Prediction and Visualization on Citi-Bike ridership data for the Mayor of New-York City

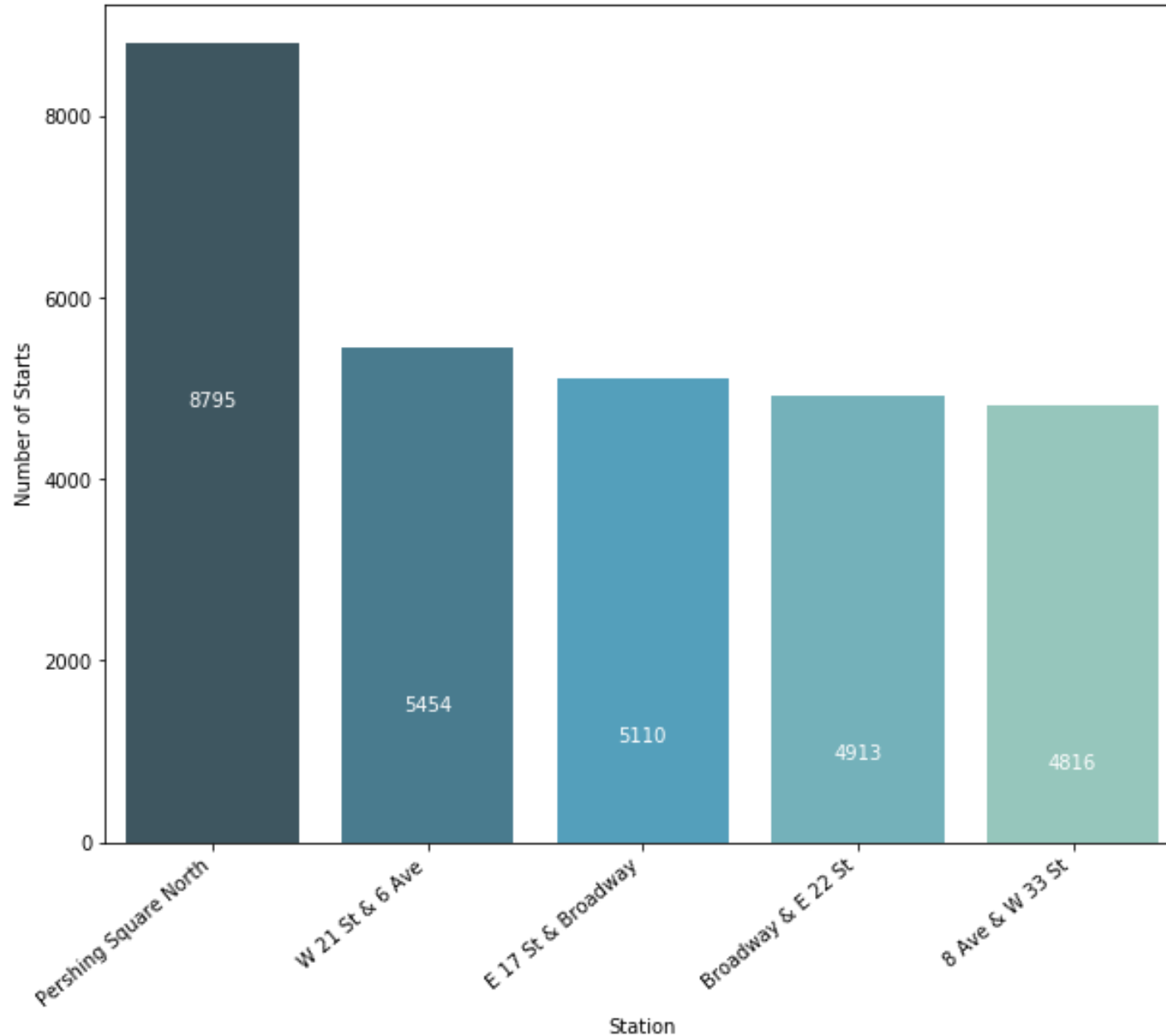## Farhad Kazemi

Q1: Top 5 Stations-missing values
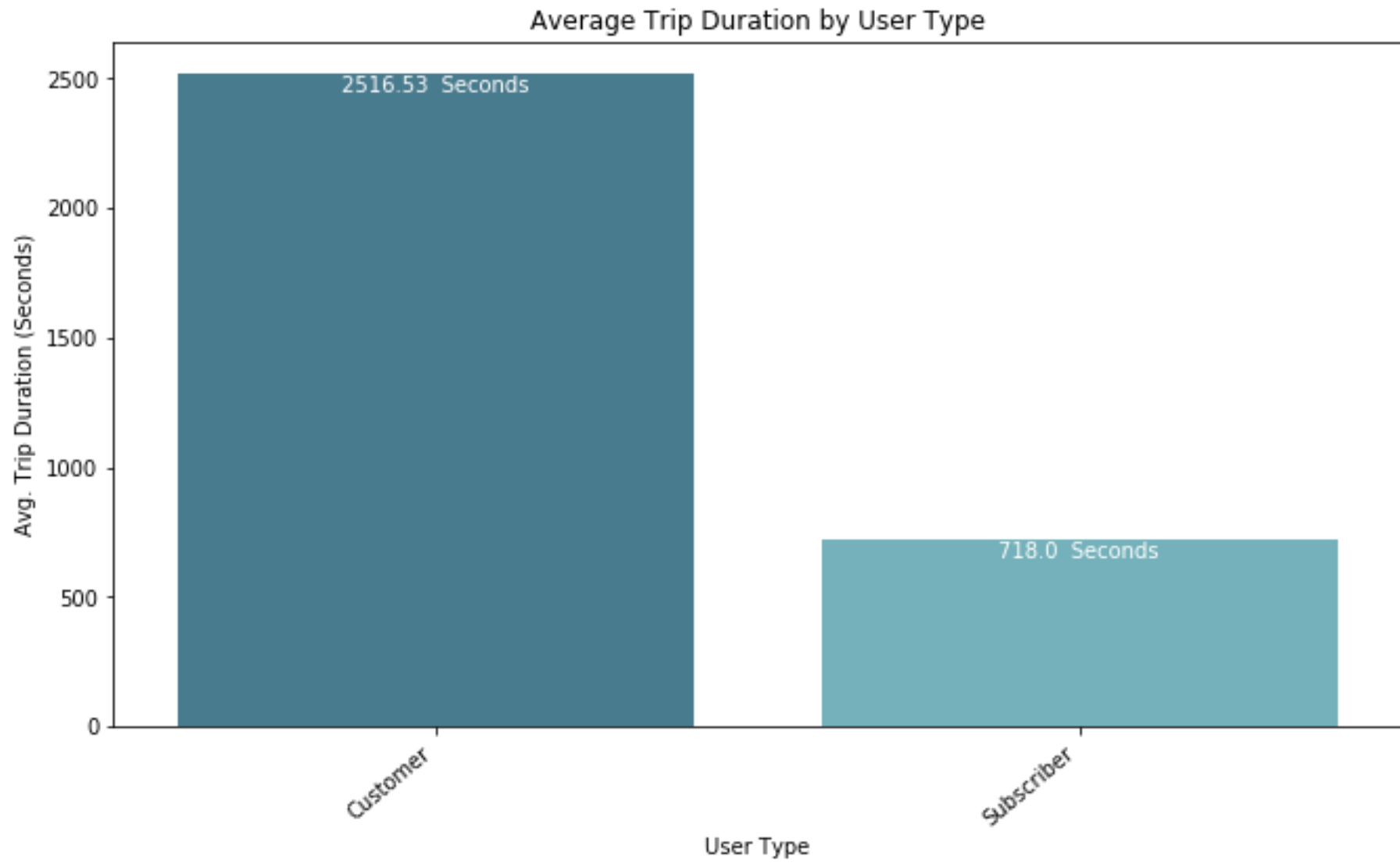
	Total	Percent
Birth Year	29076	4.001233
User Type	3193	0.439398
Gender	0	0.000000
Bike ID	0	0.000000
End Station Longitude	0	0.000000
End Station Latitude	0	0.000000
End Station Name	0	0.000000
End Station ID	0	0.000000
Start Station Longitude	0	0.000000
Start Station Latitude	0	0.000000
Start Station Name	0	0.000000
Start Station ID	0	0.000000
Stop Time	0	0.000000
Start Time	0	0.000000
Trip Duration	0	0.000000

Q1: Top 5 Stations

Top 5 Citi Bike Stations by Number of Starts
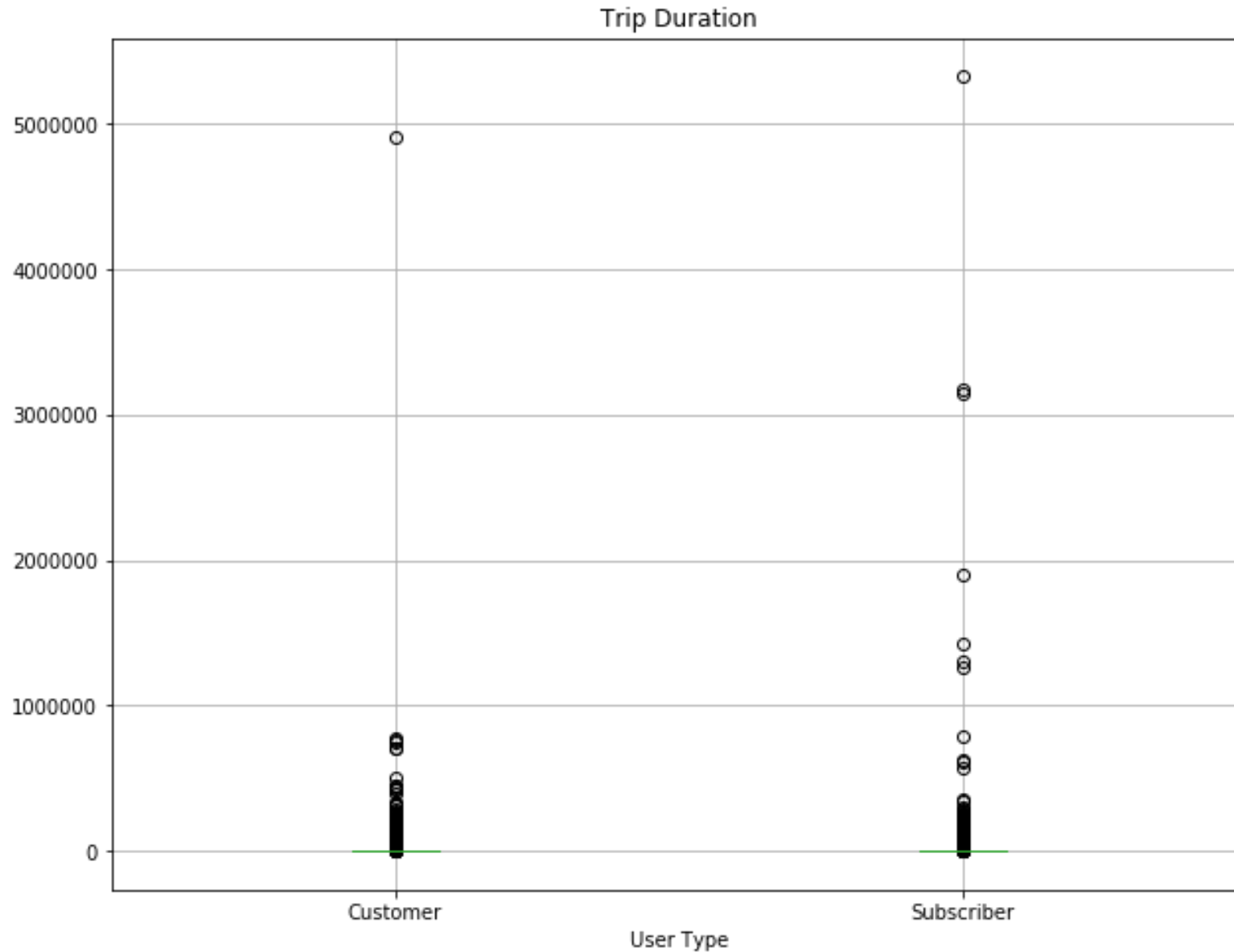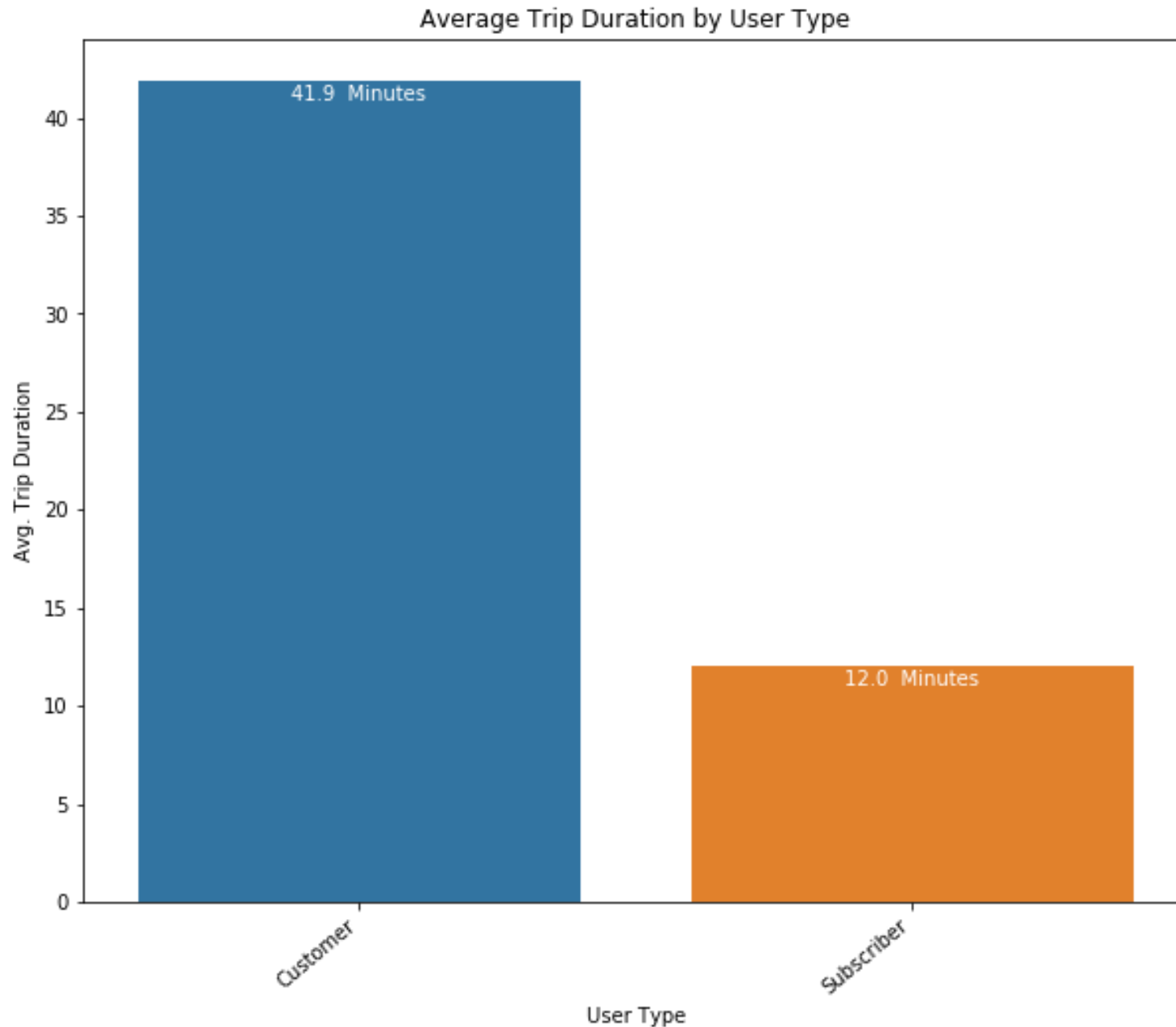
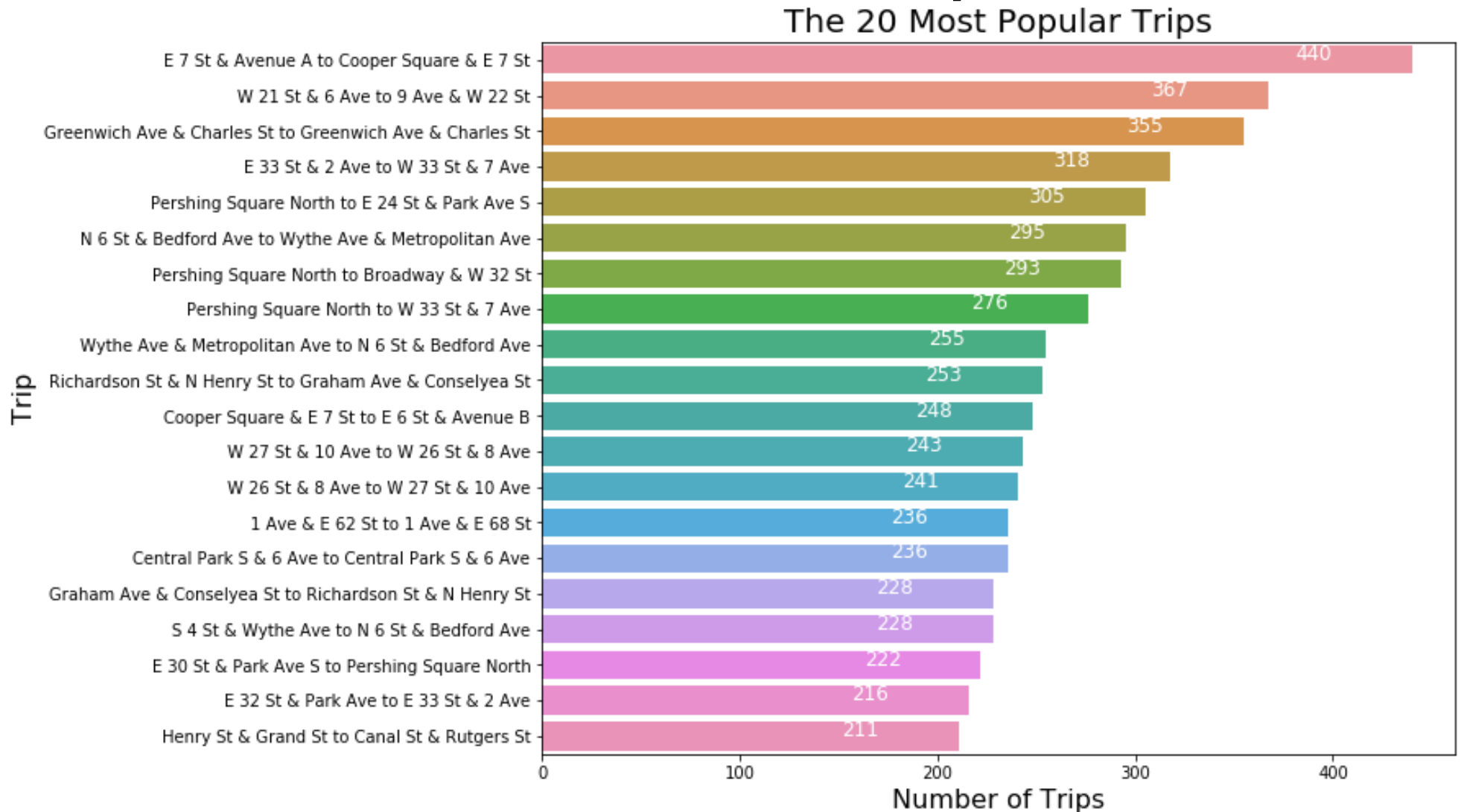# Q2: Trip Duration by User Type

Average Trip Duration by User Type

Q2: Trip Duration by User Type

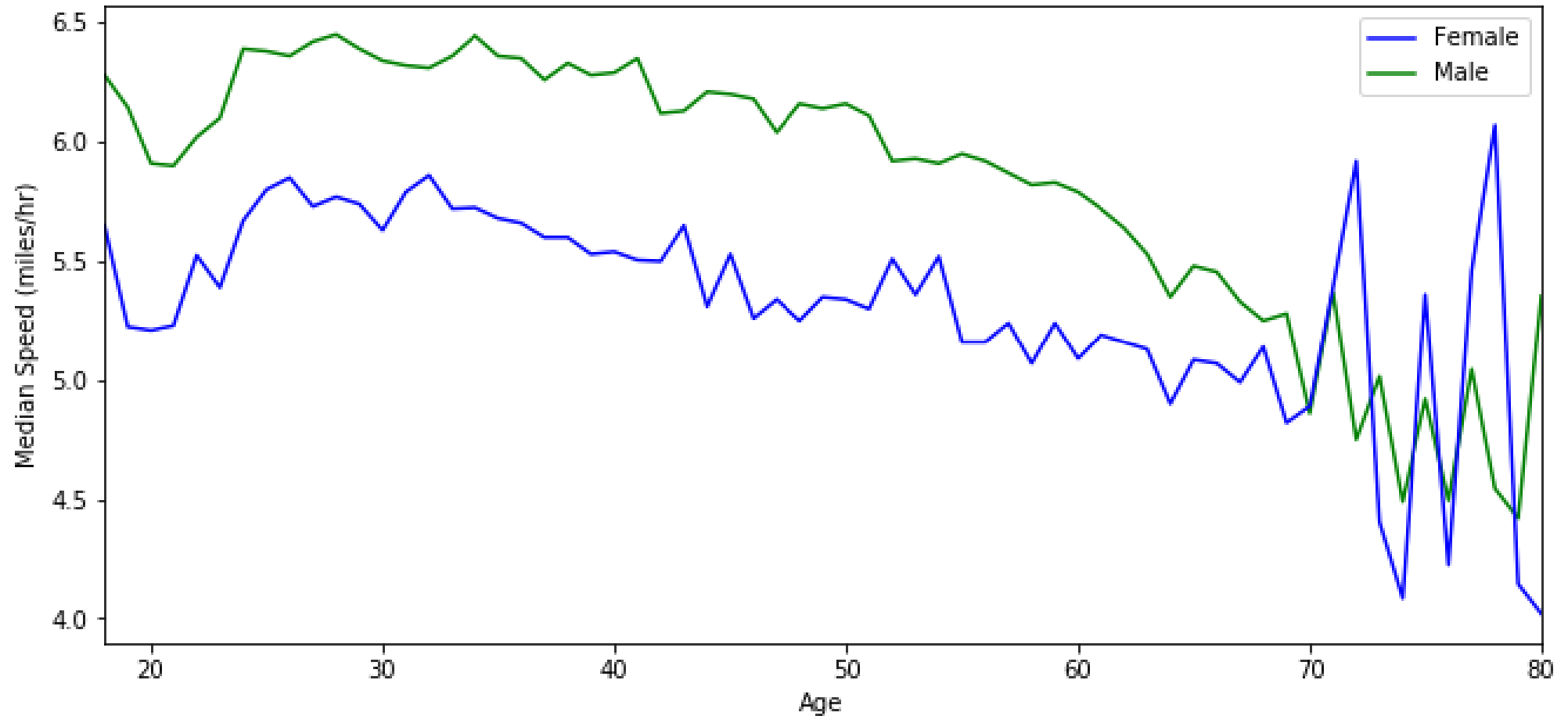Boxplot grouped by User Type

Q2: Trip Duration by User Type

Average Trip Duration by User Type

Q3: Most Popular Trip based on start station and stop station

The 20 Most Popular Trips

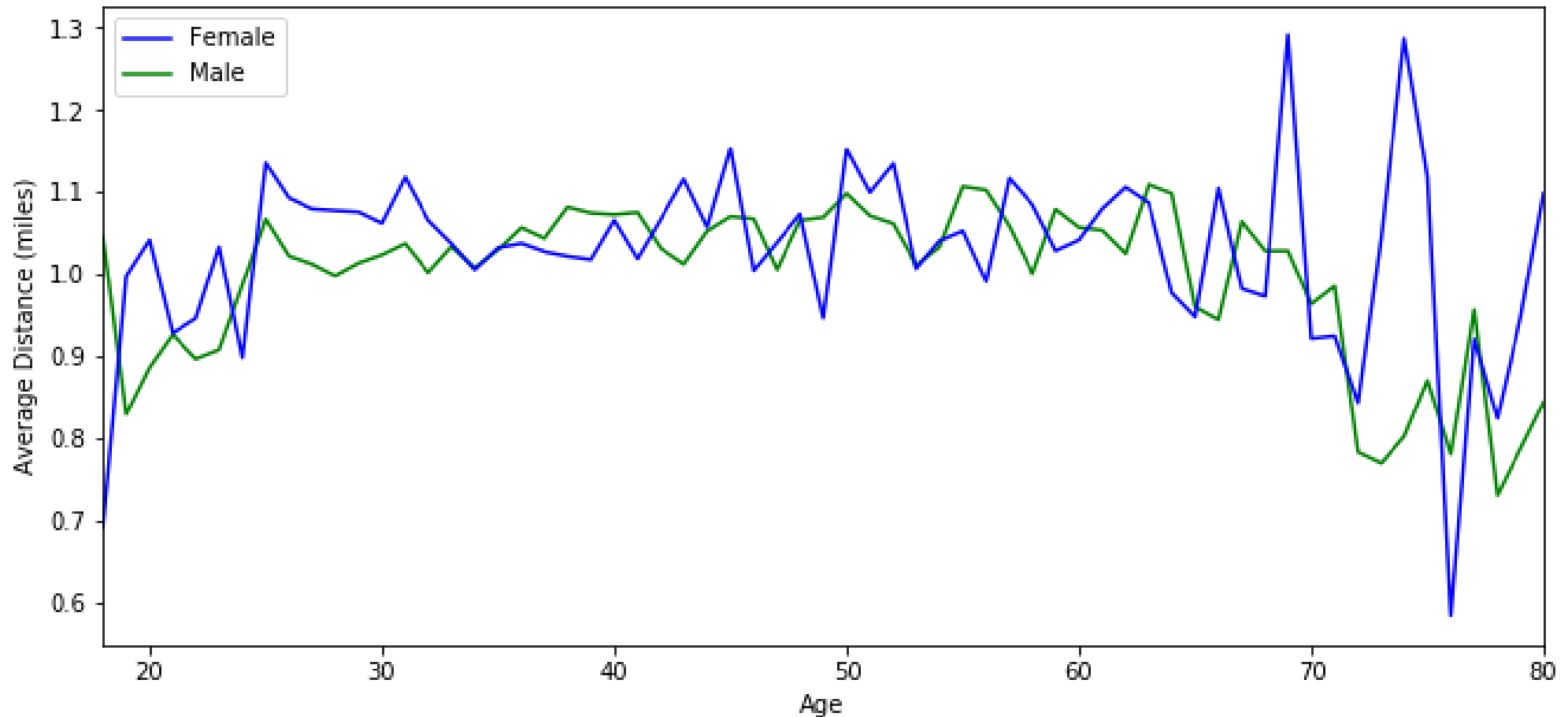| Trip	Number of Trips
E 7 St & Avenue A to Cooper Square & E 7 St	440
W 21 St & 6 Ave to 9 Ave & W 22 St	367
Greenwich Ave & Charles St to Greenwich Ave & Charles St	355
E 33 St & 2 Ave to W 33 St & 7 Ave	318
Pershing Square North to E 24 St & Park Ave S	305
N 6 St & Bedford Ave to Wythe Ave & Metropolitan Ave	295
Pershing Square North to Broadway & W 32 St	293
Pershing Square North to W 33 St & 7 Ave	276
Wythe Ave & Metropolitan Ave to N 6 St & Bedford Ave	255
Richardson St & N Henry St to Graham Ave & Conselyea St	253
Cooper Square & E 7 St to E 6 St & Avenue B	248
W 27 St & 10 Ave to W 26 St & 8 Ave	243
W 26 St & 8 Ave to W 27 St & 10 Ave	241
1 Ave & E 62 St to 1 Ave & E 68 St	236
Central Park S & 6 Ave to Central Park S & 6 Ave	236
Graham Ave & Conselyea St to Richardson St & N Henry St	228
S 4 St & Wythe Ave to N 6 St & Bedford Ave	228
E 30 St & Park Ave S to Pershing Square North	222
E 32 St & Park Ave to E 33 St & 2 Ave	216
Henry St & Grand St to Canal St & Rutgers St	211

Q4: Rider Performance by Gender and Age

Rider Performance Based on Gender and Age (Median Speed - miles/hr)

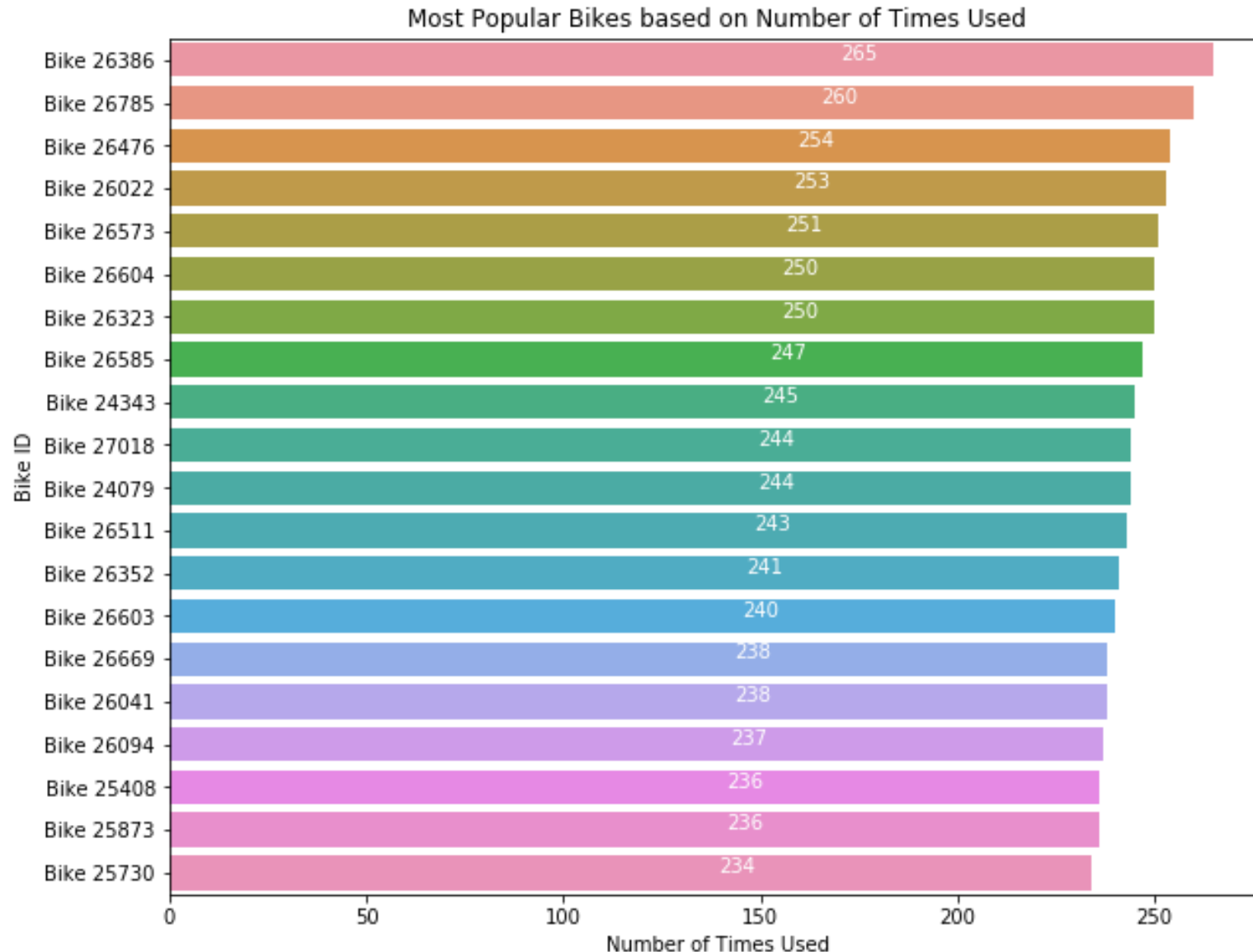# Q4: Rider Performance by Gender and Age

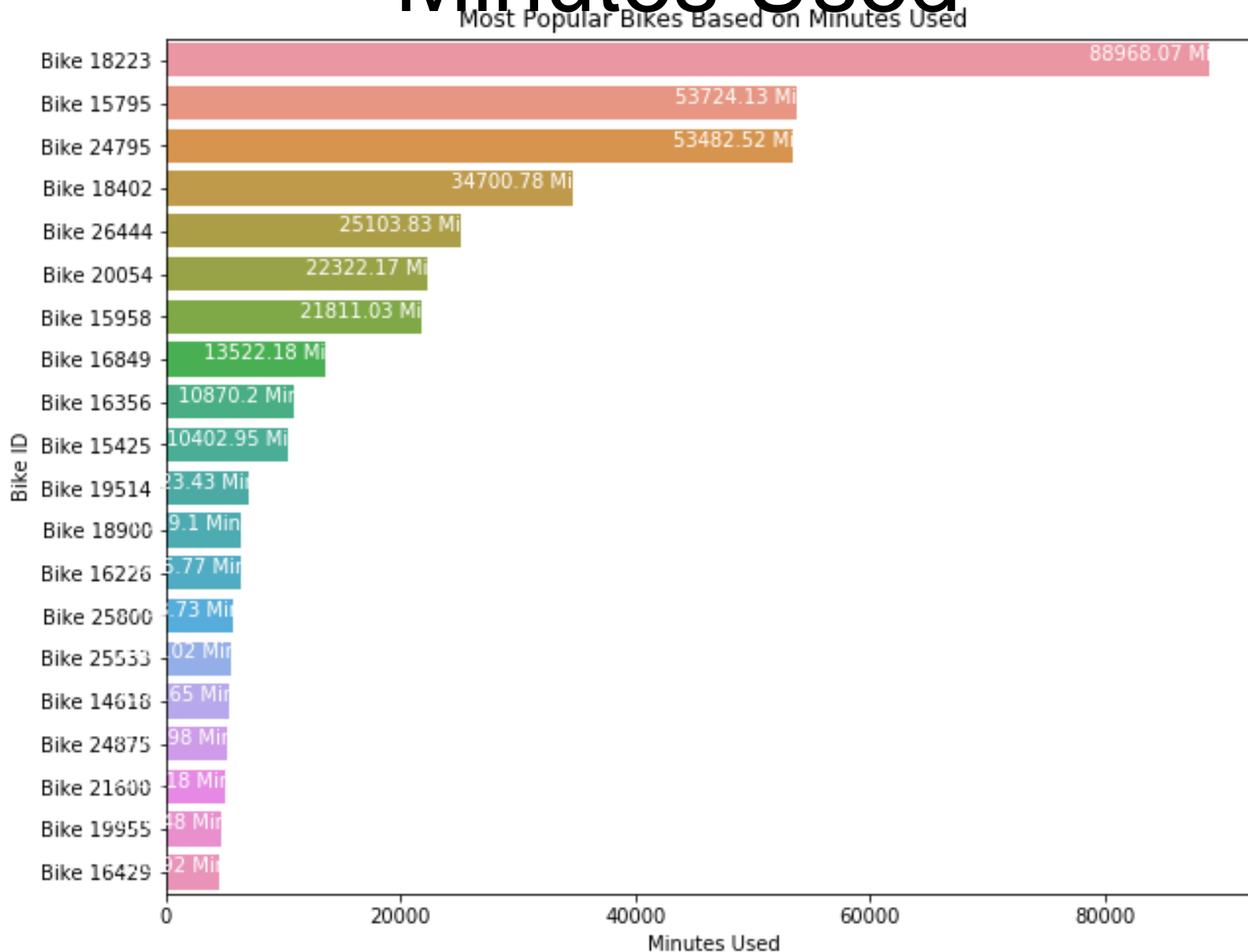Rider Performance Based on Gender and Age (Average Distance in Miles)

Q5: Busiest Bike by Times and Minutes Used

- The busiest bike in Jan 2017
    - Bike 26386

 - Times was it used.
    -265 times

 - Minutes was it in use.
    - 88968 Minutes

Q5: Busiest Bike by Times and Minutes Used

Most Popular Bikes based on Number of Times Used

Q5: Busiest Bike by Times and Minutes Used

Most Popular Bikes Based on Minutes Used

Bike ID	Minutes Used
Bike 18223	88968.07 Mi
Bike 15795	53724.13 Mi
Bike 24795	53482.52 Mi
Bike 18402	34700.78 Mi
Bike 26444	25103.83 Mi
Bike 20054	22322.17 Mi
Bike 15958	21811.03 Mi
Bike 16849	13522.18 Mi
Bike 16356	10870.2 Min
Bike 15425	10402.95 Mi
Bike 19514	23.43 Min
Bike 18900	9.1 Min
Bike 16226	6.77 Min
Bike 25800	73 Mi
Bike 25533	02 Min
Bike 14618	65 Min
Bike 24875	98 Min
Bike 21600	18 Min
Bike 19955	48 Min
Bike 16429	92 Min

Q6: Predictive Model

Prepare our data for a multivariate regression
Drop irrelevant columns

Accuracy: 0.97811

OLS Regression Results

Dep. Variable:	Gender_1	R-squared:	0.978
Model:	OLS	Adj. R-squared:	0.978
Method:	Least Squares	F-statistic:	8.192e+06
Date:	Sun, 04 Nov 2018	Prob (F-statistic):	0.00
Time:	14:08:13	Log-Likelihood:	7.5146e+05
No. Observations:	548588	AIC:	-1.503e+06
Df Residuals:	548584	BIC:	-1.503e+06
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.9986	0.004	226.600	0.000	0.990	1.007
Distance	0.0003	0.000	3.182	0.001	0.000	0.001
User Type_Subscriber	-0.0038	0.004	-0.873	0.383	-0.012	0.005
Gender_2	-0.9951	0.000	-4957.113	0.000	-0.996	-0.995

Omnibus:	1073819.150	Durbin-Watson:	2.000
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1539870245.823
Skew:	-16.102	Prob(JB):	0.00
Kurtosis:	260.547	Cond. No.	138.