```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

```
data = pd.read_csv("/content/train.csv")
data.head()
```

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome
0	LP001002	Male	No	0	Graduate	No	5849	0.0
1	LP001003	Male	Yes	1	Graduate	No	4583	1508.0
2	LP001005	Male	Yes	0	Graduate	Yes	3000	0.0
3	LP001006	Male	Yes	0	Not Graduate	No	2583	2358.0
4	LP001008	Male	No	0	Graduate	No	6000	0.0

```
data.tail()
```

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncom
609	LP002978	Female	No	0	Graduate	No	2900	0
610	LP002979	Male	Yes	3+	Graduate	No	4106	0
611	LP002983	Male	Yes	1	Graduate	No	8072	240
612	LP002984	Male	Yes	2	Graduate	No	7583	0
613	LP002990	Female	No	0	Graduate	Yes	4583	0

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 614 entries, 0 to 613
Data columns (total 13 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Loan_ID            614 non-null    object
 1   Gender             601 non-null    object
 2   Married            611 non-null    object
 3   Dependents         599 non-null    object
 4   Education          614 non-null    object
 5   Self_Employed      582 non-null    object
 6   ApplicantIncome    614 non-null    int64
 7   CoapplicantIncome  614 non-null    float64
 8   LoanAmount         592 non-null    float64
 9   Loan_Amount_Term   600 non-null    float64
 10  Credit_History     564 non-null    float64
 11  Property_Area      614 non-null    object
 12  Loan_Status        614 non-null    object
dtypes: float64(4), int64(1), object(8)
```

```
memory usage: 62.5+ KB
```

```
data.apply(lambda x: sum(x.isnull()),axis=0)
```

	0
Loan_ID	0
Gender	13
Married	3
Dependents	15
Education	0
Self_Employed	32
ApplicantIncome	0
CoapplicantIncome	0
LoanAmount	22
Loan_Amount_Term	14
Credit_History	50
Property_Area	0
Loan_Status	0

```
data['Gender'].value_counts()
```

Gender	count
Male	489
Female	112

```
data.Gender = data.Gender.fillna('Male')
```

```
data['Married'].value_counts()
```

Married	count
Yes	398
No	213

```
data.Married = data.Married.fillna('NO')
```

```
data['Dependents'].value_counts()
```

Dependents	count
0	345
1	102
2	101
3+	51

```
data.replace('3+', 3,inplace=True,limit=None)
data.replace(0, 1,inplace=True,limit=None)
```

```
data['Dependents'] = data['Dependents'].astype(float)
```

```
data.loc[:,'Dependents'].fillna(data['Dependents'].mean() )
```

	Dependents
0	0.0
1	1.0
2	0.0
3	0.0
4	0.0
...	...
609	0.0
610	3.0
611	1.0
612	2.0
613	0.0

614 rows × 1 columns

```
data['Self_Employed'].value_counts()
```

Self_Employed	count
No	500
Yes	82

```
data.Self_Employed = data.Self_Employed.fillna('No')
```


```
data.Self_Employed = data.Self_Employed.fillna('No')
```


```
data['Loan_Amount_Term'].value_counts()
```

Loan_Amount_Term	count
360.0	512
180.0	44
480.0	15
300.0	13
240.0	4
84.0	4
120.0	3
60.0	2
36.0	2
12.0	1

```
data.Loan_Amount_Term = data.Loan_Amount_Term.fillna(360.0)
```

```
data.Credit_History = data.Credit_History.fillna(1.0)
```

```
data.apply(lambda x: sum(x.isnull()),axis=0)
```

	0
Loan_ID	0
Gender	0
Married	0
Dependents	0
Education	0
Self_Employed	0
ApplicantIncome	0
CoapplicantIncome	0
LoanAmount	22
Loan_Amount_Term	0
Credit_History	0
Property_Area	0
Loan_Status	0

```
x = data[['Loan_ID','Gender','Married','Dependents','Education','Self_Employed','ApplicantIncome','Coapplic
y = data['Loan_Status']
```

```
from sklearn.model_selection import train_test_split
```

```
X_train, X_test, y_train, y_test = train_test_split(
    x, y, test_size=0.2, random_state=0)
```

```
X_train
```

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncom
90	LP001316	Male	Yes	0	Graduate	No	2958	2900
533	LP002729	Male	No	1	Graduate	No	11250	0
452	LP002448	Male	Yes	0	Graduate	No	3948	1733
355	LP002144	Female	No	NaN	Graduate	No	3813	0
266	LP001877	Male	Yes	2	Graduate	No	4708	1387
...	...	...	...	...	...	...	...	
277	LP001904	Male	Yes	0	Graduate	No	3103	1300
9	LP001020	Male	Yes	1	Graduate	No	12841	10968
359	LP002160	Male	Yes	3+	Graduate	No	5167	3167
192	LP001657	Male	Yes	0	Not Graduate	No	6033	0
559	LP002804	Female	Yes	0	Graduate	No	4180	2306

491 rows × 12 columns

```
X_test
```

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncom
454	LP002453	Male	No	0	Graduate	Yes	7085	0
52	LP001164	Female	No	0	Graduate	No	4230	0
536	LP002734	Male	Yes	0	Graduate	No	6133	3906
469	LP002505	Male	Yes	0	Graduate	No	4333	2451
55	LP001194	Male	Yes	2	Graduate	No	2708	1167
...	...	...	...	...	...	...	...	
337	LP002112	Male	Yes	2	Graduate	Yes	2500	4600
376	LP002219	Male	Yes	3+	Graduate	No	8750	4996
278	LP001907	Male	Yes	0	Graduate	No	14583	0
466	LP002500	Male	Yes	3+	Not Graduate	No	2947	1664
303	LP001977	Male	Yes	1	Graduate	No	1625	1803

123 rows × 12 columns