# A Machine Learning Approach to Forecasting Ride-Hailing Demand Using Geospatial and Weather Data

**Project Focus: Data-driven analysis of the NYC ride-hailing market.**

**Dataset Size:** Based on 14.3 million trip records.

**Key Finding 1:** Manhattan dominates the market.

**Outcome:** A machine learning model was built and validated.

**Model Capability:** Forecasts ride-hailing demand.

**Benefit:** Provides a powerful tool for strategic planning and operational efficiency.

# uber-raw-data-janjune-15.csv

**uber-raw-data-janjune-15.csv** (551.67 MB)

Detail | Compact | Column

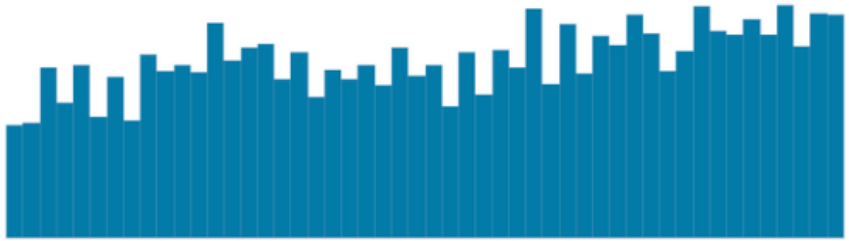
4 of 4 columns

## Dispatching_base_num

| | |
|---|---|
| B02764 | 40% |
| B02682 | 24% |
| Other (5032296) | 35% |

| | | |
|---|---|---|
| Valid | 14.3m | 100% |
| Mismatched | 0 | 0% |
| Missing | 0 | 0% |
| Unique | 8 | |
| Most Common | B02764 | 40% |

## Pickup_date



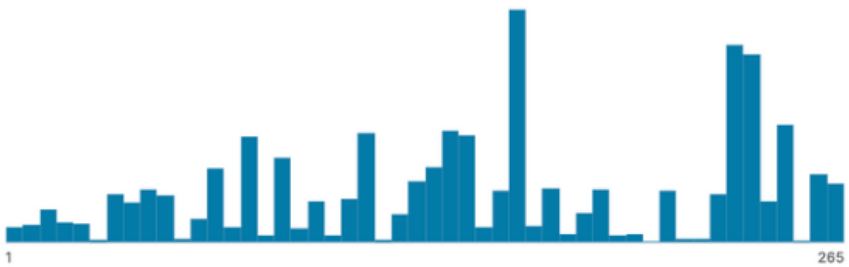

| | | |
|---|---|---|
| Valid | 14.3m | 100% |
| Mismatched | 0 | 0% |
| Missing | 0 | 0% |
| Minimum | 1Jan15 | |
| Mean | 7Apr15 | |
| Maximum | 1Jul15 | |

## Affiliated_base_num

| | |
|---|---|
| B02764 | 30% |
| B02682 | 24% |
| Other (6469460) | 45% |

| | | |
|---|---|---|
| Valid | 14.1m | 99% |
| Mismatched | 0 | 0% |
| Missing | 162k | 1% |
| Unique | 284 | |
| Most Common | B02764 | 30% |

## locationID




| | | |
|---|---|---|
| Valid | 14.3m | 100% |
| Mismatched | 0 | 0% |
| Missing | 0 | 0% |
| Mean | 152 | |
| Std. Deviation | 71.6 | |
| Quantiles | 1 | Min |
| | 265 | Max |

| | Dispatching_base_num | Affiliated_base_num | locationID | pickup_date | pickup_time | pickup_day_of_week |
|---|---|---|---|---|---|---|
| 2 | B02617 | B02617 | 141 | 5/17/2015 | 9:47:00 | Sunday |
| 3 | B02617 | B02617 | 65 | 5/17/2015 | 9:47:00 | Sunday |
| 4 | B02617 | B02617 | 100 | 5/17/2015 | 9:47:00 | Sunday |
| 5 | B02617 | B02774 | 80 | 5/17/2015 | 9:47:00 | Sunday |
| 6 | B02617 | B02617 | 90 | 5/17/2015 | 9:47:00 | Sunday |
| 7 | B02617 | B02617 | 228 | 5/17/2015 | 9:47:00 | Sunday |
| 8 | B02617 | B02617 | 7 | 5/17/2015 | 9:47:00 | Sunday |
| 9 | B02617 | B02764 | 74 | 5/17/2015 | 9:47:00 | Sunday |
| 10 | B02617 | B02617 | 249 | 5/17/2015 | 9:47:00 | Sunday |
| 11 | B02617 | B02764 | 22 | 5/17/2015 | 9:47:00 | Sunday |
| 12 | B02617 | B02617 | 263 | 5/17/2015 | 9:48:00 | Sunday |
| 13 | B02617 | B02617 | 61 | 5/17/2015 | 9:48:00 | Sunday |
| 14 | B02617 | B02617 | 229 | 5/17/2015 | 9:49:00 | Sunday |
| 15 | B02617 | B02617 | 164 | 5/17/2015 | 9:49:00 | Sunday |
| 16 | B02617 | B02617 | 237 | 5/17/2015 | 9:49:00 | Sunday |
| 17 | B02617 | B02617 | 142 | 5/17/2015 | 9:49:00 | Sunday |
| 18 | B02617 | B02617 | 188 | 5/17/2015 | 9:49:00 | Sunday |
| 19 | B02617 | B02617 | 237 | 5/17/2015 | 9:49:00 | Sunday |
| 20 | B02617 | B02617 | 224 | 5/17/2015 | 9:49:00 | Sunday |
| 21 | B02617 | B02617 | 238 | 5/17/2015 | 9:49:00 | Sunday |
| 22 | B02617 | B02682 | 242 | 5/17/2015 | 9:49:00 | Sunday |
| 23 | B02617 | B02617 | 95 | 5/17/2015 | 9:50:00 | Sunday |
| 24 | B02617 | B02617 | 141 | 5/17/2015 | 9:50:00 | Sunday |
| 25 | B02617 | B02617 | 236 | 5/17/2015 | 9:50:00 | Sunday |
| 26 | B02617 | B02617 | 233 | 5/17/2015 | 9:50:00 | Sunday |
| 27 | B02617 | B02617 | 230 | 5/17/2015 | 9:50:00 | Sunday |
| 28 | B02617 | B02617 | 162 | 5/17/2015 | 9:50:00 | Sunday |
| 29 | B02617 | B02764 | 234 | 5/17/2015 | 9:50:00 | Sunday |
| 30 | B02617 | B02617 | 161 | 5/17/2015 | 9:50:00 | Sunday |

| | LocationID | Borough | Zone | service_zone | latitude | longitude |
|---|---|---|---|---|---|---|
| 1 | 1 | EWR | Newark Airport | EWR | 40.69287997 | -74.18544993 |
| 2 | 2 | Queens | Jamaica Bay | Boro Zone | 40.6057 | -73.8713 |
| 3 | 3 | Bronx | Allerton/Pelham Gardens | Boro Zone | 40.86521003 | -73.8435548 |
| 4 | 4 | Manhattan | Alphabet City | Yellow Zone | 40.72599 | -73.98057 |
| 5 | 5 | Staten Island | Arden Heights | Boro Zone | 40.5564 | -74.1735 |
| 6 | 6 | Staten Island | Arrochar/Fort Wadsworth | Boro Zone | 40.5927 | -74.07 |
| 7 | 7 | Queens | Astoria | Boro Zone | 40.7644 | -73.9235 |
| 8 | 8 | Queens | Astoria Park | Boro Zone | 40.7785 | -73.9228 |
| 9 | 9 | Queens | Auburndale | Boro Zone | 40.7578 | -73.7834 |
| 10 | 10 | Queens | Baisley Park | Boro Zone | 40.6738 | -73.786 |

| | Dispatching_base_num | Pickup_date | Affiliated_base_num | locationID |
|---|---|---|---|---|
| 2 | B02617 | 5/17/2015 9:47 | B02617 | 141 |
| 3 | B02617 | 5/17/2015 9:47 | B02617 | 65 |
| 4 | B02617 | 5/17/2015 9:47 | B02617 | 100 |
| 5 | B02617 | 5/17/2015 9:47 | B02774 | 80 |
| 6 | B02617 | 5/17/2015 9:47 | B02617 | 90 |
| 7 | B02617 | 5/17/2015 9:47 | B02617 | 228 |
| 8 | B02617 | 5/17/2015 9:47 | B02617 | 7 |
| 9 | B02617 | 5/17/2015 9:47 | B02764 | 74 |
| 10 | B02617 | 5/17/2015 9:47 | B02617 | 249 |

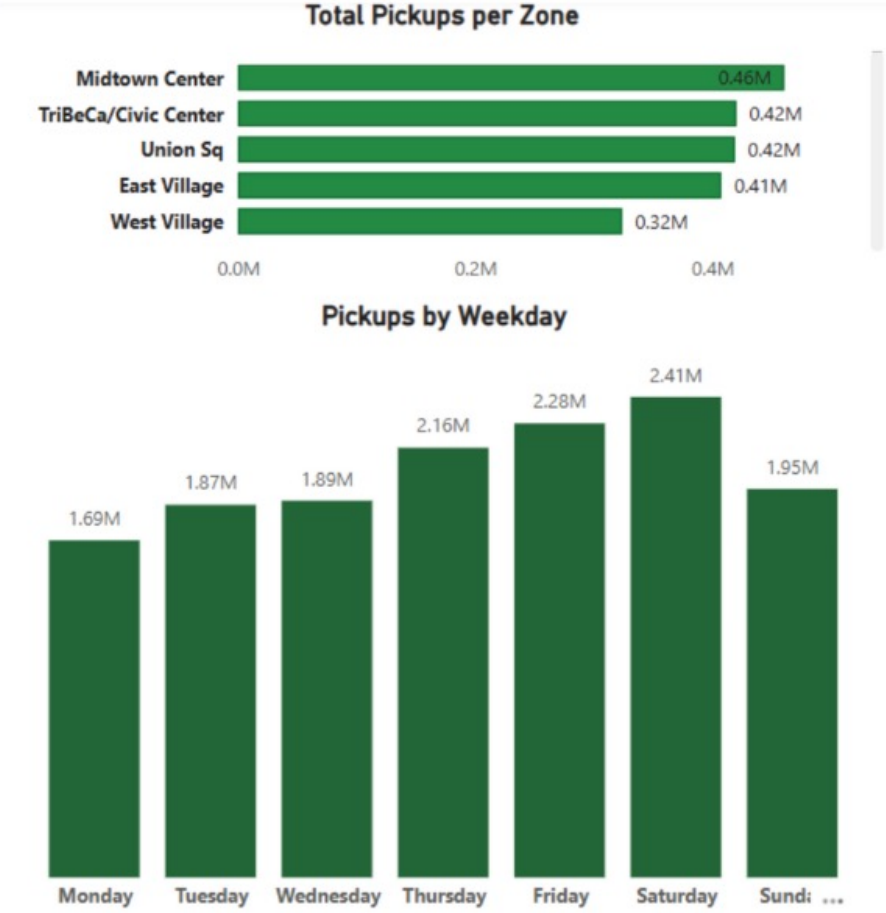
| | temperature | humidity | wind_speed | precipitation | pressure | date | hour | weather |
|---|---|---|---|---|---|---|---|---|
| 1 | -3.7 | 50 | 13.0 | 0.0 | 1018.3 | 2015-01-01 | 0 | Clear Sky |
| 2 | -3.8 | 49 | 14.1 | 0.0 | 1017.8 | 2015-01-01 | 1 | Clear Sky |
| 3 | -3.9 | 49 | 14.1 | 0.0 | 1017.1 | 2015-01-01 | 2 | Clear Sky |
| 4 | -4.0 | 52 | 14.7 | 0.0 | 1016.7 | 2015-01-01 | 3 | Clear Sky |
| 5 | -4.0 | 52 | 15.6 | 0.0 | 1016.2 | 2015-01-01 | 4 | Clear Sky |
| 6 | -4.0 | 50 | 16.1 | 0.0 | 1016.2 | 2015-01-01 | 5 | Clear Sky |
| 7 | -4.1 | 48 | 16.2 | 0.0 | 1016.5 | 2015-01-01 | 6 | Clear Sky |
| 8 | -4.2 | 46 | 13.5 | 0.0 | 1017.0 | 2015-01-01 | 7 | Clear Sky |
| 9 | -4.4 | 45 | 10.0 | 0.0 | 1017.4 | 2015-01-01 | 8 | Clear Sky |
| 10 | -4.3 | 45 | 10.6 | 0.0 | 1017.5 | 2015-01-01 | 9 | Clear Sky |
| 11 | -3.0 | 38 | 11.9 | 0.0 | 1017.8 | 2015-01-01 | 10 | Clear Sky |
| 12 | -1.7 | 32 | 17.9 | 0.0 | 1018.0 | 2015-01-01 | 11 | Clear Sky |
| 13 | -0.7 | 30 | 19.3 | 0.0 | 1016.7 | 2015-01-01 | 12 | Clear Sky |
| 14 | 0.3 | 29 | 21.3 | 0.0 | 1015.4 | 2015-01-01 | 13 | Clear Sky |
| 15 | 1.0 | 29 | 20.9 | 0.0 | 1013.7 | 2015-01-01 | 14 | Clear Sky |
| 16 | 2.1 | 31 | 21.3 | 0.0 | 1012.4 | 2015-01-01 | 15 | Mainly Clear |
| 17 | 2.4 | 35 | 20.5 | 0.0 | 1011.9 | 2015-01-01 | 16 | Clear Sky |
| 18 | 1.9 | 42 | 17.1 | 0.0 | 1011.6 | 2015-01-01 | 17 | Clear Sky |
| 19 | 0.9 | 47 | 14.3 | 0.0 | 1011.5 | 2015-01-01 | 18 | Clear Sky |
| 20 | 0.5 | 51 | 17.9 | 0.0 | 1011.6 | 2015-01-01 | 19 | Clear Sky |
| 21 | 0.3 | 53 | 19.4 | 0.0 | 1011.5 | 2015-01-01 | 20 | Clear Sky |
| 22 | 0.2 | 54 | 19.9 | 0.0 | 1011.7 | 2015-01-01 | 21 | Clear Sky |
| 23 | 0.1 | 55 | 19.6 | 0.0 | 1011.7 | 2015-01-01 | 22 | Clear Sky |
| 24 | 0.1 | 56 | 19.0 | 0.0 | 1011.9 | 2015-01-01 | 23 | Mainly Clear |
| 25 | -0.1 | 57 | 18.4 | 0.0 | 1012.7 | 2015-01-02 | 0 | Overcast |
| 26 | -0.3 | 58 | 15.6 | 0.0 | 1012.6 | 2015-01-02 | 1 | Overcast |
| 27 | -0.5 | 59 | 14.5 | 0.0 | 1012.7 | 2015-01-02 | 2 | Partly Cloudy |
| 28 | -0.5 | 62 | 14.2 | 0.0 | 1013.2 | 2015-01-02 | 3 | Partly Cloudy |
| 29 | -0.6 | 62 | 14.7 | 0.0 | 1013.8 | 2015-01-02 | 4 | Overcast |
| 30 | -0.6 | 61 | 14.6 | 0.0 | 1014.0 | 2015-01-02 | 5 | Overcast |

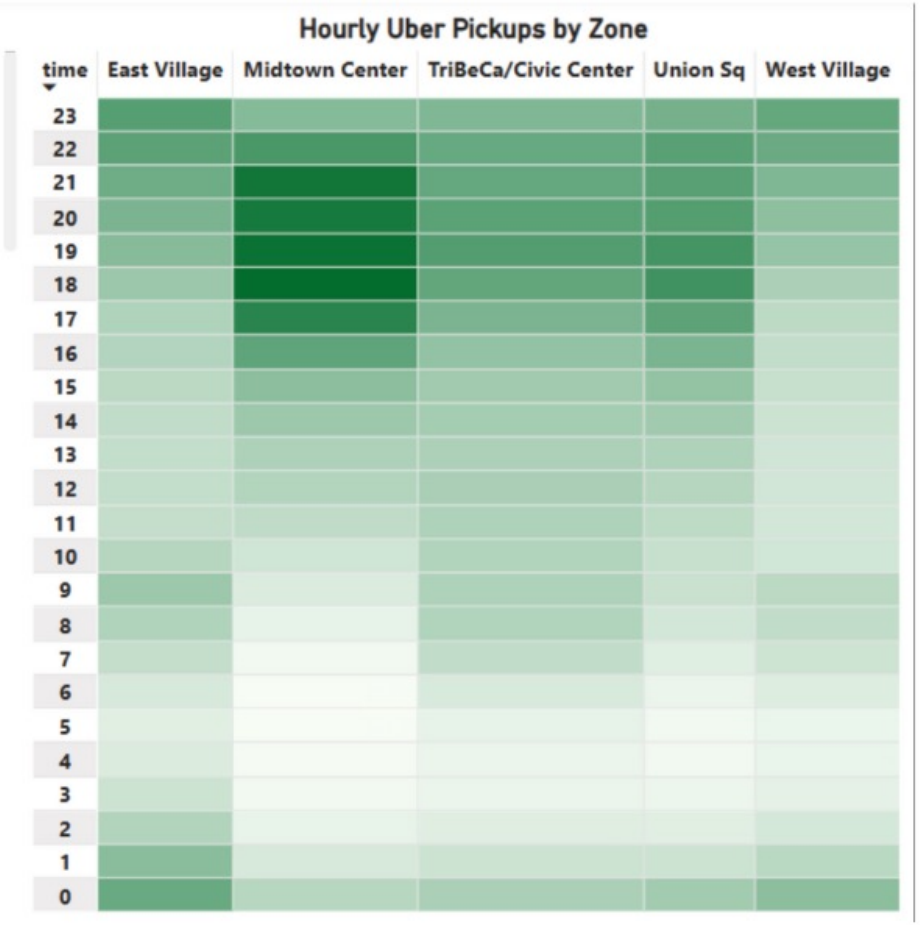
## Total Pickups per Zone

Zone demand analysis showing **Midtown Center** leads with 0.46M pickups, followed by TriBeCa/Civic Center and Union Square at 0.42M each

## Pickups by Weekday

Weekly demand pattern reveals Saturday as peak day (2.41M pickups), with Monday showing lowest activity (1.69M pickups)



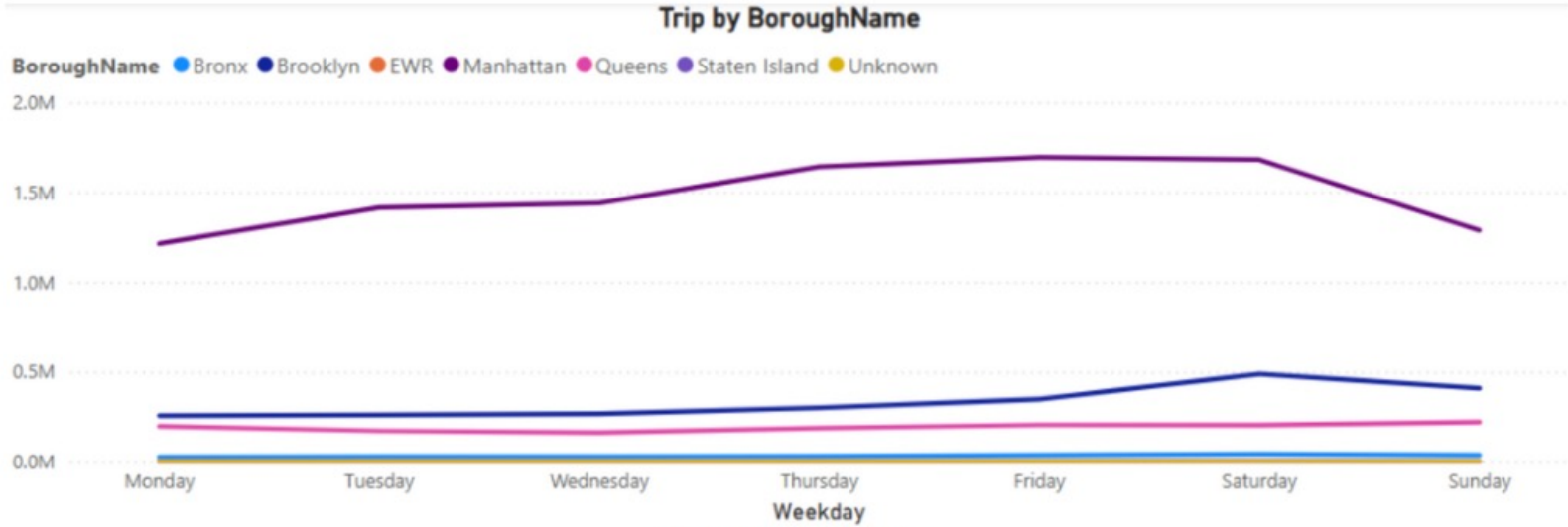

## Key Insights Summary

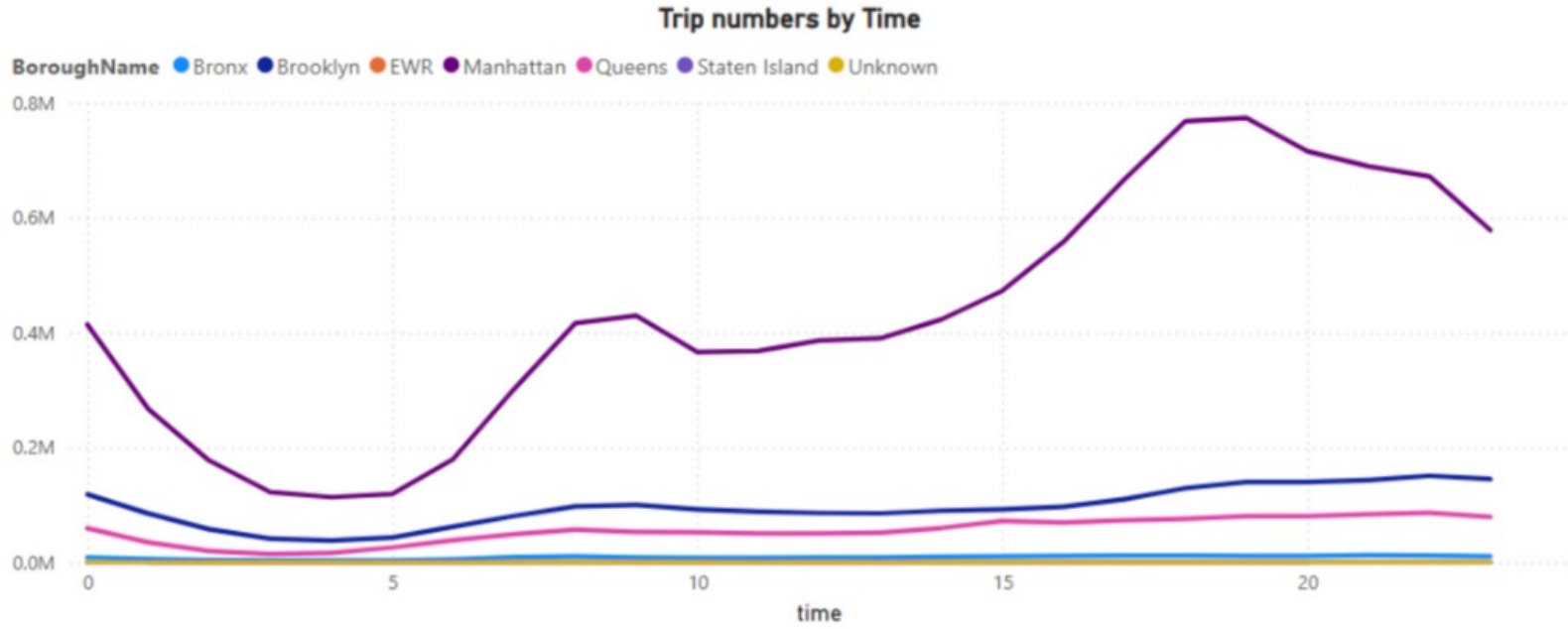
**Peak Performance**

**Geographic Leaders**

**Demand Valleys**

**Timing Strategy**




# Geographic Market Share

**Manhattan**: Clear market leader with 70%+ of total trip volume

**Brooklyn**: Secondary market with consistent but lower demand

**Outer Boroughs**: Bronx, Queens, EWR, Staten Island show minimal ridership

**Unknown Category**: Negligible impact on overall patterns

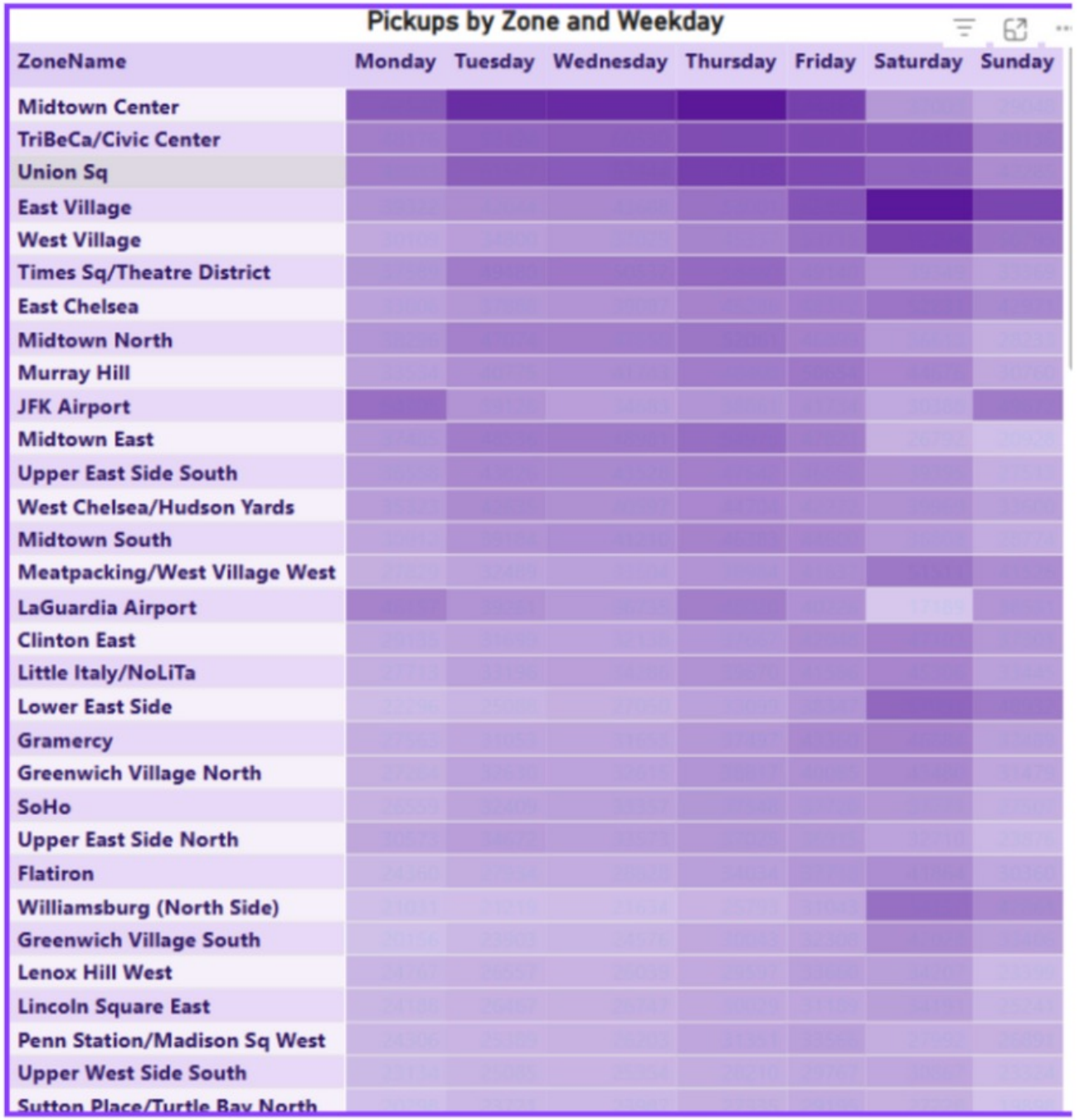
# Temporal Demand Insights

**Weekend Peak**

**Daily Rhythm**

**Off-Peak Hours**

**Consistency**




**Pickups by Zone and Weekday Heatmap**

**Top-Tier Zones (Consistent High Volume)**

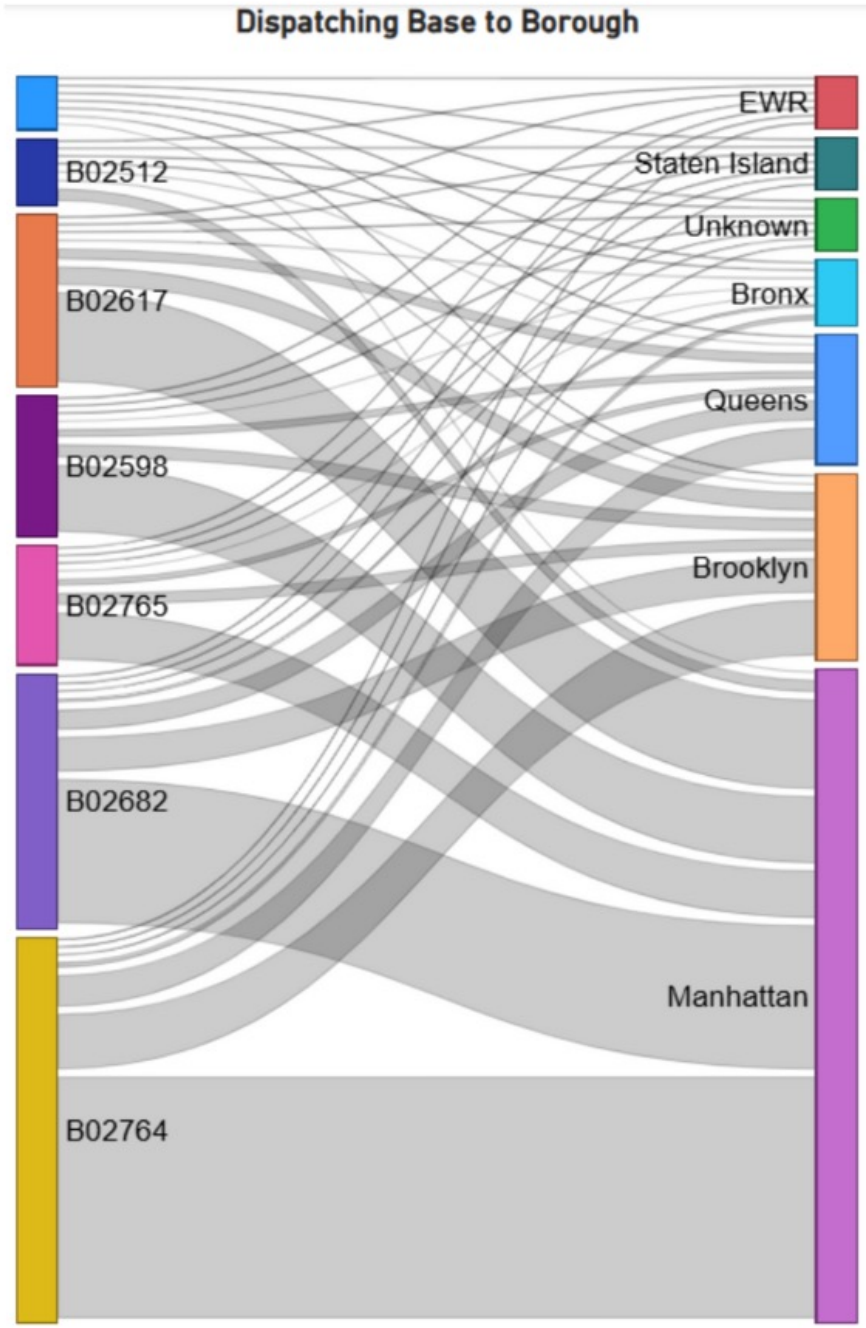
**Weekend Entertainment Zones**

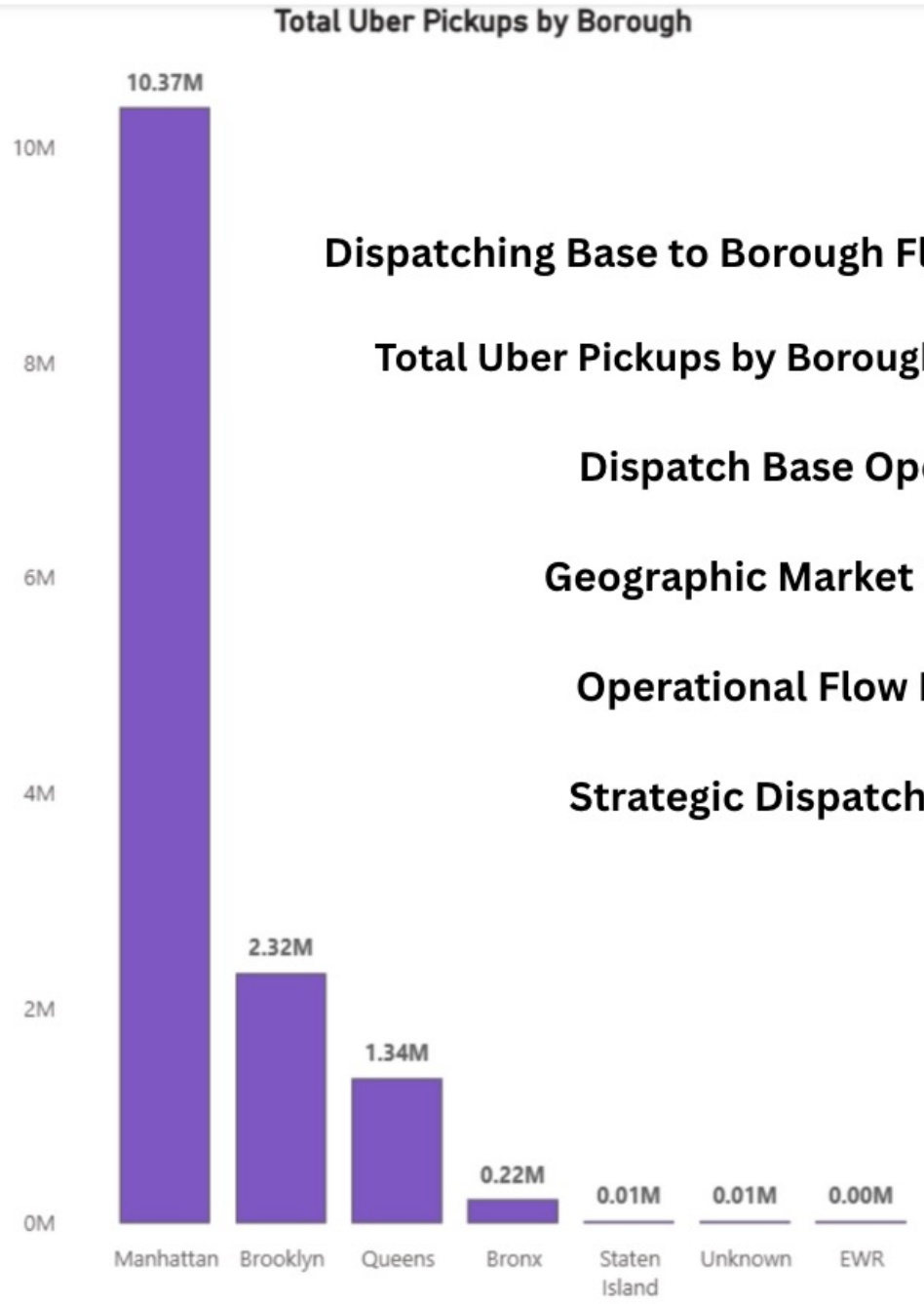
**Transportation Hubs**

**Residential Areas (Lower Volume)**

**Weekly Pattern Insights**

**Strategic Resource Allocation**




**Dispatching Base to Borough Flow (Sankey Diagram)**

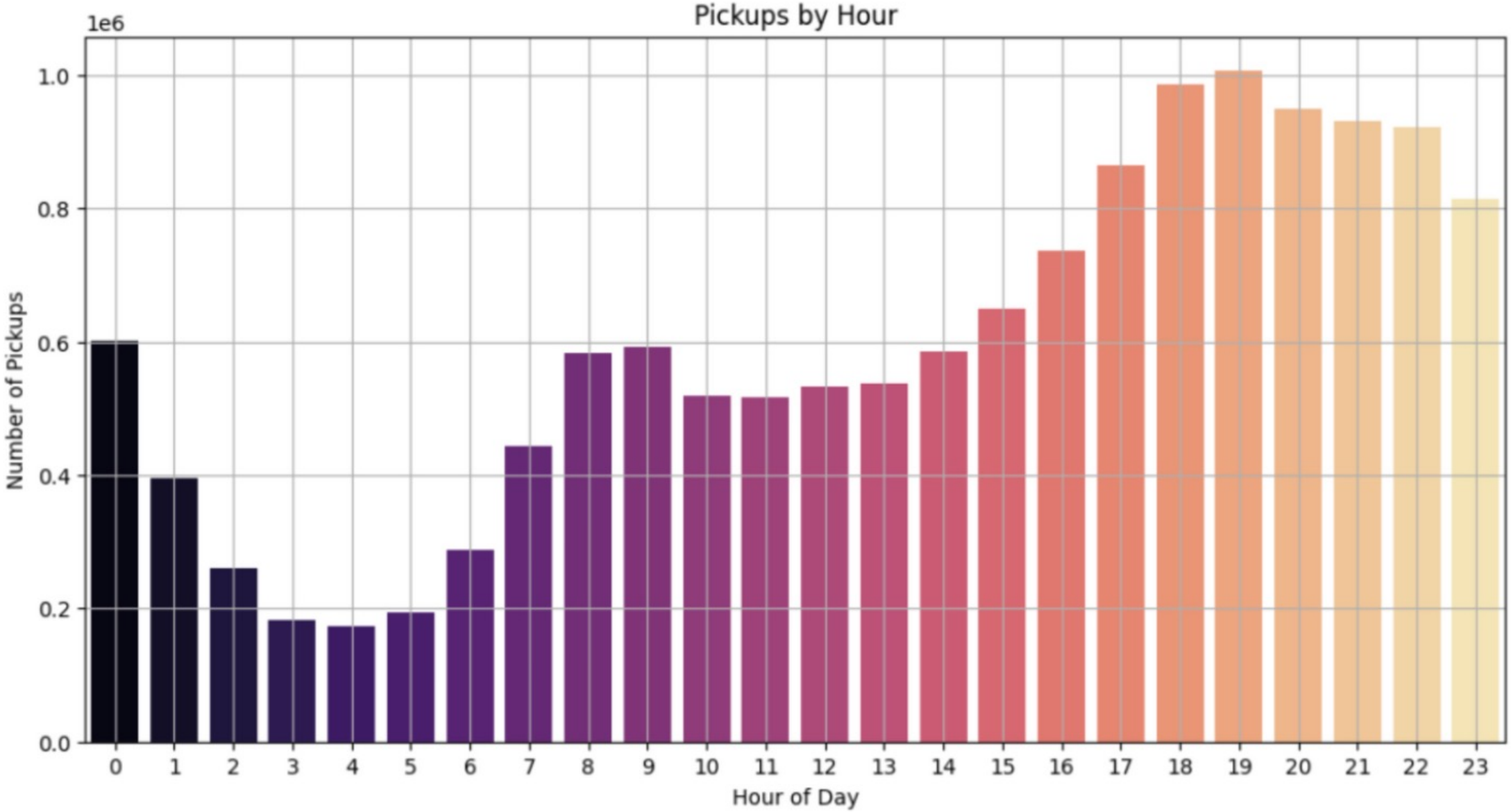
**Total Uber Pickups by Borough (Volume Analysis)**

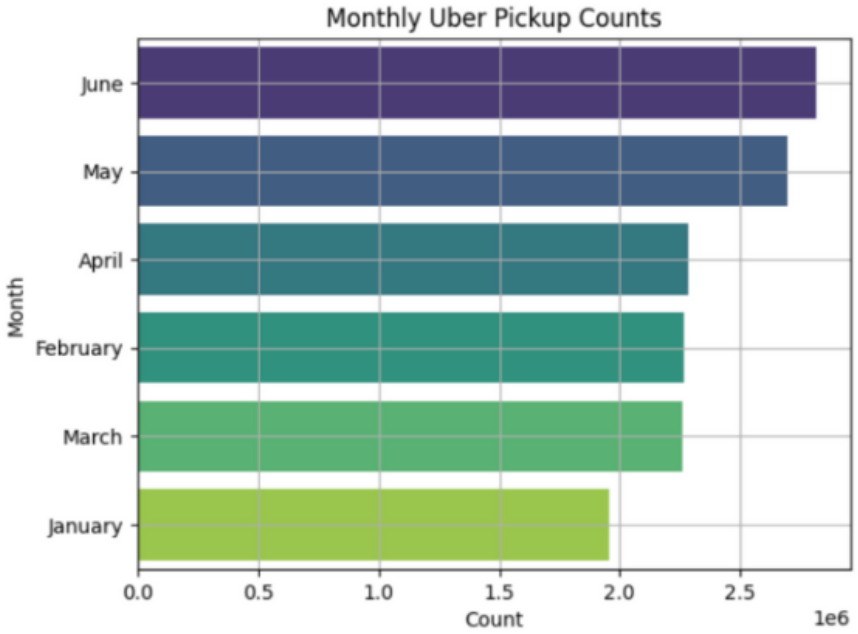
**Dispatch Base Operations**

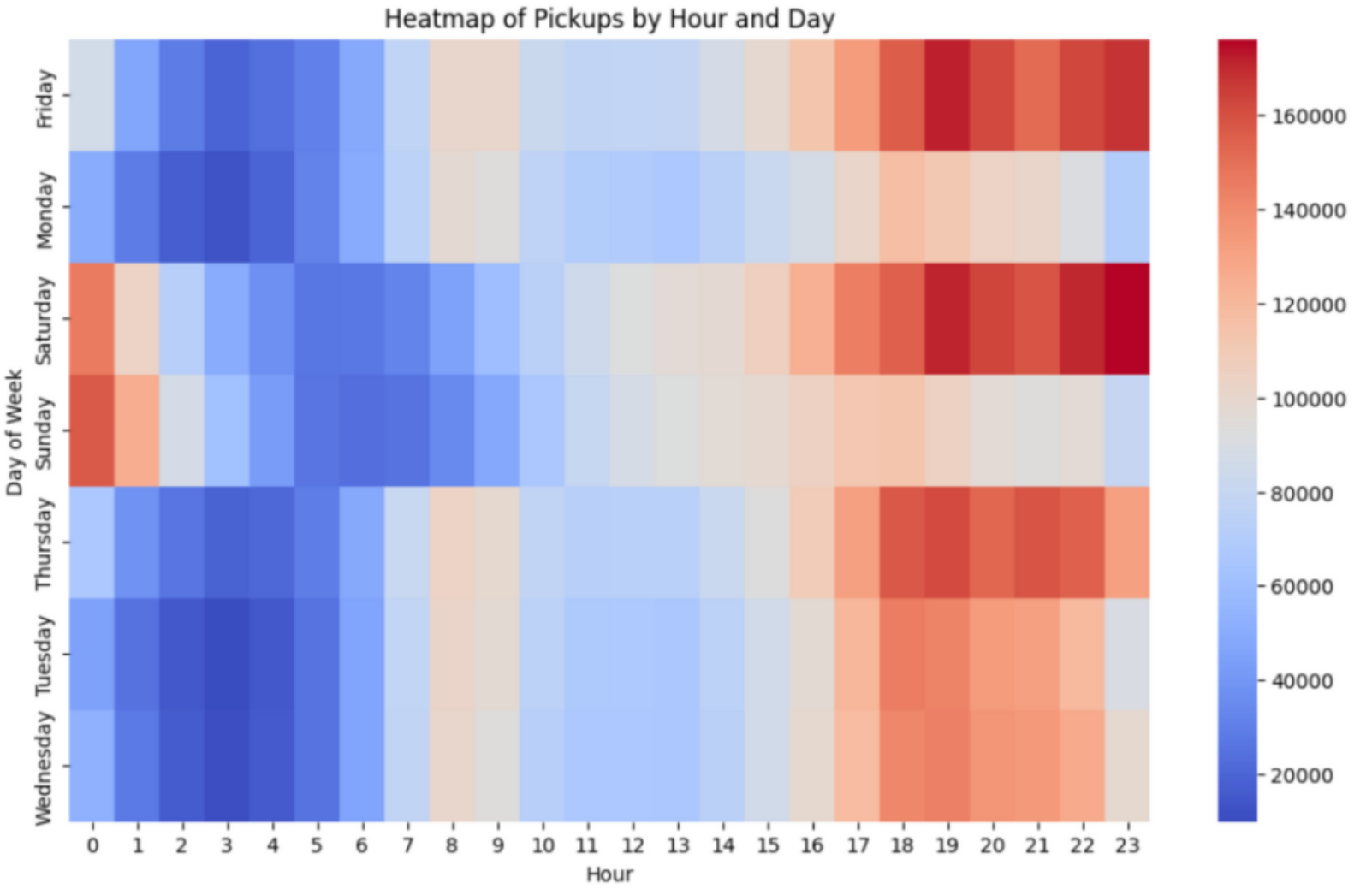
**Geographic Market Hierarchy**

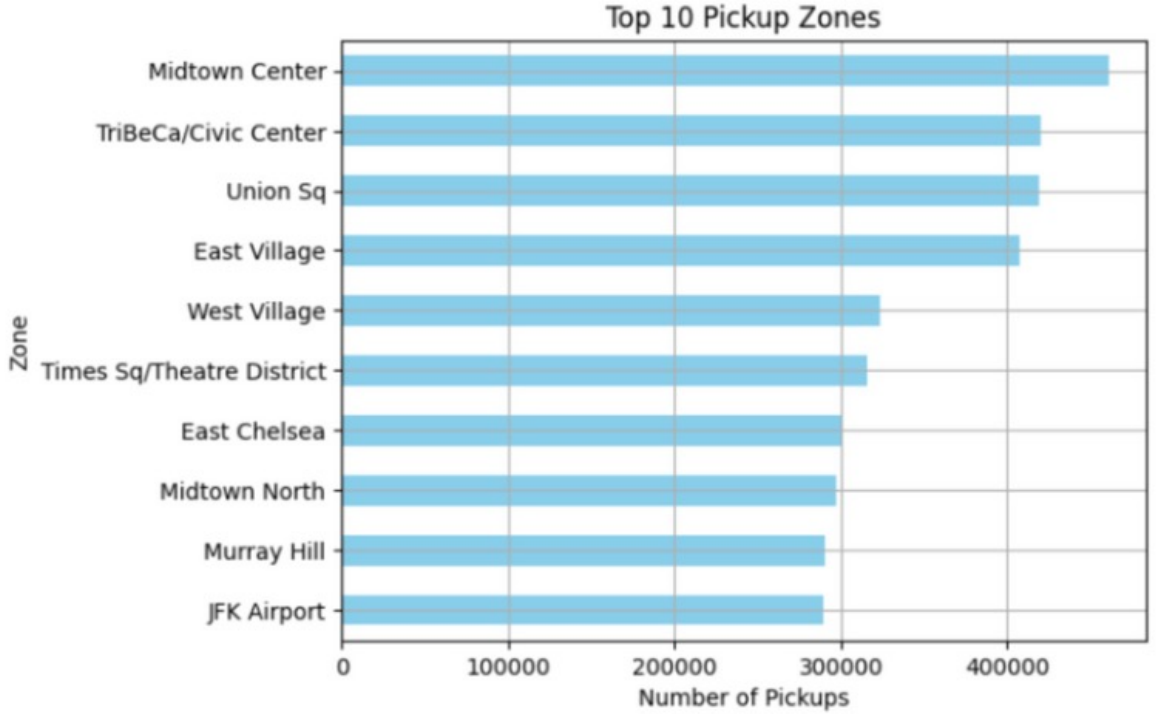
**Operational Flow Patterns**

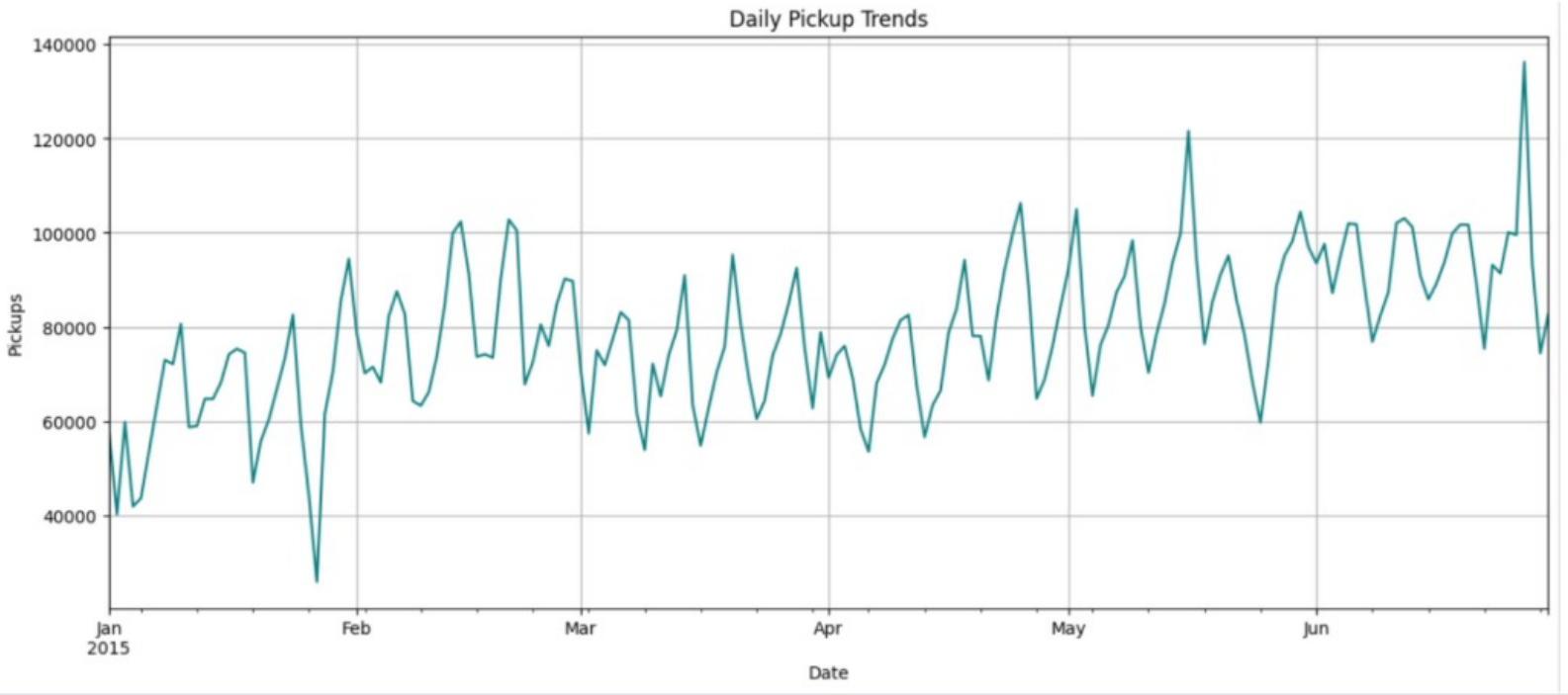
**Strategic Dispatch Insights**


Pickups by Hour
1e6
1.0
0.8
0.6
0.4
0.2
0.0
Number of Pickups
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 of Day


Monthly Uber Pickup Counts
June
May
April
February
March
January
Month
0.0
0.5
1.0
1.5
2.0
2.5
Count
1e6


Heatmap of Pickups by Hour and Day
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Day of Week
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
160000
140000
120000
100000
80000
60000
40000
20000


Top 10 Pickup Zones
Midtown Center
TriBeCa/Civic Center
Union Sq
East Village
West Village
Times Sq/Theatre District
East Chelsea
Midtown North
Murray Hill
JFK Airport
Zone
0
100000
200000
300000
400000
Number of Pickups


Daily Pickup Trends
140000
120000
100000
80000
60000
40000
Pickups
Jan
2015
Feb
Mar
Apr
May
Jun
Date

# Key US Holidays and Events

## Memorial Day

Date: May 30th

Occurs on the last Monday of May

## Valentine's Day

Date: February 14th

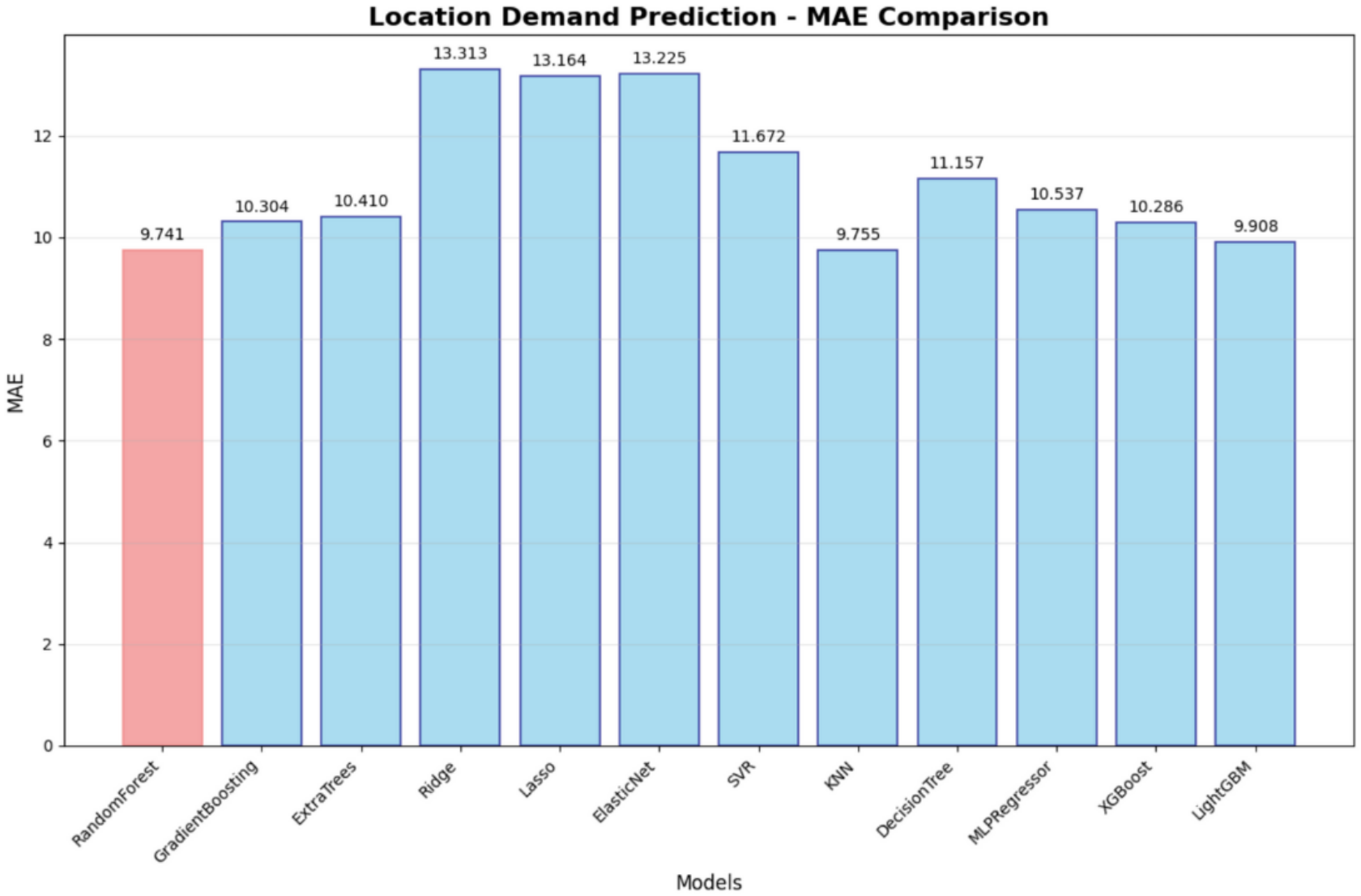
## President's Day

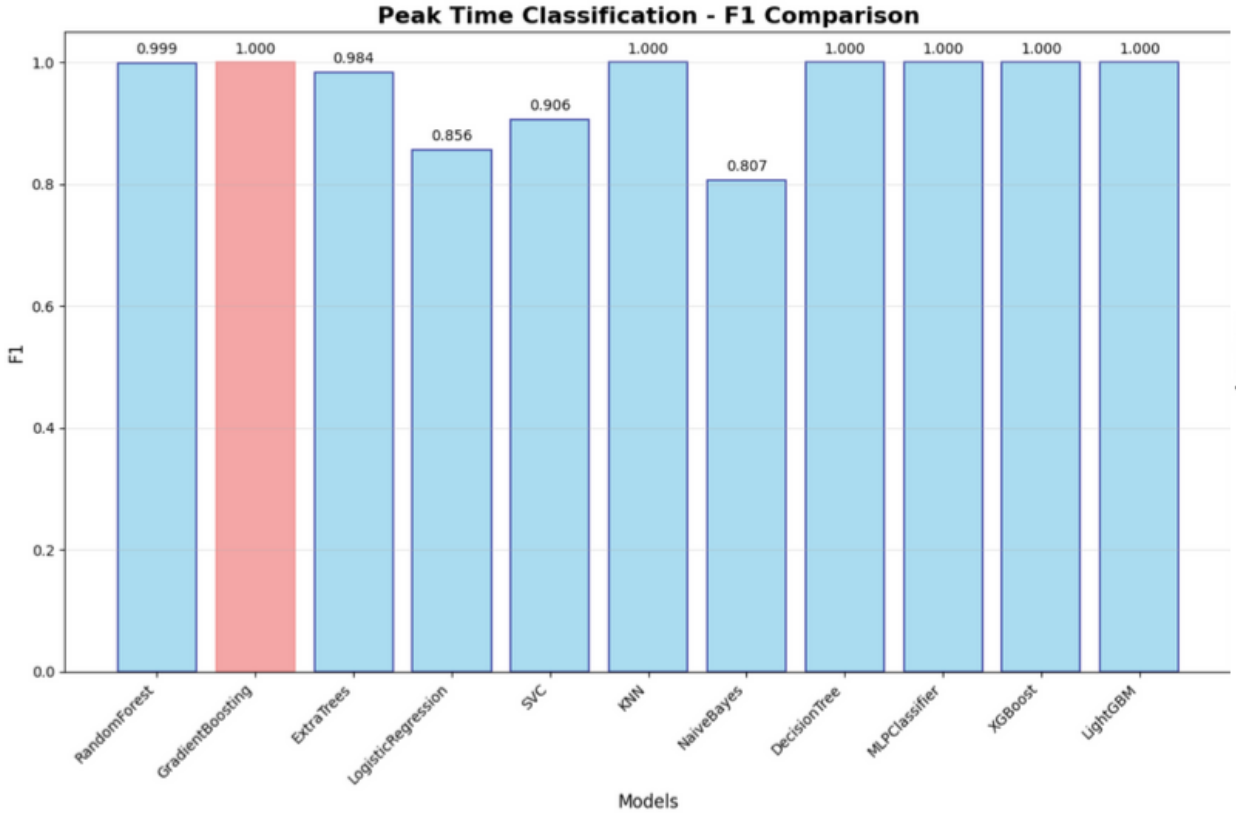
Date: February 21st

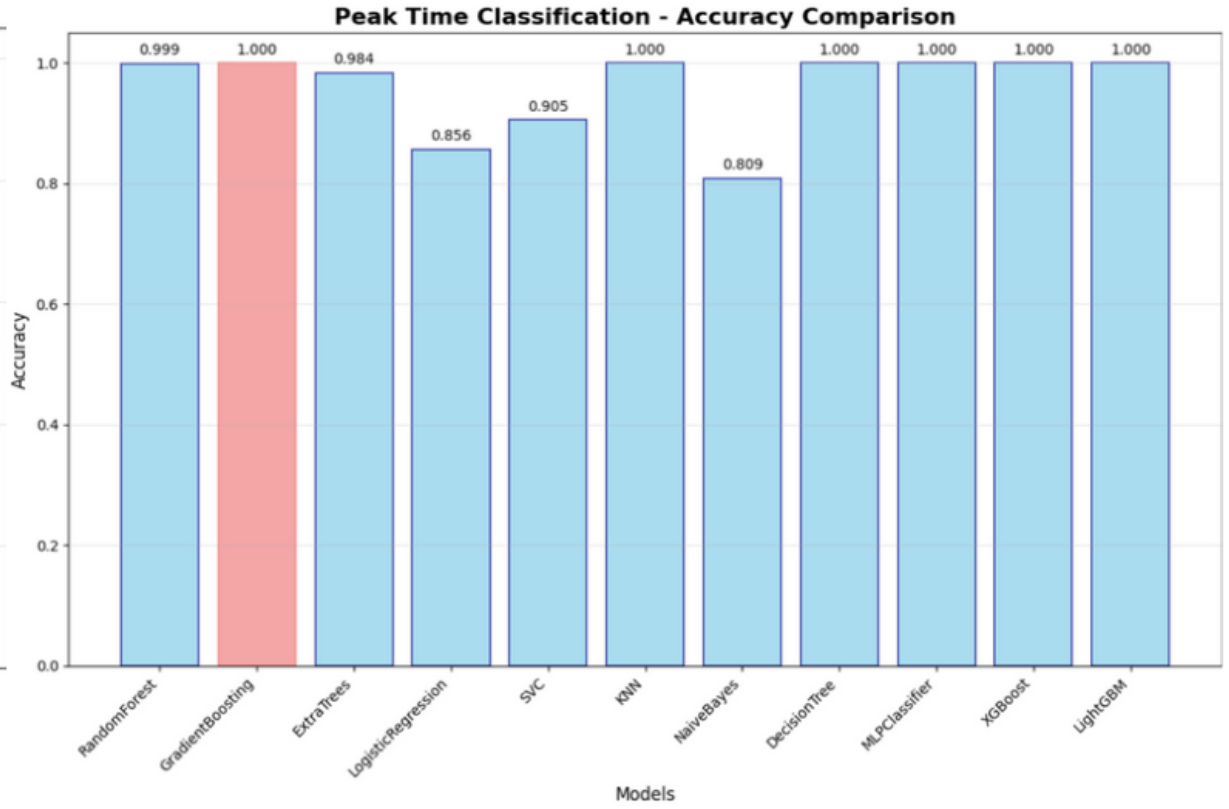
Occurs on the third Monday of February

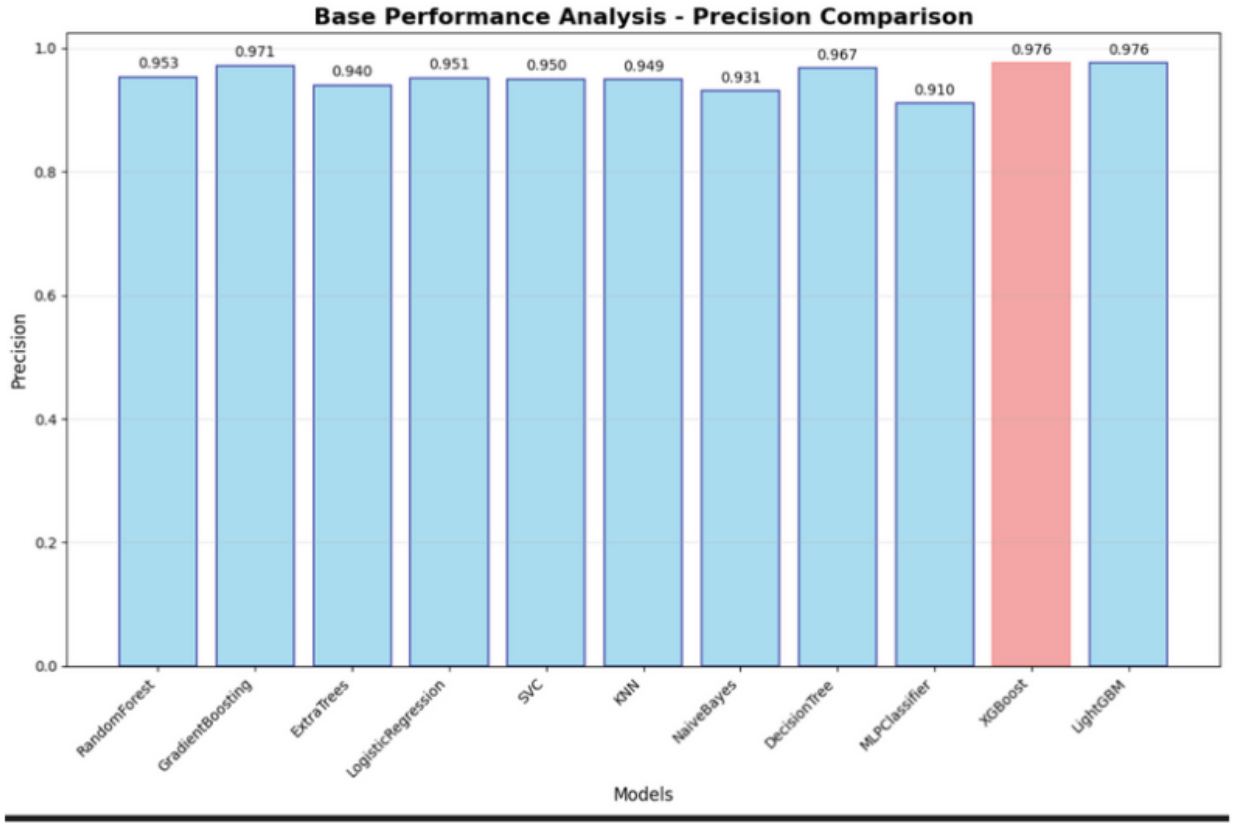
## Super Bowl Sunday (Super Bowl 56)

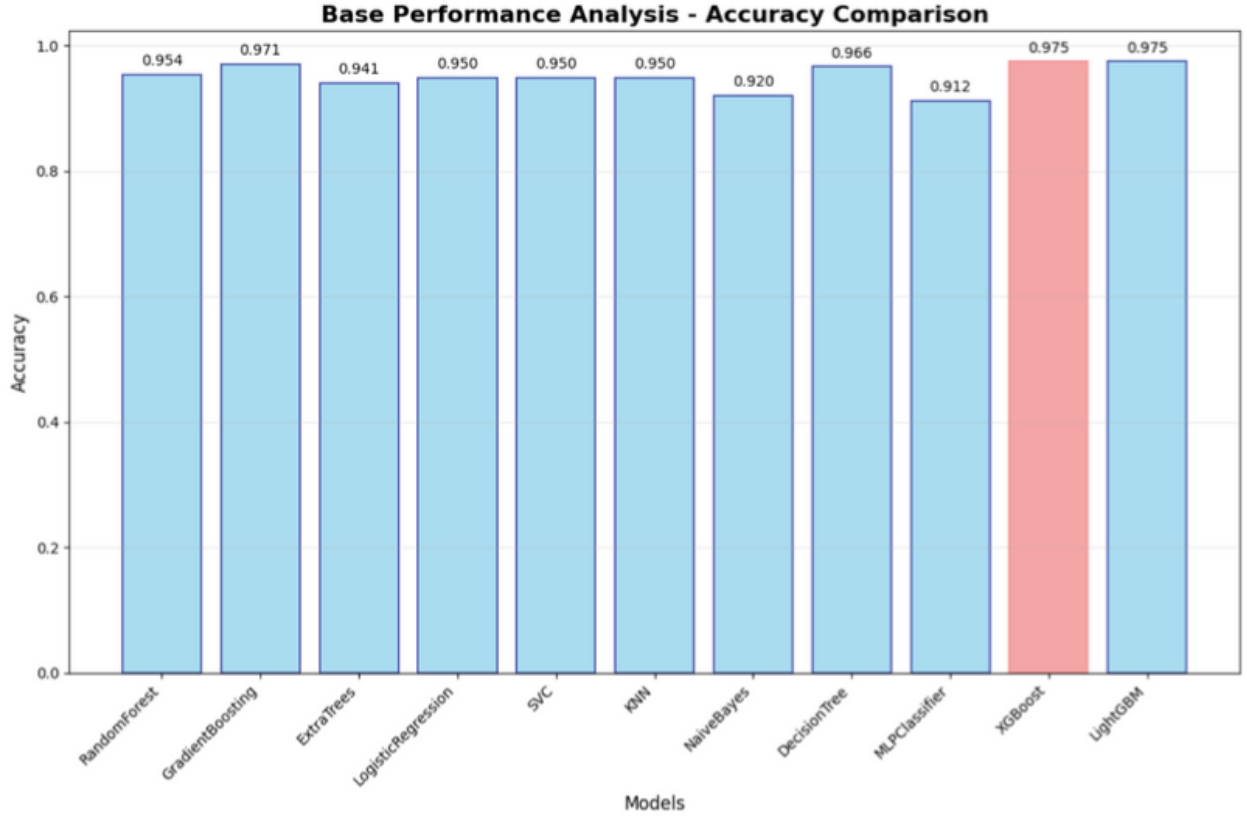
Date: February 13th


Location Demand Prediction - MAE Comparison
MAE
Models
0
2
4
6
8
10
12
9.741
10.304
10.410
13.313
13.164
13.225
11.672
9.755
11.157
10.537
10.286
9.908
RandomForest
GradientBoosting
ExtraTrees
Ridge
Lasso
ElasticNet
SVR
KNN
DecisionTree
MLPRegressor
XGBoost
LightGBM


Peak Time Classification - F1 Comparison
F1
Models
0.999
1.000
0.984
0.856
0.906
1.000
0.807
1.000
1.000
1.000
1.000
RandomForest
GradientBoosting
ExtraTrees
LogisticRegression
SVC
KNN
NaiveBayes
DecisionTree
MLPClassifier
XGBoost
LightGBM


Peak Time Classification - Accuracy Comparison
Accuracy
Models
0.999
1.000
0.984
0.856
0.905
1.000
0.809
1.000
1.000
1.000
1.000
RandomForest
GradientBoosting
ExtraTrees
LogisticRegression
SVC
KNN
NaiveBayes
DecisionTree
MLPClassifier
XGBoost
LightGBM


Base Performance Analysis - Precision Comparison
Precision
Models
RandomForest
GradientBoosting
ExtraTrees
LogisticRegression
SVC
KNN
NaiveBayes
DecisionTree
MLPClassifier
XGBoost
LightGBM
0.953
0.971
0.940
0.951
0.950
0.949
0.931
0.967
0.910
0.976
0.976


Base Performance Analysis - Accuracy Comparison
Accuracy
Models
RandomForest
GradientBoosting
ExtraTrees
LogisticRegression
SVC
KNN
NaiveBayes
DecisionTree
MLPClassifier
XGBoost
LightGBM
0.954
0.971
0.941
0.950
0.950
0.950
0.920
0.966
0.912
0.975
0.975

# Concluding Achievements:

Decoded the City's Rhythm: Successfully identified and visualized the core temporal and geospatial patterns of demand, pinpointing peak hours, days, and locations.

Engineered a Rich, Unified Dataset: Integrated millions of trip logs with complex weather and location data to create the foundation for predictive modeling.

Mastered Demand Prediction: Developed a high-performing regression model (RandomForest) to forecast location-based demand with a low Mean Absolute Error of 9.74.

Achieved Near-Perfect Peak Time Classification: Built a classification model that identifies peak travel times with over 99% accuracy and a perfect F1-score of 1.0.

Modeled Operational Performance: Created a high-precision model to analyze dispatch base activity with over 97% accuracy.

# Authors:

Mohammad Hossein Mazhari

Mohammad Taha Majlesi

Alireza Karimi

2025-july