# CHRIST (Deemed to be University) Data Science Department

**Bachelors of Science (Data Science)**

**Course**: BDS471L – Machine Learning

**Exercise No**: LAB Exercise – 5

**Date:** 17 – 02 – 2024

Use the following platform to demonstrate the following exercises.

Jupyter Notebook, Google Colabs, Any Python Editor

**Topic: Linear Regression**

1. For the given dataset identify the problem statement.
2. Handle the missing data based on CCA, simpleImputer ( Mean Median Mode, Arbitary, End of Distribution, Random Sample Imputation), categorical missing data handler, Missing Indicator method wherever applicable.
3. The data distribution verification and Linear regression assumptions have to checked and based on the results feature transformation has to be remodified.
4. Upload the top 2 R2 score results based on the various combination of techniques applied(2 separate files)

# Evaluation Scheme: (Total 10 Marks)

1. Question 2 , 4 marks
2. Question 3, checking distribution and modifying 2 marks.
3. Question 3 Checking assumptions and modifying 2 marks
4. Testing and validation 2 marks

**Total 10 marks \*2 = 20 marks, converted to 10 marks**

**NOTE: Without handling missing values or data distribution or checking assumptions test results will not be accepted and ‘0’ marks will be awarded.**

# General Instruction:

1. Create a word document and paste all the answers. File name should be your register number followed by lab No: Example: 2048001\_lab1
2. Upload the answer document in Google Classroom on or before the deadline mentioned.