ASSIGNMENT 1

1) What is the difference between Data Analysis and Machine Learning?

DATA ANALYSIS	MACHINE LEARNING
• Data analysis studies how to collect, process, and interpret data. Assumptions are made by humans.	• Machine learning refers to the study of algorithms that improve through experience. Learning happens by applying algorithms to data and applies the learning without requiring human intervention.
• Data analysis refers to reviewing data from past events for patterns.	• Machine learning is able to make assumptions, test and learn autonomously.
• It does not predict the impact of a change in a variable. Predictions rely on human interaction to query data, validate patterns, create and then test assumptions.	• Machine learning analyzes data, makes assumptions, learns and provides predictions at a scale and depth of detail impossible for individual human analysts.

2) What is Big Data?

Big data is also a data but with huge size. Big data is described in petabyte range or higher. Eg: The New York Stock Exchange is an example of Big Data that generates about one terabyte of new trade data per day.
Big data can be described by the following characteristics:
   • Volume, Variety, Velocity, Veracity, Value

Following are the types of Big Data:

   ➢ Structured - Any data that can be stored, accessed and processed in the form of fixed format is termed as a 'structured' data. Eg: An Employee table in a database is an example of Structured Data.

➢ Unstructured - Any data with unknown form or the structure is classified as unstructured data. Eg: heterogeneous data source containing a combination of simple text files, images, videos etc.

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

- Excellent problem-solving skills. Skill to collect and process data.
- Knowledge of mathematical statistics.
- Structured Query Language (SQL) to view, manage and access information you're working with.
- A programming language that will help you to work with data. Good presentation skills.

4) Most Common characteristics used in descriptive statistics?

- The Measure of Central Tendency - The measures of central tendency are used to show the center of the data set. The central tendency is estimated using the mean, median and mode.
    ➢ Mean is the average of all the data.
    ➢ Median is the middle of the entire data set.
    ➢ Mode indicates the most commonly occurring value in the data set.
- The Measure of Spread - The objective of measure of dispersion or variation is to identify the extent to which the entire data set is spread from the central tendency – specifically mean. he commonly used estimates are range, standard deviation, and variance.
    ➢ The range gives you an idea of how far apart the most extreme response scores are. To find the range, simply subtract the lowest value from the highest value.
    ➢ The standard deviation (s) is the average amount of variability in your dataset. It tells you, on average, how far each score lies from the mean. The larger the standard deviation, the more variable the data set is.
    ➢ The variance is the average of squared deviations from the mean. Variance reflects the degree of spread in the data set. The more spread the data, the larger the variance is in relation to the mean.

5) What is Quantitative Data and Qualitative Data?

- ➢ Quantitative data
    - • Quantitative data is anything that can be counted or measured.
    - • Quantitative research is based on numeric data.
    - • Quantitative variables can tell you "how many," "how much," or "how often."
    - • Quantitative research offers more objective findings.
    - • Fixed and Universal
    - • Quantitative data is analyzed using statistical analysis.
    - • Eg: Age, Height, Income, Number of errors, Percentage of lecture attended, Clinical skills performed, Speed of a car.

- ➢ Qualitative data
    - • Qualitative data is descriptive, referring to things that can be observed but not measured.
    - • Qualitative research focuses on the qualities of users - the 'why' behind the numbers.
    - • Dynamic and Subjective, open to interpretation.
    - • Qualitative data is analyzed by grouping the data into categories and themes.
    - • Eg: Gender, Marital Status, Qualification, Method of treatment, problem solving strategy used, Native language, Ethnicity, Economic status.