Documentation

Developed by Joshua Maries Colet

Title : Dmart Sales Data Analysis Dashboard

Problem Statement : DMart is one of India's most efficient retail chains, offering a vast product line to millions of customers. As the volume of sales and transactions grows, so does the complexity of retrieving meaningful insights from this data. The challenge lies in transforming raw transactional data into actionable insights for decision-making which is the core of data analysis. This project addresses that challenge by developing an interactive dashboard to analyze sales, customer behavior and various product performance factors.

INDEX

DETAILED DESCRIPTION

- This project focuses on analyzing and visualizing sales data from DMart, a major retail chain in India, by developing a dynamic and interactive dashboard using Python and Streamlit. The core objective is to uncover trends, identify sales patterns, and derive useful business insights from transactional data spanning across different product categories and time periods.

- The dashboard is structured into multiple pages, each targeting a specific analytical aspect:
- **Home Page**: Serves as a welcoming interface with branding, project overview, and instructions on how to navigate the contents of the dashboard.
- **Sales Trends**: Displays overall sales performance over time, seasonal sale analysis (summer vs. winter), and highlights the top-selling product categories using time-series plots.
- **Customer Behavior**: Analyzes purchasing patterns across different genders, customer rating distribution, and customer loyalty analysis.
- **Product Analysis**: Breaks down performance by product lines, showing quantity sold and highlighting the most popular categories and items.
- **Payment Insights**: Provides a distribution of payment methods used (such as E-wallet/UPI, Cash etc.) , revealing customer preferences and mode efficiency.
- Every data visualization present in the dashboard has been given its own description and inferences that can be made by studying them. This gives life to the data in the form of story telling and and by contextualizing it.

Data used within the problem

The dataset consists of transactional records from DMart stores, structured as a CSV file. It includes the following key features:

- Date and Time – Timestamp of the transaction
- Product Line – Category of the purchased product
- Quantity – Number of units bought
- Unit Price and Total – Price per item and total cost for the line item
- Customer Type – Indicates loyalty status (Member (availed the membership for an extra cost or Normal)
- Gender – Gender of the customer
- Payment Method – Mode of payment used
- Branch – Store location
- Rating – Customer feedback rating

Solution Plan and Design

Data Collection

- The dataset was provided in CSV format, containing sales transactions from DMart.
- The data captured various attributes like date, time, quantity, product category, revenue, gender, and customer type.

Data Cleaning & Preprocessing

- Converted the date column to datetime format
- Extracted month, day, hour for time-based analysis
- Renamed columns for consistency and readability
- Checked for nulls, duplicates, and corrected data types and removed such rows/tuples if any.

Data Analysis (DA)

- Used groupby, pivot, and filtering techniques to:

    - Find top-selling product lines
    - Analyze seasonal trends
    - Evaluate member vs non-member behavior
    - Track revenue and quantity sold by category, gender, and payment type

Dashboard Design

The Streamlit dashboard was divided into logical pages:

- Home Page – Project overview
- Sales Trends – Trends over time (monthly, seasonal)
- Customer Behavior – Gender, time-based analysis, loyalty
- Product Analysis – Best-selling categories
- Payment Insights – Payment method breakdown
- Advanced Report Generator – Custom filters and summary exports
- Customer Loyalty – Added later to compare revenue contribution by member status

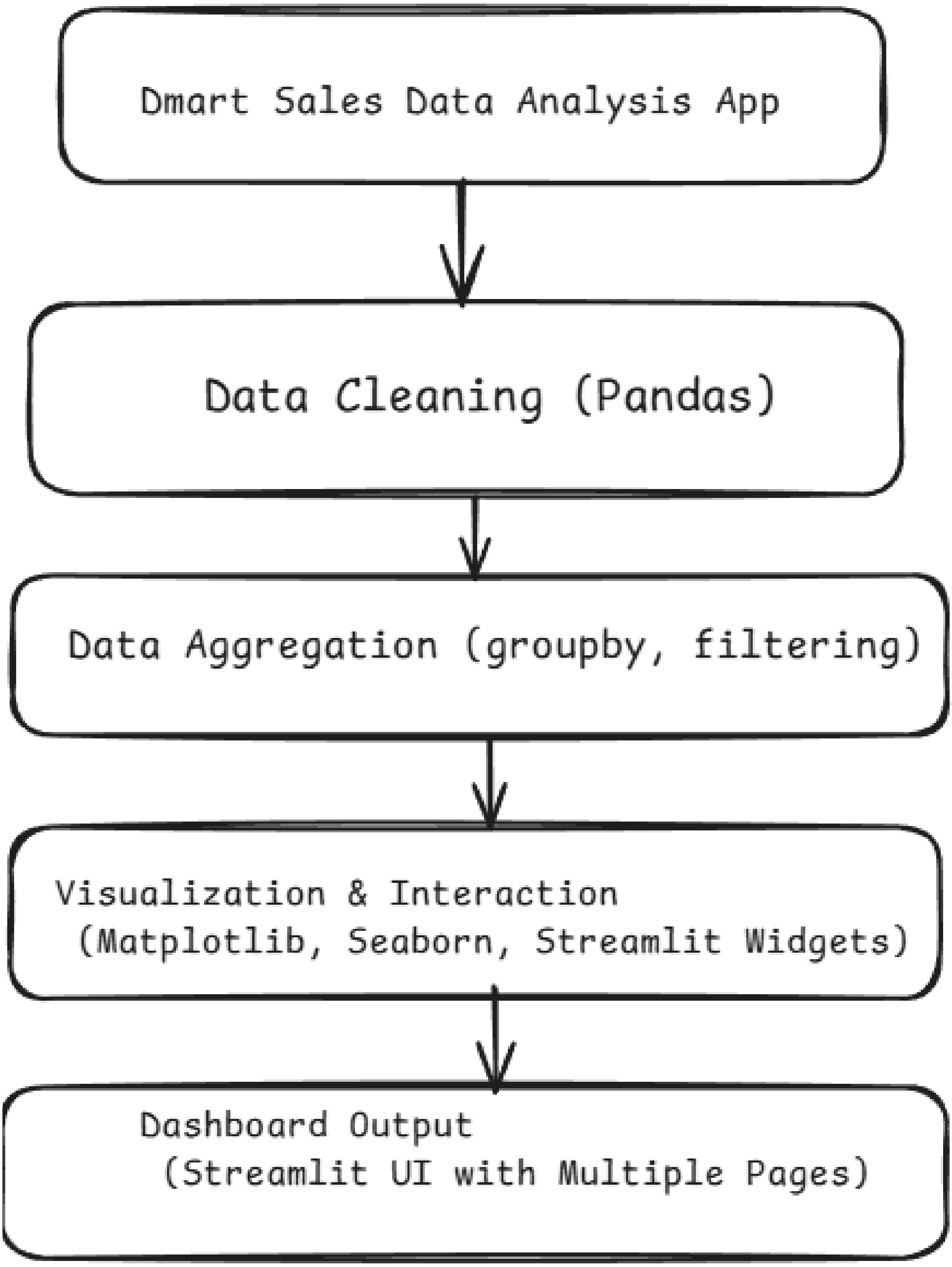
Visualization

- Used Seaborn and Matplotlib for charts:
    - Line charts for trends
    - Bar charts for product/category comparisons
    - Pie charts for customer-type breakdown

Outcome and benefits

we aim to extract meaningful insights from the given dataset which will in turn be used for major decision making.

Design



Dmart Sales Data Analysis App

Data Cleaning (Pandas)

Data Aggregation (groupby, filtering)

Visualization & Interaction
(Matplotlib, Seaborn, Streamlit Widgets)

Dashboard Output
(Streamlit UI with Multiple Pages)

Implementation , Code and Explanation

• Tech Stack

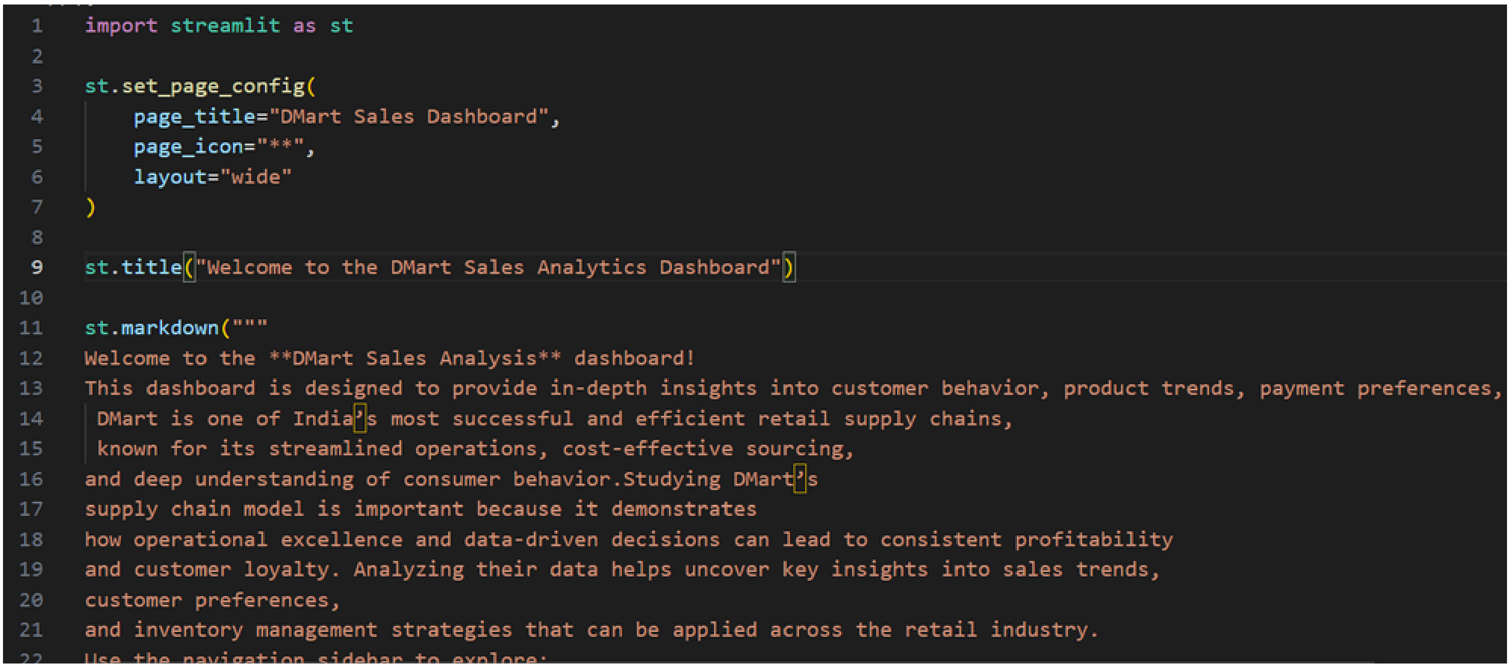
Category	Technologies
Programming Language	Python
Web Framework	Streamlit (for building interactive dashboards)
Data Visualization	Matplotlib, Seaborn
Data Storage Format	CSV (cleaned_dmart_data_final.csv)

• File Structure

```
--->  DMart_sales_analysis_dashboard/
|
├──── cleaned_dmart_data_final.csv
├────app.py
├────utils
      ────data_loader.py
```

```
├──── pages/
    ──── 1_Home.py
    ──── 2_Sales_Trends.py
    ──── 3_Customer_Behavior.py
   ──── 4_Product_Analysis.py
   ──── 5_Payment_Insights.py
```

- App.py



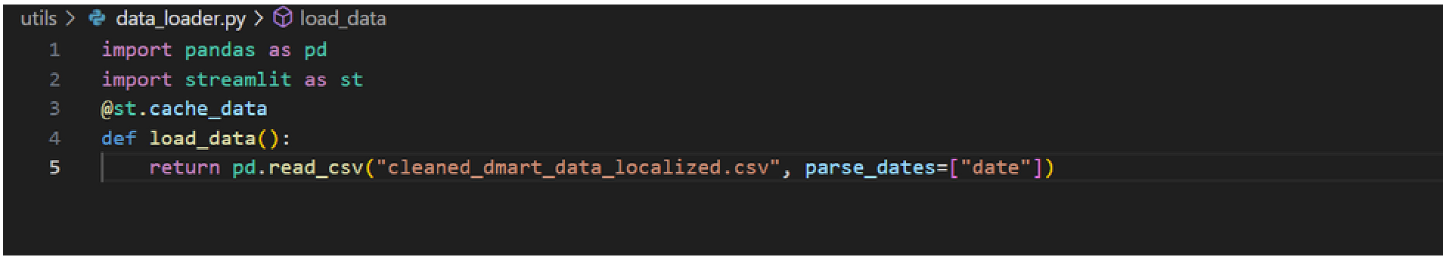
```
import streamlit as st

st.set_page_config(
    page_title="DMart Sales Dashboard",
    page_icon="**",
    layout="wide"
)

st.title("Welcome to the DMart Sales Analytics Dashboard")

st.markdown("""
Welcome to the **DMart Sales Analysis** dashboard!
This dashboard is designed to provide in-depth insights into customer behavior, product trends, payment preferences,
 DMart is one of India’s most successful and efficient retail supply chains,
 known for its streamlined operations, cost-effective sourcing,
and deep understanding of consumer behavior.Studying DMart’s
supply chain model is important because it demonstrates
how operational excellence and data-driven decisions can lead to consistent profitability
and customer loyalty. Analyzing their data helps uncover key insights into sales trends,
customer preferences,
and inventory management strategies that can be applied across the retail industry.
Use the navigation sidebar to explore:
```

Purpose:
 This is the main file that launches the dashboard when the user runs Streamlit run main.py. Therefore it acts as the entry point for the entire project.

Key Functions:

- Acts as the root from which all pages in the pages/ folder are loaded automatically (To control the order, you simply prefix your filenames with numbers.)

- **utils/Data_loader.py**

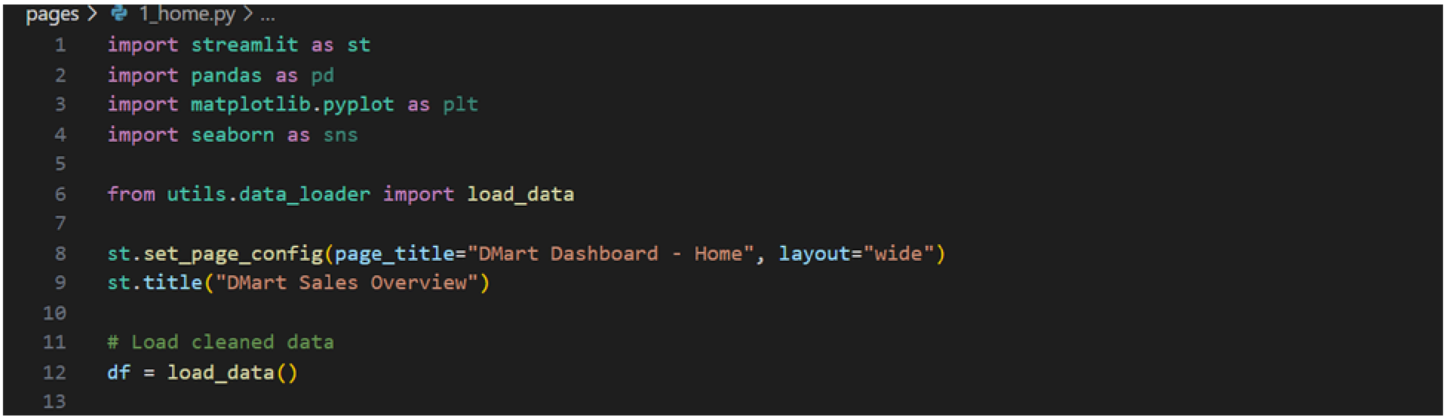


```
import pandas as pd
import streamlit as st
@st.cache_data
def load_data():
    return pd.read_csv("cleaned_dmart_data_localized.csv", parse_dates=["date"])
```

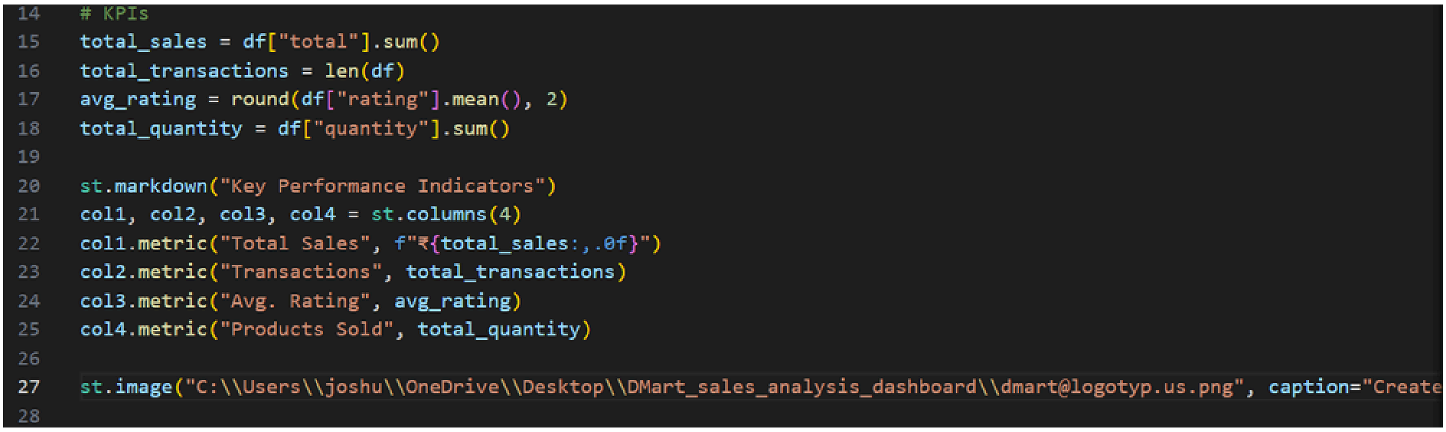
- This file handles loading and preprocessing the dataset using a single reusable function. It improves performance by caching the data and ensures consistency across all dashboard pages.
- @st.cache_data is used to load the CSV once per session, so a file having the necessity to load in the .csv file just has to import data_loader.py from utils and call the function we defined inside. This keeps the code modular and avoids repetition.

• /Pages

• 1_home.py



```python
import streamlit as st
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

from utils.data_loader import load_data

st.set_page_config(page_title="DMart Dashboard - Home", layout="wide")
st.title("DMart Sales Overview")

# Load cleaned data
df = load_data()

```



```python
# KPIs
total_sales = df["total"].sum()
total_transactions = len(df)
avg_rating = round(df["rating"].mean(), 2)
total_quantity = df["quantity"].sum()

st.markdown("Key Performance Indicators")
col1, col2, col3, col4 = st.columns(4)
col1.metric("Total Sales", f"₹{total_sales:,.0f}")
col2.metric("Transactions", total_transactions)
col3.metric("Avg. Rating", avg_rating)
col4.metric("Products Sold", total_quantity)

st.image("C:\\Users\\joshu\\OneDrive\\Desktop\\DMart_sales_analysis_dashboard\\dmart@logotyp.us.png", caption="Create

```

Purpose:

Acts as the landing page of the DMart Sales Dashboard. It introduces users to the project, its scope, and how to navigate the dashboard effectively.

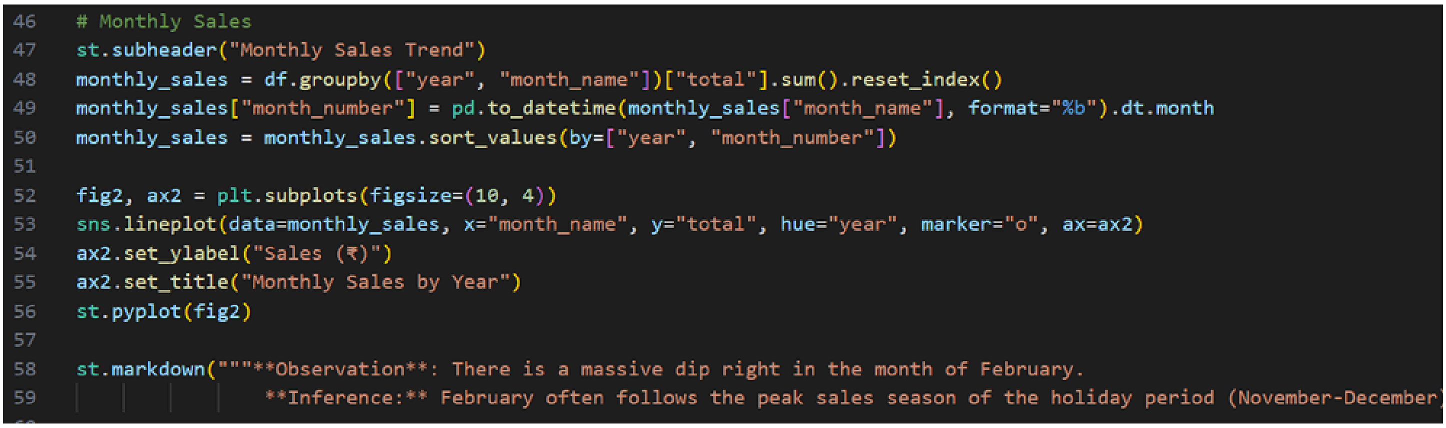
Key Functions:

- Displays the project title, author name, and brief overview using Streamlit's markdown/HTML components.
- Highlights key dashboard sections like:
  - Sales Trends

  - Customer Behavior
  - Product Performance
  - Payment Insights

- 2_Sales_trend.py

a snippet of the monthly trends program from the page which generates a line chart

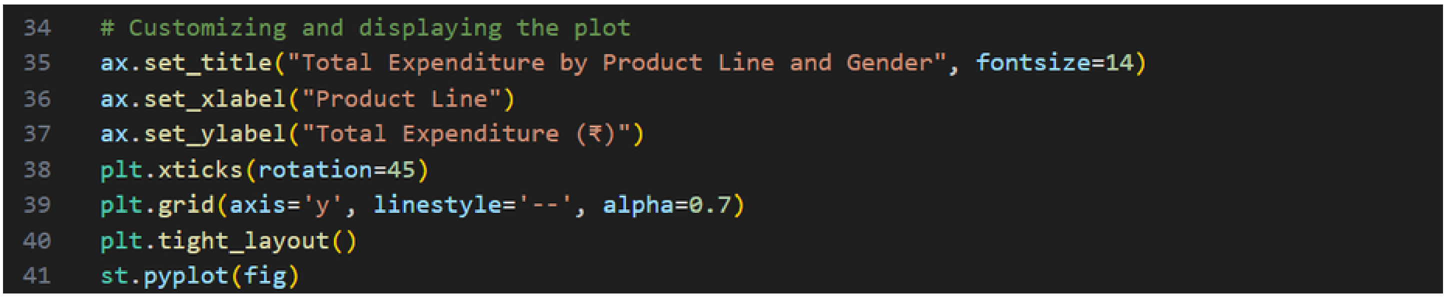


```
# Monthly Sales
st.subheader("Monthly Sales Trend")
monthly_sales = df.groupby(["year", "month_name"])["total"].sum().reset_index()
monthly_sales["month_number"] = pd.to_datetime(monthly_sales["month_name"], format="%b").dt.month
monthly_sales = monthly_sales.sort_values(by=["year", "month_number"])

fig2, ax2 = plt.subplots(figsize=(10, 4))
sns.lineplot(data=monthly_sales, x="month_name", y="total", hue="year", marker="o", ax=ax2)
ax2.set_ylabel("Sales (₹)")
ax2.set_title("Monthly Sales by Year")
st.pyplot(fig2)

st.markdown("""**Observation**: There is a massive dip right in the month of February.
                **Inference:** February often follows the peak sales season of the holiday period (November-December)
```

its functions include :

- Visualizes total revenue by month to detect seasonal patterns.
- Highlights sales during summer vs winter to identify seasonal demand.
- Shows sales by day of week and in a monthly basis to understand peak business periods.

- 3_Customer_behaviour.py



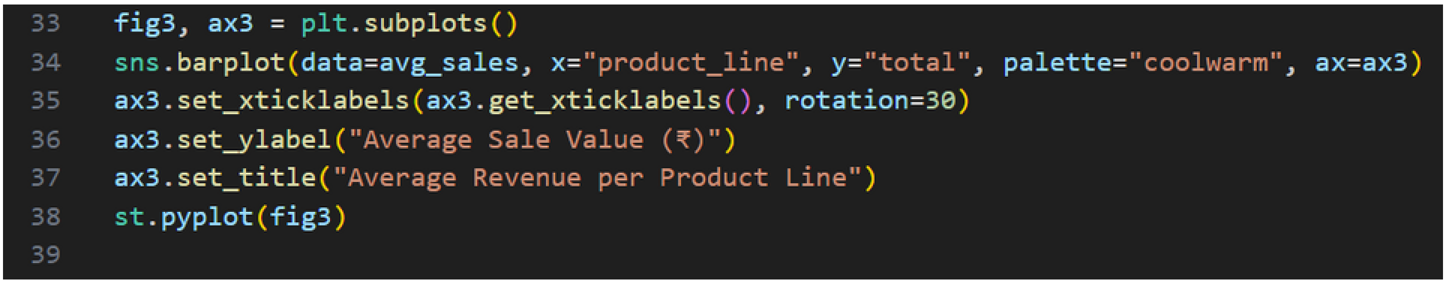
```
# Customizing and displaying the plot
ax.set_title("Total Expenditure by Product Line and Gender", fontsize=14)
ax.set_xlabel("Product Line")
ax.set_ylabel("Total Expenditure (₹)")
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
st.pyplot(fig)
```

a snippet from the page that generates a bar-graph comparing the products brought by male and female customers across all the product lines.

Functions :

- Compares **male vs female** spending across each product category.
- Shows the difference in revenue obtained from members and regular customers.
- Helps identify **target audience** for specific product lines based on gender and timing.

- 4_Product_Analysis

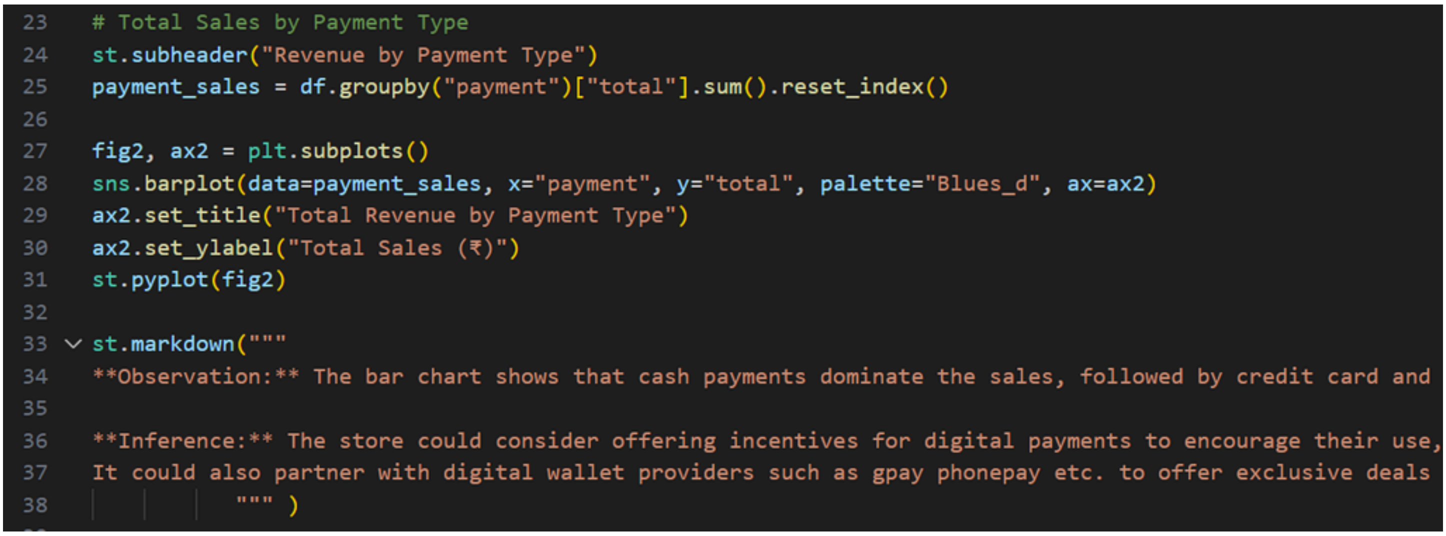


```
fig3, ax3 = plt.subplots()
sns.barplot(data=avg_sales, x="product_line", y="total", palette="coolwarm", ax=ax3)
ax3.set_xticklabels(ax3.get_xticklabels(), rotation=30)
ax3.set_ylabel("Average Sale Value (₹)")
ax3.set_title("Average Revenue per Product Line")
st.pyplot(fig3)

```

- Provides a deep dive into product line performance and sales contribution.
- Ranks top-selling categories by total quantity and revenue.

- 5_Payment_Insights.py

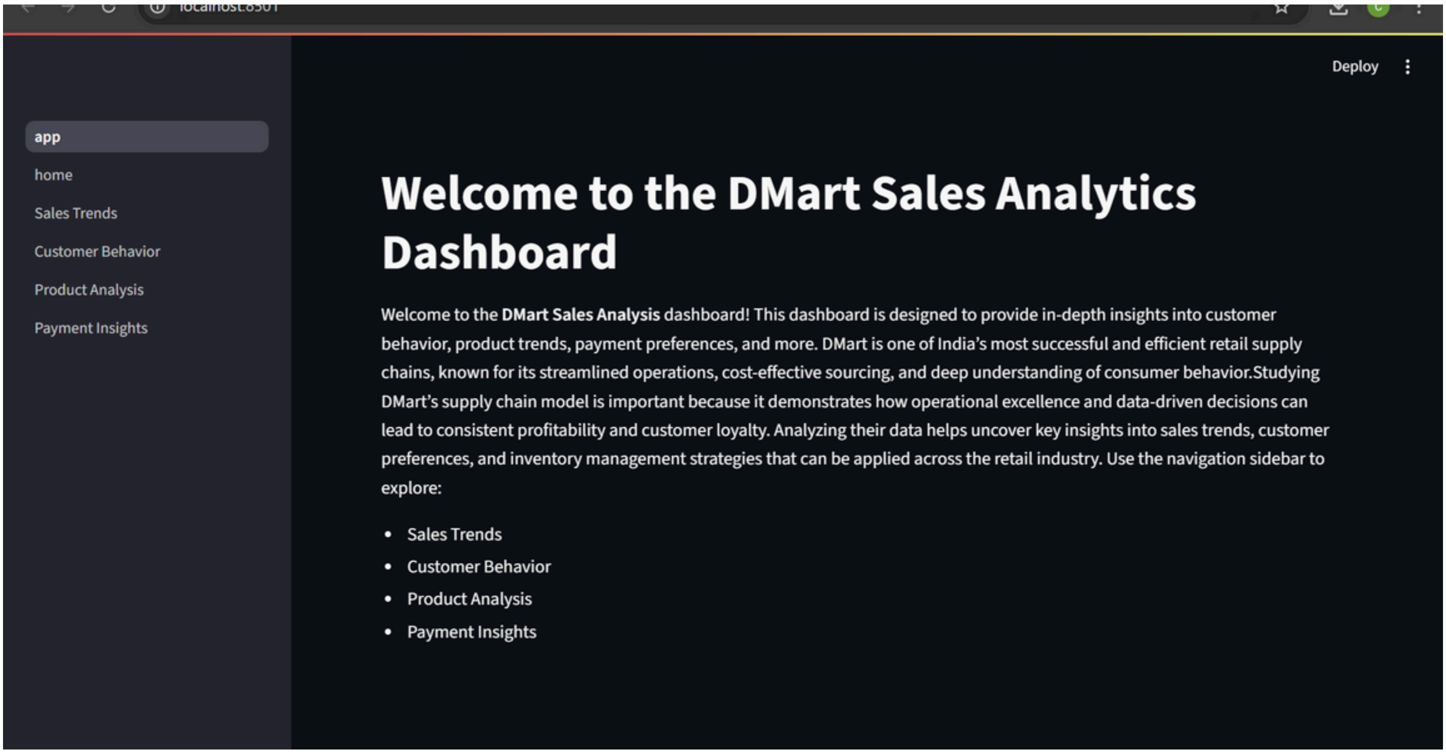
- Analyzes customer payment preferences and how sales differ across store branches.
- Shows the proportion of transactions by payment method (Cash, E-wallet, Credit Card).
- Compares total revenue per branch to identify high-performing stores.
- Helps understand regional differences in payment behavior.



```
# Total Sales by Payment Type
st.subheader("Revenue by Payment Type")
payment_sales = df.groupby("payment")["total"].sum().reset_index()

fig2, ax2 = plt.subplots()
sns.barplot(data=payment_sales, x="payment", y="total", palette="Blues_d", ax=ax2)
ax2.set_title("Total Revenue by Payment Type")
ax2.set_ylabel("Total Sales (₹)")
st.pyplot(fig2)

st.markdown("""
**Observation:** The bar chart shows that cash payments dominate the sales, followed by credit card and

**Inference:** The store could consider offering incentives for digital payments to encourage their use,
It could also partner with digital wallet providers such as gpay phonepay etc. to offer exclusive deals
            """ )
```

a code snippet from this page showing total revenue generated from different payment methods in the form of a bar-graph

- Output Screenshots

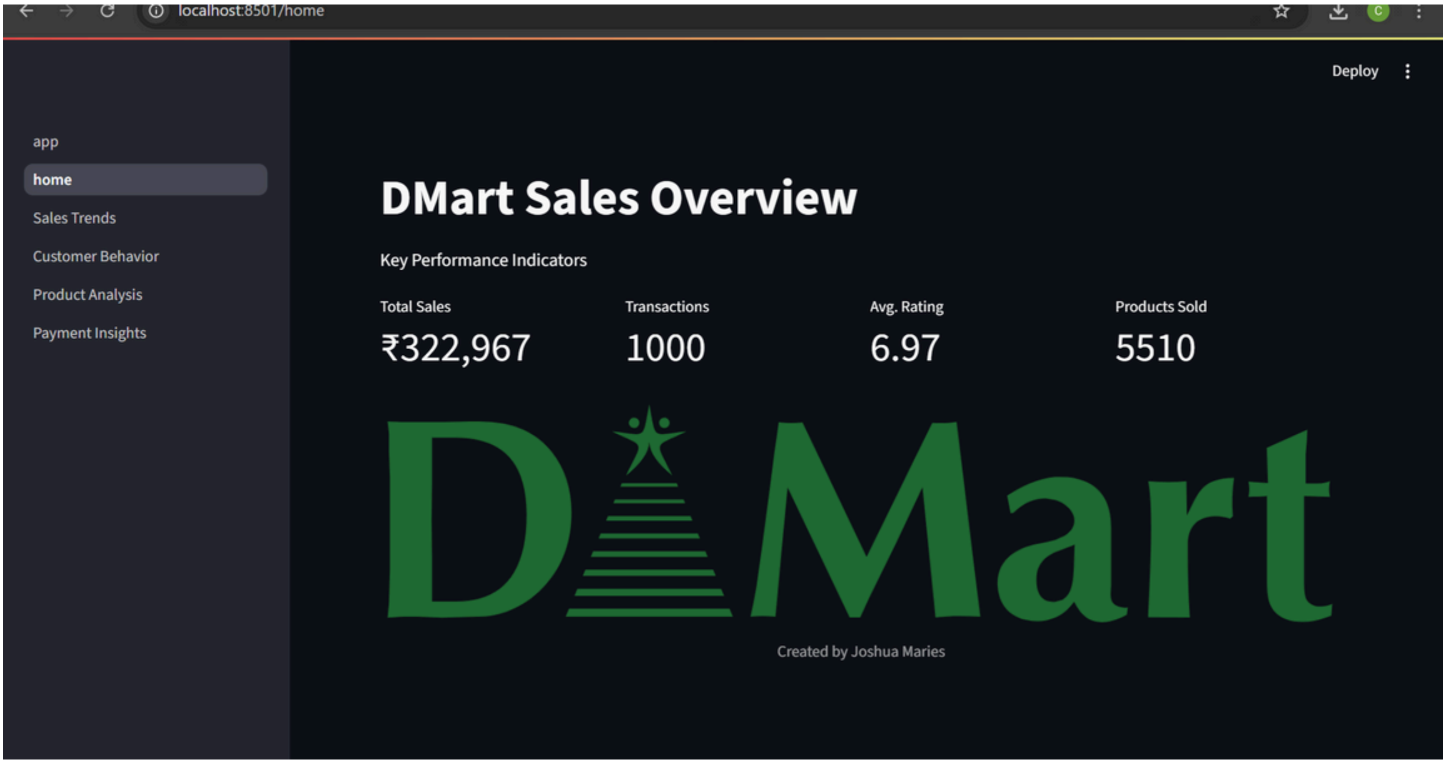


Welcome to the DMart Sales Analytics Dashboard

Welcome to the **DMart Sales Analysis** dashboard! This dashboard is designed to provide in-depth insights into customer behavior, product trends, payment preferences, and more. DMart is one of India's most successful and efficient retail supply chains, known for its streamlined operations, cost-effective sourcing, and deep understanding of consumer behavior.Studying DMart's supply chain model is important because it demonstrates how operational excellence and data-driven decisions can lead to consistent profitability and customer loyalty. Analyzing their data helps uncover key insights into sales trends, customer preferences, and inventory management strategies that can be applied across the retail industry. Use the navigation sidebar to explore:

- Sales Trends
- Customer Behavior
- Product Analysis
- Payment Insights

Home page



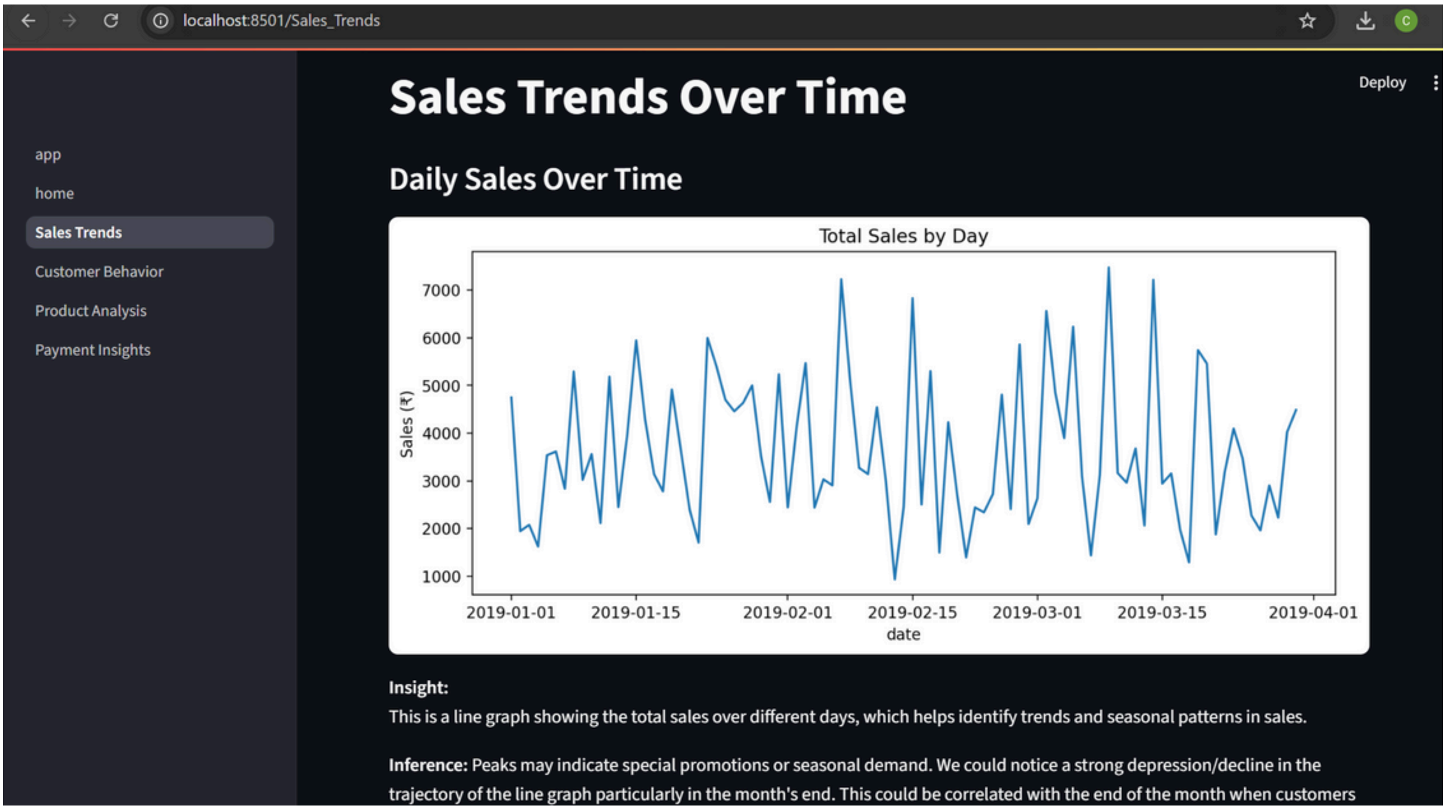
DMart Sales Overview

Key Performance Indicators

Total Sales	Transactions	Avg. Rating	Products Sold
₹322,967	1000	6.97	5510

Created by Joshua Maries

- **Sales trend page**

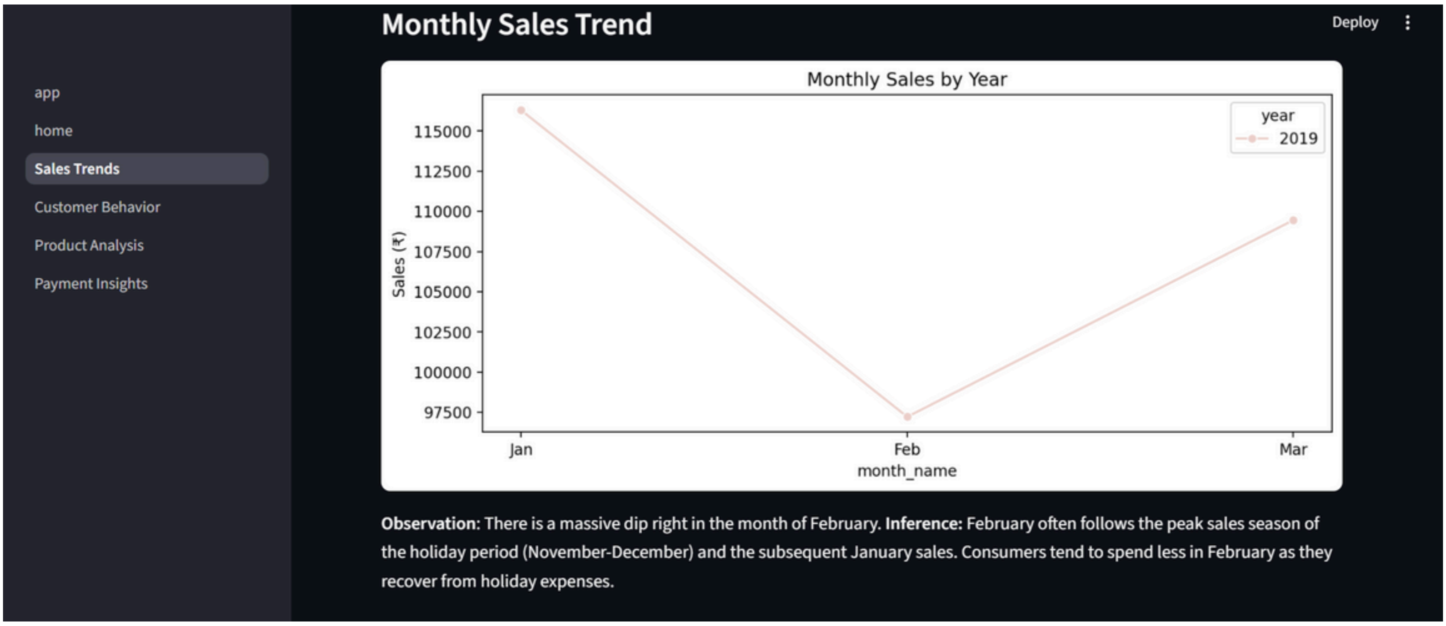


Sales Trends Over Time

Daily Sales Over Time

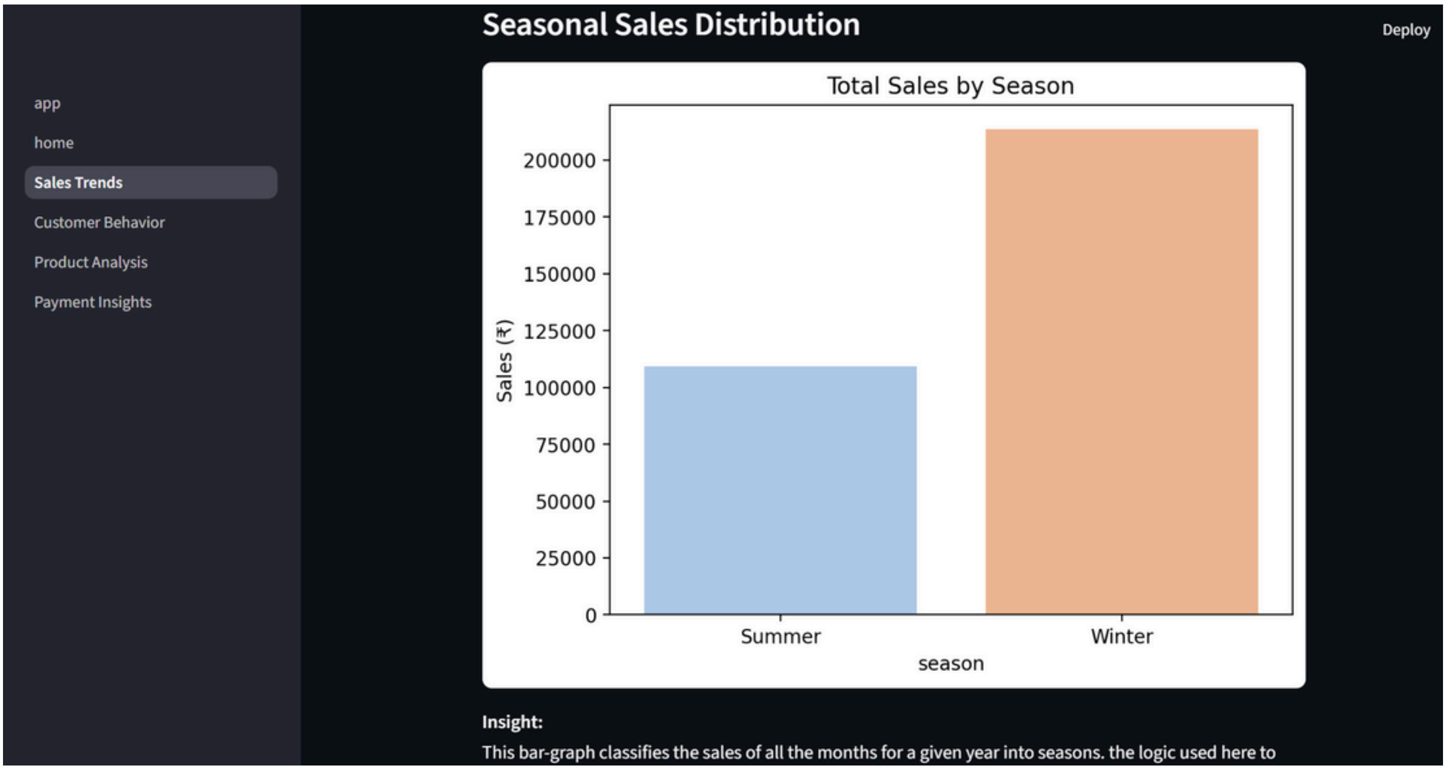
Insight:
This is a line graph showing the total sales over different days, which helps identify trends and seasonal patterns in sales.

Inference: Peaks may indicate special promotions or seasonal demand. We could notice a strong depression/decline in the trajectory of the line graph particularly in the month's end. This could be correlated with the end of the month when customers



Monthly Sales Trend

Observation: There is a massive dip right in the month of February. **Inference:** February often follows the peak sales season of the holiday period (November-December) and the subsequent January sales. Consumers tend to spend less in February as they recover from holiday expenses.

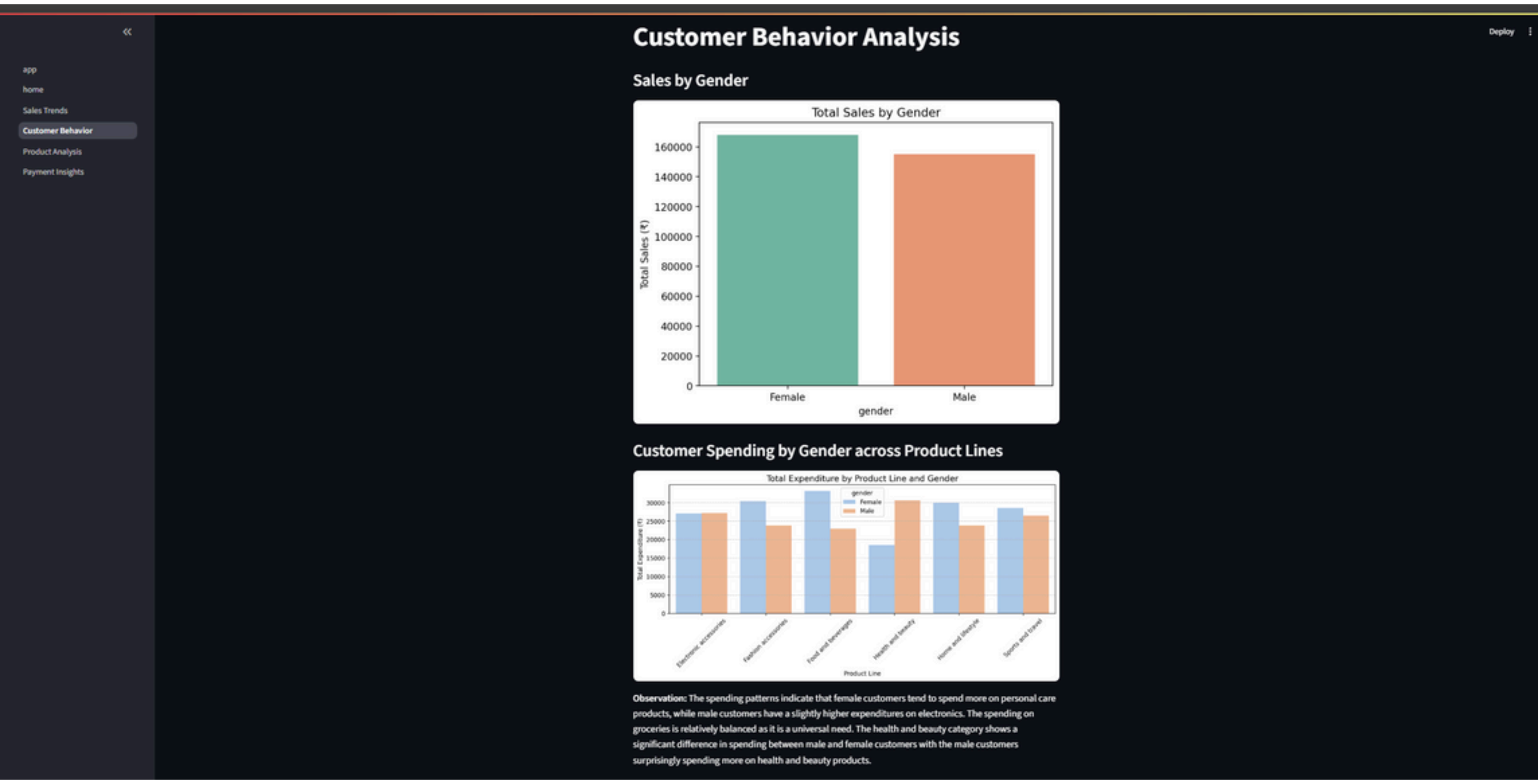


Seasonal Sales Distribution

Insight:

This bar-graph classifies the sales of all the months for a given year into seasons. the logic used here to

Customer Behavior

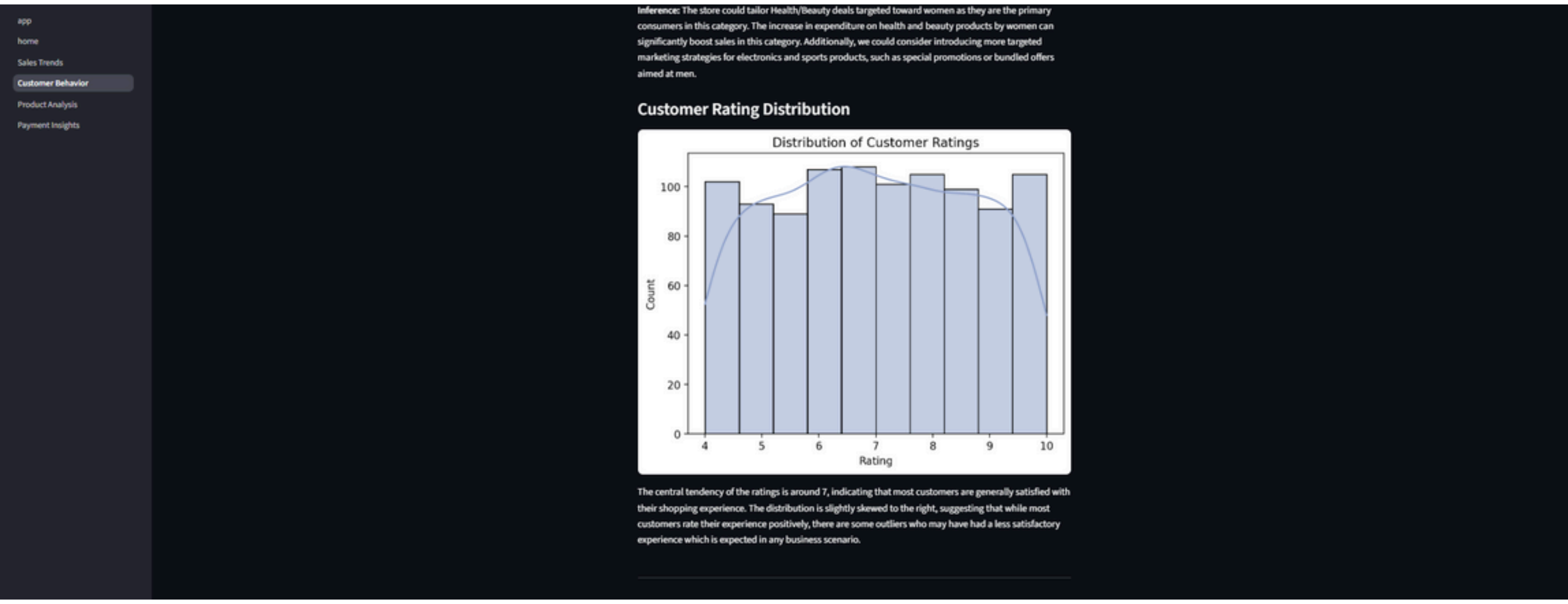


Customer Behavior Analysis

Sales by Gender

Customer Spending by Gender across Product Lines

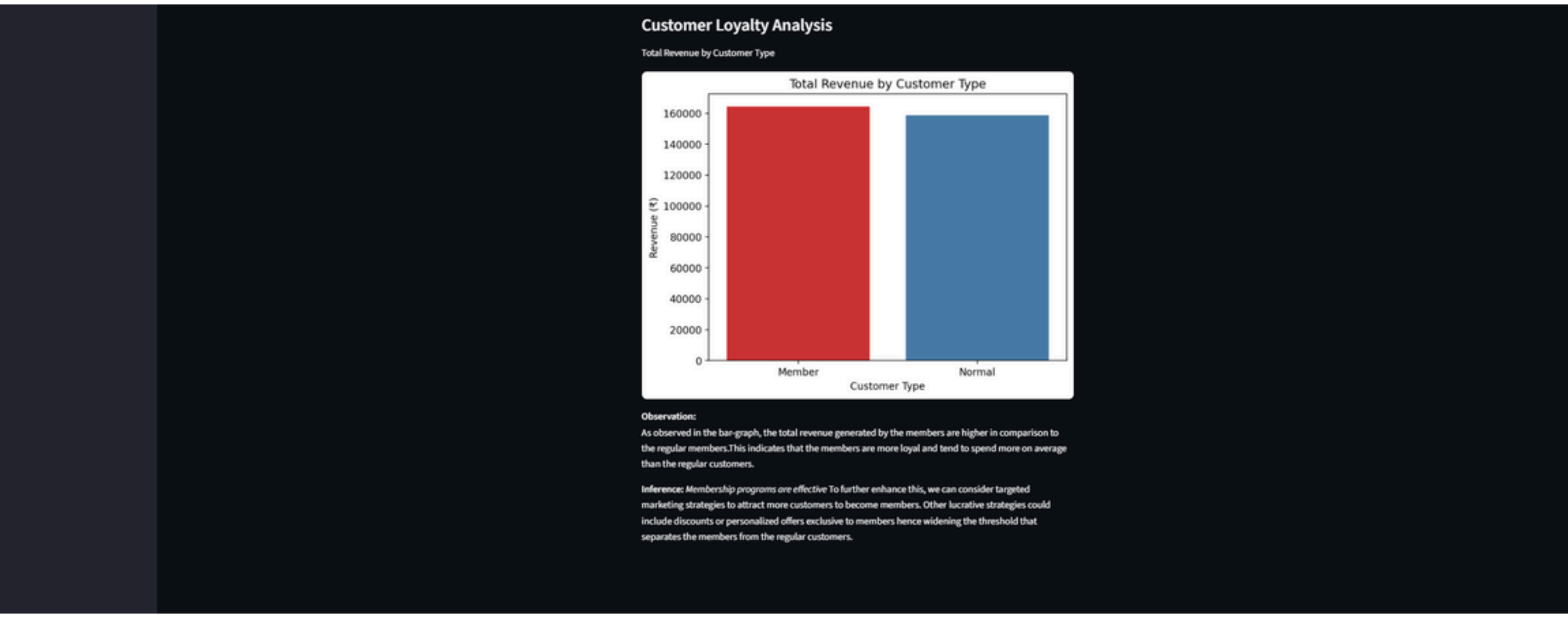
Observation: The spending patterns indicate that female customers tend to spend more on personal care products, while male customers have a slightly higher expenditures on electronics. The spending on groceries is relatively balanced as it is a universal need. The health and beauty category shows a significant difference in spending between male and female customers with the male customers surprisingly spending more on health and beauty products.



Inference: The store could tailor Health/Beauty deals targeted toward women as they are the primary consumers in this category. The increase in expenditure on health and beauty products by women can significantly boost sales in this category. Additionally, we could consider introducing more targeted marketing strategies for electronics and sports products, such as special promotions or bundled offers aimed at men.

Customer Rating Distribution

The central tendency of the ratings is around 7, indicating that most customers are generally satisfied with their shopping experience. The distribution is slightly skewed to the right, suggesting that while most customers rate their experience positively, there are some outliers who may have had a less satisfactory experience which is expected in any business scenario.



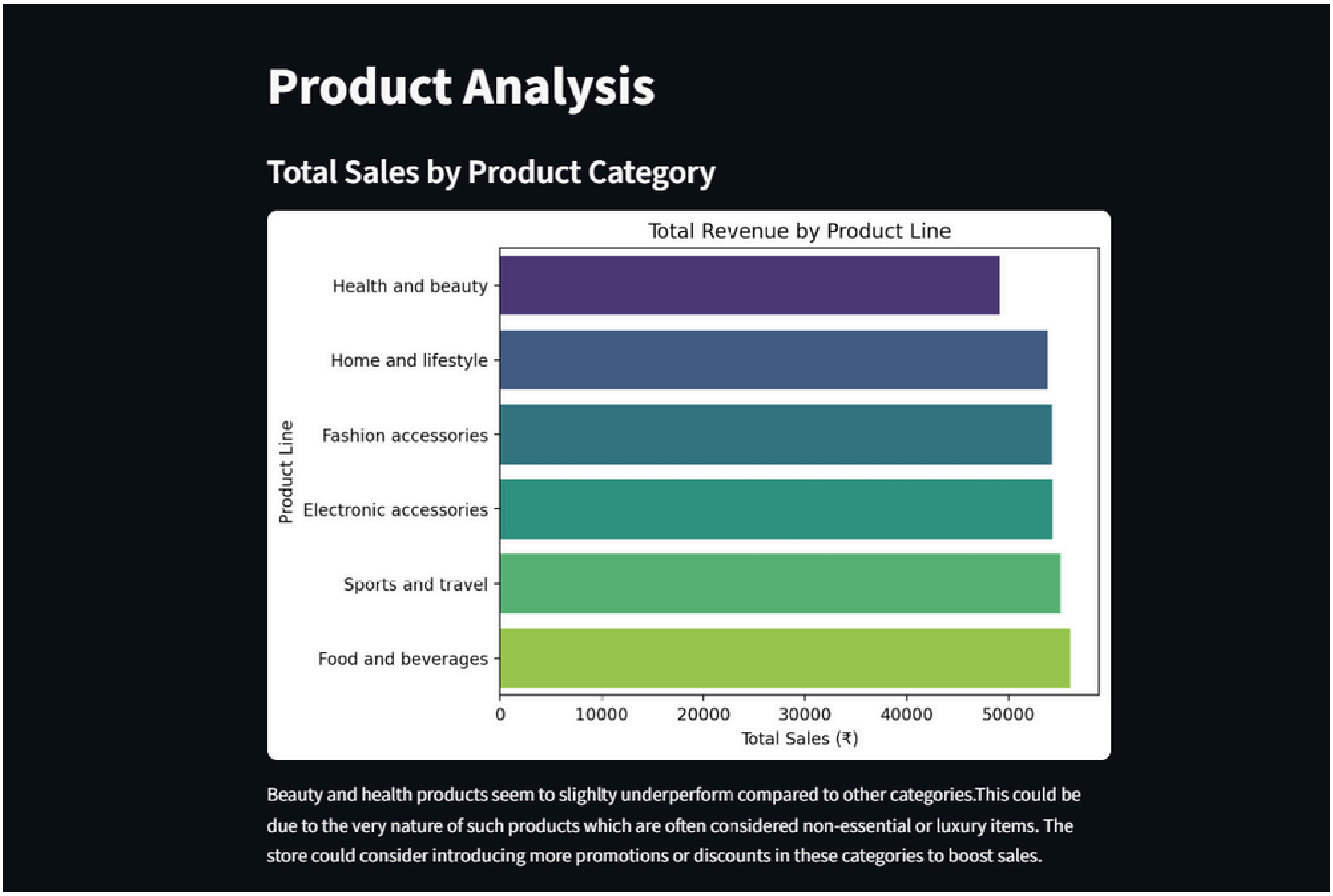
Customer Loyalty Analysis

Total Revenue by Customer Type

Observation:
As observed in the bar-graph, the total revenue generated by the members are higher in comparison to the regular members.This indicates that the members are more loyal and tend to spend more on average than the regular customers.

Inference: *Membership programs are effective* To further enhance this, we can consider targeted marketing strategies to attract more customers to become members. Other lucrative strategies could include discounts or personalized offers exclusive to members hence widening the threshold that separates the members from the regular customers.

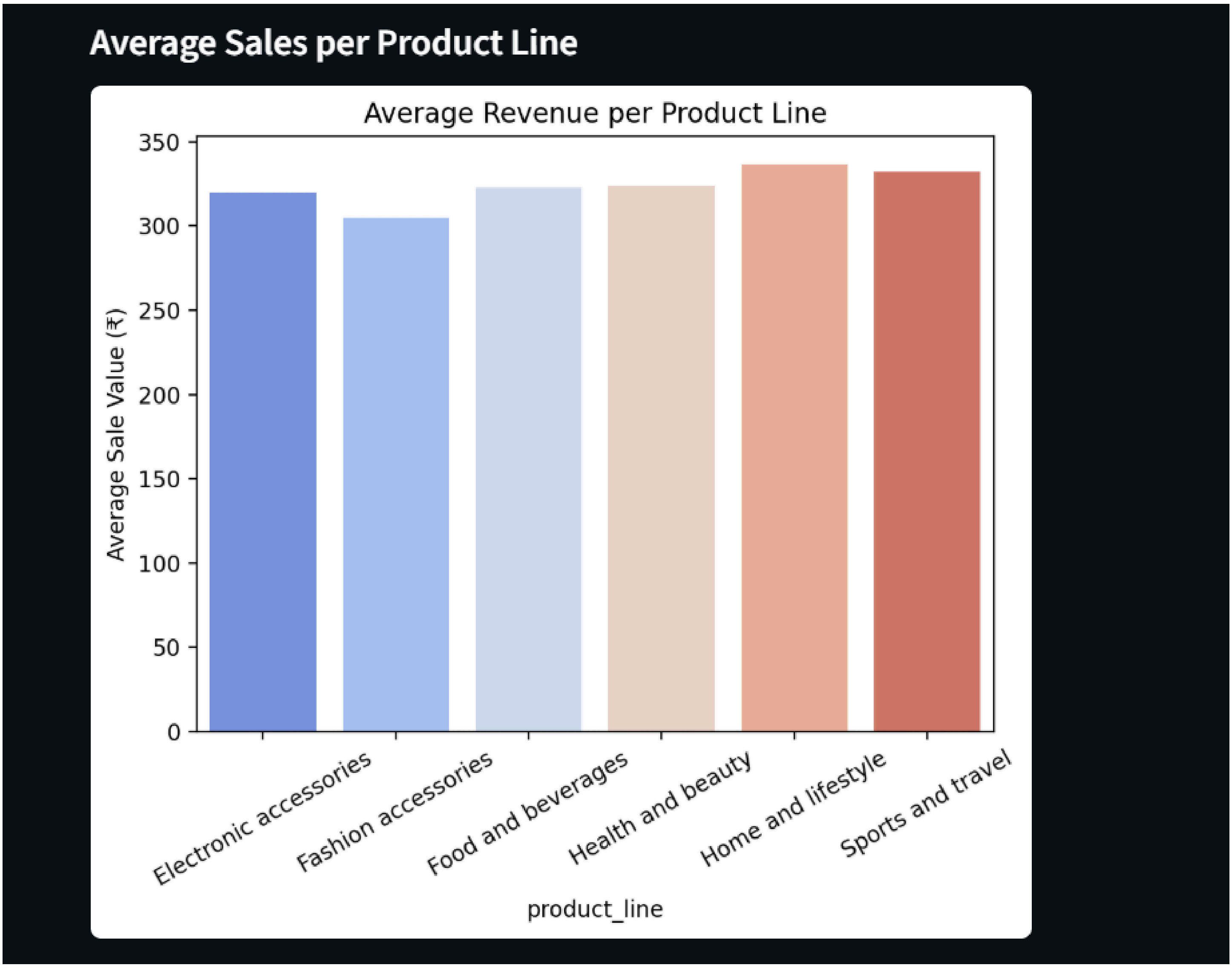
- **Product Analysis**



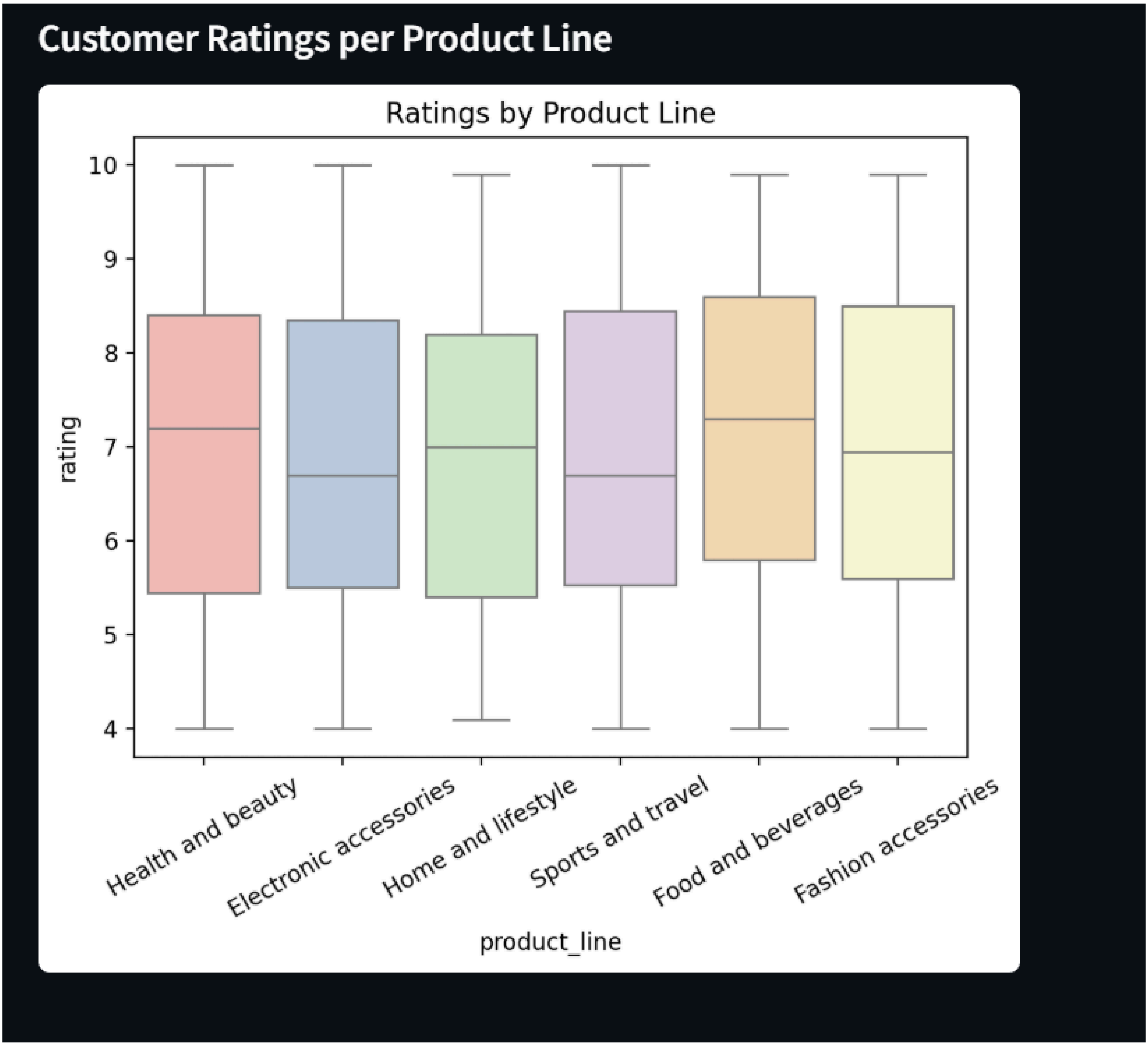
Product Analysis

Total Sales by Product Category

Beauty and health products seem to slighlty underperform compared to other categories.This could be due to the very nature of such products which are often considered non-essential or luxury items. The store could consider introducing more promotions or discounts in these categories to boost sales.



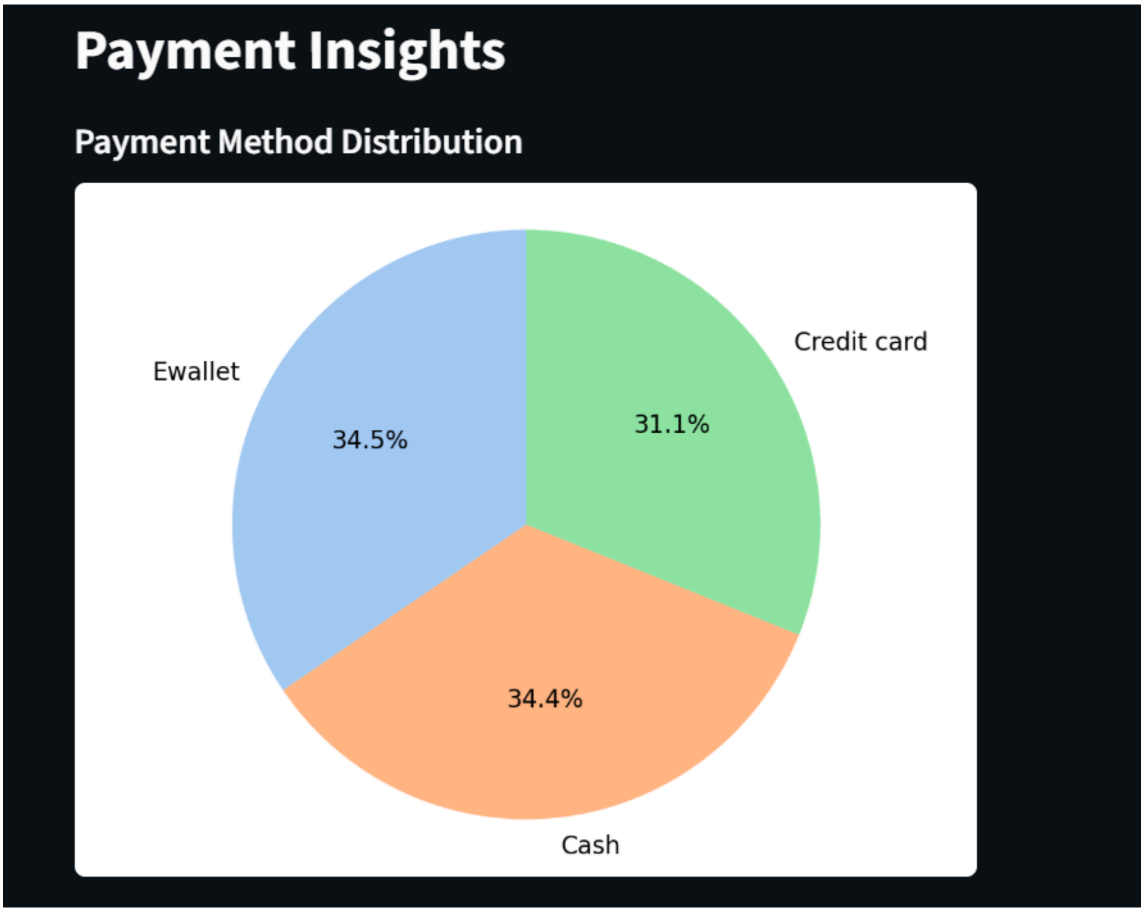
Average Sales per Product Line



Customer Ratings per Product Line

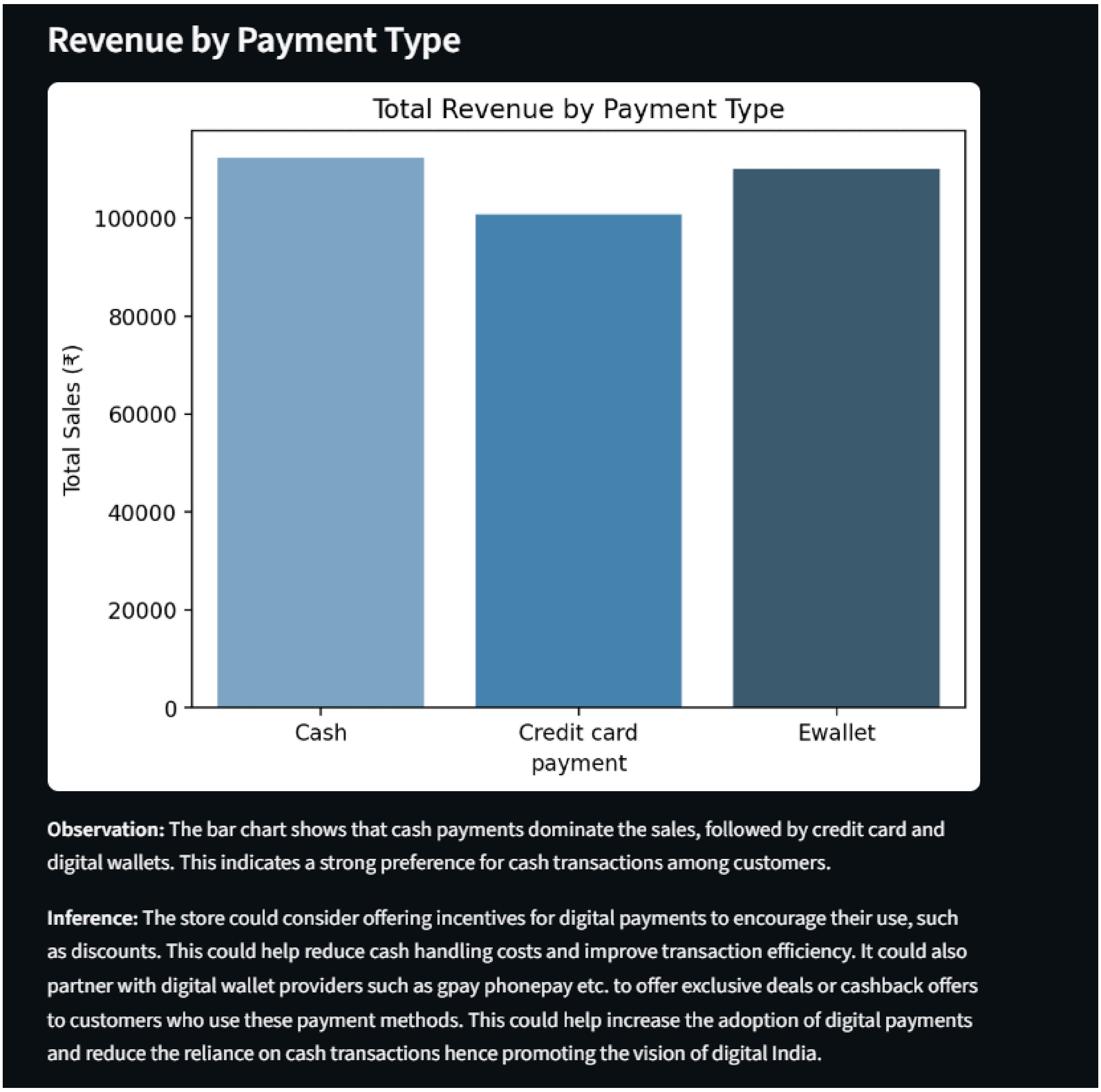
-

- **Payment Insights**



Payment Insights

Payment Method Distribution



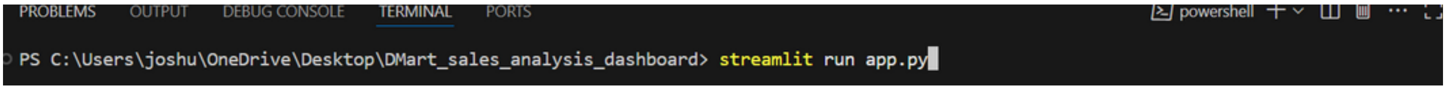
Revenue by Payment Type

Observation: The bar chart shows that cash payments dominate the sales, followed by credit card and digital wallets. This indicates a strong preference for cash transactions among customers.

Inference: The store could consider offering incentives for digital payments to encourage their use, such as discounts. This could help reduce cash handling costs and improve transaction efficiency. It could also partner with digital wallet providers such as gpay phonepay etc. to offer exclusive deals or cashback offers to customers who use these payment methods. This could help increase the adoption of digital payments and reduce the reliance on cash transactions hence promoting the vision of digital India.

Instructions on how to run it

- Make sure you have Python 3.8+ installed. Then, install the required Python libraries: (pip install -r requirements.txt) → run this in the vscode terminal
- Unzip the project folder and ensure the correct order (refer to the file structure)
- using the vscode terminal, run the streamlit command -(streamlit run app.py)



```
PS C:\Users\joshu\OneDrive\Desktop\DMart_sales_analysis_dashboard> streamlit run app.py
```

Closure

The DMart Sales Dashboard project successfully delivers a comprehensive, interactive, and user-friendly platform for exploring and analyzing retail sales data. It allows business users and analysts to:

- Track sales performance trends across time.
- Understand customer behavior based on gender, time, and product preferences.

- Evaluate product line performance by season, quantity, and revenue.
- Analyze payment method usage and branch-wise performance.
- Generate custom reports based on specific filters.
- Assess customer loyalty through revenue contribution analysis.

Scalability

- New analytical pages (e.g., customer segmentation, time-series forecasting) can be added easily using Streamlit’s multipage support.
- Modular design allows you to plug in machine learning models or API-driven reports without reworking the existing structure hence less redundancy
- More interactive data visualizations can be added along with features such authorization; giving access only to the stakeholders etc.

Bibliography

- Kaggle dataset - 🌐✅DMART Products Analysis🛒📊
- A study on seasonal effects on revenue generation in online platform - 🌐A study on the impact of seasonal sales on online sh...
- DMart Corporate Website - https://www.dmartindia.com/ For understanding the business context of DMart and its retail model.