Exploratory Data Analysis (EDA) Project

what is EDA?

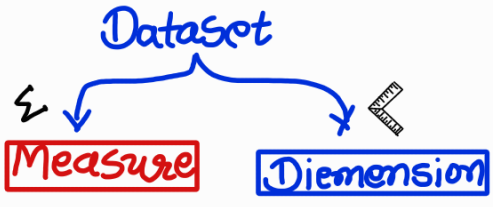
→ Understand and cover insighs about our datasets.
→ This kind of project - How to ask right quantion and How to Find answer Using SQL.

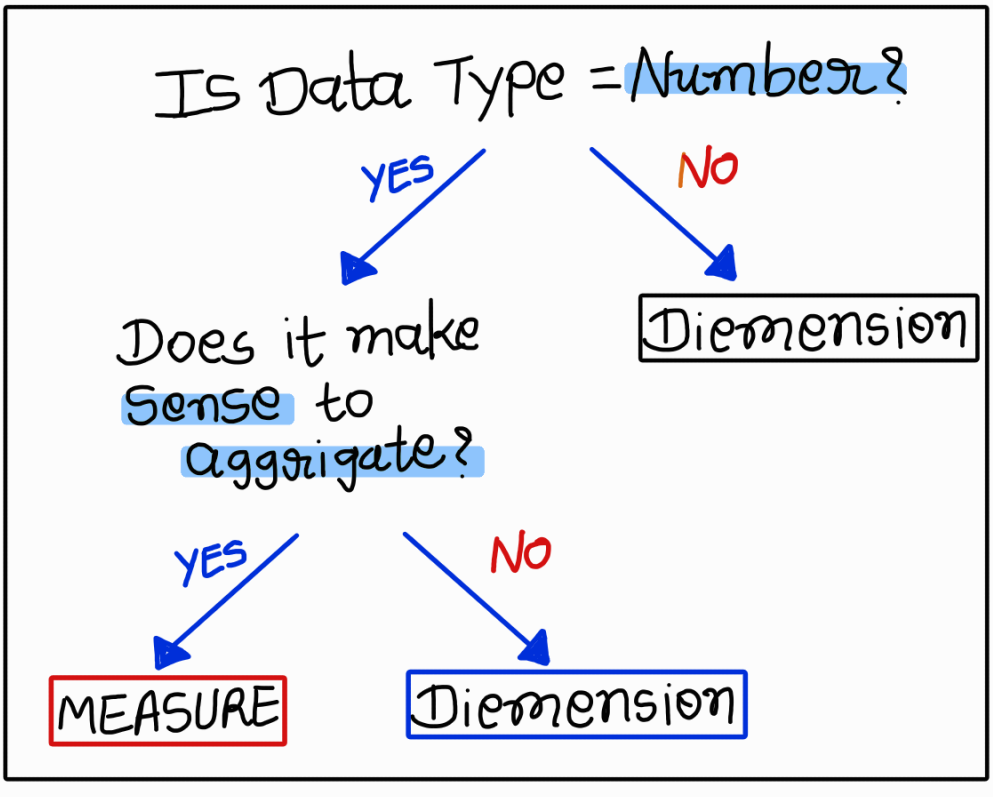
Exploratory Data Analysis (EDA)

"Understand data"

- Basic quaries
- Data profiling
- Simple aggrigations
- Subquery

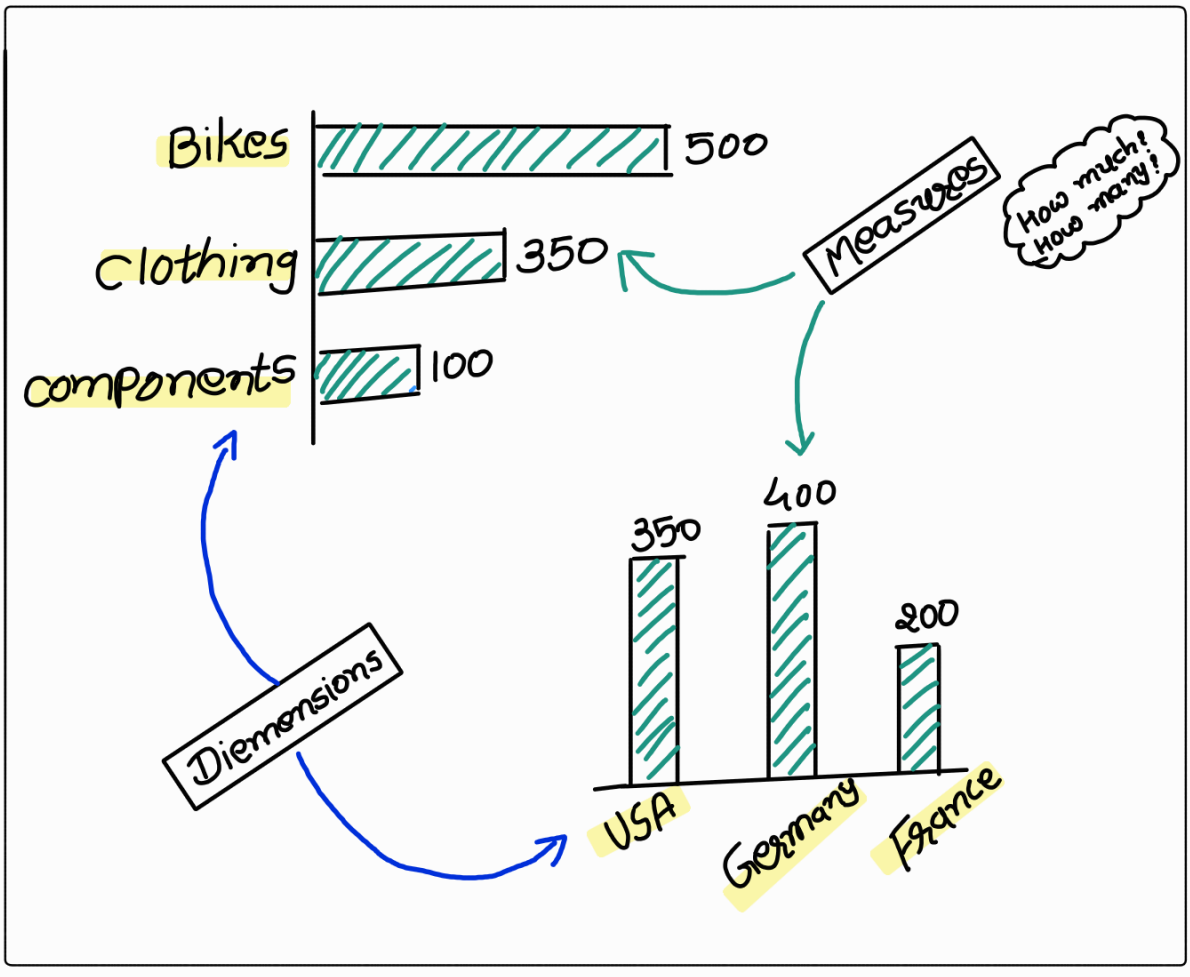
The secret MEASURE & DIMENSION





* Why need Measures & Dimensions?

→ For Grouping up data



---

1. Database Exploration

→ Explore the structure of database for basic Understanding about the tables, views, columns

---

2. Dimension Exploration

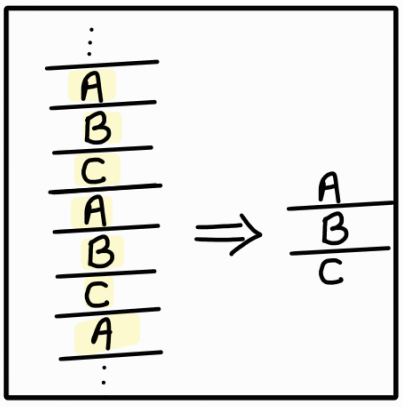
→ Identifying the unique values (or categories) in each dimension.

→ Recognizing How data might be grouped or segmented, which is usuful for later analysis.

→ Formula :-

DISTINCT [Dimension]

Ex. Distinct country
Distinct category
Distinct Product



---

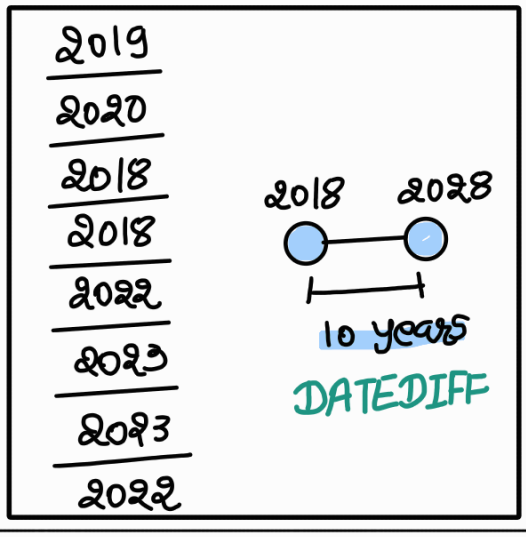
3. Date Exploration

→ Identify the earliest and latest dates (boundries).
→ Understand the scope of data and the timespan.

→ Formula :-

MIN/MAX [Date dimension]

Ex. MIN order_date
MAX Create_date
MIN Birthdate



---

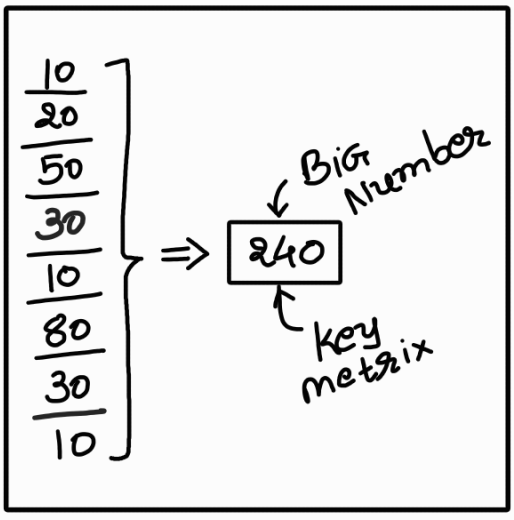
4. Measures Exploration (Big Numbers)

→ Calculate the key metric of the business (Big numbers)
→ Highest level of Aggregation | Lowest level of details

→ Formula :-

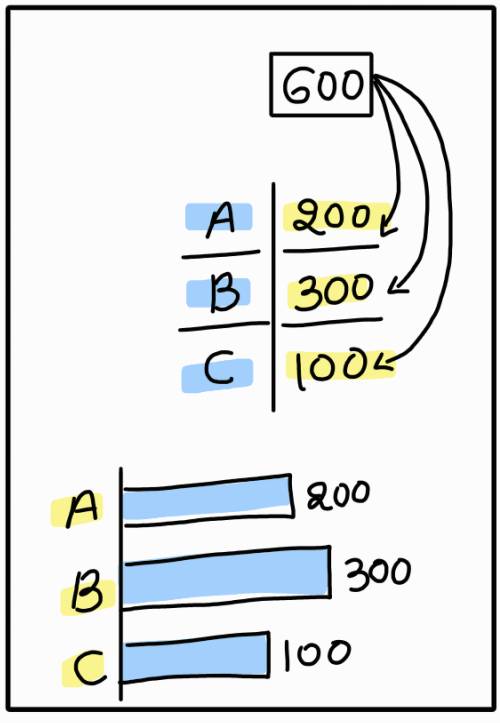
Σ [Measure]

Ex. SUM (sales)
AVG (price)
SUM (quantity)



5. Magnitude

→compare the measure values by categories.
→It helps us understand the importance of different categories.



→Formula:-

$\sum$ [Measure] By [Dimension]

Ex. Total sales by country
Total quantity by category
Total price by product
Total orders by customer

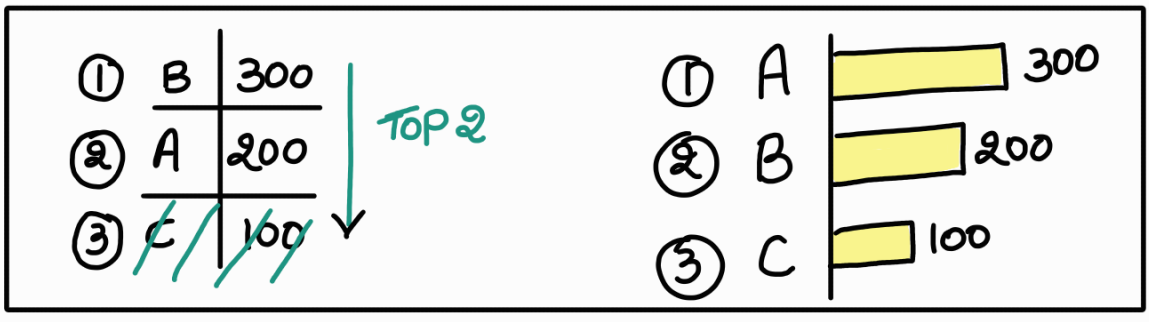
---

6. Ranking Top N-Bottom N

→order the values of dimensions by measures.
→Top N nubers | Bottom N numbers

→Rank [Dimension] By $\sum$ [Measure]

Ex. Rank category by total sales
Top 5 product by Quantity
Bottom 3 Customers By total orders



→For this use:
- TOP
- RANK()
- DENSE-RANK()
- ROW_NUMBER()