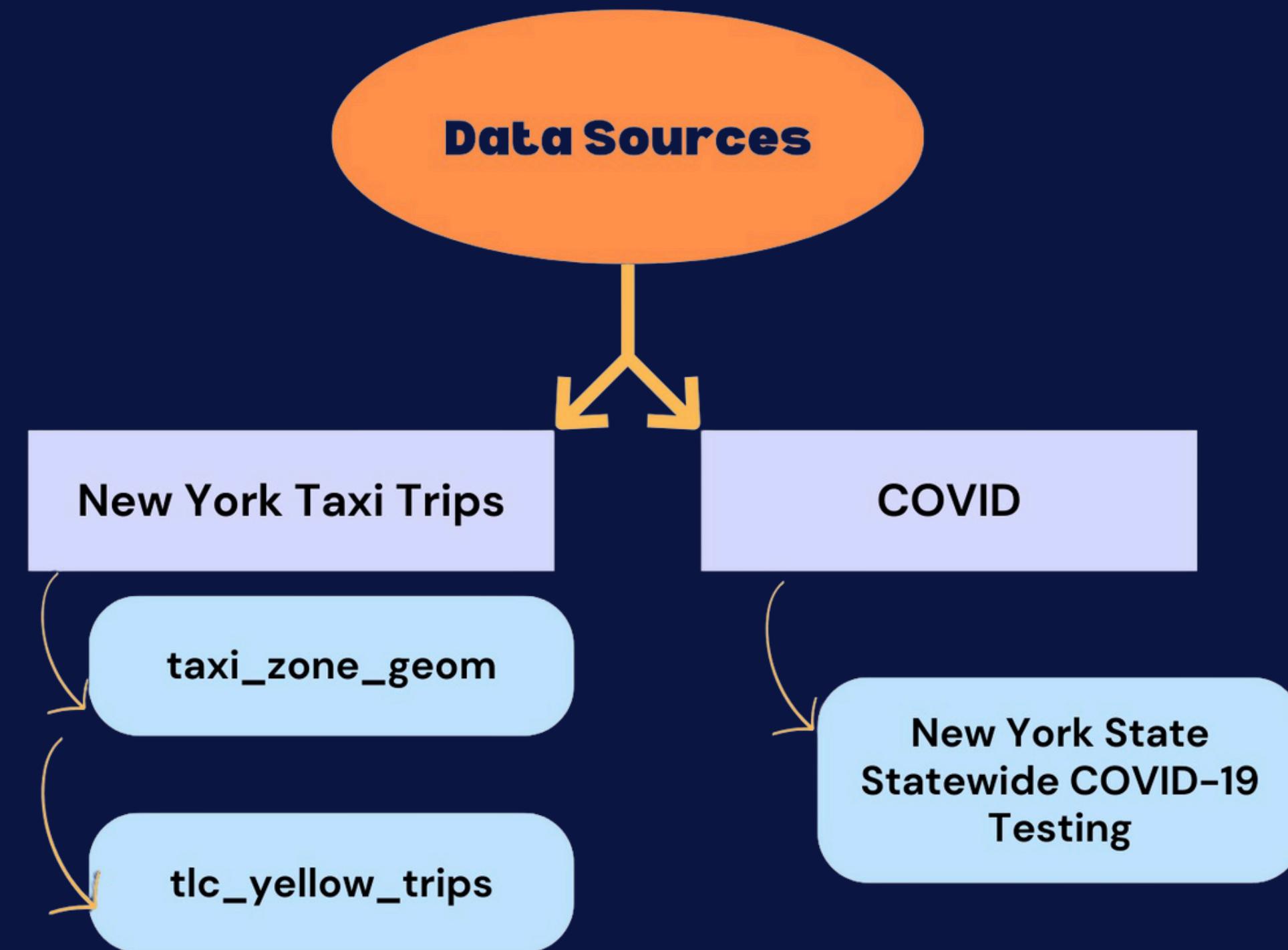


UNLOCKING INSIGHTS: NYC TAXI TRENDS AND THE IMPACT OF COVID

Group 4: Aryan Kumar, Jenn Hong, Olimpia Borgohain,
Saumya Anand, Setu Shah



DATA OVERVIEW

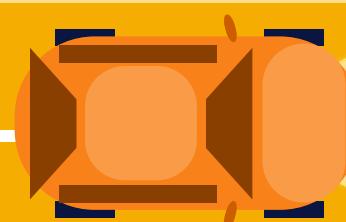
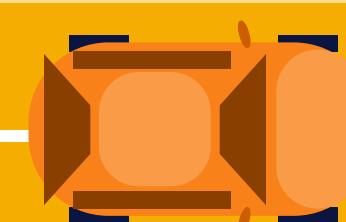
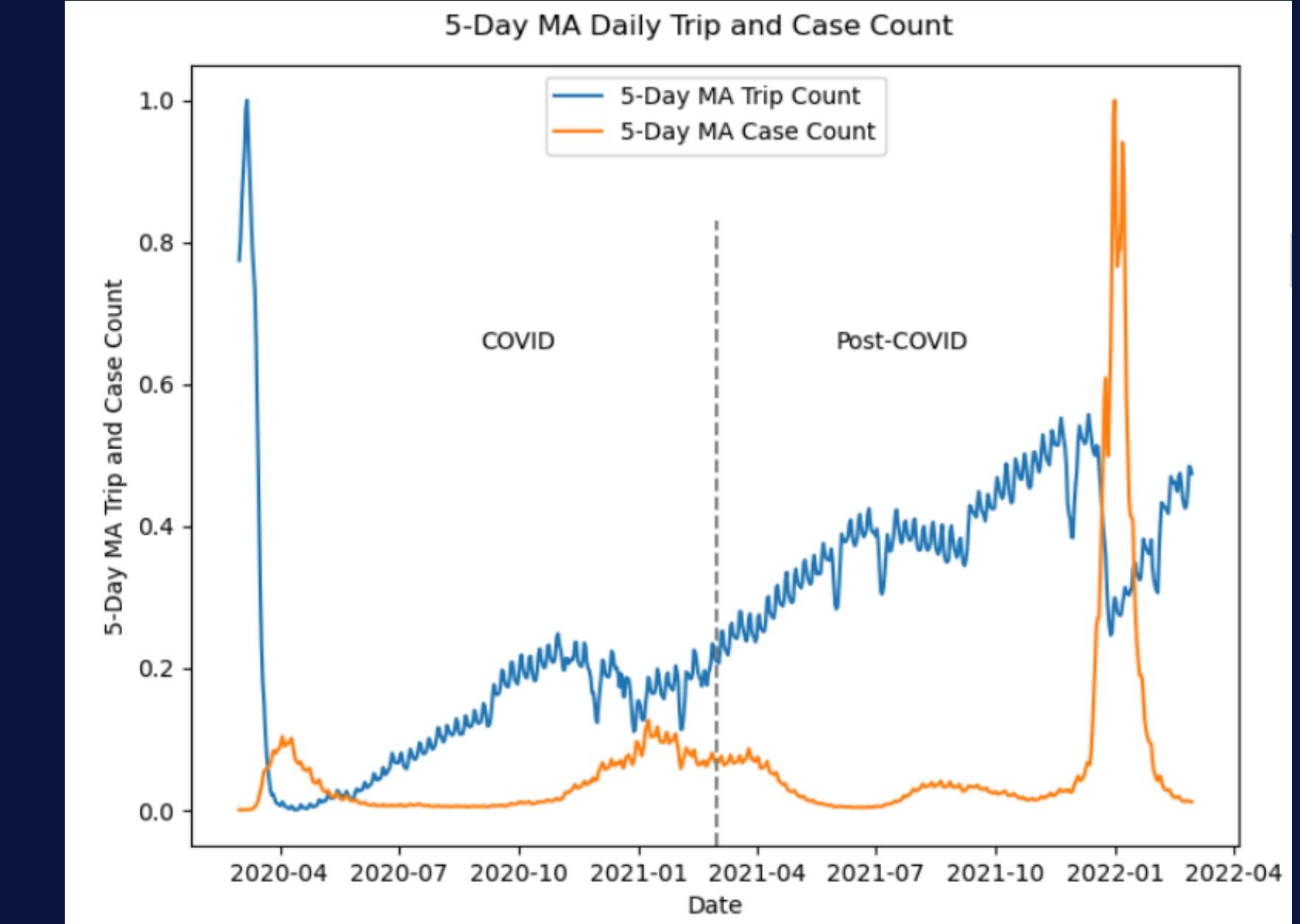
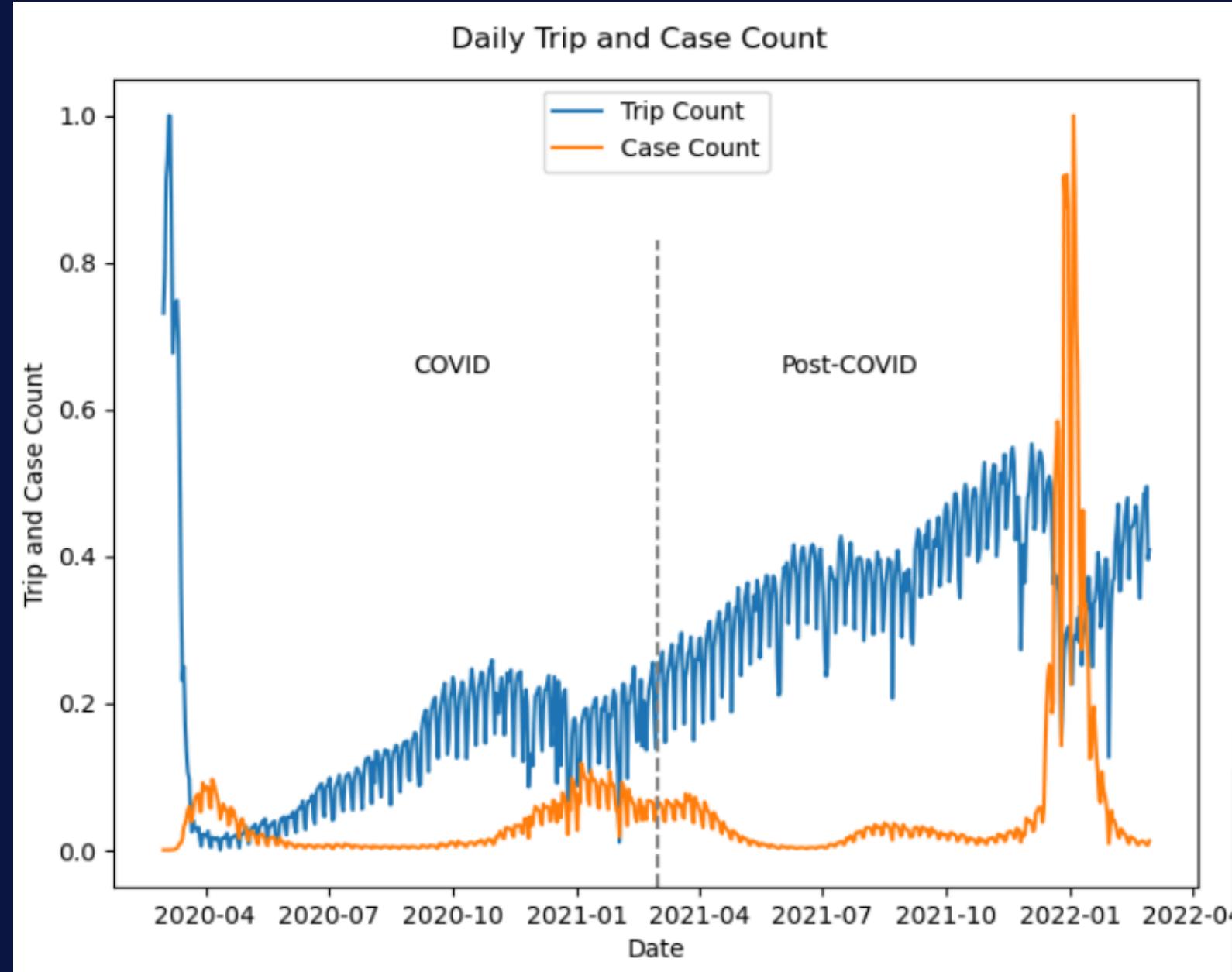


DATA CLEANING

- Data from 2019, 2020, 2021, and 2022 were chosen
- Organized into our time frames for further analysis:
 - March 2019 - March 2020: Pre-COVID
 - March 2020 - March 2021: Active-COVID
 - March 2021 - March 2022: “Post”-COVID
- Aggregated “daily_df” to improve computation efficiency

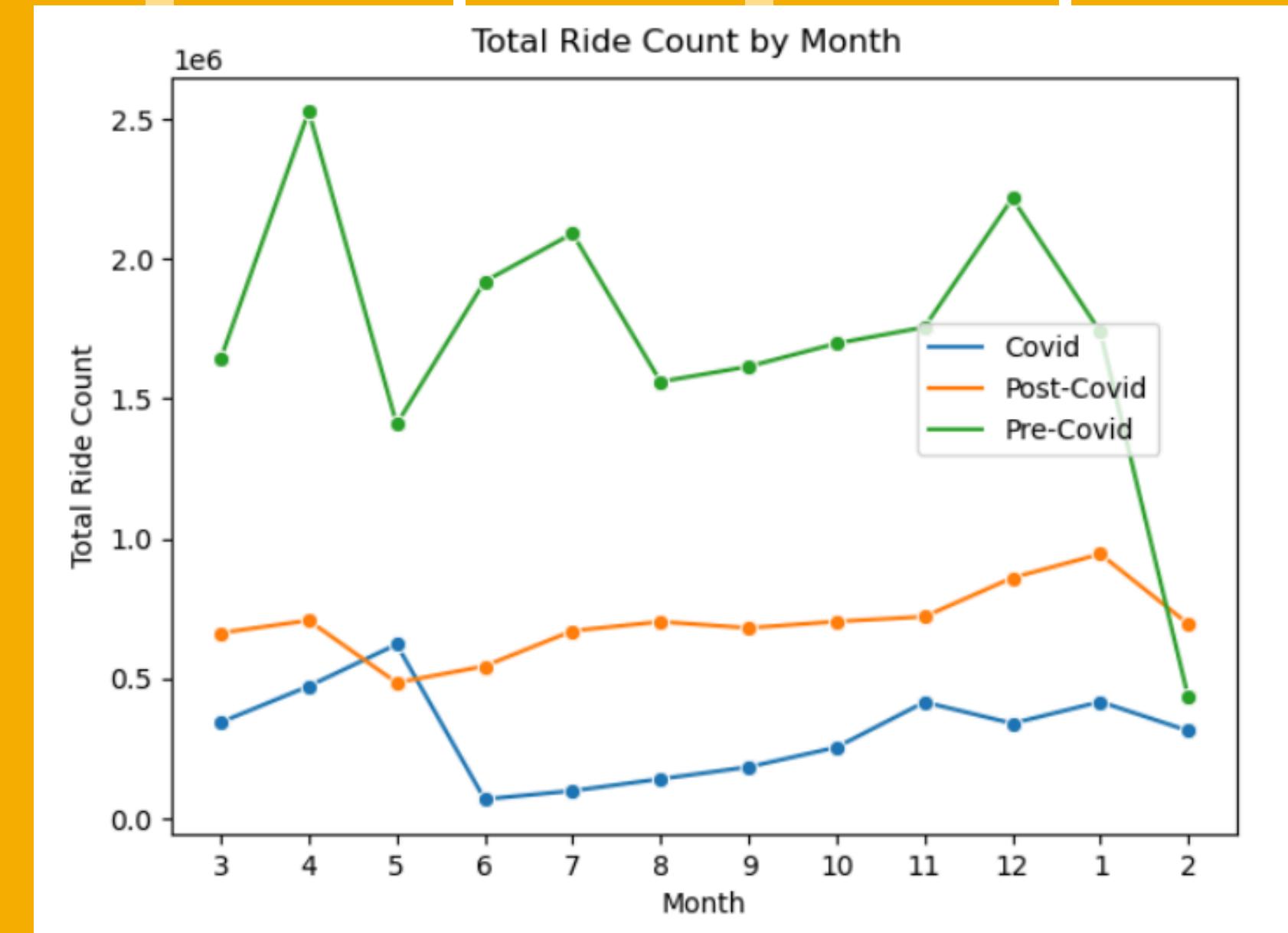


MARKETPLACE OVERVIEW



TOTAL RIDE COUNT

- Pre-COVID months are loosely linked to seasonality with warmer months, June and July, seeing more rides.
- COVID ride counts see peak in May, aligning with reopening timelines for NYC but there's a sharp drop in June.
- Post-COVID ride counts tend to trend generally upwards, suggesting a slow but steady recovery in taxi ridership.

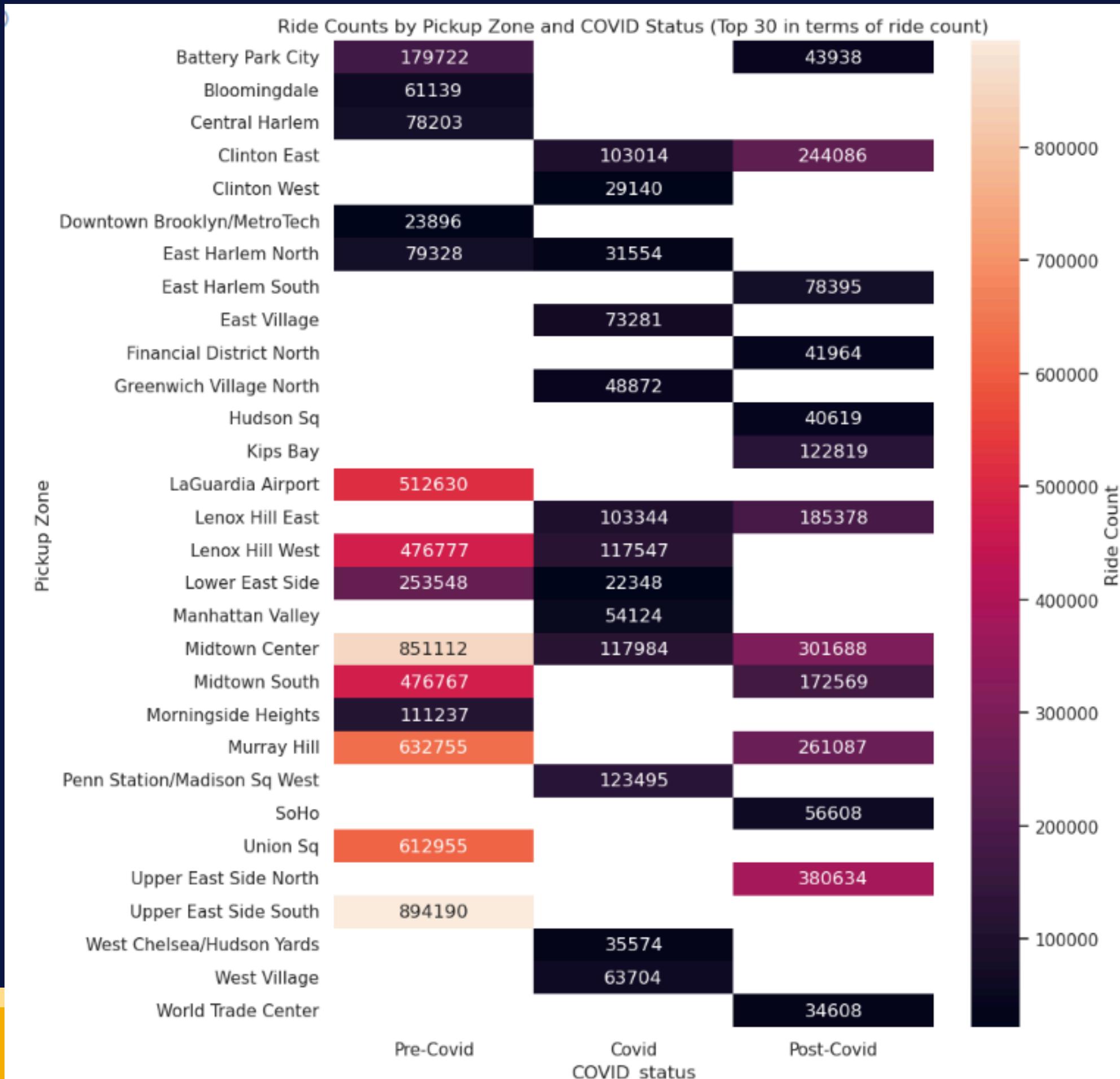


POPULAR PICKUP DAYS FOR EACH PERIOD

- **Pre-Covid Period:** Fridays had the highest number of pickups, followed by Thursdays and Wednesdays.
- **Covid Period:** Wednesdays had the most popular pickup day, closely followed by Thursdays and Fridays.
- **Post-Covid Period:** Fridays once again stood out as the peak pickup day, with Thursdays and Wednesdays following closely behind.

COVID Status	Pickup Day	Trips per day
Pre COVID	Friday	3217273
Pre COVID	Thursday	3181737
Pre COVID	Wednesday	3074036
Pre COVID	Saturday	2972683
Pre COVID	Tuesday	2966802
COVID	Wednesday	584554
COVID	Thursday	574949
COVID	Friday	572164
COVID	Tuesday	562776
COVID	Monday	506887
Post COVID	Friday	1299402
Post COVID	Thursday	1285849
Post COVID	Wednesday	1260817
Post COVID	Saturday	1221611
Post COVID	Tuesday	1206102

POPULAR ZONES AND COVID PHASES



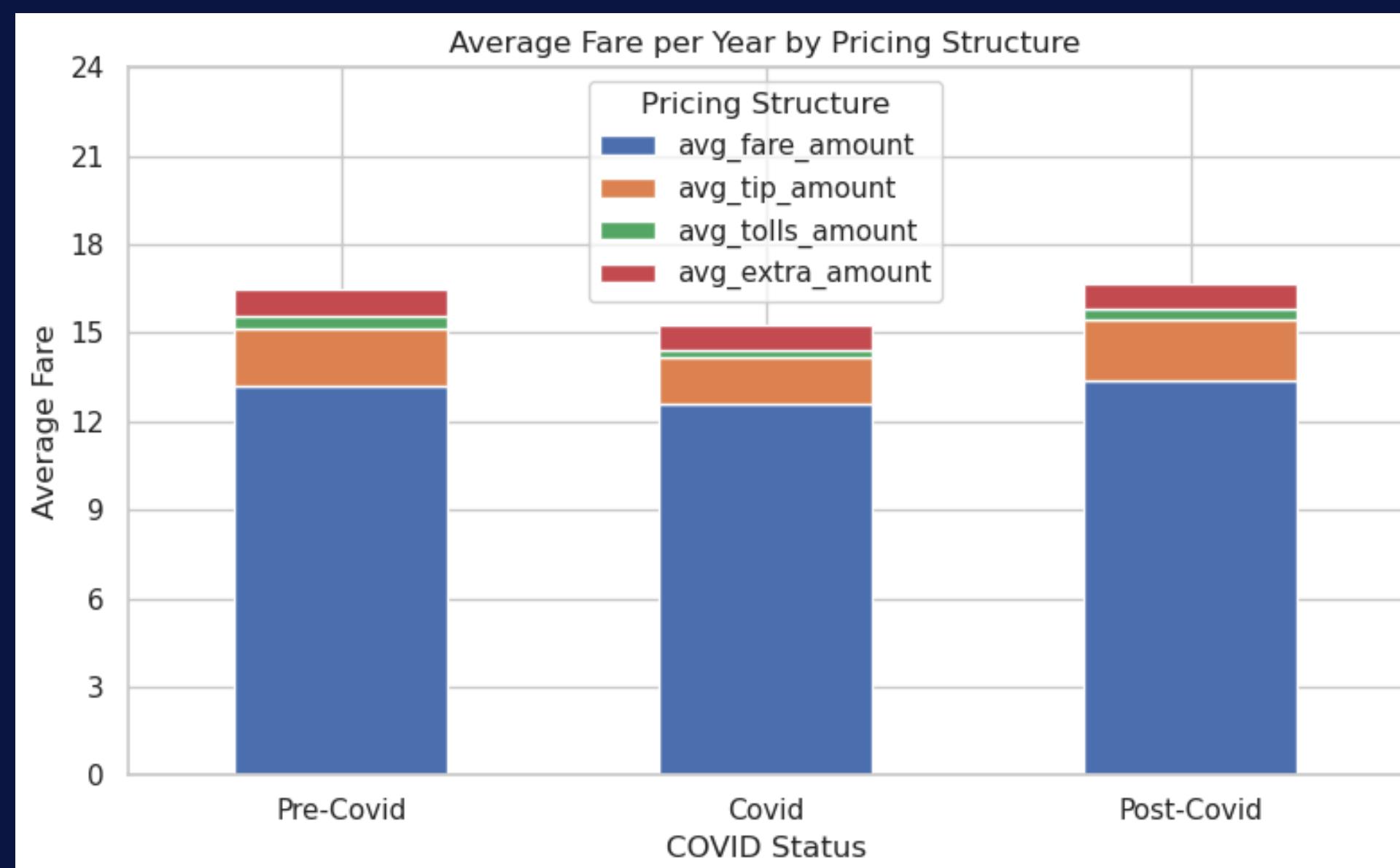
MONEY MATTERS

Avg Total Amount	→	Pre-Covid	19.52
		Covid	18.14
		Post-Covid	19.88

- Maybe due to change in pricing strategies during reduced demand
- Recovering and past initial fares!

- Breakdown follows similar pattern throughout the phases
- Average fare amount remained the highest

Pricing Structure



SO WHAT IS AFFECTING THE CHANGE IN PRICES?

We analyzed the patterns in average distance, ancillary fees and trip duration

Avg Trip Duration (mins)	Pre-Covid	17.97
	Covid	14.83
	Post-Covid	16.54

- This closely follows the pattern in change in price, indicating its effect on it.
- For NYC Taxi rides, after a base fare, the meter runs based on either time or distance travelled.

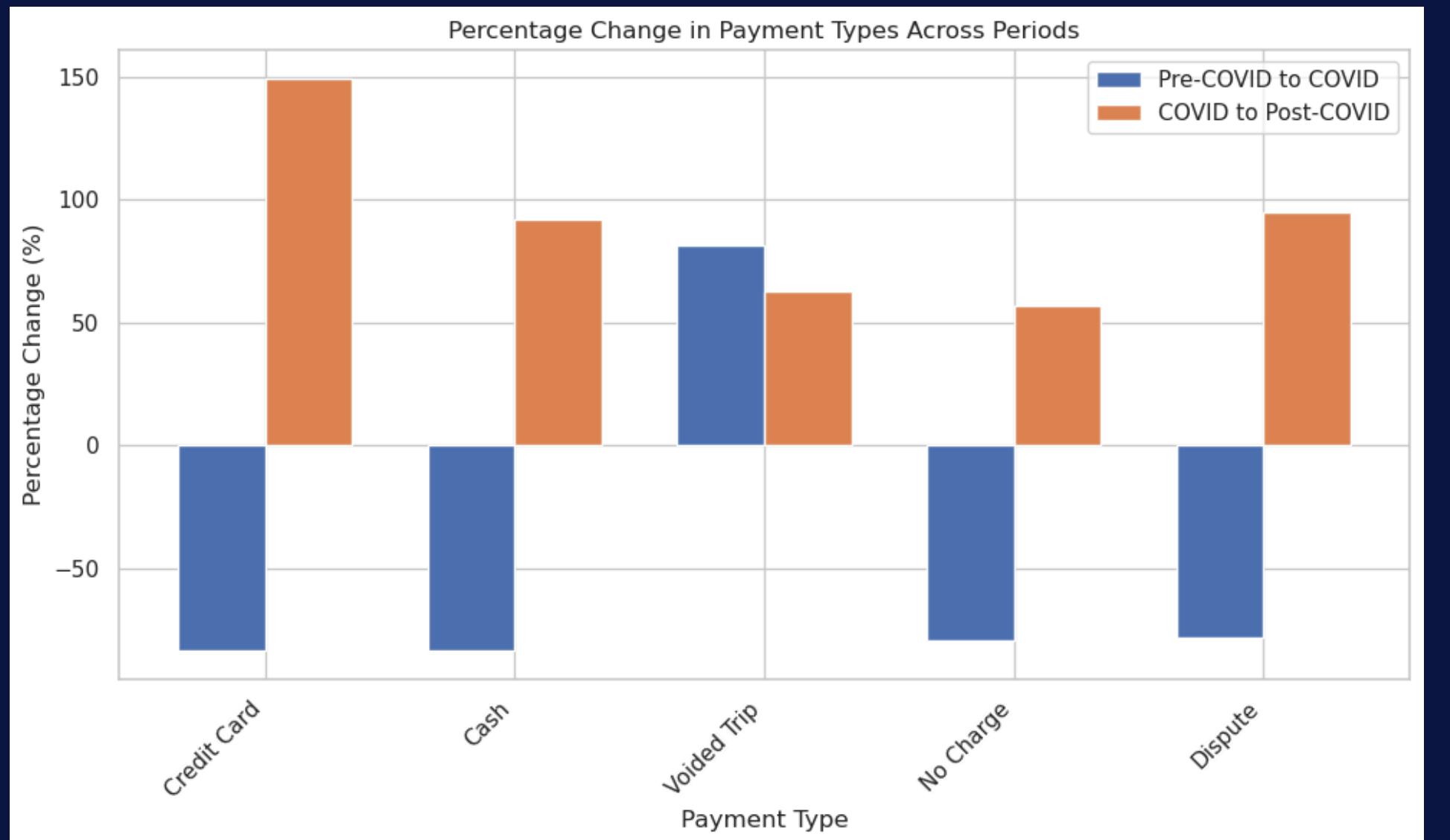
- People were travelling for longer distances during and post covid
- People moved outside cities and don't have to come to office everyday

Avg Distance	Pre-Covid	2.57
	Covid	3.93
	Post-Covid	6.38



PAYMENT TYPE BREAKDOWN ACROSS THE PHASES

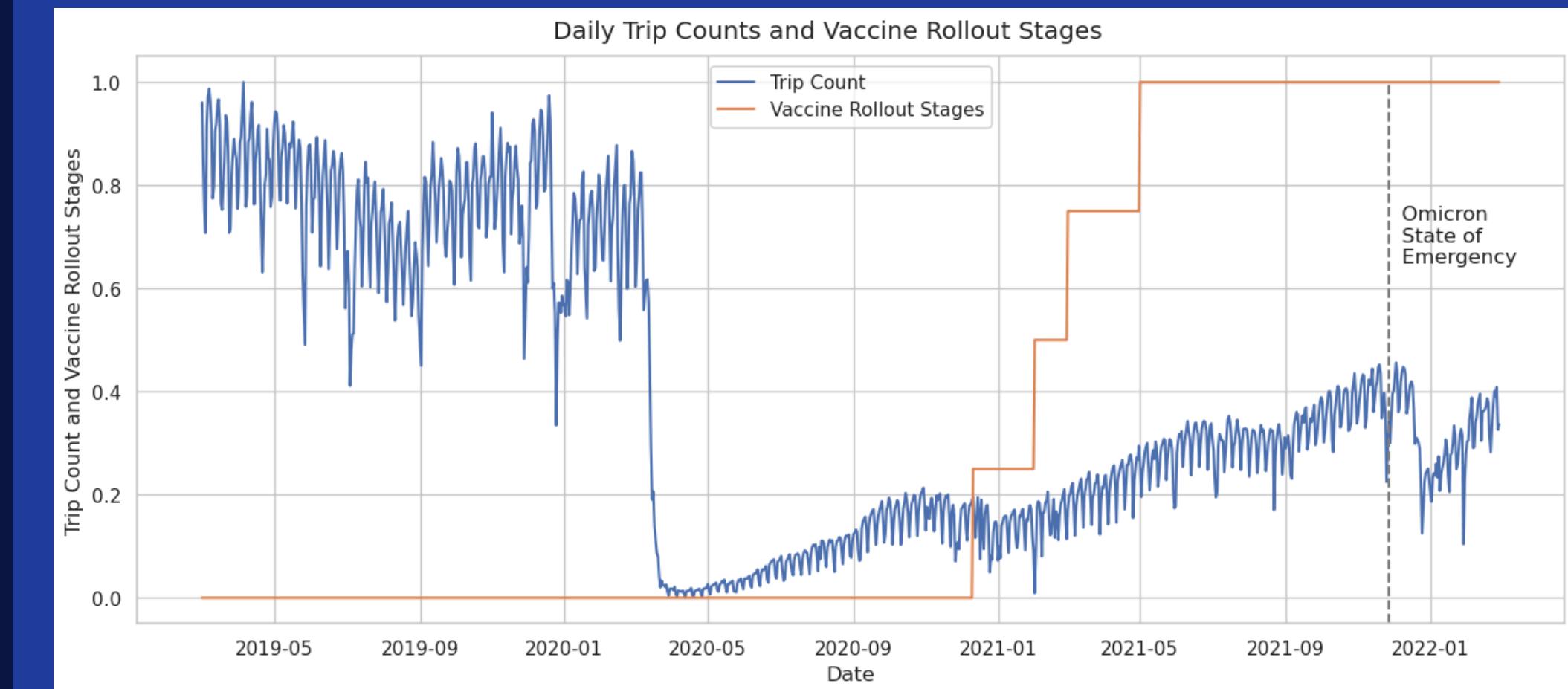
Number of transactions using Credit Card and Cash dominated. Some other types were voided trips, no charge and disputed transactions.



- Most of the payment types saw a jump from Covid to Post-Covid
- Credit cards saw highest increase. People preferred digital payments after covid

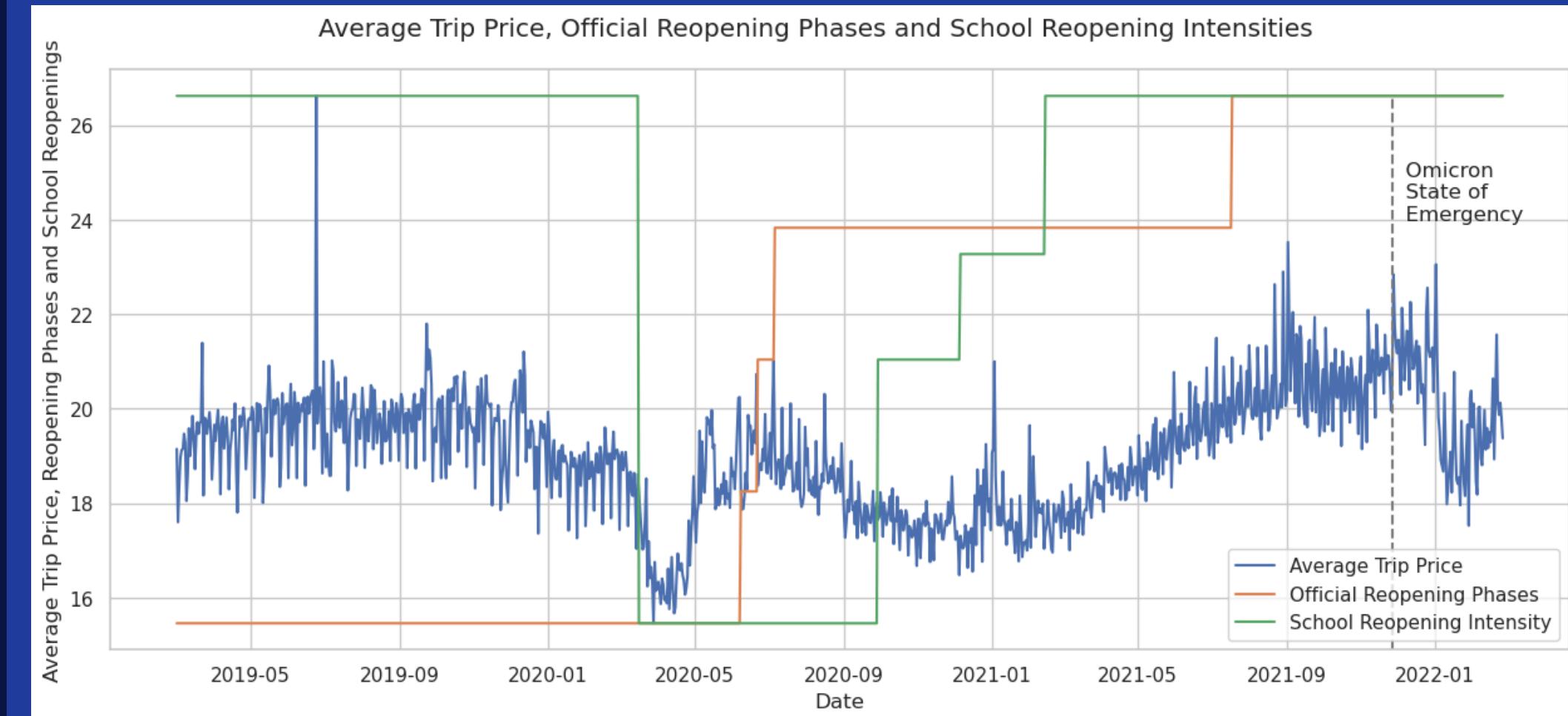
IMPACT OF VACCINE ROLLOUTS

- After the first phase of vaccine rollouts starting in November 2020, ride counts remain stable.
- However, with subsequent phases, ride counts trend upwards showing sustained recovery of the industry as things returned back to “normal.”
- We see a drop-off again when the Omicron wave hits in late 2021.



IMPACT ON AVERAGE FARES

- We observe an immediate ~15% dip in fares as the pandemic hits. However, they rebound rapidly to pre-pandemic levels as lockdown ends.
- Once more restrictions are lifted, we can see a downward trend in fares. This is because of less traffic congestion!
- From mid-2021 as things get back to normal, we see an upward trend as fare prices were hiked due to inflation.



PREDICTING DAILY DEMAND USING ML

Random
Forest →

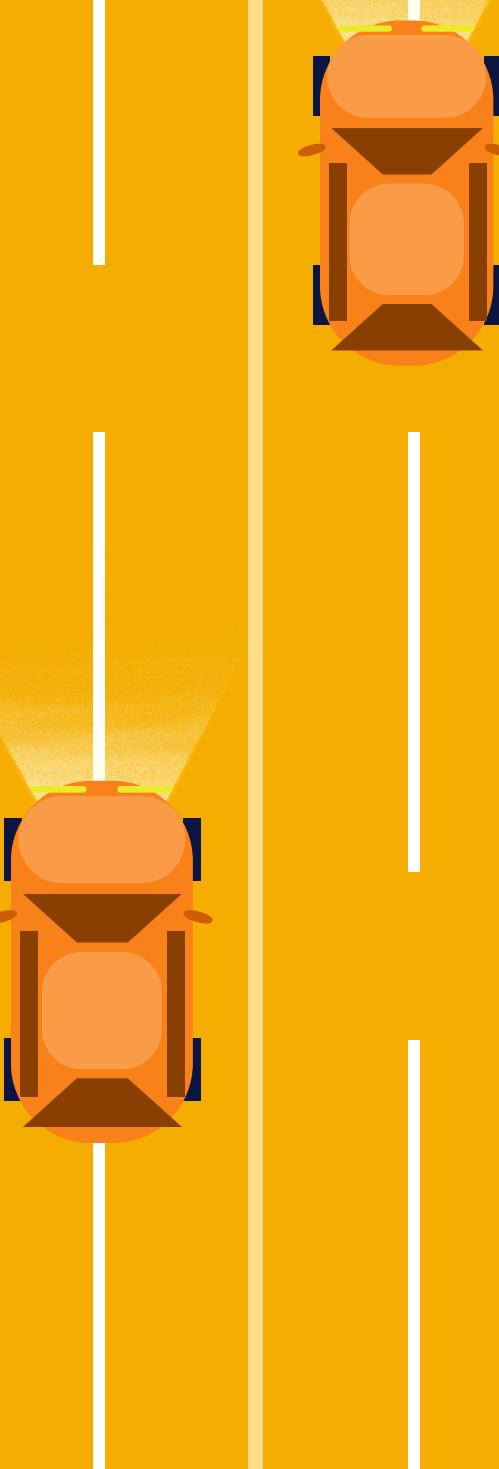
Baseline Model RMSE	3583
Fine Tuned Model RMSE	2905

- **Feature Importance**

- Passenger Counts
- Total New Positive Cases
- Average Total Cases per 100k

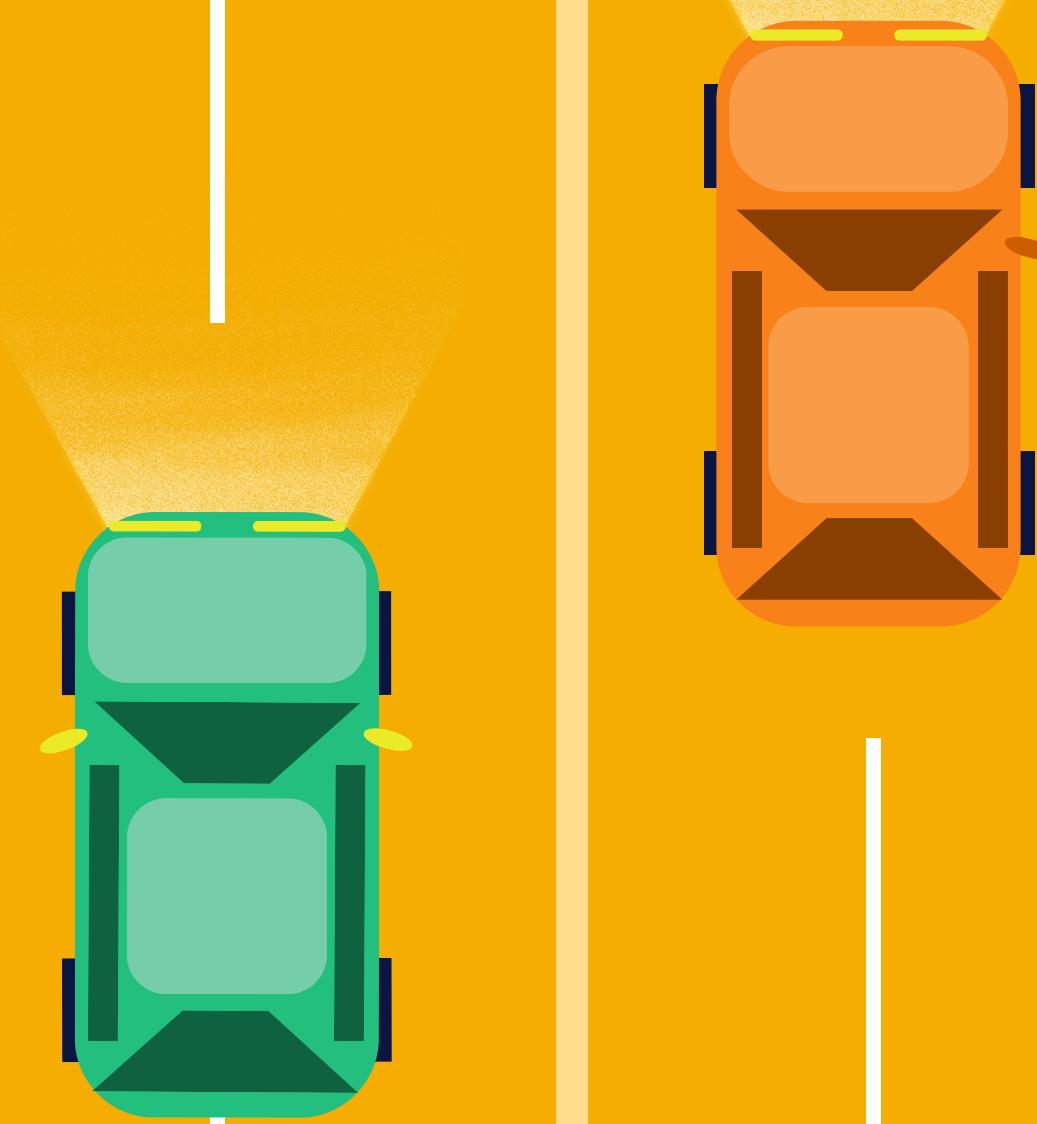
- **Limitations**

- 25% sample
- Memory limitations
- Backtesting



CONCLUSION

- **Mobility Patterns:** Average trip distance increased during COVID but increased triple post-COVID
- Fares **increased** post covid
- Average duration of trips **decreased** post covid
- **Payment Preferences:** Credit card and cash transactions remained popular throughout
- **Key Hub Analysis:** The Upper East Side remained a key hub for taxi rides
- **Commuting Habits:** Pickup hour distributions changed



CHALLENGES / LIMITATIONS



1. Size of the Dataset
2. Yellow Taxi Data
3. JSON Parsing Error
4. Monthly Averages not normalized

THANK YOU!

