SAS® GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL

Effect of Weather on Uber Ridership

Anusha Mamillapalli, Snigdha Gutha
MS in Business Analytics, Oklahoma State University

USERS PROGRAM

Effect of Weather on Uber Ridership

Anusha Mamillapalli, Snigdha Gutha

MS in Business Analytics, Oklahoma State University

ABSTRACT

Uber has changed the face of taxi ridership, making it more convenient and comfortable for riders. But, there are times when customers are left unsatisfied because of shortage of vehicles which ultimately led to Uber adopting surge pricing. It's a very difficult task to forecast number of riders at different locations in a city at different points in time. This gets more complicated with changes in weather. In this paper we attempt to estimate the number of trips per borough on a daily basis in New York City. We add an exogenous factor, weather to this analysis to see how it impacts the changes in number of trips. We fetched six month worth data (approximately 1 million records) of Uber rides in New York City ranging from January 2015 to June 2015 from github. We also gathered weather data (from Weather Underground) of New York City Borough wise for the same period of six months from Jan 2015 to June 2015. In this poster, we attempted to analyze Uber data and weather data together to estimate the change in the number of trips per borough due to changing weather conditions. We used SAS® Forecast Studio and built a model to predict the number of trips per day for the one week ahead forecast for each borough of the New York City.

METHODS

Data Preparation and Analysis

We merged three datasets namely Uber_PickUp_data, Taxi_LookUp_Zone and Weather data to obtain the final dataset as shown in Fig. 1. In the first step, we combined Uber_PickUp_Data and Taxi_Lookup_Zone datasets to bring uber rides to borough level (Merged1 in Fig. 1). We considered Uber rides information for five boroughs of New York City: Manhattan, Brooklyn, Bronx, Queens and Staten Island for our analysis. Final dataset Uber_Final is obtained by merging Merged1 dataset with Weather dataset. Information about the various weather metrics taken for analysis are listed in the Table 1.

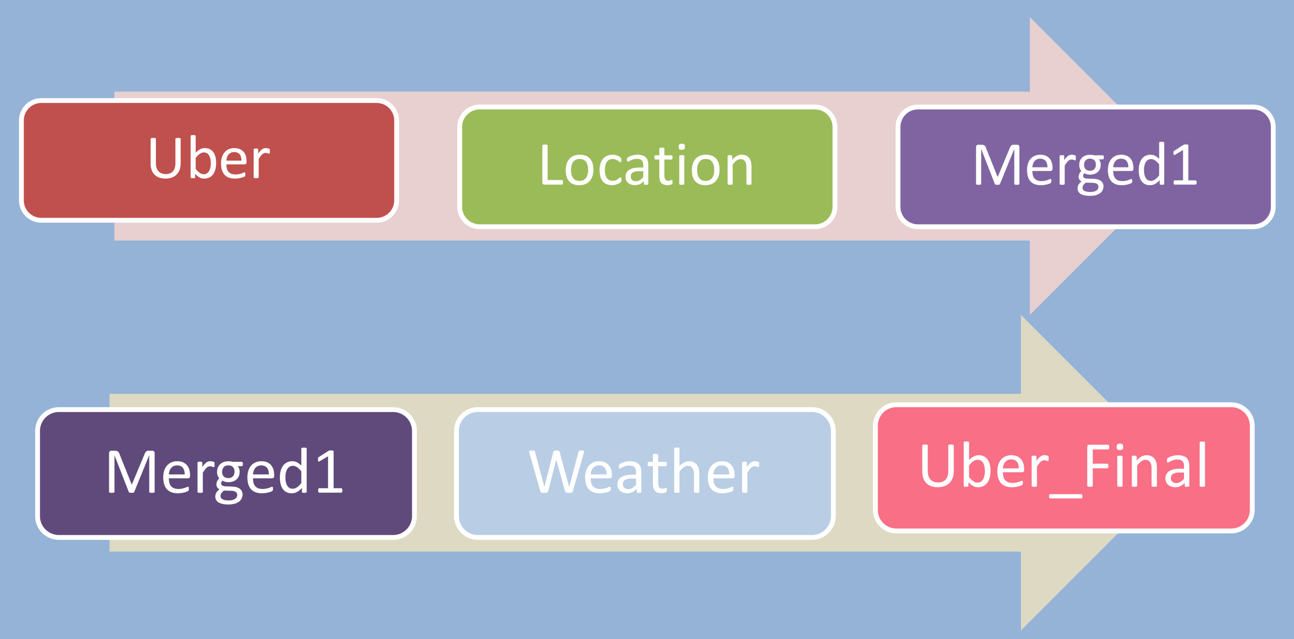


Fig. 1 (Data Preparation)

Dataset	Variables
Uber_PickUp_data	Dispatching_Base_Num, Date_of_Pickup,Time, Affiliated_Base_Num, LocationID
Taxi_Lookup_Zone	Location_ID, Borough, Zone
Weather	Date ,Temperature, Humidity, Sea_level_pressure, Precipitation (Max, Mean and Min), Cloud_Cover, Wind, Event

Table 1

RESULTS (CLICK TO EDIT)

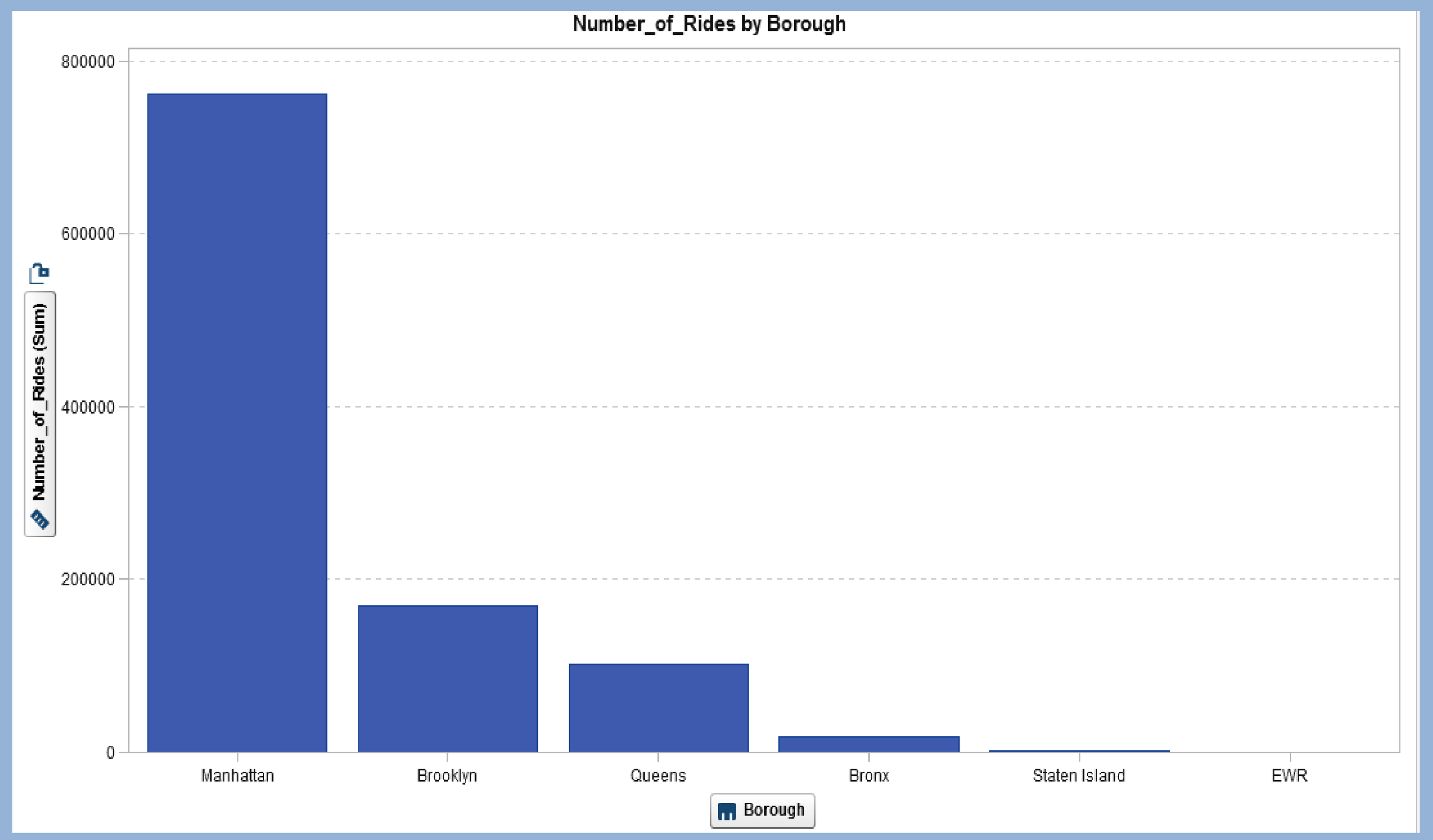


Fig. 2: Number of Rides by Borough

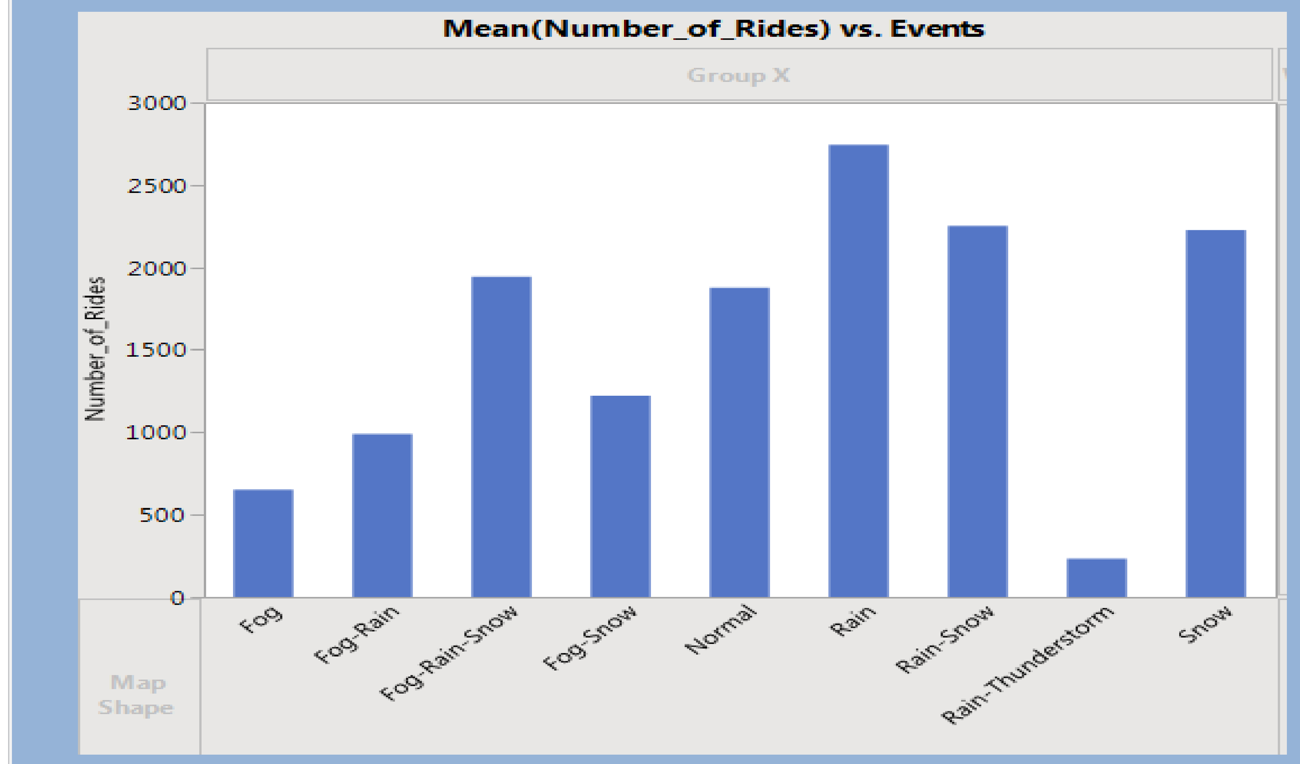


Fig. 3: Average Number of Rides by Event

The Initial exploratory analysis revealed following insights:
Manhattan has the largest user base being the most populous county and financial capital of world followed by Brooklyn and Queens. On an average a rainy day has reported more number of rides followed by snow than a normal day.

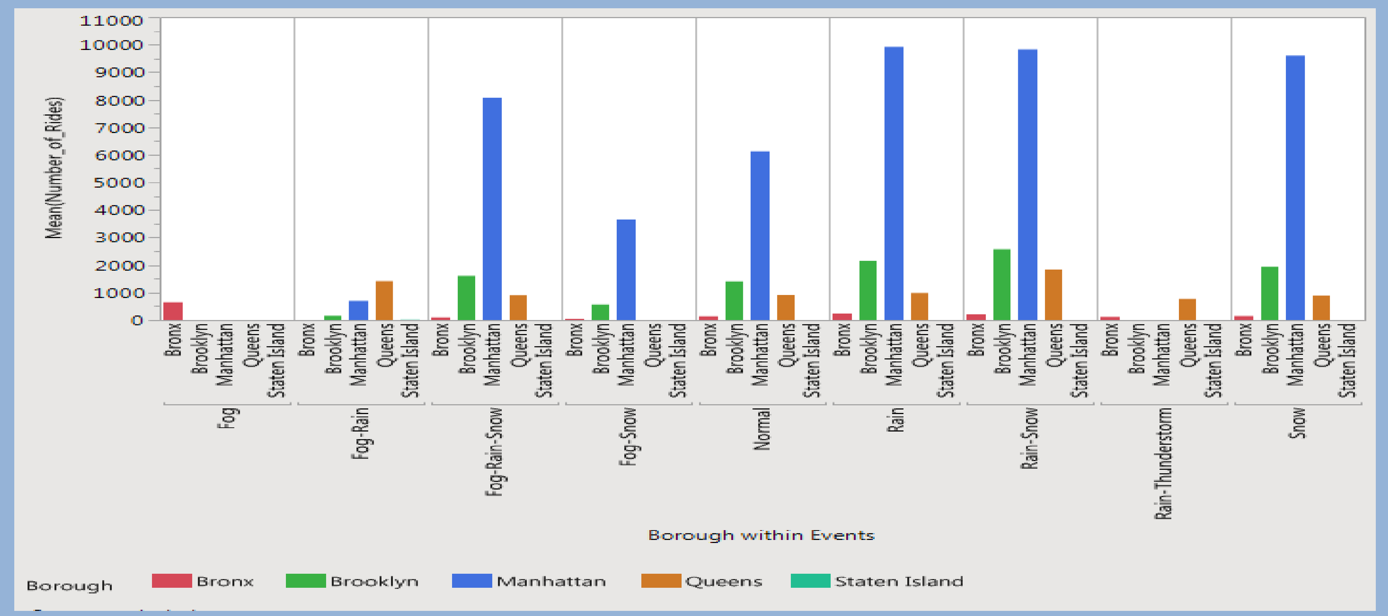


Fig. 4: Average Number of Rides by Borough and Event

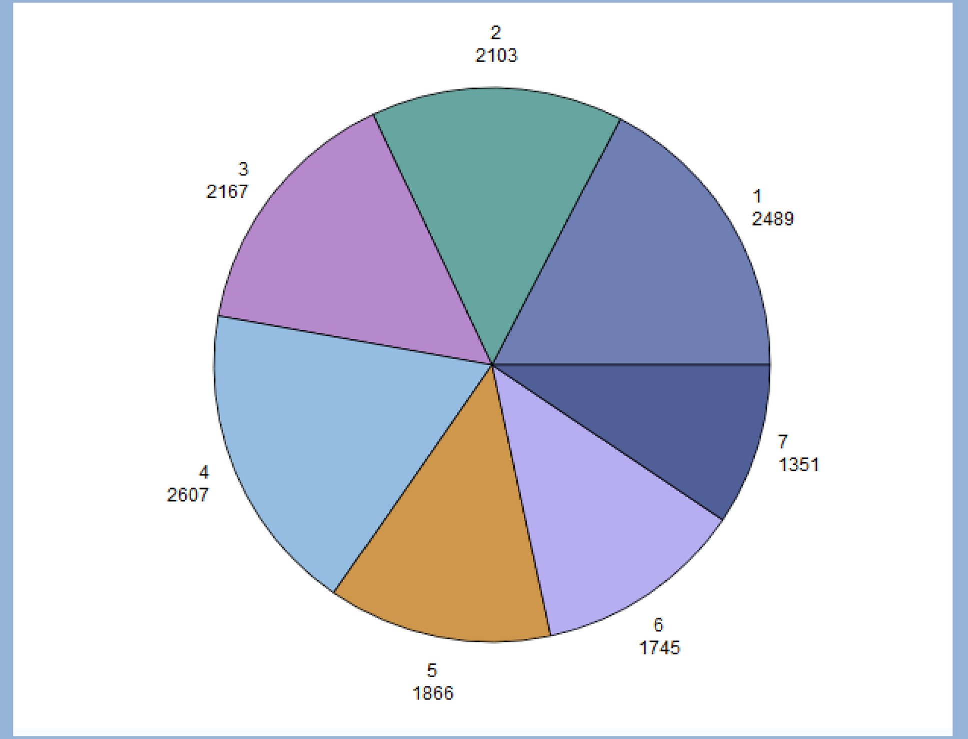


Fig. 5: Mean Number of Rides by Day of Week

Fog seems to impact the number of rides in Bronx more than any other borough. Fog-Rain and Rain-Thunderstorms are influencing the number of rides in Queens the most. Rain and Snow is increasing the number of rides in Manhattan more than any other event. Considering all five boroughs, Wednesdays are the busiest day representing mid week followed by Sunday (weekend). Saturdays have the least turnout.

Effect of Weather on Uber Ridership

Anusha Mamillapalli, Snigdha Gutha

MS in Business Analytics, Oklahoma State University

Exploratory Analysis(Contd….)

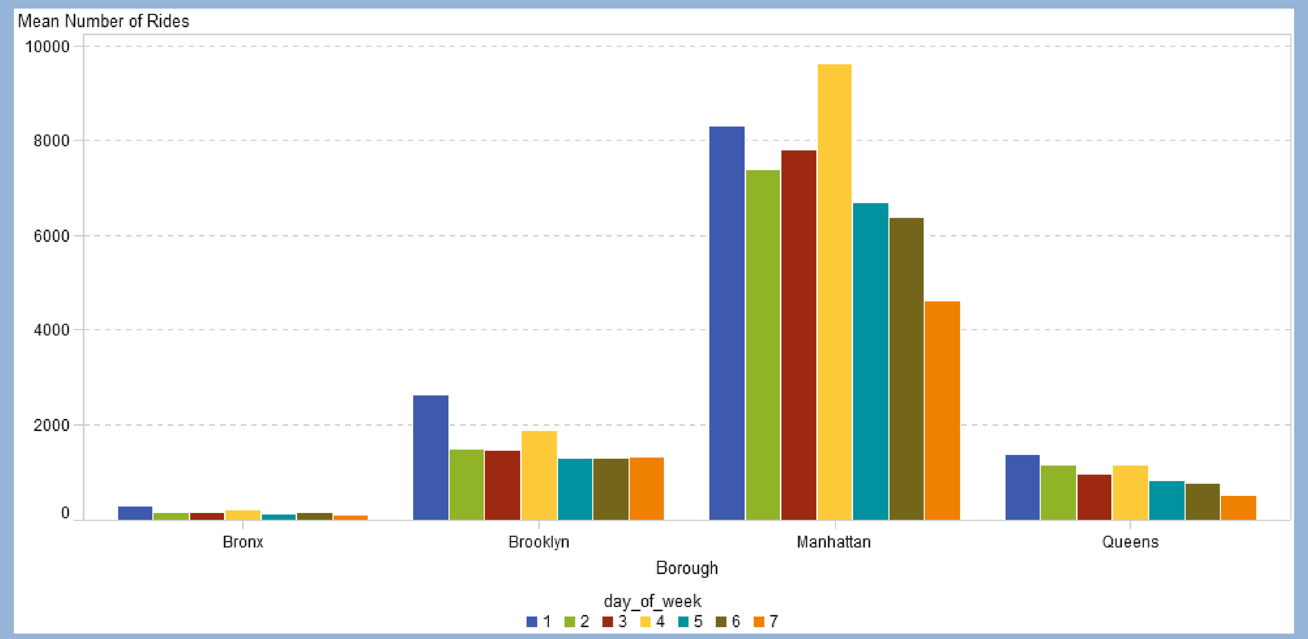


Fig. 6: Mean Number of Rides by Borough by Day of Week

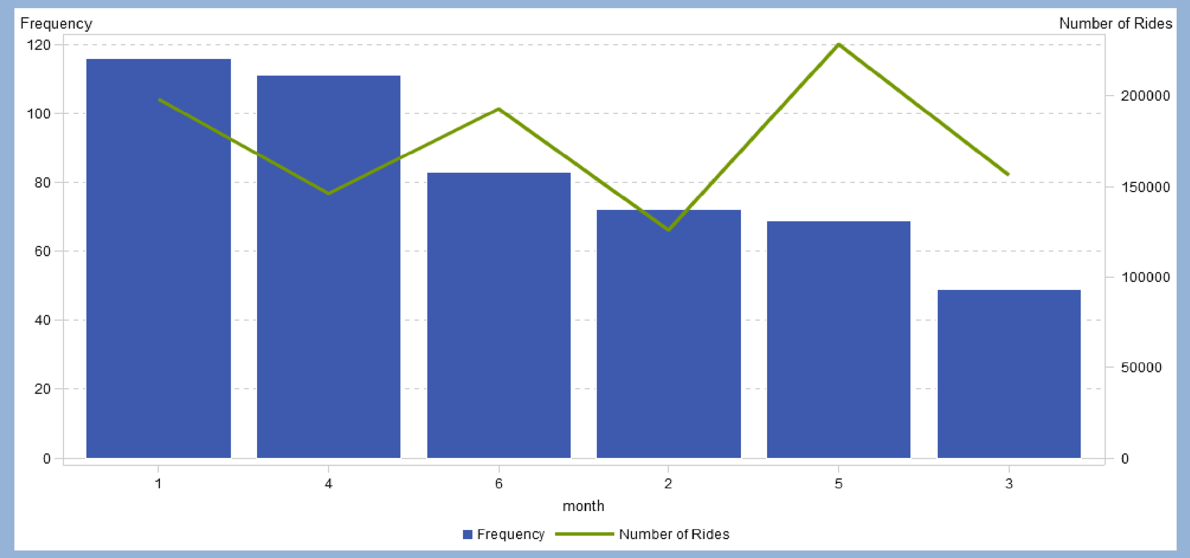
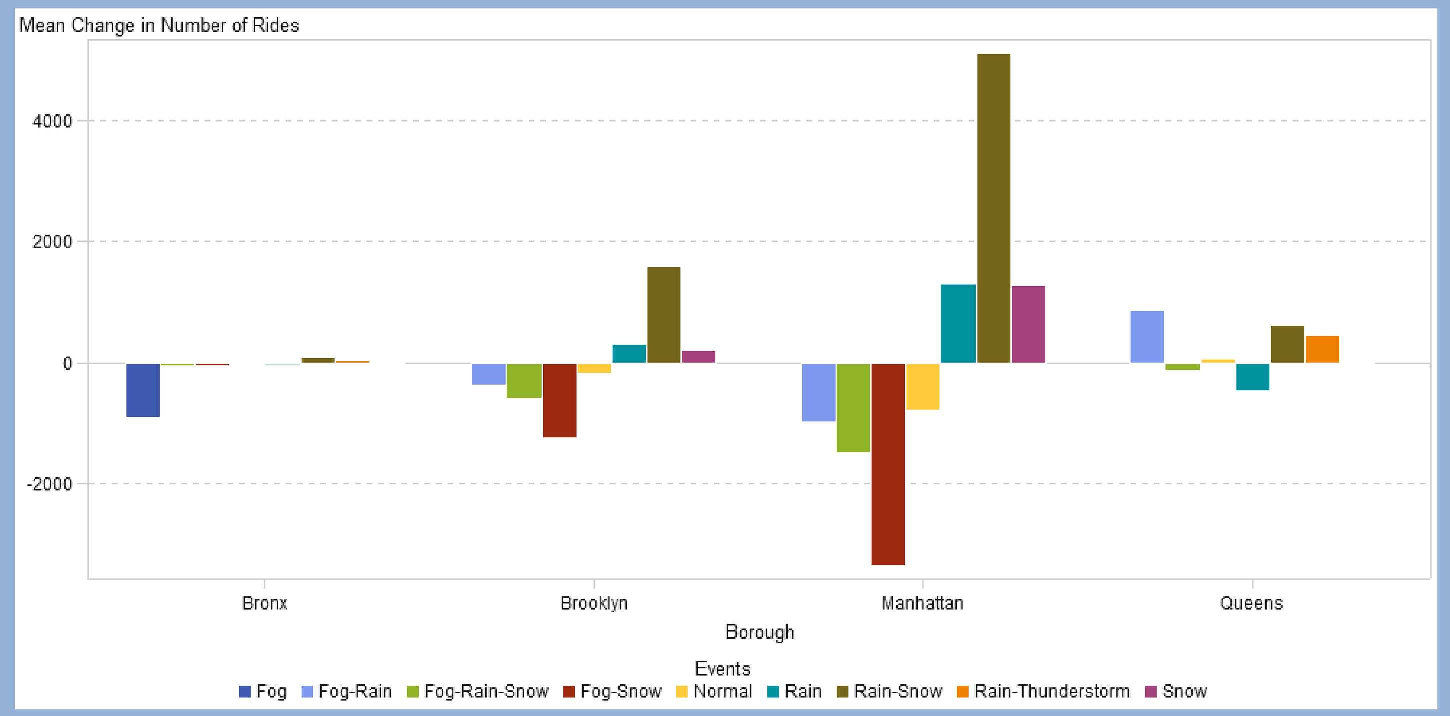


Fig. 7: Number of Rides by Month

Wednesdays are the busiest day for Manhattan followed by Sunday. Whereas Sundays turn out to be the busiest day in the other boroughs. The month of May shows the peak number of rides when compared with the other months. April has the least number of rides.



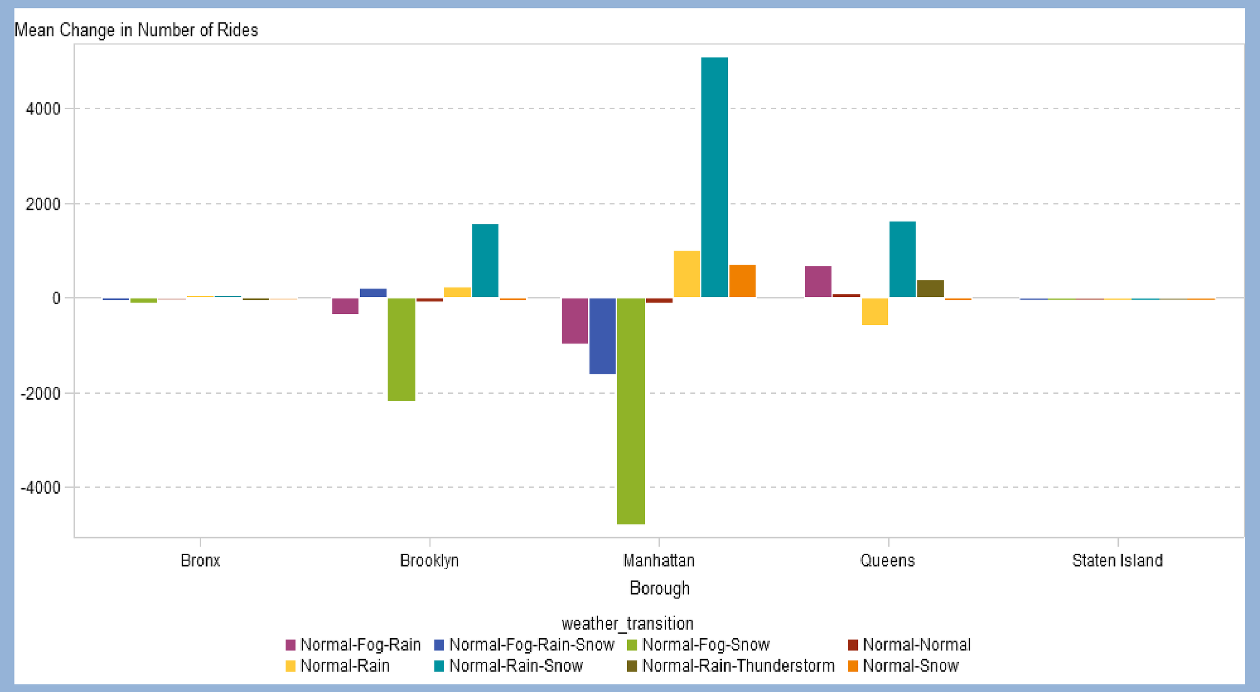


Fig. 8: Mean Change in Number of Rides by Borough by Event

Fi.g 9: Mean Change in Number of Rides by Borough by Weather Transition

Mean change in number of rides is highest on a Rainy day with Snow in every borough except Bronx. In Bronx the change in number of rides is highest on a foggy day. On an average change in number of rides is most prevalent when there is a transition of weather from normal to rain-snow or fog-snow.

Exploratory Analysis(Contd….)

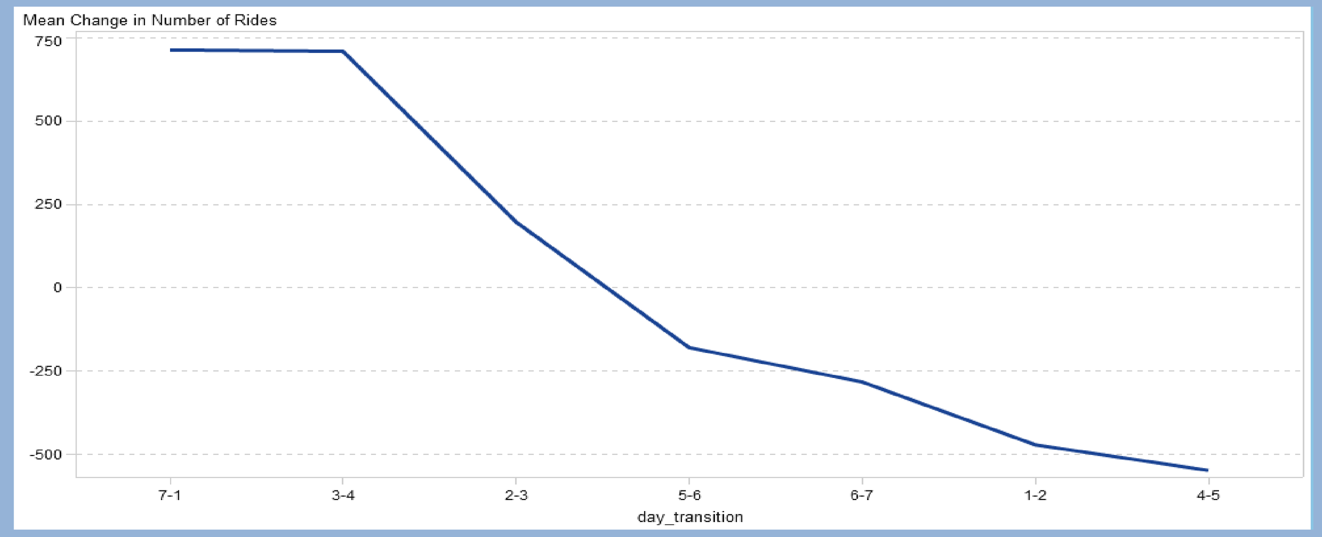


Fig. 10: Mean Change in Number of Rides by Day Transition

Mean change in number of rides is highest when compared between Saturday vs. Sunday and Tuesday vs. Wednesday. There is not much change between Wednesday and Thursday

Data Cleaning and Transformation

Uber_Pickup_Data has ridership information missing for certain days. We used average of immediate previous interval non missing value and immediate next interval non missing values to fill in the missing values. We used SAS Enterprise Guide's create time series data node to produce the final time series data on a daily level.

We used SAS Enterprise Miner's Transform Variables node to transform number of rides and weather metrics to bring the data to normality.

Time Series Forecasting

- We used SAS Forecast Studio for Desktop 14.1 version to perform time series forecasting.
- We considered one month of rides (June) as the holdout sample and borough as a by group variable to generate forecasts at borough level.
- Set the cutoff to 2% for the detection of outliers.
- ARIMA Model(Top_1) has been chosen as the best model based on holdout MAPE. Mean Temperature with a lag(1), Mean Visibility Miles, Wind direction, cloud cover and sea pressure turned out to be significant in predicting the number of rides.

Effect of Weather on Uber Ridership

Anusha Mamillapalli, Snigdha Gutha

MS in Business Analytics, Oklahoma State University

Time Series Forecasting Configuration

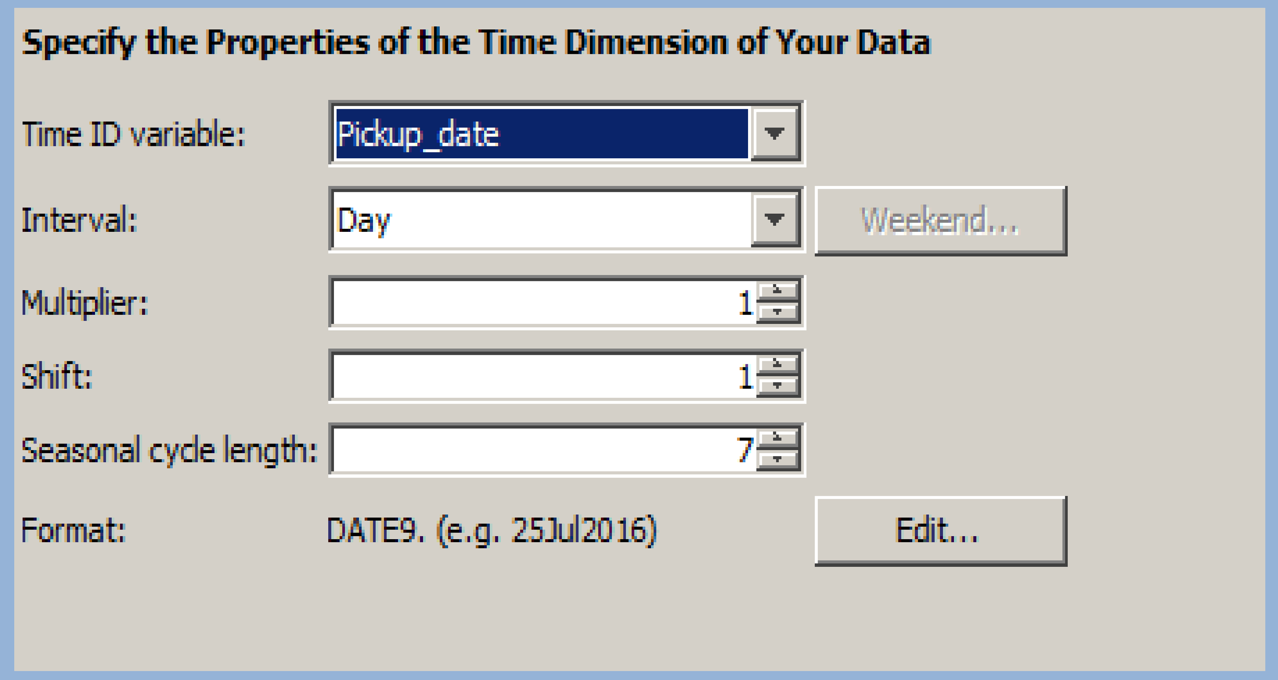


Fig. 11: Seasonality Configuration

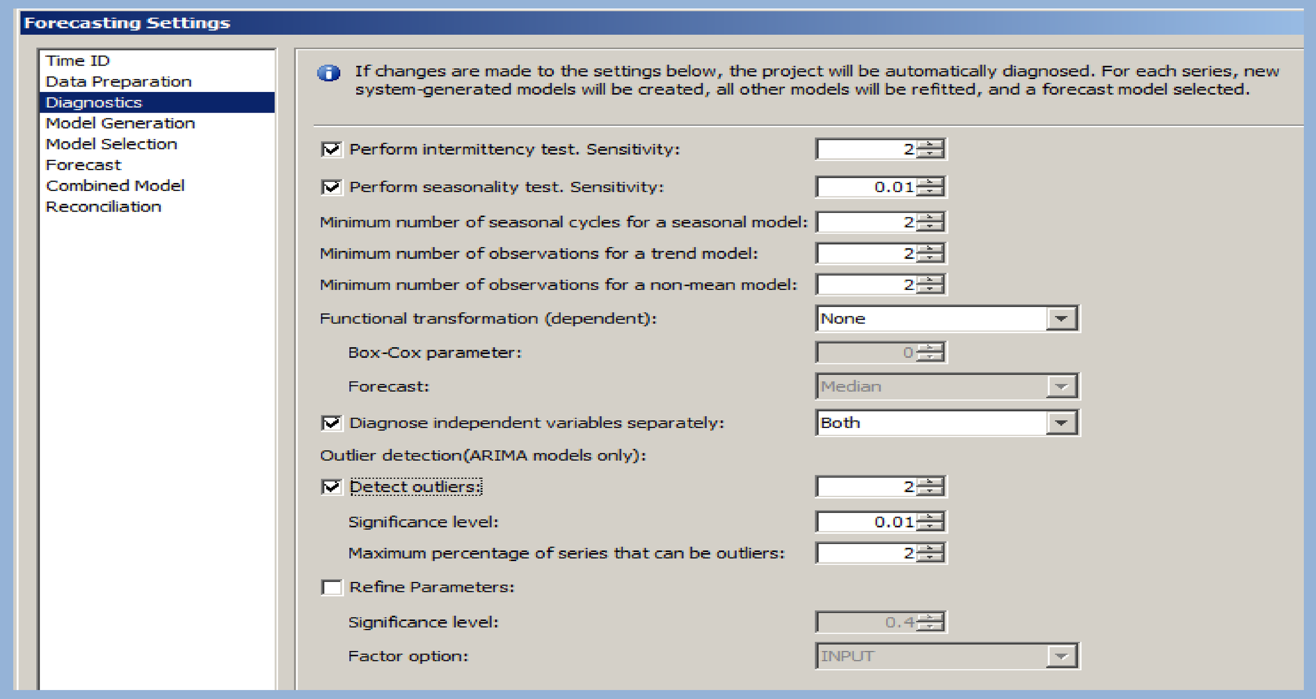


Fig. 12: Diagnostic Configuration

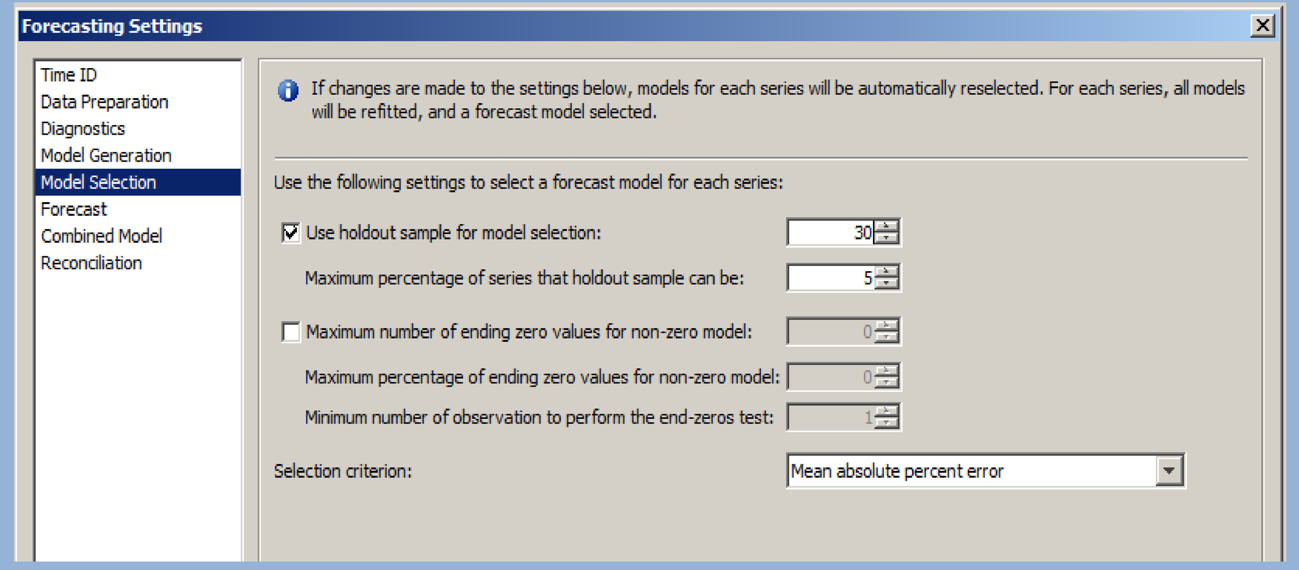


Fig 13: Hold out sample selection

BOROUGH	MAPE	Rec. MAPE
Manhattan	5.93	7.08
Queens	9.16	8.94
Brooklyn	10.89	9.14
Bronx	14.97	14.00
Staten Island	21.68	21.71

Fig 14: MAPE values for models at Borough levels

ARIMA Models selected at different borough levels

Bronx: Log_number_of_rides~ 2+ Lag(7)Mean Temperature+ Lag(1)Exp_Mean_VisibilityMiles

Brooklyn: Log_number_of_rides~2+ Lag(10)CloudCover+ Lag(1)Exp_Mean_VisibilityMiles+ Lag(1)Sqrt_Mean_Humidity+ Lag(7)Sqr_WindDirDegrees

Manhattan: Log_number_of_rides~2+ Lag(7)Sqr_windDirDegrees+ Lag(1)Exp_Mean_VisibilityMiles+ AO10MAR2015D+AO21APR2015D

Queens: Log_number_of_rides~Lag(11)Dif(7)Mean_Sea_Level_PressureIn+ AO21APR2015D+AO10MAR2015D

Staten Island : Log_number_of_rides~ Dif(1) Mean_TemperatureF+ Dif(1)Mean_Dew_PointF+ Dif(1)CloudCover+ Dif(1) SQRT_Mean_Wind_SpeedMPH+ Dif(1) SQR_WindDirDegrees

Diagnostic Plots

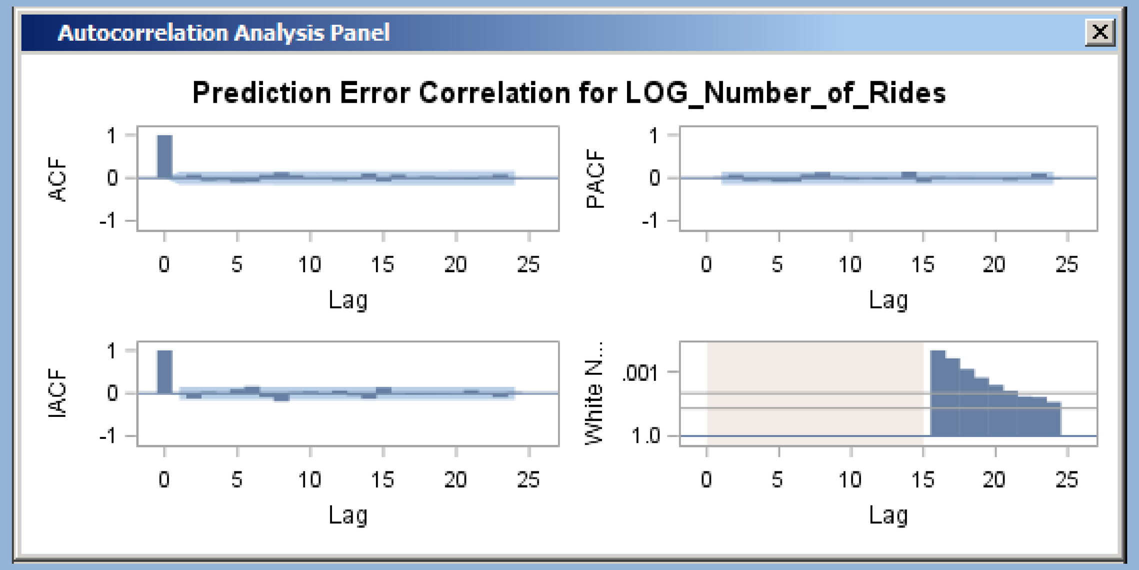


Fig. 14: Diagnostic Plots

There appears to be no significant correlation in the residuals.

Forecast Results

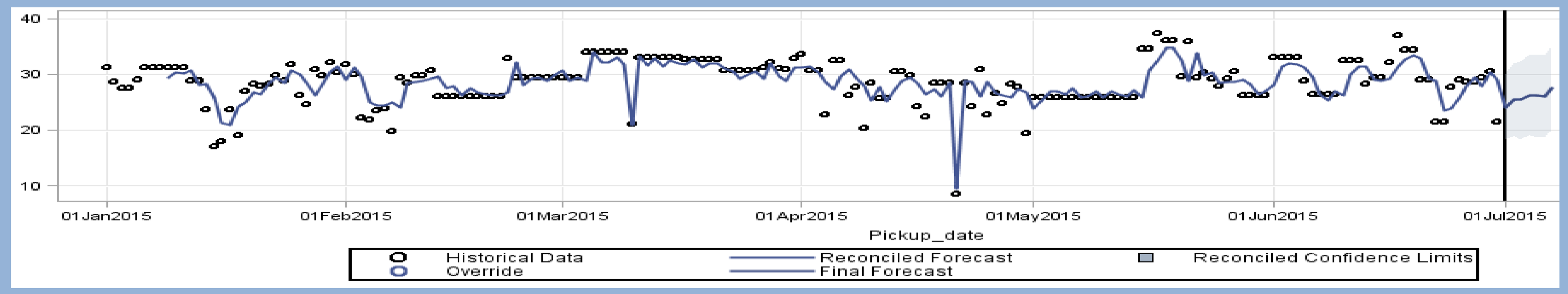


Fig. 15: Forecast Output

Pickup_date	number_of_rides
01JUL2015	35
02JUL2015	58
03JUL2015	73
04JUL2015	111
05JUL2015	81
06JUL2015	65
07JUL2015	65

Bronx Forecasted Rides

Pickup_date	number_of_rides
01JUL2015	423
02JUL2015	378
03JUL2015	433
04JUL2015	476
05JUL2015	517
06JUL2015	565
07JUL2015	1331

Brooklyn Forecasted Rides

Pickup_date	number_of_rides
01JUL2015	360
02JUL2015	575
03JUL2015	384
04JUL2015	359
05JUL2015	372
06JUL2015	410
07JUL2015	280

Queens Forecasted Rides

Effect of Weather on Uber Ridership

Snigdha Gutha, Anusha Mamillapalli

MS in Business Analytics, Oklahoma State University

Forecast Results

Pickup_date	number_of_rides
01JUL2015	996
02JUL2015	1672
03JUL2015	1722
04JUL2015	2545
05JUL2015	2623
06JUL2015	2420
07JUL2015	5124

Manhattan forecasted Rides

Pickup_date	number_of_rides
01JUL2015	4
02JUL2015	5
03JUL2015	4
04JUL2015	5
05JUL2015	5
06JUL2015	5
07JUL2015	6

Staten Island forecasted Rides

Conclusion

- The above results show that weather does influence Uber ridership, especially when it is little distracted from normal but not too extreme again. Manhattan being the most populous city is bringing a huge variation in the number of rides especially on a rainy day.
- The demand is highest during Wednesdays in Manhattan and Sundays in other boroughs.
- Staten Island has the least Uber user base, Hence the Mean Absolute Percent Error turned out to be highest as there are not many rides to predict.

Future Work

The scope of the project can be extended to hourly analysis. Analyzing at least a year worth of data will bring further insights. Demand can be more accurately predicted if the actual number of rides requested information is available along with the live number of rides.

References

http://toddwschneider.com/posts/a-tale-of-twenty-two-million-citi-bikes-analyzing-the-nyc-bike-share-system/%20-%20citibike-weather

https://newsroom.uber.com/uberdata-uber-for-style-and-comfort/

Acknowledgement

We thank Dr. Goutam Chakraborty, Professor, Department of Marketing, Director of Business Analytics program-Oklahoma State University for his continuous support and guidance.

SAS® GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL