

Consider the dataset of loangrant.csv provided to you at GCR and do the following tasks.

1. By considering the data dictionary file, do exploratory data analysis on the data (i.e. Histogram, Scatter plots, Boxplots, Outlier detection, missing values identification etc).
2. Find the missing values of every attribute and replace it with the appropriate central tendency. However, the replacement with central should be justified.
3. Find the outliers using boxplot and by employing the outliers formula (quartiles and IQR) and report how many percentage of data having outliers. Moreover, you need to replace outliers with the appropriate central tendency (may be median).
4. Do perform correlation between the attributes and identify the highly correlated attributes i.e. 0.6 or higher.
5. The ‘Year in current job’ need to be replaced using the following highlighted format

Name of new level 1	Comma-separated list of old levels	Name of new level 2	Comma-separated list of old levels	Name of new level 3	Comma-separated list of old levels	Name of new level 4	Comma-separated list of old levels	Name of new level 5	Comma-separated list of old levels
<year	< 1 year, n/a	two - four years	1 year, 2 years, 3 years, 4 years	five - seven years	5 years, 7 years, 6 years	eight - nine years	8 years, 9 years	above 10 years	10+ years

6. Write a description of every line of code and use snapshot of graphs and its description along with its associated code snippet in a file and a single PDF file need to be upload at GCR.
7. Do not forget to upload Jupiter notebook file with the name of your Student ID.
8. Prepare a 10 minutes video presentation by switching on the camera at zoom (female students are exempted from switching on the camera). Share the screen and explain all the tasks you have done during the pre-processing of the dataset. The video along with other files need to be submitted at GCR on or before the due date/time.