ANDISWA MCHUNU

PROJECT REPORT

2024

LOAN RISK ANALYSIS

TABLE OF CONTENT

EXECUTIVE SUMMARY

In order to predict the issuing of loans, this project will assess loan applications using a variety of criteria, such as applicant age, income, credit score, employment status, and other financial indicators. The project used a Random Forest model in predicting whether a loan would be approved or not. The prediction was highly accurate, at 99%, with great accuracy and recall for both approved and non-approved categories of loans. It can, therefore, be concluded that the model can be used as a reliable tool in loan risk assessment and the decision-making process for approval. Recommendations include further tuning of the model and implementation for real-time predictions.

INTRODUCTION

BACKGROUND

Financial institutions have to deal with a large number of loan applications. Hence the need for quick and accurate assessment of risks in loan approvals. The purpose of this project is to build a predictive model to assist in automating the loan approval process.

OBJECTIVES

To predict the status of loan approval based on applicant's financial stability, credit score, employment status, and many more.
Provide actionable insights to financial organizations in facilitating data-driven decisions on loan approval.

SCOPE

The analysis takes into account all the relevant variables that can affect the loan approval, and it is limited to the dataset provided, which involves financial metrics, credit history, and personal information.

DATA SOURCE

The dataset contains many variables on loan applications, basically: ApplicationDate, Age, Income, LoanAmount, CreditScore, and ApprovalStatus of the loan.

METHODOLOGY

DATA COLLECTION

The dataset sourced from Kaggle, has a total of 20 000 records that contains all the applicant demographics, financial information, and history about loan records.

DATA CLEANING

The categorical variables in this dataset were EmploymentStatus and MaritalStatus, which were one-hot encoded in a binary format.
ApplicationDate, after preprocessing, was split into individual features of Year, Month, and Day for better analysis.

TOOLS USED

Python: Cleaning the data, building the model by using Random Forest and Evaluation.
PowerQuery: Data transformation and data preparation
Power BI: This is targeted for visualization and the creation of dashboards.

ANALYTICAL TECHNIQUES

Descriptive statistics were used to summarize the data.
Data visualizations such as scatter plots and bar charts emphasized main trends.
The Random Forest algorithm generated the predictive model using a split of 80-20 between training and testing of the dataset.

DATA ANALYSIS

DESCRIPTIVE STATISTICS

Average amount borrowed: $20,000

Mean Credit Score: 680

Loans Sanctioned: 25% of the total applications.

Loans Rejected: 75%

TRENDS AND PATTERNS

Applicants with higher credit scores, in excess of 700 have been more likely to have loans approved.

People with higher AnnualIncome and more stable EmploymentStatus were more likely to get approved.

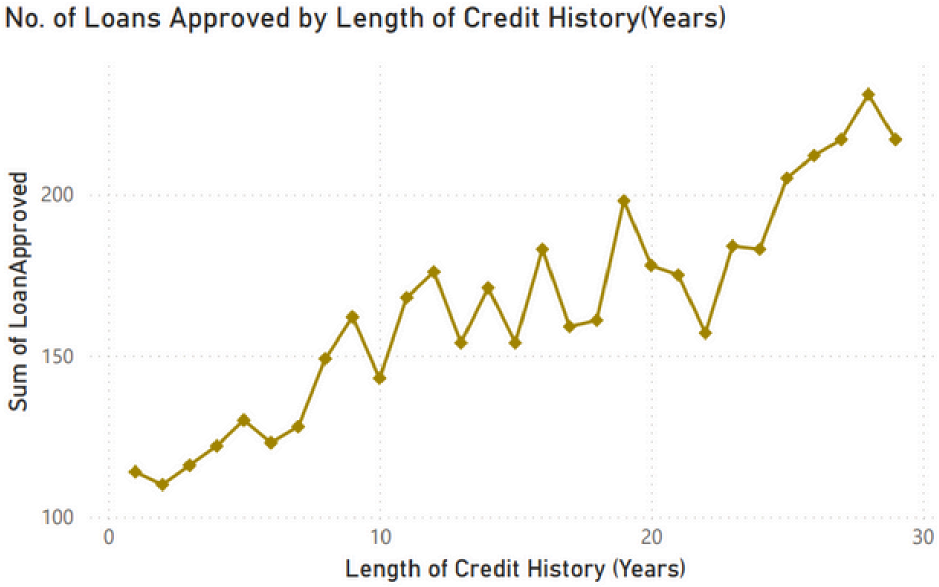
Homeowners had a higher chance of accessing the loan over tenants.

People with higher AnnualIncome and more stable EmploymentStatus were more likely to get approved.

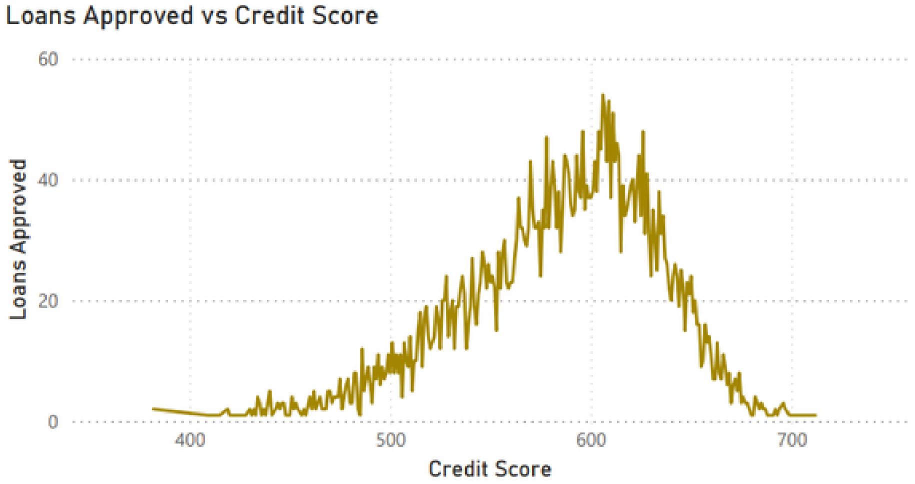
Homeowners had a higher chance of accessing the loan over tenants.

VISUALIZATIONS

Line Chart: This visualized the count of the different loan approvals granted on years of employment history

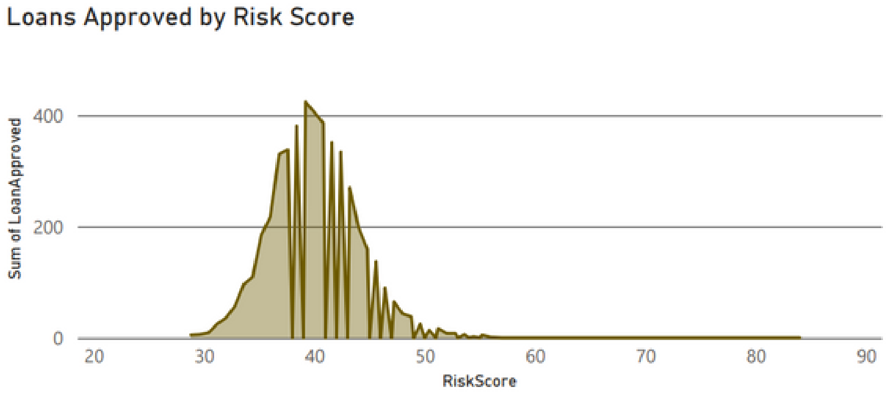


Line Chart: It reflects the relationship between credit score and the likelihood of approval for a loan

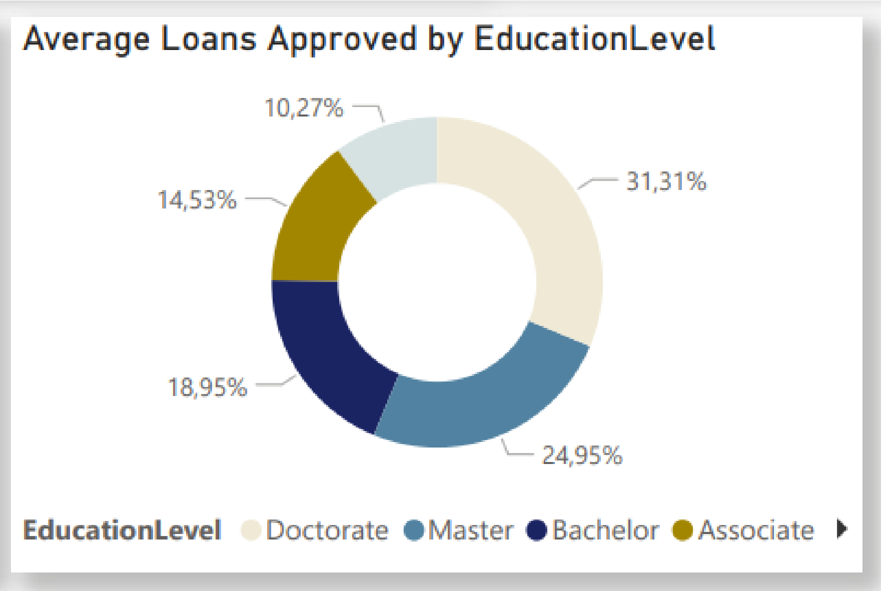


VISUALIZATIONS

Area Chart: Visualizing the distribution of loan approvals correlated with risk score.



Doughnut Chart: Visualizing the distribution of loan applications correlated with Education Level.



INSIGHTS

The debt-to-income ratio plays a important role in loan approvals

Credit Score and Loan Approved have a weak positive relationship, but it's not strong enough to make firm conclusions about their association.

Credit history is associated with a higher average approval of loan approvals.

RESULTS

The Random Forest model was able to achieve 99% overall accuracy.
The precision was 99% for Approved Loans (Class 1) and also 99% for Non-approved Loans (Class 0).

The model turned out to be very effective in predicting loan approval based on financial and demographic data.

The features that best described the approval of the loans were Credit Score, Annual Income, Loan Amount, and Employment Status.

DISCUSSION

INTERPRETATION OF RESULT

The model is highly accurate and precise meaning that it will be reliable in the prediction of approval for loans, hence helping to speed up decision-making processes by the monetary Institutes.

It is expected that credit scores and employment status have the highest impacts on loan approval decisions, following the pattern applicable in the financial industry.

LIMITATIONS

There might class imbalance biases since more rejections were recorded compared to approvals. This could affect the model's performance.
The analysis is completely dependent on data availability and may not include all real-world factors affecting loan approvals.

COMPARISON TO EXPECTATIONS

It performed well beyond expectations for its accuracy, though it was somewhat surprising how credit score was not as influential relative to other financial metrics when making decisions on loans.

RECOMMENDATIONS

MODEL DEPLOYMENT

The model can be directly used in a real-time loan approval systems to assist in making quicker decisions.

MODEL OPTIMIZATION

Further optimization of the random forest algorithm can be supported by tuning the hyperparameters so that the model performs with higher precision in cases of edge conditions.

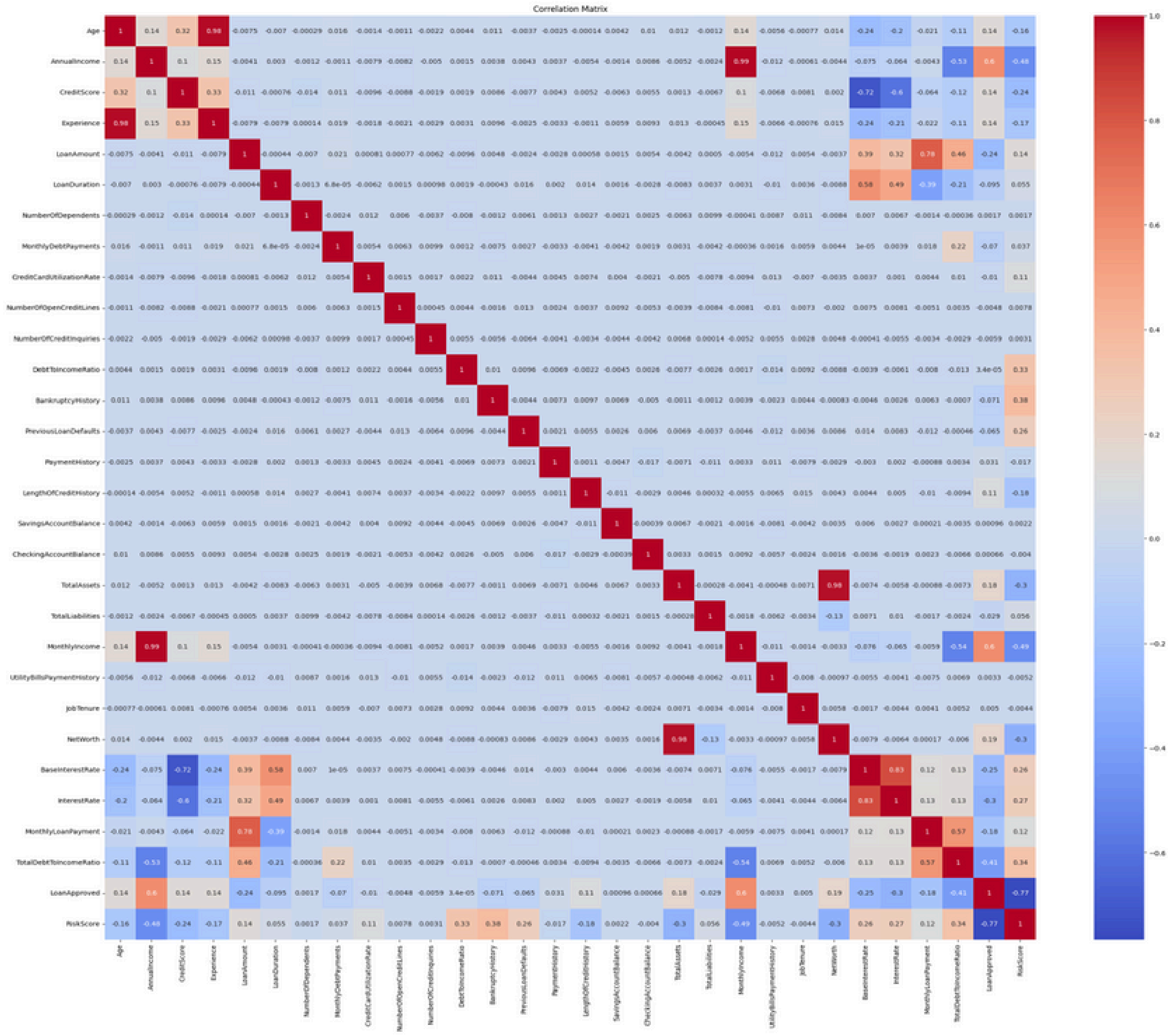
FEATURE AUGMENTATION

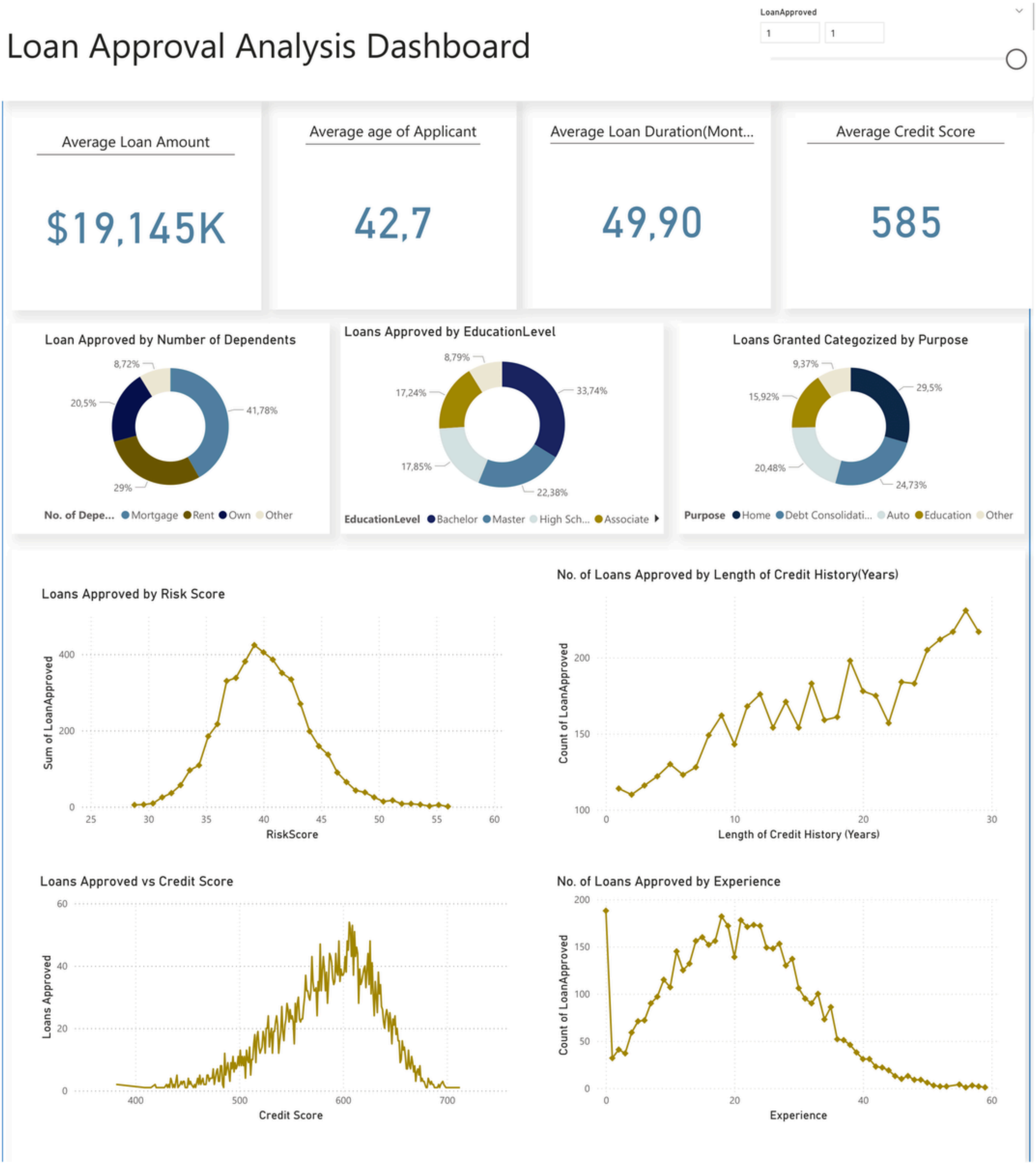
More data could be included, such as customer behavioral patterns and market trends, to make the model predictions even better.

CONCLUSION

The performed analysis successfully developed a machine learning model that can predict loan approvals with high accuracy. The insight from feature importance, such as the influence of credit scores and employment status, is a helpful guide for financial institutions in making decisions. Although the performance was quite good, improvements will be made for future work with regards to hyperparameter tuning and balance of classes for the extension of its practical applicability.

APPENDICES


Correlation Matrix
Age 1 0.14 0.32 0.98 -0.0075 -0.007 -0.00029 0.016 -0.0014 -0.0011 -0.0022 0.0044 0.011 -0.0037 -0.0025 -0.00014 0.0042 0.01 0.012 -0.0012 0.14 -0.0056 -0.00077 0.014 -0.24 -0.2 -0.021 -0.11 0.14 -0.16
AnnualIncome 0.14 1 0.1 0.15 -0.0041 0.003 -0.0012 -0.0011 -0.0079 -0.0082 -0.005 0.0015 0.0038 0.0043 0.0037 -0.0054 -0.0014 0.0086 -0.0052 -0.0024 0.99 -0.012 -0.00061 -0.0044 -0.075 -0.064 -0.0043 -0.53 0.6 -0.48
CreditScore 0.32 0.1 1 0.33 -0.011 -0.00076 -0.014 0.011 -0.0096 -0.0088 -0.0019 0.0019 0.0086 -0.0077 0.0043 0.0052 -0.0063 0.0055 0.0013 -0.0067 0.1 -0.0068 0.0081 0.002 -0.72 -0.6 -0.064 -0.12 0.14 -0.24
Experience 0.98 0.15 0.33 1 -0.0079 -0.0079 0.00014 0.019 -0.0018 -0.0021 -0.0029 0.0031 0.0096 -0.0025 -0.0033 -0.0011 0.0059 0.0093 0.013 -0.00045 0.15 -0.0066 -0.00076 0.015 -0.24 -0.21 -0.022 -0.11 0.14 -0.17
LoanAmount -0.0075 -0.0041 -0.011 -0.0079 1 -0.00044 -0.007 0.021 0.00081 0.00077 -0.0062 -0.0096 0.0048 -0.0024 -0.0028 0.00058 0.0015 0.0054 -0.0042 0.0005 -0.0054 -0.012 0.0054 -0.0037 0.39 0.32 0.78 0.46 -0.24 0.14
LoanDuration -0.007 0.003 -0.00076 -0.0079 -0.00044 1 -0.0013 6.8e-05 -0.0062 0.0015 0.00098 0.0019 -0.00043 0.016 0.002 0.014 0.0016 -0.0028 -0.0083 0.0037 0.0031 -0.01 0.0036 -0.0088 0.58 0.49 -0.39 -0.21 -0.095 0.055
NumberOfDependents -0.00029 -0.0012 -0.014 0.00014 -0.007 -0.0013 1 -0.0024 0.012 0.006 -0.0037 -0.008 -0.0012 0.0061 0.0013 0.0027 -0.0021 0.0025 -0.0063 0.0099 -0.00041 0.0087 0.011 -0.0084 0.007 0.0067 -0.0014 -0.00036 0.0017 0.0017
MonthlyDebtPayments 0.016 -0.0011 0.011 0.019 0.021 6.8e-05 -0.0024 1 0.0054 0.0063 0.0099 0.0012 -0.0075 0.0027 -0.0033 -0.0041 -0.0042 0.0019 0.0031 -0.0042 -0.00036 0.0016 0.0059 0.0044 1e-05 0.0039 0.018 0.22 -0.07 0.037
CreditCardUtilizationRate -0.0014 -0.0079 -0.0096 -0.0018 0.00081 -0.0062 0.012 0.0054 1 0.0015 0.0017 0.0022 0.011 -0.0044 0.0045 0.0074 0.004 -0.0021 -0.005 -0.0078 -0.0094 0.013 -0.007 -0.0035 0.0037 0.001 0.0044 0.01 -0.01 0.11
NumberOfOpenCreditLines -0.0011 -0.0082 -0.0088 -0.0021 0.00077 0.0015 0.006 0.0063 0.0015 1 0.00045 0.0044 -0.0016 0.013 0.0024 0.0037 0.0092 -0.0053 -0.0039 -0.0084 -0.0081 -0.01 0.0073 -0.002 0.0075 0.0081 -0.0051 0.0035 -0.0048 0.0078
NumberOfCreditInquiries -0.0022 -0.005 -0.0019 -0.0029 -0.0062 0.00098 -0.0037 0.0099 0.0017 0.00045 1 0.0055 -0.0056 -0.0064 -0.0041 -0.0034 -0.0044 -0.0042 0.0068 0.00014 -0.0052 0.0055 0.0028 0.0048 -0.00041 -0.0055 -0.0034 -0.0029 -0.0059 0.0031
DebtToIncomeRatio 0.0044 0.0015 0.0019 0.0031 -0.0096 0.0019 -0.008 0.0012 0.0022 0.0044 0.0055 1 0.01 0.0096 -0.0069 -0.0022 -0.0045 0.0026 -0.0077 -0.0026 0.0017 -0.014 0.0092 -0.0088 -0.0039 -0.0061 -0.008 -0.013 3.4e-05 0.33
BankruptcyHistory 0.011 0.0038 0.0086 0.0096 0.0048 -0.00043 -0.0012 -0.0075 0.011 -0.0016 -0.0056 0.01 1 -0.0044 0.0073 0.0097 0.0069 -0.005 -0.0011 -0.0012 0.0039 -0.0023 0.0044 -0.00083 -0.0046 0.0026 0.0063 -0.0007 -0.071 0.38
PreviousLoanDefaults -0.0037 0.0043 -0.0077 -0.0025 -0.0024 0.016 0.0061 0.0027 -0.0044 0.013 -0.0064 0.0096 -0.0044 1 0.0021 0.0055 0.0026 0.006 0.0069 -0.0037 0.0046 -0.012 0.0036 0.0086 0.014 0.0083 -0.012 -0.00046 -0.065 0.26
PaymentHistory -0.0025 0.0037 0.0043 -0.0033 -0.0028 0.002 0.0013 -0.0033 0.0045 0.0024 -0.0041 -0.0069 0.0073 0.0021 1 0.0011 -0.0047 -0.017 -0.0071 -0.011 0.0033 0.011 -0.0079 -0.0029 -0.003 0.002 -0.00088 0.0034 0.031 -0.017
LengthOfCreditHistory -0.00014 -0.0054 0.0052 -0.0011 0.00058 0.014 0.0027 -0.0041 0.0074 0.0037 -0.0034 -0.0022 0.0097 0.0055 0.0011 1 -0.011 -0.0029 0.0046 0.00032 -0.0055 0.0065 0.015 0.0043 0.0044 0.005 -0.01 -0.0094 0.11 -0.18
SavingsAccountBalance 0.0042 -0.0014 -0.0063 0.0059 0.0015 0.0016 -0.0021 -0.0042 0.004 0.0092 -0.0044 -0.0045 0.0069 0.0026 -0.0047 -0.011 1 -0.00039 0.0067 -0.0021 -0.0016 -0.0081 -0.0042 0.0035 0.006 0.0027 0.00021 -0.0035 0.00096 0.0022
CheckingAccountBalance 0.01 0.0086 0.0055 0.0093 0.0054 -0.0028 0.0025 0.0019 -0.0021 -0.0053 -0.0042 0.0026 -0.005 0.006 -0.017 -0.0029 -0.00039 1 0.0033 0.0015 0.0092 -0.0057 -0.0024 0.0016 -0.0036 -0.0019 0.0023 -0.0066 0.00066 -0.004
TotalAssets 0.012 -0.0052 0.0013 0.013 -0.0042 -0.0083 -0.0063 0.0031 -0.005 -0.0039 0.0068 -0.0077 -0.0011 0.0069 -0.0071 0.0046 0.0067 0.0033 1 -0.00028 -0.0041 -0.00048 0.0071 0.98 -0.0074 -0.0058 -0.00088 -0.0073 0.18 -0.3
TotalLiabilities -0.0012 -0.0024 -0.0067 -0.00045 0.0005 0.0037 0.0099 -0.0042 -0.0078 -0.0084 0.00014 -0.0026 -0.0012 -0.0037 -0.011 0.00032 -0.0021 0.0015 -0.00028 1 -0.0018 -0.0062 -0.0034 -0.13 0.0071 0.01 -0.0017 -0.0024 -0.029 0.056
MonthlyIncome 0.14 0.99 0.1 0.15 -0.0054 0.0031 -0.00041 -0.00036 -0.0094 -0.0081 -0.0052 0.0017 0.0039 0.0046 0.0033 -0.0055 -0.0016 0.0092 -0.0041 -0.0018 1 -0.011 -0.0014 -0.0033 -0.076 -0.065 -0.0059 -0.54 0.6 -0.49
UtilityBillsPaymentHistory -0.0056 -0.012 -0.0068 -0.0066 -0.012 -0.01 0.0087 0.0016 0.013 -0.01 0.0055 -0.014 -0.0023 -0.012 0.011 0.0065 -0.0081 -0.0057 -0.00048 -0.0062 -0.011 1 -0.008 -0.00097 -0.0055 -0.0041 -0.0075 0.0069 0.0033 -0.0052
JobTenure -0.00077 -0.00061 0.0081 -0.00076 0.0054 0.0036 0.011 0.0059 -0.007 0.0073 0.0028 0.0092 0.0044 0.0036 -0.0079 0.015 -0.0042 -0.0024 0.0071 -0.0034 -0.0014 -0.008 1 0.0058 -0.0017 -0.0044 0.0041 0.0052 0.005 -0.0044
NetWorth 0.014 -0.0044 0.002 0.015 -0.0037 -0.0088 -0.0084 0.0044 -0.0035 -0.002 0.0048 -0.0088 -0.00083 0.0086 -0.0029 0.0043 0.0035 0.0016 0.98 -0.13 -0.0033 -0.00097 0.0058 1 -0.0079 -0.0064 0.00017 -0.006 0.19 -0.3
BaseInterestRate -0.24 -0.075 -0.72 -0.24 0.39 0.58 0.007 1e-05 0.0037 0.0075 -0.00041 -0.0039 -0.0046 0.014 -0.003 0.0044 0.006 -0.0036 -0.0074 0.0071 -0.076 -0.0055 -0.0017 -0.0079 1 0.83 0.12 0.13 -0.25 0.26
InterestRate -0.2 -0.064 -0.6 -0.21 0.32 0.49 0.0067 0.0039 0.001 0.0081 -0.0055 -0.0061 0.0026 0.0083 0.002 0.005 0.0027 -0.0019 -0.0058 0.01 -0.065 -0.0041 -0.0044 -0.0064 0.83 1 0.13 0.13 -0.3 0.27
MonthlyLoanPayment -0.021 -0.0043 -0.064 -0.022 0.78 -0.39 -0.0014 0.018 0.0044 -0.0051 -0.0034 -0.008 0.0063 -0.012 -0.00088 -0.01 0.00021 0.0023 -0.00088 -0.0017 -0.0059 -0.0075 0.0041 0.00017 0.12 0.13 1 0.57 -0.18 0.12
TotalDebtToIncomeRatio -0.11 -0.53 -0.12 -0.11 0.46 -0.21 -0.00036 0.22 0.01 0.0035 -0.0029 -0.013 -0.0007 -0.00046 0.0034 -0.0094 -0.0035 -0.0066 -0.0073 -0.0024 -0.54 0.0069 0.0052 -0.006 0.13 0.13 0.57 1 -0.41 0.34
LoanApproved 0.14 0.6 0.14 0.14 -0.24 -0.095 0.0017 -0.07 -0.01 -0.0048 -0.0059 3.4e-05 -0.071 -0.065 0.031 0.11 0.00096 0.00066 0.18 -0.029 0.6 0.0033 0.005 0.19 -0.25 -0.3 -0.18 -0.41 1 -0.77
RiskScore -0.16 -0.48 -0.24 -0.17 0.14 0.055 0.0017 0.037 0.11 0.0078 0.0031 0.33 0.38 0.26 -0.017 -0.18 0.0022 -0.004 -0.3 0.056 -0.49 -0.0052 -0.0044 -0.3 0.26 0.27 0.12 0.34 -0.77 1
Age AnnualIncome CreditScore Experience LoanAmount LoanDuration NumberOfDependents MonthlyDebtPayments CreditCardUtilizationRate NumberOfOpenCreditLines NumberOfCreditInquiries DebtToIncomeRatio BankruptcyHistory PreviousLoanDefaults PaymentHistory LengthOfCreditHistory SavingsAccountBalance CheckingAccountBalance TotalAssets TotalLiabilities MonthlyIncome UtilityBillsPaymentHistory JobTenure NetWorth BaseInterestRate InterestRate MonthlyLoanPayment TotalDebtToIncomeRatio LoanApproved RiskScore
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6



Loan Approval Analysis Dashboard

LoanApproved

1 1

Average Loan Amount	Average age of Applicant	Average Loan Duration(Mont...	Average Credit Score
$19,145K	42,7	49,90	585

Loan Approved by Number of Dependents
8,72%
20,5%
41,78%
29%
No. of Depe... ● Mortgage ● Rent ● Own ● Other
Loans Approved by EducationLevel
8,79%
17,24%
33,74%
17,85%
22,38%
EducationLevel ● Bachelor ● Master ● High Sch... ● Associate
Loans Granted Categozized by Purpose
9,37%
15,92%
29,5%
20,48%
24,73%
Purpose ● Home ● Debt Consolidati... ● Auto ● Education ● Other
Loans Approved by Risk Score
Sum of LoanApproved
400
200
0
25
30
35
40
45
50
55
60
RiskScore
No. of Loans Approved by Length of Credit History(Years)
Count of LoanApproved
200
150
100
0
10
20
30
Length of Credit History (Years)
Loans Approved vs Credit Score
Loans Approved
60
40
20
0
400
500
600
700
Credit Score
No. of Loans Approved by Experience
Count of LoanApproved
200
150
100
50
0
0
20
40
60
Experience