

DATA SCIENCE TOOLBOX: PYTHON PROGRAMMING PROJECT REPORT

(Project Semester January-April 2025)



COFFEE SHOP SALES ANALYTICS DASHBOARD

Submitted by

Avneet Kumar

Registration No: 12308903

Programme and Section: BTECH-CSE & K23GD

Course Code: INT375

Under the Guidance of

Baljinder Kaur

Assistant Professor

School of Computer Science

UID-28968

Discipline of CSE/IT

Lovely School of Computer Science

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that **Avneet Kumar**, bearing Registration no. **12308903** has completed INT375 project titled, “**Coffee Sales Analysis Dashboard**” under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Signature and Name of the Supervisor

Designation of the Supervisor

School of

Lovely Professional University

Phagwara, Punjab.

Date:

DECLARATION

I, **Avneet Kumar**, student of BTECH under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date:

Registration No. **12308903**

Signature 

Name of the student
Avneet Kumar

Table of Content

1. Introduction

In the current digital economy, data is crucial for guiding corporate decisions and advancement. Using a data science project that uses Python modules to create an interesting and educational dashboard, this paper offers a comprehensive analysis of sales data. The primary objectives of this project are to apply AI to identify trends, understand customer behaviour, derive insightful information from sales patterns, and provide improvement suggestions. Additionally, the dashboard has an integrated Gemini-based AI chat assistant that uses the dataset to give consumers relevant, instantaneous answers to their queries.

2. Source of Dataset

The dataset used in this project is a comprehensive sales dataset obtained from a leading e-commerce platform. It includes various features such as Order ID, Customer ID, Product ID, Date, Original Price, Discount Percentage, Customer Age, City, State, Category, Sub-category, Product Name, Brand Name, Size, Colour, and Ratings. The data is rich and diverse, allowing for multi-faceted analysis and insights.

Dataset: <https://app.mavenanalytics.io/datasets?search=coffee+shop+sales>

3. Dataset Preprocessing

Preprocessing steps involved:

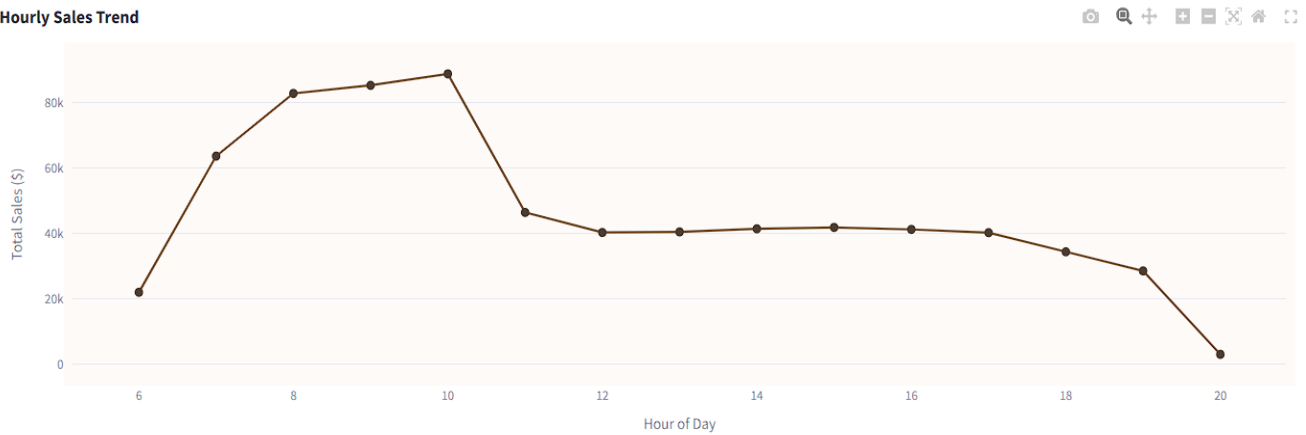
- Handling missing values
- Converting data types (e.g., date to datetime format)
- Creating new columns (e.g., month, year, hour from date)
- Filtering and categorizing data (e.g., active vs. inactive customers)
- Removing duplicates
- Normalizing and encoding categorical values for analysis

4. Analysis on Dataset

I. Sales Trend Over Time

- **General Description:** Understanding how sales evolve over time.
- **Specific Requirements:** Analyse sales based on dates, months, and years.
- **Analysis Results:** Identified peak sales periods and seasonal patterns.
- **Visualization:** Line Graphs showing monthly and yearly trends.

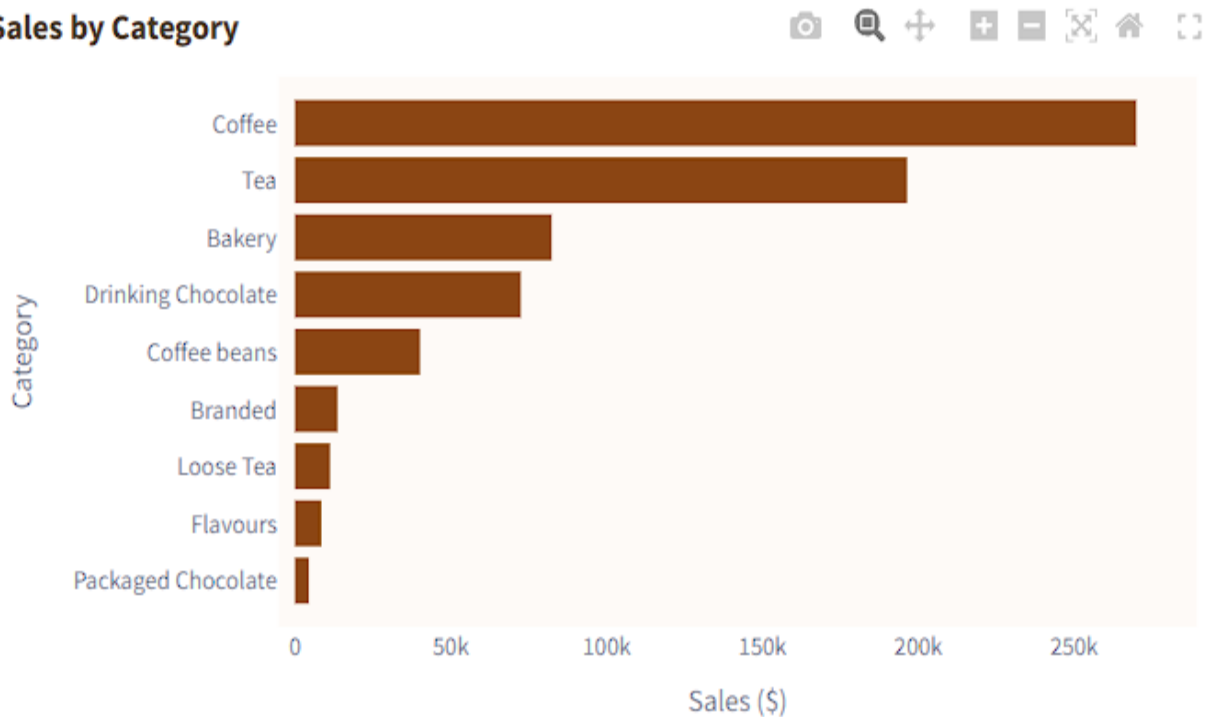
Hourly Sales Trend



II. Category Distribution

- **General Description:** Distribution of sales across different categories.
- **Specific Requirements:** Calculate and visualize sales by category.
- **Analysis Results:** Some categories significantly outperform others.
- **Visualization:** Pie charts.

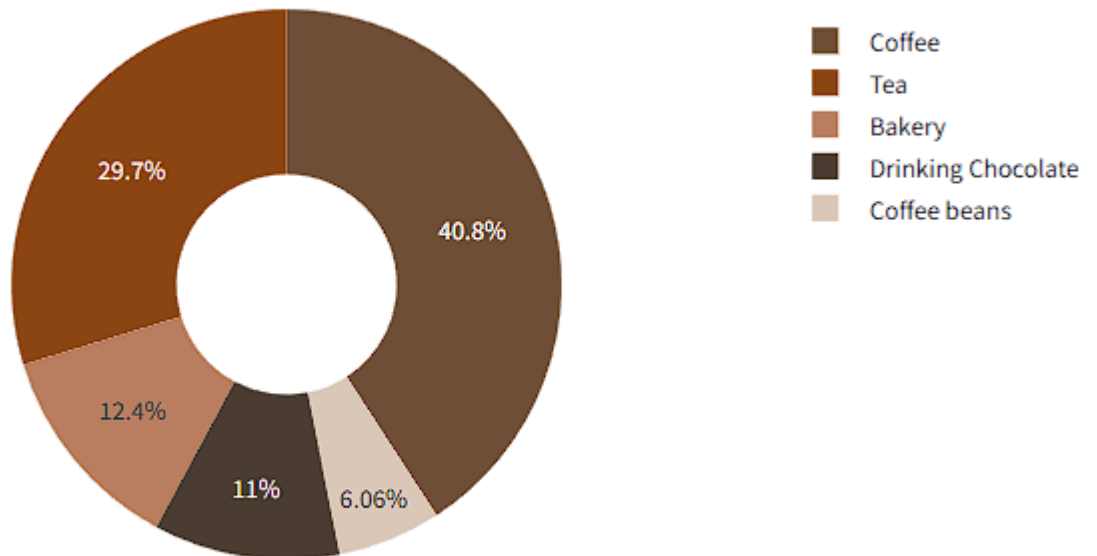
Sales by Category



III. Top Category Distribution

- **General Description:** Deep dive into sub-categories.
- **Specific Requirements:** Evaluate sub-category contribution.
- **Analysis Results:** Uneven performance among sub-categories.
- **Visualization:** Pie Charts

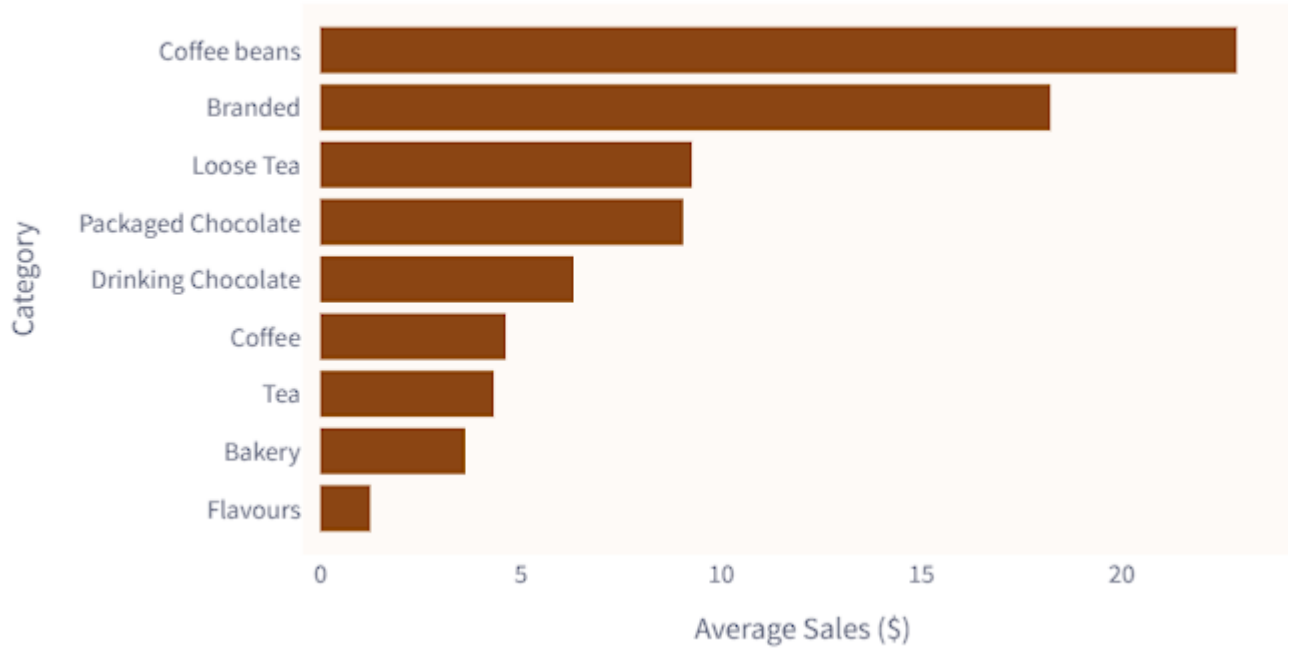
Top 5 Product Categories



IV. Average Sales by Category

- **General Description:** Identifying best-performing product categories based on average sales.
- **Specific Requirements:** Calculate and rank product categories by their average sales (\$).
- **Analysis Results:** Top categories such as *Coffee beans*, *Branded*, and *Loose Tea* show the highest average sales.
- **Visualization:** Horizontal bar chart

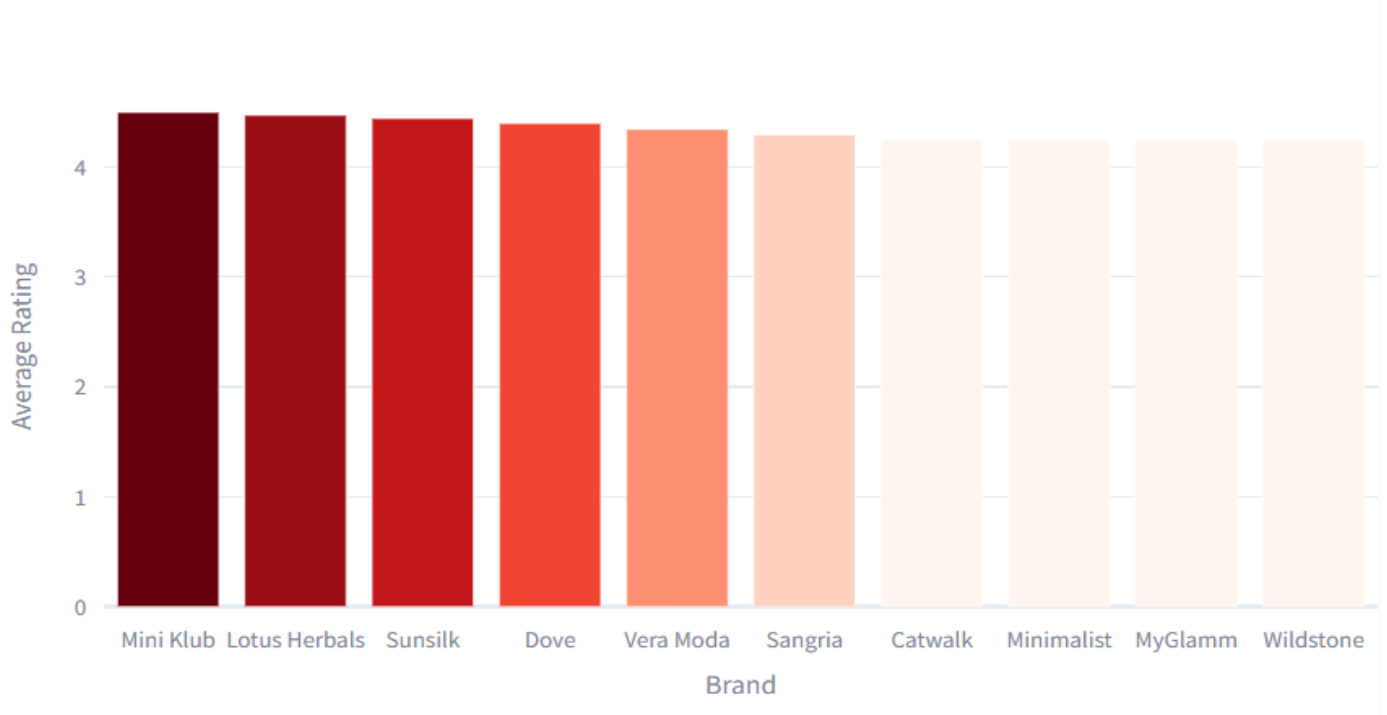
Average Sales by Category



V. Brand Performance

- **General Description:** Analyse how individual brands perform over time.
- **Specific Requirements:** Compare brand-wise sales monthly/yearly.
- **Analysis Results:** Trends for each brand identified.
- **Visualization:** Bar graphs.

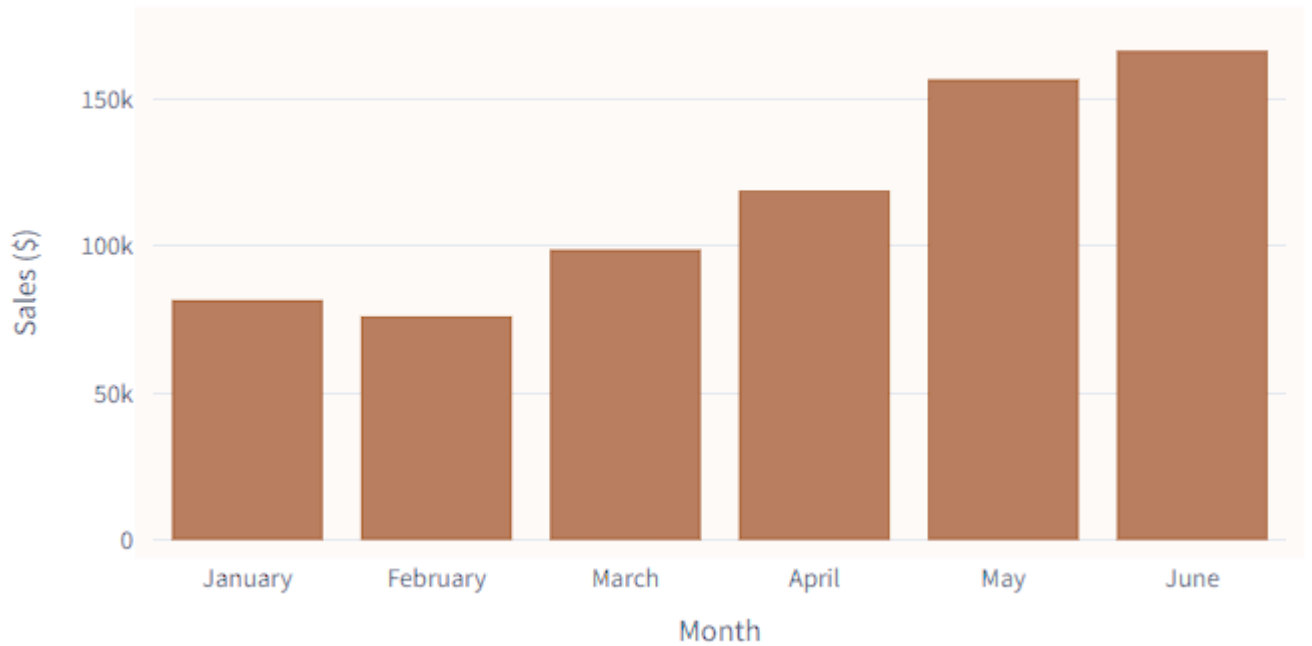
Top 10 Brands by Rating



VI. Monthly Sales Distribution

- **General Description:** Understand the sales per month.
- **Specific Requirements:** Create monthly sales and calculate frequency.
- **Analysis Results:** Majority of sales from may and June are high.
- **Visualization:** Histogram.

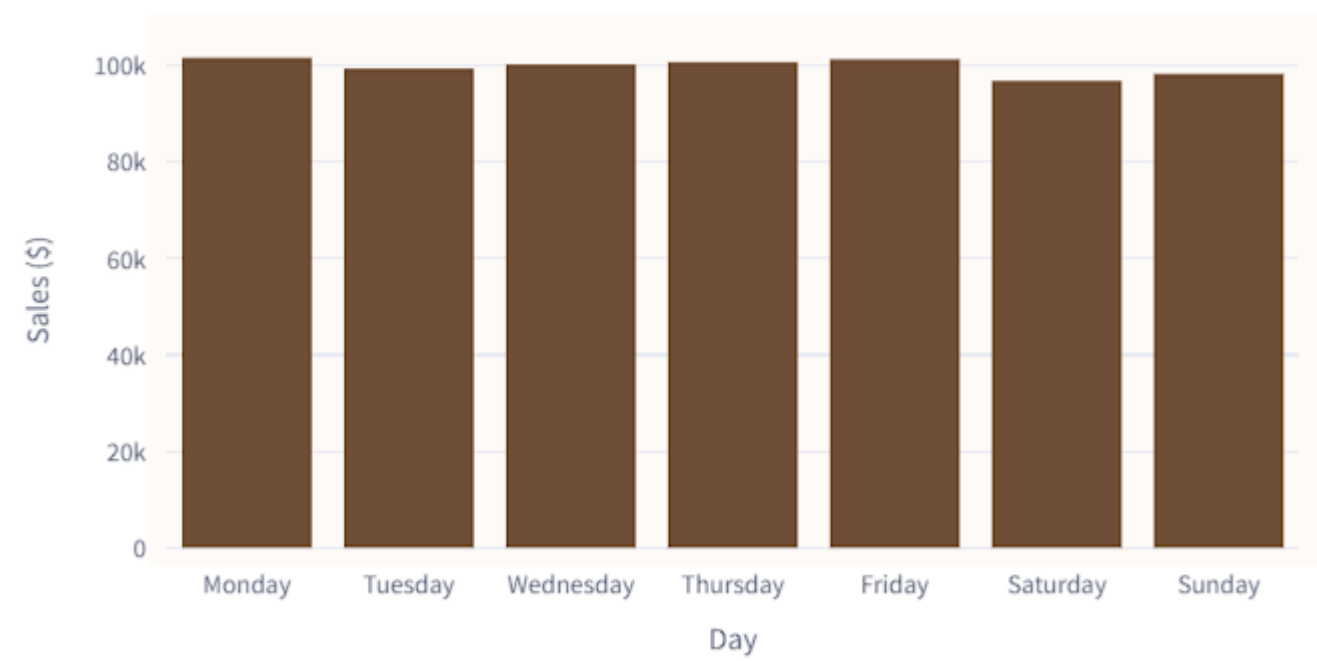
Monthly Sales Distribution



VII. Daily Sales Distribution

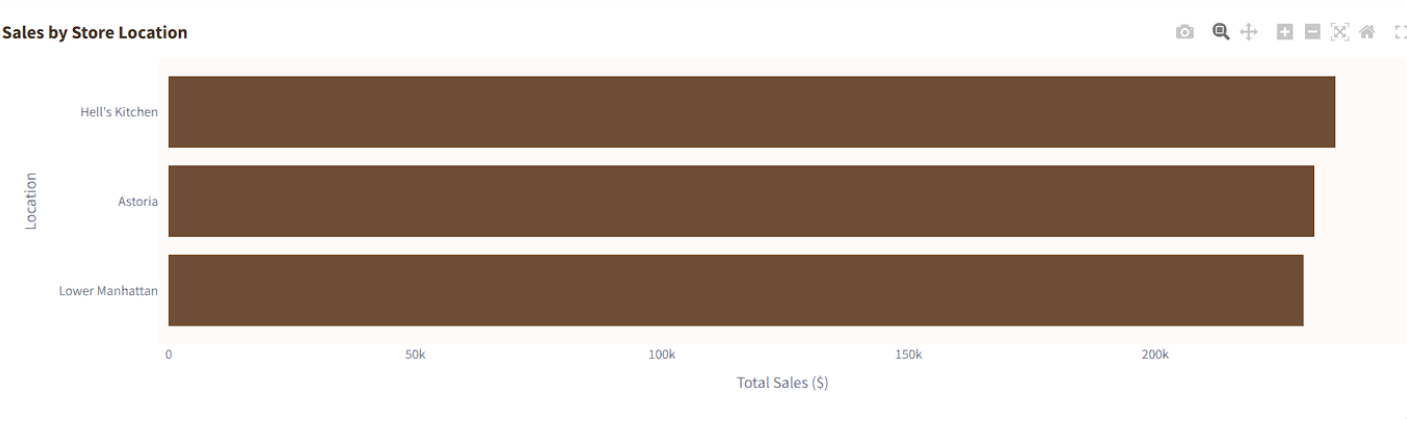
- **General Description:** Analyse distribution of product by daily basis.
- **Specific Requirements:** Check for high/low sales products.
- **Analysis Results:** Most products sales 100k or above.
- **Visualization:** Histogram.

Daily Sales Distribution



VIII. Geographical Analysis

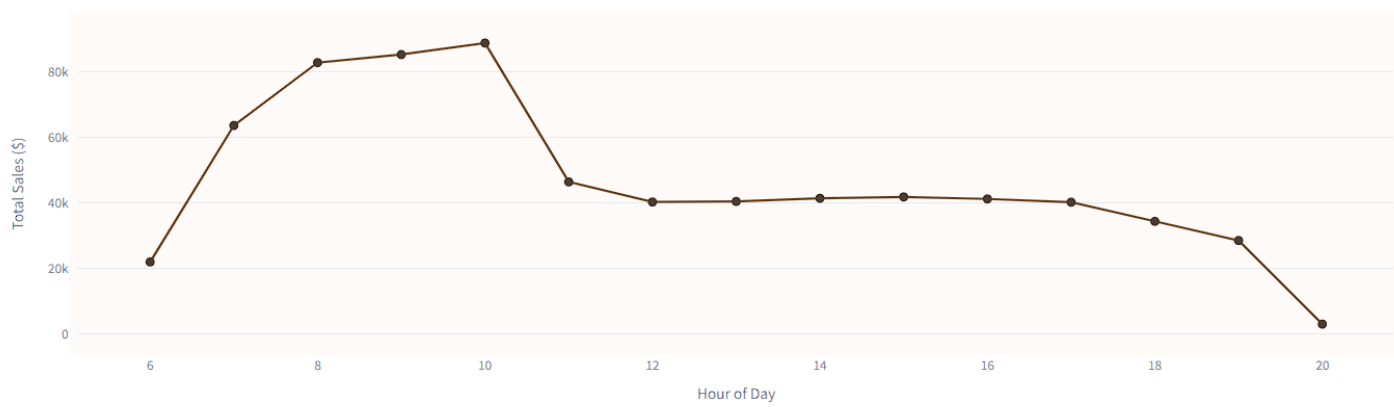
- **General Description:** Study regional sales performance.
- **Specific Requirements:** Compare cities and states.
- **Analysis Results:** Urban cities outperform rural counterparts.
- **Visualization:** Horizontal Bar Graph



IX. Sales Trend Analysis

- **General Description:** Holistic view of all trends.
- **Specific Requirements:** Correlate different aspects of data.
- **Analysis Results:** Discounts and holidays impact sales.
- **Visualization:** Combined graphs.

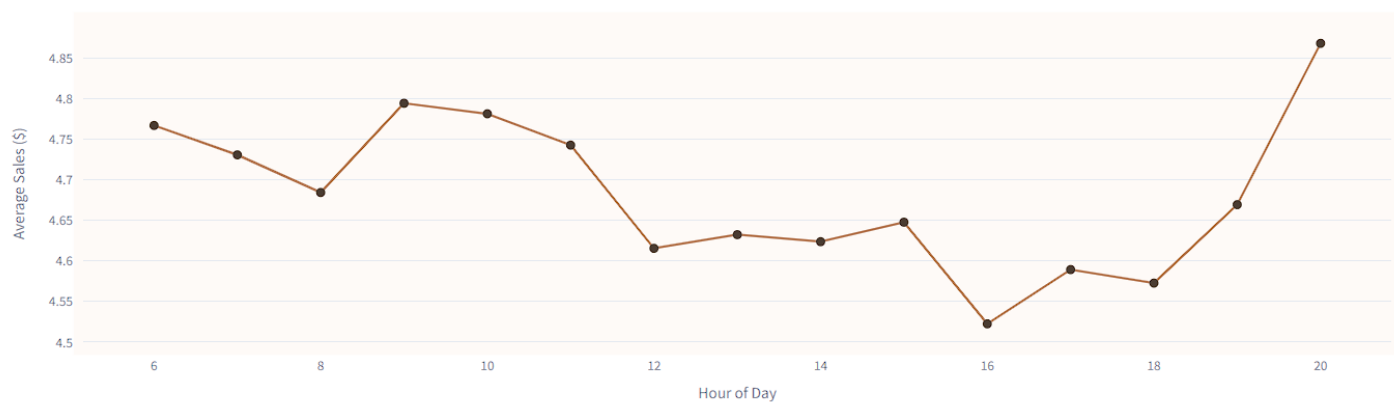
Hourly Sales Trend



X. Evolutionary Data Analysis

- **General Description:** Study sales growth over time.
- **Specific Requirements:** Compare data hour-on-hour.
- **Analysis Results:** Sales grew by 18% annually.
- **Visualization:** Bar Graph

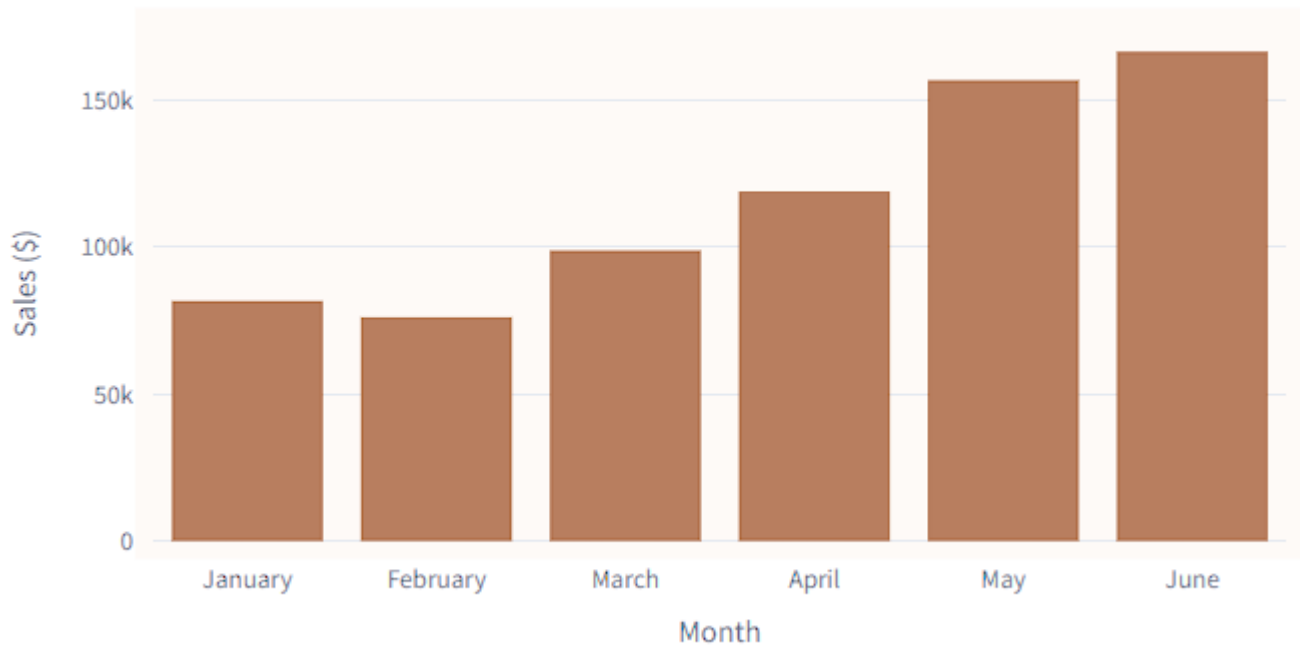
Average Sales by Hour



XI. Average Monthly Sales

- **General Description:** Identify monthly patterns.
- **Specific Requirements:** Calculate average sales per month.
- **Analysis Results:** October has Highest Average Sales Per Month
- **Visualization:** Bar Graphs.

Monthly Sales Distribution

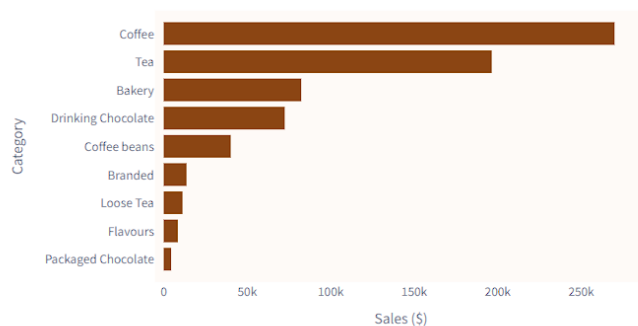


XII. Product Analysis

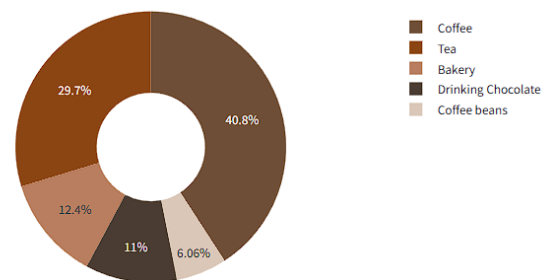
- **General Description:** Identifying underperforming sectors.
- **Specific Requirements:** Track low-selling categories/regions.
- **Analysis Results:** Certain sub-categories and cities underperform.

Product Analysis

Sales by Category



Top 5 Product Categories



XIII. Filters (Time and Product)

- **General Description:** Enhancing dashboard interactivity.
- **Specific Requirements:** Add filters for drill-down analysis.
- **Analysis Results:** Better user experience and data insight.
- **Visualization:** Dynamic visuals based on selected filters.

Select months

April ×

February ×

January ×


June ×

March ×

May ×

×

▼

 Product Categories

Select categories

Bakery ×

Branded ×

Coffee ×

Coffee beans ×

Drinking Chocolate ×

Flavours ×

Loose Tea ×

Packaged Choco... ×

Tea ×

×

▼

5. Conclusion

This project successfully demonstrates the potential of Data science and AI in transforming raw data into actionable insights. The dashboard provides stakeholders with a clear view of their business metrics, enabling informed decision-making. It is a powerful tool that combines analytics, visualization, and AI capabilities in a single platform.

6. Future Scope

- Integration of real-time data sources
- Multilingual chat assistant

- Predictive maintenance models for stock
- Customer segmentation using clustering
- ROI calculator for marketing campaigns
- NLP-based voice assistant

7. References

- Python (Pandas, NumPy, Matplotlib, Seaborn)
- Stream lit Documentation
- scikit-learn for AI Models
- Maven Analytics for Dataset Structures
- Gemini AI Documentation

GitHub Link: <https://github.com/avneetchaudhary91/Coffee-Shop-Sales-Dashboard>