Google Data Analytics: Analyzing Patterns of NYC Taxi Trips

A Project by
Nivea Rajesh Bekal
Pauline Rathni Grace

Table of Contents

New York City taxi trips dataset

Background

- The data used in the attached datasets were collected and provided to the NYC Taxi and Limousine Commission (TLC) by technology providers authorized under the Taxicab & Livery Passenger Enhancement Programs (TPEP/LPEP).
- The purpose of this is to understand more about NYC yellow taxi activities, such as which areas are more popular, at what time or day it is the busiest,in general, and which area is more generous in terms of tipping.

Objective

- This analysis can be useful in two ways
  1. For taxi drivers who want to know where or when to maximize their income.
  2. For passengers who want to avoid the traffic.
- The data set contains the data regarding several taxi trips and its duration in New York City. We will analyze the data to get insights and determine how different variables are dependent on the other.

Prepare Data

- In this Data set we have Cleaned the data and prepared the data to make some analysis, Our Pre-processing includes:
  - Dropping out null values and Out of Range Values for Latitude and Longitude columns, negative trip duration.
  - Added Zone name using new_york_taxi_trips.taxi_zone_geom Dataset.
  - Passenger count that had zero were removed.
  - Using variables such as pickup time and drop off time , time duration was calculated.
  - Likewise, driving speed was calculated using the formula and data from the distance and time duration columns.
  - Tip amount was calculated using ratio of tip amount to total amount.

Preview Data

These are few of the columns in the data that were considered for our analysis :

- **vendor_id** : A code indicating the TPEP provider that provided the record. 1= Creative Mobile Technologies, LLC; 2= VeriFone Inc
- **pickup_datetime** : The date and time when the meter was engaged.
- **dropoff_datetime** : The date and time when the meter was disengaged.
- **passenger_count** : The number of passengers in the vehicle. This is a driver-entered value.
- **trip_distance** : The elapsed trip distance in miles reported by the taximeter.
- **rate_code** : The final rate code in effect at the end of the trip. 1= Standard rate 2=JFK 3=Newark 4=Nassau or Westchester 5=Negotiated fare 6=Group ride.
- **payment_type** : A numeric code signifying how the passenger paid for the trip. 1= Credit card 2= Cash 3= No charge 4= Dispute 5= Unknown 6= Voided trip
- **fare_amount** : The time-and-distance fare calculated by the meter.
- **tip_amount** : This field is automatically populated for credit card tips. Cash tips are not included.
- **tolls_amount** : Total amount of all tolls paid in the trip.
- **total_amount** : The total amount charged to passengers. Does not include cash tips.
- **pickuplocationid** : TLC Taxi Zone in which the taximeter was engaged.
- **Dropofflocationid :** TLC Taxi Zone in which the taximeter was disengaged.

Insight 1 :
What are the most popular boroughs and zones?

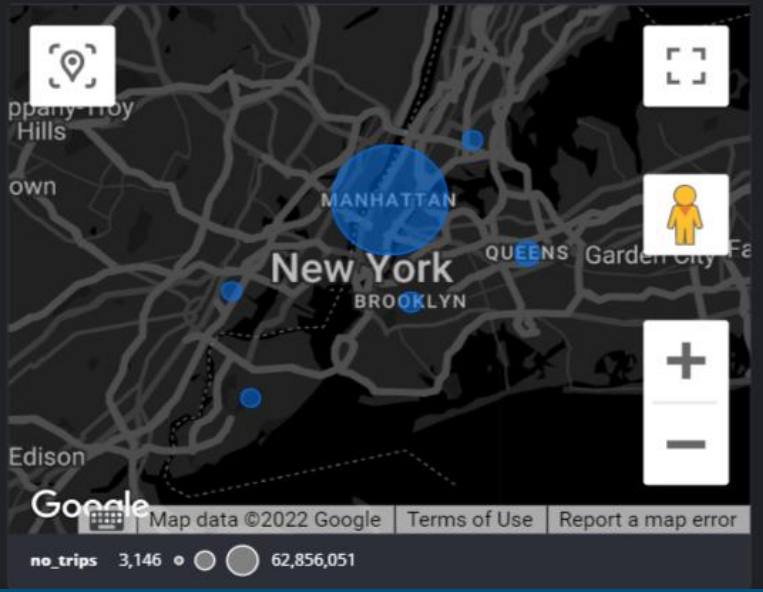
- Most of the yellow taxi operations are happening in Manhattan.
- It shows that ~92.39% of the trips started from Manhattan.
- There might be more short trips in Manhattan as the most popular routes in Manhattan are apparently still within the Upper East Side area (either from South to North or the other way round).

pickup_borough	pickup_zone_name	route_borough	route_zone_name	no_trips
Bronx	Mott Haven/Port Morris	Bronx-Bronx	Mott Haven/Port Morris-Mott H...	4033
Bronx	West Concourse	Bronx-Bronx	West Concourse-West Concour...	3836
Bronx	Mott Haven/Port Morris	Bronx-Manhattan	Mott Haven/Port Morris-East H...	1460
Bronx	West Concourse	Bronx-Manhattan	West Concourse-Central Harle...	1147
Bronx	Mott Haven/Port Morris	Bronx-Manhattan	Mott Haven/Port Morris-Centra...	952
Bronx	Mott Haven/Port Morris	Bronx-Manhattan	Mott Haven/Port Morris-East H...	679
Bronx	Mott Haven/Port Morris	Bronx-Manhattan	Mott Haven/Port Morris-Centra...	666
Bronx	University Heights/Morris Heig...	Bronx-Bronx	University Heights/Morris Heights-University Heights/Morris Heights	615
Bronx	East Concourse/Concourse Vill...	Bronx-Bronx	East Concourse/Concourse Village-East Concourse/Concourse Village	553
Bronx	West Concourse	Bronx-Bronx	West Concourse-Mott Haven/P...	553

Fig 1 : Bubble Map showing popular Boroughs

	pickup_borough	no_trips
1.	Manhattan	62,856,051
2.	Queens	3,948,818
3.	Brooklyn	1,229,489
4.	Bronx	56,784
5.	EWR	4,031
6.	Staten Island	3,146

1 - 6 / 6



Insight 2 :

What was the average speed of Yellow taxi trips in 2016?

- During the day(8 AM-9 AM), the average speed is around 11-12 MPH but at 5:00 AM the average speed almost doubles to 21 MPH.
- Intuitively this makes sense since there is likely less traffic on the road at 5:00 AM.

Row	hour	speed
1	0	15.7
2	1	16.1
3	2	16.7
4	3	17.5
5	4	20.0
6	5	21.4
7	6	17.1
8	7	13.4

Fig 2 : Combo Chart showing average speed of yellow taxis in 2016

	hour	speed
1.	05	21.4
2.	04	20
3.	03	17.5
4.	06	17.1
5.	02	16.7
6.	01	16.1
7.	00	15.7
8.	07	13.4
9.	08	11.4
10.	09	11.3

1 - 10 / 10

Insight 3 :

How does the overall pattern look like over these 6 months?

- It is observed that Mondays have lesser trips than the rests of the days, while Fridays have more trips.
- It was observed that on 31st May 2016,there was a drop, most likely caused by Memorial Day that required some streets and routes to be closed for the parades.

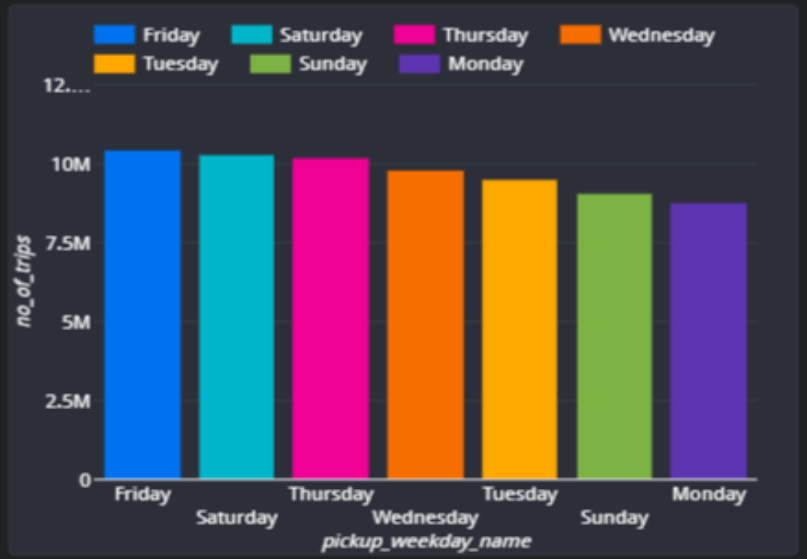
Row	pickup_borough	pickup_date	pickup_weekday_name	pickup_hour	no_of_trips
1	Manhattan	2016-01-24	Sunday	15	11686
2	Manhattan	2016-01-23	Saturday	2	9086
3	Manhattan	2016-01-23	Saturday	11	2898
4	Manhattan	2016-01-27	Wednesday	21	21410
5	Manhattan	2016-01-27	Wednesday	1	4418
6	Manhattan	2016-01-24	Sunday	23	6127
7	Queens	2016-01-24	Sunday	9	337
8	Brooklyn	2016-01-06	Wednesday	19	264
9	Manhattan	2016-01-23	Saturday	5	1295
10	Brooklyn	2016-01-06	Wednesday	15	98

pickup_borough	pickup_date	pickup_weekday_name	pickup_hour	no_of_trips
Manhattan	2016-05-31	Tuesday	0	5171
Queens	2016-05-31	Tuesday	12	1155
Bronx	2016-05-31	Tuesday	11	11
Bronx	2016-05-31	Tuesday	10	4
EWR	2016-05-31	Tuesday	18	1
Queens	2016-05-31	Tuesday	7	1257

Fig 3 : Bar graph showing which day has maximum number of trips

	pickup_weekday_name	no_of_trips
1.	Friday	10,452,797
2.	Saturday	10,300,440
3.	Thursday	10,197,085
4.	Wednesday	9,799,231
5.	Tuesday	9,487,680
6.	Sunday	9,087,654
7.	Monday	8,773,432

1 - 7 / 7

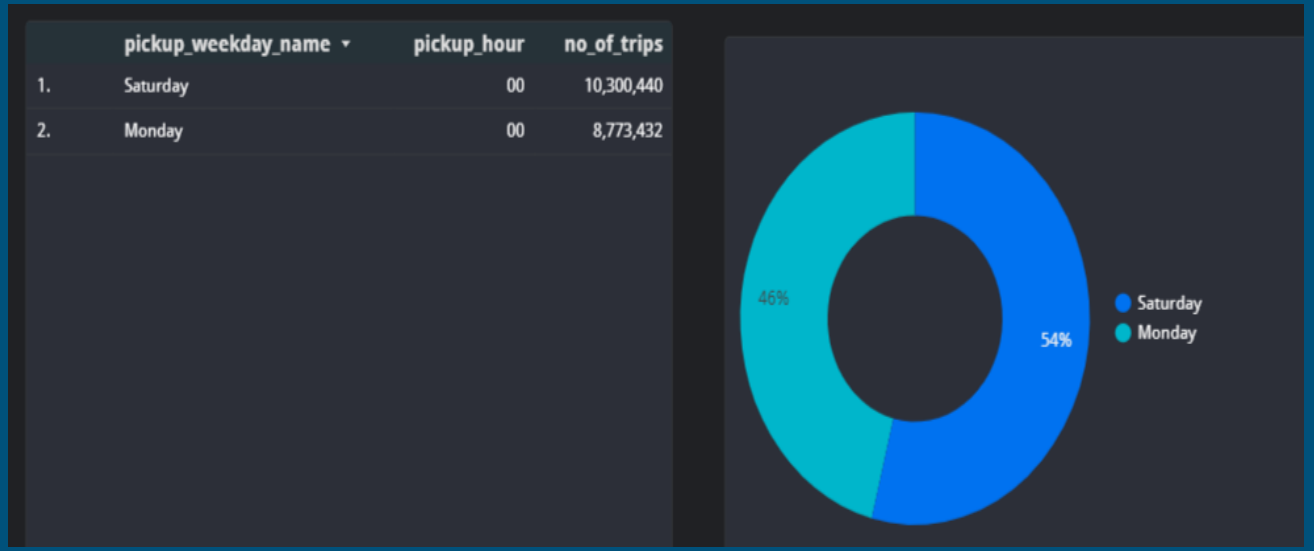


Insight 4 :
How does the hourly pattern look like in a week and in a day?

- It is observed that fewer trips on Monday midnight (00:00 am) compared to Saturday midnight when people would tend to go out on Friday night, spend a longer time outside, and get a taxi home after midnight.
- Saturday activities did start later (around 09:00 am) than Monday but got busier throughout the rest of the day. On the other hand, Monday activities started earlier (around 07:00 am), most likely when people were commuting to work, and got busier after working hours (06:00 pm).

pickup_borough	pickup_date	pickup_weekday_name	pickup_hour	no_of_trips
Bronx	2016-06-04	Saturday	21	29
Queens	2016-01-09	Saturday	4	280
Bronx	2016-01-09	Saturday	6	21
Brooklyn	2016-03-21	Monday	1	212
Brooklyn	2016-01-11	Monday	13	81
Manhattan	2016-04-30	Saturday	13	19433

Fig 4 : Doughnut chart to analyze hourly pattern and number of trips



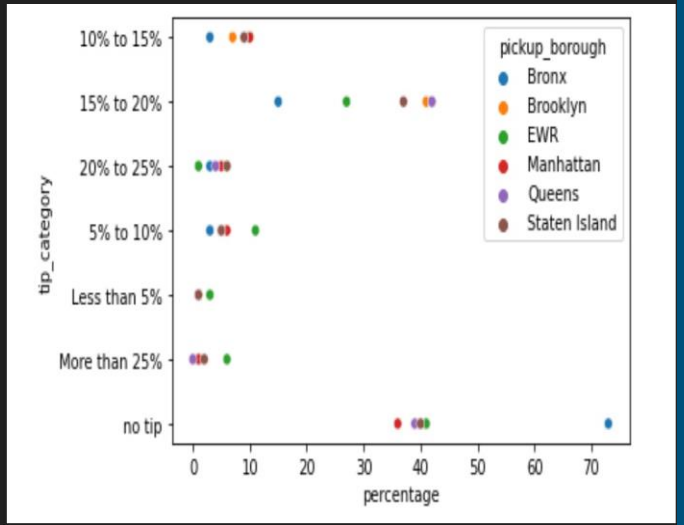
Insight 5 :
How much is the tip?

- It was observed that there were two extremes, either the passengers do not give tip at all or they gave slightly below 20%.
- The data does show that 15% to 20% is the most common tipping rate. Manhattan, Queens, and Brooklyn did tip similarly, with Manhattan being slightly more generous. There were 36% of trips without tip in Manhattan. This % is lower than the other boroughs.

pickup_borough	tip_category	no_of_trips	parent_sum	percentage
Bronx	10% to 15%	1841	56784	0.03
Bronx	15% to 20%	8741	56784	0.15
Bronx	20% to 25%	1522	56784	0.03
Bronx	5% to 10%	1811	56784	0.03
Bronx	Less than 5%	625	56784	0.01
Bronx	More than 25%	662	56784	0.01
Bronx	no tip	41582	56784	0.73

Fig 5 : Scatter chart to analyze pattern of tips

tip_category	percentage
10% to 15%	8.000000
15% to 20%	34.000000
20% to 25%	4.166667
5% to 10%	5.833333
Less than 5%	1.333333
More than 25%	1.833333
no tip	44.833333

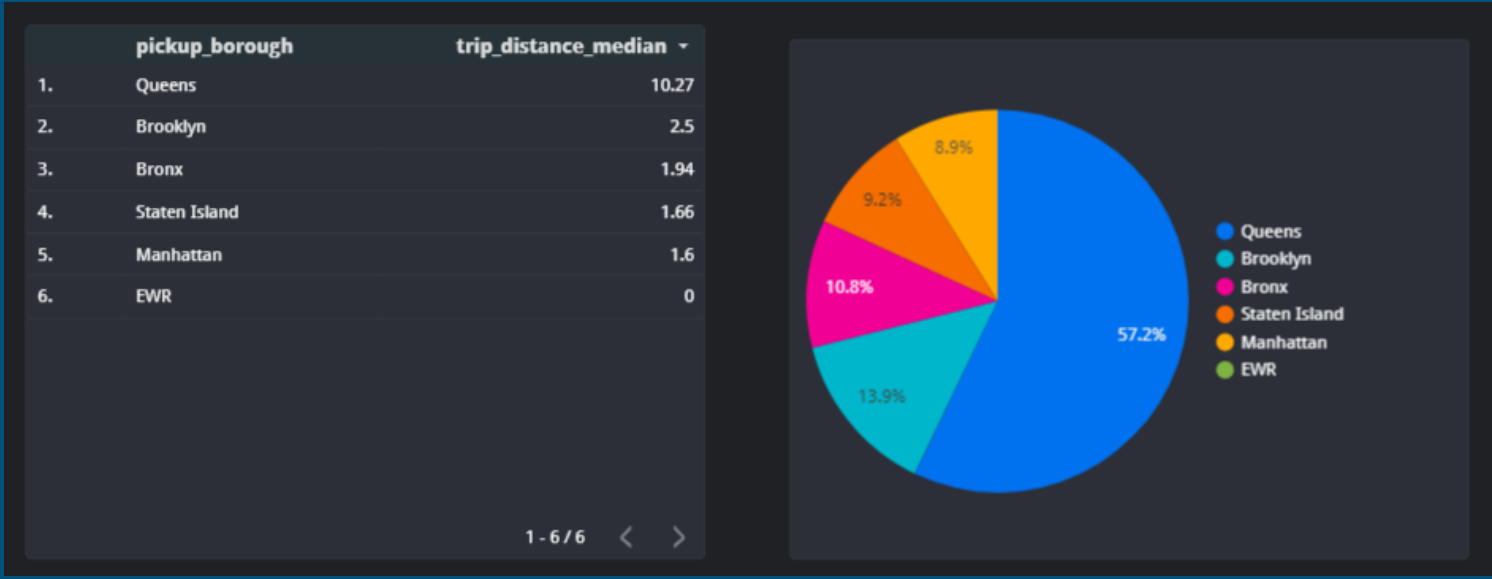


Insight 6 :
Does Manhattan really have more short trips?

- Compared to few other boroughs, Manhattan has relatively short trip duration and distance, but it is not the shortest.

pickup_borough	trip_duration_median	trip_distance_median
EWR	30	0.0
Staten Island	674	1.66
Bronx	574	1.94
Brooklyn	703	2.5
Queens	1730	10.27
Manhattan	635	1.6

Fig 6 : Pie chart to analyze trip distance



Conclusions

Herewith we share the recommendation based on insights we gained from the analysis:

1

Maximum Income

It is seen that maximum income can be obtained during Saturday nights and we also infer that taxis are frequently taken from/to Manhattan with a higher tip rate.

2

Avoid Traffic

The busy hours are generally from 8AM-10AM and in order to avoid traffic passengers are advised to leave before that. Also, it gets busier post 6PM so its advisable to avoid these hours to reach faster.

Acknowledgement

We would like to sincerely thank our Professor Charles Border for making this course interesting to us and explaining the concepts very well.

Thank you!