In this project, we are going to look at a number of houses sold in the year 2016 and 2017 in a fictional state by a well-known real estate agency. The agency have trained auditors who measure and map all the relevant features for the properties along with information related to the geography around it. The agency wants to understand the relevance of the parameters that they collect in relation to the price of the house. They have hired you to create a model which makes use of the available information to predict the monetary value of a house.

You are expected to use the data of the year 2016 to create a regression model where the price is the dependent variable. Identify the factors that are the driving factors for house prices. Using the model, you are expected to predict the selling prices of the houses sold in 2017.

**Deliverables:**

* Create an excel report that contains all the meaningful information such as relevant charts, pivot tables etc.
* Create a few hypothesis around the important variables and validate them using the data
* Mention all the variable which are highly correlated
* Build a linear regression model on the data of year 2016. Predict the price for year 2016 using this regression model, plot the regressed values against the actual values to understand the difference
* Using the above linear regression model, predict the prices of the houses sold in the year 2017. Interpret your findings from the model.

Descriptions and names of the columns (features) are given below.

Id: unique id

Date: Date house was sold

Price: Price of the sold house (Target Variable)

Bedrooms: Number of Bedrooms

Bathrooms: Number of bathrooms

Living area: Square footage of the living space

Lot area: Square footage of the lot

Number of floors: Total floors in the house

Waterfront: Whether the house is on a waterfront (1: yes, 0: no)

Number of views: number of special views

Condition: Condition of the house on a scale of 1-5 (1 being the lowest, 5 being the highest)

Grade of the house: Grade of the house based on Foundation, Drainage and Fire Prevention on a scale of 1-13 (1 being the lowest and 13 being the highest)

Area of the house: Square footage of house apart from basement

Area of Basement: Square footage of the basement

Built year: Built year

Renovation year: Year when the house was renovated

Postal code: Postal code of the house

Living\_area\_renov: Living room area currently (after renovations)

Lot\_area\_renov: Lot area currently (after renovations)

Number of Schools nearby: Number of schools in the vicinity of the house

Distance from the airport: Distance in KMs from nearest Airport