ASSIGNMENT

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

Data Analysis is the collection , transformation and organization of data in order to draw conclusions , make predictions and drive informed decision making whereas Machine learning is a type of artificial intelligence that allows computers or machines to automatically learn from data without being specifically programmed to.

Data science is the field that studies data and how to extract meaning from it while machine learning focuses on tools and techniques for building models that can learn by themselves by using data

A data analyst is someone who collects, transforms and organizes data in order to make informed decisions. A machine learning engineer builds models. They choose the most appropriate algorithm for a particular problem and try to achieve certain reproducible results by running experiments on data.

The goal of data analytics is to extract relevant information from a usually small dataset whereas the goal of machine learning is to develop software that learns by itself by extracting meaning from data.

The tools involved in data analytics involves using analytics applications on structured data whereas machine learning Involves using ML algorithms and analytical models.

The scope of data analytics includes predictive modelling, risk analytics, and other whereas machine learning includes supervised, unsupervised, semi-supervised learning.

The role of data analyst focuses on processing raw data to create meaningful insights. They work to identify trends and present them in a meaningful and easy to understand way whereas Machine learning engineers create algorithms and programs that help computers to learn automatically.

2) What is big data?

Big data is a combination of structured, semistructured and unstructured data collected by organizations that can be mined for information and used in machine learning projects, predictive modeling and other advanced analytics applications. It refers too data that is in the petabyte range or higher.

Companies use big data in their systems to improve operations, provide better customer service, create personalized marketing campaigns and take other actions that, ultimately, can increase revenue and profits.

It is described by 5 v’s

VARIETY- Data comes in all types of formats – from structured, numeric data in traditional databases to unstructured text documents, emails, videos, audios, stock ticker data and financial transactions

VOLUME - It refers to the size of Big Data. Data can be considered Big Data or not is based on the volume.

VALUE - Just because we collected lots of Data, it’s of no value unless we garner some insights out of it. Value refers to how useful the data is in decision making. We need to extract the value of the Big Data using proper analytics.

VERACITY - It refers to the assurance of **quality , credibility or accuracy** of the data. Since the data is collected from multiple sources, we need to check the data for accuracy before using it for business insights.

VELOCITY - It refers to the speed at which the data is generated , collected and analysed.

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

We should know about what is data, its different types, examples of data in real life, big data. We should also know about the statistical tools such as excel.

4) Most common characteristics used in descriptive statistics

Descriptive statistics result from gathering data from a body, group, or

population and reaching conclusions only about that group.

* Descriptive statistics summarizes or describes the characteristics of a data set.
* It consists of two basic categories of measures: measures of central tendency and measures of variability (or spread).

There are 3 main measures of central tendency – Mean, Median and Mode.

5) What is quantitative data and qualitative data?

Quantitative data can be counted, measured, and expressed using numbers. Qualitative data is descriptive and conceptual. Qualitative data can be categorized based on traits and characteristics. It is expressed in terms of language.

Examples of quantitative data include height, weight, time, price, temperature, etc. Examples of qualitative data are scents, appearance, beauty, colors, flavours, etc.