

PROJECT REPORT

DATA ANALYSIS FOR NYC YELLOW TAXI CABS



**BY:
GROUP2
(Priyanka Nagpal & Justin Han)**

Project Deliverable 2

Background:

New York City is known as the taxi capital of America and home of the classic yellow taxicab (Majaski, 2019). Currently, the traditional taxi industry faces challenges from other rideshare apps like Uber, Lyft, and Via. In a crowded city like New York, the demand for taxis is enormous, but yellow taxis struggle to keep the market shares. Thus, it is crucial for us to help yellow cab drivers to find a way to improve the market share (Crudele, 2017).

Data Source:

The data we will analyze is the trip data from New York City's yellow cab. It includes every ride recorded by meters in the cab in December 2019. Trip data is provided by the NYC Taxi and Limousine Commission (TLC) monthly. It records all rides across five boroughs of the city; it usually has over 6 million observations.

Problems addressed:

The primary goal of this project is to help yellow taxi cab drivers to increase their profits. In order to do so, major areas focused are as follows:

Q1: Overall Cost Analysis & Are yellow taxi cabs able to meet the threshold income?

Q2: How can we optimize the overall fleet for taxi drivers?

Analytical techniques:

For this analysis, we have used multiple techniques that include clustering, spatial analysis, and overall calculations of cost using simple dplyr functions and ggplots to find the relationship between variables.

Spatial Analysis:

Spatial analysis is used to visualize and identify patterns using geographical locations. For this project, we aspired to evaluate the pickup location density in New York so that we can evaluate which areas have more pickup locations & accordingly optimum fleet can be decided in conjunction with the data evaluated from pickup/drop hour of the day.

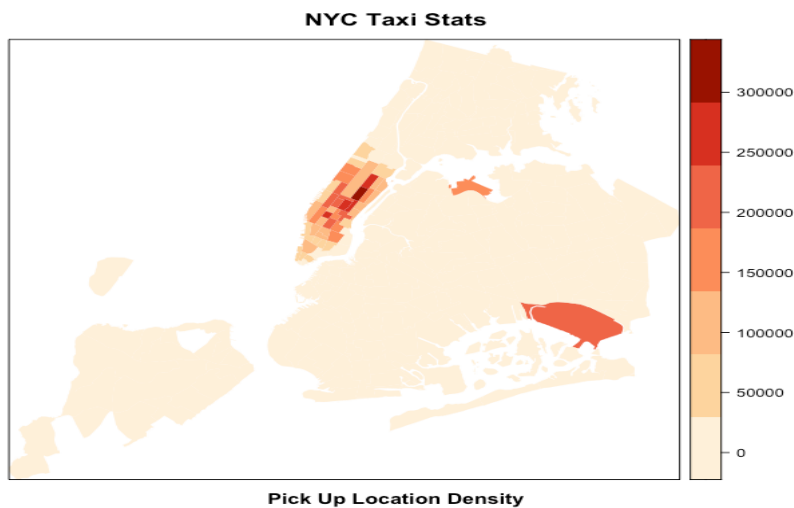
Clustering:

Clustering analysis is used for segmenting the data based on distances. For this project, to better understand how people use taxis, the clustering model was used to segment all the trips. Amongst all the clustering methods, a common problem that occurs is to determine the number of clusters. Hence, we have chosen a hierarchical cluster to segment all the trips to calculate the distances between any two observations. With this analysis, we were able to get further directions in order to drill down further.

Results

Spatial Analysis:

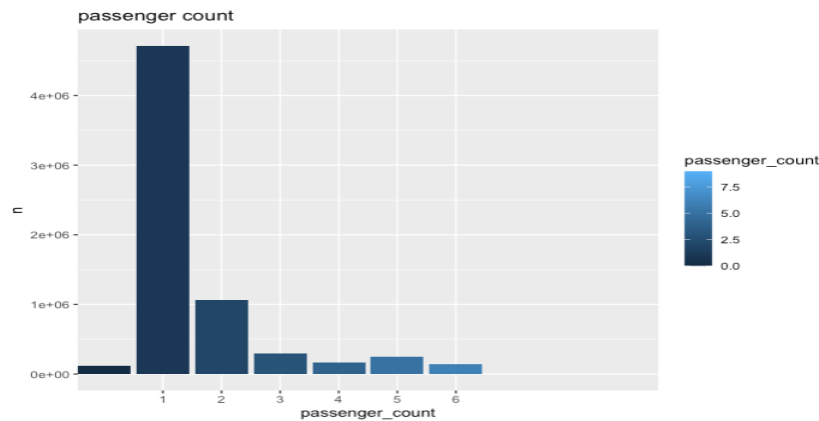
This approach was used to drive the overall NYC taxi statistics with regards to picking up location density. From the graph below, it will be appropriate to say that Upper East Side North, Upper East Side South, Midtown Center, Time Square /Theatre Districts, LaGuardia, JFK locations have the highest pickups.



Data Drill Down

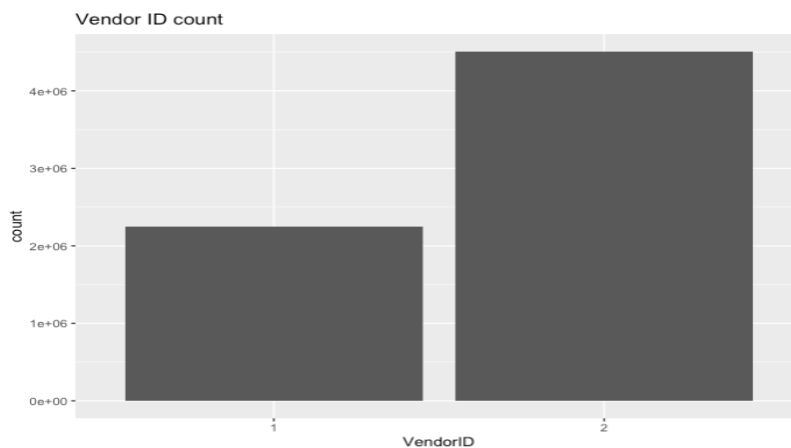
Riders by number:

The majority of the trips only have one passenger. Car sharing is a good idea since it will average the cost to each passenger, lowering the cost for each rider, thus increasing the total ridership. Uber and Lyft both have car-sharing functions where passengers can share cars with people going in the same direction. Since taxis usually pick up passenger roadside, taxi companies need to work with ride-hire apps like Arro and Curb.



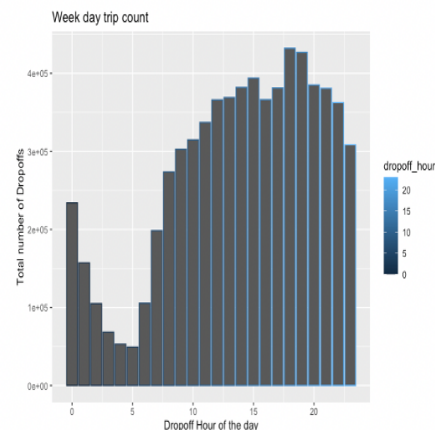
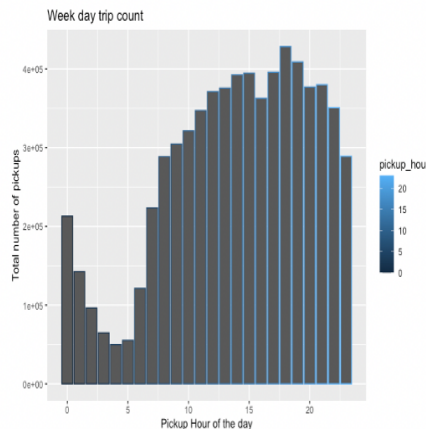
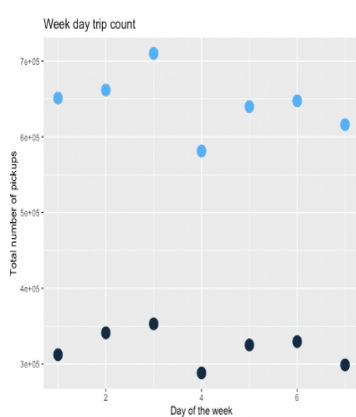
Trips by Vendor ID:

New York has two types of Taxi - green taxi (vendor ID = 1) and yellow taxi (vendor ID = 2). The green taxi is also known as a boro taxi. The boro taxis can be hailed in Manhattan north of East 96th Street and West 110th Street, and all outer boroughs (the Bronx, Brooklyn, Queens, and Staten Island) except at the airports. The vehicles can drop passengers off anywhere, but cannot pick up new passengers within the "yellow zone" (south of East 96th and West 110th Streets) or airports. By contrast, yellow cabs can pick up and drop off passengers anywhere in the city. (NYC T&LC, 2013) Since the majority of the trips are short trips within Manhattan. It is not a surprise that Vendor ID 1(boro taxi) has only half of the number of yellow cab rides.



Trips by days/pickup & drop off hours:

From plots below, we can identify that both types of taxis are busiest on Tuesdays, which is kind of surprising. Pickup and drop-off numbers gradually increased throughout the day, peaking at 6 pm.



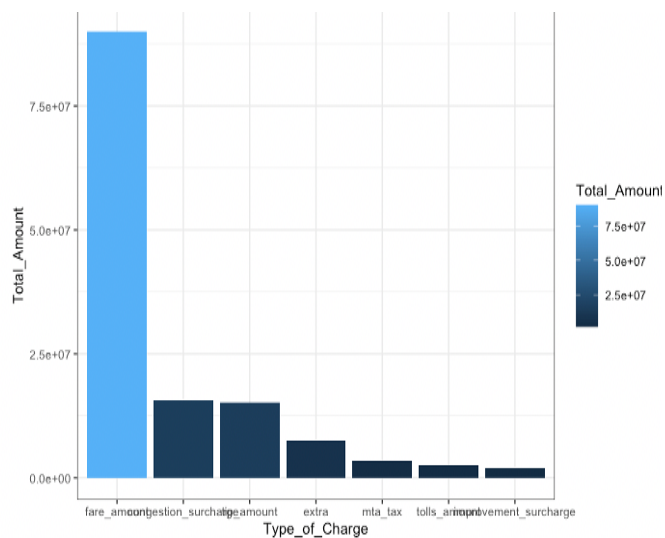
Cost Analysis:

For conducting a crisp cost analysis of yellow taxi cabs, we calculated Total Revenue for all cabs, revenue per taxi, operational costs, depreciation costs & income. Details are mentioned below:

Overall Revenue Bifurcation:

Overall revenue for one month was calculated along with revenue bifurcation. It is interesting to note that the total revenue for yellow taxi cabs is 1,36,24,6946 USD for Dec 2019. It is interesting to note that the fare amount is only 66% as rest are included in other brackets. The Tip amount is shown to be 11.20% as cash tips are not included, and these are the tips only recorded. Details can be visualized more from the graph & table (included for better insights on numbers) below:

Revenue_bifucration		
Type_of_Charge	Total_Amount(USD)	%age
fare_amount	89917684	66.00%
extra	7470098	5.48%
mta_tax	3362287	2.47%
tip_amount	15261600	11.20%
tolls_amount	2531853	1.86%
improvement_surcharge	2026132	1.49%
congestion_surcharge	15677292	11.51%
Total_revenue	136246946	

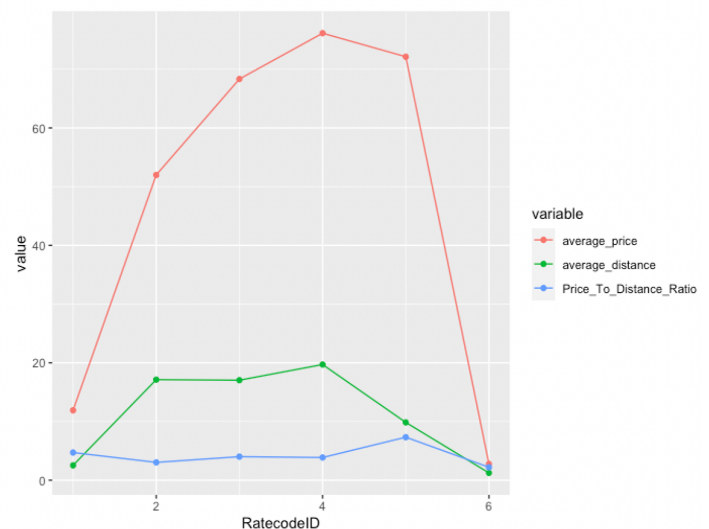


Analyzing profitable routes identified by Rate Code ID:

Dataset provides a column known as Rate code id, which has six different levels of numbers. Each rate code id is selected before starting a trip, and accordingly, the charge is calculated. (Details on

different Rate code ids are: 1= Standard rate,2=JFK,3=Newark,4=Nassau or Westchester,5=Negotiated fare,6=Group ride) For analyzing the difference between different rate code I.D.s, we calculated the mean of fare amount and the distance traveled by different rate code I.D.s. Further, it calculated the price to distance ratio in order to understand the profitable routes. Based on the outcome routes with rate code, ID 5 is more profitable. Details are depicted below in graph:

RatecodeID	average_distance	average_price	Distance to Price Ratio
1	2.53	11.90	4.70
2	17.10	52.00	3.04
3	17.00	68.30	4.02
4	19.70	76.10	3.86
5	9.84	72.10	7.33
6	1.24	2.75	2.22



Cost vs. Income analysis:

For understanding revenue versus income, it is imperative to understand the overall operating costs of a taxi. To calculate operating costs, we have considered some external sources as mentioned below (sources for it are mentioned in the references section at the end):

Details on factors used for operating costs	
Variables	Count/cost
Count of Taxis	1387
Fuel cost	2.5(miles per gallon)
Occupancy factor	1.5
Mileage	15
Insurance cost	87.5(USD per month)
Maintenance cost	166.67(USD per month)
License cost	3000(USD per month)
Depreciation cost	777.78(USD per month)

Overall Cost Calculations for single taxi:

Based on the data available, we are able to retrieve overall revenue generated by single taxi as bifurcated in the table below:

Cost Analysis per Taxi	
<i>Variable</i>	<i>Cost (in USD)</i>
Total revenue monthly per taxi	9832.355
Total Tolls/Taxes Monthly per taxi	1702.935
Average tip (Only credit card/recorded ones included)	1101.364
Monthly Revenue single Cab	7028.057

Operational Cost:

For calculating operational costs, we included monthly insurance cost, maintenance cost, fuel cost, monthly license cost, Depreciation cost and results are as indicated below:

Operational cost	4394.667	44.70%
------------------	----------	--------

Income for single taxi:

Income for single taxi was derived based on monthly revenue of single taxi cab, operational costs & deductible savings and details are as mentioned below:

Income per taxi	2946.238	29.96%
-----------------	----------	--------

Tax Percentage:

Overall a taxi is paying 17.32% of overall taxes as listed below:

Tax Percentage	1702.935	17.32%
----------------	----------	--------

Salaries for Taxi Driver:

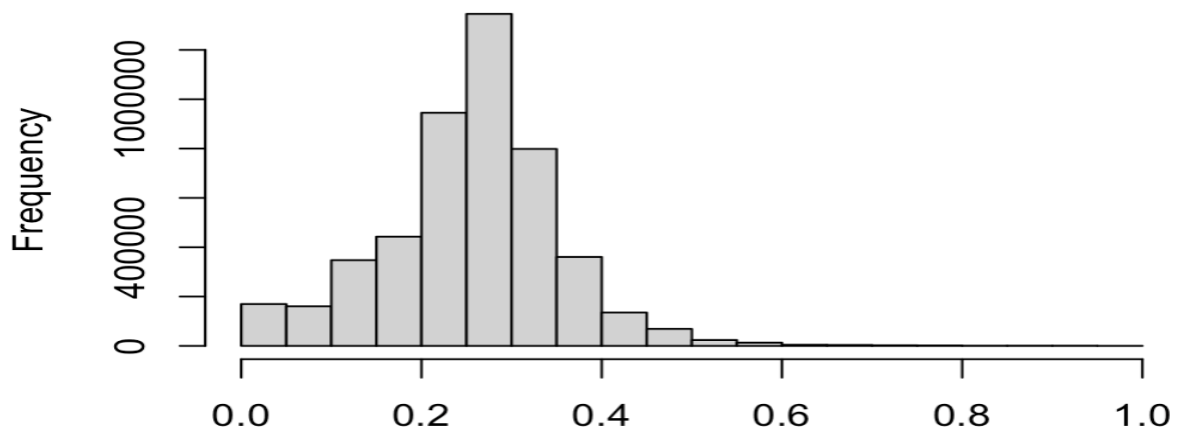
Based on salary.com, below are the brackets of the salary of taxi drivers. Considering the income of a single taxi as calculated above, overall yellow taxi cabs are in loss. However, if a person chooses to drive its car, then by considering tips, we can conclude that they can achieve a minimum bracket of the salary of a taxi driver since the amount of the tip is not included in the above figures.

Taxi Driver Salary	Cost (in USD)
Low Salary	36132
Medium Salary	44611
High Salary	53090

Conclusions:

- Based on the overall cost analysis, it is concluded that a single taxi cab can only earn 29.96% of the overall revenue.
- Overall Yellow taxi cabs are not able to even meet the lowest bracket of income of a driver's salary.

- Routes with Rate code ID 5 are more profitable routes as the distance to price ratio seems to be more profitable with this rate code I.D.
- Upper East Side North, Upper East Side South, Midtown Center, Time Square /Theatre Districts, LaGuardia, JFK locations has majority pickup locations as compared to various other locations in New York.
- The majority of the rides have single riders & based on the analysis, and group rides are considered to be the more profitable ones.
- Tuesdays come out to be the busiest day of the week as per one month's data, and more insights can be derived based on the yearly database.
- Pickup and drop-off numbers gradually increased throughout the day and are mostly at a peak at 6 pm.
- Tips that are recorded have higher percentage as visualized from graph below:



Recommendations:

- The carpool is recommended to lower the cost for individuals, attracting more people to take taxis and eventually contributing to overall revenue.
- For taking a break, Wednesdays as the recommended due to lowest trip numbers
- Evening times are busier, so adjusting the driver's time following it will help in achieving more income
- A new taxi driver should choose a yellow cab rather than a boro taxi.
- Based on pick up analysis, is it recommended to overlap the driver's working shifts in the evening time to optimize the fleet.
- Based on spatial analysis, cab drivers' Recommendation would be to move to the nearest of the identified central locations, namely Upper East Side North, Upper East Side South, Midtown Center, Time Square /Theatre Districts, LaGuardia, JFK locations, to fetch a pickup.

References

Source of dataset: <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

Majaski, C. (2020, March 07). Uber vs. Yellow Cabs in New York City: What's the Difference? Retrieved August 13, 2020, from <https://www.investopedia.com/articles/personal-finance/021015/uber-versus-yellow-cabs-new-york-city.asp>

Crudele, J. (2017, September 19). The challenges of driving a yellow cab in the age of Uber. Retrieved August 13, 2020, from <https://nypost.com/2017/09/18/the-challenges-of-driving-a-taxi-in-the-age-of-uber/>

NYC Taxi & Limousine Commission,(2013, June) "Guide to understanding the Street Hail Livery (SHL) Service Rules and Requirements
Retrieved December 18, 2013, from
https://web.archive.org/web/20180927204255/http://www.nyc.gov/html/tlc/downloads/pdf/shl_guide.pdf

Count of Taxis : Sugar, R. (2017, January 17). Uber and Lyft cars now outnumber yellow cabs in NYC 4 to 1. Retrieved August 13, 2020, from <https://ny.curbed.com/2017/1/17/14296892/yellow-taxi-nyc-uber-lyft-via-numbers>

Fuel cost : New York Times. (2006, May 10). Calculate how much your gas really costs. Retrieved August 13, 2020, from
https://archive.nytimes.com/www.nytimes.com/packages/html/business/20060510_LEONHARDT/cost_per_mile.html?scp=115

Mileage: Voelcker, J. (2010, July 29). Cities Want High-Mileage Hybrid Taxis; Judge Says It's Illegal. Retrieved August 13, 2020, from https://www.greencarreports.com/news/1047703_cities-want-high-mileage-hybrid-taxis-judge-says-its-illegal

Insurance cost: Reitz, J. (2016, August 03). Cost Of Owning A Car In New York State. Retrieved August 13, 2020, from <https://lite987.com/cost-of-owning-a-car-in-new-york-state/>

maintenance cost: NYC Taxi & Limousine Commission, (2013). The Cost Of Maintaining A Taxi Vehicle In The U.S. & The Alternative Fuel Vehicle Myth
Retrieved August 13, 2020, from http://www.tlc-mag.com/archive/pre_2013_site/tlpa_jan10.html

License cost : A, H. (2019, August 27). What is a taxi medallion worth? Retrieved August 13, 2020, from <https://www.cityandstateny.com/articles/personality/interviews-profiles/what-are-taxi-medallions-worth.html>

Depreciation cost : Chang, S. (2020, May 18). Thinking About Getting A Car In NYC? Think Again. Retrieved August 13, 2020, from <https://gothamist.com/news/thinking-about-getting-car-nyc-think-again>

Salaries: Salary.com, . (2020, July 27). Taxi Driver Salary in New York, NY.

Retrieved August 13, 2020, from <https://www.salary.com/research/salary/benchmark/taxi-driver-salary/new-york-ny>

Deductibles: .IRS, (2019, December 31). IRS issues standard mileage rates for 2020.
Retrieved August 13, 2020, from <https://www.irs.gov/newsroom/irs-issues-standard-mileage-rates-for-2020>

*Occupancy factor: Occupancy factor is considered with the fact that every time a taxi does not have rider & they may have to commute more to pick up ride so no amount is paid for that trip