Individual Assignment: Statistical Analysis 3

This deliverable has 20% weightage in the Consolidated Total score.

Deliverables:

A report **(A pdf file)** & Excel or R files.

General Instructions:

1. This is an individual assignment. The maximum marks are 50.
2. Do NOT submit **.zip** files otherwise the submission will not be considered.
3. Any late submission will attract a penalty as mentioned in the course outline.
4. The honour code for this submission is **0N**.

Dataset Description

The dataset contains house sale data for Seattle, Washington. Following are the features/ variables available.

• id: Unique ID for each home sold

• date: Date of the home sale

• price: Price of each home sold

• bedrooms: Number of bedrooms

• bathrooms: Number of bathrooms, where .5 accounts for a room with a toilet but no shower

• sqft_living: Square footage of the apartments interior living space

• sqft_lot: Square footage of the land space • floors: Number of floors

• waterfront: - A dummy variable for whether the apartment was overlooking the waterfront or not

• view: An index from 0 to 4 of how good the view of the property was

• condition: - An index from 1 to 5 on the condition of the apartment,

• grade: An index from 1 to 13, where 1-3 falls short of building construction and design, 7 has an average level of construction and design, and 11-13 have a high quality level of construction and design.

• sqft_above: The square footage of the interior housing space that is above ground level

• sqft_basement: The square footage of the interior housing space that is below ground level

• yr_built: The year the house was initially built

• yr_renovated: The year of the house's last renovation

• zipcode: What zipcode area the house is in

• lat: Latitude

• long: Longitude

• sqft_living15: The square footage of interior housing living space for the nearest 15 neighbors

• sqft_lot15: The square footage of the land lots of the nearest 15 neighbours

Questions:

1. Identify the variables that can predict house prices in Seattle. Explain why you think those variables predict house prices.

2. Create a correlation matrix of the identified variables and comment on the correlations.

3. Use regression to predict house prices by looking at the impact of the identified variables on the house prices. Explain how the identified variable impact house prices.

4. Based on your regression analysis, what factors are more important in predicting house prices? Explain.

5. Are there any potential interaction effects? Check for interaction effects and comment.

For the Report:

- **All the questions should be explained in the report and the respective R/Excel files should be attached for reference.**
- Name the report as **FirstName_LastName_Rollno.pdf**.
- Name the excel/R file as **FirstName_LastName_Rollno**

Deadline: 14th February 2021, Sunday, 11:55 pm.