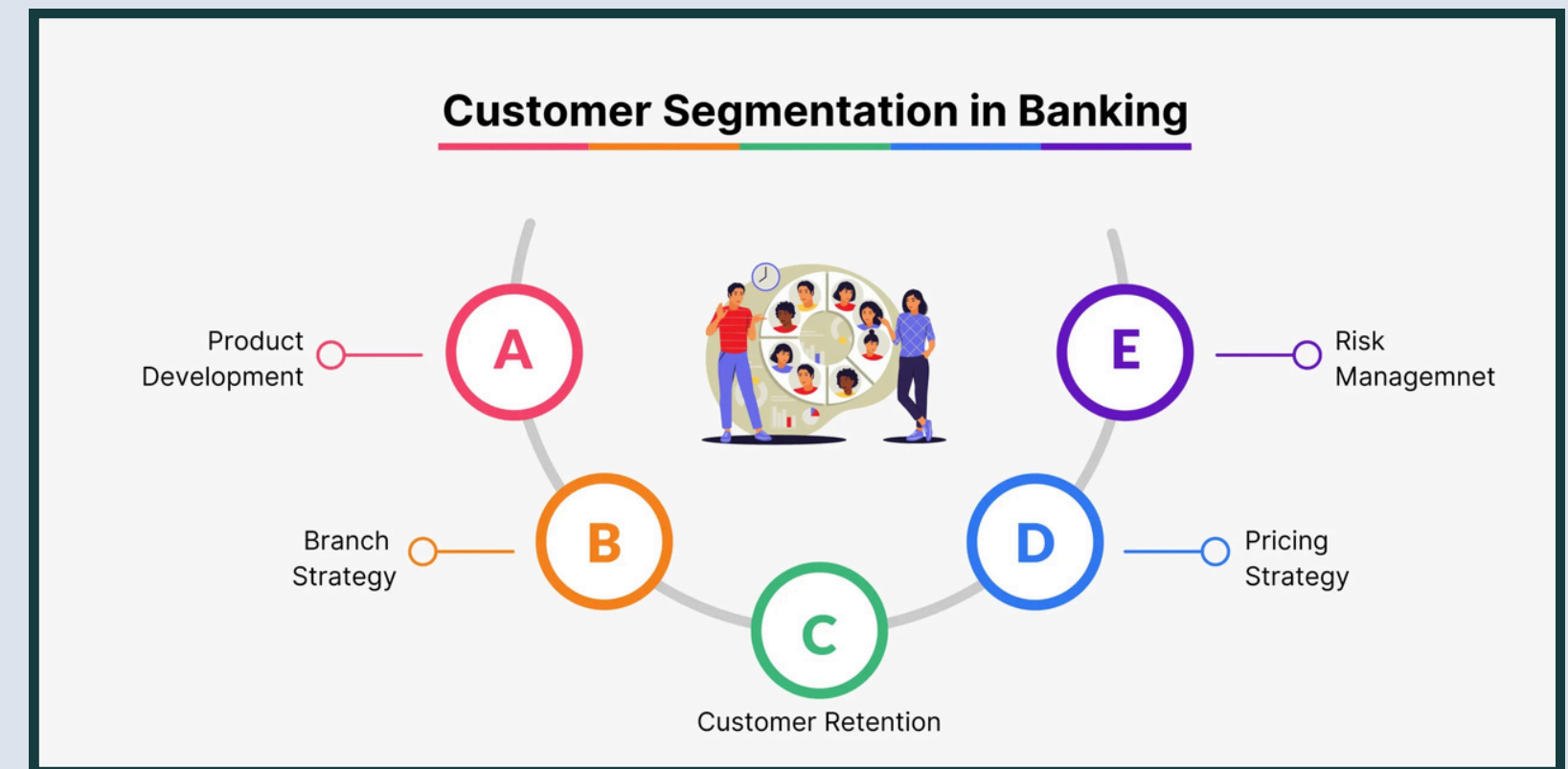# A STUDY ON CUSTOMER SEGMENTATION MODEL BASED ON RISK

Sarthak Sawant - 05

Vishakha Shelke - 06

INTRODUCTION

In today's data-driven banking environment, understanding customer behavior is crucial for managing risk and providing personalized financial services. This project focuses on customer segmentation based on behaviors, transaction patterns, and risk profiles, enabling banks to improve risk management, enhance fraud detection, and deliver more tailored products and services.

Customer Segmentation in Banking

Product Development — A
Branch Strategy — B
C — Customer Retention
D — Pricing Strategy
E — Risk Managemnet

PROBLEM STATEMENT

Banks often face challenges in identifying and managing customer risks, leading to poor credit decisions and higher default rates. To address this, the project aims to develop a risk-based customer segmentation model that groups customers by analyzing factors like credit scores, credit limits, repayment history, and spending patterns. This approach will help banks better understand their customers,make more informed decisions as well as developing tailored marketing strategies and personalized products, and reduce financial losses by targeting high-risk customers with appropriate strategies, while also optimizing offerings for low-risk customers and providing tailored solutions for those in the middle-risk category.

DATASET DESCRIPTION

Lending Club is a peer-to-peer Lending company based in the USA. The Lending Club dataset contains complete loan data for all loans issued through the 2007-2015, including the current loan status (Current, Late, Fully Paid, etc.) and latest payment information.
There are about 145 variables each consisting of around 2.26 million observations.

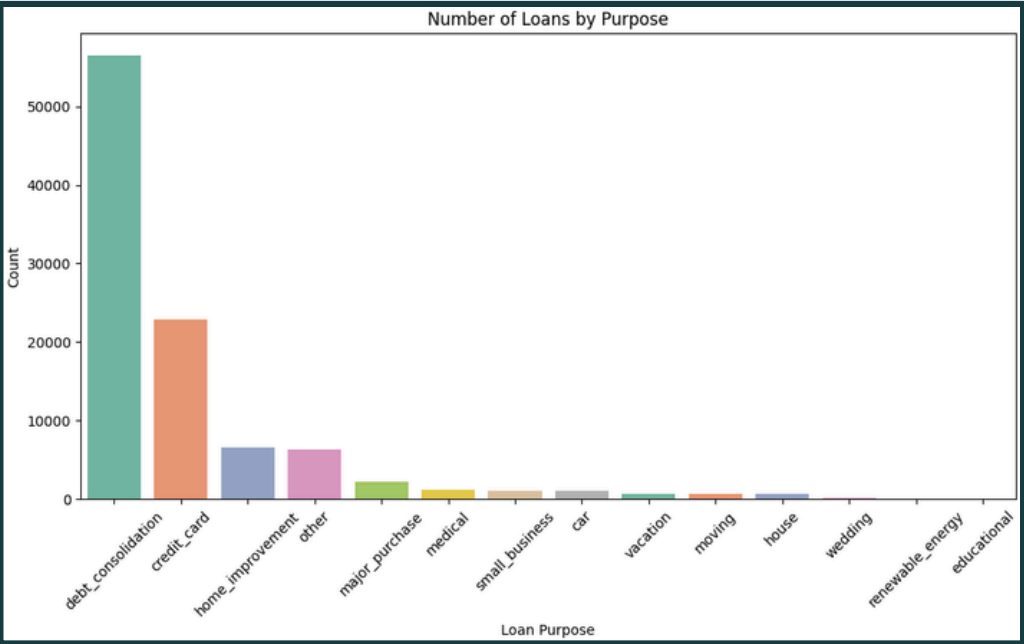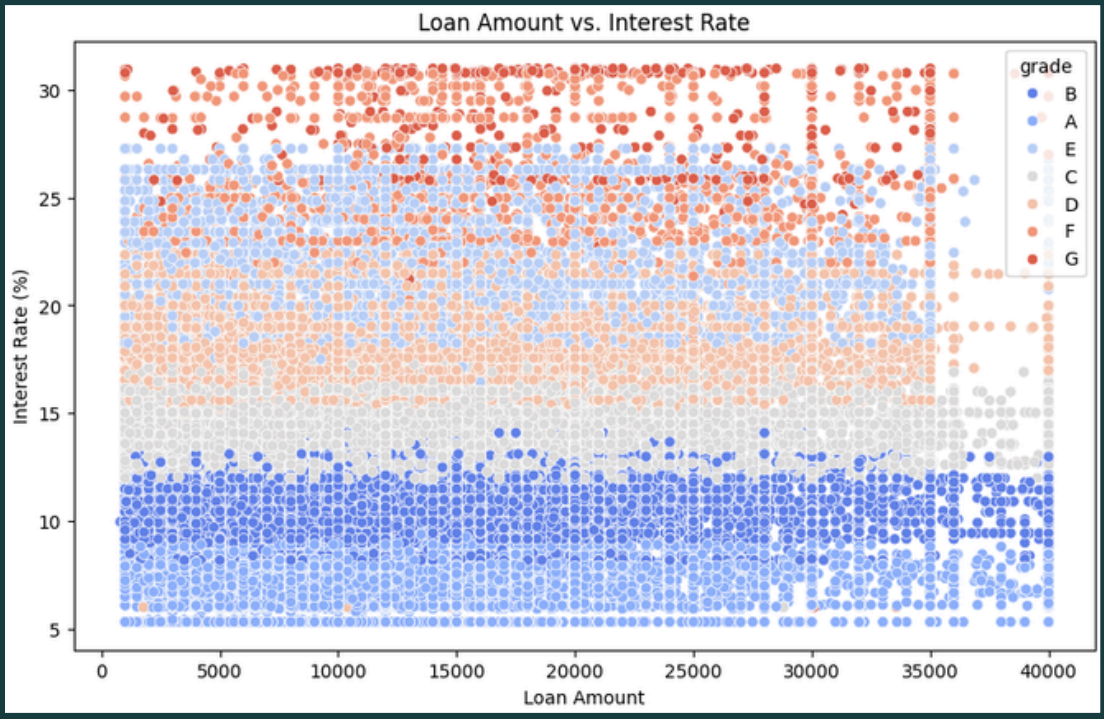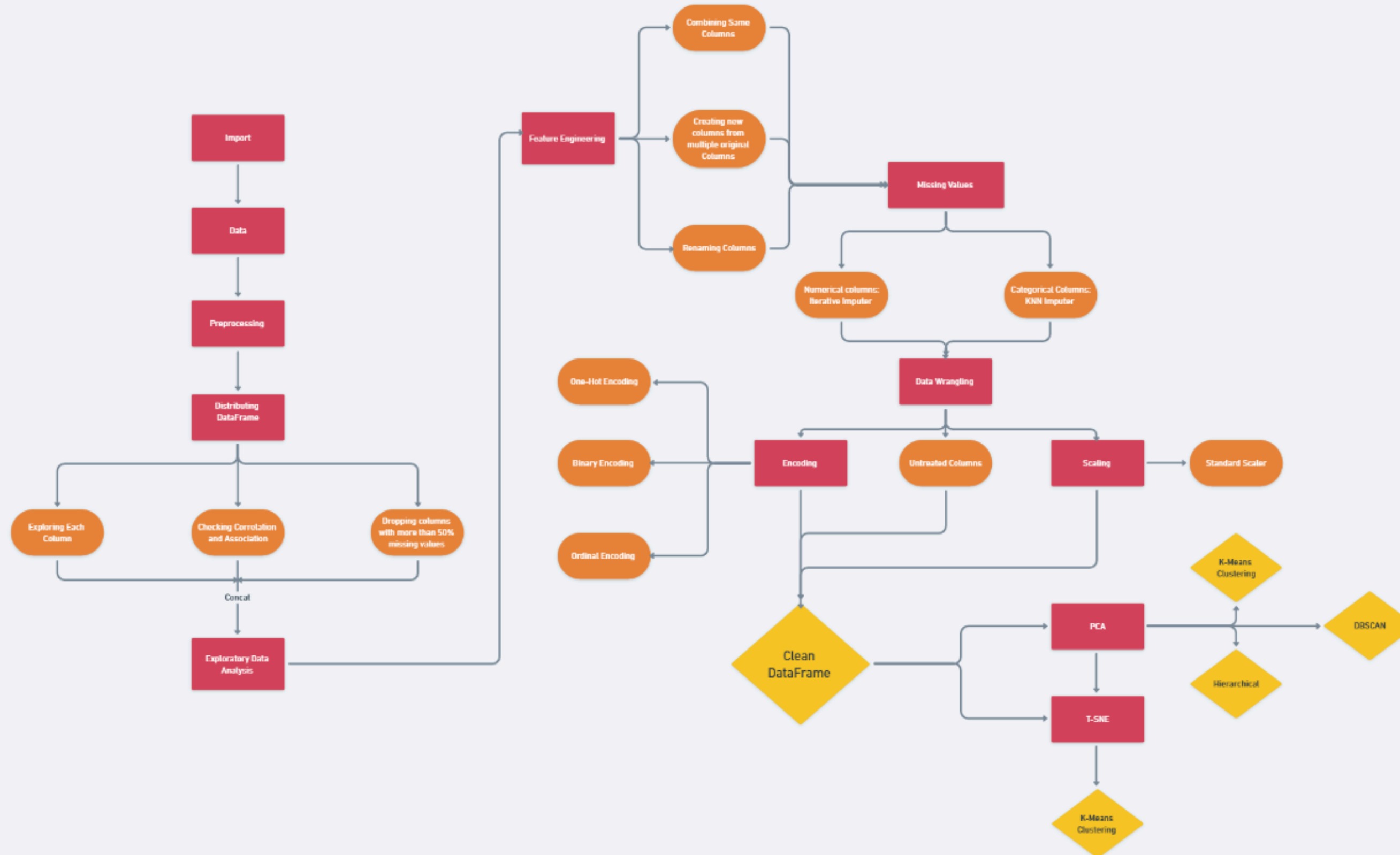|		
Borrower Information	**Loan Information**	**Credit Utilization**
Repayment History	**Income & Employment**	**Lender Information**

Note: Due to problems related to heavy computation using Google Colab, a sample of the data was considered for data analysis.
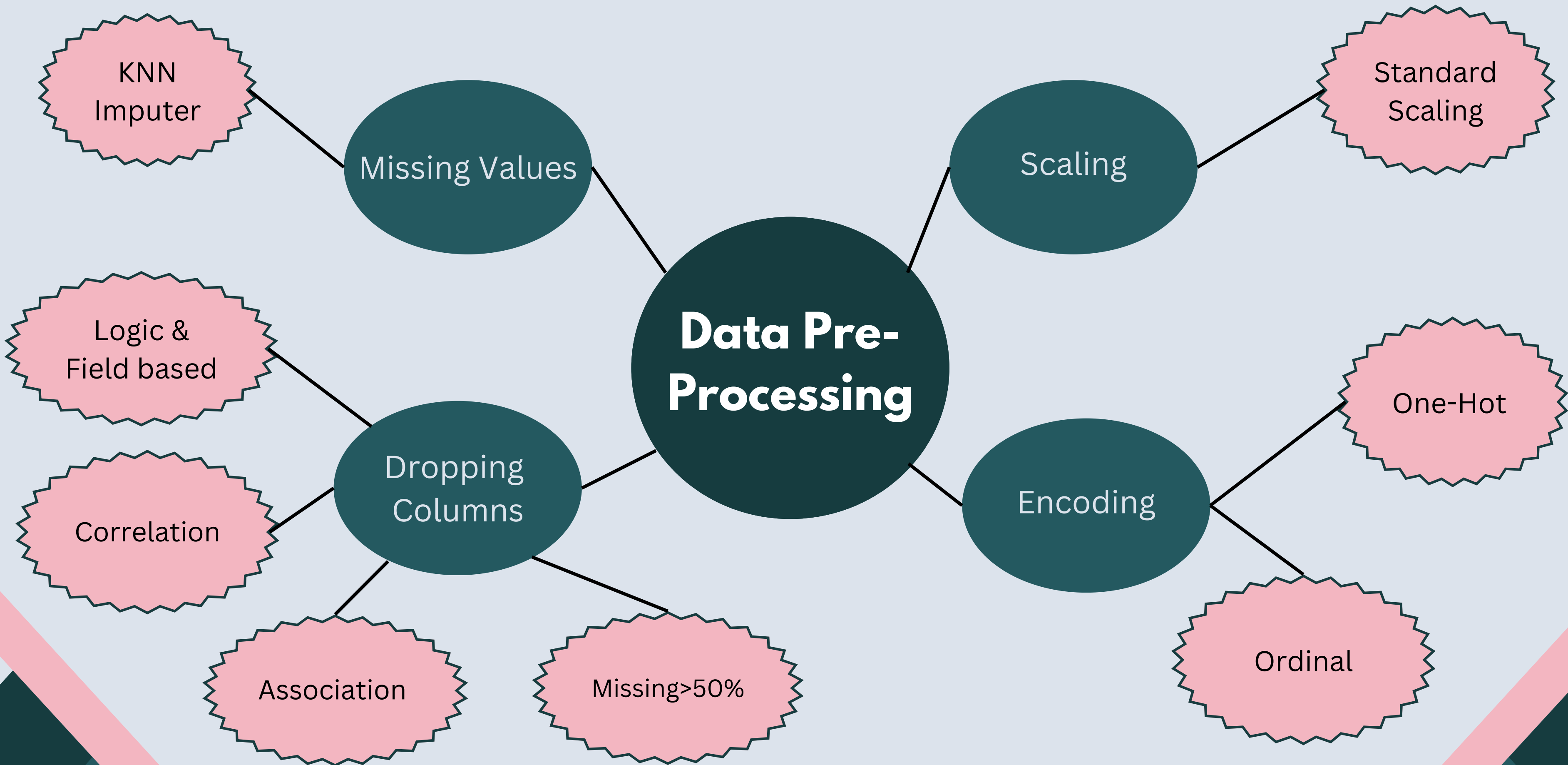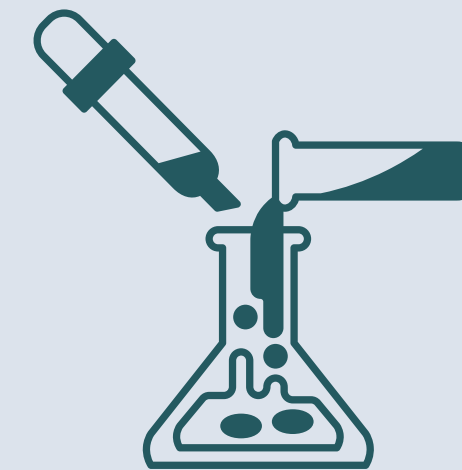
EDA
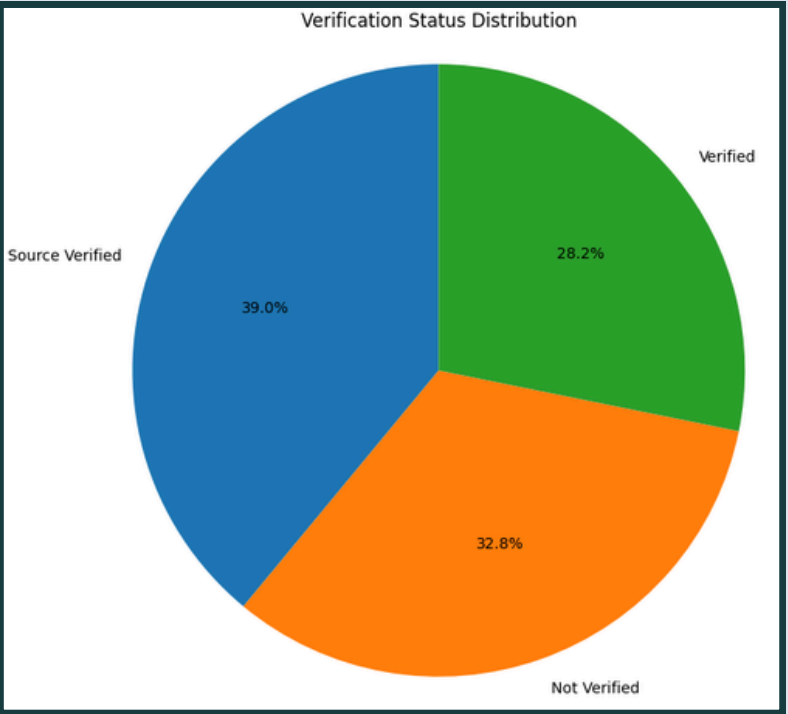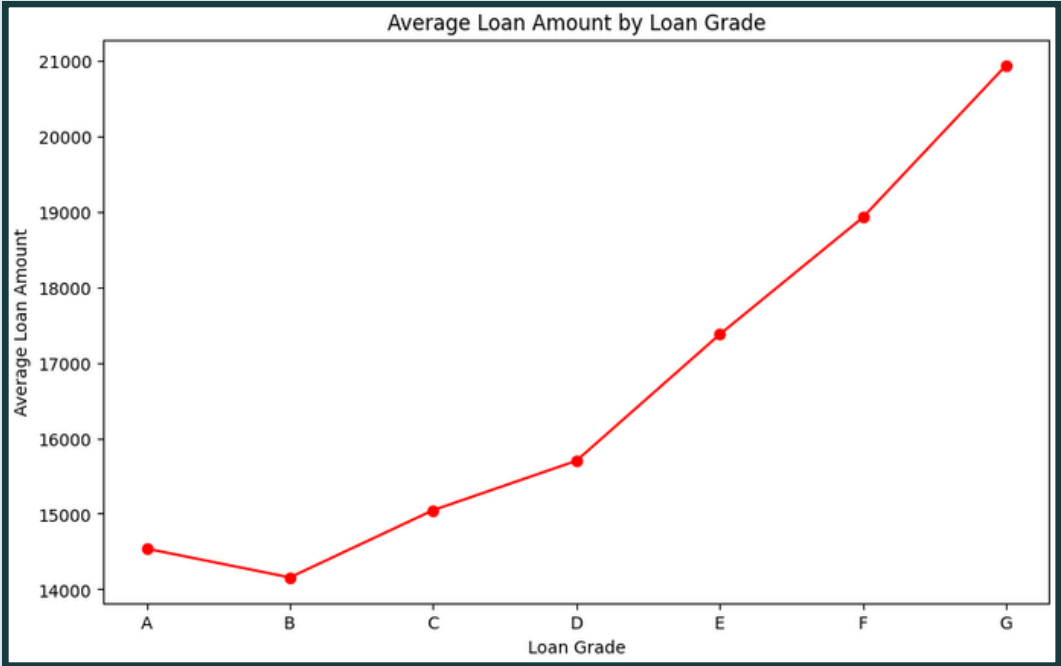
RESEARCH METHODOLOGY

DATA PRE-PROCESSING

Data Pre-Processing

Missing Values
- KNN Imputer

Scaling
- Standard Scaling

Dropping Columns
- Logic & Field based
- Correlation
- Association
- Missing>50%

Encoding
- One-Hot
- Ordinal

FEATURE ENGINEERING

Combining Similar Columns

Calculating Rates

Delinquency Flag

Calculating Ratios

Credit Age

Combining Different Columns

CLUSTERING ALGORITHM

PCA was applied for dimensionality reduction before fitting the clustering models. Multiple algorithms, including K-Means, Hierarchical Clustering, and DBSCAN, were tested, with K-Means emerging as the best performing algorithm and selected for the final model.
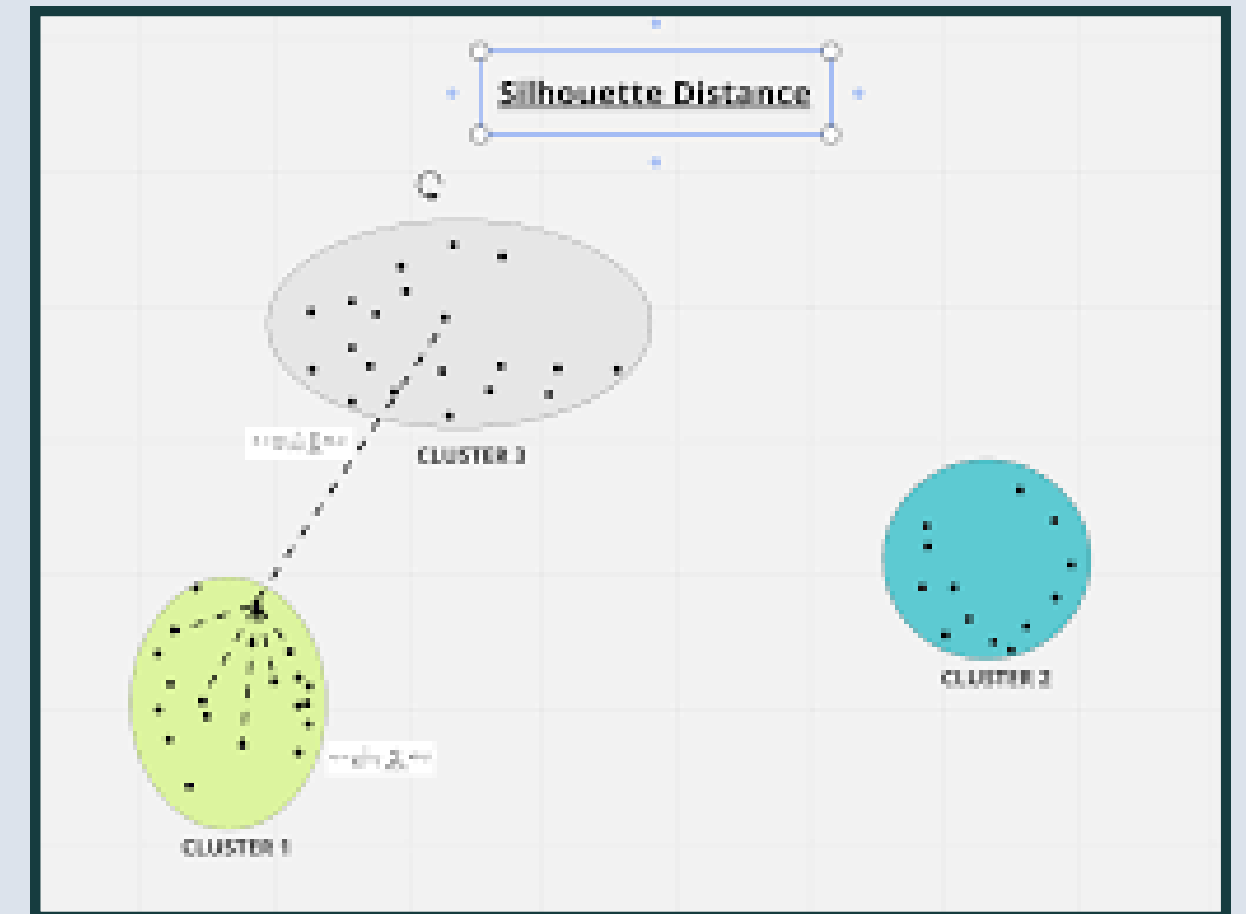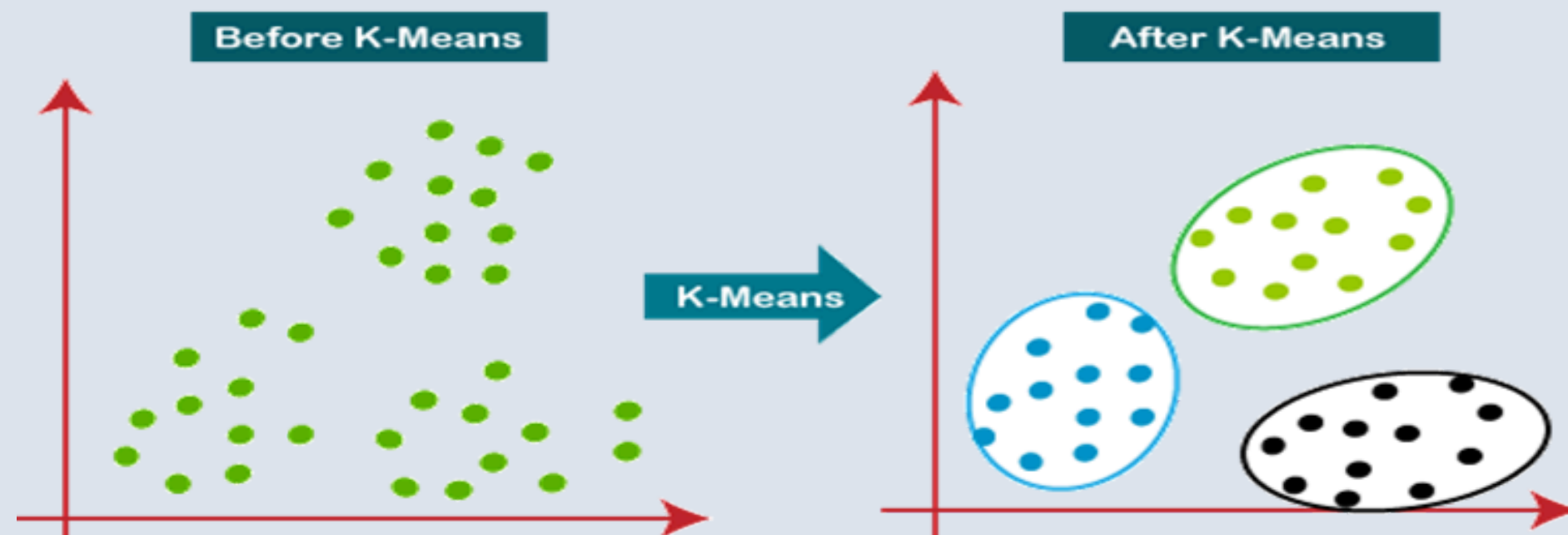

Note: Silhouette Score was used to evaluate how well data points fit within their assigned clusters. Based on this score, the optimal number of clusters was determined to be k=3.

FINDINGS

Main factors affecting PCA & Clusters:

- Loan Grade
- Credit exposure
- Delinquency flag
- Financial stability features
- Recent Account Activity
- Debt-to-Income Ratio
- Annual Income & Loan Amount

KMeans Clustering on PCA-Reduced Data (2D)
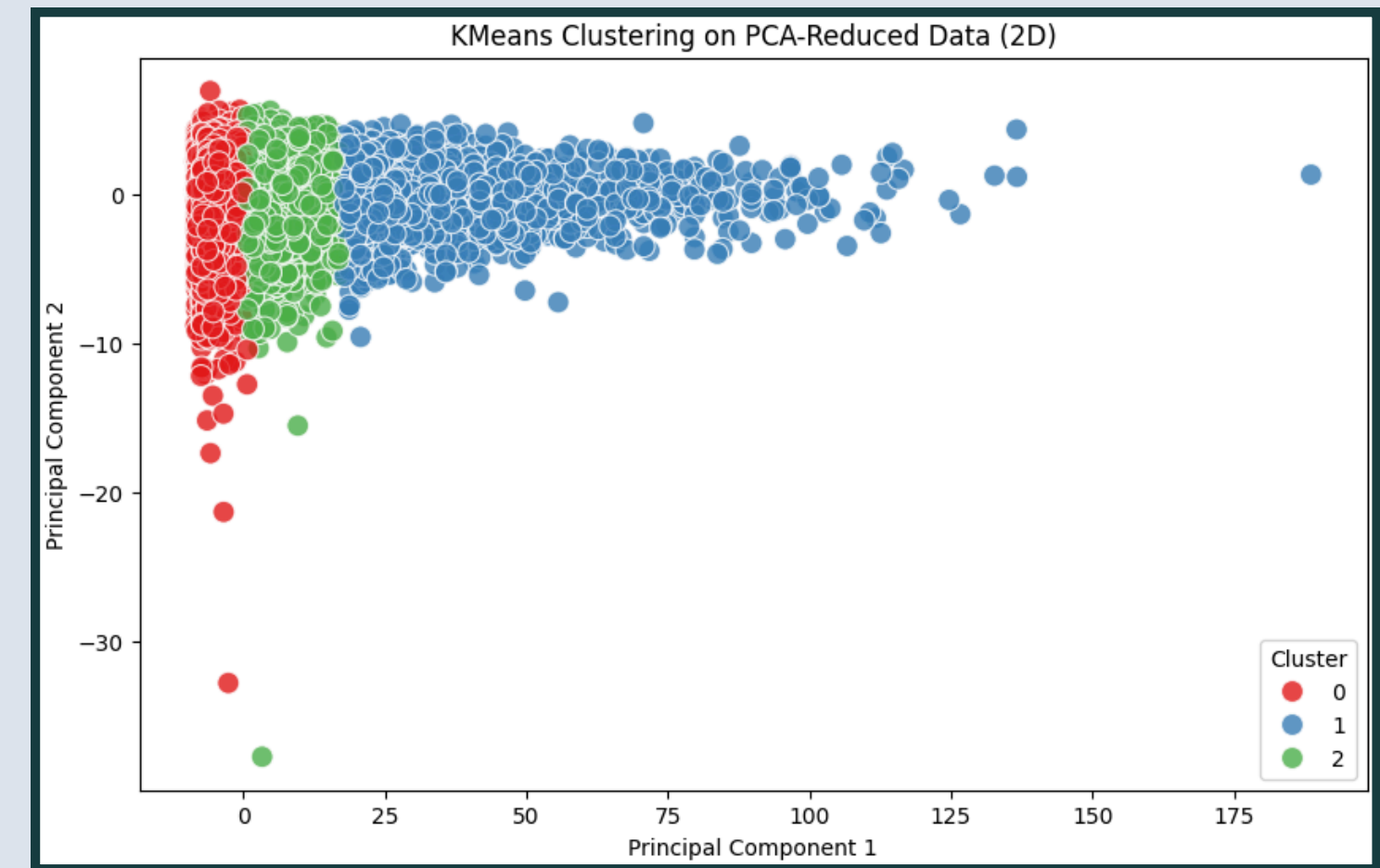
Cluster 0 : Low Risk

Cluster 2 : Medium Risk

Cluster 1 : High Risk

LIMITATIONS & FUTURE SCOPE

Algorithms of Model

Personalized Financial Products and Services

Time Related Elements

Dataset Limitations

Credit Underwriting and Risk-Adjusted Marketing

Customer Retention and Relationship Management

Geographical and Contextual Constraints

Enhanced Fraud Detection

CONCLUSION

This project aimed to create a risk-based customer segmentation model to better understand the financial behavior of customers and assess their potential credit risk. By leveraging advanced techniques such as PCA (Principal Component Analysis) and K-Means clustering, we segmented the dataset into meaningful clusters, each representing distinct risk profiles. These clusters provided insights into customers' credit usage, financial health, and overall stability, all of which are critical for risk management and lending strategies.

THANK YOU