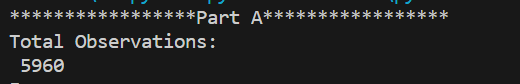
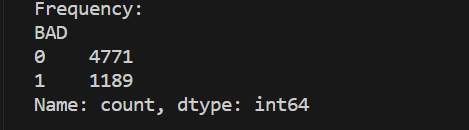
**Question 2:** The below output shown is any part is from the python code **Assignment1\_Question2.py**

**Part A**.

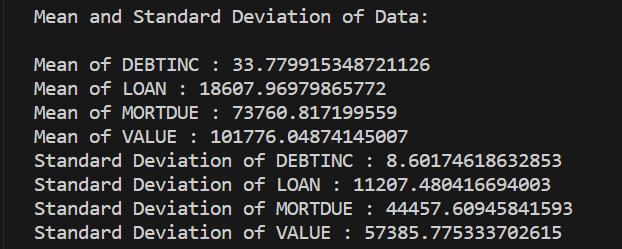
* The Dataset contains a total of 5960 data.



* The Frequency Distribution of ‘BAD’ Variable is:



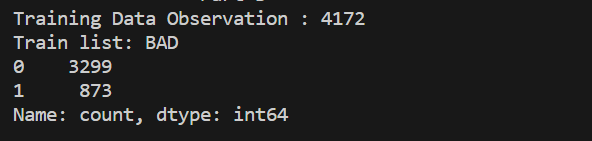
* The Mean and Standard Deviation of given variables is:



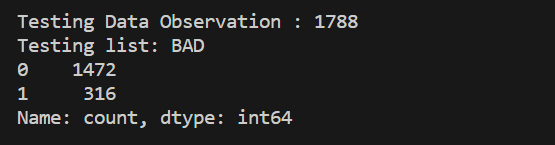
**PART B.**

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

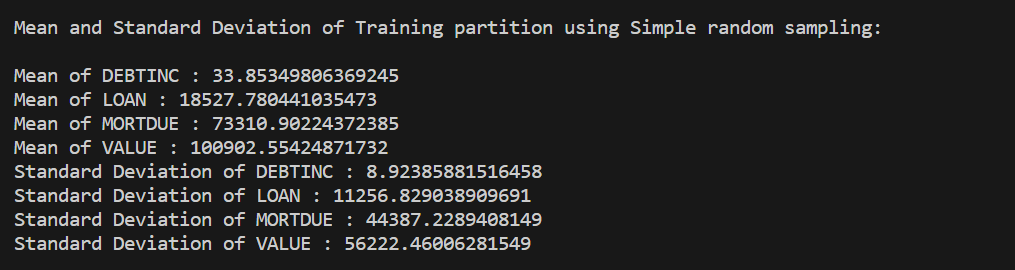
* The Training Partitioning includes **4172** **Observations**.
* The Testing Partitioning includes **1788 Observations.**
* The Frequency Distribution of is as follows:
* For Training Partition:



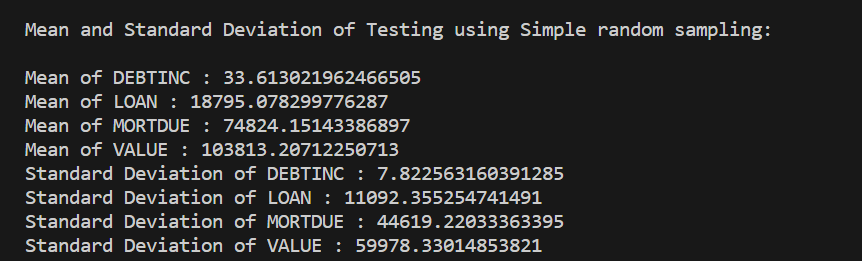
* For Testing Partition:



* Mean and Standard Deviation of Training Partition includes:



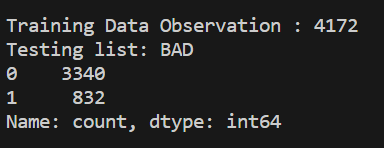
* Mean and Standard Deviation of Testing Partition includes:



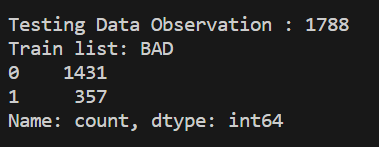
**PART C.**

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:

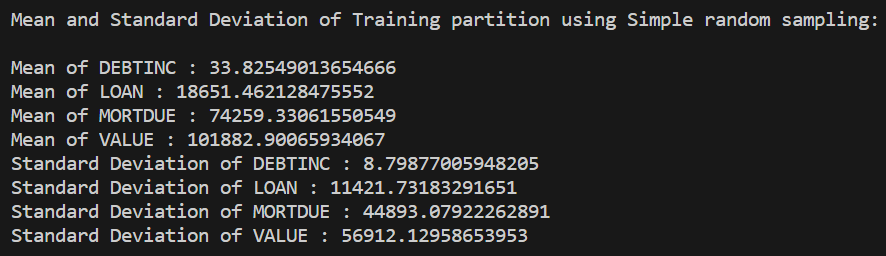
* In 'BAD,' missing values have been substituted with 99.
* In 'REASON,' missing values have been replaced with 'MISSING.
* The Frequency Distribution of ‘Bad’ in Training Partition is:



* The Frequency Distribution of ‘BAD’ in Testing Partition is:

****

* Mean and Standard Deviation of Training Partition includes:



* Mean and Standard Deviation of Testing Partition includes:

