# Subject: 21DS602

Lab Session: 01

**Notes:**

1. Please read the assignment notes carefully and comply to the guidelines provided.
2. Prepare a report based on your lab work and the Report Assignment question provided at the end of this document. Report may contain diagrams, charts, code snippets and results as applicable. Screenshots or photos of diagrams / code snippets are not allowed in the report. Submit the report to TurnItIn (details shall be provided in the lab).
3. If you have not completed the prerequisite assignments, please complete them before starting these assignments.

## Main Section (Mandatory):

**Refer to lecture portions on Linear Algebra.**

A1. Please refer to the "***Purchase Data***" worksheet of ***Lab Session1 Data.xlsx***. Please load the data and segregate them into 2 matrices A & C (following the nomenclature of AX = C). Do the following activities.

- What is the dimensionality of the vector space for this data?
- How many vectors exist in this vector space?
- What is the rank of Matrix A?
- Using Pseudo-Inverse find the cost of each product available for sale.
  *(Suggestion: If you use Python, you can use* ***`numpy.linalg.pinv()`*** *function to get a pseudo-inverse.)*

A2. Use the Pseudo-inverse to calculate the model vector X for predicting the cost of the products available with the vendor.

A3. Mark all customers (in "***Purchase Data***" table) with payments above Rs. 200 as RICH and others as POOR. Develop a classifier model to categorize customers into RICH or POOR class based on purchase behavior.

A4. Please refer to the data present in "***IRCTC Stock Price***" data sheet of the above excel file. Do the following after loading the data to your programming platform.

- Calculate the mean and variance of the Price data present in column D.
  *(Suggestion: if you use Python, you may use* ***`statistics.mean()`*** *&* ***`statistics.variance()`*** *methods).*
- Select the price data for all Wednesdays and calculate the sample mean. Compare the mean with the population mean and note your observations.
- Select the price data for the month of Apr and calculate the sample mean. Compare the mean with the population mean and note your observations.
- From the Chg% (available in column I) find the probability of making a loss over the stock.
  *(Suggestion: use lambda function to find negative values)*
- Calculate the probability of making a profit on Wednesday.
- Calculate the conditional probability of making profit, given that today is Wednesday.
- Make a scatter plot of Chg% data against the day of the week

## Optional Section:

O1. Create 2 separate square matrices from the purchase data matrix. Repeat experiments A2 & A3 with both these matrices. Do the X values obtained from the square matrices match to the one obtained from the whole purchase data matrix?

## Report Assignment:

1. Discuss the importance of rank of an observation matrix in model building for classification.
2. Discuss on regression (Ex: A2) and classification (Ex: A3) tasks. How would you differentiate between them.
3. Observing the stock data provided, record your suggestions to build a system that may be able to predict the price and Change % into future.