Data analysis with R and Python CA exercise

Consider the Sacramento real estate data file (Sacramentorealestatetransactions.csv) attached to this exercise. Create a python notebook with Jupyter to:

1. Display the structure and content of the file using the function head().
2. Display the descriptive statistics of the data contained in the file.
3. Clean the dataset (remove outliers and corrupted records).
4. Plot the scatter plot of:
    a) Number of bathrooms (baths) against price.
    b) Number of bedrooms (beds) against price.
    c) Surface of the house (sq_ft) against price.
5. Build a simple regression model explaining the price of the house from its surface.
6. Plot the scatter plot of the observations against predictions, and give the correlation coefficient between the two data series.
7. Build a multiple regression model explaining the price of the house from its number of bathrooms, its number of bedrooms, and its surface.
8. Plot the scatter plot of the observations against predictions, and give the correlation coefficient between the two data series.
9. Write a R script with Rstudio to repeat question 1 to 8.

NB: your python notebook and R script should be named using your full names separated by underscores (_)