## **1. Dataset Description**

The dataset used for this analysis, bank\_transactions\_data\_2.csv, contains simulated customer banking transactions. Each record represents a single transaction and includes the following key variables:

* Transaction Information: TransactionID, TransactionDate, TransactionType, TransactionAmount, TransactionDuration
* Customer Information: CustomerAge, CustomerOccupation, AccountBalance
* Channel & Location: Channel, Location, MerchantID, DeviceID
* Behavioral Attributes: LoginAttempts, PreviousTransactionDate

After cleaning and preprocessing, a new target variable (Transaction\_Category) was created in Python to classify transactions into High or Low categories based on the median TransactionAmount.

This final dataset was exported as bank\_transactions\_with\_category.csv for use in Tableau.

## **2. Predictive Technique 1 – CART Model (Python)**

### **Objective:**

Use a Classification and Regression Tree (CART) model to predict whether a transaction is High or Low value based on customer and transaction features.

### Process Summary:

* Imported and cleaned the dataset using Pandas.
* Converted datetime columns to proper formats.
* Dropped irrelevant identifiers (TransactionID, AccountID, IP Address, DeviceID, etc.).
* Encoded categorical columns (TransactionType, Channel, Location, CustomerOccupation) using LabelEncoder.
* Split the data (80% training, 20% testing).
* Trained a DecisionTreeClassifier (max\_depth=5).
* Evaluated accuracy, confusion matrix, and classification report.

### **Results:**

|  |  |
| --- | --- |
| **Metric** | **Result** |
| **Accuracy** | 52.3% |
| **Precision (High)** | 0.55 |
| **Recall (High)** | 0.55 |
| **F1-Score (High)** | 0.55 |

#### **Interpretation:**

* The model achieved a baseline accuracy of ~52%, showing moderate predictive performance.
* Key predictive factors included AccountBalance, TransactionDuration, and CustomerAge.
* Although performance was limited, the model established a foundation for understanding which variables influence transaction value.
* Future improvement can include feature selection, parameter tuning, and balancing class distribution.

## **3. Predictive Technique 2 – Time Series Forecasting (Tableau)**

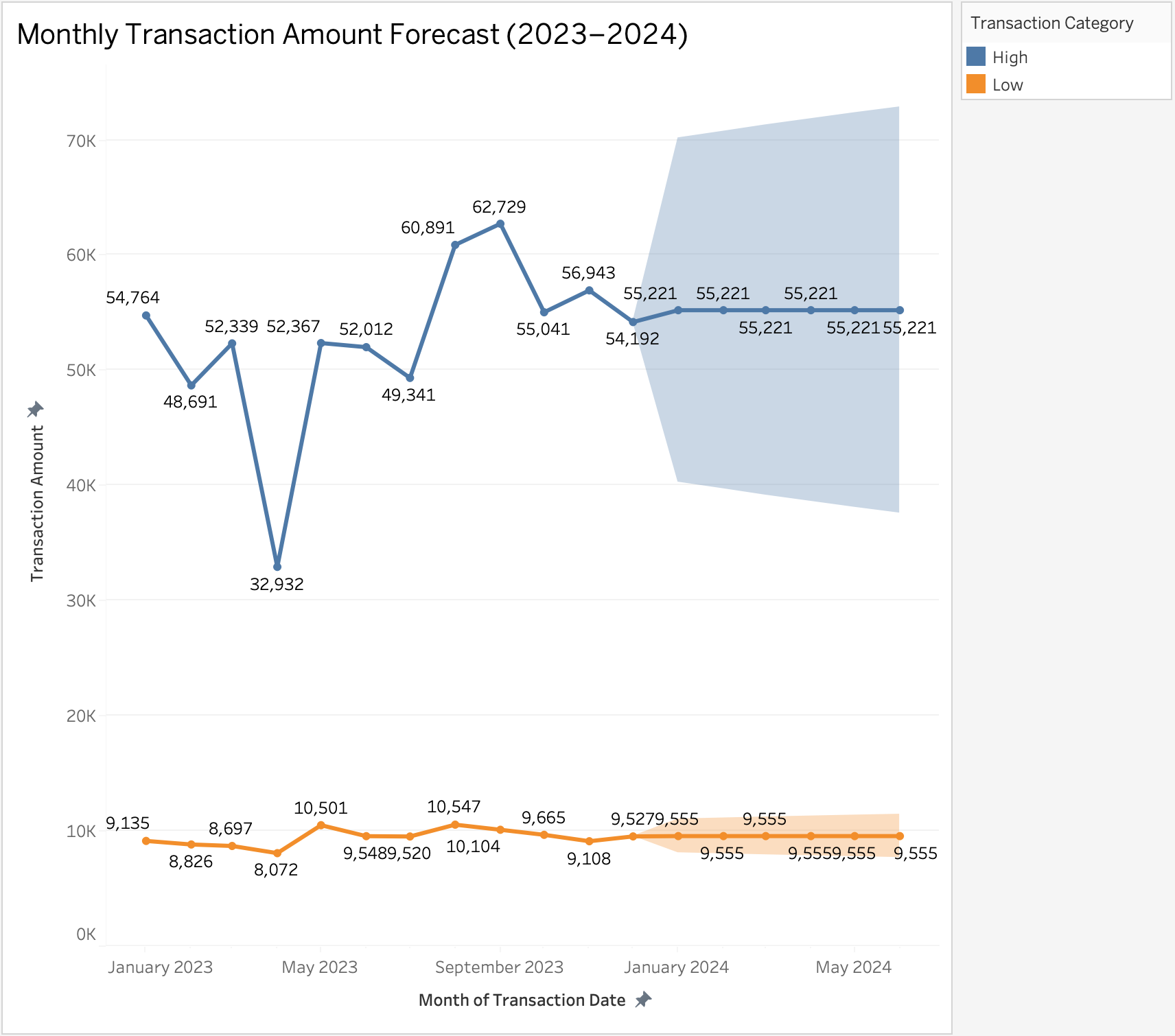
### **Objective:**

Forecast future transaction amounts over time using Tableau’s built-in time series forecasting (exponential smoothing).

### **Process Summary:**

* Imported bank\_transactions\_with\_category.csv into Tableau.
* Created a line chart:
  + MONTH(TransactionDate) → Columns
  + SUM(TransactionAmount) → Rows
* Added Tableau’s Forecast from the Analytics pane.
* Colored the lines by Transaction\_Category (High vs Low).

### **Visualization Summary:**

**Figure: Monthly Transaction Amount Forecast (2023–2024)**  


#### **Insights:**

* High-value transactions show significant month-to-month variation, peaking around June 2023 (~$62K) before stabilizing.
* Low-value transactions remain relatively steady (~$9K–$10K per month).
* The forecast predicts stable transaction volumes into early 2024, with high-value transactions expected to remain around $55K per month.
* The shaded blue region represents Tableau’s 95% confidence interval — indicating the range of uncertainty in future estimates.

#### **Interpretation:**

The forecast suggests steady customer engagement and consistent spending behavior for both categories, with high-value customers contributing the majority of total volume.

## **4. Foreseeable Challenges**

|  |  |  |
| --- | --- | --- |
| **Challenge** | **Description** | **Potential Impact / Solution** |
| **Data Import and Cleaning** | The original dataset contained mixed data types (strings, timestamps, numeric values). Encoding and parsing date columns required careful handling. | Used pd.to\_datetime() and Label Encoding to fix type mismatches. |
| **Categorical Encoding** | Columns like TransactionType and Location contained text. ML models can’t process text directly. | Used LabelEncoder() to convert text to numeric labels. |
| **Imbalanced Target Classes** | There was a near-even split but not perfect balance between High and Low transactions. | Future work could apply oversampling or class-weight balancing. |
| **Limited Predictive Strength of Features** | Many variables had weak correlation with TransactionAmount. | Feature selection or Random Forests could improve model accuracy. |
| **Time Series Constraints** | The dataset covered less than one full year of data, limiting long-term forecasting accuracy. | Forecast was generated with Tableau’s short-term model and interpreted cautiously. |
| **Software Access** | Tableau Desktop normally requires a paid license. | Used Tableau Public (Free Edition) for visualization. |

## **5. Conclusion**

This milestone combined two predictive techniques across two platforms:

1. CART (Python): Classified transactions as High vs Low based on behavioral and financial attributes.
   1. Result: 52% accuracy baseline.
   2. Key Predictors: AccountBalance, CustomerAge, TransactionDuration.
2. Time Series Forecast (Tableau): Projected future transaction amounts by month.
   1. Result: Stable trend around $55K/month for high-value transactions into early 2024.

Together, these analyses demonstrate a foundational data-driven approach to understanding customer transaction behavior and forecasting financial activity.