

---

# Improving NYC Rideshare Prediction Accuracy

Prepared by

**Javier Lopez**

**Student ID: 000697446**

August 12, 2022

---

# 1. Questions

On March 11, 2022, the United States initiated large-scale stay-at-home orders causing most of the economy to shut down. The dramatic step marked the beginning of the Covid-19 pandemic for many businesses that began seeing dramatic shifts in demand, including rideshare providers. Through our research, which focused on New York City (NYC) rideshare providers Lyft and Uber, we sought to answer four essential questions and test our hypotheses.

1. Which rideshare company experienced the most growth in demand since the beginning of the Covid-19 pandemic?

**Hypothesis:** Given previously observed data that shows Uber saw more than double the demand Lyft did, we believe Uber will see the smallest decline in demand during the Covid-19 shutdown and the largest growth through each successive phase.

2. What is the relationship between exogenous variables, such as Covid-19 cases, hospitalizations, deaths, vaccinations, temperature, precipitation, and rideshare demand?

**Hypothesis:** Based on the impact the Covid-19 pandemic had on businesses, we believe the Covid-19 related variables will significantly impact rideshare demand more than the temperature and precipitation.

3. Which predictive algorithm and exogenous variables achieve the best forecast in rideshare demand?

**Hypothesis:** Given SARIMAX's ability to forecast time series data with seasonality, we believe that the Covid-19 cases modeled with the SARIMAX algorithm will achieve the lowest root mean square error (rmse) and mean absolute percentage error (mape), yielding the most accurate forecast.

4. What is the forecasted growth for Lyft and Uber through the second quarter of 2022?

**Hypothesis:** Based on our previous belief regarding growth in demand throughout the Covid-19 pandemic, we believe that Uber will be forecasted to experience more growth than Lyft through the second quarter of 2022.

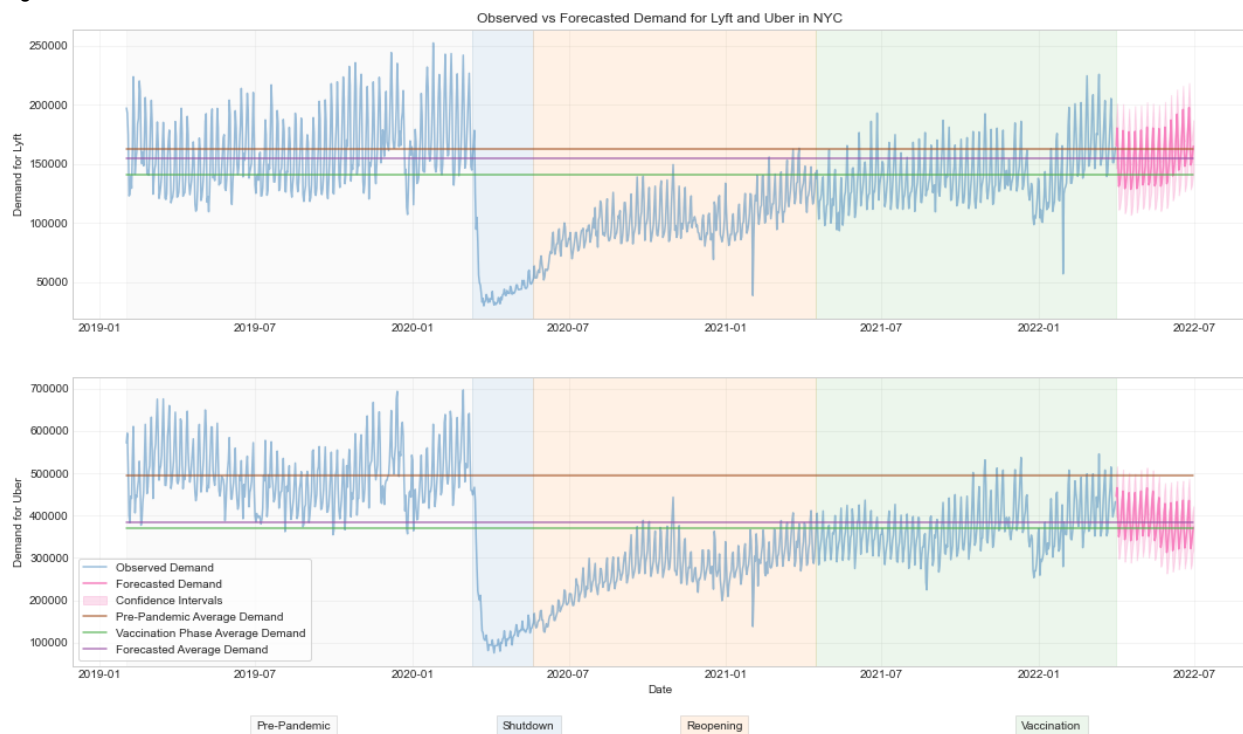
# 2. Methods

We began processing the data by aggregating all trip records into daily trip counts and splitting the time series into four key phases: pre-pandemic (Feb. 1, 2019 - Mar. 10, 2020),

shutdown (Mar. 11, 2020 - May 19, 2020), reopening (May 20, 2020 - Apr. 15, 2021), and vaccination (Apr. 16, 2021 - Mar. 31, 2022), as shown in Figure 1. Then, we began analyzing the data using descriptive statistics by calculating and plotting each variable's five-number summary statistics. We then tested the strength of the relationship between each exogenous variable and the rideshare providers by calculating the regression line and Pearson Correlation Coefficient R-value and p-value. Once we understood the trends for Lyft and Uber during each temporal phase, we built and tested six predictive models for each algorithm (SARIMAX, Prophet, and XGBoost). We trained our models with data from Feb. 1, 2019, through Apr. 30, 2021, and performed cross-validation tests from May 1, 2021, through Mar. 31, 2022.

We quantified each model's performance by logging each model's rmse and mape. These metrics measured the size and amount of errors made during cross-validation. We then analyzed the results for each model by exploring and plotting the Pandas data frame into two triple-bar charts depicting the rmse and mape. These steps helped us determine which model would best forecast Lyft and Uber demand into the second quarter of 2022. Once we forecasted demand for Lyft and Uber using each provider's best-performing model, we quantified the growth by calculating the growth percentage and average daily rides during each temporal phase to complete our research.

Figure 1. Rideshare Demand in NYC



### 3. Findings

Table 1 presents the results of the Pearson Correlation Coefficient calculations. Precipitation was the only positively correlated variable with rideshare demand, with an R-value of 0.02, but very little confidence as the sample data showed very little precipitation. The strongest relationship was between Covid-19 hospitalizations and deaths. Both variables negatively correlate with rideshare demand with an R-value of about -0.50. However, when testing the predictive models, we found that vaccinations were the best predictor for Lyft, and the temperature was the best predictor for Uber demand.

Table 1. Pearson Correlation Coefficient and P-value

		Weather, Temperature	Weather, Precipitation	Covid-19, Cases	Covid-19, Hospitalizations	Covid-19, Deaths	Covid-19, Vaccinations
Rideshare Overall	r-value	-0.17	0.02	-0.20	-0.50	-0.51	-0.13
	p-value	0.00	0.55	0.00	0.00	0.00	0.01

Table 2 presents the observed and forecasted growth percentage and average daily demand for Lyft and Uber from Feb. 1, 2019, to Jul. 31, 2022. We found that during the second half of the pre-pandemic phase, Lyft saw a 10.29% increase in demand from the first half, while Uber stagnated at just 0.17% growth. During the shutdown phase, both rideshare providers saw similar decreases in demand, averaging about a 70% loss in ride volume. During the reopening phase, Lyft more than doubled its demand, seeing a 102.26% increase. Demand for Uber rose by just 86.61%. The growth steadied during the vaccination phase, with both providers achieving an average of 37% increase in demand.

Table 2. Growth Percentage and Average Daily Rides

		Phase 0: Pre-Pandemic	Phase 1: Shutdown	Phase 2: Reopening	Phase 3: Vaccination	Q2, 2022 Forecast
Lyft	Growth Percentage	10.29%	-69.06%	102.26%	38.44%	10.23%
	Average Daily Demand	162,243	50,202	101,540	140,570	154,949
	+/- Demand	+15,908	-112,041	+51,338	+39,030	+14,379
Uber	Growth Percentage	0.17%	-70.68%	86.61%	36.40%	3.86%
	Average Daily Demand	495,462	145,249	271,048	369,715	383,990
	+/- Demand	+825	-350,213	+125,799	+98,667	+14,275

Our predictive models forecasted that Lyft would be expecting a 10.23% increase in demand and reaching an average of 154,949 rides, while Uber was forecasted to see a 3.86% increase in demand and an average of 383,990 rides. Given the difference in ride volumes between Lyft and Uber, this growth represents an average of about 14,327 more daily rides. While the model forecasted Lyft's growth rate to be higher than Uber's, it is the lowest growth rate they've had, not including the shutdown phase. Uber, however, is still expected to continue a growth rate that remains higher than the pre-pandemic phase.

## 4. Limitations

When testing the relationship between the exogenous variables and rideshare demand, we found that the smaller sample size of precipitation and Covid-19 vaccinations led to less reliable R-values. To solve this, we included the temporal phases as part of the features when training the predictive models. The largest limitation, however, was the availability of more specific vaccination data. The center for disease control (CDC) only provides vaccination data by state on their portal. The broader population in the vaccination data sample then adds the risk of introducing biases into the predictive models as there is the possibility of a higher variability from the vaccination data used to train the models and the actual vaccination rates of NYC alone.

## 5. Proposed Actions

Based on our analysis of the observed data and the forecasted growth for Lyft and Uber, we recommend two possible courses of action for each rideshare provider. In the case of Lyft, we observed a growth in demand through each pandemic phase that was greater than that of Uber. However, Lyft is forecasted to have its lowest growth period since the pre-pandemic phase. Our recommendations can be applied to either rideshare company, as they are forecasted to either have low or less-than-optimal growth over the second quarter of 2022. To turn this trend around, we recommend that Lyft and Uber:

1. Increase their marketing spending during the warmer months to ensure they stay top of mind when it matters most when consumers look for the rideshare app on their phones.
2. Explore economic variables that may impact rideshare demand and assist predictive models in higher quality forecasts for future quarters.

The first proposed action will directly impact rideshare demand, while our second proposed action will help improve forecasted demand. It is important to note that increased marketing spending can significantly impact periods of low demand. However, we recommend that Lyft and Uber consult with their marketing departments to create an action plan to optimize the best timing to invest the additional marketing spending. An increase during periods of high demand may render a bigger impact and offset the lower growth rates.

## **6. Expected Benefits**

We expect that our research can most benefit small businesses looking to optimize their forecasting models as we move further from the start of the Covid-19 pandemic. By introducing exogenous variables in our models, the two best-performing models for each predictive algorithm saw an average decrease in the rmse of 7.33% and mape of 9.09%. Based on these metrics, the increased accuracy of the predictive models that use exogenous variables from those that rely only on historic data is something to consider when deciding the best method to forecast. The forecast improvements of introducing exogenous variables can mean a small business may be able to meet customer demand and avoid incurring losses from sold-out items or not enough wait staff, to name a few.