Maximizing Fleet Utilization Through Data-Driven Demand Forecasting

NYC Yellow Taxi Analysis (Jan–Nov 2024)

Cheng Li

Feb 7th 2025

Executive Summary

- **Goal:**
  - Predict taxi demand to enhance ride-sharing resource allocation.
- **Approach:**
  - Leveraged NYC taxi data and advanced ML techniques.
- **Key Outcomes:**
  - Peak demand hours and high-demand zones identified.
  - Predictive modeling enhances fleet distribution
  - Actionable insights for dynamic pricing and fleet optimization.

Business Problem & Objectives

- Current Challenge:
  - Unused taxis during low-demand periods vs. high wait times during peak hours.
- Impact: Lost revenue, inefficient operations, and customer dissatisfaction.
- Objectives:
  - Predict ride demand effectively.
  - Strategically position vehicles based on demand patterns.
  - Maximize profitability through optimal fleet allocation.

Data Overview

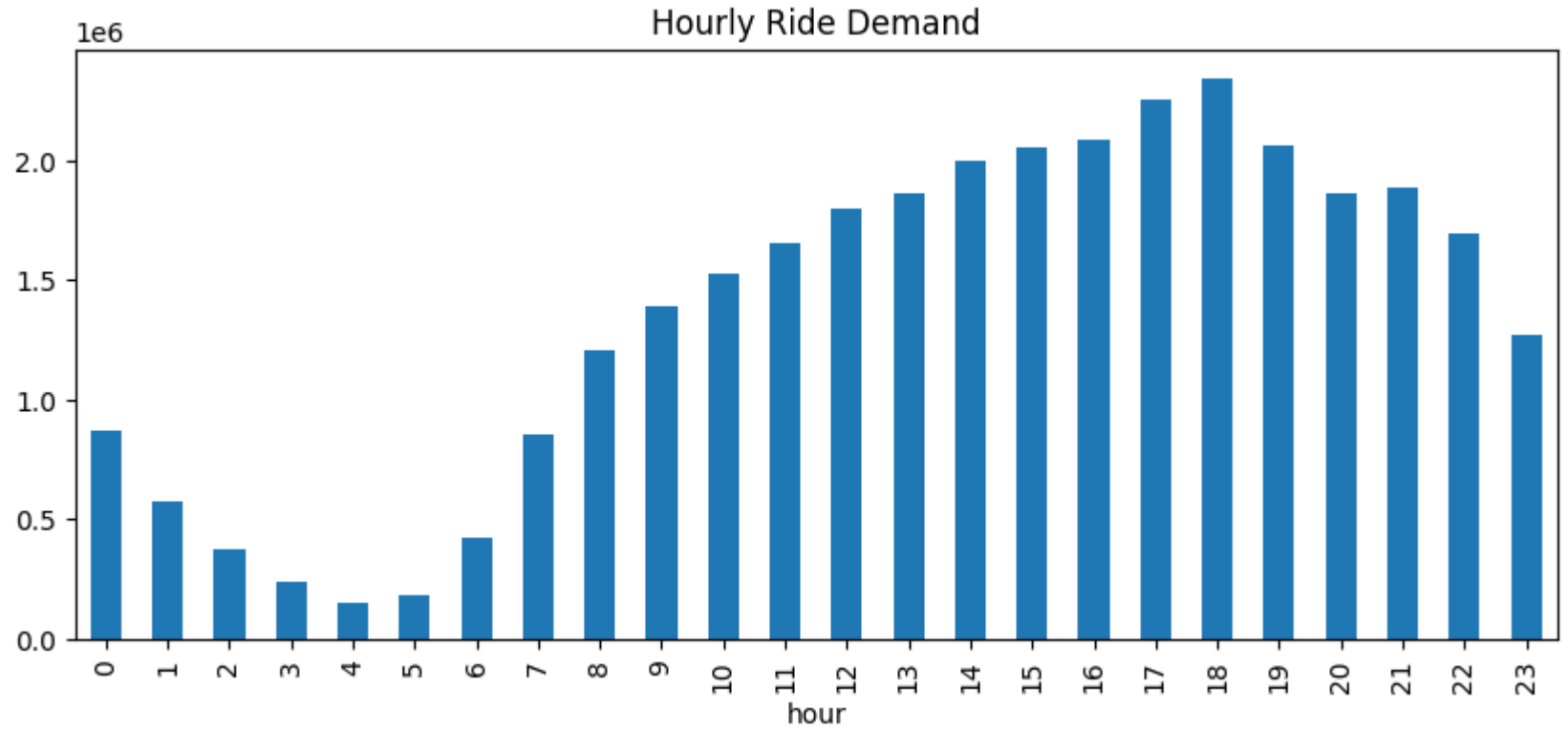
- **Data Source:** NYC Yellow Taxi trip data (Jan–Nov 2024).

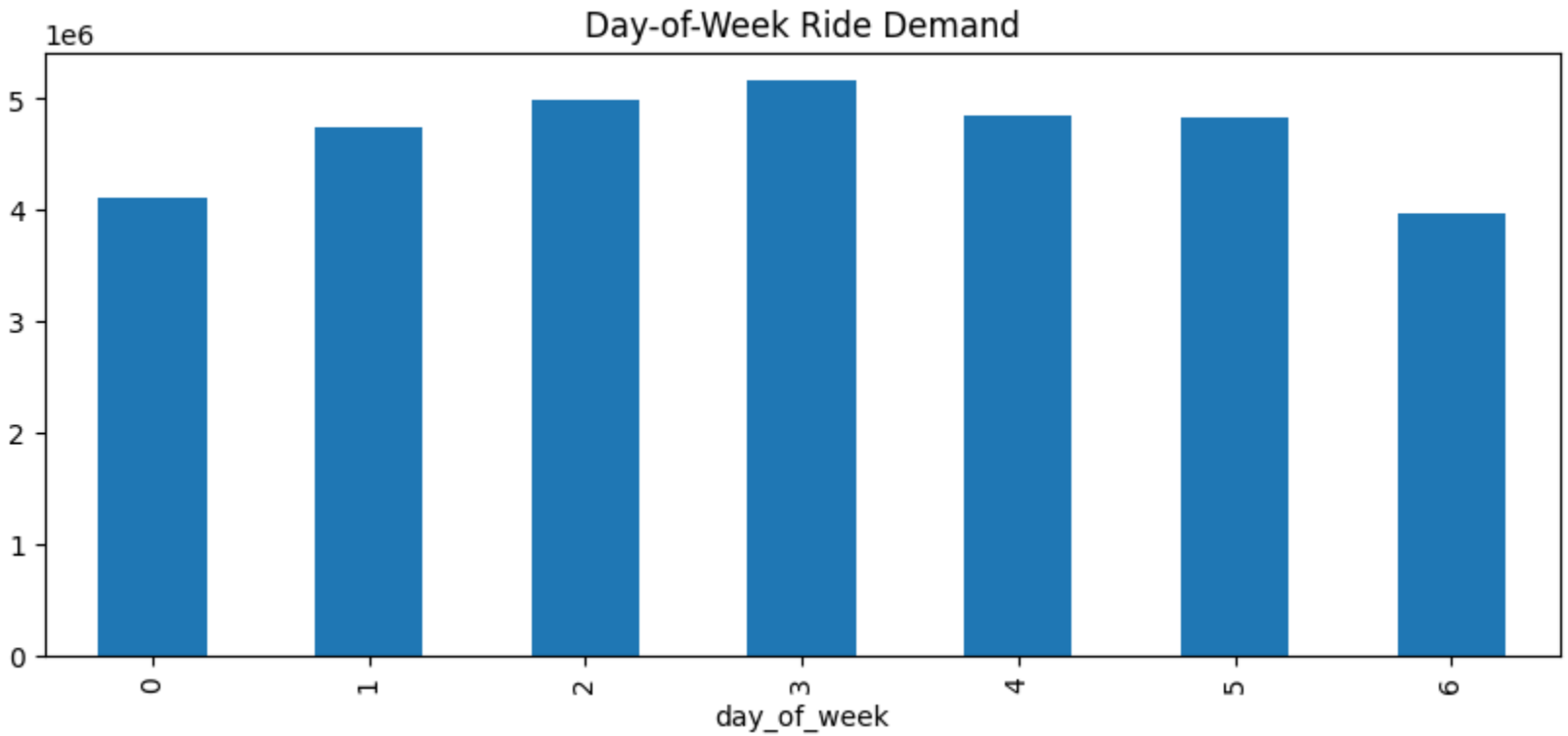
- **Methodology:**
  - Historical demand trends analyzed.
  - Predictive modeling (XGBoost, LSTM, Random Forest) used for forecasting.
  - Insights applied to optimize fleet allocation.

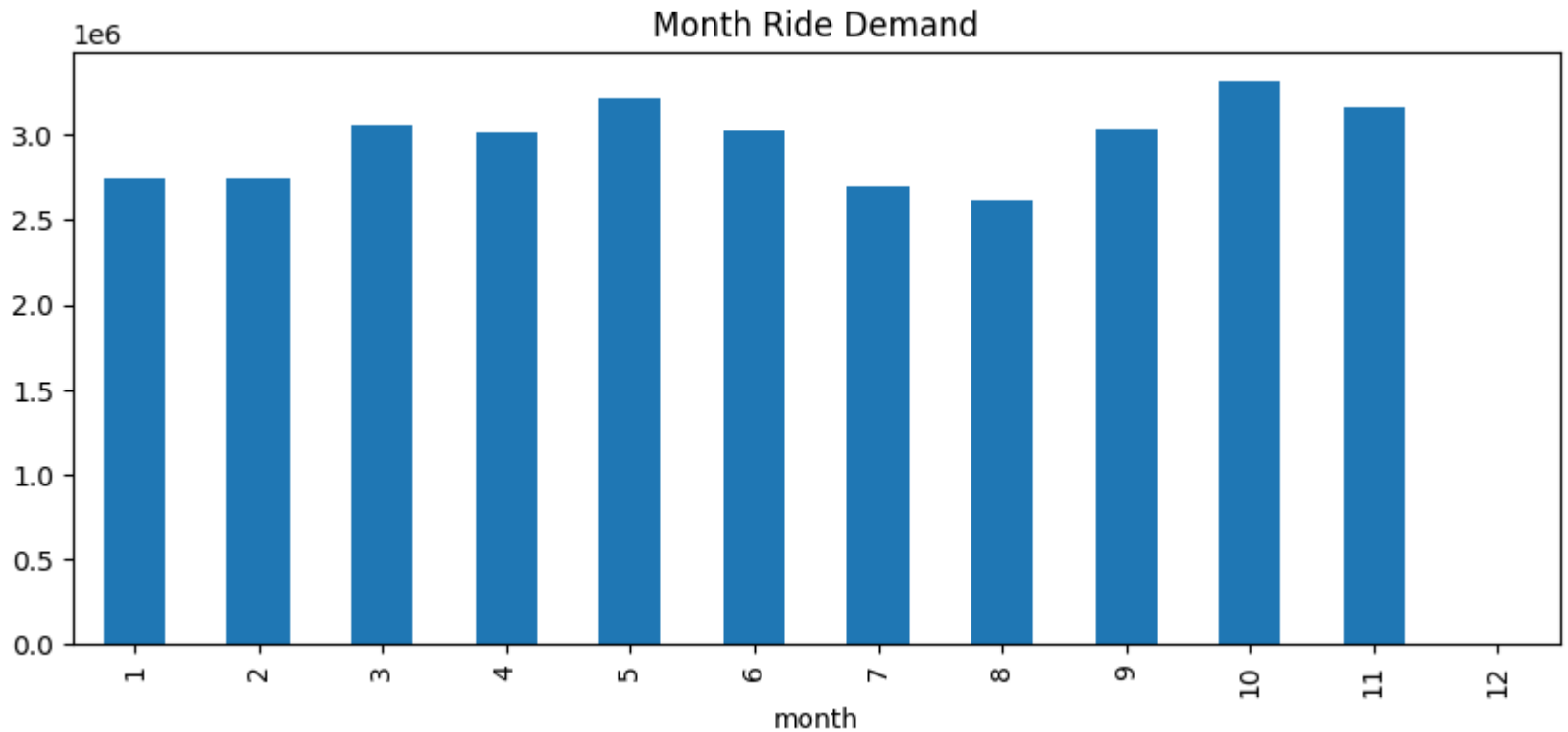
Key findings

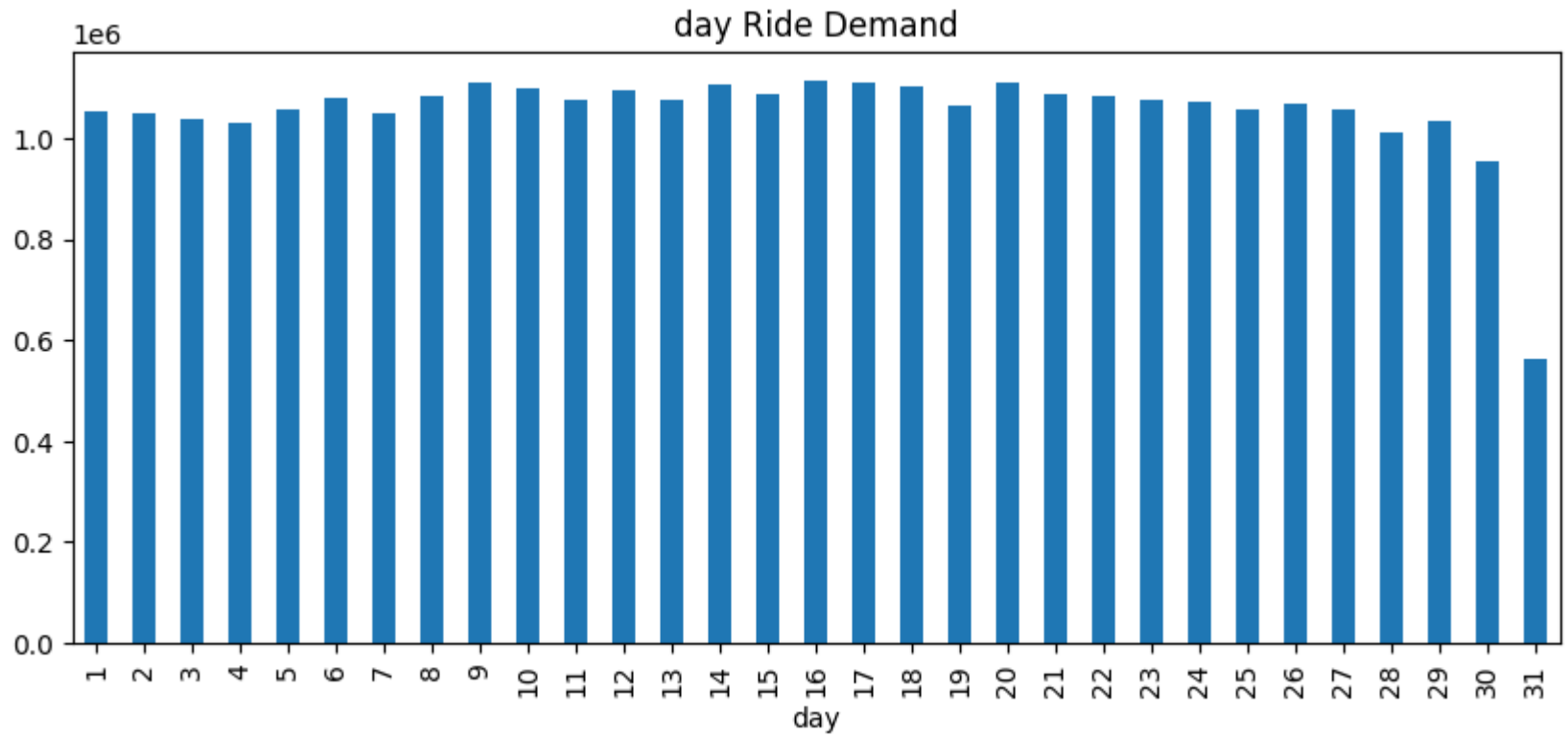
- **Heatmaps:**
  - Demand variations by time and locations.
  - High-demand pickup zones.
- **Peak vs. Off-Peak Demand Trends:**
  - Identifying critical hours for maximum efficiency.

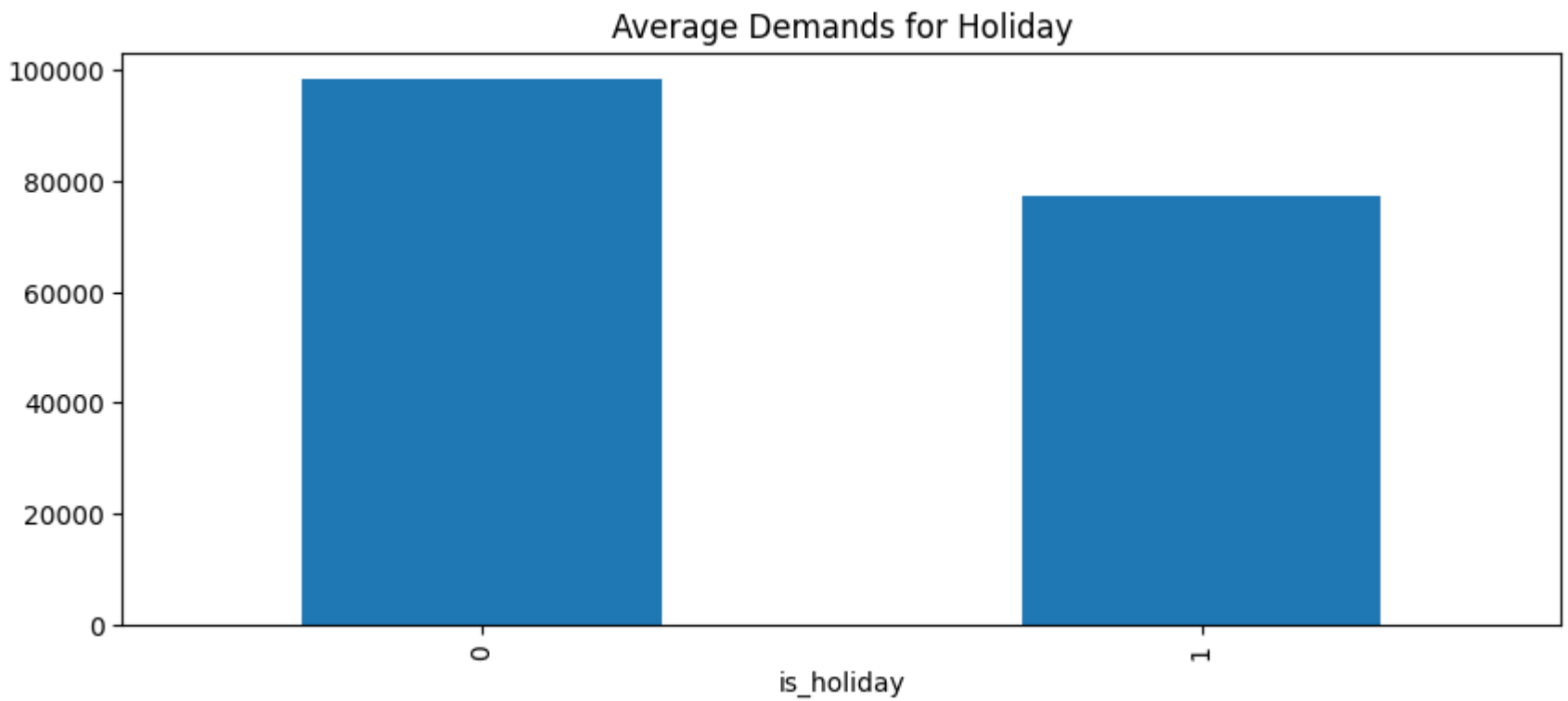
Demand variations by time



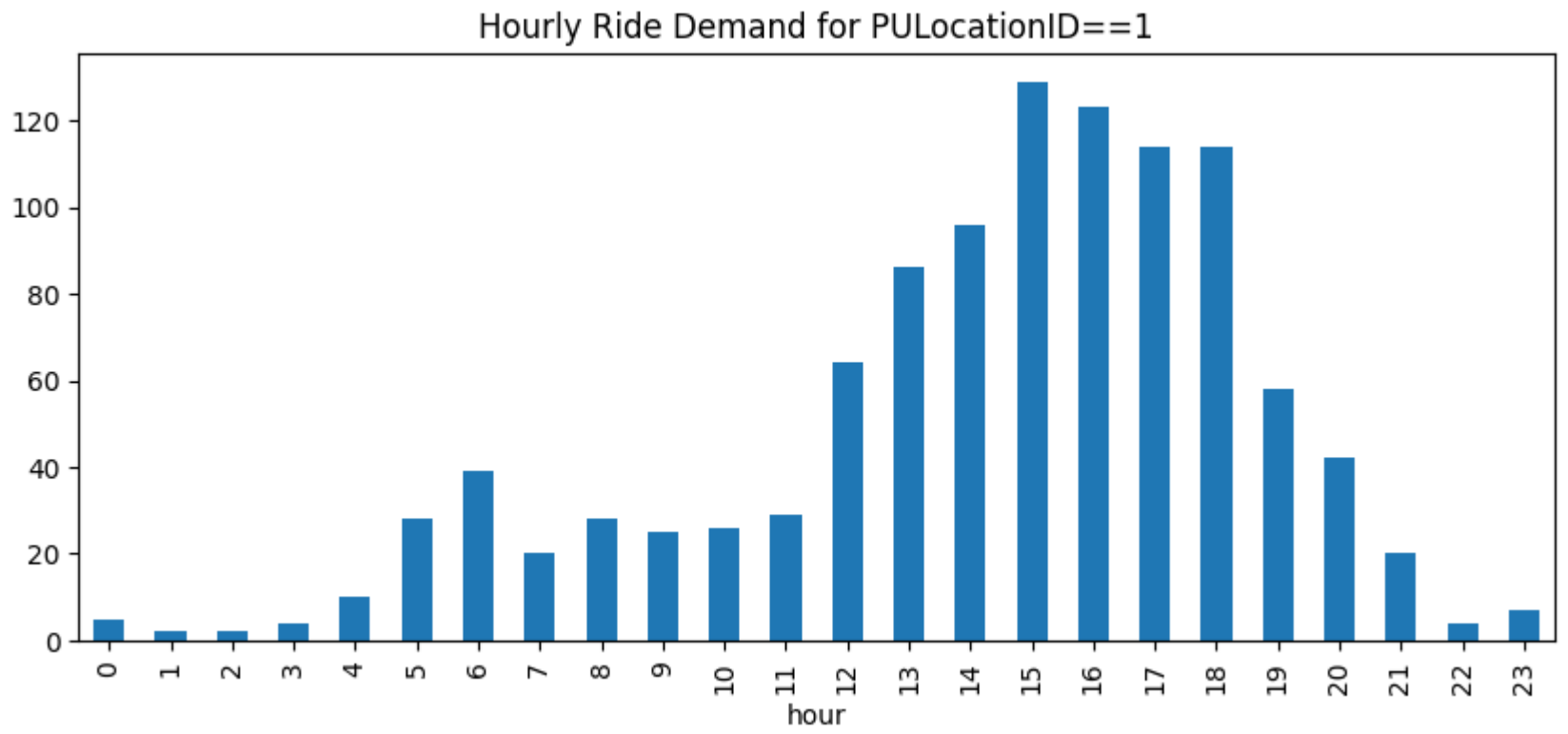

Day-of-Week Ride Demand
1e6
5
4
3
2
1
0
0
1
2
3
4
5
6
day_of_week

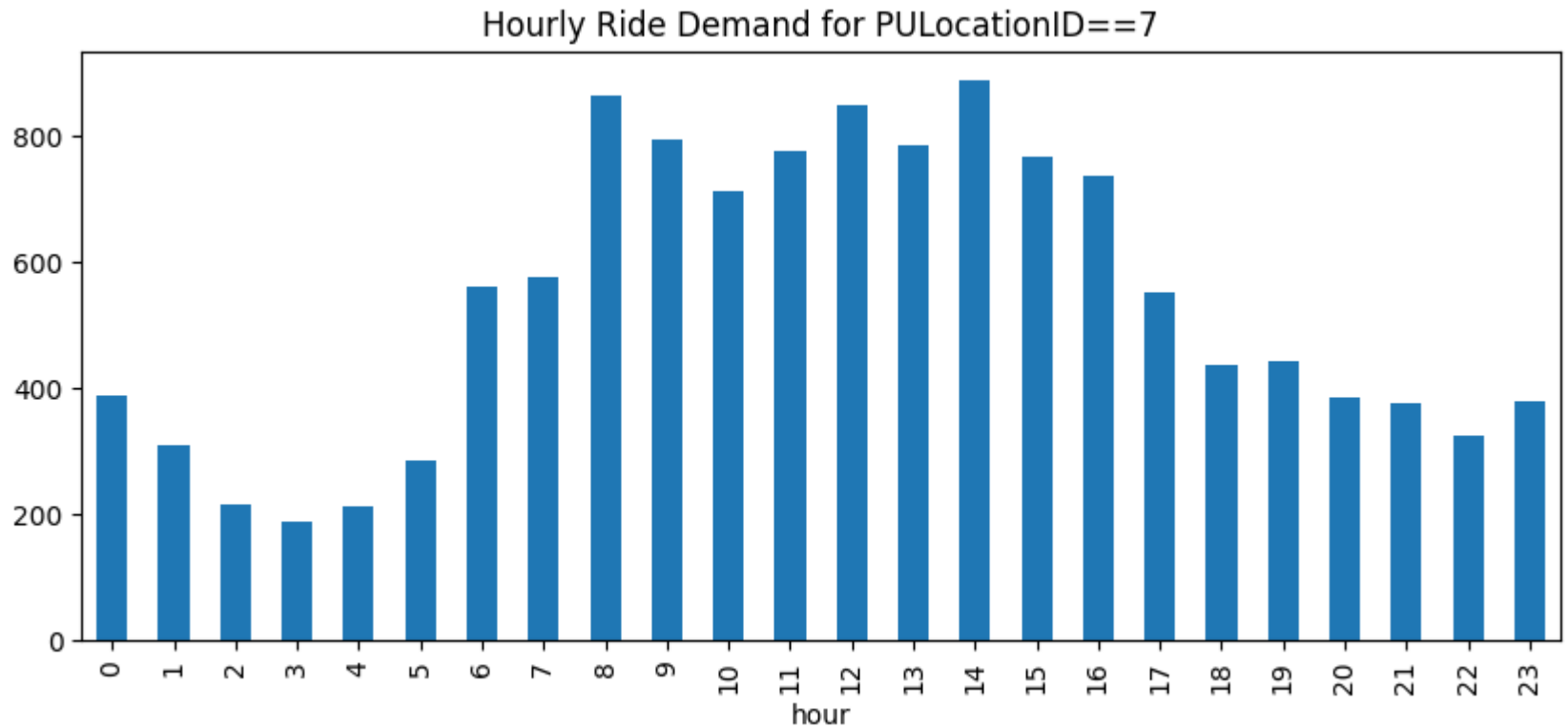

Month Ride Demand
1e6
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1
2
3
4
5
6
7
8
9
10
11
12
month

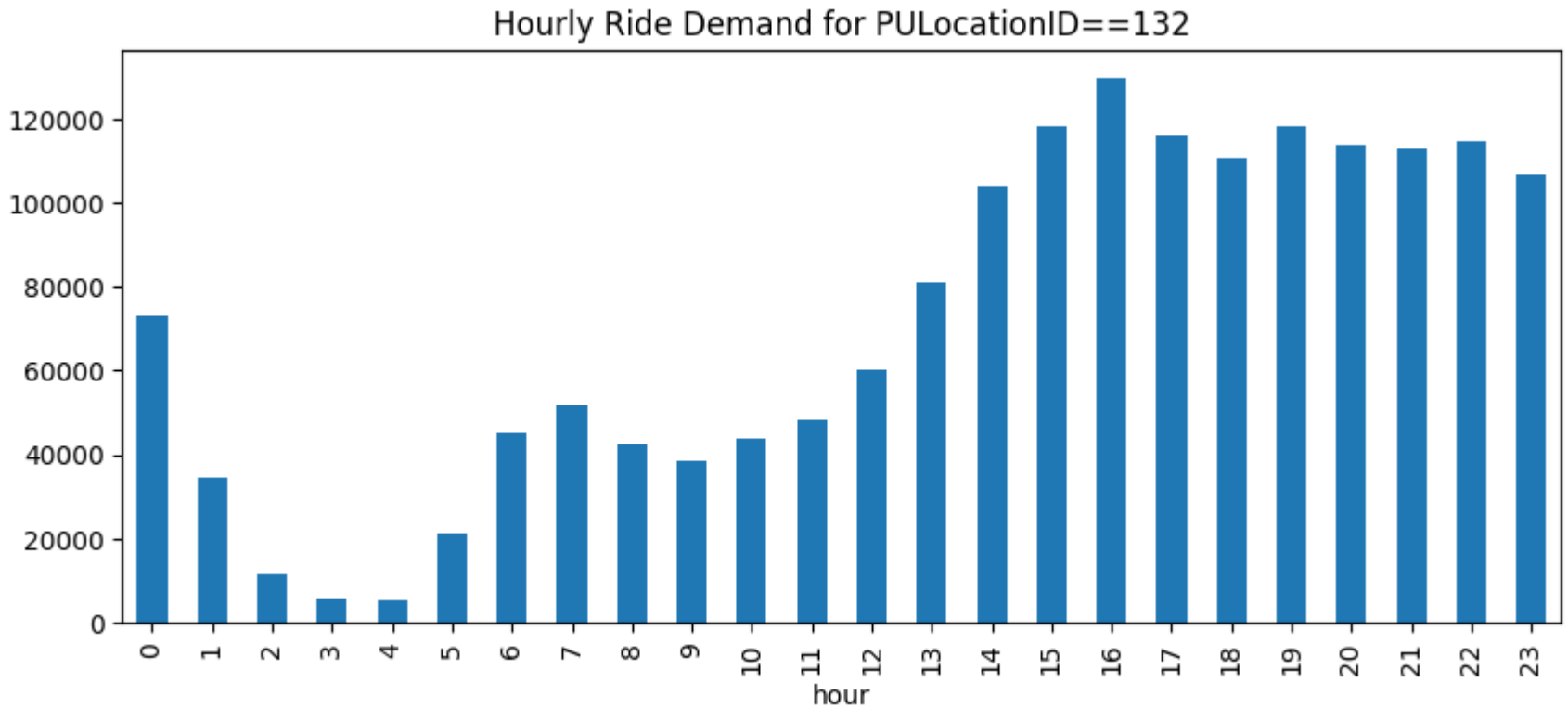

day Ride Demand
1e6
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
day


Average Demands for Holiday
100000
80000
60000
40000
20000
0
0
1
is_holiday

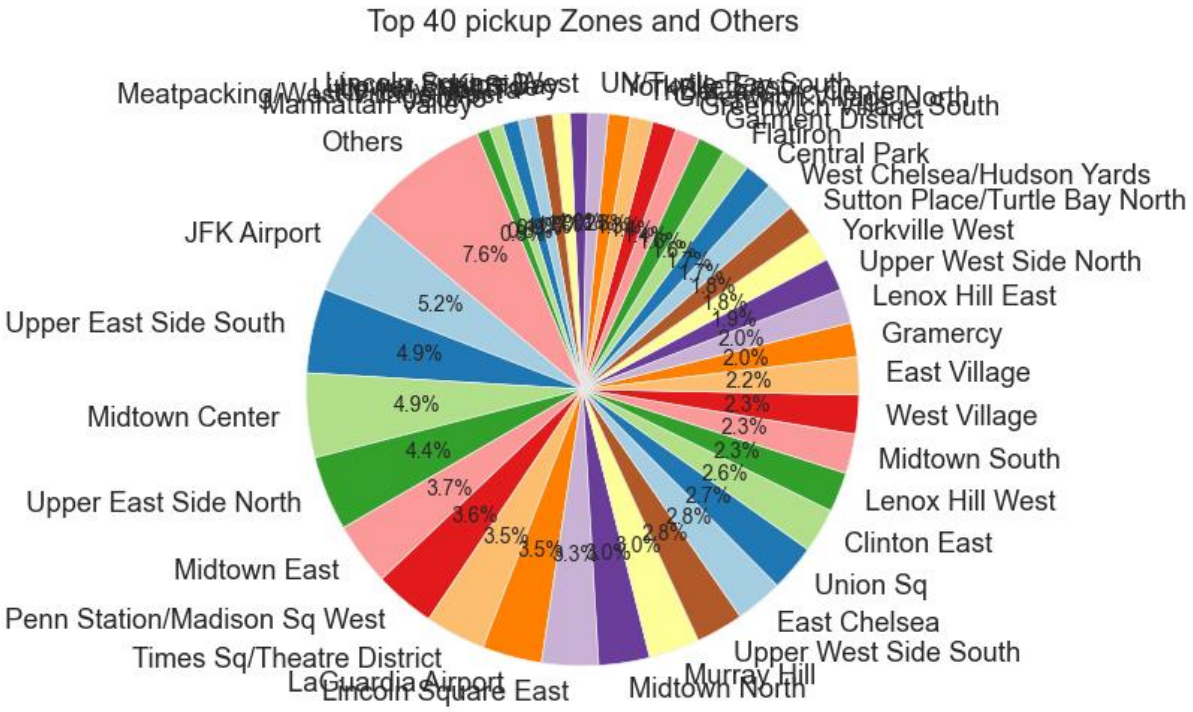
Demand variation by locations

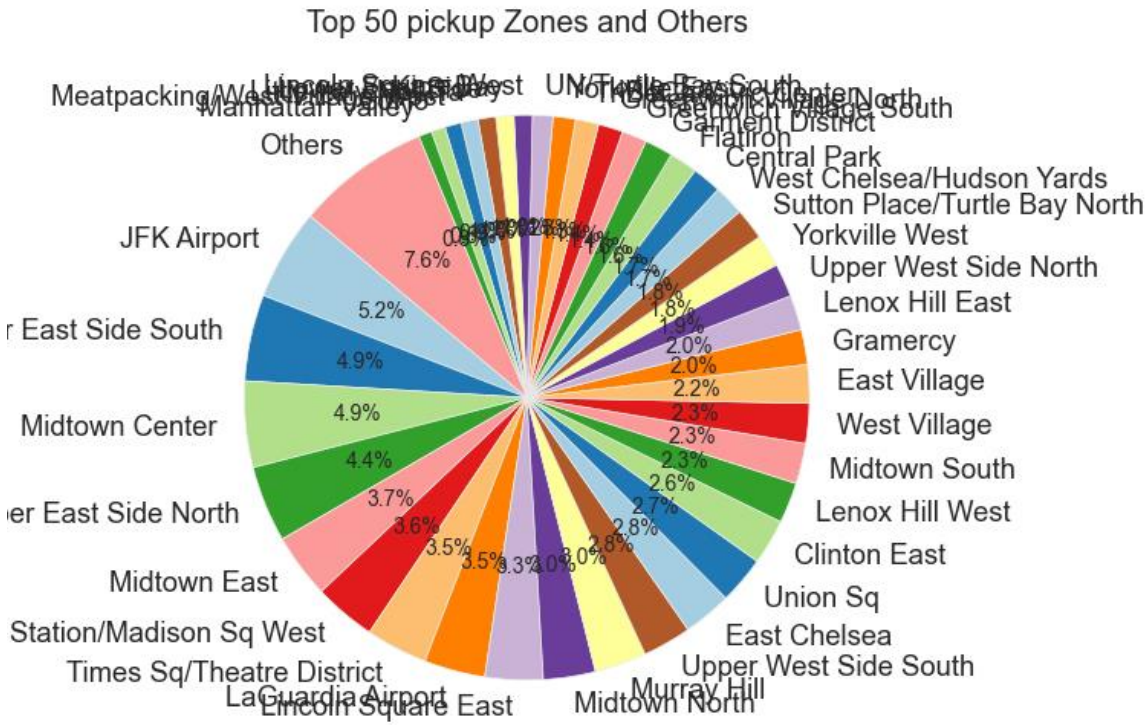



Hourly Ride Demand for PULocationID==7
800
600
400
200
0
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
hour

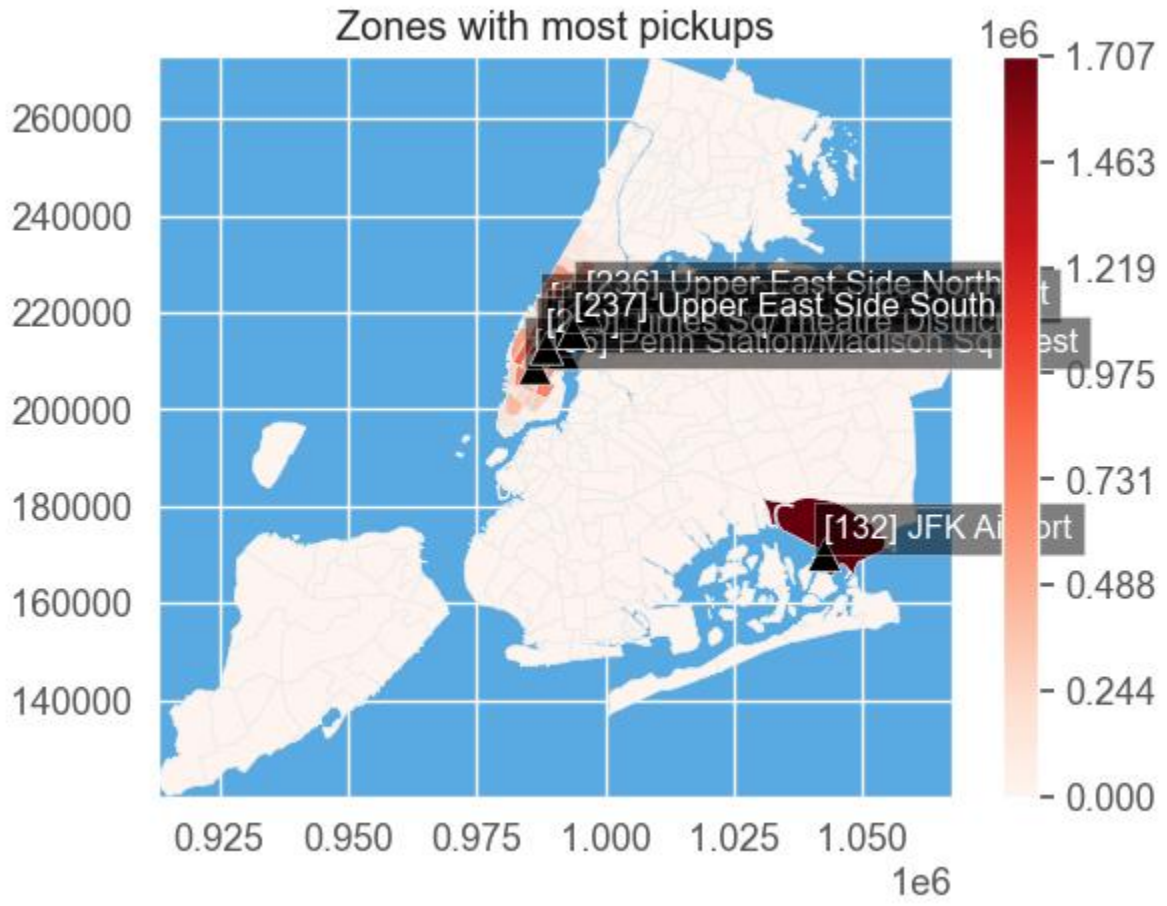

Hourly Ride Demand for PULocationID==132
120000
100000
80000
60000
40000
20000
0
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
hour

Ride demand concentrated

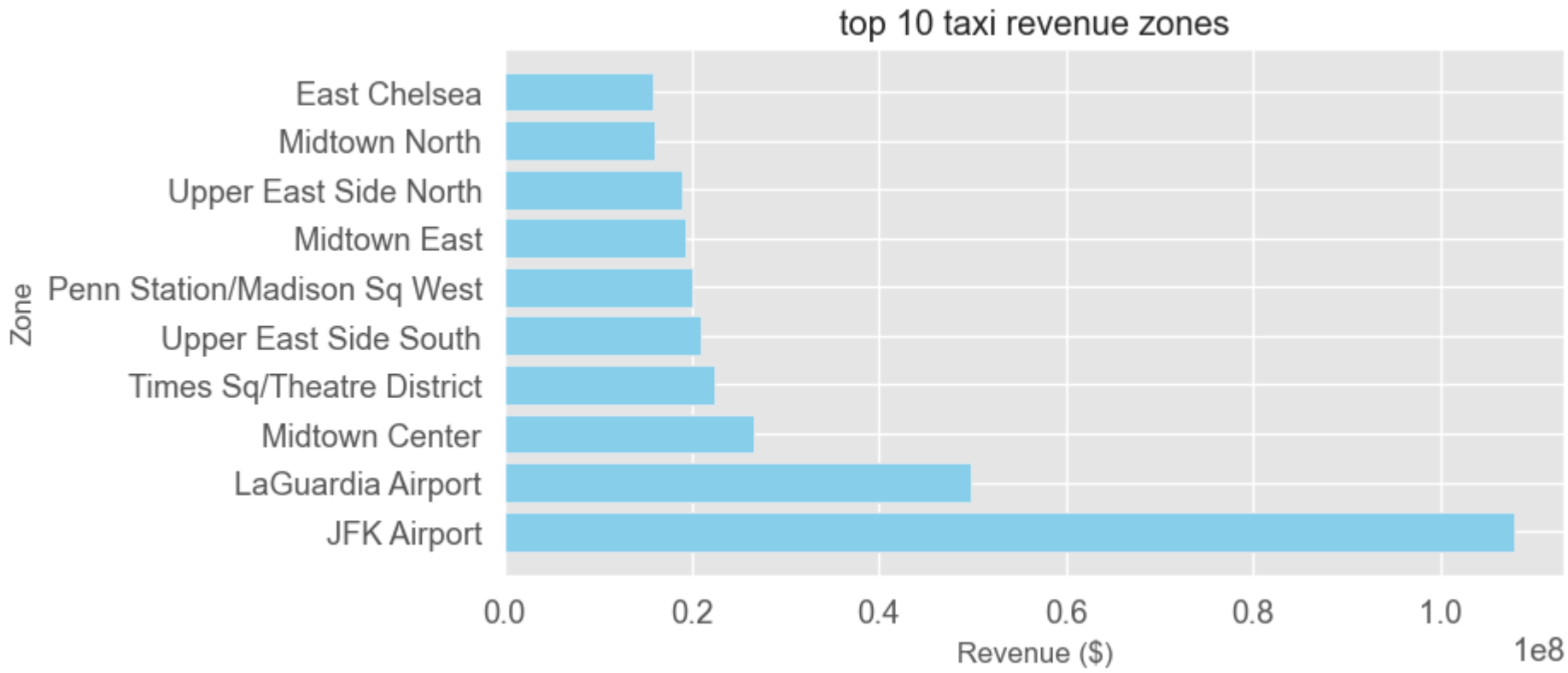




Map for top 10 pickup zones



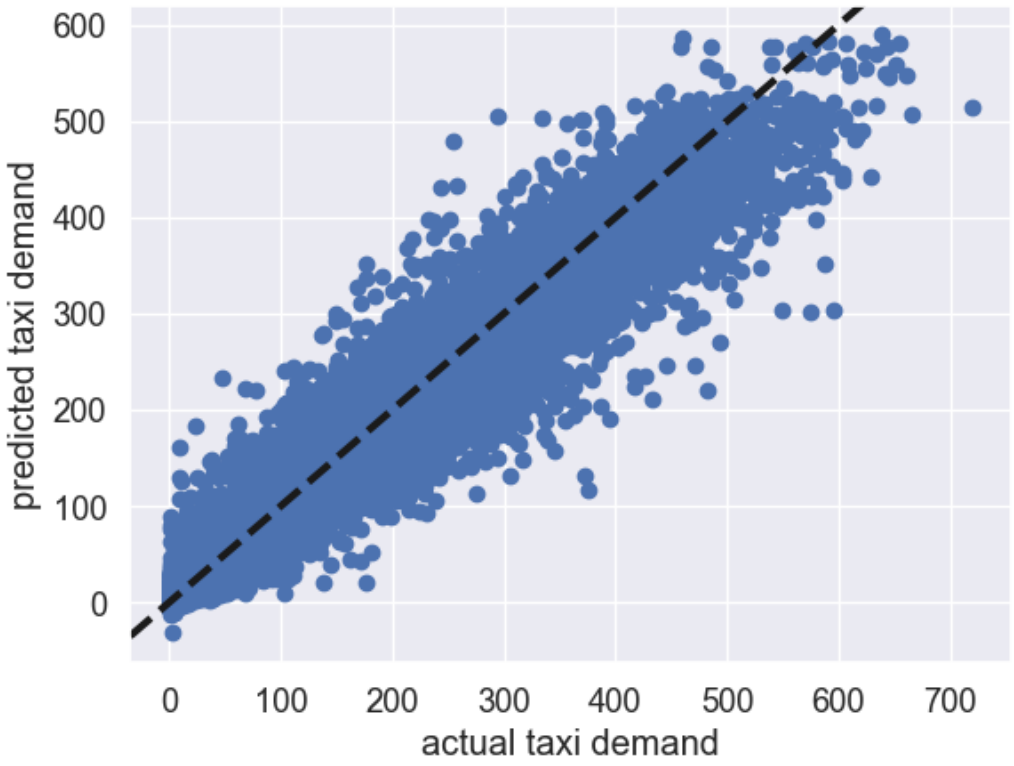
Top 10 revenue zones

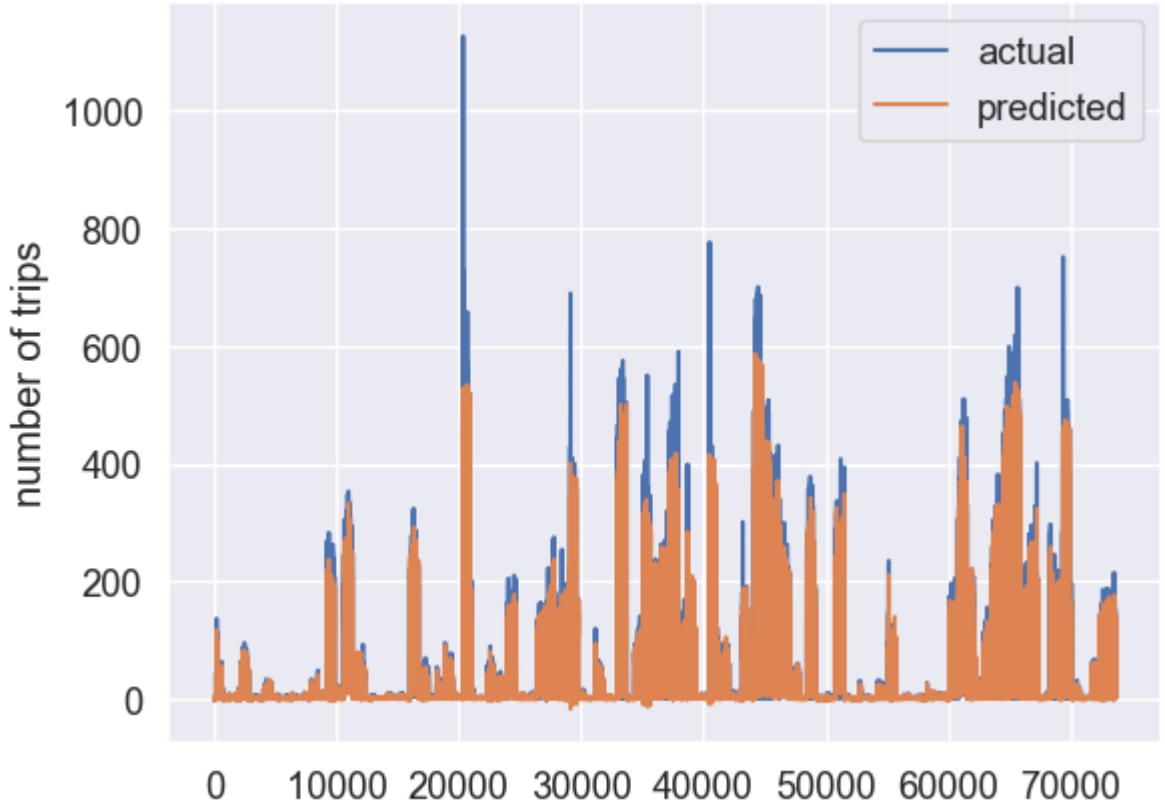


Predictive Model for Demand Forecasting

- **Model Used:** machine learning model XGBoost for forecasting.
- **Input data:**
  - Historical taxi data (date, pickup location, fare amount, passenger counts etc)
  - Weather conditions (missing in this case study)
  - Event data(missing in this case study)
- **Output**
  - Predicted hourly ride demand for all locations.

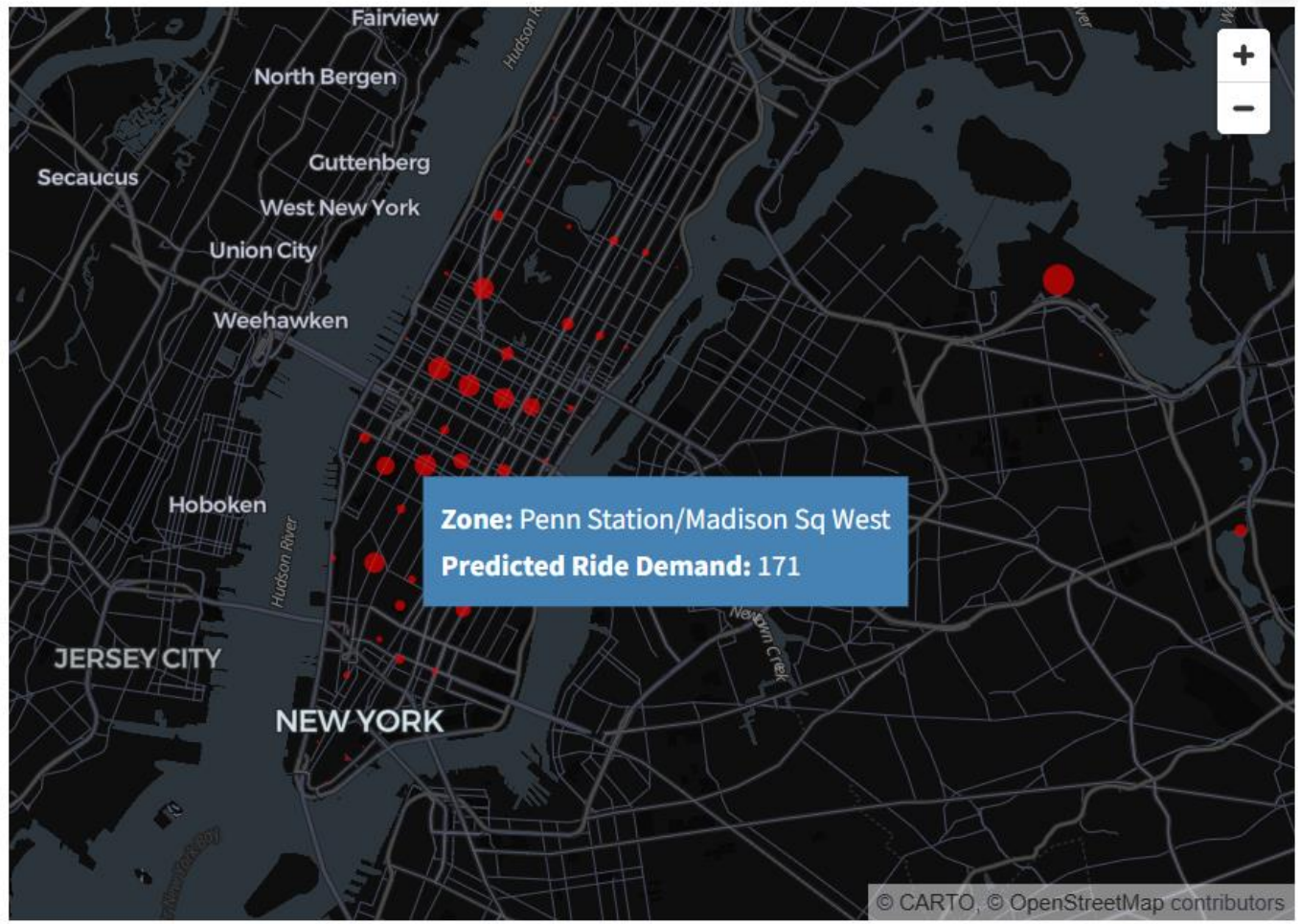
Model performance

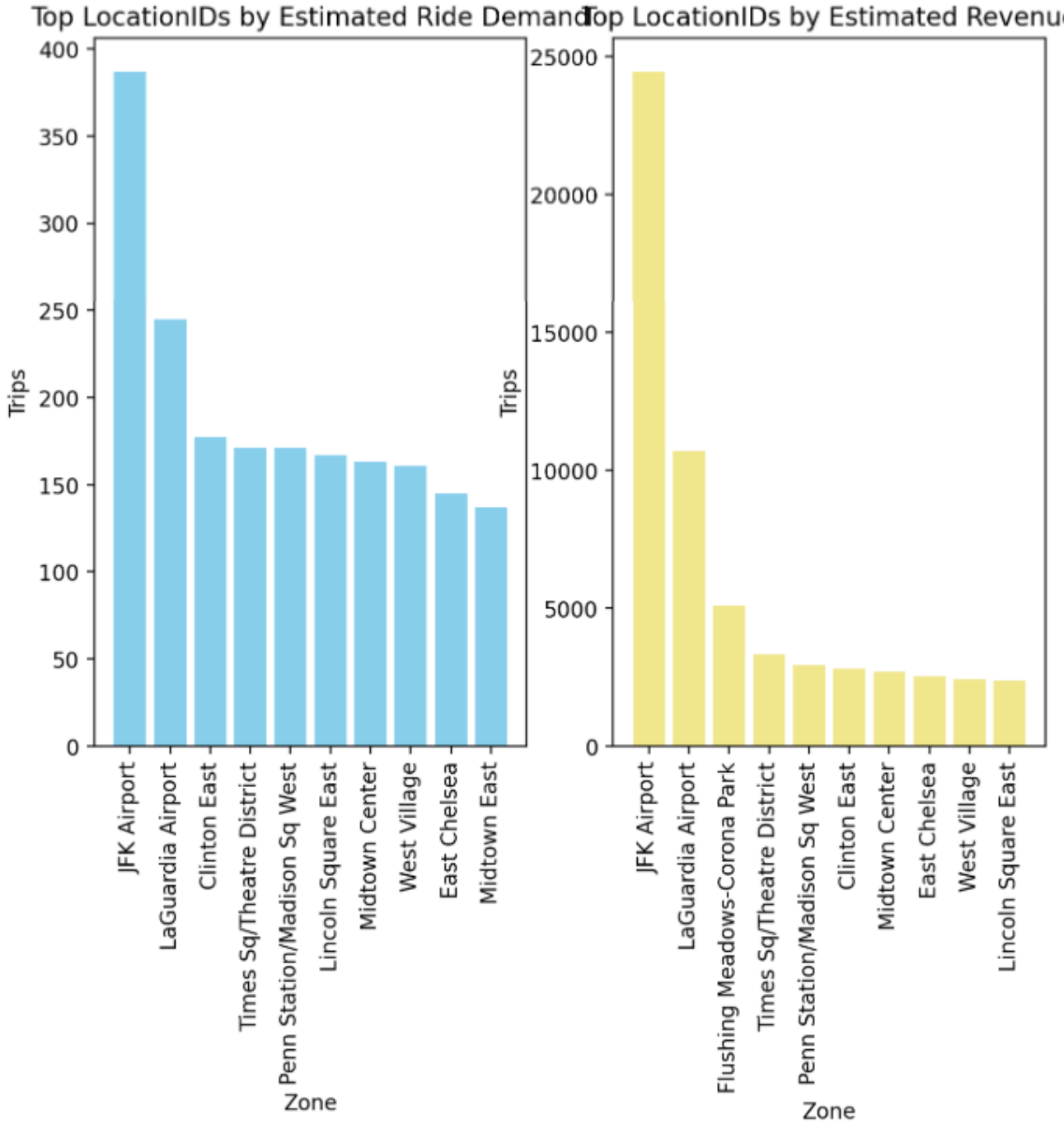




High demand also depends on other external factors, such as weather conditions, city events etc.

Overall Ride Demand Prediction Map





Top 10 locations by predicted ride demand and revenue

Revenue Impact (if we have data)

- **Revenue Uplift Potential:**
  - Optimizing fleet allocation can increase revenue by **X%**.
  - Reducing idle time could save **$Y million/year**.

- **Comparison of Scenarios:**
  - Static fleet vs. Data-driven fleet deployment.

Actionable Recommendations

Dynamic Pricing Strategy – Adjust fares based on demand.

Optimized Fleet Deployment – Send more taxis to hotspots.

Real-Time Monitoring Dashboard – Implement a data-driven decision-making tool.

Data-driven fleet optimization isn’t just an idea—it’s a competitive advantage.

Q & A