**Customer Lifetime Value Prediction Model**

**Introduction**

Customer Lifetime Value (CLV) is a key metric that helps businesses estimate how much revenue a customer is likely to generate over time. This project aims to predict CLV based on customer purchase behaviour using real e-commerce data from Olist. The objective is to support better marketing strategies by identifying high-value customers.

**🔹 Tools Used**

* **Python** (Jupyter Notebook)
* **Pandas & NumPy** for data processing
* **Scikit-learn & RandomForest** for regression modeling
* **Matplotlib** for visualizations
* **Excel** (for optional cross-checks)

**🔹 Steps Involved in Building the Project**

1. **Data Preprocessing**
   * Loaded three datasets: olist\_orders\_dataset.csv, olist\_order\_items\_dataset.csv, and olist\_order\_payments\_dataset.csv.
   * Merged datasets using order\_id to link customer info, purchase amount, and order time.
2. **Feature Engineering**
   * Extracted key RFM metrics for each customer:
     + **Recency**: Days since last purchase
     + **Frequency**: Number of unique orders
     + **Monetary**: Total amount spent
     + **AOV (Average Order Value)**: Monetary / Frequency
3. **Model Building**
   * Defined Recency, Frequency, and AOV as features.
   * Used **Random Forest Regressor** to predict the Monetary value (CLV).
   * Split data into training and testing sets (80/20).
4. **Model Evaluation**
   * Evaluated model performance using:
     + **MAE (Mean Absolute Error)**
     + **RMSE (Root Mean Squared Error)**
5. **LTV Prediction & Customer Segmentation**
   * Predicted CLV for all customers.
   * Segmented customers into four groups:
     + **Low**, **Mid**, **High**, and **Very High** based on predicted CLV.
   * Saved final results to CSV.
6. **Visualization**
   * Plotted and saved a bar chart showing **feature importance** to understand which variables impact CLV the most.

**🔹 Conclusion**

The project successfully demonstrates how customer purchase behavior can be used to predict their lifetime value using machine learning. The model and its predictions can help companies better allocate marketing resources and personalize strategies for different customer segments. This project also provides a foundation to build more advanced customer analytics models in real-world business scenarios.