BDA Project Report

Analysis of Coffee Sales Data Using Python

Student: N Lasya Priya

B.Tech, Malla Reddy University

Abstract

This report presents a detailed analysis of a Coffee Sales dataset using Python for data-driven insights and visualization. The main objective of this study is to understand customer purchase behavior, identify top-performing coffee products, analyze seasonal sales trends, and provide strategic recommendations for business improvement. Through exploratory data analysis (EDA), visual analytics, and trend evaluation, this project highlights key sales metrics and relationships that can enhance decision-making in the retail domain.

1. Introduction and Domain Description

Coffee is one of the most popular beverages worldwide, and understanding its sales patterns is critical for improving marketing, inventory, and production strategies. This project applies Python-based analytical tools to examine sales performance across different coffee types, sizes, and time periods. The dataset contains transactional information such as date, coffee type, size, quantity sold, and revenue. The study focuses on uncovering sales trends, identifying high-performing products, and analyzing customer behavior patterns. Python's Pandas, NumPy, Matplotlib, and Seaborn libraries are used to clean, analyze, and visualize the data efficiently.

2. Dataset Description

2.1 Source and Files
The analysis is based on the dataset file 'Coffe_sales.csv', which records detailed coffee sales transactions including product category, size, quantity sold, and total revenue. The dataset enables time-series and categorical analysis of sales performance.

2.2 Key Fields
- Date: The date of the transaction.
- Coffee Type: Category of coffee sold (e.g., Latte, Espresso, Cappuccino).
- Size: Product size (Small, Medium, Large).
- Quantity Sold: Units of coffee sold per transaction.
- Revenue: Total revenue generated from the sale.

- Customer Category: Type of customer purchasing the product.

The dataset was analyzed using Python DataFrames and visualized with Matplotlib and Seaborn for better interpretability.

3. Data Preparation and Methodology

The data analysis followed a structured pipeline involving multiple stages:
1. Data Ingestion: The CSV dataset was imported using Pandas.
2. Data Cleaning: Missing values were handled, data types were corrected, and duplicates removed.
3. Feature Engineering: Additional time-based features such as month and season were created for better insights.
4. Descriptive Statistics: Computed summary metrics like mean, max, and sum for sales and revenue.
5. Visualization: Created line plots, bar charts, and heatmaps to represent coffee sales patterns and customer trends.
6. Interpretation: Derived insights regarding customer preferences and sales fluctuations across seasons.

4. Observed Insights and Hidden Facts

- Espresso and Latte are the top-selling coffee types, contributing the highest revenue share.
- Sales are significantly higher during winter months, showing strong seasonal demand patterns.
- Large-size coffees generate greater revenue per transaction compared to smaller sizes.
- Customer retention patterns show consistent purchases among frequent buyers.
- Weekends display higher sales activity than weekdays, aligning with leisure consumption behavior.

5. Recommendations

For Business Managers:
- Introduce loyalty programs to retain frequent customers.
- Increase stock and promotional activities during peak winter months.

For Marketing Teams:
- Focus on promoting popular products like Espresso and Latte.
- Experiment with seasonal offers to attract new customers.

For Data Analysts:
- Extend the analysis to include customer demographics and store-level data.

- Implement forecasting models such as ARIMA or Prophet for sales prediction.

6. Limitations

- The dataset does not include customer location or demographic details.
- Missing or incomplete records required manual data cleaning.
- External market factors like pricing strategy and competition were not considered.

7. Conclusion

This project effectively demonstrates the use of Python for comprehensive sales data analysis. The insights derived help businesses understand customer behavior, optimize inventory management, and refine marketing strategies. By applying data analytics techniques, the study successfully revealed key performance indicators and sales drivers. Future enhancements could involve integrating predictive analytics, real-time dashboards, and advanced machine learning models for better forecasting.

References

1. McKinney, W. (2010). Data Structures for Statistical Computing in Python. SciPy Proceedings.
2. Dean, J. & Ghemawat, S. (2004). MapReduce: Simplified Data Processing on Large Clusters. OSDI.
3. Brown, T. (2019). Visualization Techniques in Sales Data Analysis. Data Science Review.
4. Smith, A. (2021). Consumer Behavior and Coffee Sales Analytics. Journal of Business Intelligence.