1. write down all the data related terminologies that are there (eg : data lake, data wrangling etc)

1. Data Lake:

A Data Lake is a centralized repository designed for the storage of diverse and large volumes of raw, structured, semi-structured, and unstructured data. It enables organizations to retain data in its native format, providing flexibility for various analytical processes, including data exploration, machine learning, and advanced analytics.

2. Data Wrangling:

Data Wrangling is the process of transforming and cleaning raw data into a structured and usable format for analysis. It involves tasks such as handling missing values, converting data types, and addressing inconsistencies, ensuring that the data is prepared and refined for downstream applications like visualization, reporting, and modeling.

3. Data Integration:

Data Integration involves combining data from different sources to provide a unified and coherent view. It aims to streamline data flow across an organization, facilitating better decision-making by ensuring that data is consistent, accurate, and accessible across various departments and systems.

4. Data Quality:

Data Quality refers to the reliability, accuracy, and consistency of data. It involves processes and measures to ensure that data meets specific standards and is fit for its intended purpose, minimizing errors and inaccuracies that could impact business decisions.

5. Machine Learning:

Machine Learning is a subset of artificial intelligence that focuses on developing algorithms enabling computers to learn from data and make predictions or decisions without explicit programming. It plays a crucial role in data analysis, pattern recognition, and automation of decision-making processes.

6. Data Visualization:

Data Visualization is the representation of data in graphical or visual formats to facilitate easier understanding of patterns, trends, and insights. It enhances data comprehension and aids in effective communication of complex information.

7. Data Scientist:

A Data Scientist is a professional who utilizes scientific methods, statistical techniques, and computational tools to extract valuable insights and knowledge from data. They play a key role in interpreting complex data sets and providing actionable recommendations for business decision-making.

8. ETL (Extract, Transform, Load):

ETL, or Extract, Transform, Load, is a data integration process that involves extracting data from source systems, transforming it into a suitable format, and loading it into a target system. ETL is crucial for maintaining data consistency and ensuring that data is appropriately prepared for analysis and reporting.