Store Performance Segmentation and Time-Based Sales Optimization in Retail

Final 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

Executive Summary

This report presents a comprehensive data analytics project for Rossmann, a leading European drugstore chain founded in 1972. Despite its wide reach, Rossmann faces performance gaps due to inconsistent store output, poor promotional timing, and inefficient inventory planning affecting profitability and customer experience. These challenges prompted a data-driven strategy to improve decision making.

Daily sales data from 2013 to 2015 was cleaned, structured, and enriched with time features to extract business-critical insights. The analysis cantered on two powerful techniques: store performance segmentation using K-Means clustering and threshold metrics, and time-series trend analysis to assess seasonality, holiday effects, and promotion impact. This dual pronged approach uncovered high-traffic stores with low basket value, pinpointed weak sales months, and quantified promotional uplift.

Key findings show that promotions consistently boosted monthly sales by €2,298 on average, with promo days generating €8,240 compared to €5,942 on non-promo days. Clustering analysis divided stores into three segments: high-traffic high-revenue stores, low-traffic high basket value stores, and average-performing stores enabling targeted operational strategies. Time-series trends revealed strong seasonality, with peak sales during December and week 51, and underperformance in January, February, and September. However, these low months responded well to promotions, demonstrating untapped potential.

The project delivers actionable insights to enhance store operations, targeted marketing, and demand-based inventory planning, positioning Rossmann for proactive, data-led growth.

Proof of Originality

The dataset used in this project. Titled **"Rossmann store sales"**, and it was collected from **Kaggle.com**, a popular platform for AI and data science related competition and datasets. This dataset was used in Rossmann store sales prediction competition hosted back in 2016 and provides extensive information on Walmart's sales across the Germany and some European countries. The data includes crucial variables such as store details, competitor's distance, store sales over time and promotional campaign data. The dataset was downloaded directly from source and has not been modified or manipulated externally, ensuring the originality and integrity of the data for analysis.

The link for the dataset - https://www.kaggle.com/c/rossmann-store-sales/data

Metadata and Descriptive Statistics

The Rossmann store sales dataset has two files named store.csv provides detailed records of Rossmann store data includes competitor's and promotion campaign details, sales.csv data includes sales records over time, promotions and holiday's details.

The columns in the store.csv are as follows:

Column Name	Data Type	Description
Store	Integer	Unique store ID
StoreType	Object	store categories - "a", "b", "c", "d"
Assortment	Object	assortment categories - "a", "b", "c"
CompetitionDistance	Integer	Competitor's distance from stores
CompetitionOpenSinceMonth	Integer	Competitor opening month
CompetitionOpenSinceYear	Integer	Competitor opening year
Promo2	Boolean	Does promo 2 applied on this store
Promo2SinceWeek	Integer	If promo 2 applied then it's starting week
Promo2SinceYear	Integer	If promo 2 applied then it's starting year
PromoInterval	Object	Promo interval

Descriptive Statistics:

	Store	CompetitionDistance	CompetitionOpenSinceMonth	CompetitionOpenSinceYear	Promo2	Promo2SinceWeek	Promo2SinceYear
count	1115.00	1112.00	761.00	761.00	1115.00	571.00	571.00
mean	558.00	5404.90	7.22	2008.67	0.51	23.60	2011.76
std	322.02	7663.17	3.21	6.20	0.50	14.14	1.67
min	1.00	20.00	1.00	1900.00	0.00	1.00	2009.00
25%	279.50	717.50	4.00	2006.00	0.00	13.00	2011.00
50%	558.00	2325.00	8.00	2010.00	1.00	22.00	2012.00
75%	836.50	6882.50	10.00	2013.00	1.00	37.00	2013.00
max	1115.00	75860.00	12.00	2015.00	1.00	50.00	2015.00

Numerical variable’s summary:

The dataset includes metadata for 1,115 Rossmann stores, focusing on competition and promotion-related attributes.

- Competition Distance averages around 5,405 meters, with high variability. Distances range from 20 meters to over 75,000 meters, indicating diverse competitive environments.
- Competition Open Since data is partially available (for 761 stores), with most competitors opening between 2006 and 2013.
- Promo2 Participation is nearly evenly split, with 51% of stores enrolled. Among participating stores, most joined between 2011 and 2013, typically around Week 24 of the year.

Categorical variable’s summary:

- Store Type:
  - There are 4 unique store types, with Type 'a' being the most common, representing 54% of all stores.
  - Types 'd' and 'c' account for 31% and 13% respectively, while Type 'b' is rare, comprising just 1.5% of the store network.
- Assortment Type:
  - The dataset includes 3 assortment levels, with Type 'a' (basic assortment) comprising 53% of stores and Type 'c' (extended assortment) close behind at 46%.
  - Type 'b', representing an intermediate assortment, is minimal (<1%), indicating limited use of mid-level product variety.

And columns in sales.csv are as follows:

Column	Data Type	Description
Store	Integer	Store ID
DayOfWeek	Integer	Week day of entry
Date	Date	Date of purchase
Sales	Integer	Sales happened on that day
Customers	Integer	Customer visits on that day
Open	Boolean	Is store still operational
Promo	Boolean	Is promotion applied on that day
StateHoliday	Boolean	Is the day is holiday
SchoolHoliday	Boolean	Is the day is School holiday

Descriptive Statistics:

	Store	DayOfWeek	Date	Sales	Customers	Open	Promo	StateHoliday	SchoolHoliday
0	1	5	31-07-2015	5263	555	1	1	0	1
1	2	5	31-07-2015	6064	625	1	1	0	1
2	3	5	31-07-2015	8314	821	1	1	0	1
3	4	5	31-07-2015	13995	1498	1	1	0	1
4	5	5	31-07-2015	4822	559	1	1	0	1

Numerical variable's summary:

The dataset comprises over 1 million daily sales records across Rossmann stores, capturing operational and customer behaviour variables.

- Sales have a mean of €5,774, with wide variation (up to €41,551), reflecting diverse store sizes and performance levels.
- Customer visits average 633 per day, ranging from 0 to over 7,000, indicating fluctuations in foot traffic.
- Store Open Rate is 83%, meaning most entries correspond to operating days.
- Promotions were active on 38% of the days, while School Holidays were recorded in 18% of the data.
- The Day of Week is evenly distributed, confirming comprehensive weekly coverage.

Categorical Variable's summary:

- State Holidays were observed on a relatively small portion of the dataset:
    - No holiday (0): 986,159 records (approximately 97%)
    - Holiday 'a': 20,260 records
    - Holiday 'b': 6,690 records
    - Holiday 'c': 4,100 records
- State holidays are rare events, occurring in only about 3% of total records, with type 'a' being the most common among them.
- School Holidays occurred in 181,721 records (nearly 18%), while the remaining 835,488 records (nearly 82%) fall on regular school days, indicating that school closures are less frequent and may influence customer traffic on specific dates.
- Day of the Week is evenly distributed, suggesting a consistent data collection across all weekdays:

- Monday to Saturday each have nearly 144,730 - 145,845 records
- Sunday has the lowest frequency, but still comparable, indicating balanced daily representation useful for analysing day-wise trends.

Detailed Explanation of Analysis Process

Data Cleaning and Feature Engineering process

To prepare the dataset for meaningful analysis, a comprehensive data preprocessing phase was followed:

- Date Conversion
  The Date column was converted into datetime format to enable efficient time-based operations.
- Imputing Missing Values
  - CompetitionDistance, which represents the distance to the nearest competitor, had missing entries filled using the median value. This choice preserves distribution characteristics while avoiding the influence of extreme values.
  - CompetitionOpenSinceMonth and CompetitionOpenSinceYear were imputed with default values: 1 for the month and the corresponding year from the transaction date. This approach assumes competition began at the start of the transaction year when no explicit information is available.
- Deriving Competition Duration A new datetime column, CompetitionOpenDate, was generated using the imputed year and month fields. From this, CompetitionMonthsActive was computed to capture how many months a store has been facing competition at the time of each transaction. Negative values were clipped to zero to avoid unrealistic durations.
- Handling Promo2 Campaign Information Missing values in promotional fields were addressed with:
  - Zeros for Promo2SinceWeek and Promo2SinceYear, indicating no ongoing promotions.
  - Empty strings for PromoInterval, marking stores with no seasonal promotional cycles.
- Feature Engineering
  Some new feature created with the use of pandas groupby function:

```
# Group by Store for store-level aggregates
store_kpis = df_open.groupby('Store').agg(
    AvgSales=('Sales', 'mean'),
    AvgCustomers=('Customers', 'mean'),
    SalesPerCustomer=('SalesPerCustomer', 'mean'),
    OpenDays=('Date', 'count'),
    PromoDays=('Promo', 'sum'),
    CompetitionDistance=('CompetitionDistance', 'mean'),
    CompetitionMonthsActive=('CompetitionMonthsActive', 'mean'),
    Promo2Active=('Promo2Active', 'max'),  # If active at least once
    IsPromoMonth=('IsPromoMonth', 'sum')  # Count of months with promo
).reset_index()
```

  - AvgSales was derived from mean sales each store, similarly AvgCustomer was from mean customer footfall each store.
  - SalesPerStore (Basket Size) derived by dividing AvgSales with AvgCustomer which play a significant role in threshold-based segmentation.
  - Promo2StartDate was derived from Promo2SinceYear and Promo2SinceWeek, translating the campaign start into a precise date.

- Promo2Active is a binary feature indicating whether a store is currently under Promo2 based on the transaction date.
- IsPromoMonth captures whether the current transaction month aligns with the store's scheduled promo intervals, enhancing the ability to model seasonally recurring promotions.

Store Segmentation Analysis

Each store varies in AvgCustomer, SalesPerCustomer, promotional responsiveness and competitive distance. Hence, a data-driven segmentation approach is employed to classify stores into distinct performance clusters for more targeted business planning.

❖ Metric Derivation and Feature Selection:
The segmentation is based on a combination of key performance indicators and business relevant contextual variables. Key derived metrics includes:

- Average Sales per store: Indicates revenue – generating capacity:
- Average Customer per store: Reflects footfall and potential demand.
- Sales per Customer (Basket Size): Measures revenue per visit.
- Promo Days Ratio: Proportion of days a store run promotions.
- Competition Distance: Gauges ease of access to competitor stores.
- Competition months active: Reflects maturity and competitive exposer of a store.

All metrics are standardized to ensure comparability across different scales.

❖ Segmentation and techniques:
Primary segmentation strategies used are as follow:

- The threshold-based segmentation approach involved the use of statistical cutoffs, particularly quantiles, to manually classify stores into distinct performance categories—namely high, medium, and low performers. Key performance indicators (KPIs) such as average daily sales, average customer footfall, and sales per customer were used as the basis for segmentation. By applying quantile-based thresholds (median, 30th percentile), this method allowed for targeted identification of outlier stores, such as those attracting high customer traffic but generating low revenue per customer, which indicates underutilized potential.

- K-Means Clustering—was applied for uncovering hidden patterns within the store performance data. K-Means is an unsupervised learning algorithm that partitions the dataset into K clusters, such that each store is grouped with others exhibiting similar characteristics based on selected performance indicators. In this case, features such as Average Sales, Average Customers per Day, and Sales per Customer were used to define store behaviour.

  The algorithm operates by minimizing the within-cluster variance through iterative updates of cluster centroids. The core objective of K-Means is to minimize the inertia, defined as the sum of squared distances between each data point and its assigned cluster centroid. Mathematically, the K-Means objective function is:

$$J = \sum_{i=1}^{K} \sum_{x \in C_i} ||x - \mu_i||^2$$

Where:

- K is the number of clusters,
- C_i represents the set of data points in cluster i,
- μ_i is the centroid of cluster i, and
- $||x - \mu_i||^2$ is the squared Euclidean distance between a data point x and its corresponding cluster centroid μ_i.

To determine the optimal number of clusters (K), the Elbow Method was employed. This involves calculating the total within-cluster sum of squares (WCSS) for a range of K values and plotting them against the respective K. The 'elbow point'—where the rate of decrease in WCSS sharply declines—indicates the ideal number of clusters that balances granularity with interpretability. Based on the Elbow curve generated during the analysis, K=3 was identified as the optimal value.

Applying K-Means clustering with these optimized parameters resulted in three distinct store segments:

1. High-performing stores with strong sales and customer traffic.
2. Efficient stores with moderate traffic but high basket values.
3. Average performers with balanced but modest KPIs.

This approach uncovers hidden patterns that are not apparent from traditional sorting or filtering.

Time Series Trend Analysis

To understand how sales patterns vary over time, a series of time-series analyses performed using Rossmann's daily transaction data. This process aimed to uncover seasonal trends, recurring weekly behaviours, and the influence of specific events such as promotions and holidays on store performance.

The first step involved temporal aggregation of the sales data at multiple levels monthly, weekly, and daily. Monthly average sales were computed to detect long-term seasonality, such as consistent peaks during December and dips in January and September. Weekly sales trends analysed to capture finer-grained cyclic patterns, such as periodic increases near the end of the year or during promotional campaigns. Additionally, sales were grouped by day of the week to examine behavioural differences in consumer activity across weekdays and weekends. This helped identify which days generally contributed most to overall revenue.

To measure the impact of specific events, the data was segmented based on promotional activity and holiday flags. Sales figures were compared between promotional and non-promotional days to evaluate the effectiveness of marketing efforts. A similar comparison done between regular days and state or school holidays to assess how these events influenced sales volume and customer behaviour across different time periods.

Line plots used to track overall sales trends across months and weeks, highlighting seasonality and volatility. Heatmaps generated to show interactions between month and day-of-week performance, allowing the identification of consistently high or low-performing time blocks. Area plots helped visualize cumulative sales trends and broader business movement over time.

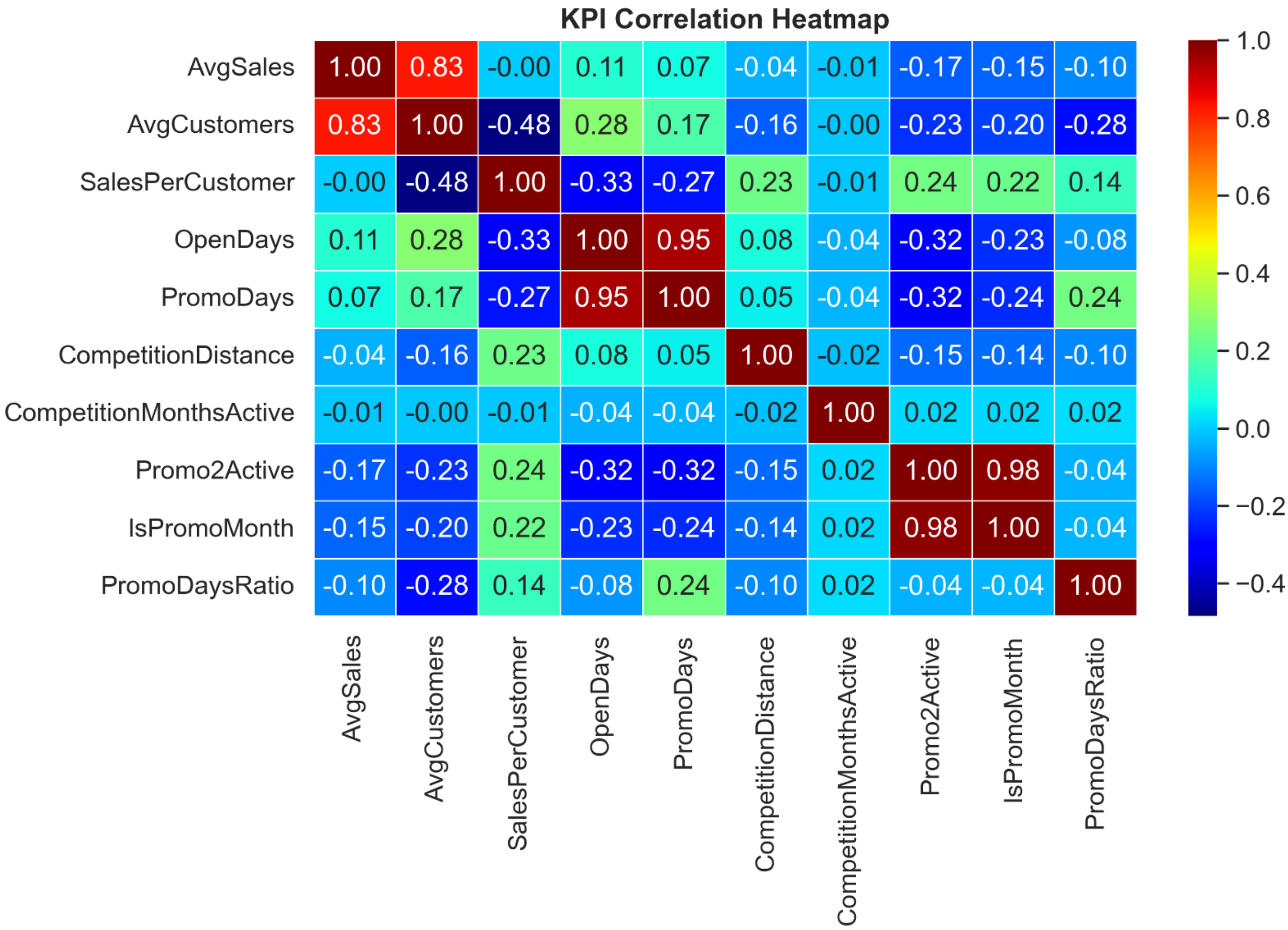
Results and Findings

Store Performance Segmentation and Optimization

Based on the analysis performed using both threshold-based metrics and unsupervised clustering (K-Means), the key insights derived as follows:

- ❖ Store Key Performance Metrics
    - Average Sales Per Store: Varied significantly, ranging from under 2703.74 to 21757.48, indicating substantial disparity in revenue generation.
    - Customer Footfall showed diversity, stores with lower sales often had comparable or higher customer counts but smaller basket sizes.
    - Basket Size (Sales/Customer) emerged as a critical metric in distinguishing high-performing stores from low performing ones as store having nearly same customer footfall have different Basket Sizes indicating some stores are underperforming.

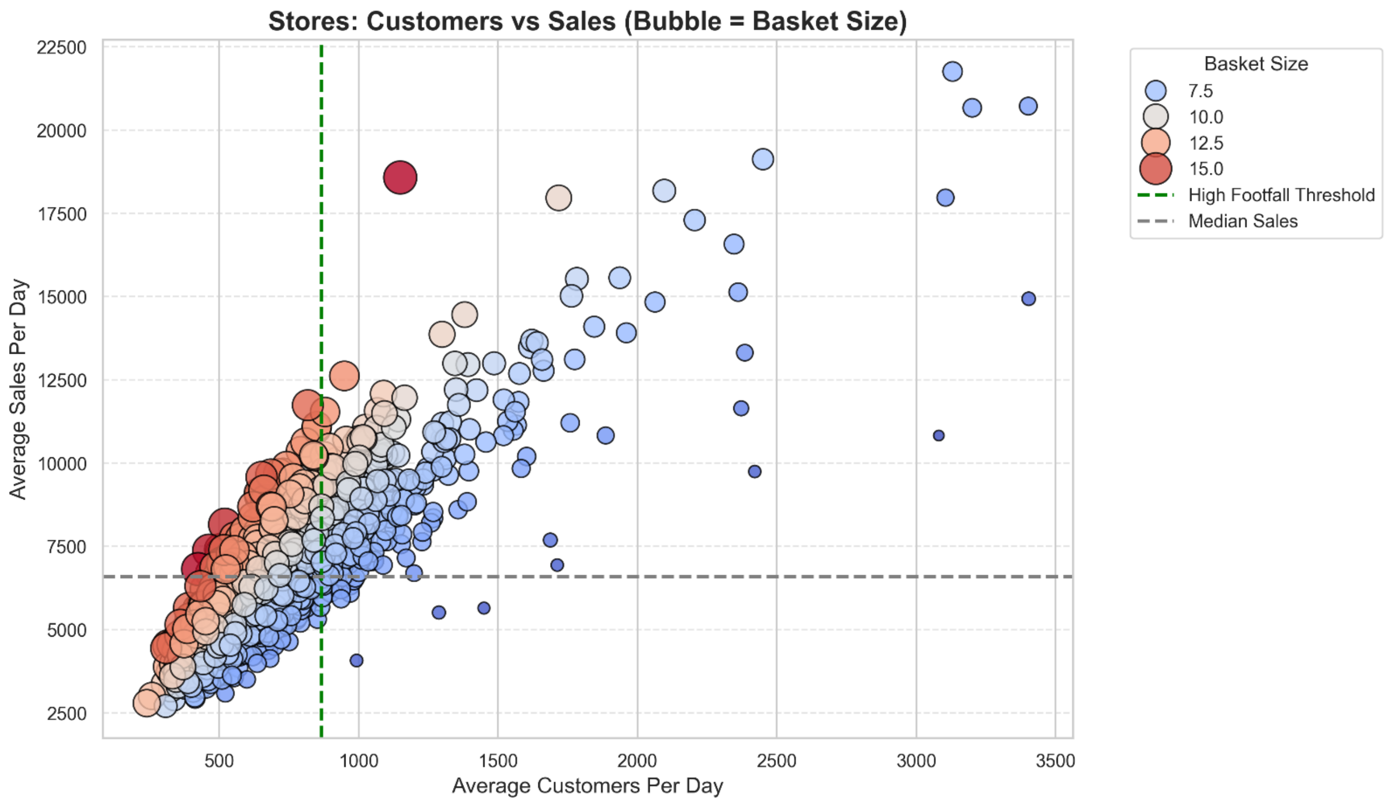
- ❖ Correlational Analysis on KPIs:
    - AvgSales and AvgCustomer exhibit as strong positive correlation (r=0.83), confirming that higher customer traffic generally leads to higher sales.
    - SalesPerCustomer is not significantly correlated with AvgSales (r=0), suggesting that increased traffic does not necessarily lead to greater basked size or spending per customer.
    - SalesPerCustomer shows a negative correlation with AvgCustomer (r = -0.48), indicating that stores with high footfall may struggle to maintain high value per customer possibly due to crowding, product availability or pricing issues.
    - Promo2Actice and SalesPerCustomer show a positive correlation (r=0.24), suggesting extended promotions may increase individual customer spend.
    - However, Promo2Acitve correlates negatively with AvgSales (r=-0.17) and AvgCustomer (r=-0.23), implying that longer promotions may not drive higher overall traffic or revenue all stores.



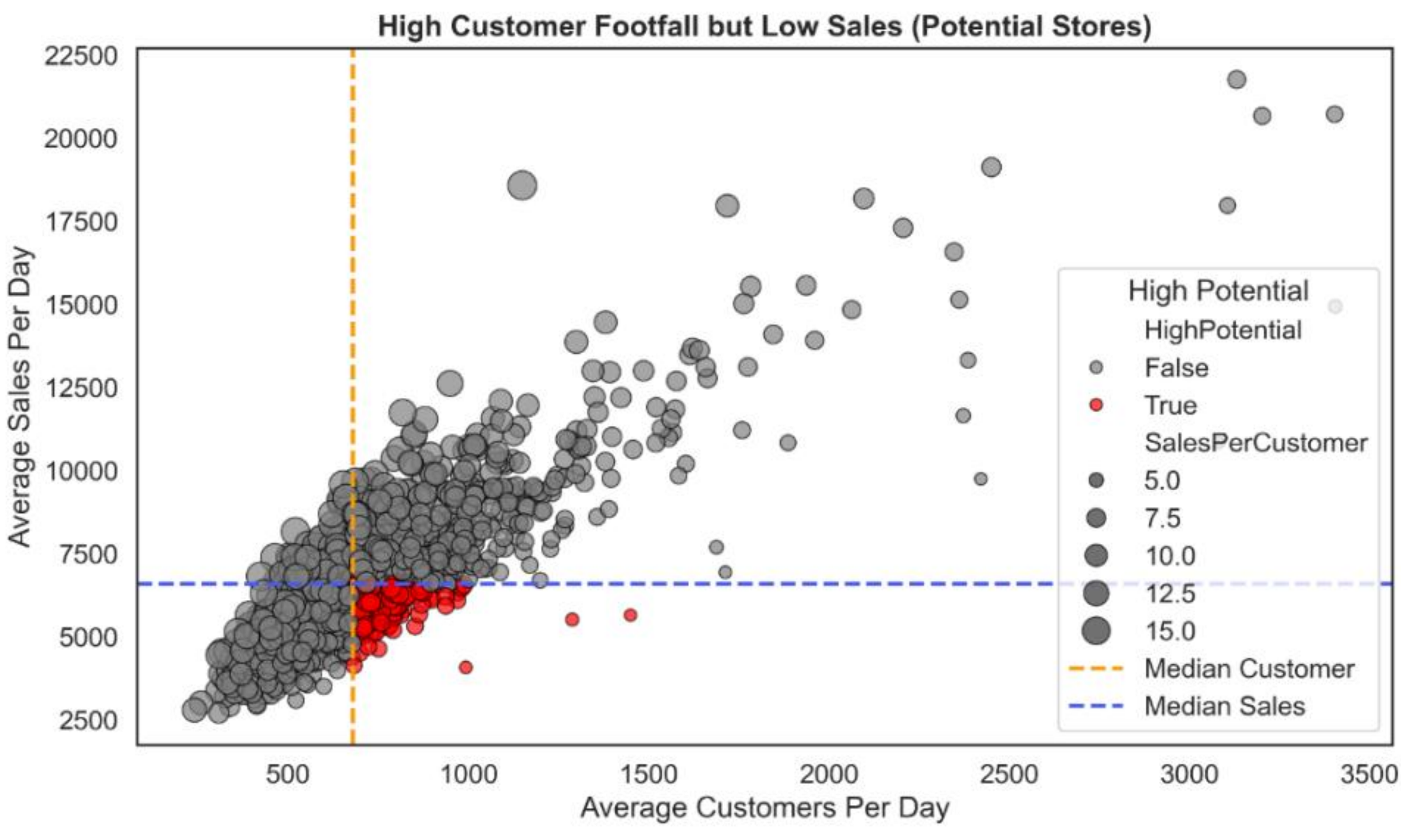
- Promodays and OpenDays are strongly correlated (r=0.95), indicating consistent promotional acitivity across the year.
- PromoDaysRatio is negatively associated with customer engagement (r = - 0.28 with AvgCustomers), pointing to diminishing returns from high-frequency promotions.
- CompetitionDistance shows a slight negative correlation with AvgCustomer (r = -0.16), indicating that store with closer competition tend to have reduced customer traffic.
- Interestingly, CustomerDistance is positively correlated with SalesPerCustomer (r=0.23), hinting that stores in less competitive regions tend to extract more value per shopper.

❖ Threshold-Based Segmentation:

A detailed threshold-based analysis revealed that 151 stores had above-average customer footfall (greater than 866.2 customers/day) but a basket value lower than €8.13, indicating underperformance in converting customer volume into revenue.



Among these, Store 769 stood out as a critical case. Despite receiving an impressive 3,081 customers per day, it achieved only €3.51 in sales per customer, significantly below the overall store median basket size. This store, categorized as StoreType 'b' with Assortment 'b', generated an average daily revenue of €10,825.90, which is above the median sales of €6,589.95, but the low basket value points to missed revenue opportunities from its large customer base.



- The goal was to isolate high-potential stores—those experiencing above-median customer footfall but generating below-median sales.
- Using median thresholds (€6,589.95 for sales and 678.67 customers/day), we identified 107 stores that meet this criterion. These stores attract substantial footfall yet underperform.

- in revenue generation, indicating potential issues such as:
- Ineffective in-store marketing or promotions
- Poor customer service or product availability
- Suboptimal store layout or checkout processes

Top High-Potential Stores (based on Basket Value):

Store	AvgCustomers	AvgSales	SalesPerCustomer	CompetitionDistance	PromoRatio
353	1451.09	5651.37	3.89	900	0.38
274	992.74	4070.83	4.10	3640	0.38
512	1287.97	5510.98	4.28	590	0.38
811	682.77	4132.04	6.05	410	0.45
530	750.81	4634.32	6.17	18160	0.39

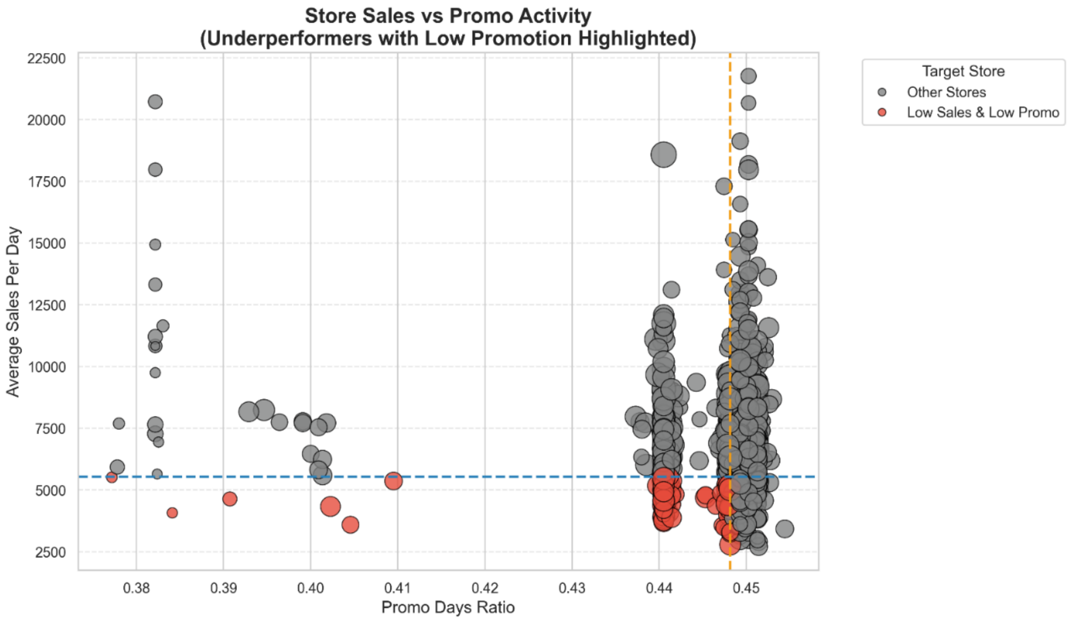
❖ Underperforming Store Cluster based on Promotion Activity

An evaluation of promotion engagement versus sales performance revealed a subset of underperforming stores characterized by both low average daily sales and limited promotional activity. Thresholds were set at the 30th percentile for both metrics:

- Low Sales Threshold: €5,538.62
- Low Promo Days Ratio Threshold: 0.45

A total of 121 stores fell into this underperforming category, suggesting potential missed opportunities where strategic promotional efforts could positively influence revenue.

These stores may be overlooked in current marketing or campaign strategies and could benefit from focused interventions such as increased advertising, localized promotions, or targeted campaigns to boost customer engagement and conversion.



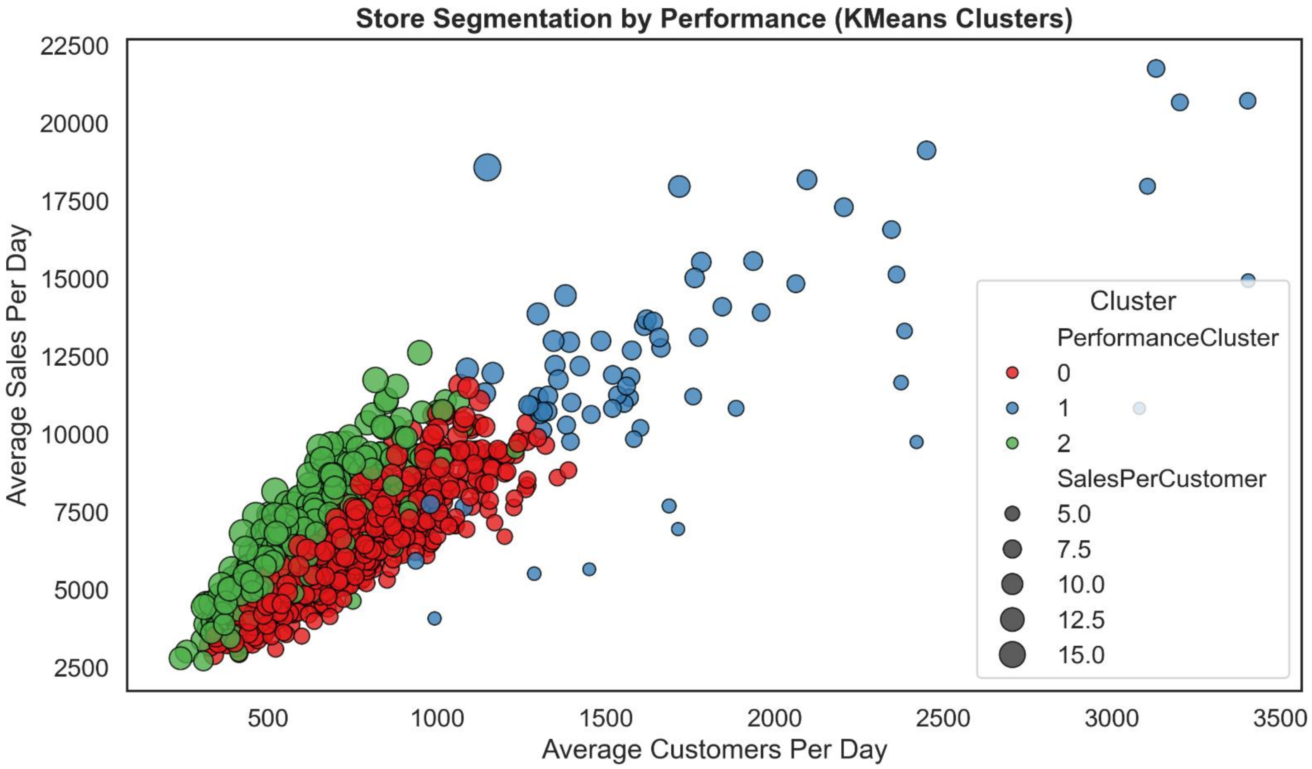
Top Underperforming Stores (Based on Average Sales):

Store	Avg Sales	Promo Days Ratio	Sales/Customer
543	2790.38	0.45	11.62
656	3245.19	0.45	8.56
970	3295.74	0.45	8.24
789	3380.01	0.45	7.2
837	3473.72	0.45	7.55

❖ Cluster-Based Segmentation Using Unsupervised Learning:

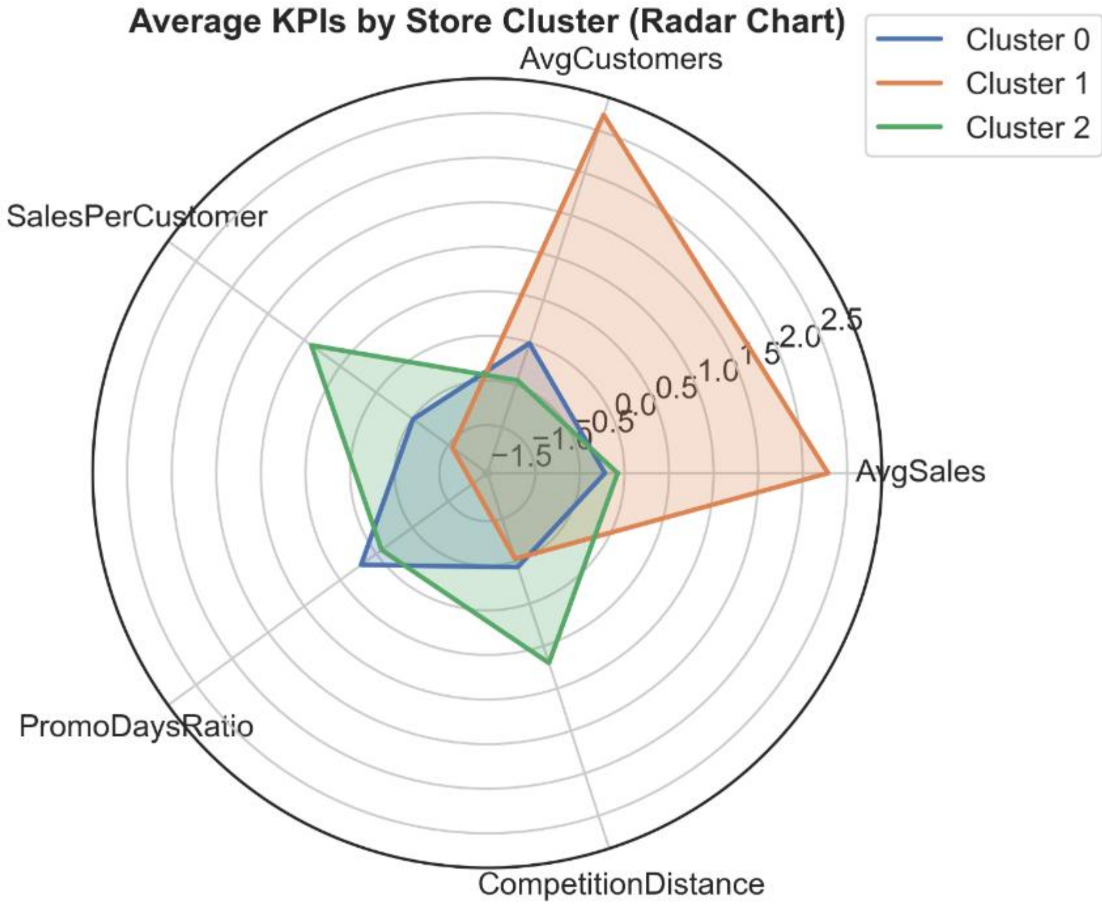
To uncover underlying performance patterns among stores, KMeans clustering was applied on key performance indicators—Average Sales, Average Customers, and Sales per Customer. The clustering algorithm identified three distinct store segments, each representing unique operational profiles:

PerformanceCluster	NumStores	AvgSales	AvgCustomers	SalesPerCustomer
0	611	6412.25	752.08	8.63
1	73	12376.5	1703.38	7.55
2	431	6753.46	597.39	11.43



- Cluster 1 represents the top-performing, high-traffic stores with the highest average sales and footfall, but slightly lower sales per customer.
- Cluster 2 comprises stores with strong basket value (11.43) but moderate customer count, indicating high individual spending despite average sales volume.
- Cluster 0 consists of majority of stores with balanced but relatively lower performance, marking them as standard or average performers.

❖ Radar chart provides a multidimensional comparison of average key performance indicators (KPIs) across the previously identified store performance clusters. The KPIs considered included:



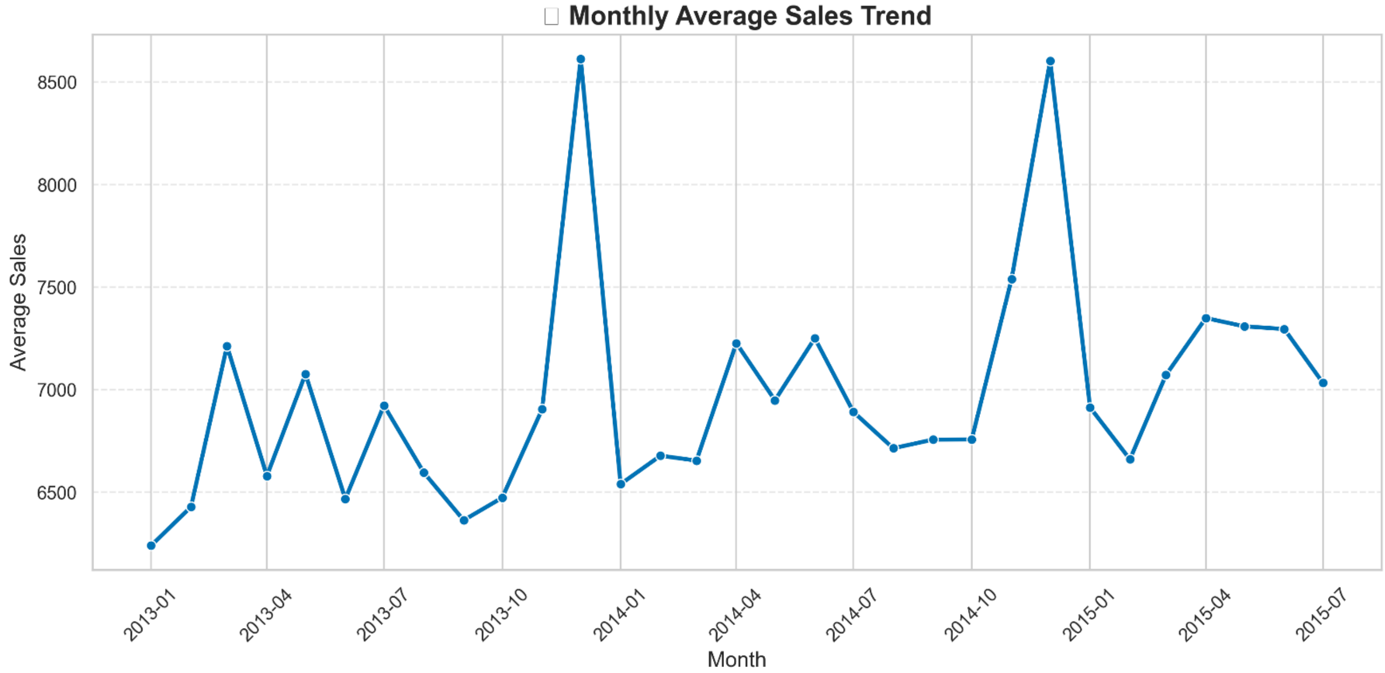
- Average Sales
- Average Customers
- Sales per Customer
- Promo Days Ratio
- Competition Distance

Key insights from the radar chart:

- Cluster 1 (High-performing stores) dominates in AvgSales and AvgCustomers, reflecting strong traffic and revenue generation but moderate SalesPerCustomer and relatively lower promotion intensity.
- Cluster 2 shows a strong edge in SalesPerCustomer, despite lower footfall, suggesting efficient conversion or higher individual spending.
- Cluster 0 maintains balanced but modest values across all KPIs, confirming its position as the average-performing segment.

Time Series Trends

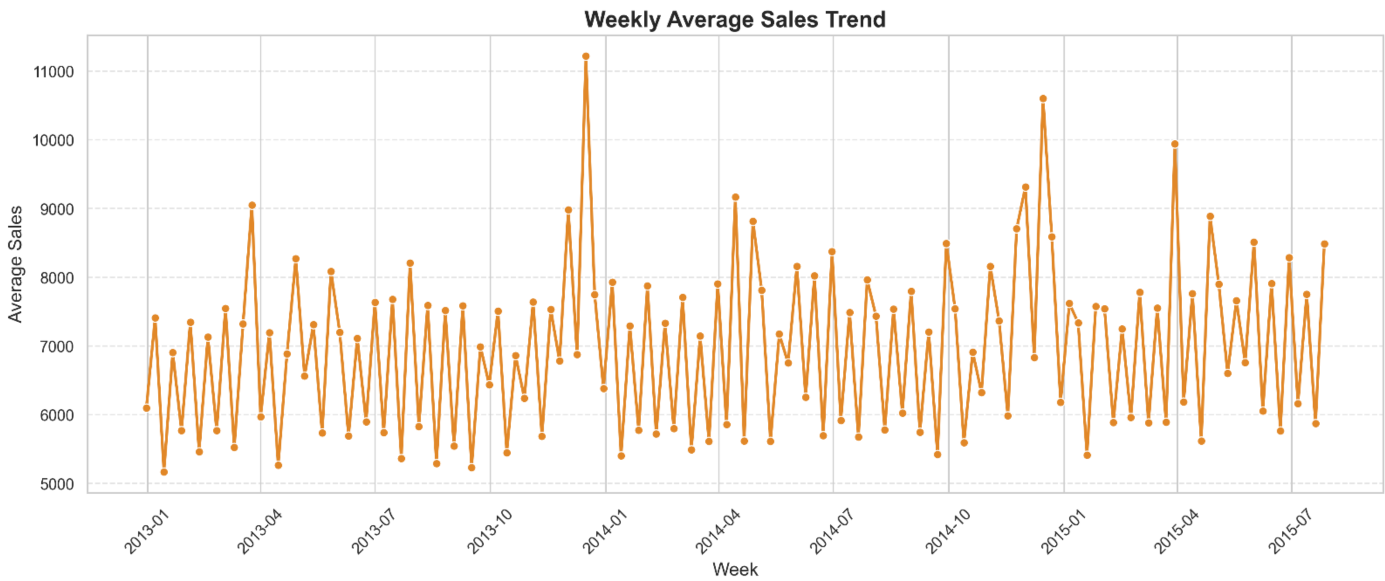
❖ Monthly Sales Trend:
Over a span of 31 months from January 2013 to July 2025, the sales patterns exhibited distinct seasonal behaviour.



- The highest average monthly sales were recorded in December 2013 at €8,613.46, likely driven by year-end shopping and holiday effects.

- In contrast, January 2013 marked the lowest monthly average at €6,240.51, possibly reflecting post-holiday dips in consumer activity.

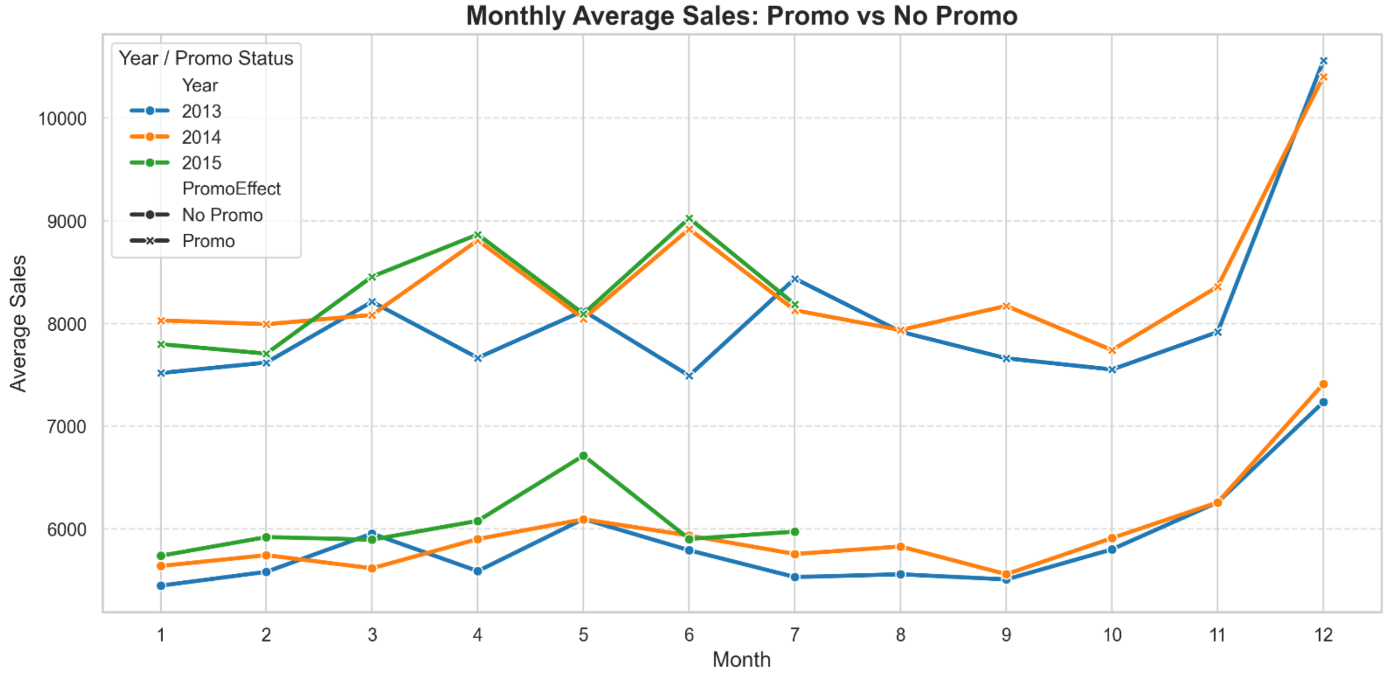
❖ Weekly Sales Trend:
Across 135 weeks, weekly trends provided a more granular view of fluctuations.



- The peak weekly average sales occurred in Week 51 of 2013, reaching €11,219.31, aligning with peak holiday periods.

- Conversely, Week 3 of 2013 recorded the lowest weekly average, with €5,167.26, highlighting early-year lulls.

❖ Effectiveness of promotional strategies over time:
compared monthly average sales between promotional and non-promotional periods across multiple years.



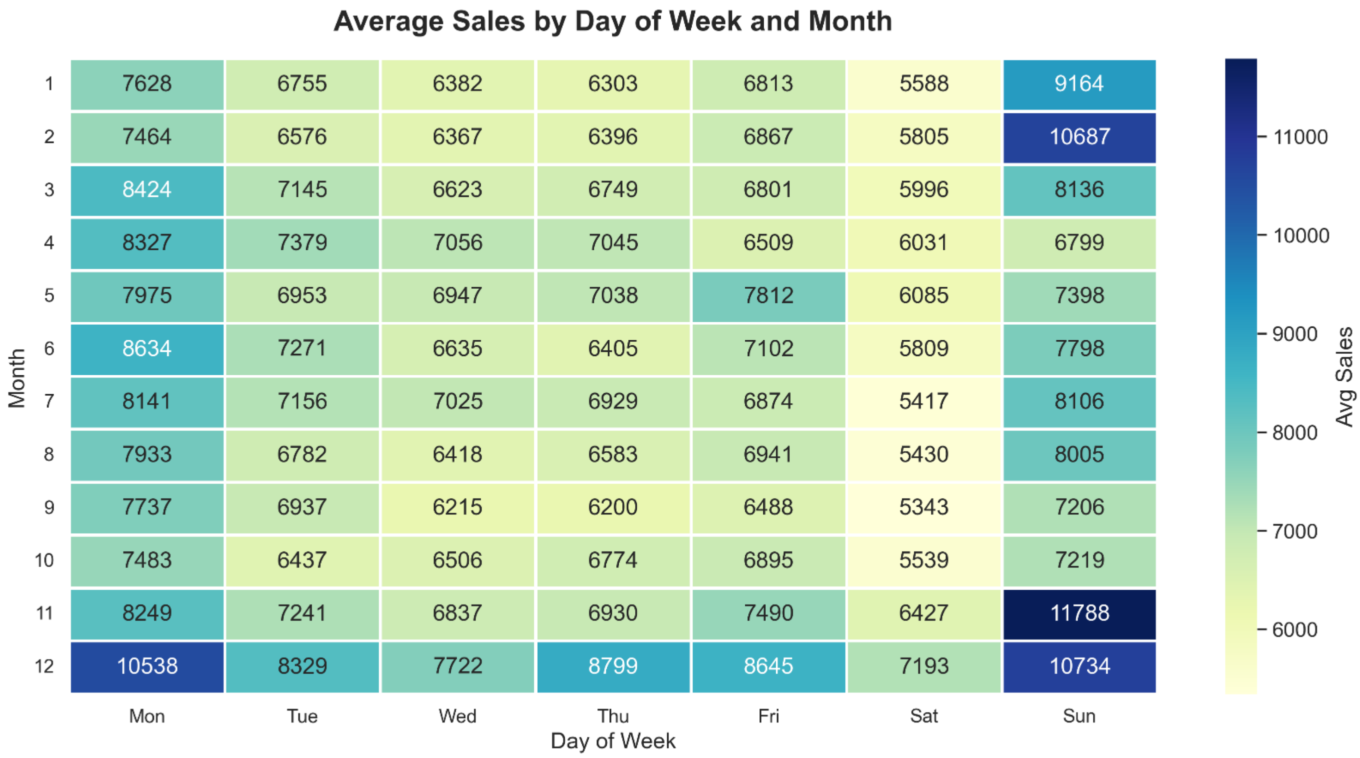
- Promotional campaigns had a significant positive impact on monthly sales, with promo days yielding an average of €8,240.19 in sales compared to €5,941.95 on non-promo days.
- The average uplift in monthly sales due to promotions was €2,298.24.
- Sales on promo days consistently outperformed non-promo days across all months, validating the effectiveness of promotional events.
- The highest sales during promo periods peaked at €10,561.61, while non-promo sales topped at €7,412.06.
- Even the lowest promo-month sales (€7,491.84) were higher than the average non-promo sales, further emphasizing their contribution.

❖ Average Sales by Day of Week and Month (Heatmap Analysis):

Analysed interaction between monthly seasonality and day-of-week sales behaviour to uncover time-based performance patterns across Rossmann stores.

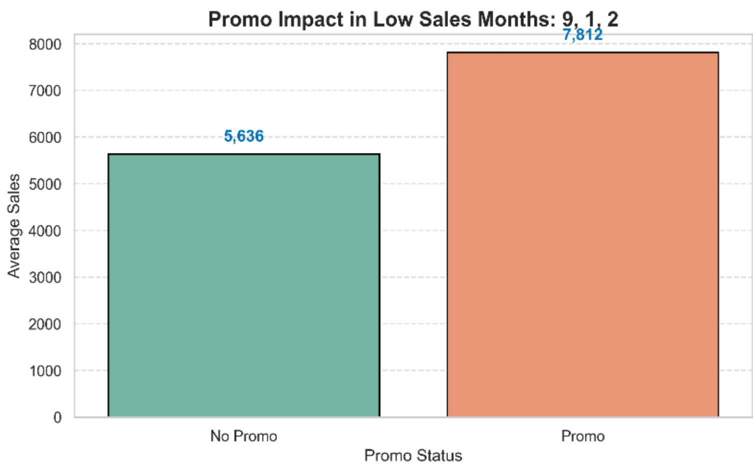
- Peak Sales Period: Sundays in November, recording the highest average sales at €11,787.91, suggesting strong year-end shopping activity or seasonal promotions.

- Lowest Sales Period: Saturdays in September, with the lowest average at €5,342.93, indicating potential opportunity for promotional interventions.

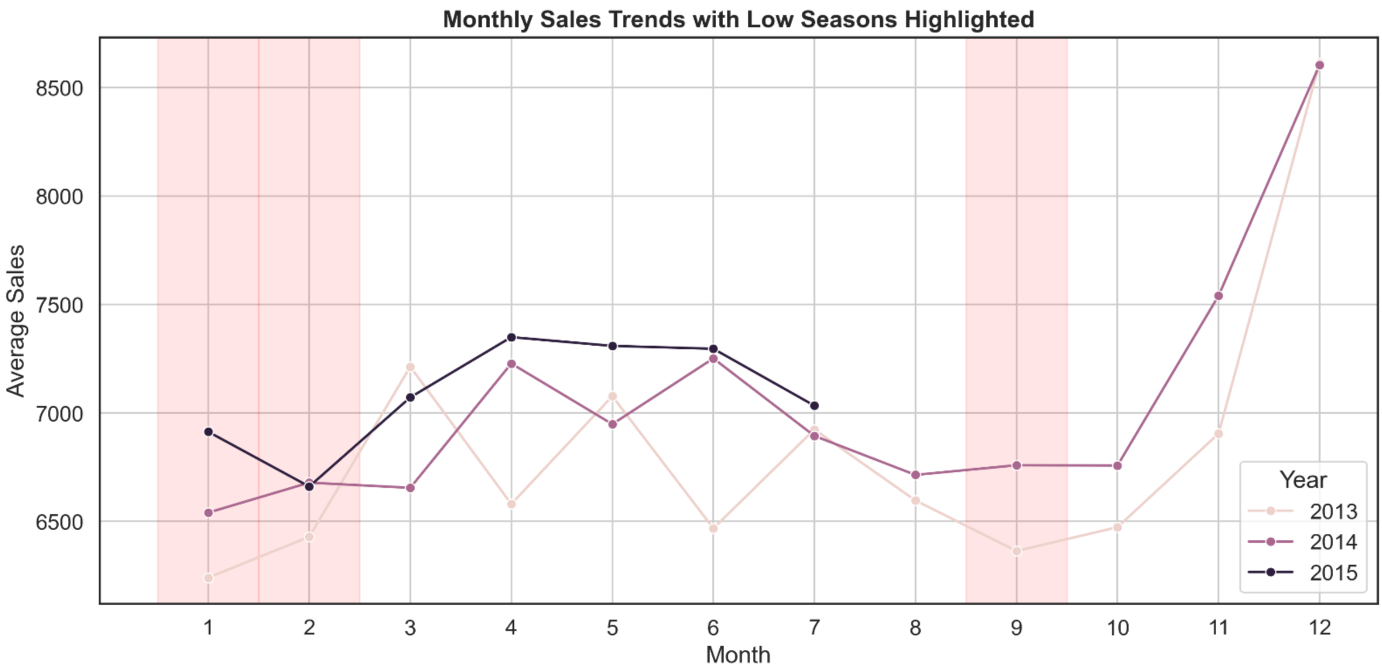


❖ Promo Effectiveness in Low-Performing Months:

- Identified Low Sales Months: January (1), February (2), and September (9), based on lowest average sales across all years.
- Sales Without Promotion: €5,868.42
- Sales With Promotion: €7,514.97
- Promo Impact: On average, promotions lifted sales by €1,646.55 in these low-performing periods.



❖ Low-Performing Months and Promotional Impact Analysis:
The analysis revealed that September, January, and February consistently recorded the lowest average sales figures, with September emerging as the lowest-performing month overall.



Promotions during low-performing months result in a 38.6% increase in average sales compared to non-promo days. This demonstrates a significant positive impact of promotional activities, particularly in traditionally low-sales periods.

Interpretation of Results and Recommendation

This analysis was structured around two core business challenges.

1. Store Performance Segmentation and Optimization
   Store-level KPIs such as average sales, customer footfall, basket size, promotional activity and competition proximity revealed significant disparities in performance across Rossmann's store network.
   - High Potential but Underperforming Stores:
     - 107 stores showed above median customer footfall but below median sales, suggesting inefficiencies due to either ineffective marketing, suboptimal product placement or service level gaps.
   - Low Sales & Low Promo Stores:
     - A group of 121 store had both low average sales and minimal promotional engagement. These stores are underutilized in terms of marketing strategy and show potential for target uplist thorough interventions.
   - Cluster Based Insights via KMeans Clustering:
     - Three performance cluster was identified
     - Cluster 0: Mid performing stores requiring strategic support and optimization.
     - Cluster 1: high sales and high footfall – top performers with consistent traffic.
     - Cluster 2: high basket value but lower footfall – efficient sales conversion.
   - Correlation Analysis:
     - Strong positive correlation (r = 0.83) between AvgCustomer and AvgSales confirms traffic's role in revenue.
     - Negative correlation (r = -0.48) between SelesPerCustomer and AvgCustomers suggest that higher Traffic may dilute individual spending efficiency.

- Promotion intensity (Promo2Active) had mixed effects, improving basket size but not always increasing total traffic or sales.

2. Promotion and Inventory Planning Using Time Series Trends
   Time series analysis was applied to uncover seasonal trends, promotional effectiveness and timing insights critical for inventory and campaign planning:

- Monthly and Weekly Sales Trends:
  - December consistently shoed peak sales, especially in week 51, aligned with holiday season demand.
  - January, February and September were identified as the lowest performing months indication seasonal slowdowns.
- Promo Effectiveness:
  - Average Monthly sales during promo periods: €8240.19 vs. non promo periods: €59441.95 – a significant €2298.24 uplift.
  - Even in low – performing months, promotions let to a 38.6% increase in sales.

- Heatmap Analysis (Months vs Day of Week):
  - Sundays in November recorded the highest sales, signalling strong weekend shopping behaviour during the festive season.
  - Saturday September had the lowest Sales, Highlighting opportunities for tactical interventions.
  - The highest sales area typically observed on Sunday and Mondays.

- Promo in low – season Months:
  - In January, February and September promotions still significantly boosted performance, reaffirming their strategic importance in mitigating seasonal demand dips.

Recommendations:

1. Enhancing Store Performance and Segmentation

   The segmentation analysis revealed a subset of stores with high customer traffic but low revenue per customer. These stores are not maximizing their potential and need focused intervention. To address this:

   Rossmann should implement strategies aimed at increasing the average basket size in high-footfall but underperforming stores. This includes introducing product bundles, improving store layout for better customer navigation, offering upselling suggestions, and training staff to assist and convert browsing into purchases more effectively. Additionally, gathering customer feedback can help uncover operational bottlenecks or service gaps that may be affecting customer satisfaction and spending.

   Cluster-based analysis identified Cluster 0 as consisting of stores with relatively balanced but modest performance. These stores may benefit from resource reallocation such as better inventory distribution, optimized staffing, and managerial oversight based on their performance cluster identity.

2. Strategic Promotion and Inventory Planning

   The time-series analysis highlighted significant seasonal trends. Months like November and December consistently recorded higher sales, while January, February, and September showed the lowest average sales. To manage these seasonal fluctuations effectively:

Rossmann should schedule major promotional events during both peak and low-performing months. While high-sales months like November–December naturally drive traffic, offering compelling discounts during traditionally low-sales months (such as January, February, and September) can help maintain consistent revenue. Special campaigns like flash sales or weekend offers can be introduced to stimulate demand during low-traffic days, especially on Saturdays which saw the lowest performance even it is weekend.

3. Improving Inventory Efficiency

   Forecasting inventory based on historical sales data can help avoid issues like overstock or stockouts. Rossmann should use monthly and weekly trend data to predict demand more accurately, especially around holidays and promotional periods. Ensuring sufficient inventory during peak weeks (e.g., Week 51) can prevent revenue loss due to stockouts, while maintaining leaner inventories in slower months reduces holding costs. Implementing demand-driven replenishment strategies will improve stock movement and customer satisfaction.

4. Measuring and Enhancing Promotion Effectiveness

   Promotions consistently showed a positive impact on sales, with promotional periods yielding an average increase of €2,298.24 in monthly sales. Rossmann should continuously monitor sales uplift during promotional periods at the store, region, and month level to evaluate effectiveness. Different types of promotions (e.g., discounts, loyalty rewards, bundles) should be tested and compared to identify which formats drive the most value.

5. Aligning Promotions with Weekly Sales Patterns

   Daily trends also revealed actionable insights. Sundays recorded the highest average sales, while Saturdays in September were notably underperforming. Rossmann can capitalize on high-traffic days like Sundays by launching key campaigns or new product offerings on these days. Meanwhile, underperforming days should be targeted with special events such as weekend discounts, family shopping days, or local customer engagement drives.