Starting with Data Science

Data

Data often refers to raw facts which further need to be analyze in order to generate information and decision in several contexts such as business, science and our everyday life. It can be a list of phone numbers or records of patients present in hospital etc.

Definition of Data Science

1. Data science aims to **extract knowledge and insights** from data using scientific methods like probability and statistics, including more than just traditional statistics.
2. The knowledge gained from data science should lead to **actionable insights** that can be applied into real-world business scenarios.
3. Data scientists must handle both **structured and unstructured data** and often need to have expertise in specific application domains such as finance, medicine, or marketing etc.

Types of Data

1- **Structured**

Data which is highly organized and easily searchable, typically stored in databases or spreadsheets i.e. List of people with their phone numbers.

2- **Semi-Structured**

Data that does not reside in a traditional database but has some organizational properties that make it easier to analyze. i.e. Collection of scientific papers in .Json format with authors, date of publication etc.

3- **Unstructured**

Data that lacks a predefined format, structure or organization, making it more challenging to collect, process, and analyze i.e. Raw video feed from a camera.

Where to get data?

About getting data, it depends on, in which format we want if want to have. For structured data, we should go for source which stores data in some specific format such as Relational databases & surveys etc. Next for Unstructured format we can get data from images, videos and logs etc. Semi-structured data can be available to us by Social Networks.

What can we do with data?

After acquiring data, we store it in our databases then preprocess in order to transform it into some structured format for further steps like for visualization and training some predictive model which requires data to be stable and standardize in order to give good results.