INFO BHARAT INTERNS – INTERNSHIP REPORT

INTERN DETAILS

Name: Arvind Kumar
Internship: Info Bharat Interns (IBI)
Domain: Data Science / Business Analytics
Duration: 4 Weeks
Project Title: Advanced Customer & Sales Analysis using Machine Learning

---

PROJECT OBJECTIVE

The objective of this project was to perform customer segmentation, sales forecasting, and churn prediction using customer transaction data. Through machine learning and data analysis, the goal was to derive actionable business insights and recommend strategic decisions.

TOOLS & TECHNOLOGIES USED

- **Google Colab** – cloud-based Python notebook for coding
- **Python 3** – main programming language
- **Pandas, NumPy** – data handling and feature engineering
- **Seaborn, Matplotlib** – data visualization
- **Scikit-learn** – machine learning modeling (GMM, preprocessing)
- **GitHub** – version control & code hosting
- **Google Docs** – report creation and formatting

DATASET DESCRIPTION

- Total records: 1000
- Features: 21
- Type: Simulated e-commerce data (Customer + Product + Transaction)
- Key columns: Customer Age, Gender, Income Level, Loyalty Score, Product Category, Purchase Quantity, Payment Method, Date of Purchase, Discount Offered

KEY STEPS EXECUTED

1. **Data Cleaning and Preprocessing**

2. **Feature Engineering**
    - Customer Lifetime Value (CLV)
    - Recency (Days since last purchase)
    - Loyalty Bucket
    - Discount Percent

3. **Exploratory Data Analysis (EDA)**
    - Boxplots, Histograms, Correlation Heatmap

4. **RFM Table Creation**
    - Recency, Frequency, Monetary

5. **Customer Segmentation**
    - Applied Gaussian Mixture Model (GMM)

6. **Segment Profiling**
    - Summarized average CLV, Recency, Loyalty, and Frequency by segment

7. **Business Recommendations**
    - Based on customer segment behavior

CUSTOMER SEGMENT PROFILES

Segment	Description
Segment 0	High-Value Loyal Customers – recent and frequent buyers with high CLV
Segment 1	Medium-Value Buyers – moderate loyalty and occasional spenders
Segment 2	Discount-Sensitive Inactive – infrequent purchases, high discount usage
Segment 3	At-Risk or Churned – low value, inactive customers

VISUAL INSIGHTS

(Include the following screenshots from your Google Colab notebook here):

CLV Boxplot

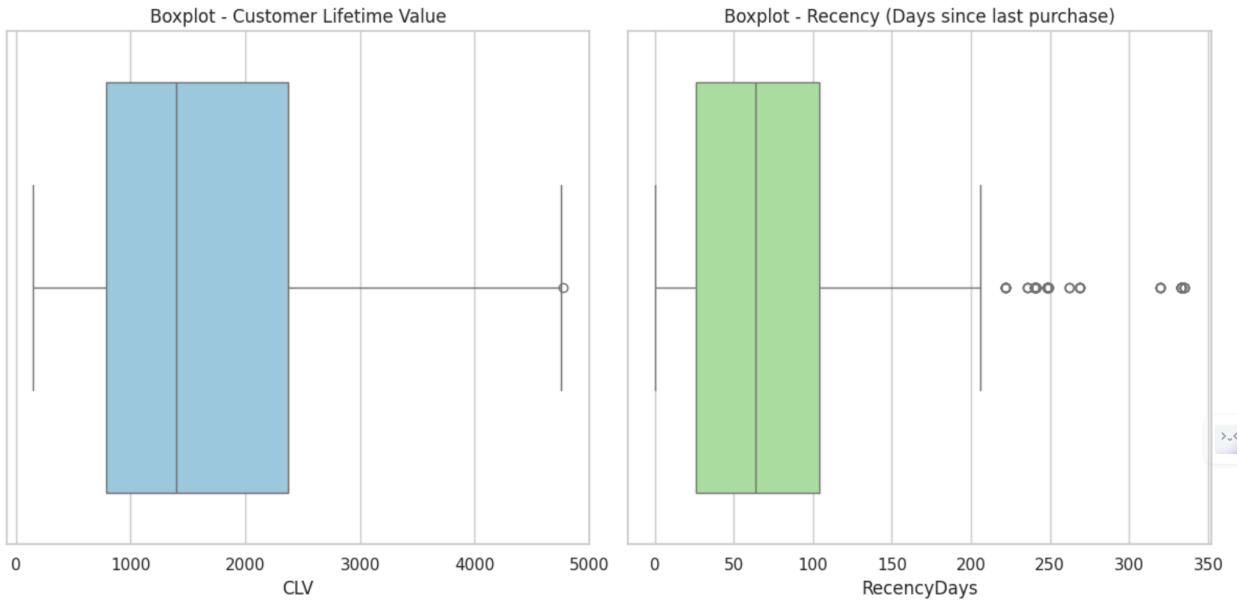


Figure 1: Boxplot – Customer Lifetime Value (CLV) and Recency (Days since last purchase)

CLV Distribution

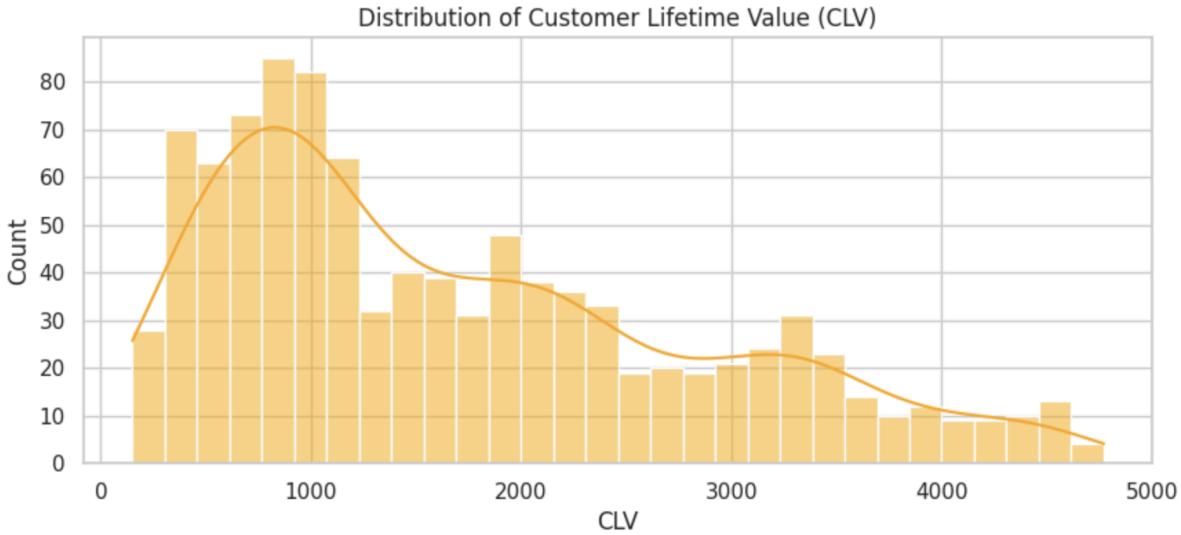


Figure 2: Distribution of Customer Lifetime Value (CLV)

Correlation Heatmap

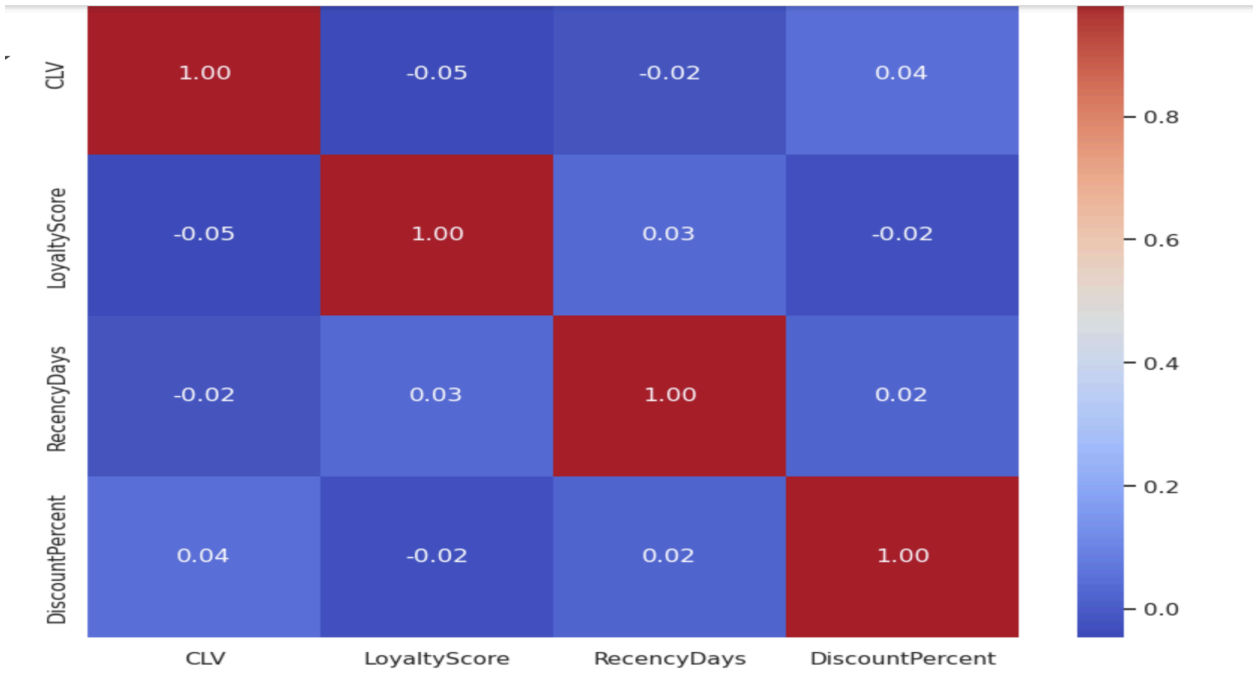


Figure 3: Correlation Heatmap of Key Engineered Features

Segment Count Plot / Pie Chart

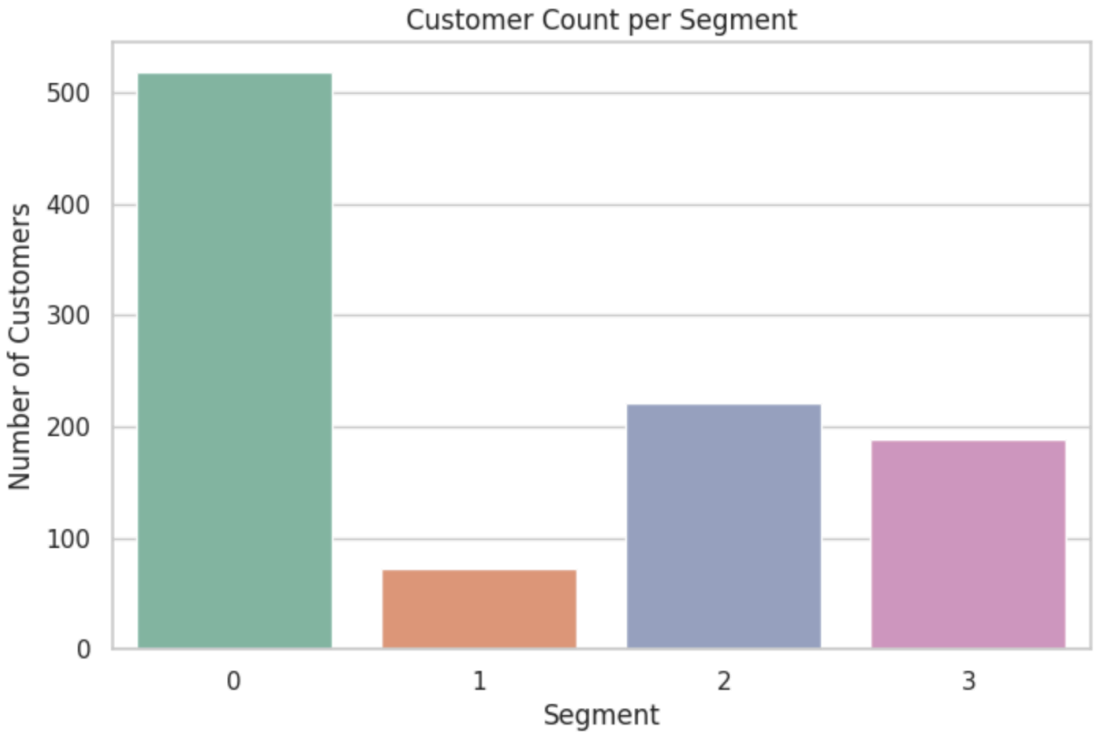


Figure 4: Customer Count per Segment (Bar Chart)

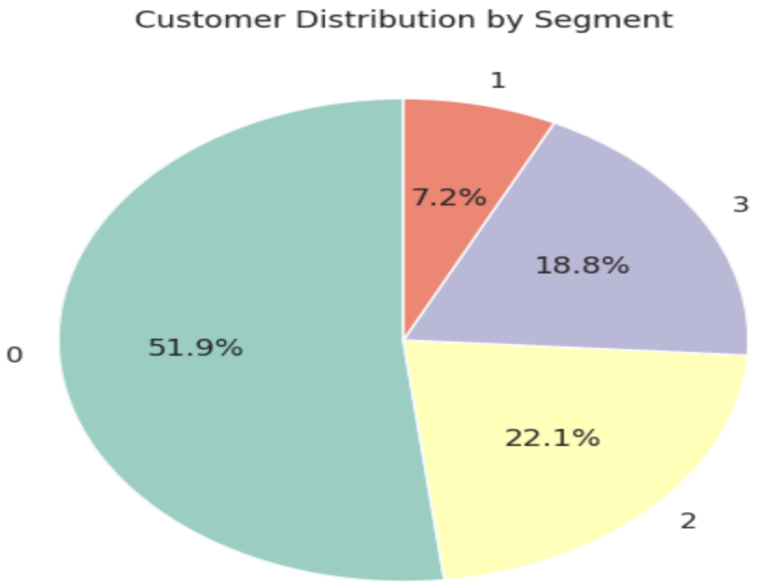


Figure 5: Customer Distribution by Segment (Pie Chart)

BUSINESS STRATEGY SUGGESTIONS

- **Segment 0:** Offer VIP loyalty programs, premium perks
- **Segment 1:** Promote frequent purchase offers and cashback
- **Segment 2:** Target with special discount campaigns and personalized coupons
- **Segment 3:** Run win-back campaigns via SMS/email with incentives

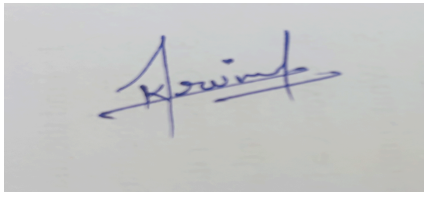
GITHUB PROJECT LINK

https://github.com/arvkumar-29/IBI-Ecommerce-Customer-Analysis

CONCLUSION

This internship project provided practical experience in advanced data analysis, feature engineering, clustering algorithms, and business insights generation. It strengthened my understanding of real-world data handling and strategic decision-making using machine learning.

SIGNATURE



Arvind Kumar
Computer Science Engineering
Info Bharat Intern – 2025