NYU

NYC TAXI FARE ANALYSIS 2023

PRESENTED BY:

Jiwoo Jeong (jj4252)
Omni jyothi Gudiguntla (og2148)
Alexander Pegot-Ogier (ap9283)
Xintong Li (xl5733)

Research Questions

Primary Question

What are some factors influencing NYC taxi fares, and how do factors such as time, passenger count, weather conditions, pickup locations contribute to fare variability?

Sub Questions

1. Can we identify temporal patterns in taxi fares based on the time of the day, day of the week, and month of the year throughout 2023?
2. How does having multiple passengers impact the total fare? Additionally, if relevant, what other factors influence passenger count?
3. How are weather conditions associated with ride frequency and fares?
4. Does the pickup location influence the total fare? If so, which pickup locations tend to generate the highest fares?
5. Can we determine the most impactful variables that are associated with ride frequency and fares?

Potential Questions

Given specific weather conditions at a particular location, is a customer more likely to choose a taxi or an Uber ride?

Data Description

Source: We use the dataset from the NYC Taxi and Limousine Commision (TLC) about taxi fares in New York City for the year 2023, obtained from the official TLC data page. Additional datasets: NYC 2023 Weather Dataset (Open Meteo Weather API), NYC Uber Rides (NYC Open Data - Rideshare dataset)

Description: This dataset provides valuable insights into taxi fare distribution across different zones, distribution across hours, components of fares, and trip distances.

Objective: The primary objective of this project is to analyze the patterns in NYC taxi fares throughout 2023, identifying key factors that influence fare amounts. By assessing the impact of each variable, we aim to provide taxi drivers insights that would help them optimize their service strategy.

Feature Name	Definition
Trip Distance	The elapsed trip distance in miles reported by the taximeter
PULocationID	TLC Taxi Zone in which the taximeter was engaged
DOLocationID	TLC Taxi Zone in which the taximeter was disengaged
Fare Amount	The time-and-distance fare calculated by the meter
Payment Type	A numeric code signifying how the passenger paid for the trip.

NYU

Question 1

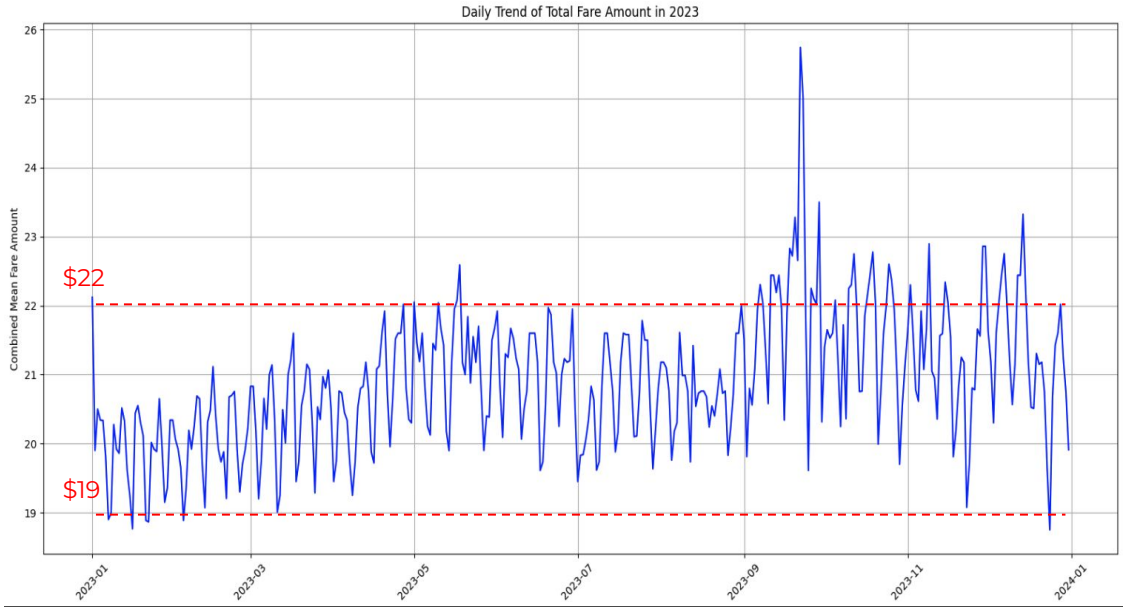
Are there temporal patterns in Taxi Fares During 2023?

NYU

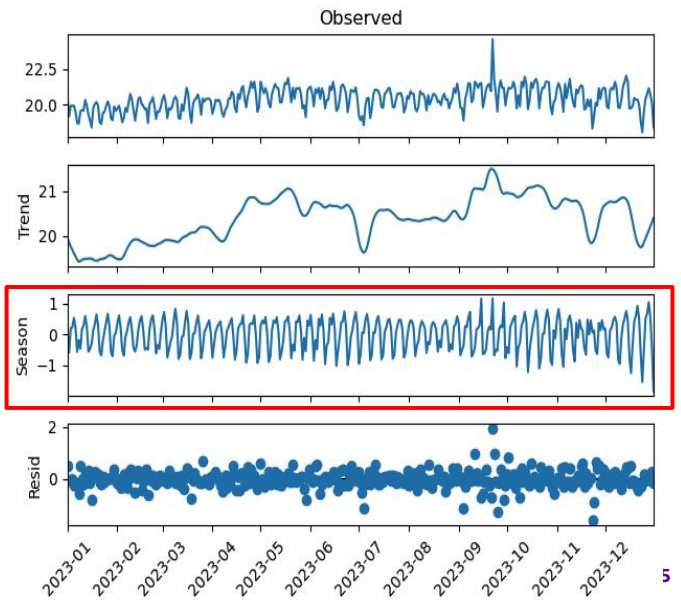
There is **no clear evidence of significant upward/downward trend** in the taxi fares, but there appears to be a **clear seasonal pattern**

- On a macro level, the daily median taxi fare in 2023 exhibits a **relatively constant trend** , with values typically fluctuating between $19 and $22
- There appears to be **periodic fluctuations** , potentially linked to time-of-week or time-of-year effects

Daily Trend of Total Fares in 2023



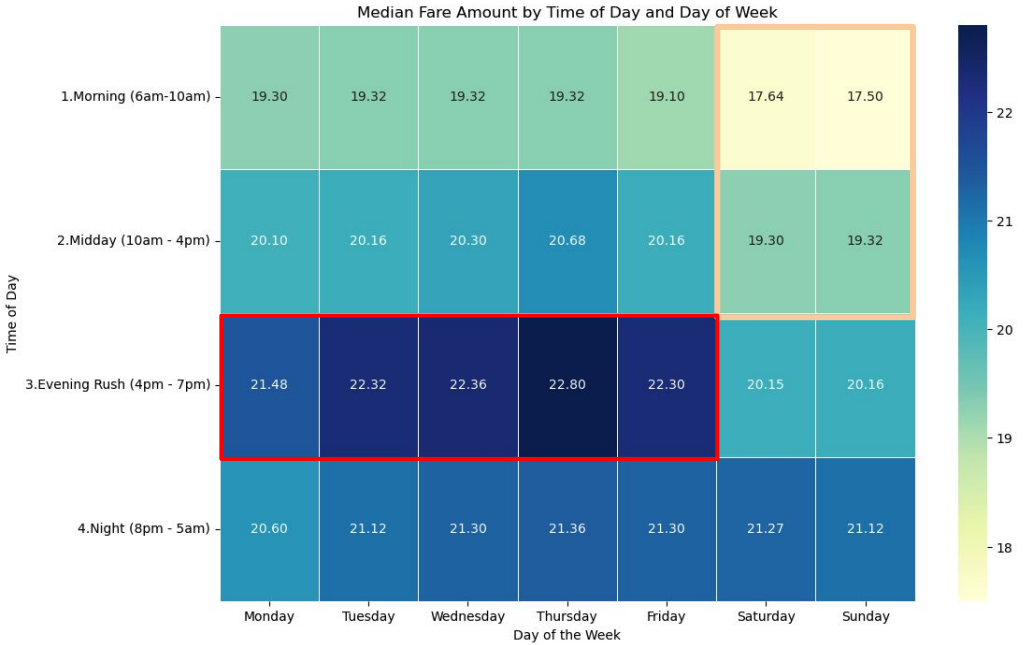
Seasonal-Trend Decomposition



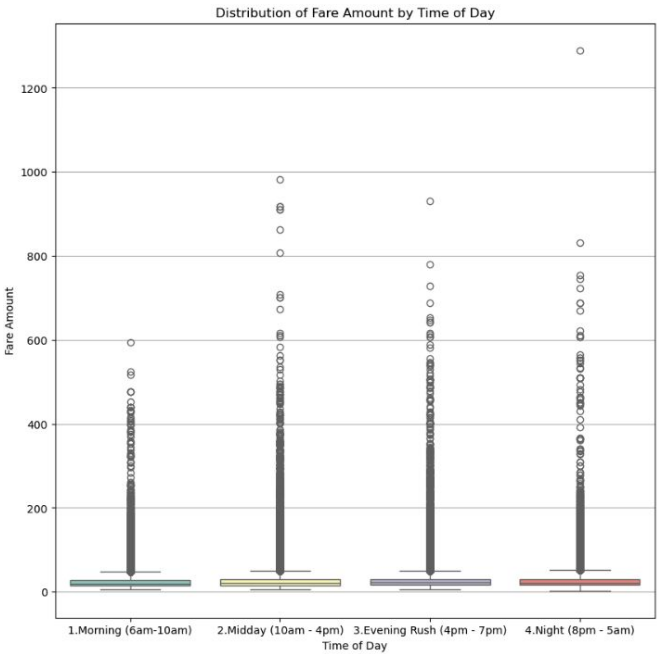
When split into different time blocks, the taxi fare tends to be **higher during weekdays** and **drop during weekends.**

- Fares tend to **peak around Thursday Evening** (4pm-7pm) and **drop significantly in the earlier hours during weekends** (6am - 4pm).
- The higher fares during weekdays could indicate that taxis in NYC are **frequently used for work-related travel** (business trips, commuting) and **may not be a primary transportation mode for leisure trips or family outings** .

Heatmap of Time of Day x Day of Week



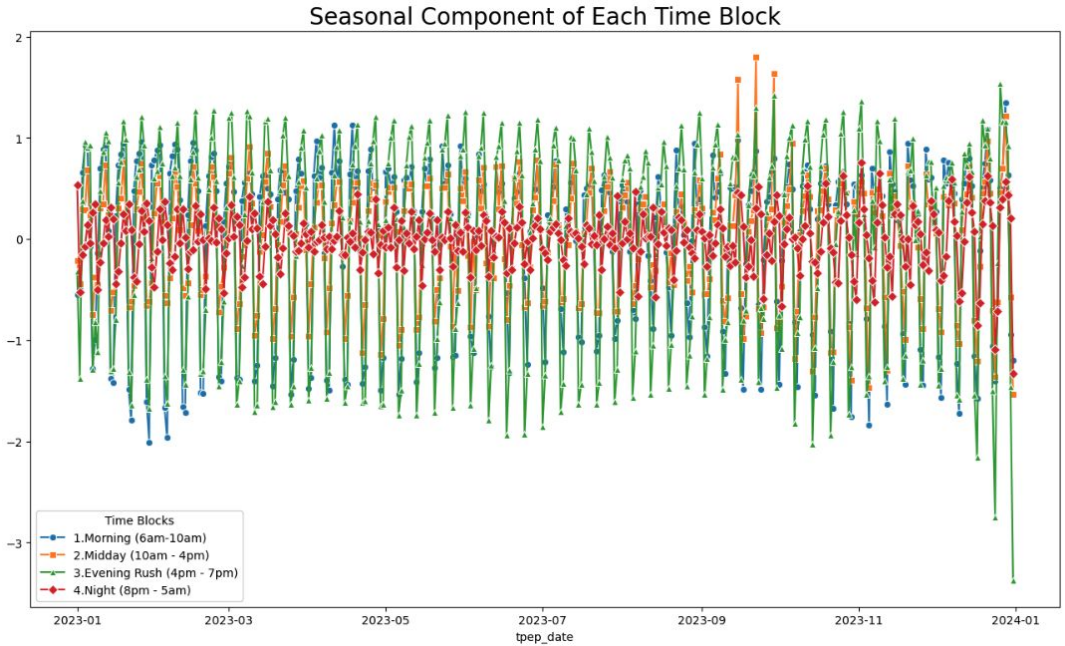
Distribution of Fare by Time of Day

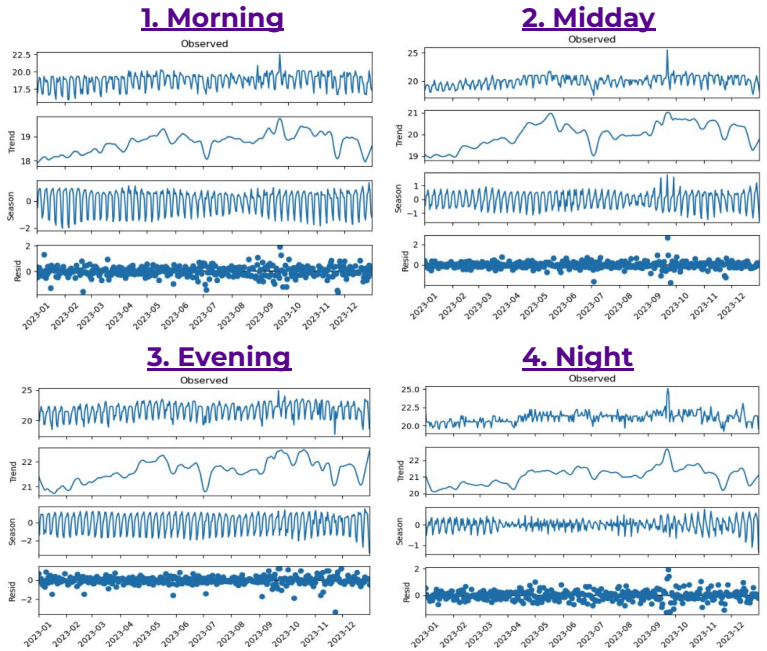


The seasonal component analysis confirms a consistent **peak in taxi fares during the Thursday Evening Rush (4pm - 7pm)**

- Across all time blocks, Thursdays consistently stand out as the peak day for fares (weekly trend).
- Evening rush hours exhibit the highest variability in taxi fares compared to other time blocks.

Seasonal Component of Each Time Block





Question 2

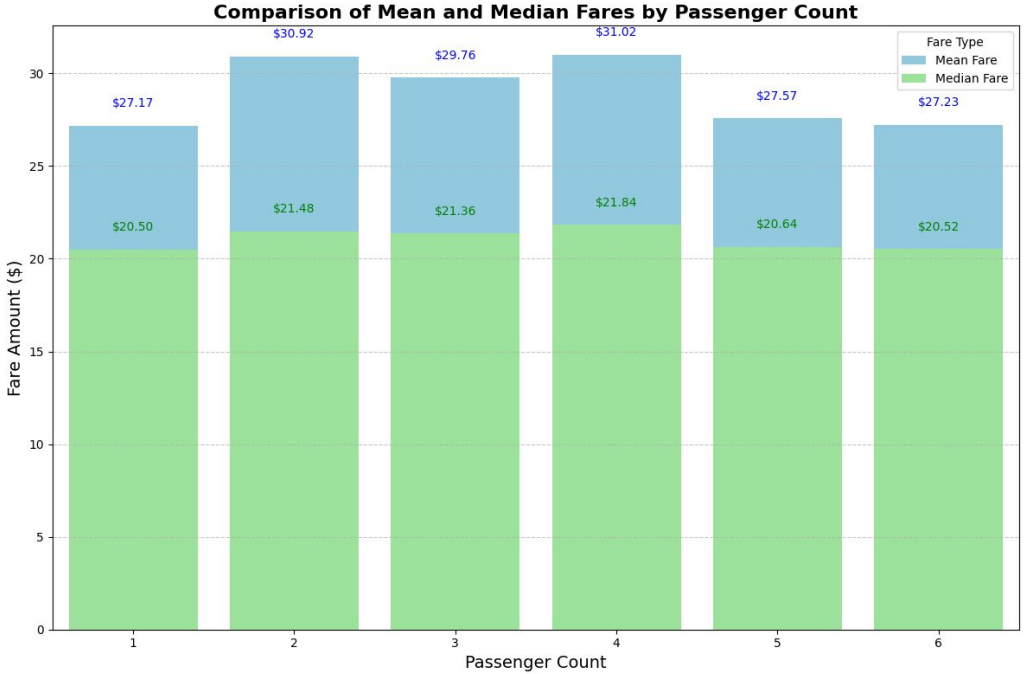
How does having multiple passengers impact the total fare?

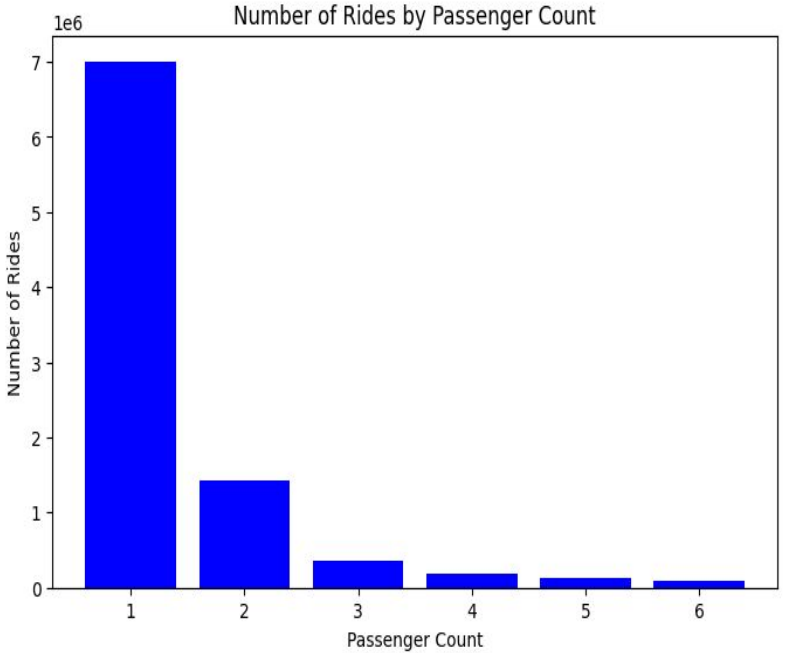
NYU

- **While fares slightly increase with passenger count, peaking at 4 passengers, decrease slightly for 5 and 6 passengers, likely due to flat-rate policies, ride-sharing discounts, or pooling arrangements however, the impact of passenger count on fare amounts remains minimal.**

- **Single-passenger rides dominate the data, likely reflecting the convenience and preference for solo travel.**

- **The mean fare is consistently higher than the median fare across all passenger counts indicating a right-skewed distribution, likely caused by outliers such as long-distance trips or premium services.**





Question 3

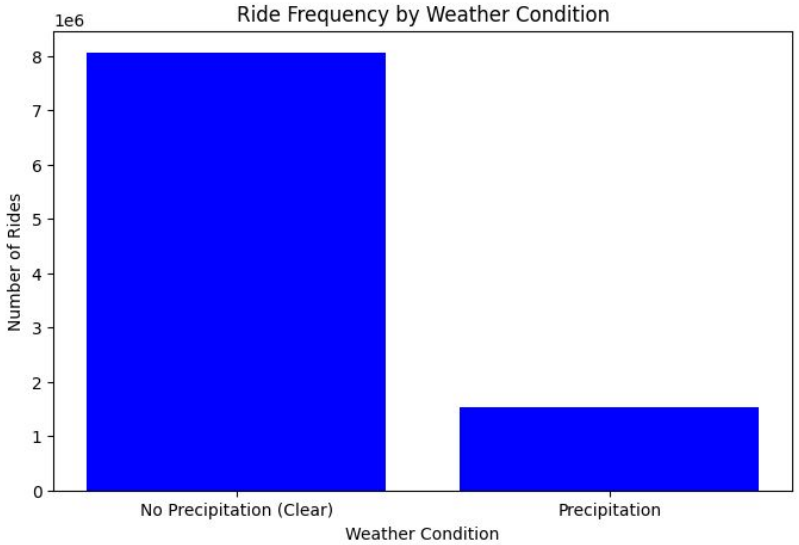
How are weather conditions associated with ride frequency and fares?

NYU

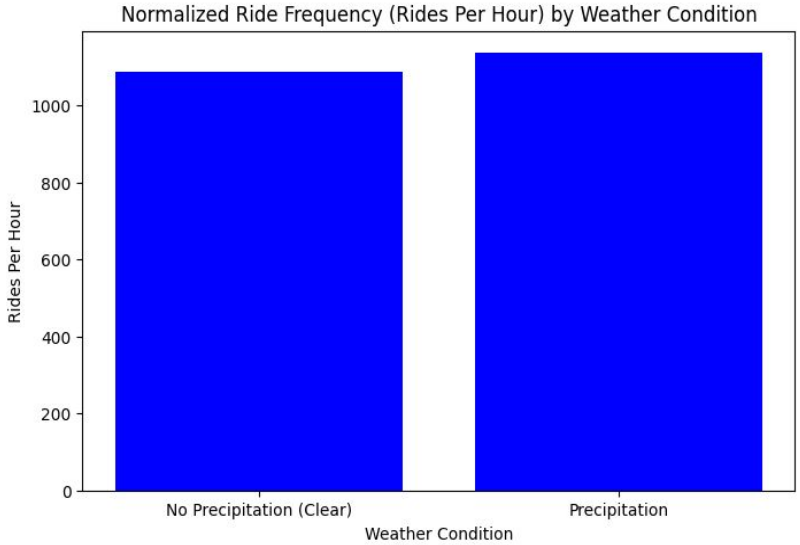
There is a significant disparity in raw ride counts of clear weather data compared to precipitation days.

- It is more common to have clear (no precipitation) days.
- Normalization to account the average rides per hour by weather condition

Bar Chart comparing ride Frequency



Bar Chart comparing ride Frequency

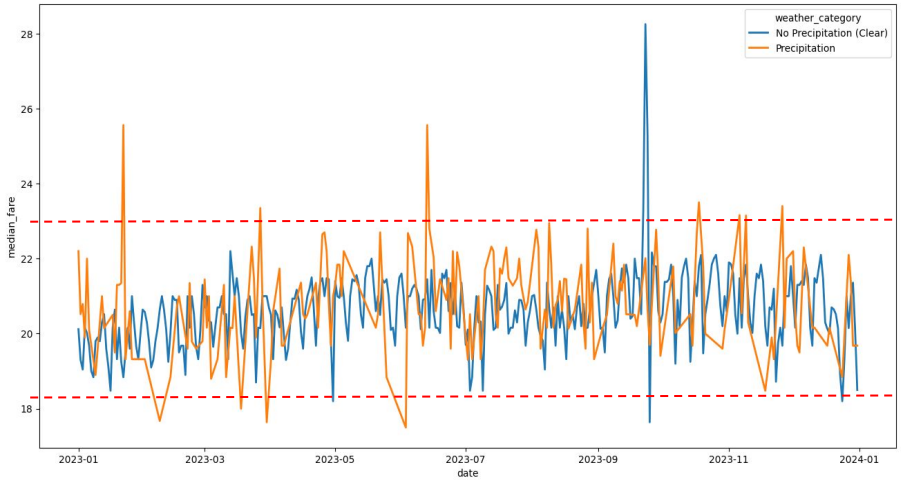


NYU

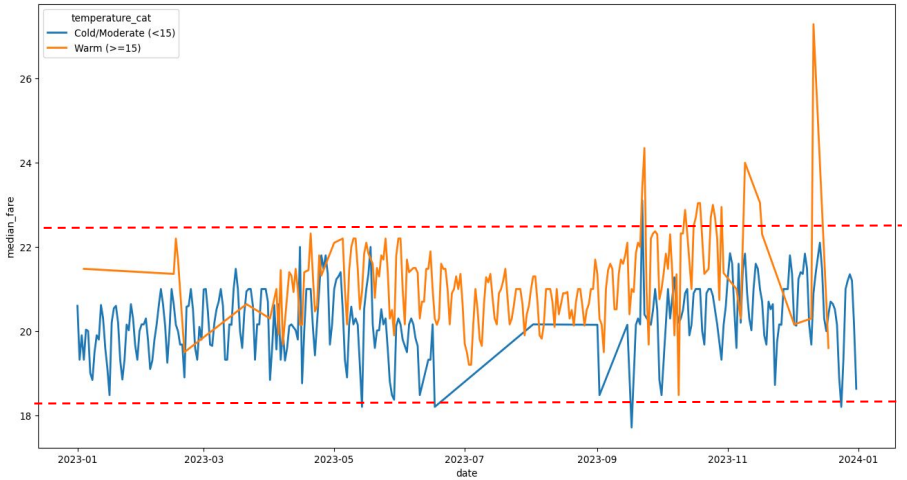
Median fares stay relatively stable regardless of weather condition or temperature

- Warmer temperature show higher volume of rides (1176 rides/hour) with greater variability (406) compared to colder temperatures (1060 rides/hour). Implying more stable patterns when weather is colder.
- Median fares remain relatively stable across temperature categories with 21$ per ride for warmer and 20.18$ for colder temperatures, reflecting minimal variability.

Median total fare per Hour by Weather Condition



Median total fare per Hour by Temperature

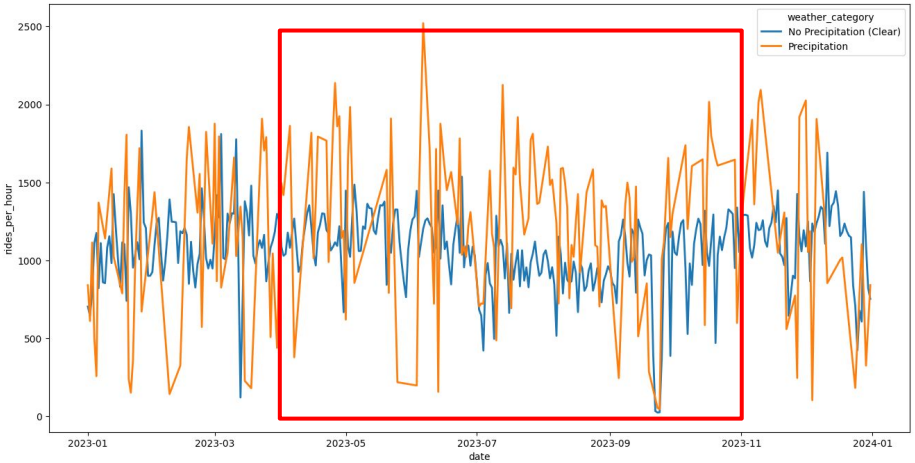


NYU

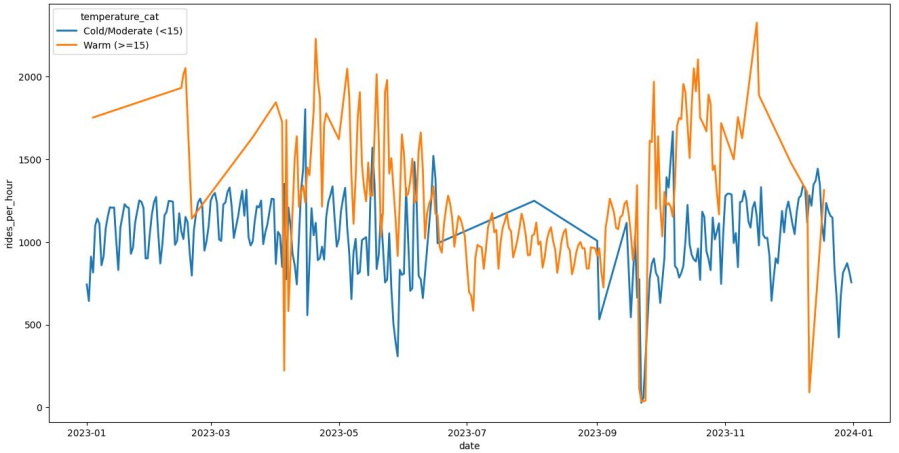
Average rides per hour have higher variability due to precipitation

- Precipitation increases variability in ride demand, as seen with higher fluctuations in rides per hour compared to clear weather.
- Warmer temperatures consistently show slightly higher especially when precipitation is present. When clear day is present, then is more stable.

Daily Rides per Hour by Weather Condition



Daily Rides per Hour by Temperature



NYU

Question 4

Does the pickup location influence the total fare? If so, which pickup locations tend to generate the highest fares

NYU

Borough-Level Analysis

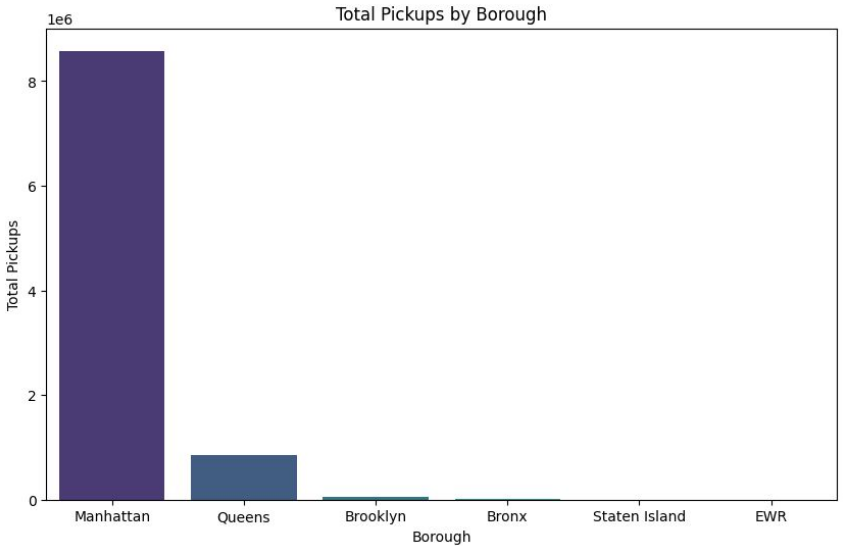
Ride Demand Patterns:

- **Manhattan** dominates pickups due to business and tourist activity.
- **Queens** follows, with airports driving high demand.
- Other boroughs (Brooklyn, Bronx, Staten Island) show lower demand.

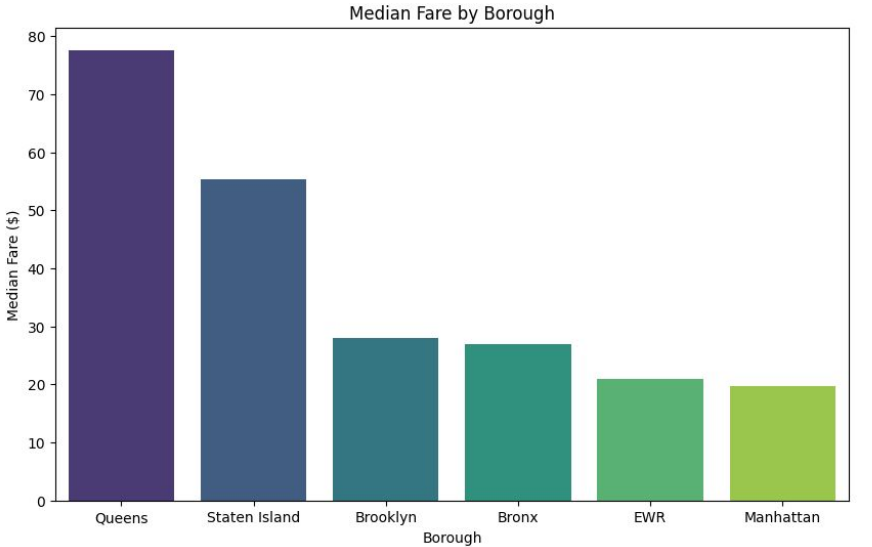
Fare Variability:

- **Highest Median Fares**: Queens (JFK & LaGuardia Airports) due to long trips and surcharges.
- **Lowest Median Fares**: Manhattan (short distances and frequent trips within the borough).

Total Pickups by Borough



Median Fare by Borough



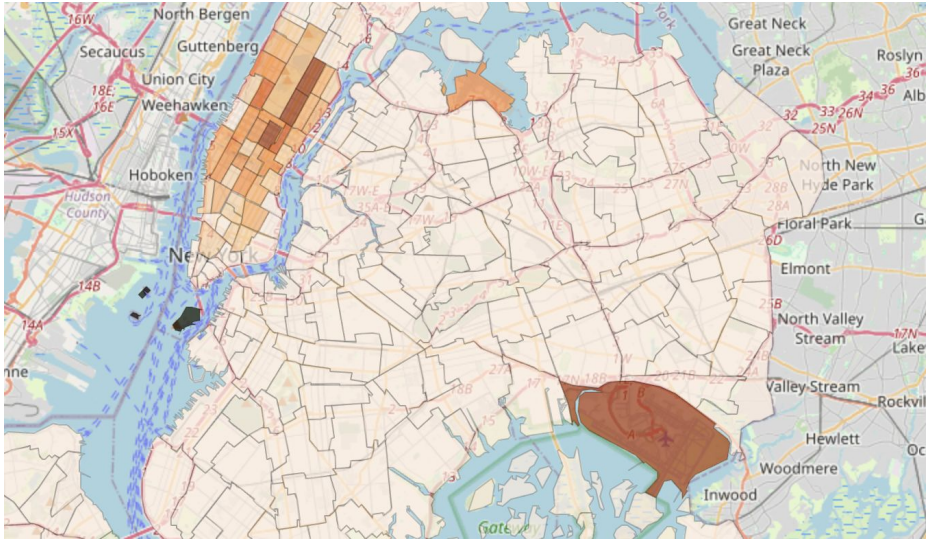
Zone-Level Analysis

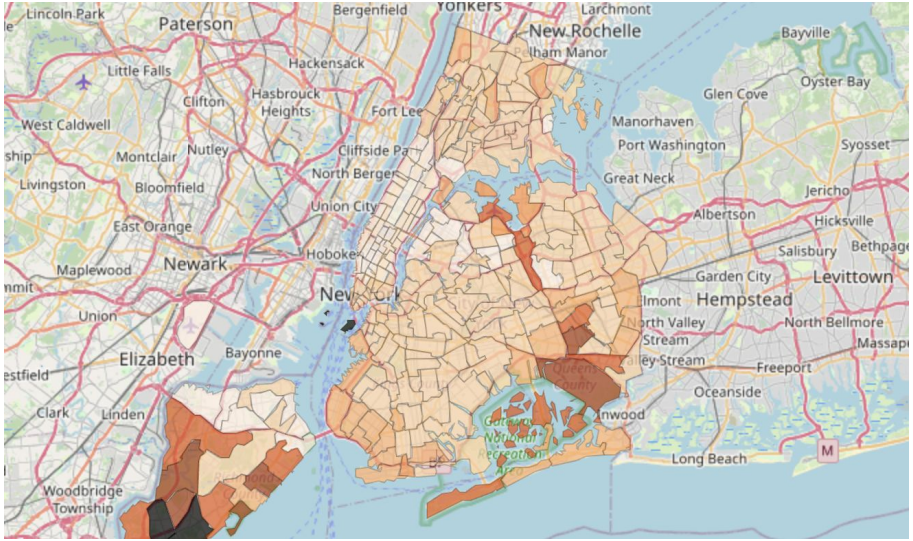
The geospatial analysis reveals distinct patterns in ride demand and fares across NYC:

- **High-Demand Zones:** Midtown and Lower Manhattan dominate in ride pickups due to business and tourism, but fares are moderate due to shorter trips.
- **High-Fare Zones:** Airports (JFK, LaGuardia) and peripheral areas like Staten Island show higher fares due to long distance trips and additional fees.

Heatmaps reveal that ride density clusters in commercial hubs and zones with high fares concentrated near airports and long-distance areas.

Interactive Heatmaps





Question 5

Can we determine the most impactful variables that are associated with ride frequency and fares?

NYU

Takeaways from previous questions

1. The total fare tends to remain relatively stable throughout the year 2023 but shows the **highest fluctuates during the Evening Rush Hours (4pm - 7 pm)**

2. The **passenger count has little to no effect** in modeling the taxi fare

3. The taxi fare tends to exhibit **higher variability during precipitation** , but appears to be less sensitive to changes in temperature conditions.

4. Pickup location is a **critical determinant** of fare amounts, with **high-demand and long-distance** boroughs or zones generating the highest fares.

Linear Regression Output

Feature	Description	Linear Regression Coefficient
Weather_Category_Precipitation	Whether the trip happened during precipitation (binary)	0.004
Temperature_cat_Warm (>=15)	Whether the temperature was warm during the trip (binary)	0.001
borough_Brooklyn	Whether the pick-up was in Brooklyn (binary)	0.765
borough_EWR	Whether the pick-up was in EWR (binary)	4.233
borough_Manhattan	Whether the pick-up was in Manhattan (binary)	-6.356
borough_Queens	Whether the pick-up was in Queens (binary)	**44.708**
borough_Staten Island	Whether the pick-up was in Staten Island (binary)	**24.940**
time_of_day_2.Midday	Whether a trip happened during Midday (binary)	0.742
time_of_day_3.Evening Rush	Whether a trip happened during Evening (binary)	**1.818**
time_of_day_4.Night	Whether a trip happened during Night (binary)	1.166
passenger_count_standardized	Standardized counts of passengers (numerical)	0.532