

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib as plt
%matplotlib inline
import statistics as st

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

In [2]:

```
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]:

```
# Summary Measures of a DataFrame Column using Statistics

print('Mean ', st.mean(df['ApplicantIncome']))    #prints mean of only ApplicantInco
print('Median ', st.median(df['ApplicantIncome'])) #prints median of only Applica
print('Mode ', st.mode(df['ApplicantIncome']))    #prints mode of only ApplicantInco
print('Standard Deviation ', st.pstdev(df['ApplicantIncome'])) #prints population
print('Standard Deviation ', st.stdev(df['ApplicantIncome']))   #prints sample stan
```

Mean 5403.459283387622

Median 3812.5

```
Mode    2500
Standard Deviation    6104.064856533892
Standard Deviation    6109.041673387178
```

```
In [4]: df.describe() #Provides summary measures for variables of type : int or float
```

```
Out[4]:
```

	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
75%	5795.000000	2297.250000	168.000000	360.00000	1.000000
max	81000.000000	41667.000000	700.000000	480.00000	1.000000

```
In [5]: print(df['Property_Area'].value_counts()) #for categorial data(works like group_
```

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

```
In [ ]:
```