# Objective of the Assignment

Your assignment is to learn statistical data analysis and draw conclusions from results. To this end, you develop a regression model that predicts or explains selling price of a house by as many as five other variables. You will be collecting real estate listing information and then applying multiple regression to fit and validate the model you developed to your real estate data.

# Data collection (Note: Data collection should be only Pertaining to a city in British Columbia Province)

First, take a look at [www.mls.ca](http://www.mls.ca/) and decide upon a city that you would like to focus on (e.g. Vancouver). Go to [www.mls.ca](http://www.mls.ca/) and click onto “Residential Properties”. It should

be a city that has at least a population size of 200,000 people. Next, decide what area or suburb you will be focusing (e.g. Kitsilano). Do a search “by map” to locate a suburb in the city of your choice. Zoom into your suburb/area until there is between 250 and 500 property listings and the property listings are shown on the right side of the webpage. Refine your search by selecting (1) Building Type = House, and (2) Style = Detached. Write down the boundaries of your data collection (e.g., Kitsilano: West of Burrard St. and North of 29th). Save your search by clicking on to the “save search” button at the top left hand corner of the webpage. This will allow you to retrieve your search listings at a later time.

Randomly select (as best as possible) 60 listings from the total number of property listings. The 60 listings will constitute your sample. MLS search listings are displayed from the lowest to highest in prices on the right side of the webpage. Try to randomly select homes across the spectrum of prices that are displayed in your search listing. For example, suppose you have a total of 250 listings of homes in your selected area. Note on the right side of the webpage, the cheapest 12 of the 250 listings are shown and with each sequential page the prices increase. Thus, for this example, there are approximately 250/12 = 21 pages of listings for the 250 houses. Therefore, try to select approximately 3 listings from each of the 21 pages (rather than selecting all prices in one page) so that the data collected is a better representation of the prices in the area. Collect the

following information (5 variables) from each of the 60 real estate property listings: Y = listing price

X1 = interior floor space (square footage of the house)

X2 = land size (square footage). This is usually represented by length of the front of the lot (in ft) X the depth of the lot (in ft). sizes range but a common lot size in Vancouver is 33 X 120. You will need to convert this to feet squared (area) before entering it into your spreadsheet (i.e., 3960)

X3 = number of bedrooms X4 = number of bathrooms

X5 = age of the building (often listed as “Built in” year date. Thus, you will need to calculate it.)

Key all data into an excel spreadsheet.

Note: some listings will not have all the data. If you are collecting data in a geographical area that does not provide the information above, please move onto another area. You will find that suburban areas around the Lower Mainland provide the above information quite readily.

Also Note: you can use your mls search and reuse it later by emailing yourself the browser URL. Simply copy and paste the URL from the mls webpage into an email and mail it to yourself so you can continue your data collection at another time.

# Use EXCEL to complete the following ( Note: all of the below have to done in excel and should provide with calculations and excel code for generating plots.)

1. Please create 5 scatter plots: Y versus each Xi for I = 1, 2, …5 and provide a comment with respect to the presence of an “approximate” linear relationship or not.
2. Determine the regression model that best predicts selling price.
3. Is the overall model significant? Carry out an overall F-test to determine this. Be certain to state the hypothesis, your decision rule and provide a concluding statement in the context of the problem.
4. Clearly state the meaning of the regression coefficients in the context of this problem.
5. Provide a proper Hypothesis Test at the 5% significance level on your estimated regression model to demonstrate if there is significant linear relationship between each independent and dependent variable.
6. How good is the fit of your model? Quote a measure from the regression output. Provide a clear statement of its meaning.
7. Check that the assumptions that are required by linear regression model are valid.
   1. Generate a residual plot on the estimated model (residuals versus fitted values) and provide a commentary of what it means.
   2. Generate a normal probability plot on the estimated model and provide a commentary of what it means.
   3. Also, check for multicollinearity by creating a correlation matrix. Correlation stronger than + or – 0.6 between any two x variables indicates that they are somewhat redundant and may cause problems in your analysis. Do you have any such potential problems?
8. Does this model seem to be a good predictor of real estate market value? Draw conclusion . State why or why not you believe the model is a good/poor predictor. Also, suggest ways that the model could be improved.

# Deliverables:

1. Research report which includes:
   1. Introduction (1 page)
   2. Data (data source, sampling, etc) – Whatever it takes to fit in
   3. Methods (discuss the variable included in the model and their hypothesized relationship with selling price of a house. Discuss briefly the regression model you estimate to test these hypotheses) (data and methods 2pages)
   4. Results (start with some descriptive results such as average price, median price, etc and provide the regression model results) and the interpretation of the results. (2-3pages)
   5. Summary and conclusion (1 page)
   6. References and in-text citations APA -7 Style ( Should be Peer reviewed and less than 10 years old).
   7. Appendix (all the results should be appended).
   8. Electronic files (collected data files, regression outputs, etc).
   9. All the requirements covered in Excel Sheet as per Excel Requirements above. This should be submitted along with the Business Research Report.