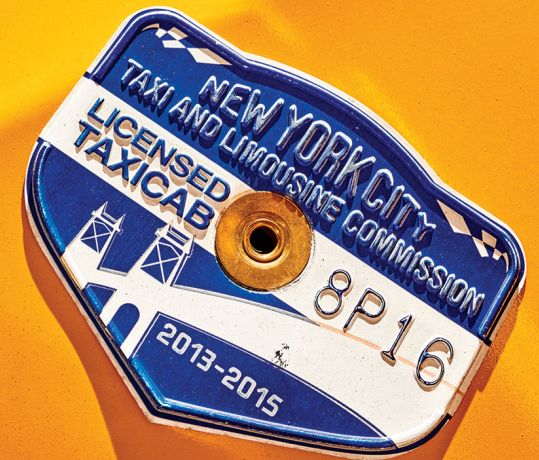# NYC Taxi Data

Abhishek Das
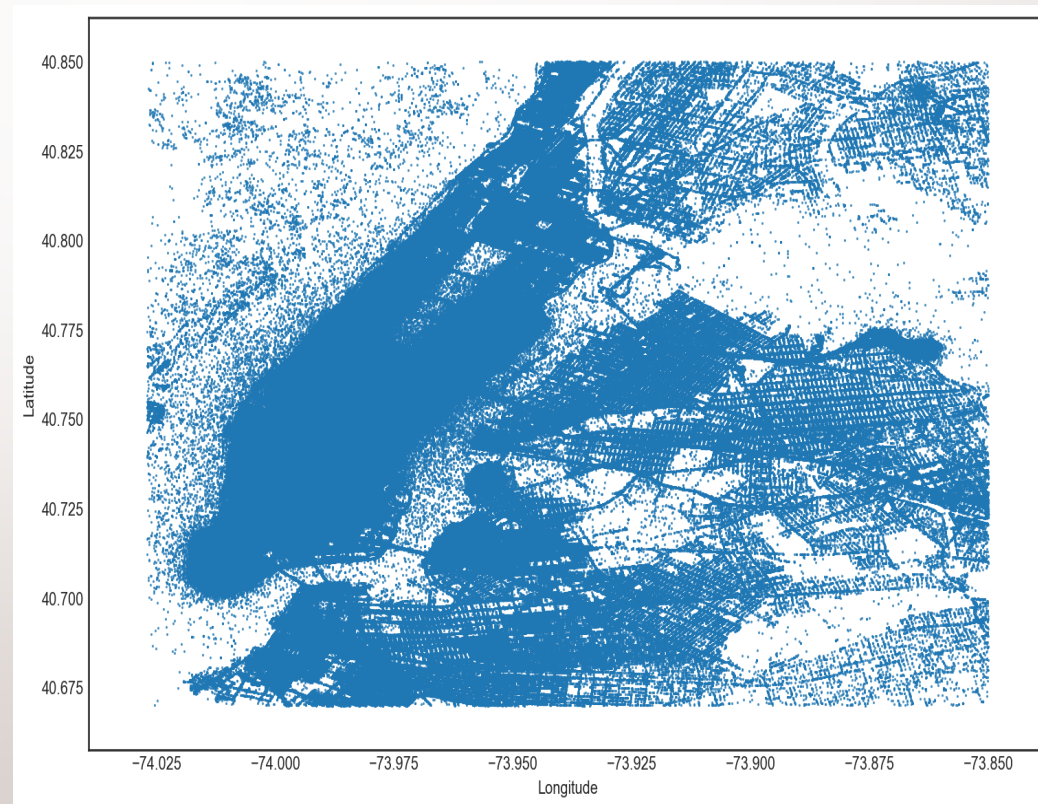
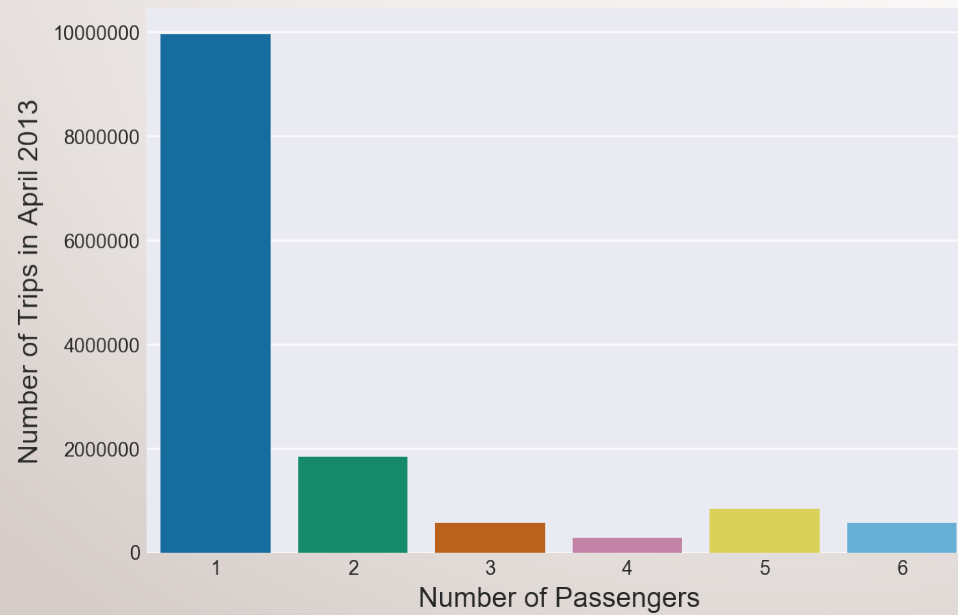# One Month

2 companies

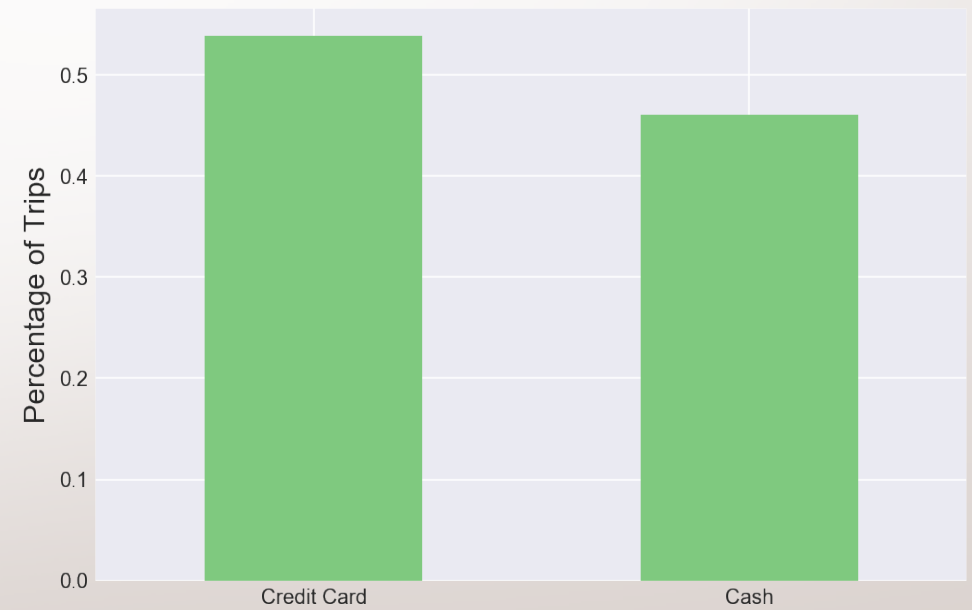13,276 taxis

34 million miles

$184 million in revenue

Passengers

Nearly 10,000,000 cab rides with a single passenger
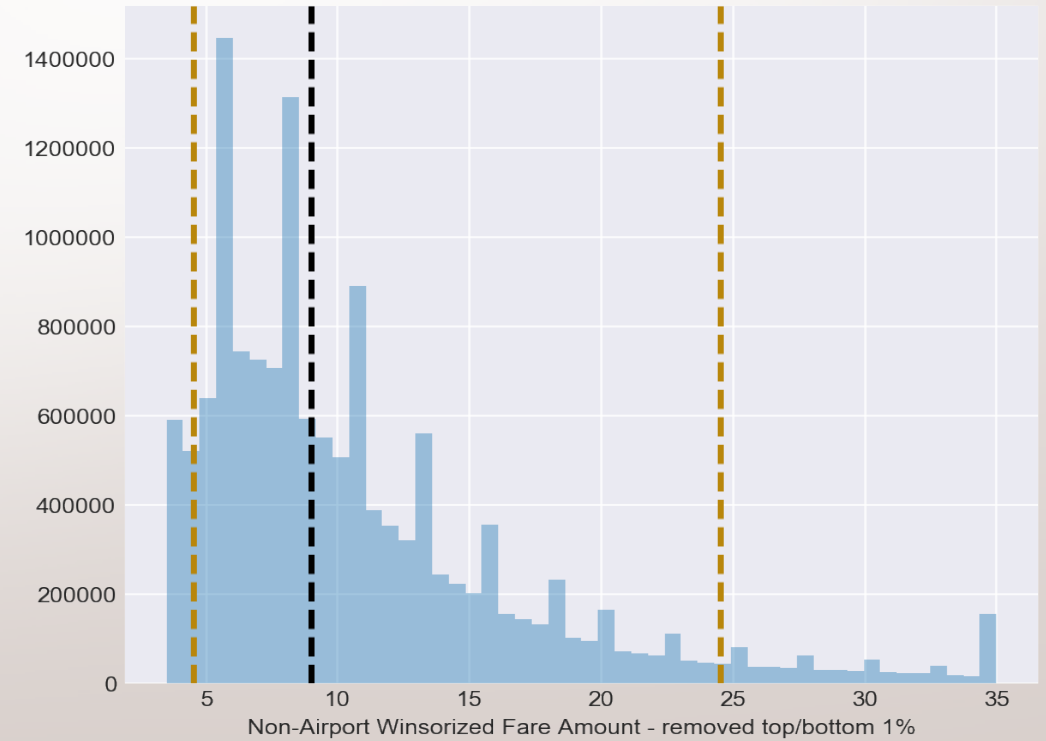
53.9% of trips are paid for with credit card

Non-Airport Fares < $35
Median Airport fares $52

Distribution of Fare Amount: Airport vs Non-Airport Fares

Airport Winsorized Fare Amount - removed top/bottom 1%

Non-Airport Winsorized Fare Amount - removed top/bottom 1%
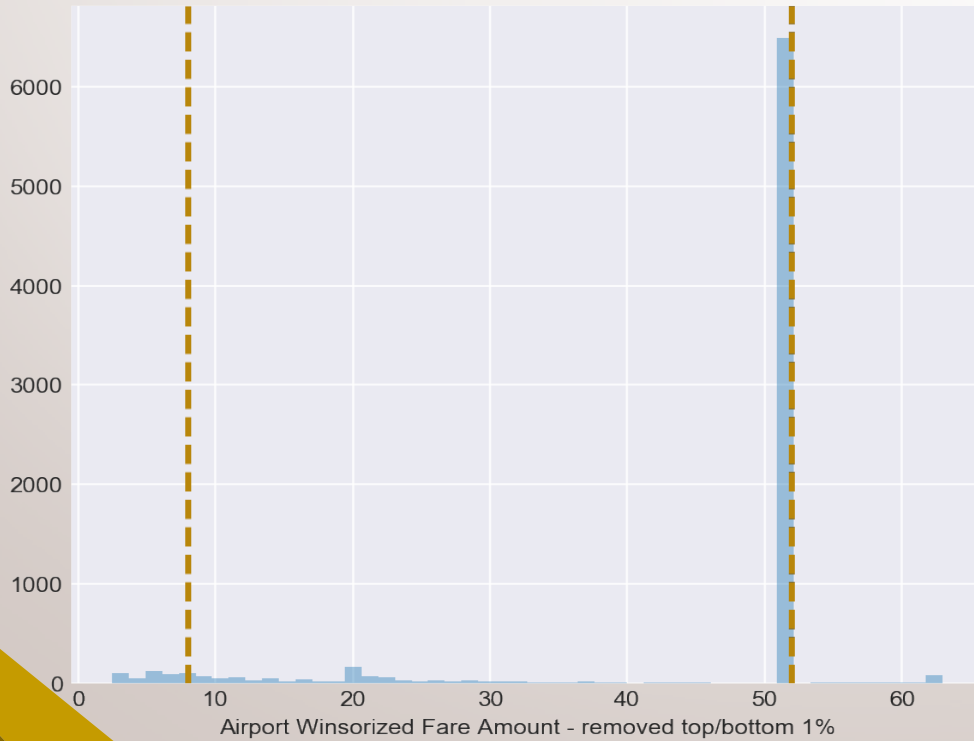
Airport tips are higher

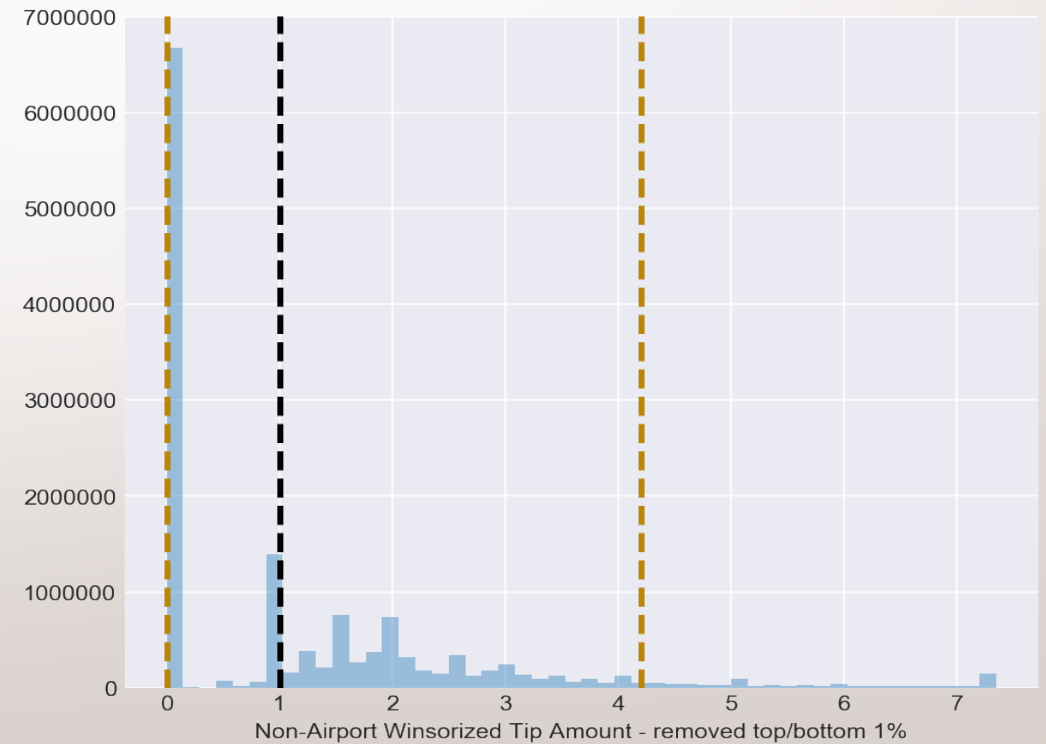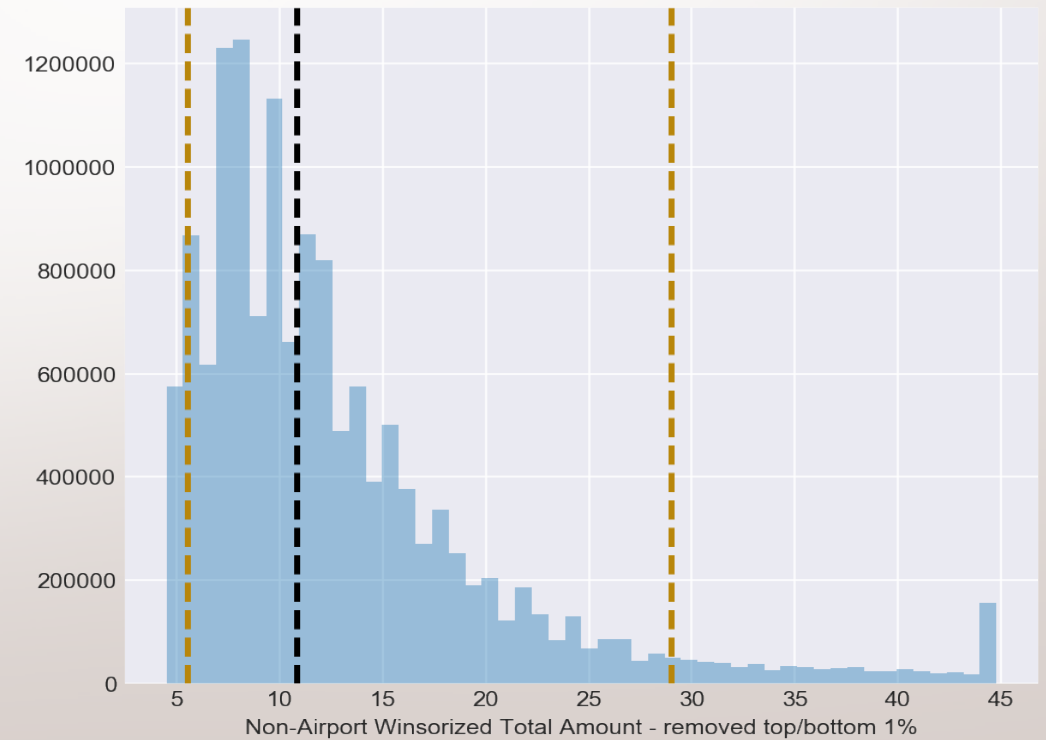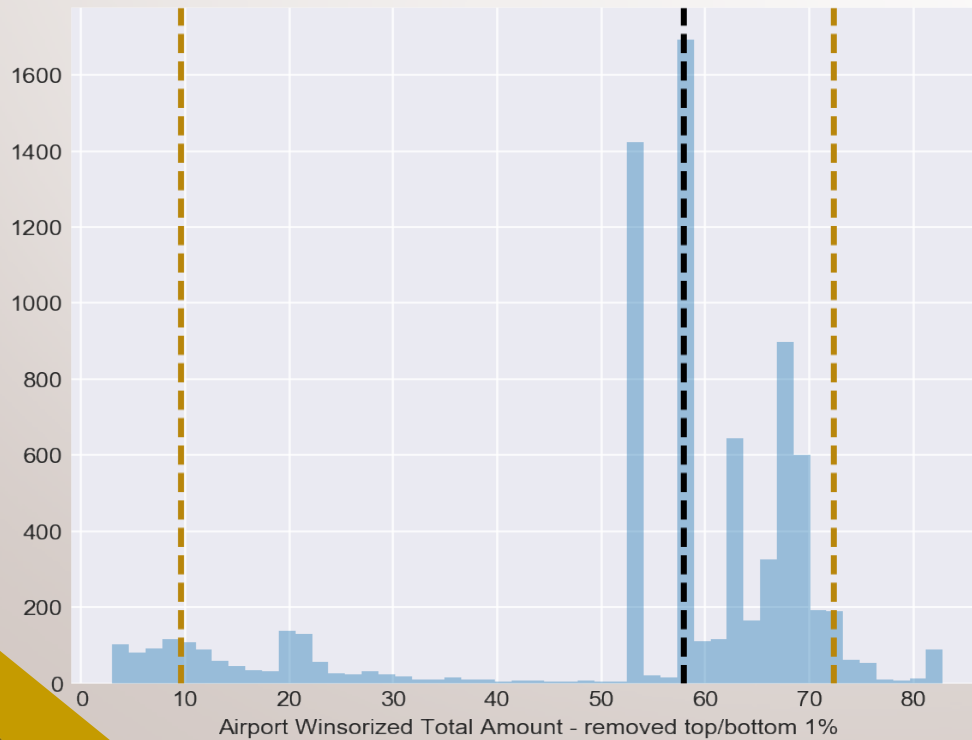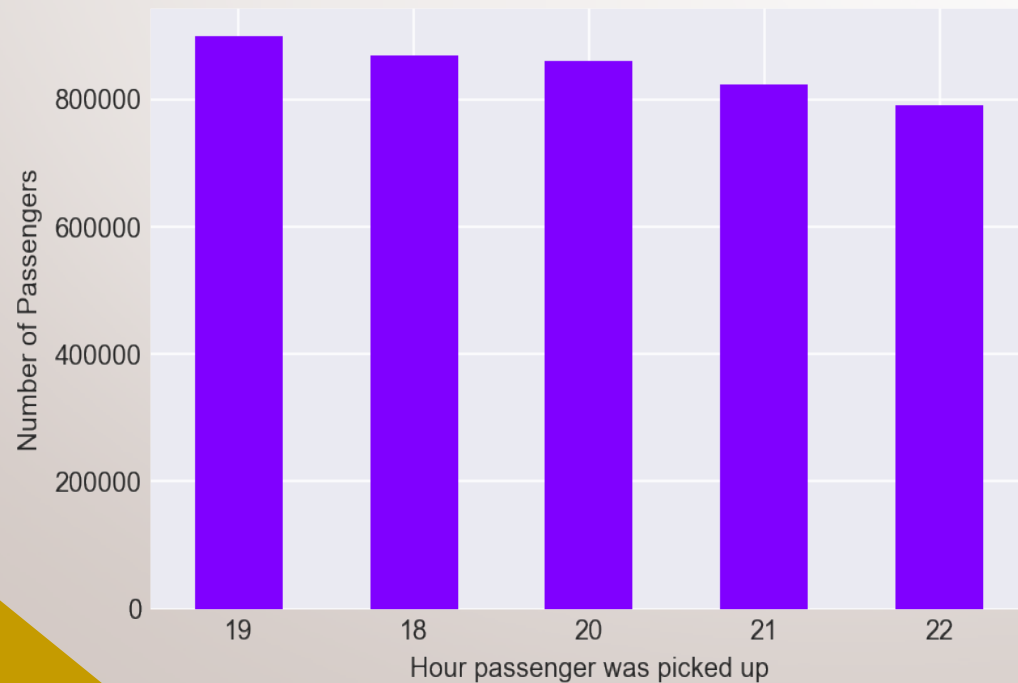Distribution of Tip Amount: Airport vs Non-Airport Fares
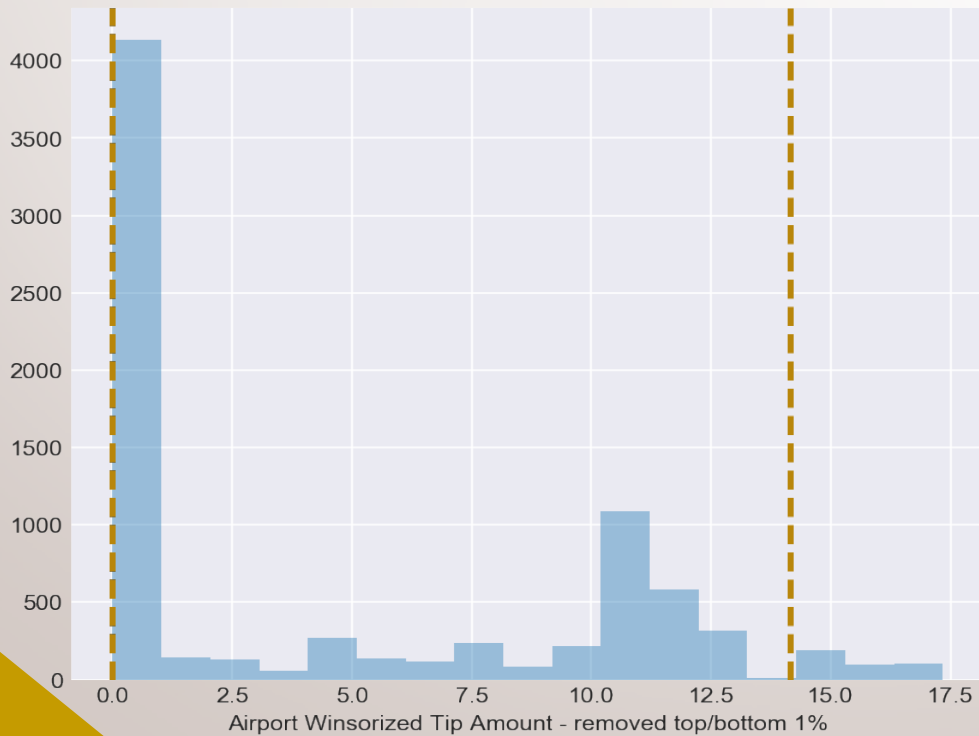
Airport Total Amounts Higher

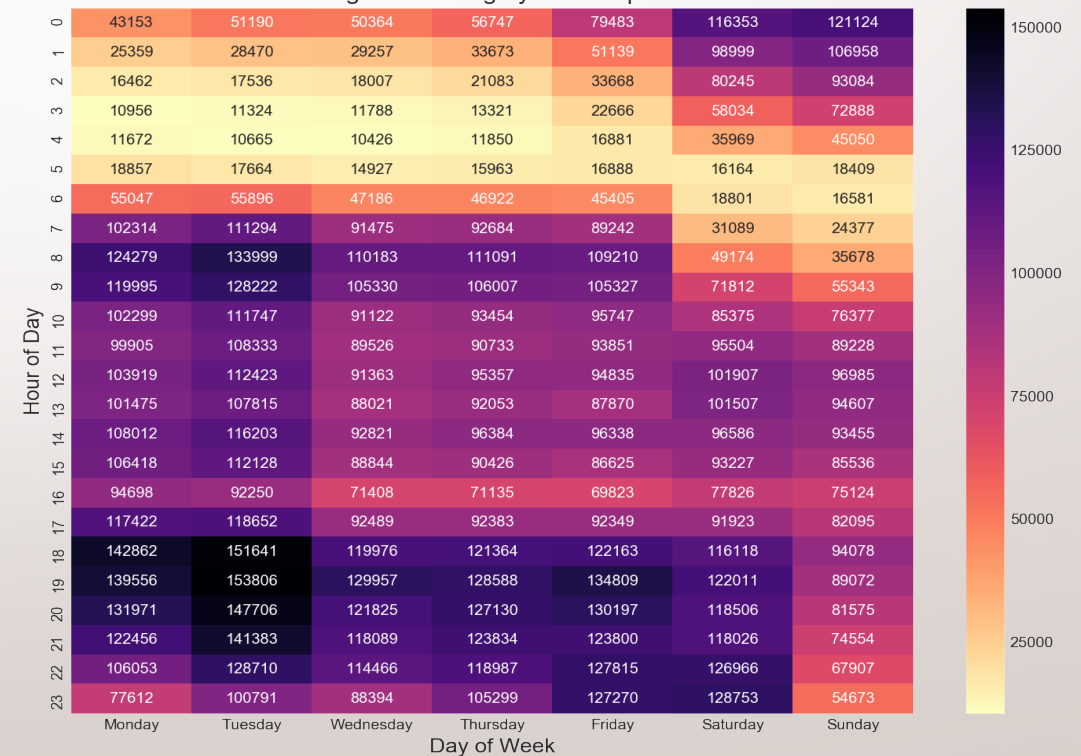Distribution of Total Amount: Airport vs Non-Airport Fares

Mon, Tue & weekend evenings are busiest

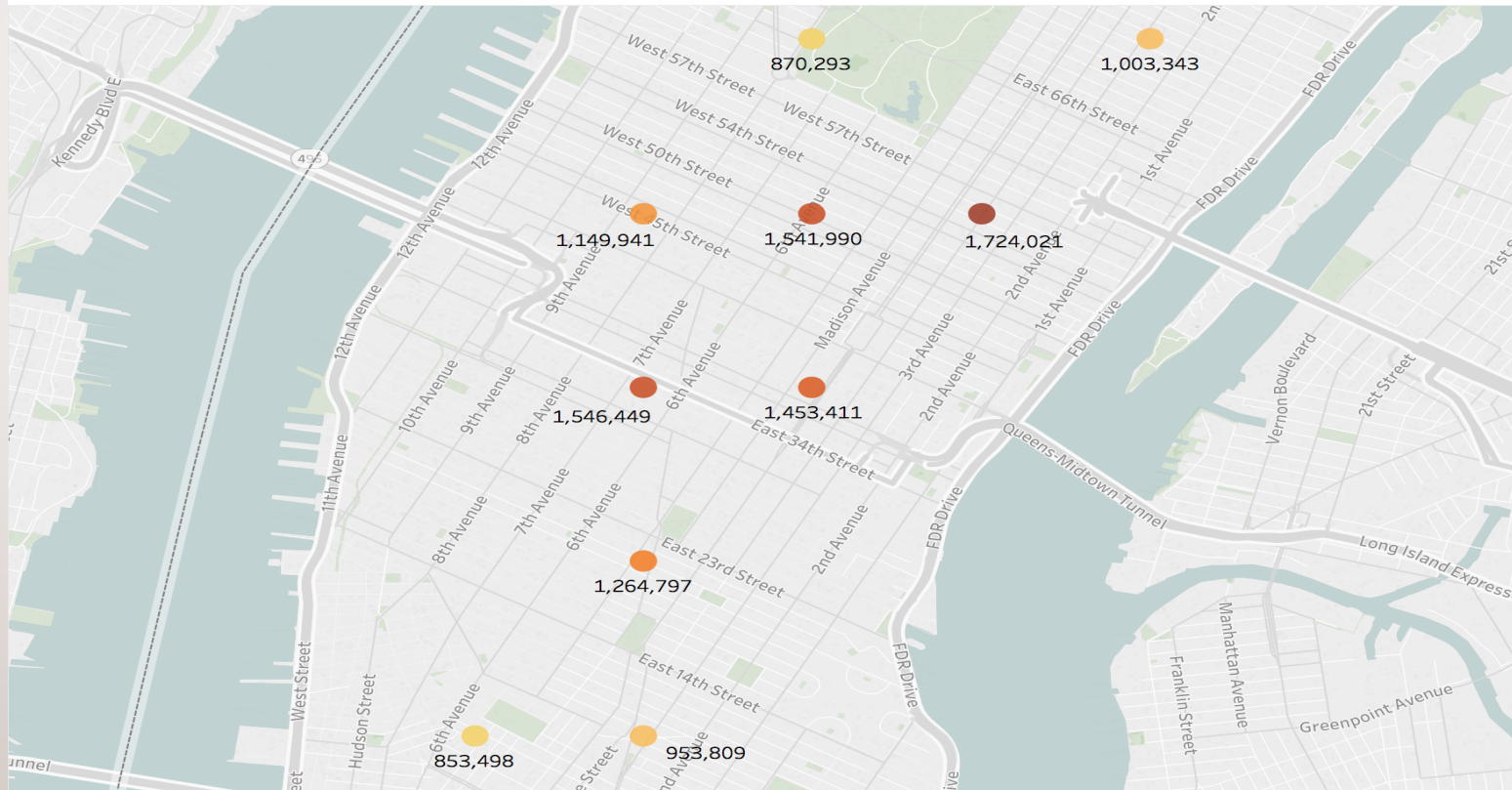When catching a cab, evenings after work/dinner are busiest

Passengers traveling by taxi in April 2013

Hour of Day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	43153	51190	50364	56747	79483	116353	121124
1	25359	28470	29257	33673	51139	98999	106958
2	16462	17536	18007	21083	33668	80245	93084
3	10956	11324	11788	13321	22666	58034	72888
4	11672	10665	10426	11850	16881	35969	45050
5	18857	17664	14927	15963	16888	16164	18409
6	55047	55896	47186	46922	45405	18801	16581
7	102314	111294	91475	92684	89242	31089	24377
8	124279	133999	110183	111091	109210	49174	35678
9	119995	128222	105330	106007	105327	71812	55343
10	102299	111747	91122	93454	95747	85375	76377
11	99905	108333	89526	90733	93851	95504	89228
12	103919	112423	91363	95357	94835	101907	96985
13	101475	107815	88021	92053	87870	101507	94607
14	108012	116203	92821	96384	96338	96586	93455
15	106418	112128	88844	90426	86625	93227	85536
16	94698	92250	71408	71135	69823	77826	75124
17	117422	118652	92489	92383	92349	91923	82095
18	142862	151641	119976	121364	122163	116118	94078
19	139556	153806	129957	128588	134809	122011	89072
20	131971	147706	121825	127130	130197	118506	81575
21	122456	141383	118089	123834	123800	118026	74554
22	106053	128710	114466	118987	127815	126966	67907
23	77612	100791	88394	105299	127270	128753	54673

Day of Week

12.36m trips in Manhattan

Ten Busiest Locations

870,293
1,003,343
1,149,941
1,541,990
1,724,021
1,546,449
1,453,411
1,264,797
853,498
953,809
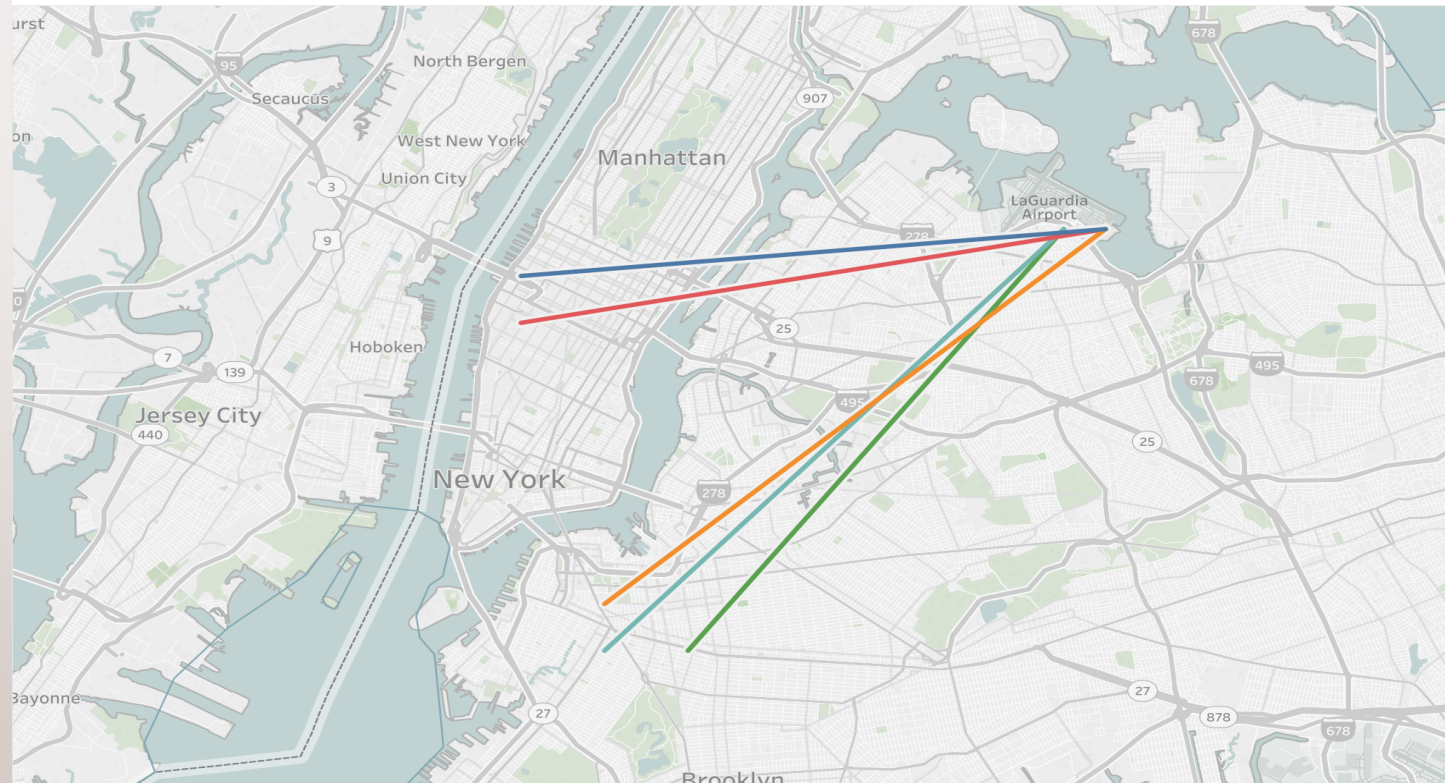
Map based on Longitude and Latitude. Color shows details about Total Trips. The marks are labeled by sum of Total Trips.
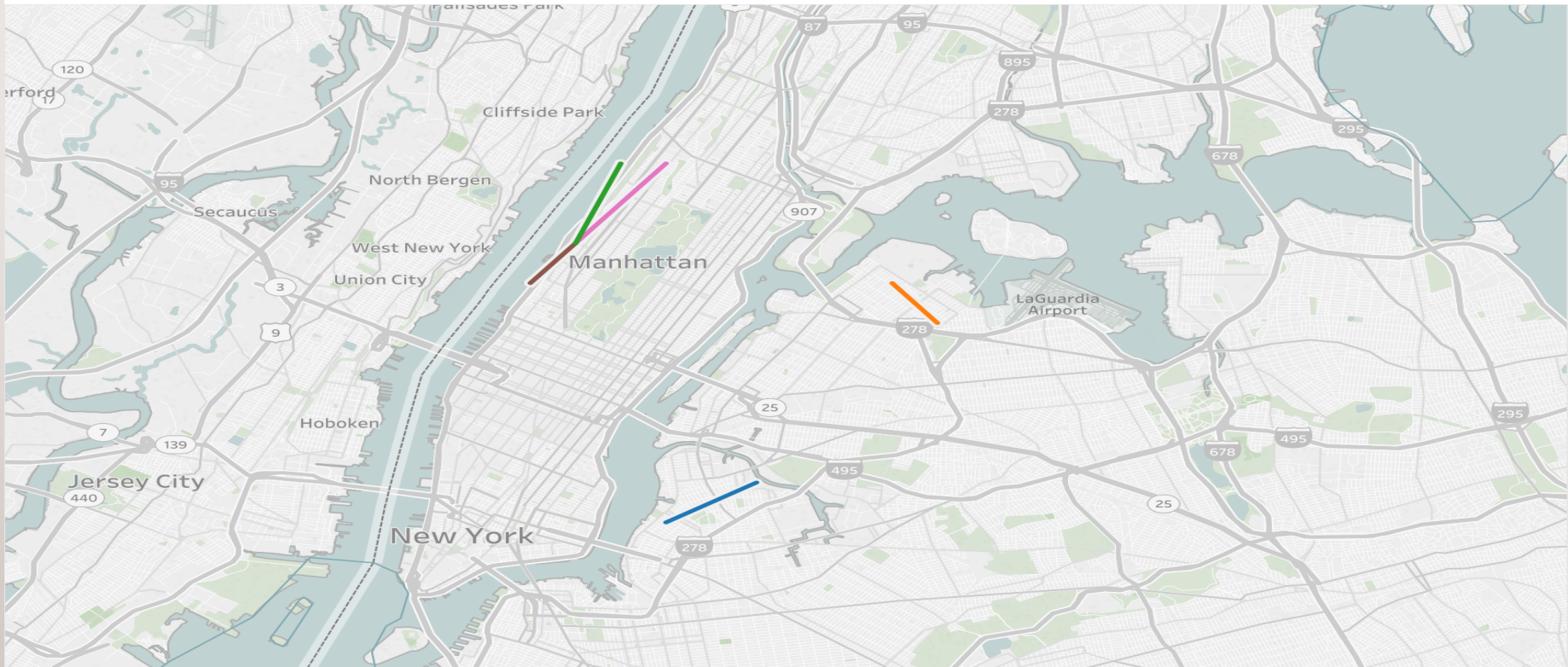
Airport travel times fluctuate most

Trips from LaGuardia Airport have the highest standard deviation of travel times

Map based on Longitude and Latitude. Color shows details about Path ID. Details are shown for Path ID.

Shorter fares reliable

Most consistent fares

Map based on Longitude and Latitude. Color shows details about Path ID. Details are shown for Path ID.

Minimum Sample Size

Margin of Error = 5%

Confidence Interval = 95% -> Z-Score = 1.96

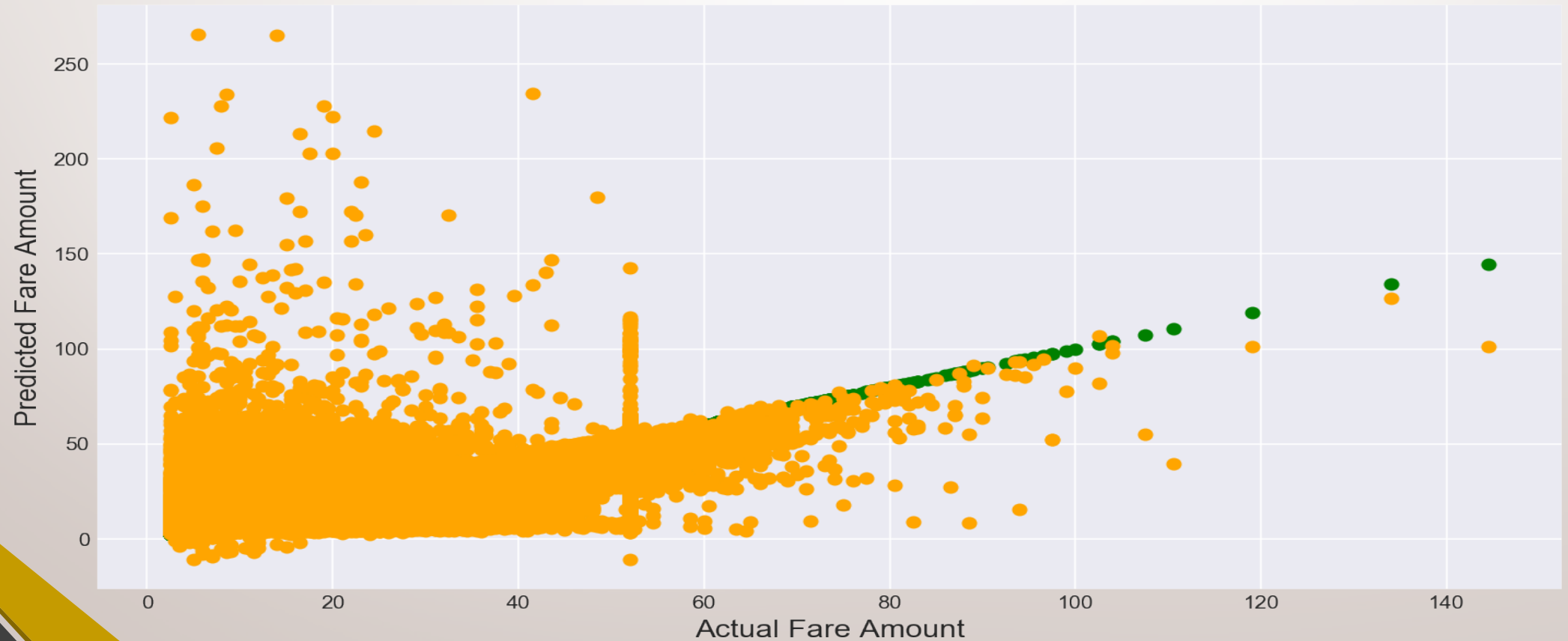**Standard Deviation = 50%**

Sample Size = (1.96*0.5/0.05)^2 = 385 trips

Features

1. Average Speed per hour
2. Trips per hour
3. Pickup Latitude/Longitude
4. Drop off Latitude/Longitude
5. Trip Distance
6. Pickup Hour
7. Day of Month
8. Taxi Shift

Predicting Taxi Fares

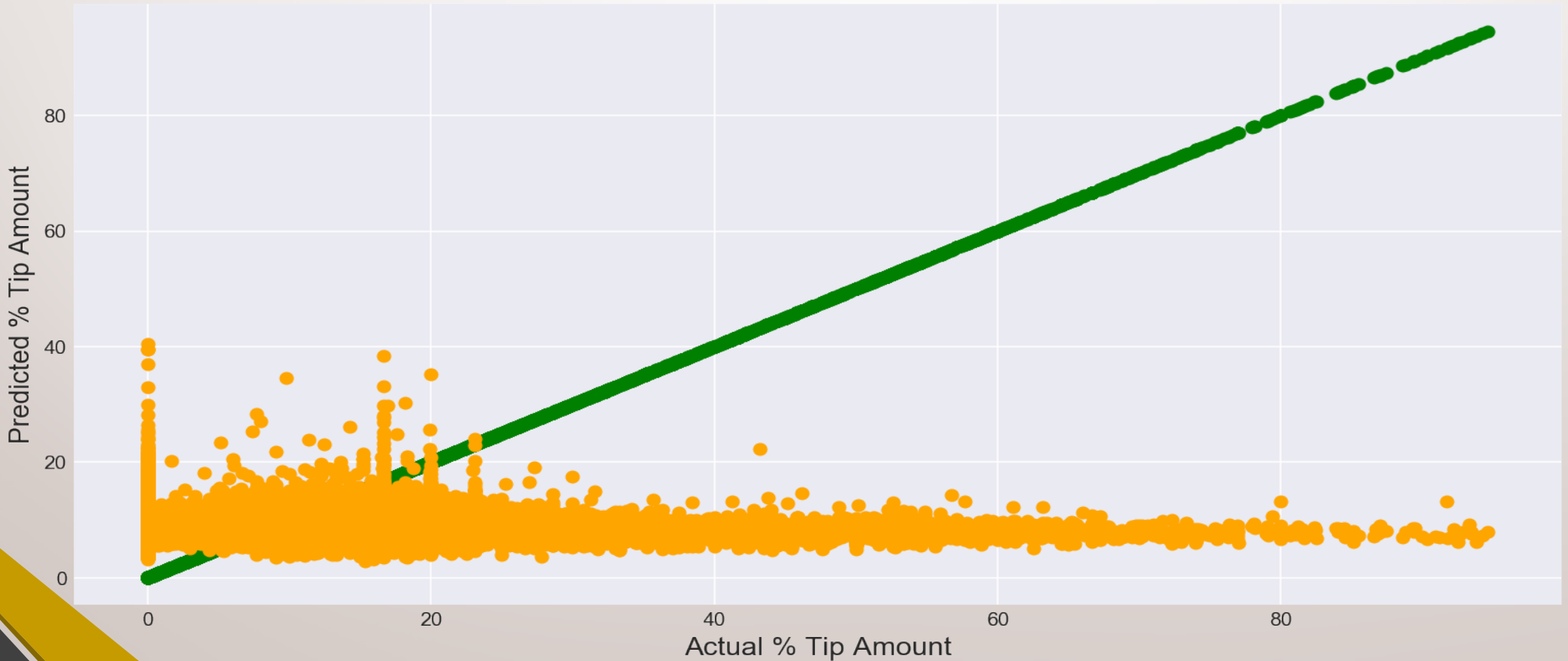

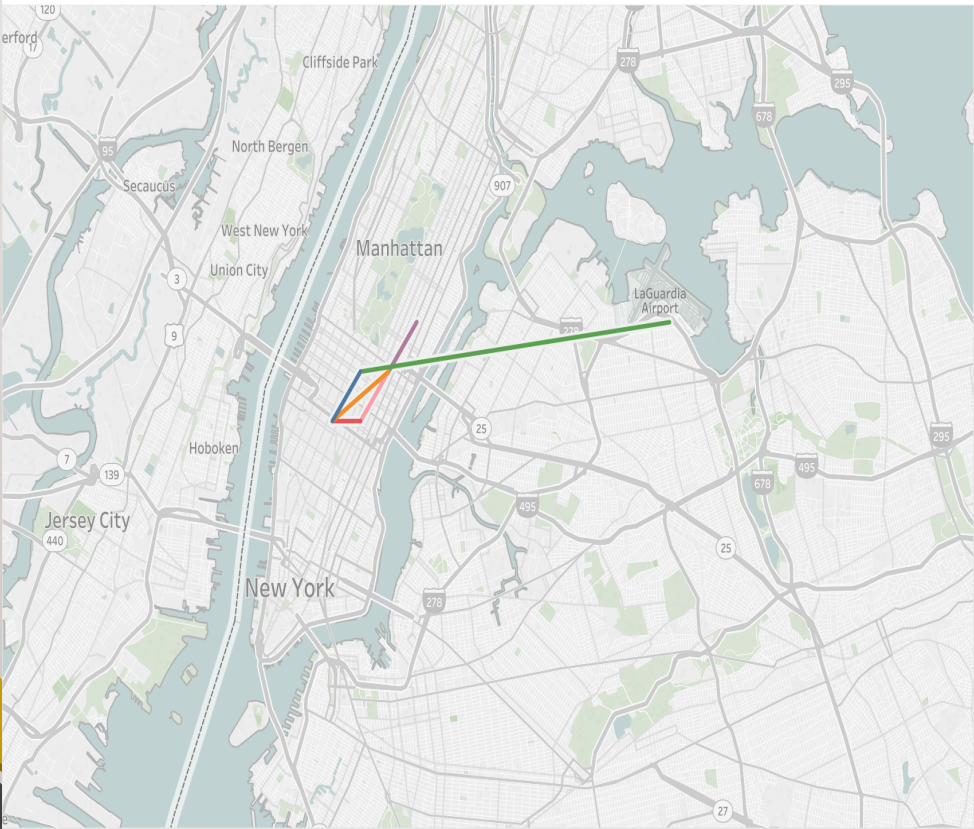OLS Prediction of NYC Taxi Non-Airport Fare Amounts has 82.5% R-Square

Predicting % Tip

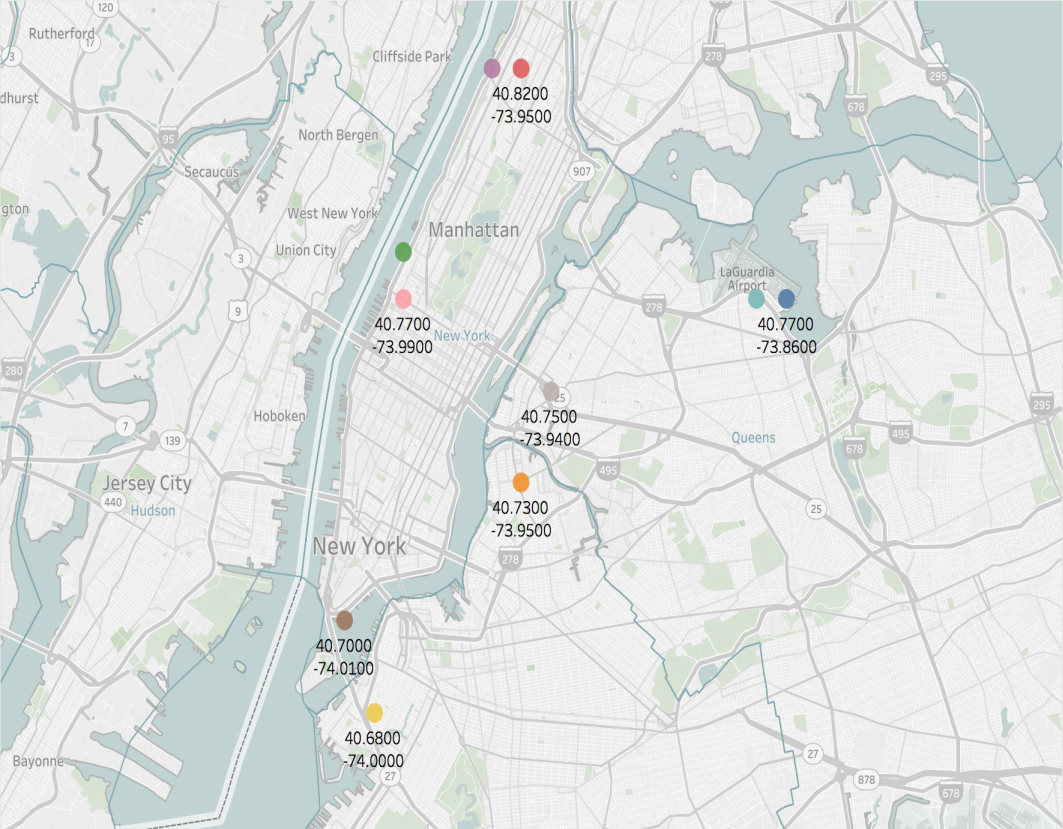OLS Prediction of NYC Taxi Tip Amounts has 1.2% R-Square

Maximize Daily Earnings

Highest Revenue

Max Revenue per Hour

Map based on Longitude and Latitude. Color shows details about Path ID.
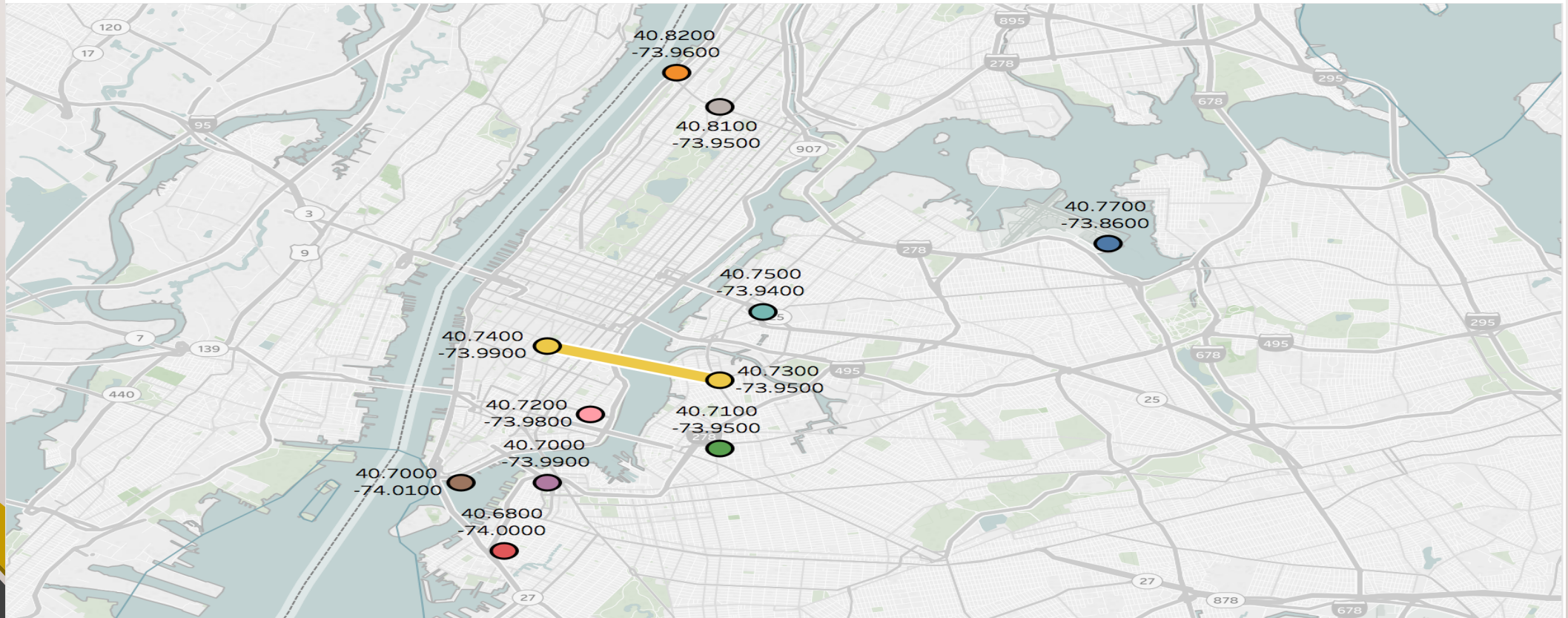
Map based on Longitude and Latitude. Color shows details about Path ID. The marks are labeled by Latitude and Longitude.
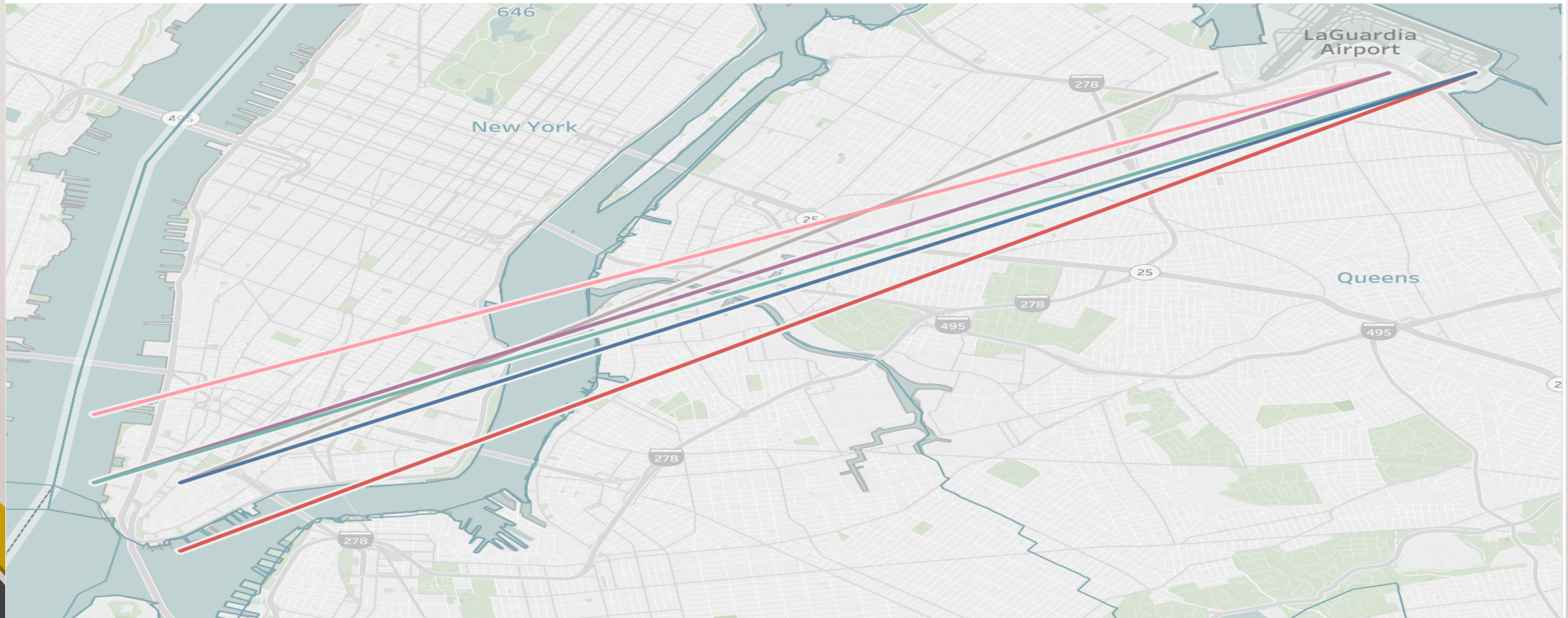
Profitable Evening Routes

Evening Profitable Routes

Map based on Longitude and Latitude. Color shows details about Path ID. The marks are labeled by average of Latitude and average of Longitude.

Non-Crowded Trips

Highest Revenue from Non-Crowded Trips

Map based on Longitude and Latitude. Color shows details about Path ID.