Customer Spending Prediction Report

1. Data Overview

- Source: Kaggle dataset (1M transactions)

- Target: Amount_spent

- Features Used: All except Amount_spent and Transaction_ID

- Preprocessing:

  - Dropped null values

  - Encoded categorical features

  - Scaled features using GPU (cuML StandardScaler)

2. Models Trained

Linear Regression (cuML)

- MAE: ~49.28

- MSE: ~4536.75

- R^2 Score: ~0.78

Random Forest Regressor (cuML)

- Log-transformed target

- Grid Search parameters:

  - n_estimators: [100, 200, 300]

  - max_depth: [10, 15, 20]

  - max_features: [0.7, 0.8, 1.0]

- Best R^2 Score: ~0.8345

- Final Evaluation:

  - MAE: 41.16

  - MSE: 3798.34

Customer Spending Prediction Report

- R^2 Score: 0.8345

3. Performance Comparison

Metric	Random Forest	Linear Regression
MAE	41.16	49.28
MSE	3798.34	4536.75
R^2	0.8345	0.78

4. Insights & Visuals

- Correlation matrix and heatmap
- Histogram of Amount_spent
- Scatter plot: Age vs. Amount_spent
- Gender distribution bar chart

5. Model Deployment

- Logistic Regression model and Scaler saved with joblib.
- Reloaded for prediction on X_test.