Pandas Assignment - Data Cleaning & Insight

Dataset Link

1. Find out the dataset dimensions
2. Find out the statistical summary of the Dataset
3. Check for the null values in the Dataset
4. Check for the noise value of the dataset
5. Handle the null values in the Dataset for each column
6. change the `child birth weight (g)` into `child birth weight (kg)`

(1000grm = 1kg)

7. Replace the 0 with “No” and 1 with “Yes” in `gestational dm` and `diabetes mellitus` column
8. Find min and max age from column `age (years)`
9. Create a column “Age_Group” having 5 age groups like 1-20, 21-40, 41-60, and 60+ for each value of `age (years)`

   (Eg: if a person's age lies between 1 to 20 assign 1-20 and so on)
10. For each “Age_Group” count the number of persons present in the dataset.
11. For each “Age_Group” find the average `mean diastolic bp (mmhg)` and `mean systolic bp (mmhg)`
12. Try to explore the data set and find which parameters (columns) has the highest correlation with each other and explain why?