In [1]:
```python
import pandas as pd
import numpy as np
import matplotlib as plt
%matplotlib inline

df = pd.read_csv("C:\\Users\\Suresh Jamadagni\\downloads\\train.csv")
```

In [2]:
```python
print(df)
```

```
       Loan_ID  Gender Married Dependents     Education Self_Employed  \
0    LP001002    Male      No          0      Graduate            No
1    LP001003    Male     Yes          1      Graduate            No
2    LP001005    Male     Yes          0      Graduate           Yes
3    LP001006    Male     Yes          0  Not Graduate            No
4    LP001008    Male      No          0      Graduate            No
..        ...     ...     ...        ...           ...           ...
609  LP002978  Female      No          0      Graduate            No
610  LP002979    Male     Yes         3+      Graduate            No
611  LP002983    Male     Yes          1      Graduate            No
612  LP002984    Male     Yes          2      Graduate            No
613  LP002990  Female      No          0      Graduate           Yes

     ApplicantIncome  CoapplicantIncome  LoanAmount  Loan_Amount_Term  \
0               5849                0.0         NaN             360.0
1               4583             1508.0       128.0             360.0
2               3000                0.0        66.0             360.0
3               2583             2358.0       120.0             360.0
4               6000                0.0       141.0             360.0
..               ...                ...         ...               ...
609             2900                0.0        71.0             360.0
610             4106                0.0        40.0             180.0
611             8072              240.0       253.0             360.0
612             7583                0.0       187.0             360.0
613             4583                0.0       133.0             360.0

     Credit_History Property_Area Loan_Status
0               1.0         Urban           Y
1               1.0         Rural           N
2               1.0         Urban           Y
3               1.0         Urban           Y
4               1.0         Urban           Y
..              ...           ...         ...
609             1.0         Rural           Y
610             1.0         Rural           Y
611             1.0         Urban           Y
612             1.0         Urban           Y
613             0.0     Semiurban           N

[614 rows x 13 columns]
```

In [3]:
```python
df.head()  #displays first 5 rows
```

Out[3]:

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	Coapplican
0	LP001002	Male	No	0	Graduate	No	5849	
1	LP001003	Male	Yes	1	Graduate	No	4583	
2	LP001005	Male	Yes	0	Graduate	Yes	3000	

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	Coapplican
3	LP001006	Male	Yes	0	Not Graduate	No	2583	
4	LP001008	Male	No	0	Graduate	No	6000	

In [3]:
```python
df.head()
to_drop = ['Dependents']
#df.drop(to_drop)
print(df)
```

```
       Loan_ID Gender Married Dependents     Education Self_Employed  \
0    LP001002   Male      No          0      Graduate            No
1    LP001003   Male     Yes          1      Graduate            No
2    LP001005   Male     Yes          0      Graduate           Yes
3    LP001006   Male     Yes          0  Not Graduate            No
4    LP001008   Male      No          0      Graduate            No
..        ...    ...     ...        ...           ...           ...
609  LP002978 Female      No          0      Graduate            No
610  LP002979   Male     Yes         3+      Graduate            No
611  LP002983   Male     Yes          1      Graduate            No
612  LP002984   Male     Yes          2      Graduate            No
613  LP002990 Female      No          0      Graduate           Yes

     ApplicantIncome  CoapplicantIncome  LoanAmount  Loan_Amount_Term  \
0               5849                0.0         NaN             360.0
1               4583             1508.0       128.0             360.0
2               3000                0.0        66.0             360.0
3               2583             2358.0       120.0             360.0
4               6000                0.0       141.0             360.0
..               ...                ...         ...               ...
609             2900                0.0        71.0             360.0
610             4106                0.0        40.0             180.0
611             8072              240.0       253.0             360.0
612             7583                0.0       187.0             360.0
613             4583                0.0       133.0             360.0

     Credit_History Property_Area Loan_Status
0               1.0         Urban           Y
1               1.0         Rural           N
2               1.0         Urban           Y
3               1.0         Urban           Y
4               1.0         Urban           Y
..              ...           ...         ...
609             1.0         Rural           Y
610             1.0         Rural           Y
611             1.0         Urban           Y
612             1.0         Urban           Y
613             0.0     Semiurban           N

[614 rows x 13 columns]
```

In [4]:
```python
df.tail()    #displays last 5 rows
```

Out[4]:

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	Coapplic
609	LP002978	Female	No	0	Graduate	No	2900	
610	LP002979	Male	Yes	3+	Graduate	No	4106	

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	Coapplic
611	LP002983	Male	Yes	1	Graduate	No	8072	
612	LP002984	Male	Yes	2	Graduate	No	7583	
613	LP002990	Female	No	0	Graduate	Yes	4583	

In [5]:
```python
print(df.tail(10))
```

```
       Loan_ID  Gender Married Dependents      Education Self_Employed  \
604  LP002959  Female     Yes          1       Graduate            No
605  LP002960    Male     Yes          0   Not Graduate            No
606  LP002961    Male     Yes          1       Graduate            No
607  LP002964    Male     Yes          2   Not Graduate            No
608  LP002974    Male     Yes          0       Graduate            No
609  LP002978  Female      No          0       Graduate            No
610  LP002979    Male     Yes         3+       Graduate            No
611  LP002983    Male     Yes          1       Graduate            No
612  LP002984    Male     Yes          2       Graduate            No
613  LP002990  Female      No          0       Graduate           Yes

     ApplicantIncome  CoapplicantIncome  LoanAmount  Loan_Amount_Term  \
604            12000                0.0       496.0             360.0
605             2400             3800.0         NaN             180.0
606             3400             2500.0       173.0             360.0
607             3987             1411.0       157.0             360.0
608             3232             1950.0       108.0             360.0
609             2900                0.0        71.0             360.0
610             4106                0.0        40.0             180.0
611             8072              240.0       253.0             360.0
612             7583                0.0       187.0             360.0
613             4583                0.0       133.0             360.0

     Credit_History Property_Area Loan_Status
604             1.0     Semiurban           Y
605             1.0         Urban           N
606             1.0     Semiurban           Y
607             1.0         Rural           Y
608             1.0         Rural           Y
609             1.0         Rural           Y
610             1.0         Rural           Y
611             1.0         Urban           Y
612             1.0         Urban           Y
613             0.0     Semiurban           N
```

In [6]:
```python
df.describe() #Provides summary measures for variables of type : int or float
```

Out[6]:

	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History
count	614.000000	614.000000	592.000000	600.00000	564.000000
mean	5403.459283	1621.245798	146.412162	342.00000	0.842199
std	6109.041673	2926.248369	85.587325	65.12041	0.364878
min	150.000000	0.000000	9.000000	12.00000	0.000000
25%	2877.500000	0.000000	100.000000	360.00000	1.000000
50%	3812.500000	1188.500000	128.000000	360.00000	1.000000

	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History
75%	5795.000000	2297.250000	168.000000	360.00000	1.000000
max	81000.000000	41667.000000	700.000000	480.00000	1.000000

In [11]:
```python
print(df['Property_Area'].value_counts())     #for categorial data(works like group_
```

```
Semiurban     233
Urban         202
Rural         179
Name: Property_Area, dtype: int64
```

In [12]:
```python
print(df['Gender'].value_counts())     #for categorial data(works like group_by)
```

```
Male      489
Female    112
Name: Gender, dtype: int64
```

In [13]:
```python
df.dtypes   #displays type of each variable
```

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

In [14]:
```python
df.columns   #displays variable names ( Column name) as a list
```

Out[14]:
```
Index(['Loan_ID', 'Gender', 'Married', 'Dependents', 'Education',
       'Self_Employed', 'ApplicantIncome', 'CoapplicantIncome', 'LoanAmount',
       'Loan_Amount_Term', 'Credit_History', 'Property_Area', 'Loan_Status'],
      dtype='object')
```

In [15]:
```python
df.columns[0]  #displays 1st column name
```

Out[15]:
```
'Loan_ID'
```

In [16]:
```python
print(df)
```

```
       Loan_ID  Gender Married Dependents     Education Self_Employed  \
0      LP001002    Male      No          0      Graduate            No
1      LP001003    Male     Yes          1      Graduate            No
2      LP001005    Male     Yes          0      Graduate           Yes
3      LP001006    Male     Yes          0  Not Graduate            No
4      LP001008    Male      No          0      Graduate            No
..          ...     ...     ...        ...           ...           ...
609    LP002978  Female      No          0      Graduate            No
610    LP002979    Male     Yes         3+      Graduate            No
611    LP002983    Male     Yes          1      Graduate            No
```

```
612  LP002984    Male       Yes           2        Graduate              No
613  LP002990  Female        No           0        Graduate             Yes

     ApplicantIncome  CoapplicantIncome  LoanAmount  Loan_Amount_Term  \
0               5849               0.0         NaN             360.0
1               4583            1508.0       128.0             360.0
2               3000               0.0        66.0             360.0
3               2583            2358.0       120.0             360.0
4               6000               0.0       141.0             360.0
..               ...               ...         ...               ...
609             2900               0.0        71.0             360.0
610             4106               0.0        40.0             180.0
611             8072             240.0       253.0             360.0
612             7583               0.0       187.0             360.0
613             4583               0.0       133.0             360.0

     Credit_History Property_Area Loan_Status
0               1.0         Urban           Y
1               1.0         Rural           N
2               1.0         Urban           Y
3               1.0         Urban           Y
4               1.0         Urban           Y
..               ...           ...         ...
609             1.0         Rural           Y
610             1.0         Rural           Y
611             1.0         Urban           Y
612             1.0         Urban           Y
613             0.0     Semiurban           N

[614 rows x 13 columns]
```

In [17]:
```python
print (df.replace('NaN',0))
```

```
       Loan_ID  Gender Married Dependents     Education Self_Employed  \
0    LP001002    Male      No          0      Graduate            No
1    LP001003    Male     Yes          1      Graduate            No
2    LP001005    Male     Yes          0      Graduate           Yes
3    LP001006    Male     Yes          0  Not Graduate            No
4    LP001008    Male      No          0      Graduate            No
..        ...     ...     ...        ...           ...           ...
609  LP002978  Female      No          0      Graduate            No
610  LP002979    Male     Yes         3+      Graduate            No
611  LP002983    Male     Yes          1      Graduate            No
612  LP002984    Male     Yes          2      Graduate            No
613  LP002990  Female      No          0      Graduate           Yes

     ApplicantIncome  CoapplicantIncome  LoanAmount  Loan_Amount_Term  \
0               5849               0.0         NaN             360.0
1               4583            1508.0       128.0             360.0
2               3000               0.0        66.0             360.0
3               2583            2358.0       120.0             360.0
4               6000               0.0       141.0             360.0
..               ...               ...         ...               ...
609             2900               0.0        71.0             360.0
610             4106               0.0        40.0             180.0
611             8072             240.0       253.0             360.0
612             7583               0.0       187.0             360.0
613             4583               0.0       133.0             360.0

     Credit_History Property_Area Loan_Status
0               1.0         Urban           Y
1               1.0         Rural           N
2               1.0         Urban           Y
3               1.0         Urban           Y
```

```
4                1.0         Urban         Y
..               ...          ...        ...
609              1.0         Rural         Y
610              1.0         Rural         Y
611              1.0         Urban         Y
612              1.0         Urban         Y
613              0.0     Semiurban         N

[614 rows x 13 columns]
```

In [19]:
```python
print ("NaN replaced with '0':")
print( df.fillna(method='pad'))
```

```
NaN replaced with '0':
       Loan_ID  Gender Married Dependents    Education Self_Employed  \
0    LP001002    Male      No          0     Graduate            No
1    LP001003    Male     Yes          1     Graduate            No
2    LP001005    Male     Yes          0     Graduate           Yes
3    LP001006    Male     Yes          0 Not Graduate            No
4    LP001008    Male      No          0     Graduate            No
..        ...     ...     ...        ...          ...           ...
609  LP002978  Female      No          0     Graduate            No
610  LP002979    Male     Yes         3+     Graduate            No
611  LP002983    Male     Yes          1     Graduate            No
612  LP002984    Male     Yes          2     Graduate            No
613  LP002990  Female      No          0     Graduate           Yes

     ApplicantIncome  CoapplicantIncome  LoanAmount  Loan_Amount_Term  \
0               5849                0.0         NaN             360.0
1               4583             1508.0       128.0             360.0
2               3000                0.0        66.0             360.0
3               2583             2358.0       120.0             360.0
4               6000                0.0       141.0             360.0
..               ...                ...         ...               ...
609             2900                0.0        71.0             360.0
610             4106                0.0        40.0             180.0
611             8072              240.0       253.0             360.0
612             7583                0.0       187.0             360.0
613             4583                0.0       133.0             360.0

     Credit_History Property_Area Loan_Status
0               1.0         Urban           Y
1               1.0         Rural           N
2               1.0         Urban           Y
3               1.0         Urban           Y
4               1.0         Urban           Y
..              ...           ...         ...
609             1.0         Rural           Y
610             1.0         Rural           Y
611             1.0         Urban           Y
612             1.0         Urban           Y
613             0.0     Semiurban           N

[614 rows x 13 columns]
```

In [20]:
```python
print (df['Loan_ID'].isnull())
```

```
0      False
1      False
2      False
3      False
4      False
       ...
609    False
```

```
610    False
611    False
612    False
613    False
Name: Loan_ID, Length: 614, dtype: bool
```

In [21]: `print (df['Dependents'].notnull())`

```
0      True
1      True
2      True
3      True
4      True
      ...
609    True
610    True
611    True
612    True
613    True
Name: Dependents, Length: 614, dtype: bool
```

In [31]: `print( df['Self_Employed'].isnull())`

```
1      False
2      False
3      False
4      False
5      False
6      False
7      False
8      False
9      False
10     False
12     False
13     False
14     False
15     False
17     False
18     False
20     False
21     False
22     False
25     False
26     False
27     False
28     False
31     False
32     False
33     False
34     False
37     False
38     False
39     False
      ...
575    False
577    False
578    False
580    False
581    False
582    False
584    False
585    False
586    False
587    False
```

```
589    False
590    False
591    False
593    False
594    False
595    False
596    False
598    False
599    False
602    False
603    False
604    False
606    False
607    False
608    False
609    False
610    False
611    False
612    False
613    False
Name: Self_Employed, dtype: bool
```

In [8]:
```python
print(df)
print ("NaN replaced with '0':")
print( df.fillna(3))
```

```
        Loan_ID  Gender Married Dependents     Education Self_Employed  \
0     LP001002    Male      No          0      Graduate            No
1     LP001003    Male     Yes          1      Graduate            No
2     LP001005    Male     Yes          0      Graduate           Yes
3     LP001006    Male     Yes          0  Not Graduate            No
4     LP001008    Male      No          0      Graduate            No
..         ...     ...     ...        ...           ...           ...
609   LP002978  Female      No          0      Graduate            No
610   LP002979    Male     Yes         3+      Graduate            No
611   LP002983    Male     Yes          1      Graduate            No
612   LP002984    Male     Yes          2      Graduate            No
613   LP002990  Female      No          0      Graduate           Yes

     ApplicantIncome  CoapplicantIncome  LoanAmount  Loan_Amount_Term  \
0               5849                0.0         NaN             360.0
1               4583             1508.0       128.0             360.0
2               3000                0.0        66.0             360.0
3               2583             2358.0       120.0             360.0
4               6000                0.0       141.0             360.0
..               ...                ...         ...               ...
609             2900                0.0        71.0             360.0
610             4106                0.0        40.0             180.0
611             8072              240.0       253.0             360.0
612             7583                0.0       187.0             360.0
613             4583                0.0       133.0             360.0

     Credit_History Property_Area Loan_Status
0               1.0         Urban           Y
1               1.0         Rural           N
2               1.0         Urban           Y
3               1.0         Urban           Y
4               1.0         Urban           Y
..              ...           ...         ...
609             1.0         Rural           Y
610             1.0         Rural           Y
611             1.0         Urban           Y
612             1.0         Urban           Y
613             0.0     Semiurban           N
```

```
[614 rows x 13 columns]
NaN replaced with '0':
      Loan_ID  Gender Married Dependents     Education Self_Employed  \
0    LP001002    Male      No          0      Graduate            No
1    LP001003    Male     Yes          1      Graduate            No
2    LP001005    Male     Yes          0      Graduate           Yes
3    LP001006    Male     Yes          0  Not Graduate            No
4    LP001008    Male      No          0      Graduate            No
..        ...     ...     ...        ...           ...           ...
609  LP002978  Female      No          0      Graduate            No
610  LP002979    Male     Yes         3+      Graduate            No
611  LP002983    Male     Yes          1      Graduate            No
612  LP002984    Male     Yes          2      Graduate            No
613  LP002990  Female      No          0      Graduate           Yes

     ApplicantIncome  CoapplicantIncome  LoanAmount  Loan_Amount_Term  \
0               5849                0.0         3.0             360.0
1               4583             1508.0       128.0             360.0
2               3000                0.0        66.0             360.0
3               2583             2358.0       120.0             360.0
4               6000                0.0       141.0             360.0
..               ...                ...         ...               ...
609             2900                0.0        71.0             360.0
610             4106                0.0        40.0             180.0
611             8072              240.0       253.0             360.0
612             7583                0.0       187.0             360.0
613             4583                0.0       133.0             360.0

     Credit_History Property_Area Loan_Status
0               1.0         Urban           Y
1               1.0         Rural           N
2               1.0         Urban           Y
3               1.0         Urban           Y
4               1.0         Urban           Y
..              ...           ...         ...
609             1.0         Rural           Y
610             1.0         Rural           Y
611             1.0         Urban           Y
612             1.0         Urban           Y
613             0.0     Semiurban           N

[614 rows x 13 columns]
```

In [9]:
```python
df = df.dropna() #drops rows with null values
```

In [24]:
```python
print(df)
```

```
      Loan_ID  Gender Married Dependents     Education Self_Employed  \
1    LP001003    Male     Yes          1      Graduate            No
2    LP001005    Male     Yes          0      Graduate           Yes
3    LP001006    Male     Yes          0  Not Graduate            No
4    LP001008    Male      No          0      Graduate            No
5    LP001011    Male     Yes          2      Graduate           Yes
..        ...     ...     ...        ...           ...           ...
609  LP002978  Female      No          0      Graduate            No
610  LP002979    Male     Yes         3+      Graduate            No
611  LP002983    Male     Yes          1      Graduate            No
612  LP002984    Male     Yes          2      Graduate            No
613  LP002990  Female      No          0      Graduate           Yes

     ApplicantIncome  CoapplicantIncome  LoanAmount  Loan_Amount_Term  \
1               4583             1508.0       128.0             360.0
```

```
2          3000              0.0        66.0       360.0
3          2583           2358.0       120.0       360.0
4          6000              0.0       141.0       360.0
5          5417           4196.0       267.0       360.0
..          ...             ...         ...         ...
609        2900              0.0        71.0       360.0
610        4106              0.0        40.0       180.0
611        8072            240.0       253.0       360.0
612        7583              0.0       187.0       360.0
613        4583              0.0       133.0       360.0

     Credit_History Property_Area Loan_Status
1              1.0          Rural           N
2              1.0          Urban           Y
3              1.0          Urban           Y
4              1.0          Urban           Y
5              1.0          Urban           Y
..             ...           ...          ...
609            1.0          Rural           Y
610            1.0          Rural           Y
611            1.0          Urban           Y
612            1.0          Urban           Y
613            0.0      Semiurban           N

[480 rows x 13 columns]
```

In [13]:
```python
import pandas as pd
import numpy as np
df = pd.read_csv("C:\\Users\\Suresh Jamadagni\\downloads\\train.csv")
df.head()
```

Out[13]:

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

In [14]:
```python
to_drop = ['Gender','Married']
#df.drop(columns=to_drop, inplace=True)
df.drop(to_drop, inplace=True, axis=1)
```

In [15]:
```python
df.head()
```

Out[15]:

	Loan_ID	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmc
0	LP001002	0	Graduate	No	5849	0.0	
1	LP001003	1	Graduate	No	4583	1508.0	1
2	LP001005	0	Graduate	Yes	3000	0.0	
3	LP001006	0	Not Graduate	No	2583	2358.0	1

	Loan_ID	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmc
4	LP001008	0	Graduate	No	6000	0.0	1

In [ ]:

In [ ]: