A Project Report on

Navigating the City That Never Sleeps: A Visual Exploration of NYC Taxi Data

Capstone Project Final Report

Submitted in partial fulfilment of requirement of

Master of Information Technology and Analytics (MITA)

Semester: Fall 2023

Batch September 2022 – December 2023

Submitted to:

Professor Michail Xyntarakis

Submitted by:

Kireeti Mantrala

sm2594

RUTGERS BUSINESS SCHOOL, NEWARK

INTRODUCTION

Embarking on an extensive quantitative examination of the New York City Taxi and Limousine Commission (TLC) trip record data, this research endeavor seeks to unveil pivotal insights in the intricate tapestry of urban transportation. TLC has released public datasets that contain data for taxi trips in NYC, including timestamps, pickup & drop-off locations, number of passengers, type of payment, and fare amount. These data sets can be found in this BigQuery public datasets link: https://bigquery.cloud.google.com/dataset/bigquery-publicdata:new_york The meticulous analysis of TLC trip records becomes a conduit for understanding the nuanced dynamics of taxi services within the city. By deciphering patterns in popular areas, a comprehensive map of high-demand zones emerges, offering taxi drivers a strategic guide for optimizing their earnings by harmonizing with the dynamic pulse of the city.

This exploration dives deeper into the temporal ebbs and flows, pinpointing peak hours and days when taxi activity reaches its zenith. This temporal roadmap empowers drivers to synchronize their schedules, transforming routine navigation into a choreography that aligns seamlessly with the city's rhythms. Simultaneously, the estimation of average trip costs unravels a financial narrative, endowing passengers with foresight that transcends the mere transactional nature of their journeys. This transparency not only aids passengers in effective budgeting but also nurtures a sense of trust in the transportation system. The inquiry extends beyond monetary considerations by delving into the intricate realm of tipping patterns. By decoding the nuances of gratitude inherent in these transactions, drivers gain insights into areas where generosity tends to flourish. For passengers, it demystifies the unspoken norms, fostering a more harmonious exchange between service providers and recipients.

The project's paramount significance lies in its positioning as a beacon, not only interpreting historical data but aligning its insights with the contemporary trajectories of urban transportation trends. Through a fusion of statistical acumen and creative data visualization, this project aspires to present a vibrant tableau that transcends the numerical realm, providing a nuanced understanding of the city's heartbeat. In this intricate dance between algorithms and human behavior, the project endeavors to contribute a transformative chapter to the ongoing narrative of New York City's dynamic taxi landscape.

ABOUT THE DATA

This analysis delves into the dynamic realm of NYC yellow taxi operations, utilizing a comprehensive dataset(full_taxi.csv) sourced from New York City Taxi and Limousine Commission (TLC). The dataset encompasses a rich array of information, ranging from drop off times, locations taxi zones, tip rates, fare amounts, trip durations, trip zones, trip distances, and various other pertinent details. To ensure precision and relevance, I meticulously cleaned the dataset using SQL and Excel to organize it into distinct tables tailored to my analytical needs, thus paving the way for compelling visuals in Tableau. These tables include a spectrum of essential insights:

1. Median Fare Amounts:
   This table captures the median values of fare amounts, offering a nuanced understanding of the central tendencies in pricing. It provides a comprehensive overview of the fare landscape and aids in discerning patterns and outliers.

2. Number of Trips by Route:
   By dissecting the dataset, I've curated a table that outlines the number of trips based on specific routes. This breakdown is instrumental in identifying the popularity of different taxi routes, shedding light on high-traffic corridors and potentially underserved areas.

3. Tip Rates:
   The tip rate table encapsulates crucial information about the generosity of passengers. Understanding tipping patterns is pivotal for taxi drivers aiming to maximize their earnings. This table highlights areas where passengers are more likely to tip generously.

4. Trend Radial Data:
   This visually dynamic table provides a radial perspective on trends related to taxi activities. By leveraging radial data, it becomes easier to discern patterns based on geographical distribution, enabling a spatial analysis of taxi hotspots and trends.

5. Trip Duration and Distance:
   Focusing on the temporal and spatial aspects of taxi journeys, this table distills insights into trip durations and distances. Understanding the average time and distance covered per trip is vital for both drivers and passengers seeking to optimize their travel plans.

In addition to the raw dataset, I have enriched the analysis by incorporating supplementary data from online sources, thereby enhancing the depth and breadth of insights derived. This comprehensive approach ensures that the analysis is not only robust but also aligned with real world dynamics, making it an invaluable resource for taxi drivers seeking to optimize their routes and for passengers looking to navigate NYC efficiently.

PROJECT OVERVIEW

The project commenced with the acquisition of extensive data pertaining to yellow taxi services. Subsequently, data cleaning was undertaken employing Excel, and SQL was employed for further segmentation of the datasets. Additional insights were gleaned from Google Big Query and relevant Medium articles, aiding in formulating a structured exploration strategy. While initial experimentation with Gephi was conducted, the decision to forgo it was based on the rationale that a comparative analysis across multiple taxi providers would be more apt, aligning with a network analysis approach. Consequently, Tableau was selected as the primary tool for visualization.

Visualization Plan:
The final visualization plan revolves around employing Tableau to communicate key insights drawn from the dataset. The focal points of the visualizations include:

1. Median Price Exploration:
   Investigating average fares for taxi trips across boroughs by summing up median prices. Navigation is incorporated to elucidate trip distance medians.

2. Trip Duration in Boroughs:
   Analyzing variations in trip durations across different New York City boroughs.

3. Tip Rates Table:
   Uncovering passenger tipping habits in pickup boroughs, categorized into percentage brackets and including a "no tip" column.

4. Top Three Boroughs:
   Identifying boroughs with the highest taxi trip frequencies and delving into the factors contributing to their popularity.

5. Manhattan 2016:
   Utilizing a line graph to illustrate the peaks and declines in taxi trips within Manhattan over the six months of 2016.

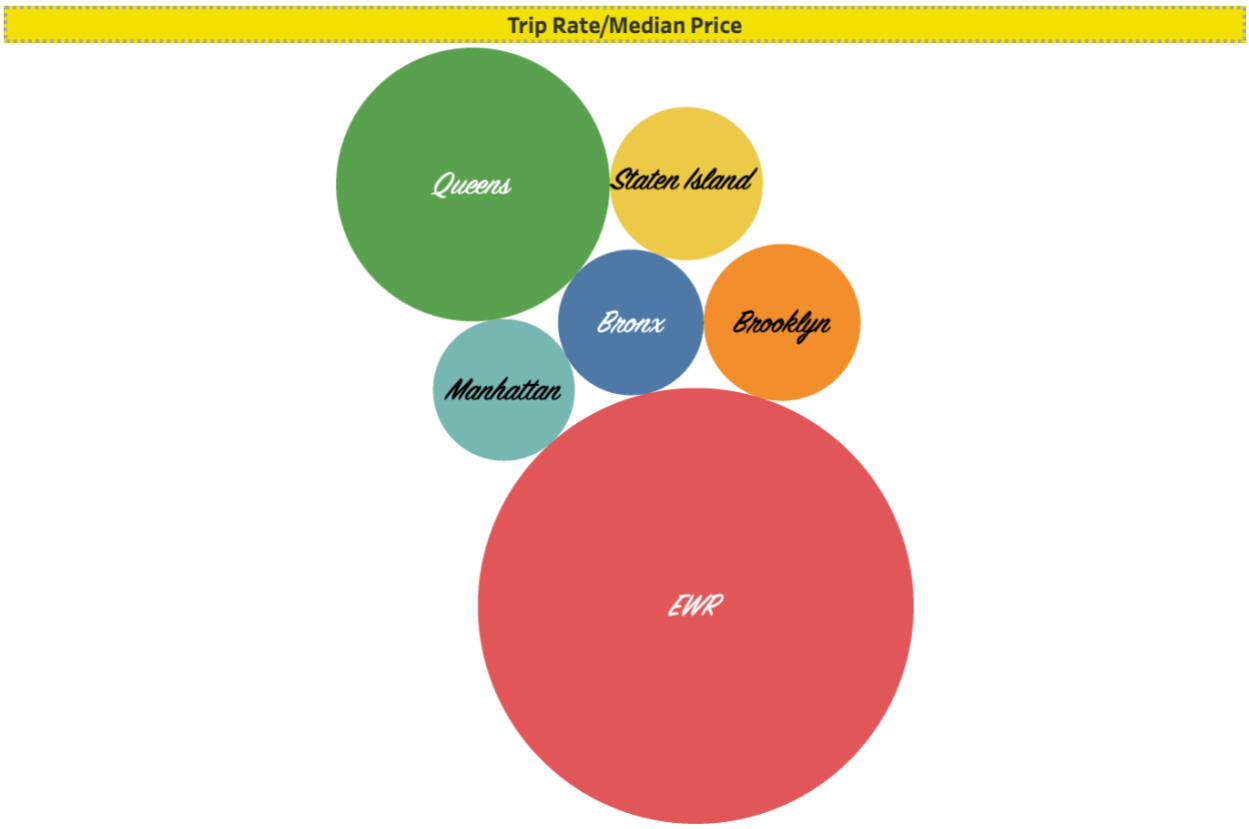
6. Week Clock:
   Presenting an hourly clock visualization for each day, depicting the day-night pattern of taxi trips.

7. By Hour Analysis:
   Offering insights into trip patterns through an analysis of average trips per hour.

8. Borough Overview:
   A map view showcasing the boroughs, with a focus on Manhattan and highlighting EWR.

The report will conclude with a detailed explanation of each visualization and the corresponding insights derived from the data. This structured approach aims to provide a comprehensive understanding of the taxi service landscape in New York City.

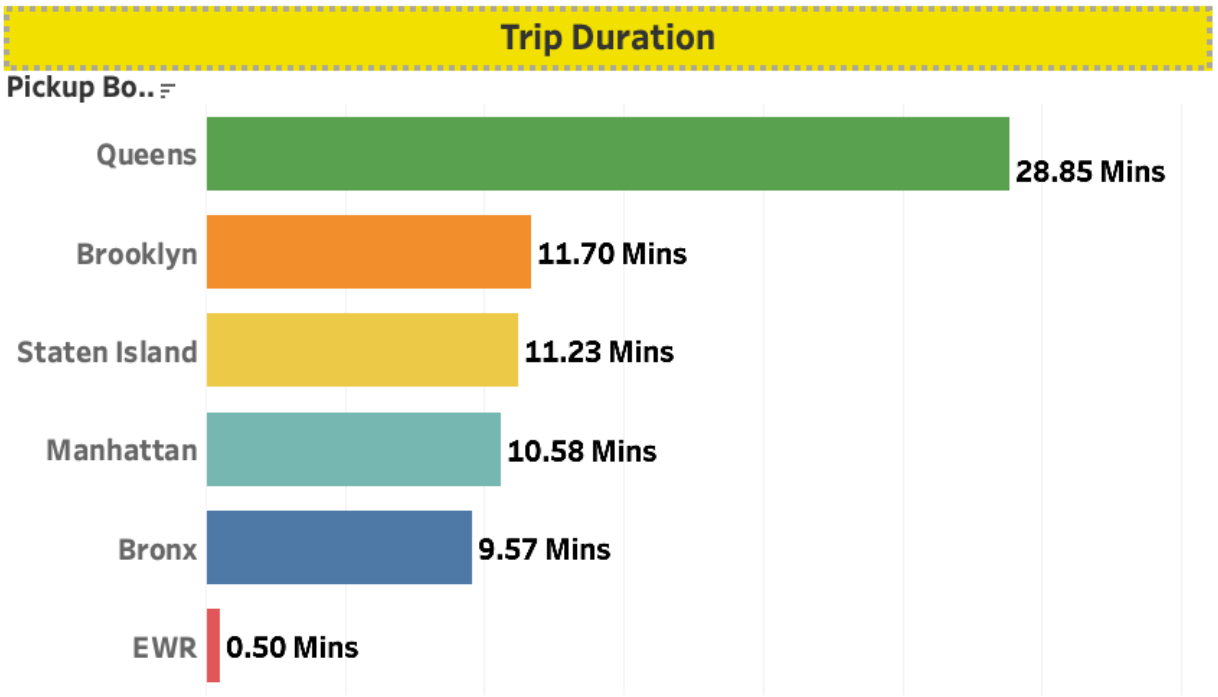
VISUALS AND INSIGHTS



The initial visualization presents a comprehensive overview of average taxi fares across the five boroughs and EWR, showcasing the aggregated median prices. Elevate the interactive experience by incorporating Tableau's hover functionality, allowing users to unveil supplementary insights effortlessly. Hovering over specific data points will unveil a detailed tooltip, revealing key information such as the median trip distance.

Notably, the highest median fare is observed at EWR, standing at $85, indicating it as the costliest destination. Conversely, Manhattan boasts the lowest median fare among the boroughs. Delving deeper into the analysis, the tooltip exposes intriguing nuances in the median trip distances. Surprisingly, EWR registers a median trip distance of 0, hinting at unique travel dynamics to and from the airport. In contrast, Queens emerges with the highest median trip distance, reaching 10.27 miles, shedding light on extensive travel within the borough and pave a path in understanding the taxis to and from JFK and LaGuardia Airports.

To enrich the understanding of travel patterns, a further exploration of trip duration is needed. Unraveling the complexities of trip dynamics through this multifaceted analysis ensures a more nuanced comprehension of taxi fare distribution across the diverse regions.



A compelling revelation emerges as Queens takes the lead with the highest median trip duration, clocking in at 28.85 minutes. This insight provides valuable context, suggesting that taxis originating from Queens navigate through journeys with longer durations compared to other boroughs. In contrast, EWR showcases a notably brief median trip duration of 0.50 minutes, indicating quick pick-ups or drop-offs at the airport. The sharp contrast in trip durations between Queens and EWR underscores the diverse dynamics at play in these distinct locations.

This disparity prompts further exploration into the factors contributing to extended trip durations in Queens. Potential influencers may include traffic patterns, geographic layout, or the nature of destinations within the borough. Such an in-depth analysis not only enriches our understanding of travel dynamics but also lays the groundwork for targeted improvements in transportation services based on the unique characteristics of each region. Let’s see an analysis on the tipping patter for the five boroughs and the EWR Airport.

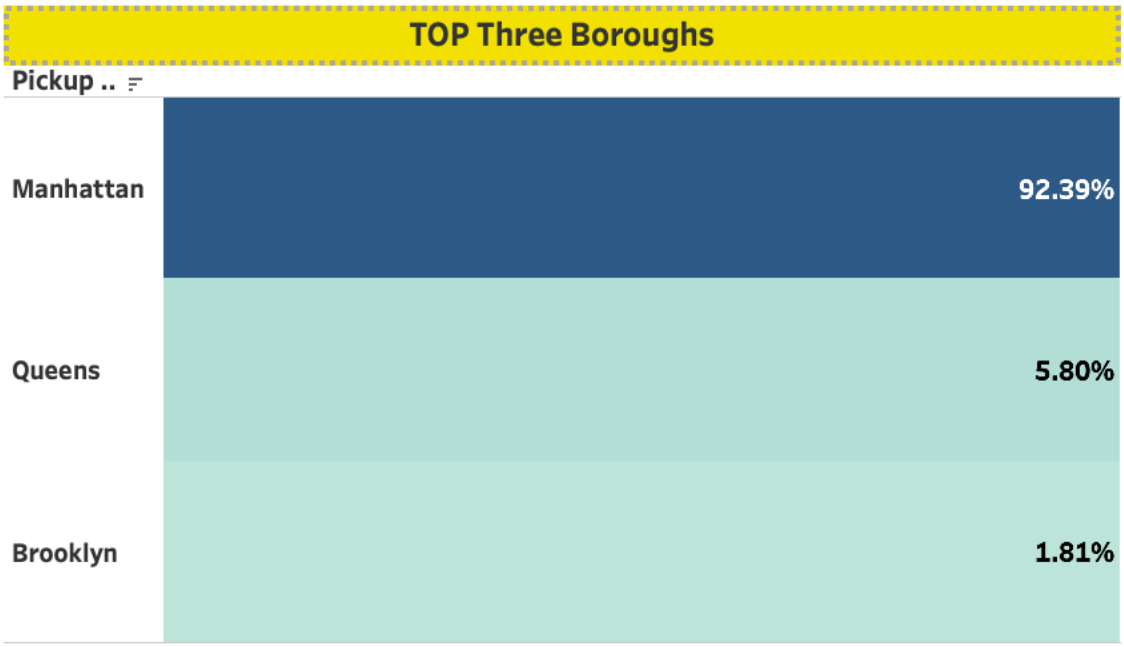
Tip Rate							
				Tip Category			
Pickup Bor..	no tip	More than 25%	Less than 5%	20% to 25%	15% to 20%	10% to 15%	5% to 10%
Bronx	73.00%	1.00%	1.00%	3.00%	15.00%	3.00%	3.00%
Brooklyn	40.00%	1.00%	1.00%	6.00%	41.00%	7.00%	5.00%
EWR	41.00%	6.00%	3.00%	1.00%	27.00%	10.00%	11.00%
Manhattan	36.00%	1.00%	1.00%	5.00%	42.00%	10.00%	6.00%
Queens	39.00%	0.00%	1.00%	4.00%	42.00%	9.00%	5.00%
Staten Island	40.00%	2.00%	1.00%	6.00%	37.00%	9.00%	5.00%

The presented visual table provides a comprehensive overview of tipping patterns across the boroughs and EWR. Notably, a prevailing trend emerges, indicating that passengers predominantly favor tipping within the range of 15%-20%. This preferred tipping bracket serves as a noteworthy observation, shedding light on the general inclination of passengers when expressing appreciation for taxi services.

Also, EWR stands out with the highest tipping rate, surpassing 25%. This revelation suggests a heightened generosity or satisfaction level among passengers traveling to or from the airport. The distinctive tipping behavior at EWR prompts an exploration into the factors influencing such a positive response, potentially tied to the nature of airport journeys or the quality of service provided.

Delving into the no-tip category, the Bronx emerges with a substantial figure, recording a 73% occurrence. This observation raises questions about the underlying dynamics, potentially linked to local customs, socioeconomic factors, or unique aspects of taxi service experiences within the borough.

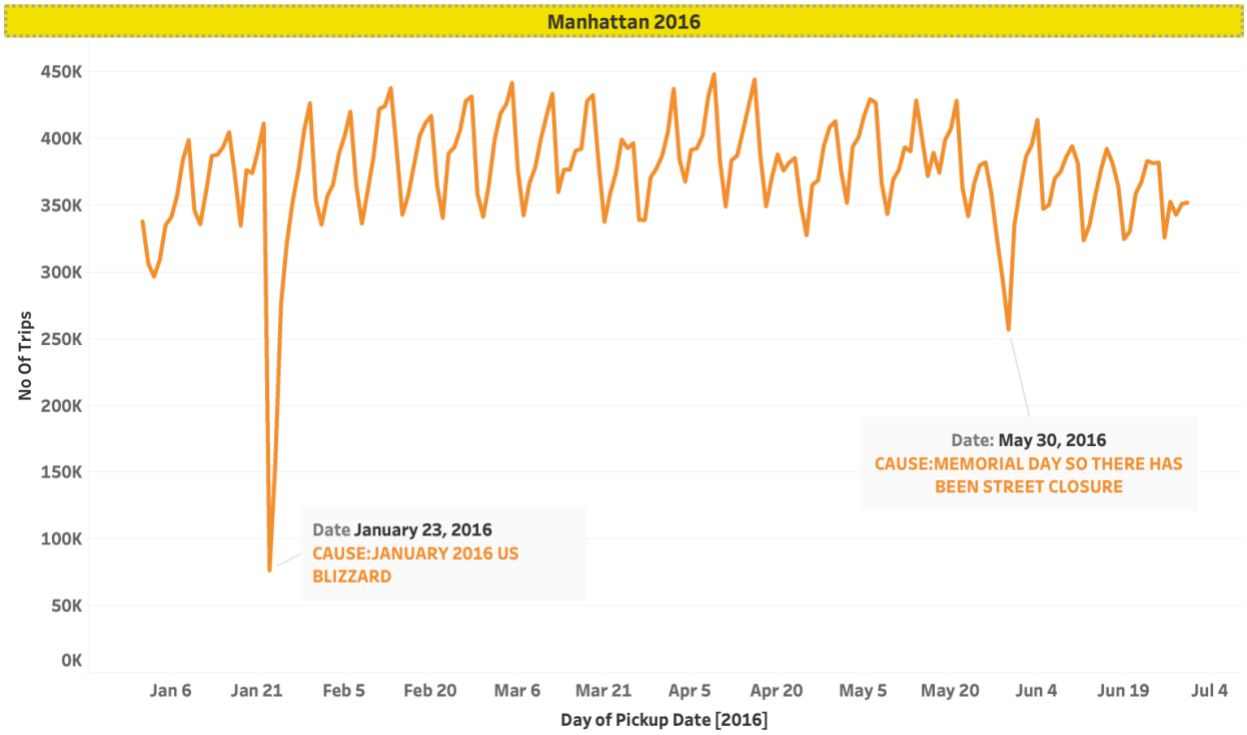
In essence, this nuanced analysis of tipping patterns not only provides insights into passenger preferences but also serves as a valuable tool for taxi service providers and policymakers seeking to enhance service quality and meet the diverse expectations across different regions.



The above visual shows the top three boroughs, Manhattan stands out with a notably high pickup rate of 92.39%, a surprisingly robust figure. This observation piqued my interest, prompting a more in-depth analysis to unravel the factors contributing to such a high pickup rate across all three boroughs and the airports.

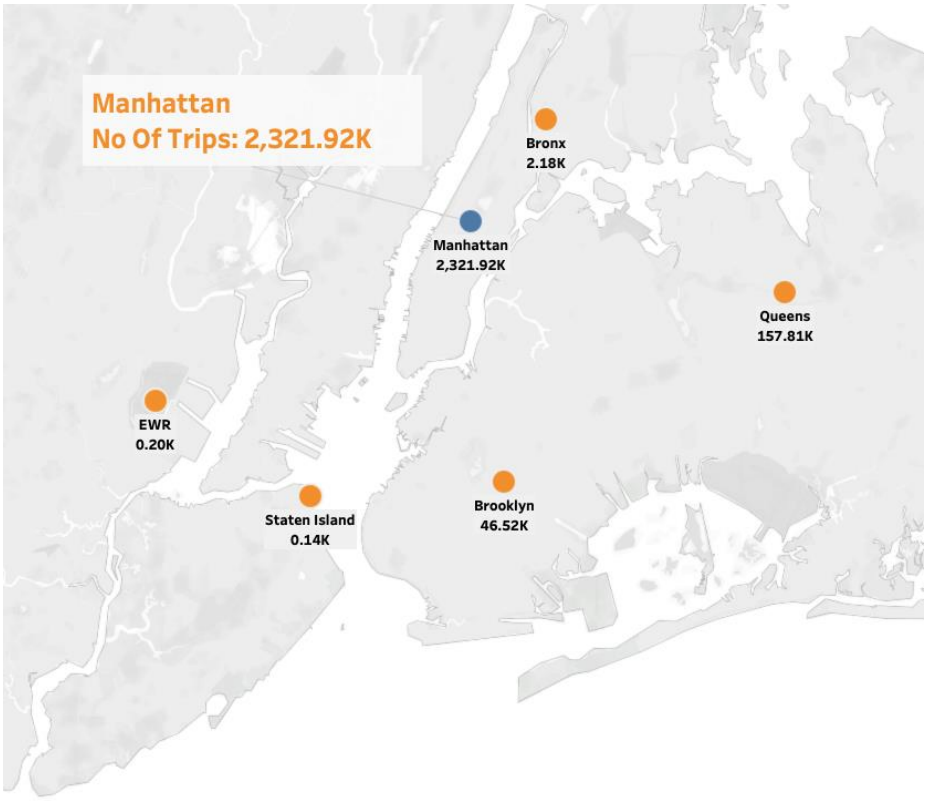
Upon delving deeper into the data and conducting thorough research, a compelling insight surfaces, the prevalence of short trips within and around Manhattan. This unique characteristic of the borough, characterized by a multitude of brief taxi journeys, offers a plausible explanation for the exceptionally high pickup rate.

To further illuminate Manhattan's trip dynamics, the subsequent three visuals delve into granular details, providing a comprehensive breakdown of trip rates and related statistics. By scrutinizing these visuals, we aim to uncover patterns, trends, and correlations that elucidate the distinctive features of Manhattan's taxi services. This meticulous analysis not only enriches our understanding of the data but also lays the groundwork for informed decision-making and strategic planning within the realm of transportation services in this dynamic urban landscape.

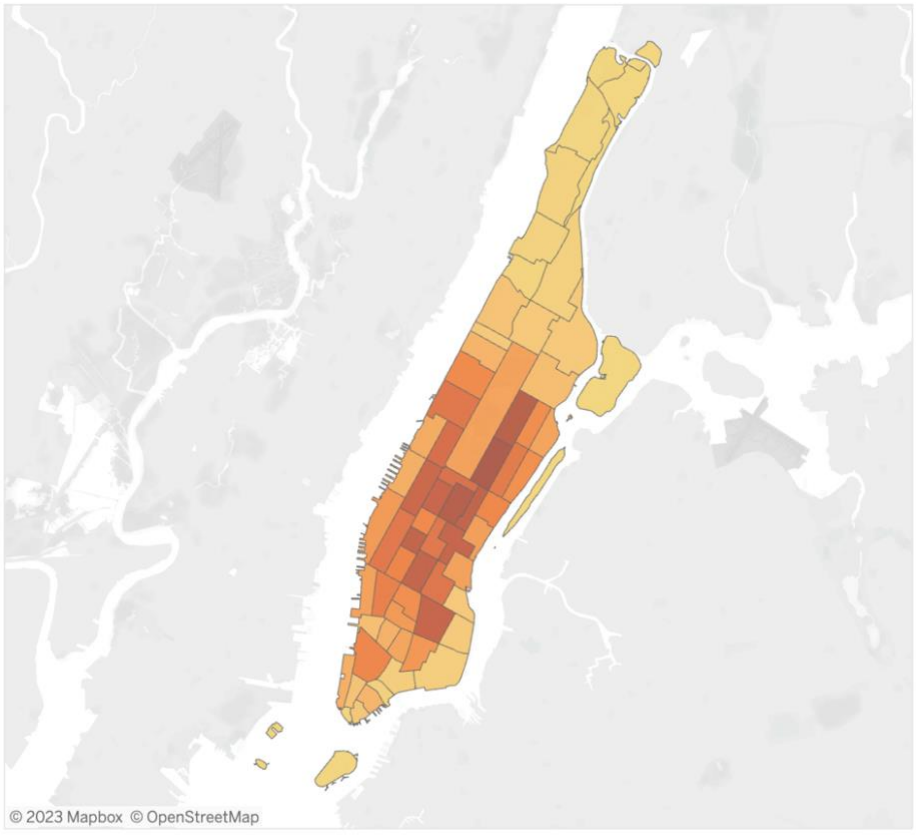


The depicted line graph illustrates the trajectory of Manhattan's trip volumes from January to July 2016, reaching a zenith of nearly 450,000 trips. Notably, two discernible downturns are evident in the graph. The first dip occurs on January 23rd, 2016, coinciding with the severe US Blizzard during that period. This weather-related event significantly impacted taxi rides, leading to a noticeable reduction in trip numbers.

The second notable decline unfolds on May 30th, 2016, attributed to Memorial Day and a concurrent street closure. The intersection of these two factors exerted an influence on the taxi ecosystem, resulting in a dip in the number of trips recorded on this particular day. This dual impact underscores the susceptibility of taxi services to both climatic conditions and urban infrastructure adjustments, affirming the interconnectedness of external factors with the observed trends in trip volumes.


Manhattan
No Of Trips: 2,321.92K
Bronx
2.18K
Manhattan
2,321.92K
Queens
157.81K
EWR
0.20K
Staten Island
0.14K
Brooklyn
46.52K

-A


© 2023 Mapbox © OpenStreetMap

-B

The above two visuals offer a detailed exploration focused on Manhattan. In the first visualization, Manhattan emerges as the preeminent borough in terms of taxi trips during 2016, registering a substantial 2,321.92k trips, the highest among all boroughs. This dominance is visually accentuated in the graphic.

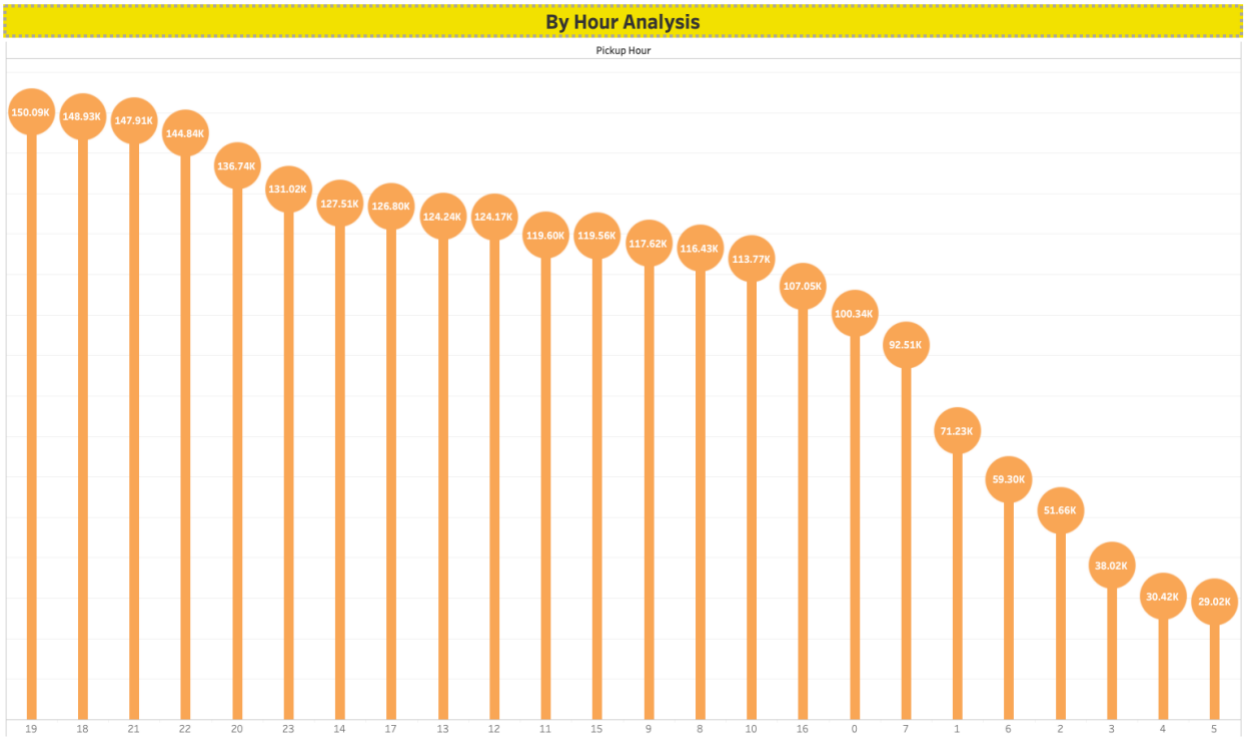
Following closely, Queens records 157.81k trips, securing the second position, while Brooklyn trails with 46.52k trips. The comparative figures underscore Manhattan's clear lead in taxi activity during the specified period.

The second visual delves into the borough's popularity by depicting common routes. Notably, the most frequented paths connect Upper East Side South to Upper East Side North. A filled map further accentuates the popularity of different zones within Manhattan, with darker shades indicating higher trip frequencies. This visualization offers a nuanced understanding of the borough's dynamics, emphasizing the concentration of taxi activities in specific areas, notably the Upper East Side, and reinforcing Manhattan's status as the epicenter of taxi demand.

A brief investigation into this substantial gap reveals a strategic pattern: taxis gravitate towards densely populated and bustling areas, such as Manhattan, where the potential for passengers is higher. This operational preference aligns with the industry's aim to maximize revenue, a strategy crucial for offsetting the exorbitant costs associated with acquiring a taxi medallion.

Notably, a medallion serves as a permit in the United States, granting taxi drivers the authority to operate. The scarcity of issued medallions, constrained by a predefined quantity, contributes to their elevated market value. This economic dynamic underscore the imperative for taxi drivers to concentrate their operations in areas with heightened passenger demand.

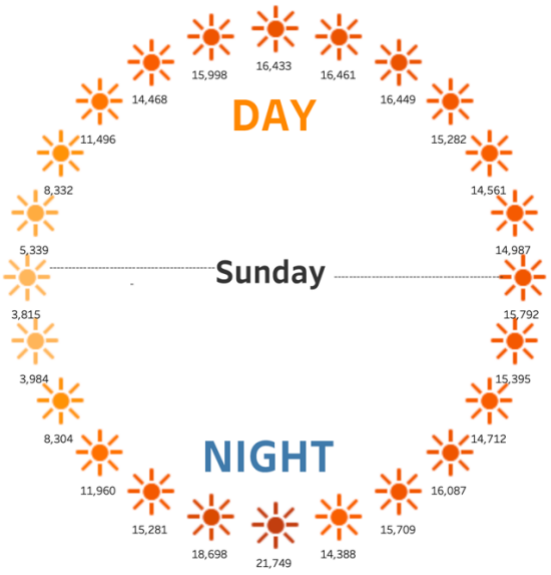
A deeper dive into the data hints at a plausible explanation for Manhattan's dominance— the prevalence of shorter trips. Research indicates that the most frequented routes within Manhattan predominantly traverse the Upper East Side. This concentration suggests a heightened dependence on taxis within this neighborhood, possibly facilitated by the abundance of taxis in the area. Whether due to convenience or local habits, residents and visitors alike in the Upper East Side seem more inclined to readily access taxis, creating a symbiotic relationship between the neighborhood's demand and the taxi supply. Compared to few other boroughs, Manhattan has relatively short trip duration and distance, but it is not the shortest. I also see an unusual result from EWR which has zero median distance, i.e. at least 50% of EWR data do not have distance information recorded. Maybe it is because EWR trips are those with negotiated fares or special Airport fares.

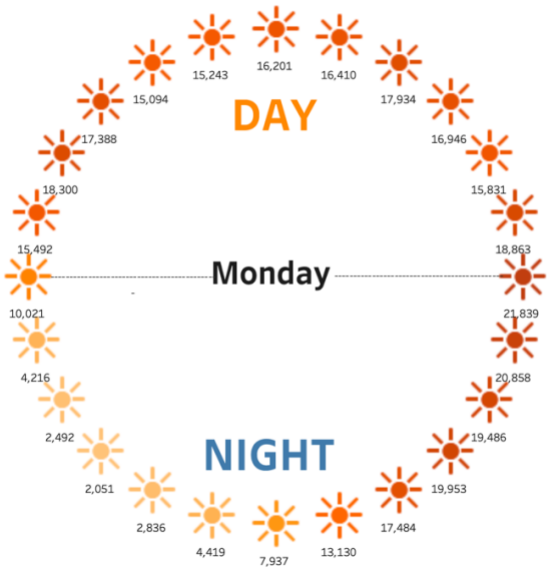


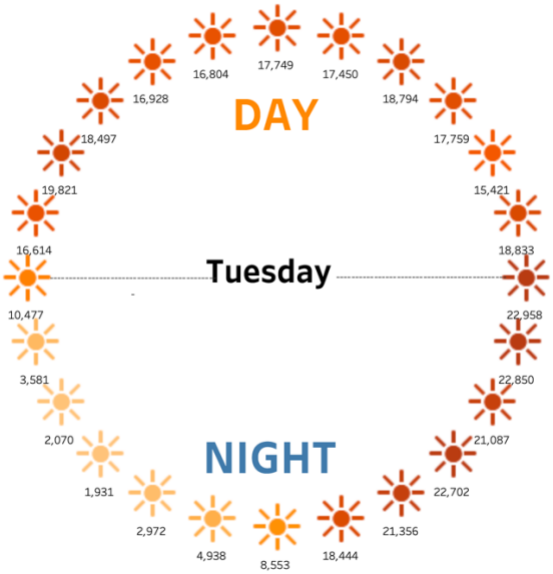
Illustrated in the lollipop graph is an hourly breakdown of the average number of trips, showcasing a notable peak during the 6-7 pm timeframe. For a more comprehensive exploration of this temporal trend, I've crafted an animated and interactive visualization using Tableau. Our forthcoming discussion will delve into the insights gleaned from this dynamic representation, providing a richer understanding of the fluctuations in trip counts throughout the day.

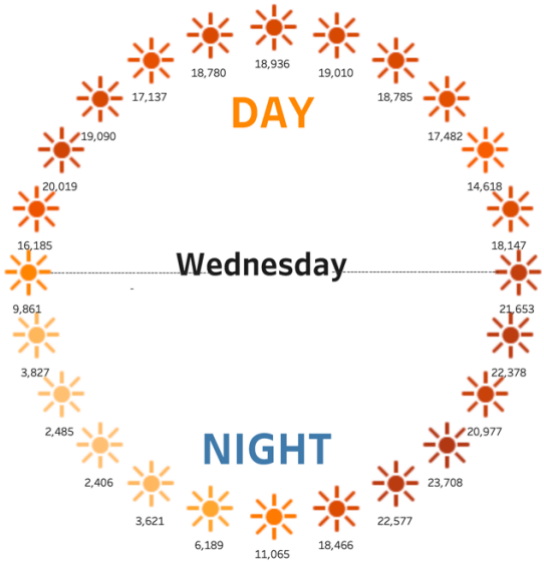
THERE IS CONTENT FURTHER

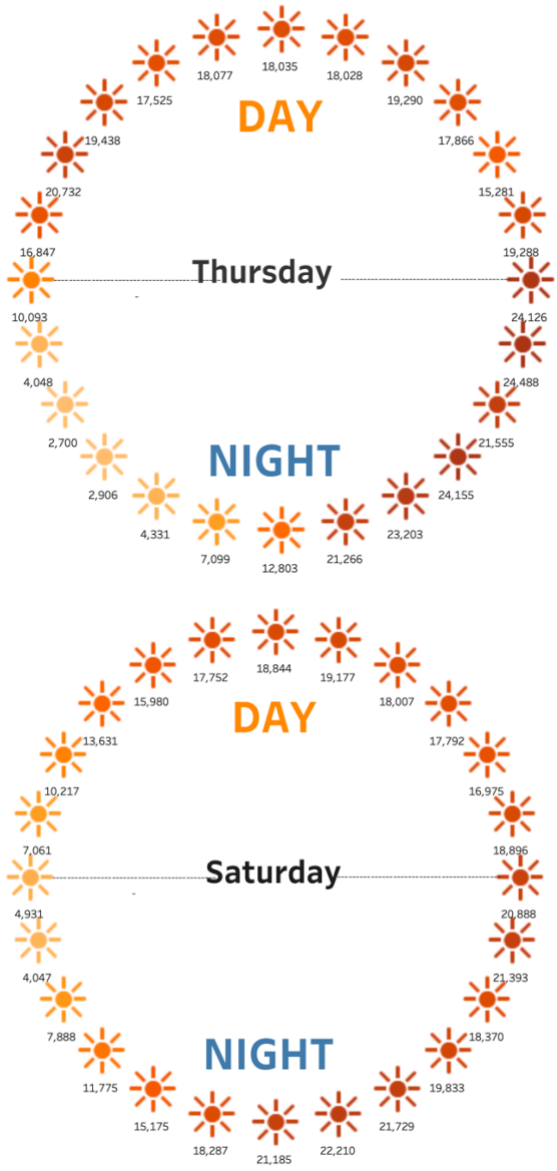
Week Clock

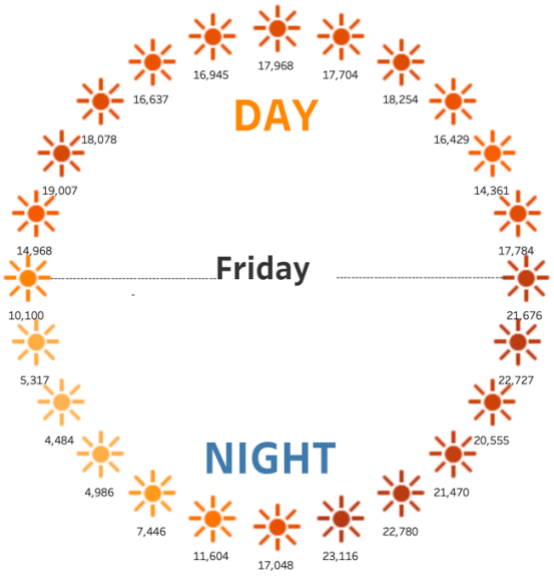

DAY
Sunday
NIGHT
15,998
16,433
16,461
16,449
15,282
14,561
14,987
15,792
15,395
14,712
16,087
15,709
14,388
21,749
18,698
15,281
11,960
8,304
3,984
3,815
5,339
8,332
11,496
14,468


DAY
Monday
NIGHT
15,243
16,201
16,410
17,934
16,946
15,831
18,863
21,839
20,858
19,486
19,953
17,484
13,130
7,937
4,419
2,836
2,051
2,492
4,216
10,021
15,492
18,300
17,388
15,094


DAY
Tuesday
NIGHT
16,804
17,749
17,450
18,794
17,759
15,421
18,833
22,958
22,850
21,087
22,702
21,356
18,444
8,553
4,938
2,972
1,931
2,070
3,581
10,477
16,614
19,821
18,497
16,928


DAY
Wednesday
NIGHT
18,780
18,936
19,010
18,785
17,482
14,618
18,147
21,653
22,378
20,977
23,708
22,577
18,466
11,065
6,189
3,621
2,406
2,485
3,827
9,861
16,185
20,019
19,090
17,137





Presented is a clock visualization, breaking down the hourly distribution of trips throughout the week, categorized into day and night periods. Each sun symbol represents an hour, with darker suns indicative of higher trip volumes, and the clock's center marking 12 pm. This is an animated visual in Tableau.

A discernible pattern emerges, revealing fewer trips on Monday midnight (00:00 am) in contrast to the heightened activity observed on Saturday midnight. This variance can be attributed to the social dynamics of the weekend, where individuals are more likely to extend their time outside on Friday nights and rely on taxis for transportation after midnight.

On Saturdays, the day's activities initiate later, commencing around 09:00 am, yet progressively intensify throughout the day. Conversely, Mondays witness an earlier start, approximately at 07:00 am, likely corresponding to the morning commute. The surge in activity for Mondays materializes

after working hours, peaking at around 06:00 pm. These distinctive trends underscore the nuanced temporal dynamics of taxi demand, influenced by both weekday routines and weekend social engagements.

LEARNINGS AND FINAL INSIGHTS

In the exploration of New York City's taxi landscape through a quantitative analysis of the Taxi and Limousine Commission (TLC) trip record data, several key insights have been uncovered. The TLC dataset, which includes information such as timestamps, pickup and drop-off locations, payment methods, and fare amounts, serves as a rich source for understanding the intricate dynamics of urban transportation. The analysis, conducted using Big Query public datasets, aims to provide not only historical interpretations but also to align its insights with the contemporary trajectories of urban transportation trends.

The geographical disparities in taxi fares and travel patterns across the city's boroughs reveal intriguing variations. EWR emerges as the costliest destination, while Queens exhibits the longest median trip distances. These insights are valuable for both taxi drivers, seeking to optimize travel plans, and passengers, looking to budget effectively.

Temporal dynamics play a crucial role in the taxi landscape, with peak activity observed during specific hours. External factors, such as weather events and street closures, significantly influence trip volumes, underscoring the interconnectedness of external elements with the taxi ecosystem.

Tipping patterns, reflecting passenger behavior, showcase a consistent preference for the 15%-20% range. Notably, EWR stands out with a higher tipping rate, suggesting heightened passenger satisfaction at airports. Conversely, the Bronx's high occurrence of no tips raises questions about local customs or service quality.

Manhattan emerges as the epicenter of taxi demand, recording significantly higher trip frequencies compared to other boroughs. Short trips within densely populated areas, particularly the Upper East Side, contribute to this dominance. The scarcity and high market value of taxi medallions drive drivers to focus on areas with elevated demand, reinforcing Manhattan's central role.

Temporal patterns, including hourly and weekly trip distributions, reveal distinct trends influenced by both weekday routines and weekend social engagements. Understanding these temporal dynamics is crucial for drivers in optimizing schedules and aligning with the city's rhythms.

The transformative power of data analysis in urban transportation is evident, providing valuable insights for taxi drivers to optimize routes and maximize earnings. Passengers benefit from transparent fare expectations and an understanding of tipping norms, fostering trust in the transportation system.

In conclusion, this research endeavor transcends mere data analysis by providing a holistic narrative of New York City's dynamic taxi landscape. It serves as a beacon, guiding stakeholders through the intricate dance between algorithms and human behavior, contributing a transformative

chapter to the ongoing narrative of urban transportation. As the city continues to evolve, these insights remain invaluable for shaping the future of taxi services and enhancing the overall experience for both drivers and passengers.

REFERENCE LINKS

1. https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page
2. https://medium.com/data-mining-the-city/change-of-yellow-taxi-ridership-a409eac8773a
3. https://www.nytimes.com/2021/06/10/nyregion/coronavirus-subway-yellow-cabs.html
4. https://medium.com/@haonanzhong/new-york-city-taxi-data-analysis-286e08b174a1
5. https://medium.com/@ShortHills_Tech/nyc-taxi-trip-duration-prediction-using-machine-learning-a92874bd761
6. https://medium.com/@dustinhsu0605/nyc-taxi-trips-challenge-646a89a05bfd
7. https://medium.com/geekculture/building-data-analysis-for-taxis-in-nyc-e91aca18a60d
8. https://www.kaggle.com/code/breemen/nyc-taxi-fare-data-exploration
9. https://www.kaggle.com/search?q=NYC+taxi