**Task1.1 - Write a short notes on the difference between Data analysis and Data Science**

Data Analysis:

* Data analysis is a subset of the broader field of data science.
* It involves the examination, cleaning, transformation, and interpretation of raw data to discover insights, identify patterns, and draw conclusions.
* The primary focus of data analysis is to extract meaningful information from data and make it more accessible for decision-making.
* Data analysts often work with structured data, using statistical techniques, charts, graphs, and visualization tools to present their findings.
* They help businesses and organizations understand historical trends and performance, aiding in optimizing processes and making data-driven decisions.
* Tools commonly used in data analysis include Microsoft Excel, SQL, and various statistical software like R or Python libraries (e.g., pandas, NumPy).

Data Science:

* Data science is a broader field that encompasses data analysis but goes beyond it, incorporating various other aspects like machine learning, artificial intelligence, and big data processing.
* It involves the entire lifecycle of data, from data acquisition and cleaning to analysis, model building, and deploying data-driven solutions.
* Data scientists not only analyze historical data but also use it to build predictive models that can forecast future trends and outcomes.
* They work with both structured and unstructured data, including text, images, and audio, to derive actionable insights and create innovative solutions.
* Data science also involves the development of algorithms and computational models for solving complex problems and automating decision-making processes.
* Data scientists often use programming languages like Python or R, along with libraries and frameworks such as TensorFlow or scikit-learn for machine learning and AI tasks.

**Task 1.2** - **Write short notes on data cleaning**

Data Cleaning:

Data cleaning, also known as data cleansing or data scrubbing, is a crucial step in the data analysis process. It involves identifying and correcting errors, inconsistencies, and inaccuracies in datasets to ensure the data is reliable and ready for analysis. Understanding data cleaning is essential for working with real-world data and producing accurate and meaningful insights.

Key Steps in Data Cleaning:

* Handling Missing Values
* Removing Duplicates
* Correcting Inaccurate Data
* Handling Outliers
* Data Formatting
* Dealing with Inconsistent Categorical Data
* Addressing Data Integrity Issues
* Handling Irrelevant Data