CHRIST
UNIVERSITY
BENGALURU, INDIA

Declared as Deemed to be University under Section 3 of UGC Act 1956

CASE STUDY REPORT
ON
"Coffee Shop- Sales Data Analysis"

Submitted By

DHILLIPKUMAR M (2348026)

I Trimester

MSc DATA SCIENCE

Department of Statistics and Data Science

CONTENTS

Abstract:

This exploratory data analysis (EDA) report presents a comprehensive analysis of coffee sales data in a local coffee shop over a specific period. The dataset contains records of daily transactions, including information on date, time, coffee item sold, quantity, and revenue generated.

The primary objective of this analysis is to gain insights into the coffee shop's sales performance, identify patterns in customer behaviour, and make data-driven recommendations for optimizing revenue and enhancing customer experience.

The EDA begins with data cleaning to ensure the dataset's integrity and consistency. Subsequently, the report delves into various aspects of the sales data, including general statistics such as total revenue and average daily revenue. The time analysis section highlights the busiest hours of the coffee shop, helping to understand peak business hours.

The popularity of coffee items is explored, revealing the top-selling items based on quantity sold and their revenue contribution. Additionally, the analysis examines revenue variations between weekends and weekdays to identify potential patterns in customer preferences.

Furthermore, customer insights are explored, including the identification of returning customers and their impact on overall revenue. Customer segmentation techniques are employed to group customers based on their purchasing behaviour.

Finally, correlation analysis is conducted to explore relationships between different variables, such as quantity sold, revenue, and time of the day.

This EDA report concludes with actionable recommendations to improve sales performance. These recommendations include promoting the top-selling items, offering targeted marketing strategies for returning customers, and optimizing pricing and promotions during peak and off-peak hours.

Overall, this exploratory data analysis provides valuable insights into the coffee shop's sales data, paving the way for data-driven decision-making to boost revenue and customer satisfaction.

Introduction:

In this report, we present a comprehensive analysis of the transactions and sales performance of a local coffee shop over a specific period. The dataset contains valuable information, including date, time, coffee items sold, quantity, and revenue generated, providing us with a deeper understanding of the coffee shop's business operations and customer behaviour.

The coffee industry has experienced significant growth in recent years, driven by the rising popularity of specialty coffee and a growing coffee culture among consumers. As competition in the market intensifies, coffee shop owners and managers seek data-driven insights to stay competitive, enhance customer experience, and optimize revenue.

The primary aim of this exploratory data analysis is to extract meaningful insights from the coffee shop's sales data, uncover hidden patterns, and identify potential areas of improvement. By analysing the data, we hope to gain valuable information on various aspects of the coffee shop's performance, such as peak business hours, popular coffee items, revenue variations between weekdays and weekends, and seasonal sales patterns.

This report is structured to take you through a step-by-step analysis of the coffee shop sales data. We will begin by cleaning the dataset to ensure data quality and consistency. Subsequently, we will delve into various exploratory analyses, utilizing various data visualization techniques to present the findings in a clear and understandable manner.

By the end of this analysis, we aim to provide actionable recommendations to the coffee shop's management, enabling them to make data-driven decisions to enhance sales performance, customer satisfaction, and overall business success.

Objectives:

The primary objective of this exploratory data analysis (EDA) report on coffee shop sales data is to gain valuable insights into the coffee shop's business performance and customer behaviour. Through a comprehensive analysis of the dataset, we aim to achieve the following objectives:

Sales Performance Analysis: Understand the coffee shop's overall sales performance by examining total revenue, average daily revenue, and revenue trends over the specified period. Identify peak sales hours and periods to optimize staffing and resource allocation.

Top-Selling Items: Identify the most popular coffee items based on quantity sold and revenue generated. Determine the contribution of each item to the overall revenue, allowing the coffee shop to focus on promoting and optimizing its best-selling products.

Revenue Variations: Analyse revenue variations between weekdays and weekends to identify any significant differences in customer behaviour and preferences. This insight will enable the coffee shop to implement targeted marketing strategies and promotions.

Recommendations: Based on the analysis and insights gained, provide actionable recommendations to the coffee shop management to enhance sales performance, customer satisfaction, and overall business success.

Data Description:

Coffee shop sales report.xlsx

Data source: Coffee shop sales dataset, Kaggle.

https://www.kaggle.com/datasets/ylchang/coffee-shop-sample-data-1113

The main dataset used for this analysis contains sales data for the coffee shop for the month of April 2019, typically comprising daily records. The dataset includes the following key variables:

- Date/Time: The timestamp of each sale transaction.
- transaction id: Unique number for all transactions.
- Product Category: The category of the product sold (e.g., coffee beverages, tea, pastries, merchandise).
- Sales outlet ID: ID number assigned for each outlet.
- Staff ID: Identification number for each employee
- Customer ID: Unique Identification number for each customer.
- Product ID: Unique number for each product.
- Product Name
- Quantity: The number of items sold in each transaction.
- Unit Price: The unit price of the item(s) sold.
- Total Sales: The total sales value of each transaction.

Additional datasets such as the following were also used: -

- Product- Having details of all products sold.
- Pastry inventory- details about the quantity sold and quantity of wastage.
- Sales targets- the minimum sales target for the outlets for minimum profit.
- Generations- categorizing all the customers by age (millennials, Gen Z, etc.)
- Customer- contact details of each customer.
- Staff- Employee details.

Methodology:

The analysis followed the following steps:

1. Data Pre-processing:

- Import and clean the raw data, handling unwanted data values if any.
- Transform the data to ensure consistency and uniformity in format.
- Perform data analysis to gain initial insights into the data distribution and characteristics, to be able to derive data.

2. Sales Performance Analysis:

- Calculate overall sales metrics, such as total revenue, average transaction value, and sales profit, etc.
- Analyse sales trends over time to identify most sales during the week and month.

3. Product Category Analysis:

- Group sales data by product category and calculate the proportion of sales contributed by each category.
- Compare sales performance across different product categories using visualizations.

4. External Factors Impact Analysis:

- Explore the relationship between external factors (if available) and sales performance.
- Conduct hypothesis testing to determine the significance of external factors on sales.

5. Data Visualization:

- Create visualizations (e.g., line plots, bar charts, heatmaps) to present the findings effectively.
- Use charts and graphs to illustrate sales trends, product performance, and any insights obtained.

6. Recommendations:

- Based on the analysis results, provide data-driven recommendations to optimize sales, enhance customer experience, and improve business performance.

7. Documentation:

- Document the entire methodology, data preprocessing steps, and analysis techniques used in the EDA process.
- Summarize the key findings and insights in the EDA report, ensuring clear communication of the results.

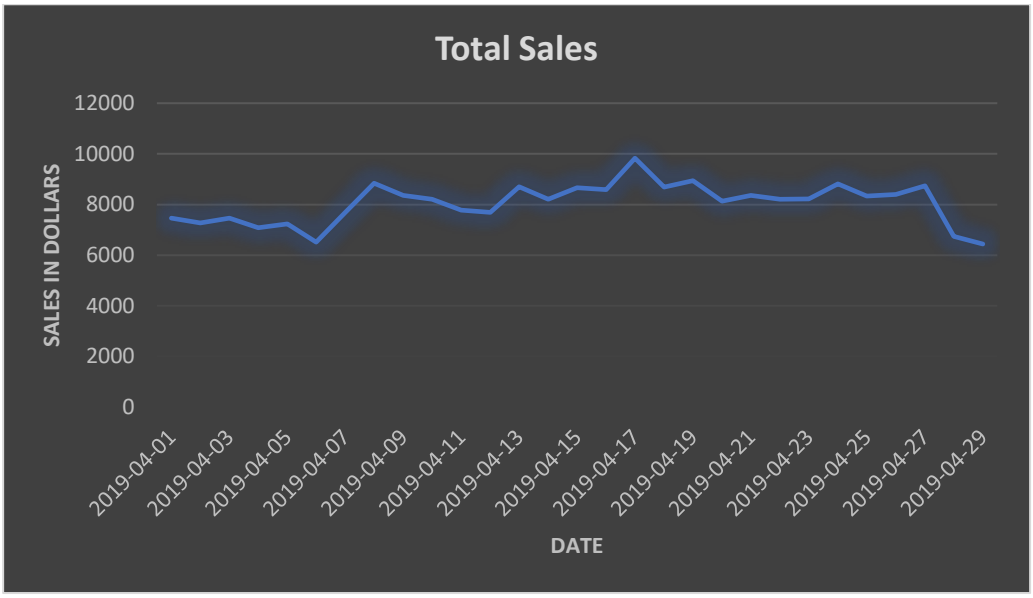
8. Customer Insights:

- Identify and group returning customers based on their purchase frequency and loyalty.
- Apply clustering algorithms, such as K-means, to segment customers with similar buying behaviours.
- Visualize the customer segments using scatter plots or parallel coordinate plots.

<u>Data analysis and representation:</u>

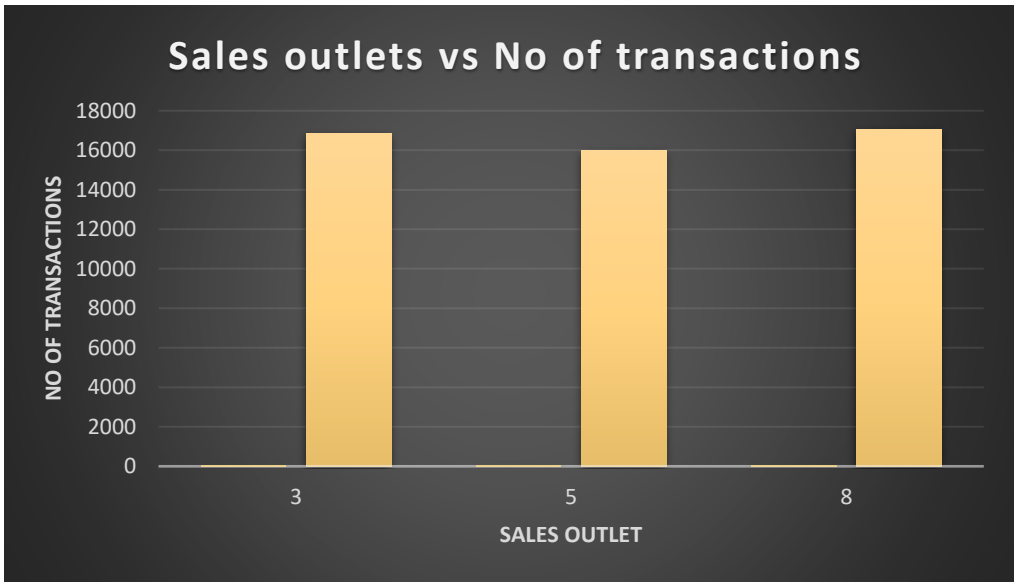
1.Daily sales trend for the month April

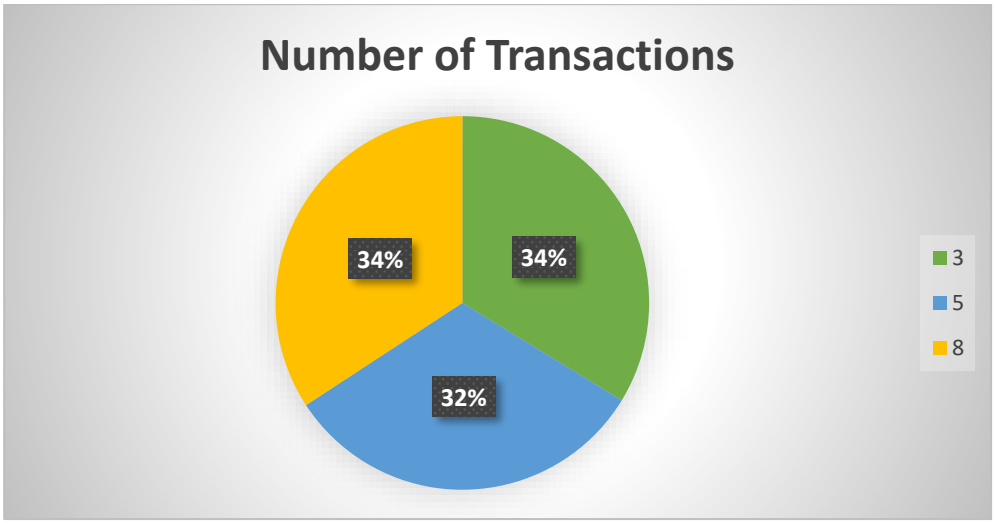
- The highest sales were on 17th and the lowest on 29th.
- The average sales per day is $ 8056.4.



2.Total number of transactions with respect to sales outlets

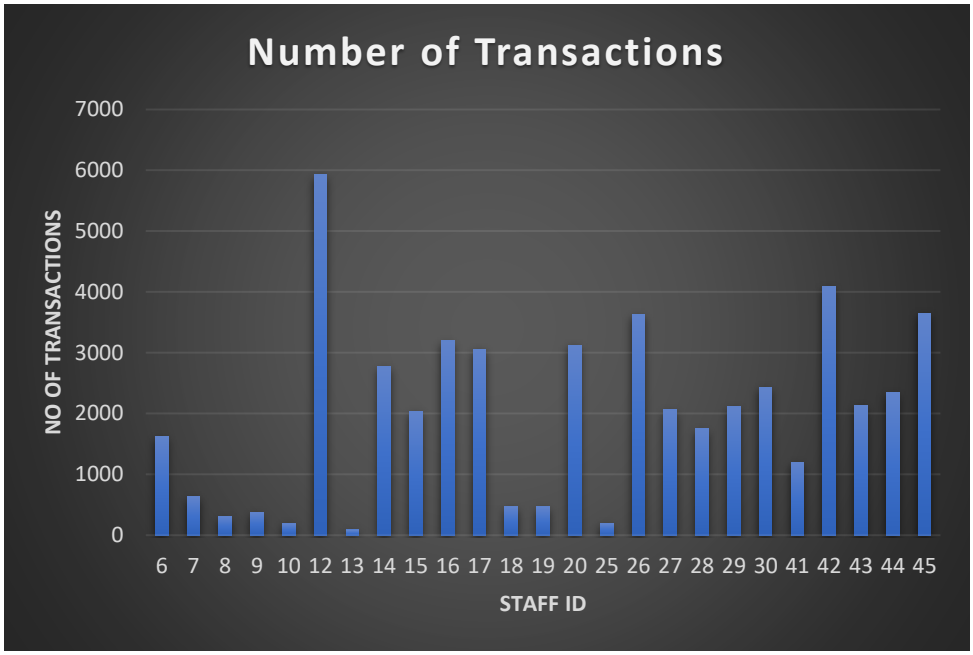
- The sales outlets have done similar no of transactions.





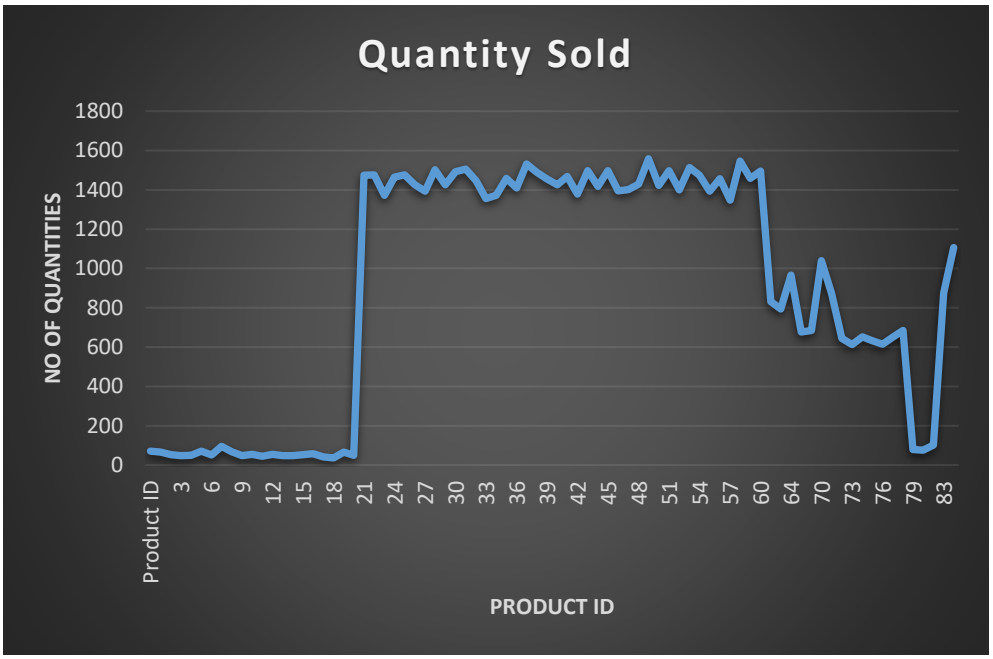
3. Number of transactions done by each staff.

- Employee with staff ID -12 is performing well.
- Employee with staff ID-13 is under performing.
- Average number of transactions done by a staff is 1995.



4.Total quantity of products sold with respective to its ID

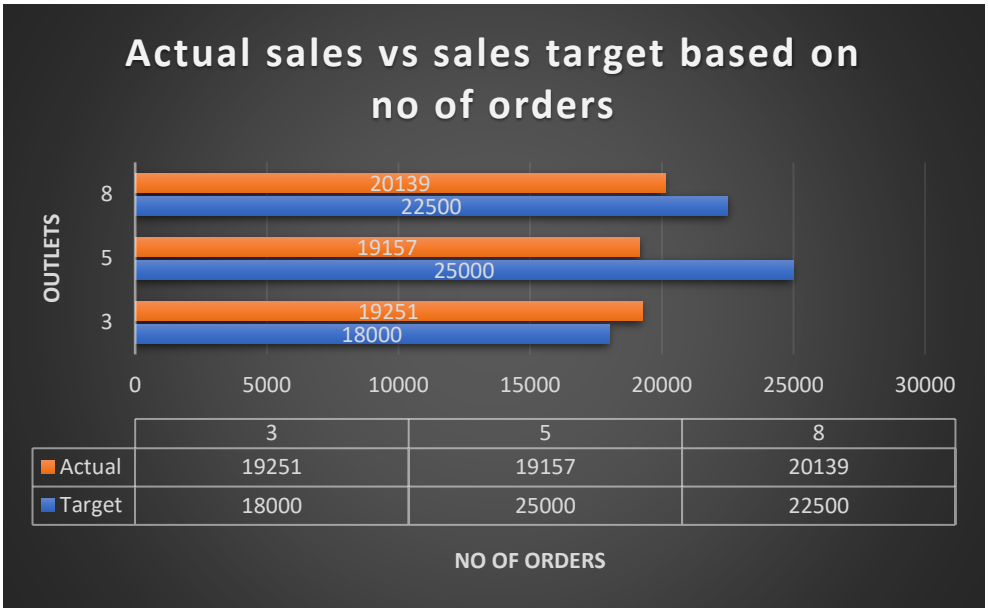
- Product Earl Grey Rg is bought most by the customers.
- Product Dark chocolate is least sold.



5. Sales target

- Outlet 3 performed well more than the target set
- Performance level:

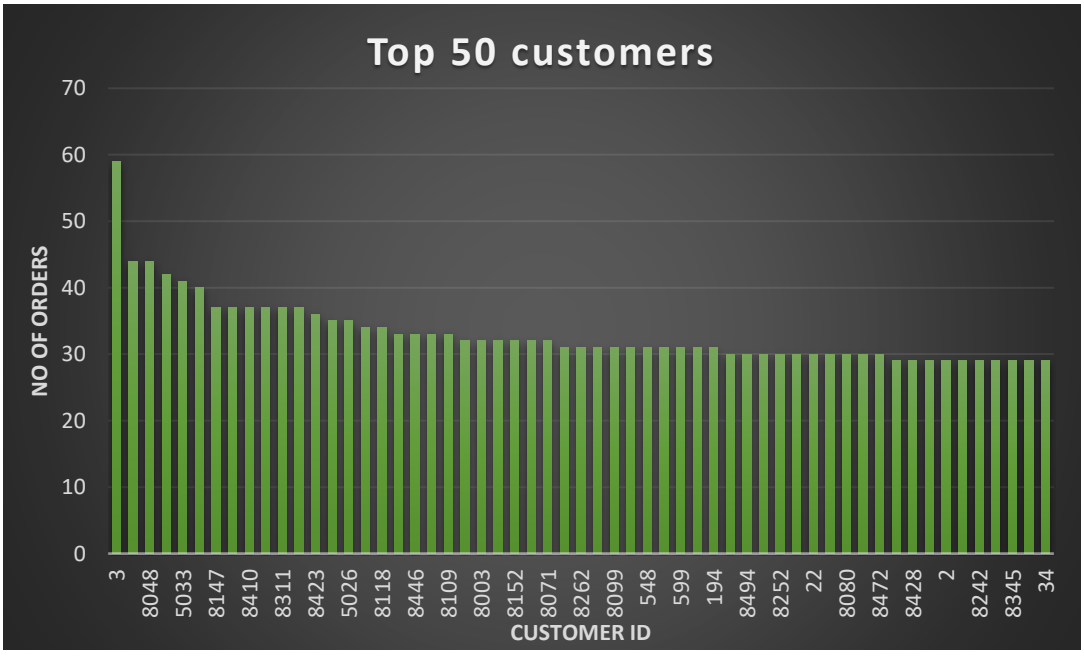
  Outlet 3 > Outlet 8 >Outlet 5



	3	5	8
Actual	19251	19157	20139
Target	18000	25000	22500

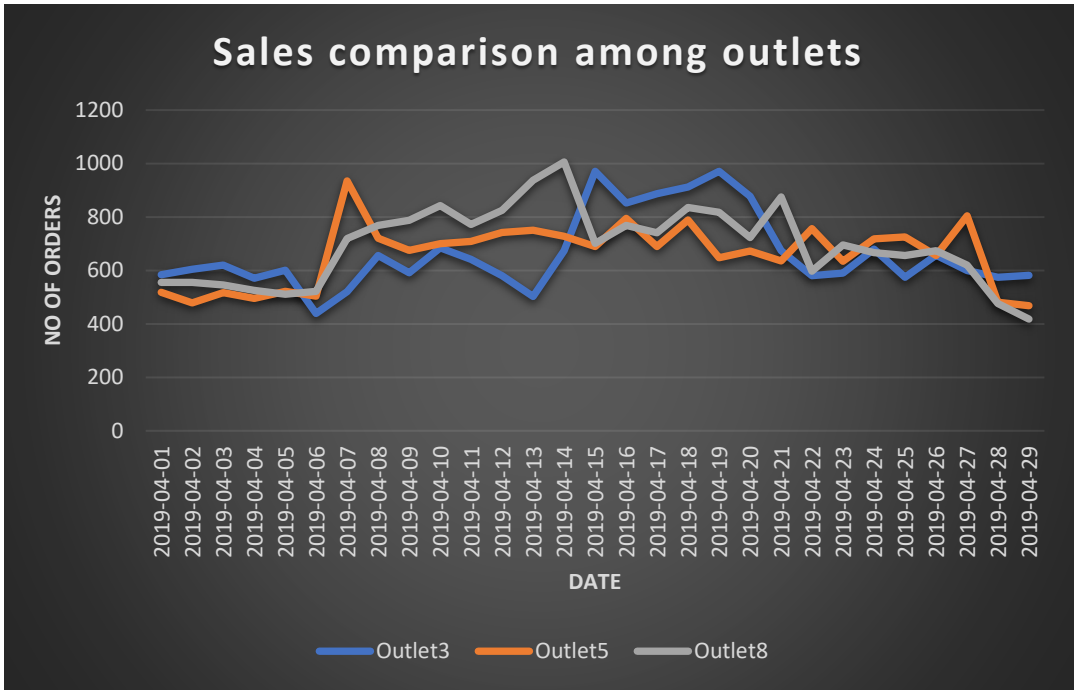
6. Customer segmentation

- Top 50 customers based on no of orders are identified.
- Customer ID =3 is the most valuable customer



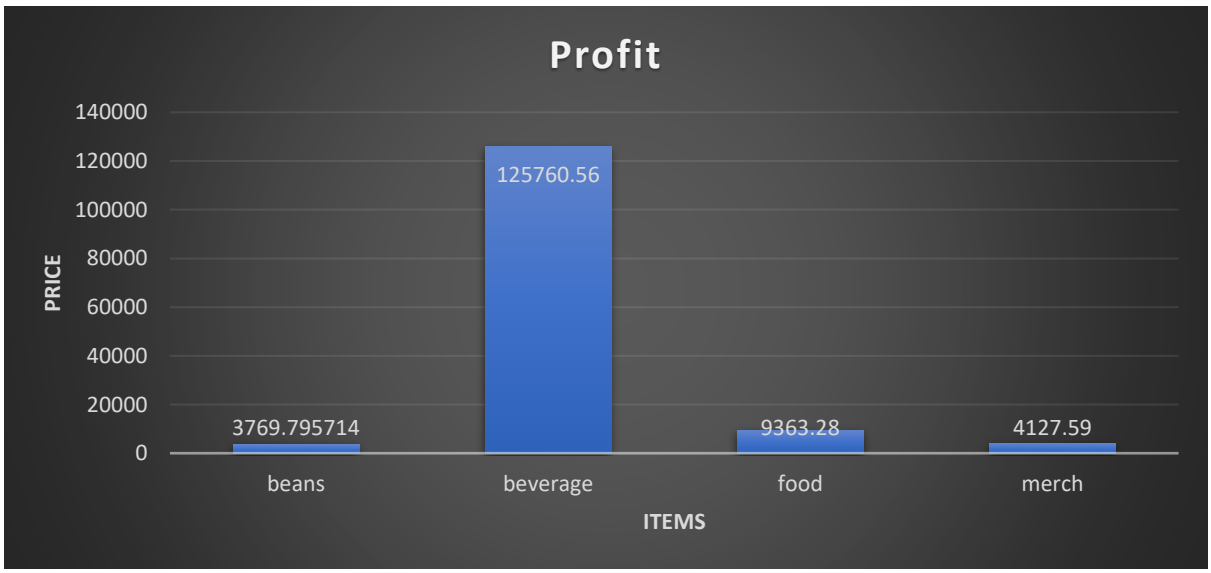
7. Sales comparison among outlets

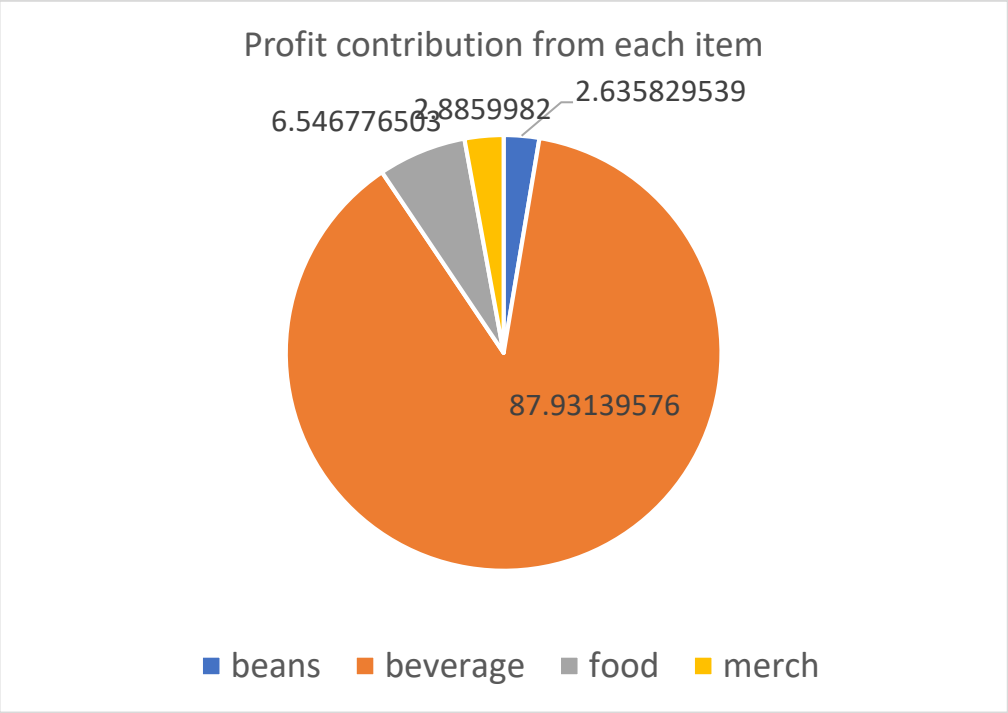
- The no of orders is high in the middle of the month from all the outlets.

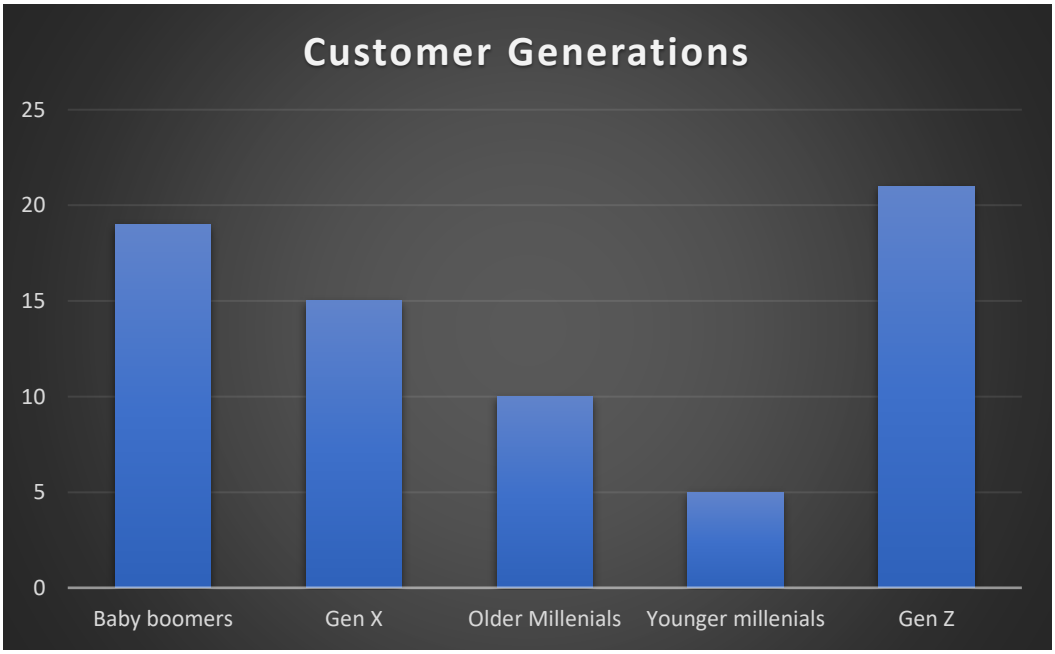


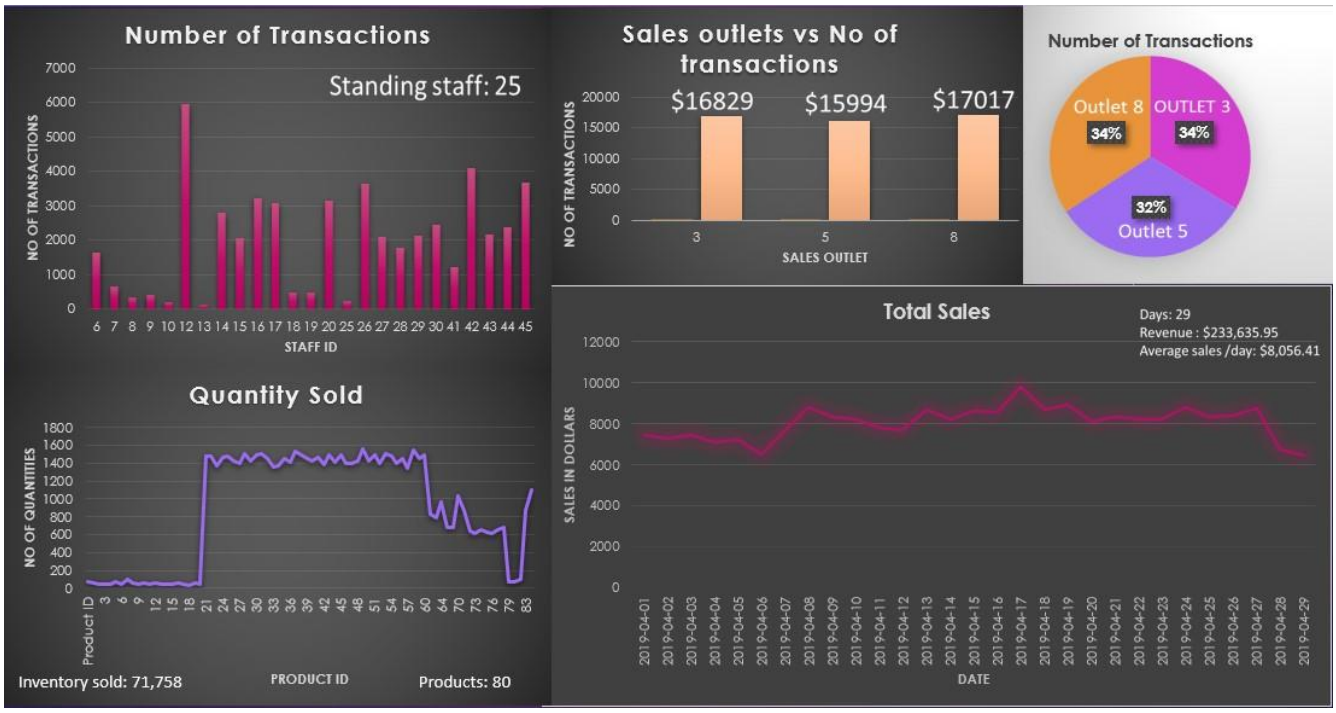
8. Profit analysis

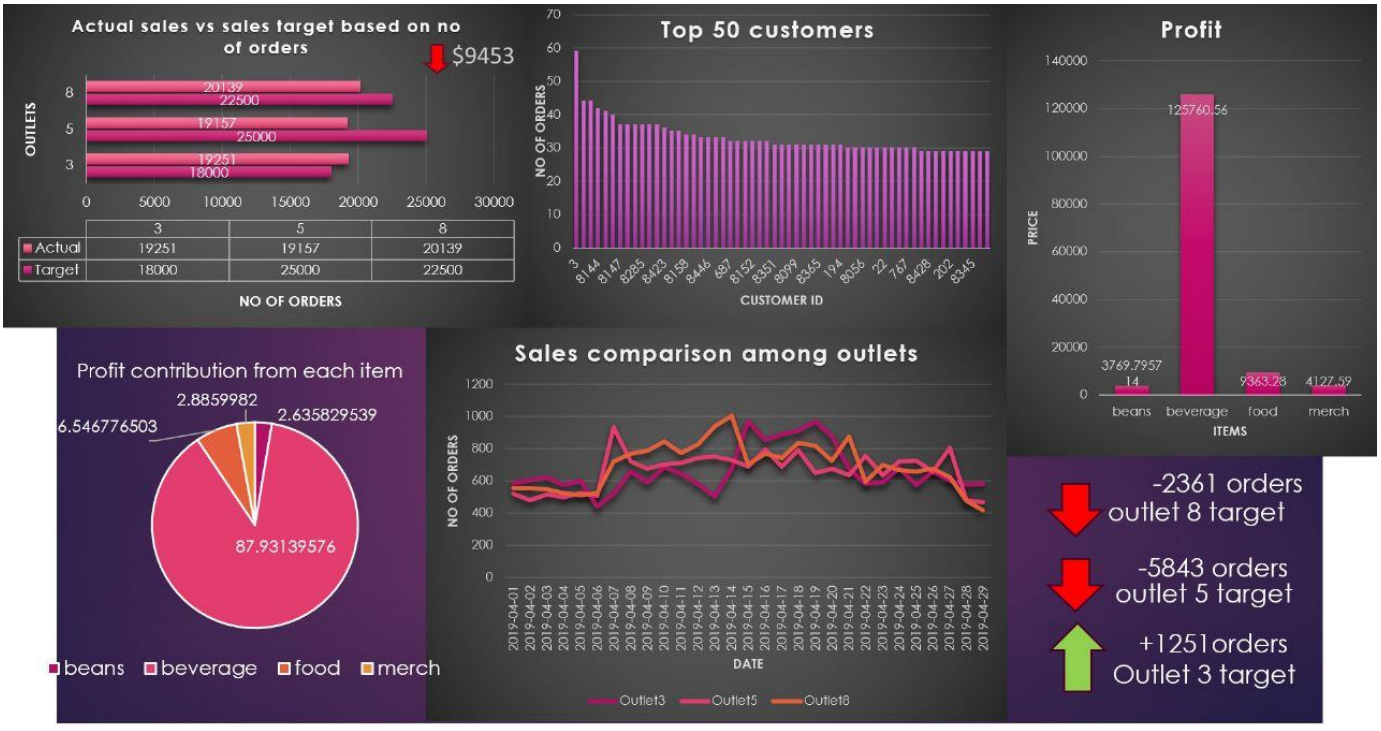
- The Beverage has the high profit of the month from all the outlets.




Profit contribution from each item
2.635829539
2.8859982
6.546776503
87.93139576
beans
beverage
food
merch


Customer Generations
25
20
15
10
5
0
Baby boomers
Gen X
Older Millenials
Younger millenials
Gen Z


Number of Transactions
Standing staff: 25
NO OF TRANSACTIONS
STAFF ID
Sales outlets vs No of transactions
$16829
$15994
$17017
NO OF TRANSACTIONS
SALES OUTLET
Number of Transactions
Outlet 8
34%
OUTLET 3
34%
32%
Outlet 5
Quantity Sold
NO OF QUANTITIES
Product ID
Inventory sold: 71,758
PRODUCT ID
Products: 80
Total Sales
Days: 29
Revenue : $233,635.95
Average sales /day: $8,056.41
SALES IN DOLLARS
DATE


Actual sales vs sales target based on no of orders
$9453
OUTLETS
NO OF ORDERS
Actual
Target
Top 50 customers
NO OF ORDERS
CUSTOMER ID
Profit
PRICE
ITEMS
Profit contribution from each item
2.8859982
2.635829539
6.546776503
87.93139576
beans
beverage
food
merch
Sales comparison among outlets
NO OF ORDERS
DATE
Outlet3
Outlet5
Outlet8
-2361 orders outlet 8 target
-5843 orders outlet 5 target
+1251orders Outlet 3 target

Recommendations and insights

1. Optimize Pricing Strategies: Analyse price elasticity based on sales data to determine the optimal pricing for products or services. Consider running pricing experiments to identify the price point that maximizes revenue and profit.

2. Focus on Top-Selling Products: Concentrate resources on promoting and enhancing the top-selling products or services. Consider bundling them with complementary items to increase the average order value.

3. Targeting High-Value Customer Segments: Segment customers based on their purchase frequency, total spending, or customer lifetime value. Tailor marketing campaigns to target high-value customer segments for better retention and increased sales.

4. Personalize Marketing Efforts: Leverage customer data from sales transactions to implement personalized marketing strategies. Use past purchase behaviour to recommend relevant products or offer exclusive deals to increase customer engagement.

5. Improving Customer Retention: Identify factors that contribute to customer churn using sales data and implement strategies to improve customer retention. Focus on customer satisfaction, post-purchase engagement, and loyalty programs.

6. Enhancing Sales Team Performance: Utilize sales data to assess the performance of individual sales representatives. Identify training needs and provide coaching to improve their effectiveness in driving sales.

Limitations:

Sales data, like any dataset, has certain limitations that should be considered when analysing and interpreting the results. Some common limitations of sales data include:

1. **Incomplete Data:** Sales data may have missing or incomplete entries, which can impact the accuracy and reliability of the analysis.

2. **Sampling Bias:** If the sales data is obtained through sampling or represents only a specific subset of the target population, it may not fully represent the entire customer base or market.

3. **Data Entry Errors:** Human errors during data entry can introduce inaccuracies into the dataset, leading to incorrect conclusions.

4**. Seasonal Variations:** Sales data may exhibit seasonal trends or cyclic patterns, which can complicate trend analysis and forecasting.

5. Lack of Context: Sales data alone may not provide a complete understanding of the underlying reasons for changes in sales performance. It may require additional data or qualitative analysis to gain insights.

Scope:

Sales data offers a wide range of applications and opportunities for businesses to improve their performance and make data-driven decisions.

1. **Sales Performance Analysis:** Sales data allows businesses to analyze their sales performance over time, identify trends, and evaluate the success of marketing and sales strategies.

2. **Customer Segmentation:** Sales data can be used to segment customers based on various criteria, such as purchasing behaviour, demographics, geography, or buying frequency. This segmentation helps in targeted marketing and personalized customer experiences.

3. **Sales Forecasting:** By analysing historical sales data, businesses can develop sales forecasting models to predict future sales trends, anticipate demand, and optimize inventory management.

4. **Pricing Optimization:** Sales data can be used to analyse the relationship between pricing and sales volume, helping businesses optimize their pricing strategies to maximize revenue and profitability.

5. **Identifying Top-Selling Products:** Sales data helps identify which products are the most popular and contribute significantly to the overall revenue. This information can guide inventory management and product development decisions.

Conclusion:

In conclusion, the analysis of the coffee shop sales dataset has provided valuable insights into the performance and trends of the coffee shop business. The sales data for the specified period has shed light on several key aspects that can inform strategic decision-making and improvements.

The top-selling products, such as Earl Grey Rg and dark chocolate Lg have emerged as the main revenue drivers. The coffee shop should continue to focus on these popular items while exploring opportunities to introduce new and innovative offerings to attract more customers.

It is crucial for the coffee shop to invest in customer retention strategies, such as loyalty programs and personalized marketing, to enhance customer satisfaction and encourage repeat business.

The findings of this analysis underscore the importance of data-driven decision-making for the coffee shop's success. To capitalize on these insights, it is essential for the coffee shop to continually monitor sales data, adapt to changing customer preferences, and stay competitive in the dynamic coffee industry.

In conclusion, by leveraging the power of sales data and implementing the recommended strategies, the coffee shop can enhance its overall performance, strengthen customer relationships, and establish itself as a beloved destination for coffee enthusiasts. By staying attentive to sales data and embracing a customer-centric approach, the coffee shop is poised for sustainable growth and continued success in the competitive coffee market.