

# **Internship tasks**

## **Section A (Theory):**

### **Q1. What is Data Science?**

**Ans:** Data science is a sub field of computer science , which deals with collecting , cleaning , analyzing of data and taking insights from data and sharing them with stakeholders to help businesses to take decisions on the basis of data .Data science is applied in fields like finance ,health care sports Politics to take decision on past available data.

### **Q2. What is the difference between Structured and Unstructured Data?**

**Ans : 1.Structured data :** Such type data which is structured in table rows and column files like csv , excel or json , or in data bases is called structured data.Structured data is mostly stored in SQL based data bases or files and data warehouses.

**2.Unstructured data :** Such type of data which is not structured in a specific formate , like social media posts Youtuble videos , images are the examples of unstructured data. Such type of data is mostly stored In NoSQL data based.

### **Q3. What are Features and Labels in a dataset?**

**Ans :** Features are the input variables In a data set which describes the data , while Label is a output variable ,these terms are mostly used in machine learning like Prediction model in which we give features to model and try to predict some unknown value

Example dataset : In titanic data set there are its columns are PassengerId , Pclass , Sex , Fare etc , so these column are also called Features in Machine learning and if suppose we want to predict that either the passenger will survived or not on the basis of Known features then this is called output variable or Label