Store Performance Segmentation and Time-Based Sales Optimization in Retail

Presenter
Mohit Raj Rathor
Roll No: 22F3003190

About the Business

Rossmann, one of Europe's leading drugstore chains with over 3,000 stores across seven countries, faces critical challenges in forecasting seasonal demand, evaluating the effectiveness of promotions, and improving visibility into store-level performance. These issues directly affect sales growth, customer satisfaction, and operational efficiency. For this project, **daily sales data was collected from the Rossmann Store Sales dataset available on Kaggle.**

Defined Problems

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.

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.

Data Description

Rossmann Sales Data

- Collected from https://www.kaggle.com/c/rossmann-store-sales/data
- Consists two csv file namely
  - Stores.csv
  - Sales.csv

Stores.csv

- provides detailed records of Rossmann store data includes competitor's and promotion campaign details
- Contains data of about 1115 stores across Germany.
- Key fields: StoreType, Assortment, CompetitionDistance.

Sales.csv

- includes sales records over time, promotions and holiday's details
- Contains sales data of different stores over the period of 2 year (2013-2015)
- About 1 Millions+ Sales record
- Key fields: Sales, Customers, Promo, StateHoliday, SchoolHoliday.

Problem Solving & Analytics Approach

Data Preparation

- Merged store and sales data.
- Remove and impute missing values in CompetitorDistance, CompetitorOpenMonth/Year etc.

Exploratory Data Analysis

- Explored distributions of customer, sales via histogram, boxplots
- To uncover relationships among fields scatterplot, and covariance matrix is created to analyze.

Feature Engineering

- Extracted day, months, week and year from Date field.
- Based on Business domain KPIs are created to analyze Performance of stores.

Segmentation Analysis

- **Threshold based:** Based on 30 percentile median, on sales and customers.
- **K-Means Cluster Based:** Uncover hidden pattern.
- Classify into following stores:
  - **High**: Strong sales & high traffic
  - **Moderate**: low traffic but high basket size
  - **Low**: balanced but modest performance

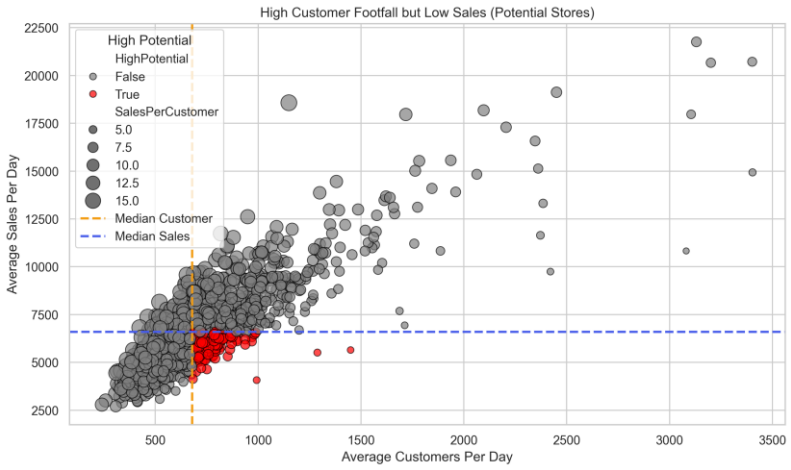
Time Series Analysis

- Aggregate sales data at multiple level monthly, weekly & daily
- Groupby day of week to analyze weekly sales performance across the months.
- Explore seaonality with the effect of promotion
- Via charts: line plots, heatmap & bar plots

Results / Findings

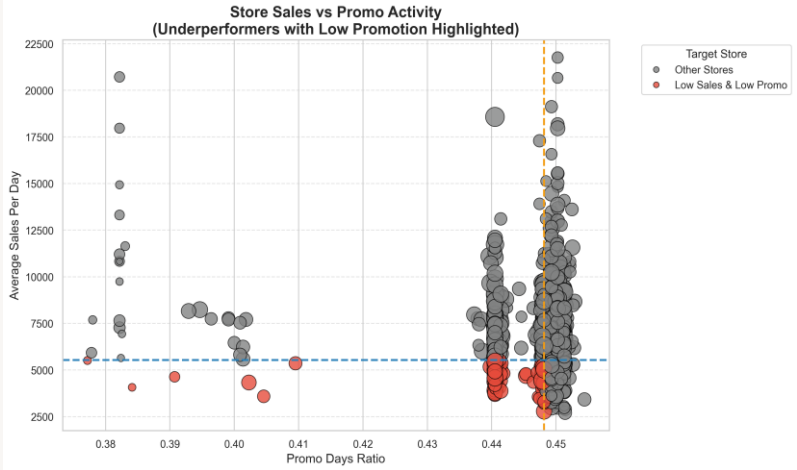
- **Store segmentation**: Underperformers, low-promo stores, 3 performance clusters.
- **Correlations**: Footfall drives sales; high traffic lowers spend per customer.
- **Seasonality**: Dec peak; Jan, Feb, Sep slowdowns.
- **Promotions**: +€2.3K uplift; +38% sales even in low months.
- **Insights**: Weekends strong; promos vital for demand dips.

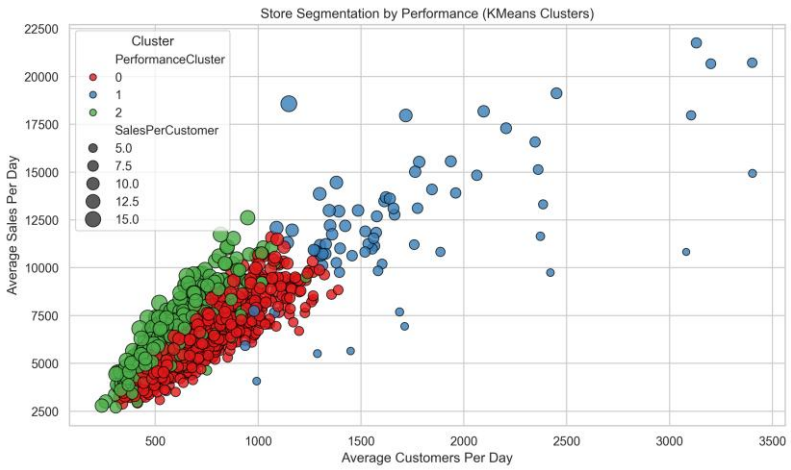
Segmentation Analysis



Threshold Based

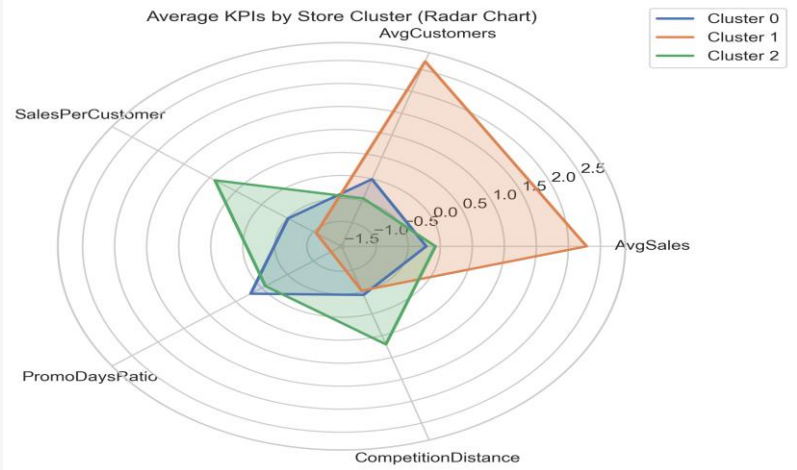
- Applied median thresholds on KPIs (sales, footfall, promos).
- Identify underperforming vs. high-potential stores.
- 107 stores had high traffic but low sales → efficiency gaps.
- 121 stores showed low sales & low promo activity.
- Clear targets for marketing & operational interventions.



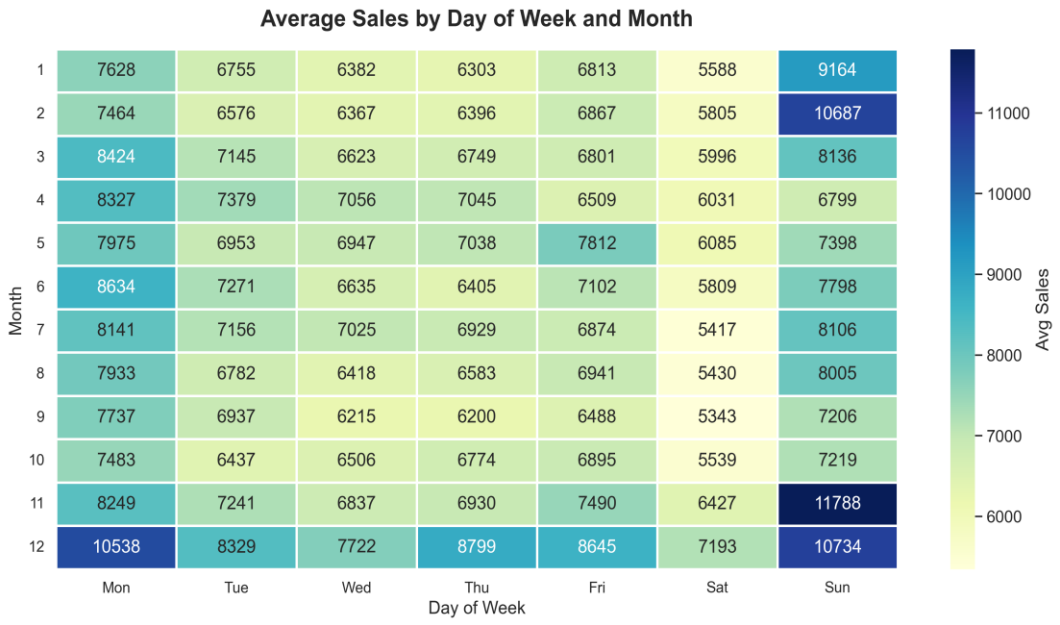


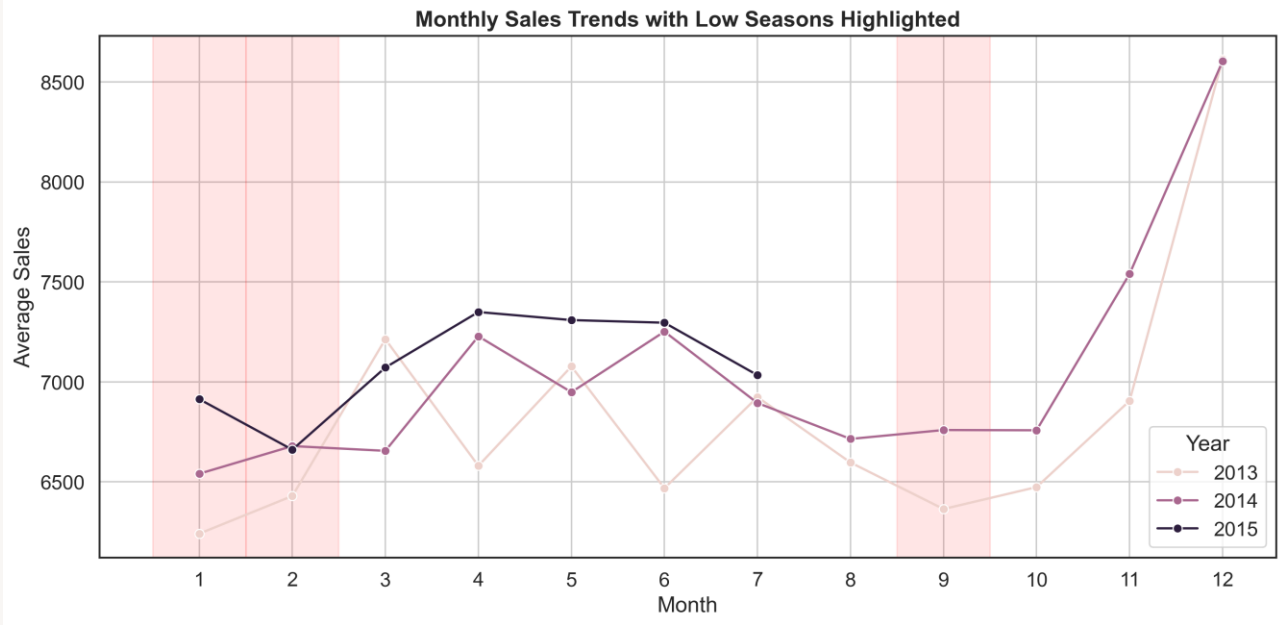
K-Means Cluster Based

- Applied K-Means on store KPIs (sales, customers, basket size, promos).
- Group stores into performance clusters for targeted strategies.
- **Clusters Identified**:
  - Cluster 0 → Mid performers needing optimization
  - Cluster 1 → High sales & high footfall (top performers)
  - Cluster 2 → High basket value, low footfall (efficient conversion)
- Enabled data-driven segmentation for tailored marketing & operational actions.



Time Series Analysis





- Created heatmap of sales across weekdays and months.
- Identify seasonal and weekly shopping behavior patterns.
- Sundays in November showed peak festive sales; Saturdays in September lowest.
- Strong weekend shopping behavior; weekdays like Mondays also contributed high sales.
- Helped optimize promo timing and staffing for peak demand periods.

- Analyzed monthly sales patterns using time-series aggregation.
- Identify seasonal and weekly shopping behavior patterns.
- Detect peak demand cycles and identify seasonal slowdowns.
- December consistently peaked (holiday demand), while Jan, Feb, and Sep were low-performing.
- Seasonal dips indicate opportunities for targeted promotions.
- Informed inventory and campaign planning to stabilize revenue.

Key Results

Threshold Segmentation

Classified stores based on sales, traffic, and promotions to identify inefficiencies.

High Potential Stores

107 outlets had strong traffic but weak sales, pointing to marketing and service gaps.

Low Sales–Low Promo Stores

121 outlets showed weak performance, requiring promotional uplift.

K-Means Clustering

Grouped stores into 3 clusters—top performers, mid-performers, and high basket–low traffic.

Correlation Insights

Strong traffic–sales link (r = 0.83) confirmed; negative traffic–basket size relation noted.

Promo Effectiveness

Promotions boosted average sales by €2,298 per month, even in low seasons.

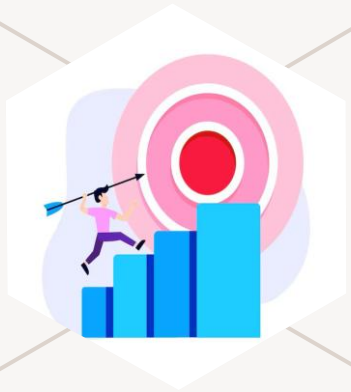
Seasonality Trends

December peaked in sales; Jan, Feb, Sep marked as low-demand months.

Heatmap Analysis

Sundays & Mondays showed highest sales, while Saturdays in September were weakest.

Recommendations



Target Underperforming Stores

Strengthen marketing and service quality in low sales–high traffic outlets → unlock untapped revenue potential.

Optimize Promotions & Timing

Align promotions with low-demand months and weak weekdays → stabilize sales and improve inventory turnover.

Leverage Customer Insights

Try product mix and campaigns by store clusters → enhance customer experience and drive sustainable growth.




Thank you

Mohit Raj Rathor

Roll number: 22f3003109

22f3003109@ds.study.iitm.ac.in