Sales Forecasting and Optimization

Presented by our data science team | Date: May 15 2025

- Atef Hamdi
- Mariam Youssef
- Mohamed Walid
- Mohammed Yaser
- Omar Reyad
- Rana Abdelrahman

Project Overview

Goal

Predict future sales using historical retail and e-commerce data.

Business Impact

Optimizes inventory management, sales strategy, and operational planning.

Project Objectives

Data Exploration & Preparation

Understand and ready the sales history for modeling.

Analysis of Trends

Identify patterns and seasonality within sales data.

Model Building & Evaluation

Create, test, and refine forecasting models.

Tool Deployment

Implement a user-friendly predictive interface.

Challenges Faced

Dataset Quality Issues

Started with poor data requiring full replacement.

Feature Engineering Difficulties

Tried different features to find those that best influence sales.

Model Selection

Evaluated multiple models to find the most accurate.

Data Collection and Preprocessing

1

New Dataset Acquisition

2

Data Cleaning

Handled missing values and removed duplicates.

3

Feature Engineering

Created time-based features like day, month, and seasonality.

4

Scaling & Transformation

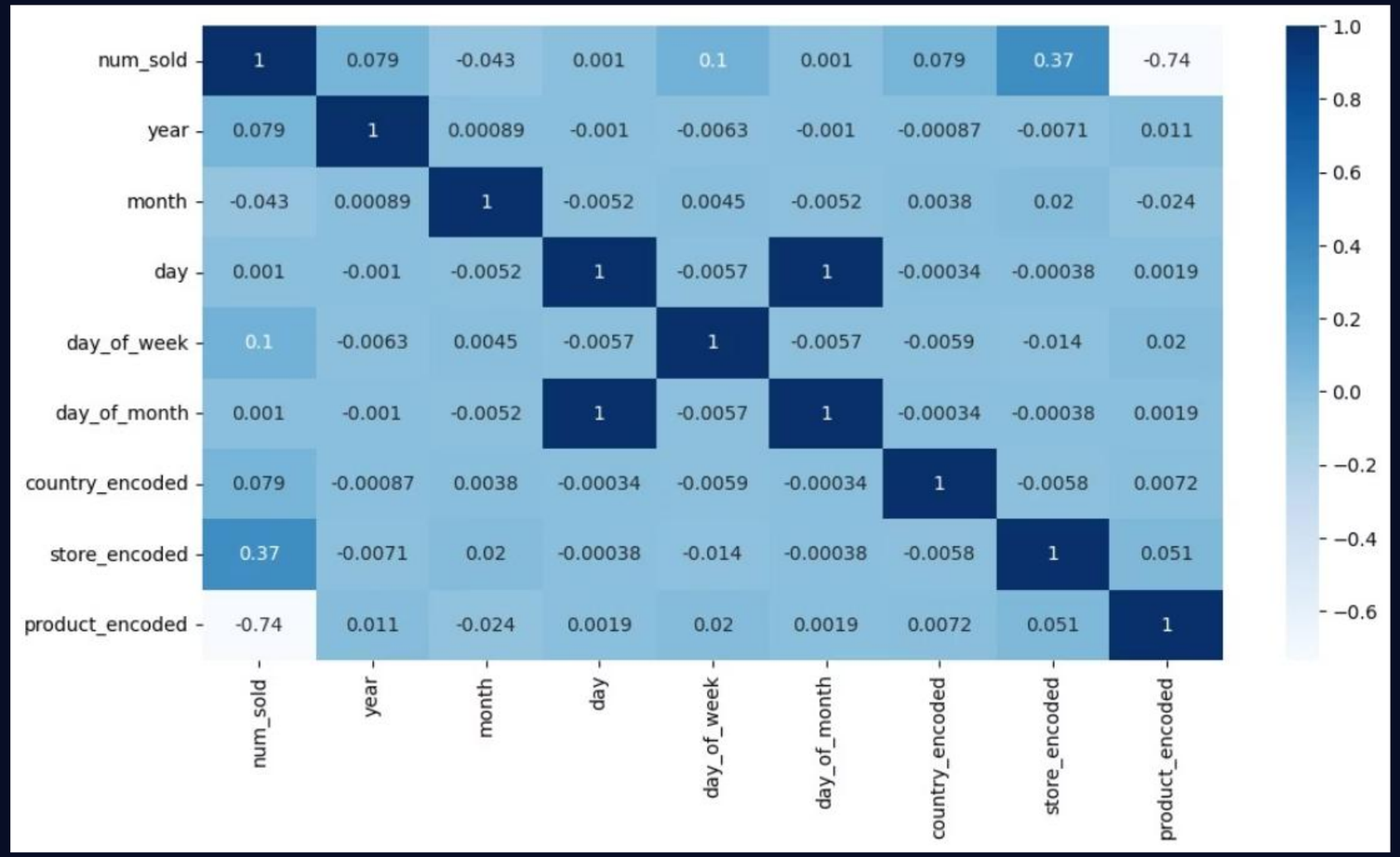
Prepared data for model input with normalization techniques.

Data Form

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 26298 entries, 0 to 26297
Data columns (total 6 columns):
 #   Column    Non-Null Count  Dtype 
---  ------    --------------  ----- 
 0   row_id    26298 non-null  int64 
 1   date      26298 non-null  object
 2   country   26298 non-null  object
 3   store     26298 non-null  object
 4   product   26298 non-null  object
 5   num_sold  26298 non-null  int64 
dtypes: int64(2), object(4)
memory usage: 1.2+ MB
```

	row_id	date	country	store	product	num_sold
0	0	2015-01-01	Finland	KaggleMart	Kaggle Mug	329
1	1	2015-01-01	Finland	KaggleMart	Kaggle Hat	520
2	2	2015-01-01	Finland	KaggleMart	Kaggle Sticker	146
3	3	2015-01-01	Finland	KaggleRama	Kaggle Mug	572
4	4	2015-01-01	Finland	KaggleRama	Kaggle Hat	911

Data Correlation



A quick look at feature correlations can offer key insights.

Data Analysis and Visualization

Exploratory Data Analysis

Detected trends, outliers, and seasonal patterns.

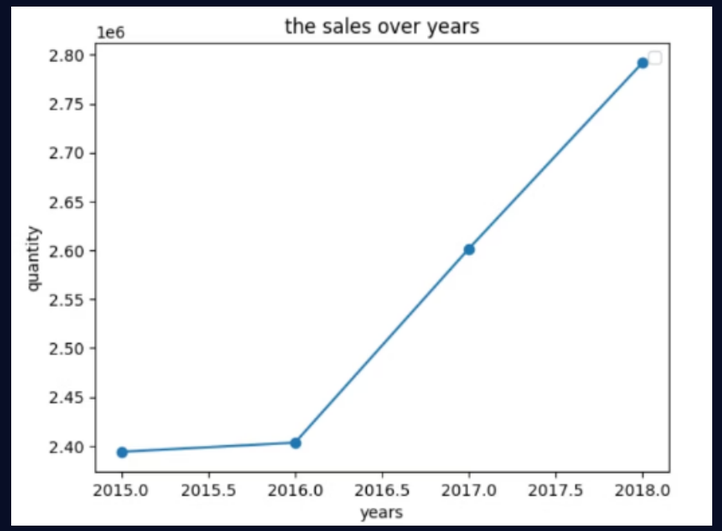
Visualizations

- Line charts
- Histograms
- Heatmaps

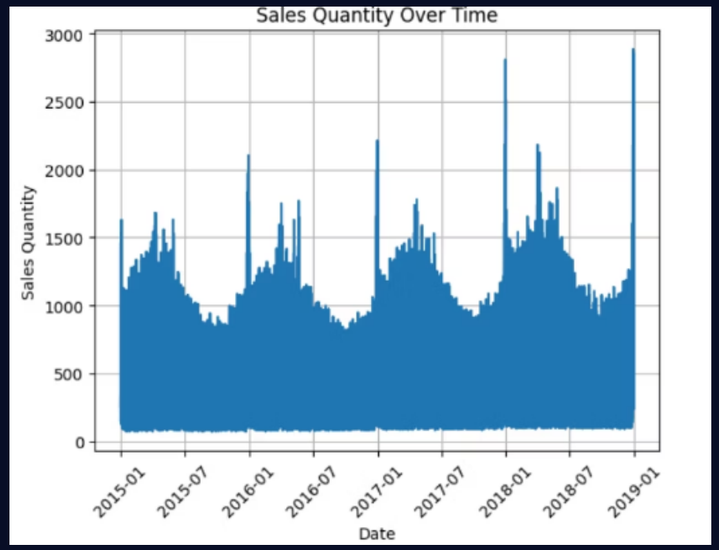
Feature Impact

Analyzed how dates and holidays influence sales.

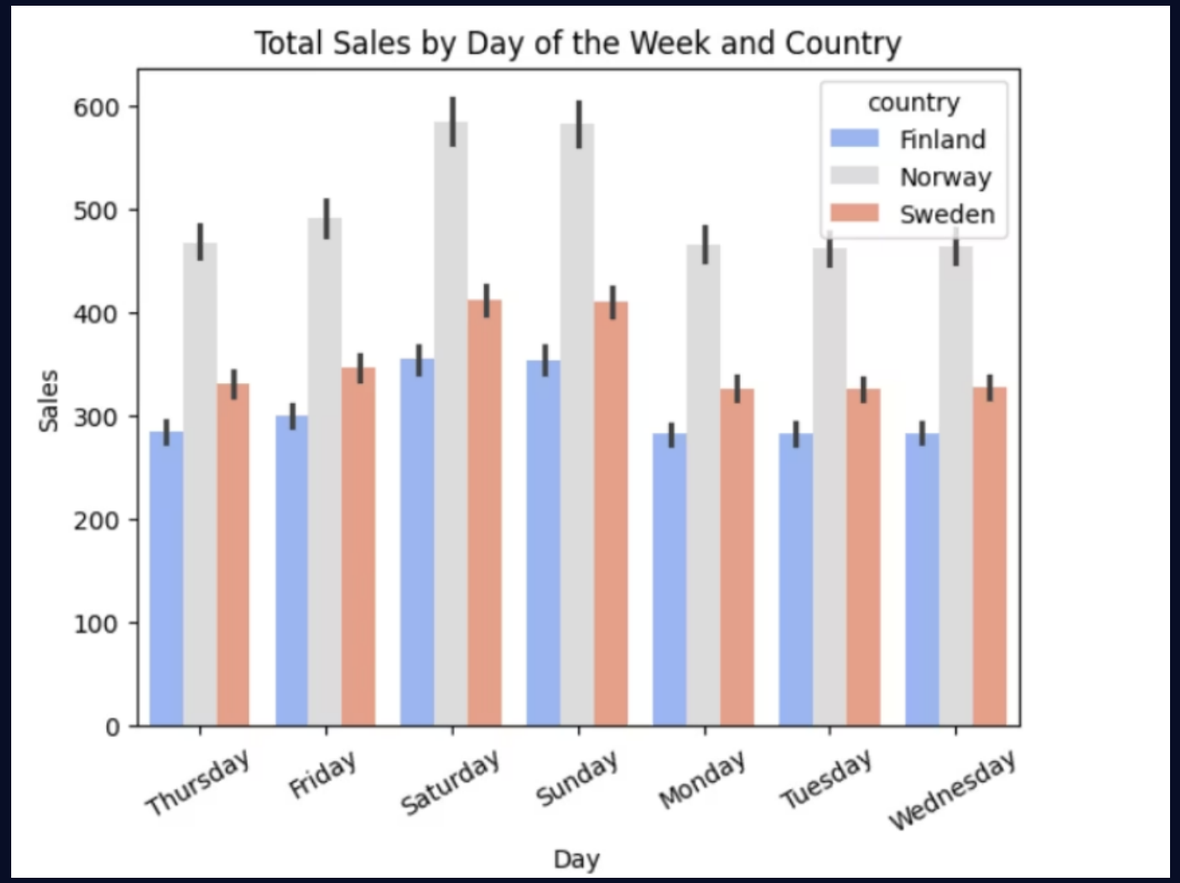
Sales Over Time



	year	num_sold
0	2015	2393991
1	2016	2403352
2	2017	2601797
3	2018	2792218

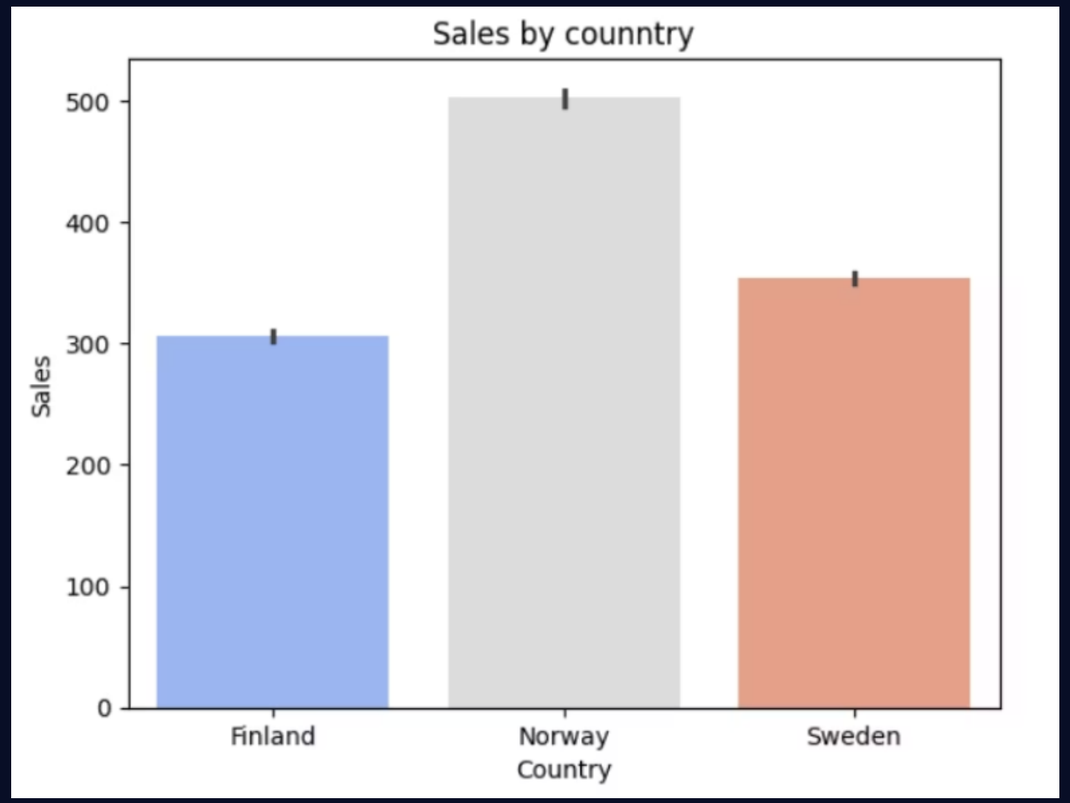


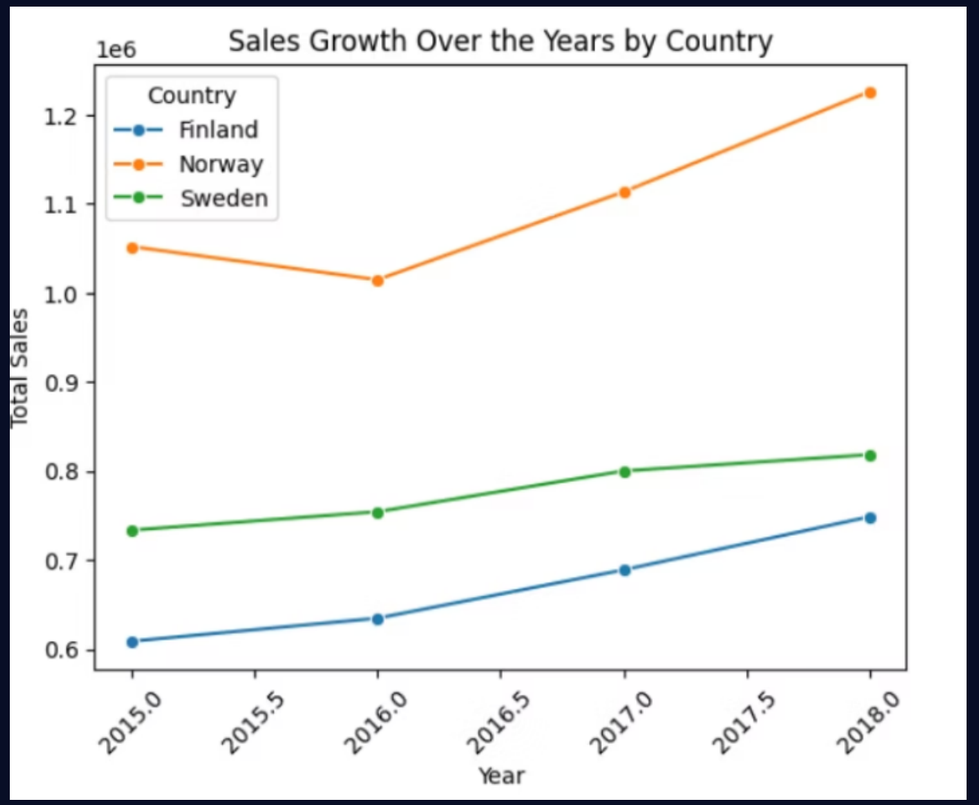
Sales by Weekday & Country



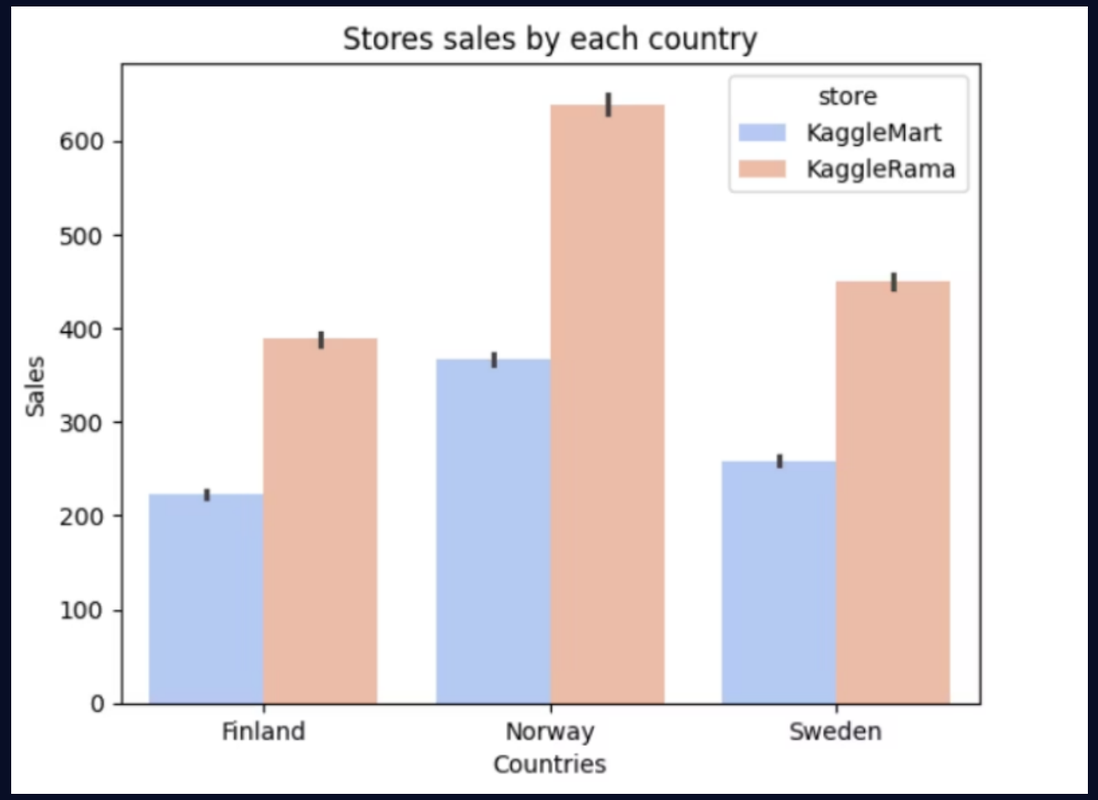
Same days, different countries–same pattern?

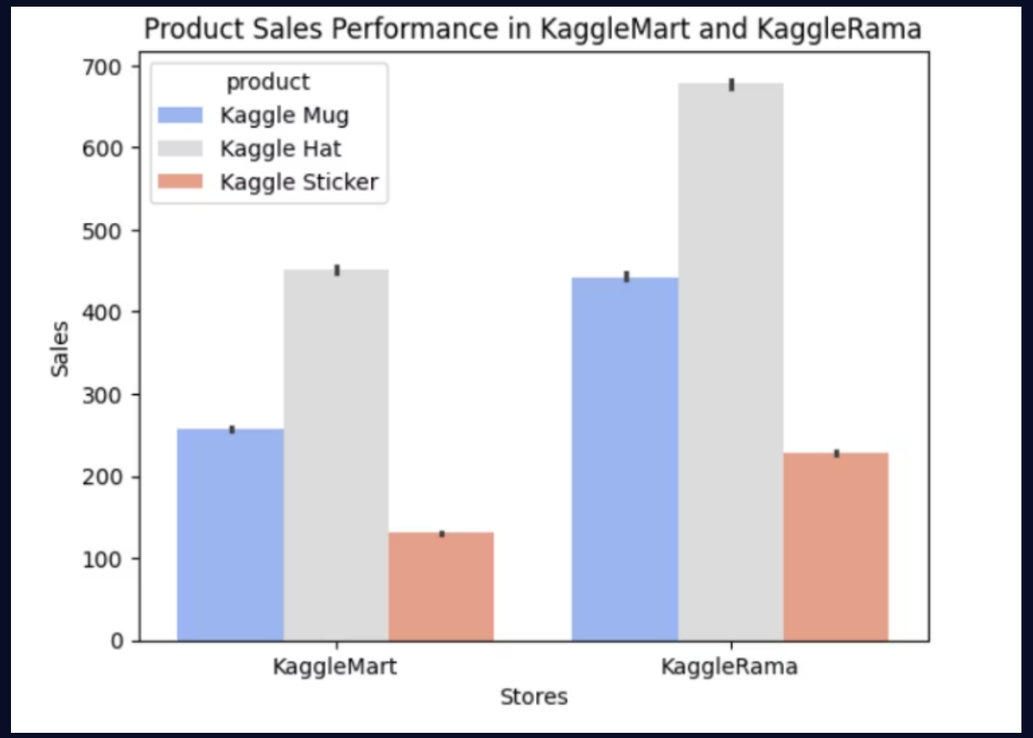
Countries Comparison

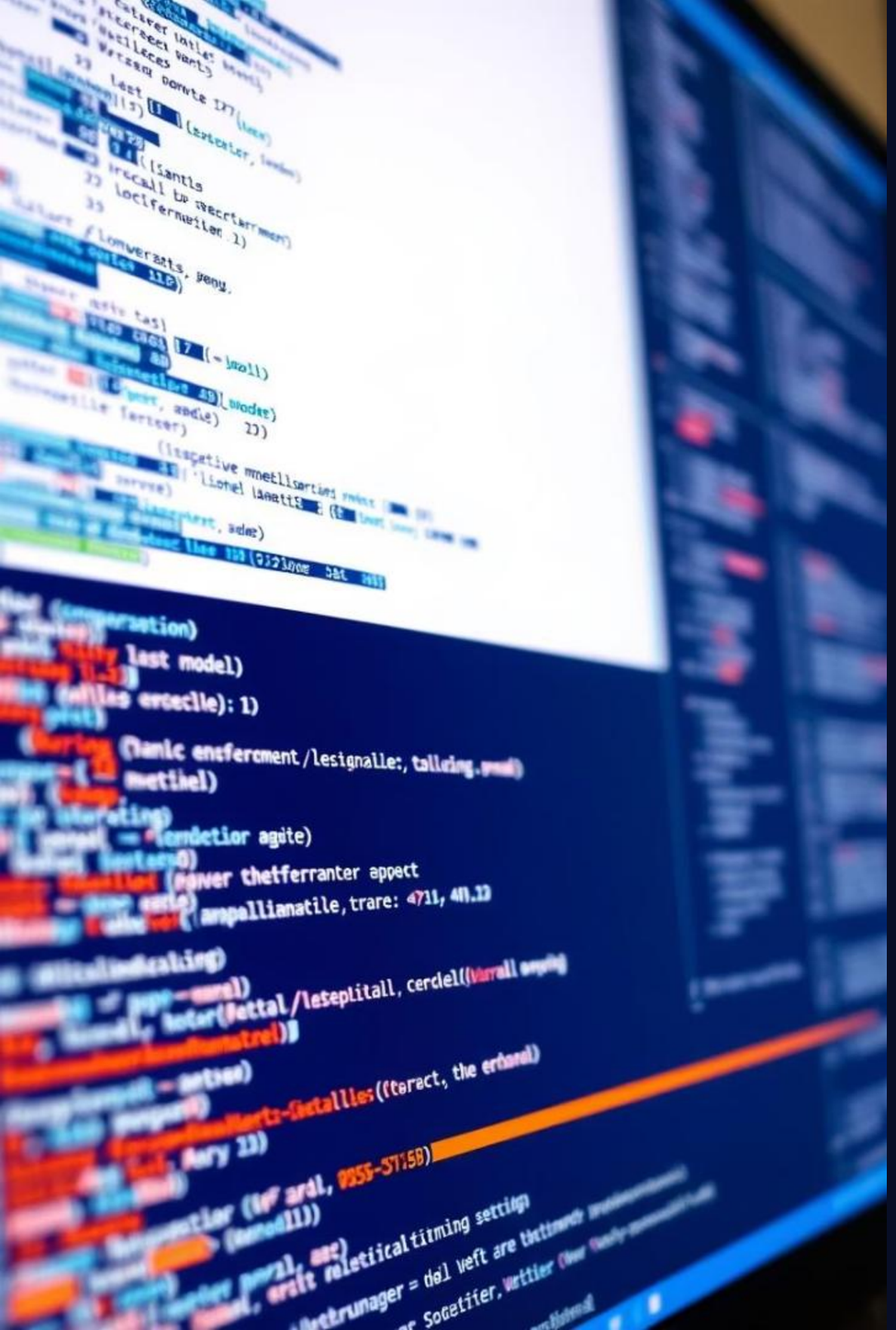




Stores & Products Comparison







Model Development and Optimization

Models Tested

- Linear Regression
- Neural Networks
- Facebook Prophet
- XGBoost (final choice)

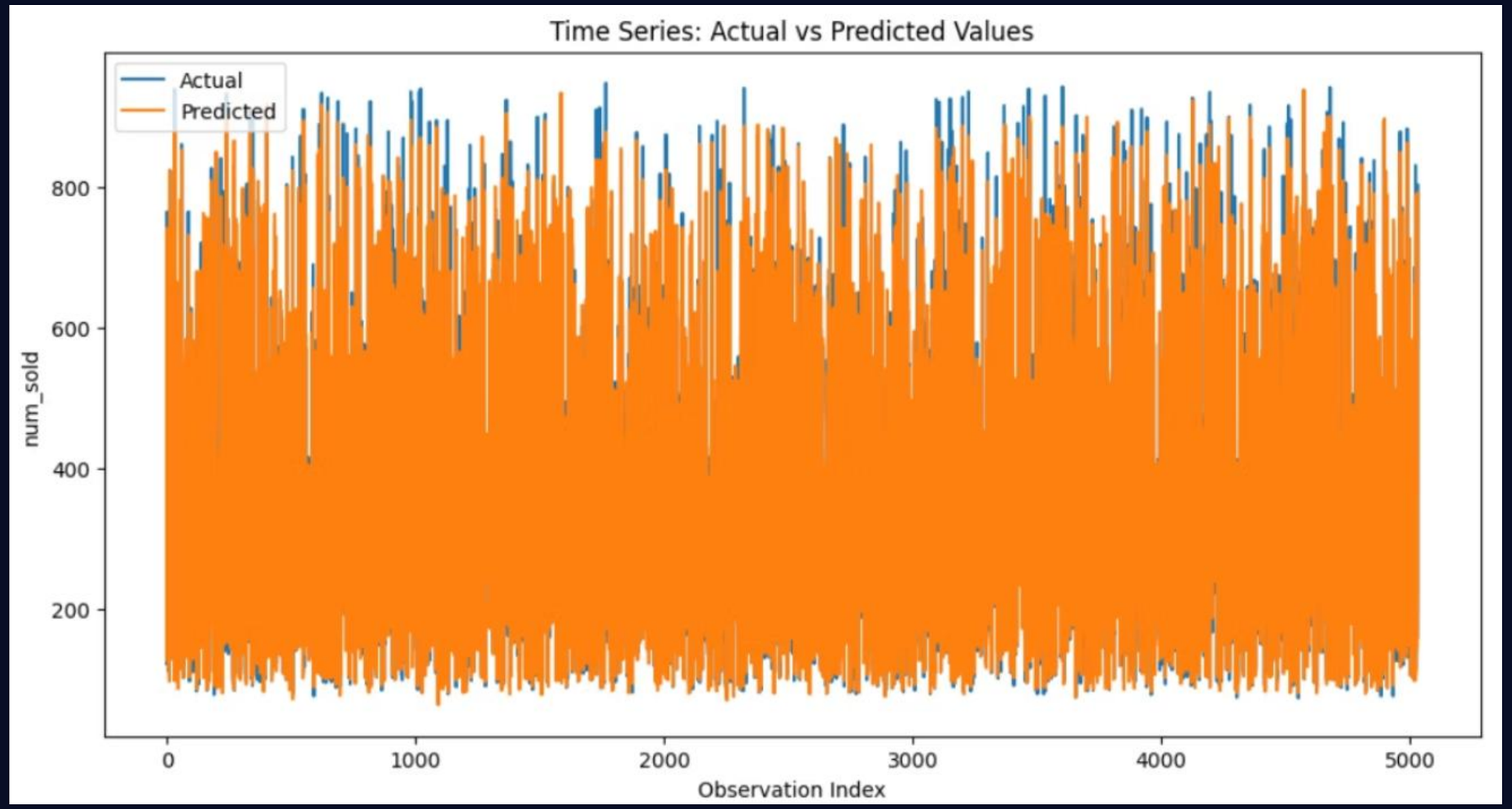
Evaluation Metrics

- RMSE
- MAE
- R2 Score

Selection Basis

XGBoost delivered the highest accuracy and robustness.

Training Results



XGBoost:

- RMSE: 25.11
- MAE: 17.17
- R2: 0.98

Deployment

Deployment Platform

Used Streamlit for interactive model interface.

Version Control

GitHub for code management and collaboration.

User Access

Provides real-time sales predictions to end users.

Key Outcomes & Next Steps

Reliable Forecasting Tool

Effective model with clear documentation.

Challenge Resilience

Adapted to data and modeling challenges.

Future Enhancements

Integrate external data for improved accuracy.

Q&A

Thank you for your time and attention. We now welcome your questions and feedback.

Open Dialogue

Foster collaborative understanding.

Clarifications

Resolve any project questions.

Next Steps

Explore future collaboration.