

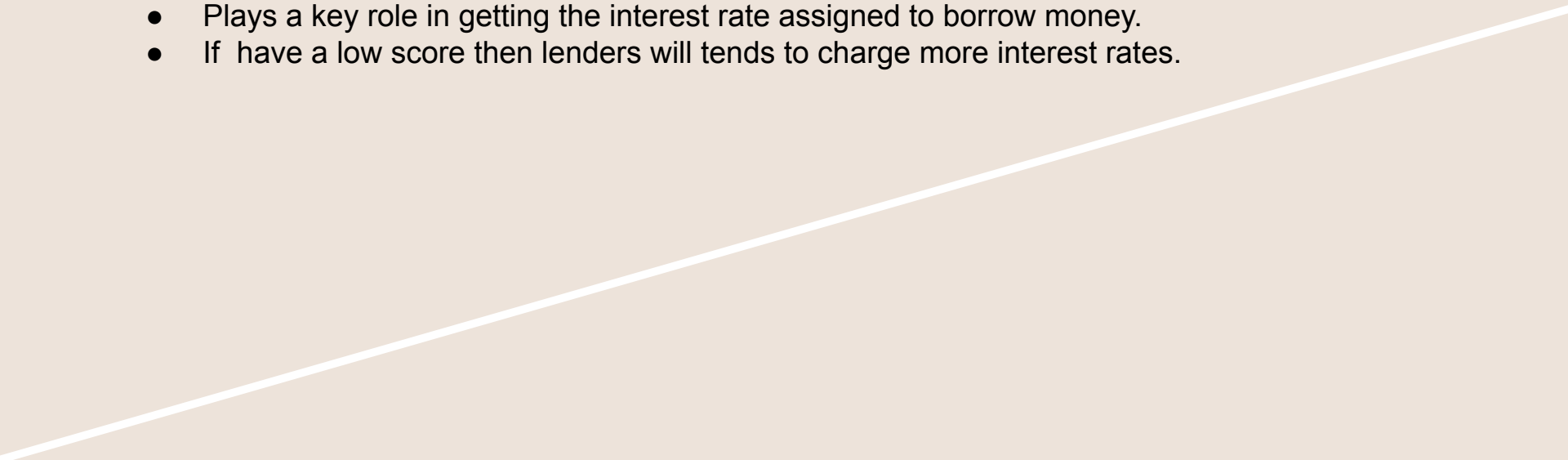
Credit Analysis

By:

Urvish Desai and Neeshma Khosa

A dark blue diagonal gradient bar that starts from the bottom left corner and extends towards the top right corner, covering the lower half of the slide.

Why Credit Score Analysis?

- One of the biggest impactful factors in getting your loan approval.
 - Plays a key role in getting the interest rate assigned to borrow money.
 - If have a low score then lenders will tends to charge more interest rates.
- 

Usage

Use By :

- Companies
- Banks
- Financial Institutions
- Individuals

Use for:

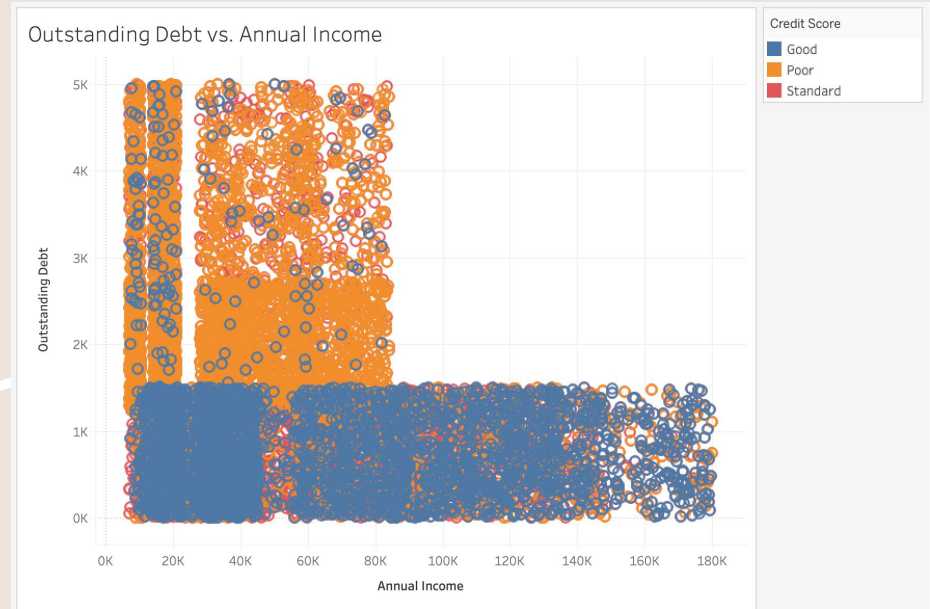
- Mortgage Approval
- Loan Approval
- Car Loans
- Estimating the Quality of a Credit Score Without Making a Hard Inquiry

Description of data

- Over the years, the company has collected basic bank details and gathered a lot of credit-related information. Our goal is to evaluate the data, clean it, then use it to predict the quality of the credit score (i.e. good, standard, or poor) for the individuals.
- There are 28 columns in it.
- Data is divided further into smaller tables in PostgreSQL in order to perform more targeted inquiries.

Data Exploration

- Tableau was used to explore relationships between variables
- A relationship between outstanding debt and annual income was found



Machine Learning Model

Data Preprocessing

- Data was preprocessed using Python in Jupyter Notebook.
- ID columns (Name, ID, Customer ID, SSN) and null values were dropped.
- Columns with string data types were encoded using OneHotEncoder.
 - Prior to encoding columns with >10 unique values were binned.
 - Occupation column was left at 15 equally distributed unique values.
- Data was then split into features and target
 - The first set had 3 targets:
 - Good credit score
 - Standard credit score
 - Poor credit score
 - The second set had 1 target column which had 2 possible values:
 - Good credit score
 - Poor credit score

Machine Learning Model cont.

Feature Engineering

- During exploratory analysis a relationship between the Outstanding Debt and Annual Income columns was shown to indicate credit score quality on some level.
- A new column was created and added to the DataFrame by dividing the Outstanding_Debt column by the Annual_Income column.

Feature Selection

- Only the identification columns were left out of feature selection as they do not provide any analytical power.
- The target column was Credit_Score.

Training and Testing

- Prior to scaling, the data was split into features and target,
- Then train_test_split was used from the sklearn library to split the features and target into a training and testing set, which was stratified.

Machine Learning Model cont.

Model Choice

- RandomForest
 - Started with a random forest model for the categorical output
 - Only had an accuracy score of 0.695
 - Presumably due to the three categorical outputs this model was insufficient
 - Ran relatively quickly
- Neural Net
 - Moved on to a NN model
 - 3 layers, ReLU activation on inner layers, Softmax activation on output layer
 - Categorical crossentropy loss metric, adam optimizer, and accuracy score used
 - Trained over 75 epochs to try and reduce overfitting
 - Loss: 0.82, accuracy: 0.65
 - Presumably due to the reduced training data size (3 possible outputs vs. 2) the data is overfitting to the training set
 - Took a long time to run

Machine Learning Model cont.

Model Choice cont.

We decided to collapse the 'Standard' category for Credit_Score into the 'Good' category to increase the training data.

- RandomForest
 - Started with a RandomForest model for the binary output
 - Only had an accuracy score of 0.83
 - Produced a good classification model
 - Ran relatively quickly
- Neural Net
 - Tried to run a NN model to see if accuracy could be increased
 - 3 layers, ReLU activation on inner layers, Sigmoid activation on output layer
 - Binary crossentropy loss metric, adam optimizer, and accuracy score used
 - Trained over 100 epochs
 - Loss: 0.56, accuracy: 0.77
 - Did not perform as well as the RandomForest model, likely due to overfitting
 - Took a long time to run

Dashboard

Tools

- Tableau Public

Interactive Elements

- Filter based on
 - credit score category
 - Occupation
 - Customer ID

Pie Chart of
Credit Score
categories

Credit score
category vs
occupation

Payment behaviour vs
credit score

Outstanding debt vs Annual income