

# **Automated Forecasting Machine & Workforce Staffing Optimizer**

**Machine Learning System for Demand Forecasting and Staffing Planning**

Karthik Javvaji

MSDS692 Data Science Practicum

Regis University

# Problem Statement

Organizations face problems with better **workforce planning** due to demand fluctuations.

## Challenges

- Service volume demand fluctuates hourly and even seasonally
- Workforce staffing has inefficiencies
- Manual forecasting processes take time
- Understaffing causes **service problems**
- Overstaffing **increases cost**

## Industries impacted:

- Customer Support Centers
- Government services (311 and 911 systems)
- Emergency dispatch centers

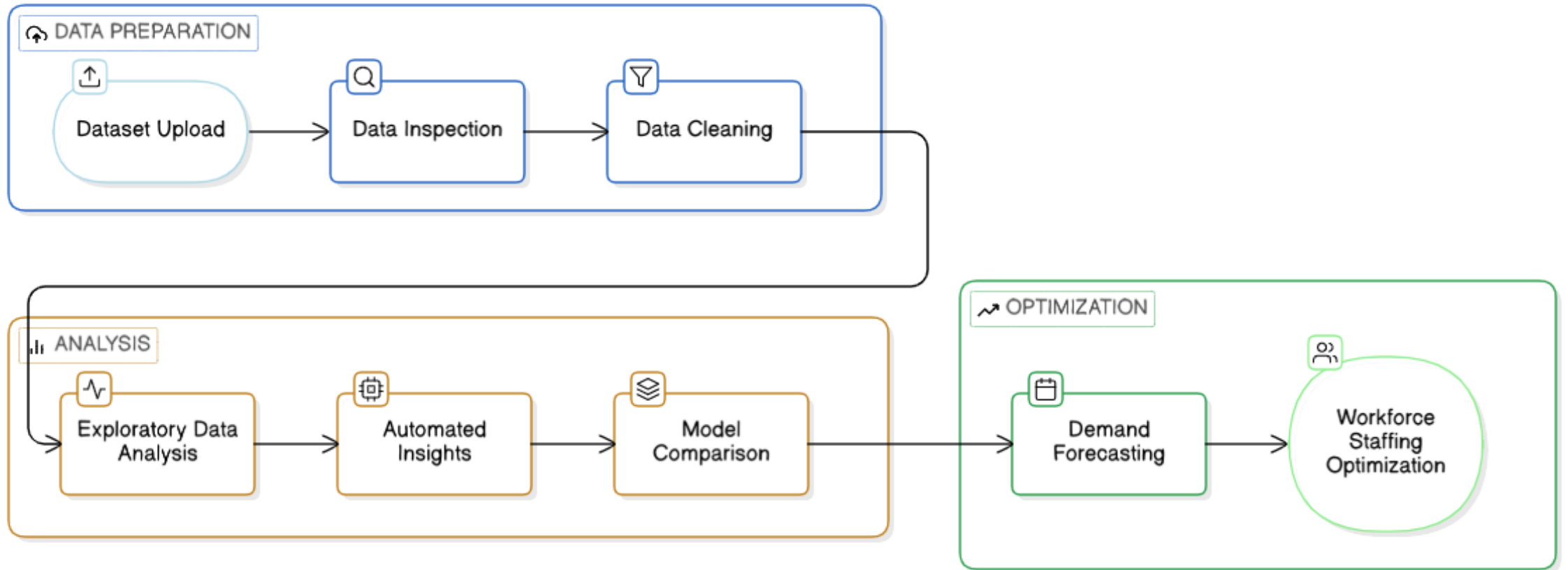
**Goal:** To create an Automated system to **forecast demand** and **Optimize Workforce Staffing**.

# Project Objectives

## To build a forecasting system capable of:

- Inspecting the data of any operational demand datasets like Call or Chat
- Cleaning and deal with missing values in the data set
- Preparing time-series data and work on the EDA visuals
- Run and compare forecasting models on it
- Choosing the best model automatically based on the accuracy
- Forecasting demand and converting forecasts into staffing needs of the team.

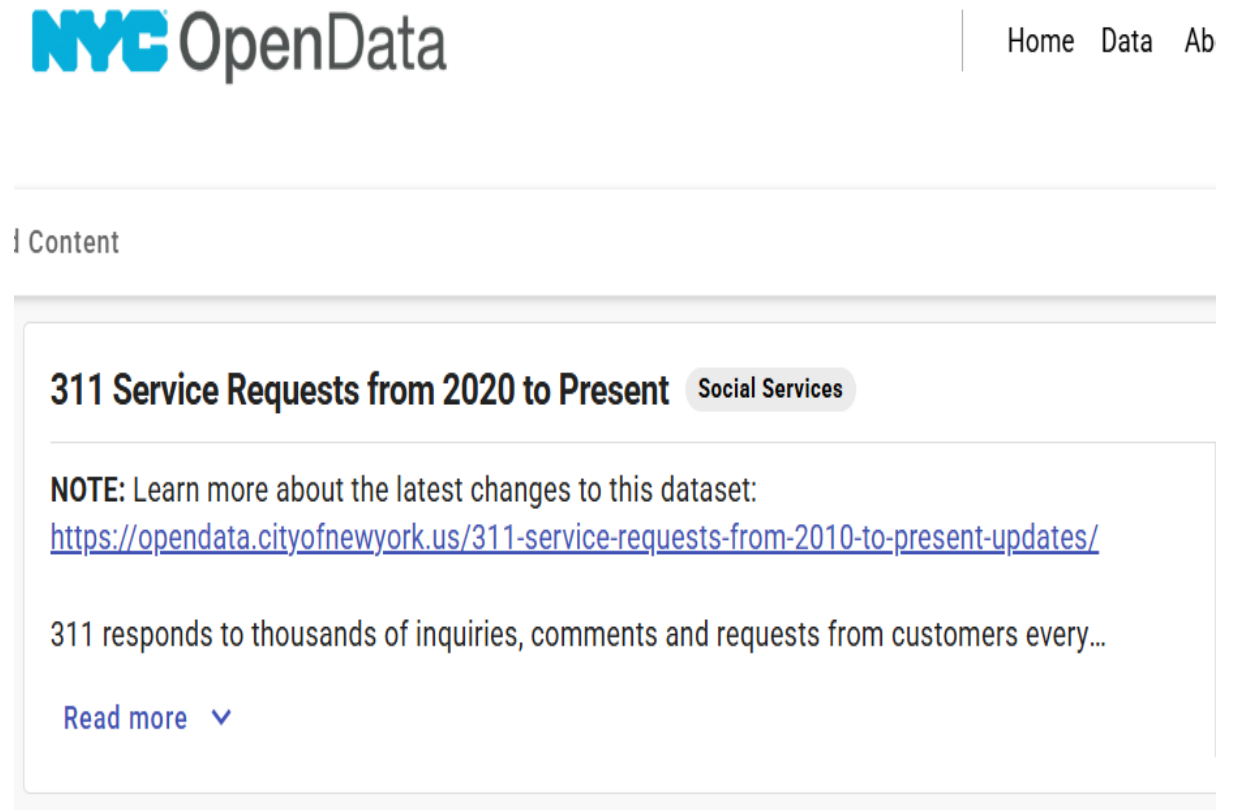
# System Architecture



# Dataset - NYC 311 Requests

Source: [NYC OpenData](#)

- ▶ Records used 255,095 service requests  
Time range: **Jan 2024 – Dec 2025**
- ▶ **Variables:** Created Date, Agency, Complaint Type
- ▶ We will use the request timestamps into **hourly demand signals** for our forecasting.
- ▶ NYC 311 captures non-emergency service requests submitted by residents to city agencies.



The screenshot shows the NYC OpenData website interface. At the top right, there are navigation links for 'Home', 'Data', and 'Ab'. The main content area features a header for the dataset '311 Service Requests from 2020 to Present' with a 'Social Services' tag. Below the header, there is a 'NOTE' section with a link to 'https://opendata.cityofnewyork.us/311-service-requests-from-2010-to-present-updates/'. The text below the link reads '311 responds to thousands of inquiries, comments and requests from customers every...'. At the bottom of the note section, there is a 'Read more' link with a downward arrow.

# Data Inspection

	Metric	Value
0	Rows	255095
1	Date min	2024-01-
2	Date max	2025-12-
3	Hour unique count	24
4	Hour min	0
5	Hour max	23
6	Missing Created Date	0

---

Key metrics:

---

Rows: **255,095**

---

Date Range:  
Jan 2024 – Dec 2025

---

Hourly Coverage:  
0 – 23 (complete 24-hour cycle)

---

Missing timestamps: **0**

---

This helps us to make sure dataset is in **valid time-series and good for modeling.**

# Data Cleaning Pipeline

	Step	Result
0	Missing values removed	0
1	Outliers capped	22
2	Smoothing applied	True

---

Automated preprocessing includes:

- Missing value removal
- Outlier detection and capping
- Time-series smoothing

---

Results:

Missing values removed: **0**

Outliers capped: **22**

Smoothing applied: **True**

---

Purpose:

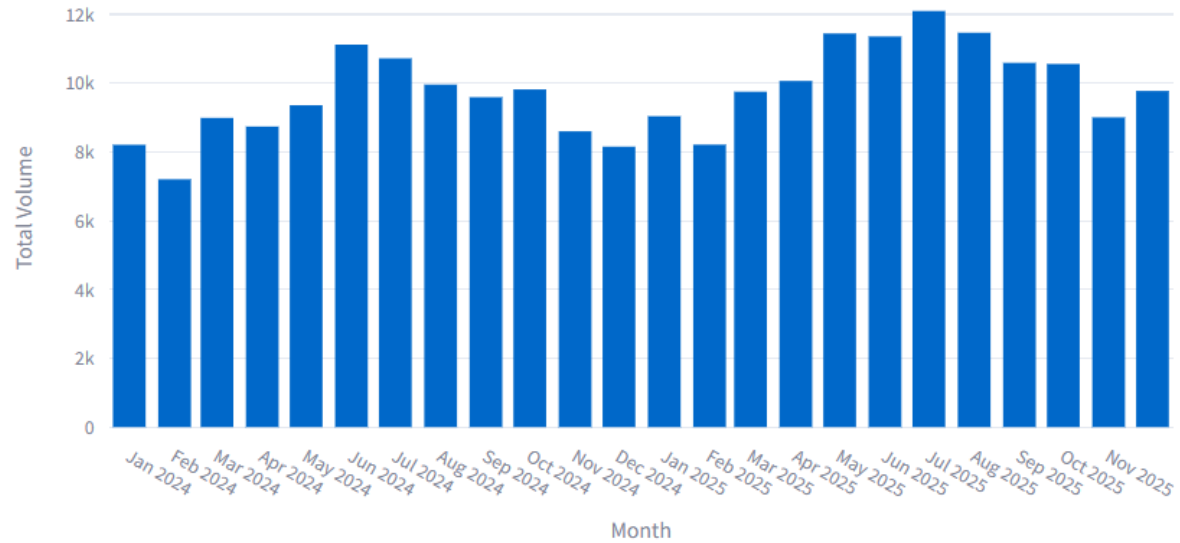
This helps the forecasting models get stable data to train on which improves accuracy

# Exploratory Data Analysis

## Monthly Demand Trend (2024–2025)

Service ticket volume shows moderate seasonality, peaking around mid-2025 with ~12K requests per month.

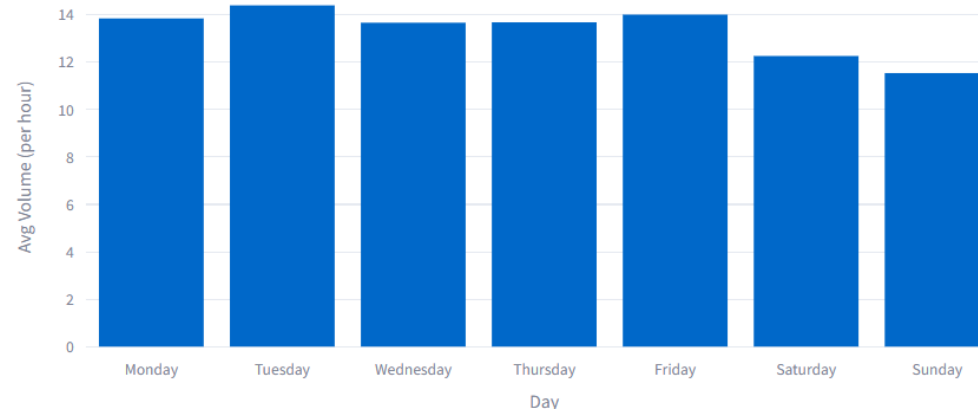
Monthly Demand (Month-Year)



## Weekly Demand Pattern

Service requests are highest on weekdays, with Tuesday showing the highest average demand.

Average Demand by Day of Week



# Automated Insights

## Automated Insights

- Peak month was Jul 2025 with 12,096 total requests.
- Demand has increased 6.5% (last 4 weeks vs previous 4 weeks).
- Peak hour is around 10:00 with 21.0 avg requests/hour.
- Highest day is Tuesday with 14.4 avg requests/hour.
- Variability: avg 13.3/hr, 95th percentile 27/hr (use this for buffer planning).
- Top driver: Illegal Parking contributes 32,492 requests (top category).
- WFM note: schedule more coverage during peak hour + peak day windows; use P95 as a service-level buffer reference.

1. It automatically generates insights by using the statistical summaries from the time-series data.
2. These pointers can help to inform staffing and capacity planning decisions at quick glance.

# Automated Model Comparison

Run Model Comparison

## Model Leaderboard

	Model	wMAPE	MAPE	RMSE	MAE
4	XGBoost	18.4908	29.0941	3.0735	2.2205
3	LinearRegression	19.2461	30.6466	3.12	2.3112
5	Prophet	51.0096	93.1678	7.4314	6.1257
1	ETS	53.831	102.6519	7.7665	6.4645
0	SeasonalNaive	57.0756	72.8779	9.8771	6.8542
2	MovingAverage	64.8628	180.5764	9.2789	7.7893

Best model selected: XGBoost

Models tested:

- XGBoost
- Linear Regression
- Prophet
- ETS
- Seasonal Naive
- Moving Average

Evaluation metrics:

- wMAPE
- MAPE
- RMSE
- MAE

The system automatically selects the **best performing model based on forecasting accuracy.**

# Automated Model Comparison

Run Model Comparison

## Model Leaderboard

	Model	wMAPE	MAPE	RMSE	MAE
4	XGBoost	18.4908	29.0941	3.0735	2.2205
3	LinearRegression	19.2461	30.6466	3.12	2.3112
5	Prophet	51.0096	93.1678	7.4314	6.1257
1	ETS	53.831	102.6519	7.7665	6.4645
0	SeasonalNaive	57.0756	72.8779	9.8771	6.8542
2	MovingAverage	64.8628	180.5764	9.2789	7.7893

Best model selected: XGBoost

## Best Performing Model: XGBoost

This model achieved the **lowest forecasting error across evaluation metrics.**

Machine learning models such as XGBoost are effective because they can capture:

- Non-linear demand patterns
- Seasonal behavior
- Hourly fluctuations

# Demand Forecast

Using the best performing model (**XGBoost**), the system generates a **6-month hourly demand forecast**.

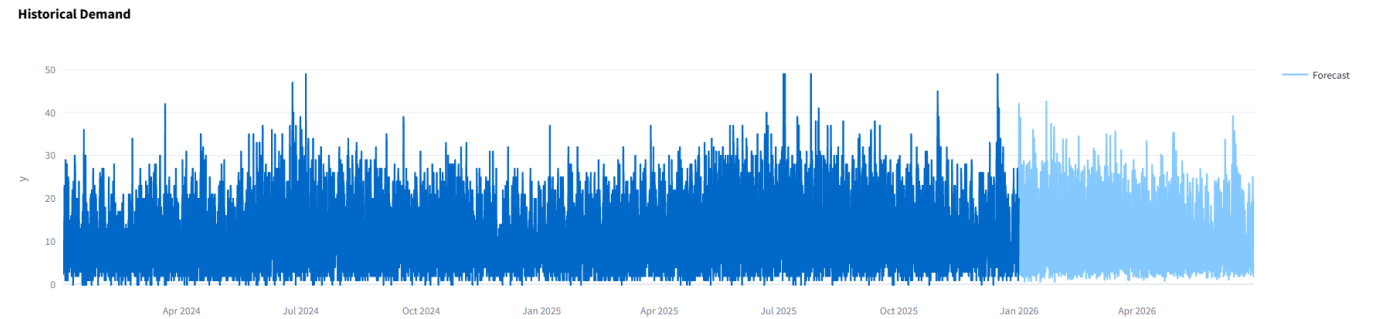
## Forecast Summary

- Average demand: **14.2 requests/hour**
- Peak demand: **42.7 requests/hour**
- Minimum demand: **0.4 requests/hour**

## Key Observation

- The forecast captures **daily demand cycles and variability**, enabling better operational planning.

## Forecast Visualization



## Forecast Summary Metrics

Average Demand

14.2

Peak Demand

42.7

Minimum Demand

0.4

# Workforce Staffing Optimization

Forecasted demand is converted into **required staffing levels** using workforce planning parameters.

## Operational Inputs

Average Handle Time (AHT): **360 seconds**

Shrinkage: **30%**

Occupancy Target: **85%**

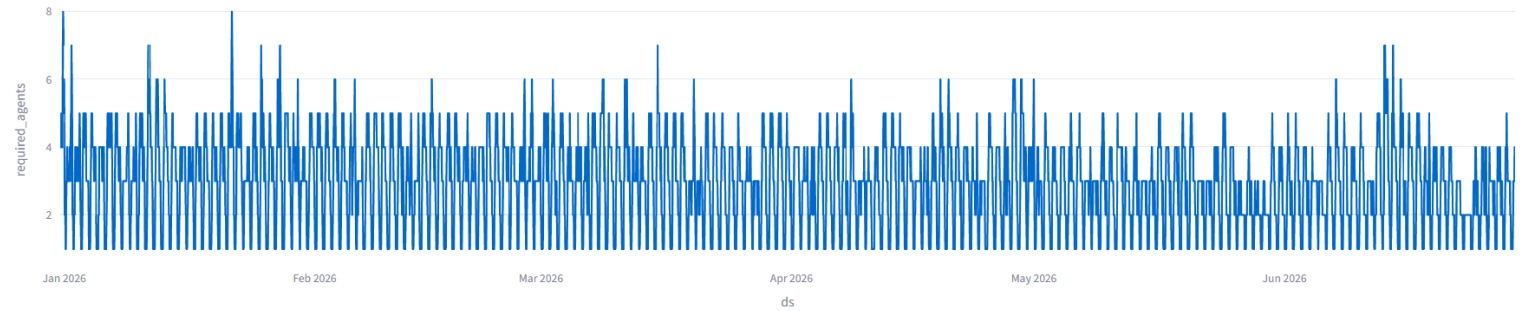
## Output

Required **agents per hour**

## Example

**21 requests → 4 agents required**

Required Staffing Over Time



	ds	yhat	required_agents
0	2025-12-31 13:00:00	18.9277	4
1	2025-12-31 14:00:00	19.9104	4
2	2025-12-31 15:00:00	20.4723	4
3	2025-12-31 16:00:00	28.9652	5
4	2025-12-31 17:00:00	24.5065	5
5	2025-12-31 18:00:00	20.4349	4
6	2025-12-31 19:00:00	20.1772	4
7	2025-12-31 20:00:00	35.8663	7
8	2025-12-31 21:00:00	42.1594	8
9	2025-12-31 22:00:00	38.7574	7

# System Validation (Test Dataset)

To test generalization, the system was applied to a second dataset.

## Dataset

- Montgomery County **911 Emergency Calls**
- Records analyzed: **663,522**

## Validation Results

- Automated **data inspection completed**
- Multiple **forecasting models evaluated**
- **Demand forecasts generated**
- **Staffing requirements calculated**

## Key Takeaway

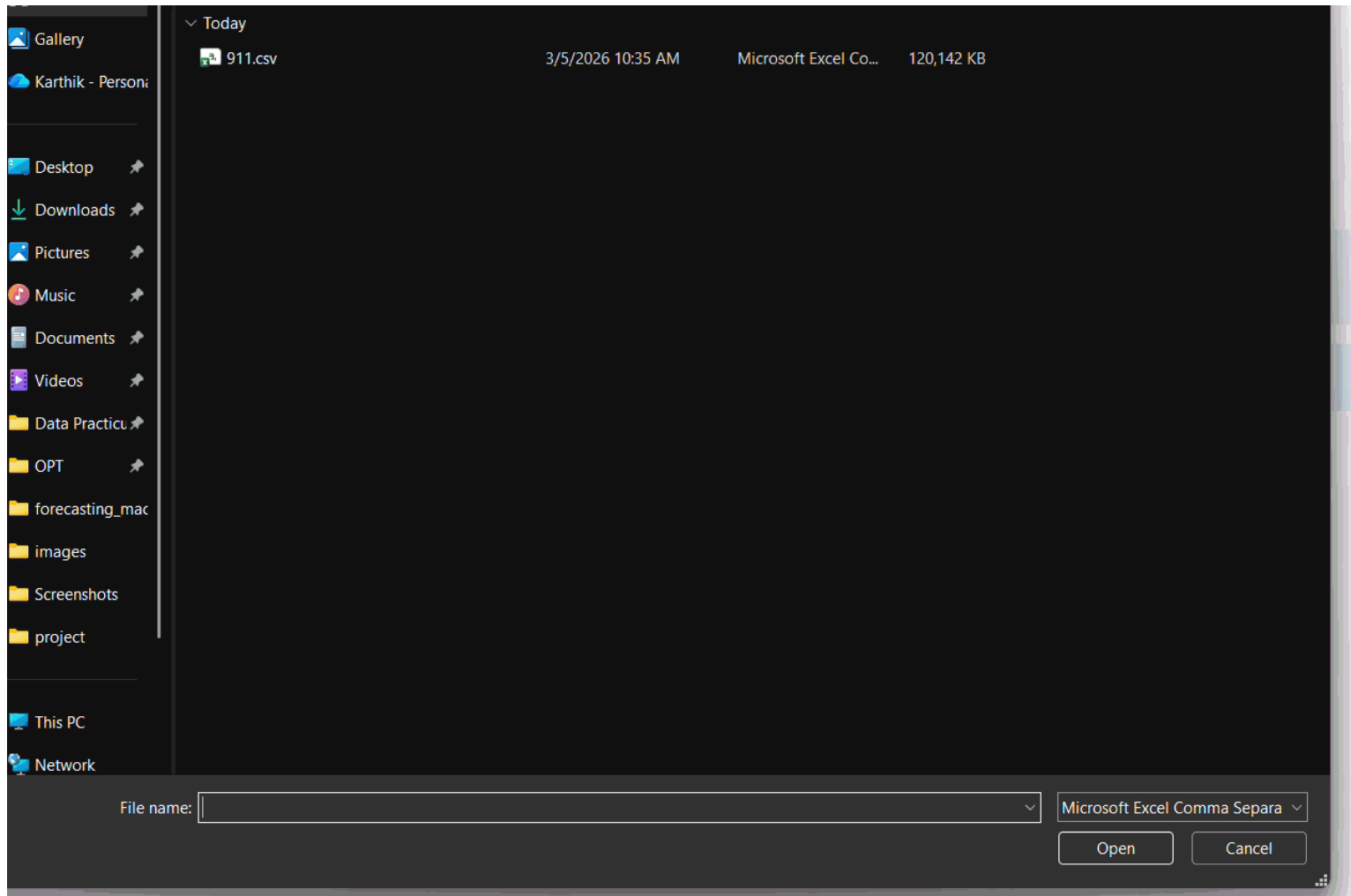
- The forecasting machine performs successfully across **different service request datasets**.

## Model Leaderboard


	Model	wMAPE	MAPE	RMSE	MAE
4	XGBoost	12.6082	14.9315	2.3276	1.6935
3	LinearRegression	12.9206	15.6952	2.3426	1.7354
0	SeasonalNaive	29.0051	30.9674	5.7183	3.8958
2	MovingAverage	40.0355	66.2175	6.4024	5.3774
5	Prophet	43.9718	52.6491	7.1497	5.9061
1	ETS	78.5357	107.2481	11.0731	10.5486

Best model selected: XGBoost

# Interface:



Deploy ⋮


Browse files 



## Conclusion

“This system helps join forecasting and operational workforce planning by automatically transforming demand data into actionable staffing recommendations without technical expertise.

## Project Summary

- This project demonstrates an end-to-end data science system that:
  - Automates demand forecasting
  - Compares multiple machine learning models
  - Selects the best model automatically
  - Converts demand forecasts into staffing recommendations
- 

## References:

- NYC Open Data. *311 Service Requests Dataset*. <https://opendata.cityofnewyork.us>
- Montgomery County Open Data. *911 Emergency Call Dataset*. <https://data.montgomerycountymd.gov>
- Vasudev, P., & Thomas, I. (2024). *Important call center KPI benchmarks by industry*. Sprinklr. Retrieved from <https://www.sprinklr.com/blog/call-center-kpi-benchmarks-by-industry/>
- Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), 785–794. <https://doi.org/10.1145/2939672.2939785>
- Taylor, S. J., & Letham, B. (2017). *Forecasting at Scale*. PeerJ Preprints. <https://peerj.com/preprints/3190.pdf>
- Hyndman, R. J., & Koehler, A. B. (2006). *Another Look at Measures of Forecast Accuracy*. **International Journal of Forecasting**, 22(4), 679–688. <https://www.sciencedirect.com/science/article/abs/pii/S0169207006000239?via%3Dihub>
- Gans, N., Koole, G., & Mandelbaum, A. (2003). *Telephone call centers: Tutorial, review, and research prospects*. **Manufacturing & Service Operations Management**, 5(2), 79–141. <https://pubsonline.informs.org/doi/10.1287/msom.5.2.79.16071>
- Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S., & Zhao, L. (2005). *Statistical analysis of a telephone call center: A queueing-science perspective*. **Journal of the American Statistical Association**, 100(469), 36–50. <https://www.tandfonline.com/doi/abs/10.1198/016214504000001808>



**Thank you**