Assignment 1

The "sec_buildings.xlsx" file contains the transaction information of the second-hand house. Please answer the following questions using python.

1. Import the "sec_buildings.xlsx" file as DataFrame in python and show the data information.
2. Extract the year from variable "built_date" and calculate the age of each house. Add the age as a new column in your DataFrame.
   (Hint: You can use Series.str method to extract the year.)
3. Use box plot or violin plot to check the data distribution status of the unit price of the houses in different regions and answer whether there exists outliers.
4. Calculate the mean, standard deviation and the skewness for the age and unit price of the houses in Xuhui and Minhang respectively. Describe what you find based on the above results.
5. Calculate the correlation coefficient between the age and the unit price of the houses and test whether these 2 variables have significant correlation.
6. Use scatter diagram to show the correlation between the age and the unit price of the houses in different regions. (Select no more than 5 regions.)