Demand Prediction

Summer 2023

CHAUFFERUR

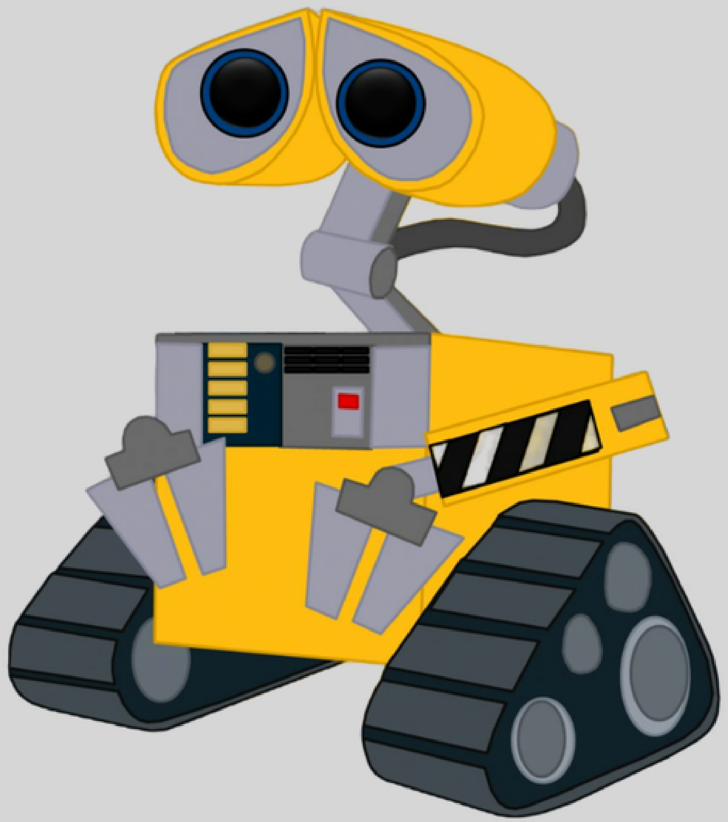
About Us

CHAUFFERUR

Our Pain

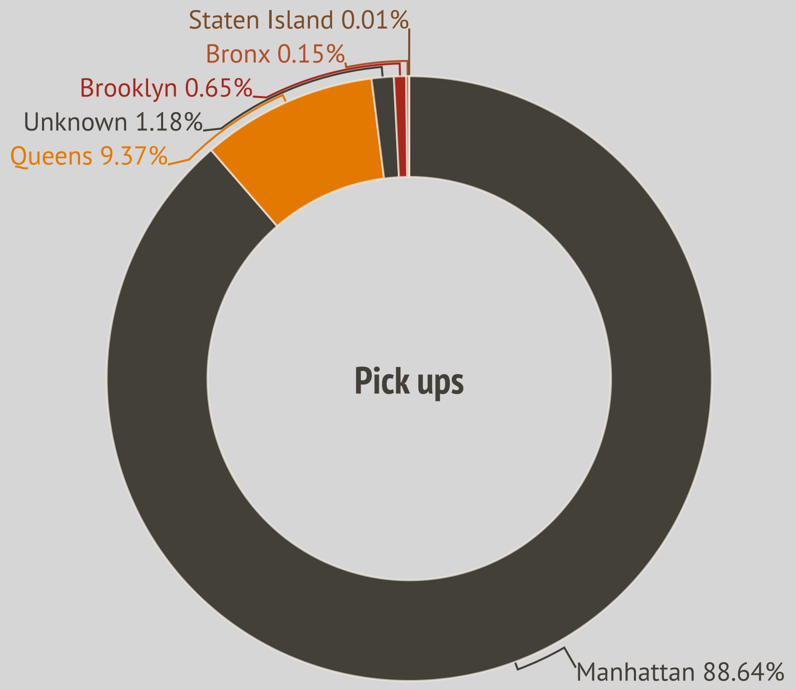
What do we want to solve?

Why it's important?

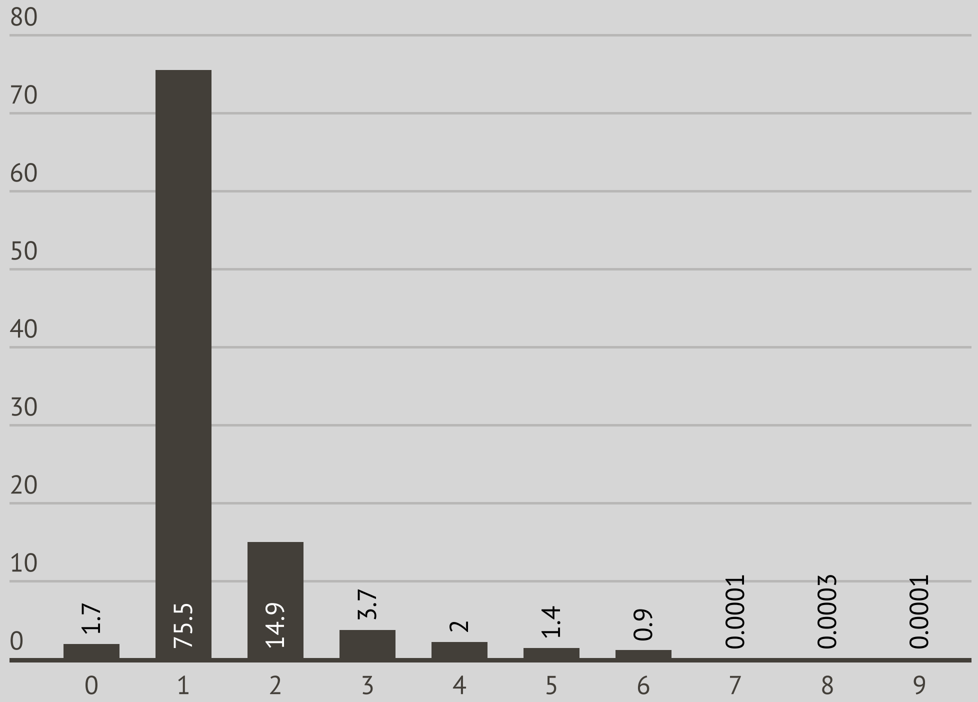


Exploratory Data Analysis

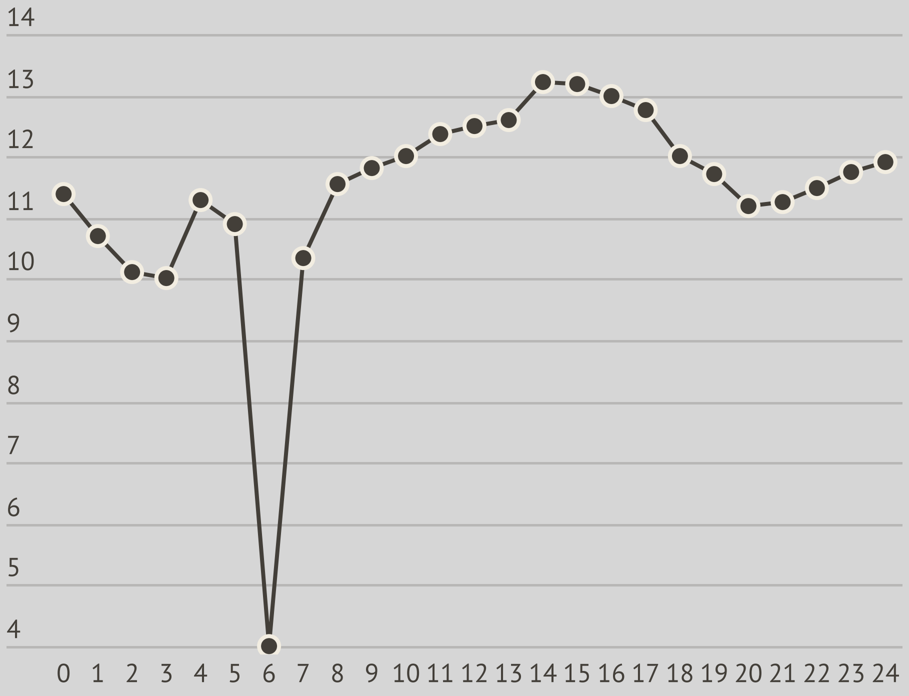
Pick ups Distribution



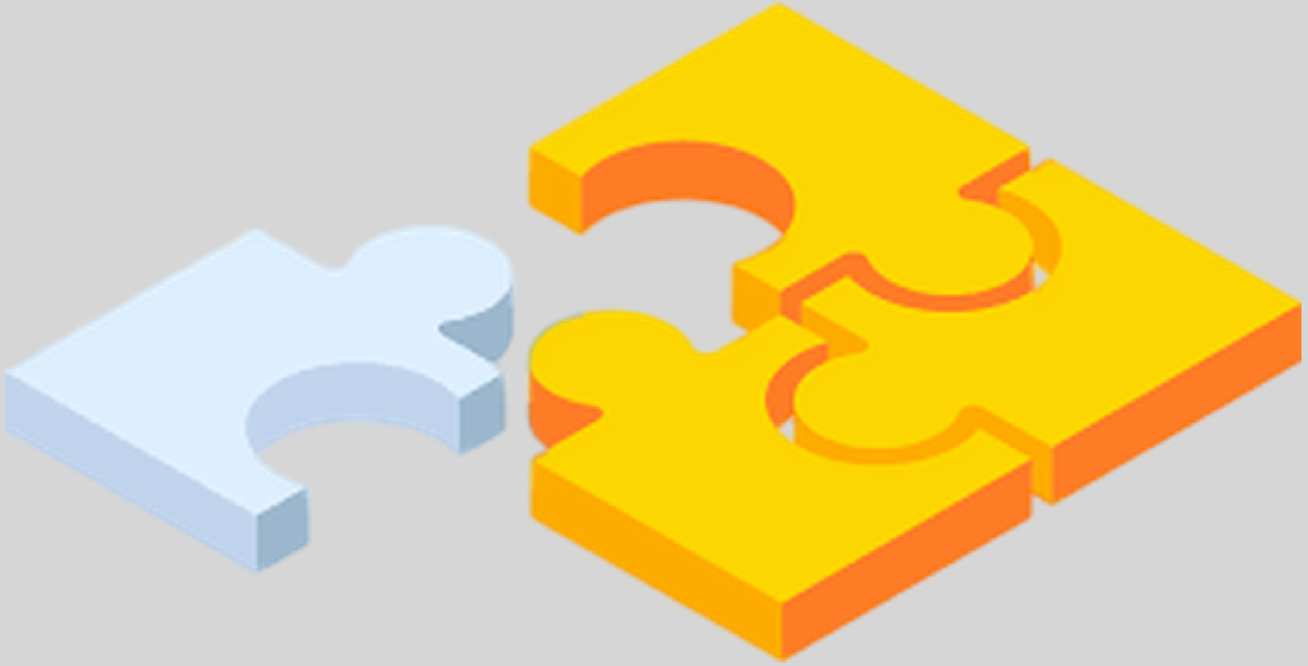
Passenger count Distribution

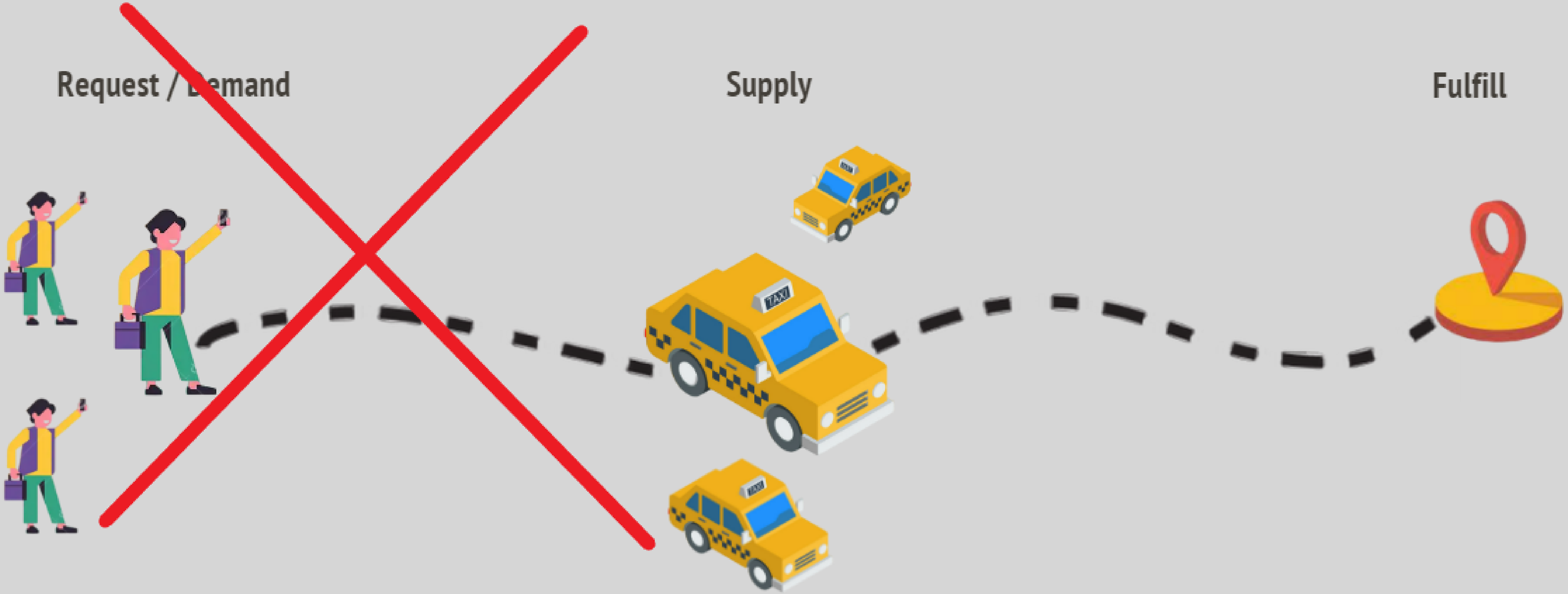


Duration Distribution in minutes



Challenges...




Request / Demand
Supply
Fulfill

Requests / Demands

Accepts / Supplies

Demand

- How we decide?

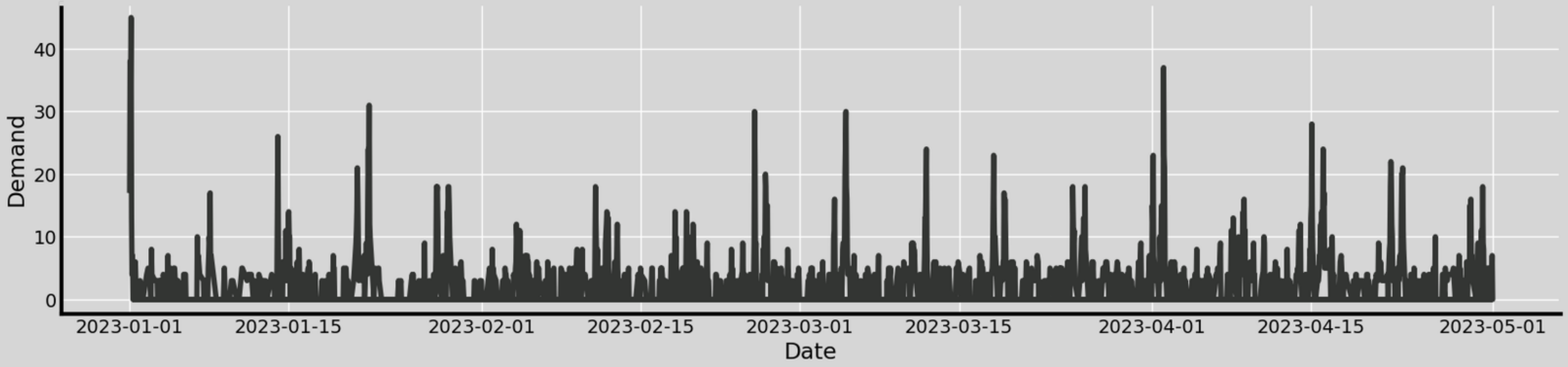
- What did we considered?

- And finally....

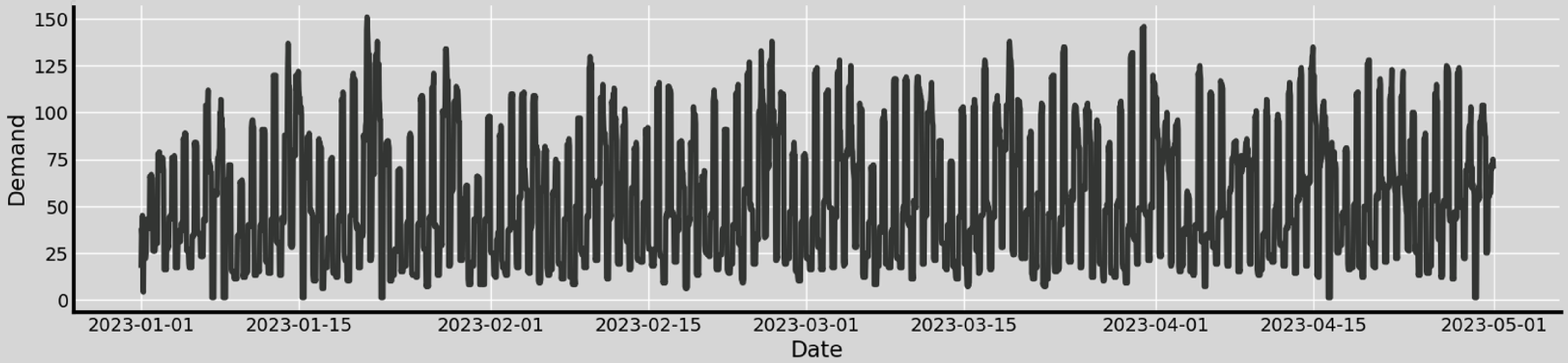
$$demand = passenger_{count} + (dropoff_{last6hours} - pickup_{last6hours})$$

Time analysis

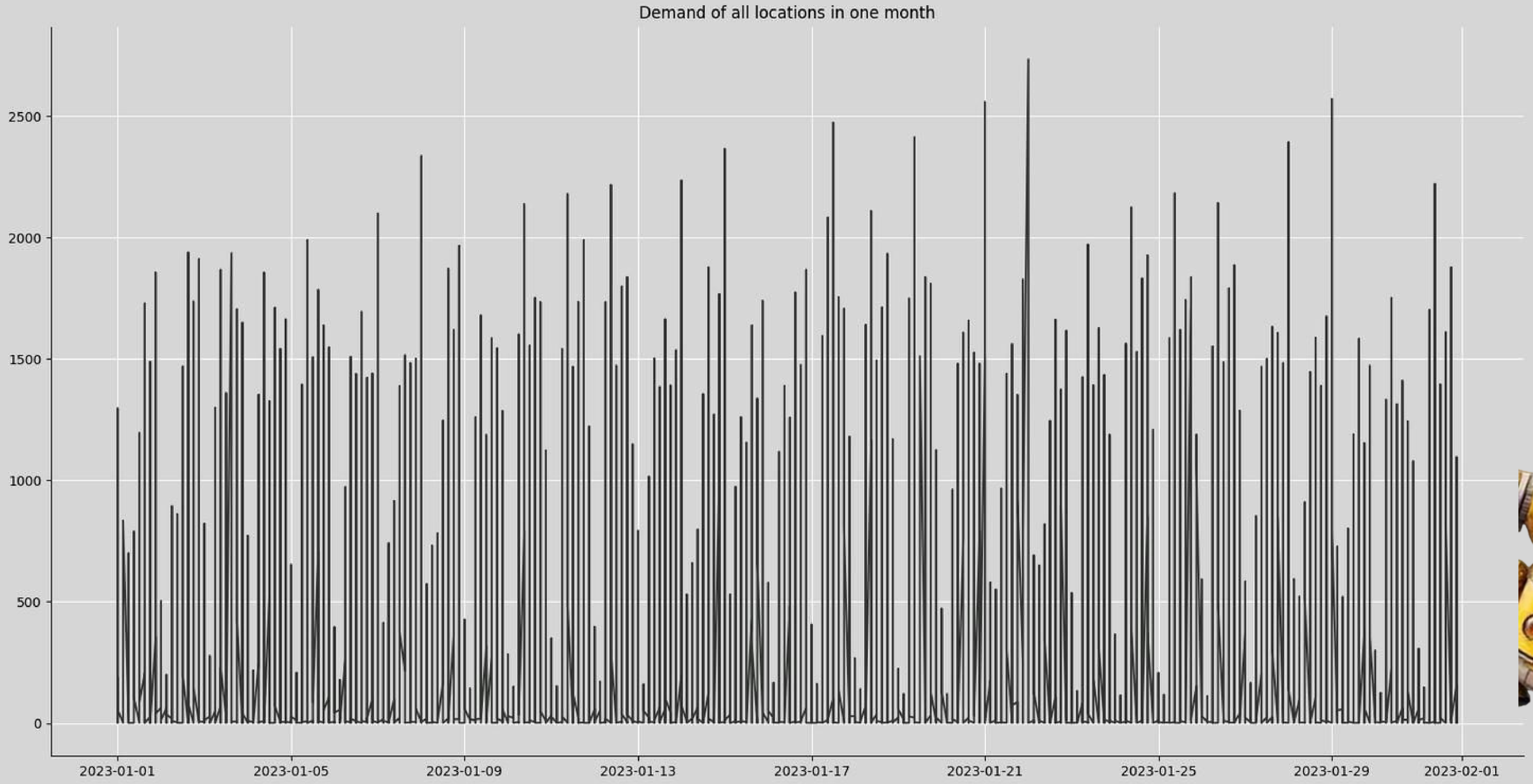
Pick up :



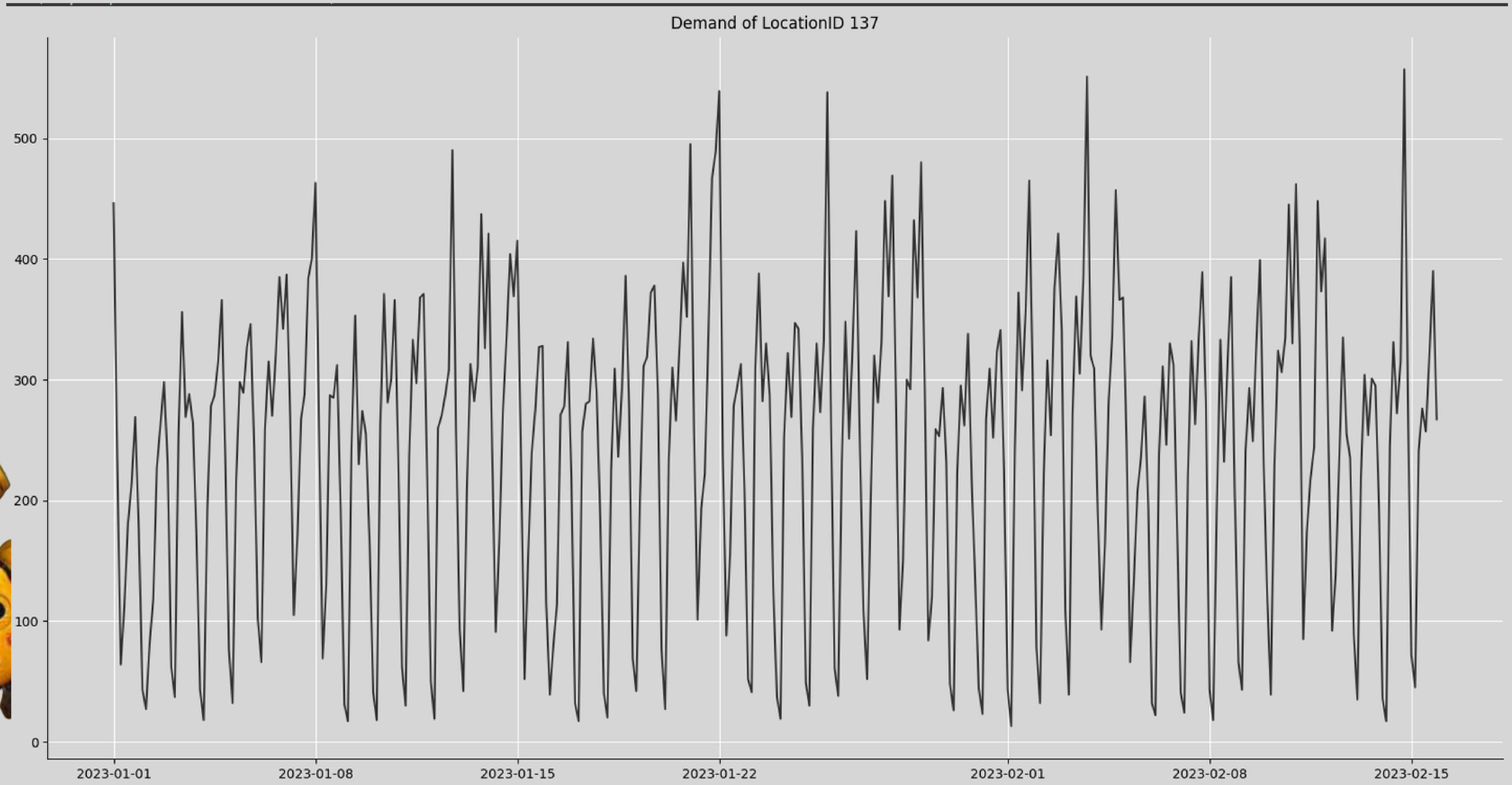
Our Demand :



Demand


Demand of all locations in one month
2500
2000
1500
1000
500
0
2023-01-01
2023-01-05
2023-01-09
2023-01-13
2023-01-17
2023-01-21
2023-01-25
2023-01-29
2023-02-01

Demand Kips bay

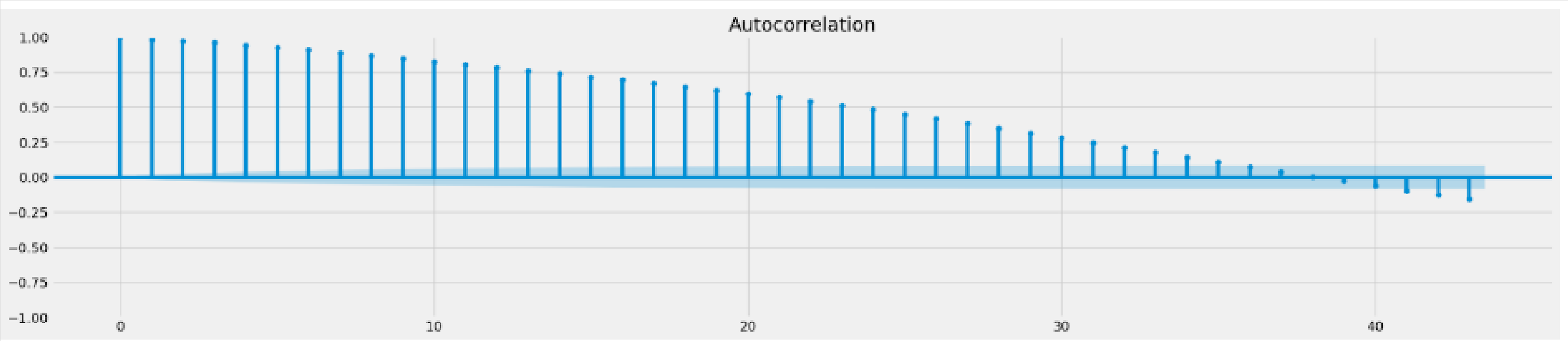


Time analysis

Stationary or non_stationary

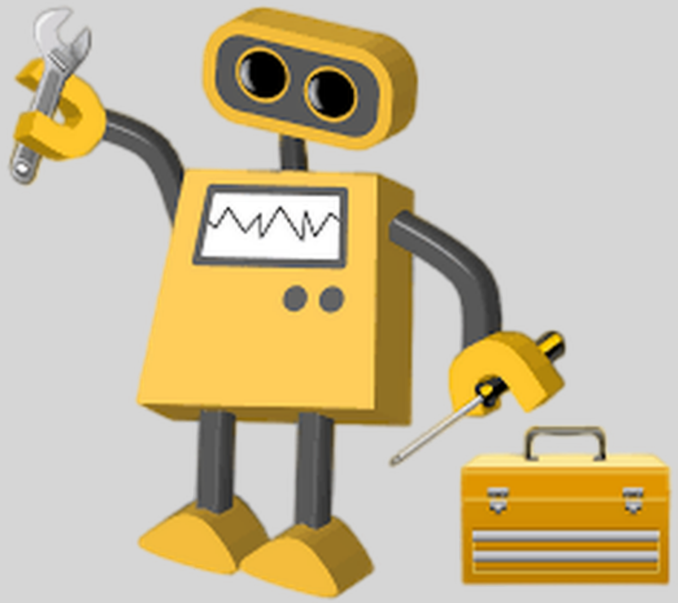
- We conducted the Augmented Dickey-Fuller Test (ADF).
- This test suggests that 42 lags **(6 hours)**

Autocorrelation

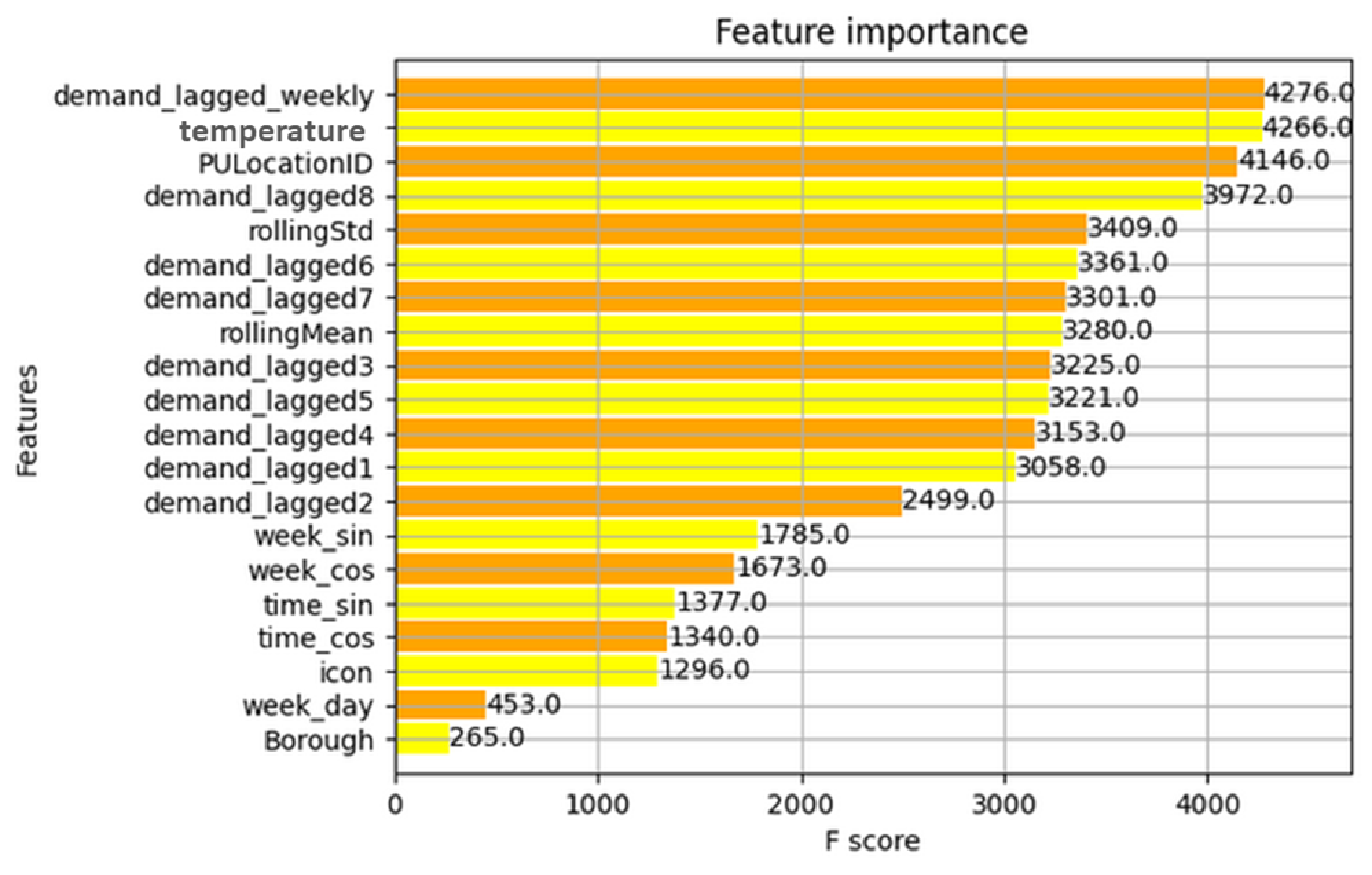


Missing Values & Outliers

Features	Missing	Outlier
Trip distance	Linear regression	Imputing with median
Total amount	Linear regression	Imputing with absolute
Fare amount	Linear regression	
Passenger count	Imputing with median	Isolation forest



Feature Importance



Feature Selection

- Demand
- Borough
- Weather
- Time
- Holiday
- Rolling window
- Demand lag

Evaluation Metrics

MAPE (for medium and high demands)

$$\frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

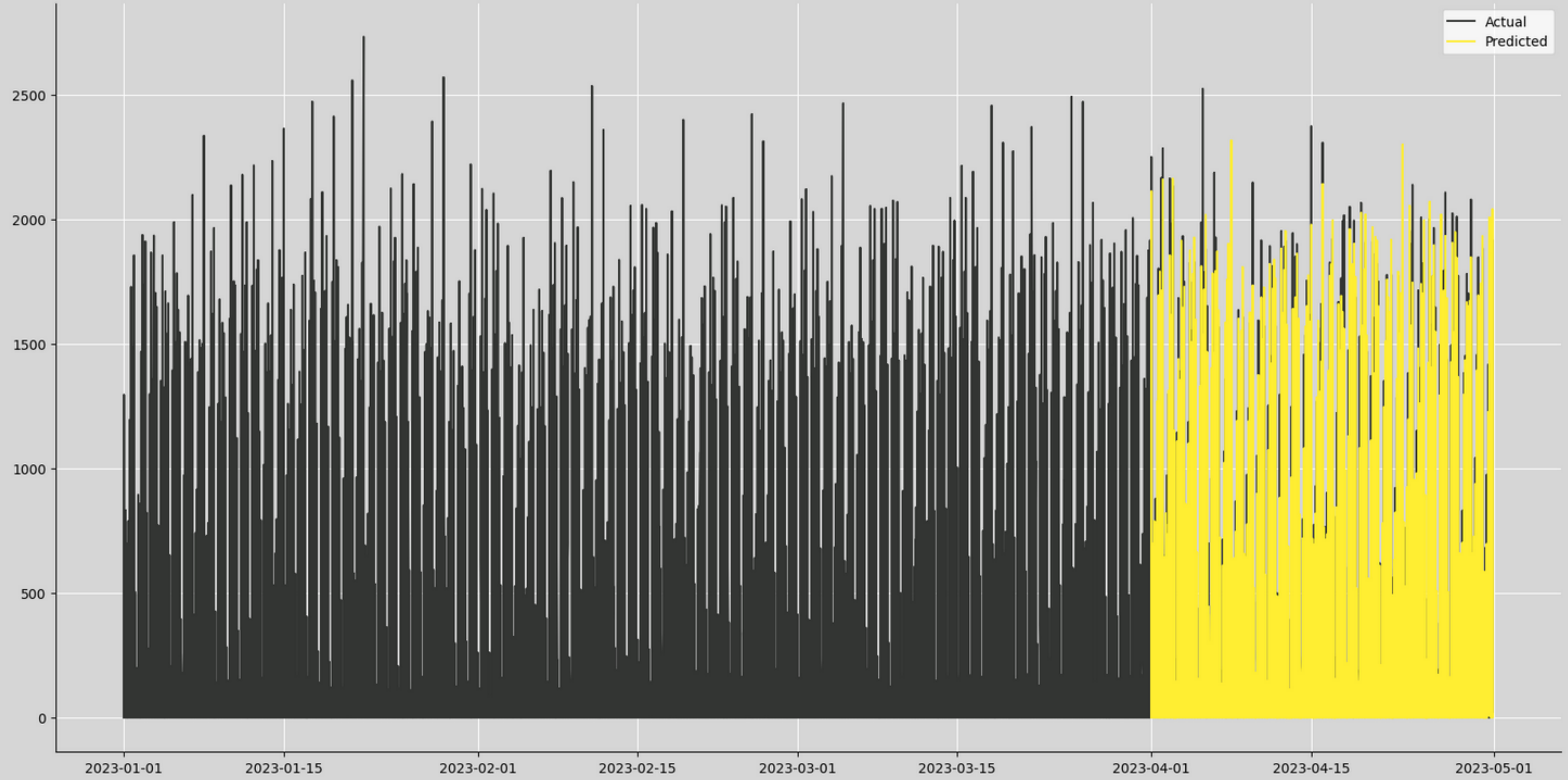
RMSE (for low demands)

$$\sqrt{\sum_{i=1}^{n}\frac{(\hat{y}_i - y_i)^2}{n}}$$

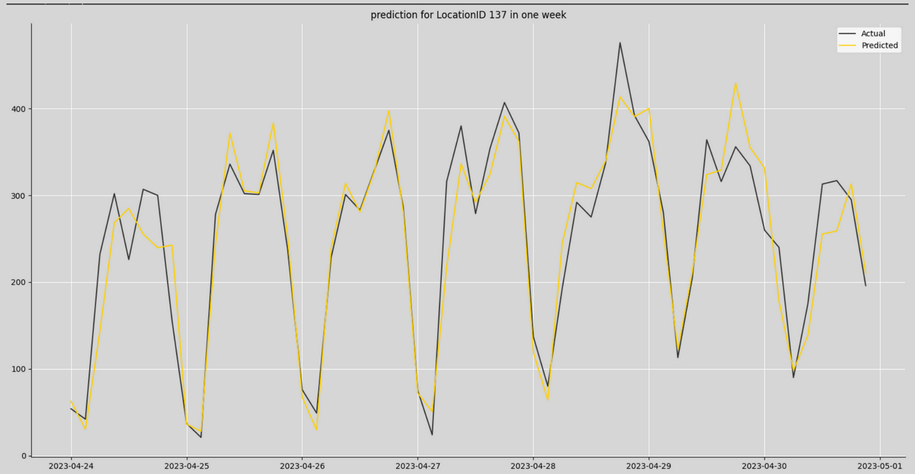
Models

Machine Learning + auto-regression

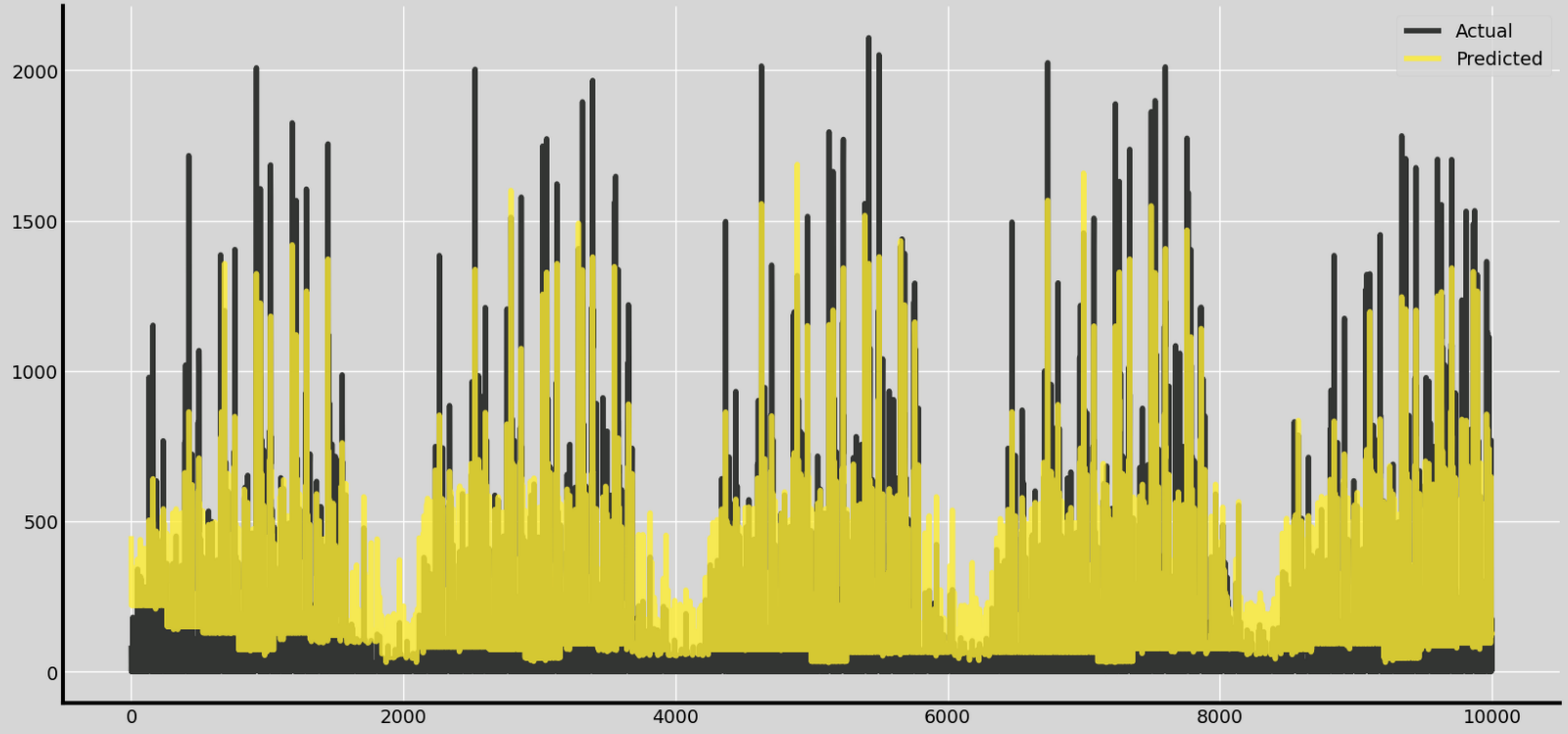
1. XGBoost



Low Demands	High Demands
RMSE	MAPE
1.68	50%

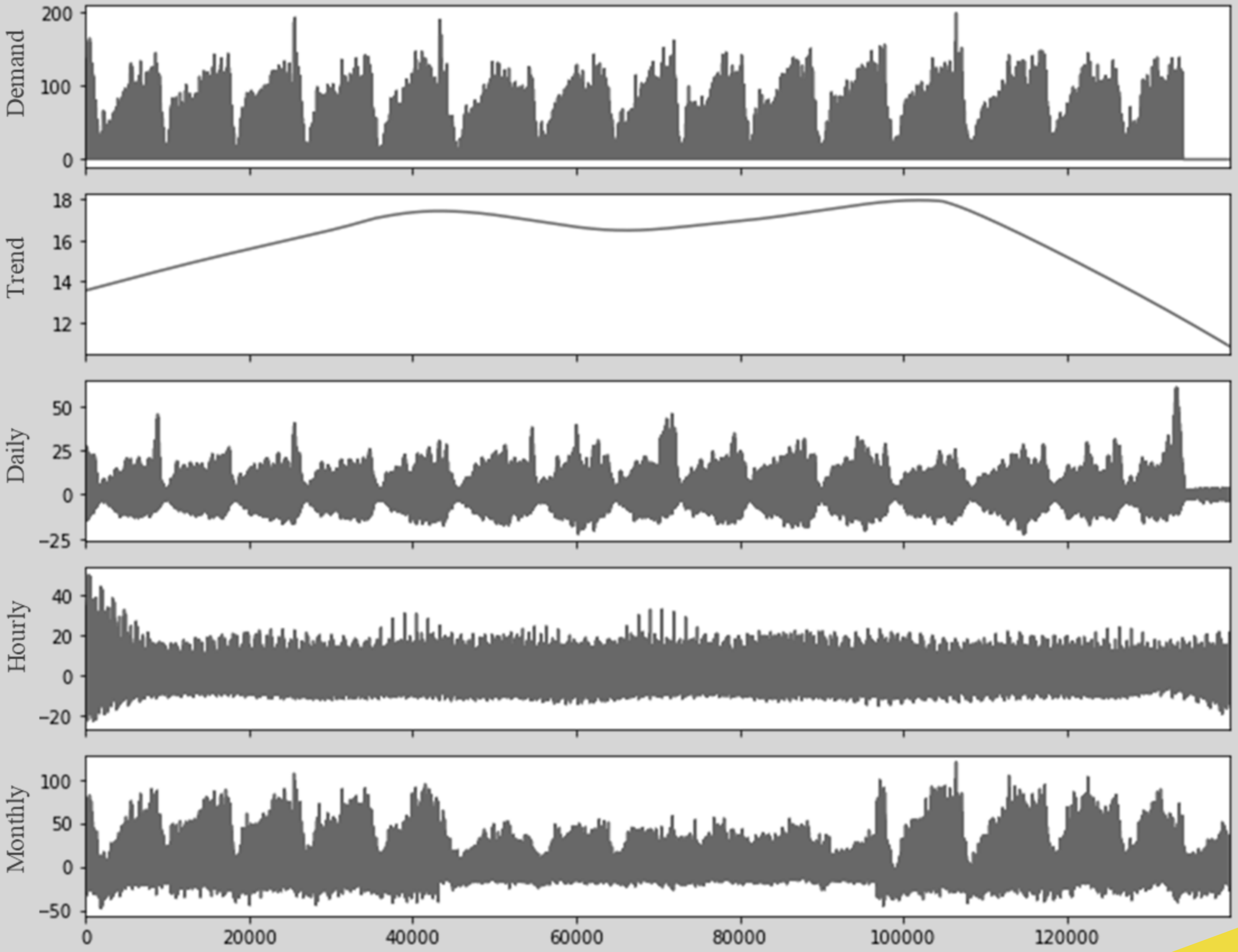


2. AdaBoost

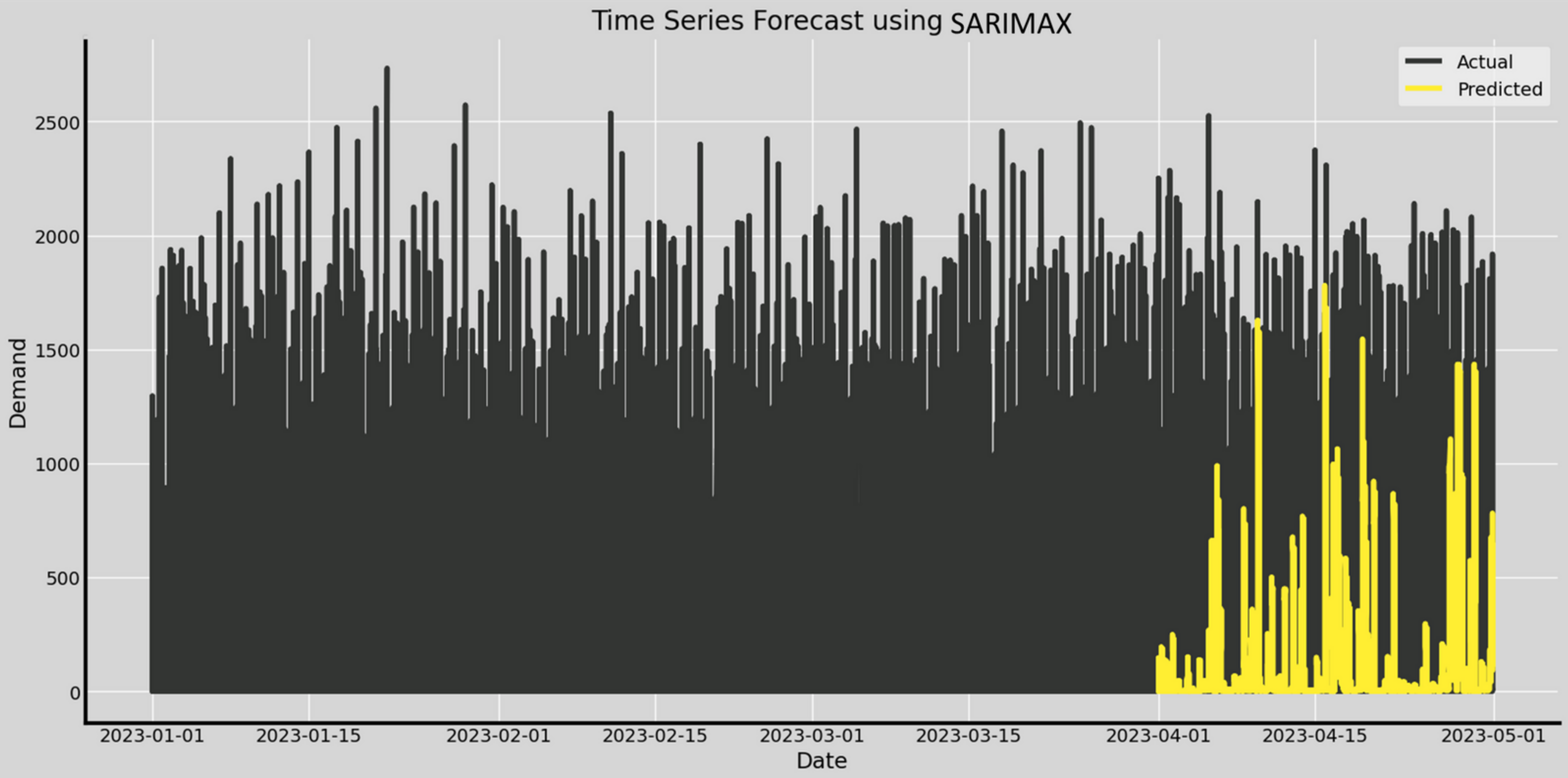


Low Demands	High Demands
RMSE	MAPE
49.33	37.4%

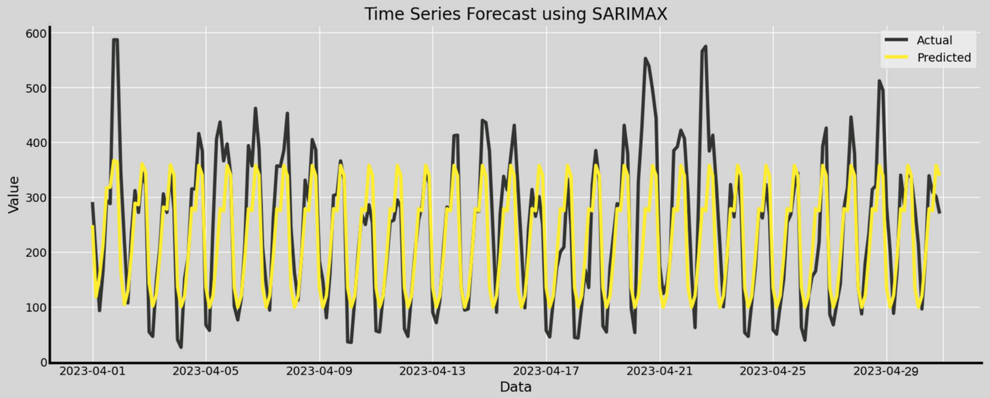
3. Auto Regressive



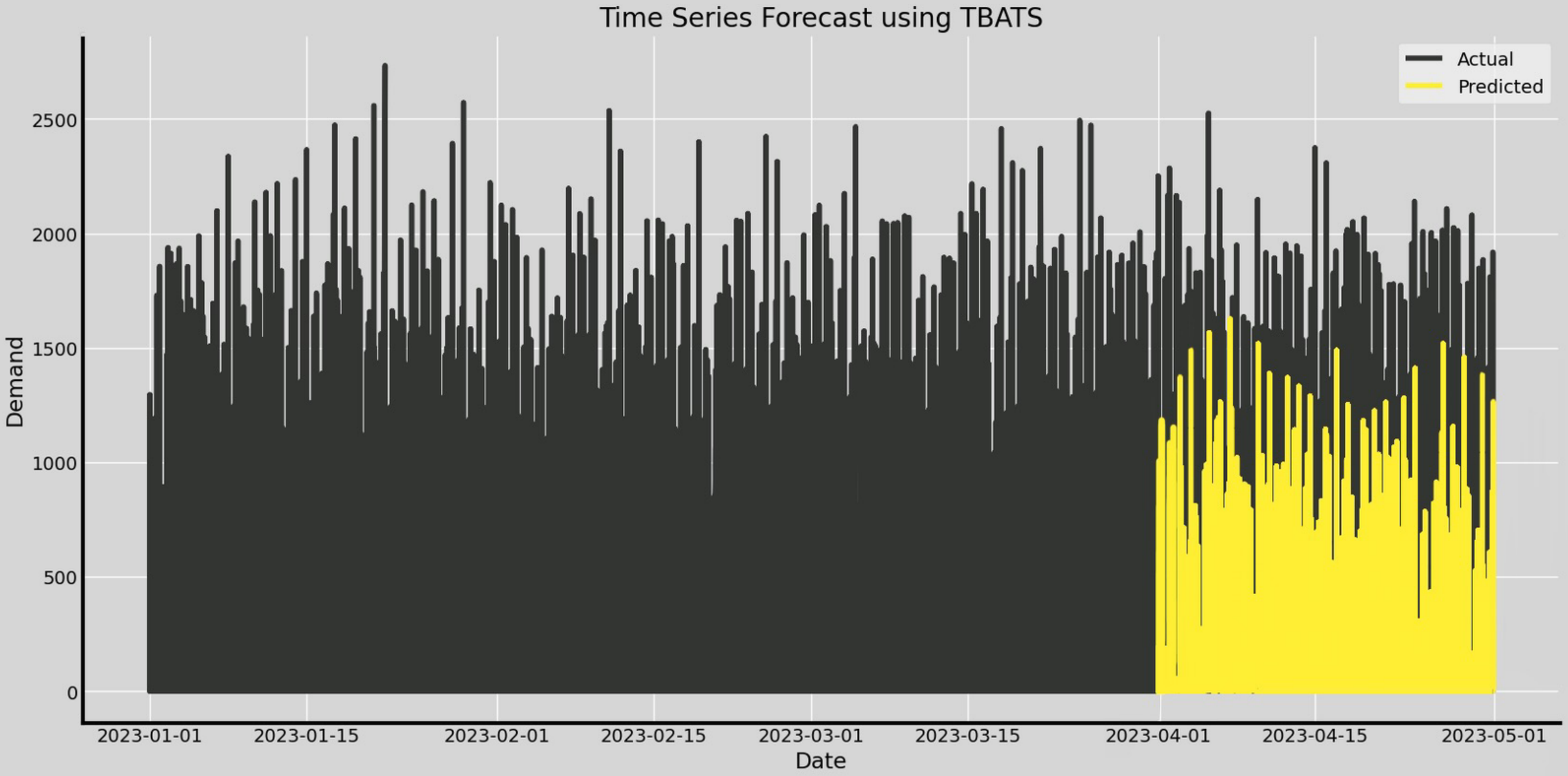
3.1.Sarimax



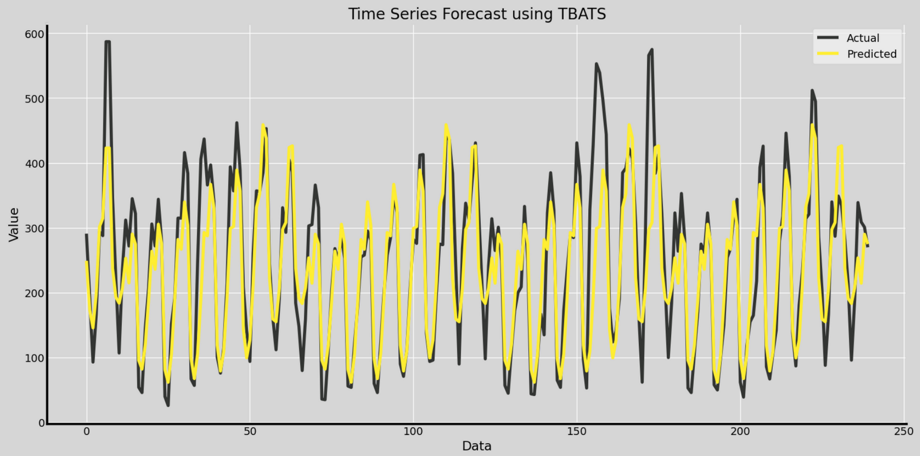
Low Demands	High Demands
RMSE	MAPE
255.3245	100%



3.2.TBATS



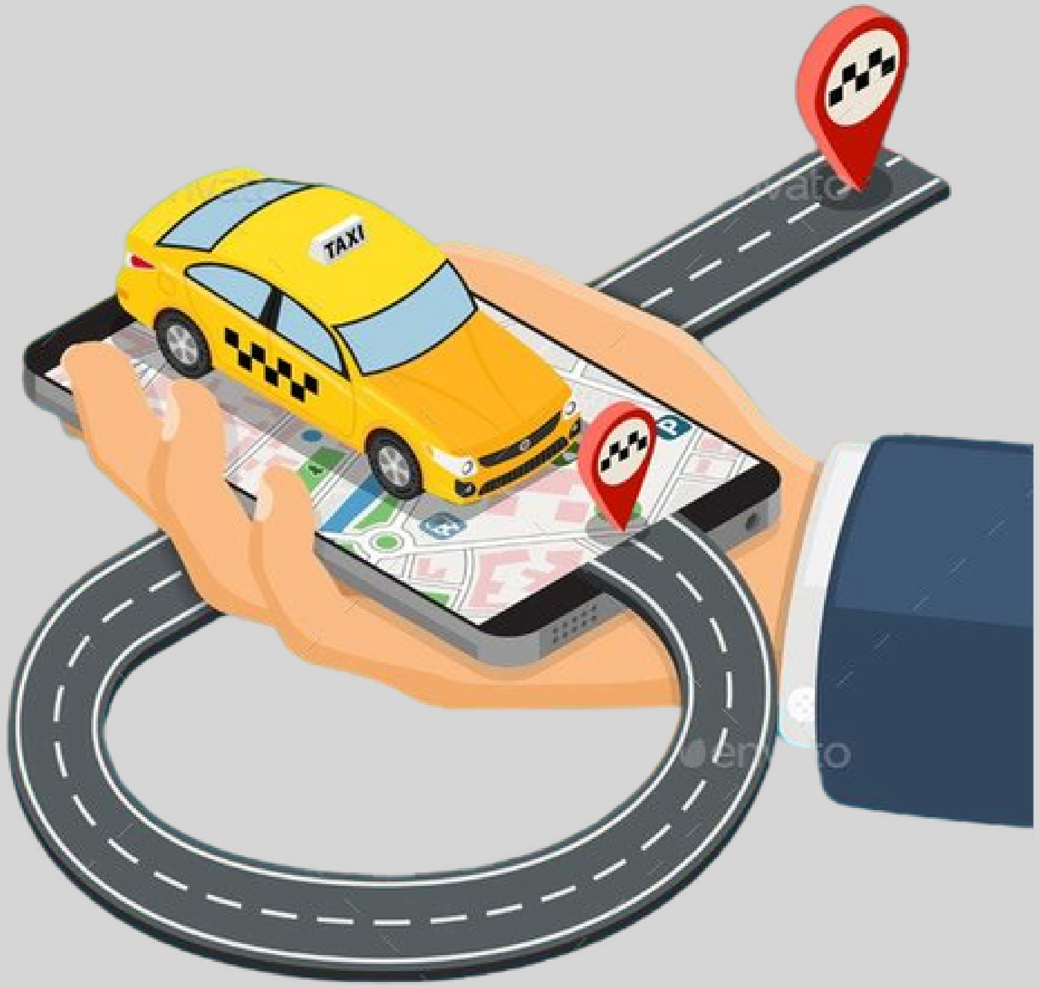
Low Demands	High Demands
RMSE	MAPE
62.5	97.65%



Prezi

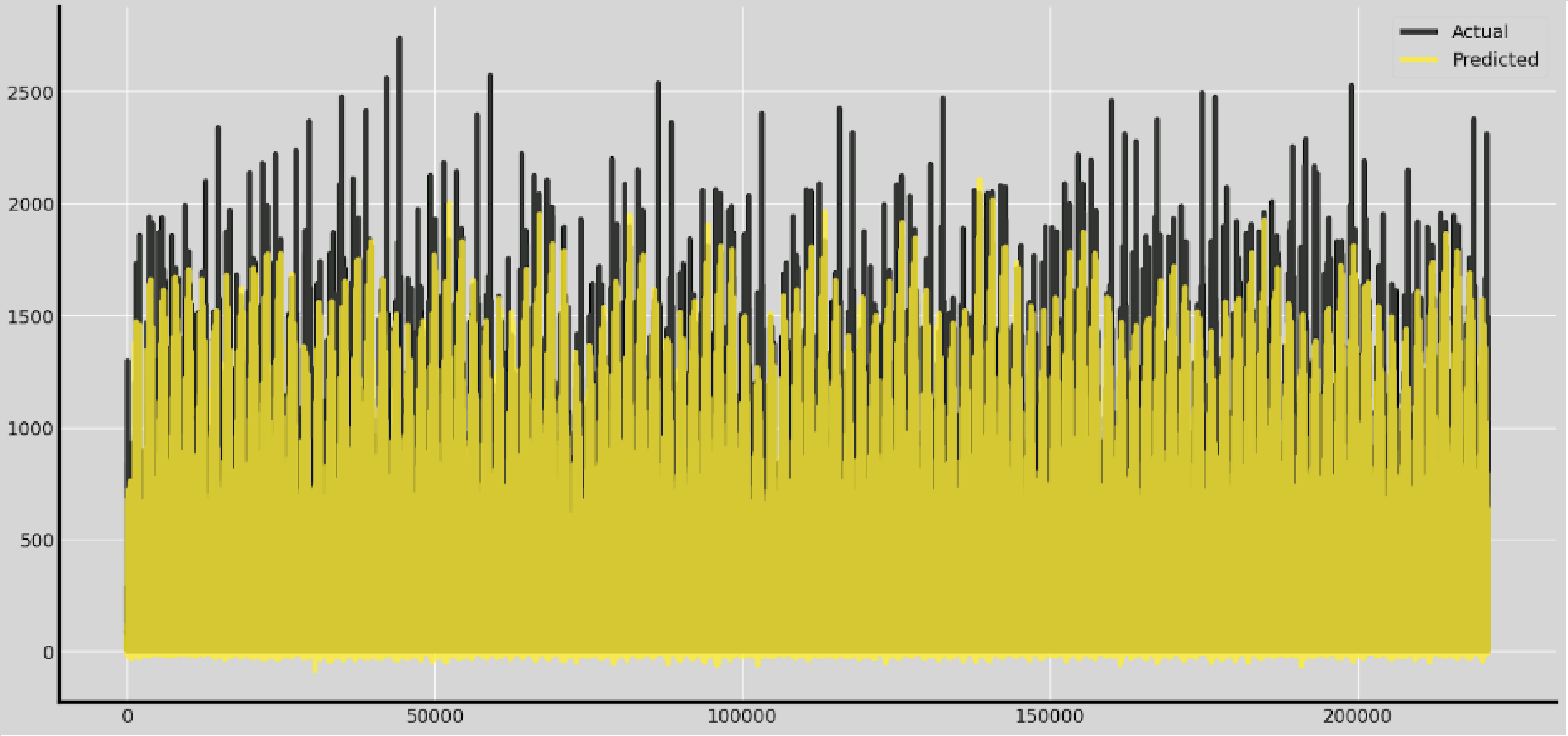
Models

Deep Learning

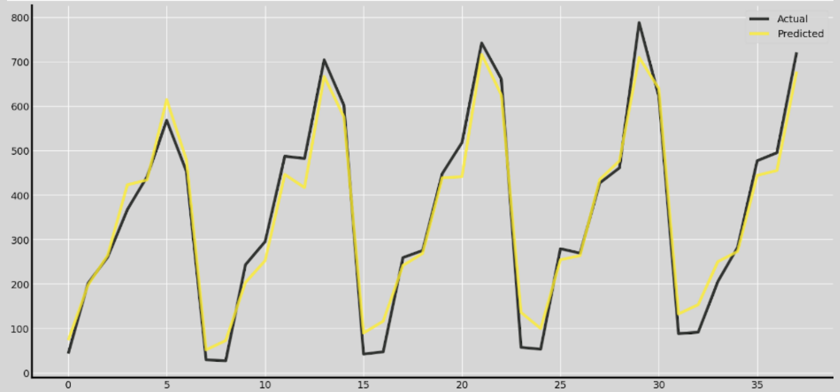


1. Deep Learning with Rolling Window

Conv1D Neural Network

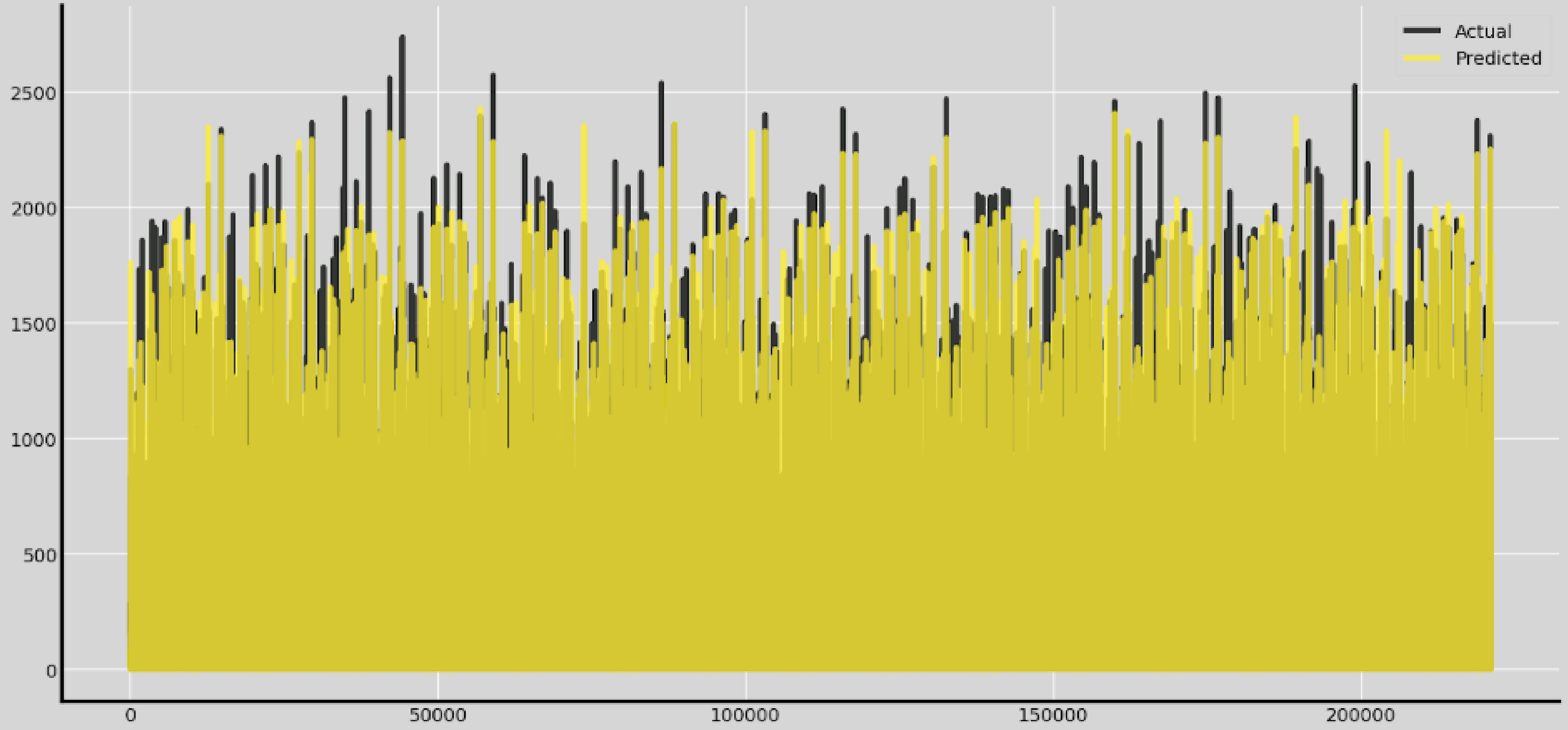


Low Demands	High Demands
RMSE	MAPE
3.81	42.9%

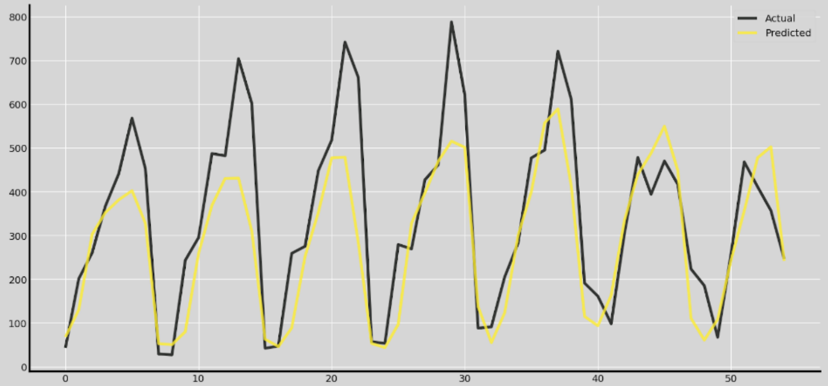


1. Deep Learning with Rolling Window

LSTM Neural Network



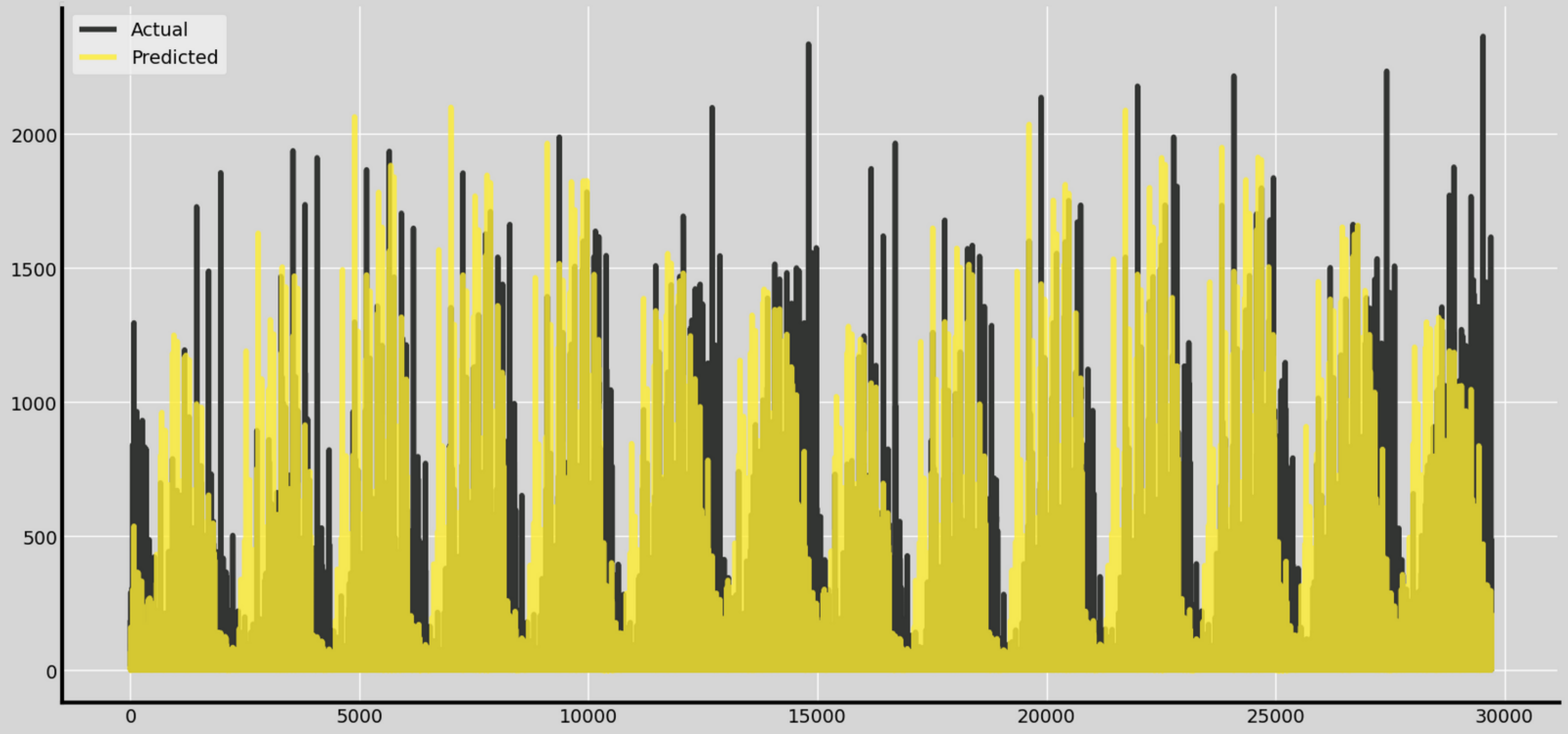
Low Demands	High Demands
RMSE	MAPE
5.28	37%



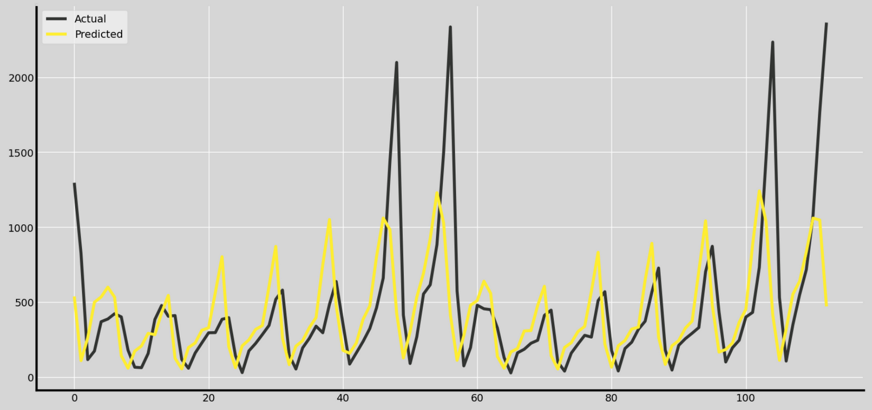
Prezi

1. Deep Learning with Rolling Window

Conv2D Neural Network



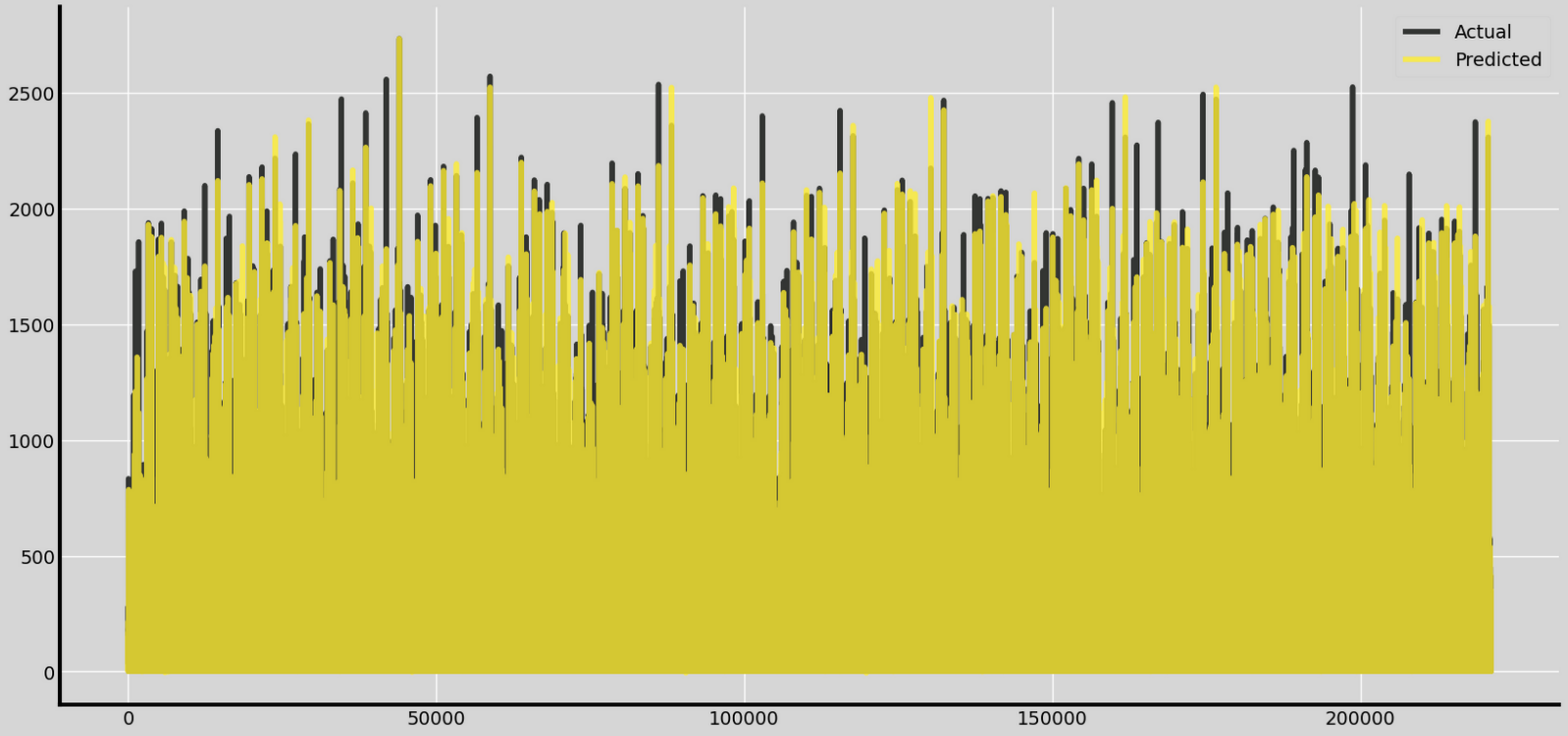
Low Demands	High Demands
RMSE	MAPE
10.778	29.8%



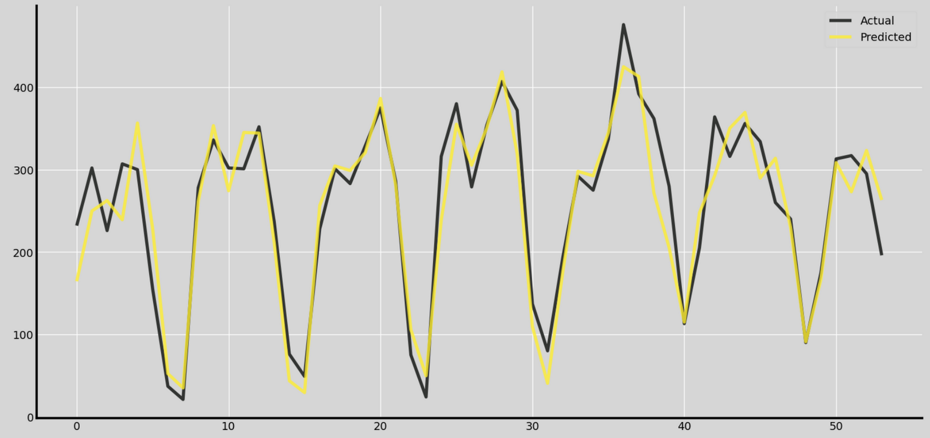
Prezi

2. Simple Deep Learning

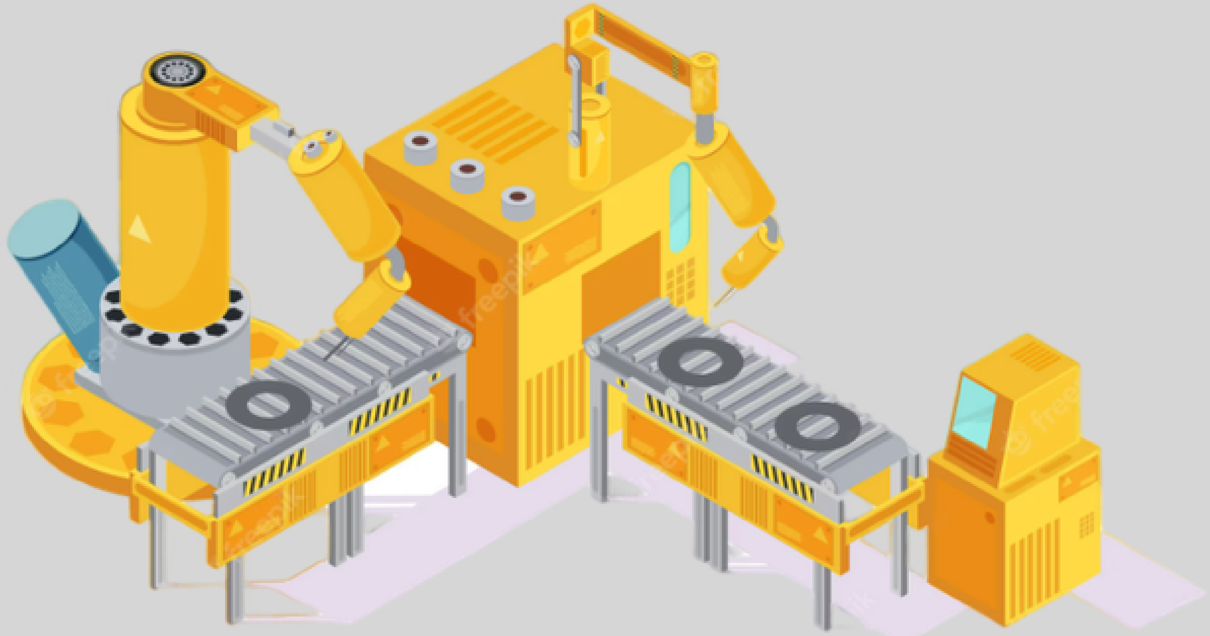
Dense



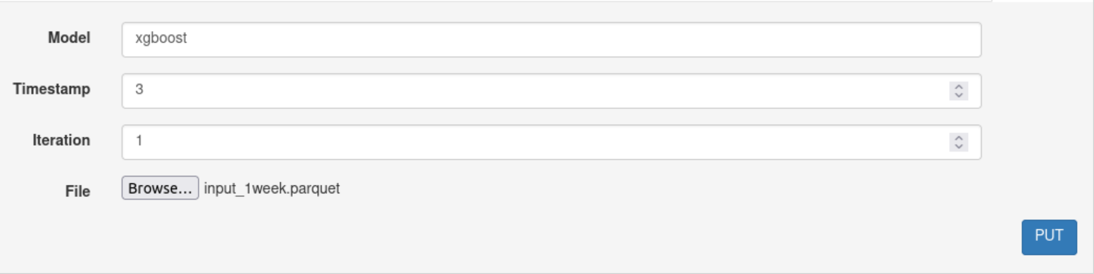
Low Demands	High Demands
RMSE	MAPE
1.75	27%

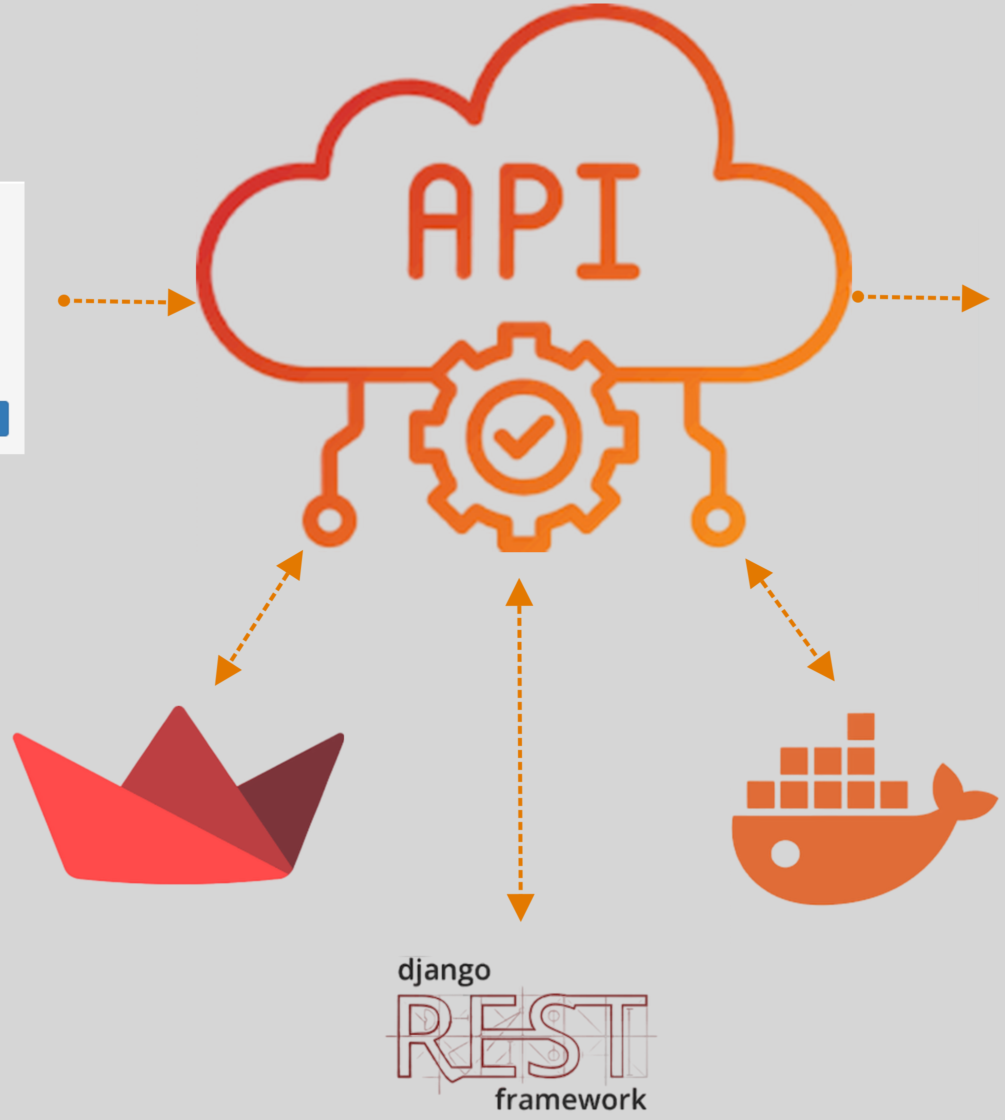


Deployment



Input

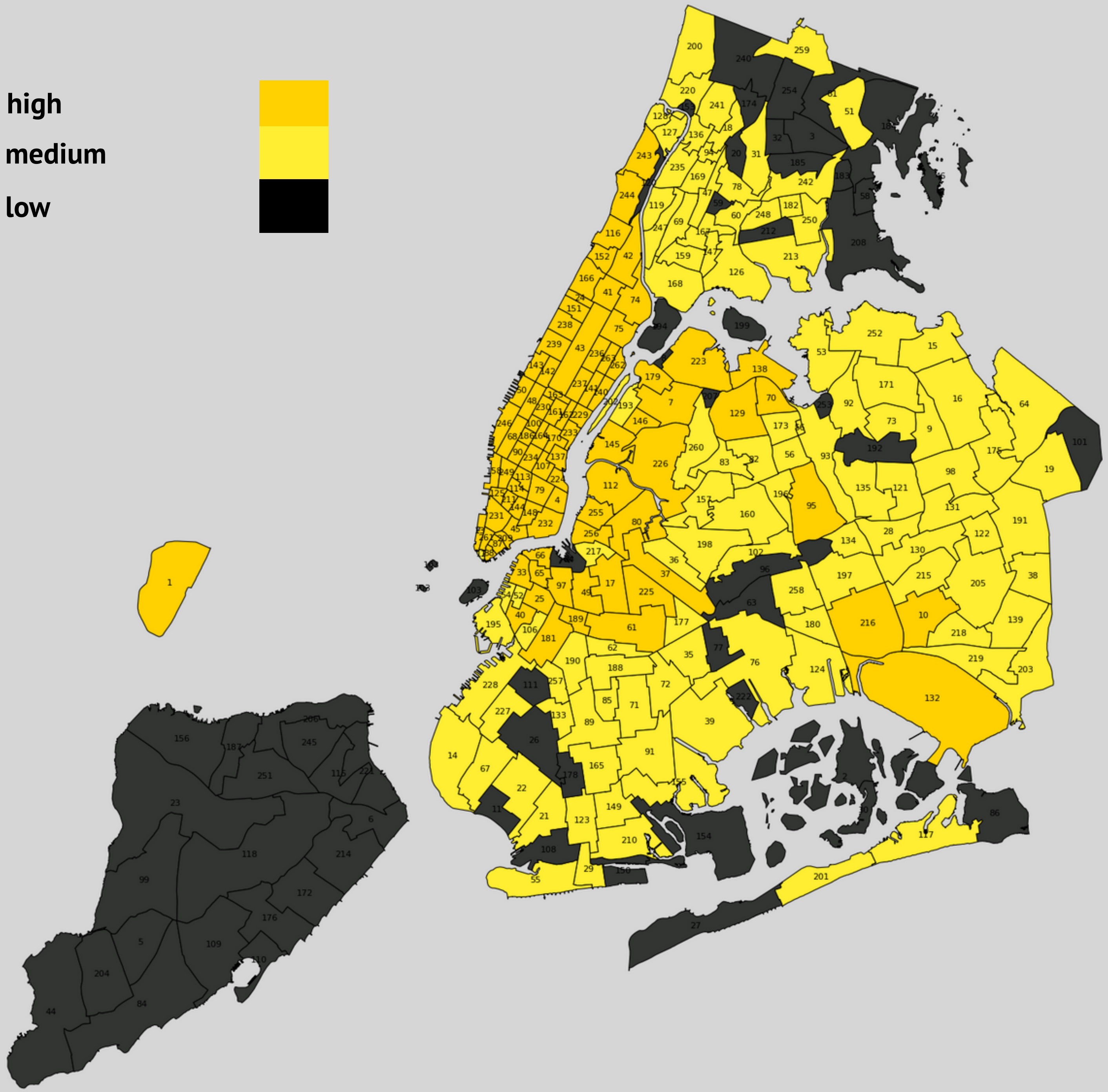




Output

Type something

New York zones based on demand on 2023-4-16



Conclusion

- Different Demand Behavior in Various Times and Locations

- Temporal and Non-Temporal Factors in Demand Forecasting

- Utilizing More Complex Models for Improved Results

- Developing Deep Learning Models for Demand Prediction in Different Locations

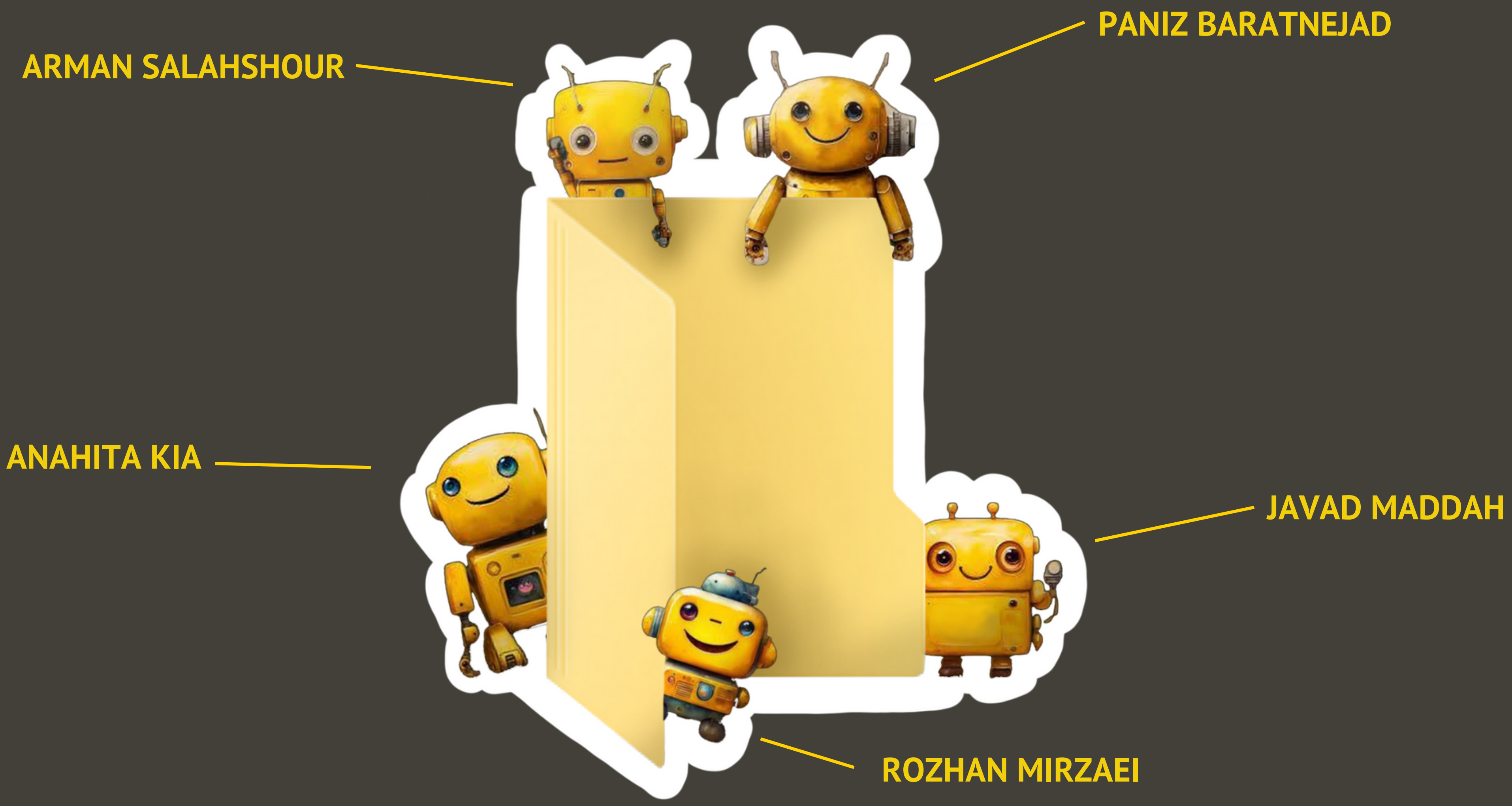
- Model Evaluation and Parameter Refinement

- Deploying the Model in Products for Marketing Team Utilization

We don't copy and paste, we create and innovate

"New folder"

Our team



Tadeh Alexani

Thank you...

Prezi