

PORTFOLIO

COFFEE SHOP SALES DATA ANALYSIS

PHYTON & TABLEAU

Taufiq Hidayat HL





CASE STUDY OVERVIEW

This project aims to analyze the evolving sales trends over time, study the performance metrics of various products, and provide valuable insights for business owners or coffee shop management.

In this project, I use Python (Google Colab), Microsoft Excel, and Tableau to describe and plot sales data, revealing important insights into sales trends, customer preferences, and overall business performance.

The dataset consists of transaction records from Maven Roasters, a fictional coffee shop based in NYC that operates in three different locations. You can download it on Kaggle.

BUSINESS QUESTION

To facilitate the analysis of this coffee shop's sales data, I have formulated several business analysis questions. These questions serve as a guide for me on the information I need, and the acquired insights will be valuable for discovering and understanding sales trends, customer preferences, improving business efficiency, and making data-driven decisions.

- 1 How has the sales performance been during a specific period?
- 2 What is the total sales during that time range?
- 3 How does the sales variation occur across different stores?
- 4 Which product category is the top-selling?
- 5 Are there any discernible trends in sales during specific hours?
- 6 Is there variation in sales trends between weekdays and weekends?



DATA PREPARATION

First, we will perform data preparation using Google Colab (Python) to load data from Kaggle and then check for any Null and Duplicate rows.

INPUT

```
import zipfile

# Define the path to your zip file
file_path = '/content/drive/MyDrive/Kaggle/archive.zip' # Replace 'your_file.zip' with your file's name

# Unzip the file to a specific destination
with zipfile.ZipFile(file_path, 'r') as zip_ref:
    zip_ref.extractall('/content/drive/MyDrive/Kaggle') # Replace 'destination_folder' with your desired folder

import pandas as pd
import numpy as np

#load data set dari Kaggle
sales = pd.read_csv('/content/drive/MyDrive/Kaggle/coffee-shop-sales-revenue.csv', sep = '|')
print(sales.shape)
print(sales.info())

#periksa data null dan duplikat data
print('Cek data null :\n', sales.isnull().sum())
print('Cek data duplikat:\n', sales.duplicated())
sales.head()
```

DATA PREPARATION

The results show that there is no data or rows containing null values, and there are also no duplicate rows.

OUTPUT

```
(149116, 11)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 149116 entries, 0 to 149115
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   transaction_id    149116 non-null   int64  
 1   transaction_date  149116 non-null   object  
 2   transaction_time  149116 non-null   object  
 3   transaction_qty   149116 non-null   int64  
 4   store_id          149116 non-null   int64  
 5   store_location    149116 non-null   object  
 6   product_id        149116 non-null   int64  
 7   unit_price        149116 non-null   float64 
 8   product_category  149116 non-null   object  
 9   product_type      149116 non-null   object  
 10  product_detail   149116 non-null   object  
dtypes: float64(1), int64(4), object(6)
memory usage: 12.5+ MB
None
```

```
Cek data null :
transaction_id          0
transaction_date         0
transaction_time         0
transaction_qty          0
store_id                 0
store_location           0
product_id               0
unit_price               0
product_category         0
product_type              0
product_detail            0
dtype: int64
```

```
Cek data duplikat:
0      False
1      False
2      False
3      False
4      False
...
149111  False
149112  False
149113  False
149114  False
149115  False
Length: 149116, dtype: bool
```

transaction_id	transaction_date	transaction_time	transaction_qty	store_id	store_location	product_id	unit_price	product_category	product_type
1	2023-01-01	07:06:11	2	5	Lower Manhattan	32	3.0	Coffee	Gourmet brewed coffee
2	2023-01-01	07:08:56	2	5	Lower Manhattan	57	3.1	Tea	Brewed Chai tea
3	2023-01-01	07:14:04	2	5	Lower Manhattan	59	4.5	Drinking Chocolate	Hot chocolate
4	2023-01-01	07:20:24	1	5	Lower Manhattan	22	2.0	Coffee	Drip coffee
5	2023-01-01	07:22:41	2	5	Lower Manhattan	57	3.1	Tea	Brewed Chai tea

DATA PREPARATION

Next, we'll change the data types and create new columns as needed for analysis.

INPUT

```
#mengubah dtype menjadi datetime
sales['transaction_date'] = pd.to_datetime(sales['transaction_date'])
sales['transaction_time'] = pd.to_datetime(sales['transaction_time'])

# Membuat kolom baru dari transaction_date dan transaction_time
sales['transaction_day'] = sales['transaction_date'].dt.day
sales['transaction_month'] = sales['transaction_date'].dt.month
sales['transaction_year'] = sales['transaction_date'].dt.year
sales['day_of_week'] = sales['transaction_date'].dt.day_name()
sales['transaction_hour'] = sales['transaction_time'].dt.hour

# Membuat Kolom total_revenue
sales['total_revenue'] = sales['transaction_qty'] * sales['unit_price']
print(sales.info())
sales.head()
```

DATA PREPARATION

The results show that 6 new columns have been added, and the data type of the 'transaction_date' column has been changed to the datetime data type

OUTPUT

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 149116 entries, 0 to 149115
Data columns (total 17 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   transaction_id  149116 non-null   int64  
 1   transaction_date 149116 non-null   datetime64[ns]
 2   transaction_time 149116 non-null   datetime64[ns]
 3   transaction_qty   149116 non-null   int64  
 4   store_id          149116 non-null   int64  
 5   store_location    149116 non-null   object  
 6   product_id        149116 non-null   int64  
 7   unit_price        149116 non-null   float64
 8   product_category  149116 non-null   object  
 9   product_type      149116 non-null   object  
 10  product_detail    149116 non-null   object  
 11  transaction_day   149116 non-null   int64  
 12  transaction_month 149116 non-null   int64  
 13  transaction_year  149116 non-null   int64  
 14  day_of_week       149116 non-null   object  
 15  transaction_hour  149116 non-null   int64  
 16  total_revenue     149116 non-null   float64
dtypes: datetime64[ns](2), float64(2), int64(8), object(5)
memory usage: 19.3+ MB
None
```



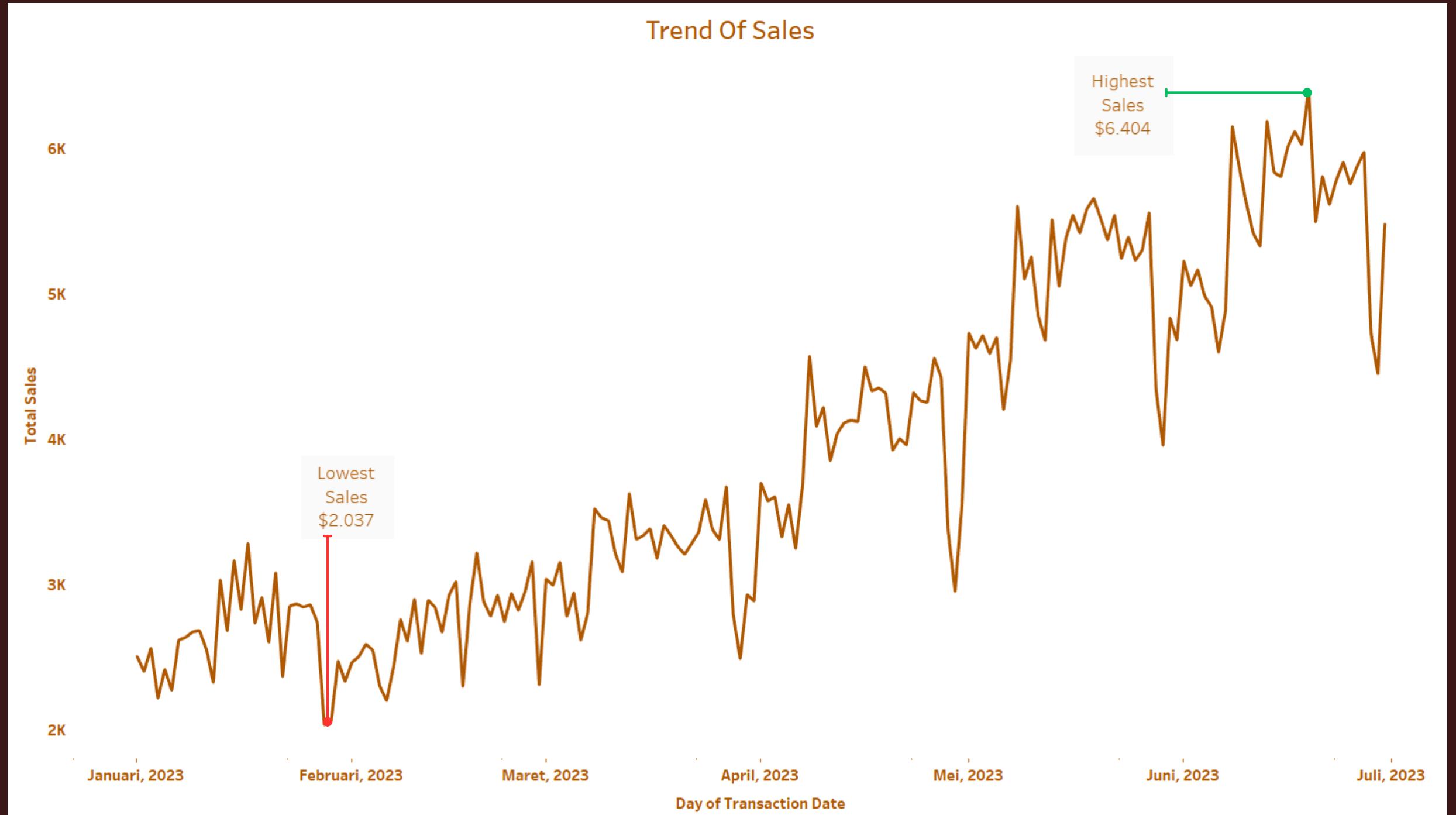
product_id	unit_price	product_category	product_type	product_detail	transaction_day	transaction_month	transaction_year	day_of_week	transaction_hour	total_revenue
32	3.0	Coffee	Gourmet brewed coffee	Ethiopia Rg	1	1	2023	Sunday	7	6.0
57	3.1	Tea	Brewed Chai tea	Spicy Eye Opener Chai Lg	1	1	2023	Sunday	7	6.2
59	4.5	Drinking Chocolate	Hot chocolate	Dark chocolate Lg	1	1	2023	Sunday	7	9.0
22	2.0	Coffee	Drip coffee	Our Old Time Diner Blend Sm	1	1	2023	Sunday	7	2.0
57	3.1	Tea	Brewed Chai tea	Spicy Eye Opener Chai Lg	1	1	2023	Sunday	7	6.2

DATA VISUALIZATION

I am using the Tableau application to create data visualizations. First, I will display the sales trend data during a specific period

The visualization results indicate that the sales trend is progressing very well, with a stable average monthly growth of \$6,404 and a growth rate of 2.43%.

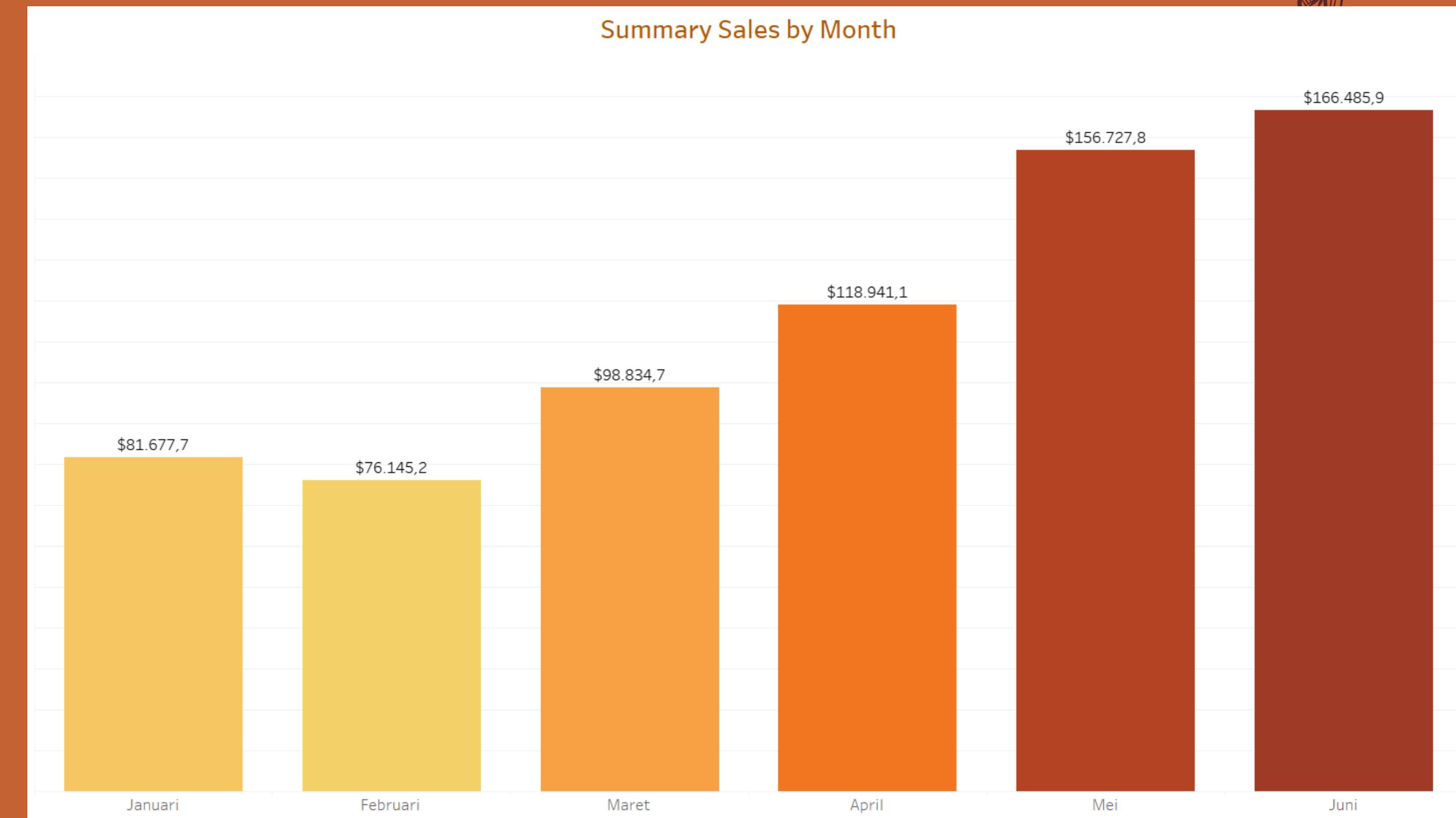
The peak sales on June 19, 2023 (\$6,404) marks a positive trend, while the lowest point on January 28, 2023 (\$2,037) provides an opportunity for strategy improvement. Evaluating the contributing factors to the growth in June is crucial to maximize performance during months with lower sales.



DATA VISUALIZATION

Visualisasi data Sales by Month

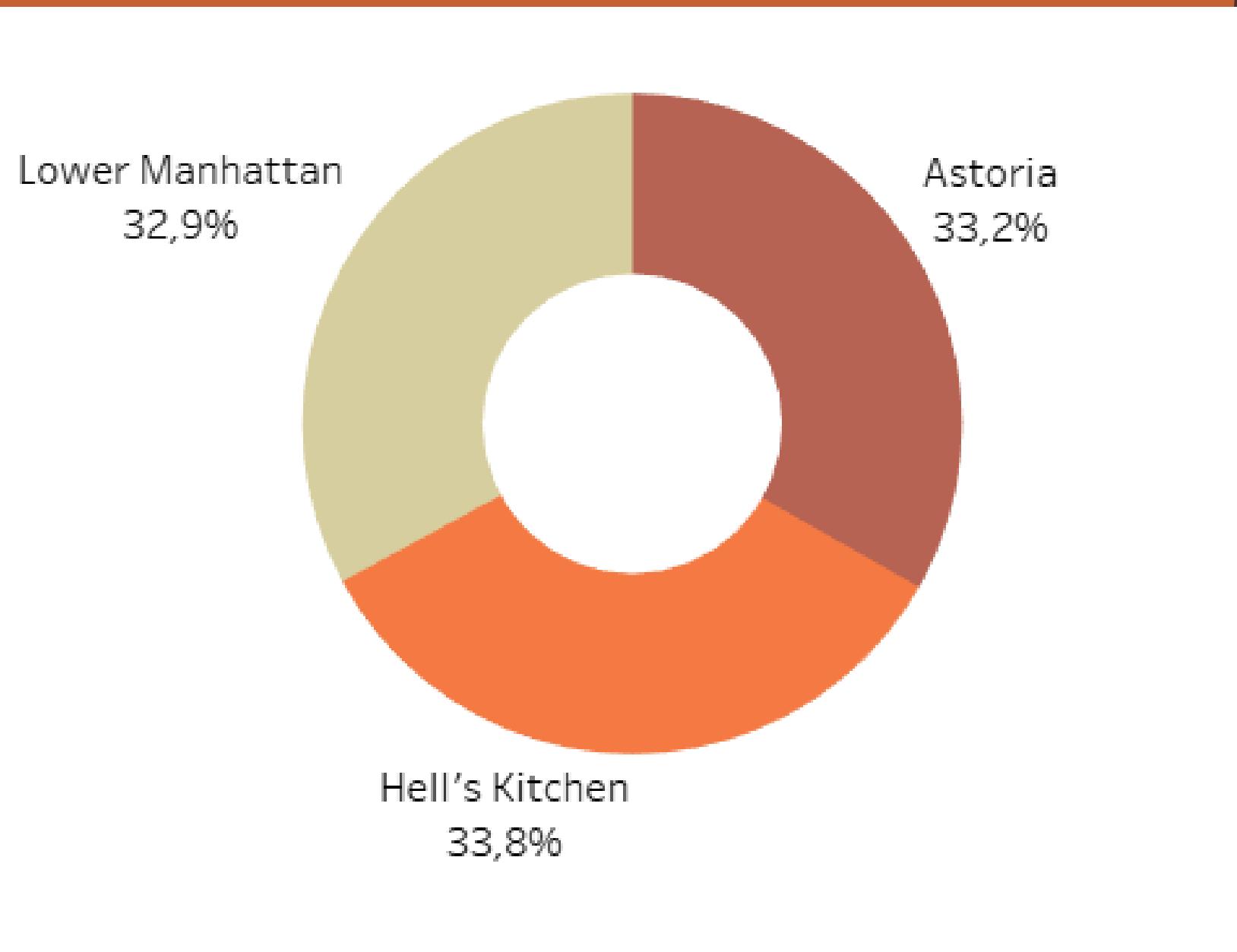
Sales consistently increased each month, with a slight dip observed in February. The total sales reached \$698.8 with a total of 149 thousand transactions. The peak occurred in June with sales reaching \$166.49 thousand, representing 23.8% of the total sales during that period



DATA VISUALIZATION

Visualisasi data Sales by Store

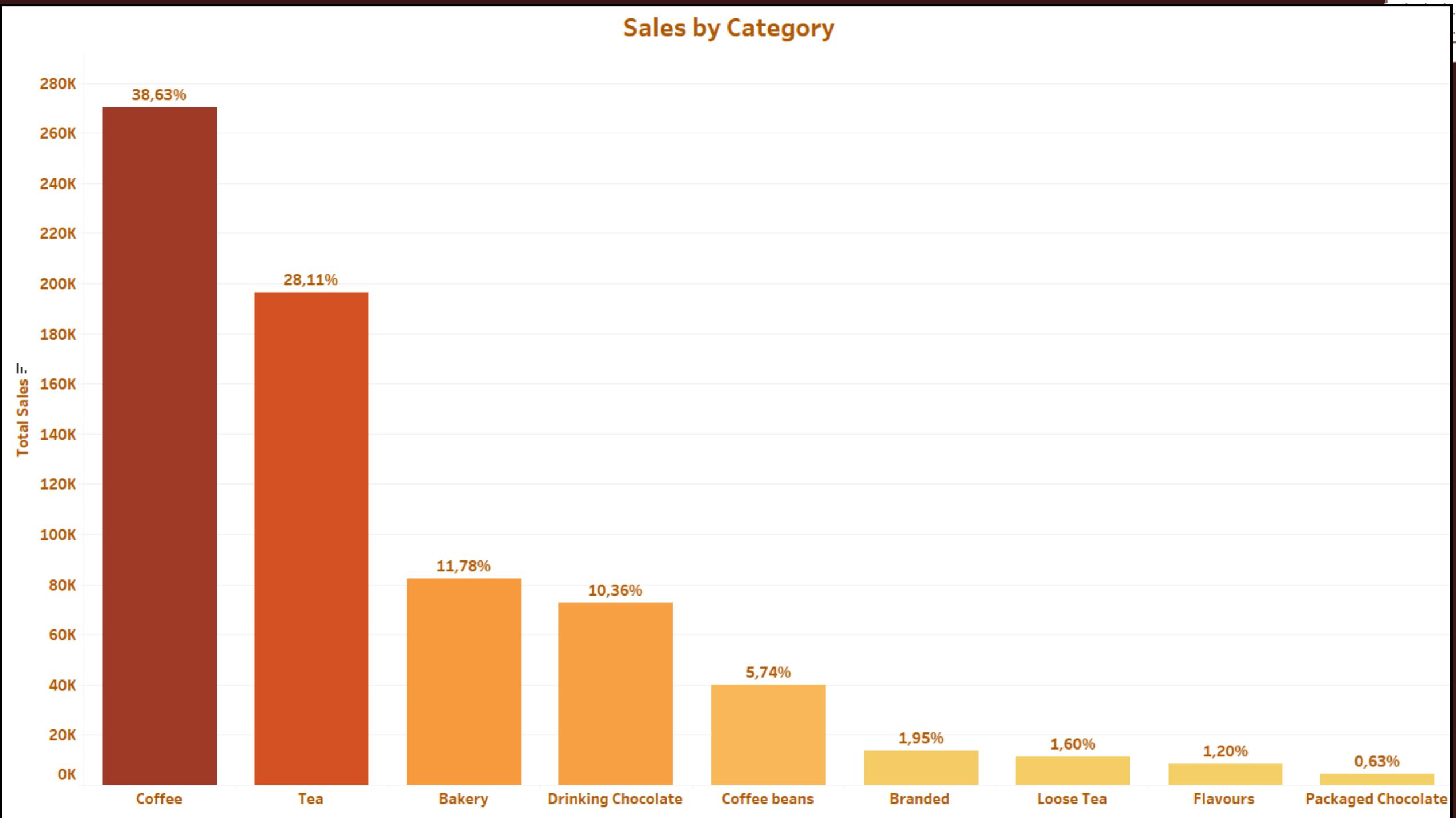
The total sales across the three stores are evenly distributed, indicating a balanced performance. This provides an opportunity to implement strategies that leverage the strengths of each store, ensuring a cohesive and effective business approach



DATA VISUALIZATION

Visualisasi data Sales by Category

The coffee product category has the highest sales contribution during the period of Jan-Jun 2023, reaching 38.6%. Meanwhile, the tea and bread categories follow with sales contributions of 28.1% and 11.8%, respectively. This indicates that **the coffee category is the most preferred by customers**, followed by tea and bread products. The four categories with the lowest sales, namely Chocolate packages, Brands, Loose tea, and Coffee beans, collectively contribute only 5%

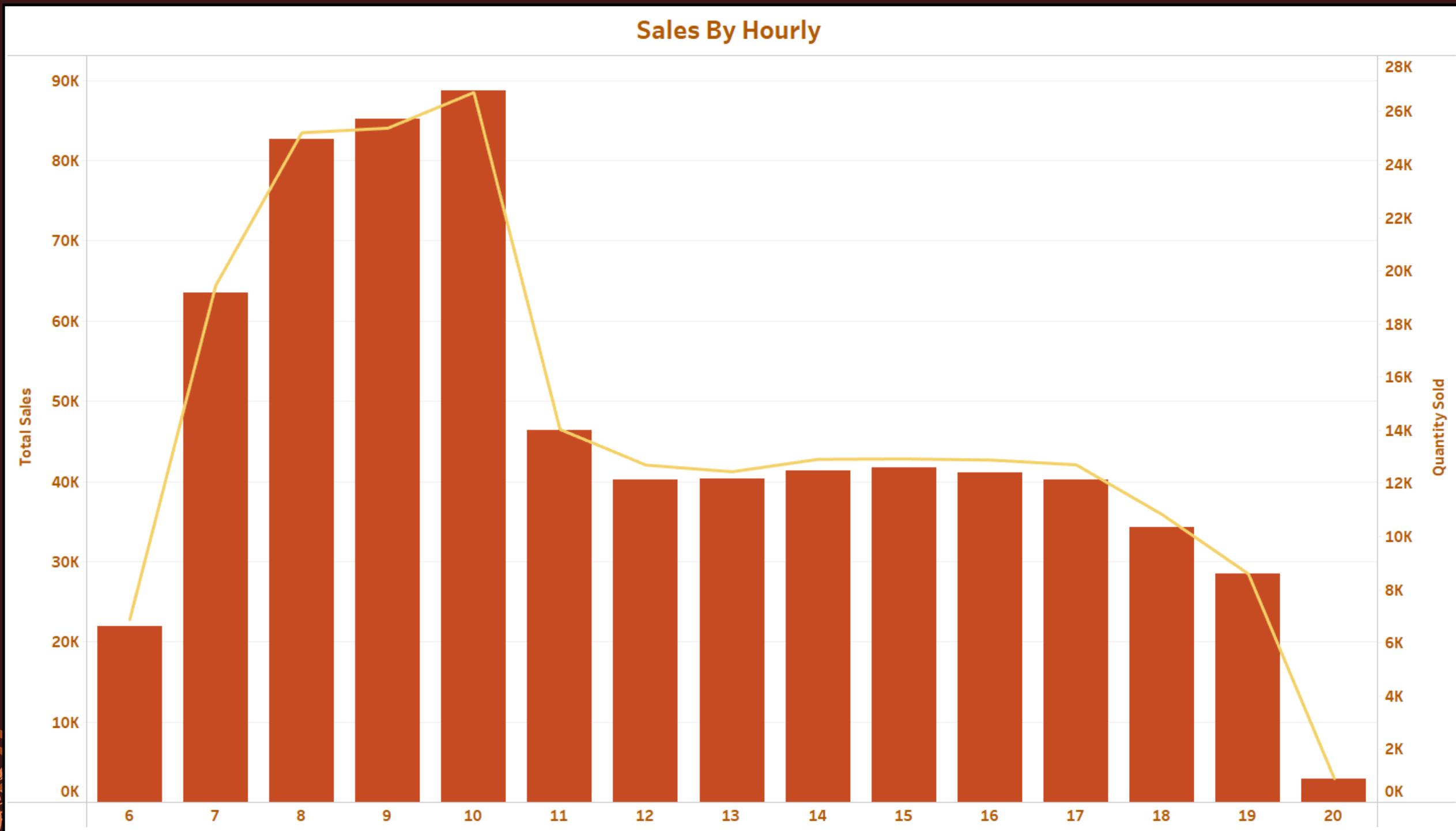


DATA VISUALIZATION

Visualisasi data Sales by Hourly

The peak of sales occurs from 07:00 AM to 11:00 AM, with featured categories such as Flavours, Tea, Coffee, and Chocolate Drinks.

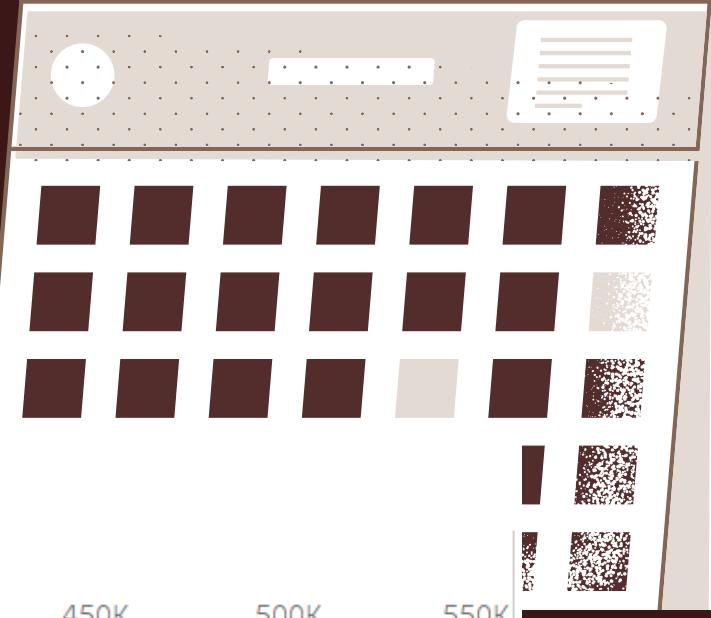
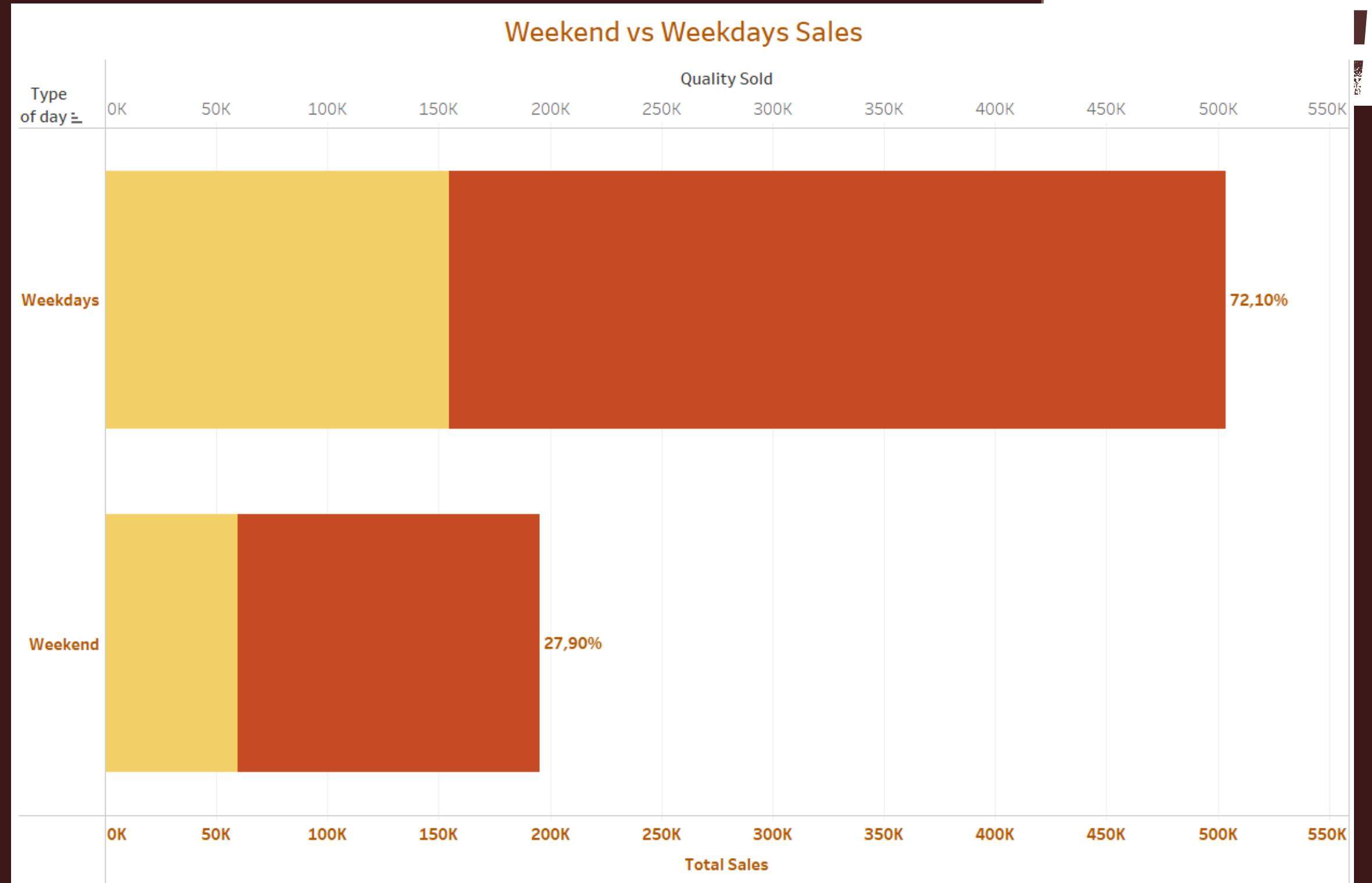
This insight can inform an effective business strategy involving a focus on promotions for the key categories in the morning, while also exploring innovative promotional initiatives to support evening sales growth



DATA VISUALIZATION

Visualization of sales trend data between weekdays and weekends.

A total of 72% of sales occur on weekdays. This can be attributed to the customer tendency to enjoy coffee as part of their daily work routine



MAKING DECISION

Based on the analysis results, the coffee shop Management can consider the following decisions

- **Monthly Promotion Strategy:** Implement monthly promotions to sustain the positive sales trend. Focus on addressing the slight dip observed in February to ensure consistent growth.
- **Product Focus and Marketing:** Given that the coffee category is the top-selling product, the owner should strategically market and promote coffee-related products. Consider innovative marketing initiatives to boost sales in other categories, especially the ones with lower contributions.
- **Peak Sales Hours Optimization:** Capitalize on the peak sales hours from 07:00 AM to 11:00 AM by emphasizing promotions and marketing efforts during this time. This includes highlighting featured categories such as Flavours, Tea, Coffee, and Chocolate Drinks.
- **Weekday Promotions:** Since 72% of sales occur on weekdays, the owner can design promotions targeting the specific preferences of customers during weekdays. This could involve creating special offers, loyalty programs, or partnerships to attract more weekday customers.



DATA VISUALIZATION

 [More Information](#)

Sales Dashboard Coffee Shop
Januari - Maret 2023

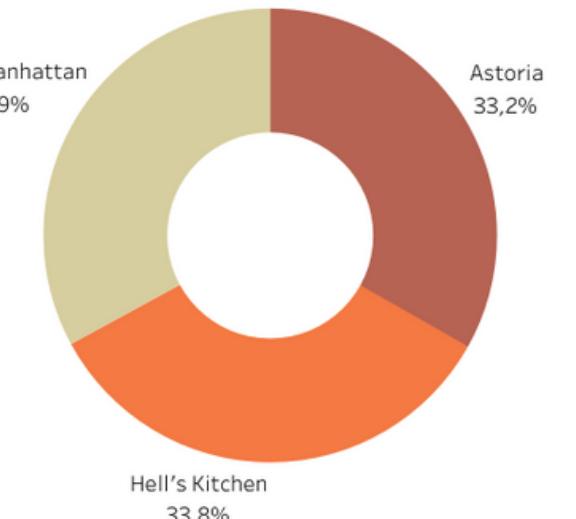
Summary Sales by Month



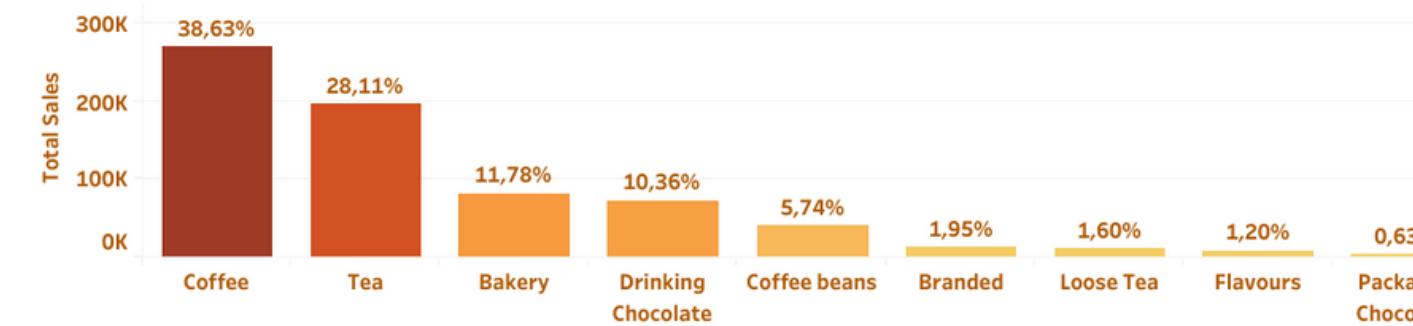
Weekend vs Weekdays Sales



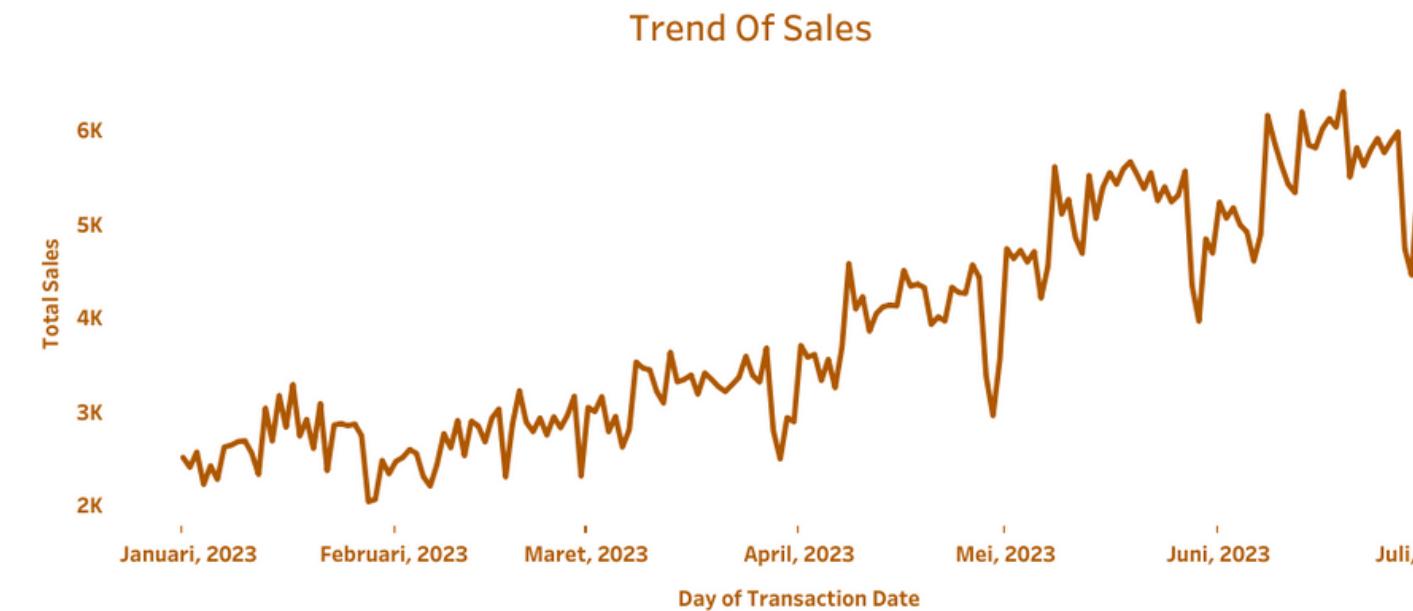
Sales by Store



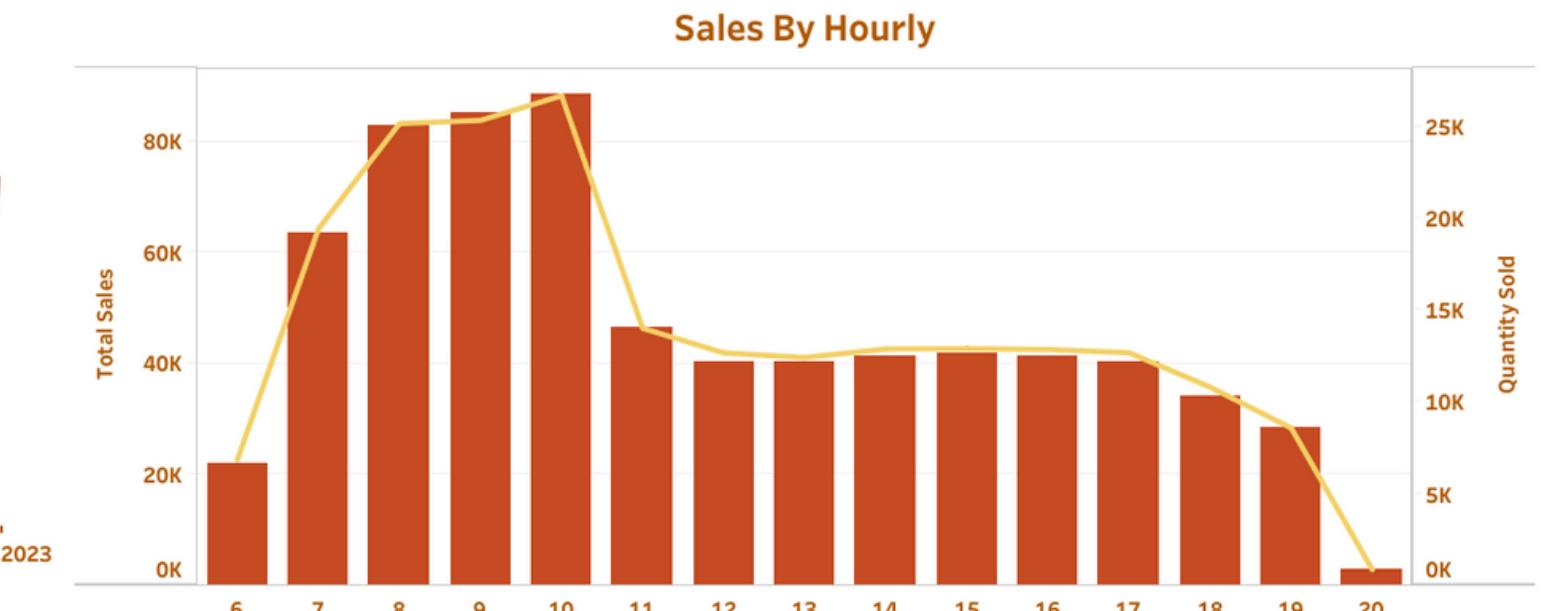
Sales by Category



Trend Of Sales



Sales By Hourly





THANK YOU !



linkedin.com/in/tfqhidayat



082251404245



taufiqhidayat1807@gmail.com