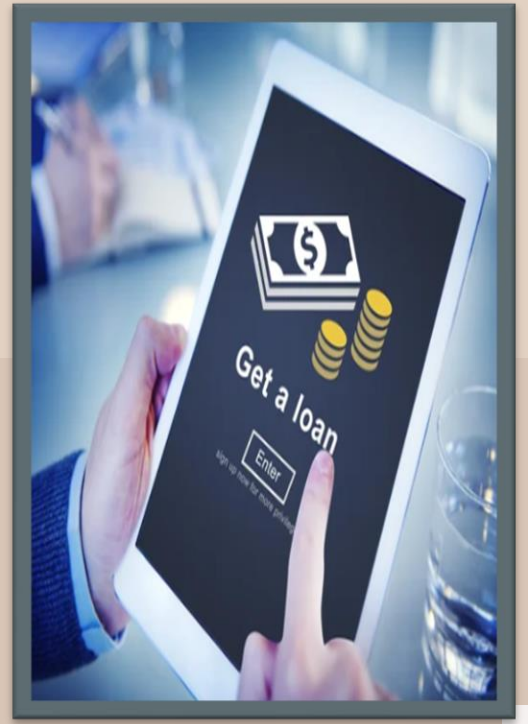
Data Science in Finance:

~Predicting Loan Defaults and Repayments~

INTRODUCTION

Let’s get rigth into it.



OVERVIEW

- The Lending Club Platform where individuals in the club loan other individuals. Someone who is certain to pay back a loan will have an easier time getting a loan with a low interest rate than someone who appears to be riskier. People who are very risky may not even get a loan offer, or may not accept the loan offer due to a high interest rate. This is bad for the club as they either loose their money or their clients. That is where I come in.

“Never a borrower or lender be; For loan oft loses both itself and friend.” In other words, you risk losing your money and your friendship if the loan isn't paid back.

Business Problem

- Does traditional creditworthiness assessments limits access to credit for some borrowers or lead to inefficiencies for lenders.
- YES
- **Solution:** leveraging the power of machine learning to move beyond the limitations of traditional creditworthiness assessments. Using AI tools such as machine learning to predict likelihood of a loan to be defaulted or fully paid.

DATA

The data that we will be working with is acquired from Kaggle. It contains includes: 56 columnentries that we will use as our predictor variables and target variables and 10,000 rows.
Our target variable is Loan_status.

The link to the dataset is given below:
https://www.kaggle.com/code/faressayah/lending-club-loan-defaulters-prediction

Process steps

1. Business understanding
2. Data mining
3. Data cleaning
4. Data exploration
5. Feature Engineering
6. Predictive Modeling
7. Data visualization

Process Steps

Business understanding:
We have used a recommendation type of analysis.(Which film would a user prefer?)

1

Data mining:
What data do we need to answer our analytical questions

2

Data Cleaning:
Ensured there are no data inconsistencies ,misspellings, outliers, outdated data and missing data.

3

Data exploration:
highlights patterns and relations in data

4

Feature Engineering:
A measurable phenomena is being observed.

5

Predictive Modelling:
Use mathematical or statistic modelsto answer question

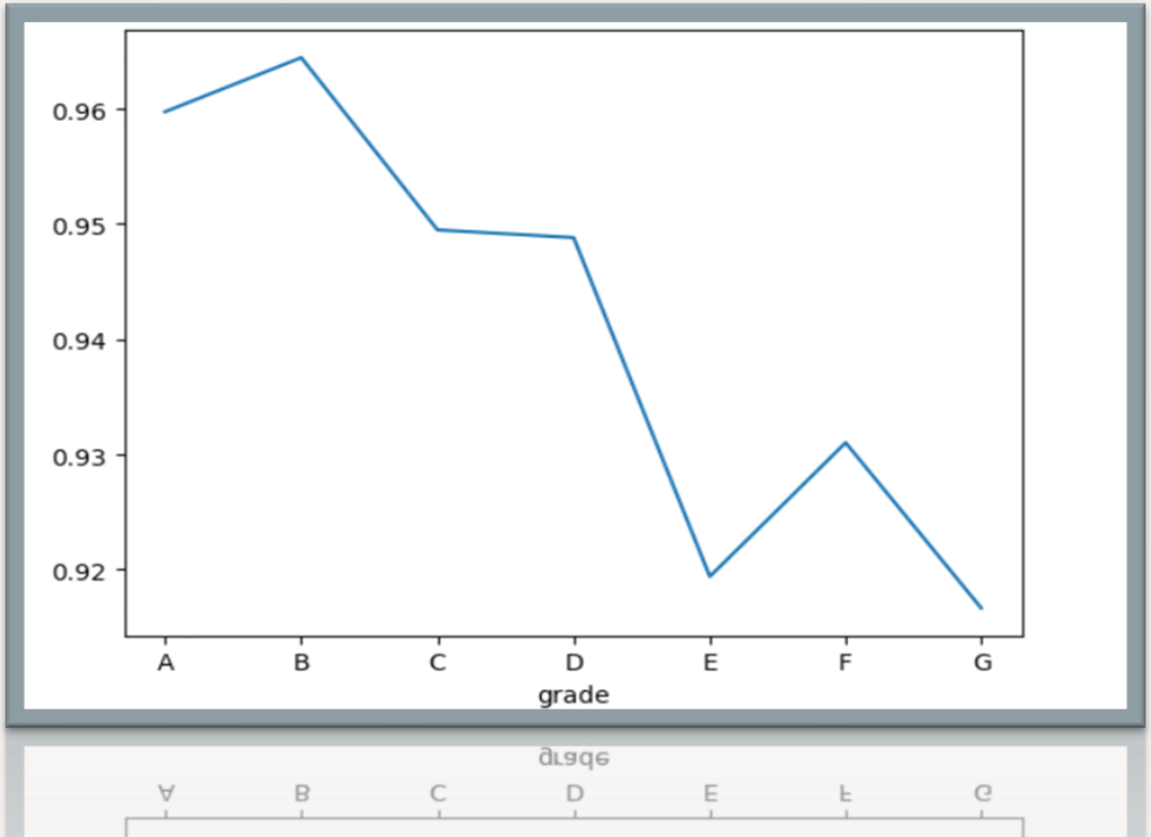
6

Data visualization:

5

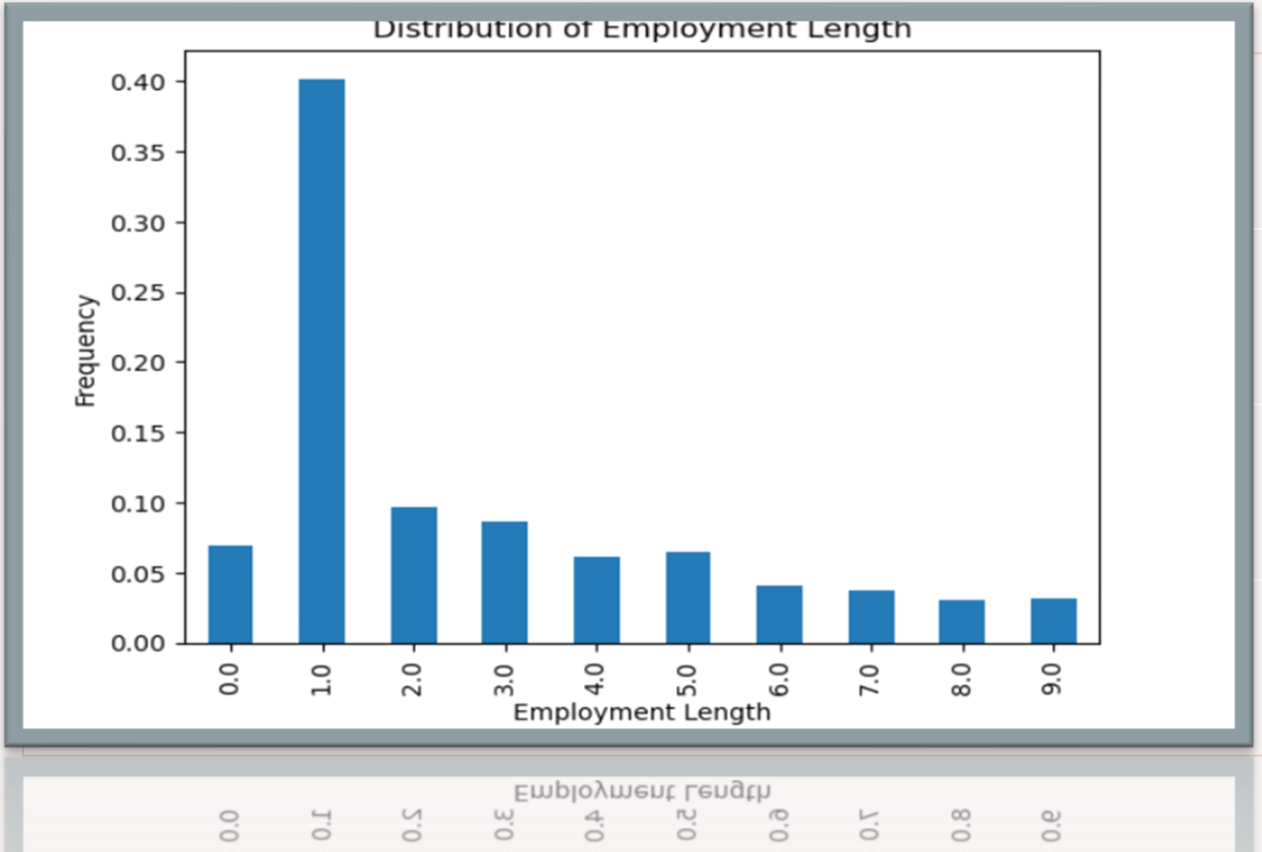
RESULTS:

First analysis:
we looked at the relationship between Loan grades : A, B, C, D, E, F, G and the default rate and we found that grade B has the highest default rate and grade G had the least.
This is also similar to the frequencies . Many people prefer loans with grade B as compared to G.



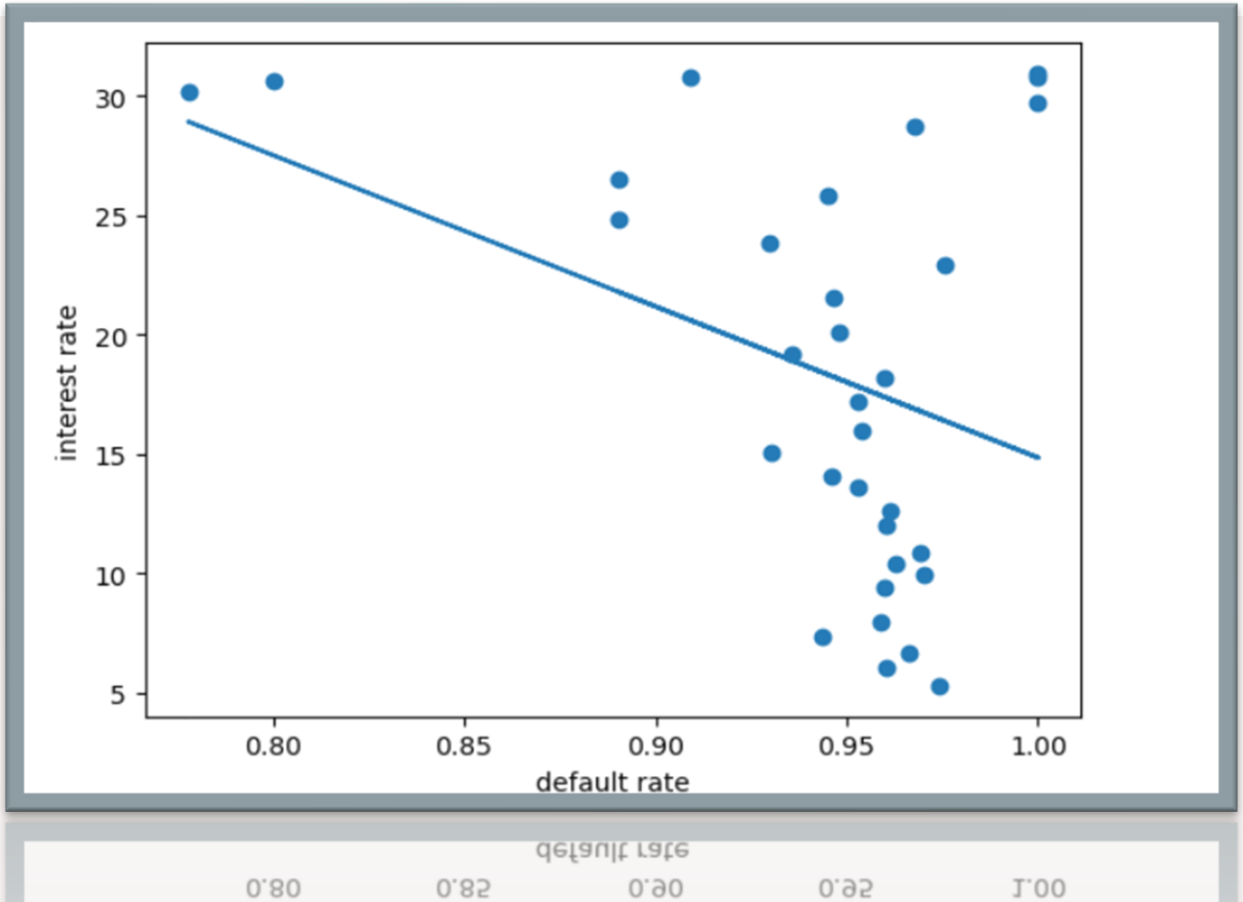
Second analysis:

It is well noted that people who have been employed for just one year are the people who seek for loans most.



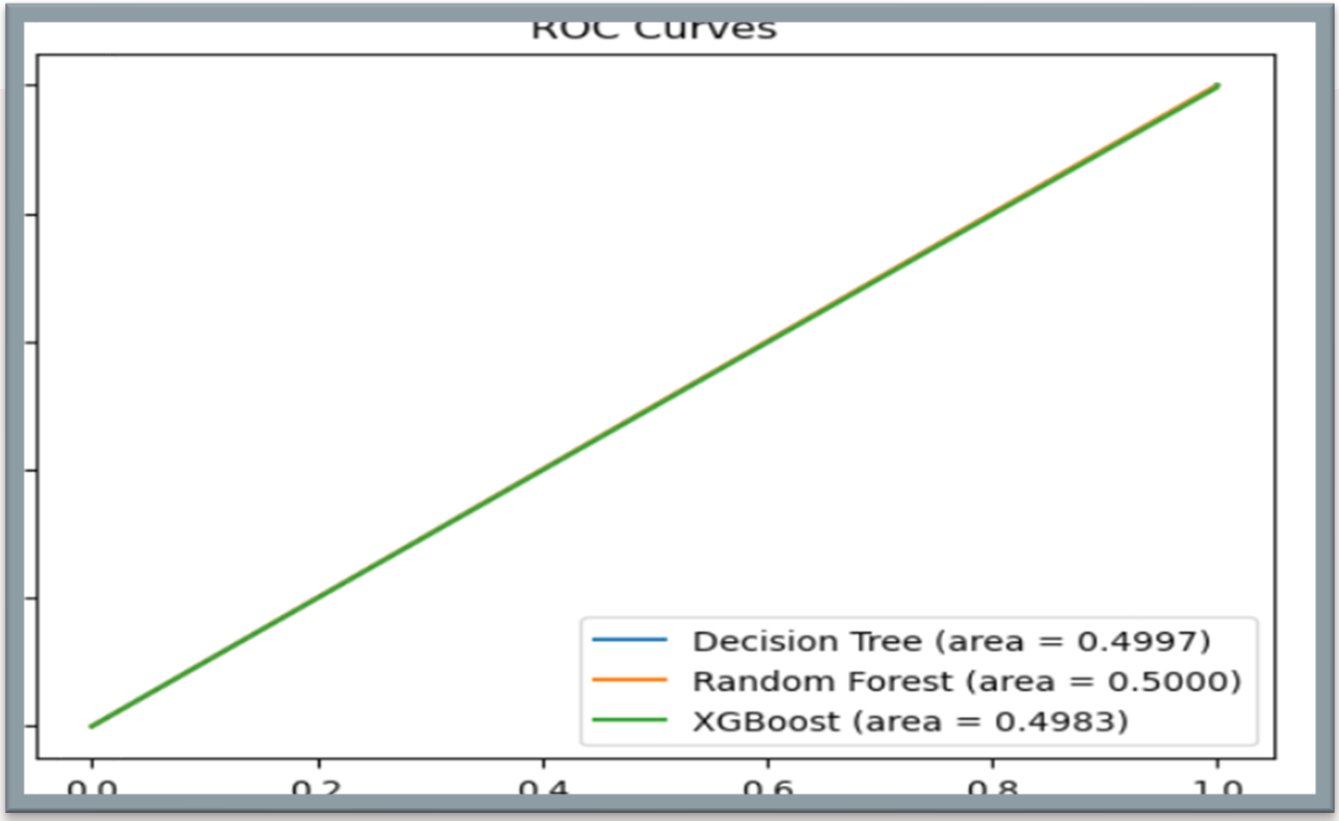
Third analysis:

interest rate is inversely related to default rate that means as the interest charged increases the default rate decreases.



Model evaluation:

From theROC curve we see that the random forest model performed best hence we recommend using it in prediction



Evaluation and Recommendation

Model recommendation

The model that performed well is the random forest model.
I recommend using the random model project when handling future loan predictions.

predictors

Having done a lot of data analysis , a few predictors seemed to stand out and should be studied closely: loan grade, length of employment, interest rate and loan purpose.

Further analysis

A further analysis should be done using F1 Score to cancel out any doubt of class imbalance in the loan status binary feature

Future Improvement Ideas

A further analysis should be done using F1 Score to cancel out any doubt of class imbalance in the loan status binary feature

Any Questions?

Thanks!

You can find me at:

- nyabagaabby@gmail. com

Contact Information

Abigael Nyatichi Nyabaga
Data Scientist

I am a data science enthusiast and I am priviledged to get this opportunity to work with Lending club on this project. I am looking forward to more of such opportunities to showcase my skills in data analysis by come up with solutions and recommendations for your problems. Below I have linked my social media platforms and details for future collaborations.

EMAIL: nyabagaabby @gmail.com

LINKEDIN:https://www.linkedin.com/in/abigael-nyabaga-46a165221

GITHUB: https://github.com/Nyabaga