

NYC Robo-Taxi Feasibility Study



By Michael Cirino, Venkat Thamma, Tom Haight, Andrew Holland



Executive Summary

Our project entails looking at Data from NYC Open Data concerning for-hire taxicab rides, and to see how feasible a Self-Driving Taxicab Business would be when the technology is available.

Our goal is to be able to work with the data, find insights on where popular pick-up and drop off locations are, and to figure out the best locations for our taxis to be in order to capture market share.

In our analysis, we saw the demand volume high enough to justify investment into a self-driving taxicab business in New York City.

“Open Data for All New Yorkers. Open Data is free public data published by New York City agencies and other partners.”



Data **Collection**, Cleanup, and Exploration

Data Collection, specifically getting the data uploaded was a huge challenge for us because of filesize.

- There are roughly 15 Million for-hire rides per quarter

Because of this we had to pair down our analysis, removing columns that we did not need, and switch from our original goal of an annual analysis into a monthly analysis.

Additionally we paired the data down to Lyft as a sample size instead of all for-hire taxicab companies.

- Lyft is about 20% of total rides



Data Collection, Cleanup, and **Exploration**

Other fun things we learned in data collection:

- **Github** has an Upload limit of ~100mb... we broke this... twice!
- **Google Drive :**
 - Errors trying to use wget tool (found gdown tool)
 - Upkeep with adding/changing ID's for files, and inability for outside users to run clean up code (found way to use request.get)
- **Negative/Null values in the data set** - Customer Refunds created “bad data” that would skew actual results, for this study these we're removed.

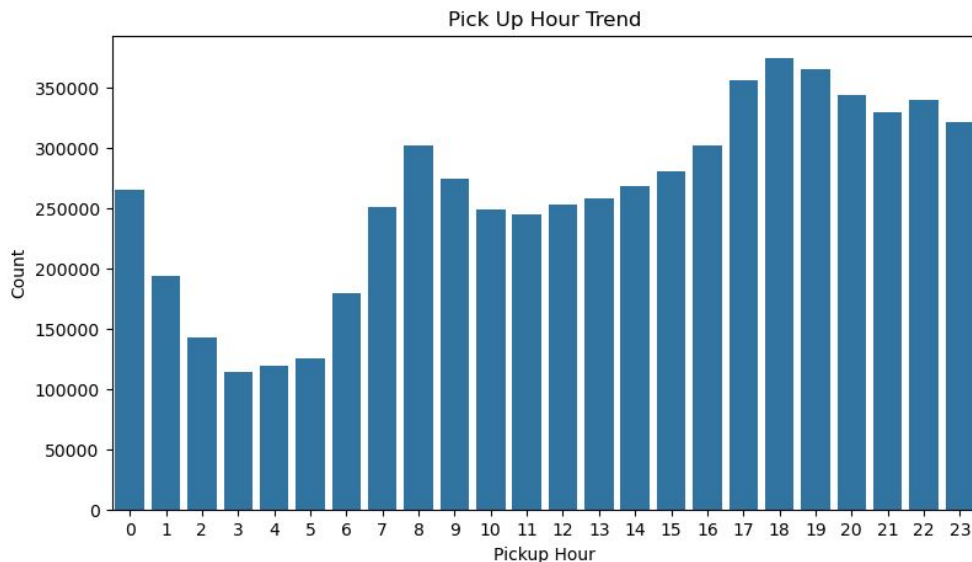
Data Collection, Cleanup, and Exploration



To better understand user usage we needed to define our dataset in different time series.

This meant looking at data through the lenses of...

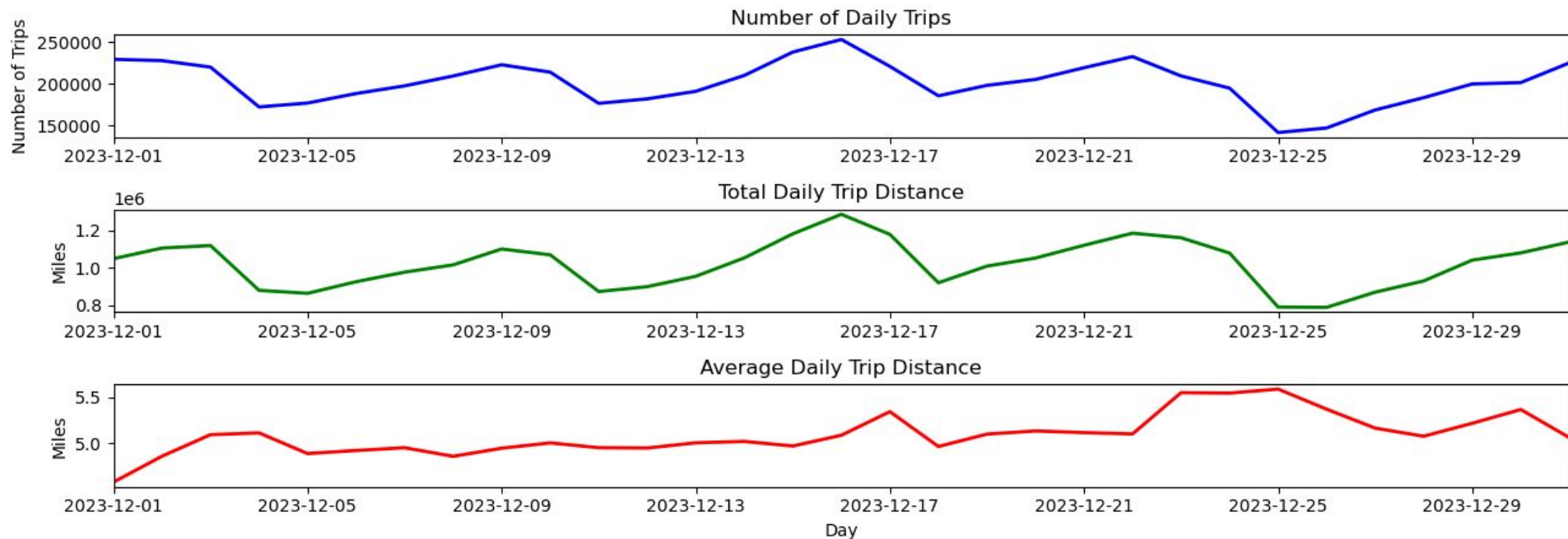
- Hourly pickup data - what times we're users most interested in getting a ride
- Daily and Weekly data - how did the trends look over a month
- Day of the Week - consolidating all the "Mondays", "Tuesdays", etc of a month to see tendencies



Key Takeaway: The majority of rides happen before and afterwork

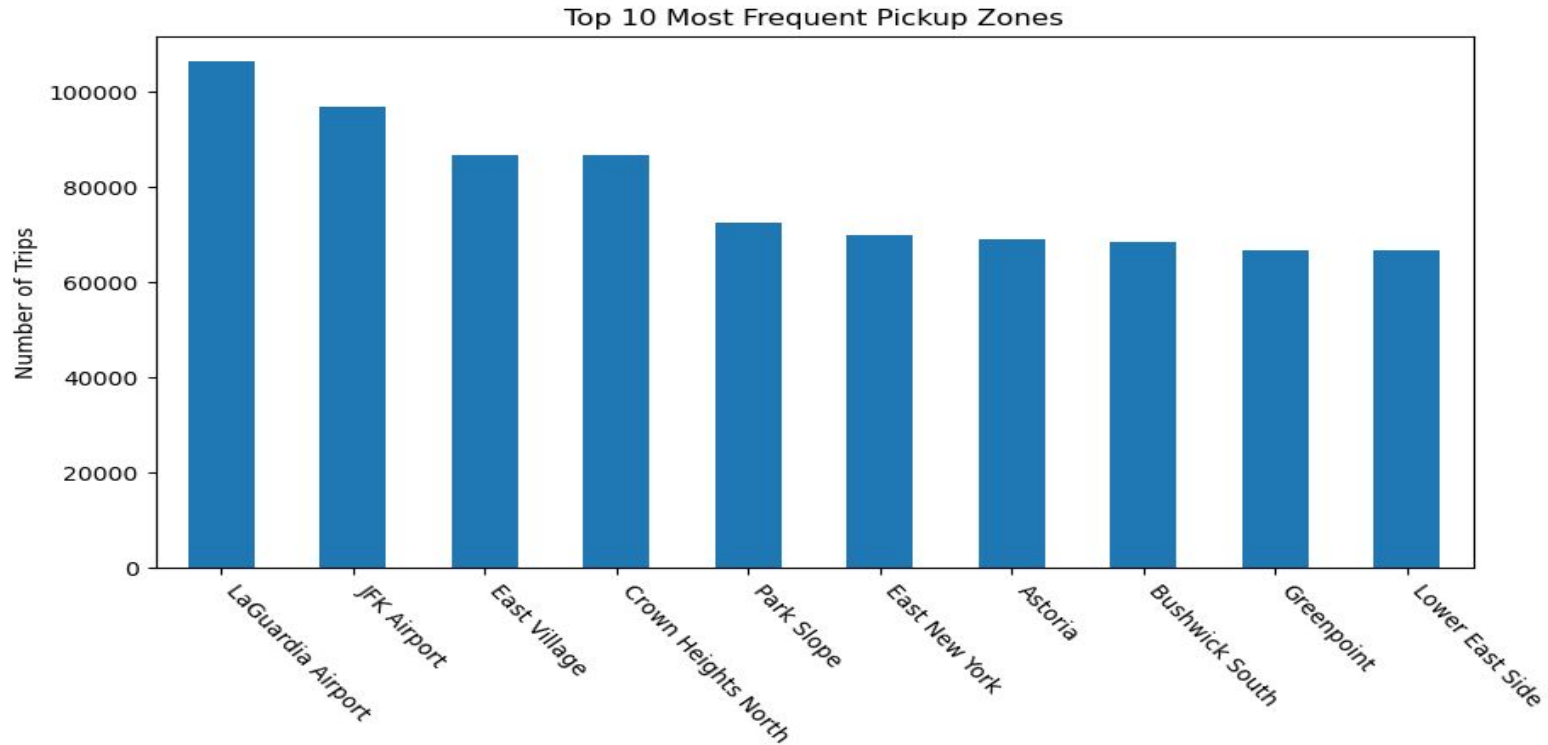
Phase 1: Examine All Data

NYC Lyfts - Dec. 2023



Key Takeaway: Average trip distance is dependent on holidays

Phase 1: Examine All Data



Key Takeaway: Airport Zones are the major driver for ride pick ups!

Phase 2:

Digging Deeper into Airports

Now that we have narrowed the data in Phase 1 from all of the rides across NYC, we see that LaGuardia and JFK Airport are the top 2 location ID's in total volume.

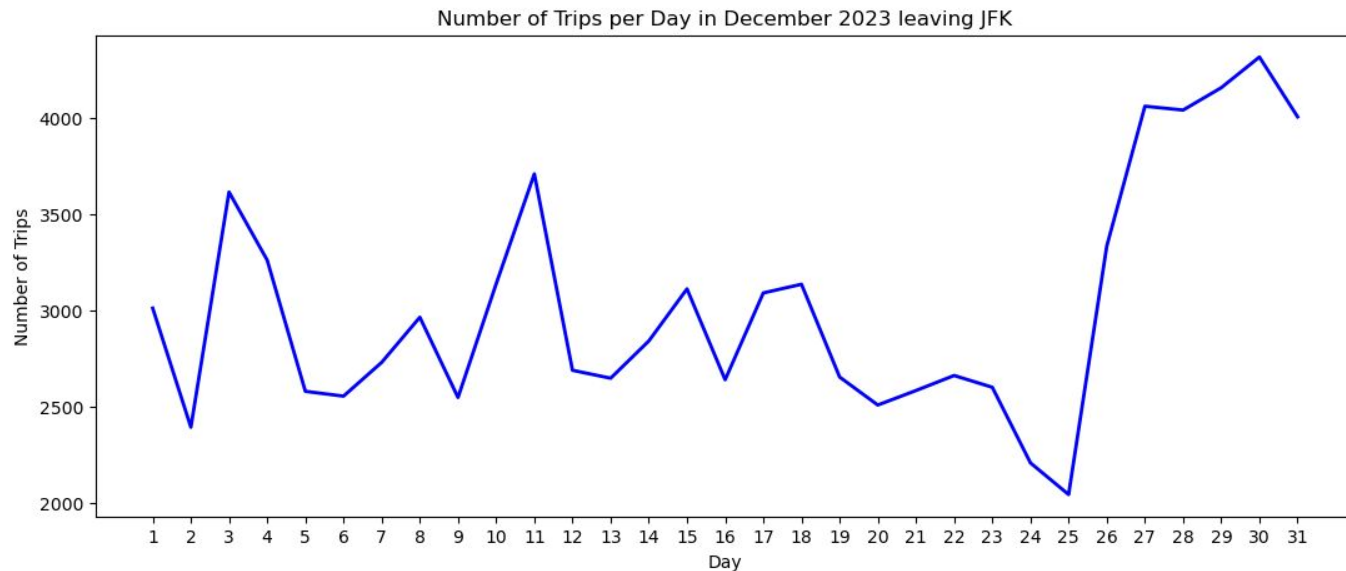
In this section of the project we will take a deeper dive into these two locations to see which one is more likely to be profitable for our feasibility study.

Phase 2: JFK Data



In Dec 2023, 93,847 rides were taken departing from JFK Airport.

Average of 3,000 per day!



Key Takeaway: Stays fairly consistent until the holiday season starts. Huge spike for NYE

Phase 2: JFK Data

Avg. Trip Time from JFK:

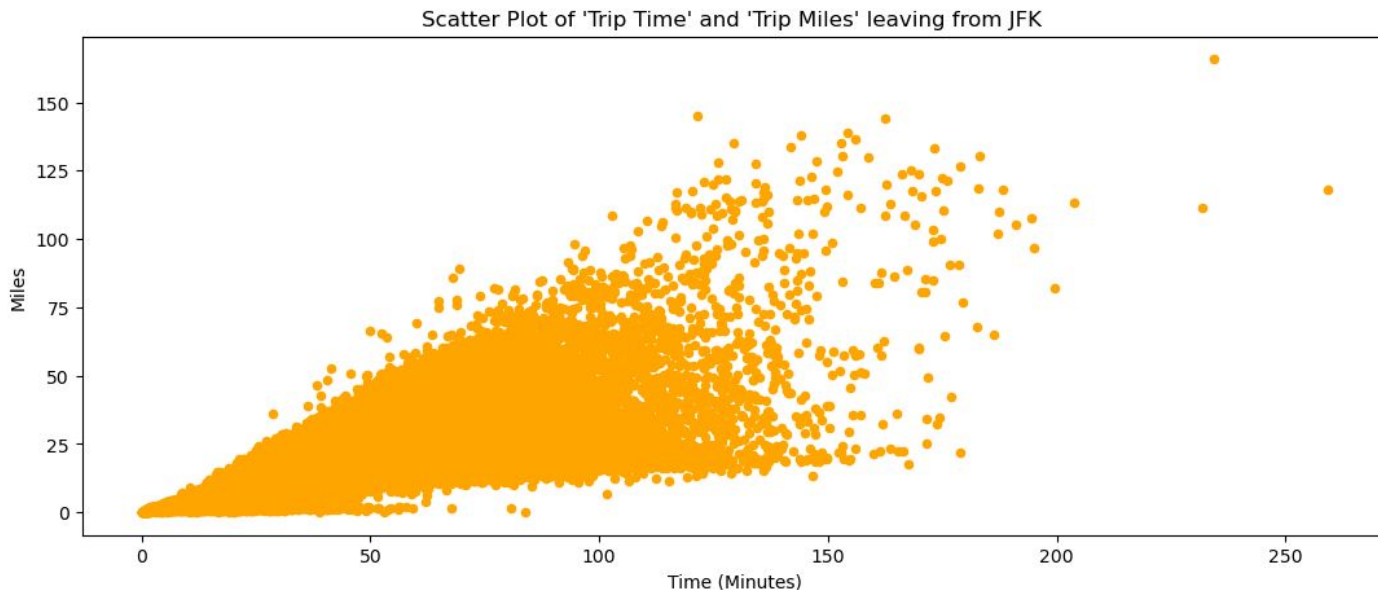
- **43.2 Minutes**

Avg Trip Miles:

- **17.30 Miles**

Avg Passenger Cost:

- **\$77.90**



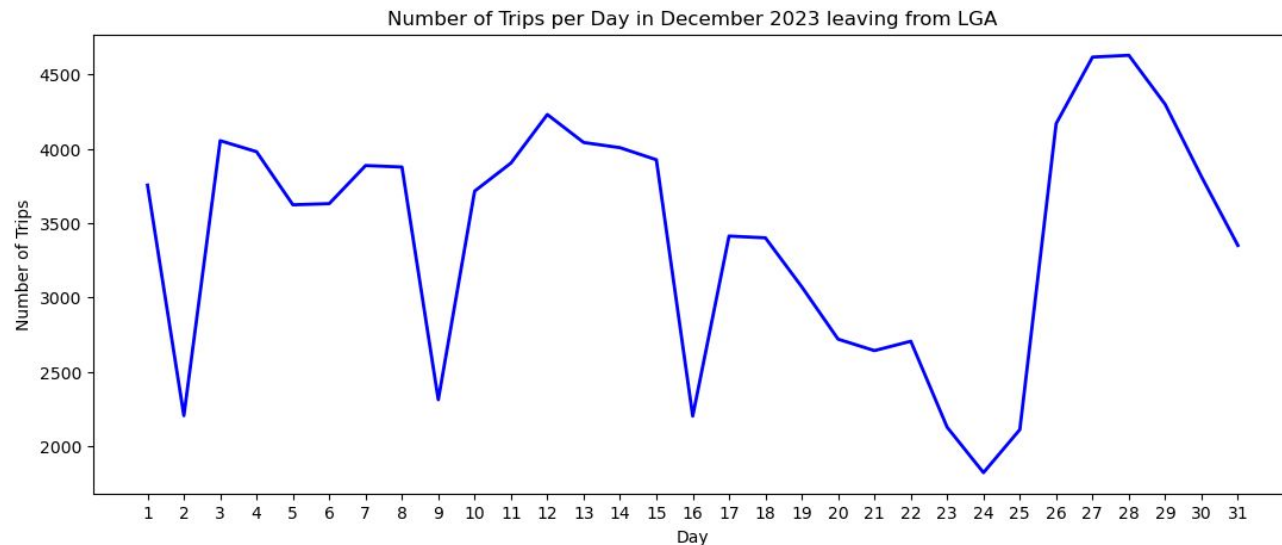
Key Takeaway: JFK has longer mile and longer duration rides on average.

Phase 2: LGA Data



In Dec 2023, 106,243 rides were taken departing from LaGuardia Airport (LGA).

Average of 3,400 per day!



Key Takeaway: Stays fairly consistent until the holiday season starts. Huge spike for NYE. Number of Trips drops on almost all Saturdays.

Phase 2: LGA Data

Avg. Trip Time from LGA:

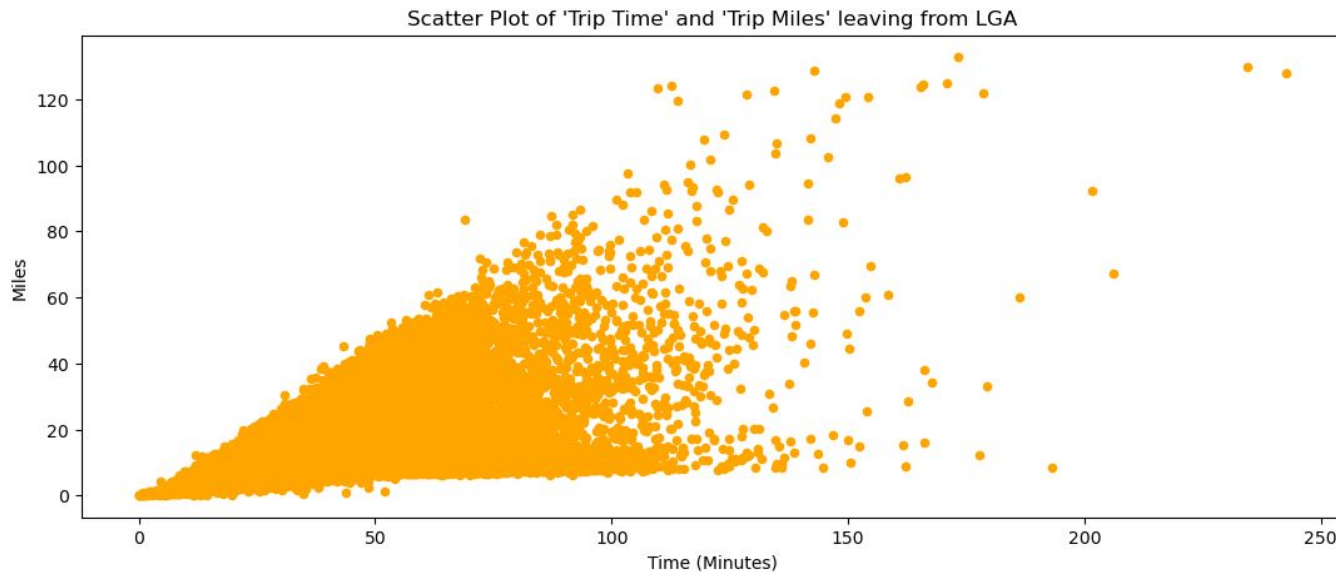
- **34.5 Minutes**

Avg Trip Miles:

- **11.11 Miles**

Avg Passenger Cost:

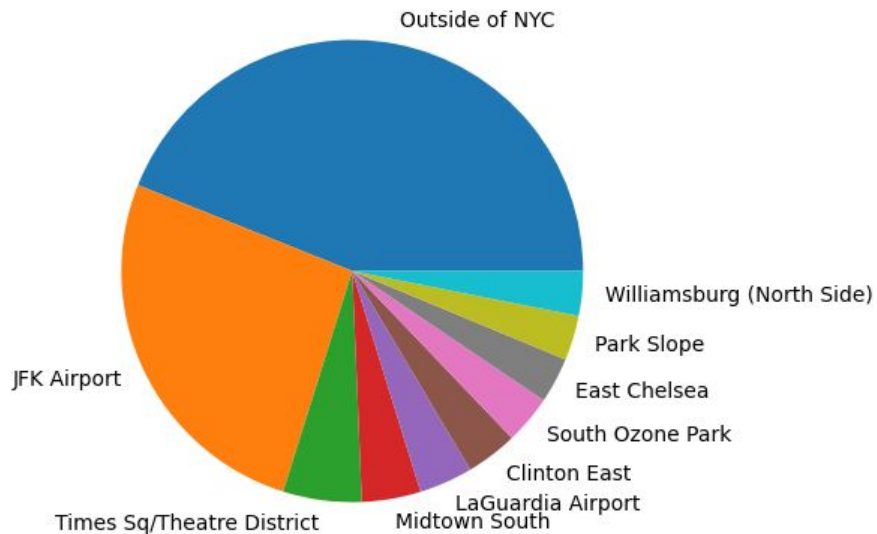
- **\$64.58**



Key Takeaway: LGA has shorter mile and shorter duration rides on average compared to JFK.

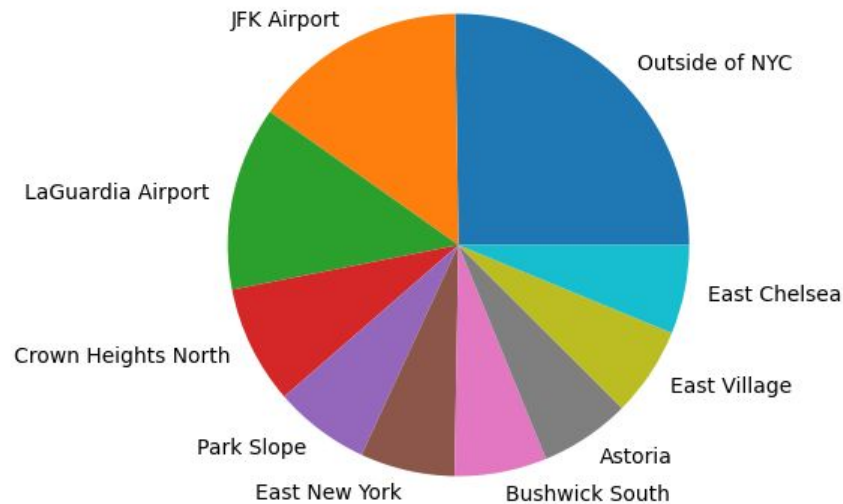
Phase 2: Airport Comparison

Top 10 Most Frequent Drop Off Zones from JFK



| | |
|---------------------|----------------|
| Avg. Rides | ~3,000 per day |
| Avg. Trip Time | 43.2 Minutes |
| Avg. Trip Miles | 17.3 Miles |
| Avg. Passenger Cost | \$77.90 |

Top 10 Most Frequent Drop Off Zones from LGA



| | |
|---------------------|----------------|
| Avg. Rides | ~3,400 per day |
| Avg. Trip Time | 34.5 Minutes |
| Avg. Trip Miles | 11.1 Miles |
| Avg. Passenger Cost | \$64.58 |



Phase 2 Summary

Both Airports are qualified locations for a taxicab hub.

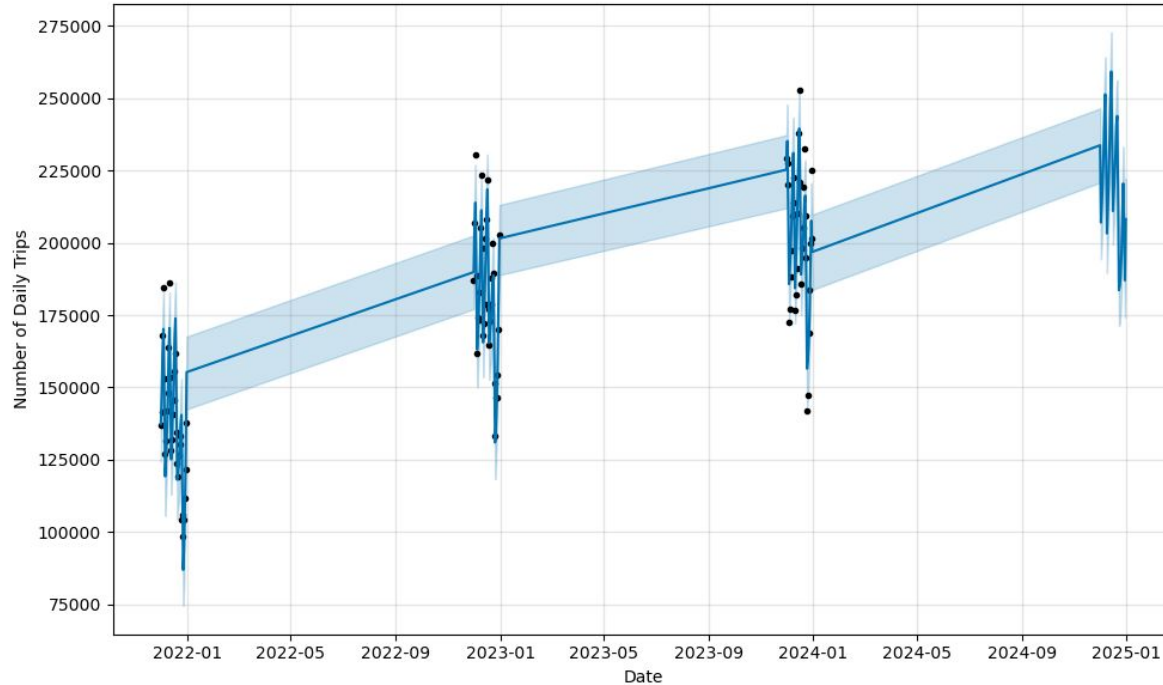
LaGuardia has higher volumes, but shorter trip duration and average fare cost, while JFK has longer rides, higher fares, but lower daily volume.

Because we can assume that there will be other players in this market, we think it is better to capture the longer, more profitable rides, and select JFK as our first hub location.

Phase 3: Predicting 2024 with Prophet

We now have a good handle on where we want to place our taxicab hub, but what does next year's ride volume look like from an overall NYC perspective

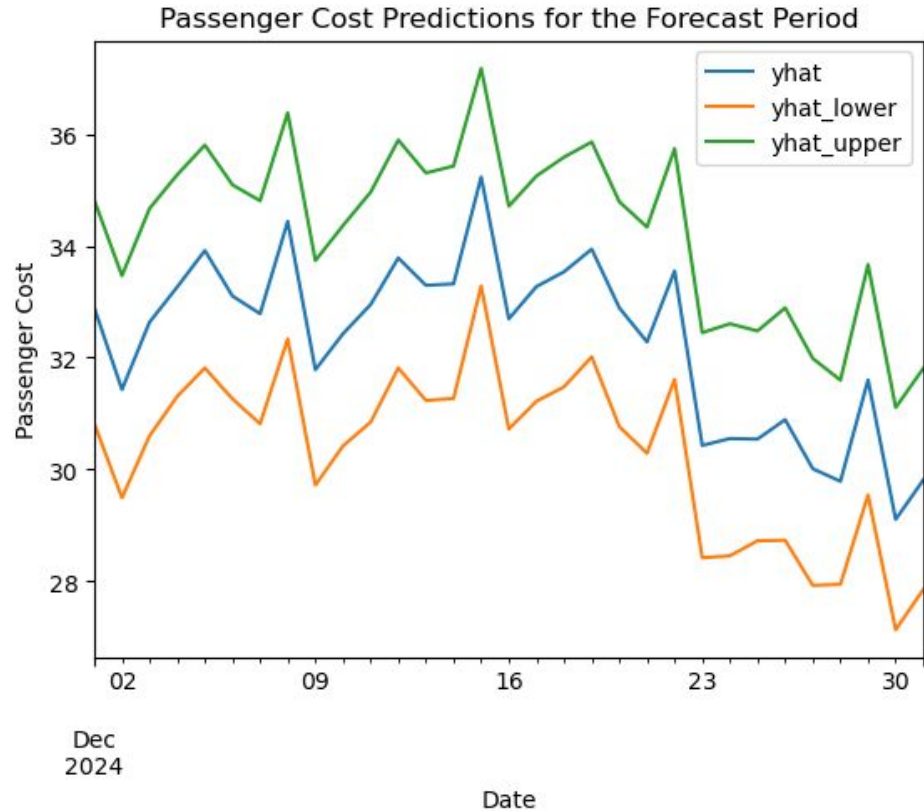
Phase 3: Predicting 2024 with Prophet



Key Takeaway: Overall number of trips predicted to increase in future forecast

Phase 3: Cost Prediction

- Passengers expected to pay between \$29 - \$35 per Lyft.
- Costs are lower the last week of December.





Conclusion

NYC is a great hub for any taxi business - 5 million+ rides per month!

Airport locations have the highest pickup/dropoff rate of any location ID in the dataset.

- Weekends, Later hour of the day also have high traffic.

JFK stands out slightly more favorable as a first location, both are strong opportunities.

According to our Prophet Analysis, trip volumes are expected to rise over time



If We Had More Time

- Uber Comparison
- Revenue Analysis
- Full Year Analysis

Thank you!

Any Questions?