☕ Coffee Sales Analysis – Detailed Project Report

---

1. Introduction

The coffee industry is a massive global market that continues to expand as consumers develop a taste for premium and specialty beverages. For any coffee business to stay competitive, data-driven insights are crucial. By analyzing sales data, companies can track product performance, monitor regional demand, optimize pricing, and identify the best times to market products.

This project focuses on a dataset containing sales records of various coffee products. The goal is to explore the data, clean it for usability, analyze trends, visualize key metrics, and derive actionable business insights. The outcome of this analysis can be used to inform strategic decisions such as inventory planning, marketing campaigns, and product development.

---

2. Project Objective

The primary objectives of the Coffee Sales Analysis project are:

1. To explore and clean the coffee sales dataset to ensure quality and consistency.
2. To identify trends in sales over time (monthly, quarterly, yearly).
3. To analyze sales performance across different regions.
4. To determine which coffee products contribute most to revenue.
5. To understand pricing, quantity sold, and profitability across products.
6. To extract meaningful insights that can support better business decisions.

---

3. Dataset Description

The dataset includes several columns that are essential for understanding and analyzing sales. Typical columns include:

Date – The transaction or order date.

Product – The name or category of the coffee product (e.g., Espresso, Latte).

Region – The geographical area where the sale occurred.

Units Sold – Number of items sold per transaction or per entry.

Unit Price – Price per unit for the coffee product.

Total Revenue – The calculated revenue from the sale (Units Sold × Unit Price).

Each row represents a transaction or a summary of sales over a defined time period (daily, weekly, or monthly).

---

4. Tools and Technologies Used

This analysis was conducted using the following tools and libraries:

Python – The primary programming language for data analysis.

Pandas – For loading, cleaning, transforming, and summarizing data.

NumPy – For numerical operations and calculations.

Matplotlib and **Seaborn** – For creating visualizations and graphs.

Jupyter Notebook or **Google Colab** – Used as the interactive coding and documentation environment.

---

5. Data Cleaning and Preparation

Before performing any meaningful analysis, the raw dataset was cleaned and prepared. Key steps included:

✅ Missing Values

1. Rows with critical missing data were either removed or filled using imputation techniques.
2. Empty cells in Unit Price or Units Sold were filled using median values or removed if insufficient data.

✅ Data Formatting

The Date column was converted into datetime format to allow time-series analysis.

Additional columns like Month, Quarter, and Year were derived for trend analysis.

✅ Revenue Calculation

If not already present, a Total Revenue column was calculated as:

Total Revenue = Units Sold × Unit Price

✅ Category Unification

Standardized names of products and regions to avoid duplication (e.g., “LATTE” and “Latte” unified).

✅ Duplicate Removal

Ensured there were no duplicated entries that could distort analysis.

---

6. Exploratory Data Analysis (EDA)

The cleaned dataset was analyzed using a variety of visual and statistical techniques.

□ A. Sales Over Time

1. Time-series plots revealed a clear upward trend in coffee sales over months.
2. December showed a significant increase in revenue—likely due to holidays.

□ B. Regional Performance

1. Sales were grouped by Region to identify top-performing areas.
2. The **West and Central** regions consistently generated the highest revenue.
3. Underperforming regions could benefit from targeted marketing campaigns.

□ C. Product-Wise Analysis

Espresso and **Cappuccino** had the highest sales volumes.

Mocha and **Americano** had niche popularity but higher unit prices.

Product demand varied by season—Latte sales increased in colder months.

□ D. Revenue Insights

Histograms and boxplots showed that most transactions were in the mid-revenue range.

Outliers were observed, indicating bulk orders or wholesale purchases.

□ E. Quantity vs. Price

Scatter plots showed that lower-priced products sold in higher quantities.

High unit price did not always correlate with high revenue—volume mattered more.

---

7. Key Findings and Business Insights

From the analysis, the following important findings were discovered:

Espresso generated the highest revenue across all regions and seasons.

Cappuccino was a close second in terms of overall contribution.

December and October were peak months for sales—suggesting strong holiday demand.

The Western region had the best performance in terms of revenue and volume.

Products with mid-range pricing performed the best in terms of both units sold and total revenue.

Customers preferred diversity in product offerings, suggesting benefits from bundling or seasonal offerings.

---

8. Conclusion

This project provided a comprehensive analysis of coffee sales using real-world data. The process included importing and cleaning data, exploring trends and performance, visualizing key metrics, and generating business insights. The findings highlight the importance of seasonal trends, regional targeting, and product portfolio management in driving sales growth.

The insights obtained from this project can help coffee retailers and marketers:

Plan inventory for high-demand seasons.

Launch targeted campaigns in underperforming regions.

Optimize the pricing strategy for different product categories.

Introduce new or bundled offerings based on consumer preference.