NYC Taxi Analysis

Ryan Knight

April 2019

Summary

I have researched June 2017 NYC taxi data to identify the best time to drive a taxi for 10 hours during that month. I decided to focus on finding opportunities to drive during high demand so that the driver would be guaranteed to stay busy and find a lot of quick, recurring high value trips. Other opportunities may exist in airport trips due to extra revenue but those trips may also take a driver out of high demand zones.

I recommend the driver drive a few hours each Thursday evening after 5pm in Manhattan with a focus on the Upper East Side. This is where demand is the highest and trips are tend to be short, high revenue trips that stay in the area.

Revenue was calculated as fare amount + extra fees + tips. This is based on the descriptions of those fields and the determination that those amounts go to the driver and other amounts like taxes go to the government. For example, the improvement surcharge is a tax based on this article:
https://www.nytimes.com/2014/05/01/nyregion/city-approves-30-surcharge-to-pay-for-accessible-taxis.html

The highest demand times exist in Manhattan on Thursday and Friday evenings after 5pm. Additionally, over 90% of those rides stay in Manhattan. Very late on those evenings there is a slight increase in rides beginning in Manhattan and leaving the borough for other areas possibly as riders return home to other boroughs. Because of this I recommend the driver focus on the 5-8pm time window so that they are more likely to stay in the high demand area.

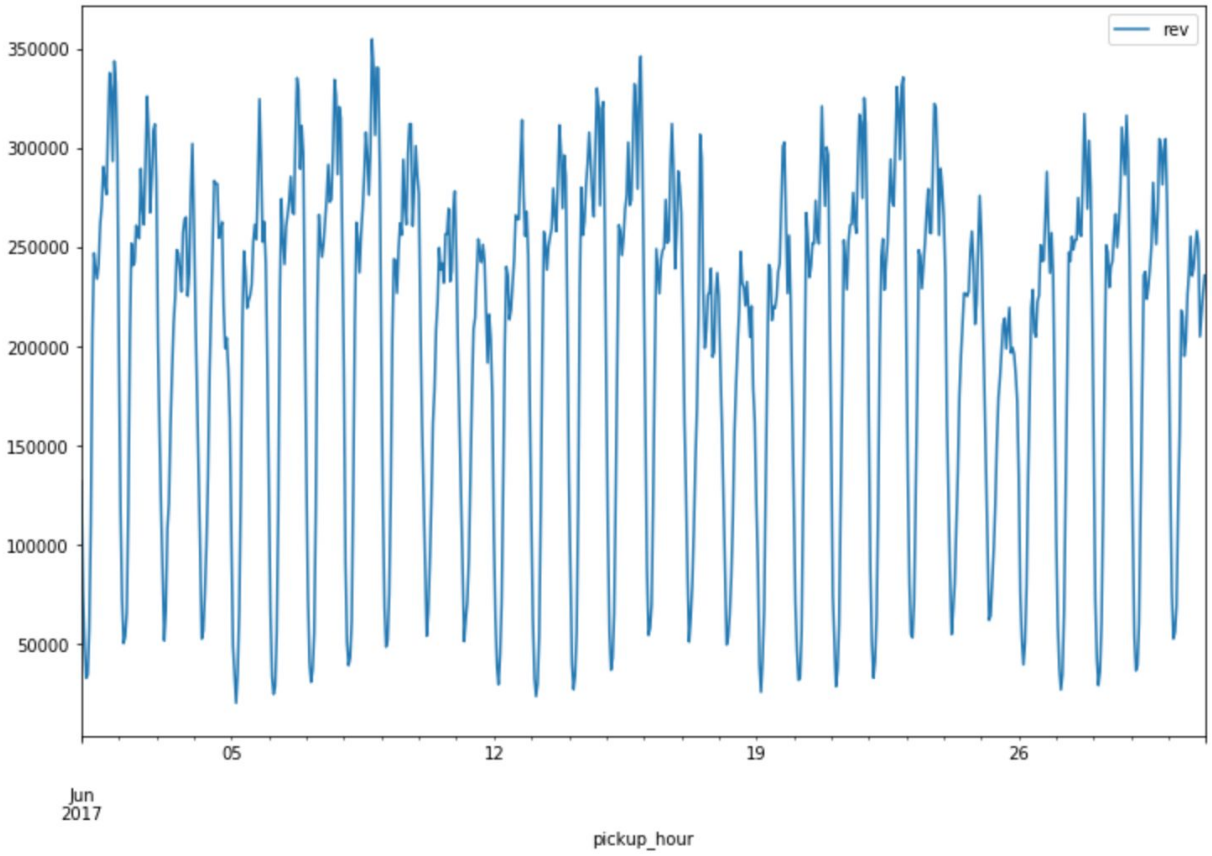
Missing data points include wait times, competition, and traffic impact. It is unknown but expected that drivers may have to wait longer at airports in long queues. Down time because of waiting or competition could negatively impact the revenue per hour. Also, the median revenue per minute is actually higher on Monday than on Thursday and Friday in Manhattan. This may be due to traffic and trips taking longer or struggles to find the next rider later in the week for a driver. More information on the impact these factors have on the ability for a driver to earn revenue consistently would be helpful.

Analysis Charts

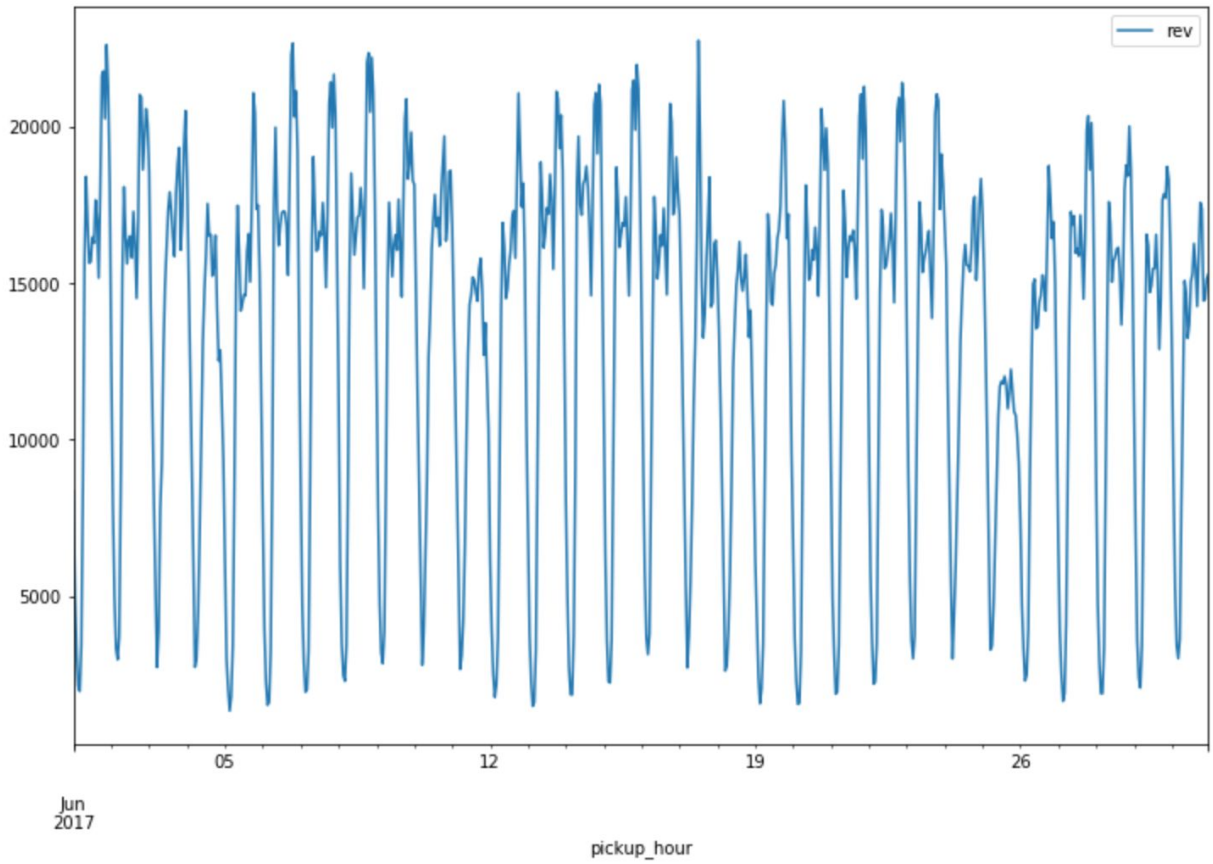
Initial data summaries show some messy data. Some trips have negative revenue and distances or times are either too low at zero or way too high. The following data is removed: trips with revenue over $200 or less than $0, trip times less than 1 minute, dropoff time before pickup time.

	rev	revpermi	revpermin	triptime	passenger_count	trip_distance	fare_amount	tip_amount	extra	revpermi
count	9,656,993.0000	9,655,547.0000	9,656,483.0000	9,656,993.0000	9,656,993.0000	9,656,993.0000	9,656,993.0000	9,656,993.0000	9,656,993.0000	9,655,547.0000
mean	15.5071	nan	nan	16.8659	1.6239	2.9786	13.2873	1.8785	0.3413	nan
std	215.2709	nan	nan	55.3363	1.2646	5.7041	215.1675	2.6962	0.4623	nan
min	-550.0000	-inf	-inf	0.0000	0.0000	0.0000	-550.0000	-74.0000	-50.5600	-inf
25%	7.9500	5.0000	0.8824	6.6833	1.0000	1.0000	6.5000	0.0000	0.0000	5.0000
50%	11.0400	6.4630	1.0545	11.2833	1.0000	1.6700	9.5000	1.3600	0.0000	6.4630
75%	17.2000	8.4375	1.2981	18.6667	2.0000	3.1000	15.0000	2.4600	0.5000	8.4375
max	630,462.3200	inf	inf	14,407.6500	9.0000	9,496.9800	630,461.8200	444.0000	22.5000	inf

Total revenue by pickup hour in the month. Clearly there is a general trend each day and a weekly cycle.

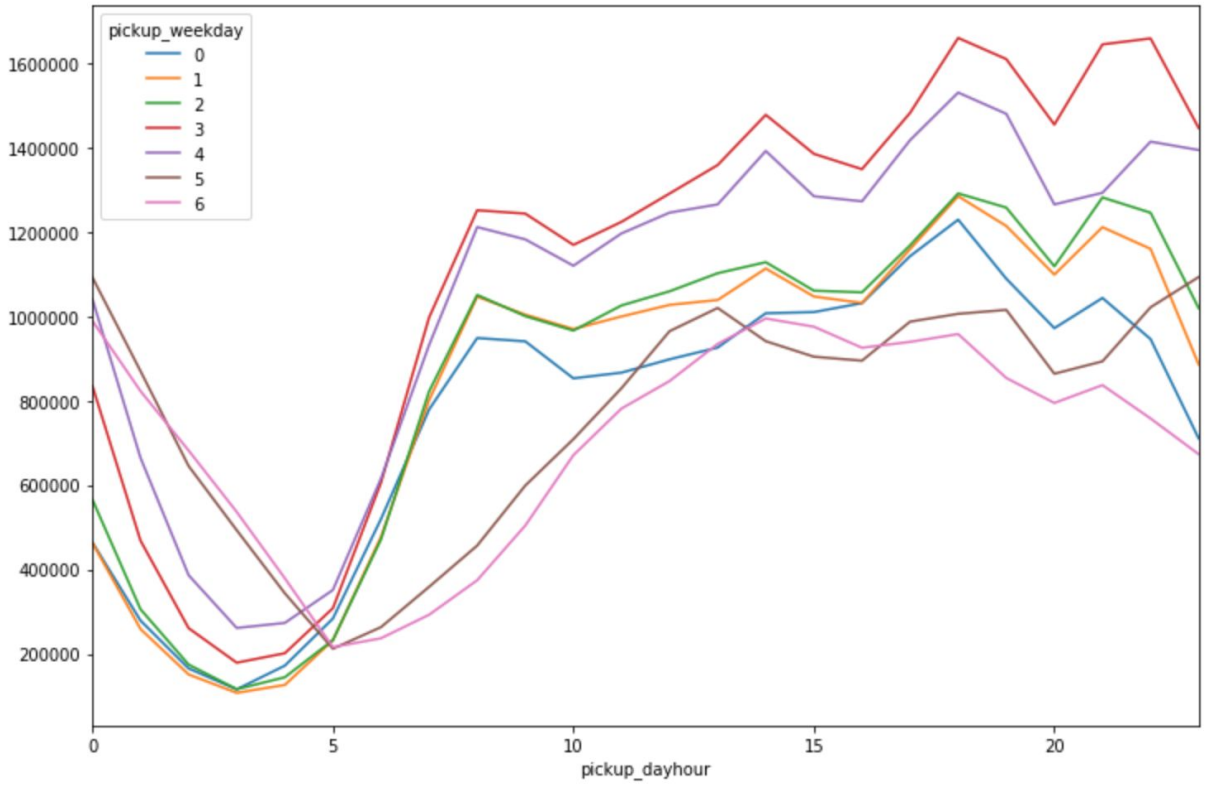


Count of trips by hour during the month has a similar pattern.



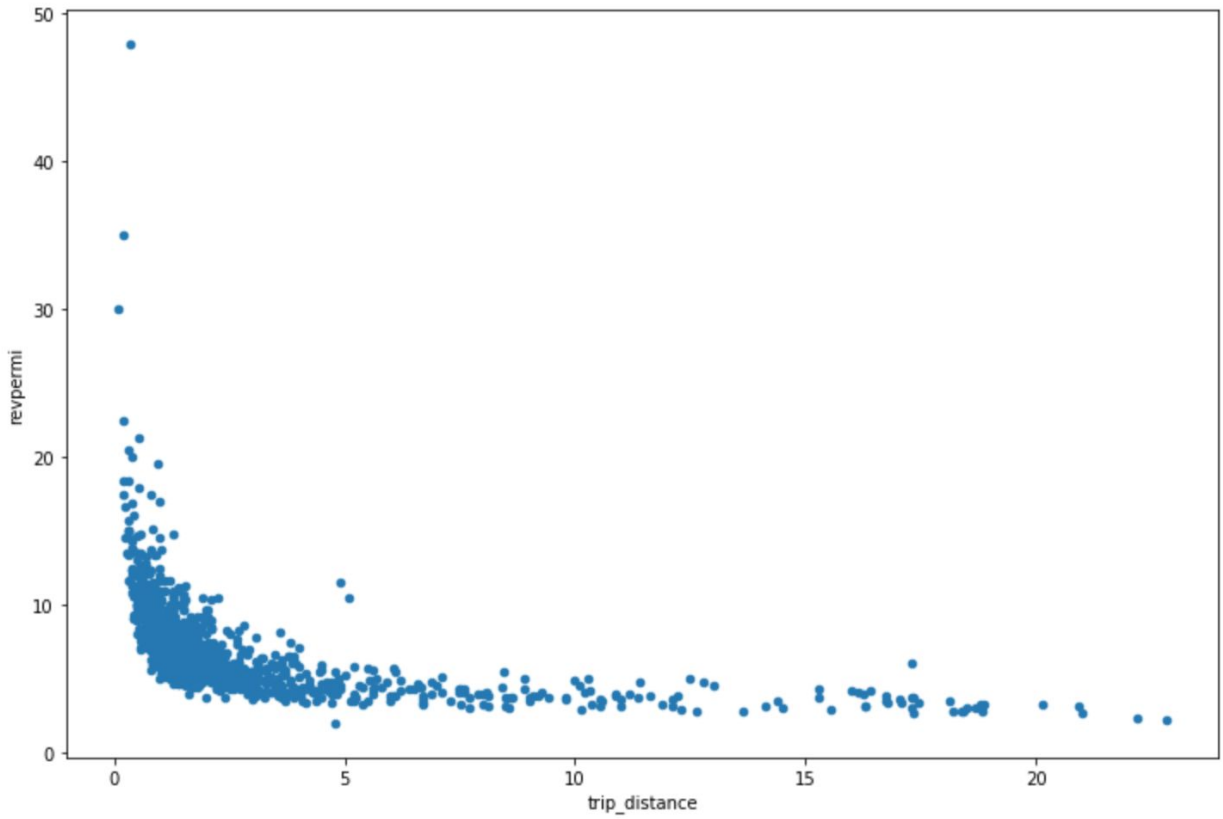
Comparing total revenue for each day of the week it looks like Thursday (day 3) has the highest demand followed by Friday (day 4).

Total Revenue by Pickup Day of Week and Hour



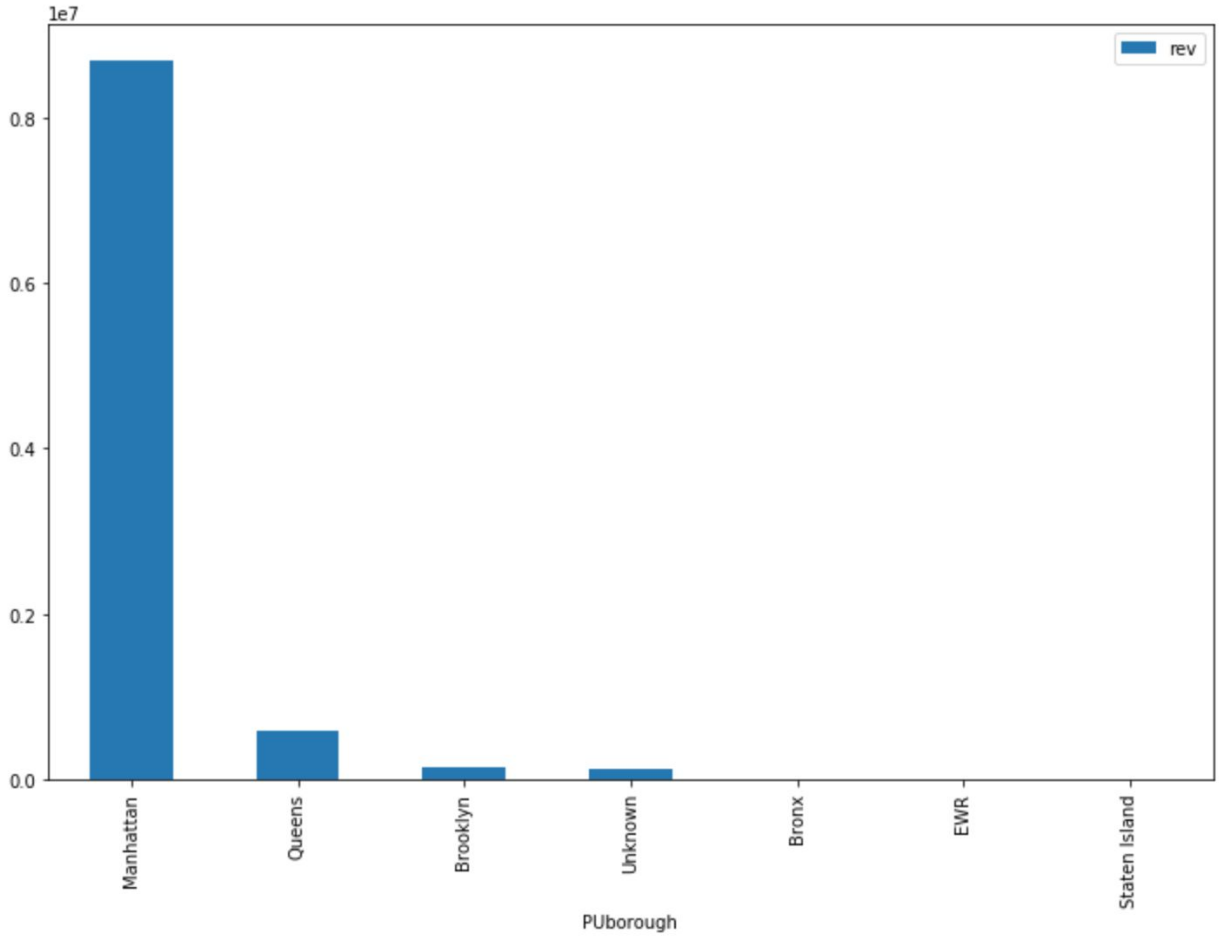
Looking at revenue per mile versus trip distance we see that shorter trips tend to pay better partly from fixed pricing that every trip receives.

Revenue per Mile versus Trip Distance



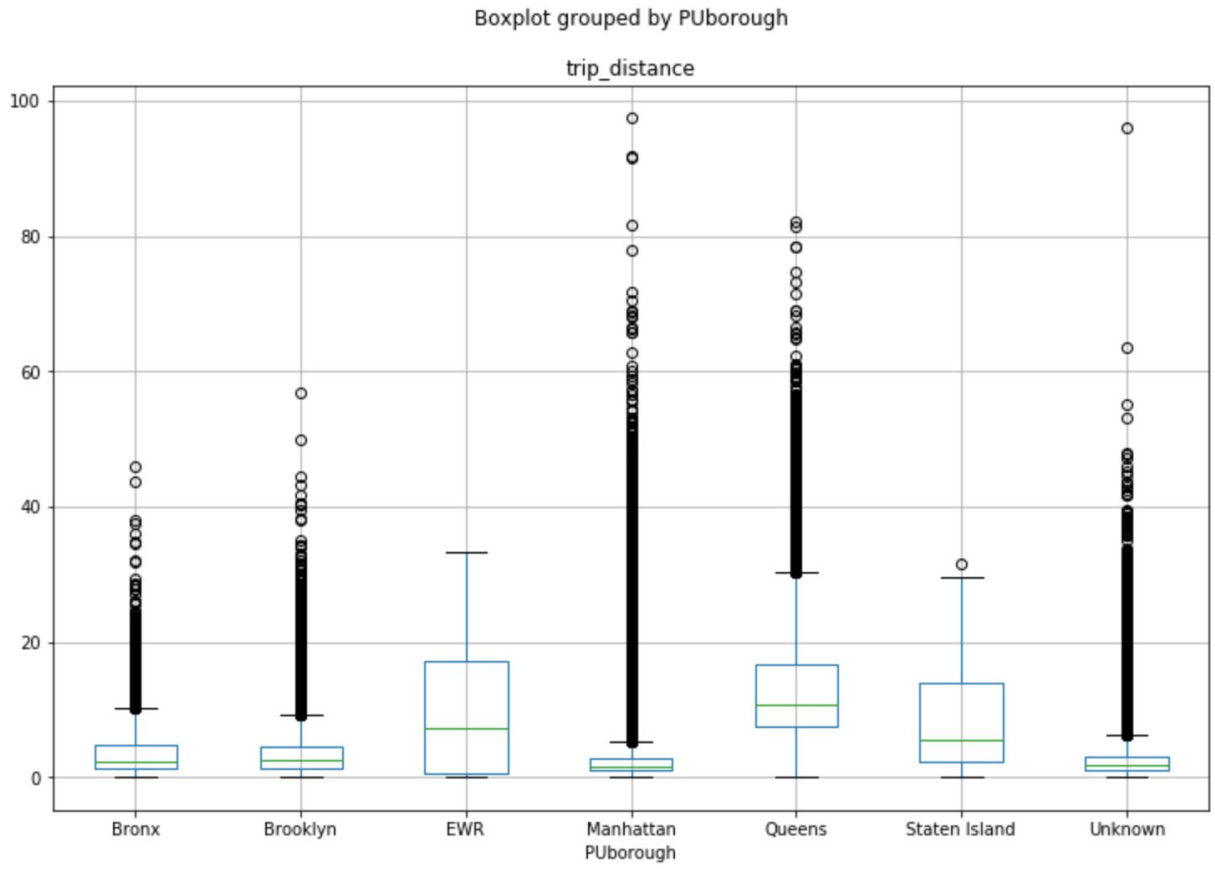
Manhattan dominates in the number of trips with over 80% of all trips. Additionally, most Queens trips are airport based trips which I have decided to avoid.

Distribution of Rides by Pickup Borough



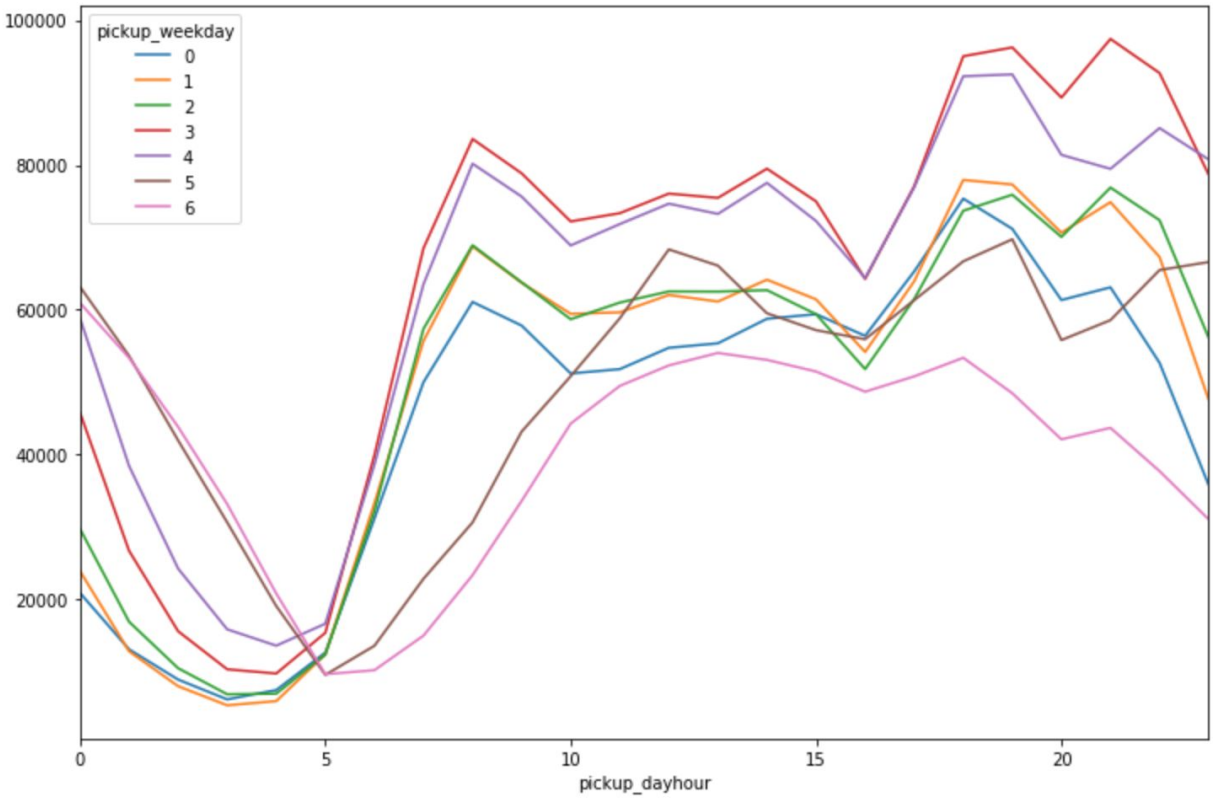
Most Manhattan trips are short trips. We hope to get lots of revenue from the quick short trips in Manhattan.

Distribution of Trip Distance for each Borough



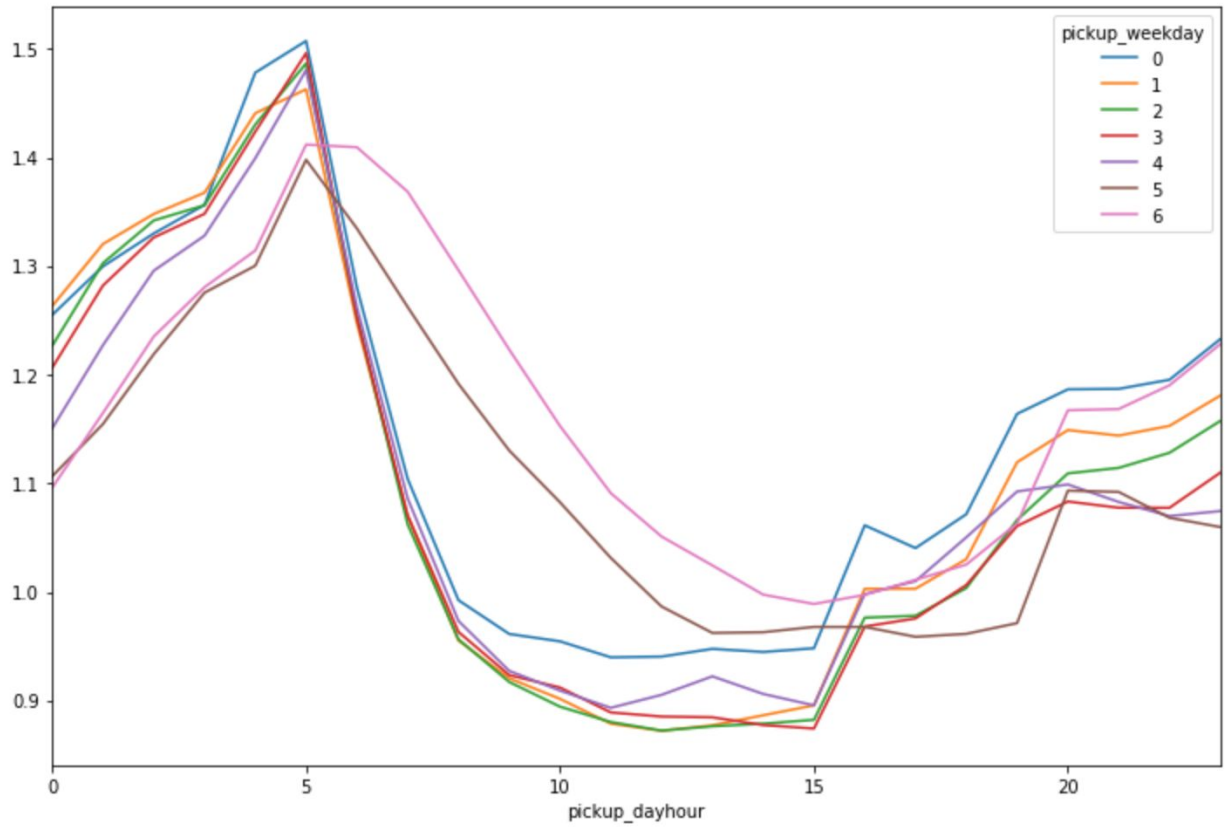
In Manhattan, highest demand is Thursday and Friday evenings.

Total Revenue by Pickup Day of Week and Hour in Manhattan



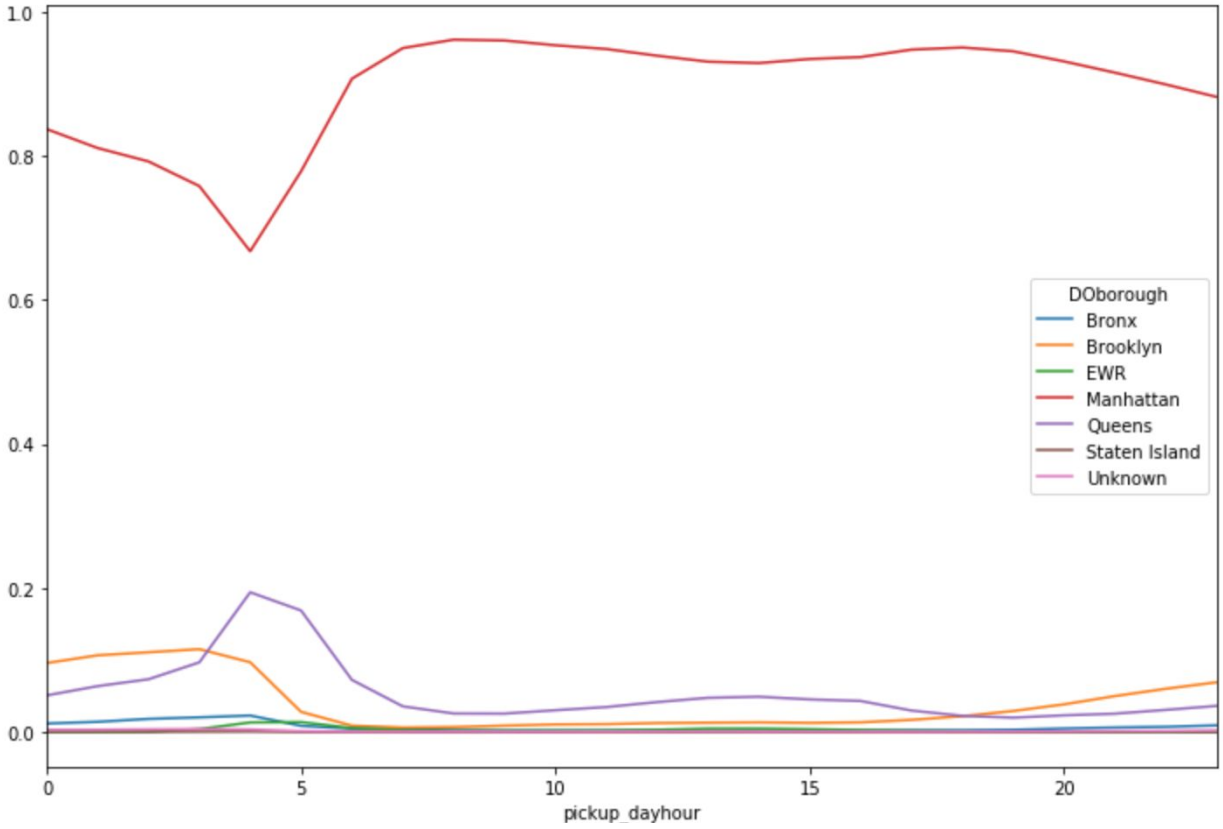
But revenue per minute is slightly better earlier in the week on Monday (day 0). More information on the impact of traffic on wait times and revenue per minute is needed to make the best possible decision. The spike in revenue per minute during early morning hours is ignored due to the low demand then but the increase in the day is interesting.

Median Revenue per Minute by Pickup Day of Week and Hour in Manhattan



This chart shows how many trips go to each borough from Manhattan. We want to avoid leaving Manhattan so will focus on the evening hours around 5-8pm due to high demand and a high percentage of rides staying in Manhattan.

Distribution of Dropoff Borough by Hour of Day for Pickups in Manhattan



Highest Manhattan revenue demand is in the Upper East side. We will start in that area to help satisfy the highest demand.

PUzone	DOzone	rev
Upper East Side South	Upper East Side North	400,790.7600
Upper East Side North	Upper East Side South	371,108.6400
Upper East Side North	Upper East Side North	252,572.7200
Upper East Side South	Upper East Side South	251,284.5100
Penn Station/Madison Sq West	Midtown Center	231,492.4800
Upper East Side South	Midtown Center	223,437.3400
Upper East Side North	Midtown Center	216,411.9800
Penn Station/Madison Sq West	Times Sq/Theatre District	214,584.6100
Upper East Side South	Midtown East	199,586.5600
Midtown Center	Upper East Side North	197,517.9100