The Impact of Weather Conditions on Public Transportation Usage in New York City

Aashish Dugar
New York University
Courant Institute of Mathematical Sciences
New York, NY, USA
ad4025@nyu.edu

Herbert Li
New York University
Courant Institute of Mathematical Sciences
New York, NY, USA
hl1785@nyu.edu

Ashwin Subramanian
New York University
Courant Institute of Mathematical Sciences
New York, NY, USA
abs716@nyu.edu

Abstract

In New York City, the demand for public transportation varies heavily based on location and time. However, the relationship between public transit demand and the weather conditions on a given day may be less obvious. In this paper, we explore the relationship between weather conditions and taxi/rideshare usage, and between weather conditions and subway usage at different times and locations across New York City in 2017. We examine whether rain, snow, or average temperature has any effect on the number of people riding taxis and subways, and whether or not there are any relationships we can draw between trends in taxi and subway usage numbers.

Keywords

New York City, NYC, taxi, Yellow Taxi, Green Taxi, For Hire Vehicle, FHV, rideshare, Uber, Lyft, subway, MTA, weather, NOAA, temperature, precipitation, analytic, Hadoop, MapReduce, HDFS, Impala, Spark, visualization, Google Sheets, Tableau

I. Introduction

In this paper, we will present our findings on how taxi, for-hire-vehicle (Uber, Lyft, limousines, etc.), and subway usage in locations across New York City in 2017 was affected by different weather conditions like rain, snow, and average daily temperature. For taxi usage, we will look at the number of passengers that are riding both yellow and green cabs at a particular location and over a particular time interval in New York City. For rideshare usage, since passenger information is not always available, we will instead look at the number of trips started during a particular time interval across different locations in New York City. The metrics we will use to analyze subway usage are number of turnstile entries and exits within a particular time interval at different subway stations across New York City. For subway usage, these metrics are used to calculate the "busyness" and "net flow" of a particular station over a particular time interval. Finally, the metrics we will analyze for weather conditions are average daily temperature, and the presence of rain, snow, and/or other forms of precipitation during the day. Multiple relationships will be studied, including but not limited to:

1. Average daily temperature vs. the number of taxi riders and for-hire-vehicle trips recorded on that day.
2. The presence of precipitation (rain/snow/hail) on a given day vs. the amount of taxi/for-hire-vehicle rides and subway usage on the same day.
3. The presence of adverse weather conditions on a given day vs. the number of turnstile entries and exits at a particular subway station on that same day.

II. Motivation

We hope that this analytic will help both the NYC Taxi and Limousine Commission (TLC) and ridesharing companies like Uber and Lyft decide at which times and at which locations to increase or decrease the number of taxis and ride-share cars in service, based on the forecasted weather conditions in New York City. We hope that this analytic will also help the MTA to decide at which times, stations, and for which subway lines to increase or decrease the frequency of trains in service, based on forecasted weather conditions. From an economic and environmental standpoint, this analytic can also help the TLC, ride-sharing companies, and the MTA save money and reduce energy consumption by decreasing the number of rides in service when demand declines due to certain weather conditions.

This analytic can also help regular transit riders as they will have more access to taxi and rideshare services when demand is high. Furthermore, consumers will have access to these forecasts and determine which mode of transportation they want to use. For example, they might want to avoid the subway when usage is forecasted to be high.

Finally, this kind of analytic can also help the local government decide whether to add additional taxi stops/stands in areas with heavy traffic depending on the weather forecast.

III. Related Work

We have carefully examined six papers that have performed similar analytics to ours. The similarities and differences are discussed for each paper.

A. Predicting Bike Usage for New York City's Bike Sharing System

In this paper, researchers used taxi, temporal, and weather data to predict the number of bike trips in NYC between point A and point B during morning rush hours between 7 and 11 AM. Their approach examined and predicted for only days when it was dry (no precipitation), and resulted in relatively accurate predictive models, with an average root-mean-square error of .42. Through their findings, the authors also saw that the number of taxi rides and the number of bike rides from point A to point B are positively correlated with one another. This analytic related to ours, but instead of predicting bike usage, we will be predicting taxi, rideshare, and subway usage. We will also be looking at temporal and weather data in a broader sense than this analytic. Rather than limiting ourselves to dry days and to certain hours in the day, we will be taking a look at daily demand and concentrate more on the effects that weather conditions have on public transportation usage [3].

B. Context-Aware Taxi Demand Hotspots Prediction

In this paper, the researchers used taxi pickup and dropoff data to determine certain "hotspots" within the Taipei metropolitan area. The researchers first filtered the data by date, time, daily weather conditions, and location. Based on these filters, they then applied various clustering techniques like DBSCAN, k-means, and agglomerative clustering to find either areas with high density, or the top *n* clusters. The results of these clustering algorithms are then shown and compared in the paper. This paper concluded that K-means was very sensitive to outliers, DBSCAN produced too little clusters and was very dependent on model parameters, and hierarchical clustering resulted in many sparse clusters. Regardless of which algorithm was used, the authors calculated a "hotness" score for every street, junction, and landmark area in Taipei, with higher scores representing places where taxi demand was high. The actual score of an area depended on a variety of conditions, which is the context-aware aspect of the paper. These conditions include time (rush hour vs. late night), weather (raining vs. dry), and day (weekday vs weekend). The authors then presented an example application of their analytic: If someone wanted to know the demand for taxis given that it was rush-hour on a perfectly sunny Monday morning for example, the authors would then use their historical "hotness" data for that area under the given conditions to predict the demand for taxis. Our analytic differs from this analytic in that we are not only considering a different city, but we are also analyzing for-hire-vehicle and subway usage in addition to taxi usage [4].

C. Mining Open Datasets for Transparency in Taxi Transport in Metropolitan Environments

The arrival of Uber changed something about the taxi transportation industry. Pricing patterns now change every minute, driven by algorithms based on offer and demand put forward by Uber. Uber's changes in pricing (surge pricing) can vary from one neighborhood to the next one in a city. Prices change in a real-time manner that is hard to trace, for authorities and the public. The case of Uber as a game changer in urban transport economics is motivating researchers to consider taxi mobility data from an economical point of view, in order to estimate and compare the financial costs incurred by customers of different taxi providers. They perform data analysis on a large, free, and open dataset of yellow taxi cab mobility records in New York City to characterize their mobility and pricing patterns. Pricing directly relates to well-known patterns observed in the past on human urban mobility. Most taxi movements are within a short distance range with longer movements occurring less frequently in the data. Yellow cabs in NYC are compared to Uber's service, Uber X, and UberPool is consistently more expensive for short trips, and has a higher revenue. Our analytic is very different from this one. Our focus is on analyzing the relationship between weather conditions and taxi, rideshare (including Uber), and subway usage, and not on how rideshare services (such as Uber) impact the economics of the taxi transportation industry. However, we hope our analytic will help taxi and rideshare services become more economical by increasing or decreasing service in response to different weather conditions [8].

D. Real-Time Prediction of Taxi Demand Using Recurrent Neural Networks

Predicting taxi demand throughout a city can help to organize the taxi fleet and minimize the wait time for passengers and drivers. The paper proposes a sequence learning model that can predict future taxi requests in each area of a city based on the recent demand and other relevant information. Remembering information from the past is critical here, since taxi requests in the future are correlated with information about actions that happened in the past. For example, someone who requests a taxi to a shopping center, may also request a taxi to return home after a few hours. The paper uses one of the best sequence learning methods, long short-term memory that has a gating mechanism to store the relevant information for future use. This method is evaluated on a data set of taxi requests in New York City by dividing the city into small areas and predicting the demand in each area. The paper shows that this approach outperforms other prediction methods, such as feed-forward neural networks. In addition, the paper shows how adding other relevant information, such as weather, time, and drop-offs affects the results. The paper analyzes four types of weather conditions - rain, snow, fog, and thunder. Our paper also analyzes these conditions, in addition to other conditions. However, our research does not involve any sequence learning model or neural networks [7].

E. Analyzing the Effect of Weather on Uber Ridership

The paper summarizes a good model of how impactful weather is to a certain mode of transportation, in this case being Uber (This is synonymous with how a drop in taking rides would be a factor in increased turnstile usage, which is an important consideration to our team's project). It provides a model that takes into consideration key features to provide a prediction on whether climate impacts Uber rides in a positive or negative way, and by how much. These features include and not limited to pick up time, pick up locations, date, temperature, humidity, wind etc. We can see the various statistics provided that help us see what each event, i.e., weather condition does to the number of rides. We can see a big drop in driving for rain/thunderstorms and fog, and spikes in regular rain and snow. Another condition listed the consideration of weekends, where we can see Sunday has a big impact on drives and Wednesday has an influence on the weekdays, possible reducing usage of the MTA. The model was an ARIMA model based on a mean absolute percentage error. This is essentially time series forecasting to see the correlation between the model predicted values and the true data. We can finally see the prediction values are quite accurate, depicting that weather is a direct influencer of rides usage. With respect to our project, if we can find a correlation between the rides vs. turnstile usage, we can determine what transportation is used most given the weather condition and day of the week [12].

F. The Impact of Rainfall on the Temporal and Spatial Distribution of Taxi Passengers

The paper primarily focuses on the temporal and spatial distribution of taxi passengers (i.e. how the taxis are spread out and scheduled to run for optimality on rainy days). The paper concludes that during non-rush hours, taxi demand drops in suburban areas as rainfall goes up. Whereas for rush hours, it significantly increases. The final recommendation is to channel taxis to areas where the rush is more so as to meet the demands and give ride opportunities. The studies were performed from an economic perspective. The introduction begins to describe the pros and cons on rainy days. Taxi drivers earn more money for the usual trip, which in a way correlates to reduced work hours. The side effects can be traffic congestion which can have a negative impact on the earning and drivers' work satisfaction. Motivation is a key mood, which is related to work hours. To retrieve the data, the GPS devices installed on the taxis can be coupled with other devices to receive vehicle information regularly such as location coordinates, direction, speed etc. Using these, we can comprehensively determine the taxi service demand. Moving on to the data cleaning and processing, there are certain characteristics taken into consideration. (This might be a good way to consider data for our project too) – *Record No; Vehicle ID; Timestamp; Longitude; Latitude; Velocity; Heading Direction* and *Running State*. Then the start and end points were extracted for each pick up and drop off, and a route was mapped accordingly that the taxi takes. Other factors such as driver profitability are also considered, where we see the speed of each driver helps determine them having a greater probability [14]. Moving to the conclusion, 2 important points were highlighted:

1. It verifies that the concept of highly profitable passengers meets the research requirement for viability and effectiveness in spatio-temporal evolution analysis of urban taxi services.
2. It demonstrates that the impact of rainy weather on taxi service demand, especially on highly profitable taxi service demand, is concentrated around the city central areas and that the most effective response for taxi drivers in Wuxi and Kunming to earn more money and limit discrepancies between taxi service demand and supply is to increase their service times.

But these come alongside 2 limitations:

1. The confidence in the results could be improved by analyzing more extensive taxi GPS data and more informative weather record features.
2. In addition, the methods and results presented in this research can be used as a practical reference for operating and dispatching urban taxi services.

IV. Design and Implementation

A. Design Diagram

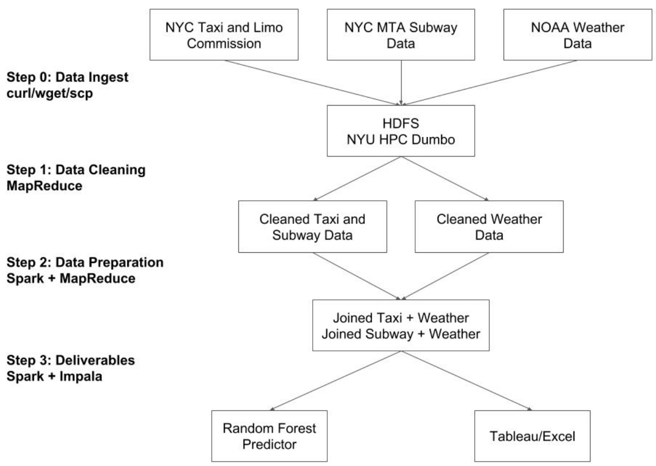

Figure 1: Design Diagram

B. Design Details

General Implementation:

1. Ingest the TLC Trip Dataset, MTA Dataset, and NOAA Weather Dataset into HDFS
2. Clean and profile the data using MapReduce. Output to HDFS.
3. Join the weather data with TLC Trip data and join the weather data with MTA data. Output to HDFS.
4. Perform analysis on the joined data using MapReduce, Spark, and Impala. Output to data visualisation tools (e.g. Tableau, Google Sheets).

Implementation Details for NYC TLC Data:

1. Ingest data using *curl* and *wget*
2. Put data into HDFS (hdfs dfs -put)
3. Remove null rows and unnecessary columns (see **Section IV-C. Description of Datasets** for retained columns)
4. Add borough and neighborhood columns from an external CSV dictionary provided by TLC, additionally for FHV data, add a ride count column.
5. Join TLC data with weather data using Spark, based on month, day of month, and year
6. Load joined data into Impala tables
7. Run SQL queries, exporting query results as CSV files
8. Import CSV files into Google Sheets for visualisations

Implementation Details for NYC MTA Data:

1. Ingest data using **scrappy**, a web scraping tool, directly into the Dumbo cluster (HDFS)
2. Insert/Remove necessary data, i.e, Timestamp to be added, Unit no. to be removed.
3. Minimize anomalous/erroneous data. (Some of the null values cannot be removed as that breaks the continuity of the station times and recordings.)
4. Get analytic necessary from the data. (namely the business and netflow per day, most busy days etc.)
5. Merge with weather data for trend analysis with the turnstile data.
6. Export data to CSV for visualizations (done via Tableau.)

Implementation Details for NOAA Weather Data:

1. Ingest data (weather_input.txt) into HDFS using secure copy from local machine into Dumbo and then hdfs dfs -put from Dumbo into HDFS.
2. Filter out data we do not need (cleaning). Ultimately, we want date, precipitation, snowfall, and average temperature, separated by commas.
3. Output cleaned data to HDFS. This data will be joined with TLC and MTA data for further analysis.

C. Description of Datasets

Here, we provide the data schemas of our final datasets before joining. Here, we distinguish between columns in the original dataset that we have retained, and columns we computed (comp.) via MapReduce and Spark.

NYC Taxi and Limousine Commision (TLC) Data:

- **Pickup Time** - Timestamp
- **Drop-off Time**
- **Trip Distance** - In miles
- **Pickup Zone ID** - Number between 1 and 265
- **Drop-off Zone ID** - Same as Pickup Zone ID
- **Pickup Borough (comp.)** - one of Manhattan, Bronx, Queens, Brooklyn, Staten Island, EWR, Unknown
- **Pickup Neighborhood (comp.)** - e.g. West Village
- **Drop-off Borough (comp.)** - same as Pickup Borough
- **Drop-off Neighborhood (comp.)** - same as Pickup Neighborhood
- **Passengers (comp.)**

NOAA Weather Data:

- **Date** - Format “MM/DD/YYYY”

- **Precipitation Amount** - in inches, includes rain and snow
- **Precipitation Flag (comp.)** - 0 or 1 denoting whether it rained/snowed (1) or not (0)
- **Snow Depth** - in inches, amount of snow on the ground
- **Snow Amount** - in inches, amount of snow that fell on that day
- **Snow Flag (comp.)** - 0 or 1 denoting whether it snowed (1) or not (0)
- **Min Temperature** - in degrees Fahrenheit
- **Max Temperature** - in degrees Fahrenheit
- **Average Temperature** - in degrees Fahrenheit
- **Average Wind Speed** - in miles per hour
- **Fog flag (comp.)** - 0 or 1 denoting whether or not fog was present on that day
- **Thunder flag (comp.)** - 0 or 1 denoting whether or not thunder was present on that day
- **Hail flag (comp.)** - 0 or 1 denoting whether or not hail was present on that day
- **Haze flag (comp.)** - 0 or 1 denoting whether or not haze was present on that day

NYC MTA Turnstile Data:

- **Period Start (comp.)** - timestamp
- **Period End (comp.)**
- **Station Name** - e.g. 14th Street Union Square
- **Line Name(s)** - e.g. N, Q, R, L, 1, 2, 3.
- **Turnstile Counter Difference (comp.)** - difference between two consecutive time intervals. To calculate busyness, we summed entries and exits and subtracted the sum of entries and exits of the previous time interval. For netflow, we subtracted entries and exits and subtracted the difference of entries and exits of the previous time interval. Positive netflow means more people entered this station than left, negative netflow means more people left this station than entered.

V. Results

We used Google Sheets for visualising the relationships between the NYC Taxi and Limousine Commission (TLC) Data and the NOAA Weather Data. Specifically, we analyzed six relationships:

1. Green Taxi Usage vs. Precipitation (rain/snow)
2. Yellow Taxi Usage vs. Precipitation (rain/snow)
3. For Hire Vehicle Usage vs. Precipitation (rain/snow)
4. Green Taxi Usage vs. Average Temperature
5. Yellow Taxi Usage vs. Average Temperature
6. For Hire Vehicle Usage vs. Average Temperature

All six analyses were done across the five boroughs of New York City (Manhattan, Brooklyn, Queens, Bronx, Staten Island) as well as Newark Airport. In all six analyses, for the most part, we discovered that average daily temperature and precipitation does not have a significant impact on Yellow Taxi, Green Taxi, and For Hire Vehicle usage in 2017. However, we did notice the following three trends.

1. Average Green Taxi usage decreased somewhat significantly in Manhattan when it snowed, and increased somewhat significantly in Brooklyn when it snowed. See **Chart 2** for more details.
2. Average For Hire Vehicle usage increased somewhat significantly in Manhattan when it snowed. See **Chart 3** for more details.

Before running the analyses, we expected Green Taxi, Yellow Taxi, and FHV usage to increase significantly in all locations when it rained or snowed, or when the average temperature was very low (below 32 degrees F). We reasoned that people may not want to walk long distances outside in inclement weather conditions. However, our expectations turned out to not be true. We have provided the following charts (charts 1-7) visualising usage numbers:

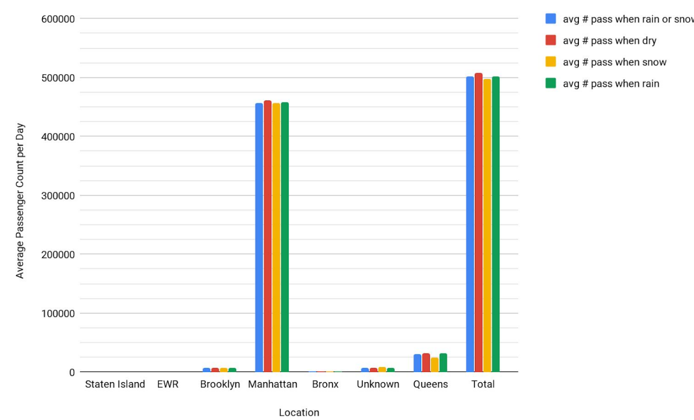

Chart 1: Yellow Taxi Usage vs Precipitation

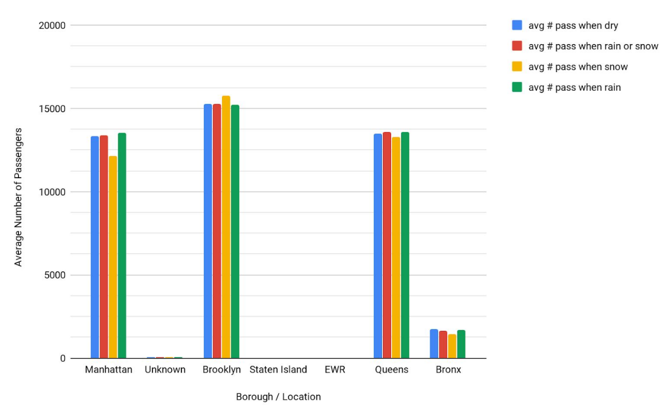

Chart 2: Green Taxi Usage vs. Precipitation

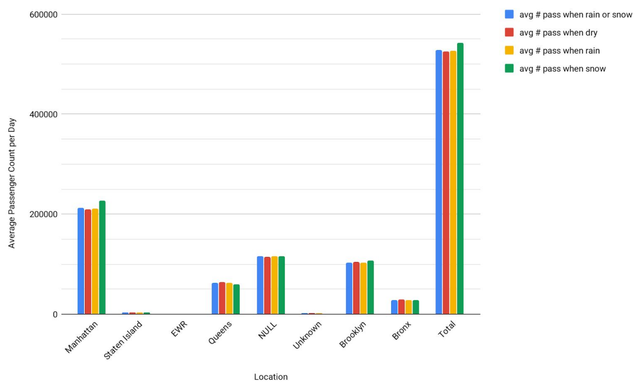

Chart 3: For Hire Vehicle Usage vs. Precipitation

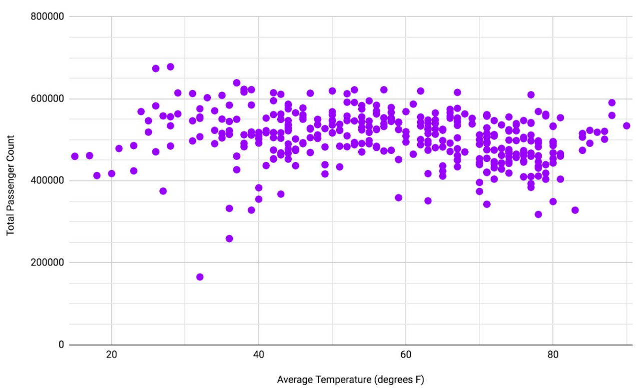

Chart 4: Total Yellow Taxi Usage vs. Temperature

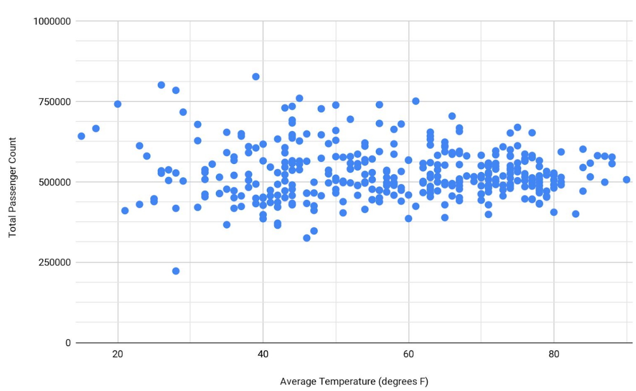

Chart 5: Total For Hire Vehicle Usage vs. Temperature

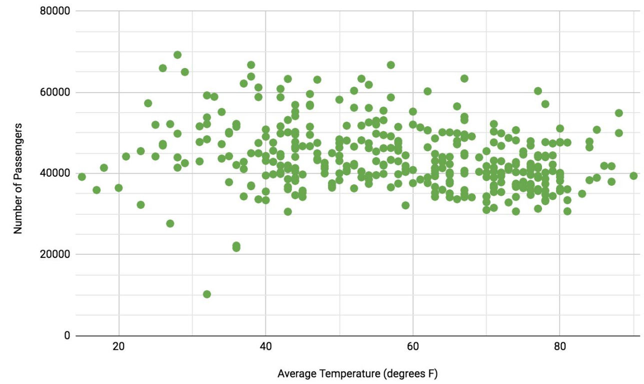

Chart 6: Total Green Taxi Usage vs. Temperature

Our results are consistent with the findings of previous research. One similar analytic was performed by Schneider, in their study "Analyzing 1.1 Billion NYC Taxi and Uber Trips, with a Vengeance" [17]. Their analytic looked at taxi trip data between 2009 and 2015, concluding that rain did not seem to affect daily taxi usage. The same holds for our findings regarding taxi ridership in 2017. However, Schneider did find that snowfall had a significant negative impact on taxi ridership, whereas we found no significant impact. This can be attributed to the fact that it only snowed 12 times in 2017 so our sample size may not be large enough to notice any trends.

With respect to the Turnstile data, there were some noticeable changes when it is compared with the weather data. The following relationships were analyzed:

1. Turnstile usage on weekdays vs. weekends
2. Turnstile usage on dry days vs. snow days
3. Turnstile usage on normal days vs. holidays
4. Turnstile usage vs. average daily temperature

These analyses were done for all the subway stations across all boroughs in NYC. We noticed the following trends:

1. The most usage comes around the time of holidays, i.e, Thanksgiving, Columbus Day etc.
2. The next highest usage is observed on Friday nights.
3. The 3rd highest usage is observed on is regular weekdays.
4. Turnstile usage increased on hotter days.
5. Immediately after reported snow or some weather condition (see figure below.), the following time period sees a steep drop in usage of turnstiles for a certain time segment, and then it resumes to significantly increase.

These results are comparable to the results found by Singhal, Kamga, and Yazici in their paper: "Impact of weather on urban transit ridership" [16]. Singhal, et. al. analyzed MTA from 2011 and found that rain and snow had an adverse impact.

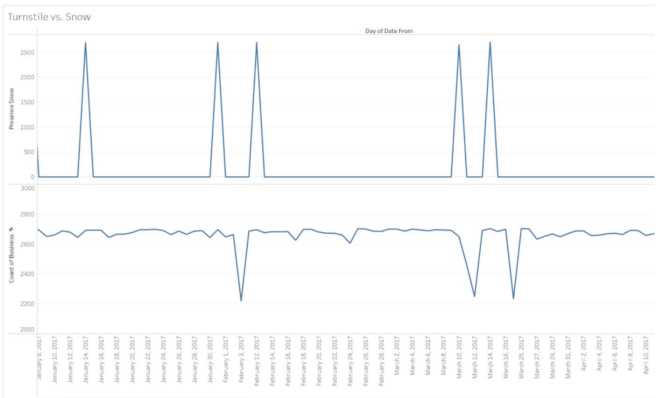

Chart 7: Subway Netflow vs. Snow

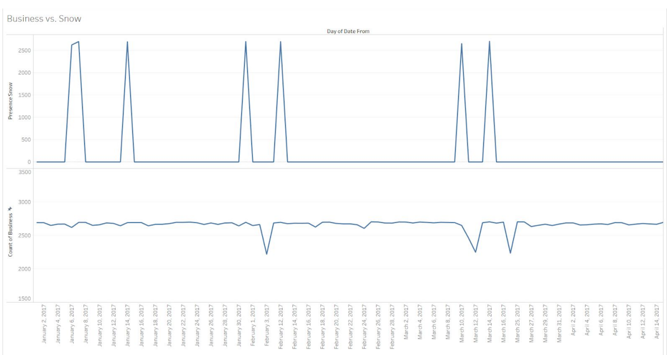

Chart 8: Subway Business vs. Snow

There wasn't a large number of snow days in 2017, only 12 days in total, and not much rain either (in total, the number of days it either rained or snowed in 2017 was 117). As a result, our analytic might be a bit skewed because of that. However, the base analyses with temperature yields solid results.

As an example application, we have also experimented with a naive linear regression model that predicts the expected number of green cab passengers at a given location and time under given weather conditions. This application serves more as a proof-of-concept, showing how the results of our analytic can be used to train a model and generate predictions. If improved and extended, something like this can be used by the MTA, ridesharing companies, or even regular New Yorkers in their daily commutes. The code for this application is also provided in our final code drop.

VI. Obstacles

During the course of our project, we ran into several obstacles. The first was the format of the NYC MTA data. Whereas the TLC dataset was formatted as PickupTime, Drop-off Time, ..., number of passengers, the MTA dataset was formatted as: Time-Of-Reading, …, CounterValue.

Counter value was an always an incrementing number, therefore, the counter could potentially read in the tens of millions. This results in anomalous events that happened last year also being an influencer to this year's results. Another challenge arose in finding the correct rows to subtract, and coming up with a partitioning scheme that fed all relevant keys to the same reducer in a MapReduce job.

When analyzing our final results, we tried to find some correlation between taxi and subway usage. However, this sort of correlation sort of assumes that if people don't ride the subway, they will take the taxi and vice-versa. As stated previously, our sample size for adverse weather conditions is also relatively small, so it hard to draw conclusions without making the assumption that the trends observed will continue in the future.

VII. Future Work

Given time, we would like to expand our analytic to look at more metrics of the data. For Taxi data, we could consider not only the relationship between weather conditions and passenger count, but also between weather conditions and trip distance and trip time. For the weather data, we could potentially not only look at average temperature and presence of precipitation, but also wind speed, presence of fog, presence of thunder, heaviness of rain/snow, and more.

Another interesting application of this analytic could be to further flesh out our predictive models. As we mentioned earlier, we experimented with generating predictions for Green Taxi usage based on a simple linear regression model, and think that there is potential to train a more complicated model to generate more accurate predictions that can be used by the MTA, ridesharing companies, or regular New Yorkers in their daily commutes.

VIII. Conclusions

After analyzing the relationships between weather conditions and Taxi/For-Hire-Vehicle ridership in 2017, we saw that neither temperature nor the presence of rain or snow affected Taxi/FHV usage to a significant degree. However, there were a few interesting findings, which are discussed in the Results section.

As for MTA usage, weather conditions in 2017 do influence the usage of the subway. On average, there was an increased usage with higher temperatures. Also, there were some fluctuations at the point of recorded snow/rainfall. This could exhibit the property of people avoiding stepping outside to prepare for the snow, proceeded by the vast usage of public instead of private transport.

IX. Acknowledgements

We would like to thank (in no particular order):

- NYC Taxi and Limousine Commission for providing taxi and rideshare data.
- NYC MTA for providing turnstile and subway data.
- NOAA for providing weather data.
- NYU HPC for providing their support and computing resources.
- Professor McIntosh for introducing the tools and techniques needed to perform this analytic.

X. References

1. Kamga, Camille & Yazici, M. Anil & Singhal, Abhishek. “Hailing in the Rain: Temporal and Weather-Related Variations in Taxi Ridership and Taxi Demand-Supply Equilibrium”. https://www.researchgate.net/profile/M_Anil_Yazici/publication/255982467_Hailing_in_the_Rain_Temporal_and_Weather-Related_Variations_in_Taxi_Ridership_and_Taxi_Demand-Supply_Equilibrium/links/00b4952cb68b74cfd3000000.pdf
2. M. A. Yazici, C. Kamga and A. Singhal. “A big data driven model for taxi drivers' airport pick-up decisions in New York City”. https://ieeexplore.ieee.org/abstract/document/6691775/?reload=true
3. Singhvi, Divya, Somya Singhvi, Peter I. Frazier, Shane G. Henderson, Eoin O'Mahony, David B. Shmoys and Dawn B. Woodard. “Predicting Bike Usage for New York City's Bike Sharing System”. https://www.aaai.org/ocs/index.php/WS/AAAIW15/paper/viewFile/10115/10185
4. Chang, H-w., Tai, Y-c., and Hsu, J.Y-j. “Context-aware taxi demand hotspots prediction”. http://agents.csie.ntu.edu.tw/system/papers/16/original/ContextAwareTaxiDemandHotspotsPrediction.pdf
5. Wallsten, Scott. “The Competitive Effects of the Sharing Economy: How is Uber Changing Taxis?”. https://techpolicyinstitute.org/wp-content/uploads/2015/06/the-competitive-effects-of-the-2007713.pdf
6. Biju Mathew. “Taxi! – Cabs and Capitalism in New York City”. https://books.google.com/books?hl=en&lr=&id=atYx81JQ2K0C&oi=fnd&pg=PR7&dq=nyc+taxi+weather&ots=wVXi3LNCIl&sig=jyQDppPPP6BHojVw5ZCHnukCDFA#v=onepage&q=nyc%20taxi%20weather&f=false
7. Jun Xu, Rouhollah Rahmatizadeh, Ladislau Bölöni, Damla Turgut. “Real-Time Prediction of Taxi Demand Using Recurrent Neural Networks”. https://ieeexplore.ieee.org/abstract/document/8082792
8. Anastasios Noulas, Vsevolod Salnikov, Renaud Lambiotte, Cecilia Mascolo. “Mining open datasets for transparency in taxi transport in metropolitan environments”. https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-015-0060-2
9. Xiujuan Xu, Benzhe Su, Xiaowei Zhao, Zhenzhen Xu, Quan Z. Sheng. “Effective Traffic Flow Forecasting Using Taxi and Weather Data”. https://link.springer.com/chapter/10.1007/978-3-319-49586-6_35
10. Fangru Wang, Catherine L. Ross. “New potential for multimodal connection: exploring the relationship between taxi and transit in New York City (NYC)”. https://link.springer.com/article/10.1007/s11116-017-9787-x
11. Chengxi Liu. “Understanding the Impacts of Weather and Climate Change on Travel Behaviour”. https://www.diva-portal.org/smash/get/diva2:928565/FULLTEXT01.pdf
12. Snigdha Gutha, Anusha Mamillapalli. “Analyzing the effect of Weather on Uber Ridership”. https://support.sas.com/resources/papers/proceedings17/1260-2017.pdf
13. Josh Grinberg, Arzav Jain, Vivek Choksi. “Predicting Taxi Pickups in New York City”. http://www.vivekchoksi.com/papers/taxi_pickups.pdf
14. Dandan Chen, Yong Zhang, Liangpeng Gao, Nana Geng, Xuefeng Li. “The impact of rainfall on the temporal and spatial distribution of taxi passengers”. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0183574
15. Sui Tao, Jonathan Corcoran, Francisco Rowe, Mark Hickman. “To travel or not to travel: ‘Weather’ is the question. Modelling the effect of local weather conditions on bus ridership”. https://www.sciencedirect.com/science/article/pii/S0968090X1730311X
16. Abhishek Singhal, Camille Kamga, Anil Yazici. “Impact of weather on urban transit ridership”, Transportation Research Part A: Policy and Practice, Volume 69, 2014, Pages 379-391, http://www.sciencedirect.com/science/article/pii/S0965856414002195
17. Schneider, Todd. “Analyzing 1.1 Billion NYC Taxi and Uber Trips, with a Vengeance.” Todd W. Schneider, November 2015, http://toddwschneider.com/posts/analyzing-1-1-billion-nyc-taxi-and-uber-trips-with-a-vengeance/#taxi-weather