CLUSTERING AND FITTING
NAME:PUDARI SAI PAVAN
ID:23100550
Github:https://github.com/Pudarisaipavan/Name--PUDARI-SAI-PAVAN-Id--23100550-.git

Introduction

The Credit Card Clustering Dataset is a reservoir of comprehensible data to explore consumers as well as their transactions involving credit cards. Through these clustering techniques, this data set can help researchers and analysts to uncover patterns for different types of customers for their expenditure by working with various behavioral patterns for conducting experiments. It has quantitative data that distinguishes spending types, frequencies, and payment patterns to analyze various customer segments. Customer segmentation strategies also benefit from this type of dataset because it will assist businesses to market themselves differently based on the needs of the segment that is being targeted. For instance, customers who spend a lot can be encouraged by the company to join elite loyalty programs, at the same time, old users who are not active will be encouraged to use their accounts more frequently through reactivation campaigns. In addition, supervising this data set would allow organizational contexts to improve customer loyalty strategies, and use churn tendencies to develop customer-specific product portfolios. The data set is useful to data scientists, marketer, and financial analysts who would like to incorporate machine learning into their business planning and execution. Its practical applicability extends to burning issues right from the modelling of promotional campaigns to maximization of credit card rewards to customer satisfaction. Thus, this dataset is the most considerable source of information to make decisions in the credit card business in the conditions of increasing competition.

Histogram of Customer Balances

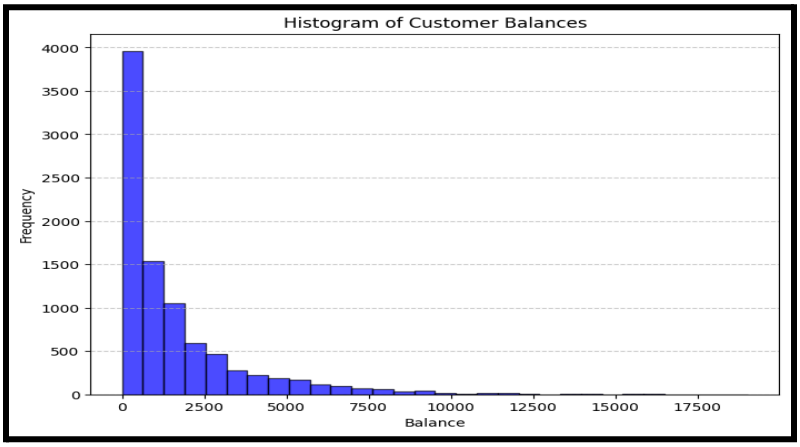

Figure 1: Histogram of Customer Balances

The above figure entitled: ‘balance distribution of customers’ by the use of a histogram. They need an input dataset, which plots the “BALANCE” column and focuses consumers on how often various ranges of balance occur. The histogram is labelled with 30 bins, blue colour for bars and black for edges to enhance the understanding. It is helpful to add other lines in the form of dashed gridlines to the diagram. To make the chart interpretable a title is provided on top of the X-axis labeled ‘Balance’ and on the Y axis labeled ‘Frequency.’ The above developed visualization will be handy in pattern searching of a balance distribution and outliers that will enhance the customers’ financial behavior understanding.

Scatter Plot: Purchases vs Payments

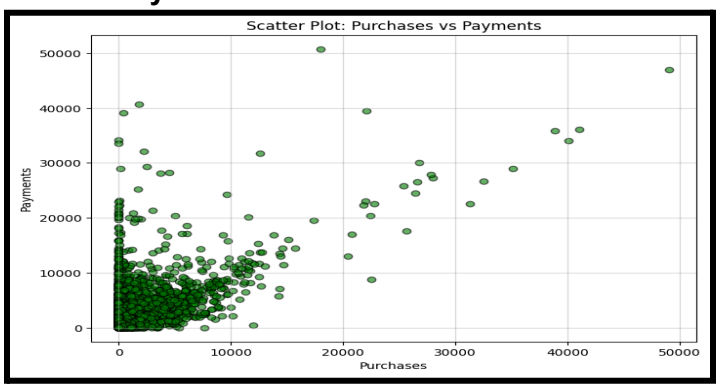

Figure 2: Scatter Plot: Purchases vs Payments

This figure gives a scatter plot that helps in comparing the variables “PURCHASES” and “PAYMENTS” in the dataset. This is represented in the sense that the green data points have black edges and that which is plotted are trends and correlations. The plot is provided with the axes, title, and grid for better distinction. The visualization makes it easier to see things like what happens if more purchases tend to be made – are higher payments made to the business as well and this is useful in understanding the customer’s behavior.

Correlation Heatmap

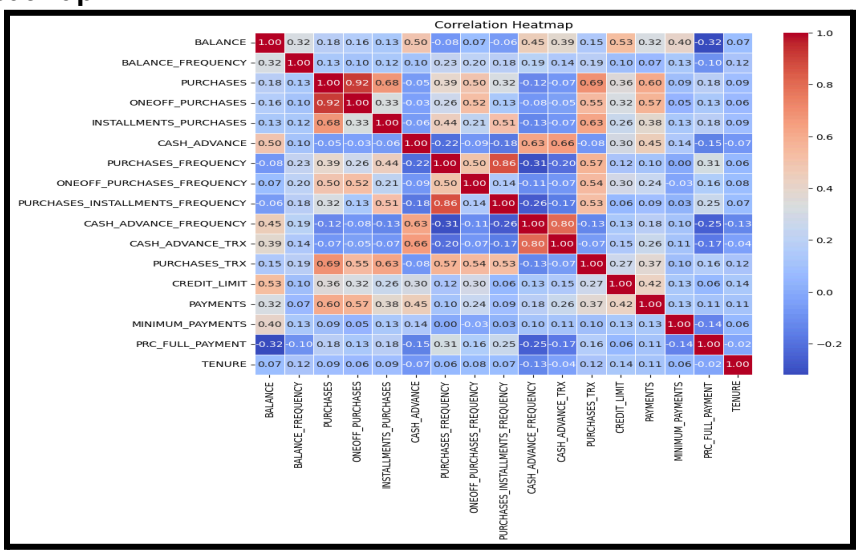

Figure 3: Correlation Heat map

The plot_correlation_heatmap(data) function will result in the correlation matrix of the selected dataset in a heatmap. It maps out the degree of association and the direction of association between variables with values written for the purpose. It the heatmap has a contrasting cmap called ‘coolwarm’ and it also has grids to distinguish the cells. This tool is useful in finding highly correlated features, feature selection, and understanding the interaction of variables in the dataset.

K-Means Clustering

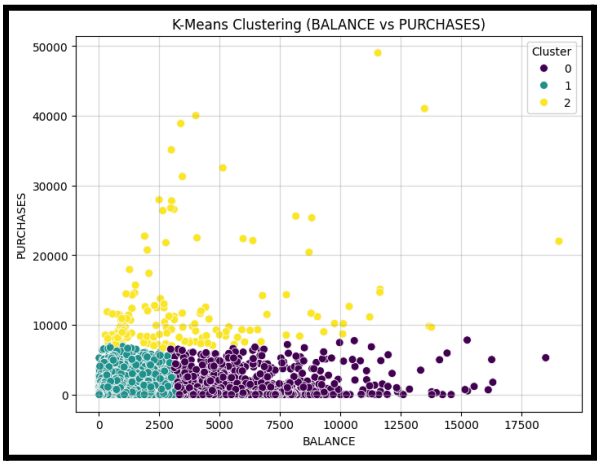

Figure 5: K-Means Clustering

This figure produces an elbow plot to find the optimal number of clusters for k-means clustering. It computes distortions (inertia) for cluster counts from 1 to 10 and plots the results. The plot points out the "elbow point," where inertia drops drastically, meaning the ideal cluster count. This method helps to choose the right k-value for effective clustering of features like "BALANCE" and "PURCHASES."

Linear Regression: Purchases vs. Payments

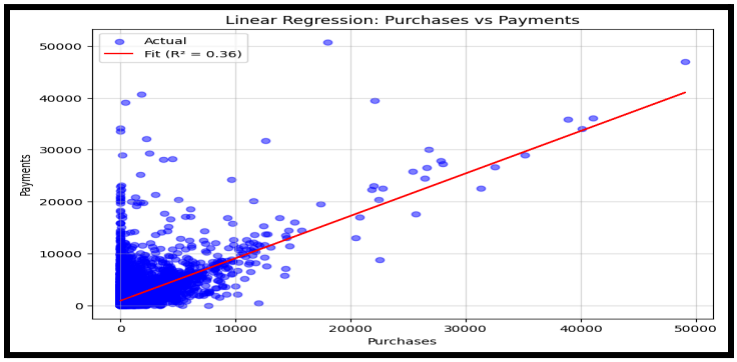

Figure 6: Linear Regression: Purchases vs Payments

The above figure represents the linear regression between "PURCHASES" and "PAYMENTS." It draws the appropriate fitted line through the data points and also reports on the R^2 of how well the fitted line explains the data. This includes axis labeling, title, and a legend for easy identification of plotted features. It can thus describe a rough measure of how good this relationship is between these two variables in terms of magnitude and direction.

Conclusion

The Credit Card Clustering Dataset offers deep insights into customer behavior, thus making effective segmentation, targeted marketing, and strategic decision-making possible. The visualization with histograms, scatter plots, heatmaps, and regression analysis helps discover patterns, optimize customer engagement, and enhance financial strategies. Therefore, it is priceless for credit card data-driven decision-making.