**Assignment 1**

**1. What is the difference between data analysis and machine learning?**

Data Analysis is a process of understanding the data, find patterns and try to obtain inferences due to which the underlying patterns are observed.

Machine Learning is when you train a system to learn those patterns and try to predict the upcoming pattern.

**2. What is Big Data?**

Big Data is a collection of data that is huge in volume, yet growing exponentially with time. It is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently. Big data is also a data but with huge size.

The New York Stock Exchange is an example of Big Data that generates about one terabyte of new trade data per day.

**3.What are the four main things we should know before studying data analysis?**

1. sql

2. Microsoft Excel

3. Critical Thinking

4. R or Python –Statistical analysis

5. Communication skill

**4. Most common characteristics used in descriptive statistics?**

This page describes graphical and pictorial methods of descriptive statistics and the three most common measures of descriptive statistics (central tendency, dispersion, and association).

Descriptive statistics can be useful for two purposes: 1) to provide basic information about variables in a dataset and 2) to highlight potential relationships between variables. The three most common descriptive statistics can be displayed graphically or pictorially and are measures of:

* Graphical/Pictorial Methods
* Measures of Central Tendency
* Measures of Dispersion
* Measures of Association

**5. What is quantitative data and qualitative data?**

Quantitativedata are measures of values or counts and are expressed as numbers. Quantitative data are data about numeric variables (e.g. how many; how much; or how often).  
Qualitative data are measures of 'types' and may be represented by a name, symbol, or a number code. Qualitative data are data about categorical variables (e.g. what type).