Good Morning

I hope your weekend was Good.

Although I have a few issues regarding the current data. What is the primary source of the data? Do you have the same data in any other file format? Is the Data provided has recently been updated or not?

Along with the questions I also found out that the data is inconsistency as it contains many missing values, by segregating and deep diving through the data. While performing the pre-processing task on the data I came across many duplicate data and understood that the complete data is not properly structured as well.

As these missing values if replaced by some random value while change the nuance of the data hence resolving these missing values is at most priority. Need to find out the cause of duplication of data as Multiple copies of the same records take a toll on computing and storing but may also produce skewed or incorrect insights when undetected.

Along with this I would need the latest production files which will provide values to the assets to the company with the more recent information.

Working with the latest production data will suffice the performance on the production. To scale up this performance separate server allocations would be recommended.

Looking Forward to Hear from you.

Thank you!

Regards

Hardik Rathod