Answer the following questions in a .jpynb file, execute your code and keep the output, then submit the .jpynb file

1. Load the data from supplied data file. Print the data dimension.

2. Continue from question 1. Display the data type of all features. If the data type is integer, print the median values of the features.

3. Continue from question 2. Print all the possible values of the feature “status” and calculate the ratio of each “status” value.

4. Is there any association between status and clue\_small\_area? Explain your results from given dataset.

5. Print the number of properties for different suburbs. Please report the pattern found in the result, if any.

6. Continue from question 5, which suburb has the biggest number of properties which are under construction?

7. Continue from question 6, which suburb has the biggest number of student apartments?

8. Create and print a data frame of the number of different status values for different year groups (based on 5 years interval).

9. Continue from question 8. Draw a histogram of number of status values against different year groups. Explain the result.

10. Based on the original dataset, exclude the clue\_small\_area feature, use the rest available features and perform clustering on all the properties and determine the number of clusters. Is this the same as the number of suburbs in the data set?

11. Continue from question 10, perform K-Means and Hierarchical clustering on the data set, report the purity score.

12. Continue from question 11, try at least three different distance metrics for K-Means and Hierarchical clustering, select the best distance metric for each corresponding clustering algorithm, explain why the chosen distance metric is the best for the given data set.

13. Apart from K-Means and Hierarchical clustering, try another clustering method, and compare the results.