Uber: Determine best location for drivers

Objective

Predict the best stop locations for the drivers to be close to possible clients, by applying exploratory Data Analysis techniques to explore the factors that affect Uber cars demand in NYC, we can predict the best location for the driver to stop by to get a better request with better money offer without wasting the car fuel by locating the cars close to the location of possible users and clients.

Data Set

The main dataset contained over 10 million observations of 4 variables which aggregated per hour and borough, and then joined with the rest of the datasets lots of observations across 19 variables. These are:

- pickups_dt: Time period of the observations.
- boroughs: NYC's borough.
- pickups: Number of pickups for the period.

- spd: Wind speed in miles/hour.
- vsb: Visibility in Miles to nearest tenth.
- temp: temperature in Fahrenheit.
- dewp: Dew point in Fahrenheit.
- slp: Sea level pressure.
- pcp01: 1-hour liquid precipitation.
- pcp06: 6-hour liquid precipitation.
- pcp24: 24-hour liquid precipitation.
- sd: Snow depth in inches.
- hday: Being a holiday (Y) or not (N).
- class: ridership class (economy-premium-accessibility-carpool)
- clients_ids : clients' ids
- clients_names : clients' names
- clients_genders: genders of clients who sent the request
- drLoc: driver location when receiving the request
- clients_num: clients number in the same ride.
- arrv_time: time the driver took to arrive to the client.

Data Analysis

1) Drop unnecessary columns:
    a. clients_names
    b. clients_genders

Note: after analysis we can find more unnecessary data such as the data affects weather.

2) The analysis is built on the following sections:

- Univariate Plots
- Univariate Analysis
- Bivariate Plots
- Bivariate Analysis
- Multivariate Plots
- Multivariate Analysis
- Final Plots and Summary
- Reflections

Reflection

The dataset used for this project included data of Uber cars' ridership in the city of New York for the first six months of 2015. As exploring it, it is noticed that, against the initial intuition, the weather variables had not any or very weak impact on the ridership.
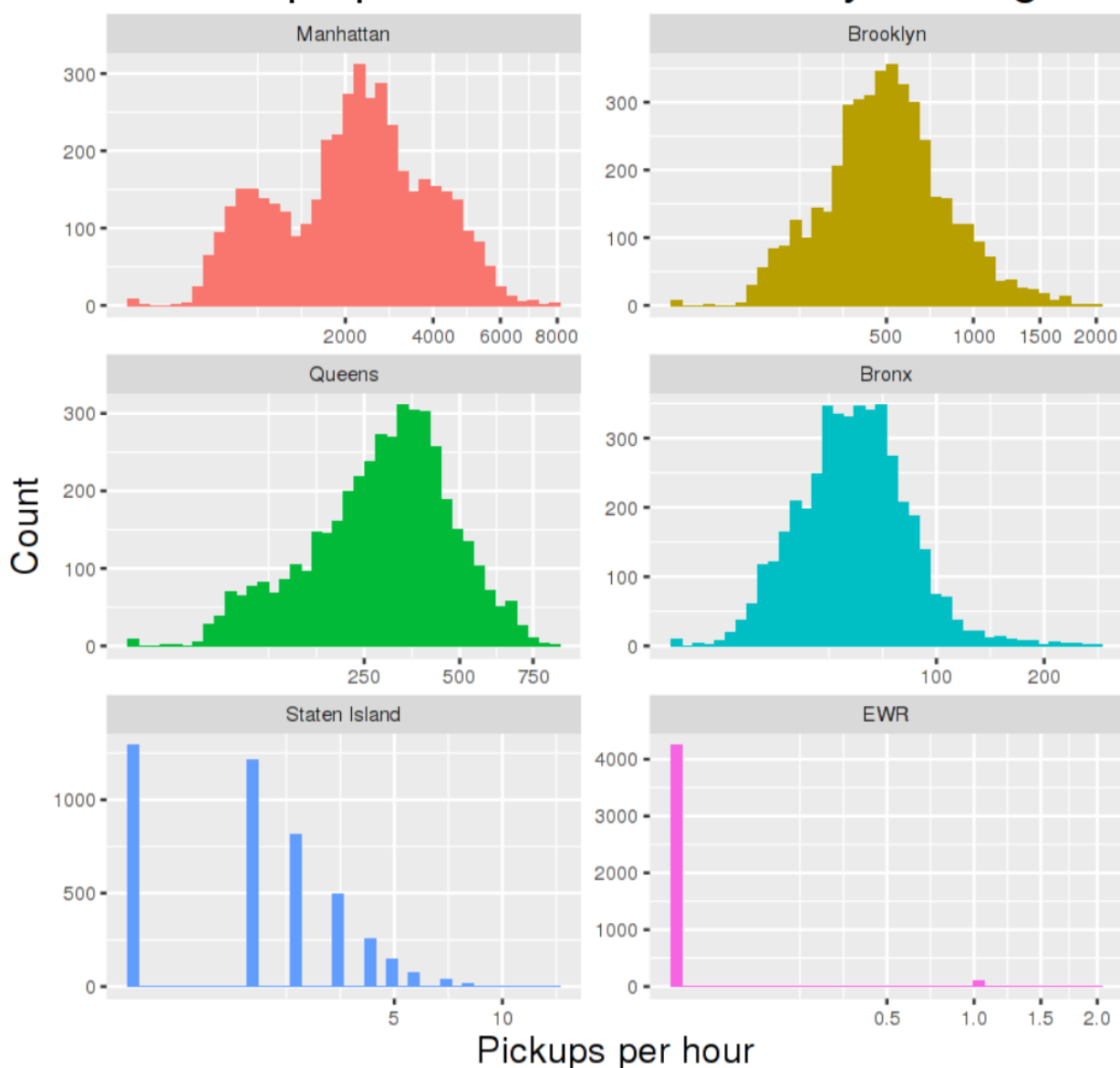Going further in analysis it was clear that the demand follows specific patterns both during the day and during the week.
Also, it is noticed that a general trend of rising demand during the six months, led the total demand from 2,000 pickups per hour to 3,500
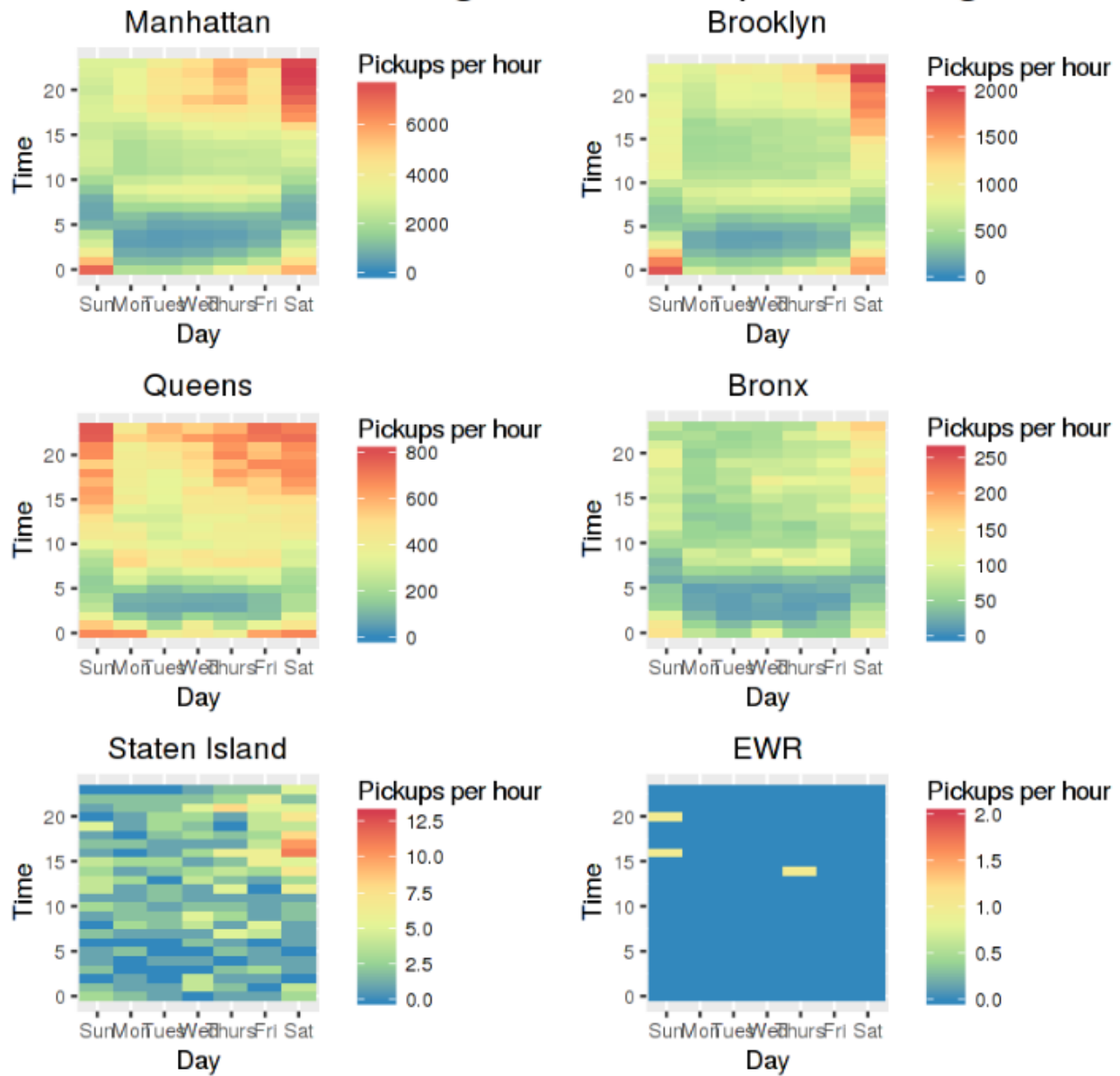
Data Visualization

Pickups per hour distribution by borough

Demand through the week per borough

Models' Performance