Objective:
In this session, you will learn to build multivariate linear regression model, to validate your model and interpret the results, and the evaluation metrics. We will also look at how to go about building a model and how to look for evidences which can guide us in building a better model.

Key takeaways:

- Building multivariate linear predictive model using lm()
- Interpreting the diagnostic plots to check for the linear regression assumptions
- Identifying influential observations and handling them.
- Checking for multicollinearity through VIF.
- Selecting features using StepAIC
- Transformation on the target variable
- Check for multiple metrics like VIF, Adjusted R- squared, p values, F statistic, Error metrics (RMSE/MAPE). What does each one of them mean and how to use them in model building.

Process:

- Reading the data and getting a feel of it
- Exploratory Data Analysis
- Data preprocessing
- Building the model
- Evaluating the model and its performance

Problem : *How to build a multivariate linear regression model*

Linear regression is such a model where there are no hyper-parameters. The model totally depends on how the data is. So better the data, better the performance of the model. Remember, 'Garbage in → Garbage out'.

- Read the 'CustomerData.csv' data into R. Our objective here is to predict 'TotalRevenueGenerated'.
- Split the data into train/validation
- Perform data preprocessing steps
- Start with a basic model by using all the attributes as it is.
- Try different models by performing residual analysis, stepAIC, VIF, p-value analysis etc., and try to select significant attributes that improve model's performance
- Check for the error metrics on train,validation, ad test datasets
- Decide on the final model