I want to buy a new car

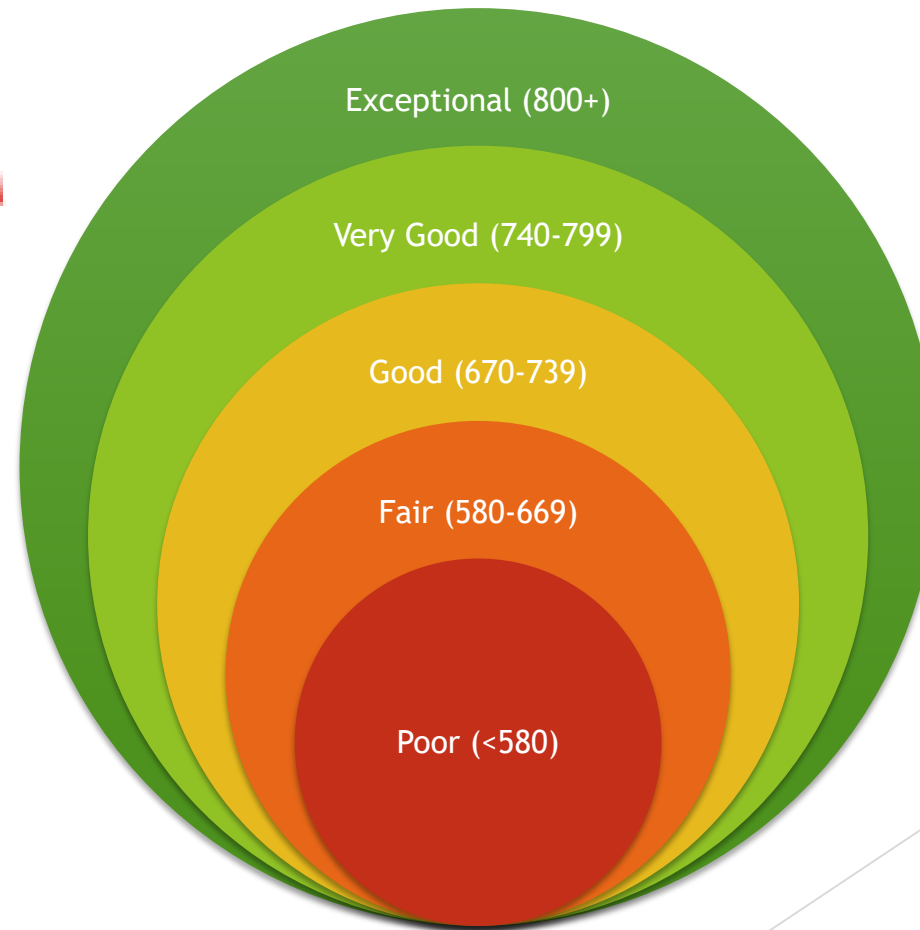

I want to buy a new house

Where do you go to get a loan?

Defining "Credit Worthiness"

- Annual Income ?
- Credit Score ?
- Home Ownership ?
- Current Loan Amount ?
- Term ?

Credit Score Scale

Exceptional (800+)

Very Good (740-799)

Good (670-739)

Fair (580-669)

Poor (<580)

779

687

726

463

760

645

732

582

540

Dataset Source

DATA CLEANING

Drop the rows with null values such Loan_ID, Customer_ID etc

```
1  dfLoan = df_train[df_train['Loan_ID'].notnull()]
```

```
1  #check if there are rows where Loan_ID is null
2  dfLoan.Loan_ID.isnull().sum()
```

```
#Replacing Loan_Status description to binary integer classification (0 for Charged O

df_train["Loan_Status"].replace(['Fully Paid', 'Charged Off'], [1, 0], inplace=True)
```

Column Row

DATASETS

- The refined train and test datasets are used in order to build the Model.

Predictive Modeling and Classification

- Logistic Regression
- Naïve Bayes
- K Nearest Neighbors
- Random Forests
- Decision Tree
- Support Vector Machine

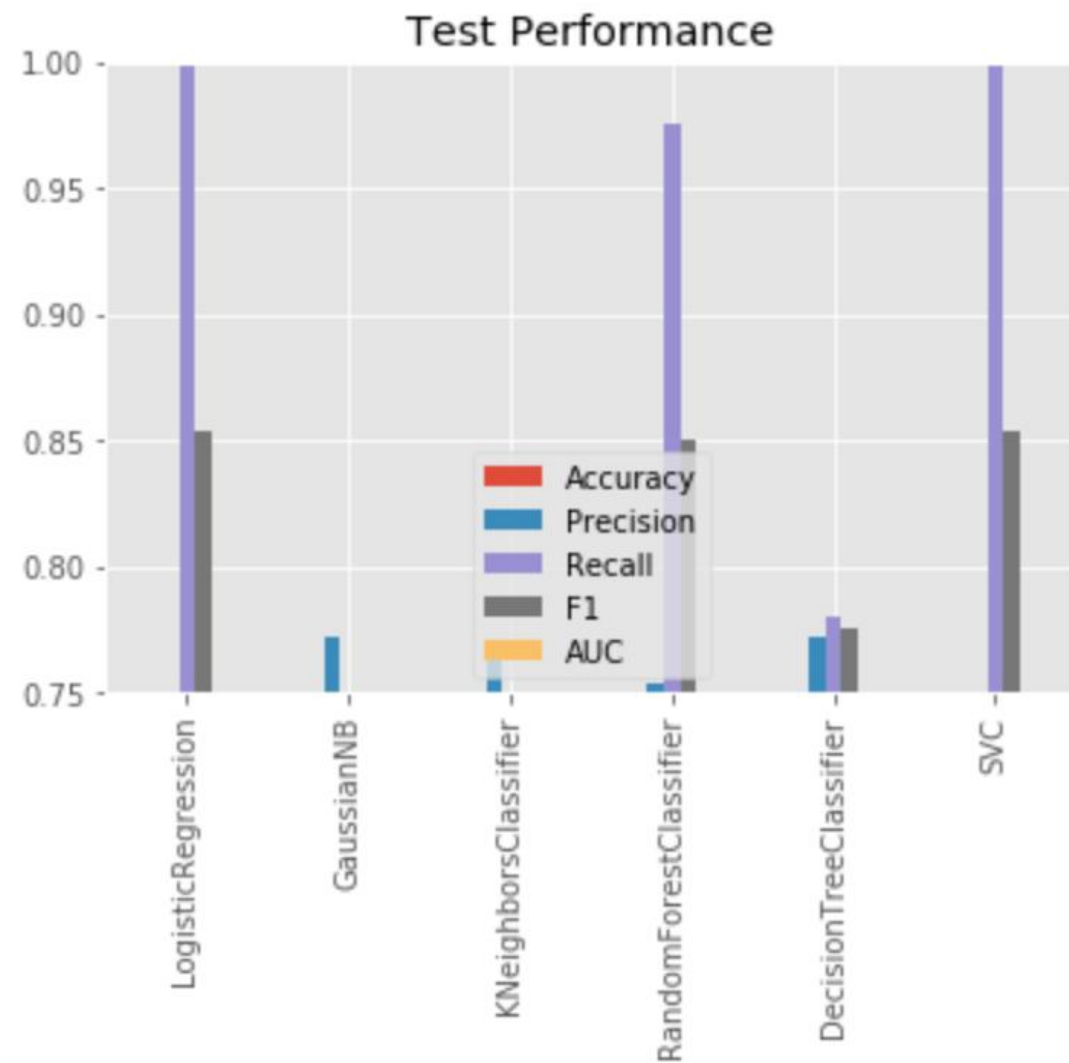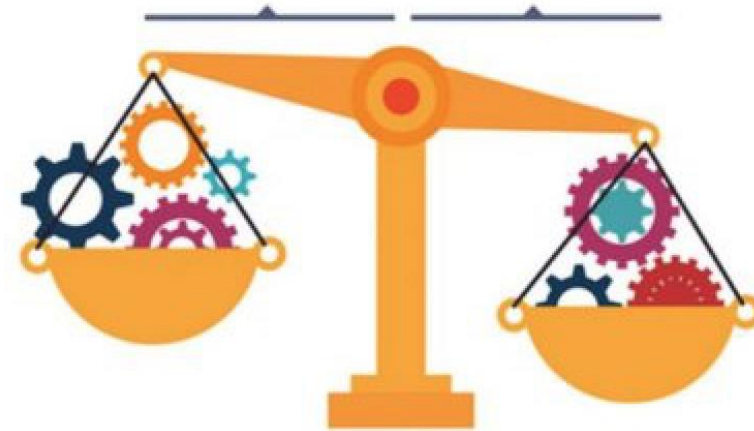We built the Machine Learning Model and fit them in dataset.

Predictive Modeling and Classification

	LogisticRegression	GaussianNB	KNeighborsClassifier	RandomForestClassifier	DecisionTreeClassifier	SVC
Accuracy	0.558005	0.511296	0.708864	0.860501	0.812796	0.558513
Precision	0.556717	0.589008	0.876905	0.886950	0.880754	0.560882
Recall	0.502307	0.042300	0.475449	0.821385	0.716868	0.475975
F1	0.528101	0.070063	0.616585	0.852903	0.790397	0.514823
AUC	0.557169	0.504257	0.705361	0.859914	0.811356	0.557274

Test Performance

Accuracy
Precision
Recall
F1
AUC

LogisticRegression
GaussianNB
KNeighborsClassifier
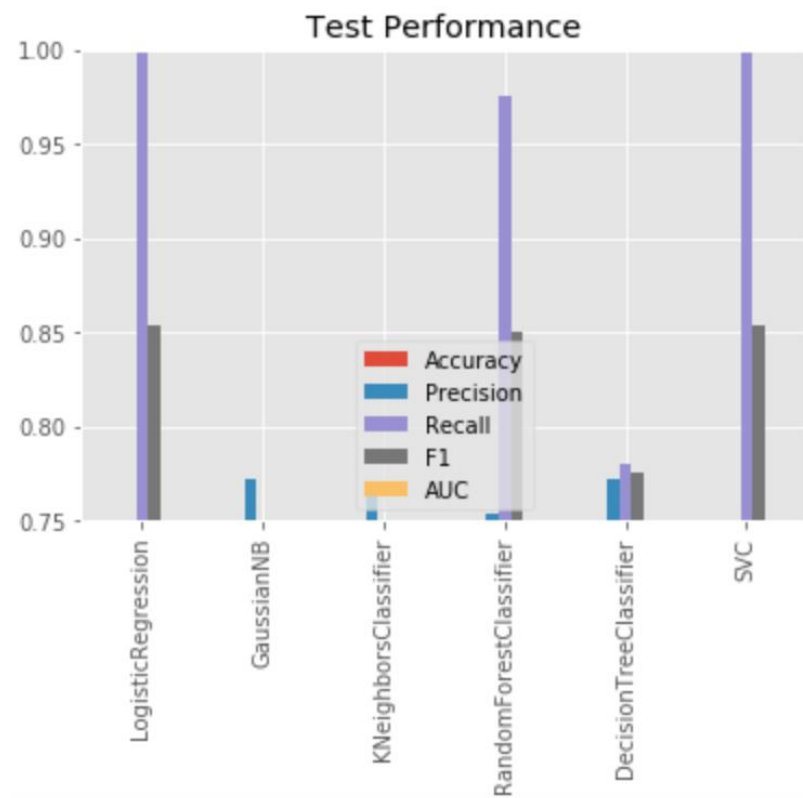RandomForestClassifier
DecisionTreeClassifier
SVC

Imbalanced Data

Balancing of Data

Based on the given dataset, the data is imbalance based on loan Status, hence, we need to balance the data to have more accurate prediction. To balance the data, we replicated the charged off multiple times until the data become balanced.

Before and After Balancing of Data

Before

Random Forest Classifier has a more balanced data after

After

Summary