Name : Rowaida Mostafa Sayed

Code : 202401981

Assignment #4

Q1- What is the difference between data analysis, data mining, and data science?

1. Data Analysis

Definition: The process of inspecting, organizing, and summarizing data to understand what is happening or what happened in the past.

Goal: To draw logical conclusions based on existing data.

Tools Used: Excel, Python (Pandas, Matplotlib), Power BI, etc.

Real-World Example: A sales company uses data analysis to determine which products sold the most in the past month or to analyze seasonal sales trends.

Simple Example: Given a monthly sales table, you calculate averages, standard deviations, and create visual charts to see general trends.

2. Data Mining

Definition: The process of discovering hidden patterns and relationships in large datasets using advanced algorithms.

Goal: To uncover previously unknown insights or correlations.

Tools Used: Weka, RapidMiner, R, Python (Scikit-learn), etc.

Real-World Example: A supermarket applies data mining to discover that customers who buy diapers often buy milk (Market Basket Analysis).

Simple Example: Analyzing purchase behavior reveals the rule: "If a customer buys bread, they are likely to buy butter."

3. Data Science

Definition: A broad field that integrates data analysis, data mining, machine learning, programming, and statistics to extract deep insights and make predictive decisions.

Goal: To build predictive and intelligent models that support future decision-making.

Tools Used: Python (NumPy, Pandas, Scikit-learn, TensorFlow), R, SQL, Jupyter Notebooks, etc.

Real-World Example: A food delivery app uses data science to predict dishes a user might want to order based on their history and similar user behavior.

Simple Example: Netflix recommends movies based on your past viewing history and the behavior of users similar to you.

Summary Table

Category	**Data Analysis**	**Data Mining**	**Data Science**
Goal	Understand the past	Discover hidden patterns	Predict the future and make smart decisions
Input	Clean data	Large datasets	Big, complex datasets
Outcome	Reports, visualizations	Rules and correlations	Predictive or intelligent models
Tools	Basic analysis tools	Advanced algorithms	Machine learning + Programming + Stats

References

Data science - Wikipedia

Data mining - Wikipedia

Data analysis - Wikipedia

Difference Between Data Mining and Data Analysis - GeeksforGeeks

Difference Between Data Science and Data Mining - GeeksforGeeks

Also from the lecture