**Applied Machine Learning**

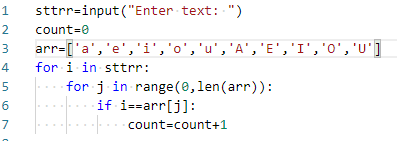
**Assignment 1**

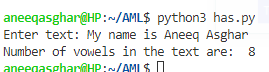
**Name:** Aneeq Asghar

**Roll number:** 17L-4525

**Section:** EE-7A

**Question 1:**

****



**CODE:**

sttrr=input("Enter text: ")

count=0

arr=['a','e','i','o','u','A','E','I','O','U']

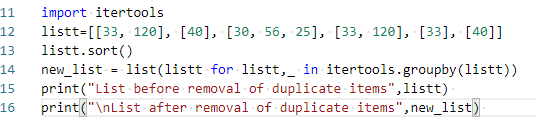
for i in sttrr:

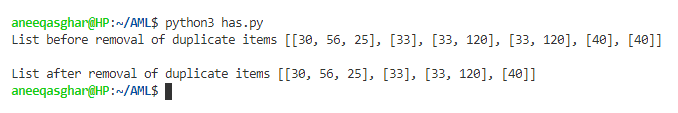
for j in range(0,len(arr)):

if i==arr[j]:

count=count+1

**Question 2:**

****

****

**Code:**

**import itertools**

**listt=[[33, 120], [40], [30, 56, 25], [33, 120], [33], [40]]**

**listt.sort()**

**new\_list = list(listt for listt,\_ in itertools.groupby(listt))**

**print("List before removal of duplicate items",listt)**

**print("\nList after removal of duplicate items",new\_list)**

**Question 3:**

**Code:**

**tuple=[(), (), ('',), ('a', 'b'), ('a', 'b', 'c'), ('d')]**

**i=0**

**x=len(tuple)**

**print("\nBefore empty tuple is removed from a list of tuples:\n",tuple)**

**while i< len(tuple):**

**#print( "Length of tuple:",len(tuple[i]),"tuple is",tuple[i])**

**if (len(tuple[i])==0):**

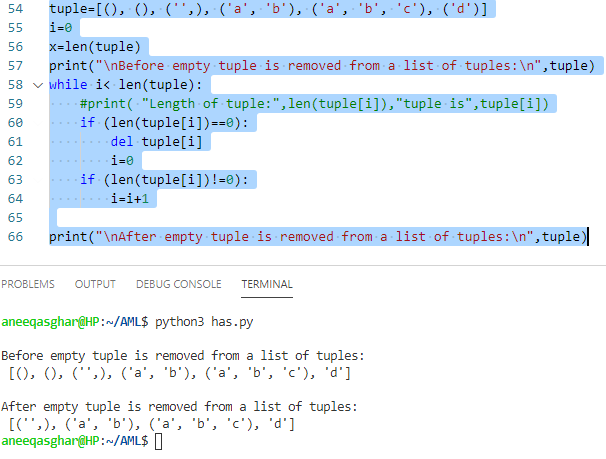
**del tuple[i]**

**i=0**

**if (len(tuple[i])!=0):**

**i=i+1**

**print("\nAfter empty tuple is removed from a list of tuples:\n",tuple)**

****

**Question 4:**

**dictonary1 = {'a': 100, 'b': 200, 'c':300}**

**dictionary2 = {'a': 300, 'b': 200, 'd':400}**

**array=['a','b','c','d']**

**Counter={'a':'','b':'','c':'','d':''}**

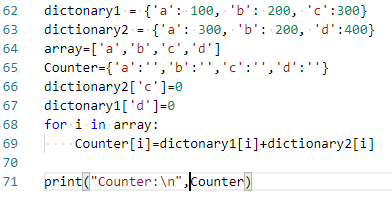
**dictionary2['c']=0**

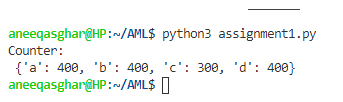
**dictonary1['d']=0**

**for i in array:**

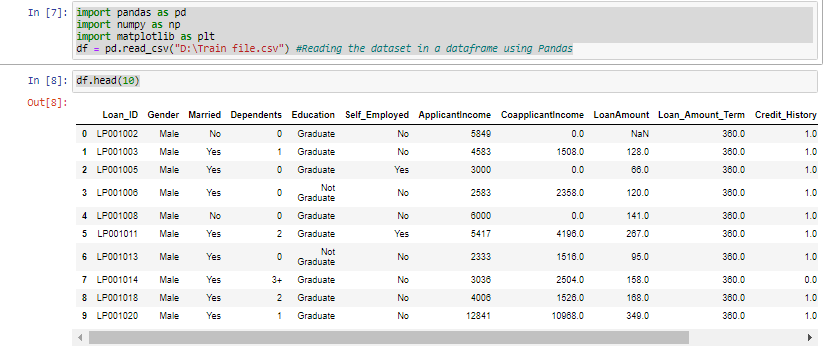
**Counter[i]=dictonary1[i]+dictionary2[i]**

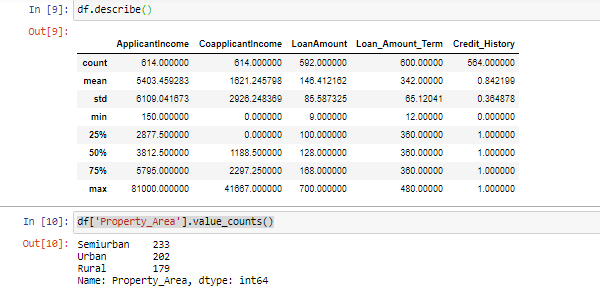
**print("Counter:\n",Counter)**

****

****

**Question 5:**





Missing values of the Loan Amount are replaced by:

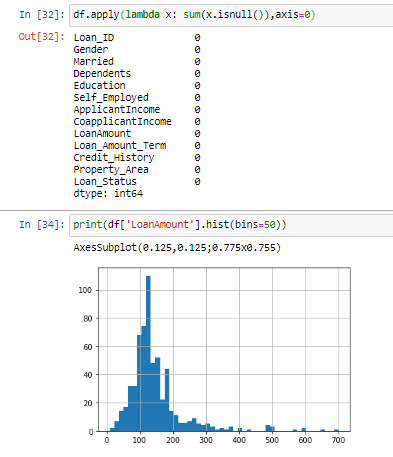
* Median of loan amount
* Standard Deviation of loan amount
* Maximum value of loan amount

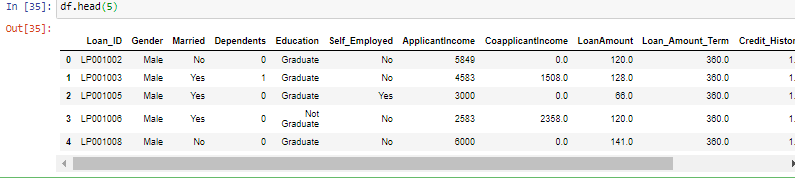
There would be six models in total:

1. Filling the missing value of Loan with median of the Loan Amount column without Self Employed
2. Filling the missing value of Loan with median of the Loan Amount column with Self Employed
3. Filling the missing value of Loan with standard deviation of the Loan Amount column without Self Employed
4. Filling the missing value of Loan with standard deviation of the Loan Amount column with Self Employed
5. Filling the missing value of Loan with maximum value of the Loan Amount column without Self Employed
6. Filling the missing value of Loan with maximum value of the Loan Amount column with Self Employed

**MODEL 1**:

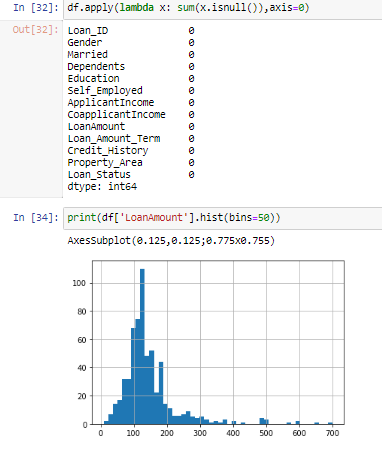
For missing values of loan are replaced by MEDIAN OF LOAN AMOUNT and missing values of Self Employed is replaced by ‘No’:

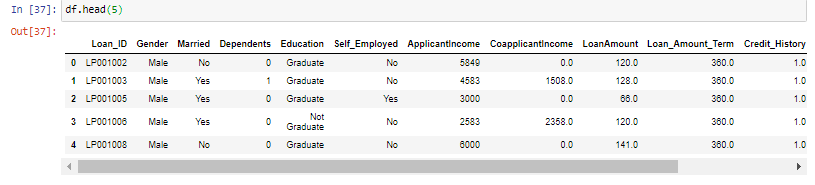




**MODEL 2:**

For missing values of loan are replaced by MEDIAN OF LOAN AMOUNT and missing values of Self Employed is replaced by ‘Yes’:

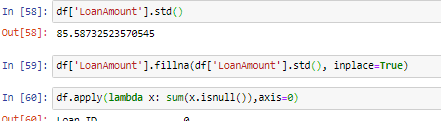


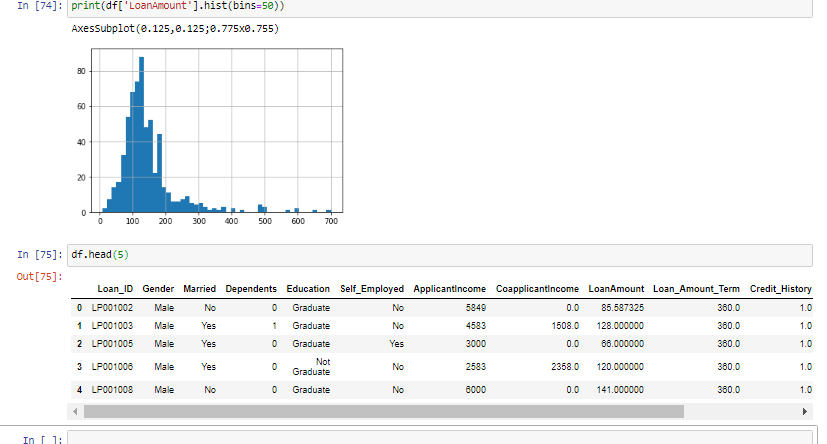


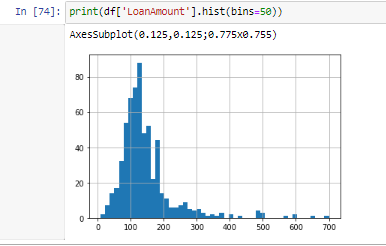
Model 1 and 2 are appropriate enough as it has all the replaced data in the center and hence, it is practical.

**MODEL 3:**

For missing values of loan are replaced by Standard Deviation of LOAN AMOUNT and missing values of Self Employed is replaced by ‘Yes’:

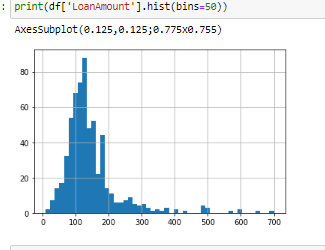


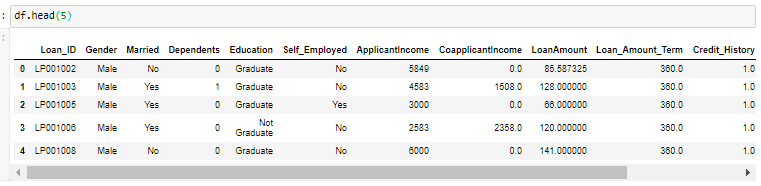




**MODEL 4:**

For missing values of loan are replaced by Standard Deviation of LOAN AMOUNT and missing values of Self Employed is replaced by ‘No’:



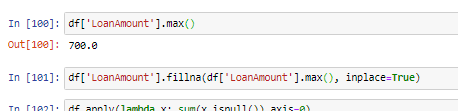


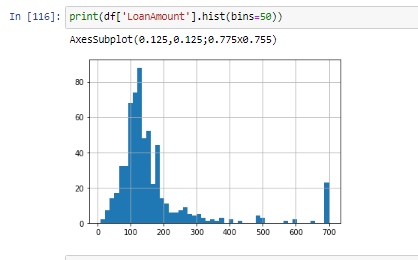


Model 3 and 4 are appropriate enough as it does not has data skewed at the ends. All the missing values are replaced by standard deviation in the middle, and enough values and hence, it is practical.

**MODEL 5:**

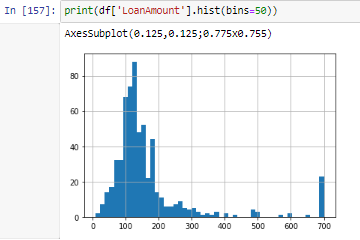
For missing values of loan are replaced by Maximum of LOAN AMOUNT and missing values of Self Employed is replaced by ‘No’:

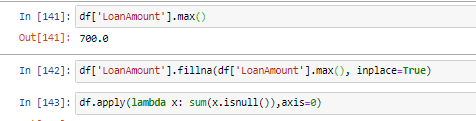




**MODEL 6:**

For missing values of loan are replaced by Maximum of LOAN AMOUNT and missing values of Self Employed is replaced by ‘Yes’:



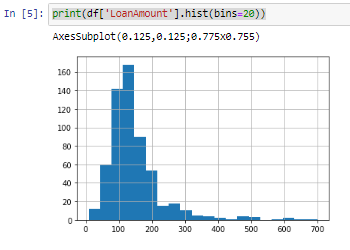


Model 5 and 6 are totally inappropriate as it has data skewed at one end. All the missing values are replaced by maximum value and has extreme values and hence, it is not practical.

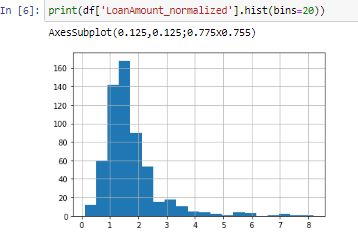
The standard deviation model is more practical as it does not has extreme values at each end and it has majority of the values in the middle.

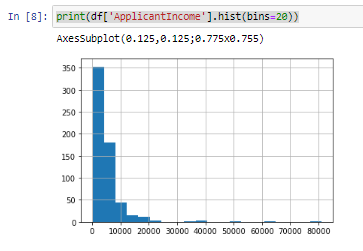
All the coding files are attached with this report.

**Question 6:**

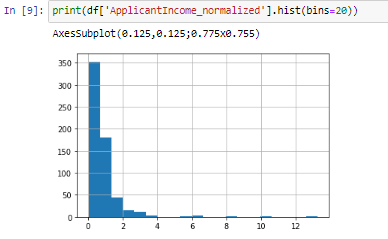












I will impute the missing values for Gender columns by using the most occurring value of the column Gender which was Male as it is a multi-class problem applied in question 5.

I will impute the missing values for Married columns by using the most occurring value of the column Married which was Yes as it is a multi-class problem applied in question 5.

I will impute the missing values for Dependents columns by using the most occurring value of the column Dependents which was 0 as it is a multi-class problem applied in question 5.

I will impute the missing values for Credit\_History columns by using the most occurring value of the column Credit\_History which was ‘N’ as it is a multi-class problem applied in question 5.

I will impute the missing values for Loan\_Amount\_Term column by using the mean of Loan\_Amount\_Term values of the column Loan\_Amount\_Term as it is a continuous problem applied in question 5.