# Customer Lifetime Value (LTV) Prediction Report

**1. Introduction**

In today's competitive market, understanding the future value of customers is vital for allocating marketing resources effectively. This project aims to predict Customer Lifetime Value (LTV) using transactional data, enabling strategic segmentation and personalized marketing.

The dataset includes customer purchase records with fields such as InvoiceDate, CustomerID, Quantity, UnitPrice, and product details. Using feature engineering and machine learning models, the goal is to estimate each customer's future monetary contribution.

---

**2. Data Preparation**

- **Source**: customer_data.csv
- **Cleaning**:
    - Removed transactions with null CustomerID
    - Excluded cancellations/returns and transactions with non-positive quantity or price
- **Final Features Engineered**:
    - **Recency**: Days since last purchase
    - **Frequency**: Number of unique transactions
    - **Monetary**: Total spend
    - **AOV** (Average Order Value): Monetary / Frequency
    - **ProductVariety**: Count of unique products purchased
    - **PurchaseInterval**: Recency divided by frequency
    - **TotalItems**: Sum of all quantities purchased

---

**3. Exploratory Data Analysis**

15 insightful visualizations were generated:

- Distributions of Recency, Frequency, Monetary value, AOV
- Correlation heatmap showing strong links between Frequency and Monetary value
- Time-based trends (monthly sales, hourly purchase patterns)
- Segmentation via KDE plots and boxplots by country
- Feature interactions: Monetary vs Recency/Frequency/Product Variety

This step confirmed clear customer behavior patterns, guiding model selection and segmentation logic.

---

### 4. Modeling Approach

- **Target Variable**: Log-transformed Monetary to reduce skew
- **Input Features**: Scaled versions of Recency, Frequency, AOV, ProductVariety, and PurchaseInterval
- **Algorithms Used**:
  - XGBoost Regressor (primary)
  - Random Forest tested during development

**Hyperparameters (XGBoost):**

- n_estimators=1000, learning_rate=0.05, max_depth=6, with early stopping

**Evaluation Metrics:**

- **MAE (Mean Absolute Error)**: Low, indicating tight error spread
- **RMSE (Root Mean Squared Error)**: Used to penalize larger errors
- **$R^2$ Score**: High explanatory power

---

### 5. Results & Outputs

- **Model Artifacts Saved**:
  - ltv_xgb_model.pkl – Trained XGBoost model
  - scaler.pkl – StandardScaler object
  - power_transformer.pkl – Log transformer for target
- **Customer Segments**:
  - Customers were segmented into Low, Medium, and High tiers using quantile-based bucketing on predicted LTV
  - Exported as customer_ltv_predictions.csv for downstream marketing integration
- **Visual Assets**:
  - 15 charts illustrating key trends and model insights were generated (not included in final folder as per instructions)

---

### 6. Conclusion

The project successfully built a predictive framework for Customer LTV using historical transaction behavior. With strong evaluation metrics and segment insights, the model offers a robust basis for data-driven customer strategy.