

Bansilal Ramnath Agarwal Charitable Trust’s

# Vishwakarma Institute of Information Technology

Business Intelligence and Data Analytics

(Assignments 2)

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Roll No. | GR No. | Batch |
| Chetan Chinchulkar | 324017 | 21810350 | E9 (Comp D) |
| Nikit Gokhe | 324022 | 21810522 | E9 (Comp D) |
| Sahil Langha | 324032 | 21810533 | E9 (Comp D) |
| Kunal Parkhe | 324040 | 21810049 | E9 (Comp D) |

GROUP MEMBERS:

# Assignment No. 02

# Aim:

To design and understand star schema, snow flake schema and fact constellation schema. Also to get familiar with tools like **Power BI/Rapid Miner.**

# Objective:

Choose a set of business processes like Sales, Customer Services, Accounting, Production, Marketing processes etc. for any organization and design star, snow flake and fact constellation schema. Also using ETL tool, extract data from various sources and perform transform and load operations on data. **(Use Power BI/Rapid Miner)**

# Theory:

1. **Star Schema:**

* Each dimension in a star schema is represented with only one-dimension table.
* This dimension table contains the set of attributes.
* There is a fact table at the center. It contains the keys to each of dimensions.
* The fact table also contains the attributes, namely dollars sold and units sold.

1. **Snowflake Schema:**

* Some dimension tables in the Snowflake schema are normalized.
* The normalization splits up the data into additional tables.
* Unlike Star schema, the dimensions table in a snowflake schema are normalized.

**Note** − Due to normalization in the Snowflake schema, the redundancy is reduced and therefore, it becomes easy to maintain and the save storage space.

1. **Fact Constellation Schema:**

* A fact constellation has multiple fact tables. It is also known as galaxy schema.
* The schema is viewed as a collection of stars hence the name Galaxy Schema.
* The dimensions in this schema are separated into separate dimensions based on the various levels of hierarchy.
* it is possible to build this type of schema by splitting the one-star schema into more Star schemes.
* The dimensions are large in this schema which is needed to build based on the levels of hierarchy.
* This schema is helpful for aggregating fact tables for better understanding.

# Input:

**Name:** Superstore

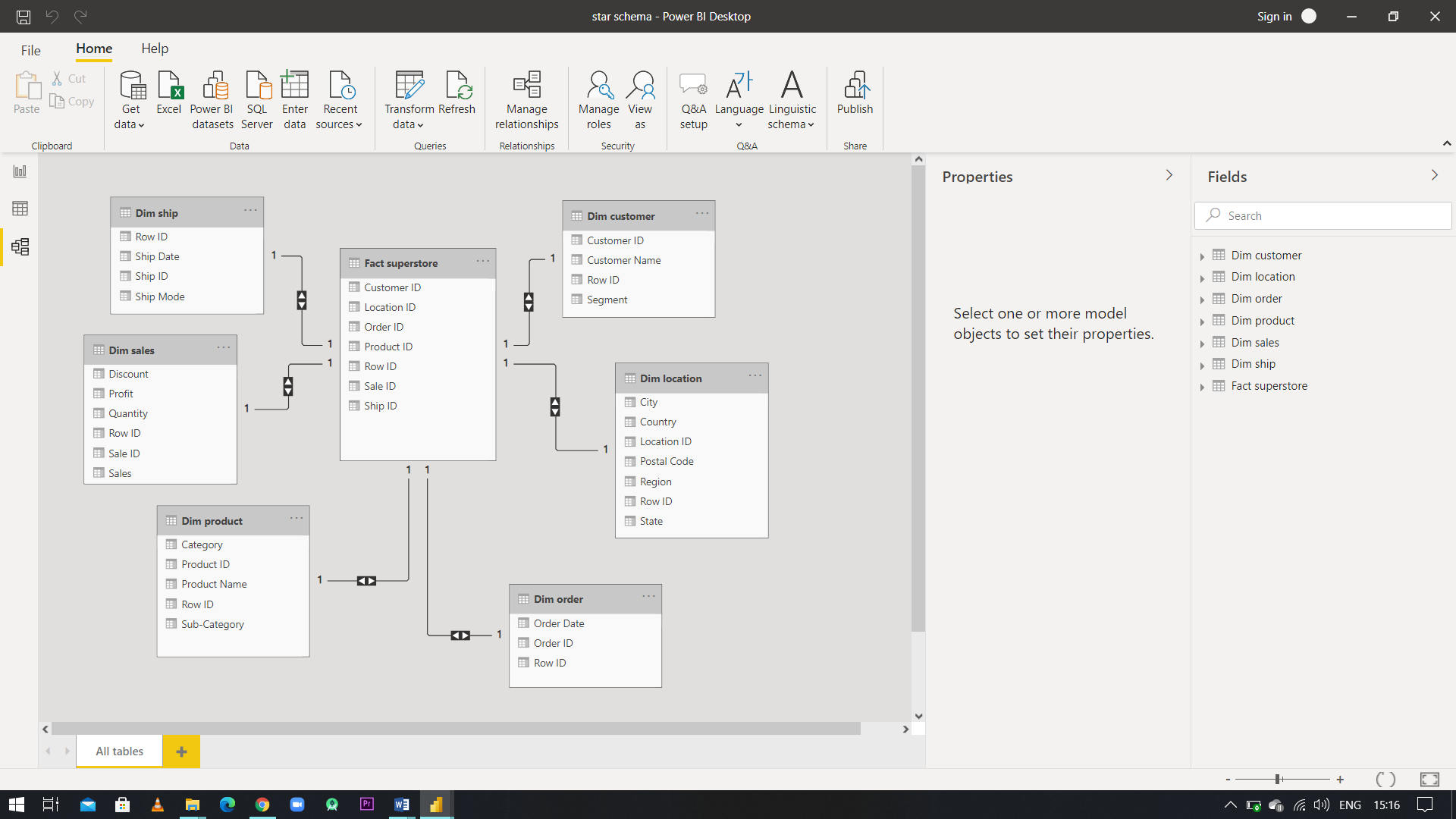
**Dataset Reference:** kaggle

**Link:**

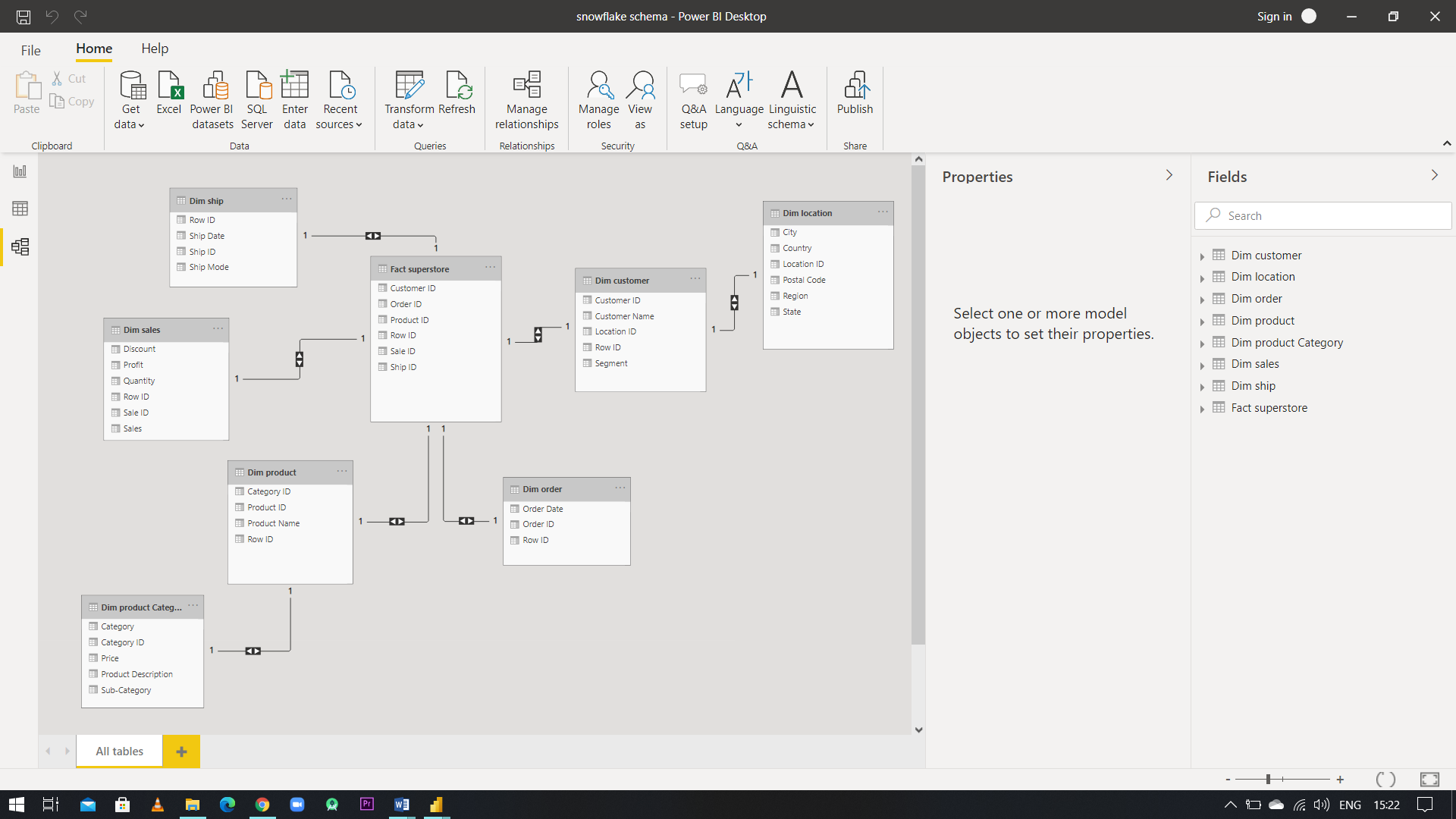
https://www.kaggle.com/keyizhang14/superstore

# Output:

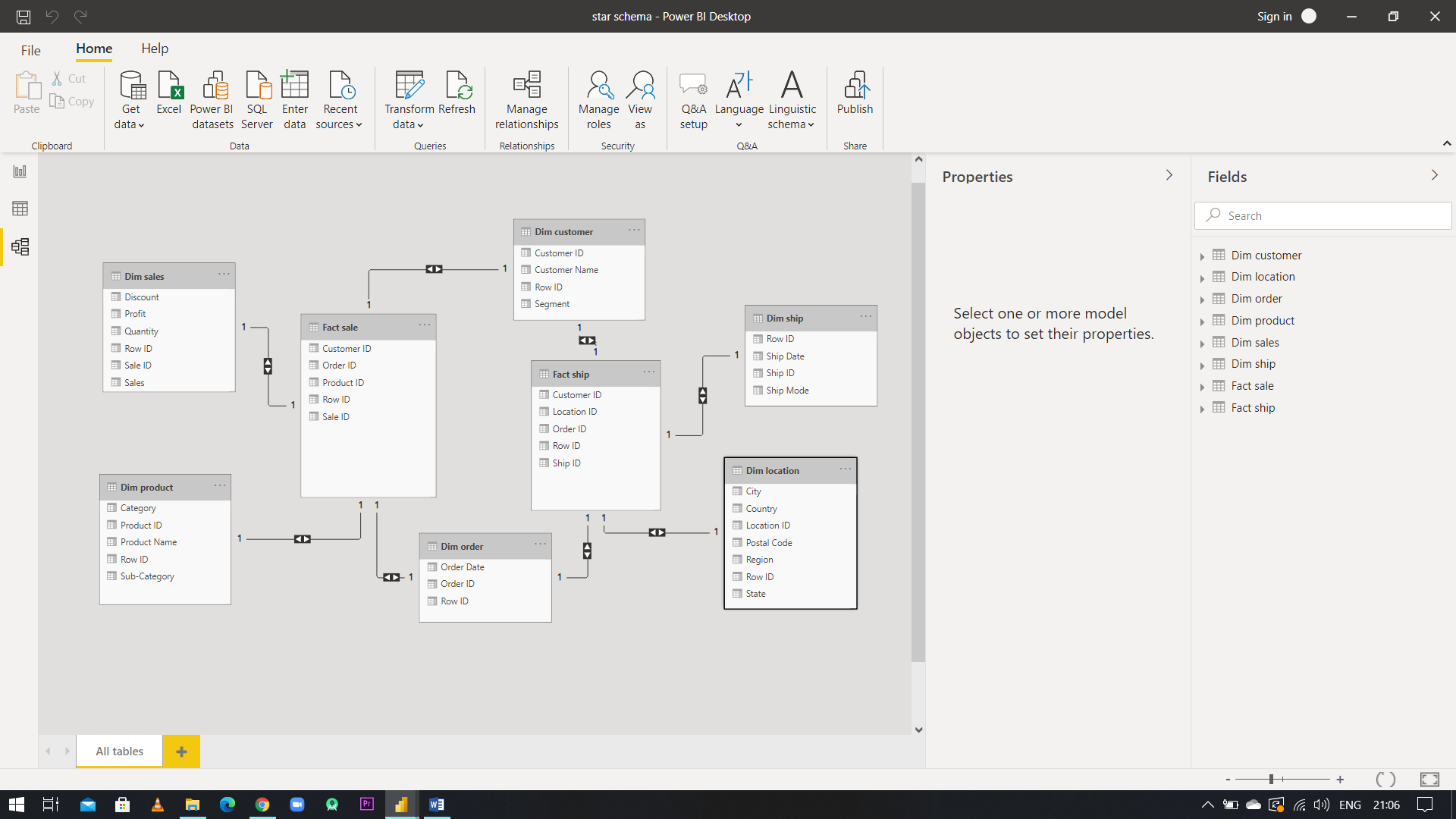
Star Schema:



Snowflake Schema:



Fact Constellation Schema:



# Conclusion:

Understood and Implemented Data Warehouse Schemas i.e.

* 1. Star Schema
  2. Snowflake Schema
  3. Fact Constellation / Galaxy Schema

On Sales dataset. Also used ETL tool and got familiar with UI and functions of Power BI.