NEXT MOST PROFITABLE SHARE STATION-

ANALYZING THE CITI BIKE SHARE STATIONS BASED OF NYC CENSUS DATA

INTRODUCTION

GOAL: Using New York City 2015 (5-year estimate) census data to analyze the next most profitable location to add a bike share station.

APPROACH: Merge the 3 datasets together, determine the distance within the boroughs and the stations. Using the stations closest to the boroughs to create a models.

RESULT: Using the Gradient Regression Boost, Queens has the highest profitable possible stations.

CITI BIKE SHARE IS THE NATIONS LARGEST BIKE SHARE PROGRAM

- 12,000 bikes and 750 stations within Manhattan, Brooklyn, Queens and Jersey City.
- Designed for quick, convenient , and affordable trips
- Have Annual Member or Day Passes
- One year of bike share is cheaper than two monthly subway passes
- Can be more convenient than owning your own bike (no locks, no storage needed)

DATA:

- Data Set 1: from Kaggle-https://www.kaggle.com/muonneutrino/new-york-city-census-data

- Data Set 2:from Citi Bike Share- https://www.citibikenyc.com/system-data

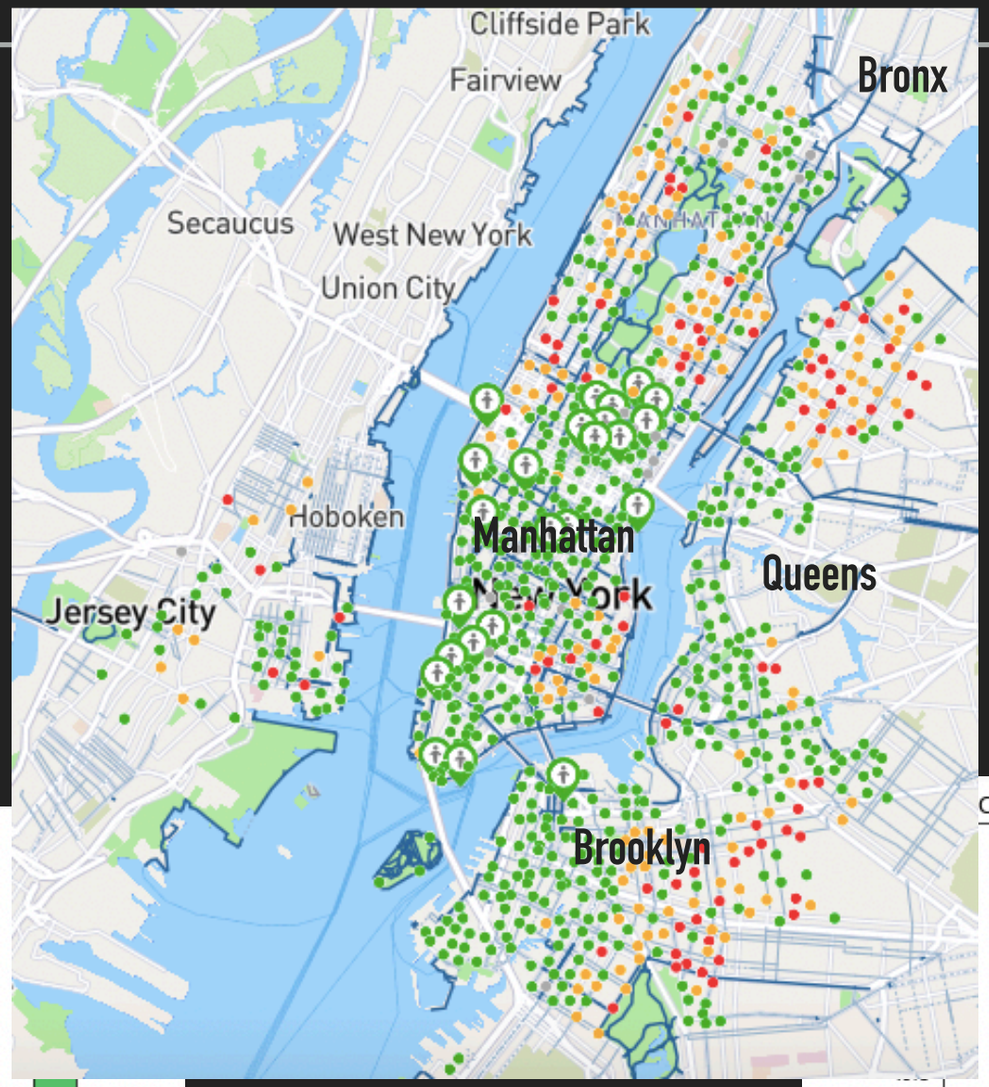
Two Datasets

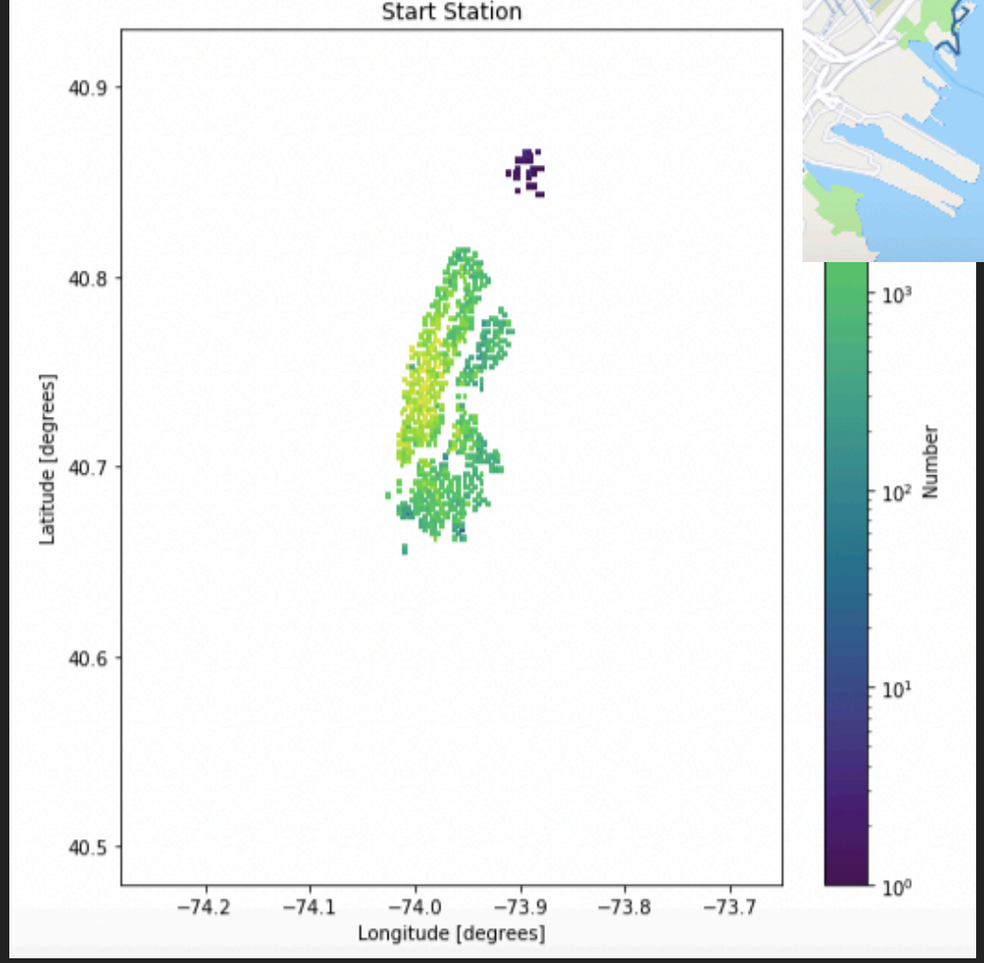
Merged:

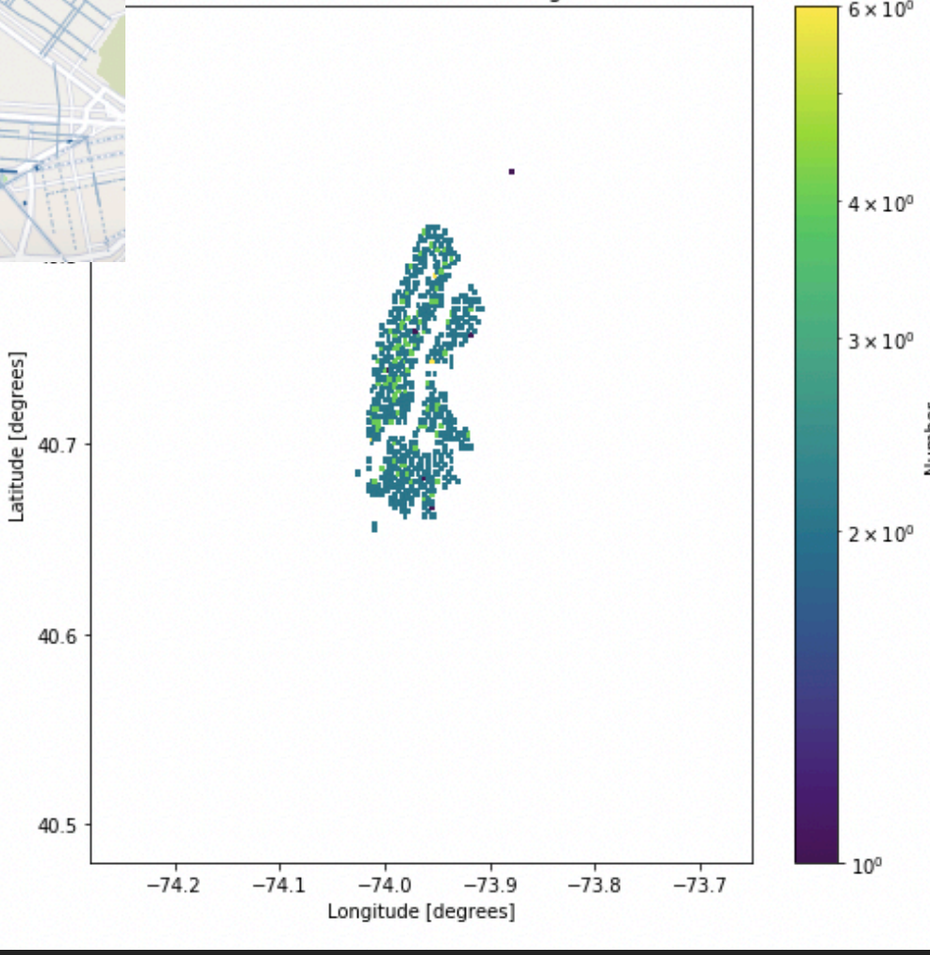
Unnamed: 0	int64
start_station_id	float64
start_station_latitude	float64
start_station_longitude	float64
tripduration	int64
tmp	object
CensusTract	int64
County_x	object
Borough	object
TotalPop	int64
Men	int64
Women	int64
Hispanic	float64
White	float64
Black	float64
Native	float64
Asian	float64
Citizen	int64
Income	float64
IncomeErr	float64
IncomePerCap	float64
IncomePerCapErr	float64
Poverty	float64
ChildPoverty	float64
Professional	float64
Service	float64
Office	float64
Construction	float64
Production	float64
Drive	float64
Carpool	float64
Transit	float64
Walk	float64
OtherTransp	float64
WorkAtHome	float64
MeanCommute	float64
Employed	int64
PrivateWork	float64
PublicWork	float64
SelfEmployed	float64
FamilyWork	float64
Unemployment	float64
Latitude	float64
Longitude	float64
BlockCode	int64
County_y	object
State	object
BlockCode_11	int64
borough_distance	float64
distance_rank	float64
dtype: object	

SHARE STATIONS- MAP OF NYC

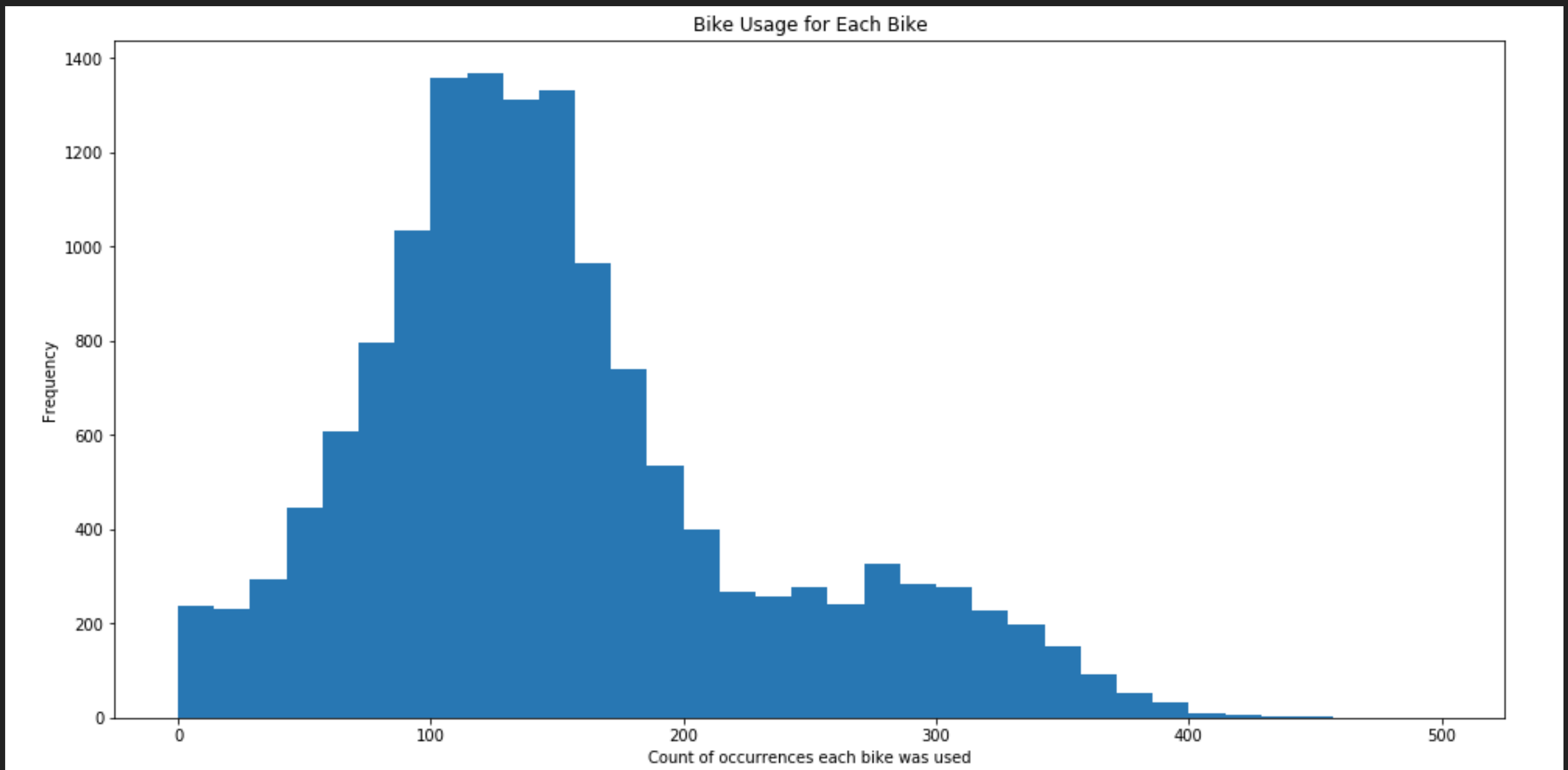
THERE ARE NO BIKE SHARE STATIONS IN THE BRONX OR STATEN ISLAND



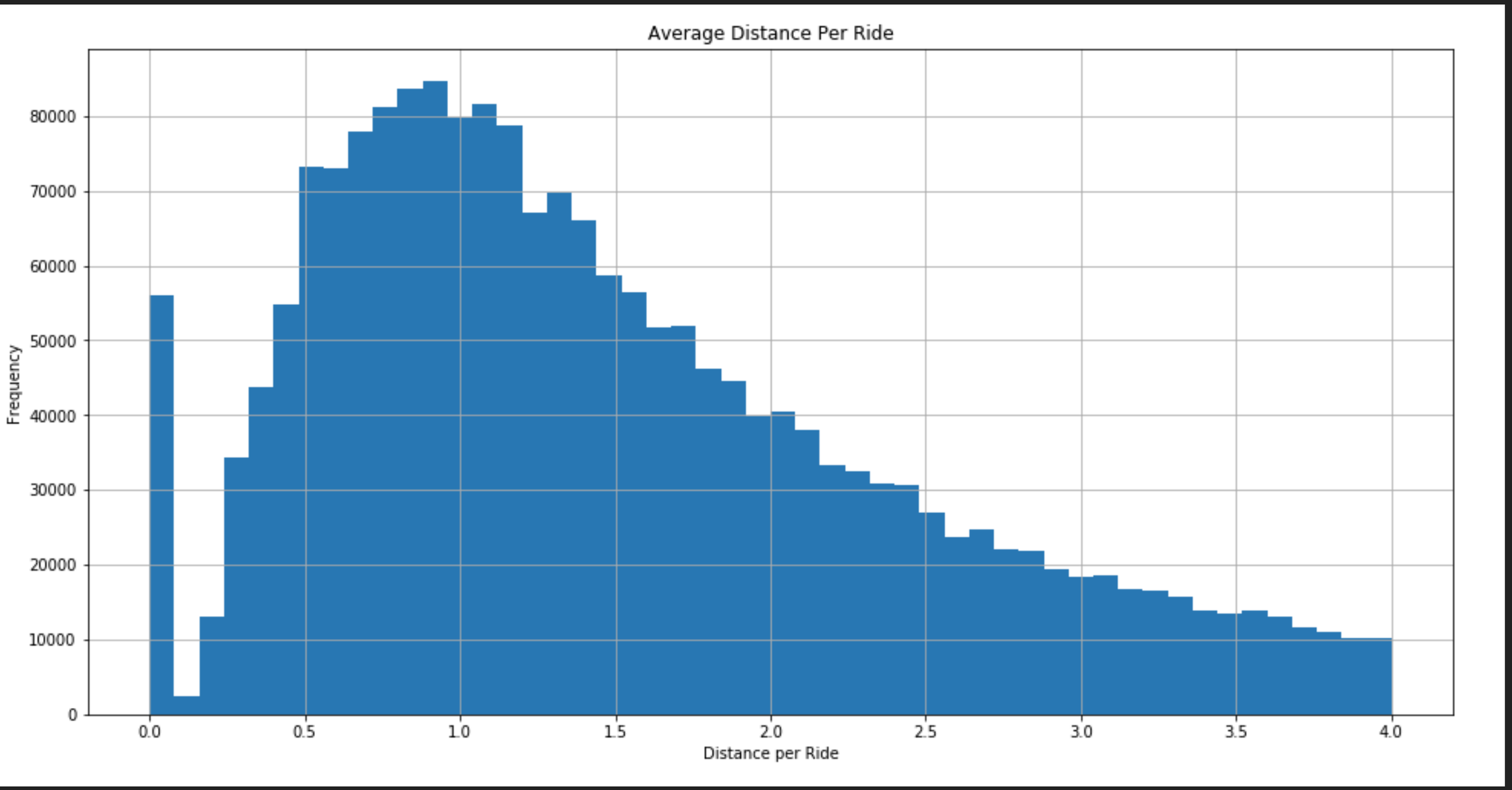




There are 14356 bikes and each have been used on average 152 times.

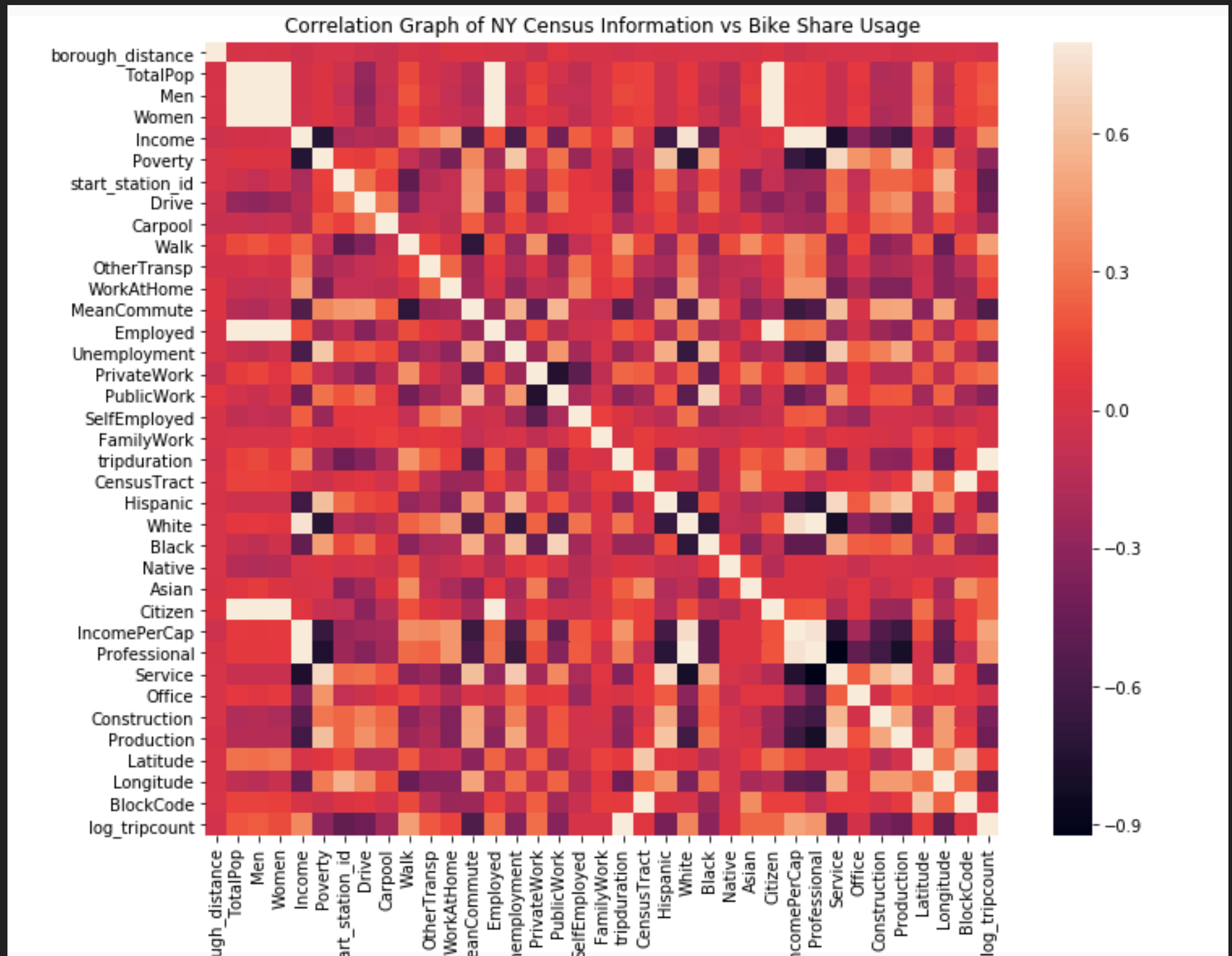


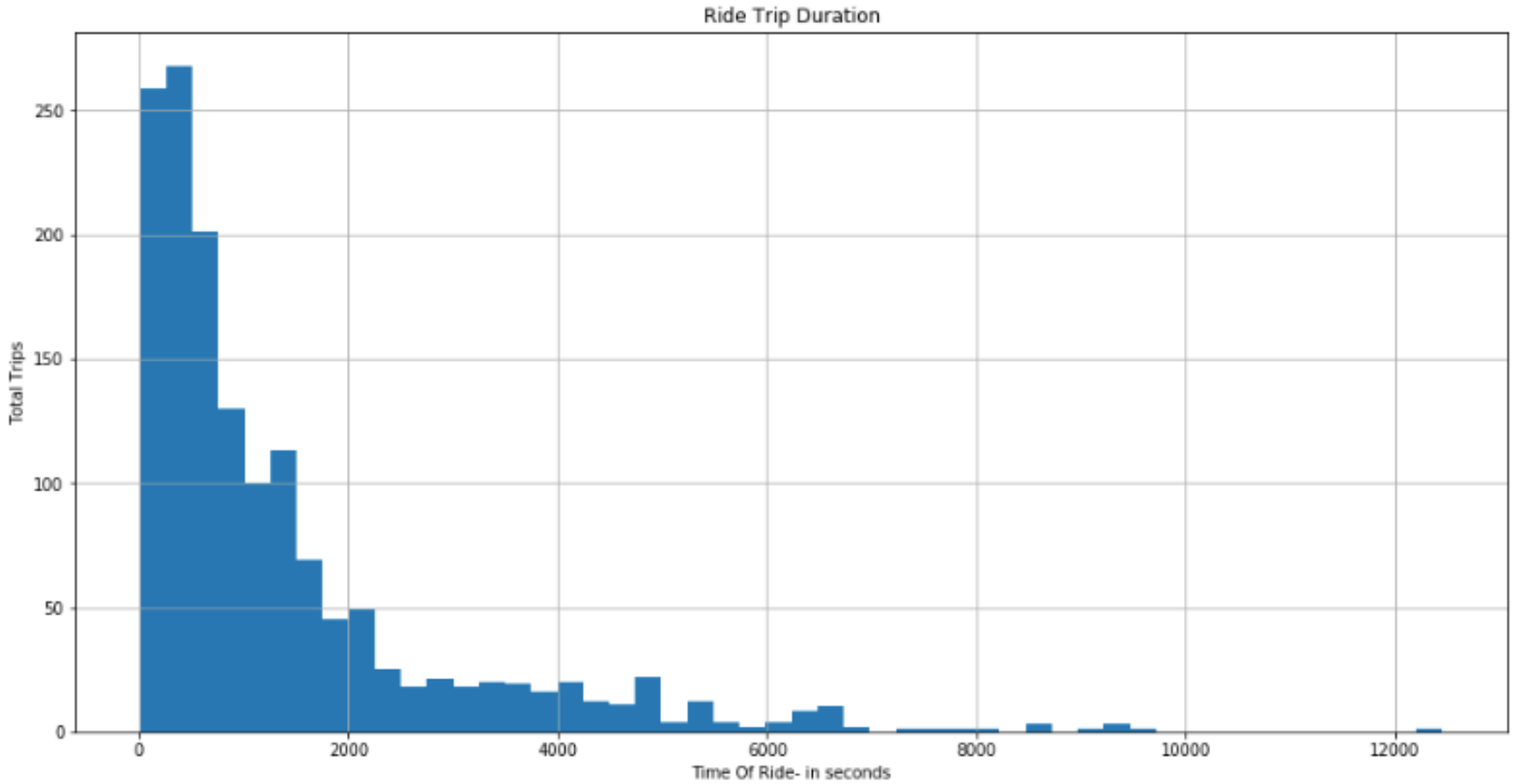
The average ride is slightly under 2km, likely riding from one station to another station



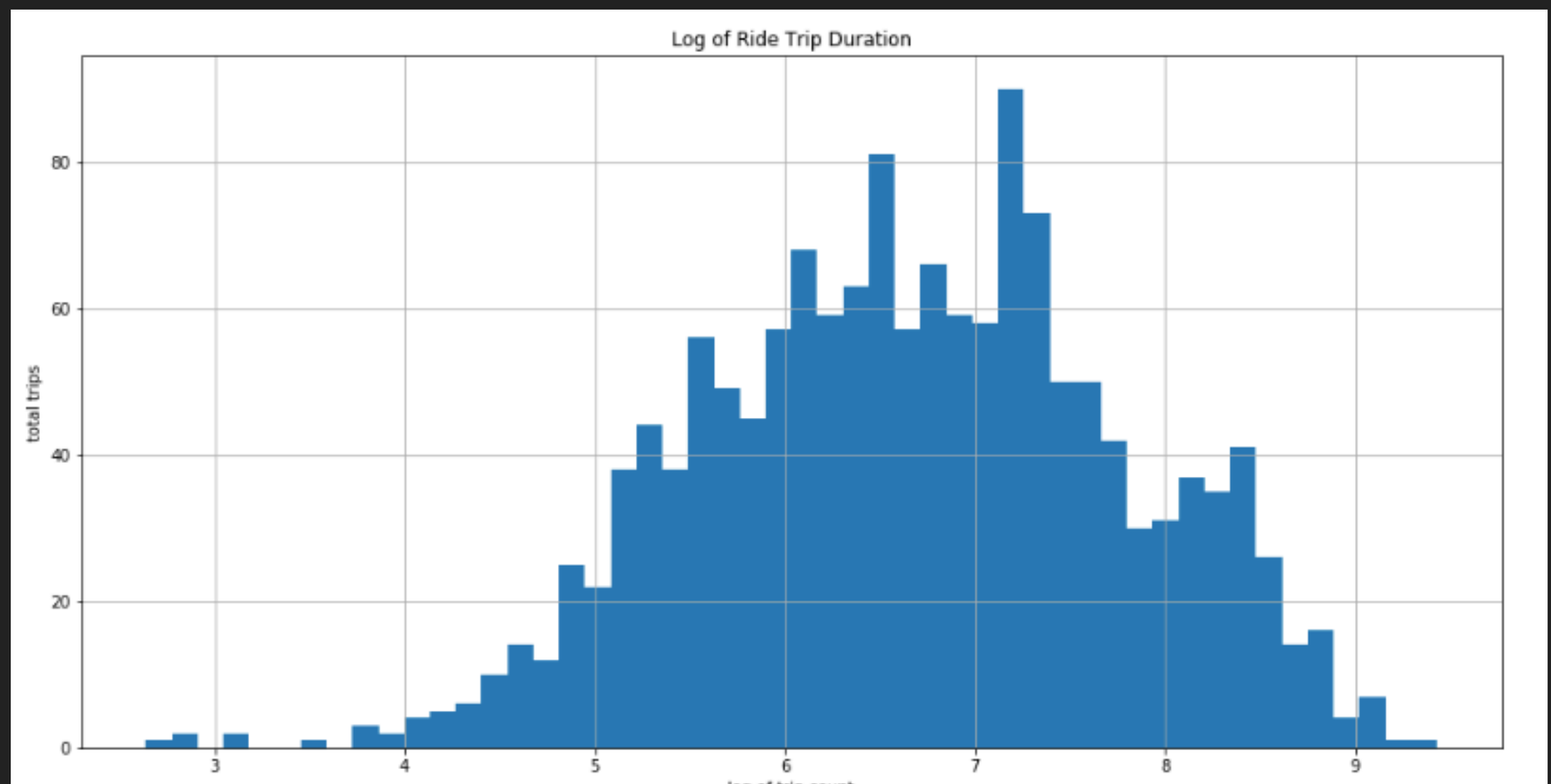
HEAT CORRELATION MAP

Showing high correlation with Income, Walking, Employed, Professional, Total Population, and Men





Uneven data shown here, so I chose to use the log function to normalize the data for use within my models.



Used both Linear Regression and R-squared Linear Regression/R-square PLSR models to try and predict the most beneficial location for additional stations, the R-squared result showed a 49% accuracy prediction rate.

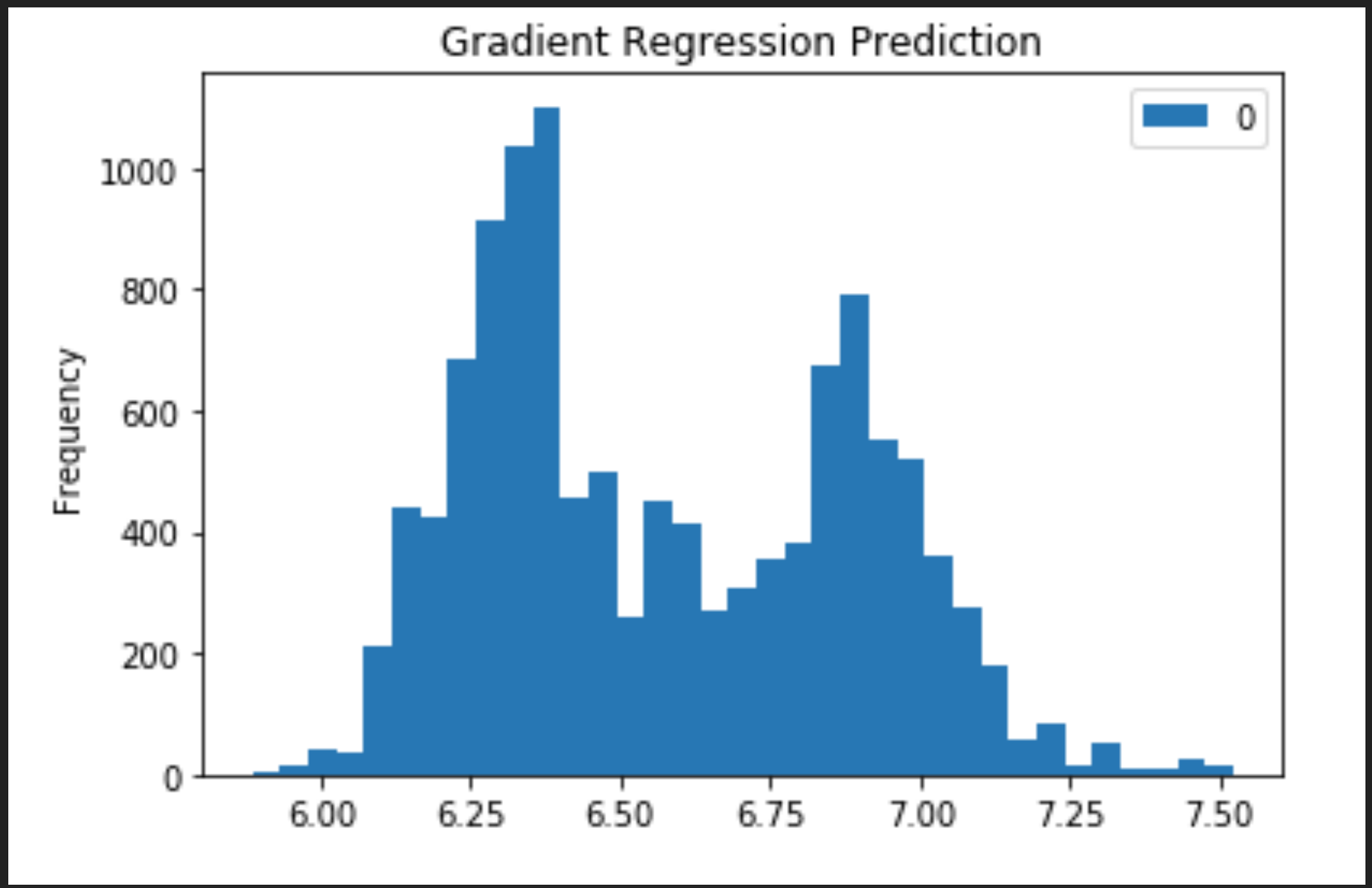
Linear Regression Means Squared Error

```
0.6607169777506211
```

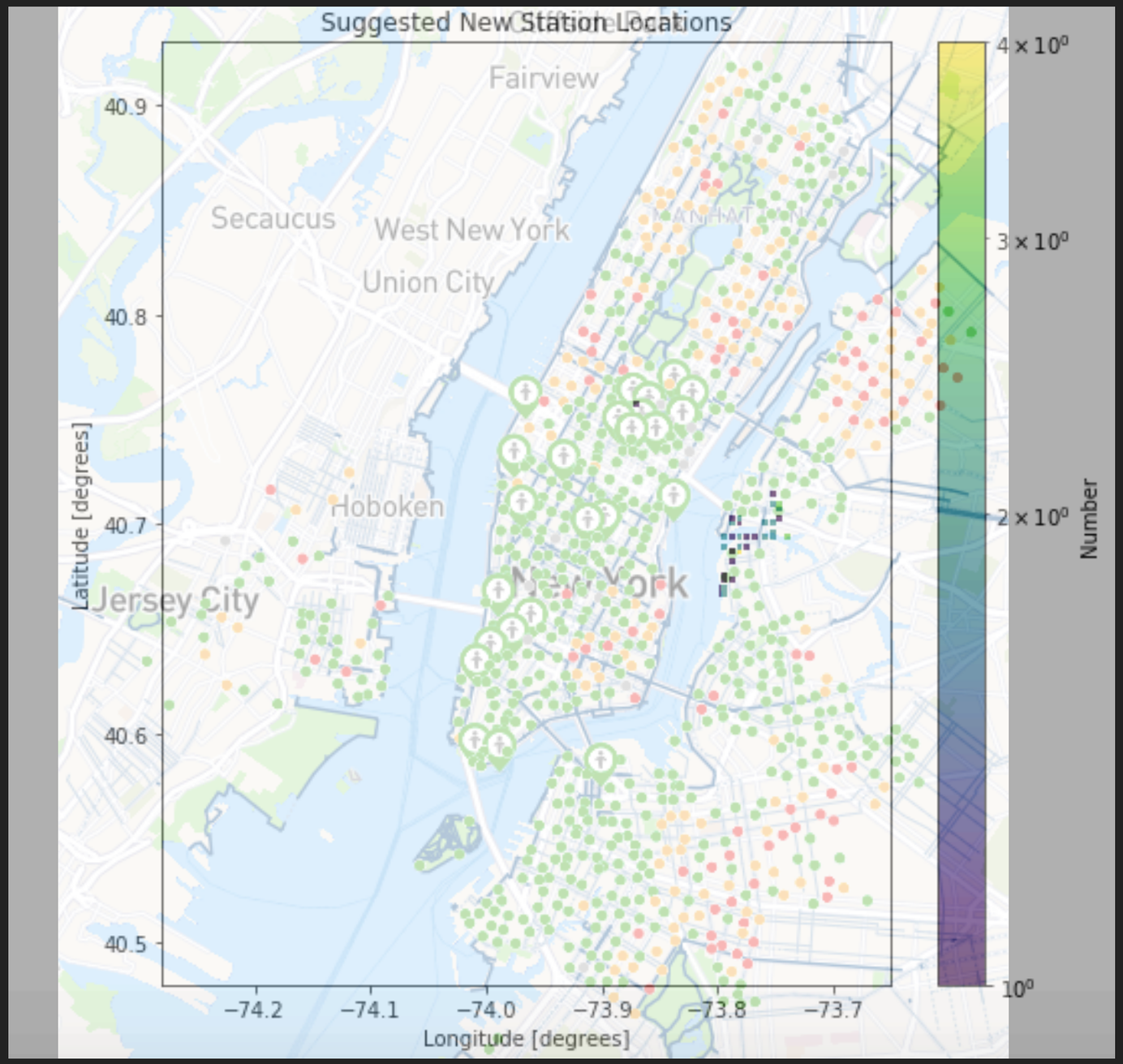
Partial Least Square vs Linear Regression

```
R-squared regression: 0.4910456463578624
R-squared PLSR: 0.48806498687104904
```

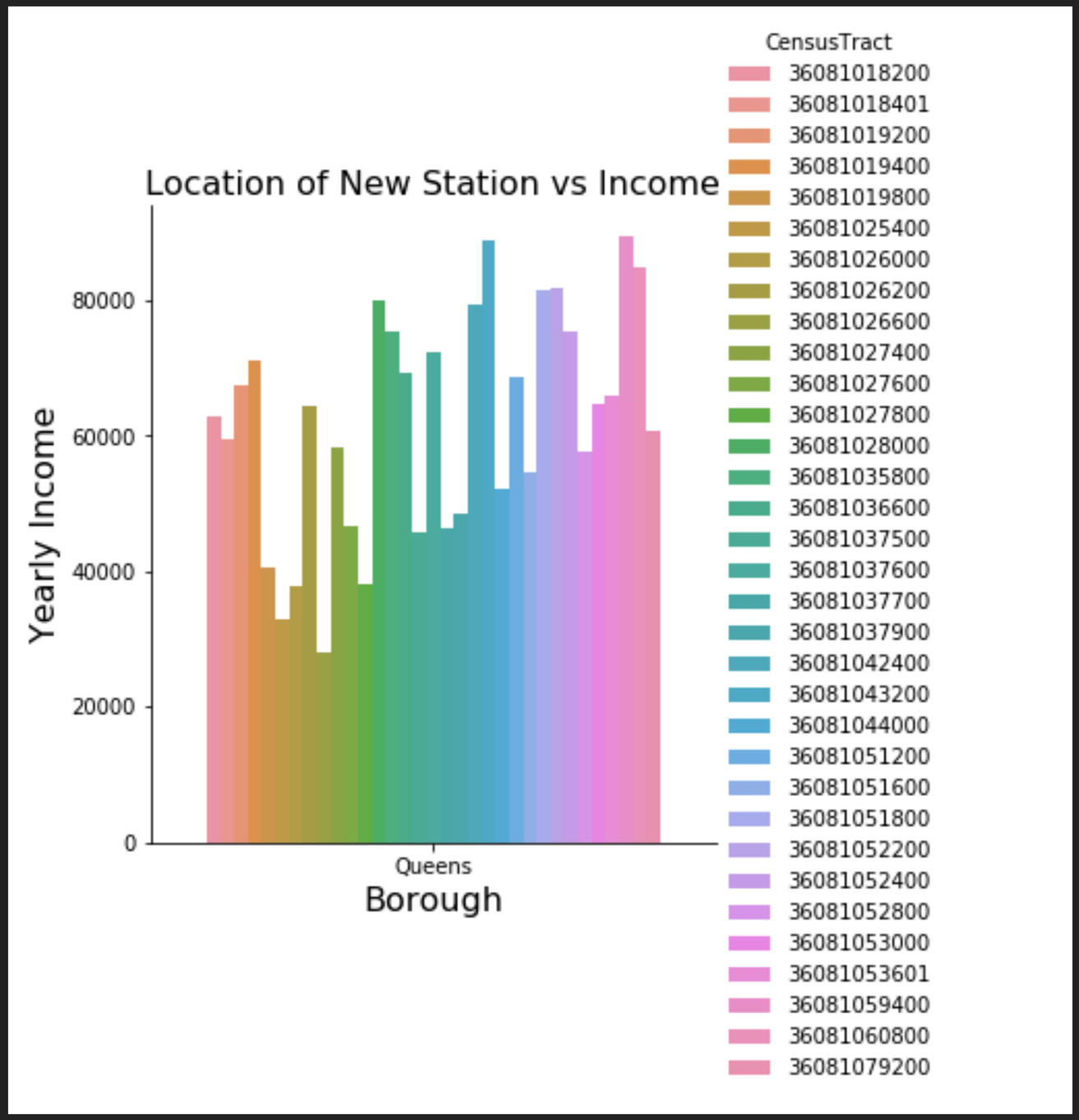
Gradient Regression proved to be the most effective prediction model for this data. With this histogram it shows the ride frequency, remembering that it is raised to the 6th from the log function. The right side of the tail is showing that there is room for growth in those boroughs, looking at about 7.25 for growth options.



Looking at the NY Census data (including income, gender, transportation preference, work type, etc) using the location based off of latitude and longitude with a gradient regression 7.25 or higher. This map shows that all the stations will fall into the Queens Borough. The blue/green dots are the locations to start with out of the 128 new suggested locations within the borough that would be the most profitable stations to start with and then add onto.



This bar graph shows that within the different census tracts/ areas of Queen the income levels vary. For those that average $50,000-60,000/ year would find the bike share stations probably most helpful as it would save them money on subway passes and the ability to ride into the city without having to find a place to keep their bike locked during the day.



“THE ONLY THING THAT LEAKS ON A BIKE IS AIR”

–Unknown

Appendix

GRADIENT BOOST REGRESSION– TRAINING MODEL

```
R2 sq:  0.616899545603221
Mean squared error: 0.50
Test Variance score: 0.62
```