Problem Statement:

Choose a set of dataset, design star, snowflake and constellation schema on the dataset. And visual analytics using tableau.

Category of the Dataset: Global Terrorism report 2016-20

Objective:

- To learn how to pre-process the data
- To learn how to design schema – star, snow-flake, fact constellation schema.
- To learn how to use the Tableau tool.
- To learn how to integrate the excel sheet with the tableau.
- To learn how to create visualizations in the Tableau tool.

DSS Tool Used: Excel, Power Bi, Tableau. We used Excel here for data Cleaning and pre-processing. Firstly, we uploaded our dataset into excel then performed basic operations, and cleaned this data then we uploaded this dataset in Power Bi to form the schema design.

Power BI is a Data Visualization and Business Intelligence tool that converts data from different data sources to interactive dashboards and BI reports. In Power BI we can load the data from sources like csv, excel files and can transform that data. Here we can easily design different styles of data mart schemas which are most widely used to develop data warehouses.

We used Tableau for data Visualization. Firstly, we load our dataset in tableau using excel sheet and then perform further operation

Dataset Used: We have used the dataset of Global terrorism database for the time period of 2016-20. This is taken from the https://www.start.umd.edu/gtd/

This dataset includes 65 columns and 51,984 rows. This data is about different terrorist attack happen in different parts of world.

- **Preview of Dataset:**

	A	B	C	D	E	F	G	H	I	J	K
1	eventid	year	month	day	approxdate	extended	country	country_txt	region_txt	provstate	city
2	201601060019	2016	1	6	January 6, 2016	0	228	Yemen	Middle East & North Africa	Adan	Unknown
3	201601060021	2016	1	6	January 6, 2016	0	228	Yemen	Middle East & North Africa	Marib	Sirwah
4	201601060022	2016	1	6		0	228	Yemen	Middle East & North Africa	Amanat Al Asimah	Sanaa
5	201601060023	2016	1	6		0	95	Iraq	Middle East & North Africa	Baghdad	Baghdad
6	201601060024	2016	1	6	January 6, 2016	0	228	Yemen	Middle East & North Africa	Taizz	Taizz
7	201601060028	2016	1	6	January 6, 2016	0	228	Yemen	Middle East & North Africa	Taizz	Taizz
8	201601060029	2016	1	6		0	4	Afghanistan	South Asia	Badakhshan	Zard Khak
9	201601060030	2016	1	6		0	95	Iraq	Middle East & North Africa	Baghdad	Baghdad
10	201601060032	2016	1	6		0	92	India	South Asia	Chhattisgarh	Mirtur
11	201601060033	2016	1	6		0	160	Philippines	Southeast Asia	Samar	Bagacay
12	201601060034	2016	1	4		0	37	Cameroon	Sub-Saharan Africa	Extreme-North	Wambache
13	201601060035	2016	1	6	January 6, 2016	0	182	Somalia	Sub-Saharan Africa	Banaadir	Mogadishu

Preprocessing:

First, we take a look at our data set, and deleted unwanted columns which are not important to our analysis. After looking at data provided, we check for a potential field where we can provide reports and visualization of the dataset.

Loading Dataset in Power BI:

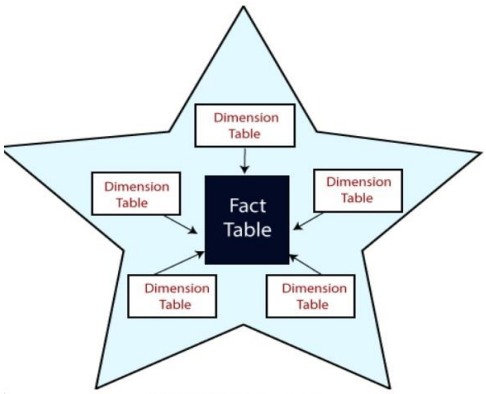
After performing preprocessing on the dataset, we moved towards the next step which is the “Loading” step. We have opened the Power BI tool and there is the Get Data option present in the home tab. Here we have given the terrorism report file that we have to load and after giving the file path it shows a preview of the dataset in the window and we have then clicked on the Load button to load the dataset into Power BI. After this we have successfully loaded the Dataset in Power BI Tool.

Now we load all other data in the same manner. The data we are concerned with here are customer data, product data , calendar and territories. These tables were selected because our fact table contains fields from the above mentioned tables as dimension attributes.

Then we clicked on edit query for input data

Schema Designing:

1) Star Schema:

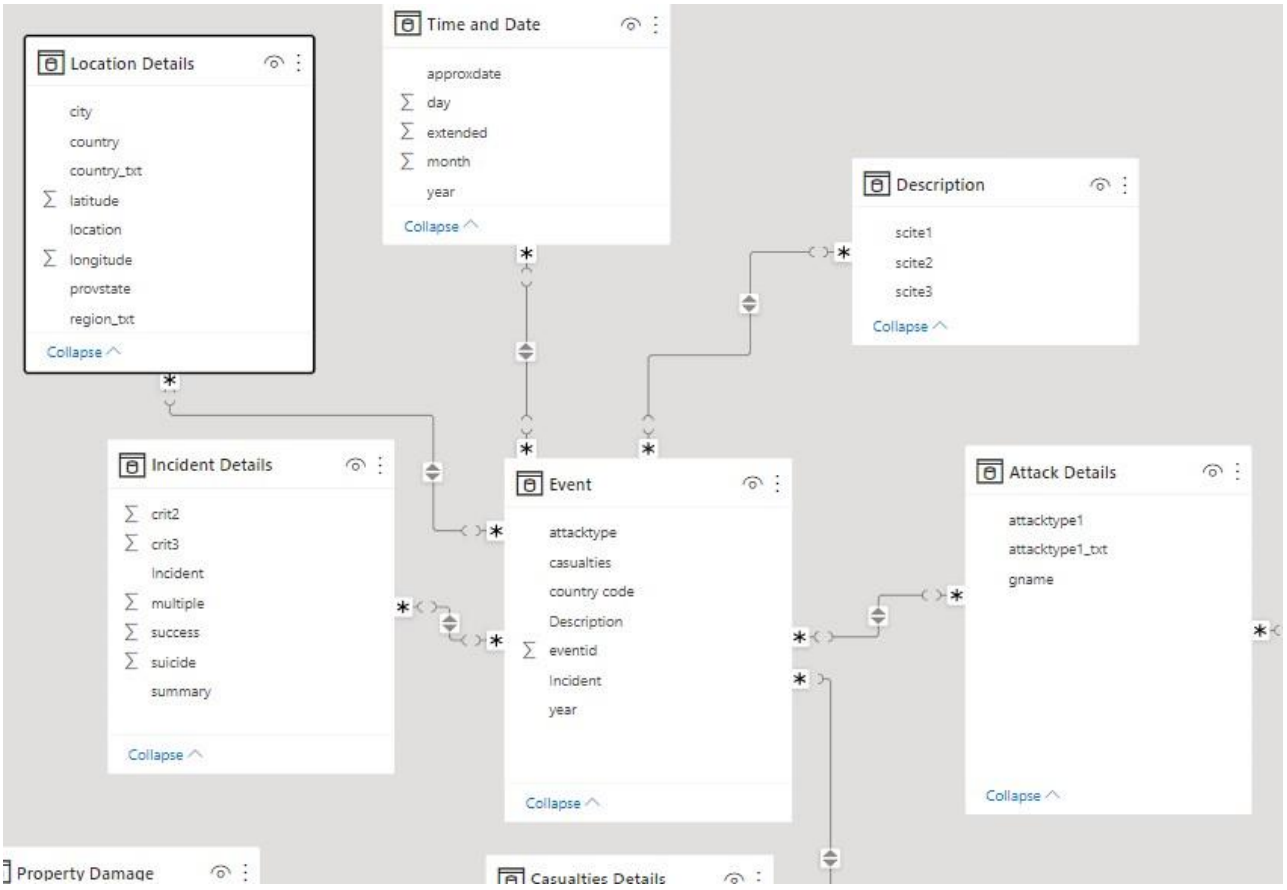


Star Schema

A star schema is the elementary form of a dimensional model, in which data are organized into facts and dimensions. A fact is an event that is counted or measured, such as a sale or log in. A dimension includes reference data about the fact, such as date, item, or customer. A star schema is a relational schema where a relational schema whose design represents a multidimensional data model.

The star schema is the explicit data warehouse schema. It is known as star schema because the entity-relationship diagram of this schema simulates a star, with points, divergent from a central table. The centre of the schema consists of a large fact table, and the points of the star are the dimension tables.

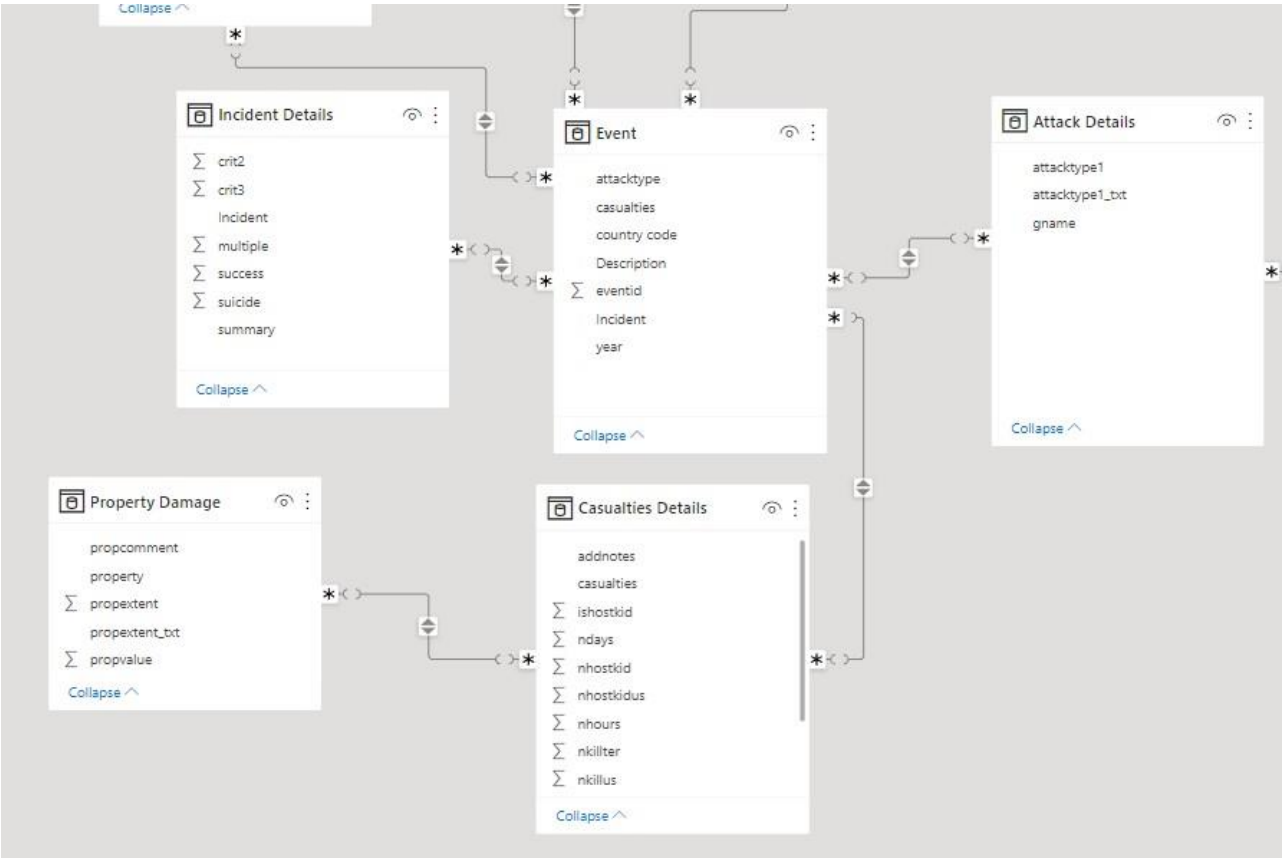
This how the final star schema looks like for our data set:



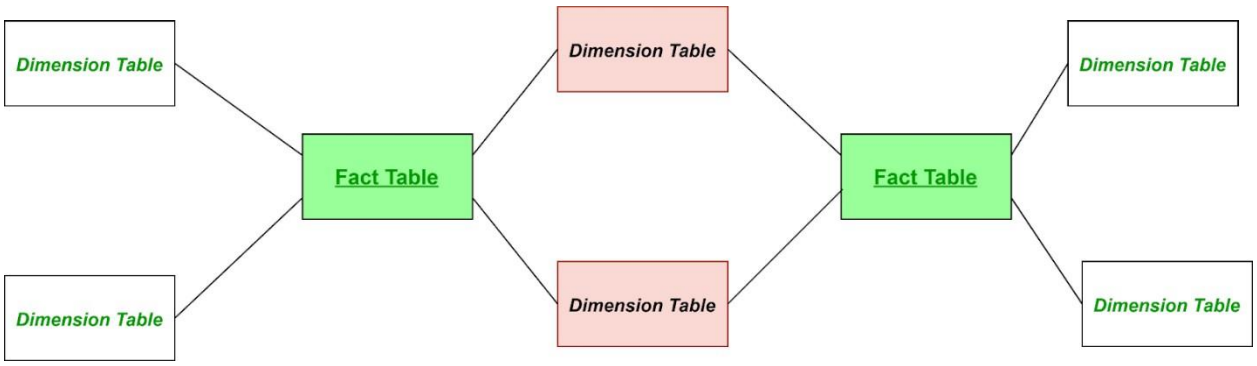
2] Snowflake Schema:

A snowflake schema is equivalent to the star schema. "A schema is known as a snowflake if one or more-dimension tables do not connect directly to the fact table but must join through other dimension tables."

This is how it would look:

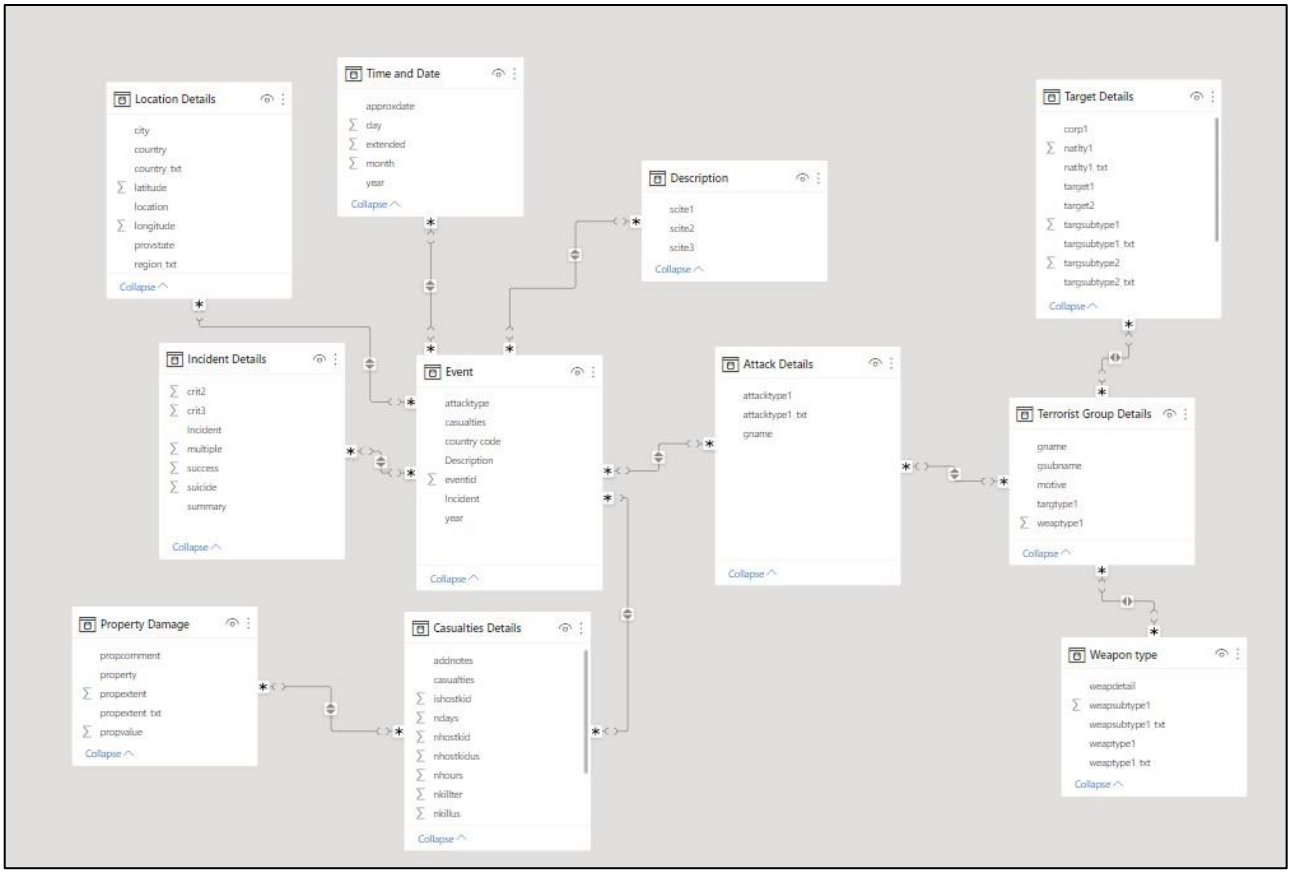


3) Fact Constellation



A Fact constellation means two or more fact tables sharing one or more dimensions. It is also called Galaxy schema.
Fact Constellation Schema describes a logical structure of a data warehouse or data mart. Fact Constellation Schema can be designed with a collection of denormalized FACT, Shared, and Conformed Dimension tables. Fact Constellation Schema is a sophisticated database design that is difficult to summarize information. Fact Constellation Schema can implement between aggregate Fact tables or decompose a complex Fact table into independent simplex Fact tables. According to the files there is another fact table which is the AW_Returns Table. To add this table into our schema we need to update our schema from snowflake to fact constellation schema.

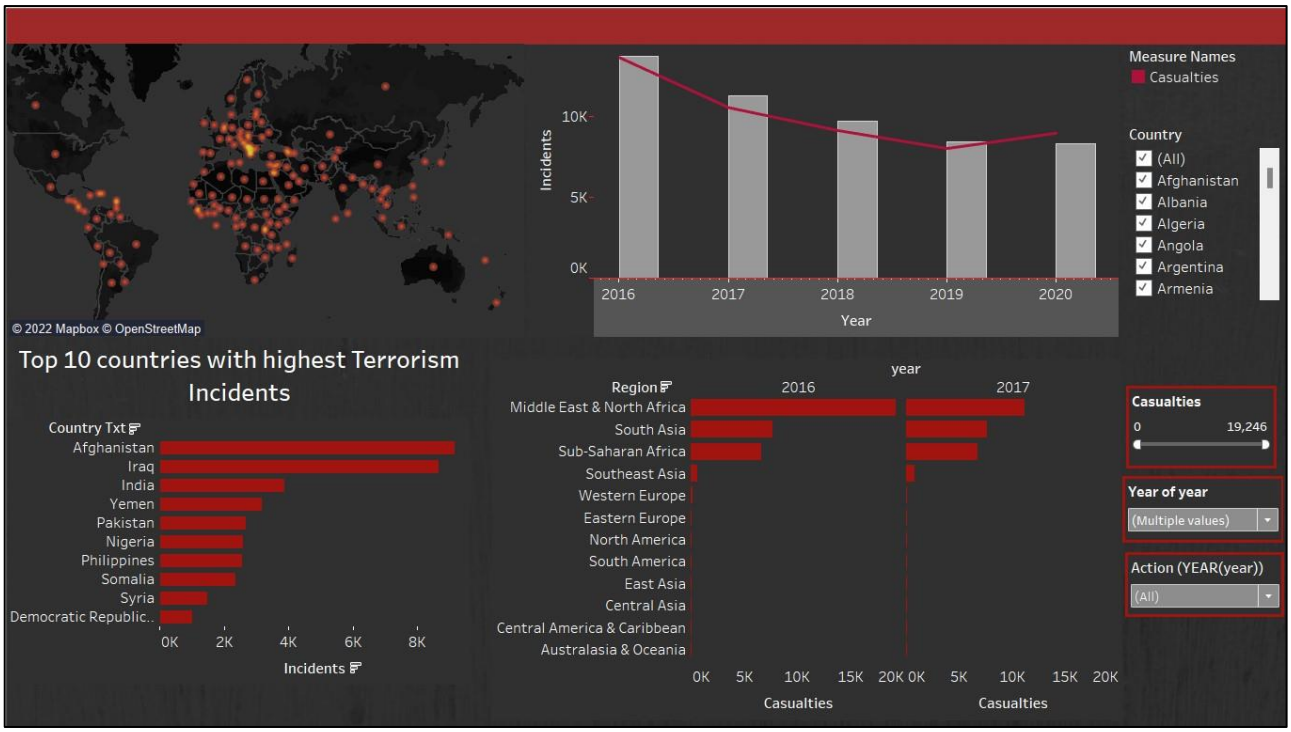
This is how the final schema will look:



• Visualization and Dashboards:

After Schemas We have implemented visualization tools on the same datasets. We can filter these visualizations according to

Take a look at this:



Charts used for Visualization:

1. Horizontal bar Chart (used for showing top 10 countries suffering from terrorism)
2. Map Chart (used to plot distribution of Terrorism per states)
3. Dual Combination Chart (used for visualization of Casualties per year)
4. Lines Chart (used for showing terrorism per region

Conclusion:

We have successfully visualized the data with a dataset (size: 21 MB) and drew useful insights from it.

From the above shown data we first pre-processed our data to load it into our Power BI. After pre-processing we created schemas. First, we made the first level schema which is the Star Schema. Then we converted that schema to Snowflake Schema to further connect our tables and then finally all our tables are connected by the Fact Constellation Schema. We also Successfully implemented visualization tools on Tableau.