Store Performance Segmentation and Time-Based Sales Optimization in Retail

A Proposal Report for the BDM Capstone Project

Submitted by

Mohit Raj Rathor

Roll No. 22f3003109

IITM Online BS Degree Program,

Indian Institute of Technology Madras, Chennai

Tamil Nadu, India, 60036

Table of contents

Declaration Statement

I am working on a Project titled "**Store Performance Segmentation and Time-Based Sales Optimization in Retail**". I extend my appreciation to "**Kaggle.com**", for providing the necessary resources that enabled me to conduct my project.

I hereby assert that the data presented and assessed in this project report is genuine and precise to the utmost extent of my knowledge and capabilities. The data has been gathered from primary sources and carefully analysed to ensure its reliability.

Additionally, I affirm that all procedures employed for data collection and analysis have been duly explained in this report. The outcomes and inferences derived from the data are an accurate depiction of the findings acquired through thorough analytical procedures.

I am dedicated to adhering to the principles of academic honesty and integrity, and I am receptive to any additional examination or validation of the data contained in this project report.

I understand that the execution of this project is intended for individual completion and is not to be undertaken collectively. I thus affirm that I am not engaged in any form of collaboration with other individuals and that all the work undertaken has been solely conducted by me. If plagiarism is detected in the report at any stage of the project's completion, I am fully aware and prepared to accept disciplinary measures imposed by the relevant authority.

I understand that all recommendations made in this project report are within the context of the academic project taken up towards course fulfilment in the BS Degree Program offered by IIT Madras. The institution does not endorse any of the claims or comments.

Signature of Candidate: (Digital Signature)

Name: Mohit Raj Rathor

Date: 12 July, 2025

Executive summary

The project is based on the analysis of retail sales of a leading European drugstore chain, Rossmann. The company operates across multiple countries and offers a wide range of products including personal care items, household goods, organic food, and health-related products. With thousands of stores and diverse customer demographics, Rossmann caters to day-to-day consumer needs with a focus on affordability and accessibility.

The case study is focused on understanding how Rossmann manages its store operations and promotions. Due to the absence of a well-defined segmentation strategy, resources are not optimally allocated across different stores. Some high-footfall stores underperform due to lack of proper promotional strategies, while other stores in low-competition areas may be over-resourced. Also, promotional events and seasonal variations are not accurately planned, which leads to either overstocking or stockouts and affects customer satisfaction and profitability.

To address these challenges, a dataset containing daily sales and store-related information has been used. The data will be analyzed using visual plots like bar charts, scatter plots, box plots, and heatmaps to understand store performance, customer behaviour, and time-based sales trends. This analysis will help in segmenting stores and identifying the most effective promotional and inventory planning strategies. The outcome will include data-backed recommendations to improve store performance and optimize operations, keeping in mind the practical scope of the project within Rossmann's broader business framework.

Organization Background

Rossmann, officially dirk Rossmann GmbH, is now of the Europe's largest drugstore chains, headquartered in Burgwedel, Germany. It was founded in 1972 by Dirk Rossman, who opened the first self-service drugstore in Hanover. Over the years, Rossmann has expanded significantly, operating more than 4,900 stores across 10 countries, including Germany, Poland, Hungary, the Czech Republic, Turkey and Spain.

The company is 60% family-owned, with the remaining 40% held by A.S. Watson Group. In Germany alone, Rossman runs over 2300 sores and employs more than 56,000 people company-wide. It offers a wide range of products, including personal care items, cosmetics, health products, household goods and organic foods.

Rossmann is especially known for its extensive private-level offerings such as ISANA, Alterra, domaol and enerBiO, which allow the company to main quality while offering competitive prices. The store is typically mid-sized and located in easily accessible areas, catering to everyday consumer needs.

With a strong logistic network and growing online presence, Rossmann continues to strengthen its position in the European retail market by combining affordability, quality and convenience in its business model.

Problem Statements

1. **Store performance segmentation and optimization:**
   Rossmann operates multiple store types with varying customer behaviour and sales patterns. Not identifying performance segments is leading to inefficient resource allocation and missed opportunities in high-potential regions.

2. **Promotion and inventory planning using time series trends:**
   Sales at Rossmann fluctuate due to promotions, holidays and seasonality. Ineffective promotional timing or poor demand forecasting is leading to overstock, stockouts and reduced profitability.

Background of the problems

1. Store Performance Segmentation and Optimization
   Rossmann operates a network of stores across various regions, each differing in size, store type, customer footfall, promotional frequency, and local competition. Despite these differences, many operational and marketing strategies remain generalized, leading to inefficient performance monitoring. Some stores with high customer inflow underperform in sales due to poor promotion targeting, inadequate product assortment, or lack of local competitive strategy. Meanwhile, smaller stores in low-competition zones may be over-resourced. Without segmenting stores based on key performance indicators like average sales, customer count, promotion responsiveness, and competition proximity, the company risks misallocating budgets and missing opportunities to scale profits or reduce losses. A segmentation strategy can help Rossmann make data-driven operational and strategic decisions.

2. Promotion and Inventory Planning Using Time-Series Trends
   Sales at Rossmann are highly influenced by promotions, public holidays, and seasonal shopping patterns. However, not having detailed time-based analysis has resulted in poor planning for promotions. For instance, promotions may be scheduled during periods of already high demand, resulting in unnecessary discounting, while missing key low-demand periods that require stimulus. Additionally, improper coordination between promotional campaigns and inventory levels can result in overstocking or stockouts. Without accurate insights into weekly, monthly, and holiday-based sales patterns, Rossmann risks reduced revenue, increased operational cost, and lower customer satisfaction due to missed expectations. Addressing these inefficiencies through time-based analytics is crucial for strategic planning and improving margins.

Problem – Solving approach

- **Methods**
  To address the identified challenges, the process will begin with cleaning and preparing the data, including handling missing values, standardizing formats, and merging store information with daily sales data.

  The next step will involve exploratory analysis using visual tools such as scatter plots, box plots, histograms, and bar charts to uncover key patterns in sales, customer behaviour, promotional activity, and the impact of competition.

  For store performance segmentation, stores will be categorized based on important business metrics such as average sales, customer count, promotional frequency, and proximity to competitors. This will help in identifying distinct performance groups and underutilized opportunities.

In time-based sales patterns, explore trends across months, weeks, and weekdays. Special focus will be given to understanding how promotions and holidays affect sales volumes, which will help in making informed decisions for future planning around promotions, inventory, and staffing.

- **Data Collection**

  The dataset used for this analysis is the Rossmann Store Sales dataset, which consists of daily sales records and store-level metadata across multiple locations. The data includes transactional information for over 1,000 stores, covering a wide time span with various operational and promotional details.

  The key variables in the dataset include:

  Store, Date, Sales, Customers, Open, Promo, StateHoliday, SchoolHoliday, StoreType, Assortment, CompetitionDistance, CompetitionOpenSinceMonth, CompetitionOpenSinceYear, Promo2, Promo2SinceWeek, Promo2SinceYear, and PromoInterval.

  These features provide insights into sales performance, customer behavior, store classification, promotional campaigns, and the competitive environment, all of which are crucial for performance evaluation and strategic planning.

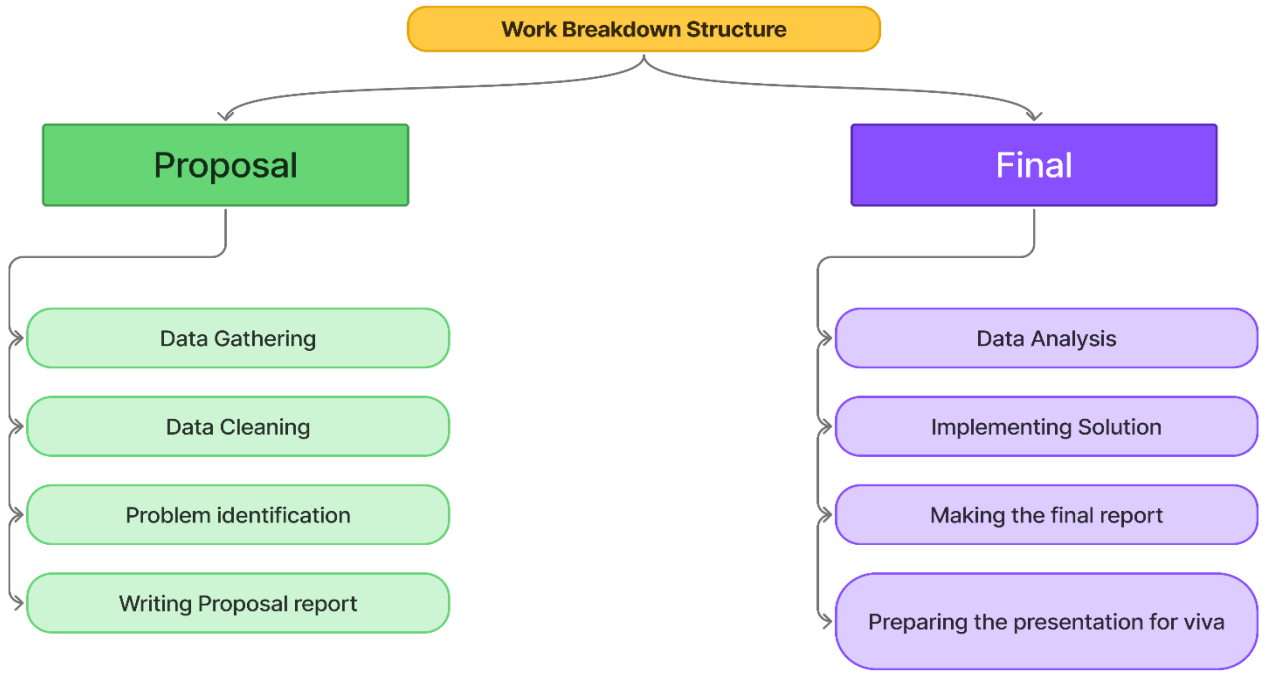
- **Analysis Tools**

  The primary tool used for this analysis is Python, due to its extensive ecosystem for data science and analytics. The Pandas and NumPy libraries will handle data manipulation and preprocessing efficiently. For visualization, Seaborn, Matplotlib will be used to create compelling charts, heatmaps, and dashboards. For clustering and scaling, Scikit-learn offers robust algorithms.

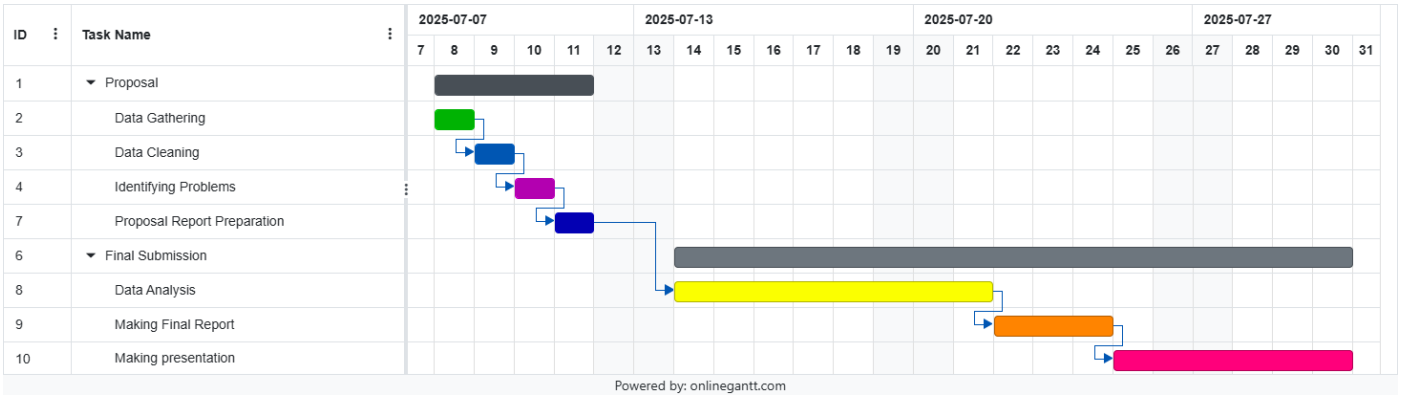
Expected Outcome

- Segment stores into performance groups (high, medium, low) to enable targeted strategic interventions.
- Improve promotional planning and timing based on insights from time-based sales trends and holiday impacts.
- Enhance store-level decision-making by identifying underperforming stores with high potential for improvement.
- Understand customer behaviour patterns and their relationship with store type, assortment, and competitive environment.
- Support inventory and staffing optimization by uncovering sales seasonality and weekly/monthly performance fluctuations.

Expected Timeline



Work Breakdown Structure



Gantt Chart