Individual Homework – Week 1

Data engineer's methodology and why they use that

Data engineer handle amount of data, so it makes them to automate everything for analysis and business pupose. Because of that, data engineers have to manage and organize data. It's important to sets up the data warehouse, the data framework, and the data pipelines for the data scientist to conduct complex analysis on. They create data pipelines that connect data from one system to another and also responsible data transformation. It can help data analyst/scientist to pull data from different systems for analysis. Data engineer must have several programming languages for automated that like Python, SQL or NoSQL and also several tools as hadoop etc. Example if data engineer want to clean or transform data, they can code in python instead of checking row per row. Also if they want to running pipeline at certain time, they can make schedule for that instead of running it manually.

"The industry of Data Engineering is taking a step forward in automating the data pipeline to confine the process that goes into transforming and collecting data. This methodology reduce the workload on Data Analytics and Machine Learning. Initially, we have seen that Data Science has adapted automation to conduct the most repetitive tasks. Now, Agile Data Engineering and DataOps tools are emerging within Data Engineering to handle the repetitive data pipeline work." (https://intellipaat.com/blog/what-is-data-engineering)

Python called Object-Oriented Language, why?

Because we can create classes and objects that can make the code is reusable and easier to understand. A class is a collection of objects or you can say it is a blueprint of objects defining the common attributes and behavior.