The Underlying relations of Traditional Taxi Services And Competitive Shared-Riding Services

Understanding The Senario Where Shared-Riding Services Succeed And Providing Recommendations For Traditional Services Stakeholders

Chenghao Li
Student ID: 1067999

August 15, 2021

1 Introduction

As the technology briskly took a step forward over the decade, the life quality of citizens has been drastically improved. It has been addressed by several resources[2, 3] that the traditional taxi and high volume for-hire vehicle(FHV) services[1] industry has been heavily affected and even overtaken by giant shared-riding transportation services known as High Volume For-Hire Vehicle(HVFHV)[2], such as Uber and Lyft since 2017 [9]. Based on monthly data provided by Taxi and Limousine Commission of New York(TLC)[1], the research aims to understand the underlying relations between Traditional Taxis and HVFHV by examine the scenarios where HVFHV dominates Taxi services. HVFHV services and traditional services are still closely related and HVFHV perform better possibly mainly rely on the utilization of internet technology. This is concluded from the finding that the number of pickups of HVFHV forms a good linear relationship with Yellow Taxi's data on a daily basis. As a result, traditional taxi services would still have chances to win back the market by applying information technologies to their business and boost their services like HVFHV services do.

2 Datasets

The research takes 5-month period data of Yellow Taxi, Green Taxi, FHV and HVFHV from January 2019 to May 2019 of New York City published by TLC[3]. All data has been pre-processed based on the criterion stated in the following subsections. Since the data of HVFHV in January is missing, when doing comparisons and modelling with HVFHV data, the timeline has been filtered only from February to May 2019. Additional external data sets have been used to identify trips under snowing conditions or occur during holidays. Also, external traffic data of NYC from August to December in 2019[4] has also been included to gain an overview of the quantitative relationships between traffic and transportation services. Additionally, some data has been aggregated on a daily and week daily basis.

[1] Including Black Cars, Luxury Limousines in New York City

[2] TLC-licensed FHV bases that dispatch more than 10,000 trips per day, referred to as High-Volume For-Hire Vehicles (HVFHV)

[3] Taxi and Limousine Commission of New York

2.1 NYC Taxi Data

The data contain the specific trip records of Taxi data, FHV data and HVFHV data. It is worth noting that Yellow Taxis are permitted to hail-riding in the central CBD area while Green Taxis are only permitted outside the CBD[1]. The data has been preprocessed according to the following rules:

2.1.1 Preprocessing

1. Data Cleanse

   Keep Valid Records:

   - from 1st Jan 2019 to 31st May 2019.
   - where trip distances are greater than 0.5 miles and less than 150 Miles.
   - which has valid location IDs.

   Remove Records:

   - where the trip Duration is less than 1 minute or greater than 10 hours.
   - which has negative noisy attributes(count, price, time).

2. Feature Engineering

   - Get Duration for each record: $Duration = EndTime - StartTime$.
   - Calculate $CarSpeed(Miles/Hour) = trip\_distance \div Duration * 60$.
   - Calculate $EarnPerTime(Dollar/Minute) = total\_amount \div Duration$.
   - Add Boolean variable classifies indicating whether it is snow on that day.
   - IQR method[4] has been used to remove outliers of Earn/Minute, Car Speed.
   - Add Boolean varaible isWeekend indicating whether the trip is at Weekend or during Holiday.
   - Add Enum value DayOfWeek to get the day of the week, "Mon, Tue, Wed" etc.
   - Daily avg_*: Calculate average number for several attributes grouped by daily data.

$$daily\_avg = (\sum_{rec=1}^{rec=n} attribute\_value) \div NumPickUps \tag{1}$$

2.2 External Datasets

2.2.1 Date Data

The date data has been used to map dates to a specific day of the week[5]. The trips occur on the weekends or during holidays have been labelled too. In addition, the date of the extreme weather(heavy snow) occurred in NYC has been used [5] to classify whether the trip is impacted by snow.

[4]data outside Q1 - 1.5*IQR or Q3 + 1.5*IQR are thought to be outlier

[5]For example, mapping 01 Jan 2019 to "Monday"

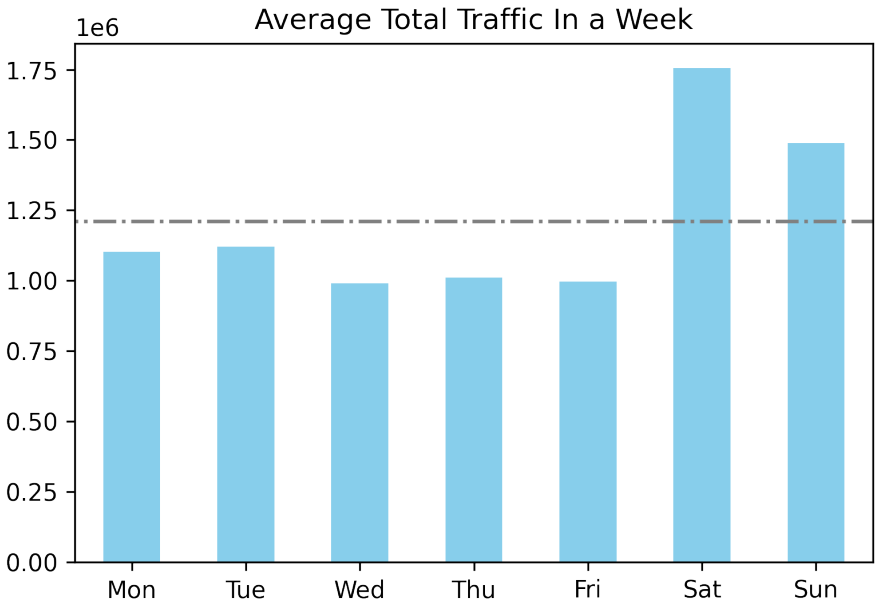


Figure 1: Traffic Pattern During A Week

2.2.2 Traffic Data

5 months data recording New York City's road traffic has been used[4]. The data has been aggregated on weekly daily basis. It is assumed that within one year, traffic statistics follow a similar pattern when grouped by on a weekly daily basis[6]. With the first 6 months data of every year is missing, under this assumption, the averaged data has been used to compare with our first two quarters transportation data. Figure1 indicates that on Weekends, traffic may be usually much higher than weekday's.

3 Daily Data Analysis

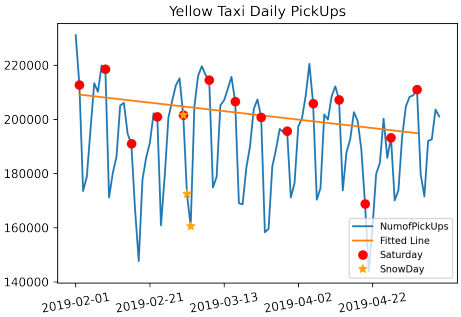
3.1 Number of PickUps Over Time

The business performances of transportation services are estimated with the daily total number of pickups. Snow affected and Saturday's trip has been addressed respectively. A regression line calculated based on Saturday's data has been drawn to visualize the trends of the pickup numbers going over time. It is worth noting that except for FHV in Figure 2c, Saturday's pick up number is generally higher than most of the other days of the week. This corresponds with the traffic data on Figure 1 that people may travel more on Saturday.

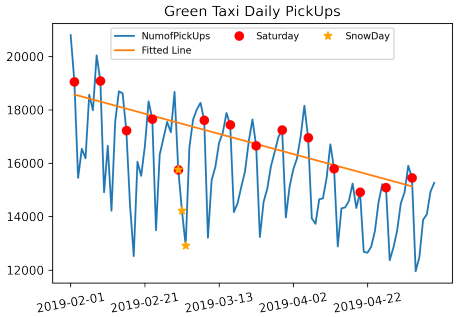
3.1.1 Traditional Transportation Services May Be Decreasing Over Time

The Taxi services show an overall decreasing business performance over the 4-month time period as shown in Figure 2a and 2b. After the extreme weather hits, the traditional FHV business remains low for nearly two months, suggesting the whole business might get a huge hit from extreme weather. Contrarily, the HVFHV business seems to perform steadily over time, seen through Figure 2d. Back in 2014, the whole industry was reported to be shared evenly by Taxis and FHV[7]. Later in 2017,

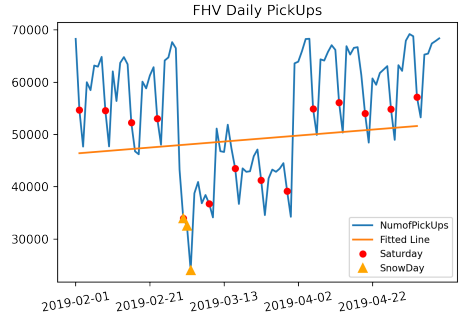
[6]It is assumed that on average, Sunday could always have the lowest traffic volume comparing to Friday's



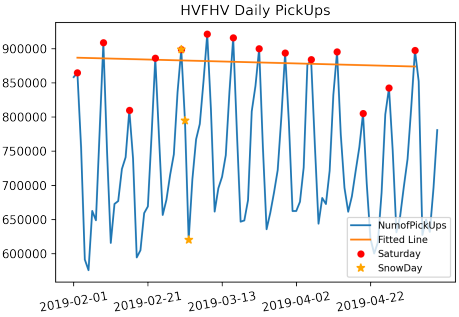
(a) Yellow Taxi Daily Pick Ups



(b) Green Taxi Daily Pick Ups



(c) FHV Daily Pick Ups



(d) HVFHV Daily Pick Ups

Figure 2: Daily Pick Ups for Each Service

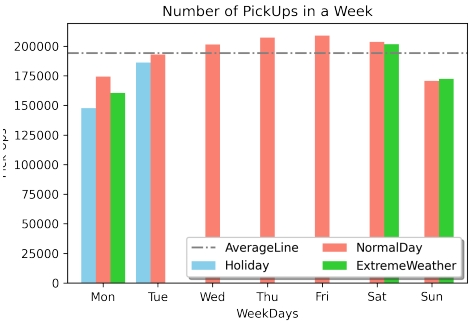
the taxis have been overtaken by HVFHV such as Uber, Lyft[9]. In 2018, to slow down the scary growth of the HVFHV services taking over the industry, the government intervened[9]. However, the intervention just slows the whole process down a little. In 2019, resources record a 36% huge increase in Uber ride's income compared to 2018's data[6]. Moreover, the exact number of pickups labelled in Figure 2 yields that the average number of daily pick ups of HVFHV is 4 to 5 times higher than traditional taxi services. With all these signals imply, the HVFHV services greatly threaten the traditional Taxi business. It is crucial for related stakeholders such as managers of traditional taxi services to contemplate the situation and take action immediately.

3.2 Weekly Analysis On Number Of Pickups

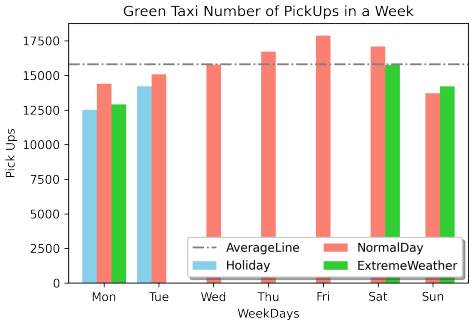
From Figure 2, the number of daily pickups of all services is time sensitive data showing an interesting fluctuate pattern from day to day. With this property been revealed, analysis is further conducted on a weekly daily basis to ensure the comparison has been done on a same level.

3.2.1 Weekend VS Weekday Preference Of Customers

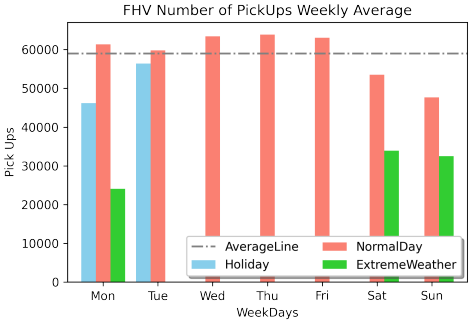
The purpose of this section is to analyse people's relative travelling preferences throughout the week. From the Figure 2, the Taxi services keep relative highest number of pickups during the working days especially on Friday but less on the Sunday and Monday. It corresponds with the traffic data that traffic peaks during the weekend from Figure1. The specific calculations have been done with aggregated data showing that on average on the weekends, Yellow Taxi, Green Taxi and FHV suffers 12.77%, 6.56% and 30.52% drop in pick up numbers compared to weekday's data respectively. However, HVFHVs takes a 14.69% huge leap on the weekends' pickups. It can be deduced that people may



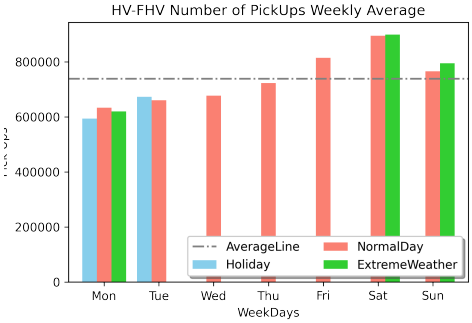
(a) Yellow Taxi Weekly Pick Ups



(b) Green Taxi Weekly Pick Ups



(c) FHV Weekly Pick Ups



(d) HVFHV Weekly Pick Ups

Figure 3: Daily Pick Ups for Each Service

prefer street hailing taxis during working days in a fast hail-and-go condition[7]. With these information in hands, when selecting travelling methods related to entertainment during holidays and weekends, NYC people may prefer shared-riding services, such as Uber and Lyft much more than traditional taxis. It is concluded that taxis businesses may not performance as well as HVFHVs on the weekends. Therefore, more weekend-topic promotions are needed for taxi services.

3.2.2 Weather Impact On Each Service

With the analysis going further, each service impacted by extreme weather (heavy snow) can be examined and analysed. From Figure 3, it is not surprising that extreme weather has an overall negative effect on most of the services. To quantified the data, exact statistics have been calculated. The green, yellow taxis and FHV services endure huge fallen with 23.11%, 24.54% and 45.2% on pick up numbers drastically. On contrary, the HVFHV's business may not be affected as the stats show a 4.63% increases instead. It could be explained that people would always sit in warm cozy rooms ordering a shared-ride service using apps on their phone, rather than standing out-door in the chilling wind hailing for taxis relying on their pure luck. As even extreme weather conditions may not have negative impacts on shared-riding services, it is even harder for Taxi and FHV's businesses trying to prevail in the industry. It is gradually unveiled that the proper use of internet technology may be one of the deterministic helpers to win this competition.

[7]No need to wait for drivers to come by but standing at the street hailing for available taxis

4 Statistical Justification

4.1 Model

Although HVFHV may outperform traditional taxis in general, the two services essentially have some correlated linear relations. The OLS[8] has been used. This further introduces the proposition that information technologies are probably acting as catalysts to HVFHV services. Traditional taxi and HVFHV still follows the similar business pattern.

$$Y = \beta X + \epsilon \tag{2}$$

4.2 Model Tuning

The total additive full model has been used to analyse linear relationships for HVFHV pickup numbers, given yellow taxi pickups, green taxi pickups, yellow taxi's average trip distance, average speed and average total earned amount. The full model has explained the observations well with R-squared value 0.823[9]. The full model usually contains attributes that may be insignificant and can cause over-fitting[10]. Therefore, several statistical steps have been taken to tune the model.

4.2.1 Model Selection And Outlier Detection

- Data has been standardized, mapping to standard normal distributions.
- The model implements default F-test rather than Likelihood Ratio Test as two are proved the same in our case[8].
- Several entries have been removed to improve the model's goodness of fitness by examining the outlier graph using cook distance.
- The model is further selected based on visualising component-component based graph and calling StepAIC[11] function to remove insignificant variables. AIC is preferred over than Likelihood Ratio Test as AIC doesn't restrict nested models to compare.
- Interaction among attributes has been tested[12] with ANOVA[13].

After running through all steps above, the additive model with selected attributes are chosen and defined as following:

$$\begin{aligned} HVFHVPicks = -0.27 + 0.92 * isHoliDay + \\ 0.86 * YellowTaxiPicks + 0.48 * YellowTaxiDistance \end{aligned}$$

As suggested in Figure 4, the model fits the data pretty well with R-squared value 0.757 showing a general positive linear relation indicating the number of pickups of HVFHV has a linear relationship with Yellow Taxi's data.

[8] Ordinary least square linear regression.

[9] One of the Analyses of Variance Family. R-squared value ranges from 0 to 1, the higher the model scores, the better the model explains data.

[10] The model explains training data too well so it cannot be generalised to real-life future data

[11] Akaike Information Criterion. Generally, a model with a Lower AIC score has higher quality.

[12] Interaction suggests if the contribution of attributes can be affected by others.

[13] Analysis of variance, a statistical technique to find out if models are better. The p-value for the full interaction model has been tested to be 0.1274, which is greater than 0.05, so we accept the Null Hypothesis that the interaction model does not perform better than the selected model

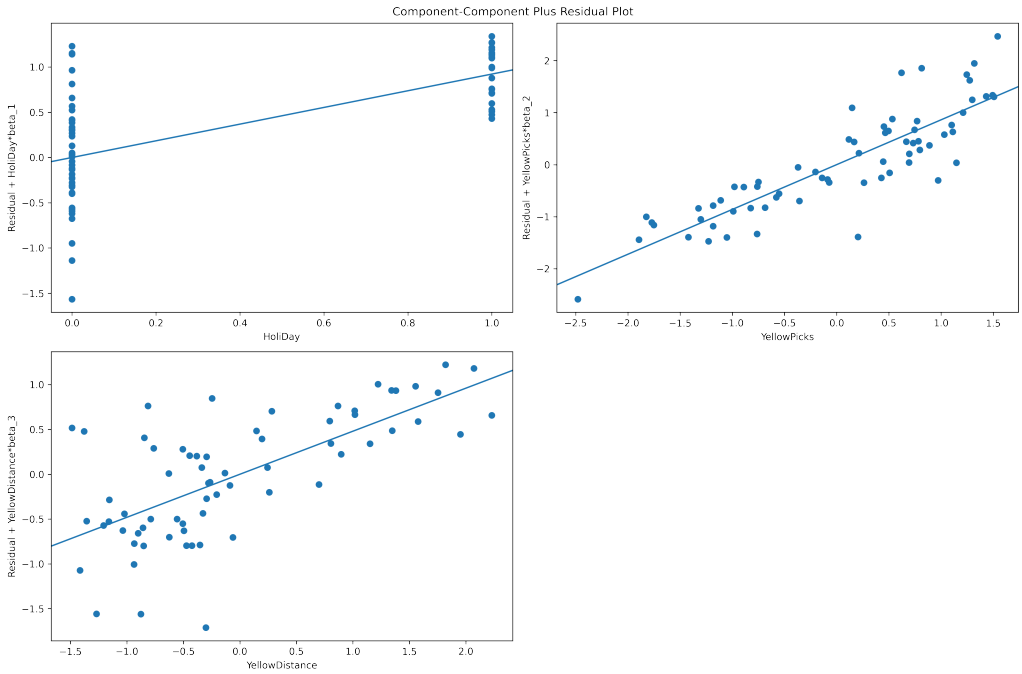


Figure 4: Model Performance

5 Discussions

It has been found that the Taxi and FHV businesses have been taken over by HVFHV services. By analysing the special case such as the impact of extreme weather, it is believed that people are more likely to prefer HVFHV more than traditional taxis or FHV services at times. With traditional transportation services' business performance gradually decreasing over time[9], the related stakeholders of management, for instance, Taxi managers, FHV services high-level officers should be aware that this situation will possibly deteriorate worse.

6 Future Directions and Limitations

The report focuses on the daily and week-daily data then constructs a linear model based on them. The future investigation could be conducted by making use of an advanced Time-series analysis model which is capable of predicting future's data based on historical records.
Additionally, since the January's HVFHV's data is missing and traffic data of NYC is not complete, by tuning the linear model discussed above, it can help to automatically impute missing values and help the NYC government to predict its incoming traffic based on transportation services data.

7 Conclusions

This report aims to analyse the underlying relations between traditional taxi services and HVFHV services. By visualising and analysing the business performance of each services, it is believed that the shared-riding services such as Uber, Lyft, take vast advantages in the competition especially under extreme weather's impact. For related stakeholders of traditional services, it is crucial to start a reformation on their services.

7.1 Recommendations

From technical levels, it is vital to adapt Taxis, FHV services with advanced technologies by designing easily used and customer favourable mobile Apps, introducing algorithmic rating systems to make their service more convenient and high quality.

Moving onto marketing levels, as the analysis indicating from Figure 3, Taxi and FHV's promotion campaign should focus more on leisure holiday-travelling topic to increase their trips volume. The idea that taking Taxis during holidays is somehow a better choice rather than calling a Uber or drive their own cars should be promoted. This can be done by giving away specific free coupons that can only be used during holidays.

Last but not least, traditional taxis and FHV should take advantages of their inherit safety and liable services recognised by public within history.

All in all, With the linear relationship between Taxi and HVFHV been found, it is safe to conclude that the competition between traditional transportation services and shared-riding services is highly possible to be an example of how contemporary technologies boost the traditional industries. Eventually, the competitions between those giant service providers will benefit the local citizens enjoying rides of higher quality and safety with lower cost.

References

[1] "Taxi Fare." Taxi Fare - TLC. Accessed August 3, 2021. https://www1.nyc.gov/site/tlc/passengers/taxi-fare.page.

[2] Lana Pepić. The sharing economy: Uber and its effect on taxi companies, 2018. https://www.researchgate.net/publication/329898519_The_sharing_economy_Uber_and_its_effect_on_taxi_companies

[3] Akira Takihara Wang, Calvin Zhuoqun Huang, Chaoyi Chen, Yong See Foo. A Game Playing Agent To Model Taxi Drivers, 2019.

[4] Department of Transportation (DOT). Traffic Volume Counts (2014-2019). https://data.cityofnewyork.us/Transportation/Traffic-Volume-Counts-2014-2019-/ertz-hr4r. Accessed on August 5, 2021.

[5] National Centers for Environmental Information. Storm Events Database. https://www.weather.gov/okx/stormevents. Accessed on August 5, 2021.

[6] Business Wire. Uber Reports First Quarter 2019 Results. https://www.businesswire.com/news/home/20190530005819/en/Uber-Reports-First-Quarter-2019-Results.. Accessed on August 8, 2021.

[7] Jamison Dague. Taxis, Taxes, And The MTA Funding Gap. https://cbcny.org/sites/default/files/REPORT_TaxisAndTaxes_08262015.pdf. Accessed on August 10, 2021.

[8] Yan Lu, Guo YiZhang, The equivalence between likelihood ratio test and F-test for testing variance component in a balanced one-way random effects model. https://www.tandfonline.com/doi/abs/10.1080/00949650802695664?scroll=top&needAccess=true&journalCode=gscs20. Accessed on August 6, 2021.

[9] The New York Tiems. Uber Hit With Cap as New York City Takes Lead in Crackdown. https://www.nytimes.com/2018/08/08/nyregion/uber-vote-city-council-cap.html. Accessed on August 12, 2021.

8 Appendix

Additional figures generated during the research.

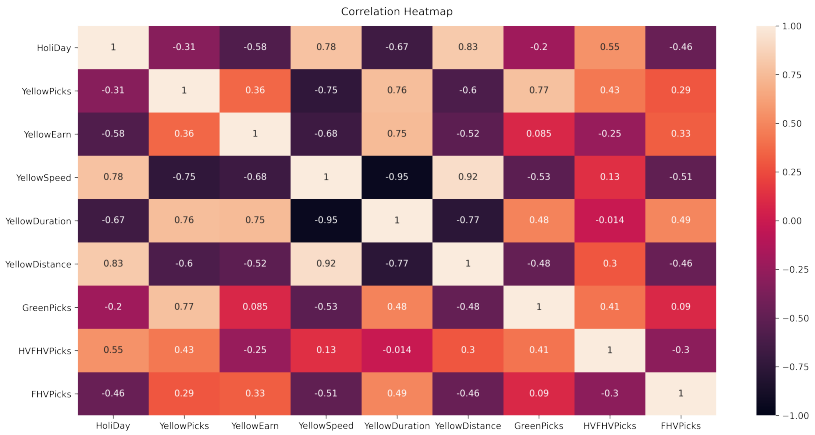


Figure 5: Correlation HeatMap

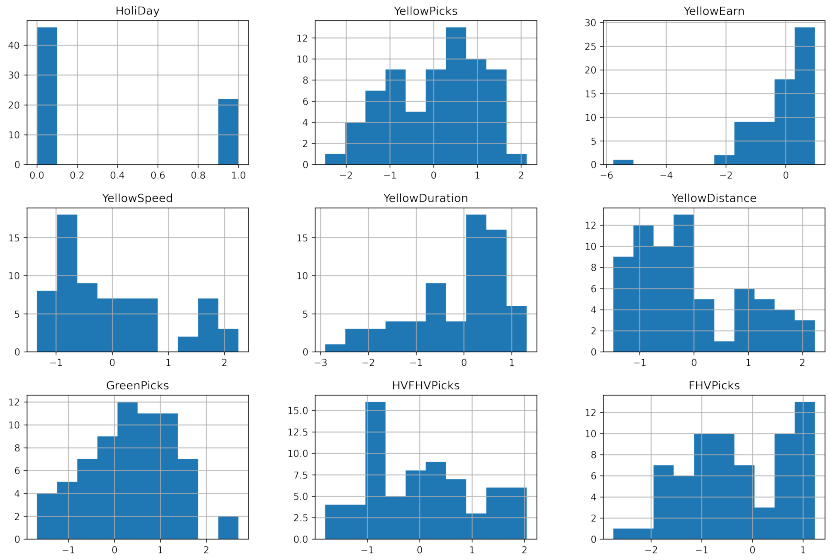


Figure 6: Attribute's Distribution

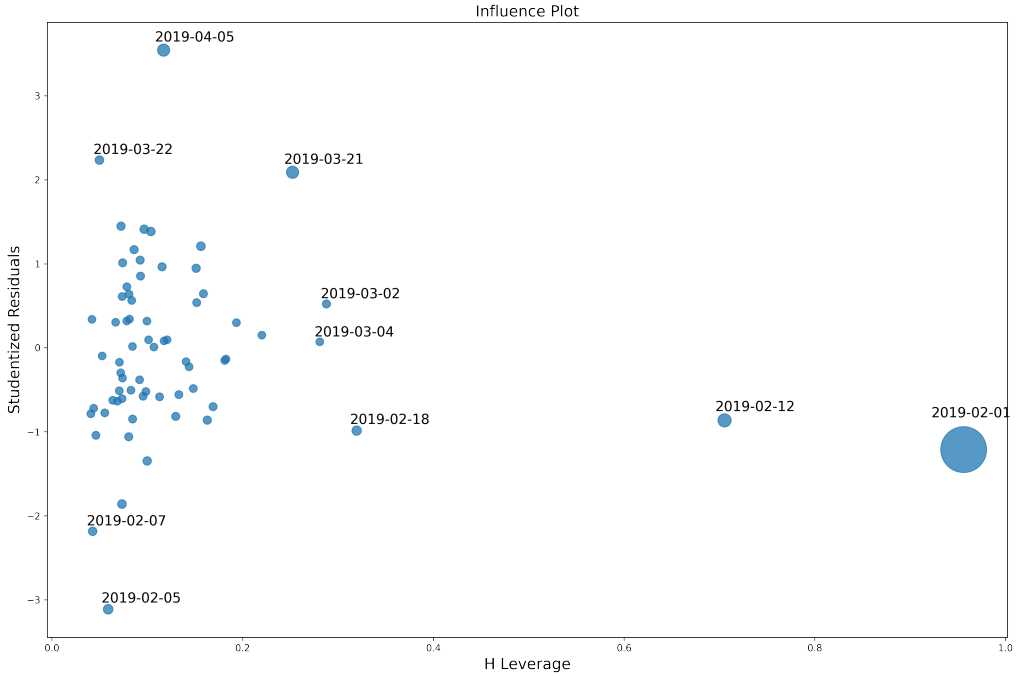
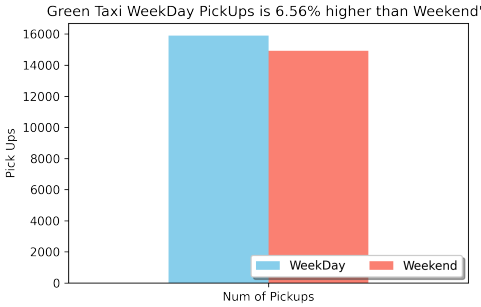
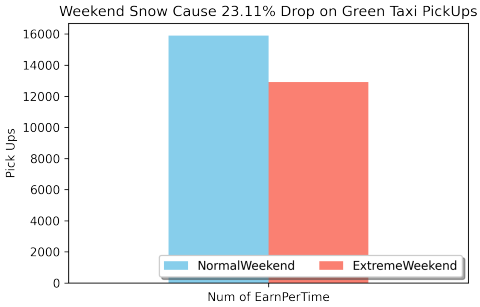


Figure 7: Influence Plot for Outlier Detection

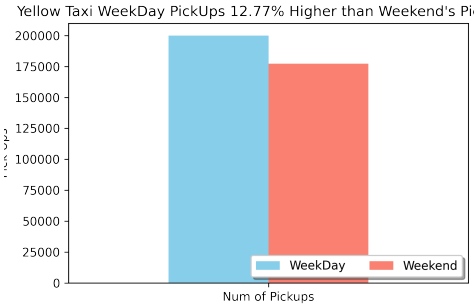


(a) Yellow Taxi Weekly Pick Ups

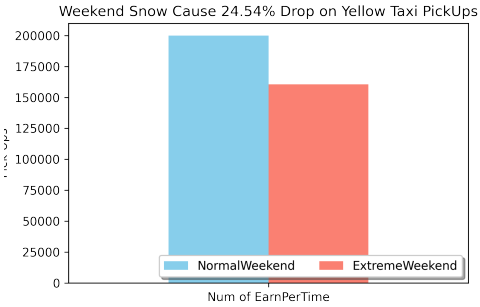


(b) Green Taxi Weekly Pick Ups

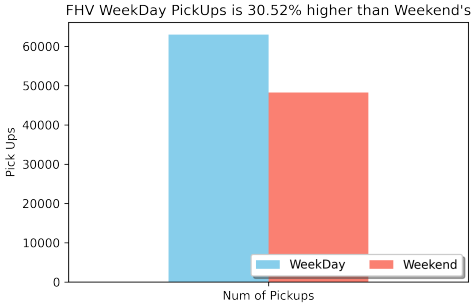
Figure 8: Daily Pick Ups for Each Service



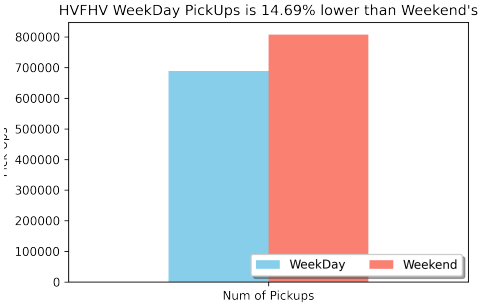
(a) Yellow Taxi Weekday VS Weekend



(b) Yellow Taxi Weekend VS Snow Weekend

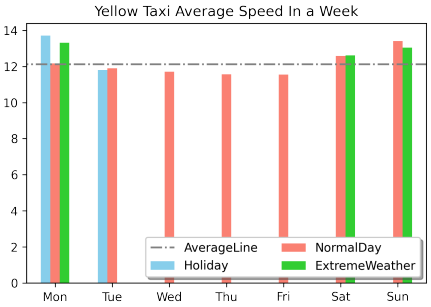


(c) FHV Weekday VS Weekend

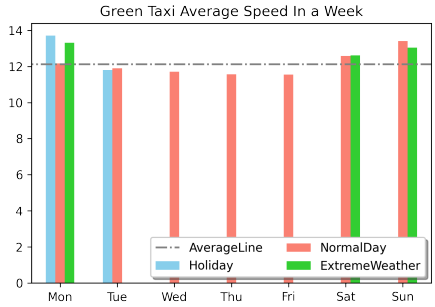


(d) HVFHV Weekday VS Weekend

Figure 9: Appendix Plot

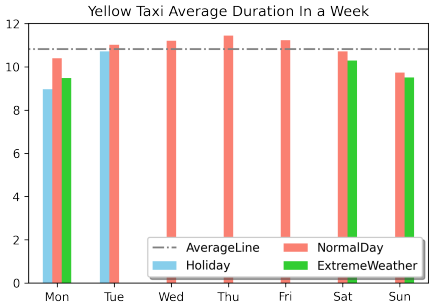


(a) Yellow Taxi Weekly Daily AVG Speed

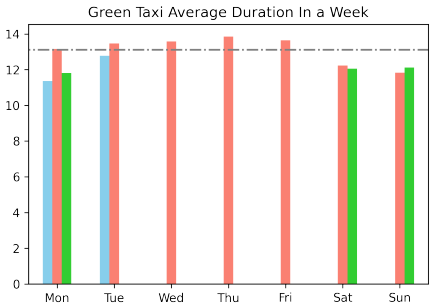


(b) Green Taxi Weekly Daily AVG Speed

Figure 10: Daily Pick Ups for Each Service



(a) Yellow Taxi Weekly Daily AVG Duration



(b) Green Taxi Weekly Daily AVG Duration

Figure 11: Appendix Plot