**Task 1  
Data Roles and Responsibilities**

**Roles:**

There are three main roles in data domain.

1. Data Scientist
2. Data Engineer
3. Data Analyst

**Data Scientist:**

Data Scientists act as the engine of every data team. They gather data, clean and process it and perform predictive analysis on it by using statistics, math and solve business related problems.  
They also extract meaningful insights from data generated by customer’s behavior

**Example:**

Auto recommendation process of social media apps, YouTube, Netflix etc..

**Skills:**

Strong Programming Knowledge in SQL, R and Python

Probability and Statistics

Machine Learning

Data Wrangling and Cleaning.

Neural Networks and NLP

Business Knowledge

Communication Skills

Problem Solving Mindset

**Data Engineer:**

The key role of a Data Engineer is to convert raw data into usable data, which includes collection, transformation, sorting, design and implementation, storing, ETL (Extract, Transform, Load) and making data readily available to users (Data Scientist and Data Analysts) securely.  
This whole process is called a Data Pipeline.  
Data Engineers also do Data Warehousing and create Data Lake, Data Marts etc and also do reporting to Data Scientists and Business Stakeholders.

**Example:**

Below are certain use cases of Data Engineering in Finance Sector  
Risk Management, Customer Relationship Management, Fraud Detection and prevention.

**Skills:**

Strong Command on Programming Languages like Python and SQL  
Good Knowledge of Linux OS and Commands  
Database Management System  
Data Warehousing  
Distributed Systems (Spark with Python for Big Data)

Cloud (AWS, GCP, Azure)

**Must Learn Tools:**

1. Orchestration: Airflow  
   Computing: DataBricks/Snowflake
2. CICD: Jenkins
3. Streaming: Kafka
4. Containers: Dockers

**Data Analyst:**

The job of Data Analyst is to collect, clean and transform data into meaningful insights by visualizing data and making dashboards, also presenting those dashboards to business stakeholders to take informed decisions.

**Skills:**

1. Python and SQL.
2. Excel and Power BI/Tableau
3. Basic Statistics and Math.
4. Good Communication Skills

**Conclusion**

In the data domain, each role requires specific technical skills and a problem-solving mindset to effectively manage and utilize data. Data Scientists focus on gathering, cleaning, and analyzing data to extract insights and perform predictive analysis. Data Engineers build and manage data pipelines, ensuring data is available and secure for use by others. Data Analysts transform data into actionable insights through visualization and dashboards for business decision-making.