FutureCart: AI-Driven Demand Prediction for Smarter Retail.

NAME-SWASTIK ROY CHOUDHURY

Project Overview

Objective:

Develop an AI-driven demand prediction model for the retail sector.

Key Focus:

Address dynamic consumer behavior, seasonal trends, and external factors.

Deliver actionable insights through cutting-edge time series modeling techniques.

Approach:

Comprehensive EDA and effective data preprocessing.

Integration of advanced statistical and machine learning models.

Goal:

Create a scalable, adaptable system for accurate forecasts and meaningful insights.

Impact:

Smarter inventory management, pricing strategies, and optimized retail operations.

Problem Statement

Challenges:

>Volatility in consumer demand leads to overstocking, understocking, and financial inefficiencies.
>Traditional methods fail to address time-sensitive factors and external shocks.

Solution:

>Design a domain-specific, intelligent demand prediction system.
>Employ ARIMA, SARIMA, and multivariate regression models to capture retail trends.

Outcome:

>Enhanced operational efficiency, reduced waste, and increased profitability.
>Empower retailers with data-driven foresight to innovate and thrive.

Outcomes

Deliverables:

Reliable forecasting system with a user-friendly interface for visualizing trends and insights.

Key Benefits:

>Enhanced Decision-Making: Informed inventory, pricing, and promotion strategies.

>**Operational Efficiency**: Reduced costs from overstocking/understocking.

>**Consumer Satisfaction:** Better inventory alignment with demand, minimizing stockouts.

>Scalability: Adaptable to diverse product lines and seasonal variations.

>Behavioral Insights: Discover hidden patterns for strategic planning.

Impact:

Demonstrates how AI and data analytics transform retail by converting raw data into actionable intelligence.

Data Sources and Project Workflow

Data Collection:

❍ I sourced historical sales data, clicks, impressions, and time-related variables from credible repositories.

❍ Focused on ensuring data accuracy and representativeness to capture consumer behavior effectively.

❍ Documented all data sources to maintain transparency and reproducibility.

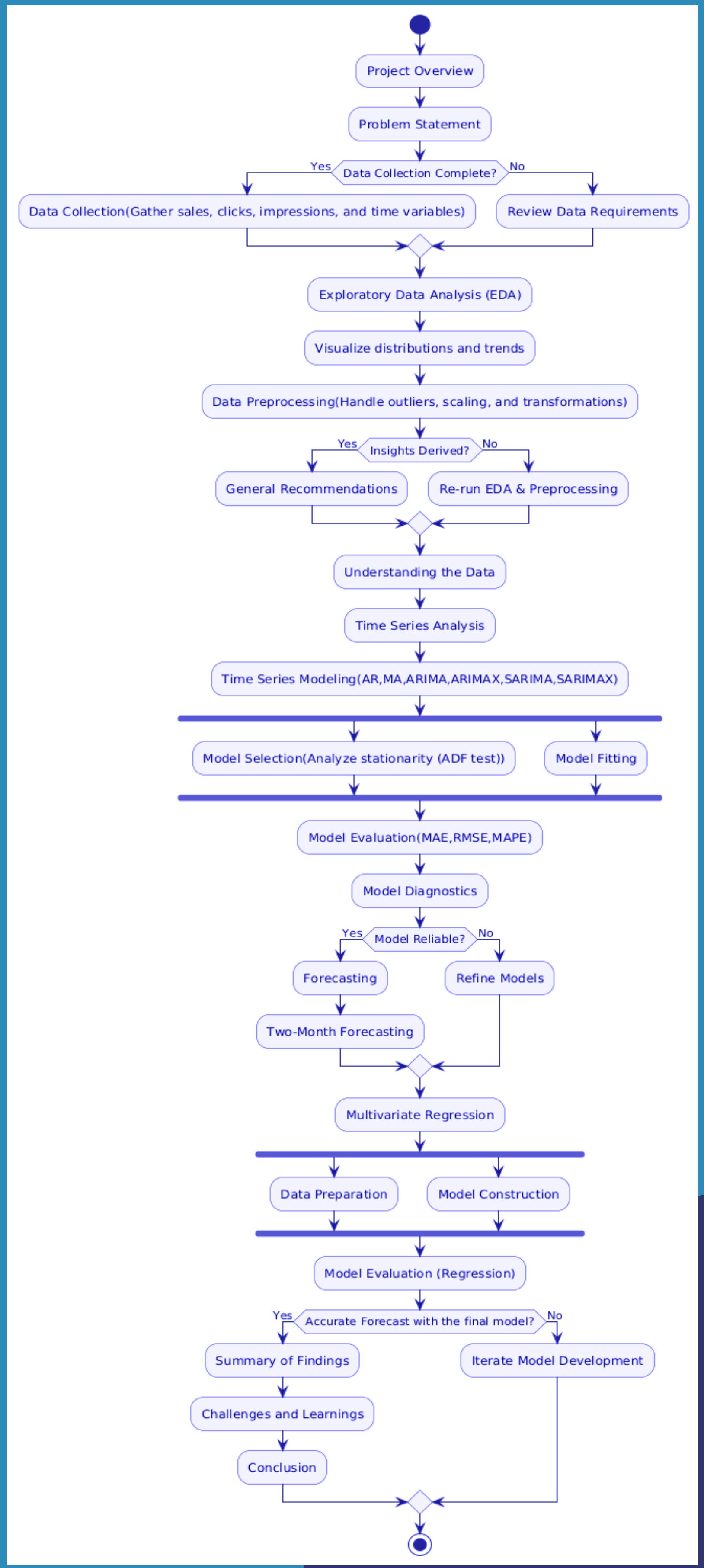
Project Workflow:

❍Mapped the entire workflow of my project, including data collection, EDA, preprocessing, modeling, and forecasting.

❍ Used a systematic flow to ensure clarity and logical sequencing of each phase.

Libraries Used

1. IPython - Interactive computing.
2. concurrent - Concurrent task management.
3. google - Integration with Google services.
4. itertools - Efficient looping constructs.
5. matplotlib - Data visualization.
6. multiprocessing - Parallel task execution.
7. numpy - Numerical computations.
8. pandas - Data manipulation and analysis.
9. seaborn - Advanced data visualization.
10. sklearn - Machine learning algorithms.
11. statsmodels - Statistical modeling and time series analysis.
12. warnings - Handling and filtering warnings.



Exploratory Data Analysis (EDA)

- Conducted a comprehensive analysis to uncover patterns and trends.

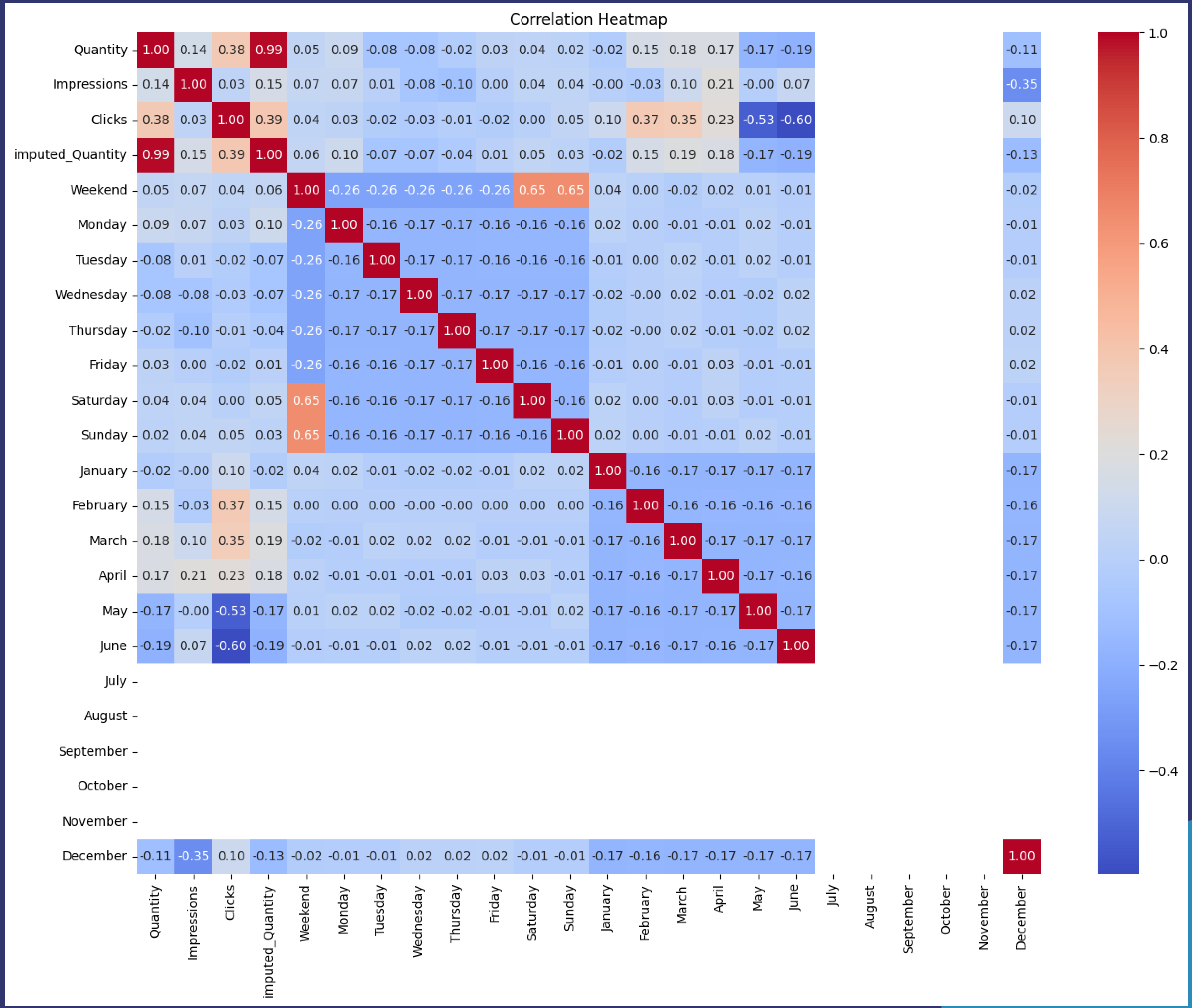
- Visualized sales data, clicks, and impressions to identify:

  >Seasonal trends (e.g., holiday spikes).

  >Correlations between clicks and impressioans with sales demand.

  >Anomalies, such as outliers linked to specific campaigns or promotions.

- Tools used: matplotlib and seaborn for detailed visualizations.



Data Preprocessing & Cleaning

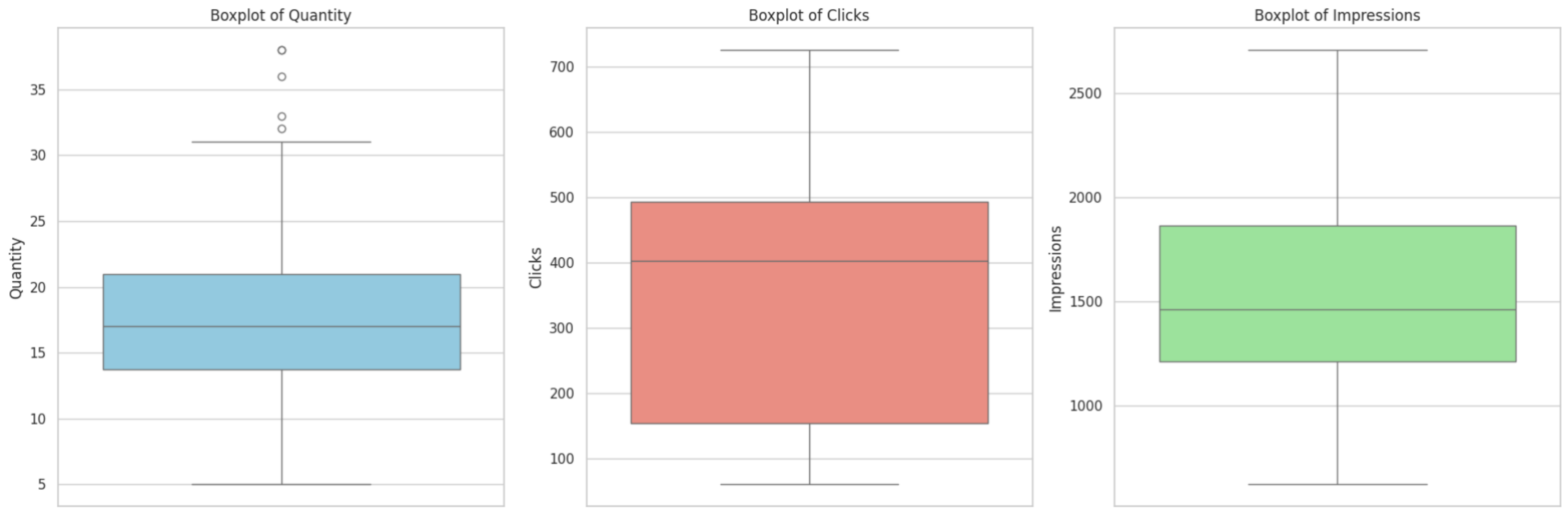
Data Cleaning:

Addressed missing values with mean/median imputation for numerical variables.

Applied mode or predictive methods for categorical variables.

Outlier Detection:

Used interquartile range (IQR) and Z-scores to identify and decide on retaining or removing outliers.



Feature Engineering:

Added derived metrics like click-through rates (CTR) and impression-to-conversion ratios to enrich the dataset.

Data Transformation:

Applied normalization and log transformation to skewed variables for better model performance.

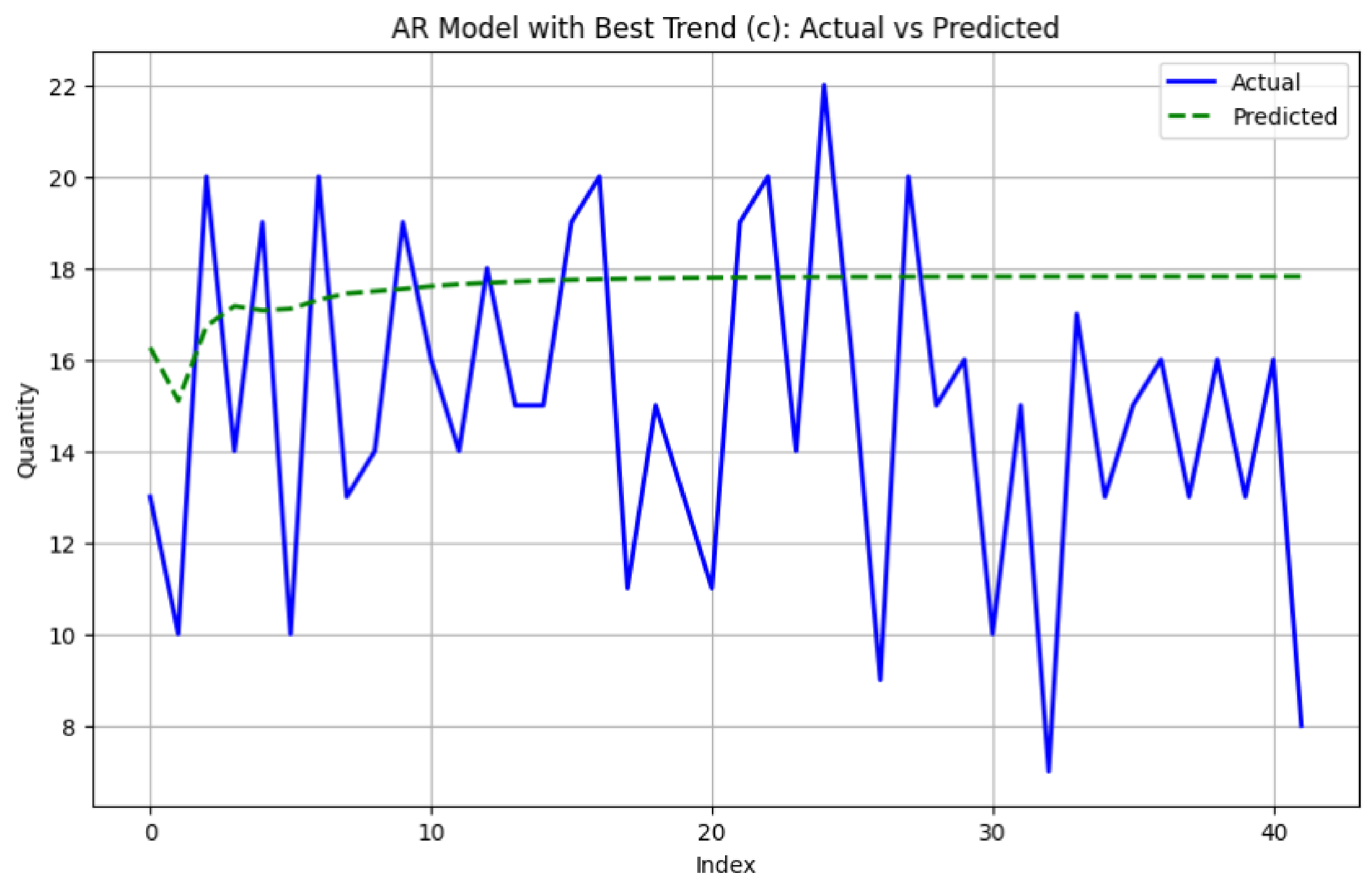
Time Series Modeling (AR & MA)

Autoregressive (AR):

>>Utilized past values to predict future demand.

>>Determined lag order using PACF (Partial Autocorrelation Function).

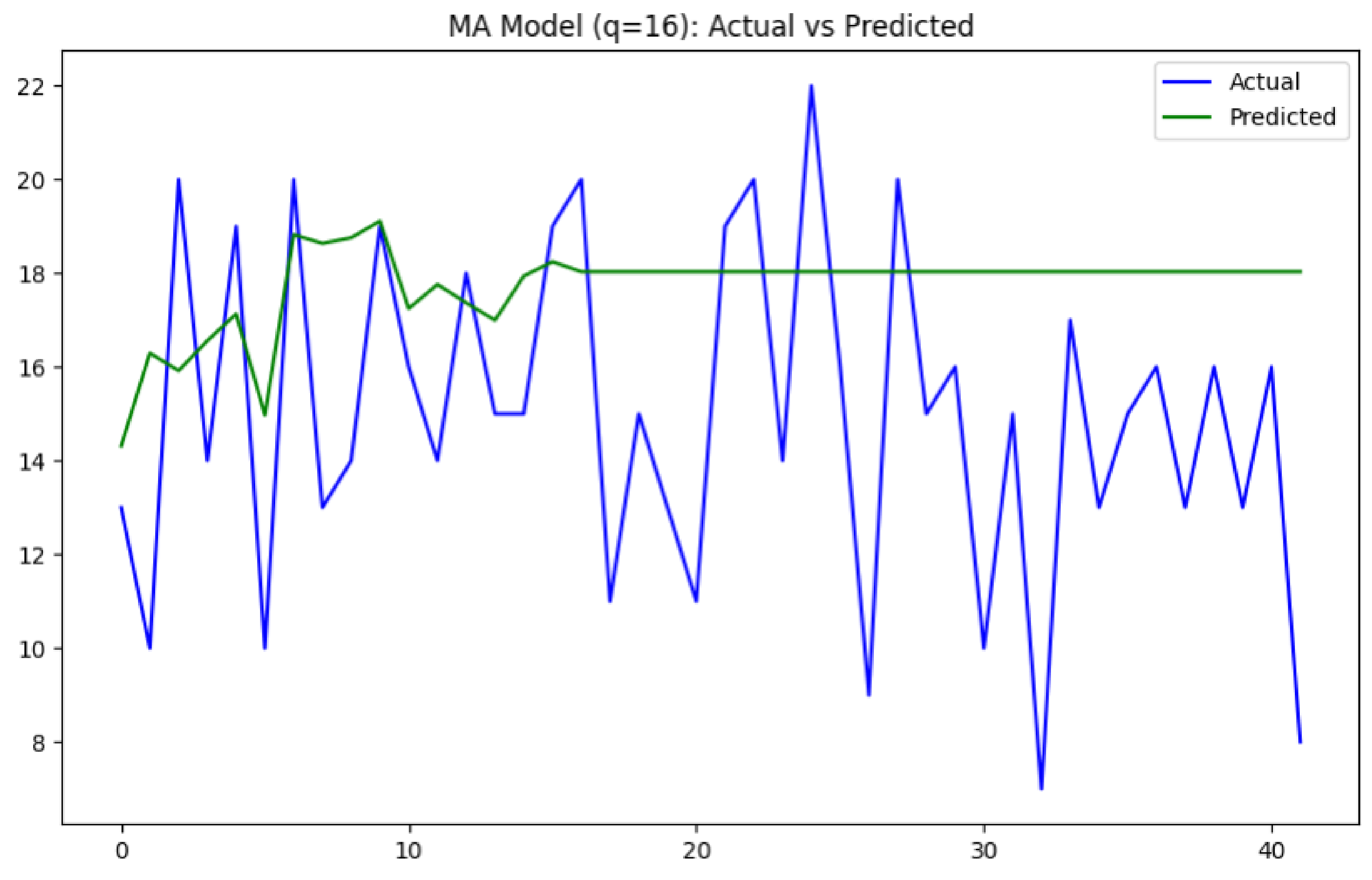
>>Effective for datasets with strong temporal dependencies.



Moving Average (MA):

>>Modeled the relationship using past error terms to smooth irregularities.

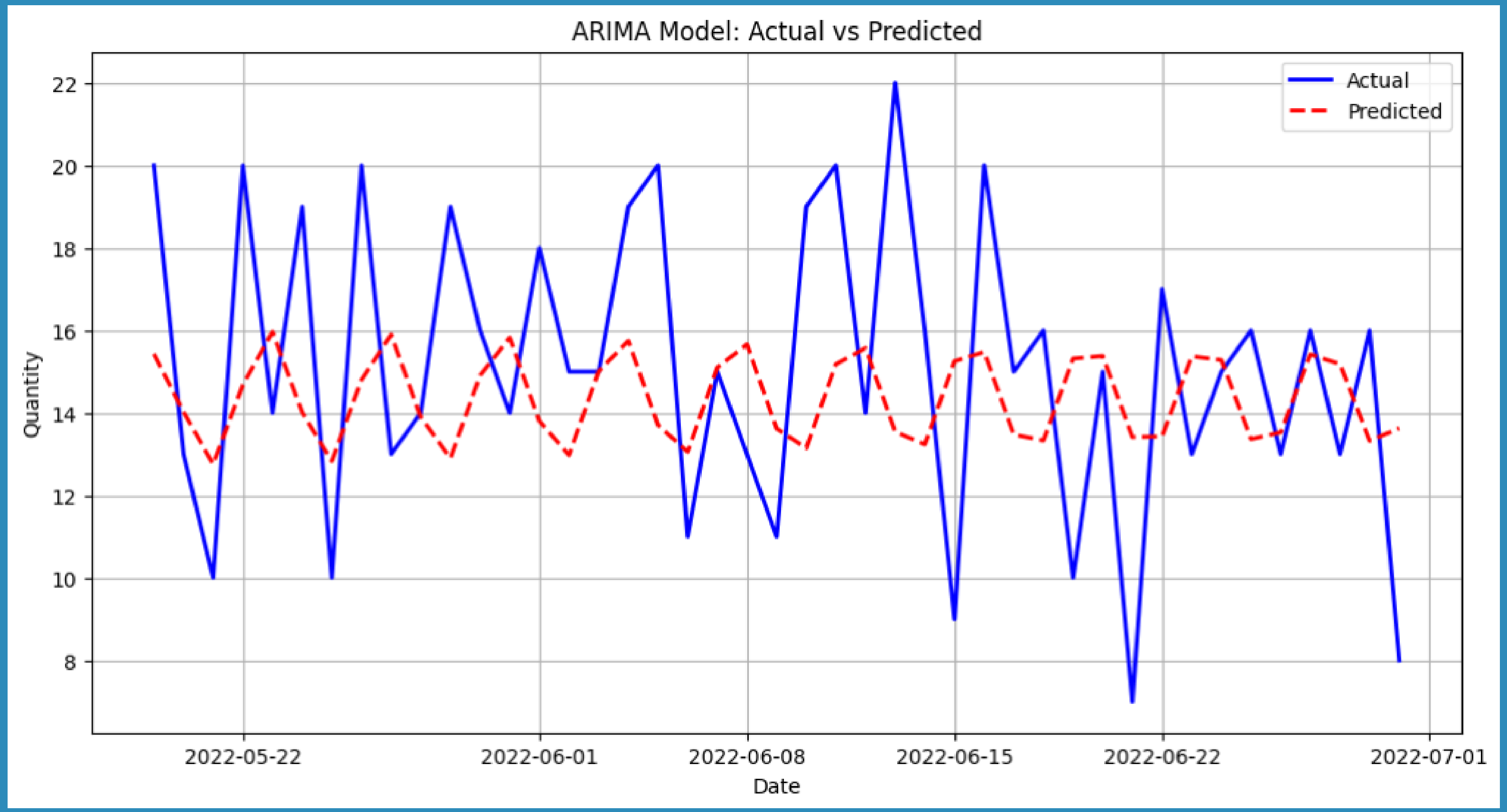
>>Applied ACF (Autocorrelation Function) to determine lag order.



ARIMA & ARIMAX

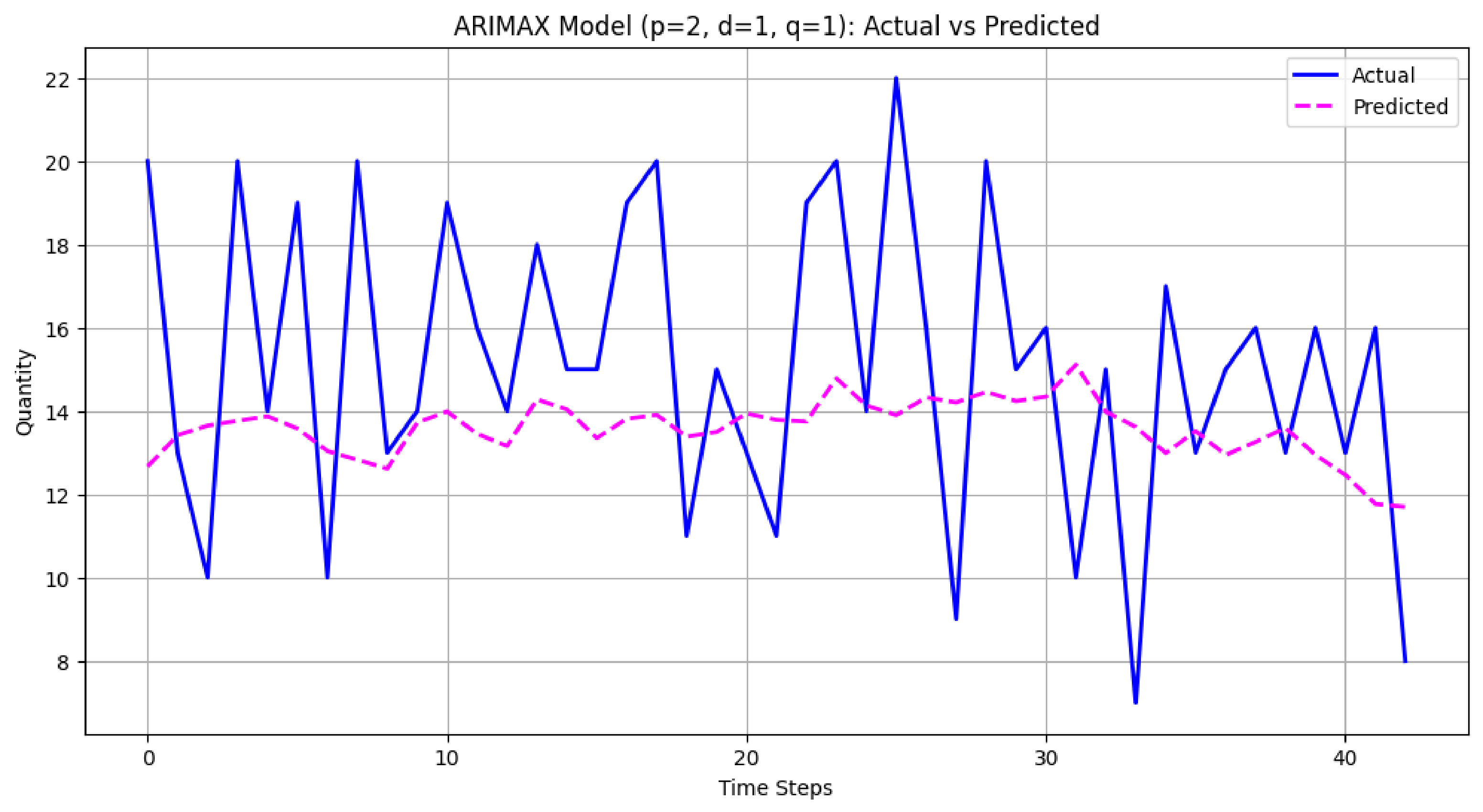
ARIMA (Autoregressive Integrated Moving Average):

>>Addressed non-stationary data by applying differencing and capturing long-term trends.
>>Tuned parameters (p, d, q) using grid search for optimal performance.



ARIMAX (ARIMA with Exogenous Variables):

>>Incorporated external predictors (clicks and impressions) to improve accuracy.
>>Enhanced demand predictions by leveraging the interplay between marketing efforts and sales.

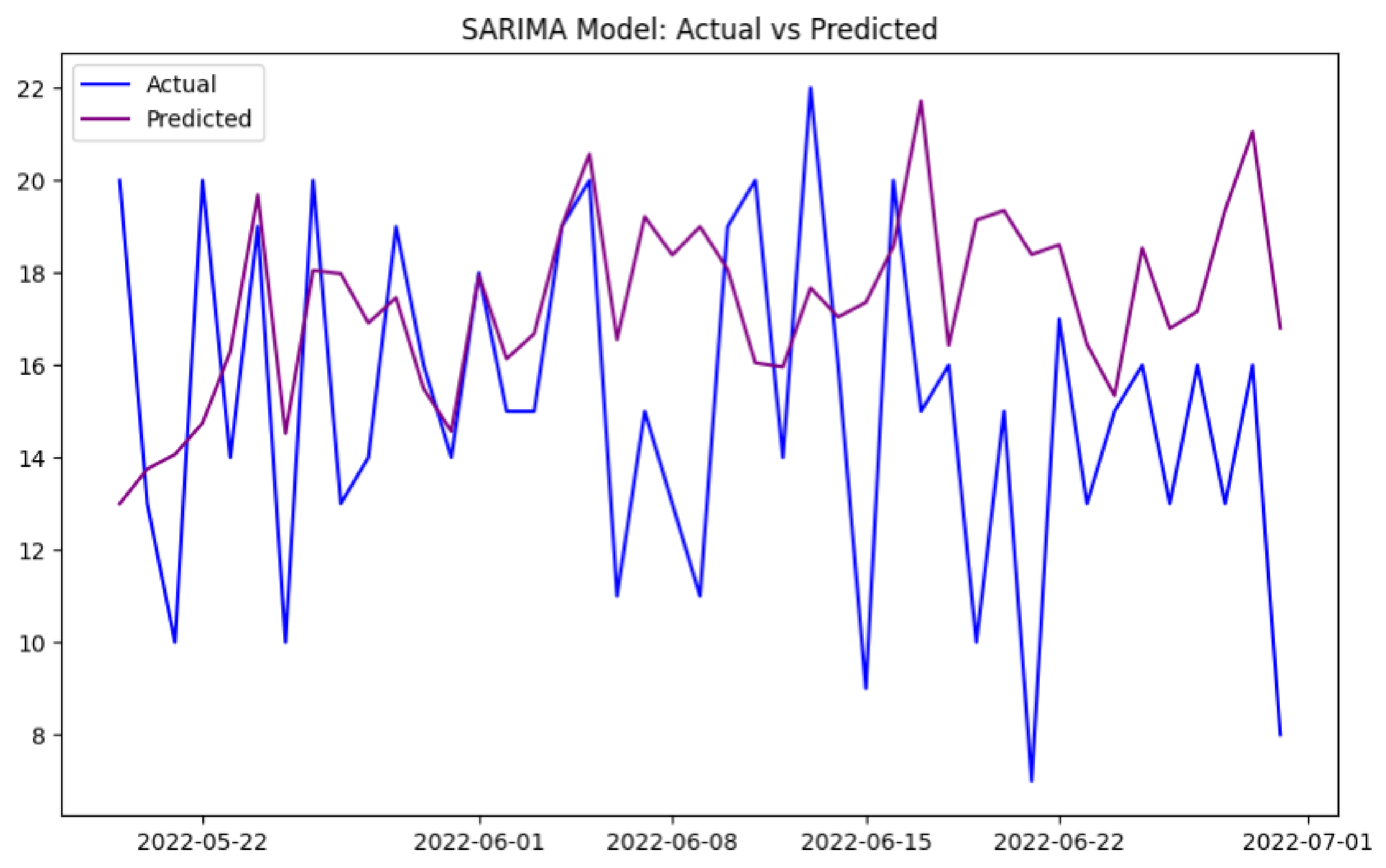


SARIMA & SARIMAX

SARIMA (Seasonal ARIMA):

>>Added seasonal differencing to ARIMA for periodic patterns.

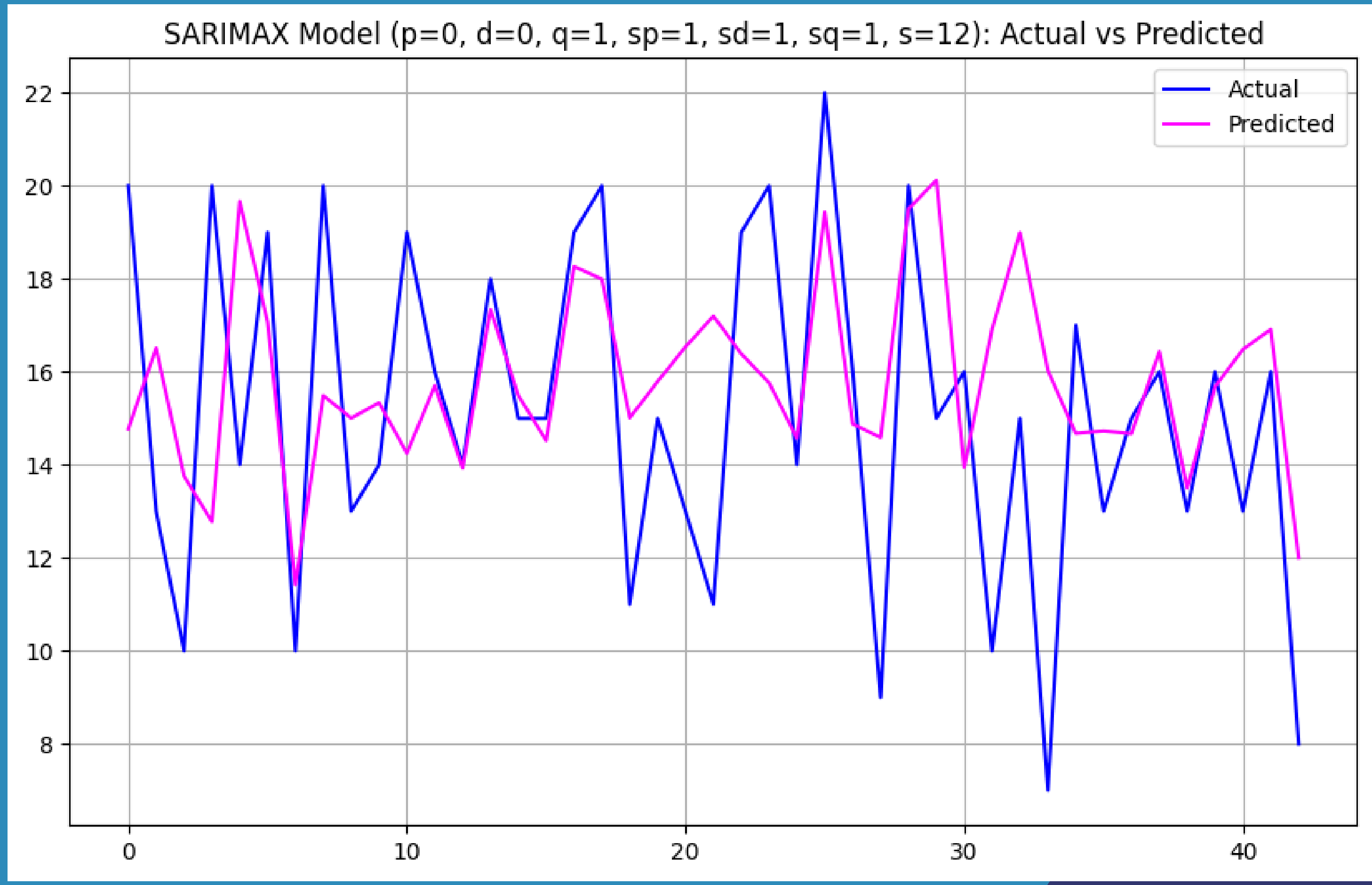
>>Tuned seasonal parameters (P, D, Q, S) to align with demand cycles like holidays and promotions.



SARIMAX (Seasonal ARIMA with Exogenous Variables):

>>Combined SARIMA's seasonal capabilities with external variables.

>>Provided the most accurate predictions by capturing seasonality and marketing impacts..



METRICS TABLE

To evaluate the performance of the models comprehensively, I calculated key metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

Metrics Table (Models in Rows):

	Models /Error Metrics	MAE	RMSE	MAPE	R^2	Differencing Mean
0	AR	3.6954	4.4155	0.3105	-0.5137	-0.017857
1	MA	3.6802	4.4960	0.3114	-0.5693	-0.017857
2	ARIMA	3.1370	3.7640	22.35%	-0.0771	-0.017857
3	ARIMAX	3.1487	3.8930	21.65%	-0.1521	-0.017857
4	SARIMA	2.9200	3.6482	0.2181	-0.0118	-0.017857
5	SARIMAX	2.7669	3.5569	21.63%	0.0382	-0.017900

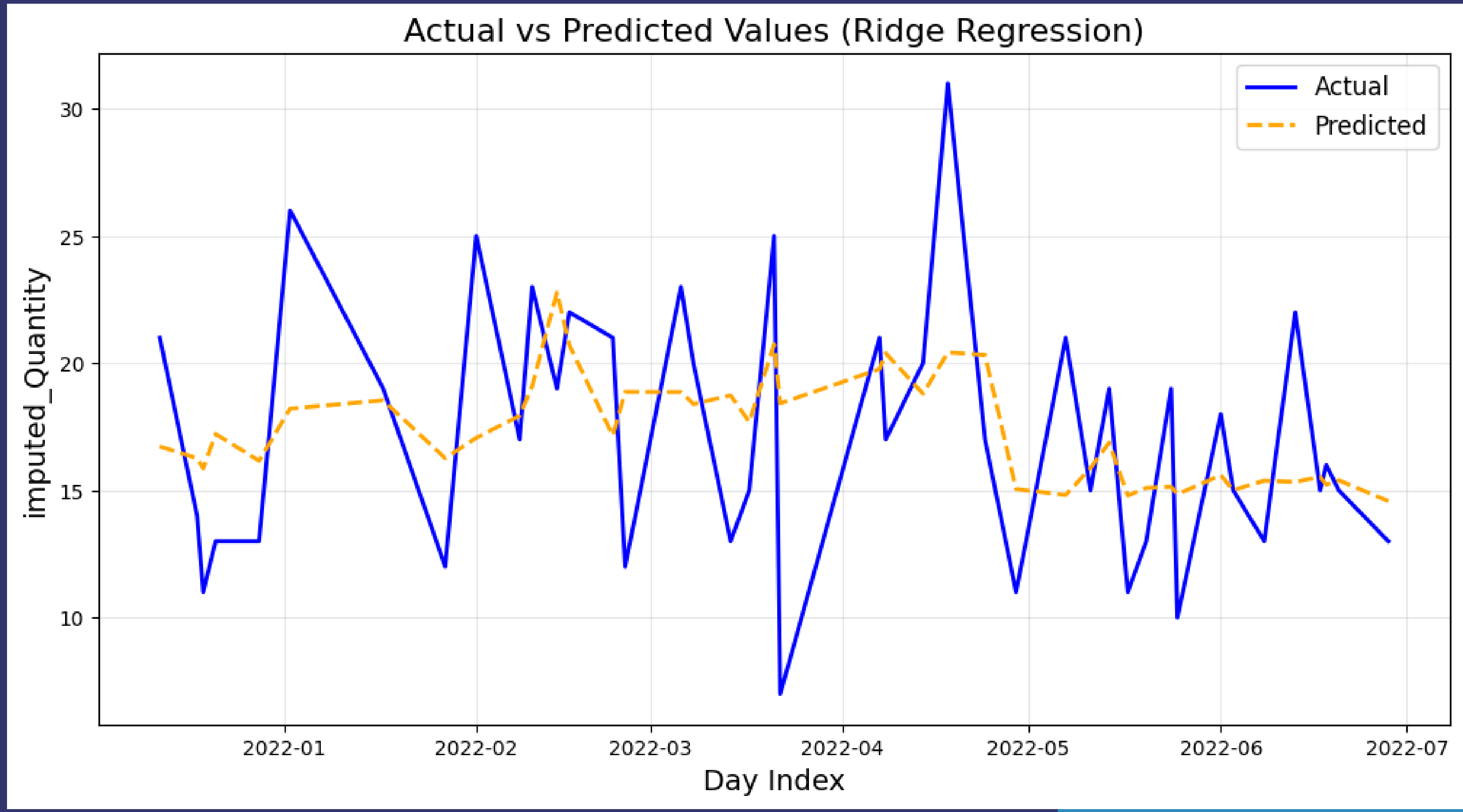
MULTIVARIATE REGRESSION

>>Integrated multiple predictors to capture dynamic relationships over time.

>>Created lagged versions of variables (clicks, impressions) to account for delayed effects.

>>Applied scaling and stationarity checks to ensure robust results.

>>Used the dynamic regression model to analyze temporal dependencies and derive actionable insights.



FORECASTING

Forecasting Results:

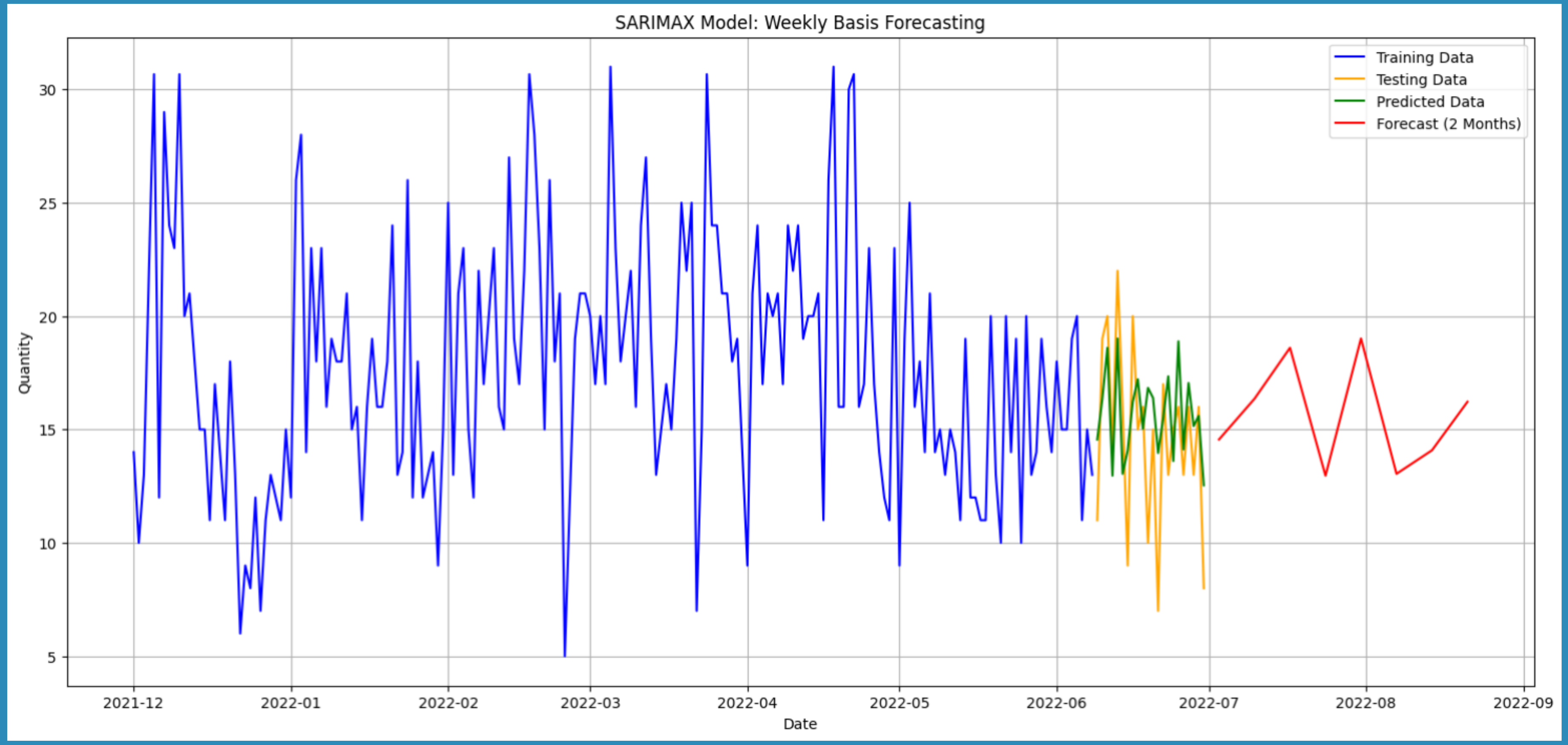
>Generated a two-month demand forecast for inventory optimization.
>Predicted seasonal spikes and identified the correlation between ad impressions and sales.

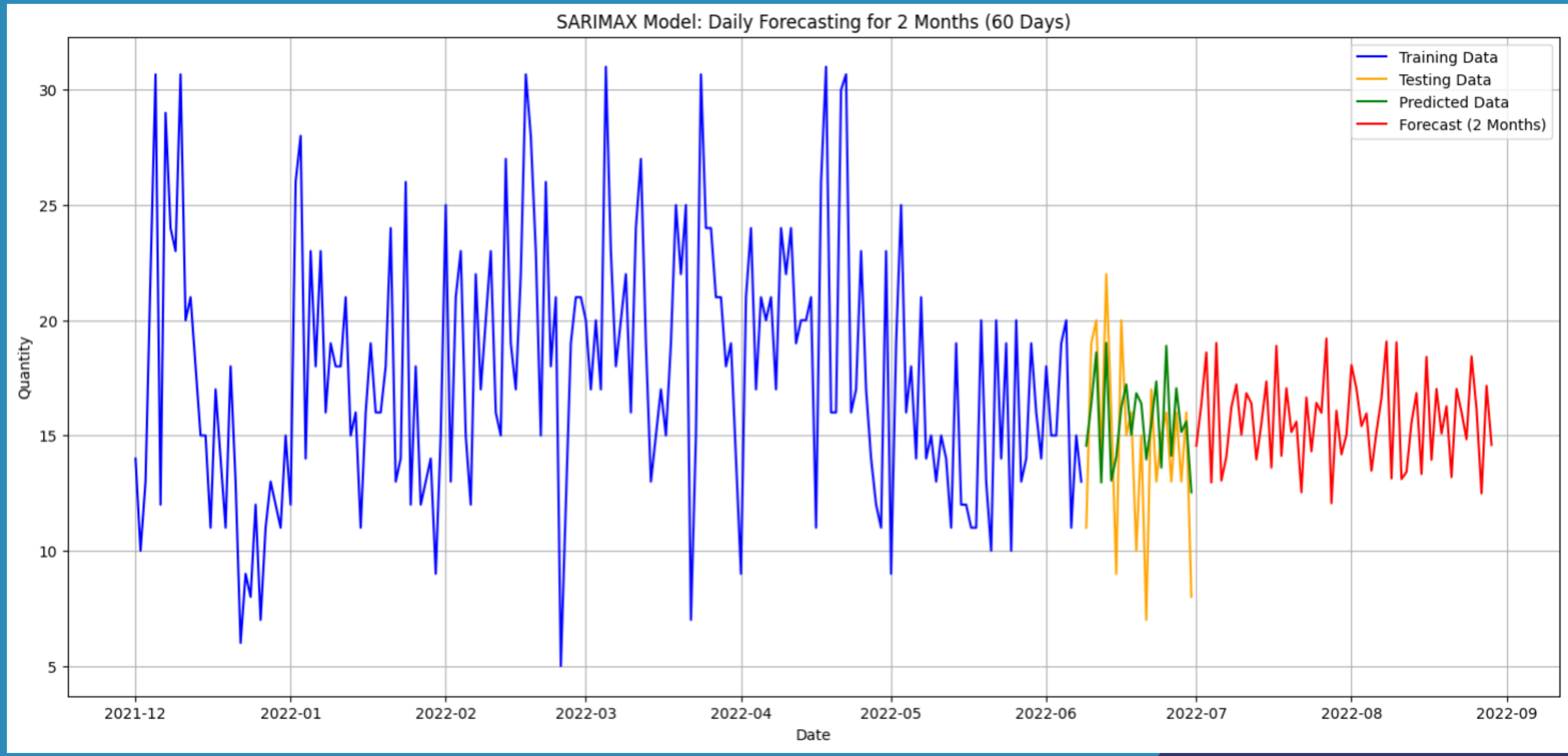
Summary of Models:

>AR and MA: Effective for short-term forecasts with minimal noise.
>ARIMA and ARIMAX: Captured trends and the impact of external variables.
>SARIMA and SARIMAX: Best suited for handling seasonality and complex patterns.
>Multivariate Regression: Added depth by integrating lagged external predictors.

Visualization:

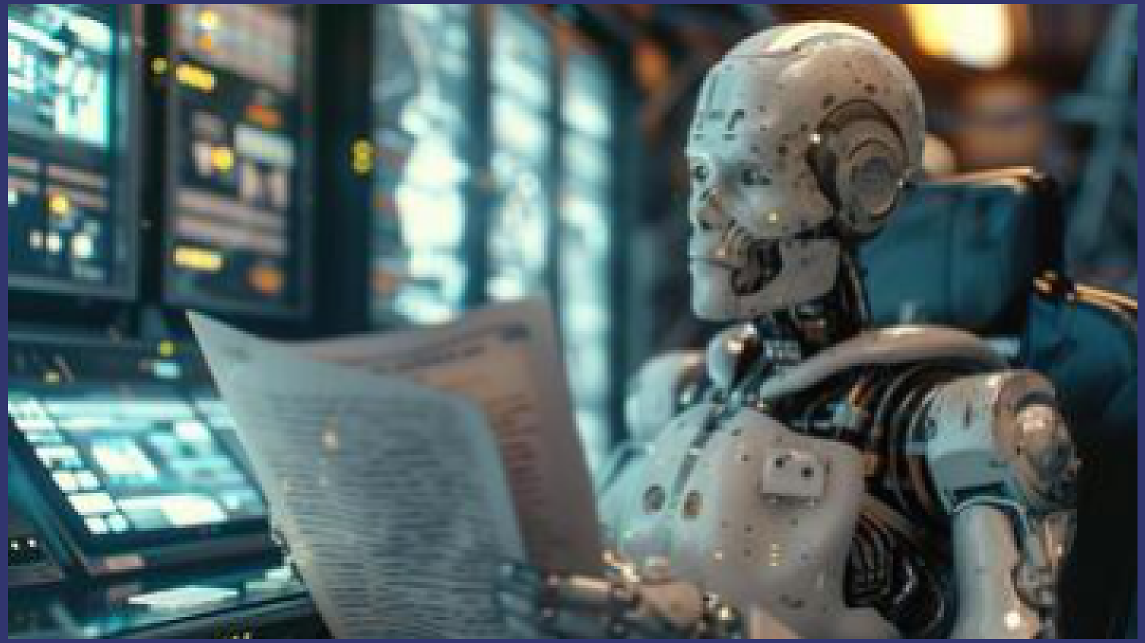
Plotted predicted vs. actual values for validation.





Challenges Faced & Solutions

Challenges:
>Handling data sparsity in certain product categories.
>Ensuring stationarity for time series modeling.
>Balancing model complexity and interpretability.



Solutions:
>Aggregated sparse data at higher levels for reliability.
>Applied differencing and transformations to achieve stationarity.
>Opted for SARIMAX to balance accuracy with interpretability for stakeholders.

Conclusion

Key Takeaways:
>AI-driven demand prediction offers actionable insights for smarter retail operations.
>SARIMAX emerged as the best-performing model for capturing seasonal trends & external influences.

Impact:
>Improved inventory management, reduced waste, and enhanced customer satisfaction.
>Enabled targeted marketing campaigns for higher ROI.

Future Scope:
>Explore advanced models like LSTMs for longer-term predictions.
>Integrate additional variables like economic indicators for more holistic insights.

THANK YOU !