

Question I:

Imagine that you decide to drive a taxi for 10 hours each week to earn a little extra money. Explain how you would approach maximizing your income as a taxi driver.

Since I am limited to only 10 hours each week to earn a little extra money I would look to optimize the timeblocks with the highest earnings per hour (\$/hr). In the earnings calculation I will only consider completed trips on credit card or cash, I can dig into other categories to see how they impact the overall earnings, but due to time limitation I will drop those. The earning calculation would be: earnings = fare amount + extra + tips (tips are scaled up since we do not have data for cash tips, making assumption that people tip the same on both cash and card)

In order to do this I would like to better understand the detailed demand:

- \$ per hour by overall dow(day of the week)-hour and by taxi zone dow-hour
- Number of completed rides by overall day- hour and by taxi zone dow-hour
- Average fare by overall dow-hour and by taxi zone dow-hour

Since we are limited by the data and do not know how many drivers are on the road at which time and how many total hours are being worked by all taxi drivers I will check the following proxy numbers as a proxy:

- Total trips and total fares completed by dow-hour to better understand when to drive
- Distribution of trips completed by Taxi Zone to better understand the location of where to drive

Normalized Trip Count								Normalized Total Earnings							
hour	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	hour	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	2.2%	2.2%	2.6%	3.1%	4.0%	5.5%	6.3%	0	2.6%	2.3%	2.8%	3.2%	4.2%	6.0%	6.0%
1	1.3%	1.2%	1.5%	1.8%	2.6%	4.6%	5.4%	1	1.5%	1.3%	1.5%	1.8%	2.7%	4.8%	5.0%
2	0.9%	0.7%	0.9%	1.0%	1.6%	3.5%	4.5%	2	0.9%	0.8%	0.9%	1.0%	1.5%	3.5%	4.1%
3	0.6%	0.5%	0.6%	0.7%	1.0%	2.6%	3.4%	3	0.6%	0.5%	0.6%	0.7%	1.0%	2.7%	3.2%
4	0.7%	0.5%	0.6%	0.6%	0.9%	1.7%	2.2%	4	0.9%	0.6%	0.7%	0.7%	1.0%	1.8%	2.2%
5	1.2%	1.1%	1.0%	1.0%	1.1%	0.9%	1.0%	5	1.5%	1.1%	1.1%	1.1%	1.3%	1.1%	1.2%
6	2.9%	2.7%	2.6%	2.6%	2.5%	1.2%	1.1%	6	2.8%	2.4%	2.3%	2.2%	2.4%	1.4%	1.3%
7	4.6%	4.6%	4.7%	4.4%	4.1%	1.9%	1.6%	7	4.3%	4.1%	4.0%	3.8%	3.7%	1.9%	1.7%
8	5.6%	5.7%	5.6%	5.3%	5.2%	2.6%	2.4%	8	5.3%	5.3%	5.2%	4.8%	4.9%	2.5%	2.2%
9	5.3%	5.3%	5.2%	5.0%	4.9%	3.6%	3.4%	9	5.2%	5.1%	4.9%	4.8%	4.7%	3.2%	3.0%
10	4.7%	5.0%	4.8%	4.6%	4.5%	4.2%	4.5%	10	4.7%	4.9%	4.7%	4.5%	4.5%	3.8%	3.9%
11	4.7%	5.0%	5.0%	4.7%	4.7%	4.8%	5.0%	11	4.7%	5.1%	5.0%	4.7%	4.7%	4.5%	4.6%
12	5.0%	5.1%	5.1%	4.9%	4.8%	5.6%	5.2%	12	4.9%	5.2%	5.2%	4.9%	4.9%	5.2%	5.0%
13	5.0%	5.1%	5.1%	4.9%	4.8%	5.5%	5.5%	13	5.0%	5.2%	5.3%	5.1%	4.9%	5.5%	5.4%
14	5.4%	5.3%	5.1%	5.2%	5.1%	5.0%	5.4%	14	5.4%	5.5%	5.4%	5.5%	5.4%	5.0%	5.7%
15	5.4%	5.1%	4.9%	4.8%	4.7%	4.8%	5.3%	15	5.4%	5.2%	5.1%	5.1%	5.0%	4.8%	5.6%
16	5.2%	4.5%	4.3%	4.2%	4.3%	4.7%	5.1%	16	5.2%	4.8%	4.7%	4.7%	4.7%	4.8%	5.4%
17	5.9%	5.4%	5.1%	5.0%	5.1%	5.2%	5.3%	17	6.0%	5.6%	5.4%	5.4%	5.4%	5.3%	5.5%
18	6.8%	6.5%	6.0%	6.1%	6.0%	5.5%	5.5%	18	6.6%	6.4%	6.1%	6.2%	5.9%	5.4%	5.6%
19	6.4%	6.4%	6.2%	6.2%	6.1%	5.8%	5.1%	19	5.9%	6.1%	6.0%	6.0%	5.8%	5.5%	5.1%
20	5.7%	5.9%	5.8%	5.9%	5.4%	4.8%	4.6%	20	5.4%	5.6%	5.5%	5.6%	5.1%	4.7%	4.7%
21	5.9%	6.3%	6.4%	6.4%	5.3%	5.0%	4.8%	21	5.9%	6.2%	6.4%	6.4%	5.2%	4.9%	5.0%
22	5.0%	5.8%	6.1%	6.2%	5.7%	5.5%	4.2%	22	5.3%	6.0%	6.2%	6.4%	5.7%	5.6%	4.5%
23	3.6%	4.2%	4.8%	5.3%	5.5%	5.7%	3.5%	23	4.0%	4.5%	5.1%	5.6%	5.6%	6.0%	4.0%

Borough	completed trips	percent trips
Manhattan	8,575,567	91%
Queens	549,929	6%
Brooklyn	145,984	2%
Unknown	128,831	1%
Bronx	7,296	0%
Staten Island	76	0%
EWB	75	0%

I would pick my top ten hours driving in Manhattan in the following time frame:

- Monday 5-10pm
- Tuesday, Wednesday, & Thursday 6-11pm
- Friday 6-8pm

Question II:

If you could enrich the dataset, what would you add? Is there anything in the dataset that you don't find especially useful?

I would like to add the following information to the dataset:

- The total hours that all the drivers are on the road or a pure number of drivers on the road at each hour, so I can get a better understanding of when the demand is not met
- Total potential requests of when a passenger attempted to get a trip, but were not able to
- Uber's messaging on the best time to drive since they want to optimize drivers' and their own earnings

The following data was not especially useful:

- Passenger count - could potentially look for larger groups to pick up if there is a significant upcharge on this item