DATA MODELING:

* What is dm?

To design the structure of the data is known as data modelling

* why dm?

To organize the data so that we can get the meaningful insights from the data and to provide a proper structure so that the relationship between the attributes can be known easily.

* Types of dm?

Relational database:

It is used to give the data a shape i.e it converts the data into the rows and columns and uses schema to define the relationship between them.

It normalizes the data which removes the redundancy from the same.

dimensional database:

It is used in data warehousing where there is 1 fact table and have multiple dimension tables.

Entity - Relationship database:

Showing the relationship between the entities .

It helps in understanding the structure of the database

DATA INTEGRATION:

What is di?

It collects data from multiple resources and combine it to get the insights.

It is stored in the datawarehouse.

Data cleaning is also the part of data integration.

It comes under the ETL !

it extracts the data, transform the data as per the current system data format and normalize it and then load it in the database or the system.

For date inconsistent pattern :

If date column has any different values from the actual format so we have to correct it by giving it the value of the exact next column…

For that we use library..

**import datetime**

Character encoding:

When our data is in different places suppose countries so we have to encode and decode it as per the region it is stored in …so for that we **import charset\_normalizer.**

Inconsistent data:

When there is redundancy in the data so for cleaning we use fuzzywuzzy library…

**import fuzzywuzzy**

**from fuzzywuzzy import process**

**import charset\_normalizer**

Data visualization:

Using seaborn library

Sns.lineplot

Sns.barplot

Sns.heatmap—to analyze the correlation between the variables

Sns.scatterplot

Sns.swarmplot --for the categorical variables

Sns.regplot—to add one regression line(linear ) we use this

Sns.lmplpot—to add 2 regression lines