

Capstone Project Proposal

Chandima Attanayake

BA 64099 – Capstone Project in Business Analytics

Sales Forecasting for a Retail Chain

Table of Contents

1. Executive Summary.....	1
2. Project Overview	1
3. Objective	1
4. Data Source.....	1
5. Methodology	2
6. Tools and Technologies	3
7. Deliverables	3
8. Expected Impact	3
9. Work Format	3

1. Executive Summary

This capstone project focuses on developing a robust and accurate sales forecasting model for a retail chain using historical sales data from the Rossmann Store Sales dataset (Kaggle). The primary aim is to assist retail decision-makers in planning inventory, workforce, promotions, and logistics. Through advanced analytics and machine learning techniques, the project will explore and model time-dependent sales behavior across multiple stores and provide actionable forecasting tools.

2. Project Overview

This project focuses on how it is forecasting sales for a retail chain by analyzing historical sales data. In this case, reliable sales forecasting plays a key role in operational areas of the business including inventory management, workforce planning, supply chain coordination, and overall strategic decision-making. The analysis will utilize the Rossmann Store Sales dataset, which is publicly available on Kaggle and includes a comprehensive collection of sales records from a variety of stores over time.

A company like Rossmann pharmacy chain with over 3,000 store locations across several countries also require accurate sales forecasting to maintain operational efficiency and financial stability. This project aims to develop a practical solution to the following problem.

How can future store level sales be forecasted using historical data to support more effective business planning and decision-making?

3. Objective

The objective of this project is to develop a machine learning-based forecasting model that predicts future sales based on historical patterns.

- Cleaning and preprocessing the Rossmann dataset
- Exploring and analyzing patterns and seasonality in sales
- Building predictive models (e.g., Linear Regression, XGBoost, LSTM, etc.)
- Evaluating model performance using RMSE, MAE, and MAPE
- Providing actionable insights for retail operations and strategy

4. Data Source

Dataset - Rossmann Store Sales – Kaggle

Platform - Kaggle

Link - <https://www.kaggle.com/competitions/rossmann-store-sales>

Key features include

- Date-wise sales data for over 1,000 stores
- Number of customers
- Promotional periods and holidays
- Store-specific information (store type, competition, etc.)
- Weather, location, and school holidays

5. Methodology

Phase 1 - Data Exploration & Preprocessing

- Handle missing values, outliers, and data inconsistencies
- Merge multiple CSV files (e.g., train.csv, store.csv)
- Create time-based features (month, day, year, week, holiday flag)
- Visualize sales trends, seasonality, and anomalies

Phase 2 - Model Development

Model Candidates -

- Baseline - Moving average, linear trend model
- Machine Learning - XGBoost, Random Forest, Gradient Boosting
- Time Series - ARIMA, SARIMA, Prophet (Facebook)
- Deep Learning - LSTM (for sequential modeling with longer-term dependencies)

Phase 3 - Evaluation

- Evaluation metrics - RMSE, MAE, MAPE
- Cross-validation (time series split)
- Hyperparameter tuning with GridSearchCV or Bayesian optimization

Phase 4 - Insights and Visualization

- Store-wise performance comparison
- Impact analysis of promotions and holidays
- Dashboard or report summarizing key patterns and forecasts

6. Tools and Technologies

- Programming Language - Python (Jupyter Notebooks) or R
- Libraries/Packages - pandas, scikit-learn, XGBoost, Prophet, matplotlib/seaborn, TensorFlow/Keras (for LSTM)
- Project Management - GitHub and/or project documentation via Jupyter Markdown or RMarkdown

7. Deliverables

- Cleaned dataset with feature engineering
- Forecasting models with performance metrics
- Final report (including methodology, code snippets, insights, and recommendations)
- Visualization dashboard or static report for business stakeholders
- Presentation slides summarizing the project

8. Expected Impact

By providing a scalable, store-level forecasting framework, this project can support -

- Improved inventory and workforce planning
- Better promotional scheduling
- Cost savings through reduced overstock and stockouts
- Enhanced decision-making for store operations and finance teams

9. Work Format

This capstone project will be completed individually.