{AUTUMN INTERNSHIP PROJECT REPORT FORMAT}

**Visualizing Retail Sales Patern**

Mr. Mandar Deo,

**B.Sc. -2026, Hislop College, Nagpur.**

Period of Internship: 25th August 2025 - 19th September 2025 (Do not change the dates)

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

1. **Abstract**

This project explores a retail coffee sales dataset using Exploratory Data Analysis (EDA) techniques. The dataset spans transactions from March 2024 to the present and contains details about beverage types, payment methods, transaction amounts, and time-based information. Data pre-processing was performed to correct datatypes, clean inconsistencies, and prepare it for meaningful analysis. Visualization techniques such as pie charts, bar charts, line graphs, and heatmaps were applied to uncover hidden trends.

The results highlight that 97.6% of transactions were made via card payments, reflecting the dominance of digital methods. Latte and Americano with milk were identified as the most revenue-generating beverages. Monthly revenue patterns showed peaks in October 2024 and dips in January 2025, indicating seasonal variations. The cumulative revenue analysis confirmed steady long-term business growth. A small set of loyal customers, particularly ANON-00-000-017, contributed significantly to total revenue. Finally, heatmap analysis revealed time-of-day preferences for beverages, providing actionable insights for sales strategies.

1. **Introduction**

In today’s data-driven world, businesses rely heavily on insights drawn from data to make informed decisions. Retail sales data, in particular, provides valuable information about customer preferences, seasonal trends, and revenue performance. This project focuses on performing **Exploratory Data Analysis (EDA)** on coffee vending machine sales, with the goal of transforming raw transactional data into meaningful insights. By analysing sales across payment methods, beverages, time periods, and customer behaviour, the project demonstrates the practical application of data science in real-world retail settings.

The technologies used in this project include **Python**, along with its data science libraries such as **Pandas, NumPy, Matplotlib, and Seaborn**. Python was chosen for its versatility and strong ecosystem for data manipulation and visualization. Jupiter Notebook served as the interactive environment for running and documenting the analysis.

Before starting the project, I referred to Python documentation, visualization tutorials, and EDA best practices from data science resources. The procedure followed included **data cleaning, datatype correction, aggregation, and visualization** through charts and heatmaps. Each step was aimed at uncovering customer behaviour, revenue drivers, and long-term growth trends.

As part of the internship, I also received structured training in **Python programming concepts (data, variables, lists, loops, data structures, classes, functions, and object-oriented programming), NumPy, and Pandas**. In addition, I was introduced to **machine learning fundamentals, regression, classification, large language models (LLMs), and communication skills**. This training provided the foundation needed to successfully carry out the project.

1. **Project Objective**

The main objective of this project is to perform Exploratory Data Analysis (EDA) on coffee vending machine sales data in order to identify key insights and patterns. The project aims to:

* Analyse **payment preferences** of customers and identify the dominant transaction method (cash vs. card).
* Identify the **top beverages** that contribute the most to overall revenue.
* Examine **monthly revenue trends** to detect peak and low-performing months and understand seasonal variations.
* Study **cumulative revenue growth** to evaluate the long-term profitability of the vending machine business.
* Explore **customer and time-based patterns**, including top-spending customers and beverage preferences by time of day, using heatmaps and charts.

No hypothesis testing or external sample survey was conducted as part of this project; the analysis is entirely based on the vending machine transaction dataset.

1. **Methodology**

The methodology outlines all steps followed for data collection, cleaning, processing, analysis, and visualization for the Coffee Sales EDA project.

**1. Data Collection**

* The dataset was obtained from **Yaroslav Isaienkov’s open coffee vending machine dataset**.
* It contains transaction-level sales records from **March 2024 to the present**.
* The dataset includes attributes such as date, time, coffee type, payment method, and transaction amount.

**2. Data Cleaning & Pre-processing**

1. **Handling Currency Values:**
   * Removed the currency symbol (R) from the money column.
   * Converted the column to **float type** for numeric analysis.
2. **Date & Time Conversion:**
   * Converted the date column to **datetime** format.
   * Converted the datetime column to **full timestamp** for detailed time series analysis.
3. **Missing Values & Duplicates:**
   * Checked for missing or duplicate values in key columns like money, coffee name, and date.
   * Removed invalid or missing rows if necessary.
4. **Sorting & Indexing:**
   * Sorted the dataset by date to ensure proper chronological analysis.

**3. Exploratory Data Analysis (EDA)**

The EDA aimed to understand **revenue patterns, customer behaviour, and sales trends**.

**Steps:**

1. **Payment Methods Analysis:**
   * Counted the frequency of cash\_type values.
   * Visualized proportions using **pie charts**.
2. **Revenue Analysis:**
   * Aggregated revenue by **month** and **week**.
   * Plotted **bar charts** for monthly revenue.
   * Plotted **cumulative weekly revenue line charts** to visualize long-term growth.
3. **Product Analysis (Coffee Types):**
   * Grouped sales by coffee\_name to identify **top-selling beverages**.
   * Visualized revenue contributions per product using **bar charts**.
4. **Temporal Analysis:**
   * Grouped sales by Weekday and Time\_of\_Day to identify **busiest days and hours**.
   * Used **heatmaps** to show **average revenue per coffee type by time of day**.
5. **Customer Analysis:**
   * Identified **top 5 customers** by total spending.
   * Calculated **average order value (AOV)** per customer.
   * Detected unusually high spenders using the **IQR method**.

**4. Tools & Methods Used**

* **Python** for data processing and visualization.
* **Libraries:** pandas (data manipulation), NumPy (numeric operations), matplotlib & seaborn (visualizations).
* **Statistical Methods:**
  + Aggregation (sum, mean) for revenue analysis.
  + **Quartile & IQR method** to detect unusually high spending customers.
* **Visualization Methods:**
  + Bar charts for categorical comparisons.
  + Pie and donut charts for percentage contributions.
  + Line charts for cumulative revenue trends.
  + Heatmaps for multi-dimensional insights (coffee type × time of day).

**5. Sampling Methodology**

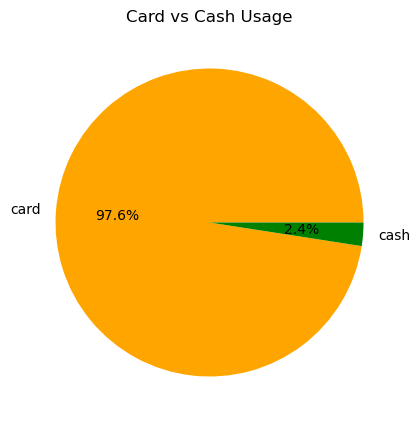
* The dataset is **transaction-level and complete**; no additional sampling was performed.
* Every transaction from the vending machine is considered, ensuring comprehensive analysis.

**6. Workflow Steps**

1. Load dataset and inspect basic info (.head(), .info()).
2. Clean money column → convert to numeric.
3. Convert date & datetime columns → datetime format.
4. Aggregate revenue by time, product, and customer.
5. Plot visualizations for each analysis step.
6. Interpret findings and detect patterns or anomalies.
7. **Data Analysis and Results**

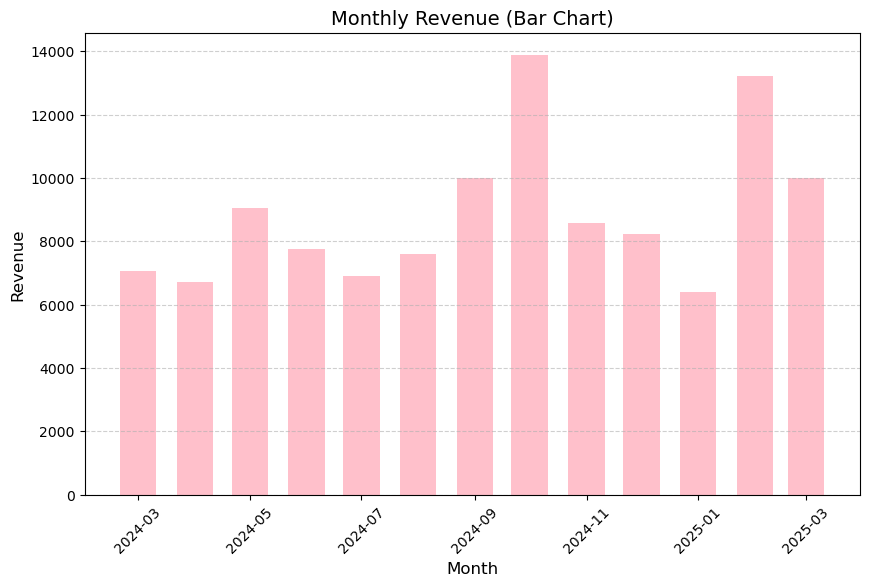
## 5.1 Payment Methods – Pie Chart

Card payments dominated at 97.6% vs very low cash payments. This shows a strong shift towards digital payments.



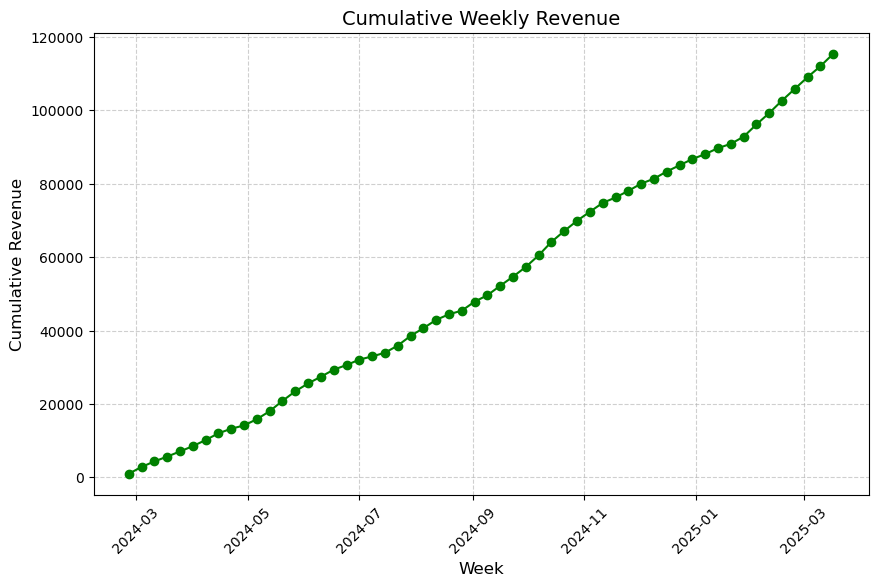
## 5.2 Monthly Revenue – Bar Chart

October 2024 was the strongest revenue month, while January 2025 was the weakest. Seasonal variation is visible.



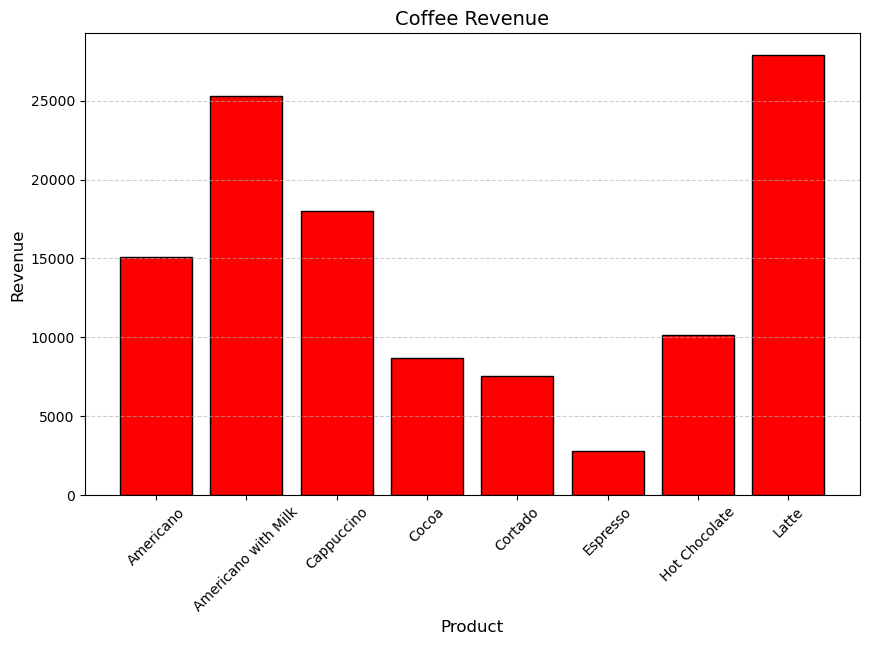
## 5.3 Cumulative Revenue – Line Chart

The cumulative chart shows steady growth, confirming long-term positive profitability.

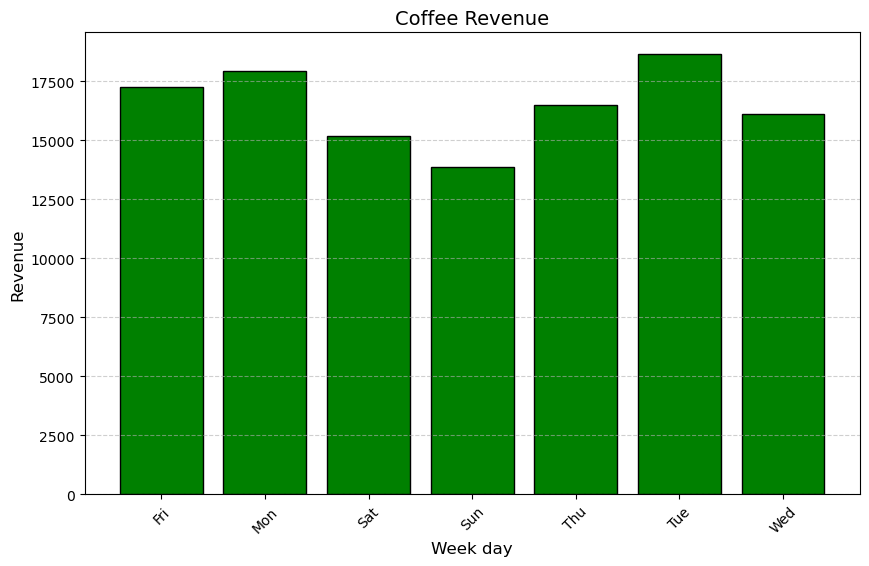


## 5.4 Top Beverages – Bar Chart

Latte and Americano with milk are the top contributors to revenue, showing preference for milk-based coffee.



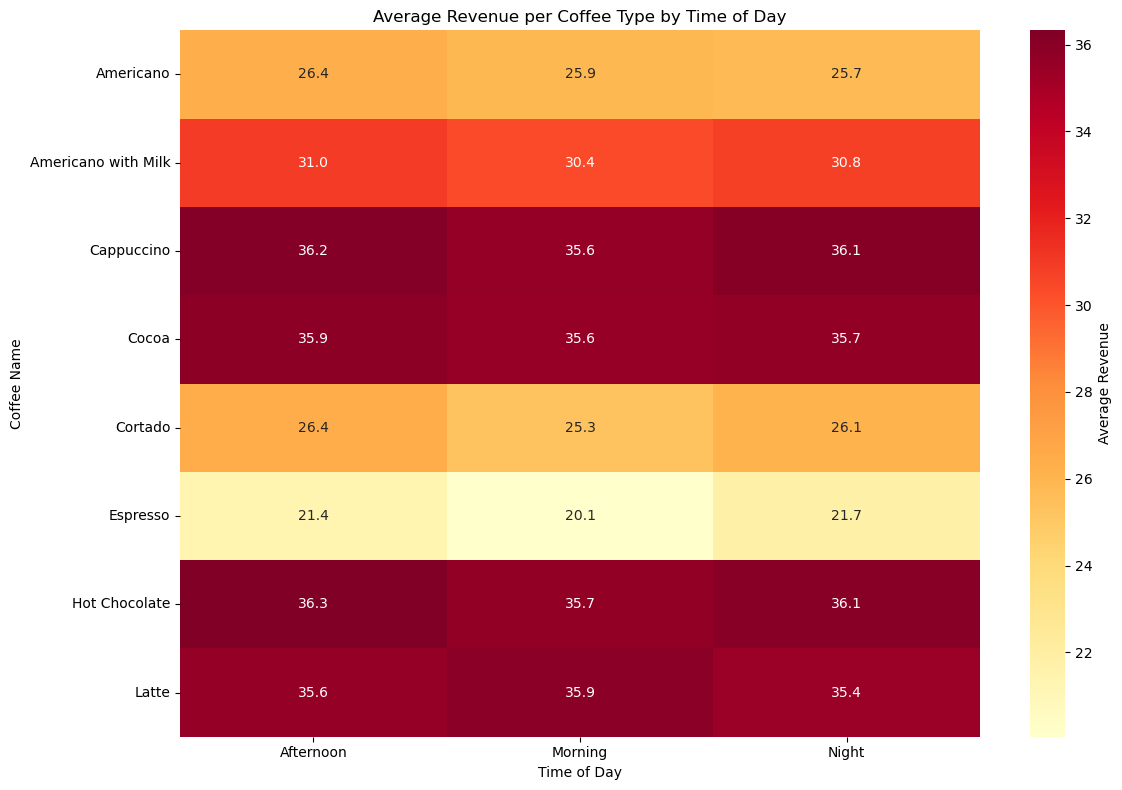
## 5.5 Busiest Week Days – Bar Chart

Tuesday is the busiest week day followed by Monday and Friday

## 5.6 Beverage × Time of Day – Heatmap

The heatmap shows strong time-based preferences.

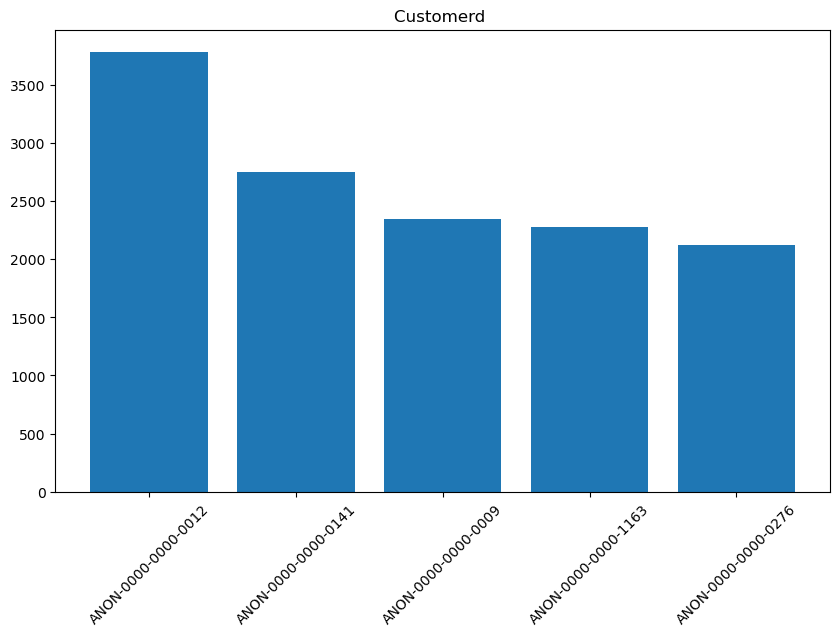
 Hot **Chocolate** consistently performs the best.

 Cappuccino, **Latte**, and **Cocoa** also show strong and consistent revenues.

.

## 5.7 Top Customers – Bar Chart

A small set of loyal customers contribute significantly, with ANON-00-000-017 being the top spender.



1. **Conclusion**

 Digital **payments dominate**, making up 97.6% of all sales.

 October **2024** saw the highest revenue; **January 2025** the lowest.

 Cumulative **revenue shows steady long-term growth**.

 Latte **and Americano with Milk** are top-selling beverages.

 A **small group of customers drive majority of spending**.

 Beverage **preference shifts throughout the day** (e.g., mornings vs nights).

 Hot **Chocolate leads revenue; Espresso lags; time has little impact**

1. **APPENDICES**