New York City Yellow Taxi Data

Neal Sodhi
Zhanqiu Guo

Background and Problem Statement

New York City is famous for its yellow taxis.

They are a government sponsored transportation force utilized by many New Yorkers.

The Taxi & Limousine Commision (TLC) of New York City publishes monthly data of taxi trips, including:

Pickup Locations, Dropoff Locations, Fares, Tips, Durations, Times of Day, Tolls, Fees, Surcharges, Number of Passengers, Payment Methods

Background and Problem Statement (Pt. 2)

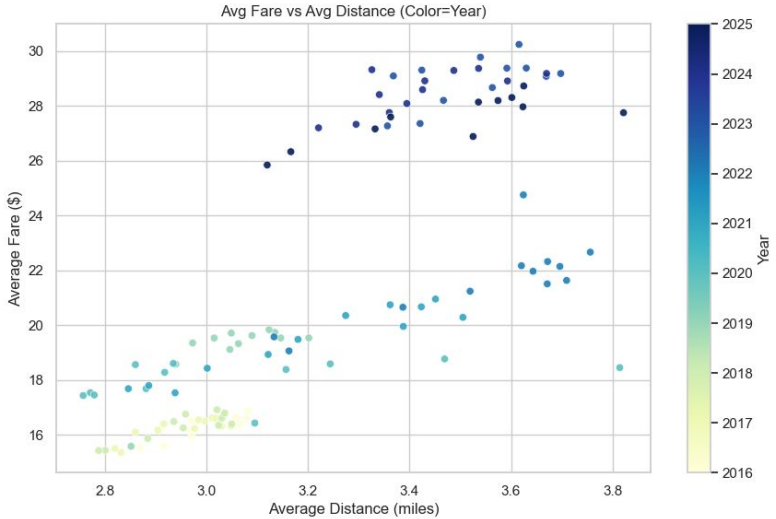
1. **Taxi Drivers** want to maximize their personal profit as best they can (via efficient routes and tip potential).
2. **Taxi Riders** want fair prices, low wait times, and safety.
3. **Government Regulators** wants to ensure accessibility across the city, prevent price gouging, control traffic rates, and stabilize driver income.

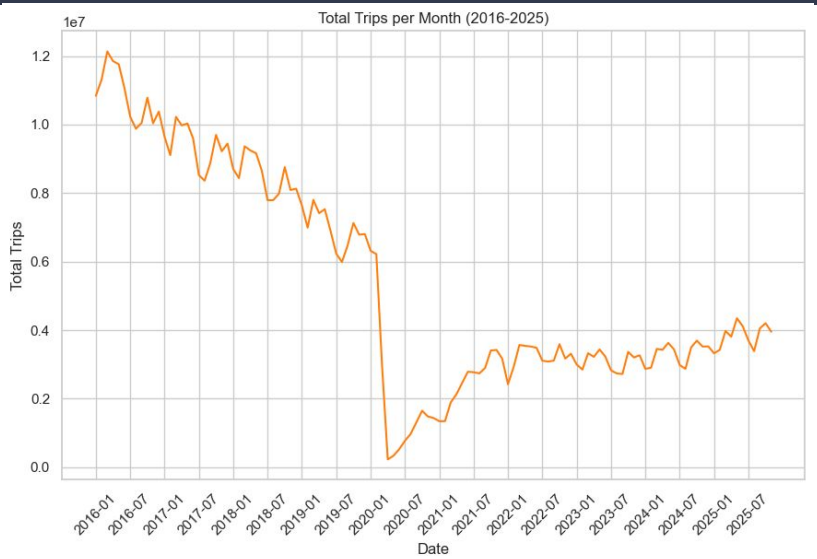
The result of this is a **complex equilibrium** that we aim to discover using Machine Learning:

We seek to identify where the system appears balanced, where disparities arise, and how pricing outcomes differ across the five boroughs.
We aim to learn the structure of the market and evaluate how well it aligns with fairness objectives for drivers, riders, and regulators.

Demand Drop & Rising Costs

- The Pre-Pandemic Decline (2016~2020)
- The COVID Crash (2020~2021)
- Stagnant Recovery (2021~2025)

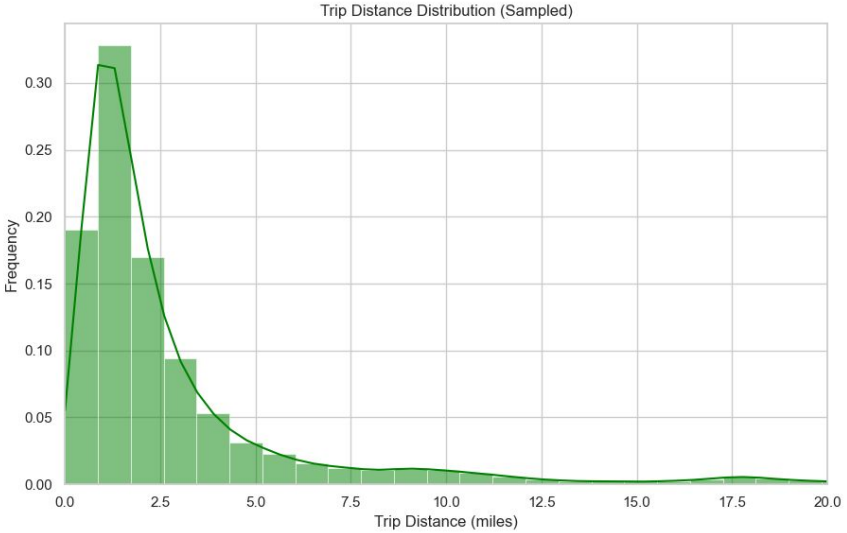


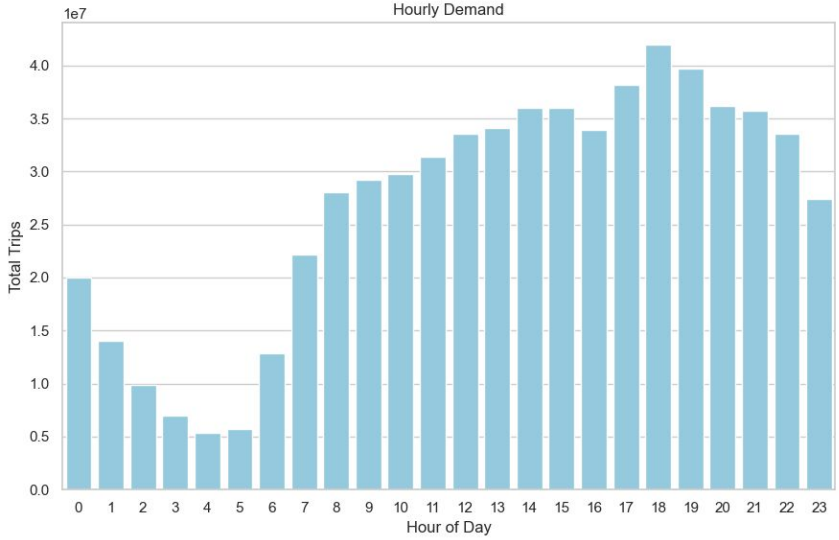


- Inflation and Rate Hikes
- Trip Distance Changes

Daily Travel Pattern

- Travel Vally (1 AM~6 AM)
- Morning Rush (8 AM~9 AM)
- Evening Rush (5 PM~7 PM)





- “Last Mile” Solution: majority of trips between 1~2 miles

New York City Taxi Fare Calculation

Fare: Fixed cost + Linear with regard to miles + Linear with regard to time + Additional fees, surcharges, and tolls

Important: Cost determined by the New York City Taxi & Limousine Commision (TLC)

Destination is not known until pickup, unlike with common rideshare apps today (Uber, Lyft).

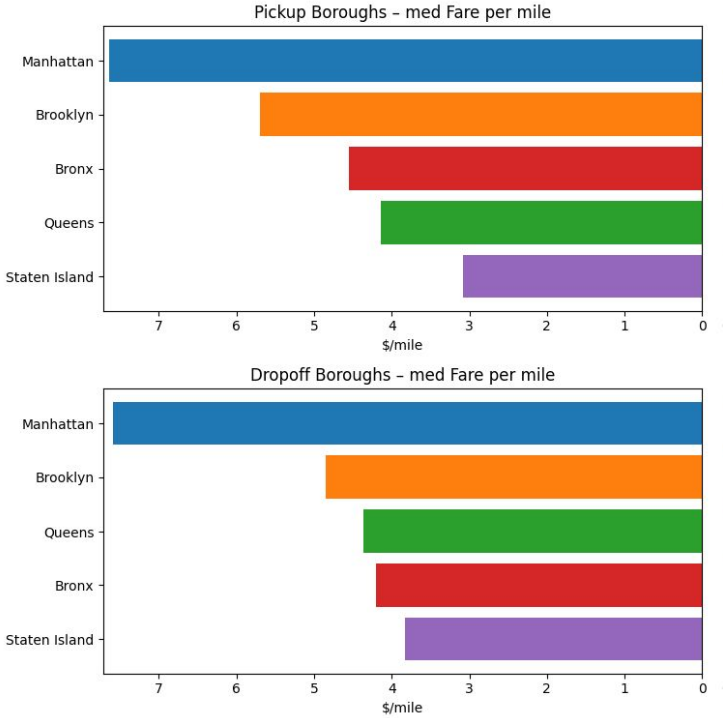
However, cab drivers can decline rides upon learning the destination.

Tips are entirely determined by the customer.

Fares Per Mile Analysis

Most **urban** boroughs have highest median fare per mile.

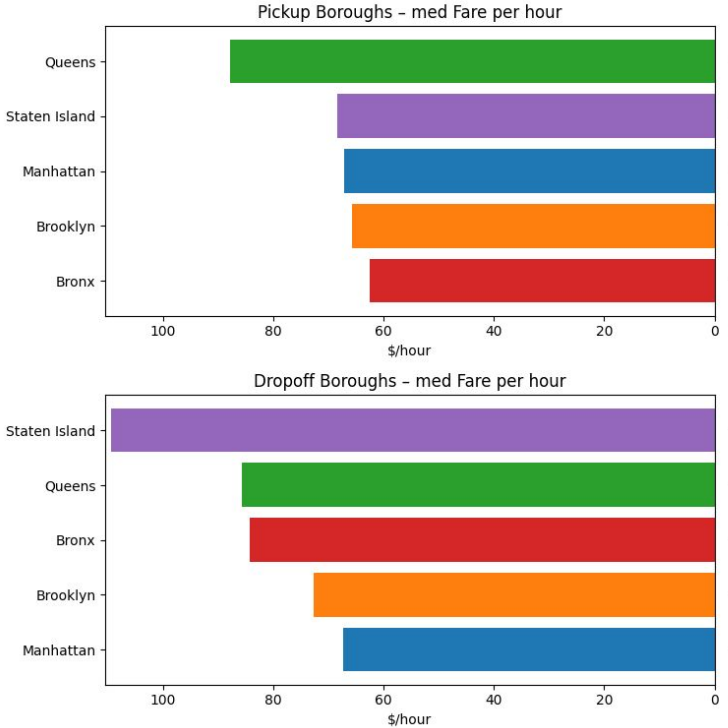
Consistent across pickup and dropoff.



Fares Per Hour Analysis

Most **suburban** boroughs have highest median fare per hour.

Consistent across pickup and dropoff.



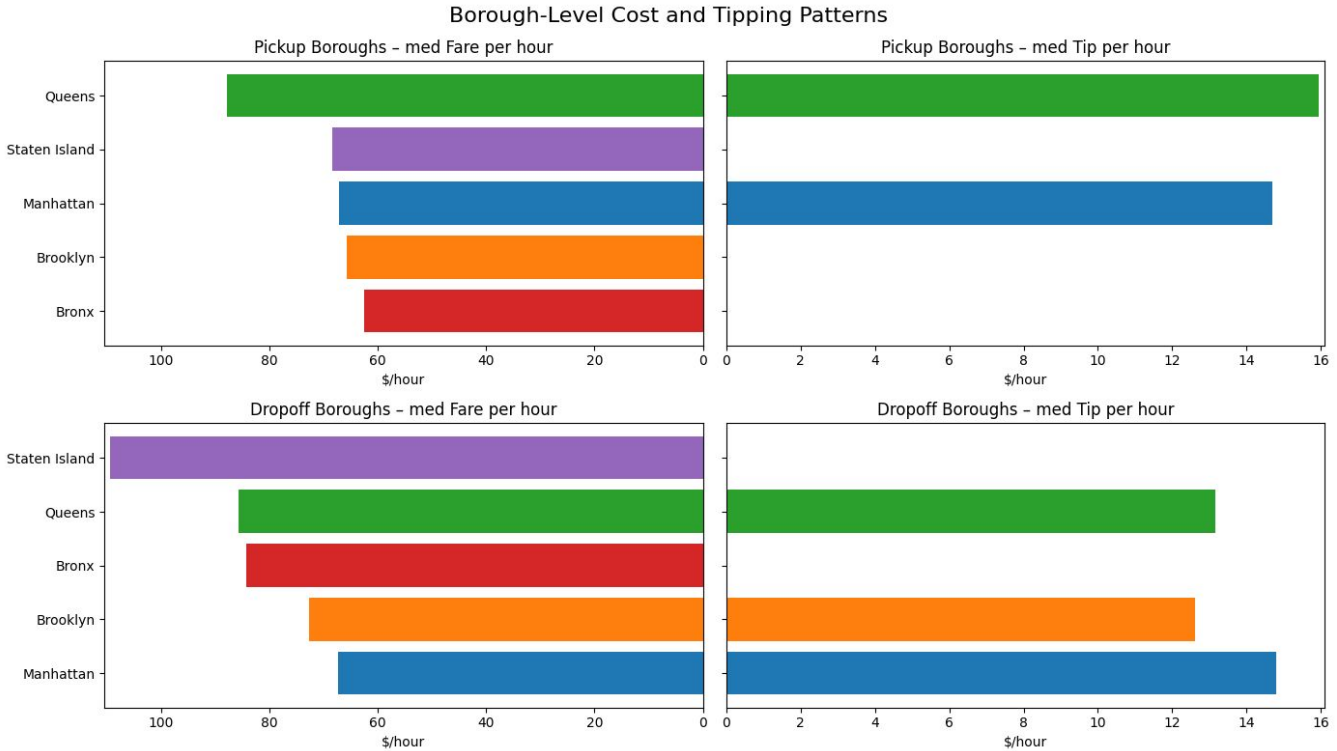
Tips Per Hour Analysis

Drivers know that not all customers will tip the same.

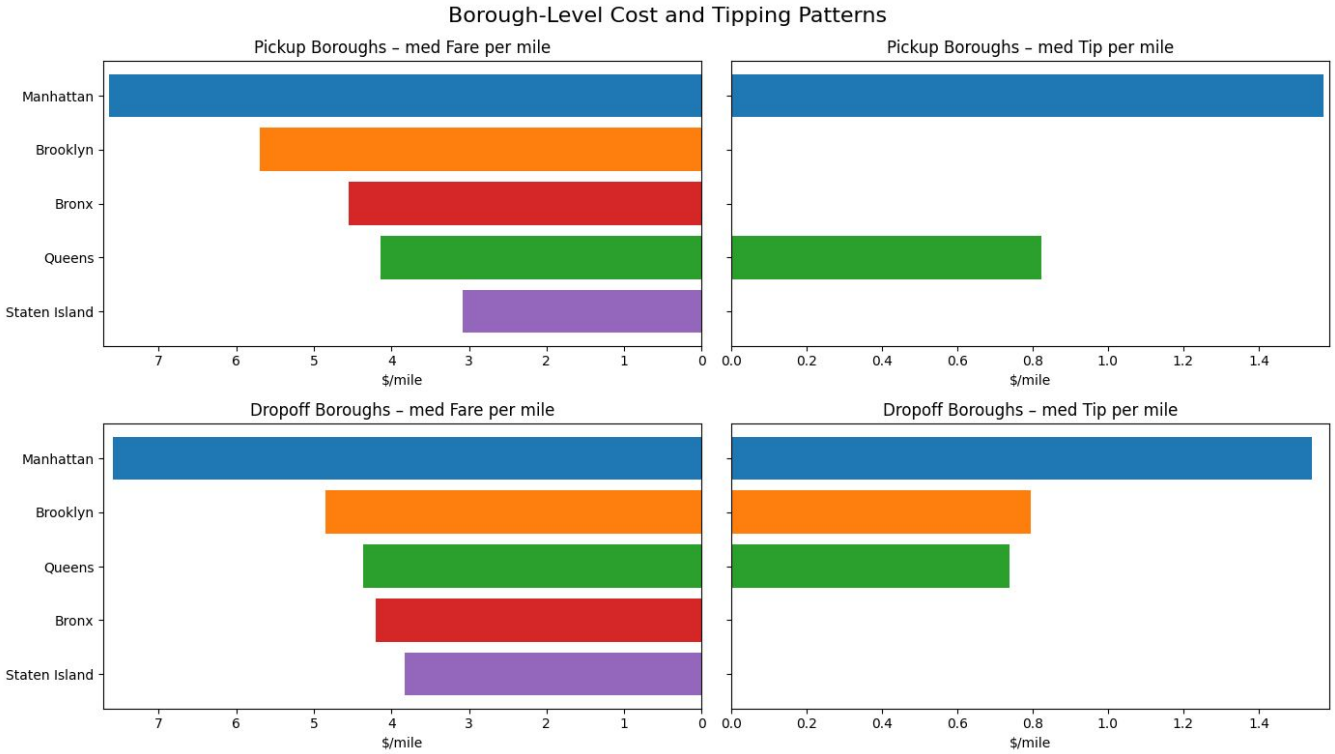
They may adjust their behavior or decline rides accordingly.

Queens (green) is where airports are located.

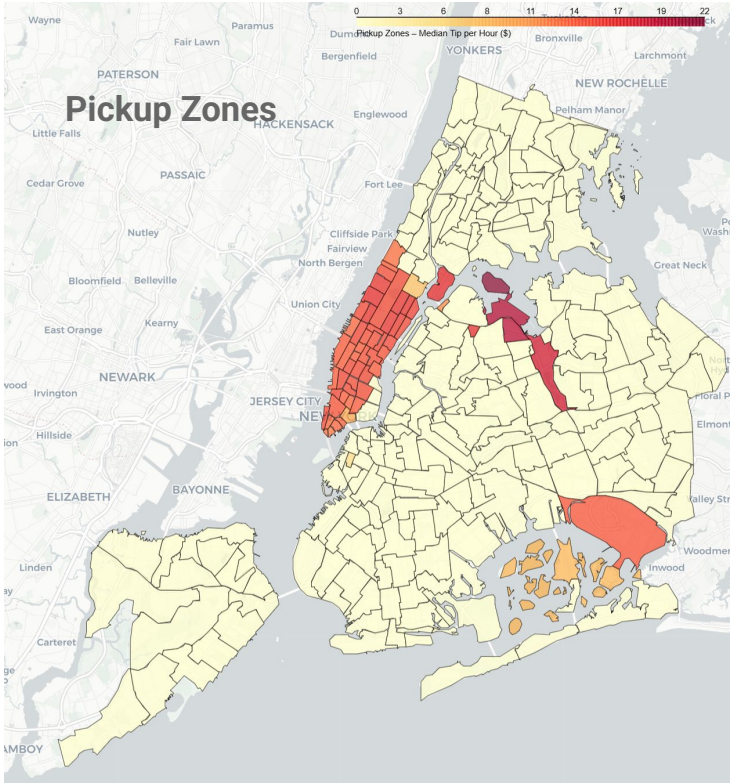
Bronx (red) is generally considered the poorest borough.

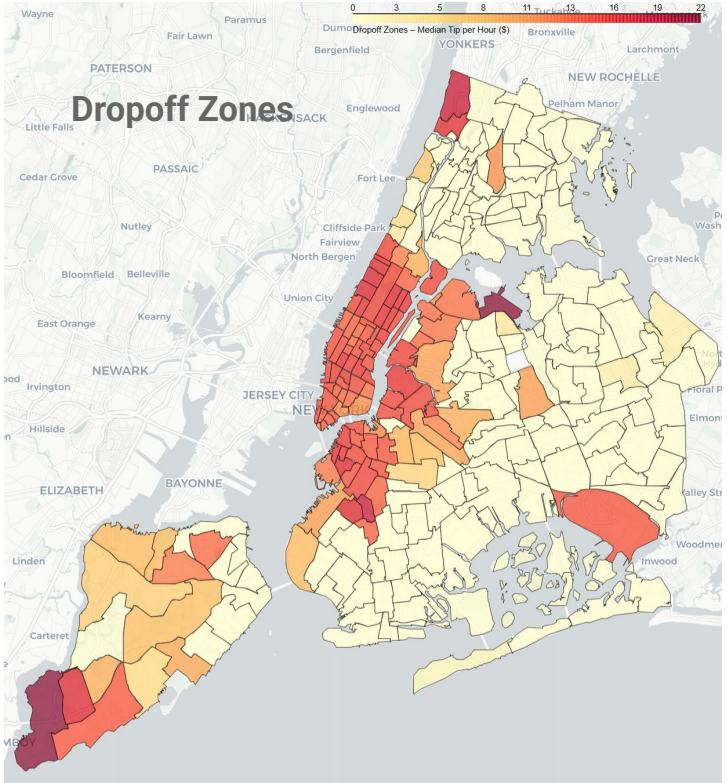


Tips Per Mile Analysis



Geographic Map of Median Tip Revenue Per Hour



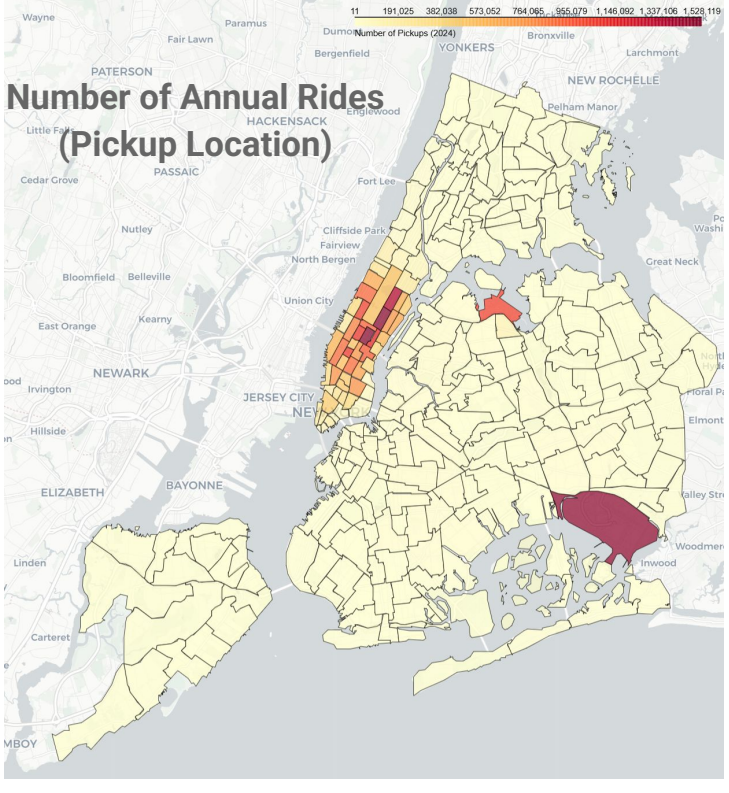


New York City Taxi Distribution

Yellow Taxis are authorized to start rides in any of the five boroughs of New York City.

95% of taxi pickups happen below 96th street in Manhattan or at the airports.

This leaves the outer boroughs with less access to government taxis.



New York City Taxi Distribution (Pt. 2)

New York City's administration introduced "Boro Taxis" in 2011 to address this problem.

These taxis are only allowed to pickup in upper Manhattan and the outer four boroughs (excluding airports).

Phased Out: Over 7500 in 2015; Less than 900 by 2023.

TLC issuing more permits to private rideshare drivers instead.

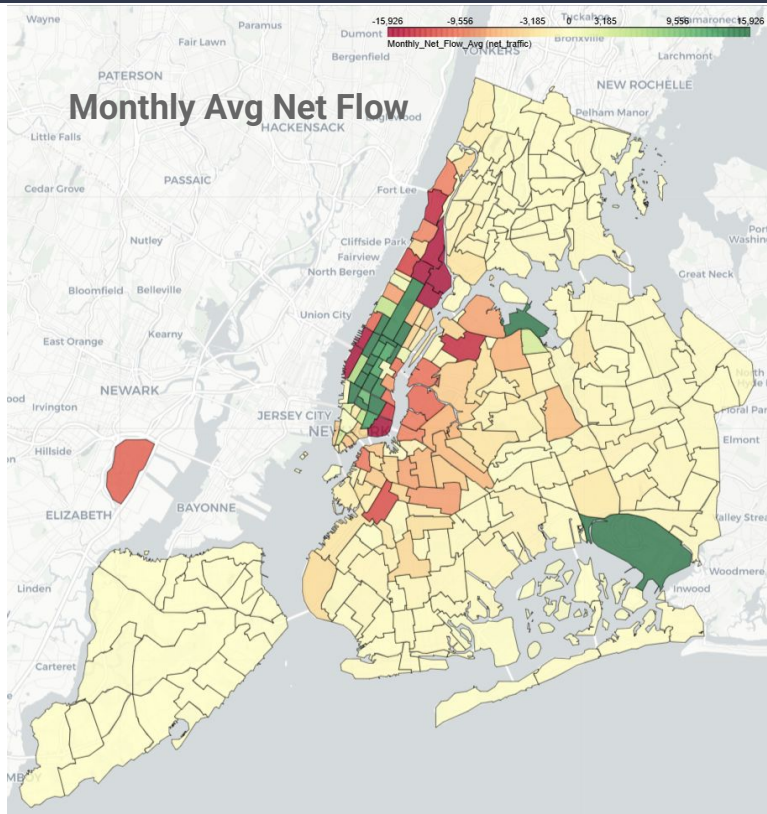
Private rideshare pricing not determined by the government.

New York City Taxi New Flow

- Green Zones (Source): High demand in JFK/LGA Airports and Midtown Manhattan. Generate more pickups than drop-offs
- Red Zones (Sink): Residential Boroughs (Brooklyn, Queens) and Upper Manhattan. Absorb trips but generate fewer return fares

Business Insight:

1. High efficiency in Green Zones due to rapid turnover
2. Red Zones indicate high probability of empty return trips, reducing driver profitability per hour

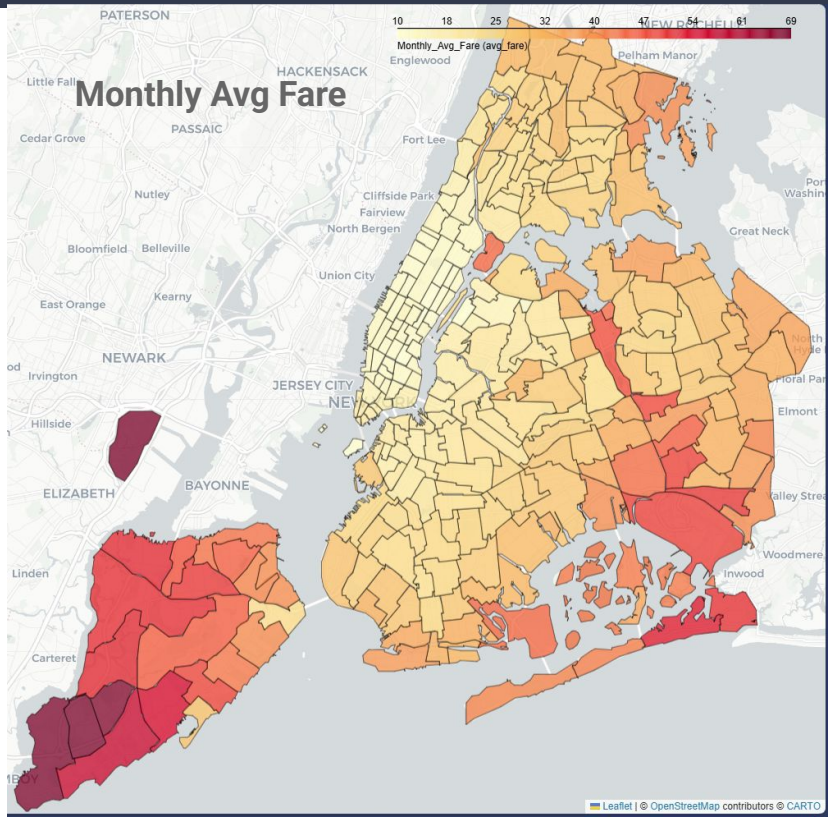


Pricing Geography

- Light Yellow (Low Fare): Manhattan —— high-frequency, short-distance trips.
- Deep Red (High Fare): Staten Island, Far Rockaway, and EWR Airport —— long-haul, cross-borough, or airport trips.

Business Insight:

1. Trade-off between trip frequency and distance
2. High-fare zones are geographically large but contribute little to transaction volume



Revenue Heatmap

- Cluster 1 (Midtown): driven by volume, the sheer number of short trips outweighs the low average fare
- Cluster 2 (JFK Airport): driven by value, high average fares with consistent demand

Business Insight:

1. 80% of revenue is likely generated in less than 20% of the geographic area
2. Operational resources and driver incentives should prioritize Manhattan/Airport corridor to maximize platform Gross Transaction Value