Oluseyi Ajayi

Home | About | Contact

Loan Default Prediction

01

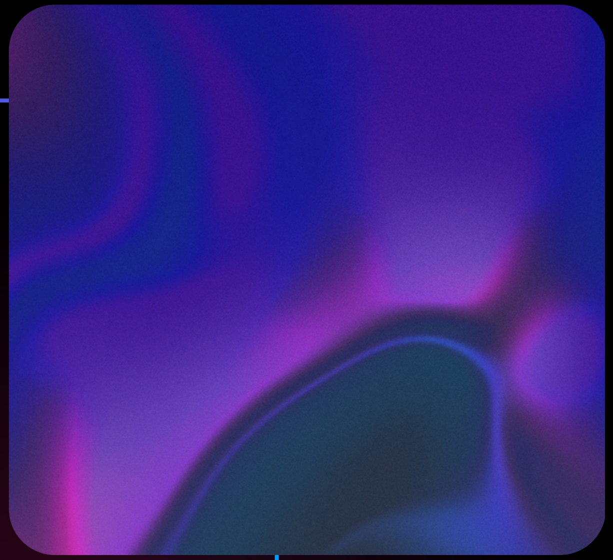
02

03

Start Slide

May 2025

Lot of organisation facs significant financial risk when borrower fails to repay their loans as at when due. The need for us to be able to predict this likelihood can not be overemphasized, hence the need for a loan default model.



Learn More

Introduction

May 2025

01

02

03

Dataset Overview

i analyze a loan origination dataset comprising over 3000 records of personal loan booked between January 2023 and November 2023. The dataset contains key variables like loan amount, income, CRC data and other bureaus and also demography and behavioural fatures

Learn More

01

02

03

About the Dataset

We have the dataset distributed into

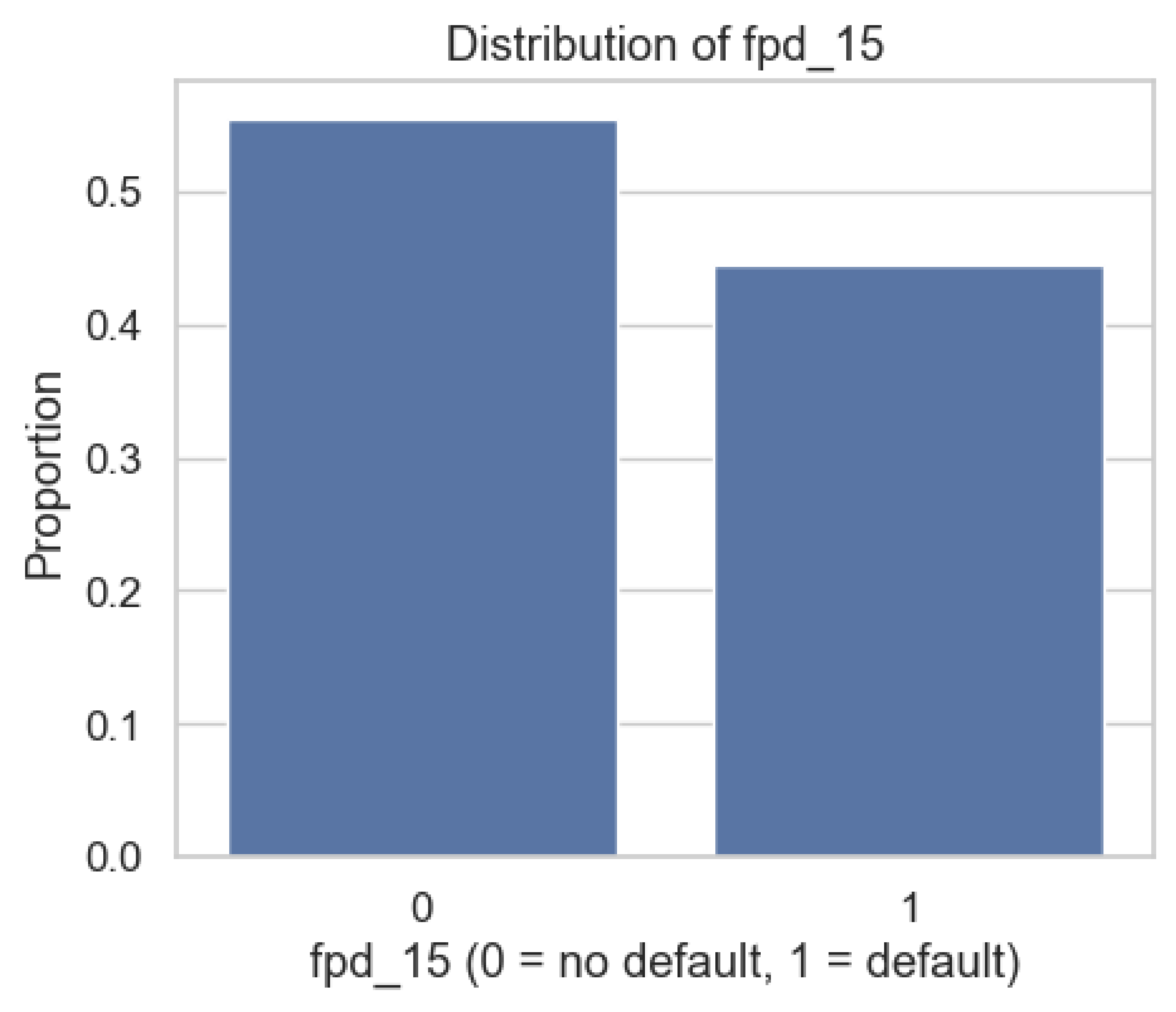
1. Numeric features
2. Categorical Features
3. Date

Variable Type	Fields
Numeric	LoanAmount, data.Request.Input.CB2.MaxDPD, data.Request.Input.CB2.CurrentDPD, data.Request.Input.CB2.Outstandingloan, data.Request.Input.CB1.MaxDPD, data.Request.Input.CB1.CurrentDPD, data.Request.Input.CB1.Outstandingloan, data.Request.Input.Customer.TotalExistingExposure, data.Request.Input.Customer.Income.Final, data.Request.Input.Customer.NumberOf data.Request.Input.Customer.TimeAtAddressMM, data.Request.Input.SalaryService.MonthlyElectricitySpending4, data.Request.Input.SalaryService.MinimumBalance, data.Request.Input.SalaryService.MinimumCredit, data.Request.Input.SalaryService.AvgNumDebitMn, data.Request.Input.SalaryService.MonthlyCashFlow2, data.Request.Input.SalaryService.MonthlyCashFlow3, data.Request.Input.PrevApplication.LoanAmount, data.Request.Input.PrevApplication.LoanTerm, data.Request.Input.PrevApplication.InterestRate, data.Request.Input.Application.RequestedLoanTerm, data.Request.Input.SalaryService.OpeningBalance, fpd_15
Categorical	data.Request.Input.CB2.Labelling, data.Request.Input.CB1.Labelling, data.Request.Input.Customer.EducationStatus, data.Request.Input.Customer.Gende data.Request.Input.Customer.AddressLGA, data.Request.Input.Customer.Employment.EmployerLGA, data.Request.Input.Customer.Employment.IndustrySector, data.Request.Input.Customer.Employment.BusinessSector, data.Request.Input.Customer.ResidentialStatus, data.Request.Input.Customer.Mar data.Request.Input.Customer.ExistingCustomer, data.Request.Input.BVN.StateOfOr
Date	CreationDate, data.Request.Input.Customer.DateOfBirth

EDA Highlights

Target Feature Balance

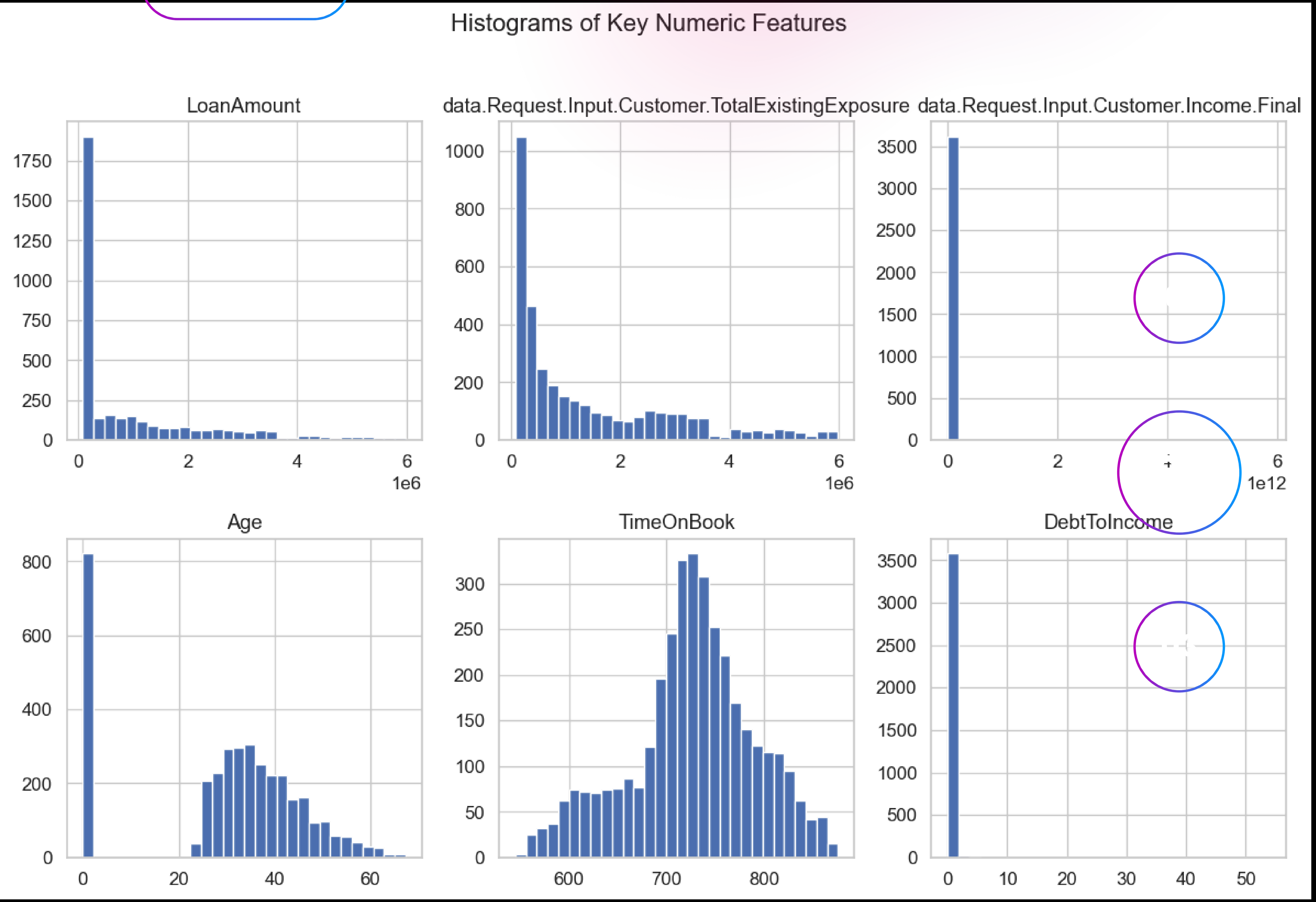
This shows the target(fpd_15) distribution proportion showing default as 1 and no default as 0



EDA Highlights

Numeric Features distributions

This shows the feature distributions for numeric columns like Age, TimeOnBook, DebtToIncome,Loan amount, etc

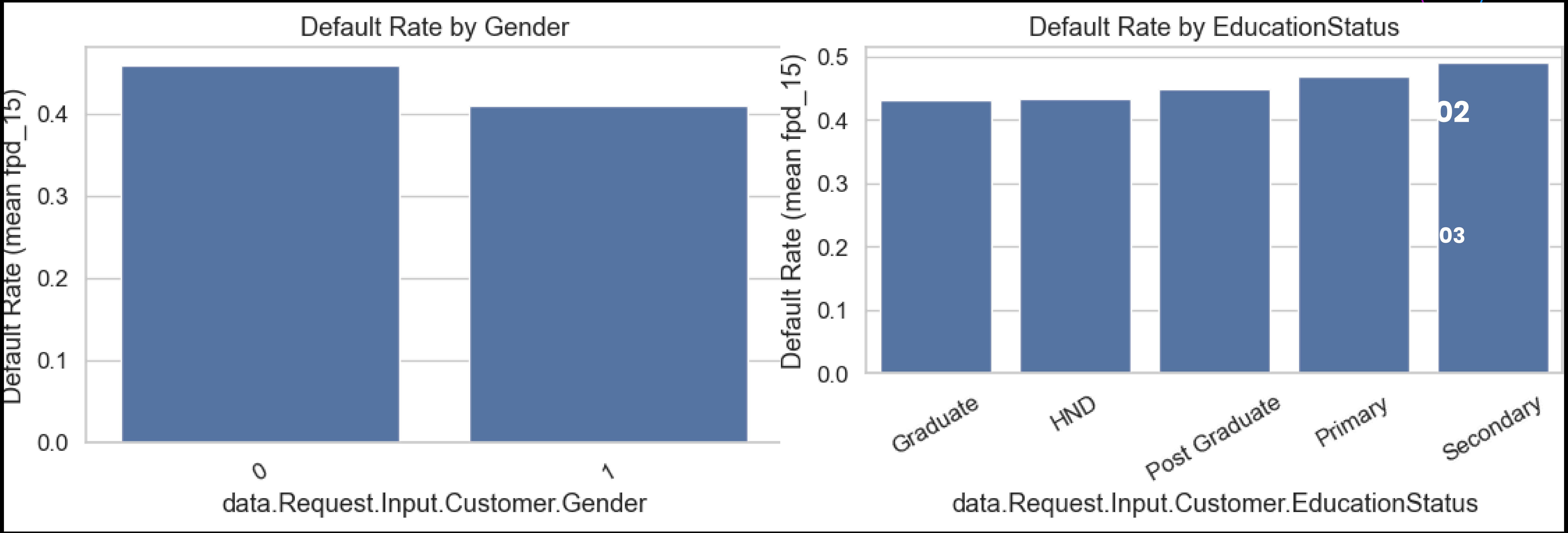


EDA Highlights

Default Rate by some categorical features

Showing relationship with Gender, EducationStatus ,marital status, CB labels

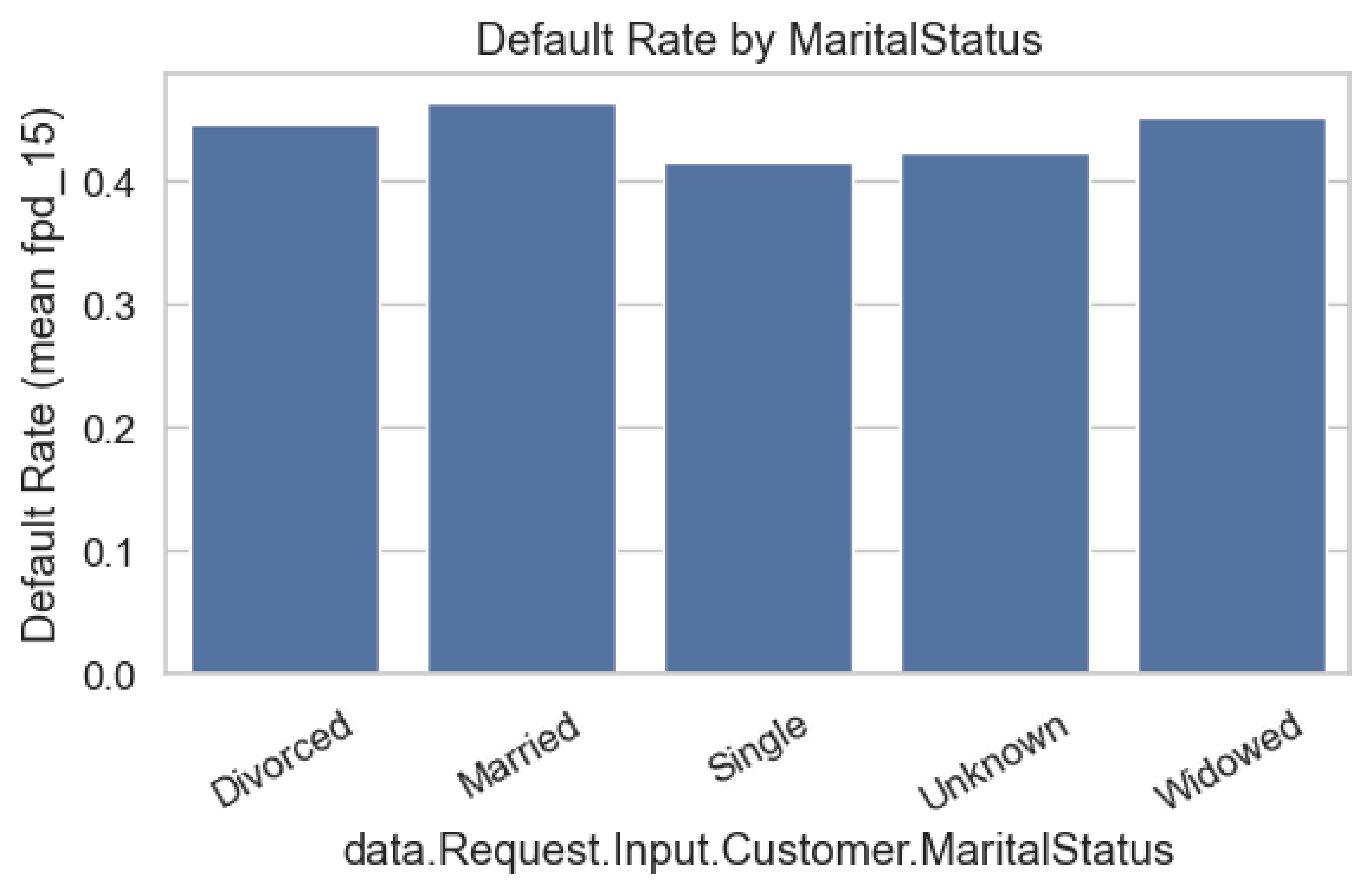
01

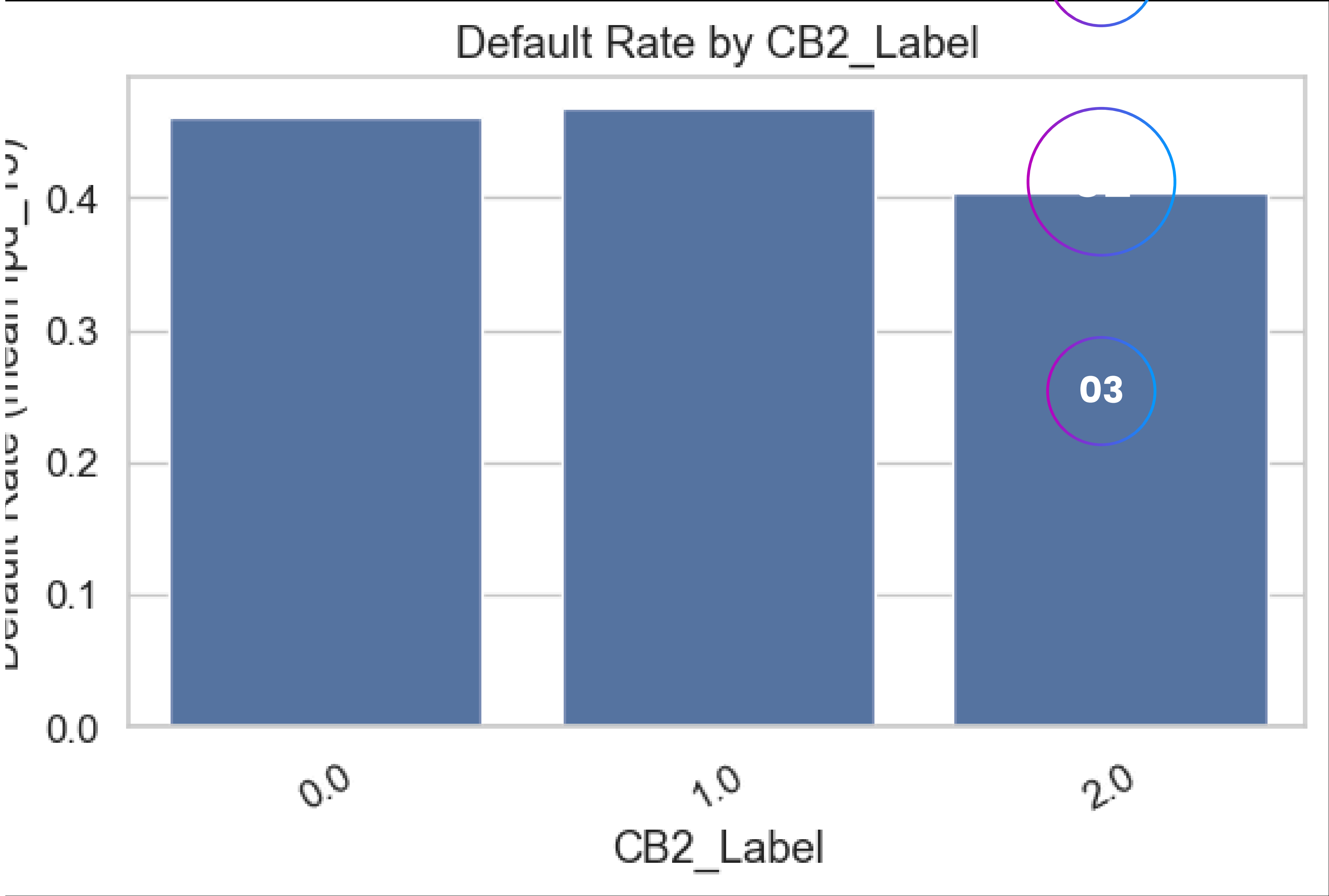


EDA Highlights

Default Rate by some categorical features

Showing relationship with Gender, EducationStatus ,marital status, CB labels



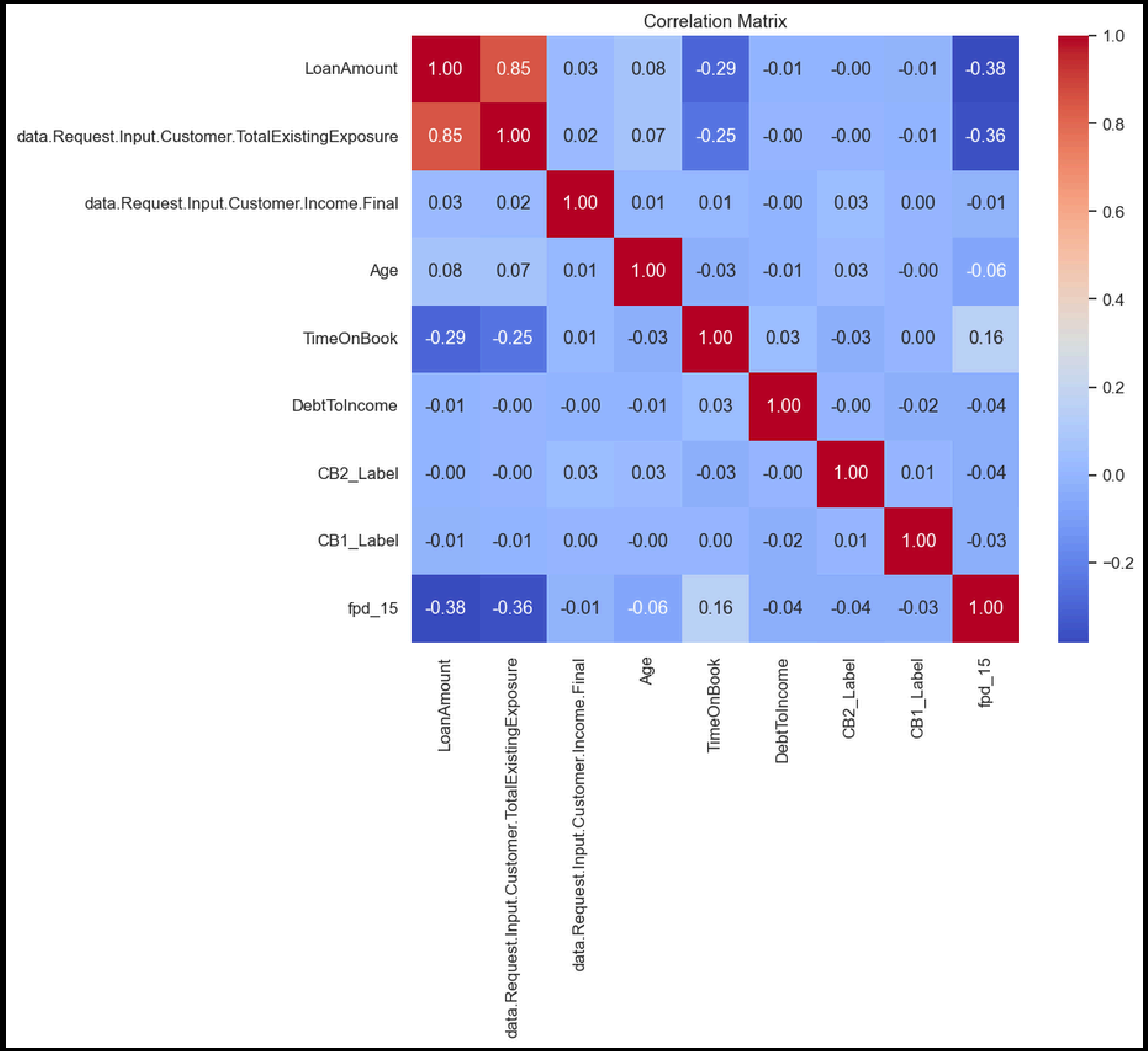


EDA Highlights

Oluseyi Ajayi Home About

Correlation Heatmap

This heatmap shows a strong positive correlation between loan amount and existing exposure, and a negative correlation between loan amount and first payment default. Other variables shows low multicollinearity which makes it suitability for regression or credit risk modeling.



01

02

03

EDA Highlights

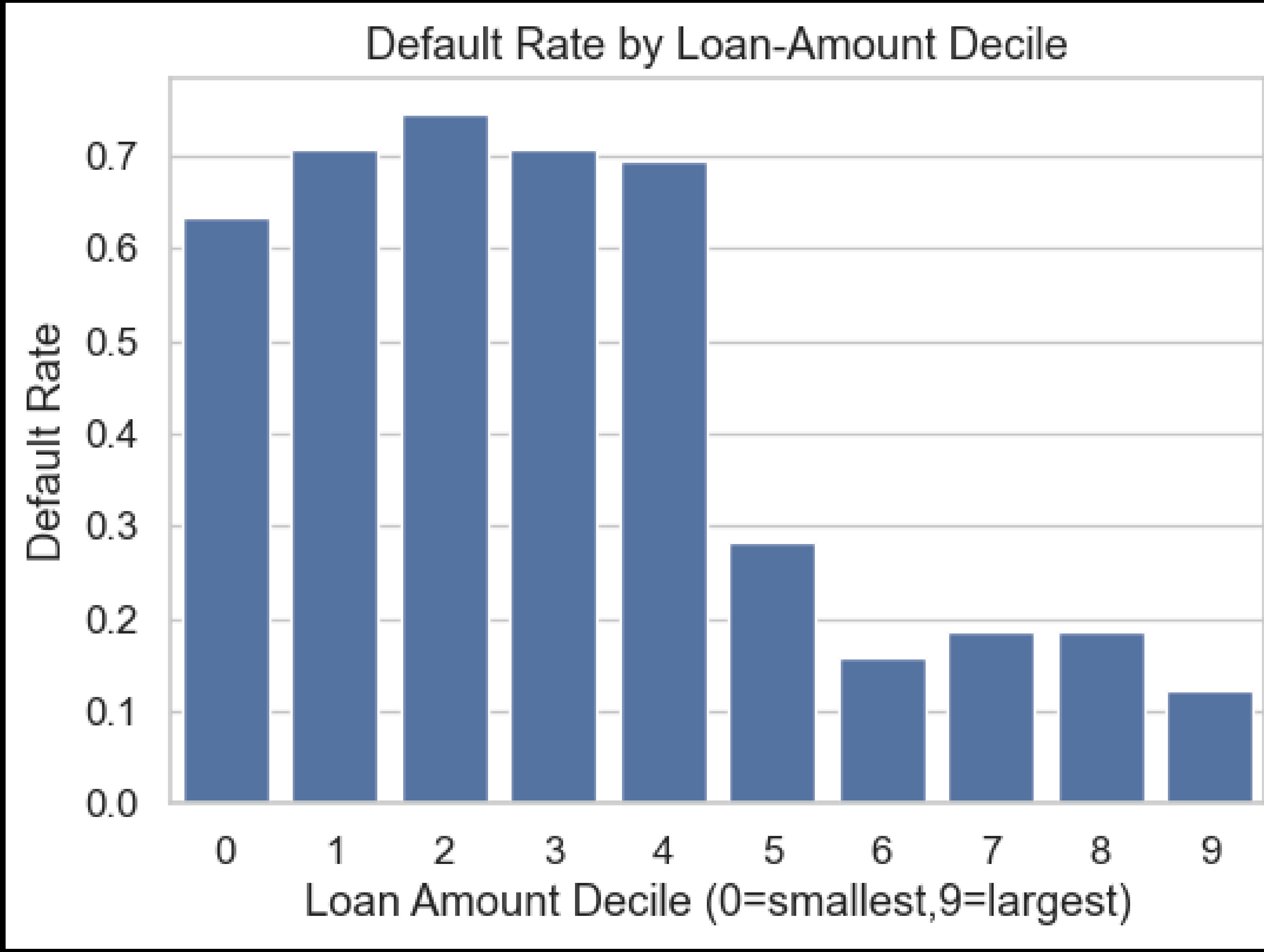
Oluseyi Ajayi

Home

About

Default Rate analysis

Showing default rate by loan amount and it trends overtime



01

02

03

EDA Highlights

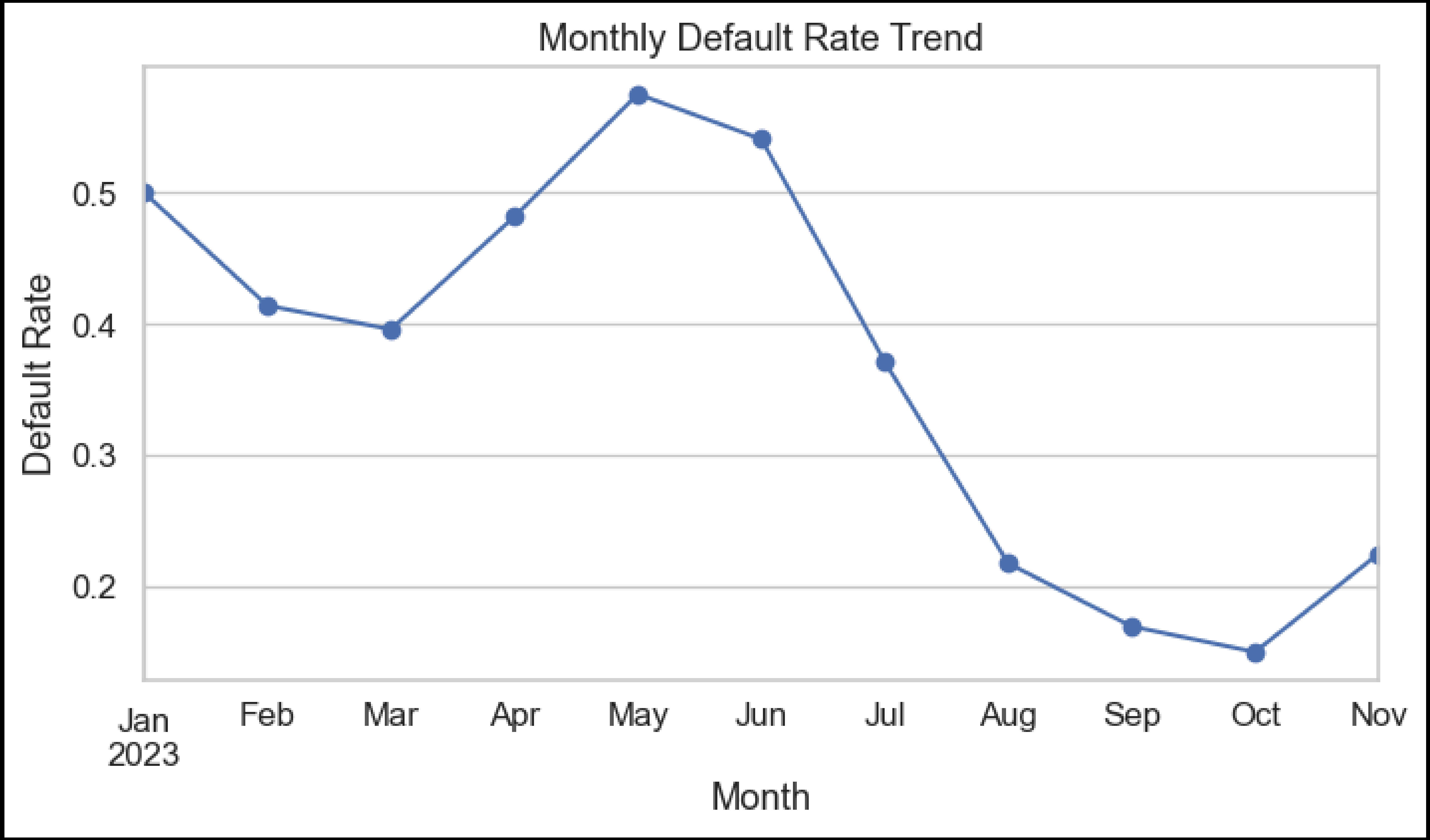
Oluseyi Ajayi

Home

About

Default Rate analysis

Showing default rate trends overtime



01

02

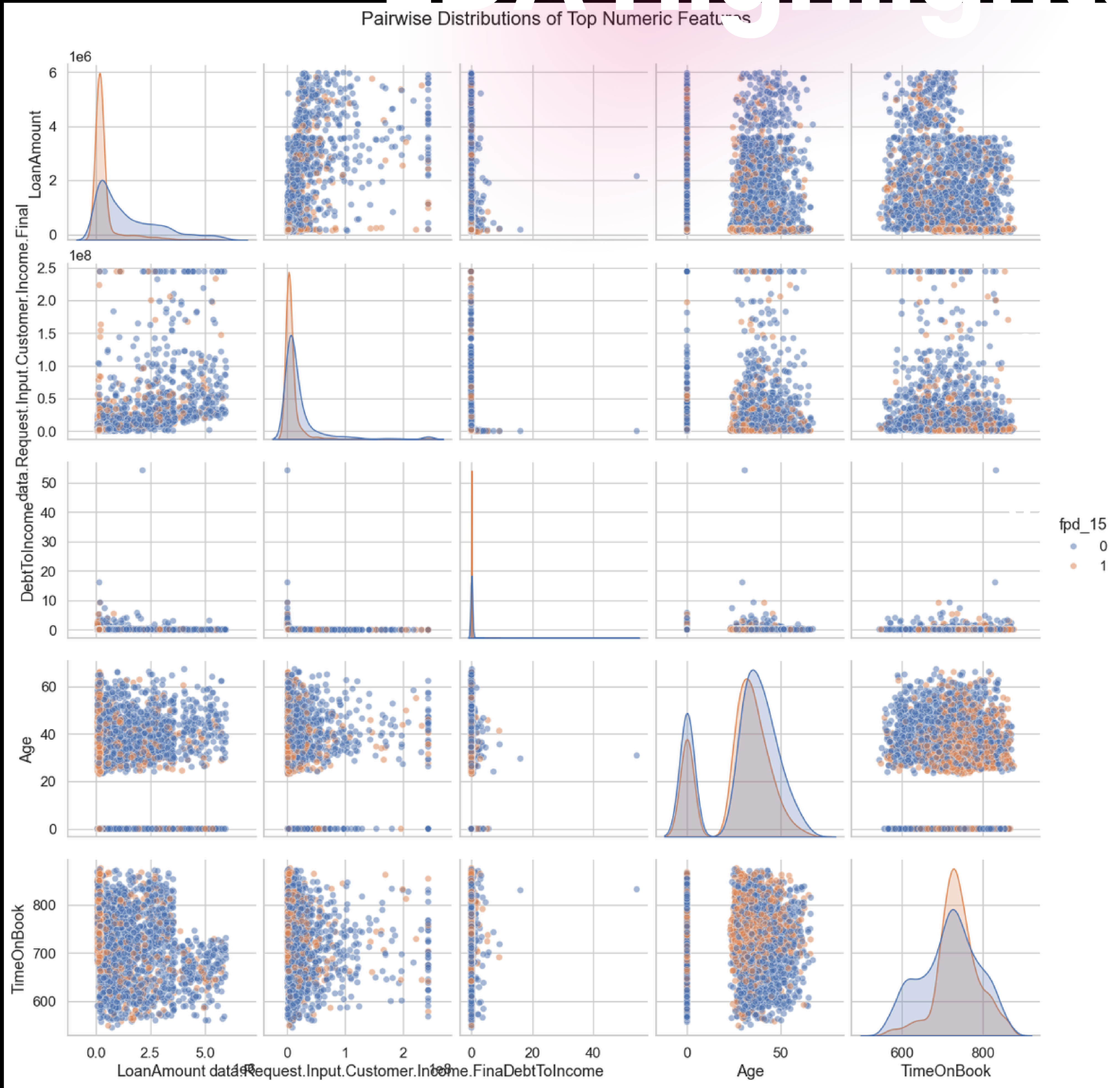
03

EDA Highlights

Home

Pairplot for Top Numeric Features

Shows the distrbution of some top numeric features like LoanAmount, Income, DebtToIncome, Ag, TimeOnBook.



Oluseyi Ajayi

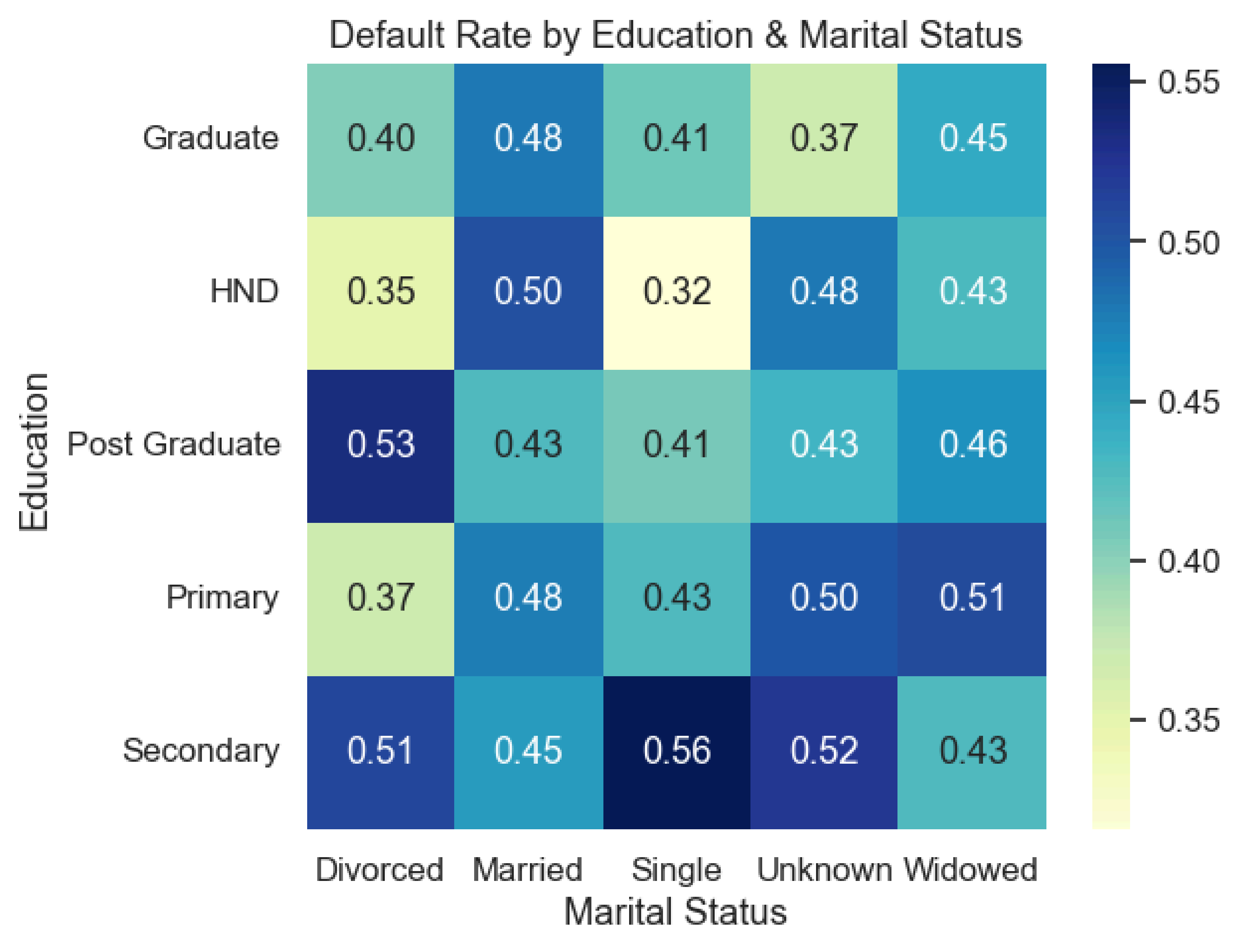
Oluseyi Ajayi

Home

About

EDA Highlights

Default Rate analysis



Showing default rate by education and marital status and can see the highest and lowest default rate , 0.56 and 0.32 respectively

01

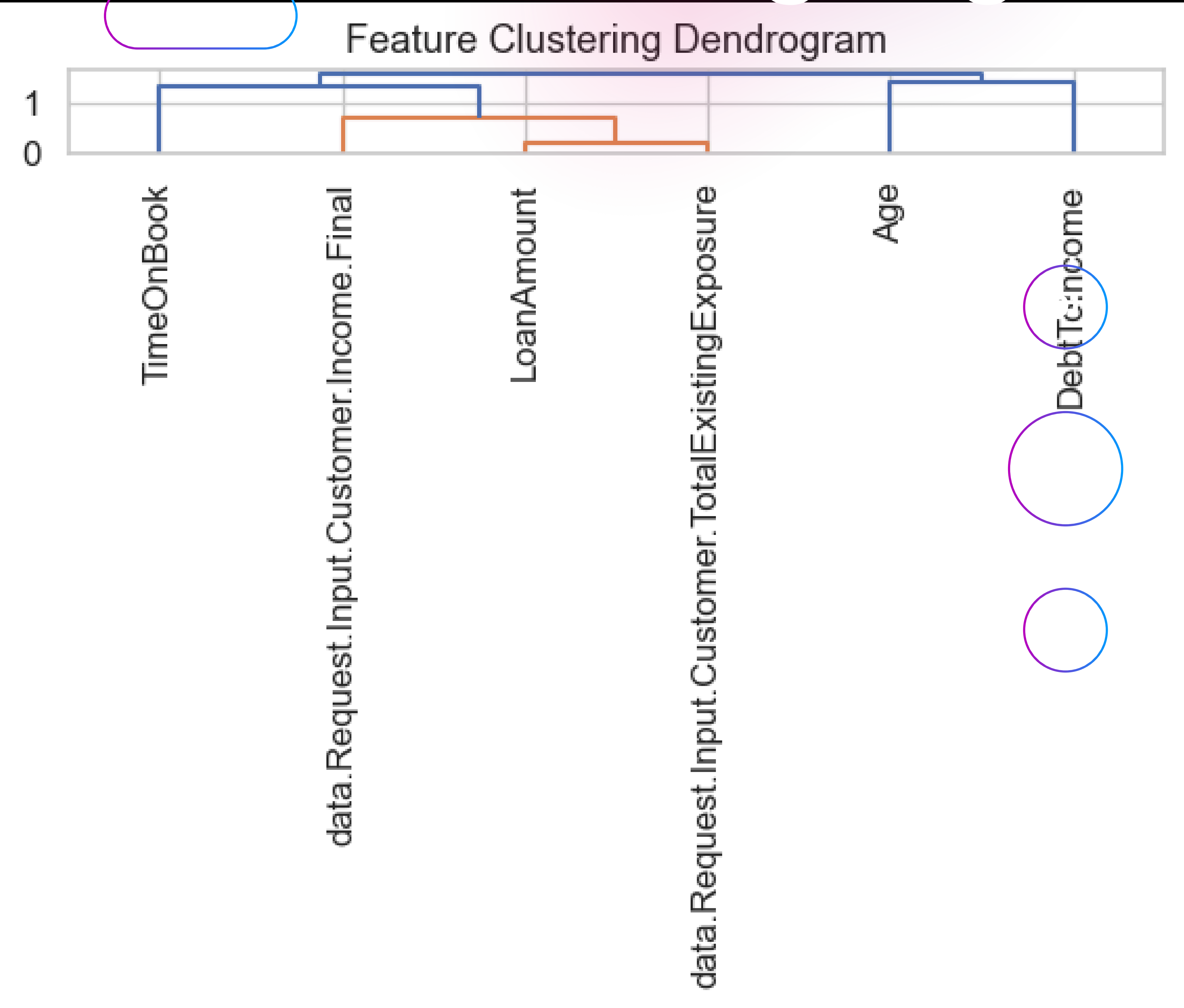
02

03

EDA Highlights

Features Selection Analysis

showing features correlation to help feature selection process as we see LoanAmount and TotalExistingExposure are very correlated



EDA Highlights

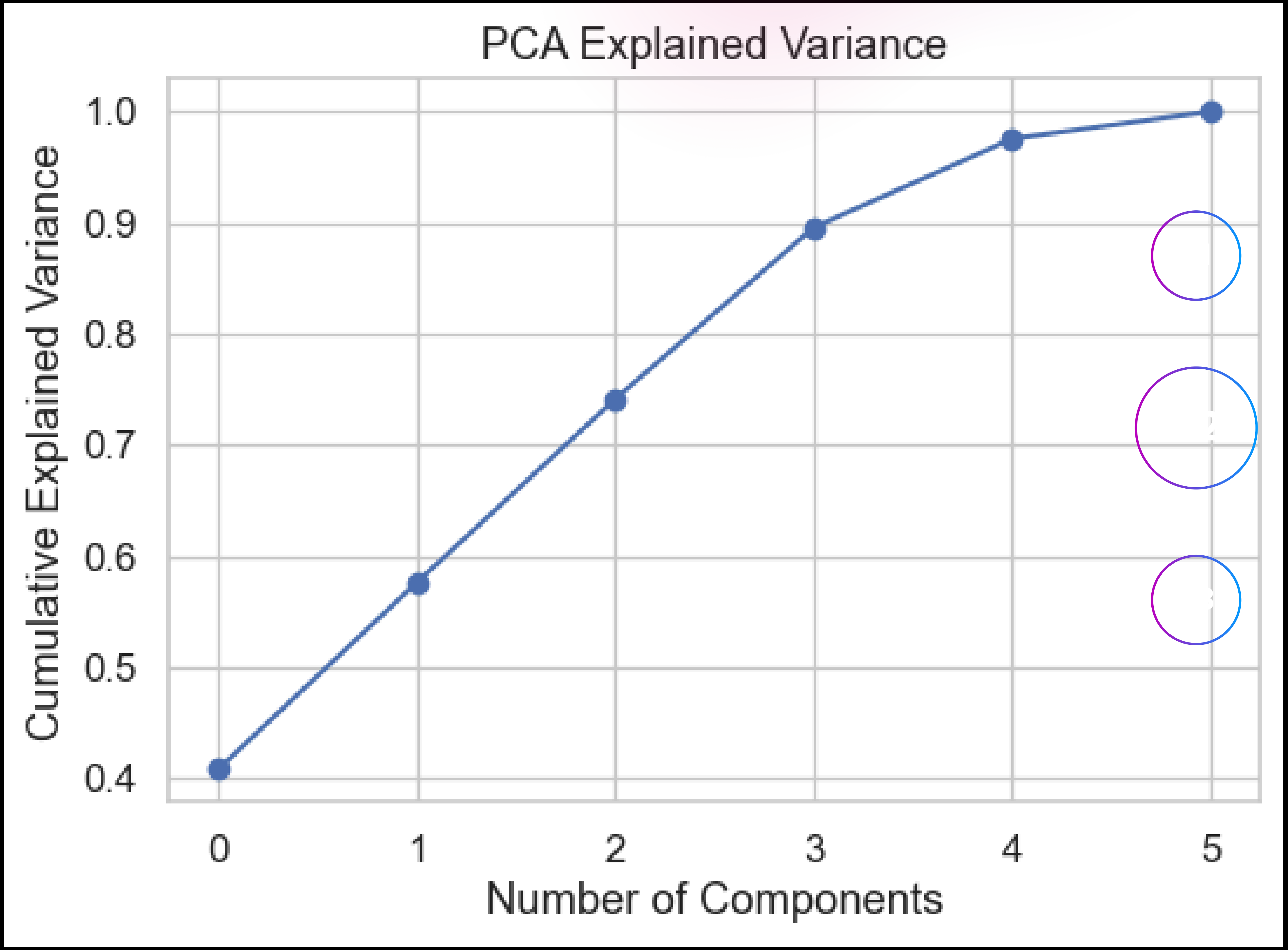
Oluseyi Ajayi

Home

About

Features Selection Analysis

showing PCA based on number of component and possible losses



Feature Engineering

i did some feature engineering to get new important features like TimeOnBook, DebtToIncome and age

```python
# a) Age in years
df["Age"] = (
    today - df["data.Request.Input.Customer.DateOfBirth"]
).dt.days.div(365).round(1)

# b) Time on book in days
df["TimeOnBook"] = (
    today - df["CreationDate"]
).dt.days

# c) Debt-to-Income ratio
df["DebtToIncome"] = (
    df["data.Request.Input.Customer.TotalExistingExposure"]
    / df["data.Request.Input.Customer.Income.Final"]
).replace([np.inf, -np.inf], np.nan).fillna(0)
```

Oluseyi Ajayi

Model Development

i developed the model using the XGBClassifier with learning rate 0.1 and n-estimator at 200 which gave an

ROC-AUC: 0.82

Precision: 0.45

Recall: 0.60

F1 Score: 0.51

Findings

1. Higher DebtToIncome and Shorter TimeOnBook would result in high default risk
2. Applicants from LGA X exhibited abou 20% higher default rate

Recommendations

Would explore Deep Learning libraries like TabNet to detect complex interations

2. An automated data pipeline for constant retraining and refresh

Recommendations

Thank You

FOR YOUR ATTENTION

01
02
03