# NYC'S GREEN TAXI INSIGHTS 2019 – 2020


# Table of contents


# 01

# Dataset Overview

# NYC Taxi Explanation

There are **2 types** of taxi in NYC, with same fare structure but have the differences in pick-up points:

- **Yellow taxi:** tradition taxi, are able to pick up passengers anywhere in NYC.
- **Green taxi:** established in 2013, are able to pick up passengers in Upper Manhattan, the Bronx, Brooklyn, Queens (excluding LaGuardia Airport and John F. Kennedy International Airport), and Staten Island.

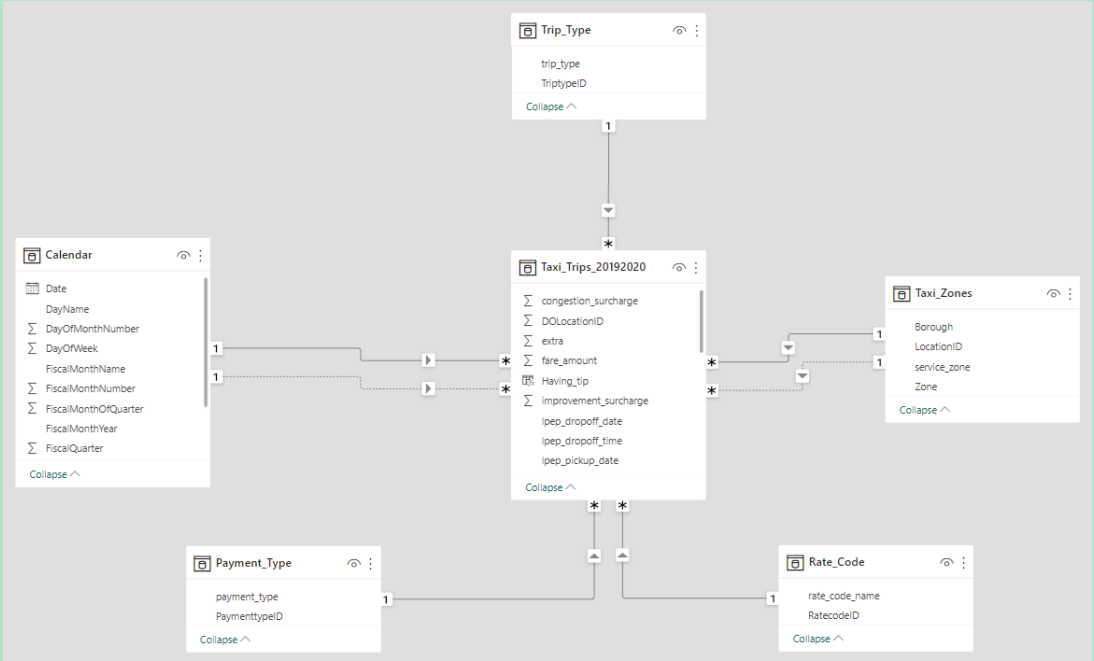
# About the dataset

The dataset provides trip records from all green taxis in NYC from 2019 to 2020. Records include fields capturing pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts.

Currency: USD($)

# Information of the dataset

- All data were recorded from 01.01.2019 to 31.12.2020.
- There are 6 tables after processing data (there were 3 tables in the beginning).
- 7,734,505 rows.
- 20 columns.




# Link of dataset:

https://www.mavenanalytics.io/data-playground?accessType=open&order=-fields.dateUpdated&search=NYC%20taxi%20trips

# Data dictionary (1)

Overall, there are 20 attributes in the dataset, in which each attribute provides taxi information like pick-up, drop-off location, time, fares and trip types.

| Field | Description |
|---|---|
| VendorID | A code indicating the LPEP provider that provided the record (1= Creative Mobile Technologies, LLC; 2= Verifone Inc.) |
| lpep_pickup_datetime | The date and time when the meter was engaged |
| lpep_dropoff_datetime | The date and time when the meter was disengaged |
| store_and_fwd_flag | This flag indicates whether the trip record was held in vehicle memory before sending to the vendor, aka “store and forward,” because the vehicle did not have a connection to the server (Y= store and forward trip; N= not a store and forward trip) |
| RatecodeID | 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) |
| PULocationID | TLC Taxi Zone in which the taximeter was engaged |
| DOLocationID | TLC Taxi Zone in which the taximeter 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 |
| fare_amount | The time-and-distance fare calculated by the meter |

# Data dictionary (2)

Overall, there are 20 attributes in the dataset, in which each attribute provides taxi information like pick-up, drop-off location, time, fares and trip types.

| Field | Description |
|---|---|
| extra | Miscellaneous extras and surcharges (this only includes the $0.50 and $1 rush hour and overnight charges) |
| mta_tax | $0.50 MTA tax that is automatically triggered based on the metered rate in use |
| tip_amount | Tip amount (automatically populated for credit card tips - cash tips are not included) |
| tolls_amount | Total amount of all tolls paid in trip |
| improvement_surcharge | $0.30 improvement surcharge assessed on hailed trips at the flag drop |
| total_amount | The total amount charged to passengers (does not include cash tips) |
| 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) |
| trip_type | A code indicating whether the trip was a street-hail or a dispatch that is automatically assigned based on the metered rate in use but can be altered by the driver (1= Street-hail; 2= Dispatch) |
| congestion_surcharge | Congestion surcharge for trips that start, end or pass through the congestion zone in Manhattan, south of 96th street ($2.50 for non-shared trips in Yellow Taxis; $2.75 for non-shared trips in Green Taxis) |


# 02

# Dataset Preprocessing

# Data preprocessing

**Data clean**

*(SQL)*

**Step1**

*Check data type and remove unnecessary columns*

**Step2**

*Check and replace null value*

**Step3**

*Remove records that have negative values*

**Data process**

*(SQL,PowerBI)*

**Step1**

*Import & split columns*

**Step2**

*Create dim table*

**Step3**

*Create data model*


# 03

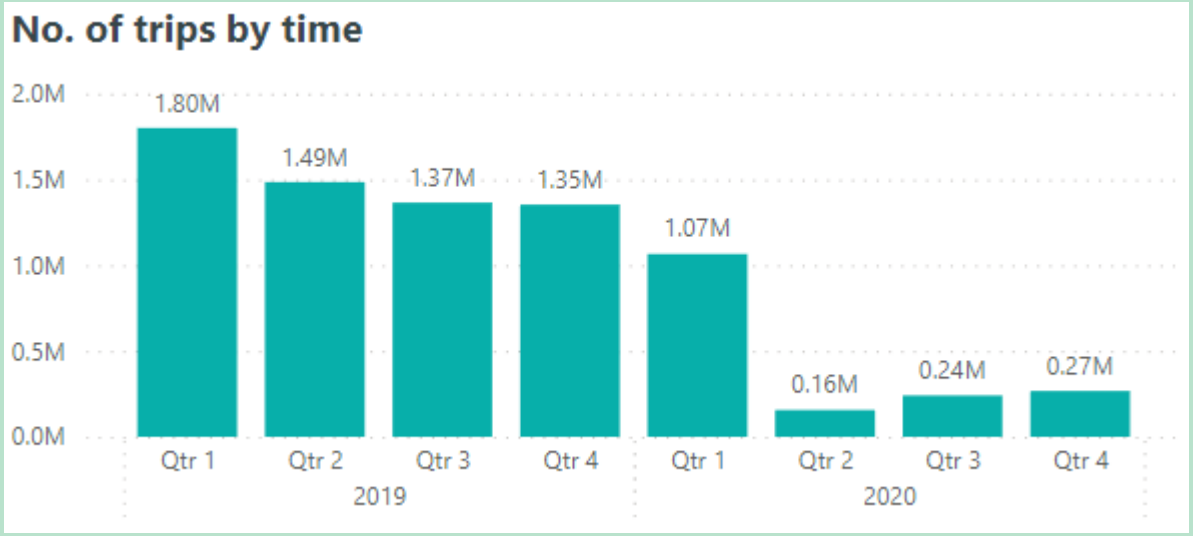
# Data Insight

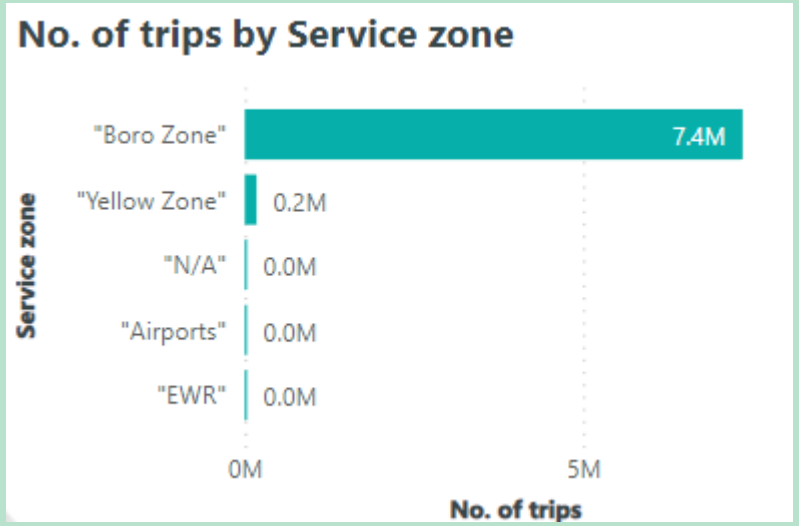
# Taxi Trip Overview

# Overview about Taxi trip

- The dataset provides total number of green taxi trips in NYC during 2019 and 2020, there were total of nearly 8 million trips occurred during the period, serving total 8.9 million passengers with nearly 139M total payment.
- Total trips dropped significantly in 2020, approximately 70% decline.
- The average payment per trip was 17.95 USD.
- Obviously, since the data about Green Taxi so the most chosen Service Zone was Boro Zone.

| 7.73M | 8.90M | 1.15 | 138.84M | 17.95 |
|---|---|---|---|---|
| No. of taxi trips | Total passengers | Avg passengers per trip | Total payment | Avg payment per trip |

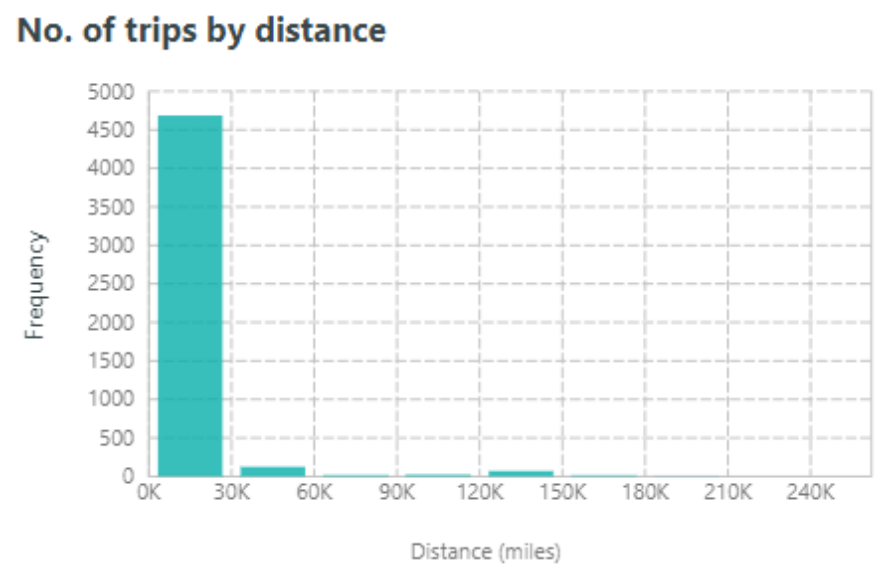


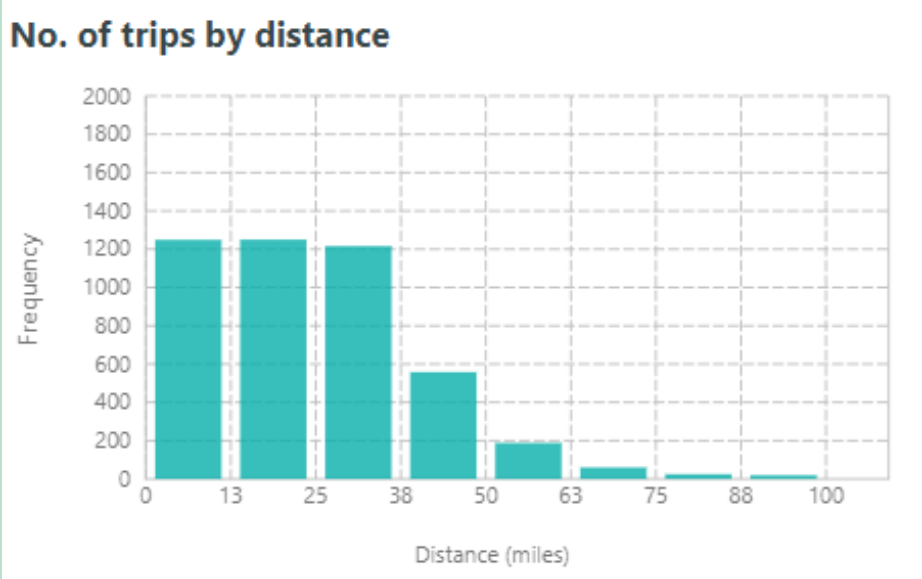



# Trips & Distance

# Distance traveled, by zone and time

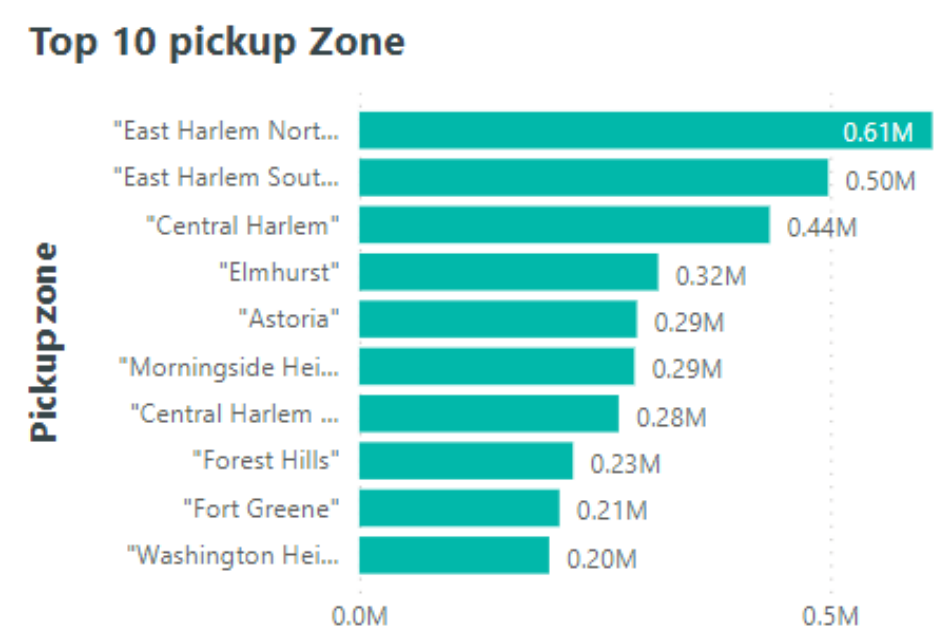
- 4 pm to 8 pm was the most commuted time, as this tend to be outside working hours.
- Regarding the distance, people tend to travel around less than 38 miles.

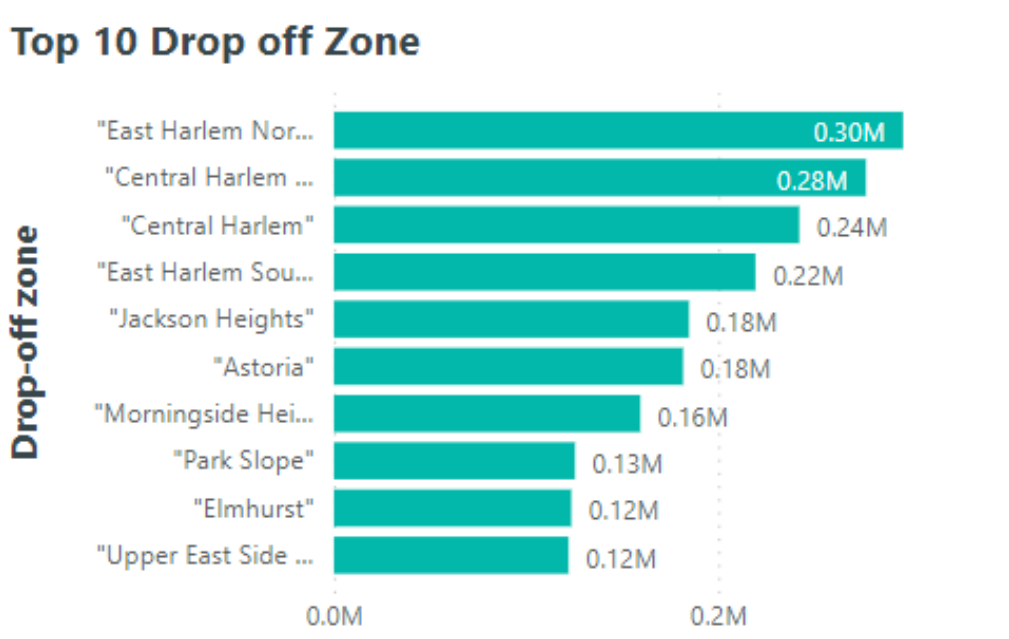


# Top 10 Pick up/ Drop off Locations

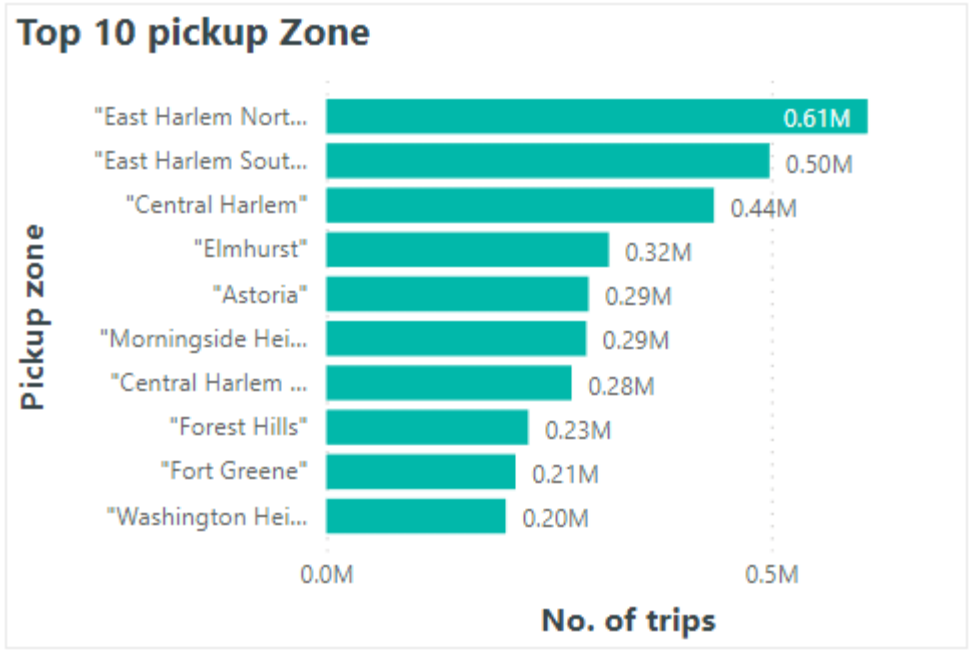
- Top 10 pickup zones: 6 areas in upper Manhattan, 3 zones in Queens and 1 area in Brooklyn. In which the top 3 pickup zones are all belonged to the Upper Manhattan area.
- Top 10 drop off zone: Top 5 drop off points are all belonged to the Upper Manhattan area, while the remaining locations are belonged to Queens and Brooklyn

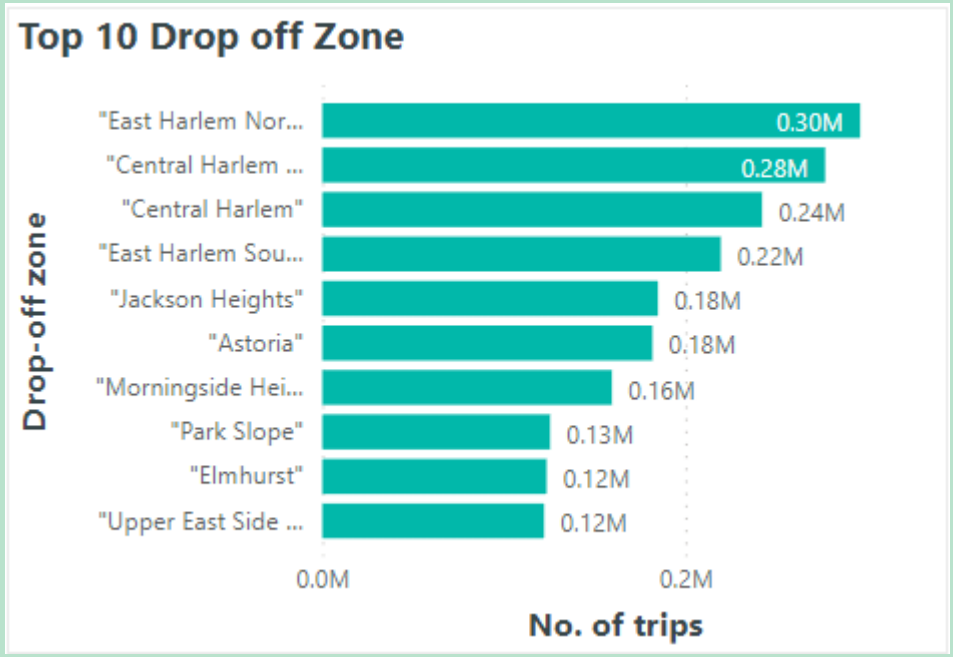


# Top 10 Trips by Green Taxi in NYC

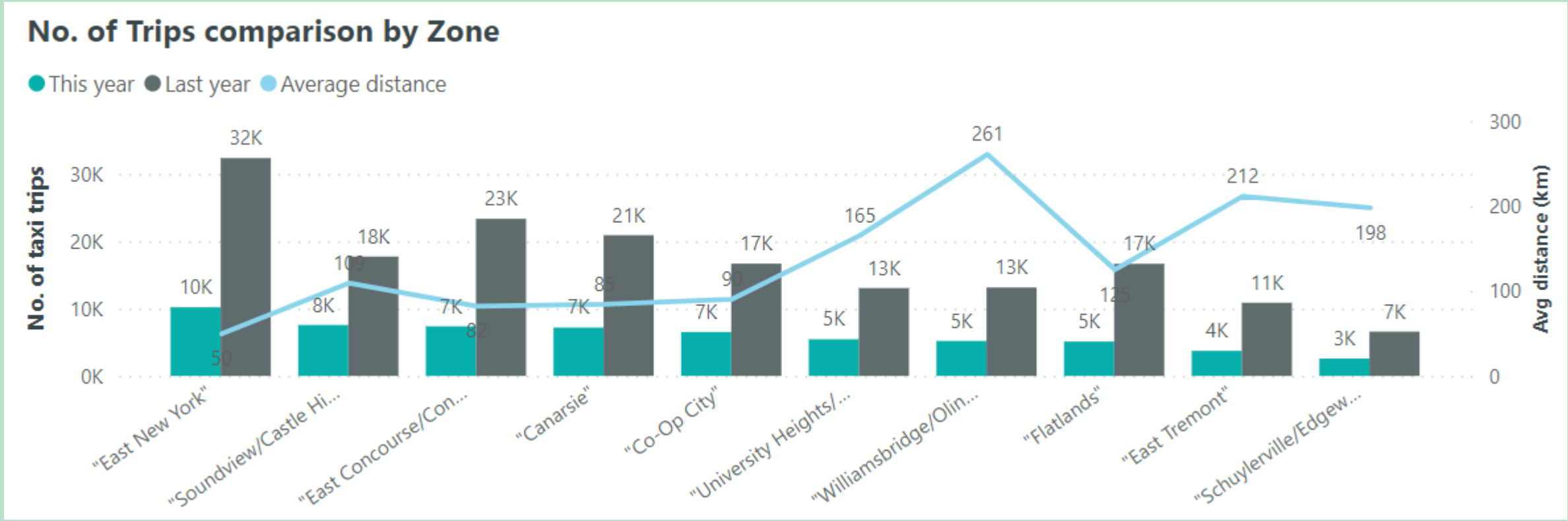
- Top 10 pickup zones: 6 areas in upper Manhattan, 3 zones in Queens and 1 area in Brooklyn. In which the top 3 pickup zones are all in upper Manhattan
- Top 10 drop off zone: Top 5 drop off points are all in upper Manhattan and the others belong to Queens and Brooklyn



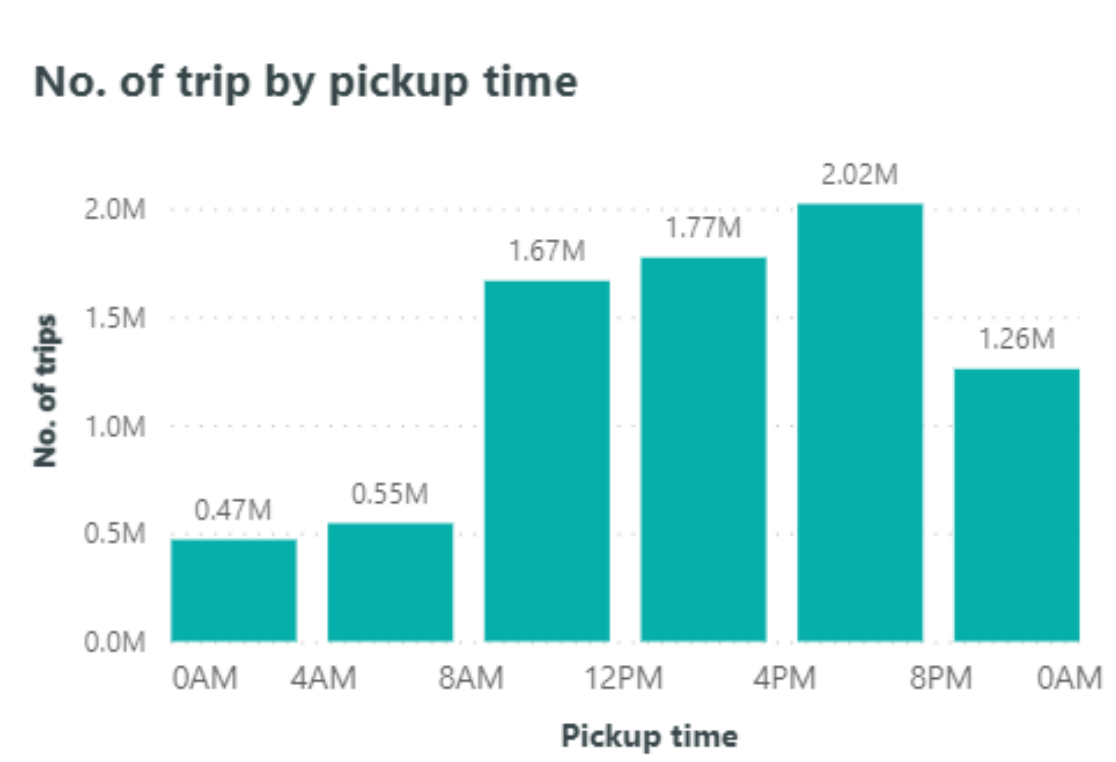
# Number of trips comparison by zone

- There was a significant drop in terms of total trips in 2020 in all zone, meaning COVID-19 caused a negative impact on commuting by taxi behavior, as most zones witnessed more than 50% drop.



# Number of trips comparison by Time

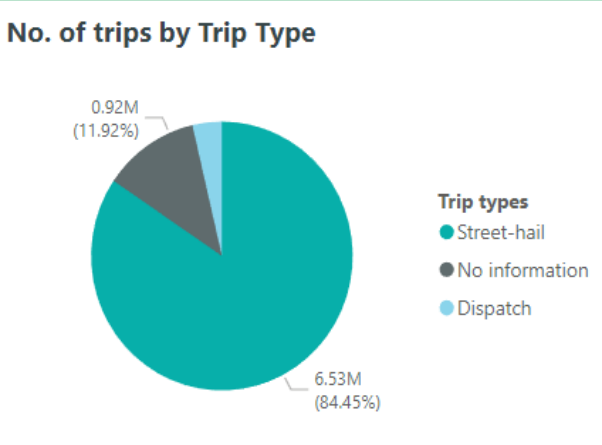
- Top 3 popular periods for Green Taxi: 8 AM to 12 PM, 12 PM to 4 PM, 8 PM to 12 AM.

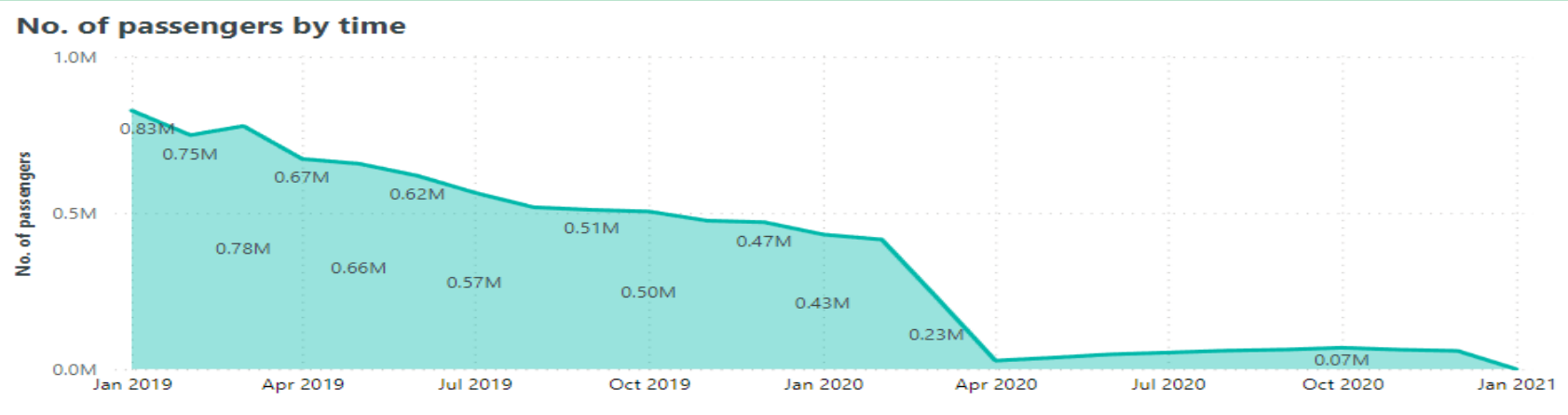


# Passengers Insights

# Passengers characteristics

- ❑ Most trips were ordered by hailing on the street, with the average commuting time of 6.7 minutes
- ❑ The total served passengers went down substantially in relation to COVID-19.




**6.70**

Avg time (mins) per trip

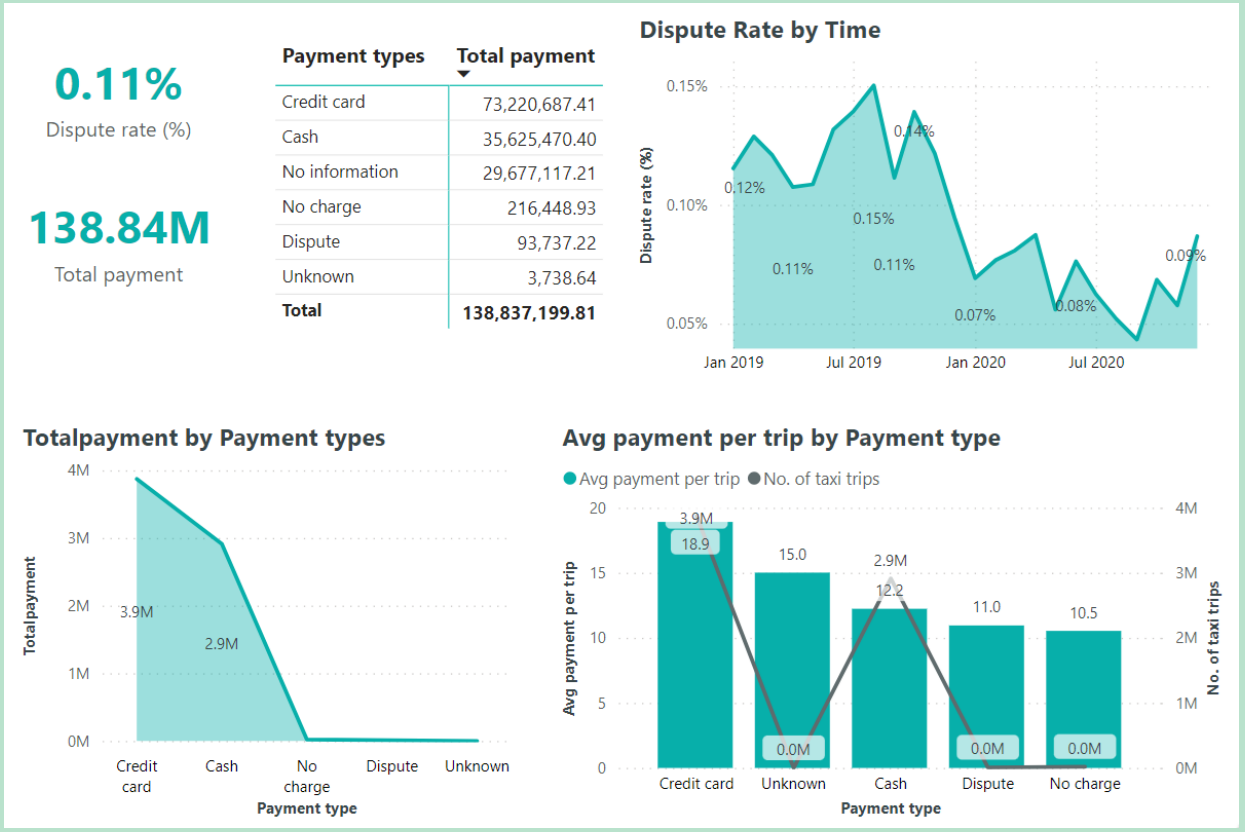
**1.15**

Avg passengers per trip

# Payment Insights

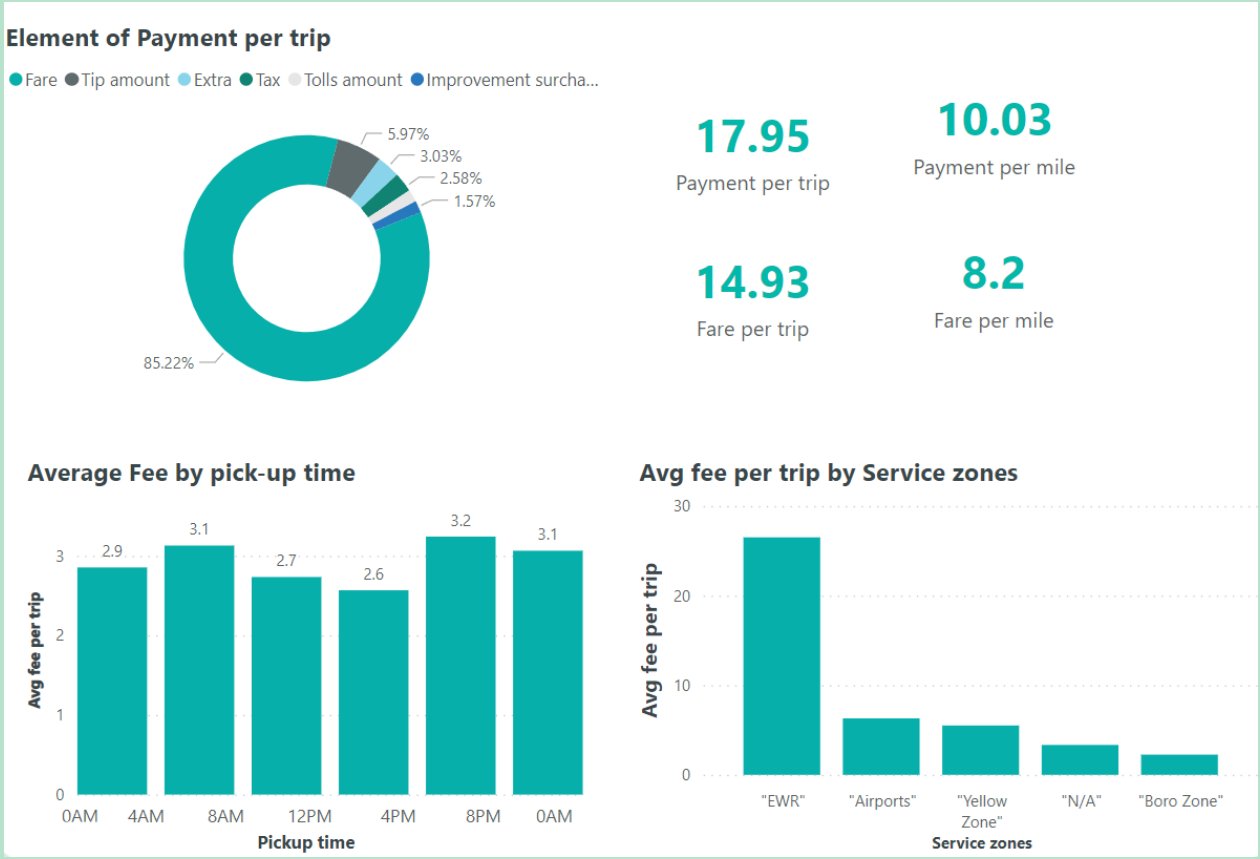
# Payment behaviors

- Green Taxi generated 138.84 million USD in NYC during 2019-2020, with very low dispute rate.
- Credit card was the most chosen payment method, while Cash is the second.



# Payment behaviors (2)

- The average amount of payment per trip is 17.95 USD, of which 14.93 USD is the fare, the remaining 3 USD is the surcharge.
- For travel fare over 1 mile, the passenger must pay 8.2 USD, and when adding the surcharges, the passenger must pay 10.03 USD per mile
- The surcharge (including all extra charges not included in the rate) is the highest during peak hours (4-8pm), but there is no significant difference compared to the rest of the hours.
- Two areas with the highest surcharges are EWR and Airports, both of which have airports

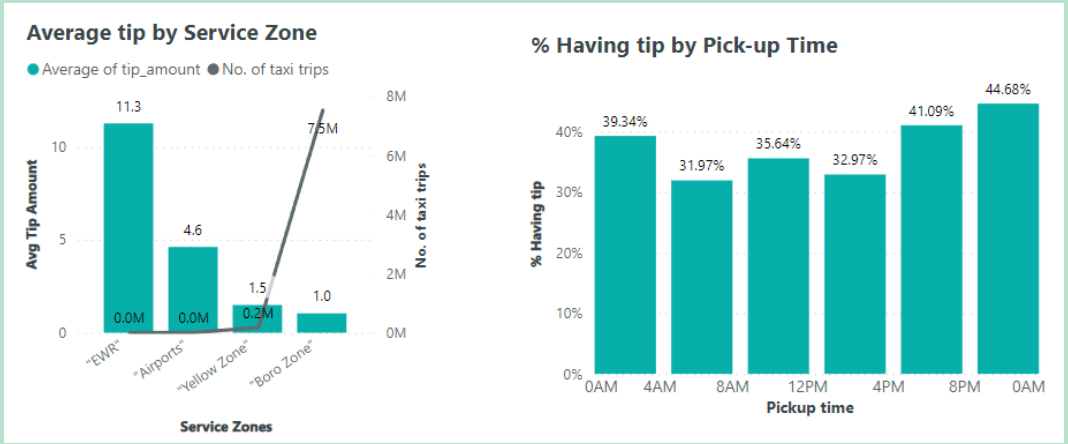


# Tipping behaviors by customers

- ❏ Nearly 38% of taxi trips are paid with tips, the average tip is 2.78 USD / trip
- ❏ The most common tip is in the range of 1-3 USD/trip, a higher level of 4-6 USD is incurred but relatively little
- ❏ The highest average tip is EWR and airport
- ❏ The percentage of passengers who tips to drivers falls a lot during rush hours or night trips

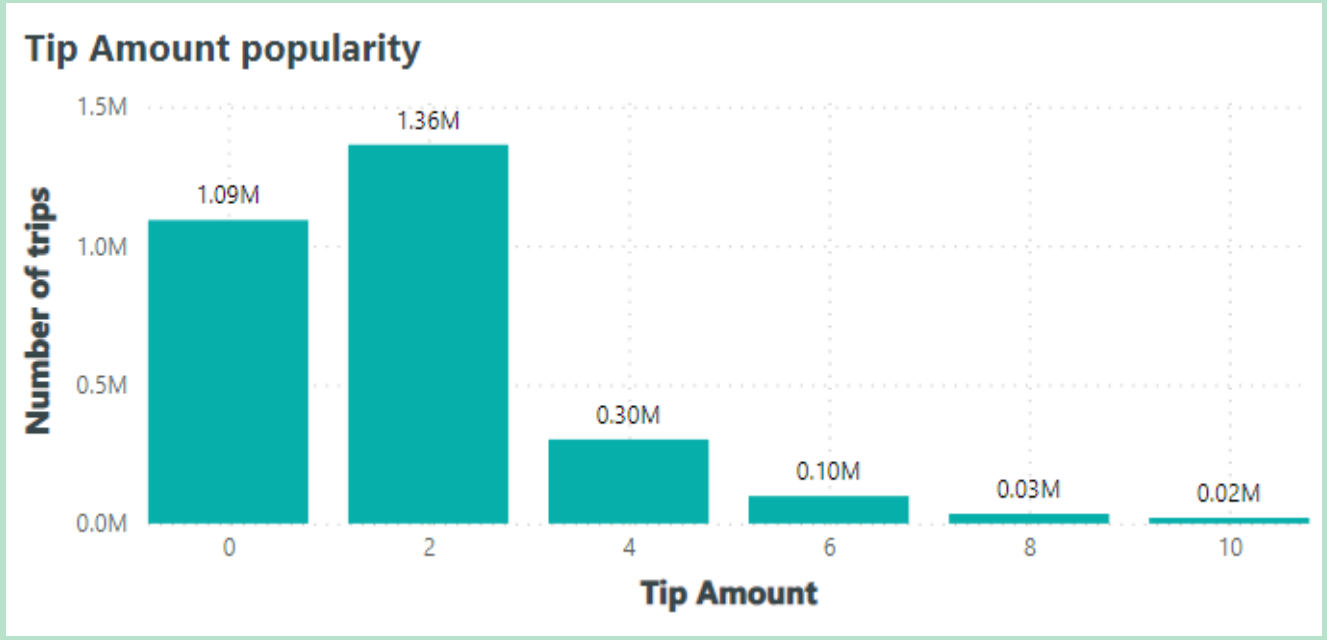
**37.89%**
% trips having tip

**2.78**
Avg tip amount/trip









# Conclusions

- Green taxis are limited in terms of selecting pick-up locations, leading to further distance traveled.
- Regulations over pick up locations could lead to more consumed energy and wasted time, as the drivers would have to wander on the road for considerable distance, and time before reaching the allowed pick up locations.
- The optimized location for taxi drivers are neighboring areas, close to Central Manhattan, which contradict to their original mission; since Green Taxi was created to serve neighboring areas – whose places do not have traffic jam.


# 04

# Recommendations

# Recommendations

## Trip optimization

- Better coordianation during rush hours by taxi brands.
- Adjust pick up location policy.

## Payment method optimization

- Reduce cash payment
- Provide more incenvtives (discount, etc) for payment by credit card

## Passenger optimization

- Improve service quality (convenience, intra trip experience, driver attitudes)
- Payment method variety
- Optimize route