Yellow Taxi: Iconic symbol of New York

Today's Agenda

- Introduction
- What is the research problem ? Why does the problem matter to them?
- How do we help them to tackle the challenge?
- Business Suggestion from EDA and Logistic Regression

Introduction

- Problem Statement

  How can the yellow taxi industry in New York City address the challenge of decreasing demand?

  - Rapid growth of Uber & Lyft, challenging the traditional yellow taxi industry
  - Significant decrease in yellow taxi trips from 175 million (2013) to 71 million (2020)
  - Need to understand factors driving these changes.

- Target Audience:
  - Yellow taxi drivers
  - Yellow taxi companies

NYC Taxi & Limousine Commission

- Mission:
  - Efficient allocation of cabs
  - Data-driven recommendations based on analysis.
  - Secure the future of the iconic yellow taxi in New York City

Methodology:

1

Data Source:
Utilized three datasets:

1. TLC Trip Record Data (Yellow Taxi, Nov 2022),
2. Taxi Zone Lookup Table, 3
3. High-Volume For-Hire Vehicle Records (Uber/Lyft)

Focused on Manhattan due to its highest concentration of yellow cab rides

2

Exploratory Data Analysis (EDA):

- Assess the service gap between yellow taxis and Uber/Lyft

3

In-depth Analysis:
Comparing Yellow Taxi and Uber/Lyft
Metrics used: Fare and rides
Demand across various zones in Manhattan

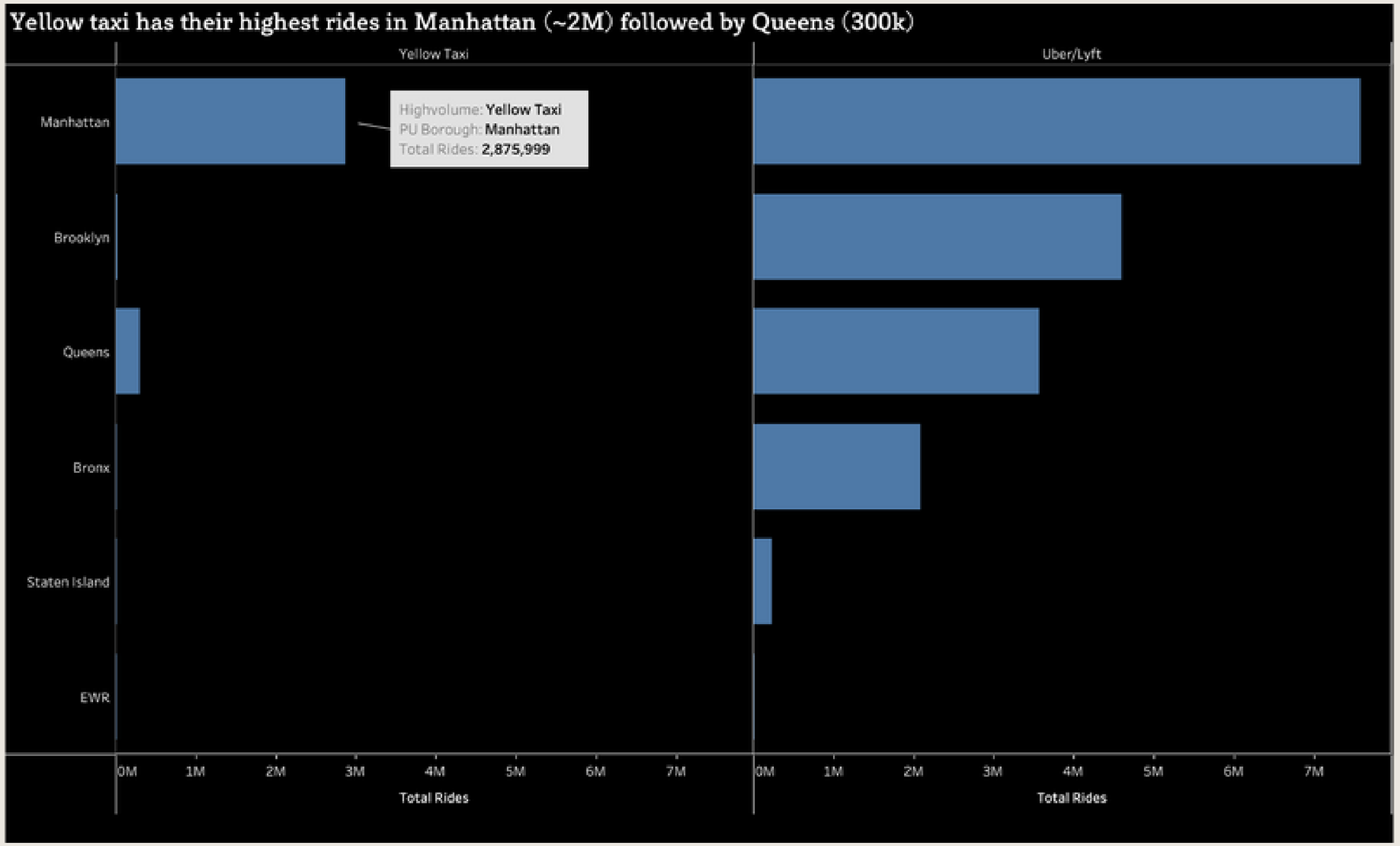
- Demand patterns with respect to time

4

Logistic Regression Model:

- Identify factors influencing a trip to be a yellow taxi or Uber/Lyft

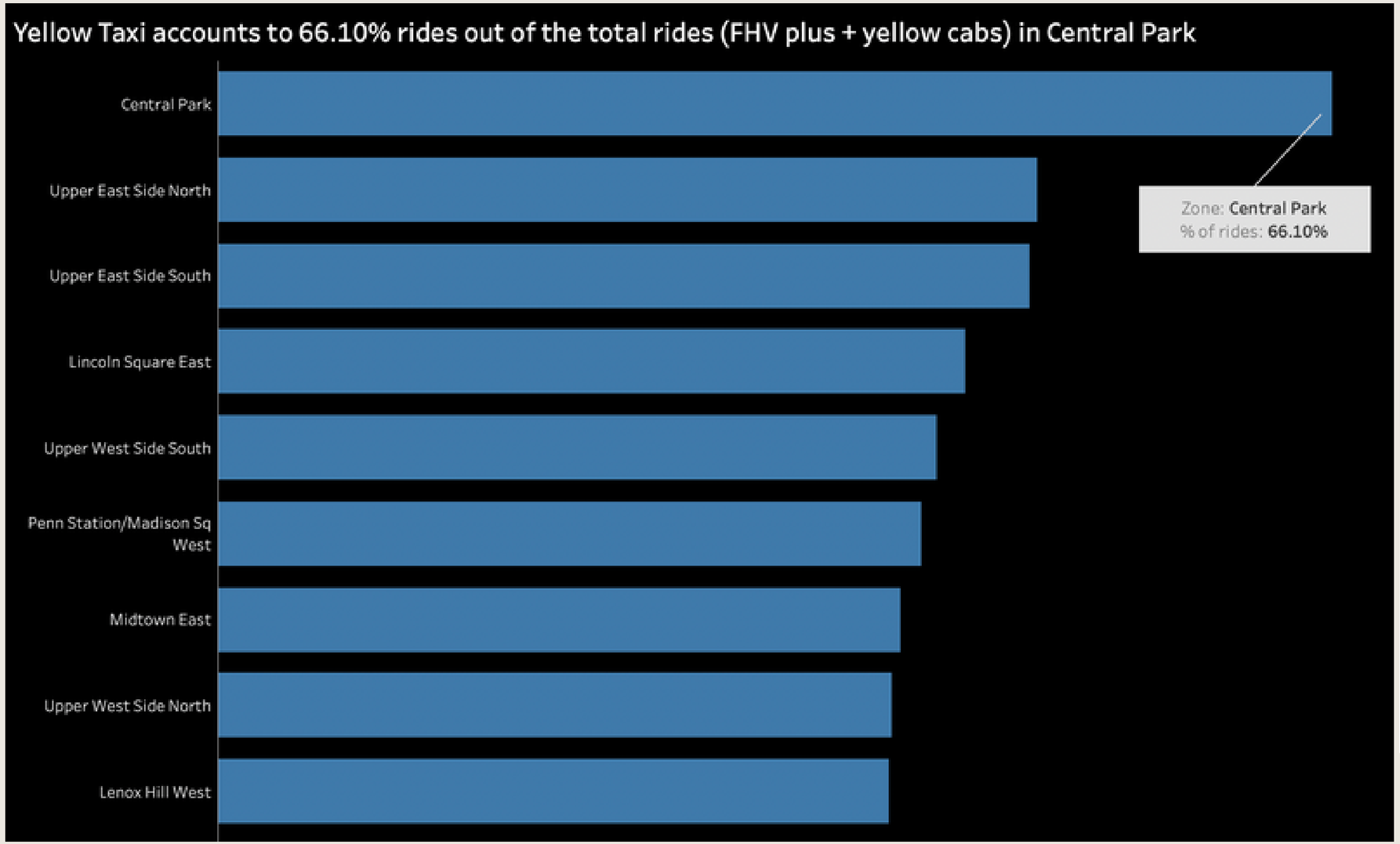
Total rides across boroughs for Yellow Taxi and FHV



Key Findings

- Yellow taxis majorly function in Manhattan and Queens (~2M and 300K respectively)
- Uber/Lyft functions across almost all the boroughs with Manhattan and Brooklyn having the highest number of rides (7586146 and 4602459 respectively)

Zone-wise percentage of trips for yellow cabs



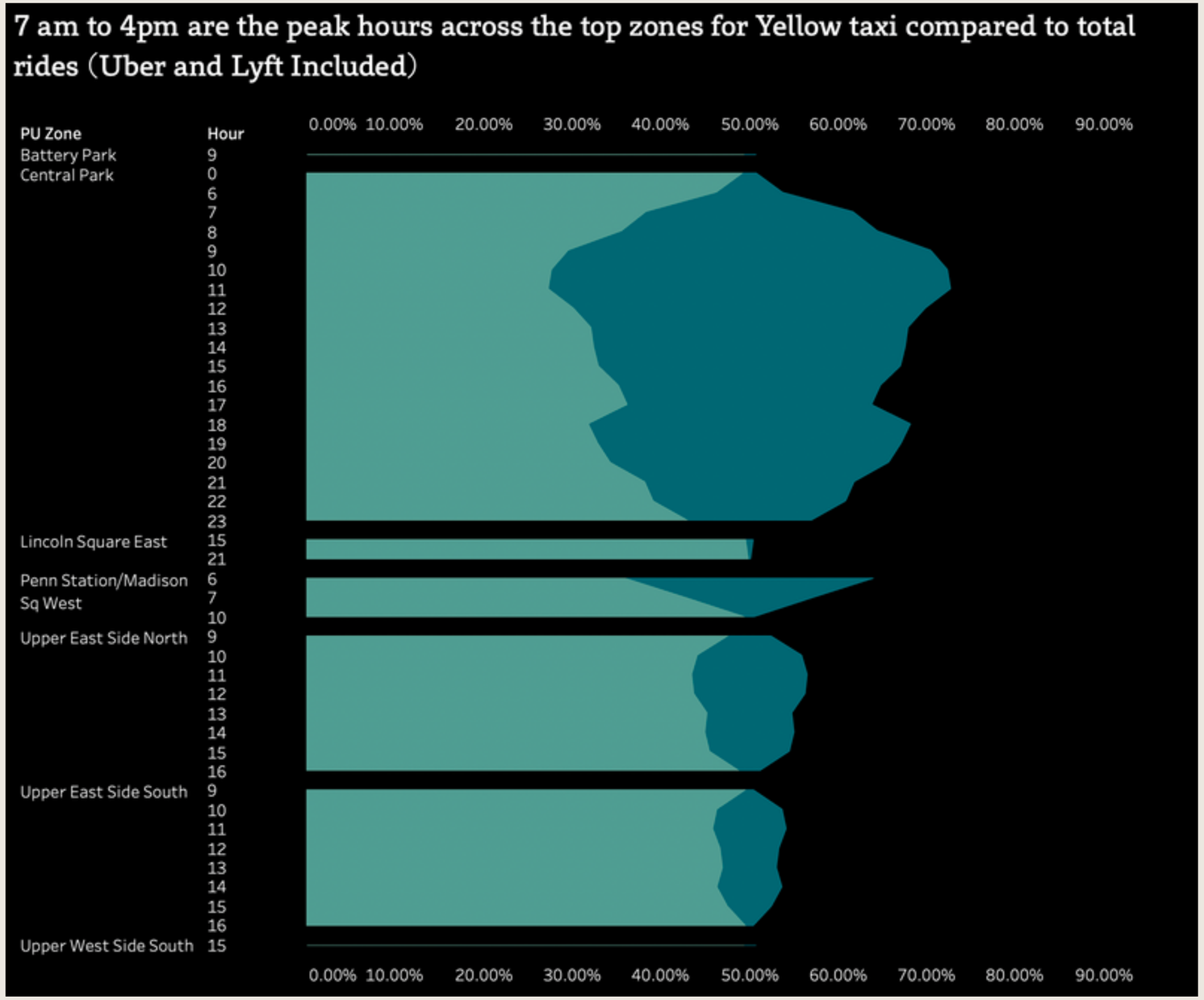
Key Findings:

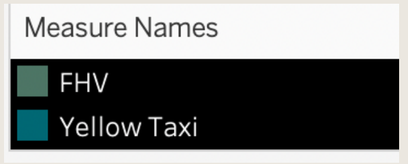
Yellow cabs accounts to higher % of rides(66.10) out of total rides in central Park

It is the second highest in Upper East Side North (48.65%)

Yellow cabs have third highest % of rides in Upper East Side South (48.16%)

Hour and zone-wise percentage of trips for yellow cabs and Uber/Lyft

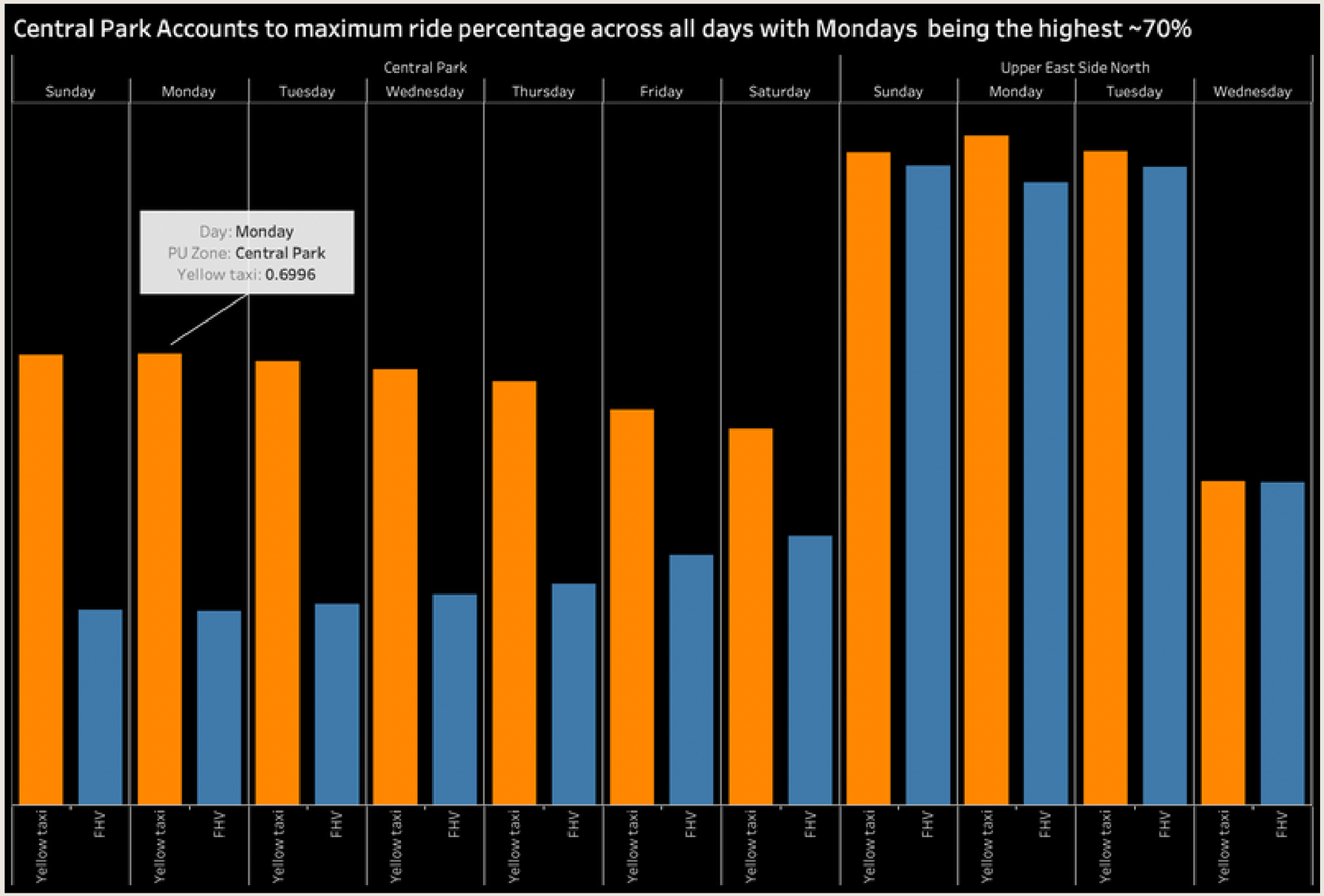




Key Findings

- In central park Yellow cabs have the most number of rides compared to FHV across all hours of the day peaking at 11 am (72.70%)
- Interesting observation is that in Penn Station the % of rides peaks to 64% out of total rides in that area at 6 am

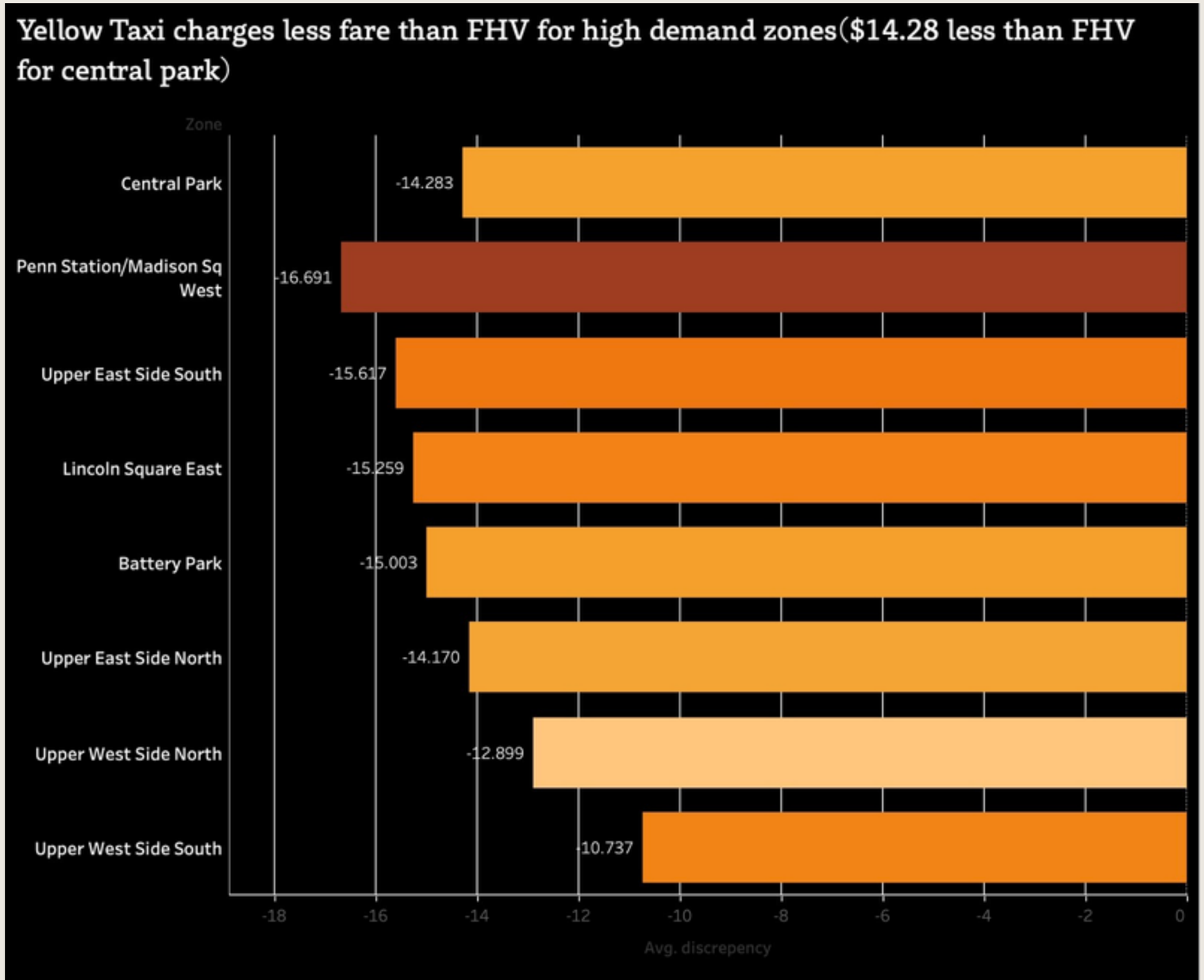
Day-wise analysis of top zones for yellow cabs and FHV



Key Findings

- Central Park the highest % of rides across days of week peaking to ~70% during Mondays
- Upper East Side has the second highest % of rides

Fare comparisons between yellow taxis and ride-hailing services



Key Findings

- Rides in Penn Station by yellow taxi charge $16.91 less than FHV rides
- Yellow cabs charge $14.28% less than FHV in Central Park

- Upper East Side has the second highest % of rides

Logistic Regression

- Independent variables: trip_miles, base_passenger_fare, dummies for days, for time of day, and for each Borough.
- Dependent variable: High Volume Flag: 1 -> Uber/Lyft and 0 -> Yellow taxi
- Month: November

Highlights

Measures of Performance:

- F1 Score: 0.93
- Accuracy: 0.88
- Precision: 0.88
- Recall: 0.98

Significant Variables:

- Manhattan: -4.897 (p-value=0.000*)
- EWR -3.172 (p-value=0.002*)
- Queens -2.970 (p-value=0.003*)

Negative coefficient: Probability of the ride being from Uber/Lyft decreases.

Logistic Regression

Model: Manhattan

- F1 Score: 0.89
- Significant Variables:
  - +Central Harlem North
  - +East Harlem North
  - +Hamilton Heights
  - +Inwood
  - +Marble Hill
  - +Roosevelt Island
  - +Washington Heights

Model: Queens

- F1 Score: 0.96
- Significant Variables:
  - -East Elmhurst
  - -JFK Airport
  - -LaGuardia Airport
  - -Flushing Meadows-Corona Park
  - -Queensbridge/Ravenswood
  - -Saint Michaels Cemetery/Woodside:

Model: EWR

No significant variables due to small data size.

- Likelihood of a ride being from Uber/Lyft or Yellow taxi is not significantly influenced by the distance or fare -> influenced by the Borough of origin.
- Manhattan: Increase the availability of Yellow Taxis in the neighborhoods with significant coefficients where Uber and Lyft have higher presence.
- Queens: Focus on providing better services and convenience for passengers traveling from these locations, specially the airports.

Recommendations

Dynamic Pricing Strategy:

-> Zones where yellow taxis are charging less, consider allocating more taxis to meet the potential increase in demand.

-> Match their prices to Uber and Lyft

Optimize Service Coverage and Availability:

-> Allocate cabs in areas with high demand, such as Central Park, Upper East Side North, and Upper East Side South.

-> Focus on peak hours (7 am to 4 pm) in top zones

Target Key Boroughs and Zones:

-> Prioritize service in Manhattan and Queens, where rides are more likely to be Yellow Taxi.

-> Enhance service quality and convenience for passengers traveling to and from statistically significant zones: Central Harlem, as well as JFK Airport and LaGuardia Airport, since these are areas that Uber/lyft dominate.

Data-driven Resource Allocation:

-> Use the analysis of day-wise and hour-wise demand to allocate resources more efficiently.

-> Monitor and analyze trip data to identify emerging trends and adjust resource allocation