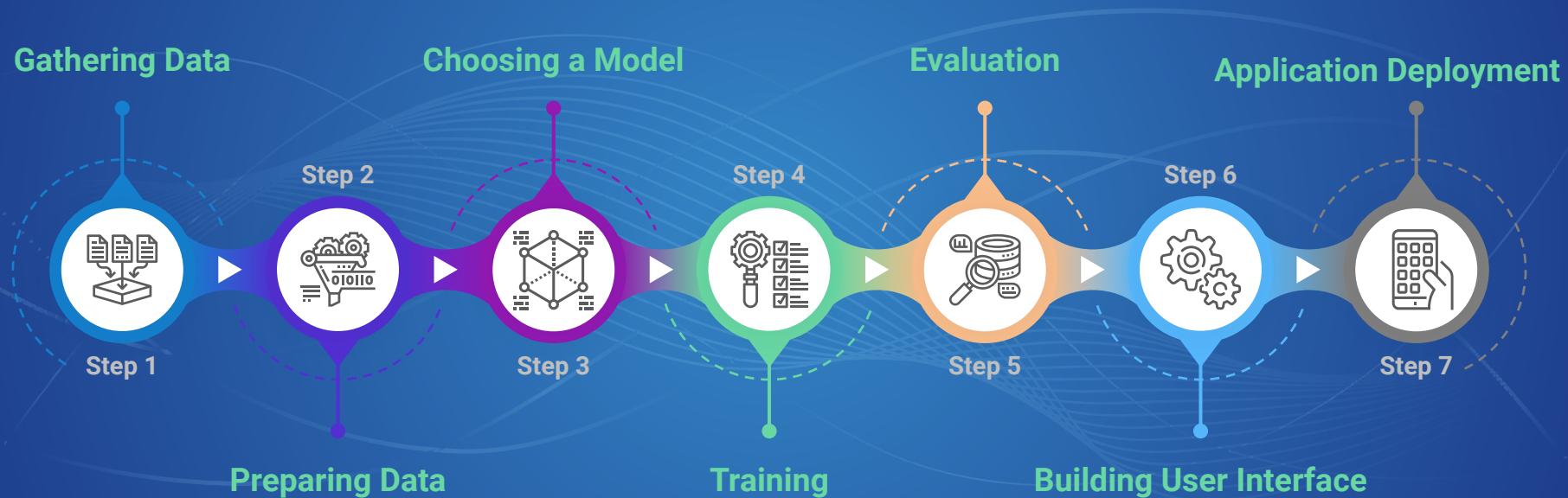


# Machine Learning Loan Application



Team: British Green, Karina Horna, Julian Louden, Reinhard Ndlovu, Amy Philippon and Kris Shaw

# Loan Application - Project Timeline



# Question and Assumptions

## Problem to Solve

- Processing loans is a core business function of many financial institutions.
- Automate the process of identifying the ideal loan applicant.
- Build a model to determine if an applicants loan will get approved or rejected.

## Assumptions: Most Impactful Metrics for Approved Loans

- High correlation between debt to income ratio, credit utilization rates, credit history and number of credit problems
- Income of \$60,000+
- Credit score of 650+
- No bankruptcies or liens
- Low delinquency rate

# Data Preprocessing for Loan Application Data Set

## Getting The Dataset



## Importing Datasets

Kaggle - Loan Application



## Importing Libraries

Jupyter Notebook  
Pandas  
Numpy  
Sklearn  
Seaborn  
RandomForestClassifier  
Matplotlib  
Pyplot  
Pickle



## Data Cleaning

Removing non-numeric characters  
Changing data types from string to integers  
Mean-filling missing data  
`dropna` (for text missing in columns)



## Encoding Categorical Data

`pd.get_dummies()`

- Short/Long Term
- Rent/Mortgage/Own
- Purpose of Loan
- Utilization Rate
- Debt to Income Ratio



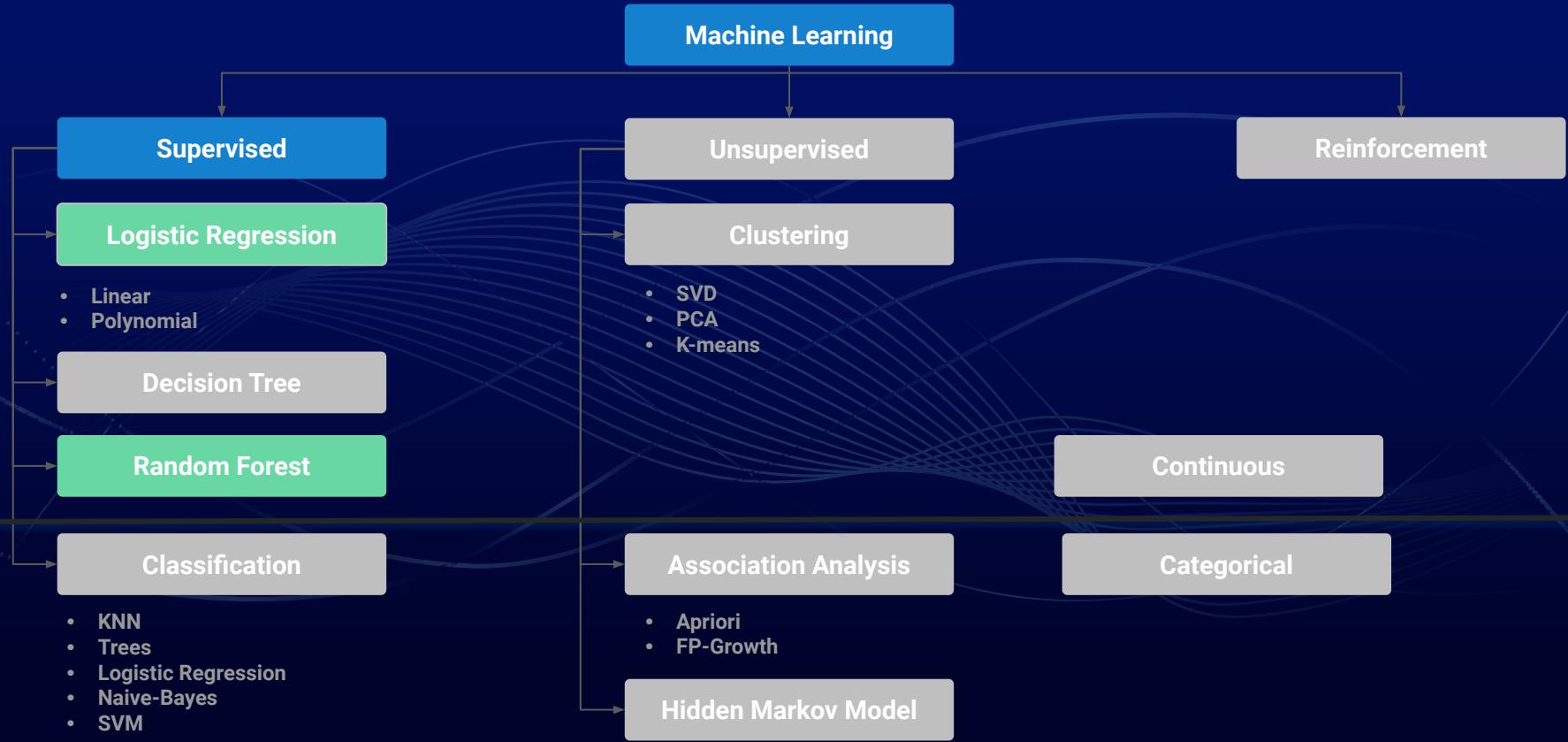
## Splitting Dataset Into Training And Test Set

## Feature Scaling

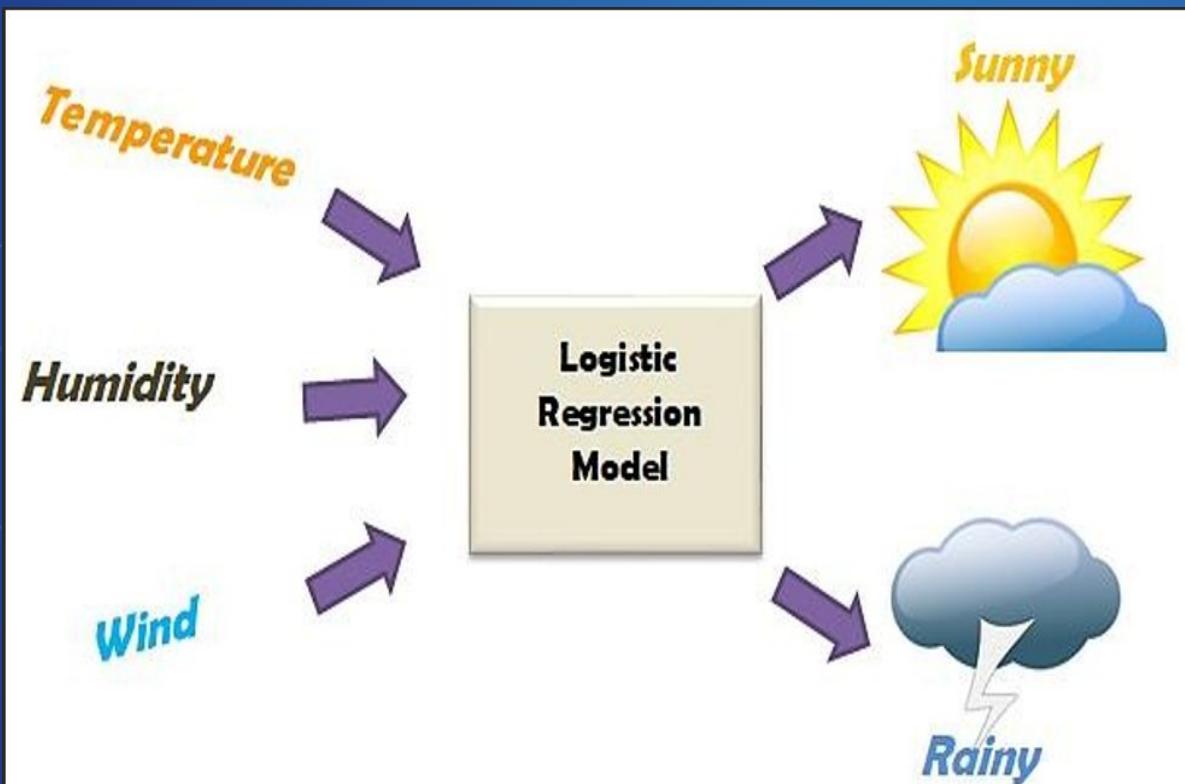
*not needed in this model*



# Choosing a Machine Learning Model



# Logistic Regression

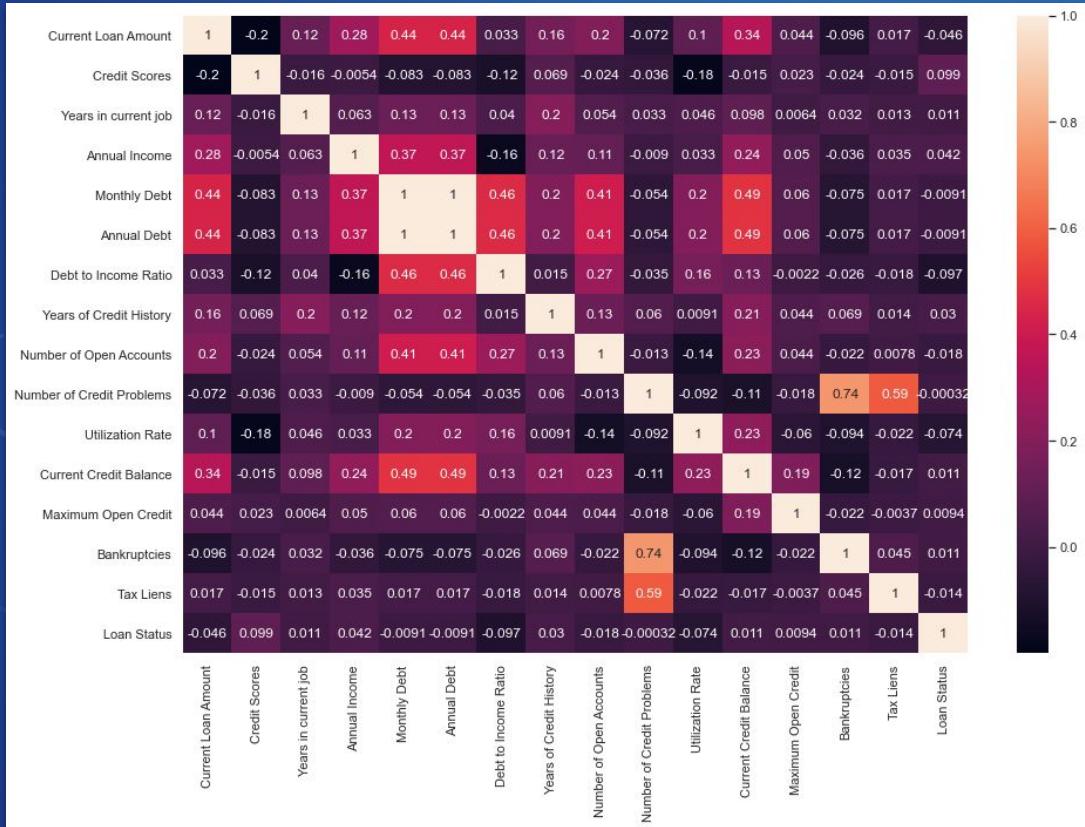


- What is Logistic Regression?
- What is Logistic Regression used for?
- What is the outcome or product of a Logistic Regression?
- The reason we chose Logistic Regression.

# Flask Setup

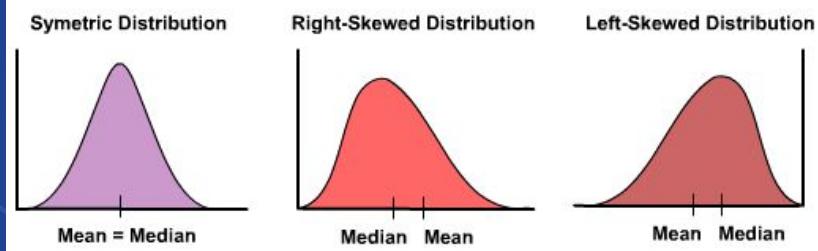
- Connection to index.html
  - Loading to pickle
  - Utilization of variables that are called on our form.
  - Setting conditions for the loan status after form submission

# Correlation Matrix

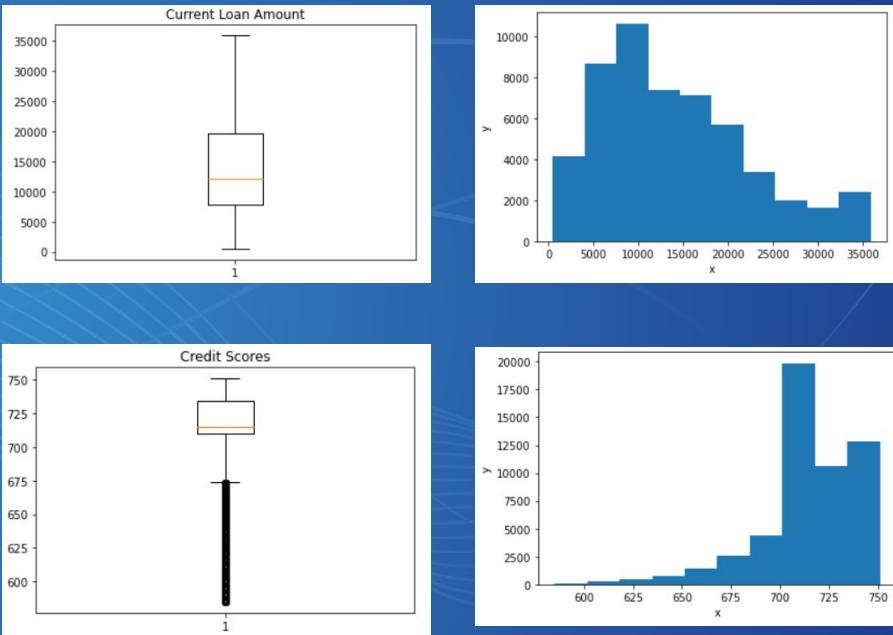


- Correlation matrix is consistent with hypothesis
- With 1 representing high correlation between debt to income ratio, credit history, number of credit problems, and credit utilization rates.

# Normalizing Data When the Model Isn't Working

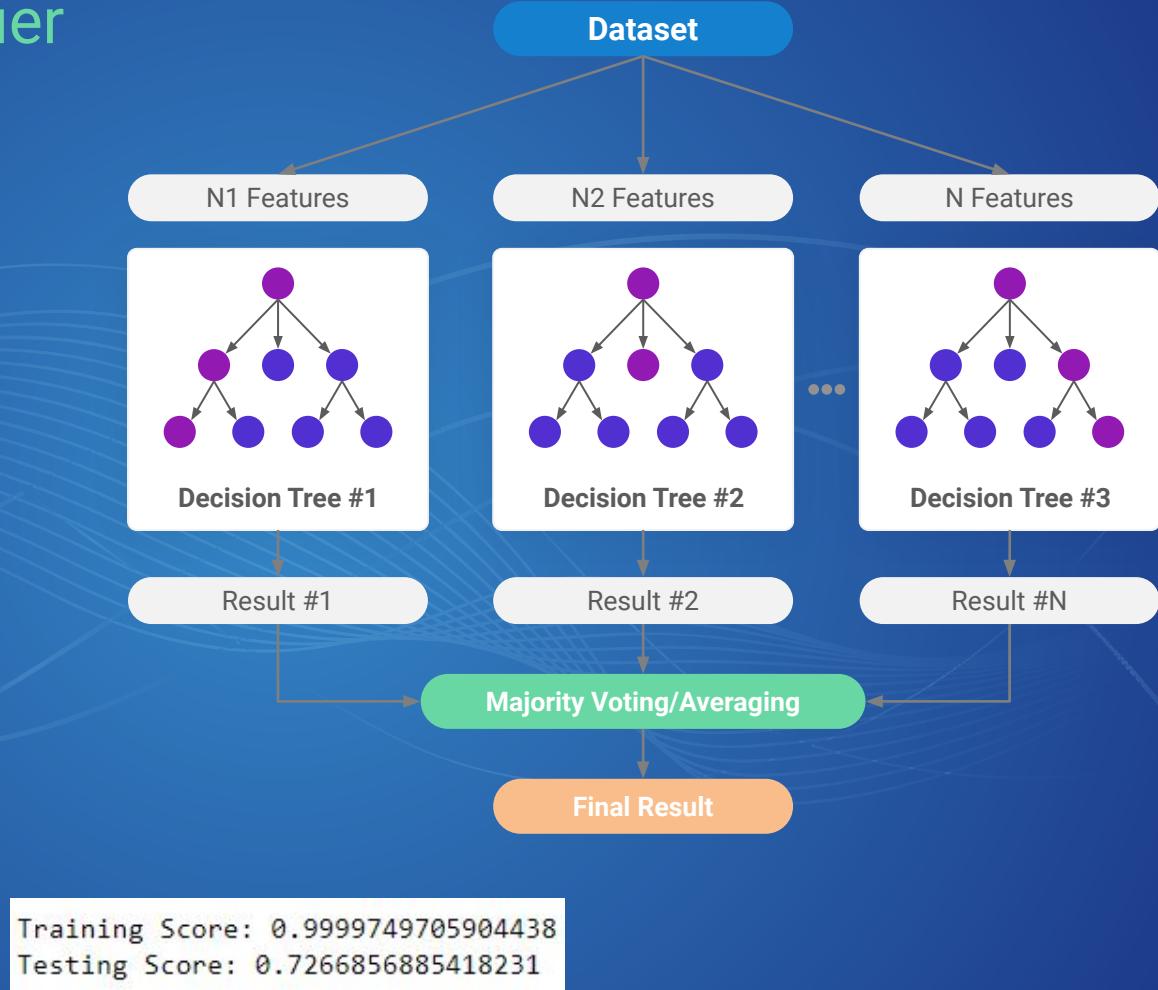


Common transformations of this data are:  
square root, cube root, or logarithmic



# Random Forest Classifier

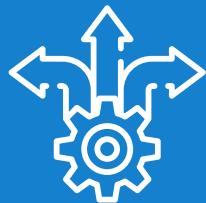
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 53271 entries, 0 to 53270
Data columns (total 44 columns):
 #   Column           Non-Null Count Dtype  
 --- 
 0   Current_Loan_Amount      53271 non-null int64  
 1   Credit_Scores            53271 non-null int64  
 2   Years_in_current_job    53271 non-null int64  
 3   Annual_Income            53271 non-null int64  
 4   Monthly_Debt             53271 non-null float64 
 5   Annual_Debt              53271 non-null float64 
 6   Debt_to_Income_Ratio     53271 non-null float64 
 7   Years_of_Credit_History  53271 non-null float64 
 8   Number_of_Open_Accounts  53271 non-null int64  
 9   Number_of_Credit_Problems 53271 non-null int64  
 10  Utilization_Rate          53271 non-null float64 
 11  Current_Credit_Balance   53271 non-null int64  
 12  Maximum_Open_Credit       53271 non-null int64  
 13  Bankruptcies              53271 non-null int64  
 14  Tax_Liens                 53271 non-null int64  
 15  Term_Long_Term            53271 non-null uint8  
 16  Term_Short_Term           53271 non-null uint8  
 17  Home_Ownership_Home_Mortgage 53271 non-null uint8  
 18  Home_Ownership_Own_Home    53271 non-null uint8  
 19  Home_Ownership_Rent        53271 non-null uint8  
 20  Purpose_Business_Loan     53271 non-null uint8  
 21  Purpose_Buy_House          53271 non-null uint8  
 22  Purpose_Buy_a_Car          53271 non-null uint8  
 23  Purpose_Debt_Consolidation 53271 non-null uint8  
 24  Purpose_Educational_Expenses 53271 non-null uint8  
 25  Purpose_Home_Improvements   53271 non-null uint8  
 26  Purpose_Medical_Bills      53271 non-null uint8  
 27  Purpose_Other               53271 non-null uint8  
 28  Purpose_Take_a_Trip         53271 non-null uint8  
 29  Purpose_major_purchase     53271 non-null uint8  
 30  Purpose_moving              53271 non-null uint8  
 31  Purpose_other                53271 non-null uint8  
 32  Purpose_renewable_energy    53271 non-null uint8  
 33  Purpose_small_business      53271 non-null uint8  
 34  Purpose_vacation             53271 non-null uint8  
 35  Purpose_wedding              53271 non-null uint8  
 36  Debt_to_Income_Ratio_G/B_Bad DTI 53271 non-null uint8  
 37  Debt_to_Income_Ratio_G/B_Excellent DTI 53271 non-null uint8  
 38  Debt_to_Income_Ratio_G/B_Good DTI   53271 non-null uint8  
 39  Debt_to_Income_Ratio_G/B_Very_Bad DTI 53271 non-null uint8  
 40  Utilization_Rate_Good/Bad_Bad UR 53271 non-null uint8  
 41  Utilization_Rate_Good/Bad_Excellent UR 53271 non-null uint8  
 42  Utilization_Rate_Good/Bad_Good UR   53271 non-null uint8  
 43  Utilization_Rate_Good/Bad_Very_Bad UR 53271 non-null uint8
```



# Data Prep for Tableau Visualizations

Outline of the process used to create our data vis criteria and analysis buckets for our visualizations

D1



## Decision Vis Prep

- After reviewing our data we decided we needed to add calculated fields.
- We added criteria buckets to ensure we could extract actionable insights from our data.

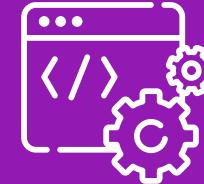
D2



## Analysis Buckets Created

- Loan amount buckets
- Annual income buckets
- Credit score buckets
- Years in current job
- Utilization rate
- Debt to income ratio
- Years of credit history

D3



## Data Visualized

- Loan status
- Loan type breakdown
- Loan amount vs purpose
- Years in current job
- Credit score vs # of credit problems
- Annual income vs job history

# UI Design and Build

## Machine Learning Loan Application



Processing loans are a critical part of the core business for banks and various other financial institutions. The main profit comes from the loans interest. Financial institutions grant loans after they verify, validate and approve an applicant. Approving an applicant is risky as the company issuing the loan doesn't have assurance if the applicant will or has the ability to pay back the loan.

For this project we wanted to build a predictive model using machine learning to help automate the process of targeting the right applicants and determining if they will be approved for a loan.

Hypothesis: We predict that there will be a high correlation between debt to income ration, credit utilization rates, credit history and number of credit problems for the applicants that get approved.

Variables for Exploration: Credit utilization rates, debt to income ratio, credit history, years in current job and number of credit problems.

HTML to add Tableau



```
29   <div class="col-lg-12 col-md-12">
30     <div class="box" style="padding-bottom: 20px;">
31       <h3 class="title">Machine Learning Application </h3>
32       
33       <p>Processing loans are a critical part of the core business for banks and various other financial institutions. The main profit comes from the loans issued by an applicant. Approving an applicant is risky as the company issuing the loan doesn't have assurance if the applicant will or has the ability to pay back the loan. For this project we wanted to build a predictive model using machine learning to help automate the process of targeting the right applicants and determining if they will be approved for a loan. There will be a high correlation between debt to income ration, credit utilization rates, credit history and number of credit problems for the applicants that get approved.</p>
34     </div>
35   </div>
36   <div class="box">
37     <h3 class="title">MODELS USED</h3>
38     <hr>
39     <div class="row">
40       <p><br>Scikit-learn <br>Python Pandas<br>Python Matplotlib<br>HTML/CSS<br>Bootstrap <br>Tableau <br>Feature Engineering for Machine Learning<br></p>
41   </div>
42   <div class="row">
43     <div class="col-lg-12 col-md-12 ">
44       <div class="box" >
45         
46     </div>
47   </div>
48 </div>
```

HTML for body



```
<div class="row" style="padding-bottom: 70px; ">
<div class="col-6 ">
  <div class="title ">TABLEAU
    <div class="tableauPlaceholder" id="viz1646593614492" style="position: relative"><object class="tableauViz" style="display:none;"><param name='host_url' value='https://public.tableau.com/javascripts/api/viz_v1.js' /><param name='host_name' value='public.tableau.com' /><param name='content' value='viz1646593614492' /><param name='js_url' value='/static/api/viz_v1.js' /></object><div><script type="text/javascript">
      var divElement = document.getElementById('viz1646593614492');
      var vizElement = divElement.getElementsByTagName('object')[0];
      vizElement.style.width = '1016px';
      vizElement.style.height = '991px';
      var scriptElement = document.createElement('script');
      scriptElement.src = 'https://public.tableau.com/javascripts/api/viz_v1.js';
      vizElement.parentNode.insertBefore(scriptElement, vizElement);
    </script></div>
  </div>
<div class="title ">ANALYSIS
  
<div class="title ">MACHINE LEARNING</div>
</div>
```

# UI Design and Build

## Loan Approval Prediction

### Credit Score

## Credit Score

## Annual income

## Annual Income

### Current Loan Amount

### Current Loan Amount

**Submit**

## HTML for user input

# App.py code

# Web App Deployment

- Steps for deployment
  - Created a repo for the application
  - Prepped the application with configuration files
  - Created the Heroku application
  - Prepped the Heroku database and deployed the app
- Challenges
  - Case sensitive

# Live Demo

