

Question 2: The below output shown is any part is from the python code **Assignment1_Question2.py**

Part A.

→ The Dataset contains a total of 5960 data.

```
*****Part A*****  
Total Observations:  
5960
```

→ The Frequency Distribution of 'BAD' Variable is:

```
Frequency:  
BAD  
0    4771  
1    1189  
Name: count, dtype: int64
```

→ The Mean and Standard Deviation of given variables is:

```
Mean and Standard Deviation of Data:  
  
Mean of DEBTINC : 33.779915348721126  
Mean of LOAN : 18607.96979865772  
Mean of MORTDUE : 73760.817199559  
Mean of VALUE : 101776.04874145007  
Standard Deviation of DEBTINC : 8.60174618632853  
Standard Deviation of LOAN : 11207.480416694003  
Standard Deviation of MORTDUE : 44457.60945841593  
Standard Deviation of VALUE : 57385.775333702615
```

PART B.

Using the simple random sampling method with a **random seed of 202303484**, we have split the data into two parts:

- The Training Partitioning includes **4172 Observations**.
- The Testing Partitioning includes **1788 Observations**.
- The Frequency Distribution of is as follows:
- For Training Partition:

```
Training Data Observation : 4172
Train list: BAD
0      3299
1       873
Name: count, dtype: int64
```

- For Testing Partition:

```
Testing Data Observation : 1788
Testing list: BAD
0      1472
1       316
Name: count, dtype: int64
```

- Mean and Standard Deviation of Training Partition includes:

```
Mean and Standard Deviation of Training partition using Simple random sampling:

Mean of DEBTINC : 33.85349806369245
Mean of LOAN : 18527.780441035473
Mean of MORTDUE : 73310.90224372385
Mean of VALUE : 100902.55424871732
Standard Deviation of DEBTINC : 8.92385881516458
Standard Deviation of LOAN : 11256.829038909691
Standard Deviation of MORTDUE : 44387.2289408149
Standard Deviation of VALUE : 56222.46006281549
```

- Mean and Standard Deviation of Testing Partition includes:

```
Mean and Standard Deviation of Testing using Simple random sampling:

Mean of DEBTINC : 33.613021962466505
Mean of LOAN : 18795.078299776287
Mean of MORTDUE : 74824.15143386897
Mean of VALUE : 103813.20712250713
Standard Deviation of DEBTINC : 7.822563160391285
Standard Deviation of LOAN : 11092.355254741491
Standard Deviation of MORTDUE : 44619.22033363395
Standard Deviation of VALUE : 59978.33014853821
```

PART C.

When using the stratified random sampling method with a random seed of 202303484, we have categorized the data based on 'BAD' and 'REASON' variables. To account for missing values, we have replaced them as follows:

- In 'BAD,' missing values have been substituted with 99.
- In 'REASON,' missing values have been replaced with 'MISSING.'
- The Frequency Distribution of 'Bad' in Training Partition is:

```
Training Data Observation : 4172
Testing list: BAD
0      3340
1       832
Name: count, dtype: int64
```

- The Frequency Distribution of 'BAD' in Testing Partition is:

```
Testing Data Observation : 1788
Train list: BAD
0      1431
1       357
Name: count, dtype: int64
```

- Mean and Standard Deviation of Training Partition includes:

```
Mean and Standard Deviation of Training partition using Simple random sampling:

Mean of DEBTINC : 33.82549013654666
Mean of LOAN : 18651.462128475552
Mean of MORTDUE : 74259.33061550549
Mean of VALUE : 101882.90065934067
Standard Deviation of DEBTINC : 8.79877005948205
Standard Deviation of LOAN : 11421.73183291651
Standard Deviation of MORTDUE : 44893.07922262891
Standard Deviation of VALUE : 56912.12958653953
```

- Mean and Standard Deviation of Testing Partition includes:

```
Mean and Standard Deviation of Testing partition using Simple random sampling:

Mean of DEBTINC : 33.672169362879785
Mean of LOAN : 18506.48769574944
Mean of MORTDUE : 72588.81952586207
Mean of VALUE : 101526.44314888763
Standard Deviation of DEBTINC : 8.118999087220887
Standard Deviation of LOAN : 10693.305091448932
Standard Deviation of MORTDUE : 43407.7146176993
Standard Deviation of VALUE : 58492.924614461386
```