**RStudio Lab Week 3 – Multiple Linear Regression**

# *Notes*

The ‘multreg.csv’ dataset shows the price (in thousands of rands) of 60 properties in the Knysna region. The dataset also contains the size of each plot in hectares, the floor size of the house in square meters, the number of trees on the plot, the distance in km of the plot from the Knysna lagoon and whether or not the property has a pool. The task in this lab is to build a multiple regression model of the price of the property as a function of all these given explanatory variables, as well as an interaction between the plot size and floor area.

*Datasets*

All questions relate to the ‘multreg.csv’ data.

# *Instructions*

1. Make sure that you have downloaded the dataset ‘multreg.csv’.
2. Copy the dataset into your student drive and make sure you have opened RStudio.
3. Set your working directory to your student drive.
4. Insert a new R Script.
5. Import the multrreg.csv dataset into RStudio.
6. Generate the correlation matrix between all the numeric variables by applying the **cor()** function to the first 5 columns of the data.
7. Perform a multiple linear regression using Price as the Dependent variable and all the other variables as Explanatory variables. Also, include an interaction term for plot size and floor area. Save your model in an object called fit. Hint: *formula = dependent variable ~ explanatory 1 + explanatory 2 + … + explanatory n + explanatory 1 \* explanatory n.* The last term in this formula denotes an interaction between explanatory 1 and explanatory n.
8. Use **summary()** and **confint()** on the fit object to generate the summary of regression results and 95% confidence intervals for the coeﬃcients.
9. Use the **predict()** function to find out what the estimated house prices for each house in the dataset is.