

## Importing Data-Set

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

## Loading Data

```
In [2]: train=pd.read_csv("train.csv")
test=pd.read_csv("test.csv")
```

```
In [3]: train.head()
```

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area	Loan_Status
Out [3]:	0	LP001002	Male	No	0	Graduate	No	5849	0.0	NaN			360
	1	LP001002	Male	Yes	1	Graduate	No	4583	1508.0	128.0			360
	2	LP001005	Male	Yes	0	Graduate	Yes	3000	0.0	66.0			360
	3	LP001006	Male	Yes	0	Not Graduate	No	2583	2358.0	120.0			360
	4	LP001008	Male	No	0	Graduate	No	6000	0.0	141.0			360

```
In [4]: test.head()
```

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area	Loan_Status
Out [4]:	0	LP001015	Male	Yes	0	Graduate	No	5720	0	110.0			360
	1	LP001022	Male	Yes	1	Graduate	No	3076	1500	126.0			360
	2	LP001031	Male	Yes	2	Graduate	No	5000	1800	208.0			360
	3	LP001035	Male	Yes	2	Graduate	No	2340	2546	100.0			360
	4	LP001051	Male	No	0	Not Graduate	No	3276	0	78.0			360

```
In [5]: train.shape
```

```
Out [5]: (614, 13)
```

```
Out [6]: test.shape
```

```
Out [6]: (367, 12)
```

```
In [7]: #dropping the Loan_ID column
train.drop('Loan_ID',axis=1,inplace=True)
```

```
In [8]: #dropping the Loan_ID column
test.drop('Loan_ID',axis=1,inplace=True)
```

## Observations

```
In [9]: categorical_features_train = train.select_dtypes(include='object').columns
categorical_features_train
```

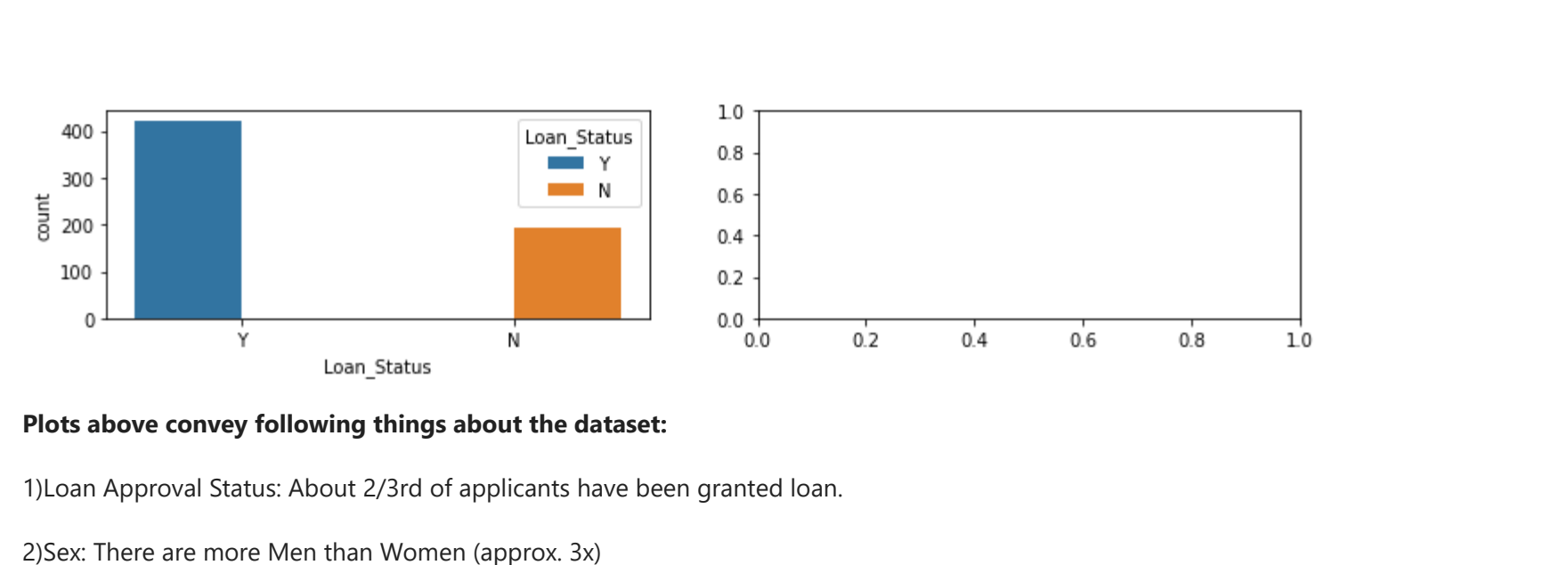
```
Out [9]: Index(['Gender', 'Married', 'Dependents', 'Education', 'Self_Employed',
'Property_Area', 'Loan_Status'],
dtype='object')
```

```
In [10]: continuous_features_train = train.select_dtypes(exclude='object').columns
continuous_features_train
```

```
Out [10]: Index(['ApplicantIncome', 'CoapplicantIncome', 'LoanAmount',
'Loan_Amount_Term', 'Credit_History',
dtype='object')
```

## For Categorical Features

```
In [11]: fig,axes=plt.subplots(4,2,figsize=(12,15))
for idx,col in enumerate(categorical_features_train):
row,col = idx//2,idx%2
sns.countplot(x=col,data=train,hue='Loan_Status',ax=axes[row,col])
plt.subplots_adjust(hspace=1)
```



Plots above convey following things about the dataset:

- 1)Loan Approval Status: About 2/3rd of applicants have been granted loan.
- 2)Sex: There are more Men than Women (approx. 3x)
- 3)Marital Status: 2/3rd of the population in the dataset is Married. Married applicants are more likely to be granted loans.
- 4)Dependents: Majority of the population have zero dependents and are also likely to be accepted for loan.
- 5)Education: About 5/6th of the population is Graduate and graduates have higher proportion of loan approval
- 6)Employment: About 5/6th of population is not self employed.
- 7)Property Area: More applicants from Semi-urban and also likely to be granted loans.
- 8)Applicant with credit history are far more likely to be accepted.
- 9)Loan Amount Term: Majority of the loans taken are for 360 Months (30 years).

## Feature Engineering

```
In [12]: train.dtypes
```

```
Out [12]: Gender          object
Married          object
Dependents       object
Education         object
Self_Employed    object
ApplicantIncome  int64
CoapplicantIncome int64
LoanAmount       float64
Loan_Amount_Term float64
Credit_History   object
Property_Area     object
dtype: object
```

```
In [13]: test.dtypes
```

```
Out [13]: Gender          object
Married          object
Dependents       object
Education         object
Self_Employed    object
ApplicantIncome  int64
CoapplicantIncome int64
LoanAmount       float64
Loan_Amount_Term float64
Credit_History   object
Property_Area     object
dtype: object
```

```
In [14]: train.isnull().sum()
```

```
Out [14]: Gender          11
Married          13
Dependents       10
Education         0
Self_Employed    23
ApplicantIncome   0
CoapplicantIncome 0
LoanAmount        5
Loan_Amount_Term  6
Credit_History   29
Property_Area     0
dtype: int64
```

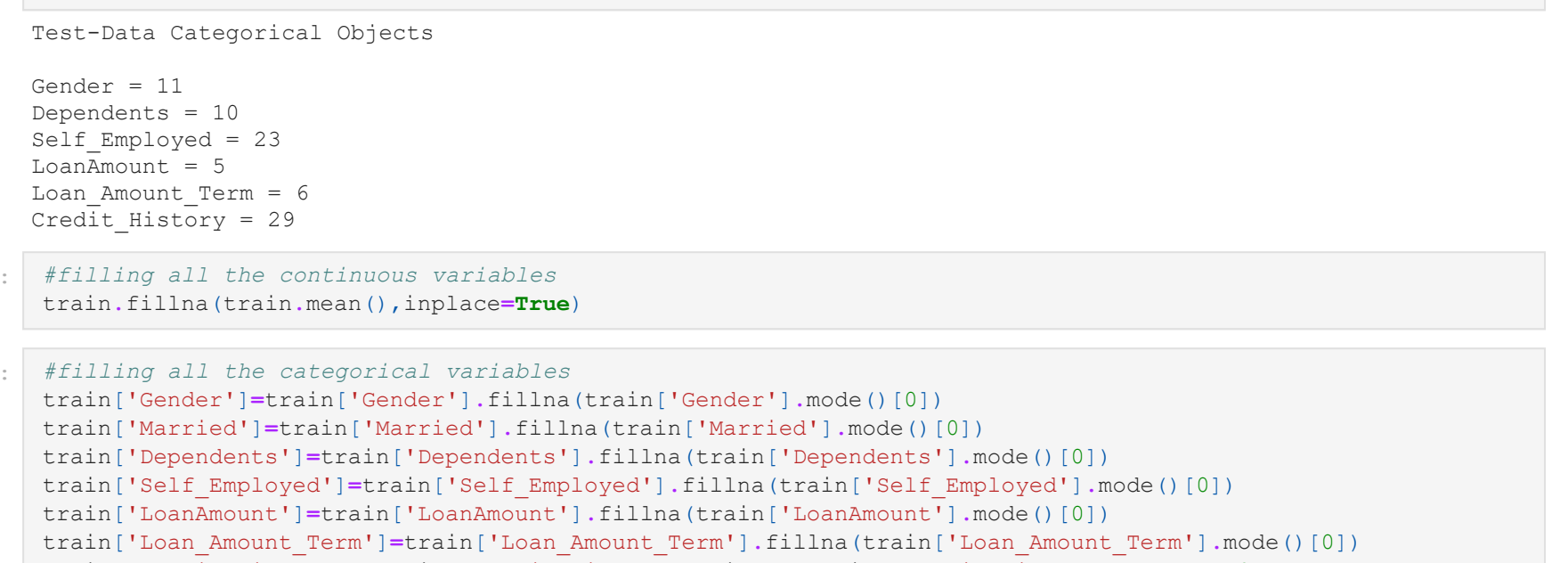
```
In [15]: test.isnull().sum()
```

```
Out [15]: Gender          11
Married          10
Dependents        9
Education         0
Self_Employed    23
ApplicantIncome   0
CoapplicantIncome 0
LoanAmount        5
Loan_Amount_Term  6
Credit_History   29
Property_Area     0
dtype: int64
```

## Checking null values

```
In [16]: sns.heatmap(train.isnull(),yticklabels=False,chart=False)
```

```
Out [16]: <AxesSubplot>
```



```
In [17]: print("Train Data Categorical Objects")
print()
```

```
Out [17]: categorical_features_train = train.select_dtypes(include='object').columns
for categorical_features_train in train:
if train[categorical_features_train].isnull().sum()>0:
print(categorical_features_train,"=",train[categorical_features_train].isnull().sum())
```

```
Train Data Categorical Objects
Gender = 11
Married = 13
Dependents = 10
Education = 0
Self_Employed = 23
ApplicantIncome = 0
CoapplicantIncome = 0
LoanAmount = 5
Loan_Amount_Term = 6
Credit_History = 29
Property_Area = 0
dtype: object
```

```
In [18]: print("Test Data Categorical Objects")
print()
```

```
Out [18]: categorical_features_test = test.select_dtypes(include='object').columns
for categorical_features_test in test:
if test[categorical_features_test].isnull().sum()>0:
print(categorical_features_test,"=",test[categorical_features_test].isnull().sum())
```

```
Test Data Categorical Objects
Gender = 11
Dependents = 10
Self_Employed = 23
ApplicantIncome = 0
CoapplicantIncome = 0
LoanAmount = 5
Loan_Amount_Term = 6
Credit_History = 29
Property_Area = 0
dtype: object
```

```
In [19]: #filling all the continuous variables
train.fillna(train.mean(),inplace=True)
```

```
In [20]: #filling all the categorical variables
train['Gender']=train['Gender'].fillna(train['Gender'].mode()[0])
train['Married']=train['Married'].fillna(train['Married'].mode()[0])
train['Dependents']=train['Dependents'].fillna(train['Dependents'].mode()[0])
train['Self_Employed']=train['Self_Employed'].fillna(train['Self_Employed'].mode()[0])
train['LoanAmount']=train['LoanAmount'].fillna(train['LoanAmount'].mode()[0])
train['Loan_Amount_Term']=train['Loan_Amount_Term'].fillna(train['Loan_Amount_Term'].mode()[0])
train['Credit_History']=train['Credit_History'].fillna(train['Credit_History'].mode()[0])
```

```
In [21]: #test data-set
test.fillna(test.mean(),inplace=True)
train['Gender']=train['Gender'].fillna(train['Gender'].mode()[0])
train['Married']=train['Married'].fillna(train['Married'].mode()[0])
train['Dependents']=train['Dependents'].fillna(train['Dependents'].mode()[0])
train['Self_Employed']=train['Self_Employed'].fillna(train['Self_Employed'].mode()[0])
train['LoanAmount']=train['LoanAmount'].fillna(train['LoanAmount'].mode()[0])
train['Loan_Amount_Term']=train['Loan_Amount_Term'].fillna(train['Loan_Amount_Term'].mode()[0])
train['Credit_History']=train['Credit_History'].fillna(train['Credit_History'].mode()[0])
```

```
In [22]: print("For Train Data-Set")
for categorical_features_train in train:
if train[categorical_features_train].isnull().sum()>0:
else:
break
print("No Null Values")
```

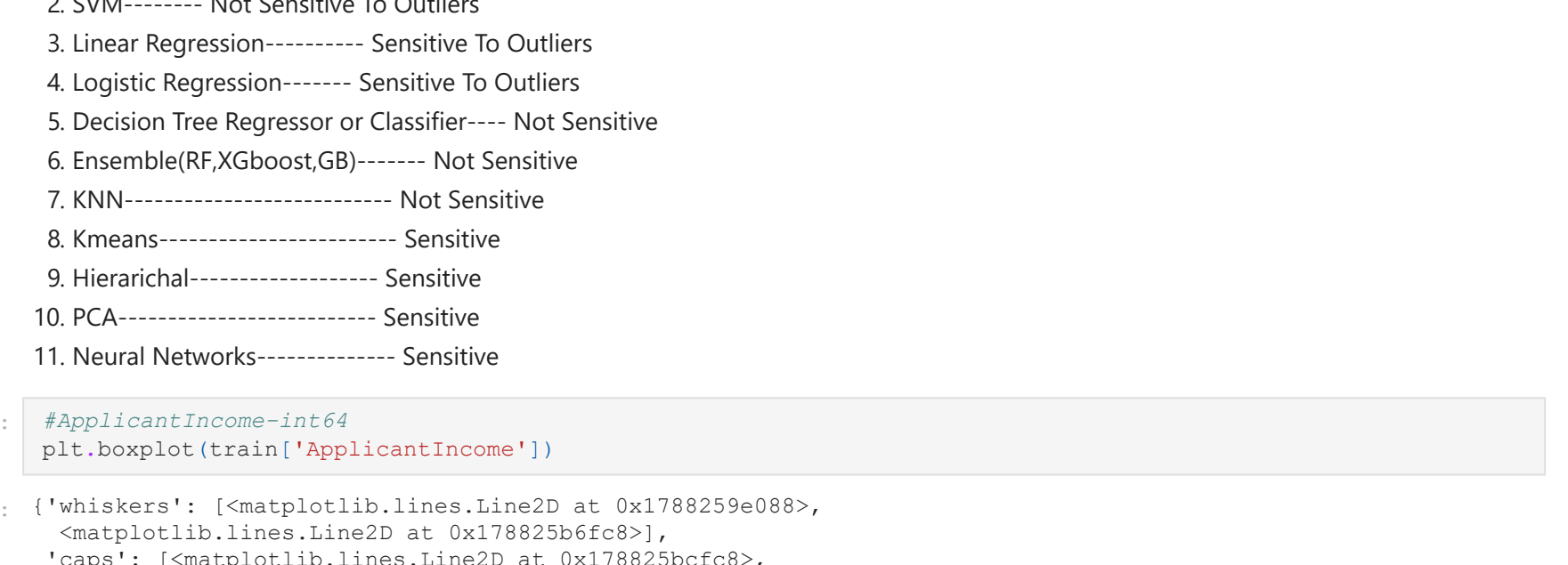
```
For Train Data-Set
No Null Values
```

```
In [23]: print("For Test Data-Set")
print()
```

```
Out [23]: for categorical_features_test in test:
if test[categorical_features_test].isnull().sum()>0:
else:
break
print("No Null Values")
```

```
For Test Data-Set
No Null Values
```

```
In [24]: #In the below graph the white colour represent the null values
sns.heatmap(train.isnull(),yticklabels=False,chart=False)
```



```
Out [24]: <AxesSubplot>
```

```
In [25]: train.isnull().sum()
```

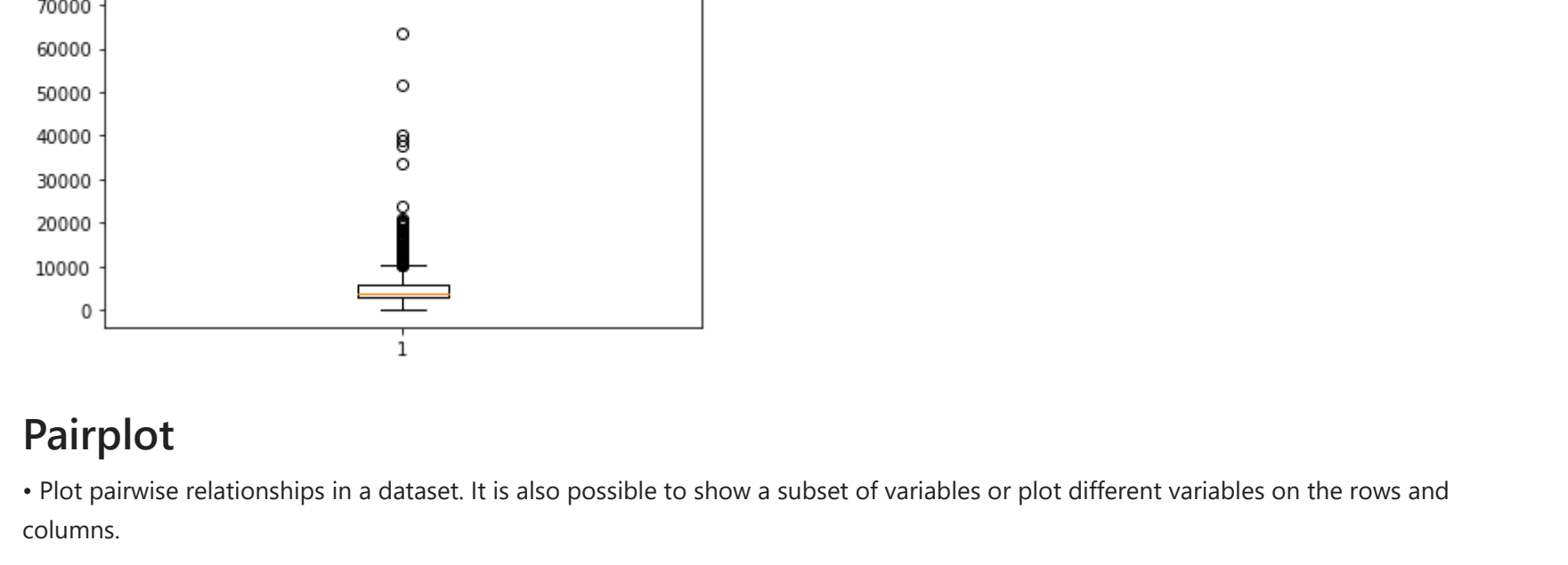
```
Out [25]: Gender          0
Married          0
Dependents        0
Education         0
Self_Employed    0
ApplicantIncome   0
CoapplicantIncome 0
LoanAmount        0
Loan_Amount_Term  0
Credit_History   0
Property_Area     0
Loan_Status       0
dtype: int64
```

## Handling Outliers

1. Naive Bayes Classifier---- Not Sensitive To Outliers
2. SVM----- Not Sensitive To Outliers
3. Linear Regression----- Sensitive To Outliers
4. Logistic Regression----- Sensitive To Outliers
5. Decision Tree Regressor or Classifier---- Not Sensitive
6. Ensemble(RF,Xgboost,GB)----- Not Sensitive
7. KNN----- Not Sensitive
8. Kmeans----- Sensitive
9. Hierarchical----- Sensitive
10. PCA----- Sensitive
11. Neural Networks----- Sensitive

```
In [26]: #ApplicantIncome=int64
plt.boxplot(train["ApplicantIncome"])
```

```
Out [26]: ['whiskers': [matplotlib.lines.Line2D at 0x1788259e088],
'matplotlib.lines.Line2D at 0x178825b6f08',
'caps': [matplotlib.lines.Line2D at 0x17882672e88],
'box': [matplotlib.lines.Line2D at 0x17882672e88],
'box': [matplotlib.lines.Line2D at 0x17882672e88],
'medians': [matplotlib.lines.Line2D at 0x17882672e88],
'fliers': [matplotlib.lines.Line2D at 0x17882672e88],
'means': []]
```



```
In [28]: #LoanAmount=float64
plt.boxplot(train["LoanAmount"])
```

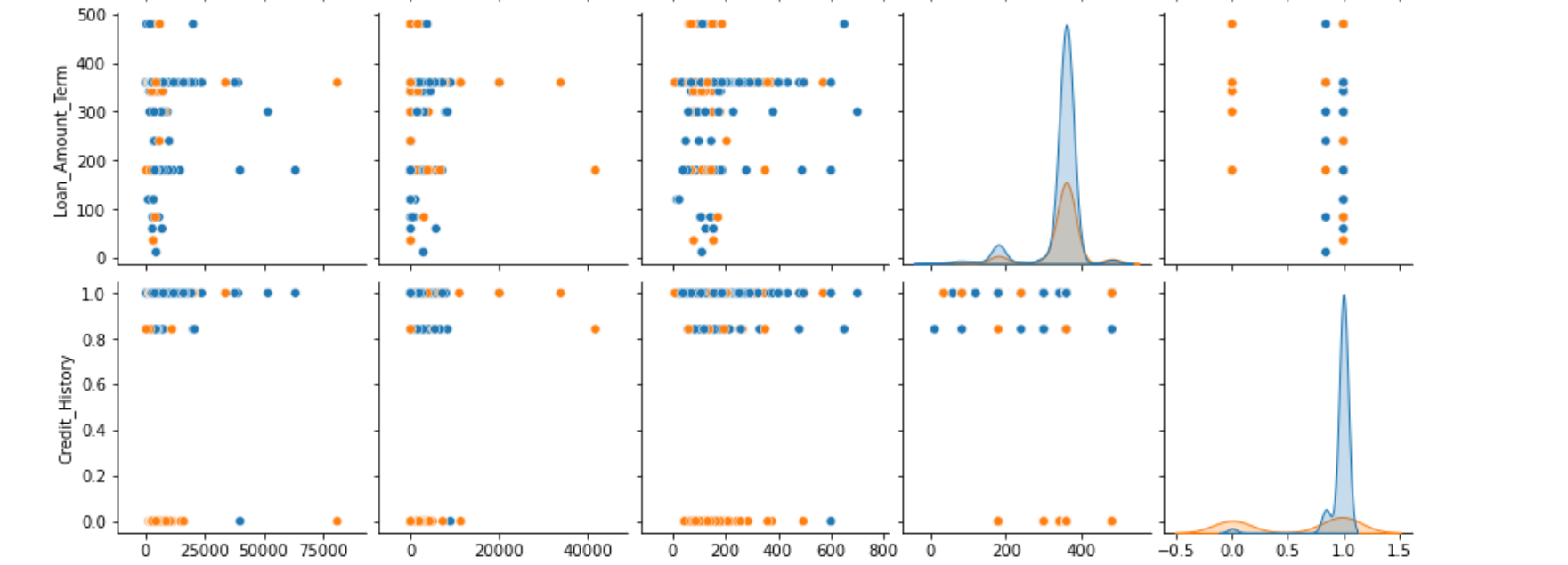
```
Out [28]: ['whiskers': [matplotlib.lines.Line2D at 0x1788262a0a08],
'matplotlib.lines.Line2D at 0x1788262a0a08',
'caps': [matplotlib.lines.Line2D at 0x1788262a0a08],
'box': [matplotlib.lines.Line2D at 0x1788262a0a08],
'box': [matplotlib.lines.Line2D at 0x1788262a0a08],
'medians': [matplotlib.lines.Line2D at 0x1788262a0a08],
'fliers': [matplotlib.lines.Line2D at 0x1788262a0a08],
'means': []]
```

## Pairplot

Plot pairwise relationships in a dataset. It is also possible to show a subset of variables or plot different variables on the rows and columns.

```
In [29]: sns.pairplot(train,hue='Loan_Status')
```

```
Out [29]: <AxesGrid: PairGrid at 0x178826a7e08>
```



By observing the above graphs we can say that there is a more overlapping between the 2 features . so it's not better to use the logistic regression here it's better to use KNN

```
In [30]: #combining the train and test data-set
train_copy=train.copy()
test_copy=test.copy()
train_test=pd.concat([train,test],axis=0)
train_test.shape
```

```
Out [30]: (981, 12)
```

```
In [31]: train_test.head()
```

	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area	Loan_Status
Out [31]:	0	Male	No	0	Graduate	No	5849	0.0	146.412162			360.0
	1	Male	Yes	1	Graduate	No	4583	1508.0	128.000000			360.0
	2	Male	Yes	0	Graduate	Yes	3000	0.0	66.000000			360.0
	3	Male	Yes	0	Not Graduate	No	2583	2358.0	120.000000			360.0
	4	Male	No	0	Graduate	No	6000	0.0	141.000000			360.0

## Handling Categorical Variables

```
In [32]: for i in train_test.select_dtypes(include='object').columns:
print(i,"=",len(train_test[i].unique()))
```

```
Out [32]: Gender = 2
Married = 2
Dependents = 4
Education = 2
Self_Employed = 2
Property_Area = 3
Loan_Status = 3
```

```
In [33]: train_test_copy=train_test.copy()
train_test = pd.get_dummies(train_test.drop_first=True))
```

```
Out [33]: train_test.isnull().sum()
```

```
Out [34]: ApplicantIncome      0
CoapplicantIncome      0
LoanAmount             0
Loan_Amount_Term       0
Credit_History         0
Gender_Male            0
Married_Yes           0
Dependents_1          0
Dependents_2          0
Dependents_3+         0
Education_Not Graduate 0
Self_Employed_Yes     0
Property_Area_Semiurban 0
Property_Area_Urban   0
Loan_Status_Y         0
dtype: int64
```

```
In [35]: train_test.shape
```

```
Out [35]: (981, 15)
```

```
In [36]: train_test.head()
```

	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Gender_Male	Married_Yes	Dependents_1	Dependents_2	Dependents_3+	Education_Not Graduate	Self_Employed_Yes	Property_Area_Semiurban	Property_Area_Urban	Loan_Status_Y
Out [36]:	0	5849	0.0	146.412162	360.0	1.0	1	0	0	0	0	0	0	0	0
	1	4583	1508.0	128.000000	360.0	1.0	1	1	1	1	1	1	1	1	1
	2	3000	0.0	66.000000	360.0	1.0	1	1	1	1	1	1	1	1	1
	3	2583	2358.0	120.000000	360.0	1.0	1	1	1	1	1	1	1	1	1
	4	6000	0.0	141.000000	360.0	1.0	1	1	1	1	1	1	1	1	1

```
In [37]: train.head()
```

	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area	Loan_Status
Out [37]:	0	Male	No	0	Graduate	No	5849	0.0	146.412162			360.0
	1	Male	Yes	1	Graduate	No	4583	1508.0	128.000000			360.0
	2	Male	Yes	0	Graduate	Yes	3000	0.0	66.000000			360.0
	3	Male	Yes	0	Not Graduate	No	2583	2358.0	120.000000			360.0
	4	Male	No	0	Graduate	No	6000	0.0	141.000000			360.0

```
In [38]: train_test.dtypes
```

```
Out [38]: ApplicantIncome      int64
CoapplicantIncome      float64
LoanAmount             float64
Loan_Amount_Term       float64
Credit_History         int64
Gender_Male            uint8
Married_Yes           uint8
Dependents_1          uint8
Dependents_2          uint8
Dependents_3+         uint8
Education_Not Graduate uint8
Self_Employed_Yes     uint8
Property_Area_Semiurban uint8
Property_Area_Urban   uint8
Loan_Status_Y         uint8
dtype: object
```

```
In [39]: train.isnull().sum()
```

```
Out [39]: Gender          0
Married          0
Dependents        0
Education         0
Self_Employed    0
ApplicantIncome   0
CoapplicantIncome 0
LoanAmount        0
Loan_Amount_Term  0
Credit_History   0
Property_Area     0
Loan_Status       0
dtype: int64
```

```
In [40]: #dropping the data-set
train=train.iloc[614:]
test=test.iloc[614:]
```

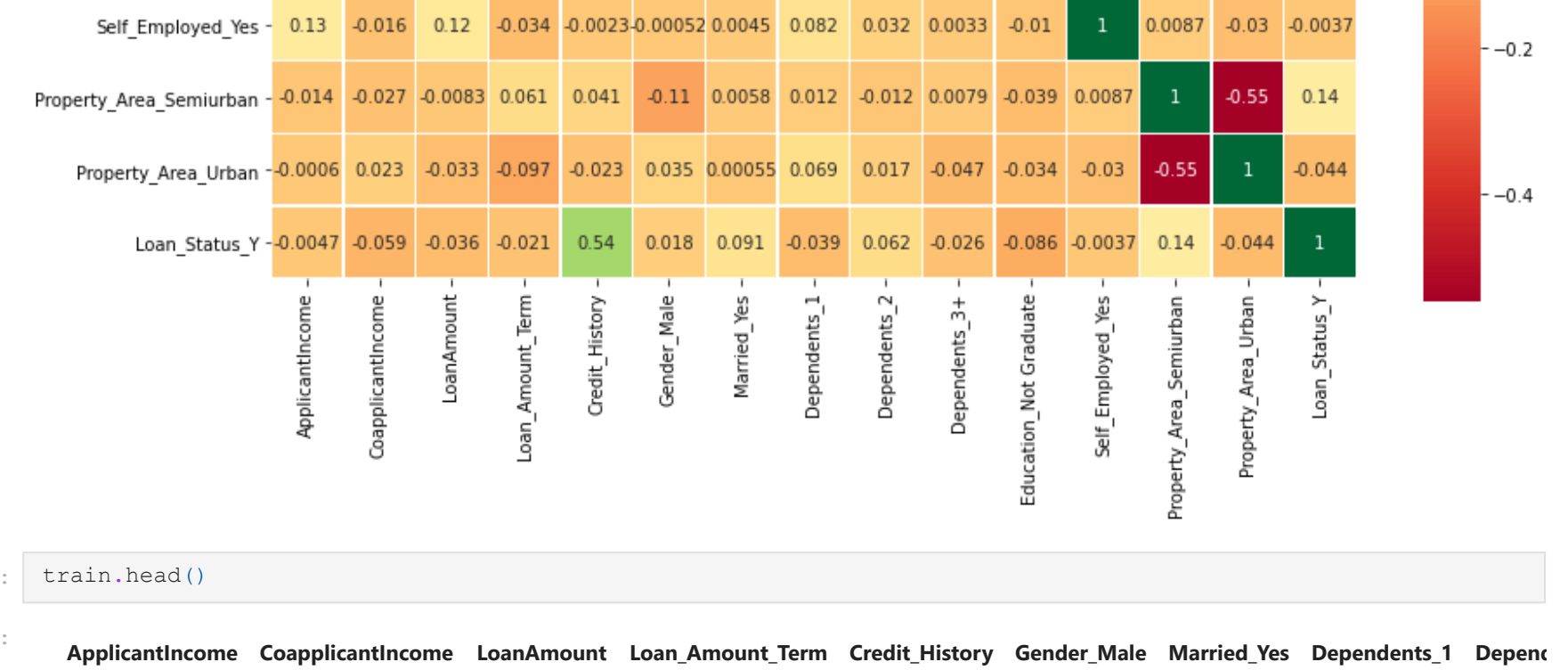
```
In [41]: (train.shape),(test.shape)
```

```
Out [41]: ((614, 15), (367, 15))
```

```
In [42]: #choosing co-relation
colormap=plt.cm.RdBu
plt.figure(figsize=(14,12))
plt.title('Pearson Correlation of Features',y1=0.5,size=15)
```

```
Out [42]: sns.heatmap(train.astype(float).corr(),linewidth=0.1,vmask=1.0,
square=True, cmap='RdBu',linecolor='white',annot=True)
```

```
Out [42]: <AxesSubplot:title='center':Pearson Correlation of Features>
```



```
In [43]: train.head()
```

	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Gender_Male	Married_Yes	Dependents_1	Dependents_2	Dependents_3+	Education_Not Graduate	Self_Employed_Yes	Property_Area_Semiurban	Property_Area_Urban	Loan_Status_Y
Out [43]:	0	5849	0.0	146.412162	360.0	1.0	1	0	0	0	0	0	0	0	0
	1	4583	1508.0	128.000000	360.0	1.0	1	1	1	1	1	1	1	1	1
	2	3000	0.0	66.000000	360.0	1.0	1	1	1	1	1	1	1	1	1
	3	2583	2358.0	120.000000	360.0	1.0	1	1	1	1	1	1	1	1	1
	4	6000	0.0	141.000000	360.0	1.0	1	1	1	1	1	1	1	1	1

```
In [44]: train.isnull().sum()
```

```
Out [44]: ApplicantIncome      0
CoapplicantIncome      0
LoanAmount             0
Loan_Amount_Term       0
Credit_History         0
Gender_Male            0
Married_Yes           0
Dependents_1          0
Dependents_2          0
Dependents_3+         0
Education_Not Graduate 0
Self_Employed_Yes     0
Property_Area_Semiurban 0
Property_Area_Urban   0
Loan_Status_Y         0
dtype: int64
```

```
In [45]: test.drop(['Loan_Status_Y'],axis=1,inplace=True)
```

```
In [46]: X = train.drop(columns='Loan_Status_Y')
y = train['Loan_Status_Y']
```

## Using CV

```
In [47]: from sklearn.model_selection import cross_val_score
def cross_val_score(model,X,y, scoring='accuracy',cv=10):
return cvxv
```

## Logistic Regression

```
In [48]: from sklearn.linear_model import LogisticRegression
log_classifier=LogisticRegression()
```

```
Out [48]: print("Train Accuracy =",cv(log_classifier, cv=10).mean())
Train Accuracy = 0.8921531795435378
```

```
In [49]: log_classifier.fit(X,y)
pred=log_classifier.predict(test)
print("Test Accuracy =",pred.mean())
Test Accuracy = 0.8337874659400545
```

## K-Nearest Neighbours

```
In [50]: from sklearn.neighbors import KNeighborsClassifier
k=KNeighborsClassifier(n_neighbors=18)
print("Train Accuracy =",cv(k).mean())
Train Accuracy = 0.8200728048786399
```

```
In [51]: k.fit(X,y)
knn_pred=k.predict(test)
print("Test Accuracy =",knn_pred.mean())
Test Accuracy = 0.9237057220708447
```

## Decision Trees

```
In [52]: from sklearn.tree import DecisionTreeClassifier
tree_classifier=DecisionTreeClassifier(random_state=0,max_depth=3)
cv(tree_classifier,cv=10).mean()
```

```
Out [52]: 0.89976540443222
```

```
In [53]: tree_classifier.fit(X,y)
tree_pred=tree_classifier.predict(test)
tree_pred.mean()
```

```
Out [53]: 0.8283378746594006
```

## Random Forest

```
In [54]: from sklearn.ensemble import RandomForestClassifier
random_classifier = RandomForestClassifier(n_estimators=4
```



X[2] <= 0.5  
gini = 0.803  
samples = 375  
value = [175, 440]

X[3] <= 0.426  
gini = 0.455  
samples = 243  
value = [134, 248]

X[4] <= 0.941  
gini = 0.274  
samples = 332  
value = [38, 194]

X[13] <= 0.5  
gini = 0.451  
samples = 239  
value = [129, 247]

X[1] <= 2507.0  
gini = 0.278  
samples = 5  
value = [5, 1]

X[1] <= 331.5  
gini = 0.467  
samples = 25  
value = [22, 13]

X[3] <= 420.0  
gini = 0.249  
samples = 107  
value = [16, 181]

## SVM

```
In [57]: from sklearn.svm import SVC
svm_classifier = SVC()
print("Train Accuracy =" , cv(svm_classifier, cv=10).mean())

Train Accuracy = 0.829034746304472
```

```
In [58]: svm_classifier.fit(X,y)
svm_pred=svm_classifier.predict(test)
print("Test Accuracy =" , svm_pred.mean())

Test Accuracy = 1.0
```

## AdaBoost

```
In [59]: AdaBoostLearn.ensemble import AdaBoostClassifier
ada_classifier=AdaBoostClassifier(learning_rate=0.5)
cv(ada_classifier,cv=10).mean()
```

Out[59]: 0.8940568173117258

```
In [60]: ada_classifier.fit(X,y)
ada_pred=ada_classifier.predict(test)
ada_pred.mean()
```

Out[60]: 0.814713984577657

## Gradient Boost

```
In [61]: from sklearn.ensemble import GradientBoostingClassifier
gb = GradientBoostingClassifier(n_estimators=50, max_features=2)
cv(gb,cv=10).mean()
```

Out[61]: 0.8931462739836444

```
In [62]: gb.fit(X,y)
gb_pred=gb.predict(test)
gb_pred.mean()
```

Out[62]: 0.8310621702997275

## Using Tensorflow

```
In [63]: from keras.models import Sequential
from keras.layers import Dense
from keras.layers import Dropout
from keras.layers import Regularizers
import pandas as pd
```

Using TensorFlow backend.

```
In [64]: import tensorflow as tf
```

```
In [65]: train=pd.read_csv("train.csv")
```

```
In [66]: train.head()
```

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term
	LP001002	Male	No	0	Graduate	No	5849	0.0	NaN	360
1	LP001003	Male	Yes	1	Graduate	No	4583	1508.0	128.0	360
2	LP001005	Male	Yes	0	Graduate	Yes	3000	0.0	66.0	360
3	LP001006	Male	Yes	0	Not Graduate	No	2583	2358.0	120.0	360
4	LP001008	Male	No	0	Graduate	No	6000	0.0	141.0	360

```
In [67]: train.isnull().sum()
```

```
Out[67]: Loan_ID      0
Gender          13
Married         3
Dependents     35
Education       0
Self_Employed  32
ApplicantIncome 0
CoapplicantIncome 0
LoanAmount     22
Loan_Amount_Term 14
Credit_History 50
Property_Area   0
Loan_Status     0
dtype: int64
```

## Handling Categorical Features

```
In [68]: from sklearn.preprocessing import LabelEncoder
train.iloc[:,0] = LabelEncoder().fit_transform(train.iloc[:,0].astype('str'))
train.iloc[:,1] = LabelEncoder().fit_transform(train.iloc[:,1].astype('str'))
train.iloc[:,2] = LabelEncoder().fit_transform(train.iloc[:,2].astype('str'))
train.iloc[:,3] = LabelEncoder().fit_transform(train.iloc[:,3].astype('str'))
train.iloc[:,4] = LabelEncoder().fit_transform(train.iloc[:,4].astype('str'))
train.iloc[:,5] = LabelEncoder().fit_transform(train.iloc[:,5].astype('str'))
train.iloc[:,11] = LabelEncoder().fit_transform(train.iloc[:,11].astype('str'))
train.iloc[:,12] = LabelEncoder().fit_transform(train.iloc[:,12].astype('str'))
```

Out[68]: 1614, 13

In [69]: train.head()

```
Out[70]: Loan_ID Gender Married Dependents Education Self_Employed ApplicantIncome CoapplicantIncome LoanAmount Loan_Amount_Term
0      0      1      1      0      0      0      5849      0.0      NaN      360
1      1      1      1      1      0      0      4583      1508.0     128.0     360
2      2      1      1      0      0      1      3000      0.0      66.0      360
3      3      1      1      0      1      0      2583      2358.0     120.0     360
4      4      1      0      0      0      0      6000      0.0      141.0     360
```

In [71]: train.shape

Out[71]: (614, 13)

```
In [72]: train.isnull().sum()
```

```
Out[72]: Loan_ID      0
Gender          0
Married         0
Dependents      0
Education        0
Self_Employed   0
ApplicantIncome 0
CoapplicantIncome 0
LoanAmount      22
Loan_Amount_Term 14
Credit_History  50
Property_Area    0
Loan_Status      0
dtype: int64
```

```
In [73]: import numpy as np
train['LoanAmount'].replace(np.NaN, train['LoanAmount'].mean())
train['Loan_Amount_Term'].replace(np.NaN, train['Loan_Amount_Term'].mean())
```

```
Out[73]: 0      360.0
1      360.0
2      360.0
3      360.0
4      360.0
...
609    360.0
610    360.0
611    360.0
612    360.0
613    360.0
Name: Loan_Amount_Term, Length: 614, dtype: float64
```

```
In [74]: train=train.dropna()
train.isnull().sum()
```

```
Out[74]: Loan_ID      0
Gender          0
Married         0
Dependents      0
Education        0
Self_Employed   0
ApplicantIncome 0
CoapplicantIncome 0
LoanAmount      0
Loan_Amount_Term 0
Credit_History  0
Property_Area    0
Loan_Status      0
dtype: int64
```

```
In [75]: dataset=train.values
pd.DataFrame(dataset)
```

```
Out[75]:      0      1      2      3      4      5      6      7      8      9     10     11     12
0      0      1      1      0      0      0     4583     1508     360.0      0.0      0.0
1      1      1      1      1      0      0      4583     1508.0     128.0     360.0
2      2      1      1      0      0      1      3000      0.0      66.0      360.0
3      3      1      1      0      1      0      2583     2358.0     120.0     360.0
4      4      1      0      0      0      0      6000      0.0      141.0     360.0
...
524    609.0      0.0      0.0      0.0      0.0      0.0      710     360.0      1.0      0.0
525    610.0      1.0      1.0      0.0      0.0      0.0      400     360.0      1.0      0.0
526    611.0      1.0      1.0      0.0      0.0      0.0      807.0     240.0     360.0      1.0
527    612.0      1.0      2.0      0.0      0.0      0.0      758.0      0.0     187.0     360.0
528    613.0      0.0      0.0      0.0      0.0      0.0     1330     360.0      1.0      0.0
```

529 rows x 13 columns

## Scaling Features

```
In [76]: from sklearn import preprocessing
min_max_scaler = preprocessing.MinMaxScaler()
X_scale = min_max_scaler.fit_transform(X)
pd.DataFrame(X_scale)
```

```
Out[76]:      0      1      2      3      4      5      6      7      8      9     10     11     12     13
0      0.070489  0.000000  0.198860  0.435990  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
1      0.054830  0.036192  0.172214  0.743590  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0
2      0.035520  0.000000  0.062489  0.743590  1.0  1.0  0.0  0.0  0.0  1.0  1.0  0.0  1.0
3      0.030390  0.056592  0.160637  0.743590  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  1.0
4      0.072356  0.000000  0.191027  0.743590  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  1.0
...
609    0.034014  0.000000  0.089725  0.743590  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
610    0.048930  0.000000  0.044863  0.358974  1.0  1.0  1.0  0.0  1.0  0.0  0.0  0.0  0.0
611    0.097984  0.005760  0.033111  0.743590  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  1.0
612    0.091936  0.000000  0.257598  0.743590  1.0  1.0  1.0  0.0  1.0  0.0  0.0  0.0  1.0
613    0.054830  0.000000  0.179450  0.743590  1.0  1.0  0.0  0.0  0.0  0.0  1.0  1.0  0.0
```

614 rows x 14 columns

## Splitting Our Data

```
In [77]: from sklearn.model_selection import train_test_split
X_train, X_val_and_test, Y_train, Y_val_and_test = train_test_split(X_scale, y, test_size=0.3)
```

```
In [78]: pd.DataFrame(X_train)
```

```
Out[78]:      0      1      2      3      4      5      6      7      8      9     10     11     12     13
0      0.253766  0.000000  0.198860  0.743590  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0
1      0.041707  0.000000  0.104197  0.358974  1.0  1.0  1.0  0.0  0.0  0.0  1.0  0.0  0.0
2      0.096255  0.000000  0.384949  0.743590  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  1.0
3      0.064232  0.034824  0.225760  0.743590  0.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  1.0
4      0.068237  0.006407  0.247467  0.743590  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0
...
424    0.037304  0.049536  0.137482  0.743590  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  1.0
425    0.023055  0.046296  0.094067  0.743590  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  1.0
426    0.020656  0.042456  0.124457  0.743590  1.0  1.0  1.0  0.0  0.0  0.0  1.0  0.0  0.0
427    0.056710  0.000000  0.186686  0.743590  1.0  1.0  1.0  0.0  0.0  0.0  1.0  1.0  0.0
428    0.046592  0.000000  0.166425  0.743590  1.0  1.0  1.0  0.0  1.0  0.0  1.0  1.0  0.0
```

429 rows x 14 columns

```
In [79]: X_val, X_test, y_val, Y_test = train_test_split(X_val_and_test, Y_val_and_test, test_size=0.5)
```

```
In [80]: print(X_train.shape, X_test.shape, x_val.shape, Y_train.shape, Y_test.shape, y_val.shape)
```

```
In [81]: (429, 14) (93, 14) (92, 14) (429,) (93,) (92,)
```

```
In [82]: #the word dense is used to define a layer of connected neurons here we use 3 layers and input_shape refers to #inputs we are using
```

```
model = Sequential([
    Dense(32, activation='relu', input_shape=(14,)),
    Dense(32, activation='relu'),
    Dense(1, activation='sigmoid'),
])
```

```
In [82]: model.compile(optimizer='sgd',
                    loss='binary_crossentropy',
                    metrics=['accuracy'])
```

```
In [83]: hist=model.fit(X_train,Y_train,
                    batch_size = 32, epochs = 100,
                    validation_data = (x_val,y_val))
```

```
Train on 429 samples, validate on 92 samples
Epoch 1/100
429/429 [=====] - 0s 579us/step - loss: 0.7406 - accuracy: 0.3263 - val_loss: 0.7219 - val_accuracy: 0.3696
Epoch 2/100
429/429 [=====] - 0s 78us/step - loss: 0.6994 - accuracy: 0.4755 - val_loss: 0.6837 - val_accuracy: 0.5109
Epoch 3/100
429/429 [=====] - 0s 76us/step - loss: 0.6691 - accuracy: 0.6760 - val_loss: 0.6581 - val_accuracy: 0.7174
Epoch 4/100
429/429 [=====] - 0s 82us/step - loss: 0.6493 - accuracy: 0.6830 - val_loss: 0.6404 - val_accuracy: 0.6739
Epoch 5/100
429/429 [=====] - 0s 77us/step - loss: 0.6354 - accuracy: 0.6807 - val_loss: 0.6272 - val_accuracy: 0.6848
Epoch 6/100
429/429 [=====] - 0s 79us/step - loss: 0.6251 - accuracy: 0.6783 - val_loss: 0.6180 - val_accuracy: 0.6739
Epoch 7/100
429/429 [=====] - 0s 79us/step - loss: 0.6181 - accuracy: 0.6783 - val_loss: 0.6133 - val_accuracy: 0.6739
Epoch 8/100
429/429 [=====] - 0s 78us/step - loss: 0.6124 - accuracy: 0.6783 - val_loss: 0.6053 - val_accuracy: 0.6739
Epoch 9/100
429/429 [=====] - 0s 86us/step - loss: 0.6078 - accuracy: 0.6783 - val_loss: 0.6010 - val_accuracy: 0.6739
Epoch 10/100
429/429 [=====] - 0s 84us/step - loss: 0.6042 - accuracy: 0.6783 - val_loss: 0.5979 - val_accuracy: 0.6739
Epoch 11/100
429/429 [=====] - 0s 83us/step - loss: 0.6014 - accuracy: 0.6783 - val_loss: 0.5951 - val_accuracy: 0.6739
Epoch 12/100
429/429 [=====] - 0s 84us/step - loss: 0.5988 - accuracy: 0.6783 - val_loss: 0.5929 - val_accuracy: 0.6739
Epoch 13/100
429/429 [=====] - 0s 83us/step - loss: 0.5969 - accuracy: 0.6783 - val_loss: 0.5912 - val_accuracy: 0.6739
Epoch 14/100
429/429 [=====] - 0s 84us/step - loss: 0.5950 - accuracy: 0.6783 - val_loss: 0.5896 - val_accuracy: 0.6739
Epoch 15/100
429/429 [=====] - 0s 84us/step - loss: 0.5932 - accuracy: 0.6783 - val_loss: 0.5877 - val_accuracy: 0.6739
Epoch 16/100
429/429 [=====] - 0s 84us/step - loss: 0.5916 - accuracy: 0.6783 - val_loss: 0.5861 - val_accuracy: 0.6739
Epoch 17/100
429/429 [=====] - 0s 79us/step - loss: 0.5896 - accuracy: 0.6783 - val_loss: 0.5847 - val_accuracy: 0.6739
Epoch 18/100
429/429 [=====] - 0s 77us/step - loss: 0.5881 - accuracy: 0.6783 - val_loss: 0.5832 - val_accuracy: 0.6739
Epoch 19/100
429/429 [=====] - 0s 77us/step - loss: 0.5867 - accuracy: 0.6783 - val_loss: 0.5815 - val_accuracy: 0.6739
Epoch 20/100
429/429 [=====] - 0s 81us/step - loss: 0.5852 - accuracy: 0.6783 - val_loss: 0.5802 - val_accuracy: 0.6739
Epoch 21/100
429/429 [=====] - 0s 82us/step - loss: 0.5836 - accuracy: 0.6783 - val_loss: 0.5787 - val_accuracy: 0.6739
Epoch 22/100
429/429 [=====] - 0s 77us/step - loss: 0.5822 - accuracy: 0.6783 - val_loss: 0.5772 - val_accuracy: 0.6739
Epoch 23/100
429/429 [=====] - 0s 79us/step - loss: 0.5806 - accuracy: 0.6783 - val_loss: 0.5759 - val_accuracy: 0.6739
Epoch 24/100
429/429 [=====] - 0s 74us/step - loss: 0.5791 - accuracy: 0.6783 - val_loss: 0.5743 - val_accuracy: 0.6739
Epoch 25/100
429/429 [=====] - 0s 81us/step - loss: 0.5774 - accuracy: 0.6783 - val_loss: 0.5728 - val_accuracy: 0.6739
Epoch 26/100
429/429 [=====] - 0s 77us/step - loss: 0.5761 - accuracy: 0.6783 - val_loss: 0.5714 - val_accuracy: 0.6739
Epoch 27/100
429/429 [=====] - 0s 79us/step - loss: 0.5743 - accuracy: 0.6783 - val_loss: 0.5700 - val_accuracy: 0.6739
Epoch 28/100
429/429 [=====] - 0s 81us/step - loss: 0.5727 - accuracy: 0.6783 - val_loss: 0.5685 - val_accuracy: 0.6739
Epoch 29/100
429/429 [=====] - 0s 75us/step - loss: 0.5712 - accuracy: 0.6807 - val_loss: 0.5671 - val_accuracy: 0.6739
Epoch 30/100
429/429 [=====] - 0s 79us/step - loss: 0.5696 - accuracy: 0.6807 - val_loss: 0.5658 - val_accuracy: 0.6739
Epoch 31/100
429/429 [=====] - 0s 84us/step - loss: 0.5679 - accuracy: 0.6807 - val_loss: 0.5643 - val_accuracy: 0.6739
Epoch 32/100
429/429 [=====] - 0s 85us/step - loss: 0.5664 - accuracy: 0.6807 - val_loss: 0.5628 - val_accuracy: 0.6739
Epoch 33/100
429/429 [=====] - 0s 84us/step - loss: 0.5648 - accuracy: 0.6876 - val_loss: 0.5612 - val_accuracy: 0.6739
Epoch 34/100
429/429 [=====] - 0s 82us/step - loss: 0.5631 - accuracy: 0.6876 - val_loss: 0.5598 - val_accuracy: 0.6739
Epoch 35/100
429/429 [=====] - 0s 72us/step - loss: 0.5613 - accuracy: 0.6900 - val_loss: 0.5582 - val_accuracy: 0.6957
Epoch 36/100
429/429 [=====] - 0s 79us/step - loss: 0.5598 - accuracy: 0.6970 - val_loss: 0.5567 - val_accuracy: 0.6957
Epoch 37/100
429/429 [=====] - 0s 83us/step - loss: 0.5584 - accuracy: 0.6993 - val_loss: 0.5553 - val_accuracy: 0.7065
Epoch 38/100
429/429 [=====] - 0s 84us/step - loss: 0.5570 - accuracy: 0.7040 - val_loss: 0.5539 - val_accuracy: 0.7065
Epoch 39/100
429/429 [=====] - 0s 81us/step - loss: 0.5551 - accuracy: 0.7110 - val_loss: 0.5523 - val_accuracy: 0.7065
Epoch 40/100
429/429 [=====] - 0s 86us/step - loss: 0.5538 - accuracy: 0.7179 - val_loss: 0.5509 - val_accuracy: 0.7065
Epoch 41/100
429/429 [=====] - 0s 84us/step - loss: 0.5523 - accuracy: 0.7273 - val_loss: 0.5492 - val_accuracy: 0.7065
Epoch 42/100
429/429 [=====] - 0s 85us/step - loss: 0.5507 - accuracy: 0.7226 - val_loss: 0.5477 - val_accuracy: 0.7065
Epoch 43/100
429/429 [=====] - 0s 81us/step - loss: 0.5489 - accuracy: 0.7273 - val_loss: 0.5462 - val_accuracy: 0.7065
Epoch 44/100
429/429 [=====] - 0s 87us/step - loss: 0.5474 - accuracy: 0.7296 - val_loss: 0.5446 - val_accuracy: 0.7065
Epoch 45/100
429/429 [=====] - 0s 85us/step - loss: 0.5459 - accuracy: 0.7319 - val_loss: 0.5429 - val_accuracy: 0.7065
Epoch 46/100
429/429 [=====] - 0s 84us/step - loss: 0.5443 - accuracy: 0.7413 - val_loss: 0.5414 - val_accuracy: 0.7065
Epoch 47/100
429/429 [=====] - 0s 81us/step - loss: 0.5428 - accuracy: 0.7483 - val_loss: 0.5400 - val_accuracy: 0.7065
Epoch 48/100
429/429 [=====] - 0s 81us/step - loss: 0.5411 - accuracy: 0.7529 - val_loss: 0.5386 - val_accuracy: 0.7174
Epoch 49/100
429/429 [=====] - 0s 86us/step - loss: 0.5397 - accuracy: 0.7552 - val_loss: 0.5370 - val_accuracy: 0.7174
Epoch 50/100
429/429 [=====] - 0s 77us/step - loss: 0.5381 - accuracy: 0.7576 - val_loss: 0.5353 - val_accuracy: 0.7174
Epoch 51/100
429/429 [=====] - 0s 72us/step - loss: 0.5365 - accuracy: 0.7552 - val_loss: 0.5341 - val_accuracy: 0.7283
Epoch 52/100
429/429 [=====] - 0s 82us/step - loss: 0.5348 - accuracy: 0.7576 - val_loss: 0.5323 - val_accuracy: 0.7391
Epoch 53/100
429/429 [=====] - 0s 81us/step - loss: 0.5333 - accuracy: 0.7576 - val_loss: 0.5307 - val_accuracy: 0.7500
Epoch 54/100
429/429 [=====] - 0s 81us/step - loss: 0.5319 - accuracy: 0.7622 - val_loss: 0.5291 - val_accuracy: 0.7500
Epoch 55/100
429/429 [=====] - 0s 81us/step - loss: 0.5301 - accuracy: 0.7622 - val_loss: 0.5281 - val_accuracy: 0.7609
Epoch 56/100
429/429 [=====] - 0s 84us/step - loss: 0.5286 - accuracy: 0.7646 - val_loss: 0.5263 - val_accuracy: 0.7609
Epoch 57/100
429/429 [=====] - 0s 86us/step - loss: 0.5274 - accuracy: 0.7646 - val_loss: 0.5248 - val_accuracy: 0.7609
Epoch 58/100
429/429 [=====] - 0s 81us/step - loss: 0.5258 - accuracy: 0.7669 - val_loss: 0.5232 - val_accuracy: 0.7609
Epoch 59/100
429/429 [=====] - 0s 81us/step - loss: 0.5240 - accuracy: 0.7669 - val_loss: 0.5219 - val_accuracy: 0.7826
Epoch 60/100
429/429 [=====] - 0s 103us/step - loss: 0.5228 - accuracy: 0.7786 - val_loss: 0.5203 - val_accuracy: 0.8043
Epoch 61/100
429/429 [=====] - 0s 135us/step - loss: 0.5213 - accuracy: 0.7809 - val_loss: 0.518 - val_accuracy: 0.8043
Epoch 62/100
429/429 [=====] - 0s 137us/step - loss: 0.5187 - accuracy: 0.7809 - val_loss: 0.516 - val_accuracy: 0.8152
Epoch 63/100
429/429 [=====] - 0s 135us/step - loss: 0.5172 - accuracy: 0.7832 - val_loss: 0.514 - val_accuracy: 0.8152
Epoch 64/100
429/429 [=====] - 0s 142us/step - loss: 0.5156 - accuracy: 0.7892 - val_loss: 0.513 - val_accuracy: 0.8152
Epoch 65/100
429/429 [=====] - 0s 135us/step - loss: 0.5145 - accuracy: 0.7879 - val_loss: 0.512 - val_accuracy: 0.8261
Epoch 66/100
429/429 [=====] - 0s 137us/step - loss: 0.5133 - accuracy: 0.7879 - val_loss: 0.511 - val_accuracy: 0.8261
Epoch 67/100
429/429 [=====] - 0s 135us/step - loss: 0.5120 - accuracy: 0.7995 - val_loss: 0.509 - val_accuracy: 0.8261
Epoch 68/100
429/429 [=====] - 0s 128us/step - loss: 0.5107 - accuracy: 0.7995 - val_loss: 0.508 - val_accuracy: 0.8370
Epoch 69/100
429/429 [=====] - 0s 133us/step - loss: 0.5097 - accuracy: 0.7995 - val_loss: 0.507 - val_accuracy: 0.8370
Epoch 70/100
429/429 [=====] - 0s 140us/step - loss: 0.5084 - accuracy: 0.7972 - val_loss: 0.506 - val_accuracy: 0.8370
Epoch 71/100
429/429 [=====] - 0s 130us/step - loss: 0.5075 - accuracy: 0.7995 - val_loss: 0.504 - val_accuracy: 0.8370
Epoch 72/100
429/429 [=====] - 0s 133us/step - loss: 0.5063 - accuracy: 0.7995 - val_loss: 0.503 - val_accuracy: 0.8370
Epoch 73/100
429/429 [=====] - 0s 129us/step - loss: 0.5049 - accuracy: 0.7995 - val_loss: 0.501 - val_accuracy: 0.8370
Epoch 74/100
429/429 [=====] - 0s 133us/step - loss: 0.5039 - accuracy: 0.7995 - val_loss: 0.500 - val_accuracy: 0.8370
Epoch 75/100
429/429 [=====] - 0s 130us/step - loss: 0.5029 - accuracy: 0.7995 - val
```