Exploratory Data Analysis (EDA) Credit Card Churn Project

1. Understand the Structure of the Data

- Dataset Source: Gigasheet Credit Card Customers

- Dataset Shape: ~1,000 rows ~20 columns

- Variable Types:

- Numerical: Customer_Age, Credit_Limit, Total_Trans_Ct, Total_Trans_Amt, Total_Revolving_Bal, etc.

- Categorical: Gender, Marital_Status, Income_Category, Card_Category, Education_Level, Attrition_Flag (target)

- No datetime fields were present in this dataset.

2. Summarize Main Characteristics

Descriptive Statistics:

Customer_Age (Mean: 46.3, Median: 46.0, Std Dev: 8.0, Min: 26, Max: 73)

Total_Trans_Ct (Mean: 64.9, Median: 67.0, Std Dev: 20.0, Min: 10, Max: 139)

Total_Trans_Amt (Mean: 4404.0, Median: 4200.0, Std Dev: 3400.0, Min: 500, Max: 20000)

3. Distribution of Variables

- Histograms used for age, transaction amount, etc.

- Boxplots revealed outliers in Total_Revolving_Bal and Total_Trans_Amt

- Churned customers had lower Total_Trans_Ct and Total_Trans_Amt

4. Detect Patterns and Relationships

- Correlation:

- Avg_Open_To_Buy Credit_Limit (strong)

- Total_Trans_Amt Total_Trans_Ct (strong)

- Segmentation:

- Churn slightly higher in lower income brackets

- Even churn distribution across genders

- Fewer relationships = higher churn

5. Spot Anomalies and Outliers

- Outliers found in Total_Revolving_Bal and Total_Trans_Amt

- Values were valid, not removed but flagged

6. Check for Missing Data

- No significant missing data

- No imputation needed

7. Visualize the Data

- Count plots: Churn vs. Gender, Income

- Boxplots: Churn vs. transaction metrics

- Histograms of key numeric features

- Correlation heatmap guided feature selection

EDA Summary:

- Clean dataset with informative features

- Lower transaction activity linked to churn

- Strong feature candidates: Total_Trans_Ct, Total_Trans_Amt, Total_Relationship_Count