**ASSIGNMENT-1**

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

|  |  |
| --- | --- |
| DATA ANALYSIS | MACHINE LEARNING |
| It is the process of deriving meaningful insight from the data through various analytical method and tools. | In this computer programs are developed to make decisions/learn by themselves using algorithms programmed into them. |
| Skills used:   * Statistical Analysis * SQL(Structured Query Language) * Knowledge of R & Python | Skills used:   * Deep learning * Natural Language Processing(NLP) * Computer vision |

**2.What is Big Data?**

* When we have data in big size is known as “Big Data”.
* In different sizes from terabytes to zettabytes.
* It includes Structured ,semi structured ,unstructured data from different sources.
* It includes 5v’s-Variety,Volume,Value,Veracity,Velocity.

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

* Descriptive statistics
* Inferential statistics
* Distribution
* Hypothesis testing

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

* It summarizes or describes the characteristics of data.
* It consists two categories:
* Measure of Central Tendency-Focus on the average or middle values of data set ,whereas measures of variability focus on the dispersion of data.
* Measure of Spread-Aid in analyzing how dispersed the distribution is for a set of data.

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

Quantitative data: It is divided into two categories,

* Nominal data-For naming a things and used for labeling variables.{example: gender ,hair color ,ethnicity}.
* Ordinal data-It is used for arranged in order.{example: ordinal numbers(first, second, third)letter grades(A,B,C) and star reviews}.

Qualitative data: It is divided into two categories,

* Discrete data-It involves only integer values{example: no. of students ,no. of workers}.
* Continuous data-It only possible when it having infinite values.{example: speed of cars ,height of children}.