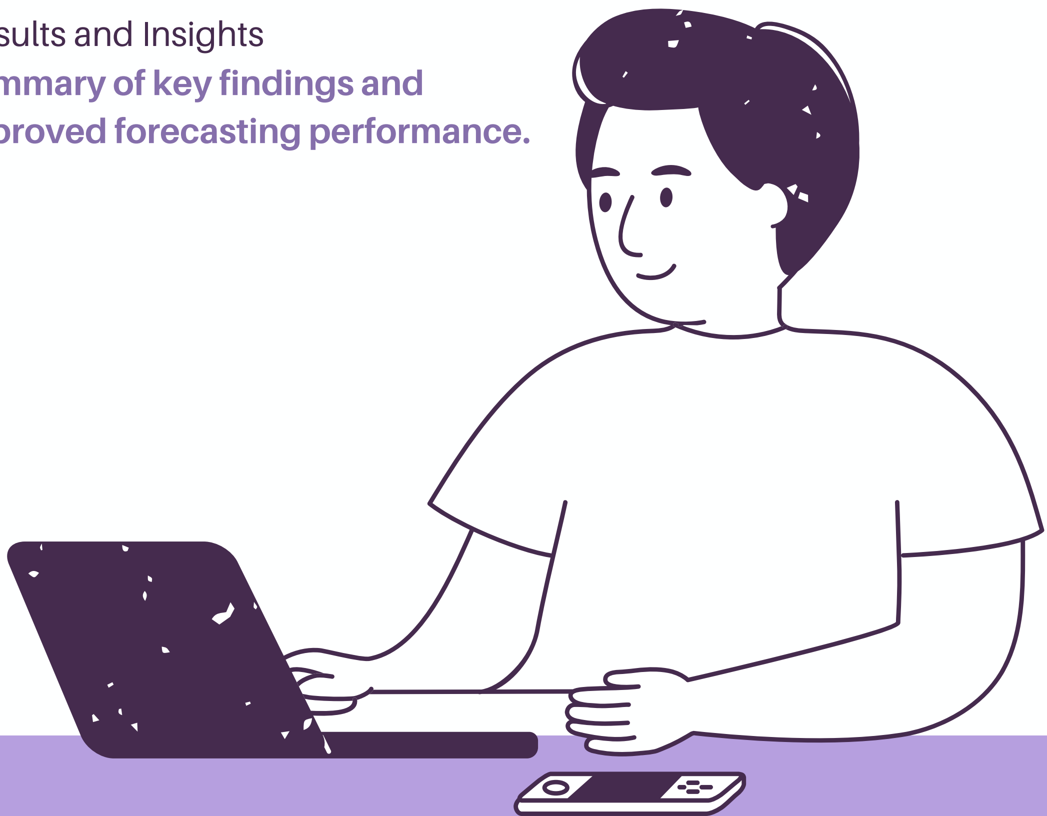
Milestone 2 report

Advanced Data Analysis and Feature Engineering

Project : Sales Forecasting and Demand Prediction

Contents



Introduction

Brief Summaries of the Report

This milestone focuses on performing advanced data analysis and feature engineering to enhance the accuracy and performance of the forecasting model.

The primary goal is to identify key trends, patterns, and external influences that affect sales behavior.

By transforming raw data into meaningful insights and engineered features, the model becomes more robust and capable of producing reliable predictions for future demand.

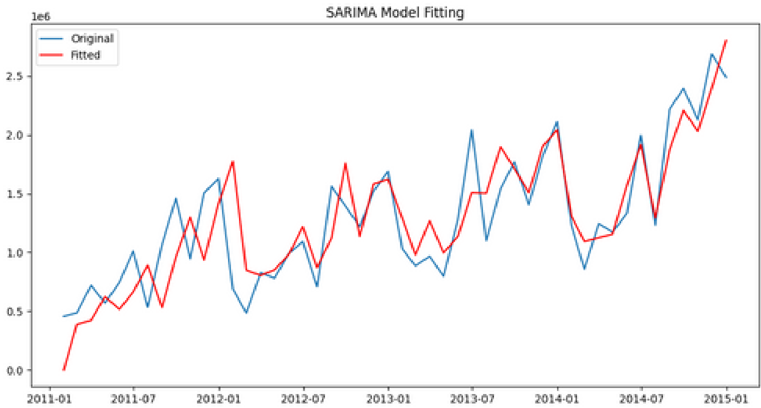
Advanced Data Analysis

Time Series Analysis

2.1

A detailed time series analysis was conducted to identify long-term trends, seasonal variations, and cyclic patterns in the sales data.

The data revealed periodic fluctuations corresponding to monthly and holiday-based variations, indicating strong seasonality in consumer behavior.



Stationarity Testing

2.2

To ensure suitability for time series forecasting, the Augmented Dickey-Fuller (ADF) test was applied.

The results confirmed that the original data was non-stationary due to clear trends and seasonal effects.

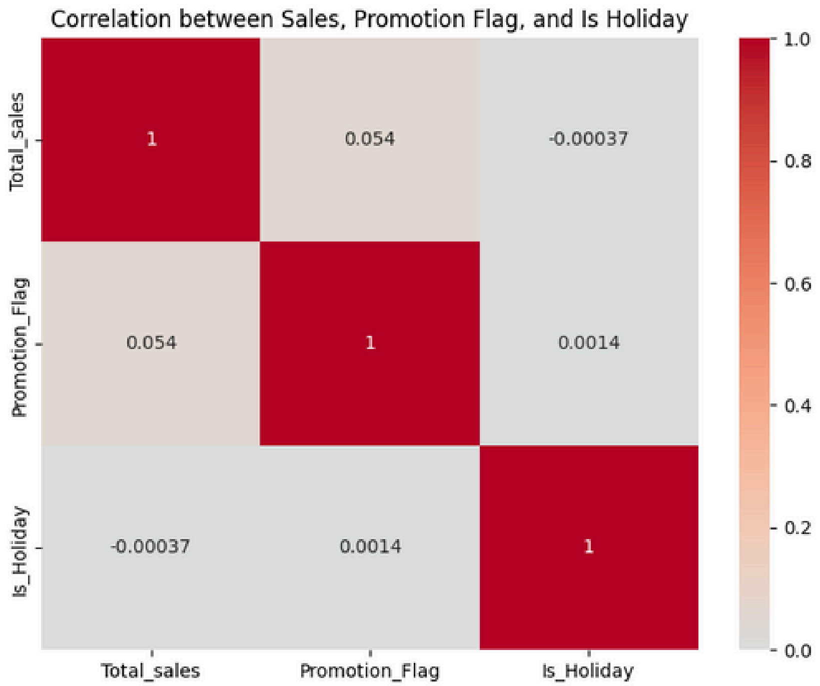
Differencing and transformation techniques were applied to achieve stationarity, preparing the dataset for accurate modeling using algorithms like ARIMA and Prophet.

Correlation Analysis

2.3

A correlation matrix was generated to explore the relationships between features such as sales, promotions, holidays, and weather conditions.

The analysis showed a strong positive correlation between promotional activities and sales spikes, while adverse weather conditions occasionally reduced consumer activity, influencing overall demand.

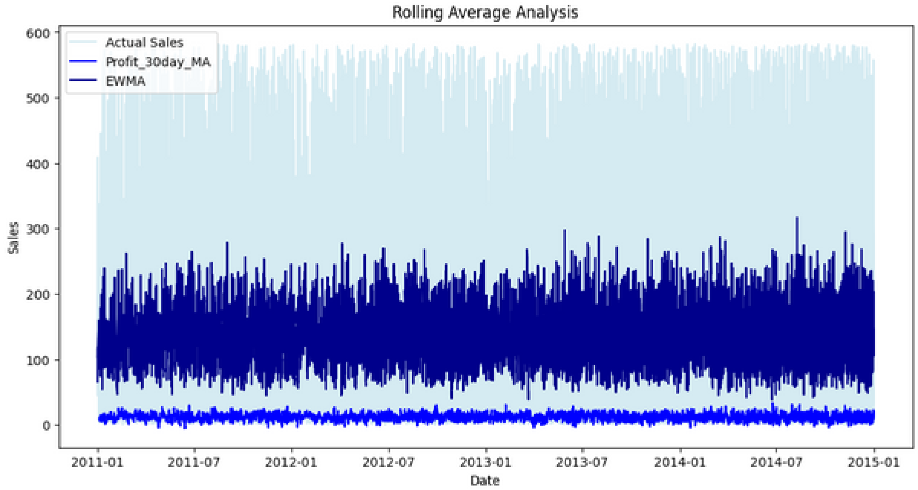


Feature Engineering

Time-Based Feature

3.1

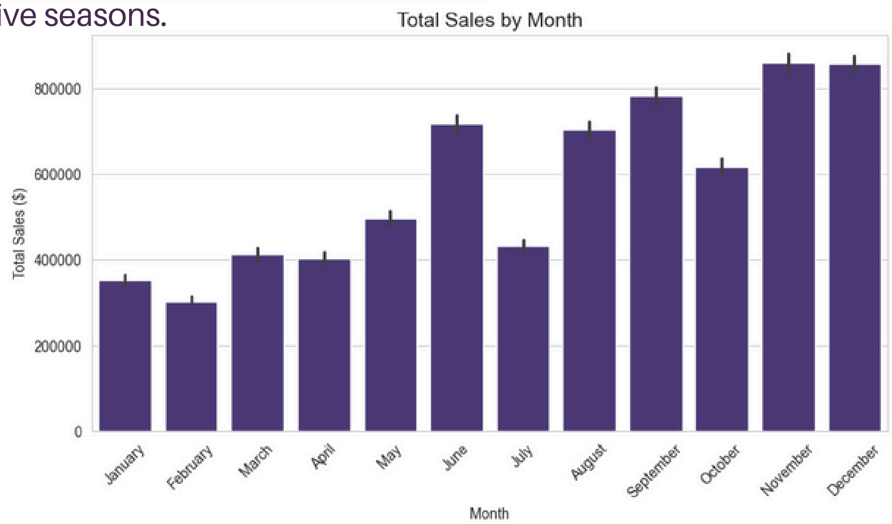
New temporal features were created to capture time-dependent relationships. These included rolling averages, lag features, and moving windows that allowed the model to understand recent trends and short-term fluctuations in sales.



Seasonal Components

3.2

Seasonal indicators such as month, day of the week, and holiday flags were added to capture periodic effects. These components helped the model recognize recurring sales patterns around specific months and festive seasons.



Data Transformation

3.3

The dataset underwent normalization and encoding processes to ensure that all features contributed effectively to model training.
Aggregations like monthly and weekly totals were introduced to reduce noise and improve interpretability.

External Variables

3.4

External factors such as weather, promotions, and economic indicators were integrated into the dataset.
These features added valuable context, allowing the model to account for environmental and market-driven influences that affect consumer demand.

Data Visualization

Historical Trend Analysis 4.1

Visualizations such as line charts and seasonal decomposition plots were created to highlight long-term trends and recurring seasonal behavior.

These visual insights provided a clear view of sales dynamics over time.

External Factor Impacts 4.2

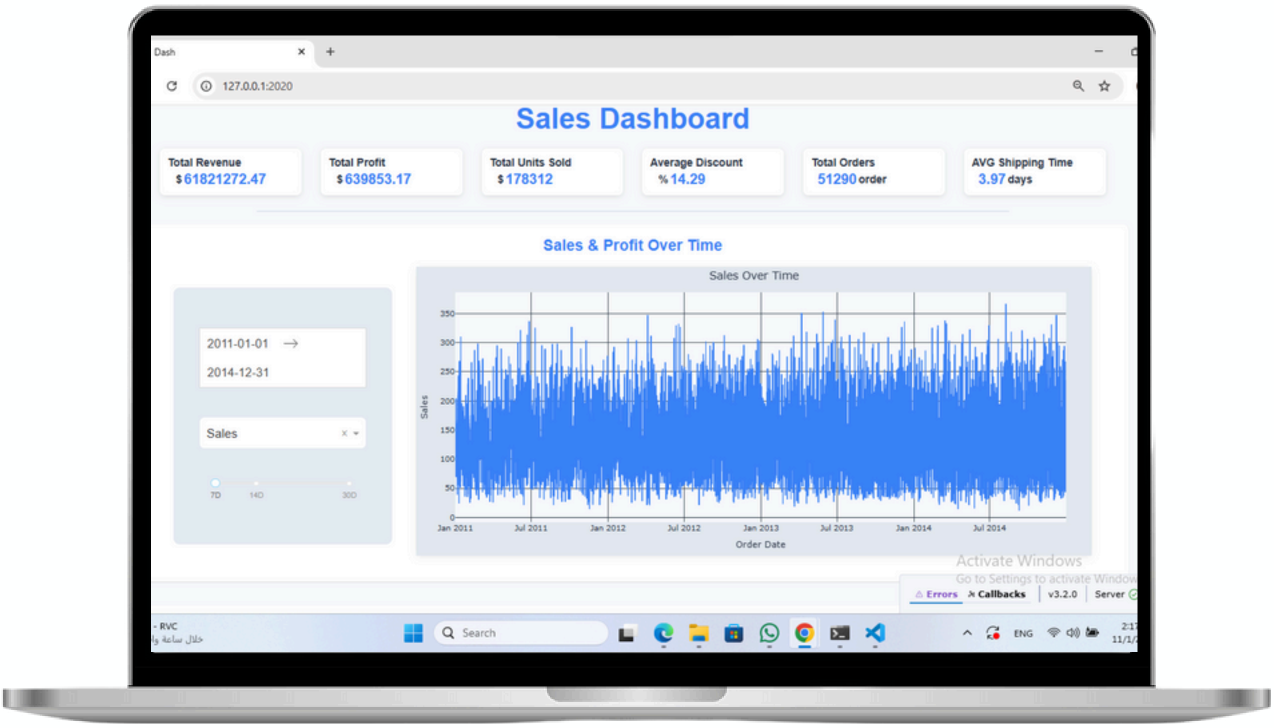
Advanced visualizations demonstrated the effect of promotions and weather on sales performance.

For instance, graphs indicated a significant increase in demand during promotional periods and a decrease during extreme weather events.

Interactive Dashboards 4.3

Interactive dashboards were designed to explore different time periods, categories, and external factors.

These tools enable decision-makers to dynamically assess how various influences impact overall sales trends



Results and Insights

5

The advanced analysis and feature engineering processes significantly improved the forecasting model's accuracy.
The inclusion of lag features, rolling means, and external variables helped the model capture both short-term fluctuations and long-term seasonal effects.
The insights derived from correlation and visualization analyses provided a deeper understanding of what drives sales performance.

Next Step

6

- Develop and optimize forecasting models to predict future sales.
- Experiment with models such as ARIMA, ETS, Random Forest, and LSTM.
- Train and validate models using time-aware data splits.
- Evaluate model accuracy with MAE, RMSE, and R^2.
- Select and fine-tune the best-performing model for final deployment.

References

7

- Python Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, statsmodels, fbprophet
- Dataset: Historical sales, weather, and promotional data
- Statistical Methods: ADF Test, Correlation Analysis, Rolling Window Techniques