Data Engineering-Final Project

Part-2

Creating database, schema and checking if it exists.

```
--Create a Database for the Term Project
CREATE OR REPLACE DATABASE DE_Final;
--Create a Schema within the above DB for the Term Project
CREATE OR REPLACE SCHEMA DE_Final.STAGE_DATA;
-- Checking for schema if not ezist.
CREATE SCHEMA IF NOT EXISTS STAGE_DATA;
--so the schema exists.
--Now we will be making the dimension tables,

Results

**Chart*

STAGE_DATA already exists, statement succeeded.
```

-Create 3 Dimension Tables in your Stage Schema in Snowflake.

• Customer dim

```
10
         CREATE TABLE Customer_Dim (
 11
            CustomerID INT PRIMARY KEY,
 12
             FirstName VARCHAR(255),
 13
            LastName VARCHAR(255),
 14
             Email VARCHAR(255),
 15
             Phone VARCHAR (20),
 16
            CustAddress VARCHAR(255),
 17
            CustState VARCHAR(50),
 18
            CustCity VARCHAR(50),
 19
            CustZip VARCHAR(10));
 20
→ Results

✓ Chart

   status
   Table CUSTOMER_DIM successfully created.
```

Product_Dim-

```
21
         --creating product dimension table.
 22
         CREATE TABLE Product_Dim (
 23
             ProductID INT PRIMARY KEY,
 24
             Brand VARCHAR (255),
 25
             Name VARCHAR (255),
 26
             Description TEXT,
 27
             Price DECIMAL(10, 2),
             MSRP DECIMAL(10, 2),
 28
 29
             Cost DECIMAL(10, 2)
 30
 31
→ Results

✓ Chart

   status
   Table PRODUCT_DIM successfully created.
```

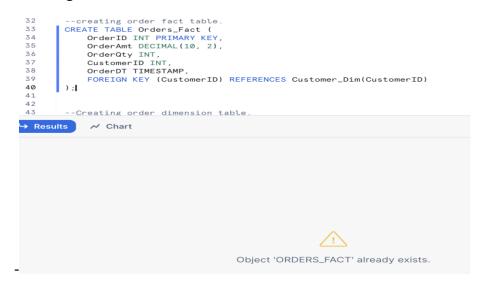
Order_dim-

```
--Creating order dimension table.
44
       CREATE TABLE OrderDetail_Dim (
45
            OrderDetailID INT PRIMARY KEY,
46
            ProductID INT,
47
            OrderID INT,
48
            FOREIGN KEY (ProductID) REFERENCES Product_Dim(ProductID),
49
            FOREIGN KEY (OrderID) REFERENCES Orders_Fact(OrderID)
50
51
53
       --reading the Customer_Dim table data loaded into the snowflake
→ Results

→ Chart

  status
  Table ORDERDETAIL_DIM successfully created.
```

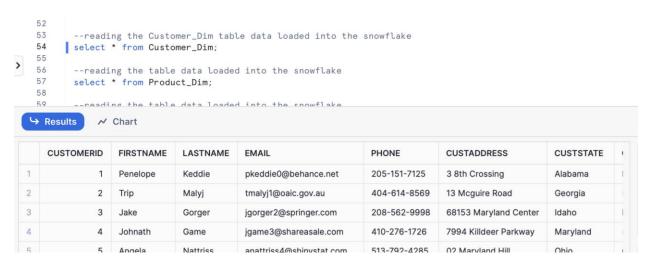
Creating Fact Table-



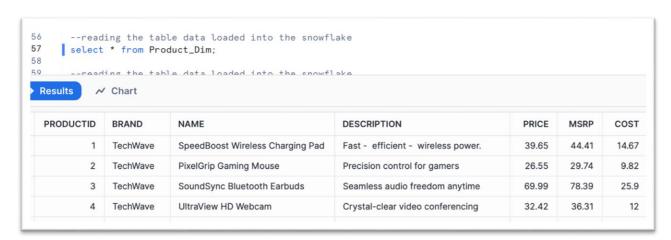
Performing Data Cleaning and transformation.-

Reading data from all the tables

Customer table



Product_Dim



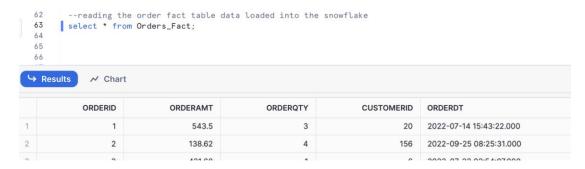
Order_Detail_Dim

```
--reading the table data loaded into the snowflake
60
     select * from OrderDetail_Dim;
61
62
       --reading the order fact table data loaded into the snowflake
63
       select * from Orders_Fact;
64
65
→ Results

✓ Chart

                           ORDERDETAILID
                                                                    PRODUCTID
                                                                                                          ORDERID
                                                                           213
                                                                                                               16
                                       2
                                                                           60
                                                                                                                6
```

Fact Table



Checking for Null values in Customer_Dim.

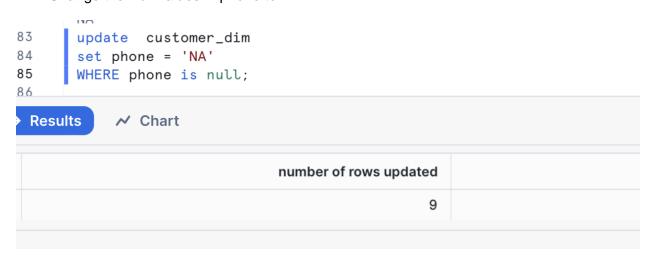
```
67
       --Checking for missing values in customer_dim
68
       -- Customer_Dim
69
      SELECT
70
           COUNT(*) AS total_rows,
71
           COUNT(CustomerID) AS customer_id_count,
72
           COUNT(FirstName) AS first_name_count,
73
           COUNT(LASTNAME) AS last_name_count,
74
           COUNT(EMAIL) AS email_count,
75
           COUNT(PHONE) AS Phone_count,
76
           COUNT(CUSTADDRESS) AS Address_count,
77
           COUNT(CUSTSTATE) AS custstate_count,
78
           COUNT(custcity) AS city_count,
79
           COUNT(custzip) AS zip_count,
80
      FROM Customer_Dim;
81
 Results

✓ Chart

  TOTAL ROWS
               CUSTOMER_ID_COUNT FIRST_NAME_COUNT LAST_NAME_COUNT
                                                                          EMAIL COUNT
                                                                                        PHONE COUNT
                                                                                                        ΑI
         200
                              200
                                                 200
                                                                    200
                                                                                   200
                                                                                                  191
```

Only the customer_dim table has null values in 5 columns. We will be doing more analysis and data cleaning on customer_dim. --sales fact table, Orderdetail table, productdim does not have any null values

I will Change the Null values in phone to NA.



Checking for Duplicates in Customer Dim.



Deleting the Duplicates in Customer.



· Checking if the row is deleted. Query Produced no result.

```
99
       -- checking if the customer_id is deleted
00
      select * from customer_dim
01
      where customerid = '100';
02
03
       --we see that certain states have abbreviations in custstate column, this needs to be changed to full r
 Results

→ Chart

     CUSTOMERID FIRSTNAME
                                 LASTNAME
                                               EMAIL
                                                          PHONE
                                                                     CUSTADDRESS
                                                                                       CUSTSTATE
                                                                                                      CUSTCITY
```

Outers produced no recults

Changing the state abbreviations to full names-

```
--updating state abbreviations to full name, there are 3 states which needs to be updated update customer_dim set custstate = 'Texas'
WHERE custstate = 'TX';

--updating for AZ to Arizona update customer_dim set custstate = 'Arizona'
WHERE custstate = 'AZ';

update customer_dim set custstate = 'California'
WHERE custstate = 'California'
WHERE custstate = 'CA';
```

Quality Check-

```
155
       --looking for duplicates on customer_dim
156
       -- Customer_Dim
157
      SELECT
158
          CustomerID, FirstName, LastName, Email, Phone, CustAddress, CustState, CustCity, CustZip,
159
          COUNT (*)
160
       FROM Customer_Dim
161
       GROUP BY CustomerID, FirstName, LastName, Email, Phone, CustAddress, CustState, CustCity, CustZip
162
       HAVING COUNT(*) > 1;
163
→ Results

✓ Chart

      CUSTOMERID FIRSTNAME
                             LASTNAME
                                         EMAIL
                                                  PHONE
                                                           CUSTADDRESS
                                                                         CUSTSTATE
                                                                                      CUSTCITY
TOO
164
        --looking for duplicates on Order_fact_dim
165
        -- Order_Fact Dim
166
       SELECT
167
           OrderDetailID, ProductID, OrderID,
168
            COUNT (*)
169
        FROM OrderDetail_Dim
170
        GROUP BY OrderDetailID, ProductID, OrderID
171
        HAVING COUNT(*) > 1;
172
→ Results

✓ Chart

                     ORDERDETAILID
                                                      PRODUCTID
                                                                                    ORDERID
                                                Query produced no results
    --looking for duplicates on Product_dim
    -- Product Dim
    SELECT
       ProductID, Brand, Name, Description, Price, MSRP,
    Cost,
         COUNT (*)
    FROM PRODUCT_DIM
    GROUP BY ProductID, Brand, Name, Description, Price, MSRP,
    Cost
    HAVING COUNT(*) > 1;
sults

✓ Chart
```

Query produced no results

PRICE

DESCRIPTION

BRAND

NAME

PRODUCTID

Creating Customer Dim in Prod table

```
191
        --Creating Customer_Fact_Table for Prod_DE_Final.
192
        CREATE TABLE IF NOT EXISTS CustomerDim (
193
            CustomerID INT PRIMARY KEY,
194
            FirstName VARCHAR(255),
195
            LastName VARCHAR(255),
196
            Email VARCHAR(255),
197
            Phone VARCHAR(20),
198
            CustAddress VARCHAR(255),
199
            CustState VARCHAR(50),
200
            CustCity VARCHAR(50),
201
            CustZip VARCHAR(10),
202
            RecordInserted TIMESTAMP DEFAULT CURRENT_TIMESTAMP(),
203
            RecordSource VARCHAR(255)
        );
204
→ Results

✓ Chart

   status
   CUSTOMERDIM already exists, statement succeeded.
```

Inserting data from staged to prod.

```
206
       --Insertinmg data from staged to prod in customer dim
207
       USE DATABASE DE_Final;
208
       INSERT INTO Prod_DE_Final.CustomerDim (CUSTOMERID, FIRSTNAME, LASTNAME, EMAIL, PHONE, CUSTADDRESS, CUSTSTATE,
       CUSTCITY, CUSTZIP
210
211
       SELECT CUSTOMERID, FIRSTNAME, LASTNAME, EMAIL, PHONE, CUSTADDRESS, CUSTCITY, CUSTSTATE, CUSTZIP
212
       FROM STAGE_DATA.CUSTOMER_DIM;
213
Results

✓ Chart

                                                                                               number of rows inserted
                                                                                              Query Details
                                                                                              Query duration
```

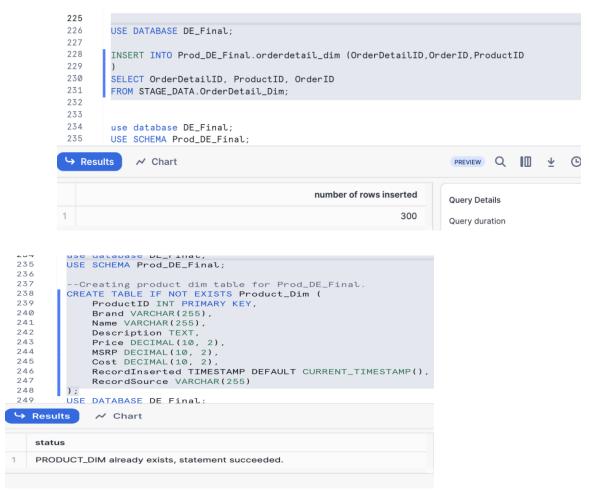
Creating order detail dim in prod.

```
213
214
        --creating orderdetail dim for Prod_DE_Final.
215
        CREATE OR REPLACE TABLE orderdetail_dim (
216
        OrderDetailID number(19,0) not null,
217
        OrderID number(19,0) not null,
218
        ProductID number (19,0) not null,
219
        copy_date TIMESTAMP NOT NULL,
220
        primary key (OrderDetailID)
221
        );
222
223

✓ Chart

→ Results
   status
   Table ORDERDETAIL_DIM successfully created.
```

Inserting Data from staged to prod in order detail dim.



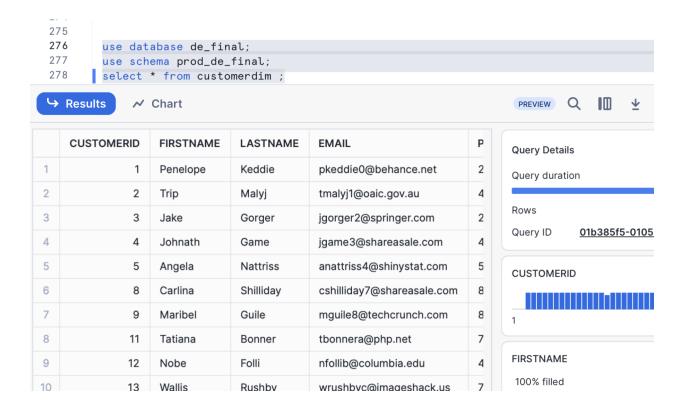
Creating Product Dim table and inserting data into it from staged.,

```
237
238
239
          --Creating product dim table for Prod_DE_Final.
CREATE TABLE IF NOT EXISTS Product_Dim (
ProductID INT PRIMARY KEY,
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
                Brand VARCHAR(255),
               Name VARCHAR(255),
               Description TEXT,
               Price DECIMAL(10, 2).
               MSRP DECIMAL(10, 2),
Cost DECIMAL(10, 2),
                RecordInserted TIMESTAMP DEFAULT CURRENT_TIMESTAMP(),
               RecordSource VARCHAR (255)
         USE DATABASE DE_Final;
         INSERT INTO Prod_DE_Final.Product_Dim (ProductID,Brand,Name,Description,Price,MSRP,Cost
          SELECT ProductID, Brand, Name, Description, Price, MSRP, Cost
          FROM STAGE_DATA.Product_Dim
                                                                                            PREVIEW Q III ± © □

✓ Chart

                                                           number of rows inserted
                                                                                           Query Details
                                                                                           Query duration
                                                                                                                            686ms
                                                                                           Rows
                                                                                           Query ID
                                                                                                          01b385f2-0105-1033-0...
```

• Selecting data from customerdim in the prod schema.



```
281
        use database de_final;
282
        use schema prod_de_final;
283
        SELECT
284
            COUNT(*) AS total_rows,
285
            COUNT(CustomerID) AS customer_id_count,
286
            COUNT(FirstName) AS first_name_count,
287
            COUNT(LASTNAME) AS last_name_count,
288
            COUNT(EMAIL) AS email_count,
289
            COUNT(PHONE) AS Phone_count,
290
            COUNT(CUSTADDRESS) AS Address_count,
291
            COUNT(CUSTSTATE) AS custstate_count,
292
            COUNT(custcity) AS city_count,
→ Results

✓ Chart

                                                          LAST_NAM
   TOTAL_ROWS
                 CUSTOMER_ID_COUNT
                                      FIRST_NAME_COUNT
           398
                                398
                                                    398
```

Successful Connection to Tableau.

