

You are given a dataset. Use it to answer the following questions:

housing.csv

The dataset is primarily centered around the housing market of London. It contains a lot of additional relevant data:

- Area
- Date
- Yearly average house prices
- Yearly number of houses
- Yearly percentage of households that recycle
- Yearly life satisfaction
- Yearly median salary of the residents of the area
- Yearly mean salary of the residents of the area
- Yearly number of jobs
- Yearly number of people living in the area
- Area size in hectares

- A. Import the dataset
- B. What is the length of the dataset?
- C. Remove the following columns from the dataset. Use inplace = True
 - a. Borough_flag
 - b. Unnamed
 - c. Code
- D. Display the first 10 rows of the dataset
- E. Display the datatype of each column in the dataset
- F. How many null values are in each column?
- G. Use a line graph to show the total number of houses in each year.
 - a. The line graph should be 10 by 8 in size
 - b. The line should be green in color
 - c. The date should be on the x-axis while the total number of houses should be on the y-axis
 - d. Label the x-axis and y-axis appropriately
 - e. Label the line graph appropriately
 - f. Rotate the dates in your graph by 70 degrees
 - g. Briefly interpret your graph
- H. Use a barplot to show the average number of houses in each area over the years.
 - a. It should be 15 by 8 in size
 - b. Use the color red
 - c. Rotate the names of the areas in your graph by 90 degrees
 - d. List the four areas with the highest number of houses
- I. Create a subset of the housing data containing only data from the area with the highest average number of houses in G above
 - a. Use a scatterplot to check the relationship between the number of jobs and the number of houses in that area
 - b. Explain your observation from your scatterplot