market of For-Hire-Vehicle and have a more comprehensive dataset covering the ride information compared to the normal For-Hire-Vehicle category.

Research Question

We find it insightful to study data of different types of taxis during a certain time span to compare the usage differences and unveil the mobility pattern of commuters in NYC. Considering the scope limitation of this project, we don't analyze the whole dataset from 2009-2023 but make some careful selection of the taxi types and time frame while establishing the research questions. Our research questions are formulated as follows:

1) ***What is the difference of usage between Green Taxi and HVFHV on weekdays and weekends between 2019 and 2022?***
2) ***For Green Taxi and HVFHV, which type of taxi is preferred on different timeslots in one day?***
3) ***Regarding the travel distance difference between Green Taxi and HVFHV, do commuters have a preference for taking a certain type of taxi for long/short distance travel?***

A lot of research has revealed the conspicuous impact that HVFHV has made on the taxi market and influenced commuter behaviors across USA[6][7][8][9]. Our research questions focus on the passage usage differences between Green Taxi and HVFHV. This is not a random selection. An analysis by the Taxi and Limousine Commission(TLC) of NYC has indicated that 95% of pick-ups of the traditional yellow taxi occurred in Manhattan below 96th Street and at JFK and LaGuardia Airport. Therefore the city launched the Green Taxi program(officially Five Borough Taxi Plan) as a complement to provide more taxi services in outer boroughs and improve the service quality of traditional taxis[10]. Meanwhile, HVFHV is also freely operated in all city boroughs instead of clustering in Manhattan. It can be argued that the operating areas of Green Taxi resemble the ones of HVFHV. And thus the comparison between Green Taxi and HVFHV is plausible because they can be regarded as direct competitors. The data of HVFHV is only available from 2019, and the data of the year 2023 is not completely published yet, hence we only carry out analysis on four-year data of Green Taxi and HVFHV from 2019 to 2022.

The establishment of our research questions aims to provide commercial or academic viewpoints on urban mobility patterns by analyzing a large taxi ride data set of NYC. The rest of this report is structured as follows. The next session discusses the related literature on taxi ride data research. Session 3 describes our utilized data set and we present our methodological framework. In Session 4 we implement big data analysis and try to uncover the results. Finally, Session 5 summarizes our research project and points to future work prospects.

II. Related Work

Previous studies have shown that an increased competition has unethical effect on behaviour of humans. With the introduction of the Green Taxis, the yellow taxis faced more competition in some areas. In order to meet a target income, yellow taxi drivers did not take the optimal route. Fraudulent behaviour came to light in the areas with competition when taxis needed to transport passengers from hotels or during hours when there was a change in driving shifts [11]. With the introduction of For Hire Vehicle, it is likely to assume that unethical behavior by taxi drivers has increased. Another studies have shown that, in general, Uber taxis are additive rather than a replace for yellow and green cabs city wide. However, in some instances, Uber taxis replace traditional taxis. These instances appear to be during the morning rush hour in the central business district of Manhattan [12]. by contrast, the authors [6] shows that there is a shift in consumer preferences towards ride-hailing. Despite Uber leading to an increase in customer complaints about taxi service, consumers still prefers Uber. Moreover, the author shows the entry of Uber into New York resulted in a decrease in the use of traditional taxi rides (yellow taxis). From this paper, we can conclude that, although customers may not be very satisfied with Uber, there is still a significant change in the usage of taxis, with consumers leaning towards Uber.

In addition, a study by Nistal and Regidor recognizes the importance of travel goals with reference to Uber and taxi services.[13] It explores the typical reasons people choose Uber or taxis for transportation, including private matters, going to events, and daily commutes to and from work or school. Moreover, it emphasizes how frequently Uber is used for a variety of purposes in contrast to conventional taxis, suggesting that there may be a difference in the kind of journeys that are made using each service(*). However, the study does not explicitly differentiate between short and long distance trips for Uber and taxi services.

A thorough investigation into the spatial dynamics of the Uber and taxi industries in New York City was carried out by Correa et al. in 2021.[14]Their study focuses on creating equilibrium models to examine how ride-sharing services, especially Uber, interact with traditional taxis on metropolitan road networks. The GPS trajectories and trip records from Uber and Medallion taxis, along with geographical maps of the Manhattan NTA area, were among the extensive datasets that the authors employed. They were able to model the spatiotemporal dynamics within the Uber mmarket, including passenger arrivals, fleet allocation, and the matching of Uber vehicles with passengers, by using this large datasets. Additionally, the research offers significant statistics on the average distance driven by taxis and Uber during the morning hours. Taxi averaged 1.92 miles (with standard deviation of 1.45 miles) while Uber traveled an average 2.23 miles (with a standard deviation of 2.3 miles). However, the statistics were only calculated with data from January through June of 2015. If the data covered a wider range of time, it would be more useful to examine whether people's preferences changed when it came to short-and long-distance travel modes.

Poulsen, Dekkers et al. in 2016 made a thorough examination of Uber and Green Cabs in New York City, with a focus on the outer boroughs[15]. It turns out that on weekends, Uber and Green Cabs are equally popular. Uber's growth, however, appears to be happening more quickly, particularly in the districts close to Manhattan, suggesting that these locations are preferable. Conversely, Green Taxis typically have a stronger presence in less affluent areas. The lack of a clear link between the growth of Uber and Green Cabs in various regions raises the possibility that they serve distinct market niches. This thorough analysis provides the New York Taxi Commission with insightful information by emphasizing the influence of sociodemographic variables on transportation preferences and market developments in the city.

Correa and Moyano in 2023 explores the patterns and trends

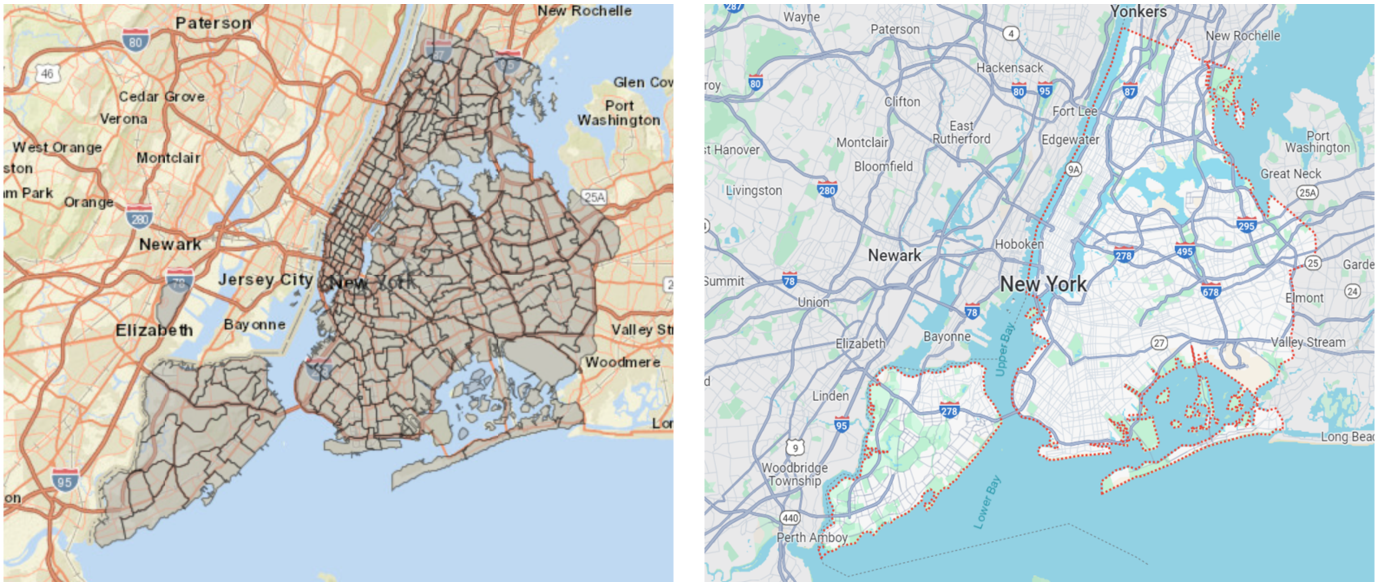


Fig. 1. The Official NYC Text Zone and City Range Map

of yellow and green cab and Uber pickups throughout the five boroughs of New York City, with an emphasis on a Zip Code-based forecasting analysis[16]. It shows a clear change in favor of Uber, especially between April and September, when it overtook taxis in popularity. While green cabs are seeing an increase in patronage in the outer boroughs, Uber's expansion is more noticeable, particularly in lower-class neighborhoods. In order to forecast ridership trends for both Uber and taxis, the study uses cutting-edge machine learning techniques like XGboost and random forest. This gives researchers a better grasp of what kind of vehicles people choose to travel in cities. This method, which combines advanced forecasting tools with Uber data, represents a significant improvement over earlier research.

Another study focuses on data on Uber and taxi pickups, and it finds that there has been a noticeable increase in the usage of Uber, with a 223.3% rise in trips between 2014 and 2015, explored by Correa and Diego in 2022[17]. By comparison, there was a 1.0% decrease in the use of taxis. Compared to Manhattan, the surrounding boroughs like the Bronx and Staten Island had a greater increase in Uber usage. The study highlighted how Uber is becoming more and more popular in New York City than traditional taxi services by using sophisticated spatial analysis algorithms to estimate demand based on a variety of socioeconomic and transportation parameters.

III. Research Methodology

Data Overview

The whole NYC taxi dataset contains data of four types of taxis from 2010-2023. Our project focuses on studying the usage differences between Green Taxi and HVFHV. Besides, HVFHV data is published from February 2019, so we don't have any data regarding HVFHV in January 2019, but we can argue that the lack of one-month data of HVFHV will not cause much bias on the results of our research questions. Our dataset is retrieved from the official data publish platform of TLC, where they publish the data of each taxi type monthly in parquet format. The data of each taxi type contains abundant data attributes to depict the trip information. For each trip record, there is a timestamp respectively for pick-up and drop-off, so that we can trace within which hour on which day the trip occurred, which makes it possible to figure out the answers to RQ1 and RQ2. Moreover, each trip record includes the pick-up location and drop-off location, revealing more travel patterns of the passengers for RQ1. Every pick-up and drop-off location is labeled with a zone number. TLC has published an openly-accessed city taxi zone and the list of zone numbers to clarify the detailed location that every zone number refers to, and from Google Maps we can extract the city range of NYC. This is shown in figure 1. Last but not least, another important attribute of every trip record, namely the trip distance, is used to tackle RQ3.

Methodological Framework

It would be inefficient and cumbersome to manually download all the necessary data month by month into our HDFS. In the beginning, we propose to load the data directly into PySpark through the API of NYC Open Data Platform for the following analysis, instead of downloading and storing the data in HDFS. This helps to avoid the hassle of storing data in HDFS. But later we find it unfeasible because firstly the RESTful API of NYC Open Data Platform has set limitation for the amount of exported data at each request(1000 records per request but most datasets contains more than 1 million records) and moreover the size of dataset makes it slow to

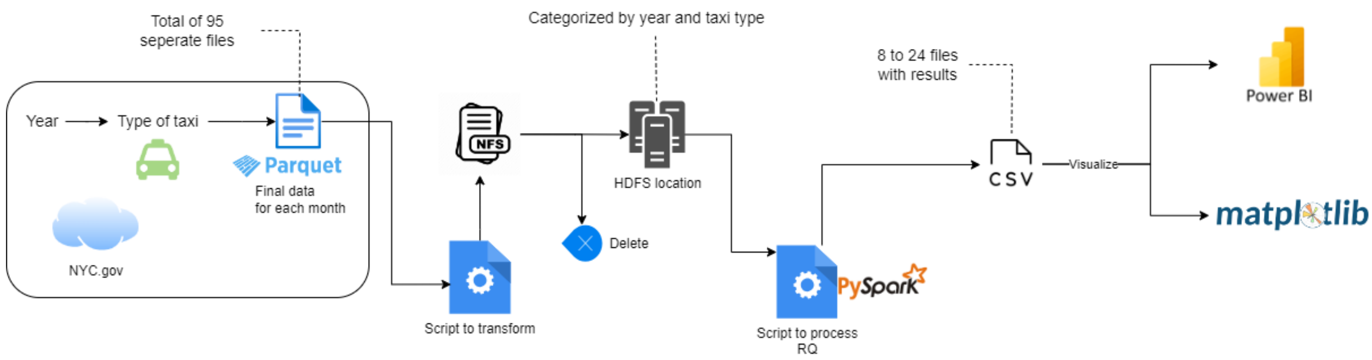


Fig. 2. Methodological Framework

load data every time from the internet to analyze, thus we have to to find another solution. In order to simplify this process, We write a Python program to automate it. The basic workflow of our program is to loop a process in which we use wget to pull one-month data into our NFS firstly, then use subprocess to transfer the one-month data into HDFS, and then delete the data file in NFS. The program can be easily reused as long as the user replaces the old URLs with URLs of any new download sources and it will pull data files of all the input URLs into our HDFS recursively. By doing so it is also avoided that too much data is stored in NFS simultaneously. The data is categorized by year and taxi type during data pull-in to facilitate data management and retrieval. Then we carry out analysis of the data in accordance with our research questions in PySpark framework. After the results are obtained, they are printed into CSV files in HDFS, because CSV file is compatible with a lot of environments and thus is simply convenient for further analysis, data export and visualization. For each research question, 8-24 CSV Files are generated to represent the results. In the end, we utilize PowerBI and Matplotlib for data visualization. The flowchart of our methodological framework is shown in Figure 1.

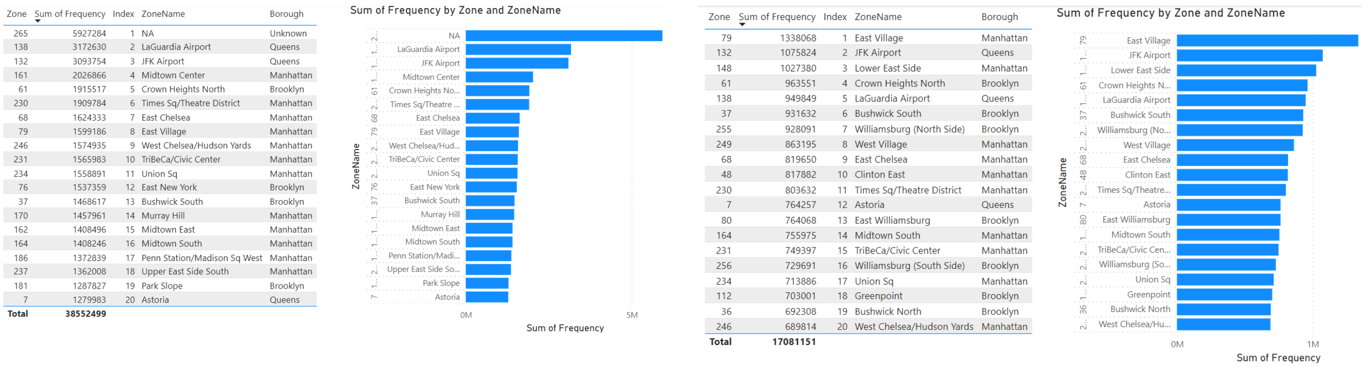
Data Preprocessing

After the pull-in, data analysis is implemented with Spark framework in HDFS environment. For RQ1 & 2, the monthly data needs to be merged into a yearly schema to analyze from an annual perspective. Here we conduct some data preprocessing because the original data is a bit dirty. It contains different datatypes from month to month on the same data attribute, hence we have to covert the datatypes of the same column into the same one in every month. By doing so we can merge the monthly data to yearly data into one dataframe to analyze the data of a certain year. Then we do some data cleaning on the data frame by eliminating columns like "mta_tax", "improvement_surcharge" and "VendorID" because these attributes either have the same value in every record or have nothing to do with our research questions. Next we add a new column on the dataframe to label whether each record is on the weekday or the weekend using *dayofweek* function of SQL. For RQ3, since we only need the attribute of "trip_distance", so we only select and load this column into a dataframe for the following analysis and eliminate all the other attributes, which lightens the data size and saves computational time.

Data Analysis

1) Research Question 1: For RQ1, we define the pick-up location(point of departure) and drop-off location(destination) of each trip as the dimensions we would like to study and compare on "usage", and we have data attributes "PULocationID" and "DOLocationID" accordingly in each record. On weekday and weekend, for each year, two taxi types, We determine to calculate the 20 pick-up and drop-off locations with the highest frequency respectively. Therefore, we can identify that on weekdays and weekends respectively, in each year, for each taxi type, what are the most popular pick-up locations and drop-off locations of the passengers.
We turn the datatype of values in "DOLocationID" and "PULocationID" into an array so that *explode* method can be applied. Meanwhile We also trim all the values of columns "PULocationID" and "DOLocationID" to make sure there is no whitespace before using explode method to turn all the elements in column "PULocationID" and "DOLocationID" into rows. Thereafter we implement *filter* to extract the records on weekdays or weekends and use *groupBy* and descent order to execute the calculation of frequencies.

2) Research Question 2: For RQ2, a similar method is implemented but to the trip frequency during different timeslots. Each trip record has two timestamp attributes, "lpep_pickup_datetime" and "lpep_pickup_datetime", corresponding to pick-up time and drop-off time respectively, but we define the starting time, namely the pick-up time as the occurrence time of the trip and leave out the drop-off time. We segment one hour as one timeslot so there are 24 timeslots per day. Using the pick-up timestamp of each record, again the trip frequency of every timeslot, of each type of taxi, and in each year, can be calculated precisely. Here we ignore a small amount of records with a null pick-up time and concentrate on the valid records. Because the fleet size of Green Taxi is much smaller than the HVFHV one(5-10



Zone	Sum of Frequency	Index	ZoneName	Borough
265	5927284	1	NA	Unknown
138	3172630	2	LaGuardia Airport	Queens
132	3093754	3	JFK Airport	Queens
161	2026866	4	Midtown Center	Manhattan
61	1915517	5	Crown Heights North	Brooklyn
230	1909784	6	Times Sq/Theatre District	Manhattan
68	1624333	7	East Chelsea	Manhattan
79	1599186	8	East Village	Manhattan
246	1574935	9	West Chelsea/Hudson Yards	Manhattan
231	1565983	10	TriBeCa/Civic Center	Manhattan
234	1558891	11	Union Sq	Manhattan
76	1537359	12	East New York	Brooklyn
37	1468617	13	Bushwick South	Brooklyn
170	1457961	14	Murray Hill	Manhattan
162	1408496	15	Midtown East	Manhattan
164	1408246	16	Midtown South	Manhattan
186	1372839	17	Penn Station/Madison Sq West	Manhattan
237	1362008	18	Upper East Side South	Manhattan
181	1287827	19	Park Slope	Brooklyn
7	1279983	20	Astoria	Queens
Total	**38552499**			

Zone	Sum of Frequency	Index	ZoneName	Borough
79	1338068	1	East Village	Manhattan
132	1075824	2	JFK Airport	Queens
148	1027380	3	Lower East Side	Manhattan
61	963551	4	Crown Heights North	Brooklyn
138	949849	5	LaGuardia Airport	Queens
37	931632	6	Bushwick South	Brooklyn
255	928091	7	Williamsburg (North Side)	Brooklyn
249	863195	8	West Village	Manhattan
68	819650	9	East Chelsea	Manhattan
48	817882	10	Clinton East	Manhattan
230	803632	11	Times Sq/Theatre District	Manhattan
7	764257	12	Astoria	Queens
80	764068	13	East Williamsburg	Brooklyn
164	755975	14	Midtown South	Manhattan
231	749397	15	TriBeCa/Civic Center	Manhattan
256	729691	16	Williamsburg (South Side)	Brooklyn
234	713886	17	Union Sq	Manhattan
112	703001	18	Greenpoint	Brooklyn
36	692308	19	Bushwick North	Brooklyn
246	689814	20	West Chelsea/Hudson Yards	Manhattan
Total	**17081151**			

Fig. 3. Drop-off Location Frequency on Weekday and Pick-up Location Frequency on Weekend of HVFHV in 2022

million dispatches vs ¿100 million dispatches per year), in each timeslot, HVFHV has dominantly more trips than Green Taxi. Therefore, it's severely biased to simply compare the number of trips of two types of taxi in the timeslot. To mitigate the bias, we propose to use the proportion of Green Taxi trips to HVFHV in every timeslot and show the variation trend during the day so that we can see at which timeslot the popularity of Green Taxi or HVFHV rises or drops. And we study the trend of each year from 2019 to 2022

3) Research Question 3: For RQ3, it's key to first clarify the definitions of "long distance" and "short distance". Kim, Seok, et al. and Wu et al. have identified that for in commuting, trips with distance of more than 15 kilometers can be defined as long-distance trip while trips with distance of less than 5 kilometers can be defined as short-distance trips and the ones in between are medium-distance trips[18][19]. With the thresholds, we can measure the proportions of three different types of trips for each type of taxi from 2019-2022. We also eliminate some outliers with trip distance of less than 0 or more than 500km in the records. We explore some important statistical indicators of the trip distances, such as mean, medium and probability distribution, from which we infer the preferences of the passengers on taxi types for different trip distances.

IV. Result and Discussion

Research Question 1

Figure 3 are one visualization example of our results, namely the drop-off location frequency of HVFHV on weekdays and the pick-up location frequency of HVFHV on weekends in 2022. The full visualization dashboard for two taxi types in four years is linked in the appendix. Here the zone code 265 and zone name NA refer to locations outside the city range of NYC. Our complete analysis shows that the usage patterns do vary from weekday to weekend. The most popular pick-up locations of HVFHV on weekdays are the airports. During 2020-2021 they become the residential neighborhoods in Brooklyn. As for the reason, We infer that the number of air trips has dramatically declined during these two years because of the pandemic and lockdown. On the weekend the nightlife area of Manhattan is always the most popular pick-up location of HVFHV. The CBD of Manhattan doesn't rank dominantly compared to the airports, residential neighborhoods and nightlife areas. Meanwhile, the most popular pick-up locations of Green Taxi are almost the same as the drop-off locations. They all cluster in Manhattan, ranking top 5 no matter on weekdays or on weekends. And interestingly, the usage patterns of Green Taxi and HVFHV are also quite different. It is recognized that the most popular drop-off locations of HVFHV are outside NYC no matter it's on weekdays or weekends. In addition, While going to destinations outside of New York City and the main airports, most passengers take HVFHV. It's an enormous advantage for HVFHV to Green Taxi because the latter can only run within the city range and not to JFK Airport, Newark Airport or LaGuardia Airport. From figure 1, considering the CBD in Manhattan is just adjacent to the State of New Jersey and the State of New York, this makes sense because a lot of commuters work in Manhattan but live in the nearby states because of lower rent and living expense. The results provide potential commercial references for TLC and the service vendors of HVFHV in planning their fleet dispatch schedule and location.

Research Question 2

Figure 4 shows the usage of Green Taxi spread over the day. Based on this figure, we can see that in 2019, the Green Taxi was preferred in the morning and the afternoon. The HVFHV are popular in the evening and night. One explanation for this result could be that customers would probably not prefer to hail on the street when it is dark outside. However, there is not enough evidence to assume that the latter reason is the only one because the Green Taxi also has their own app service.

Furthermore, figure 5 shows that the popularity of the Green Taxi is declining. This visualization shows the usage of the Green Taxi over the years from 2019 to 2022. What once was a popular choice for customers is now losing ground to the HVFHV. The Green Taxi had more than 6 million rides in 2019, but in 2022 it fell below the 1 million rides (0,8 million).

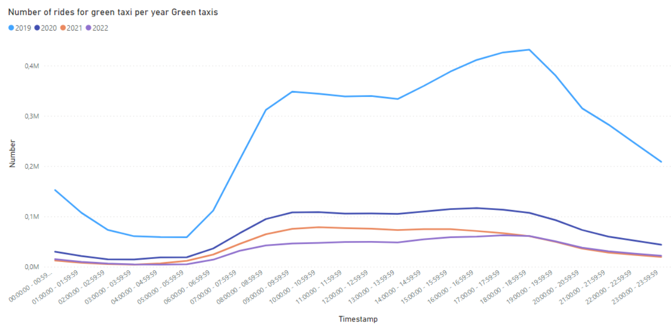


Fig. 4. Usage of Green Taxi in 2019 Across Timestamp

This is a decline of more than 85%. The corona pandemic hit the Green Taxi industry very hard, and the Green Taxi has not been able to recover from this. Despite the decrease in the number of rides, the Green Taxi is still popular in the morning and the afternoon.

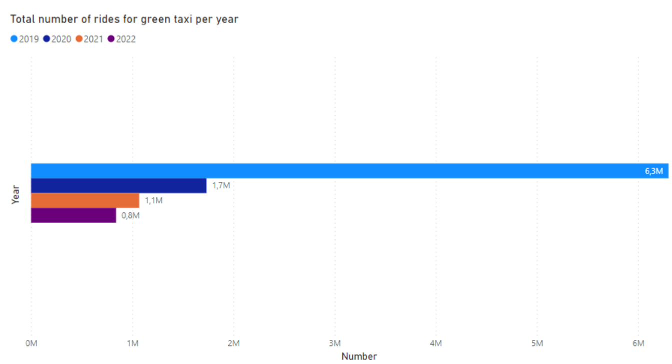


Fig. 5. Total Number Rides Green Taxi

On the other hand, figure 6 shows the number of rides for the HVFHV taxis spread over the time period. We can also see that the usage of the HVFHV has not changed. There is still some similarity in the usage of the taxis.

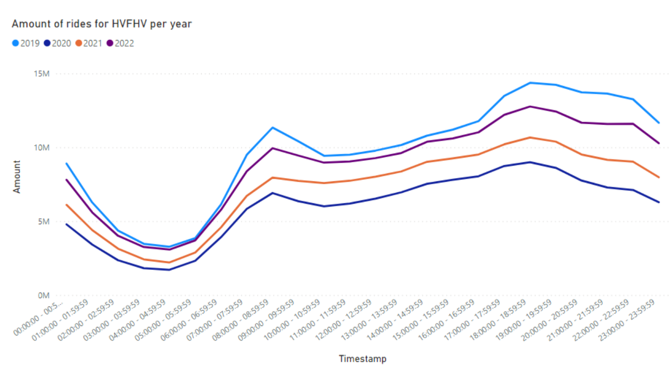


Fig. 6. Total HVFHV Accross Time Stamp

Moreover, even the HVFHV had some problems with the pandemic as we can see in figure **??**, which shows the number of rides per year. However, HVFHV managed to recover fast

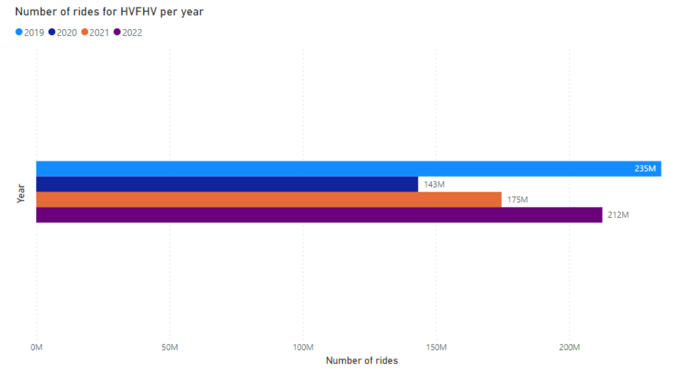


Fig. 7. Total Number Rides HVFHV

from the pandemic, because the number of rides in 2022 is almost equal to the number of rides in 2019. From 2019 to 2022 there was only a decrease of 9,4%.

Even though the data has been collected from the New York City Taxi website, it cannot be guaranteed that the data is completely correct. Furthermore, the data does not show a way to find how many taxis there are. If one vendor has more taxis available then this vendor could also be more accessible for ride pickups, potentially leading to greater dominance in the market. This could be one of the reason why the HVFHV recovered faster than the Green Taxis, perhaps the Green Taxis lost employees.

Research Question 3

For both Green Taxi 8 and HVFHV 8, short distance is much hotter than the other two categories; for the long distance and medium trips, people prefer high volum-fhv than Green Taxi. In addition, people are more into hailing a taxi on the street for a short distance trip rather than using their phones to book an Uber.

The Green Taxi trip distance distribution is depicted in the figure 9. The histogram indicates a high frequency of trips over shorter distances, with a notable peak at the lower end of the spectrum. As the distance grows, the frequency gradually drops. The finding is supported by the vertical lines that depict the mean and median distances of 5.86 km and 3.28 km, respectively.

As the distance increases, the histogram rapidly tapers down from a high frequency of trips at shorter distances. This suggests that the vast majority of travels are short, with the lowest frequency occurring at the shorter end of the distance spectrum. There are more longer trips, but they are less frequent, as indicated by the fact that the median (3.28 km) is less than the mean (5.8 km). This further shows that the distribution is right-skewed.

It appears that people choose Green Taxi for short-distance travel because there are more journeys below the 5 km threshold than those beyond it. This is a typical pattern for the use of taxis in urban areas, as people frequently take short, fast trips for shopping, commuting, and exploring the city.

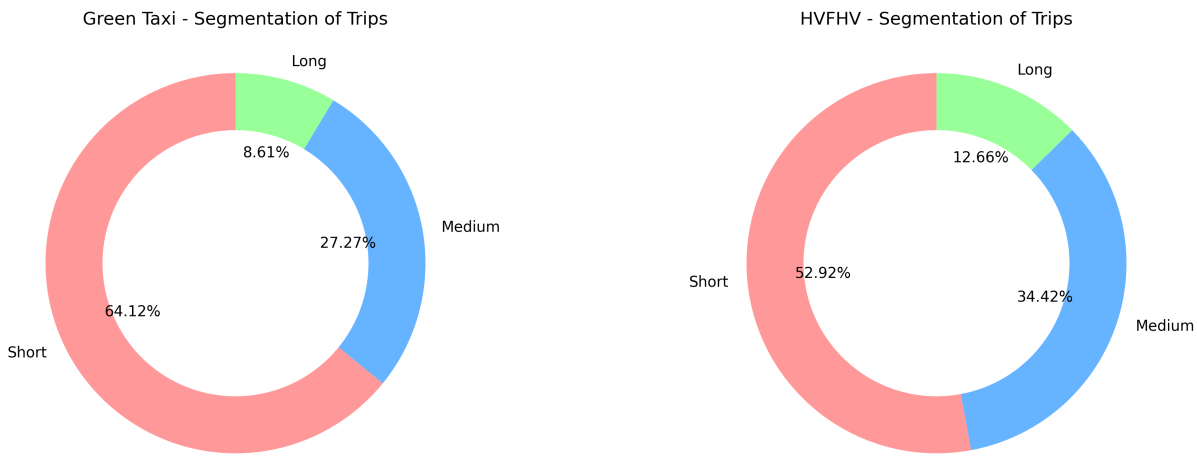


Fig. 8. Proportion of Trips Regarding to Travel Distance: Green Taxi and HVFHV

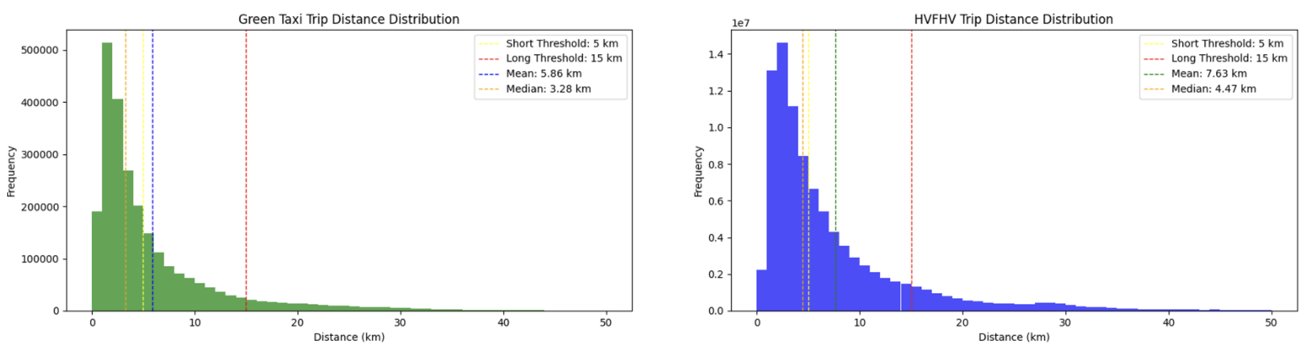


Fig. 9. Green Taxi & HVFHV Trip Distribution

From the HVFHV trips distribution presented in figure 9, we can see a clear peak below 5 km and a noticeable concentration of trips within short distances. The vertical lines that show the mean and median trip distances, at 7.63 km and 4.47 km, respectively, both lie closer to the short trip threshold, supporting the trend of shorter journeys.

The high percentage of trips below the 5 km threshold and the significant decline in trips above the 15 km threshold suggest that consumers mainly use HVFHV services for short to medium distances.

V. Conclusion

It is foreseeable that in the near future, the favor from passengers to HVFHV will consolidate the leading position in the taxi market of HVFHV. They are stronger with a larger fleet and with a more flexible running schedule and range. And they are invincible on the transportation to outside NYC and the main airports. However, we also identify that the usage patterns of Green Taxi are considerably different from those of HVFHV, regarding drop-off locations and pick-up locations. Therefore, TLC should focus on the so-called "differentiation competition" strategy with Green Taxi, and cover the areas where HVFHV does not prevail with Green Taxi. These may not be the most popular locations but in aggregation, they still consist of a large amount of traffic flow. This "long tail" effect is expected to finally generate profitable revenue that matches the fleet size of Green Taxi.

What once was a popular choice, is nowadays losing popularity. The Green Taxi lost more than 85% of their rides after the pandemic (2019-2022). Even though the Green Taxi has lost a lot of clients, it still has its own peak hours which are between the morning and the afternoon. On the other hand, the HVFHV has almost the same number of rides before the pandemic. In 2022, the HVFHV had only a decrease of 9.4% compared to 2019.

Both Green Taxis and HVFHV services are predominantly used for shorter trips in the city, with most travel distances falling under 5 kilometers. The distribution pattern for Green Taxi shows a frequent use for brief distances, with average trips being longer than the median, indicating fewer but consistent longer trips. Similarly, HVFHV services exhibit a preference for short to intermediate travel lengths, as demonstrated by the commonality of trips within a shorter range and a drop-off in frequency for longer distances. This trend underlines a typical urban travel behavior favoring quick and convenient trips.

Reference

[1] M. Veloso, S. Phithakkitnukoon, and C. Bento, "Urban mobility study using taxi traces," in *Proceedings of the 2011 international workshop on Trajectory data mining and analysis*, 2011, pp. 23–30.

[2] B. Jiang, J. Yin, and S. Zhao, "Characterizing the human mobility pattern in a large street network," *Physical Review E*, vol. 80, no. 2, p. 021 136, 2009.

[3] A. Chechkin, R. Metzler, J. Klafter, and V. Gonchar, "Introduction to the theory of lévy flights," in Sep. 2008, pp. 129–162, ISBN: 9783527622979. DOI: 10.1002/9783527622979.ch5.

[4] D. Brockmann, L. Hufnagel, and T. Geisel, "The scaling laws of human travel," *Nature*, vol. 439, no. 7075, pp. 462–465, 2006.

[5] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *nature*, vol. 453, no. 7196, pp. 779–782, 2008.

[6] G. Willis and E. Tranos, "Using 'big data'to understand the impacts of uber on taxis in new york city," *Travel Behaviour and Society*, vol. 22, pp. 94–107, 2021.

[7] J. Cramer and A. B. Krueger, "Disruptive change in the taxi business: The case of uber," *American Economic Review*, vol. 106, no. 5, pp. 177–182, 2016.

[8] G. Dudley, D. Banister, and T. Schwanen, "The rise of uber and regulating the disruptive innovator," *The political quarterly*, vol. 88, no. 3, pp. 492–499, 2017.

[9] D. Gabel, "Uber and the persistence of market power," *Journal of Economic Issues*, vol. 50, no. 2, pp. 527–534, 2016.

[10] N. TLC. "Why boro taxis?" (2013), [Online]. Available: https://web.archive.org/web/20131219021315/http://www.nyc.gov/html/tlc/html/passenger/shl_passenger_background.shtml (visited on 01/26/2023).

[11] I. Ahmadi, "Face/off: The adverse effects of increased competition," *Quantitative Marketing and Economics*, vol. 21, no. 2, pp. 183–279, 2023, Cited by: 0. DOI: 10.1007/s11129-023-09262-z. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85160863410&doi=10.1007%2fs11129-023-09262-z&partnerID=40&md5=ce7bb1bc14e7736969ec1ec7206bcb92.

[12] H. Mammen Kristin Shim Suk, "New york city taxis in an uber world," 2018. DOI: https://doi.org/10.33423/jmdc.v12i2.1254. [Online]. Available: https://academicworks.cuny.edu/si_pubs/205.

[13] P. D. Nistal and J. R. F. Regidor, *Comparative study of uber and regular taxi service characteristics*, University of the Philippines, Available online at: https://ncts.upd.edu.ph/tssp/wp-content/uploads/2016/08/Nistal-Regidor.pdf, Quezon City 1101, No Date.

[14] D. Correa, J. Y. J. Chow, and K. Ozbay, "Spatial-dynamic matching equilibrium models of new york city taxi and uber markets," *Journal of Transportation Engineering, Part A: Systems*, vol. 147, no. 9, p. 04 021 048, 2021. DOI: 10.1061/JTEPBS.0000550. eprint: https://ascelibrary.org/doi/pdf/10.1061/JTEPBS.0000550. [Online]. Available: https://ascelibrary.org/doi/abs/10.1061/JTEPBS.0000550.

[15] L. K. Poulsen, D. Dekkers, N. Wagenaar, *et al.*, "Green cabs vs. uber in new york city," in *2016 IEEE International Congress on Big Data*, IEEE, 2016, pp. 222–229.

[16] D. Correa and C. Moyano, "Analysis and prediction of new york city taxi and uber demands," *Journal of Applied Research and Technology*, vol. 21, pp. 886–898, 2023.

[17] D. Correa, *Exploring the taxi and uber demands in new york city: An empirical analysis and spatial modeling*, Aug. 2017. [Online]. Available: https://ssrn.com/abstract=4229042.

[18] H. Kim, H. Seok, S. I. You, and C. Lee, "An empirical analysis for mode choice in a short-distance trip with personal rapid transit," *Journal of Advanced Transportation*, vol. 2020, 2020.

[19] H. Wu, H. Fan, and S. Wu, "Exploring spatiotemporal patterns of long-distance taxi rides in shanghai," *ISPRS International Journal of Geo-Information*, vol. 6, no. 339, 2017.

VI. Appendix

You can check our dashboards for the results of RQ1 freely via the following links.

Green Taxi 2019
Green Taxi 2020
Green Taxi 20212022
HVFHV 20192020
HVFHV 20212022

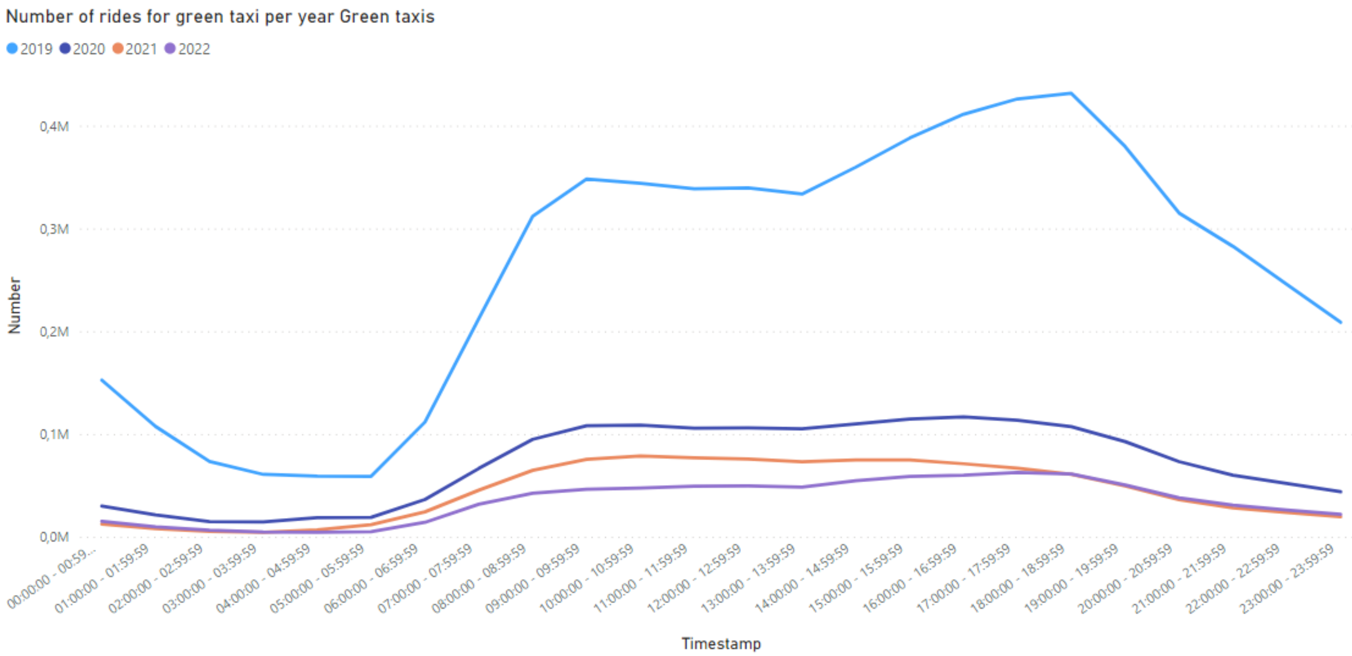


Fig. 10. Usage of Green taxi in 2019 across timestamp

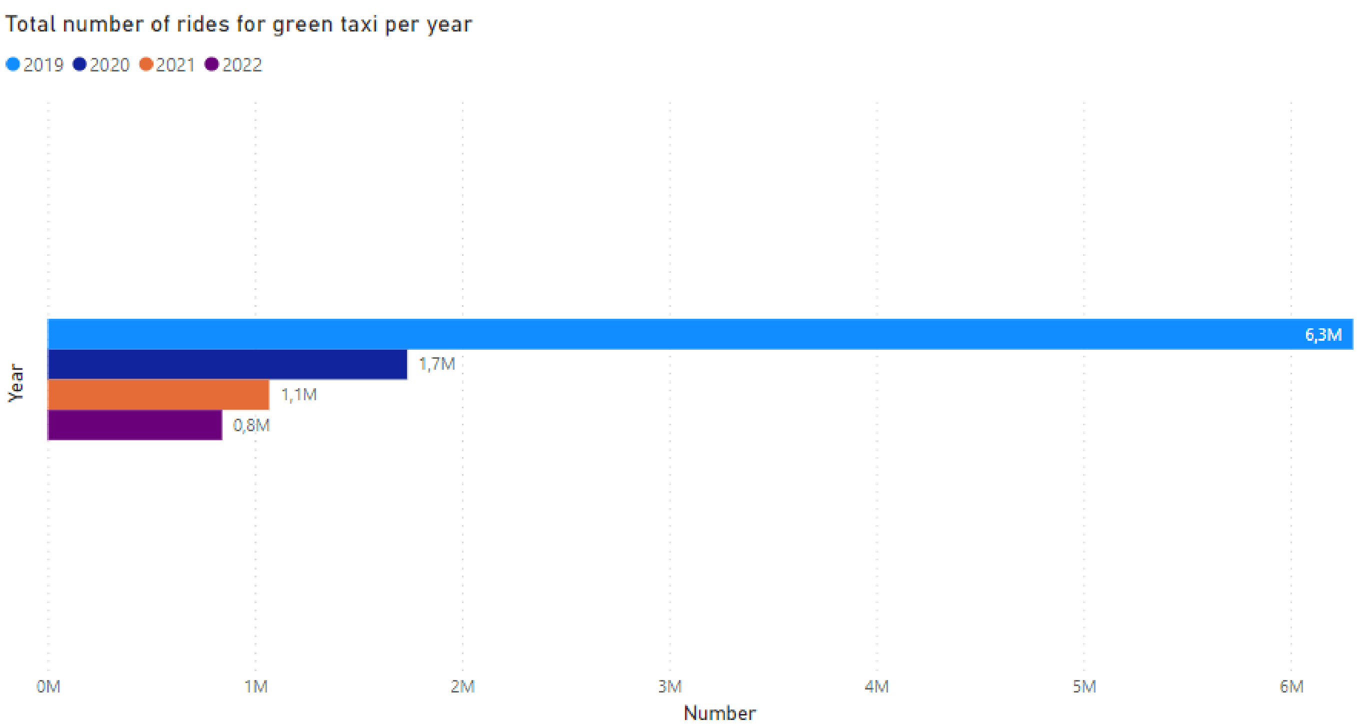


Fig. 11. Total number rides Green Taxi

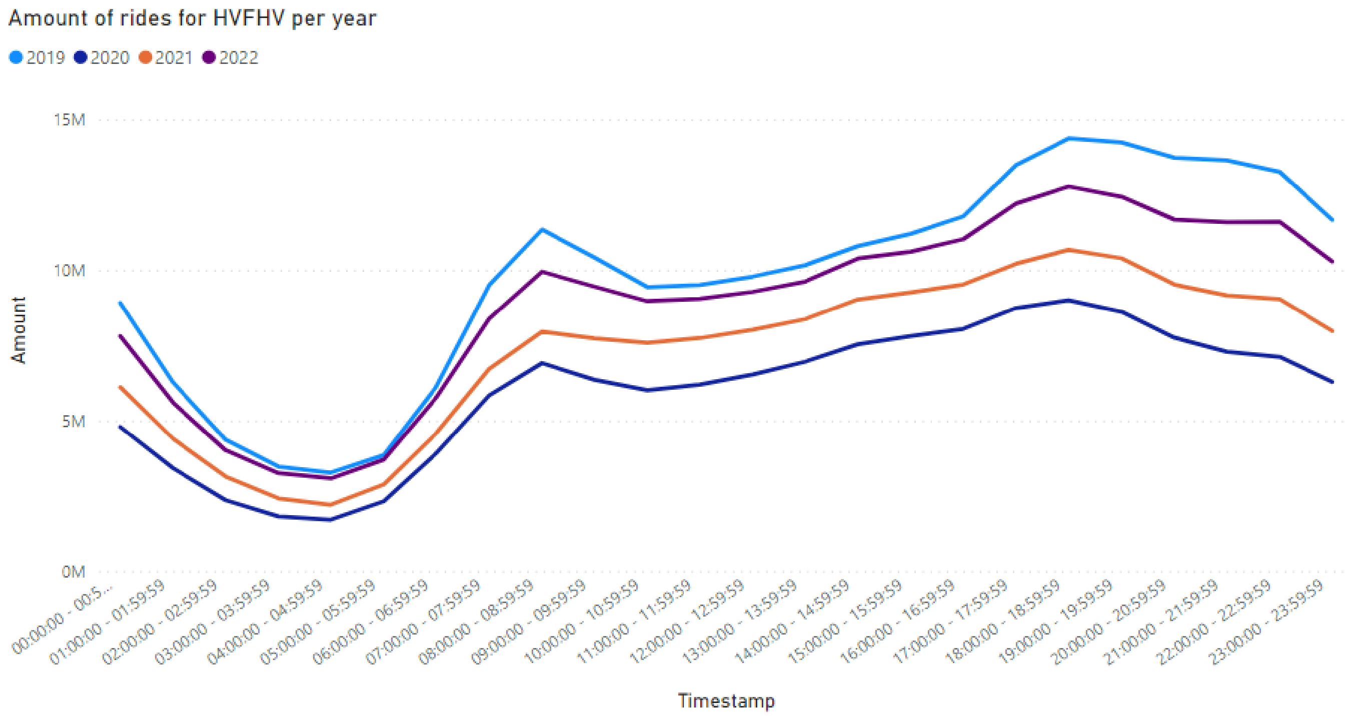


Fig. 12. Total HVFHV accross time stamp

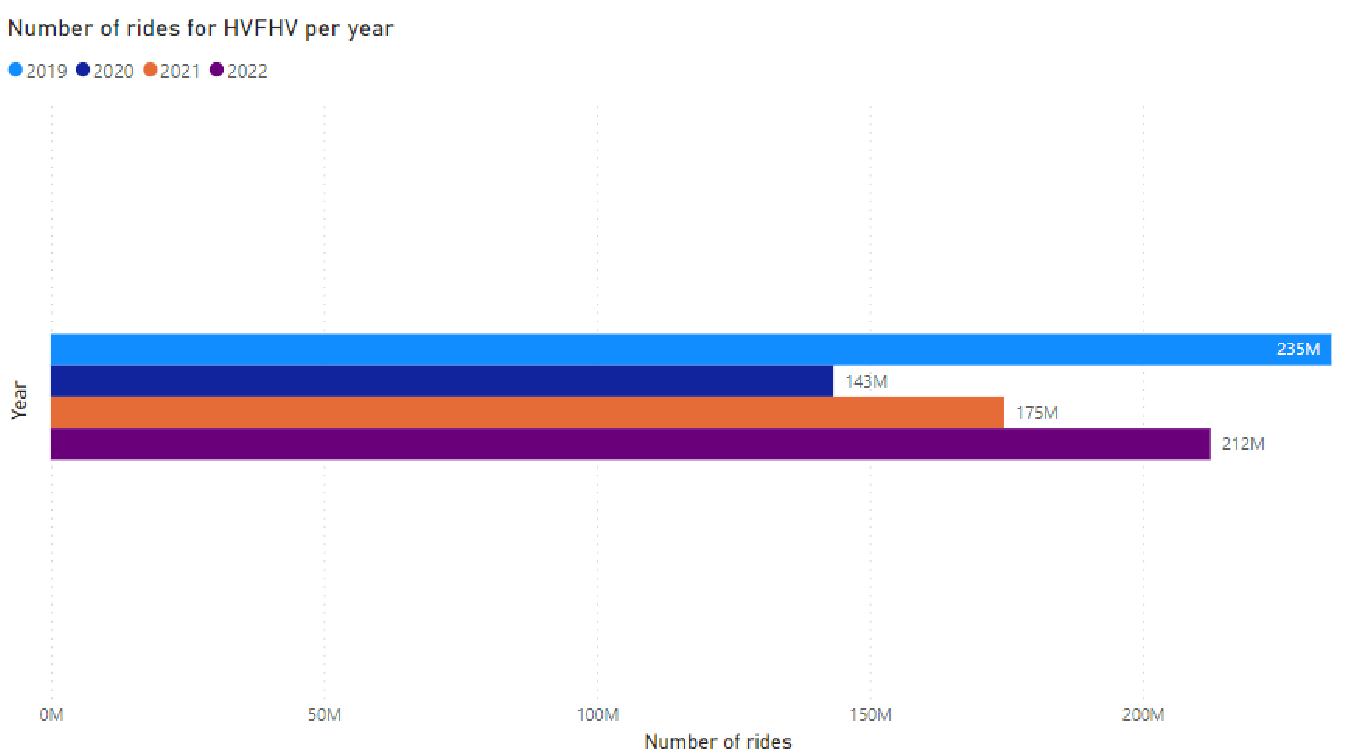


Fig. 13. Total number rides HVFHV