{Autumn Internship Project Report}

Visualizing Internship Program 2025

Preprocessing and visualizing Coffee sales data

SUMIT HALDER

SECTION 1

BANGABASI COLLEGE

Period of Internship: 25th August 2025 - 28th September 2025

Report submitted to: IDEAS – Institute of Data Engineering, Analytics and Science Foundation, ISI Kolkata

1. Abstract

This project looks at how to use a retail coffee dataset to analyze sales data. The dataset keeps track of sales across products, customers, and areas, which helps us understand trends in revenue, customer preferences, and how well products are doing. We found seasonal buying patterns, the best-selling categories, and the revenue distribution by region through data preprocessing, exploratory analysis, and visualization. Line charts and bar graphs are examples of visualizations that showed changes in customer demand and sales contribution. The results show how Python-based data analysis can help businesses in the retail sector make better decisions by finding useful information.

2. Introduction

Sales data is important for figuring out how well a business is doing, how customers act, and when money comes in and goes out. We looked at a coffee sales dataset for this project to find patterns in sales and performance indicators. The main goal was to learn about monthly revenue trends, popular products, where sales were happening in different parts of the country, and how much customers were contributing. The analysis gives retail and e-commerce companies a starting point for making decisions based on data.

Relevance

- Sales analytics help businesses make the most of their inventory, marketing, and prices.
- Knowing what customers and areas like helps businesses make better targeted plans.
- Coffee sales are a good example of how consumer demand changes with the seasons.

Technology Used

- NumPy for data manipulation and Python with Pandas.
- For visualization, used Seaborn and Matplotlib.
- The coding environment is Jupyter Notebook.

Procedure

- loading and examining of data.
- Cleaning and formatting include classifying variables, converting dates, and dealing with missing values.

- exploratory analysis including identifying trends, high-revenue products, and regional sales contributions.

3. Project Objective

- To analyze overall revenue trends in coffee sales.
- To identify the top-performing products and categories.
- To explore region-wise and customer-wise sales contributions.
- To visualize seasonal or monthly demand fluctuations.
- To highlight actionable insights for business improvement.

4. Methodology

1) **Data Collection**
    a. Coffee sales dataset provided during internship (Excel/CSV format).
2) **Data Preprocessing**
    a. Converted date columns to datetime.
    b. Extracted Month, Year, and Day for trend analysis.
    c. Handled duplicate and missed values.
3) **Visualization Techniques**
    a. **Line Charts:** Monthly revenue and sales quantity.
    b. **Bar Graphs:** Top customers and product performance.
    c. **Grouped Analysis:** Seasonal trends across months and regions.
4) **Tools Used**
    a. NumPy, Matplotlib, Python, Pandas, and Seaborn.

5. Data Analysis and Results

Part A: Coffee Sales Trends

- **Line Plot:** Revenue over time shows consistent growth with seasonal fluctuations.
- **Monthly Trends:** Certain months show higher demand, reflecting seasonal effects.

- **Grouped Analysis:** Clearly defined sales cycles with year-to-year variations.

Part B: Product and Customer Insights

- **Top Products:** Espresso and Arabica blends generated the highest revenue.
- **Customer Analysis:** A small number of valuable clients made a substantial contribution to overall sales.
- **Bar Chart:** Information about customer loyalty was obtained by ranking the top 10 clients by revenue.

Part C: Regional Distribution

- **Bar Chart:** Sales varied by region, with [Region X] contributing the maximum revenue share.
- Regional comparisons showed opportunities for expansion in underperforming markets.

Conclusion

According to the analysis, coffee sales had discernible peaks during months with high demand and followed distinct seasonal cycles. Products with an Arabica base became the most popular, and a small number of devoted clients generated a sizable amount of income. Uneven distribution was revealed by regional analysis, indicating room for growth in underperforming markets. All things considered, the project showed how well Python-based data analytics can produce useful business insights from sales data.

6. APPENDICES

DATA SOURCE: Coffe_sales.csv

PROJECT LINK:

https://colab.research.google.com/drive/19ag4eEFuQu6aFUNUuCr2quaq5BvYu6Dh?usp=sharing

GITHUB LINK:

https://github.com/sumit-halder12/preprocessing-and-Visualizing-coffee-sales-data-.git