**Data Analytics Coursework**

Steps taken to clean data.

Replace age text with numerical equivalent.

Replace ages less than minimum credit application age and maximum living age with the medium age value as an imputation.

# Check if the age is less than 18 or more than 122, and if so, replace it with 33, otherwise leave it as is.

if cells['Applicant Age'].value < 18 or cells['Applicant Age'].value > 122:

return 33

else:

return cells['Applicant Age'].value

Binned Savings values as none-observed, low, moderate, moderate-high, High

Binned employment term as N/A for unemployed, short, medium, medium-long and long

Open refine automatically Converted credit amount to integer thus removing the integer separating commas.

Age binning to get the nominal datatype.

# Jython script for OpenRefine to categorize age

def categorize\_age(age):

"""

Categorize age into respective bins: Young Adult, Adult, Senior.

Parameters:

age (int): The age to categorize.

Returns:

str: The age category.

"""

if age < 30:

return "Young Adult"

elif 30 <= age < 50:

return "Adult"

else:

return "Senior"

# Use the function to categorize the 'Applicant Age' column

categorized\_age = categorize\_age(value)