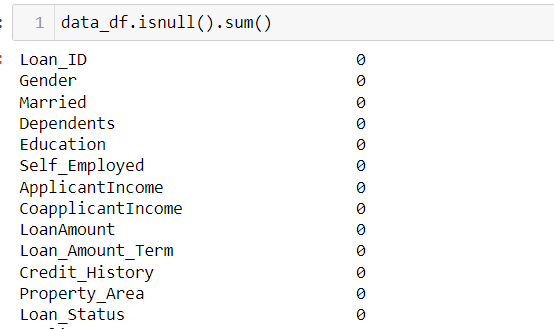
# Loan Eligibility Prediction: Consumer Creditworthiness Analysis

MileStone-3

# Missing value treatment:

* The Missing values for the categorical features are replaced with the Mode.
* The Missing values for the numeric features are replaced with the Median.

Output:



# Outlier Detection and Treatment

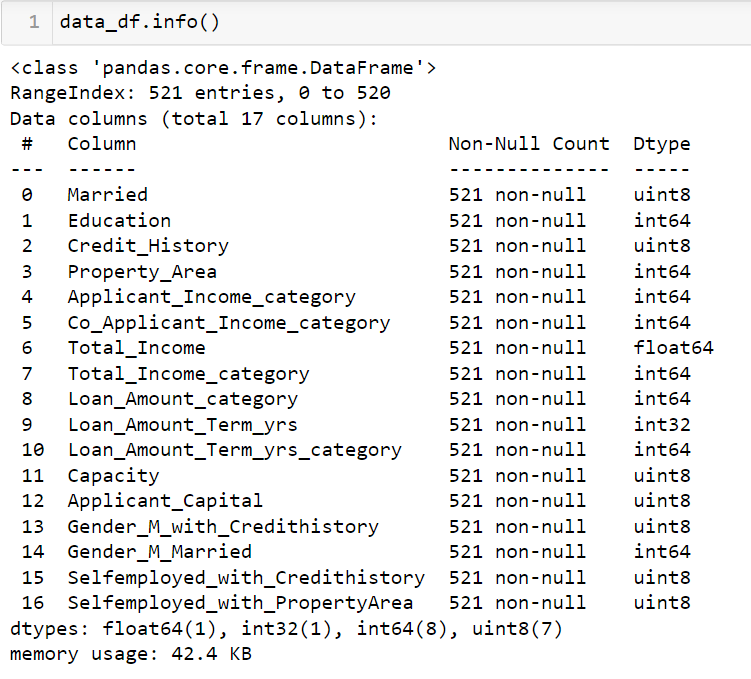
* The columns Applicant Income, CoapplicantIncome, Loan Amount were having outliers.
* The Outliers are detected using the following method.
  1. Boxplot method (Visualization)
  2. IQR method
* The Outliers are not treated as there could be possibility of high income and Loan categories.

# Encoding

* The Categorical features are encoded by Label Encoding and One hot encoding.
  1. Nominal features encoded using Label encoding
  2. Ordinal Features encoded using One Hot encoding

Output:





# Scaling:

* The Continuous features which are present in the dataset does not affect the Target variable.
* They will not be used further.
* Scaling would not be required.

# Feature Engineering:

|  |  |  |
| --- | --- | --- |
| S. No | Feature | Description |
| 1 | Total Income | Applicant Income +Co Applicant Income |
| 2 | Loan Amount term in Years | Loan Amount term in months / Years |
| 3 | Applicant Income Category | Applicants Income divided into bins (Low, Avg, High, very high) |
| 4 | Co Applicant Income Category | Co Applicants Income divided into bins (Low, Avg, High, very high) |
| 5 | Total Income Category | Total Income divided into bins |
| 6 | Loan Amount Category | Loan Amount divided into bins |
| 7 | Loan Amount Term Years Category | Loan Amount Term divided into bins |
| 8 | EMI | 1. Assuming the rate of interest = 10%  2. To calculate the rate of interest per month = (10/100)/12 = 0.0083  3. EMI formula : EMI = [P x R x (1+R) ^ N] / [(1+R) ^ (N-1)]  4. N = loan amount term in months  5. P = Principal amount |
| 9 | Monthly\_Balance\_Income | Monthly Income-EMI |
| 10 | DTI | Debt to Income Ratio = Month EMI / Monthly gross Income |
| 11 | Capacity | ((Total Income/Loan Amount)>1) and (credit history =1) |
| 12 | Capital | Property in Semiurban and Urban and Credit History = 1 |
| 13 | Gender\_M\_with\_Credithistory | Gender = M and CreditHistory = 1 |
| 14 | Gender\_F\_with\_Credithistory | Gender = F and CreditHistory = 1 |
| 15 | Gender\_M\_Married | Gender = M and Married = Yes |
| 16 | Gender\_F\_Married | Gender = F and Married = Yes |
| 17 | Selfemployed\_with\_Credithistory | Selfemployed = No and Credithistory = 1 |
| 18 | Selfemployed\_with\_PropertyArea | Selfemployed = No and PropertyArea = “Urban”, “SemiUrban” |
| 19 | Dependents Range | 0-1, 1-2, 2-3 |
| 19 | Total\_Inc\_Edu | Total Income = High and Very High and Education = Graduate |

For All the above features the Hypothesis testing is done. The results are as shown below.

