Customer Lifetime Value Prediction Project

Internship Project | Role: Data Analyst
Author: Paluru Tejasri
Date: June 2025

---

1. Introduction

This project focuses on predicting Customer Lifetime Value (CLV) using historical transaction data. CLV is a vital metric for businesses to identify and retain high-value customers. I applied RFM analysis—Recency, Frequency, and Monetary—to engineer features and trained a regression model to predict future customer value.

---

2. Dataset Used

- **Source:** Kaggle (Online Retail Dataset)
- **Records:** 541,909 rows | 8 columns
- **Fields:** InvoiceNo, InvoiceDate, Quantity, UnitPrice, CustomerID, Country
- **Format:** Excel (.xlsx)

---

3. Step-by-Step Workflow

1. **Cleaned data:** Removed missing values and cancelled orders
2. **Engineered 'TotalAmount' column:** Quantity × UnitPrice
3. **Created RFM table per customer**
4. **Trained a linear regression model** using Recency and Frequency to predict Monetary value
5. **Evaluated model performance** with MAE and RMSE
6. **Visualized Actual vs Predicted LTV** with scatter plots and bar charts
7. **Exported results:** Saved predictions and top customers to Excel and saved charts as PNG images

---

4. Tools & Libraries Used

- Jupyter Notebook (Anaconda)
- Python (pandas, numpy, matplotlib, scikit-learn)
- Excel files for data management

---

5. What I Learned

I learned how to work with real-world datasets, perform end-to-end data cleaning, feature engineering, modeling, and interpretation. This project helped build confidence in applying data analytics practically and improved my Python and data visualization skills.

6. Acknowledgement

I sincerely thank my internship team and mentors for giving me the opportunity and guidance to work on this project. Their support helped me understand the real-world application of data analysis.

7. Future Scope

This project can be improved further by:
- Trying advanced models like XGBoost or Random Forest
- Adding customer segmentation
- Building interactive dashboards using tools like Power BI or Tableau
- Using time-series approaches for CLTV prediction

8. Conclusion

Through this project, I understood how customer data can be used to derive valuable business insights. Predicting Customer LTV can help businesses focus more on valuable customers and optimize marketing efforts.

9. Final Note

This report reflects my internship project work as a Data Analyst intern and demonstrates my ability to handle a complete data science project lifecycle—from raw data to actionable insights.