

Sales Department

❖ Project Summary:

This project aims to predict future sales for the Sales Department based on historical data, enabling better decision-making, inventory planning, and revenue forecasting. The project is divided into several key tasks:

1. Understand the Problem Statement and Business Case:

The first step involves gaining a deep understanding of the sales challenges faced by the department, identifying business goals, and defining how predictive sales insights can enhance decision-making.

2. Data Exploration, Visualization, and Analysis:

Historical sales data will be explored for trends, seasonality, and patterns. Data visualization techniques will help to identify key insights and outliers, and initial analysis will set the foundation for model development.

3. Facebook Prophet for Predictive Modeling:

We will use the Facebook Prophet, a robust forecasting tool from Facebook's data science team, will be used to predict future sales trends based on historical data. This tool is well-suited for capturing seasonality, holidays, and demand fluctuations in sales forecasting.

4. Model Training:

The final step will involve training the predictive model using Facebook Prophet on historical sales data. The model will be fine-tuned to achieve high accuracy in forecasting future sales, ensuring its reliability for real-world business use.

This structured approach will provide valuable sales forecasts, helping to optimize resources and improve sales strategies.

❖ Problem and Business Case:

To stay competitive and drive growth, companies must harness the power of AI and Machine Learning to build predictive models that accurately forecast future sales. These models not only analyze historical sales data but also factor in seasonality, demand fluctuations, holidays, promotions, and competitive dynamics. By doing so, they enable businesses to make informed decisions and stay ahead of market trends.

In this project, as a data scientist within the Sales Department, you've been provided with data from 1,115 stores. The primary objective is to predict future daily sales based on various store-related features. By developing a robust sales forecasting model, the company aims to improve inventory

management, optimize marketing strategies, and allocate resources more effectively. Ultimately, this will help the business meet customer demand efficiently, reduce costs, and enhance overall profitability.

This statistical summary gives a snapshot of the sales and customer distribution for the dataset you're working with. Here's a clear breakdown:

Key Insights:

1. Sales Distribution:

- The average sales per day is 5,774 Euros, with significant variation ranging from 0 to a maximum of 41,551 Euros.
- 50% of the sales are below 5,744 Euros, while 25% of the stores generate less than 3,727 Euros in sales per day.
- The high standard deviation (3,850 Euros) suggests a wide variation in daily sales across stores.

2. Customer Visits:

- Stores see an average of 633 customers per day, with a minimum of 0 and a maximum of 7,388.
- 50% of the stores serve fewer than 609 customers daily, indicating that half of the stores have low to moderate foot traffic.
- The standard deviation of 464 customers shows a significant difference in customer traffic across stores.

3. Store Operations:

- Stores are open on average 83% of the time, with some days where stores are closed (17%), potentially explaining the zero sales and customer counts on certain days.
- About 38% of the time, the sales period coincides with school holidays, which may impact foot traffic and sales.

4. Promotional Activity:

- On average, stores run promotions 38% of the time, with a wide variability (some stores run promotions almost continuously while others rarely do).
- 50% of stores do not run promotions on any given day, highlighting the importance of promotional activity in driving customer traffic and sales.

5. Day of the Week:

- The dataset includes all days of the week, with the average day of operation being closer to the middle of the week (Day 4, Thursday).
- Sales and customer traffic likely fluctuate across different days, which could affect future predictive models, given that some days might be significantly busier than others.

These insights are critical for understanding store performance, promotional impact, and how external factors like school holidays influence daily sales and customer activity.

These figures provide valuable insights into the range of sales performance and customer visits across the 1,115 stores. The presence of days with zero sales or customers suggests either store closures or missing data points, which might require further investigation during data preprocessing. The large variance between the minimum and maximum values points to potential seasonality or promotional effects that could be incorporated into your predictive model.

=====

Key Insights:

1. Store Count and Competition Distance:

- The dataset contains 1,115 stores, with competition distance ranging from 20 meters to 75,860 meters.
- The average competition distance is approximately 5,405 meters, indicating that most stores face nearby competition.

2. Competition Open Since:

- Data on when competition started is available for 761 stores, with competition opening as early as 1900 and as recent as 2015.
- The average competition began around July 2008, suggesting that many stores have had nearby competition for over a decade.

3. Promo2 (Continuous Promotion):

- About 51% of stores participate in the Promo2 promotional program.
- For stores participating in Promo2, the program typically started around week 24 of the year, with an average start year of 2012.
- Promo2 has been active for several years, with the most recent start year being 2015.

4. Promo2 Since:

- Promo2 start weeks range from week 1 to week 50, showing the program's staggered introduction across stores.
- The majority of Promo2 stores started the program around 2012, with most between 2011 and 2013.

5. Variability in Data:

- There is significant variability in the competition distance, with some stores facing very close competition, while others are in more isolated locations.
- The wide range in competition opening years, from 1900 to 2015, shows that some stores have long-standing competition, while others face more recent challenges.

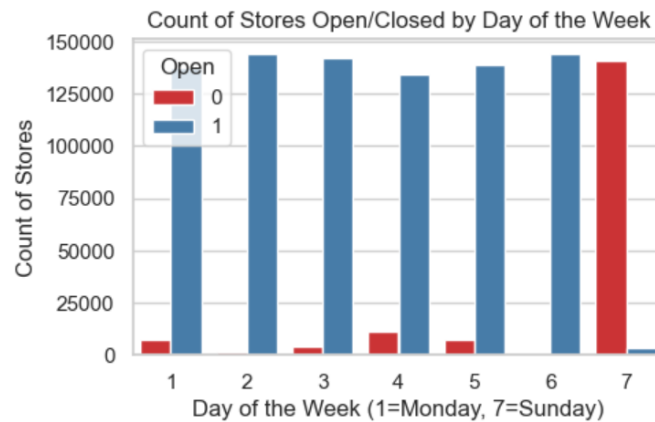
These points summarize the distribution of competition and promotional factors that could influence sales performance. Understanding these variables will be crucial for building a predictive model that accounts for competition and promotional effects.

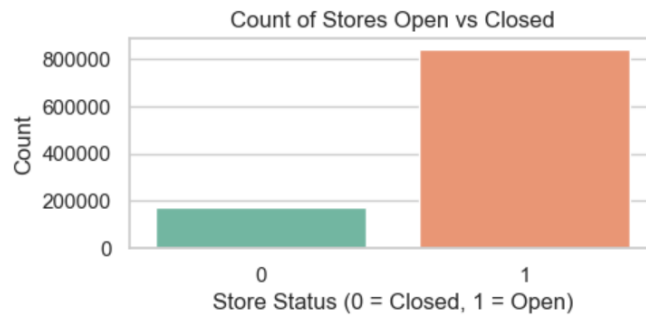
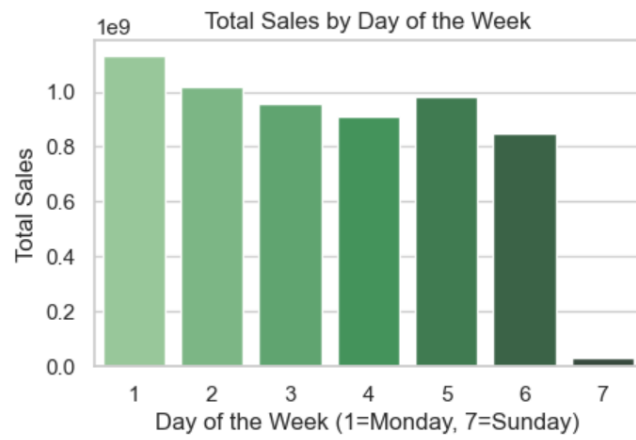
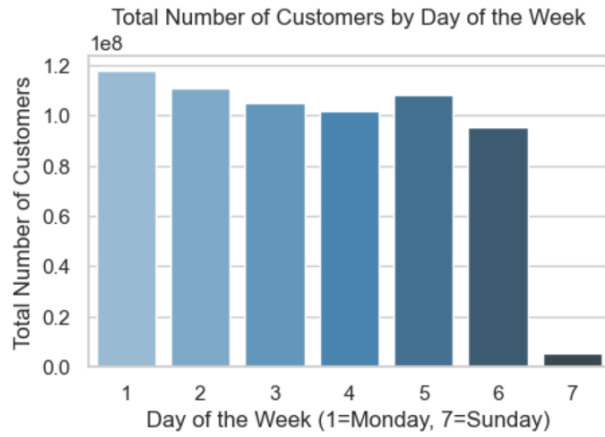
```
sales_train_df.head(5) # This file consists of all the transections
```

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

❖ Impact of Store Closures on Sales Data

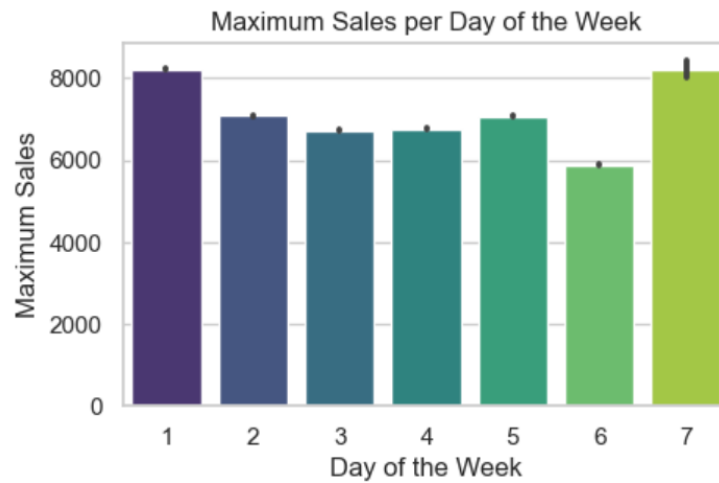
This project analyzes store closures and their effect on sales and customer data. Most stores are closed on Sundays, leading to no customer activity or sales on that day. The focus will be on understanding how this impacts overall sales trends and performance during operational days.





❖ Impact of Sunday Store Openings on Sales

This project analyzes the effect of opening stores on Sundays. Data shows that Sunday could become the highest day for sales and customer traffic, highlighting its potential as a key day for business growth.

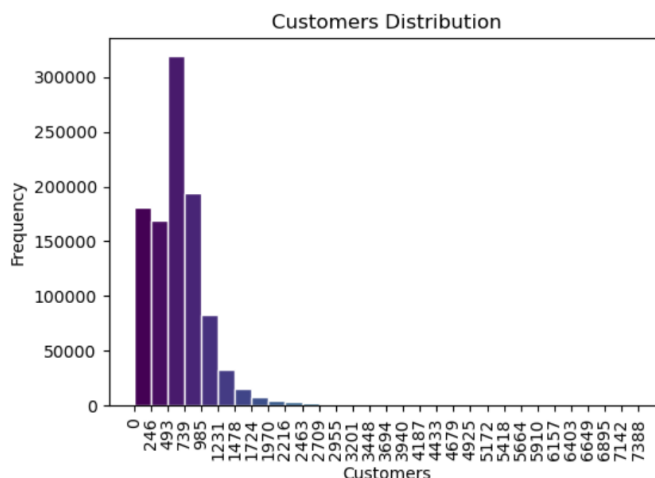


❖ Filtering Sales Data for Open Stores

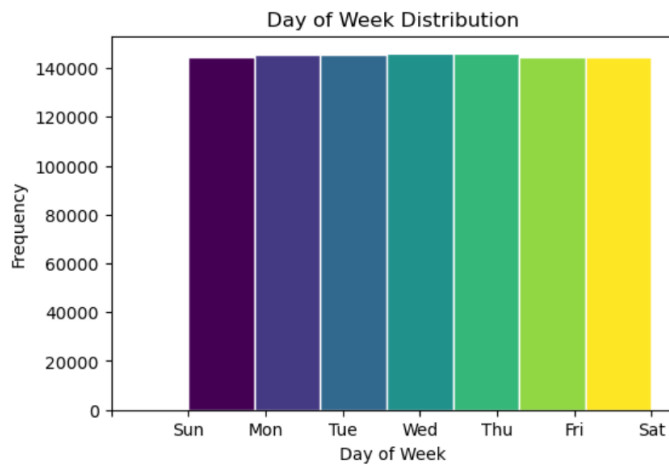
This project focuses on filtering the sales data to include only records from days when stores are open. By utilizing the feature indicating store status (open or closed), we aim to analyze sales performance and customer traffic specifically during operational hours, providing a clearer view of business activity.

Summary of Key Points:

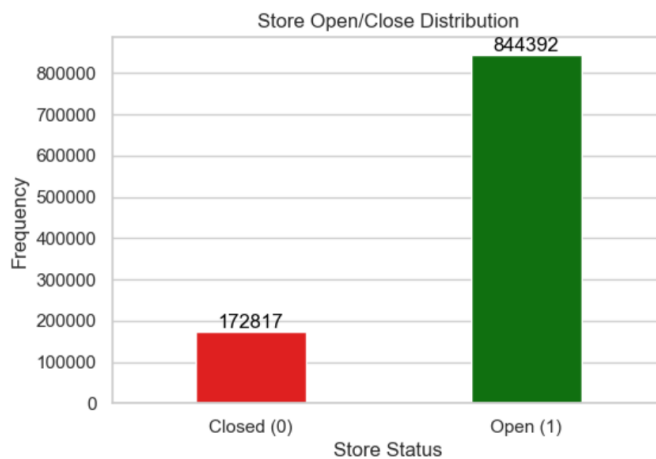
- **Customer Traffic:** The average number of customers per day is around 600, with a maximum observed value of 4,500 (noting that the extreme outlier of 7,388 customers is not visible in the general data distribution).



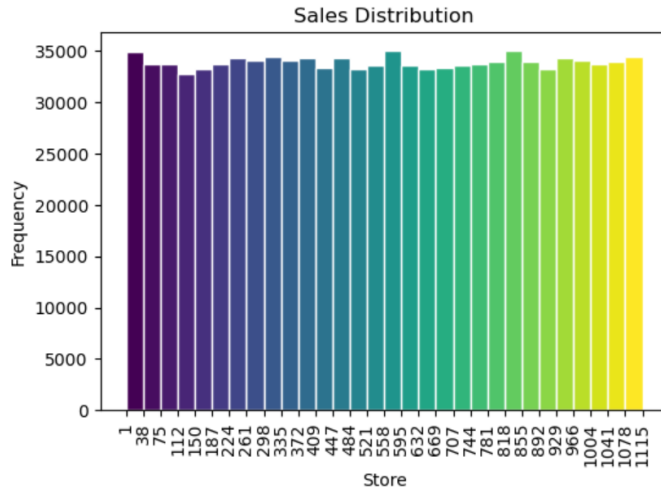
- **Daily Distribution:** Data is evenly distributed across the days of the week, with approximately 150,000 observations for each day, totaling around 1.1 million observations.



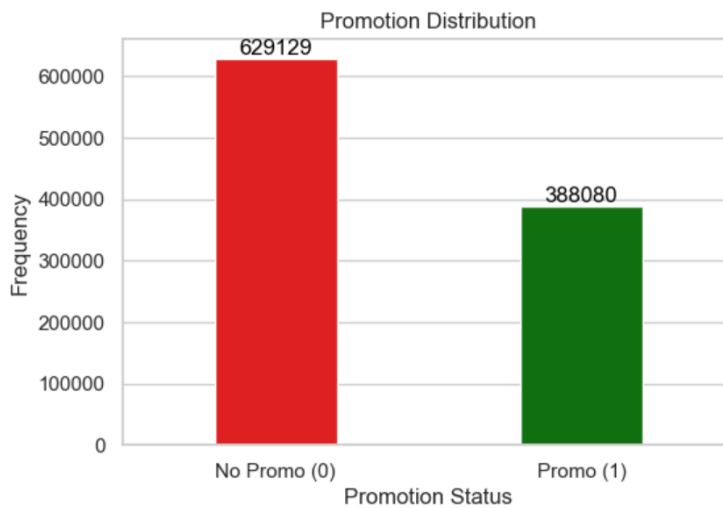
- **Store Operations:** Stores are operational about 80% of the time, indicating consistent availability for customers.



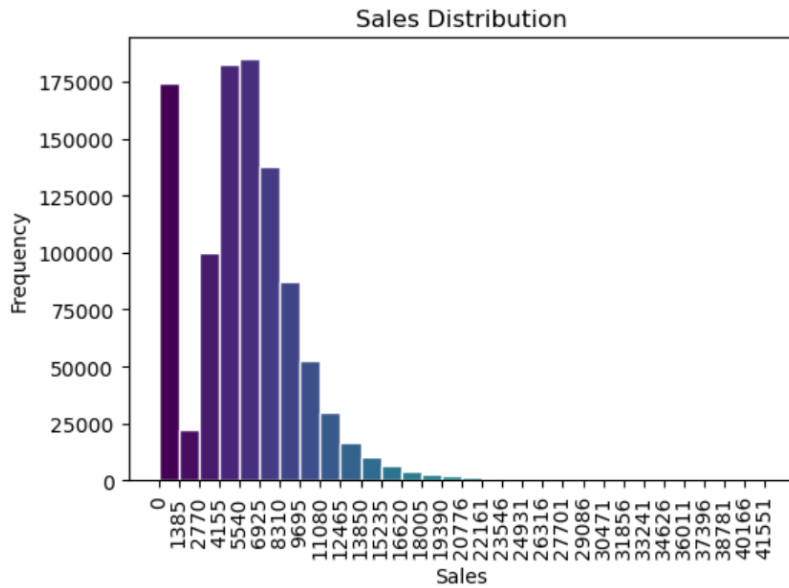
- **Data Distribution:** The dataset shows no bias, as sales and customer data are equally distributed among all stores.



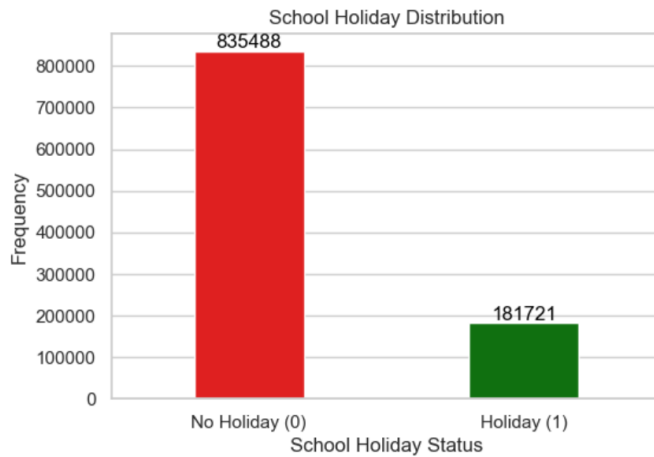
- **Promotional Activity:** Promo #1 was active approximately 40% of the time, suggesting that promotions are a significant factor influencing sales.



- **Sales Performance:** Average daily sales range between 5,000 to 6,000 Euros, reflecting a healthy revenue stream for the stores.



- **Impact of School Holidays:** School holidays account for about 18% of the time, which may influence customer footfall and sales during these periods.



- These insights provide a solid foundation for understanding store performance dynamics and the factors influencing sales and customer traffic.

=====

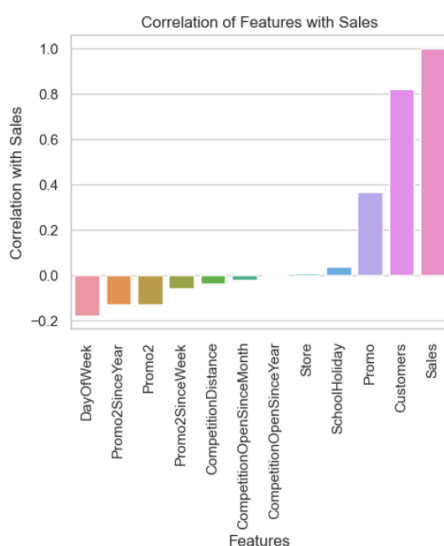
❖ Open Store observation

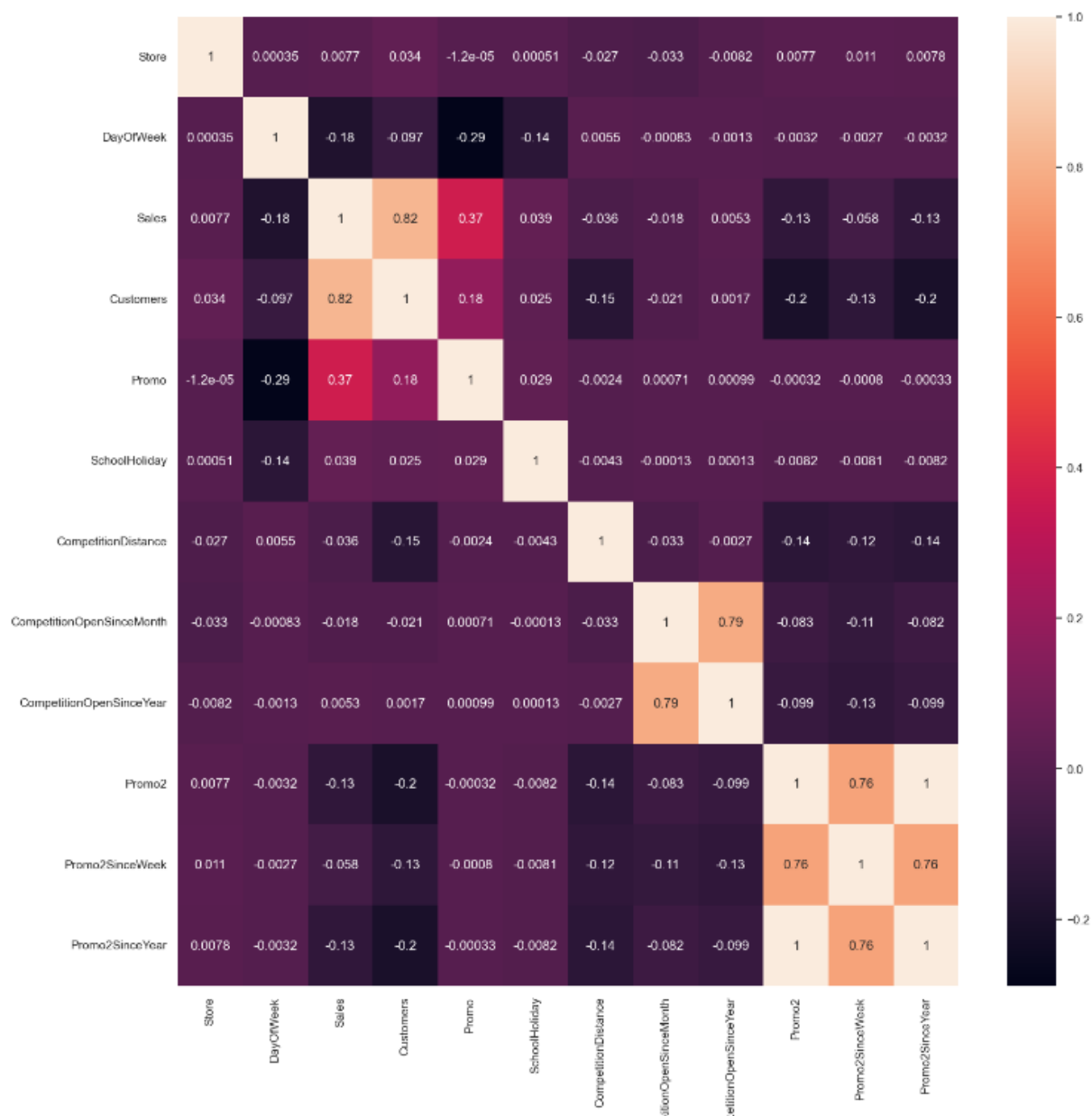
- We focus only on the stores that are open during the observation period. When stores are closed, both customer counts and sales figures are zero, skewing the data. By filtering out closed stores, we ensure our analysis reflects true customer engagement and the effectiveness of promotional activities.
- After filtering out closed stores, the average sales increased to 6,955 Euros, and the average number of customers rose to 762. This improvement indicates that operational stores are performing better, highlighting the positive impact of open hours on sales and customer engagement. These figures suggest that maintaining store availability is crucial for maximizing revenue and attracting more customers.
- Average sales = 6955 Euros.
- Average number of customers = 762 (went up)

❖ The correlation analysis

The correlation analysis reveals that both customers and promo are positively correlated with sales, indicating that increased foot traffic and promotions significantly boost revenue. However, Promo2 does not show any meaningful impact on sales, suggesting it may not be as effective or needs further optimization to drive results.

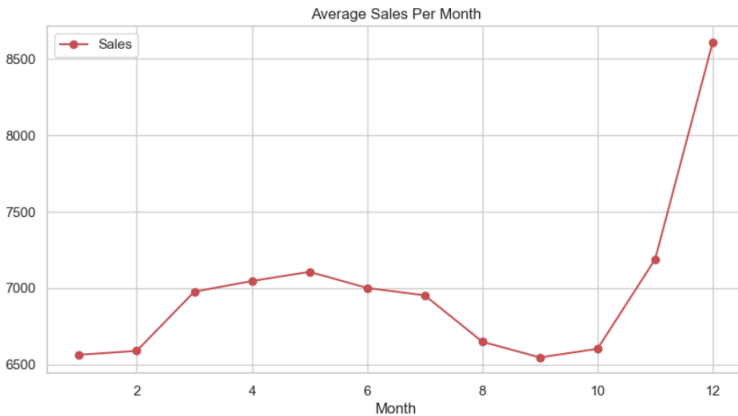
The negative correlation between DayOfWeek and Sales (correlation = -0.178736) indicates that sales tend to decrease on certain days of the week. A negative correlation means that as the day of the week increases (e.g., Monday = 1, Sunday = 7), sales decrease.



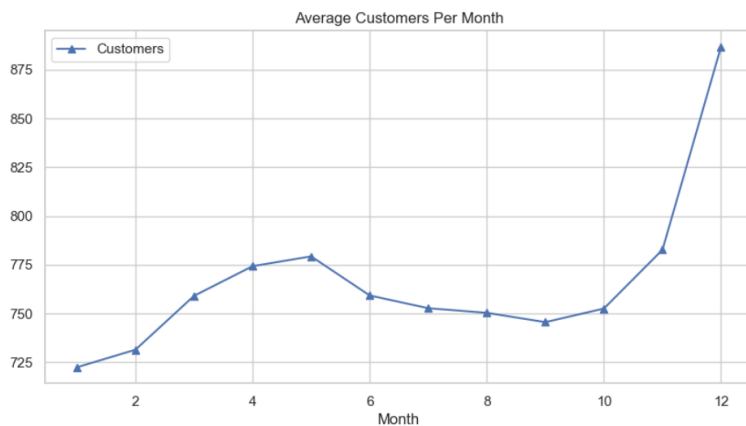


Monthly Sales Trends:

We observe clear seasonal trends. Both average sales and the number of customers show a significant peak around the Christmas timeframe, indicating a surge in holiday shopping. This seasonal spike suggests that the festive season drives higher store traffic and boosts sales during the final months of the year.

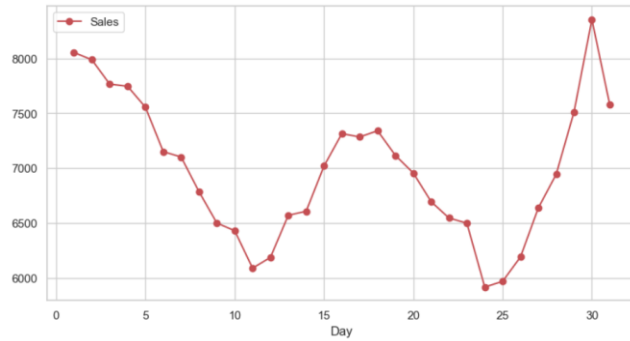


Monthly Customer Trends:

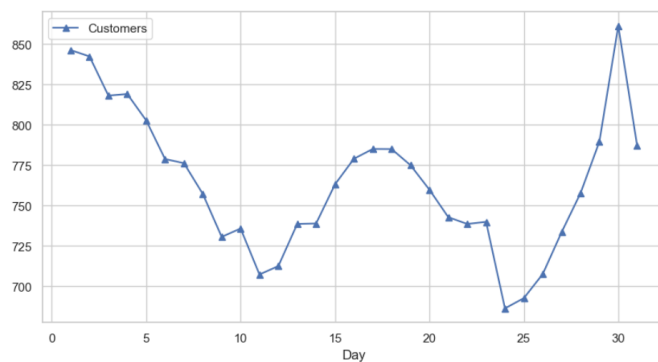


Sales Trends by Day of the Month

Examining the data on a day-by-day basis reveals that the minimum number of customers typically occurs around the 24th of the month. In contrast, the highest customer traffic and sales are observed around the 30th and 1st of the month. This trend could be related to salary cycles or end-of-month shopping habits, where customers tend to shop more as the month concludes and a new one begins.



Customer Trends by Day of the Month



Average Sales Per Day of the Week

Analyzing the average sales by day of the week reveals clear patterns:

Sunday (Day 7): Highest average sales at 8,224 Euros.

Monday (Day 1): Strong performance with 8,216 Euros.

Tuesday to Thursday (Days 2-4): Moderate sales, ranging from 6,728 to 7,088 Euros.

Friday (Day 5): Sales rise again, averaging 7,072 Euros.

Saturday (Day 6): The lowest average sales at 5,874 Euros.

This indicates that weekends, especially Sundays, tend to be the most lucrative, while Saturdays experience a surprising dip in sales.



❖ Average Customers Per Day of the Week

Looking at the average number of customers by day of the week:

Sunday (Day 7): Highest foot traffic with 1,441 customers.

Monday (Day 1): Strong start to the week with 855 customers.

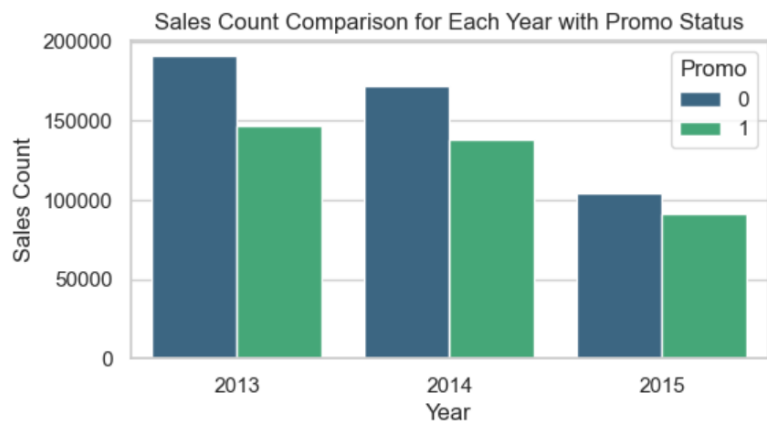
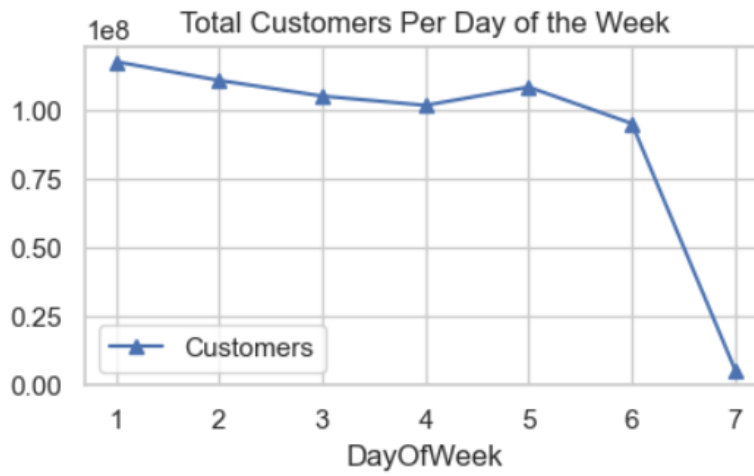
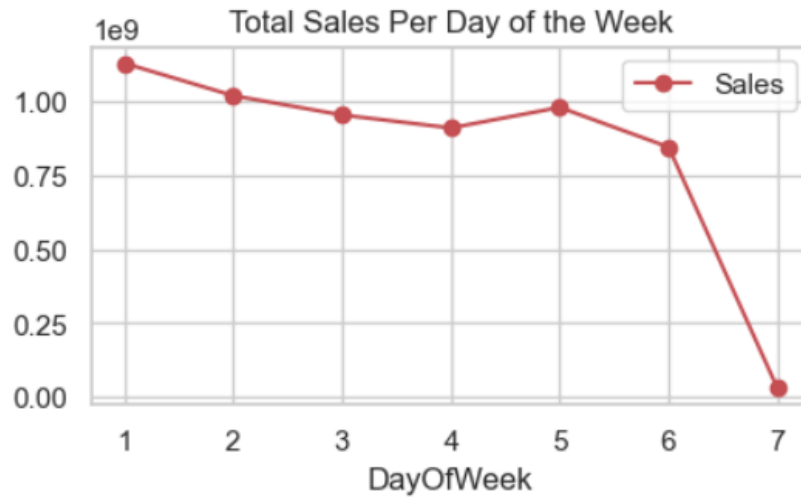
Tuesday to Thursday (Days 2-4): Customer counts range between 740 to 770.

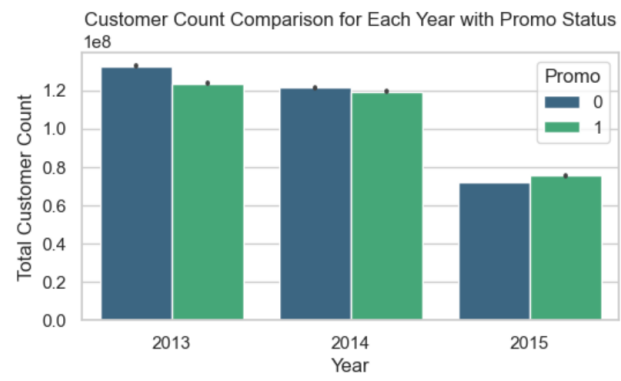
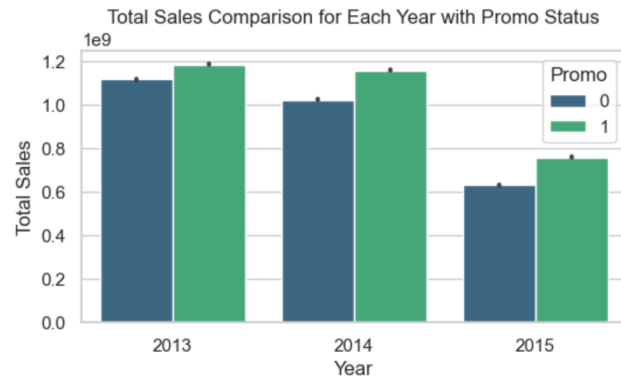
Friday (Day 5): A slight increase to 782 customers.

Saturday (Day 6): Lowest foot traffic at 660 customers.

This shows Sundays attract the most customers, while Saturdays see the least, indicating a significant variance in customer behavior throughout the week.







60-Day Sales Forecasting for a Specific Store

This project involves forecasting sales for the next 60 days using two key parameters: **sales** and **date**. By analyzing the historical sales data against time, we aim to project future sales trends for a specific store. This focused approach will provide a clear and concise forecast to aid in business planning and decision-making.

