

Data set (Data Warehouse)

gold schema

Numeric? Make sense to aggregate? (conditions)

NO

Yes

Dimension

Category (nvarchar)

Product

Birthdate

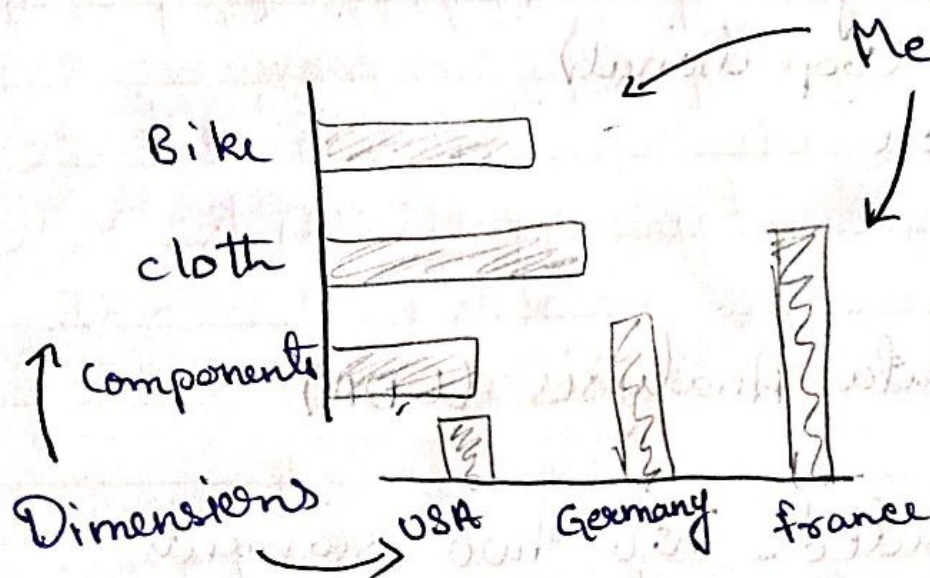
ID

Measure

Sales (int)

Quantity

Age



- # Data Analytics (~~Roadmap~~)
- # Exploratory Data Analysis (EDA) <sup>Steps</sup> (~~Roadmap~~)
- Step 1 - Database Exploration
- 2 - Dimensions Exploration
  - 3 - Date "
  - 4 - Mass Measures " Big Numbers
  - 5 - Magnitude
  - 6 - Ranking, Top N - Bottom N

## # Exploratory Data Analysis (EDA)

~~Step~~ Data set is divided into two segments

Dimension and ~~Mass~~ Measure



## 1. Data Exploration

- Explore all objects in the Database

`select * from information - schema - tables;`

- Explore all columns in the Database

`select * from information - schema - columns.  
where table - name = 'write - table - name'.`

## 2. Dimensions Exploration

- Distinct [Dimension]

Ex:- `Dis-select Distinct [colname] from table-name`

## 3. Date Exploration

- MIN/MAX [Date Dimension]

Ex:- `Select Min(birthdate) as oldest-man,  
Max(birthdate) as youngest-man.`

`From table-name`

## 4. Measures Exploration

- calculate [Measure]

Ex:- `Select Sum/count/Avg (col.name) from table name`



## 5. Magnitude Analysis .

[Measure] By [Dimension]

Ex:- total Sales By Country .

total Quantity By Category .

Select country, count(customer-key) as total-customers  
From customer gold.dim-customers

Group by country

Order by total-customers

## 10. Rank Analysis .

Rank [Dimension] by [Measure]

Ex:- Rank Countries By total sales .

Top 5 products By Quantity .

Select top 5 . p.product-name, sum(f.salesamt) as total-revenue  
From gold.fact-sales as p .

left join gold.dim-products as .

on p.product-key as

on p.Product-key = f.product-key .

Group by . p . product-name

Order by

total-revenue . Desc ,