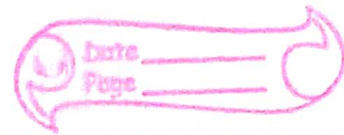Assignment No. 1

Q.1 Explain first phases of data science process?
→ ① The first phase of data science process is problem statement or understanding the problem.

② It's really very important to declare or formulate your problem statement very clearly and precisely.

③ Your whole model and it's working depend on your statement.

④ Many scientist considers this as the main and much important step of data science.

⑤ So make sure what's your problem statement and how well can it add value to business or any other organisation

⑥ Sometimes the customers will make a clear request, while other may ask you to solve a very broad problem.

⑦ The first step in these situation is to identify clear objectives and concrete difficulties.

Q.2 Describe Term Data Science? with its Need?
→ * Data Science :→

① Data Science is a deep study of the massive amount of data, which involves extracting meaningful insights from raw, structured and unstructured data that is processed using the scientific method, different technologies and algorithm.

③ It encompasses a wide range of task, including data cleaning and preparation, data visualization, modelling, ML & etc.

② It is interdisciplinary field that allows you to extract knowledge from structured or unstructured data.

* Need of data science :→

① Data science can help you to detect fraud using ML algorithm.

② Allows to build intelligence ability in machines.

③ It helps you to recommend the right product to right customer to enhance your business.

④ It enables you to take better and faster decision.

⑤ with data science technology, we can convert the massive amount of raw and unstructured data into meaningful insights.

Q.3 Summarize the importance of linear algebra in data science.

→ ① Linear algebra is a fundamental branch of mathematics that has important application in data science.

② Some key reasons why linear algebra is important in data science are :

a) Data representation : →

   linear algebra provides a powerful toolset for representation and manipulation data in a concise and efficient manner.

b) ML algorithm : →

   many machine learning algo like linear regression, logistic regression & etc. rely on linear algebra. This use to make prediction and their models.

c) Optimization : →

   linear algebra is used to solve optimization problems in data science, such as finding the best parameter for ML model. It also provide a framework for solving problem efficiently.

d) Data visualization : →

   linear algebra can be used to transform high-demensional data into a lower-dimensional space that can be visualized easily.

Ques.4 Describe in detail structured thinking of data science.

→ Structured thinking in data science involves breaking down complex problem into smaller, more manageable component and using logical step by step approach to solve them. It solve using several key steps:

a) Define the problem
b) Collect and preprocess data
c) Exploratory data analysis
d) choose an appropriate model
e) Train and validate the model
f) Evaluate and interpret results.
g) communication finding.

   Structured thinking is critical for data science because it helps data scientists break down complex problem into smaller, more manageable components, and ensures that they follow a logical step by step approach to solve them.

   This approach can help improve the accuracy and efficiency of data analysis and lead to more informed decision-making.

Q.5 Summarize numpy arrays.

→① Numpy array are a fundamental data structure in numpy library for python.

② They are used to store & manipulate large, homogeneous and multidimensional arrays of data.

③ Numpy arrays are similar to python lists, but they are more efficient in terms of memory usage and computation speed.

④ Numpy array are created using the numpy.array() function and can be one-dimensional, or multi dimensional.

⑤ One of the key benefit of numpy array is their ability to perform vectorized operations.

⑥ Numpy array are widely used in scientific computing, data analysis and machine learning.

⑦ Example:

```
import numpy as np
li = [1,2,3,4]
numpyArr= np.array(li)
print(numpyArr)
```

output:
[1 2 3 4]