|  |  |
| --- | --- |
| Name | Vamsi Krishna Chamarthi |
| Assignment | Lab Assignment 01 |

**Task-1**

Here is the detailed explanation of identified issues for each variable

* int\_tgt: 21 values are negative.
* rfm1 to rfm12: There are no observed issues with this variable.
* demog\_age : 3 values are negative and there are 2 values greater than 120.
* demog\_homeval: 1 value is negative
* demog\_inc:There are no observed issues with this variable.
* demog\_pr: There are no observed issues with this variable.
* demog\_genf:There are 3 invalid entries.
* demog\_genm:There are 4 ivalid entries.
* demog\_ho: There are no observed issues with this variable.

**Task-2**

All the values in the dataset are modified as mentioned in which all the values which are negative are left as missing and values with yes are standardized as YES or NO.

**Task-3**

Original dataset was duplicated and renamed as bank\_chamarthi\_revised. Data cleaning involved handling missing values by inputting the mean for numerical columns and the mode for categorical columns. Out-of-range values were marked as missing.

**Task-4:**

The highest correlation with the target variable INT\_TGT was found with rfm2.A scatter plot was created to visualize this relationship, which showed a negative correlation.

**Task-5:**

Outliers were identified using the IQR method across all numerical variables. The number of outliers for each variable was reported. Box plots for INT\_TGT and rfm2 provided insights into the distribution and outliers in the dataset.