INTRODUCTION:

In today's competitive market, understanding customer behaviour and predicting their potential value is crucial for strategic decision-making. This project aims to predict the **Customer Lifetime Value (LTV)** using historical purchase data. The predicted LTV helps businesses identify high-value customers, optimize marketing efforts, and enhance customer retention strategies.

ABSTRACT:

This project presents a machine learning approach to estimate the future monetary value a customer is expected to generate. Using transaction-level data, we engineered features like **Recency**, **Frequency**, and **Average Order Value (AOV)**. A **Random Forest Regression** model was trained to predict LTV. The model's performance was validated using MAE and RMSE metrics, and customers were segmented into three value tiers: **High**, **Mid**, and **Low**. The final output is a labeled dataset for actionable marketing insights.

TOOLS USED:

Python: Core programming language

Pandas: Data manipulation and feature engineering

NumPy: Numerical calculations

Matplotlib/Seaborn: Data visualization

Scikit-learn: Model training and evaluation

Random Forest: Regression algorithm for prediction

Excel: Final result export

Jupyter Notebook: Development and testing environment

STEPS INVOLVED IN BUILDING THE PROJECT:

1. Data Collection

- Used a synthetic dataset of 5,000 transactions across 500 customers.
- Each record included: CustomerID, OrderID, OrderDate, and OrderAmount.

2. Data Preprocessing

- Converted date fields to datetime.
- Checked for missing values and ensured proper formatting.

3. Feature Engineering

- **Recency**: Days since the last transaction.
- **Frequency**: Number of unique transactions.
- **AOV** (Average Order Value): Total Amount / Frequency.

4. Model Development

- Used **Random Forest Regressor** to predict Monetary value (LTV).
- Split dataset into training and test sets (80:20 ratio).
- Trained model and evaluated performance using:
    - **Mean Absolute Error (MAE)**
    - **Root Mean Squared Error (RMSE)**

5. Customer Segmentation

- Based on predicted LTV, customers were labeled as:
    - **High Value**: Top 25%
    - **Mid Value**: 25–50%
    - **Low Value**: Bottom 50%

6. Output and Visualization

- Final predictions saved in CSV and Excel formats.
- Visualized LTV distribution and customer segment counts.

CONCLUSION:

This project successfully demonstrated how machine learning can be leveraged to estimate Customer Lifetime Value based on past purchase behaviour. The use of Random Forest regression proved effective in handling non-linear relationships in the data. The segmentation of customers into tiers provides valuable insights for targeted marketing, resource allocation, and customer relationship management. This solution is scalable and can be adapted to real-world retail and e-commerce scenarios.