**Python Project for California Housing Data Set**

In this project, I have provided you with a California dataset to answer the following questions. Before starting the project, please explore the dataset online also. All these questions are mandatory. It is a ‘must’ to give an introduction of the project. You must put comments after codes to explain the step. Each result must be explained after getting the result and figures. If you miss any of these, marks will be deducted. These steps are a must to write a good project. You must explain your project and features in the introduction section. Please explain which feature is nominal, ordinal, discrete or continuous.

Please address the following questions:

1. What is the average median income of the data set and check the distribution of data using appropriate plots. Please explain the distribution of the plot.

2. Draw an appropriate plot to see the distribution of housing\_median\_age and explain your observations.

3. Show with the help of visualization, how median\_income and median\_house\_values are related?

4. Create a data set by deleting the corresponding examples from the data set for which total\_bedrooms are not available.

5. Create a data set by filling the missing data with the mean value of the total\_bedrooms in the original data set.

6. Write a programming construct (create a user defined function) to calculate the median value of the data set wherever required.

7. Plot latitude versus longitude and explain your observations.

8. Create a data set for which the ocean\_proximity is ‘Near ocean’.

9. Find the mean and median of the median income for the data set created in question 8.

10. Please create a new column named total\_bedroom\_size. If the total bedrooms is 10 or less, it should be quoted as small. If the total bedrooms is 11 or more but less than 1000, it should be medium, otherwise it should be considered large.