Case Study: Data Manipulation using Pandas

Import both the data sets "Cust_data" & "Cust_Demo" into python and solve the below questions using functions from "pandas" module.

1. Create sub set of **cust_demo** as "**cust_s1**" with **ID, age, Gender** and **Location** variables using the condition marital_status = "Married" and Own_house = 1 and age>28.
2. Sort the **cust_s1** dataset using location in ascending order and age in descending order with in the location.
3. Rename the variable "**NumberOfDependents**" as "**No_of_dependents**" in **cust_demo** data.
4. Remove duplicates from the both data sets Cust_data & Cust_Demo.
5. Create new variable as "No_of_30_plus_DPD = No_of_30_59_DPD+No_of_90_DPD+ No_of_60_89_DPD", "No_of_60_plus_DPD = No_of_90_DPD+ No_of_60_89_DPD" in **cust_data**.
6. Perform below joins between cust_data & cust_demo.
   a. Create data set "cust_leftjoin" using Left Join
   b. Create data set "cust_rightjoin" using right Join
   c. Create data set "cust_innterjoin" using inner Join
   d. Create data set "cust_fulljoin" using full Join
7. Create two data sets by taking random sample from cust_demo with 5% data in first data set and 10000 observations in second data set.
8. Calculate "number of customers" and "number of married customer" using combined data set comes from left join of cust_data and cust_demo
9. Create summary report with below column by combination of Gender, Serious_delinquency.
   a. Number of customers
   b. Percentage of customers
   c. Average Revolving utilization
   d. Standard deviation with in the utilization
   e. Average monthly income
   f. Maximum Monthly income
   g. Standard deviation with in the income
   h. Average Age
   i. Average Dependents
10. Create summary report with below information
    a. Number of married customers
    b. Percentage of married customers