
Yellow Taxi
Maximize Income
BRONX
9/11
0201

Indent

- Objective
- Dataset
- EDA
- Model and Results
- Conclusion

Objective

Maximize your income as a taxi driver

How would you enrich dataset?

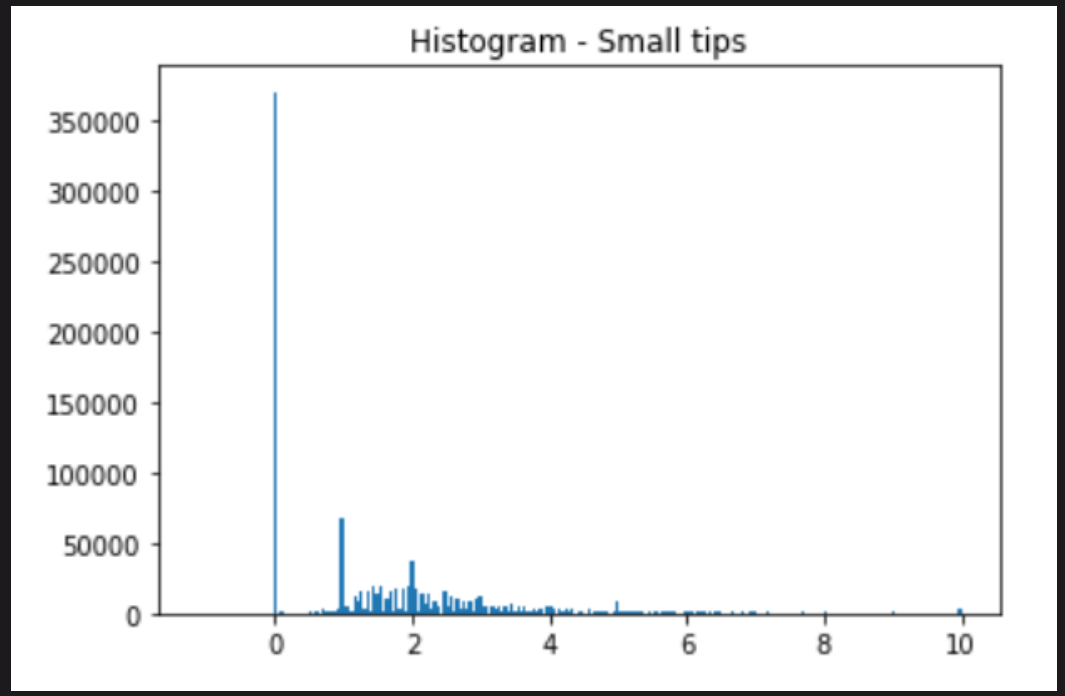
What you don’t find useful

Dataset

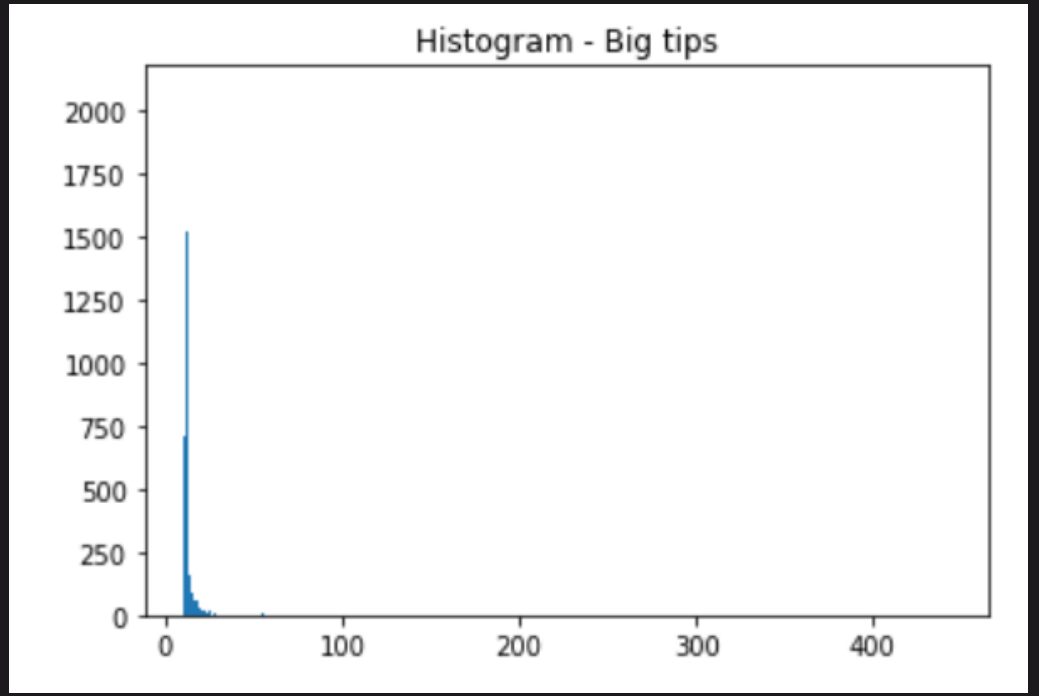
- Yellow Taxi data – June 2017
- 10,48,575 rides
- 266 locations

EDA results – Tip Distribution

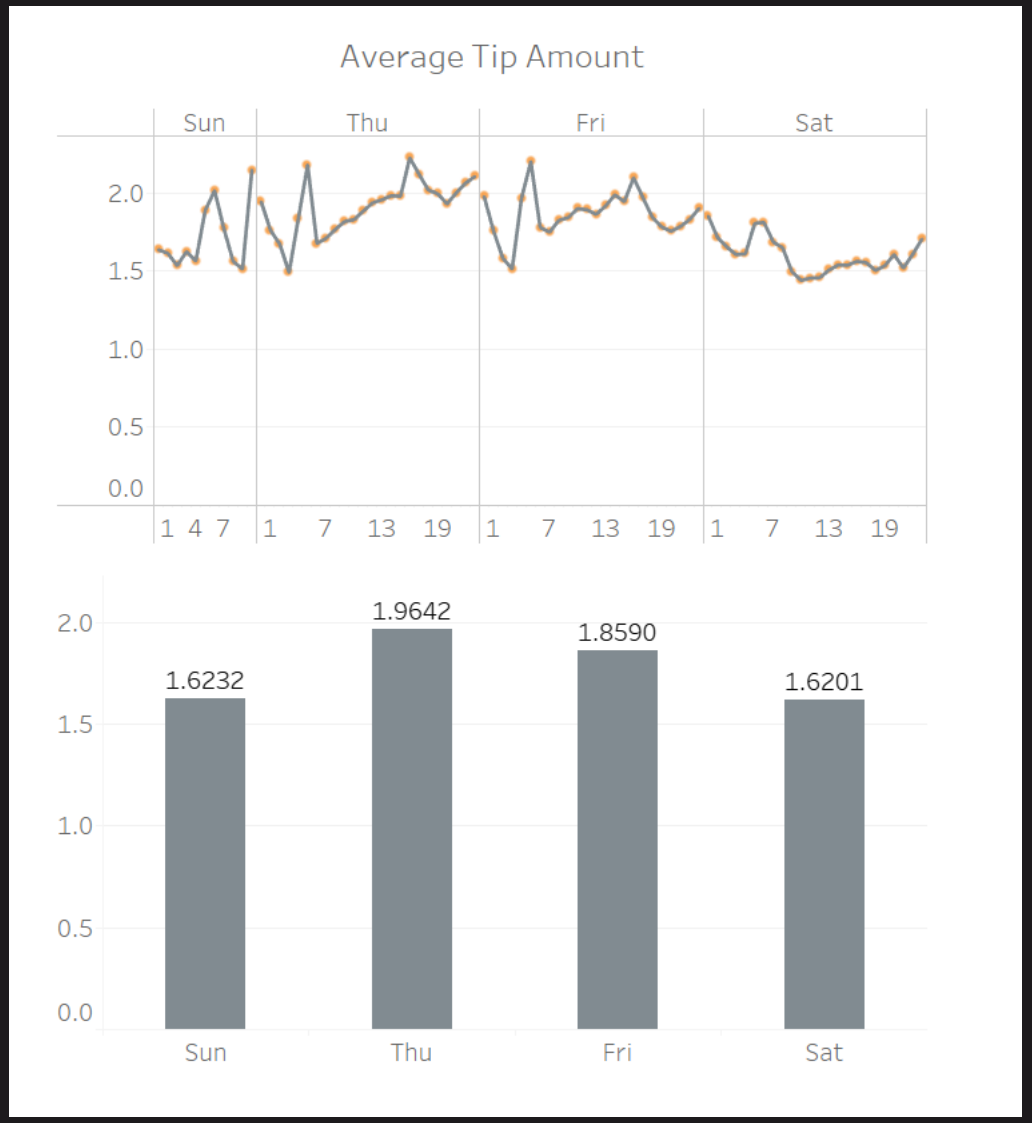
tip <= $10

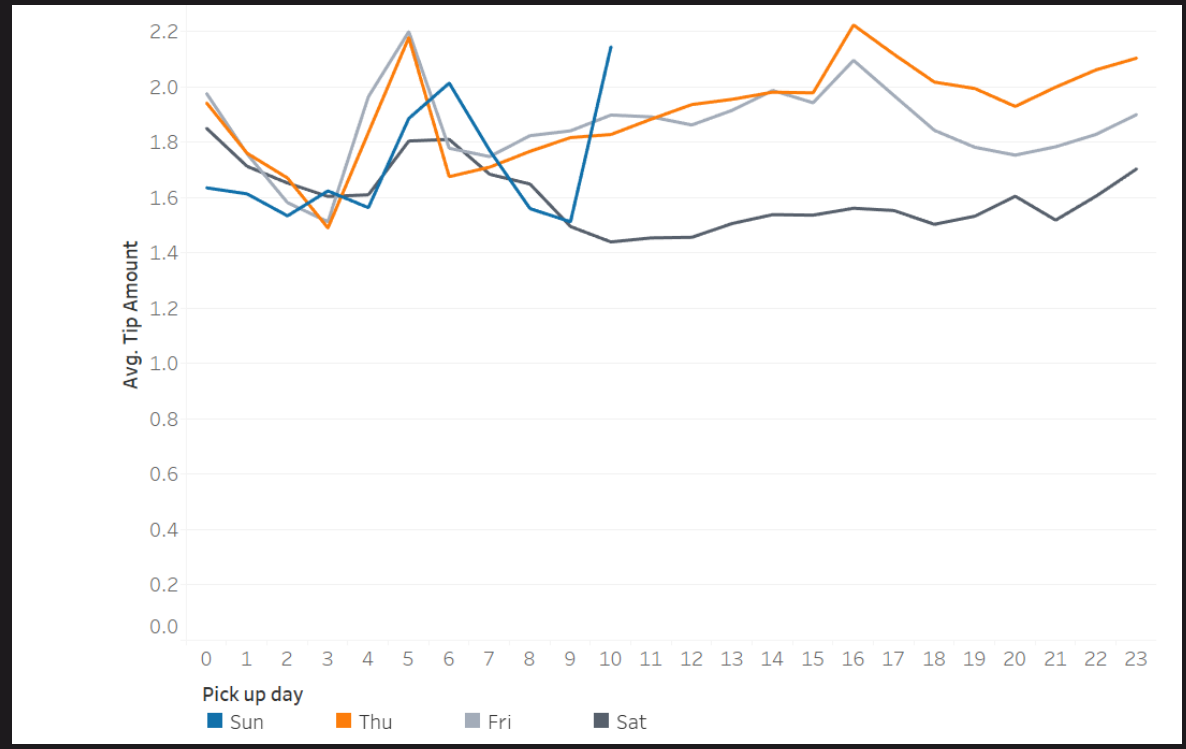


tip > $10



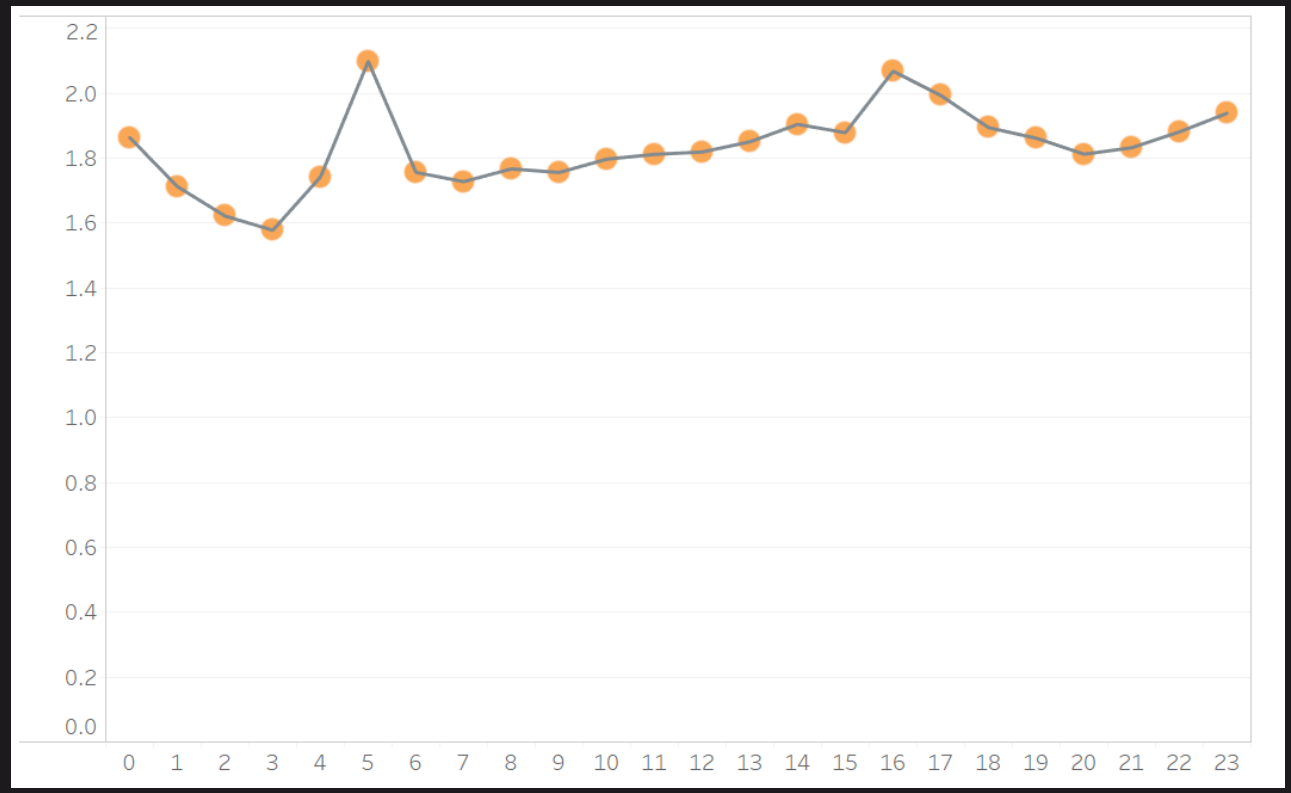
EDA - Average Tip Amount Distribution



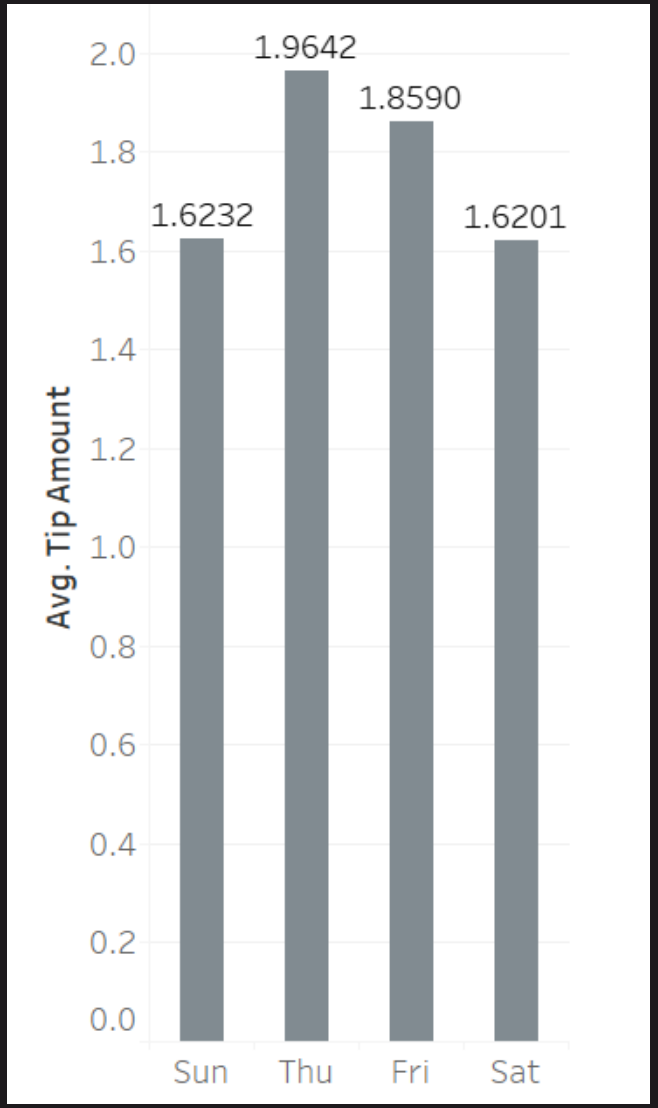


Average Tip Amount

By Pick Up hour



By Day of Week



Average tip amount by Pick Up Location

Zone pick	Borough pick	
South Beach/Dongan Hills	Staten Island	18.85
East Flushing	Queens	14.75
Kew Gardens	Queens	14.38
Stapleton	Staten Island	10.00
Van Nest/Morris Park	Bronx	8.95
Newark Airport	EWR	8.18
Howard Beach	Queens	7.54
Grymes Hill/Clifton	Staten Island	7.50
Highbridge Park	Manhattan	7.19
Coney Island	Brooklyn	7.19
Springfield Gardens South	Queens	7.18
Baisley Park	Queens	6.70
South Jamaica	Queens	6.35
Laurelton	Queens	5.85
Flushing Meadows-Corona Park	Queens	5.74
Bellerose	Queens	5.62
LaGuardia Airport	Queens	5.57
Allerton/Pelham Gardens	Bronx	5.43
JFK Airport	Queens	5.31
Hillcrest/Pomonok	Queens	5.15
South Ozone Park	Queens	4.99
Saint Albans	Queens	4.90
Springfield Gardens North	Queens	4.78
Rosedale	Queens	4.46
Richmond Hill	Queens	4.43

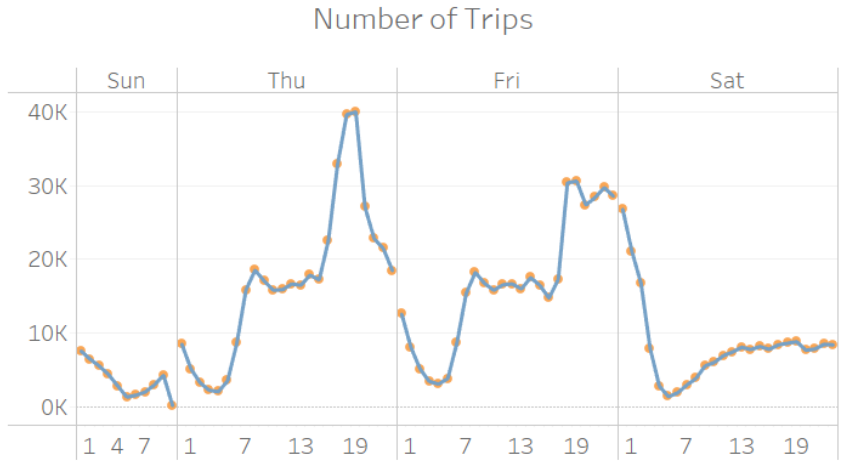
Top 25 pick up zones

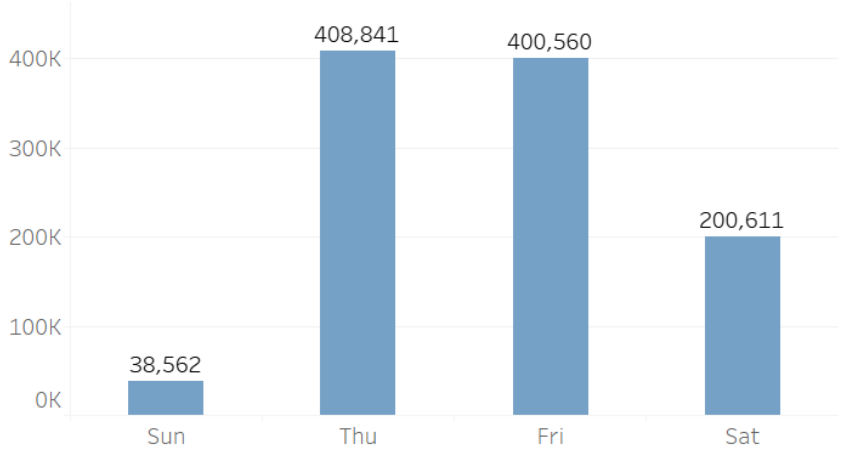
EDA results – Number of Trips

By Zones

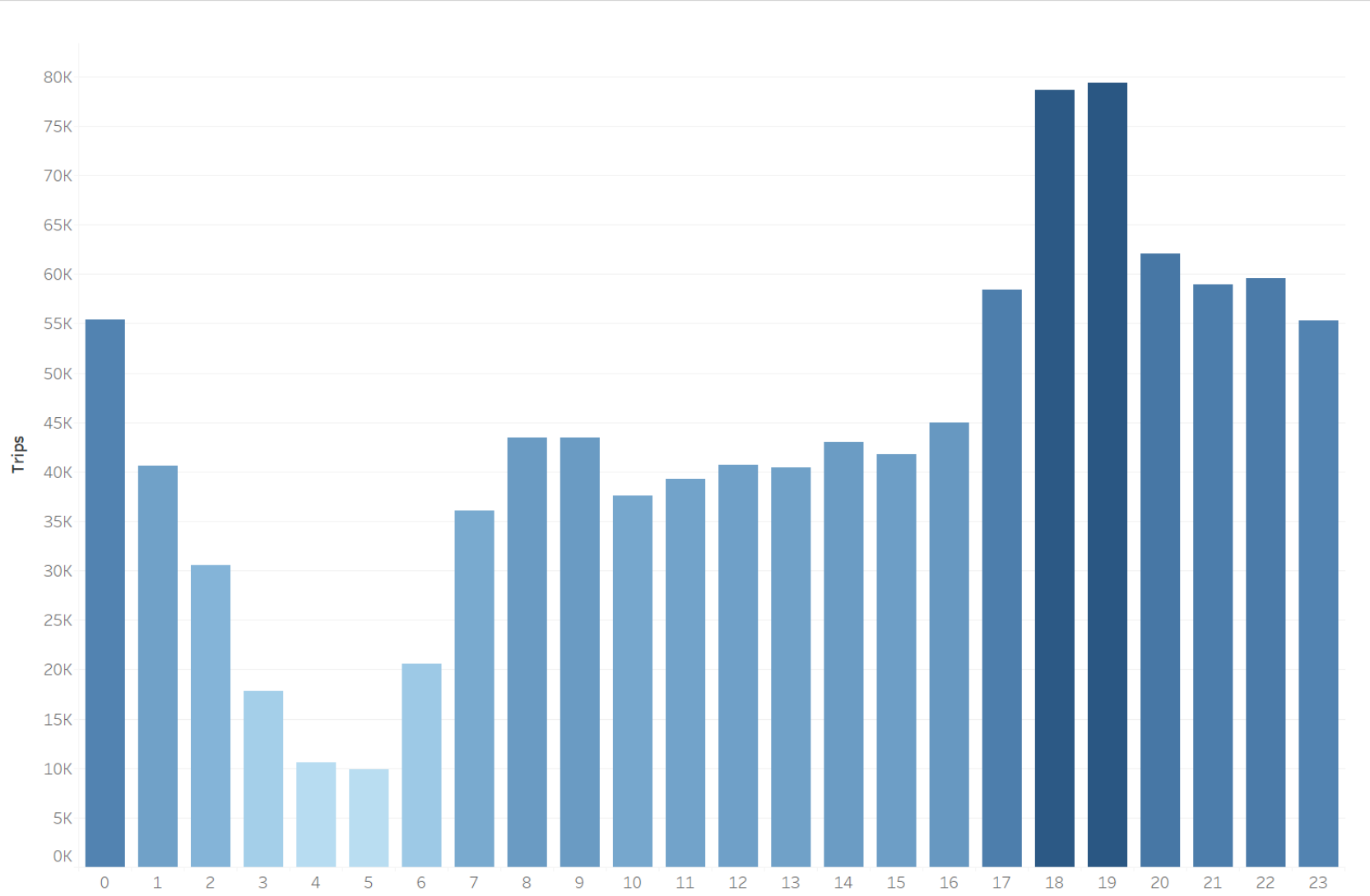
Zone pick	Borough pick	
LaGuardia Airport	Queens	24,791
JFK Airport	Queens	21,904
Baisley Park	Queens	201
Flushing Meadows-Corona Park	Queens	113
South Ozone Park	Queens	88
Newark Airport	EWR	72
South Jamaica	Queens	63
Kew Gardens	Queens	43
Springfield Gardens South	Queens	42
Richmond Hill	Queens	25
Van Nest/Morris Park	Bronx	17
Howard Beach	Queens	10
Bellerose	Queens	5
Highbridge Park	Manhattan	5
Rosedale	Queens	5
Coney Island	Brooklyn	4
Springfield Gardens North	Queens	4
Allerton/Pelham Gardens	Bronx	3
Hillcrest/Pomonok	Queens	3
Laurelton	Queens	2
Saint Albans	Queens	2
East Flushing	Queens	1
Grymes Hill/Clifton	Staten Island	1
South Beach/Dongan Hills	Staten Island	1
Stapleton	Staten Island	1

By Day and Time



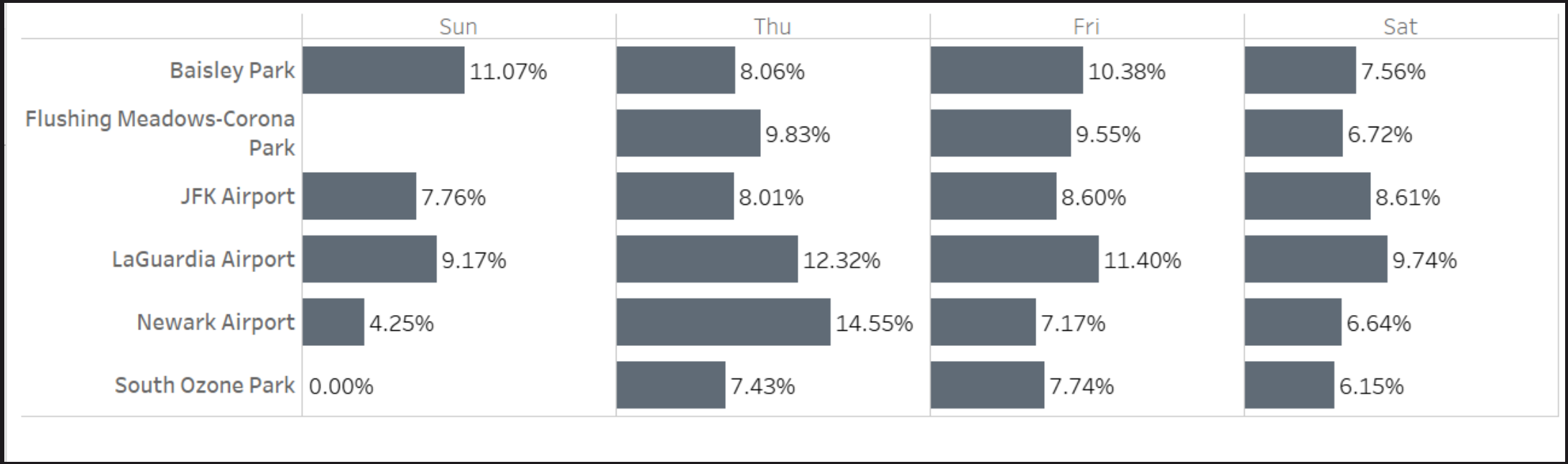


Number of Trips by Pick up hour



Peak hours in our dataset : 18, 19

Further analysis of our top 6 pick up locations



More than 10% tips at

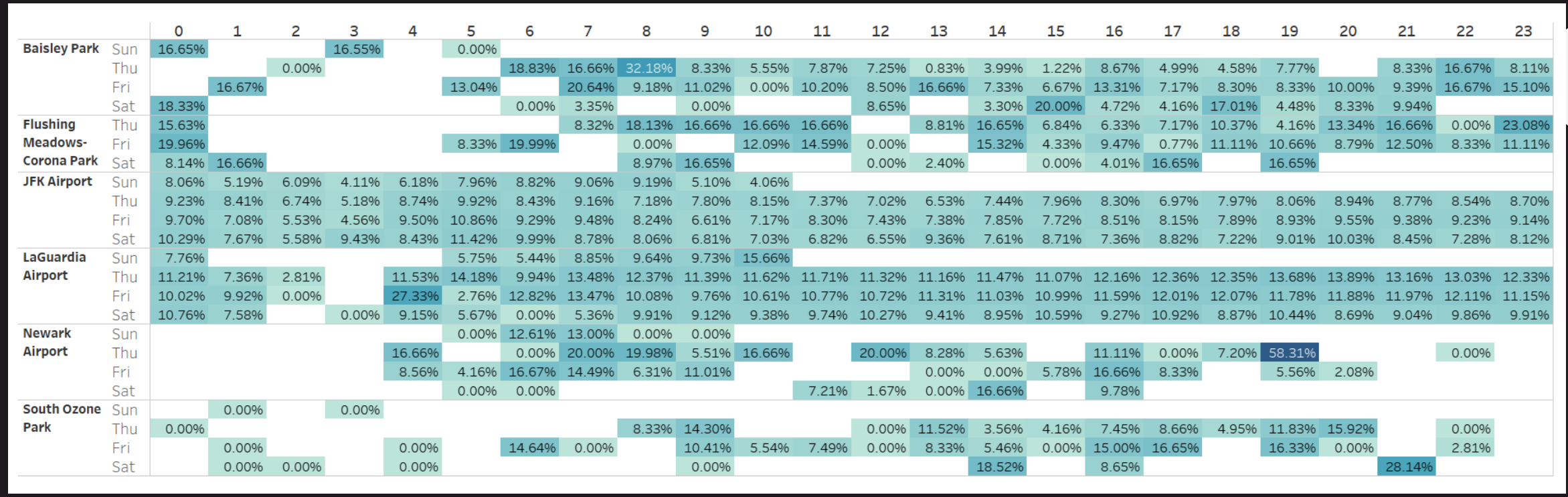
- Baisley Park on Friday and Sunday
- La Guardia Airport on Thursday and Friday
- Newark Airport on Thursday

Further analysis of our top 6 pick up locations

	Baisley Park	Flushing Meadows-Corona P..	JFK Airport	LaGuardia Airport	Newark Airport	South Ozone Park
0	17.77%	11.97%	9.76%	10.46%		0.00%
1	16.67%	16.66%	7.51%	7.46%		0.00%
2	0.00%		6.20%	1.40%		0.00%
3	16.55%		6.14%	0.00%		0.00%
4			8.53%	14.98%	12.61%	0.00%
5	6.52%	8.33%	10.24%	7.78%	1.39%	
6	12.55%	19.99%	9.08%	7.23%	6.98%	14.64%
7	14.22%	8.32%	9.22%	13.03%	15.49%	0.00%
8	17.54%	12.06%	7.87%	10.97%	4.66%	8.33%
9	6.92%	16.66%	6.83%	10.29%	7.70%	9.11%
10	3.33%	14.38%	7.64%	10.88%	16.66%	5.54%
11	9.42%	15.63%	7.73%	11.14%	7.21%	7.49%
12	7.81%	0.00%	7.13%	11.01%	10.84%	0.00%
13	7.16%	5.61%	7.26%	10.99%	4.14%	9.92%
14	5.33%	15.65%	7.64%	11.11%	8.91%	6.38%
15	5.90%	5.14%	8.00%	10.98%	5.78%	2.78%
16	10.08%	7.39%	8.25%	11.73%	11.95%	10.25%
17	5.33%	5.15%	7.53%	12.13%	4.17%	11.32%
18	7.13%	10.81%	7.87%	12.11%	7.20%	4.95%
19	7.48%	8.66%	8.47%	12.83%	31.93%	12.95%
20	9.16%	11.06%	9.35%	12.64%	2.08%	7.96%
21	9.35%	13.33%	9.06%	12.33%		28.14%
22	16.67%	5.55%	8.70%	12.36%	0.00%	2.11%
23	11.61%	14.10%	8.89%	11.59%		

Late evenings fair better tips than afternoon.

Further analysis of our top 6 pick up locations



		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Baisley Park	Sun	16.65%			16.55%		0.00%																		
	Thu			0.00%				18.83%	16.66%	32.18%	8.33%	5.55%	7.87%	7.25%	0.83%	3.99%	1.22%	8.67%	4.99%	4.58%	7.77%		8.33%	16.67%	8.11%
	Fri		16.67%				13.04%		20.64%	9.18%	11.02%	0.00%	10.20%	8.50%	16.66%	7.33%	6.67%	13.31%	7.17%	8.30%	8.33%	10.00%	9.39%	16.67%	15.10%
	Sat	18.33%						0.00%	3.35%		0.00%			8.65%		3.30%	20.00%	4.72%	4.16%	17.01%	4.48%	8.33%	9.94%		
Flushing Meadows-Corona Park	Thu	15.63%							8.32%	18.13%	16.66%	16.66%	16.66%		8.81%	16.65%	6.84%	6.33%	7.17%	10.37%	4.16%	13.34%	16.66%	0.00%	23.08%
	Fri	19.96%					8.33%	19.99%		0.00%		12.09%	14.59%	0.00%		15.32%	4.33%	9.47%	0.77%	11.11%	10.66%	8.79%	12.50%	8.33%	11.11%
	Sat	8.14%	16.66%							8.97%	16.65%			0.00%	2.40%		0.00%	4.01%	16.65%		16.65%				
JFK Airport	Sun	8.06%	5.19%	6.09%	4.11%	6.18%	7.96%	8.82%	9.06%	9.19%	5.10%	4.06%													
	Thu	9.23%	8.41%	6.74%	5.18%	8.74%	9.92%	8.43%	9.16%	7.18%	7.80%	8.15%	7.37%	7.02%	6.53%	7.44%	7.96%	8.30%	6.97%	7.97%	8.06%	8.94%	8.77%	8.54%	8.70%
	Fri	9.70%	7.08%	5.53%	4.56%	9.50%	10.86%	9.29%	9.48%	8.24%	6.61%	7.17%	8.30%	7.43%	7.38%	7.85%	7.72%	8.51%	8.15%	7.89%	8.93%	9.55%	9.38%	9.23%	9.14%
	Sat	10.29%	7.67%	5.58%	9.43%	8.43%	11.42%	9.99%	8.78%	8.06%	6.81%	7.03%	6.82%	6.55%	9.36%	7.61%	8.71%	7.36%	8.82%	7.22%	9.01%	10.03%	8.45%	7.28%	8.12%
LaGuardia Airport	Sun	7.76%					5.75%	5.44%	8.85%	9.64%	9.73%	15.66%													
	Thu	11.21%	7.36%	2.81%		11.53%	14.18%	9.94%	13.48%	12.37%	11.39%	11.62%	11.71%	11.32%	11.16%	11.47%	11.07%	12.16%	12.36%	12.35%	13.68%	13.89%	13.16%	13.03%	12.33%
	Fri	10.02%	9.92%	0.00%		27.33%	2.76%	12.82%	13.47%	10.08%	9.76%	10.61%	10.77%	10.72%	11.31%	11.03%	10.99%	11.59%	12.01%	12.07%	11.78%	11.88%	11.97%	12.11%	11.15%
	Sat	10.76%	7.58%		0.00%	9.15%	5.67%	0.00%	5.36%	9.91%	9.12%	9.38%	9.74%	10.27%	9.41%	8.95%	10.59%	9.27%	10.92%	8.87%	10.44%	8.69%	9.04%	9.86%	9.91%
Newark Airport	Sun						0.00%	12.61%	13.00%	0.00%	0.00%														
	Thu					16.66%		0.00%	20.00%	19.98%	5.51%	16.66%		20.00%	8.28%	5.63%		11.11%	0.00%	7.20%	58.31%			0.00%	
	Fri					8.56%	4.16%	16.67%	14.49%	6.31%	11.01%				0.00%	0.00%	5.78%	16.66%	8.33%		5.56%	2.08%			
	Sat						0.00%	0.00%					7.21%	1.67%	0.00%	16.66%		9.78%							
South Ozone Park	Sun		0.00%		0.00%																				
	Thu	0.00%								8.33%	14.30%			0.00%	11.52%	3.56%	4.16%	7.45%	8.66%	4.95%	11.83%	15.92%		0.00%	
	Fri		0.00%			0.00%		14.64%	0.00%		10.41%	5.54%	7.49%	0.00%	8.33%	5.46%	0.00%	15.00%	16.65%		16.33%	0.00%		2.81%	
	Sat		0.00%	0.00%		0.00%					0.00%					18.52%		8.65%					28.14%		

I would select location and time where tip percentage is greater than 15%

Maximize your income as a taxi driver

- As mentioned in the problem statement , I want to work 10 hours each week, therefore I would want these 10 hours to be most productive.
- I would be looking up areas and days and hour of the day where I can get rides with minimum waiting time and maximum income

Approach 1

- I can first select locations where waiting time is minimum or select locations with maximum pickups
- After selecting locations I can look for optimum day of the week and time of the day to work based on tip amount or tip amount percentage (of total amount)

Approach 2

- For best utilization of time, I can check total amount earned by trip duration. For example money earned per minute of trip. And analyze data so as to maximize this metric

Maximize your income as a taxi driver

Pick areas with frequent pick ups

- Airports and Parks seemed to be popular pick up location in our dataset
  - Airports
    - LaGuardia Airport
    - JFK Airport
    - Newark Airport
  - Parks
    - Baisley
    - Flushing Meadows-Corona
    - South Ozone

Maximize your income as a taxi driver

Choose location where customers pay maximum tip

- Customers are likely to tip more in Staten Island

Borough pick	
Staten Island	12.117
Bronx	8.420
EWR	8.177
Manhattan	7.190
Brooklyn	7.188
Queens	5.460

- Though Queens is better if you check Zone wise

Zone pick	Borough pick	
South Beach/Dongan Hills	Staten Island	18.85
East Flushing	Queens	14.75
Kew Gardens	Queens	14.38
Stapleton	Staten Island	10.00
Van Nest/Morris Park	Bronx	8.95
Newark Airport	EWR	8.18
Howard Beach	Queens	7.54
Grymes Hill/Clifton	Staten Island	7.50
Highbridge Park	Manhattan	7.19
Coney Island	Brooklyn	7.19
Springfield Gardens South	Queens	7.18
Baisley Park	Queens	6.70
South Jamaica	Queens	6.35
Laurelton	Queens	5.85
Flushing Meadows-Corona Park	Queens	5.74
Bellerose	Queens	5.62
LaGuardia Airport	Queens	5.57
Allerton/Pelham Gardens	Bronx	5.43
JFK Airport	Queens	5.31
Hillcrest/Pomonok	Queens	5.15
South Ozone Park	Queens	4.99
Saint Albans	Queens	4.90
Springfield Gardens North	Queens	4.78
Rosedale	Queens	4.46
Richmond Hill	Queens	4.43

Maximize your income as a taxi driver

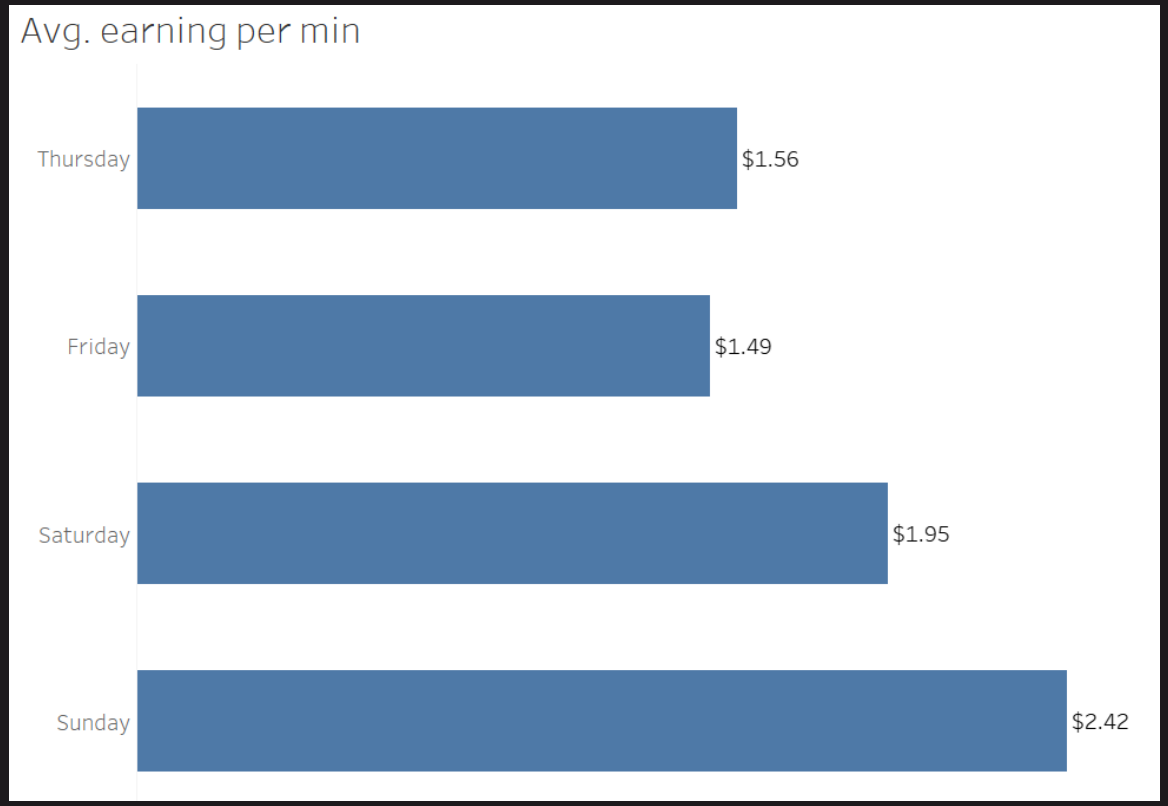
Choose time and day when customers pay maximum tip consider all locations

- Drivers earn highest tips when customer is picked up around **5 am**
- Drivers earn highest tips when customer is picked up on **Thursday**
- Combining day and Time analysis –
  - **4 pm on Thursday & 5 am on Thursday and Friday** customers pay highest

Maximize your income as a taxi driver

Pick day by average earnings per min (selected pickups only)

- Average Earning per min =
  Total Amount / Trip Duration in mins
- Weekends have better average than Weekday (our dataset only has data for Thurs , Fri, Sat and Sun. Therefore, this analysis may be biased)
- This analysis is only for our top 6 pickup locations



Maximize your income as a taxi driver

Using Data Science - Getting feature importance

- After initial data analysis, I ran random forest model with tip percentage as our target variable
- This is to understand the impact of features on the target variable
- Due to computing limitation I have run the code for 100,000 records. Therefore, results might differ on entire dataset
- As per these results, passenger_count (95%) and trip_distance (5%) are most important to get higher tip percentage
- One can say that when more people travel in group they tend to tip more than individual customers

Enrich dataset

- Features added in dataset:
  - Pick up hour
  - Pick up day
  - Trip duration
  - Avg. earning
  - Tip percentage

How would I enrich the dataset

- Adding longitude and latitude information of pick ups would help in better geo analysis
- The data is missing information on cash tips. I would add that for better selection of hours and days the driver should work
- The current dataset only contains information for days – Thurs, Fri, Sat, Sun. I would add full weeks data for better analysis of weekend vs weekdays
- Once we have geo locations we can also do analysis for optimizing the route taken to save time

What you don’t find useful

- Payment Method
  - The dataset has maximum Credit Card Payment rides.
  - This field doesn’t give us any important information for maximizing income
- Store_and_fwd_flag
- Service zone in the taxi zone lookup table

Conclusion

Recommended time and Location

- **Newark Airport –**
  Thursday late evenings – 8 pm to 1 am
  Friday early mornings - 4 am to 7 am
- **Baisley Park –**
  Saturday 3 pm – 4pm
- **South Ozone Park –**
  Saturday 9 pm – 10 pm