EDA SLIDES PRESENTATION - NANDHINI

AGENDA

Introduction

EDA Plots

Insights



EDA

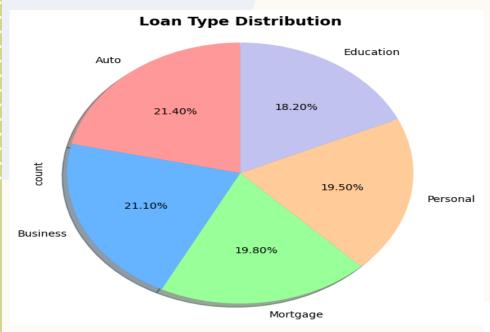
(EXPLORATORY DATA ANALYSIS) IS A CRITICAL STEP IN THE DATA ANALYSIS PROCESS. IT INVOLVES ANALYZING AND SUMMARIZING THE MAIN CHARACTERISTICS OF A DATASET, OFTEN THROUGH VISUALIZATION AND DESCRIPTIVE STATISTICS. EDA HELPS YOU UNDERSTAND THE DATA BETTER AND IDENTIFY PATTERNS, ANOMALIES, OR INSIGHTS THAT COULD INFLUENCE THE DIRECTION OF YOUR ANALYSIS OR MODELING.

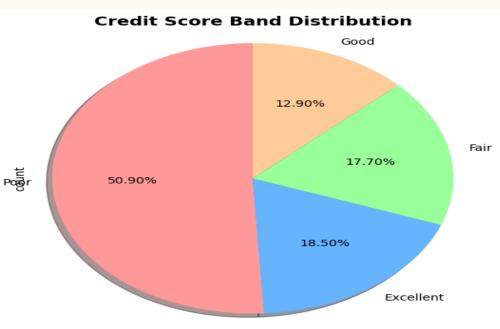
OBJECTIVE

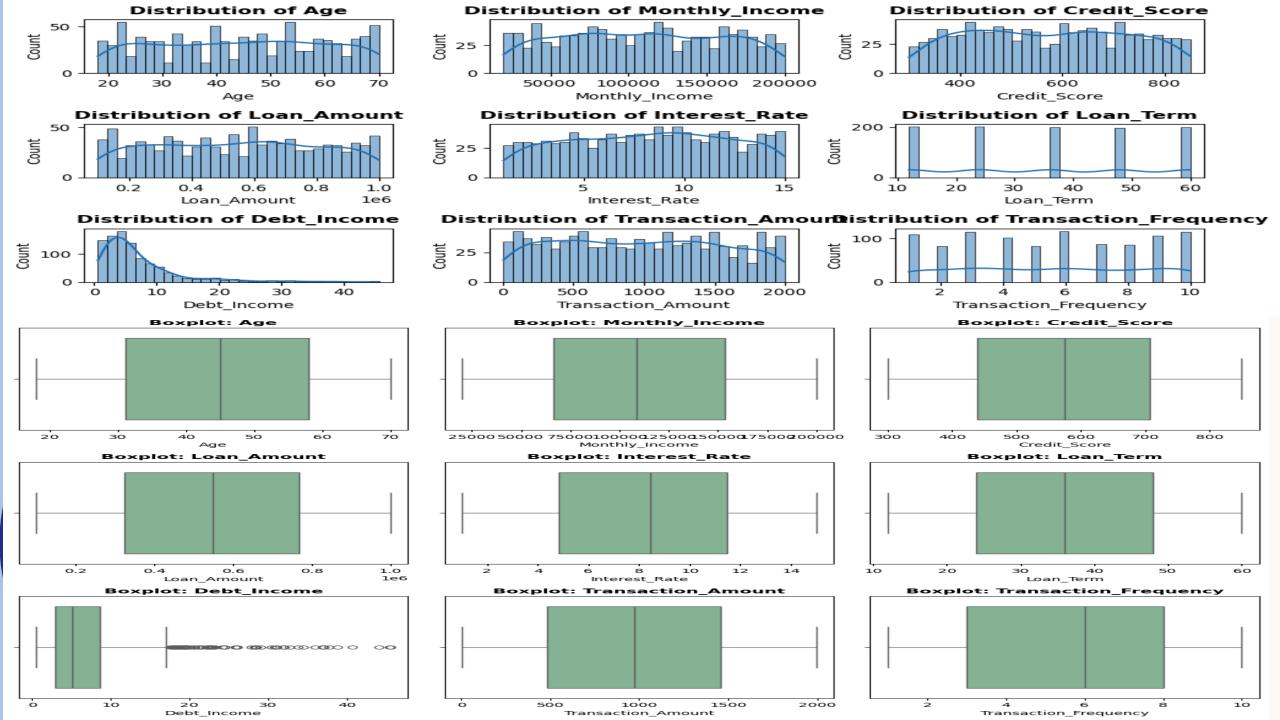
- 1. Understand data distributions and relationships.
- 2. Identify features contributing to predictions.

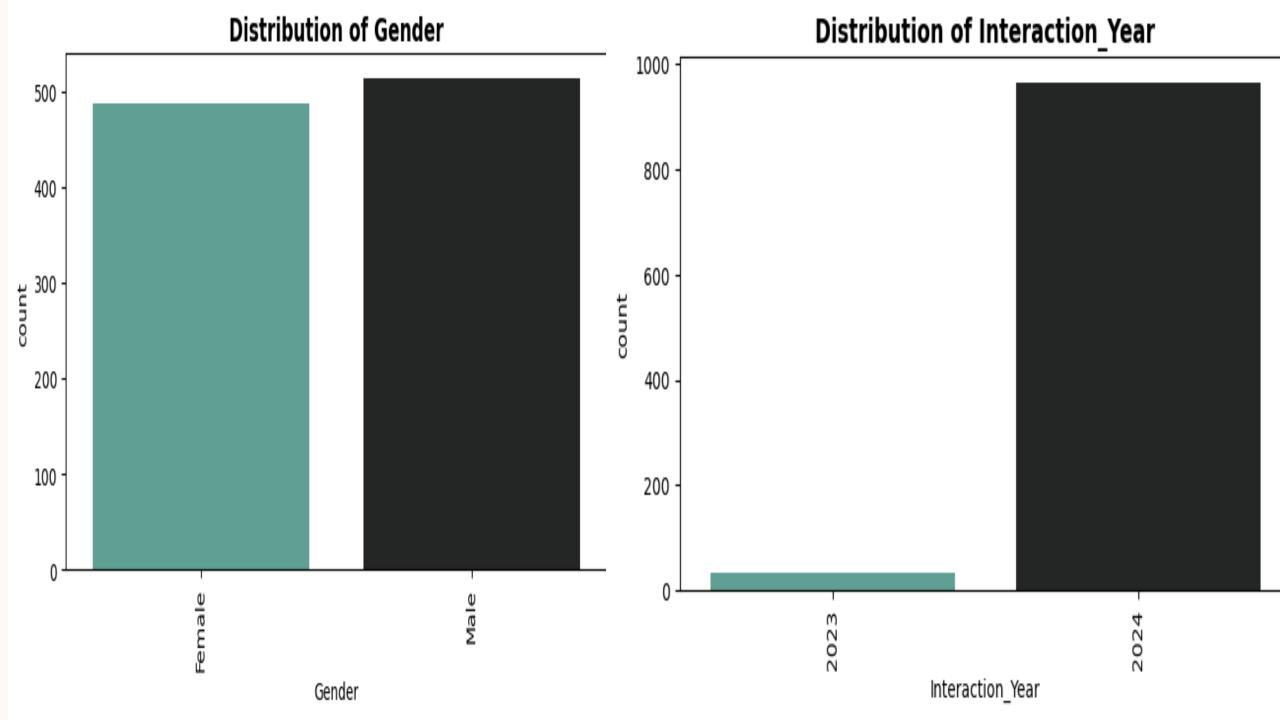


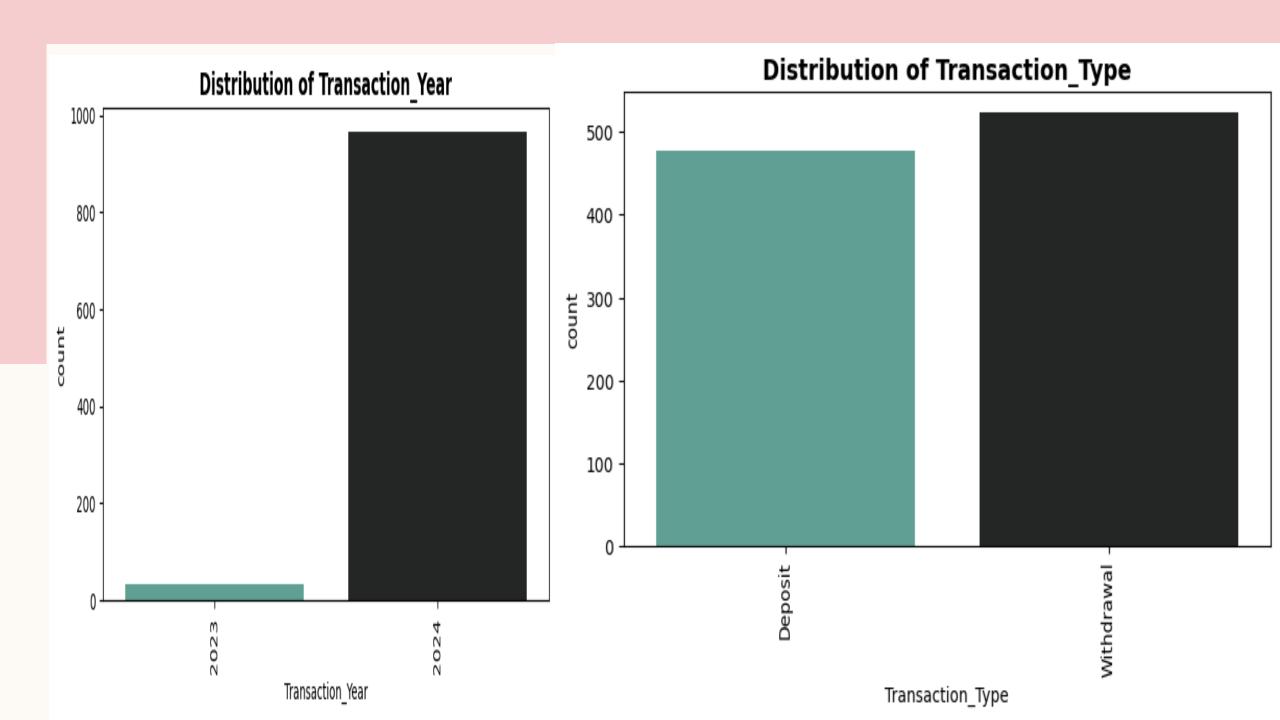
EDA PLOTS

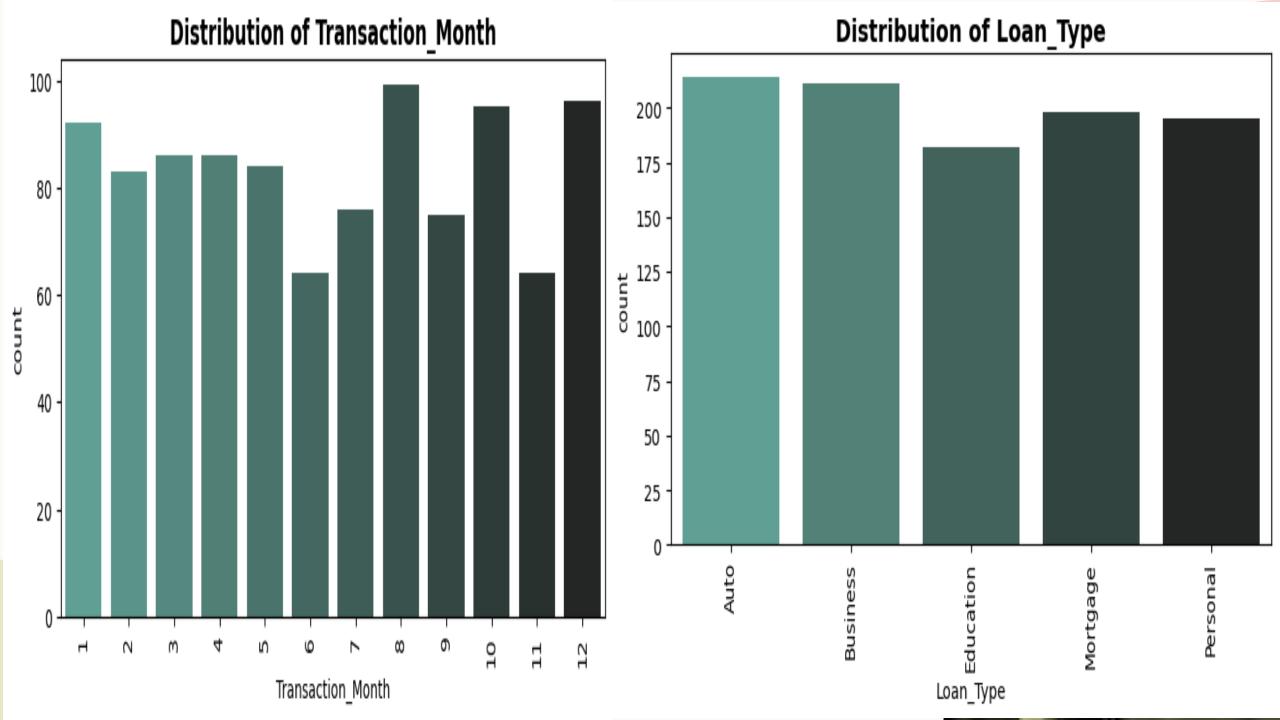


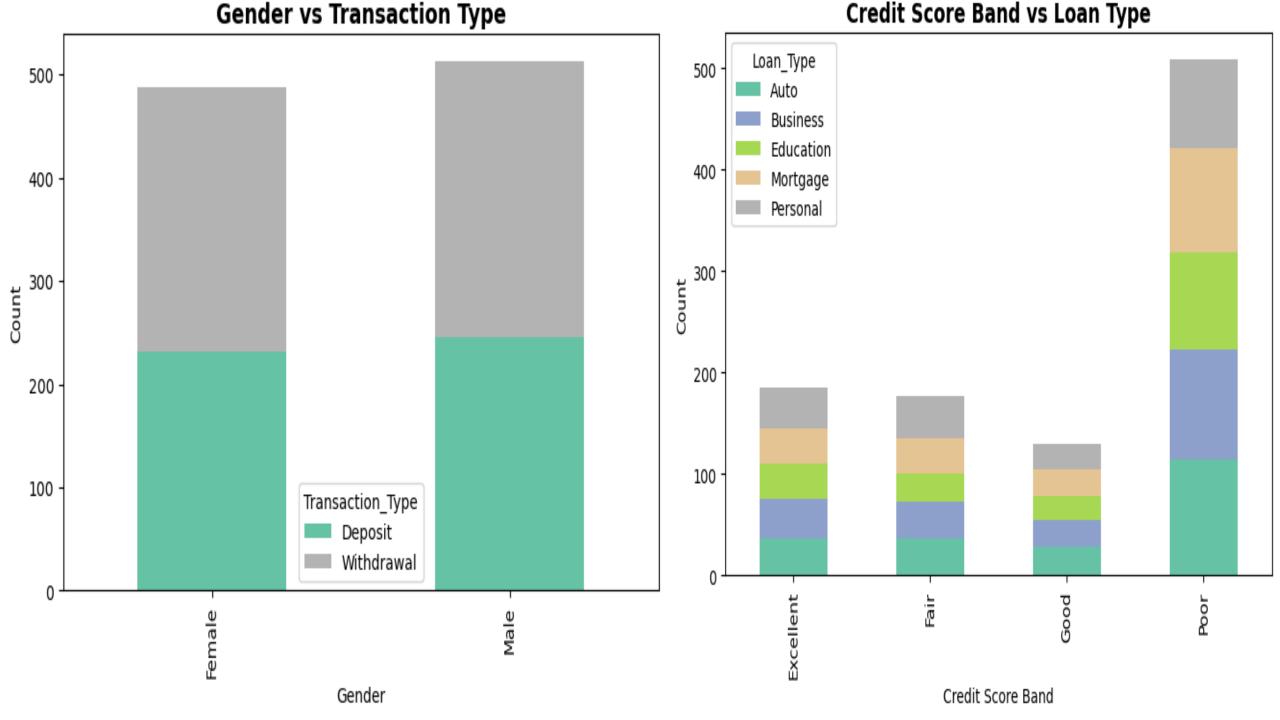










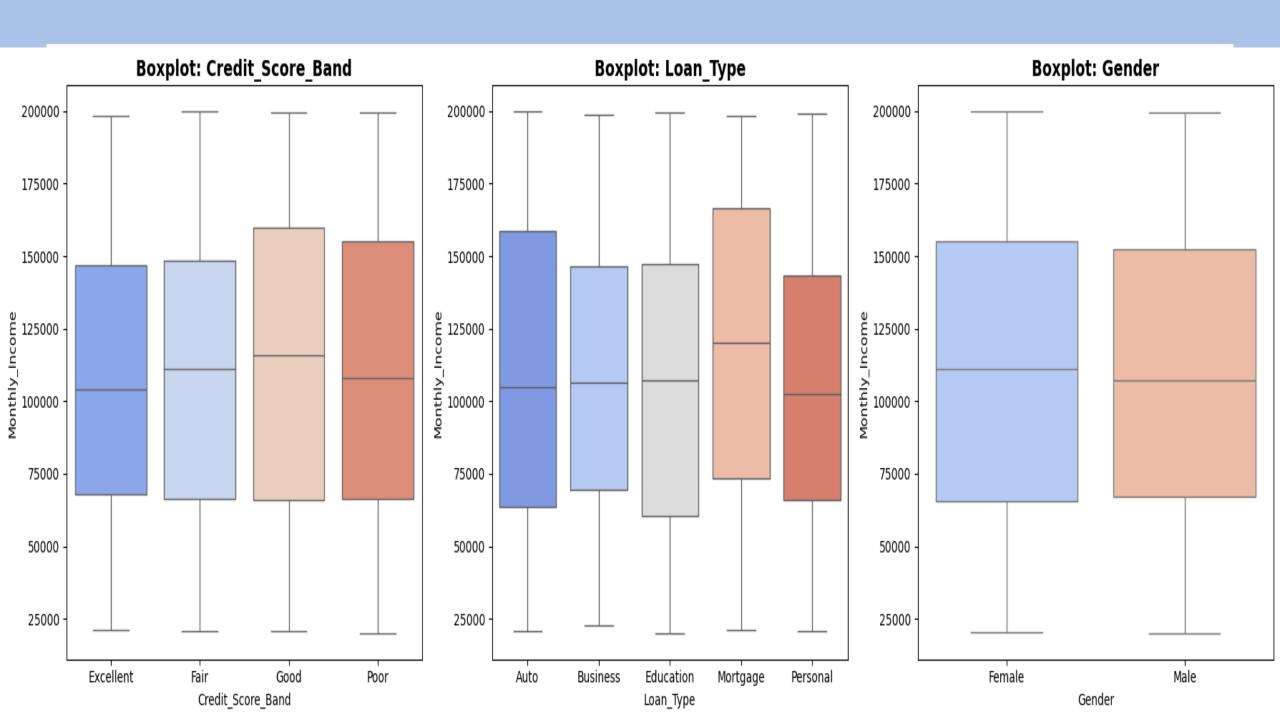


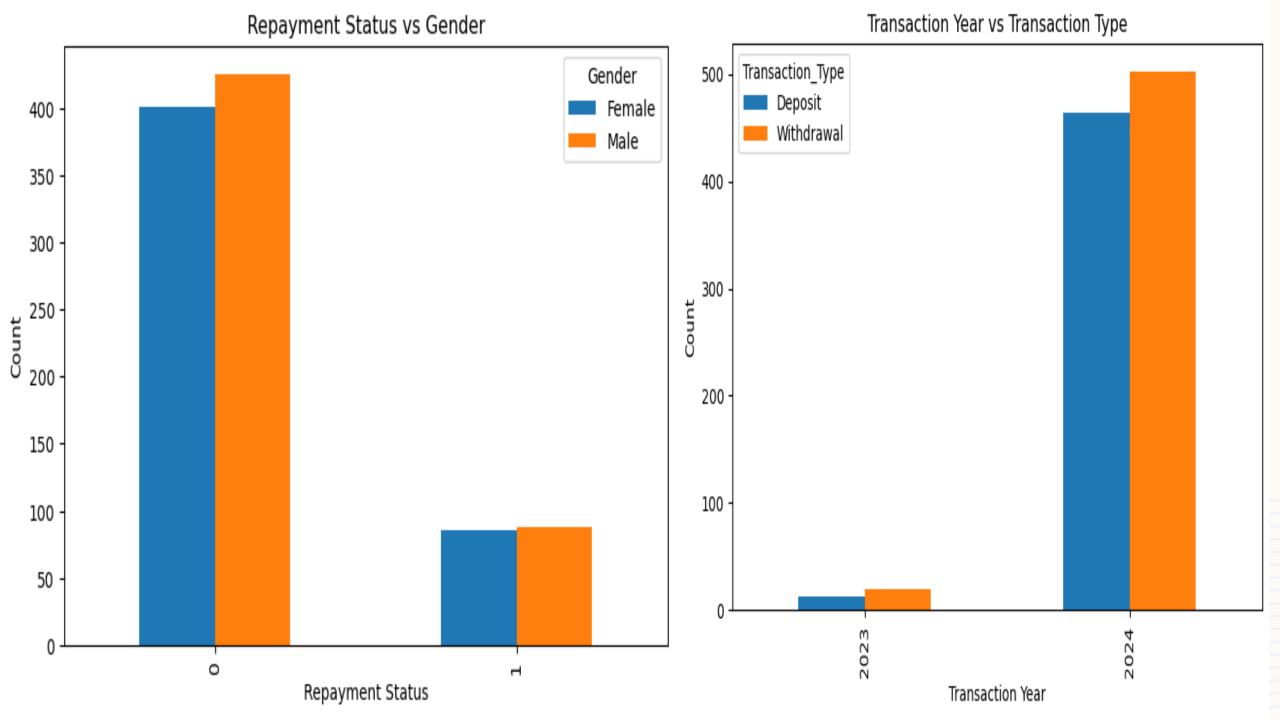
Correlation Heatmap

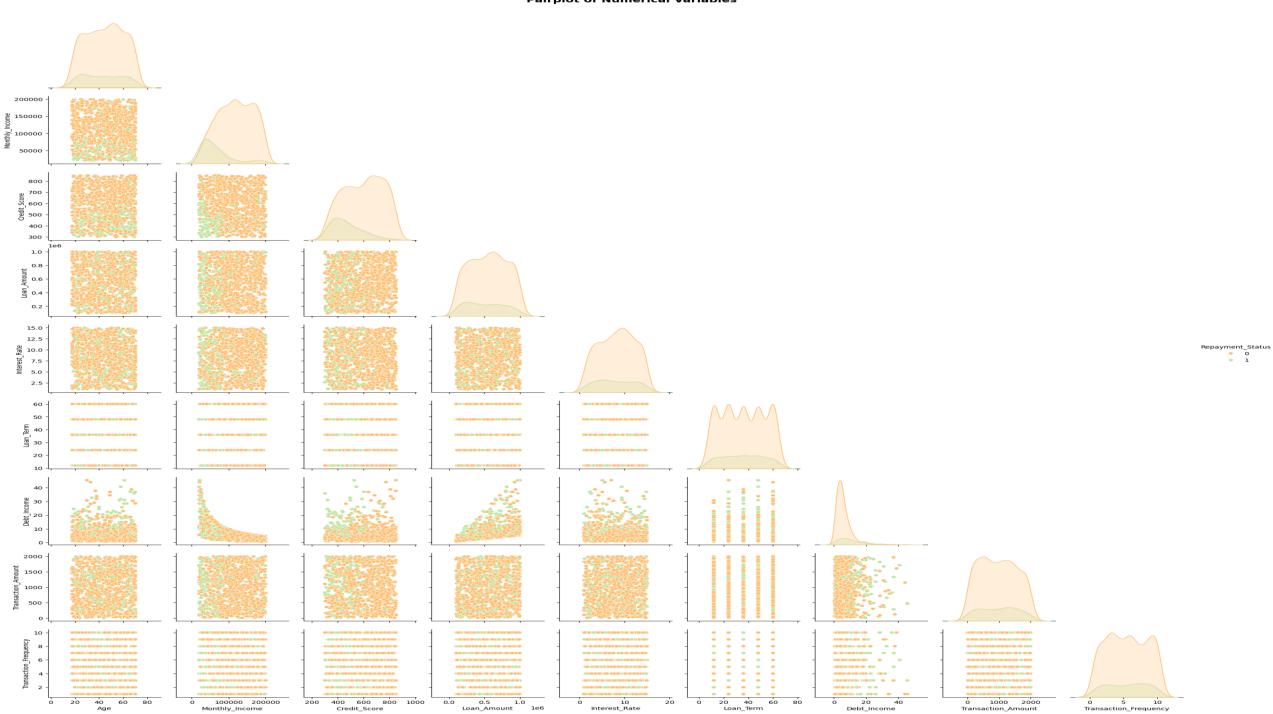
Age -	1	0.018	0.074	-0.028	-0.049	-0.062	-0.0059	-0.0076	-0.037	0.00084
Monthly_Income -	0.018	1	0.0028	0.034	0.067	-0.0059	-0.64	0.011	-0.38	0.0057
Credit_Score -	0.074	0.0028	1	0.059	-0.0068	-0.029	0.046	0.019	-0.32	-0.043
Loan_Amount -	-0.028	0.034	0.059	1	0.058	-0.0057	0.49	0.037	-0.063	0.015
Interest_Rate -	-0.049	0.067	-0.0068	0.058	1	0.0017	-0.04	0.026	-0.028	-0.01
Loan_Term -	-0.062	-0.0059	-0.029	-0.0057	0.0017	1	0.028	0.028	-0.011	-0.013
Debt_Income -	-0.0059	-0.64	0.046	0.49	-0.04	0.028	1	0.011	0.29	-0.011
Transaction_Amount -	-0.0076	0.011	0.019	0.037	0.026	0.028	0.011	1	0.0082	0.0045
Repayment_Status -	-0.037	-0.38	-0.32	-0.063	-0.028	-0.011	0.29	0.0082	1	-0.0055
Transaction_Frequency -	0.00084	0.0057	-0.043	0.015	-0.01	-0.013	-0.011	0.0045	-0.0055	1
	Age -	Monthly_Income -	Credit_Score -	Loan_Amount -	Interest_Rate -	Loan_Term -	Debt_Income -	Transaction_Amount -	Repayment_Status -	Transaction_Frequency -

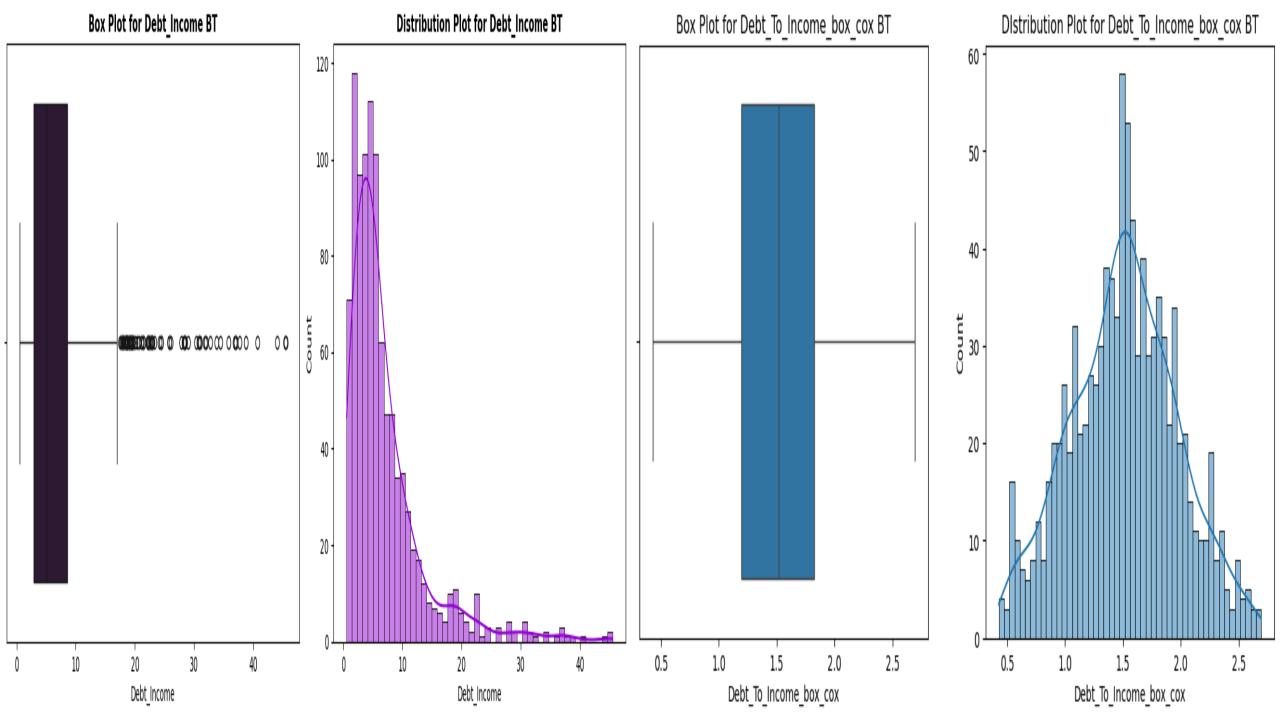
- 1.00 - 0.75 - 0.50 - 0.25 - 0.00 -0.25-0.50 - -0.75

- -1.00









INSIGHTS

- This pie chart displays the distribution of loan purposes among customers. It shows that the Mortgaes are the most common loan type followed by Education, Auto, Business and Personal
- This pie chart displays the distribution of credict score band among customers. It shows that the most of the customers are with poor credit score band than excellent, fair and good
- This histplot displays the distribution of numerical columns. It represents that Age, Monthly Income, Credit Score, Loan Amount, Interest Rate, Loan Term, Transaction Amount, Transaction Frequency are symmetric, whereas Debt_Income tends to be positively skewed.
- The Boxplot indicates outlier detection, showing that only the Debt_Income feature displays outliers where all other features appear to be free of outliers.

- The countplot displays the distribution of categorical features.
 Countplot for distribution of Credit Score Band:
 - The plot displays that most of the customers has Poor credit score band compared to Excellent, Fair, Good.
 - Countplot for Distribution of Loan Type:

The Mortgage loan type has the highest count, followed by Education, Auto, Business and Personal loans.

- Countplot for Distribution of Gender:

More Males have taken loans compared to females.

- Countplot for Distribution of Transaction Type:

Most of the transactions were Withdrawals compared to Deposits.

- Countplot for Distribution of Transaction Year:
- 2024 has the highest transaction count.
- Countplot for Distribution of Transaction Month:
 July has the highest transaction count compared to other months.
- Countplot for Interaction Year:
- 2024 shows high number of interactions than 2023.
- Countplot for Distribution of Interaction Day Name:

Tuesdays has the highest interaction count than other days.

- Countplot for Distribution of Interaction Type

The counts for clicked and purchased are higher than for viewed.

- The heat map displays the relationships between different variables related to loan data. This plot displays that
 - The Monthly Income and Debt Income are strong negative correlations.
 - The Debt Income and Repayment Status are strong positive correlations.
 - The Loan Amount and Debt Income are strong positive correlations.
- Customers with higher credit scores tend to have higher monthly income regardless of gender or loan type.
- Customers with higher incomes are more likely to have Mortgage loans.
- Gender has minimal impact on monthly income.
- The crosstab shows the distribution of Transaction Types:
 - Male have a higher count of deposit transactions compared to female.
 - Female have a higher count of withdrawal transactions compare to male.
- The crosstab shows the distribution of Loan Types across different Credit Score Brands:
 - Customer with Poor Credit Score Band has the highest overall number of Loans followed by Excellent, Good and Fair.

- The Crosstab displays the distribution of Repayment Status by Gender:
 - Both Female and Male have a similar number of individuals with Repayment Status 0.
 - Male have a slightly higher proportion of individuals with Repayment Status 1 compared to Female.
- The Crosstab displays the distribution of Transaction Type by Transaction Year:
 - In 2024, withdrawal transactions slightly outnumber deposits.
 - In 2023, withdrawal exceeds deposits.
- Boxplot displays outliers in the Debt Income.
- Histogram illustrates that the Debt Income is positively skewed.
- After the transformations, both outliers and skewness have been addressed.

THANK YOU