Python questions - regression

Create a Jupyter notebook, name it `python_questions.ipynb` to load the data present on file `regression_data.xls` for `house_price_regression`. Fill the notebook with the Python code needed to answer each question. This last file should be delivered.

1. Select all the data from table `house_price_data` to check if the data was imported correctly.

2. Drop the column `date` from the data frame, as we would not use it in the analysis. Select all the data from the table to verify if the command worked. Limit your returned results to 10.

3. How many rows of data do you have?

4. Find the unique values of the following columns:

   - What are the unique values in the column `bedrooms`?
   - What are the unique values in the column `bathrooms`?
   - What are the unique values in the column `floors`?
   - What are the unique values in the column `condition`?
   - What are the unique values in the column `grade`?

5. Arrange the data in decreasing order by the price of the house. Return only the IDs of the top 10 most expensive houses in your data.

6. What is the average price of all the properties in your data?

7. In this exercise use a simple `groupby` to check the properties of some of the categorical variables in our data

   - What is the average price of the houses grouped by bedrooms? The returned result should have only two columns: `Bedrooms` and `Average price`.
   - What is the average `sqft_living` of the houses grouped by bedrooms? The returned result should have only two columns, `Bedrooms` and `Average_sqft_living`.
   - What is the average price of the houses with a waterfront and without a waterfront? The returned result should have only two columns, `Waterfront` and `Average_price`.
   - Is there any correlation between the columns `condition` and `grade`? Also, create a plot to visually check if there is a positive correlation or negative correlation or no correlation between both variables.

- Get the number of houses in each category (ie number of houses for a given `condition`) to assess if that category is well represented in the dataset to include it in your analysis. For eg. If the category is under-represented as compared to other categories, ignore that category in this analysis

8. One of the customers is only interested in the following houses:

- Number of bedrooms either 3 or 4
- Bathrooms more than 3
- One Floor
- No waterfront
- Condition should be 3 at least
- Grade should be 5 at least
- Price smaller than 300000

For the rest of the things, they are not too concerned. Write code to find what are the houses available for them?

9. Your manager wants to find out the list of properties whose prices are twice more than the average of all the properties in the database. Write code to show them the list of such properties.

10. Most customers are interested in properties with three or four bedrooms. What is the difference in average prices of the properties with three and four bedrooms? In this case you can simply use a `groupby` to check the prices for those particular houses.

11. What are the different locations where properties are available in your database? (distinct zip codes).

12. Show all the properties that were renovated.

13. Provide the details of the property that is the 11th most expensive property in your database.