OSIRI UNIVERSITY

SCHOOL OF DATA SCIENCE AND INFORMATION SYSTEMS

**Project Title: Retail Demand Forecasting & Inventory Optimization**

Course: CSS 503 Data Science & Analytics

Group 8 Members

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Names** | **Email** | **Phone** |
| 1. | Ibrahim Olanrewaju Mufu | [ibmufu@osiriuniversity.org](mailto:ibmufu@osiriuniversity.org) | +234 706 187 3846 |
| 2. | Mihret Fikadu Wondimu | [mihret@osiriuniversity.org](mailto:mihret@osiriuniversity.org) | +251 91 038 5877 |
| 3. | Wendwesen Endale | [weworkenh@osiriuniversity.org](mailto:weworkenh@osiriuniversity.org) | +251 93 485 9047 |
| 4. | Temitope Ademola | [temademola@osiriuniversity.org](mailto:temademola@osiriuniversity.org) | +234 807 730 1197 |
| 5. | Gabriel Simeon Ikwor | [gabrielsimeon@osiriuniversity.org](mailto:gabrielsimeon@osiriuniversity.org) | +234 813 505 4970 |
| 6. | Peace Toluwani Ogunmodede | [Peogunmodede@osiriuniversity.org](mailto:Peogunmodede@osiriuniversity.org) | +234 708 588 6827 |
| 7. | Obanleowo Yusuf Olalekan | [yuobanleowo@osiriuniversity.org](mailto:yuobanleowo@osiriuniversity.org) | +234 805 543 9319 |
| 8 | Mary |  |  |

|  |  |
| --- | --- |
| **Lecturer** | Dr. Noble Achumba |
| **Data Submitted** | July 27, 2025 |

1. **Executive Summary**

Rossmann operates over 3,000 drug stores across seven European countries, requiring each store manager to forecast daily sales up to six weeks in advance to manage inventory, staffing, and promotions effectively. Sales volume is driven by multiple factors—**promotions, competition, holiday effects, seasonality, and local market dynamics**—leading to considerable variability in manually produced forecasts. Leveraging the public Rossmann Store Sales dataset from the 2015 Kaggle competition, this project will apply the full data science lifecycle to build and evaluate time series and machine learning forecasting models (e.g., ARIMA, Prophet, XGBoost). By comparing model outputs to actual sales, we aim to deliver a robust, data-driven forecasting tool that reduces prediction error, enhances inventory planning, and supports consistent decision-making across all Rossmann stores (kaggle.com).

**1.1 Introduction & Background**

Rossmann GmbH is a leading retailer in the health and beauty sector, with over 3,000 stores in Germany, Austria, Poland, Czech Republic, Hungary, Turkey, and Slovenia. To maintain high service levels and cost efficiency, store managers must predict six weeks’ worth of daily sales. The Rossmann Store Sales Kaggle dataset comprises three years of historical sales records for 1,115 German stores, including fields for sales, customer counts, store open/closed flags, promotion indicators, school and state holidays, store type, assortment level, and competition distance/opening date. This rich dataset captures the multifaceted drivers of demand and provides an ideal testbed to evaluate advanced forecasting techniques.

Manual forecasting at the store level often leads to inconsistent and inaccurate predictions due to the complexity of factors influencing sales. Our project addresses this challenge by systematically cleaning and exploring the dataset, engineering relevant features (lagged sales, rolling windows, calendar events), and training multiple models to capture both linear and nonlinear patterns. Through rigorous back-testing and error analysis, we will identify the most accurate and scalable models to standardize forecasting across all stores. This data-driven approach promises to improve inventory management, reduce stockouts and overstock, and ultimately increase operational efficiency and profitability for Rossmann.

**1.2. Context**

Rossmann is face increasingly complex market dynamics, including fluctuating customer demand, supply chain disruptions, workforce planning and high competition. In such an environment, accurately predicting customer demand and aligning inventory levels and workforce accordingly has become more crucial than ever. This project takes place in the context of digital transformation in retail, where data-driven decision-making is becoming the norm.

**1.3. Problem Statement:**

Rossmann store with thousands of individuals and managers predicting sales based on their unique circumstances, suffer from getting accurate forecast prediction which affect its workforce planning, stocking and productivity—leading to ineffective customer satisfaction, poor productivity and employee motivation.

**1.4. Rationale**

By developing a robust demand forecasting system, businesses can boost the satisfaction of their customers through proper workforce planning, inventory planning, improved productivity and motivation. This not only enhances profitability by reducing excess inventory and adequate staffing – improving turnover rates and customer experience through better service levels. A data science-driven approach allows for continuous learning and adaptation to changes in demand patterns.

**1.3 Objectives**

**Develop forecasting models:** Build and compare time series and machine learning models (e.g. ARIMA/SARIMA, ETS, Prophet, LSTM, XGBoost) to predict product-level retail demand.

**Optimize inventory decisions:** Use forecasting outputs to design inventory policies (e.g. reorder points, safety stock levels) that minimize total costs (holding + stockout costs).

**Apply data science lifecycle:** Follow a systematic methodology (e.g. CRISP-DM) through data acquisition, preprocessing, modeling, and evaluation**.**

**Validate and deploy:** Evaluate model performance with appropriate metrics (e.g. MAPE, RMSE) and provide an easy-to-interpret dashboard or reports for stakeholders.

**Document insights:** Deliver a comprehensive analysis and recommendations, demonstrating how predictive analytics can improve workforce scheduling and productivity, and inventory management.

**Dataset Summary: Rossmann Store Sales**

* Training Data: 1,017,209 rows; 1,115 stores; daily sales (2013–Oct 2015).
* Test Data: 410,653 rows; forecasting period (Oct 2015–Dec 2015).
* Key Features:

|  |  |  |
| --- | --- | --- |
| **Column** | **Type** | **Description** |
| Store | Integer | Store identifier |
| DayOfWeek | Integer | Day of week (1=Mon…7=Sun) |
| Date | Date | Observation date |
| Sales | Integer | Sales amount (€) |
| Customers | Integer | Number of customers |
| Open | Integer | Store open (1) or closed (0) |
| Promo | Integer | Active promotion (1) or not (0) |
| StateHoliday | Category | State holiday indicator (a, b, c, 0) |
| SchoolHoliday | Integer | School holiday flag (1) or not (0) |

* Store Metadata: store type, assortment, competition distance, promotion intervals.

**Reference & Data Sources**

FlorianKnauer and Will Cukierski. Rossmann Store Sales. https://kaggle.com/competitions/rossmann-store-sales, 2015. Kaggle.