



# Uber Service in NYC

A study on what affects Uber ridership in NYC

By: Jorge Londono, 3/2/19

# The Problem

- NYC Public transit is used by millions of people to commute every day
- Private companies such as Uber provide services that are more expensive than public transportation. Why are people paying more for these services?

# The Problem (2)

- Are there any factors that we can measure to see what specifically can make Uber ridership increase or decrease

# Data Set Components

- The data set used to explore this problem is a combination of 3 data sets:
  - ➔ Uber pickups by time and date from two ranges in 2014 and 2015
  - ➔ Weather data( conditions, amount of precipitation, temperature) for specific hours and dates.
  - ➔ Holiday dates from the 2010-2020.

# Methods of analysis

- With this data set we can analyse this data through:
  - Statistical measurements (medians, averages)
  - Data visualizations
  - An algorithm that we can use to see what factors can predict Uber ridership

# Data Wrangling

- The one thing all of the 3 separate sets having in common is the date. So when we unite all of these data sets all of the observations must abide by the correct date
- A binary weekend variable (was it a weekend or not) was created to help us explore the effects weekends have on Uber Ridership

# Data Wrangling (2)

- Sometimes the weather conditions had multiple descriptors ( both foggy and rainy)
- To fix this, we measure the mean ridership of each of the two components ( mean ridership of foggy weather, and mean ridership of rainy weather).
- Then we reclassify the variable to whatever its mean ridership is closest to

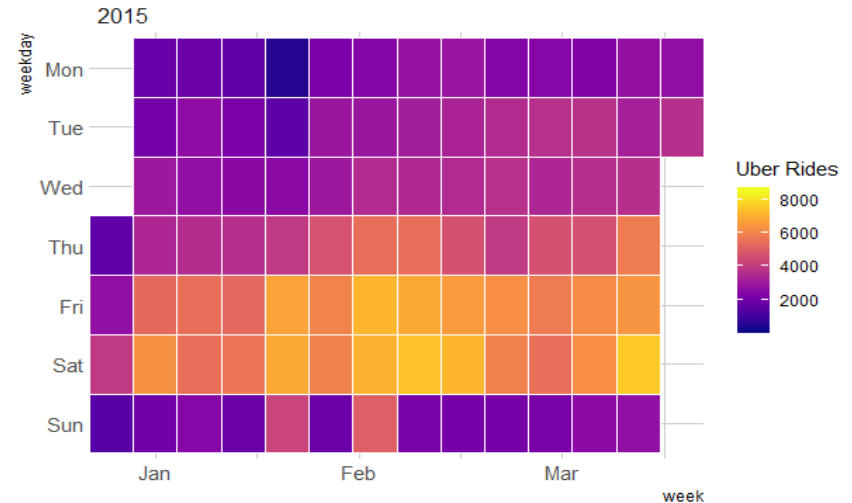
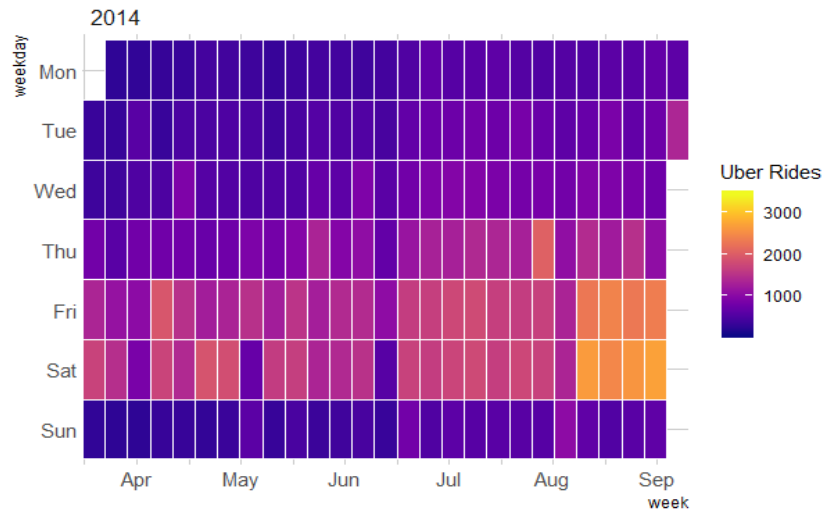


# Findings: Holidays

- Holidays had a dramatic effect on ridership. For example, the average daily rides for the month of April in 2014 was 784, while the daily total for Good Friday in 2014 was 18,074.

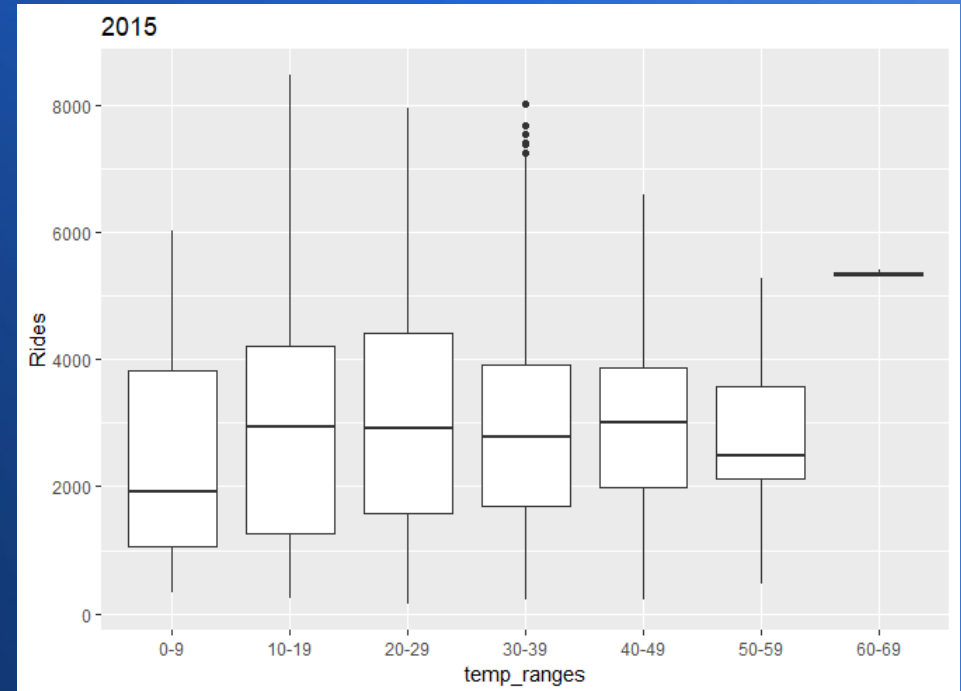
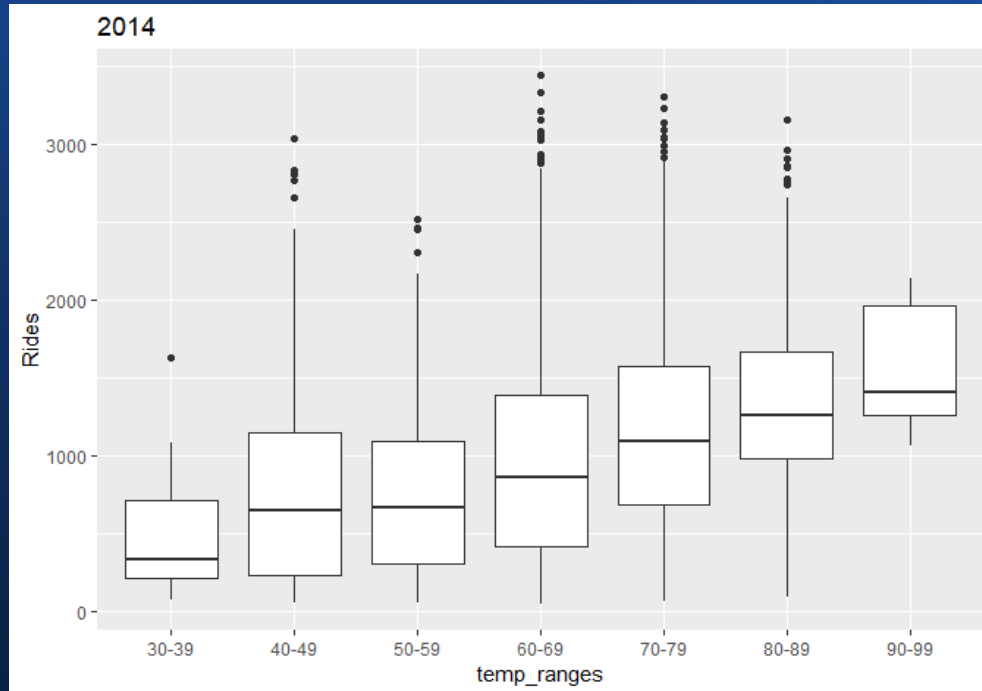


# Findings: Days of the Week



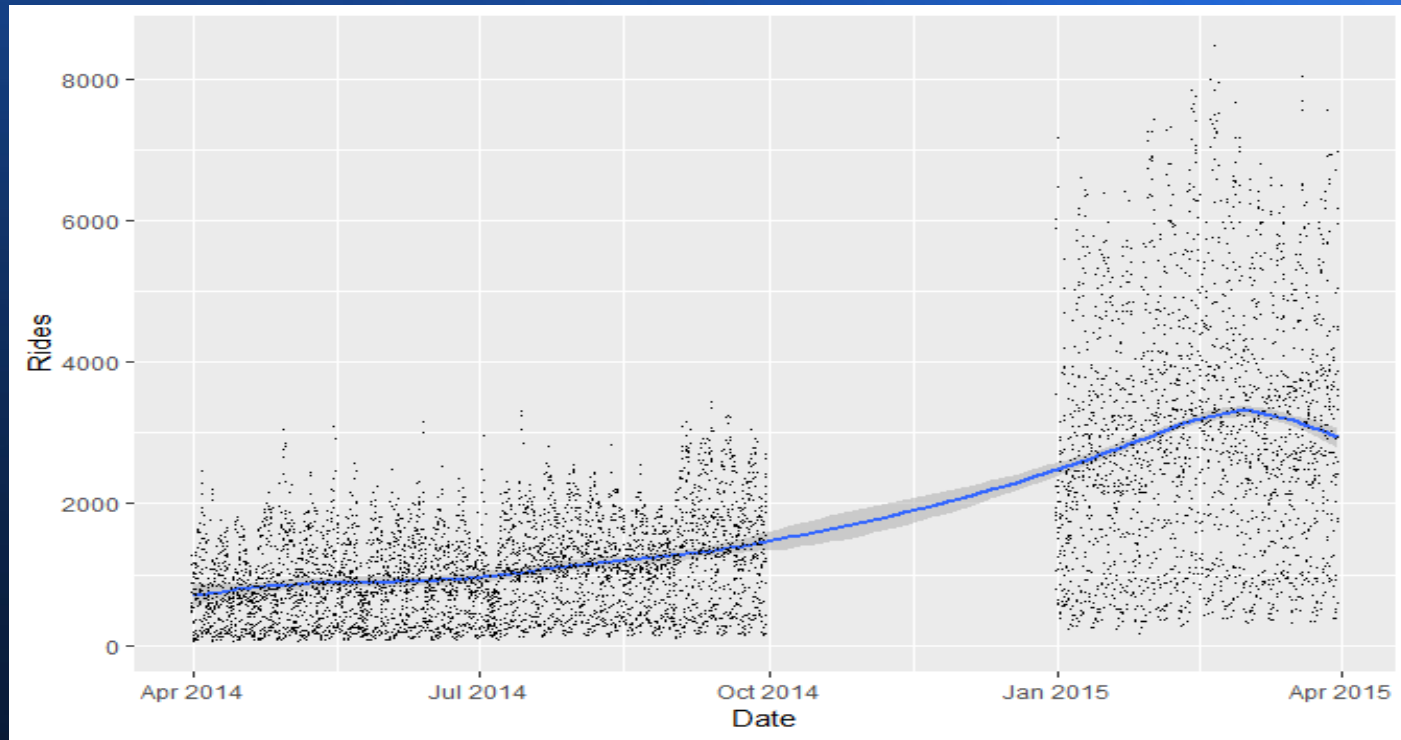
- There is increased Ridership on Thursdays, Fridays, and Saturdays

# Findings: Temperature



- When it was colder out, ridership was low, and as temperatures rose, so did the ridership.

# Findings: Trend Over Time



- As times goes on there is an increase in ridership

# Machine Learning

- Since ridership is a continuous variable, I will be using a supervised linear regression model to predict the rides per given:

- Date
- Hour
- Temperature
- Holiday (yes or no)
- Weekend (yes or no)
- Precipitation
- Weather Condition

# Training and Testing Sets

- The data for this algorithm will be trained on our data set which has the ranges of :

April 1st – September 30th 2014 and January 1st – March 31st 2015

- The data for this algorithm will be tested on a data set which has the range of :

April 1st – July 30th 2015

# Results

- With the training set, the linear regression model had an adjusted R-squared value of 0.7874.
- The coefficient with the largest absolute t - value is the Date (65.406)
- The hour with the largest absolute t value is 18:00 or 6PM (30.864)
- The weather condition with the highest absolute t value is clear weather (3.503)

## Results (2)

- With the testing set, my linear regression model had an R - Squared value of 0.8756. This value shows a fairly strong relationship between the features of the model and Uber Ridership.
- The Root Mean Square Error value while using the testing set is 0.1785.



# Conclusion & Recommendations

Uber should be able to specifically increase services on:

- Holidays
- Clear Weather
- 6:00 PM

# Areas of Further Study

- How do the features in the linear regression model relate to public transportation usage?
- What areas of New York City experience more Uber pickups?
- How do MTA delays and service changes affect Uber ridership?