Data warehousing and business intelligence project report

Topic: Analysis of sales for a retail store

By:

Sankalp Ram Saoji

X17154171

MSc. In Data analytics

School of Computing

National college of Ireland

Contents

Introduction:

In this world of competition, it is very difficult for a retail business to sustain in the market and move forward towards the goal of acquiring maximum customers and maximizing the profit. Therefore, there's lot of room for analyzing the consumer needs and their trend. This project focuses on the analysis of sales of different categories of products for a retail superstore, with respect to their customer data and considering different store locations.

This project is mainly based on the concepts of data warehouse and business intelligence queries which uses the data collected and efficiently performs analysis of sales for the superstore. In this project we are also forecasting the sales with consideration of customer choices and store locations.

The project was implemented with various excellent tools from Microsoft such as SQL server management studio (SSMS), SQL server Integration Services (SSIS), SQL Server Analysis Services (SSAS). We will proceed to discuss the sources of data, followed by the extract, transform, load (ETL) process, how we deployed the cube and then final outcomes which are the business intelligence queries.

Sources of data:

Choice of data plays a crucial role in implementation of the data warehouse project. The implementation of project is made up of four data sources of which three are structured datasets and one is an unstructured dataset.

Structured data:

1. This project uses a structured dataset extracted from Tableau community website mentioned below, the data consists of the different categories, sub categories and names of products sold. The data at which product is sold is also mentioned. The data also contains the measures such as what quantity sold, profit, sales in dollars.
   https://public.tableau.com/s/resources?qt-overview_resources=1

2. Second data used was downloaded from dunnhumby website from which we obtained the data about the customers who purchased in the superstore. From this dataset we got customer's names, age and their income.
   https://www.dunnhumby.com/sourcefiles

3. We have used mockaroo website and created a customized dataset for the superstore which contains store number and its location. Location of the store data contains its city, state, region, country and postal code.
   https://mockaroo.com/

Unstructured data:

1. We have extracted the unstructured data from the twitter tweets. We have used the Twitter API to extract data. We have performed sentiment analysis on Twitter using the R code in R studio. We have extracted sentiment of the customers on the different categories of products. Sentiment analysis contains positive, negative and neutral counts. Finally, these sentiments from twitter are converted to CSV file format.

Assumption while using data source:

- The three structured data sources that are store, product and customer datasets are assumed to be of a franchised retail superstore.
- The unstructured data from twitter sentimental analysis is scrapped from the most recent twitter posts and are fully independent from historical or previous years data.

Data warehouse architecture:

We have studied about the two main approaches of a data warehouse first is the “Ralph Kimball” approach which is a bottom up approach and second is the “Bill Inmon” approach which is a top down approach. In our project we have used the Kimball approach for the implementation of the data warehouse.

Advantages of Kimball approach over Inmon approach:

- Does not require data to be in normalized form.
- It is relatively faster than the Inmon’s top down approach.
- Kimball’s approach requires less space and is a low maintenance data warehouse.
- Implementation of data warehouse is simple.
- Star schema can be used in Kimball’s approach.

Implementation of Kimball approach in our data warehouse architecture:

In this project different dimension data marts were designed as per the business requirements and sources of data for the data warehouse. Data mart refers to a single business department while the data warehouse has a wider scope. Using the Kimball approach the different data marts were created for store, product, customer, date with the use of raw data from the different sources mention above. Dimension tables and fact table were generated and populated using the Kimball's approach. Star schema was set in the data warehouse for analysis of product sales. Star schema was used because it is very feasible for our business process and is easy to implement. Star schema is efficient for our analysis of product sales as it uses less time to execute and has simple design (Kimball & Ross, 2013).

Technologies used in data warehouse architectures:

- SSIS is used for integration and loading the dimension table and fact table to the SQL server.
- SSAS is used for building and deploying the cube.
- SSMS is the SQL server where data is stored.
- SQL command were used for creating and dimension tables and populating them.
- R studio was used to extract unstructured data from Twitter using twitter API and the R code.
- Microsoft excel was used to clean the structured data, removing the NULL values and special characters were done using excel.
- Tableau is a visualization software used to analyze and visualize our business intelligence queries.

Data warehouse data model:

Dimensional design process for data warehouse:

1. Select the business process:

Business process is a low-level activity performed by an organization, for this project the business process is to analyze the sales of product at different store locations. Further we will drill down to different categories of product and we also emphasis on forecast on sales values in the year 2020. Twitter sentiments are also compared and analyzed in our business process. This is how Kimball approach was used to select the business process (Kimball & Ross, 2013).

2. Declare the grain:

Kimball approaches states that declaring the grain is a very crucial step in the dimensional design process. Most of the dimension modeler try to evade this step as they seem this as an unnecessary step. In our project the primary key of the fact table will we having unique values and these values will be associated with the foreign key values of our dimensions that are store, product, customer, date and sentiment. Each dimension has its own unique value primary key (Kimball & Ross, 2013).

3. Identify the dimensions:

All the descriptive data for a business process is stored in the dimension tables. Dimension table should be robust. Dimension table answers "who, what, where, when, why, how" associated with the business event. In our business process dimension are store dimension which tells us where the stored is located. Product dimension tell us about what are the different products sold. Dimension customer tell us about who are customers their name, age, income. Date dimension is used to find when is the product sold. We have a dimension sentiment which stores the positive, negative and neutral sentiments from Twitter (Kimball & Ross, 2013).

4. Identify the facts:

Kimball approach states fact table consist values which tells us all about the measures of the business process. Typically, fact table are the additive figures. Example are cost in dollars, quantity sold. Considering the business process and the sources of data we have selected fact table which contains foreign keys of all dimension tables and measures such as sales in dollars, profit earned, quantity sold and discount applied (Kimball & Ross, 2013).

Benefits of star schema in our project:

- No need for normalization, dimension tables need not to be in normalized form.
- Star schema queries are simple in contrast snowflake schema queries are complex for our sales business process.
- Star schema is implemented as there is no dependency between the dimension tables.
- It simplifies the future actions for a business user to analyze business intelligence queries.
- Star schema performance is faster as it has less joins between the fact table and dimension tables.

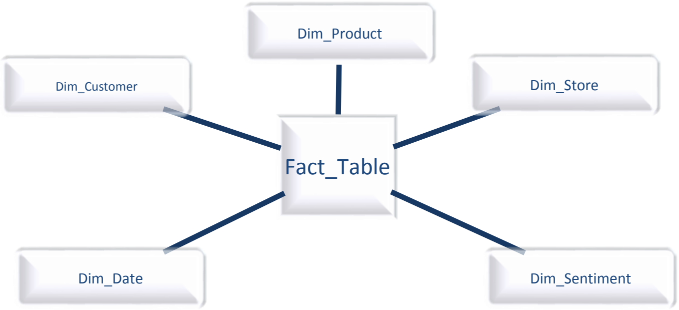


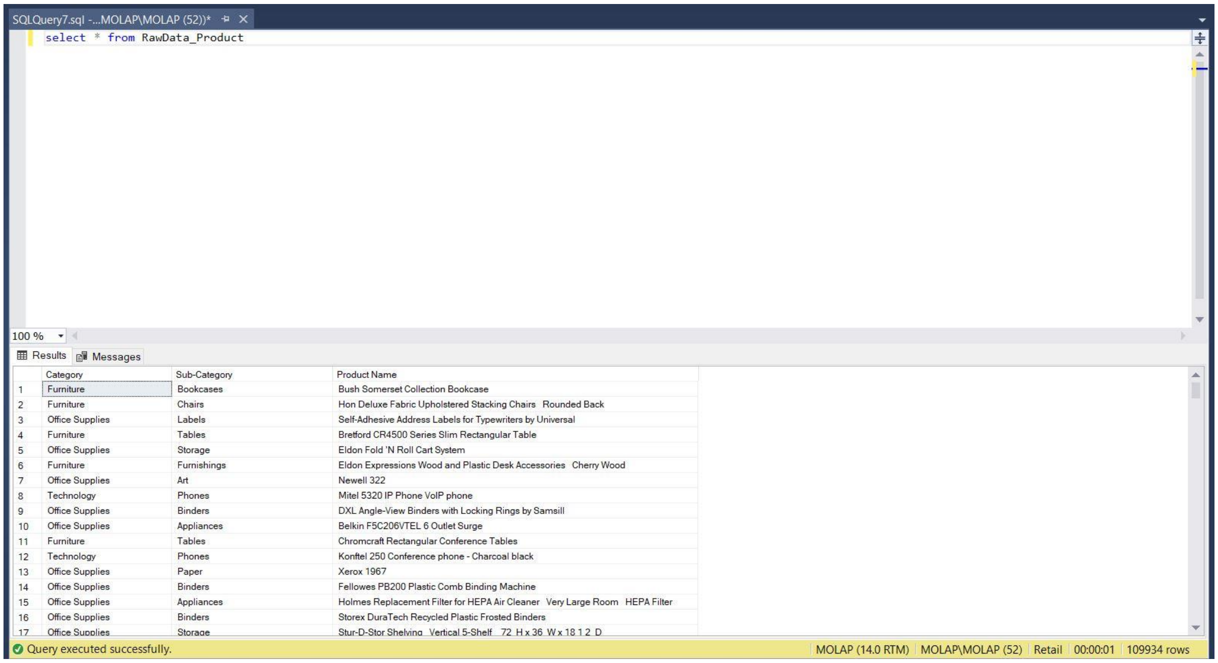
Fig. 1 Star schema model for this project

This is the star schema for our project which consists of one fact table and five-dimension tables that are for store, product, customer, date, sentiment.

Importing raw data and creating dimension tables:

The raw data is imported using a data flow task in SSIS and then using the flat file import the data is fetched and a raw data table is created using a create SQL command. Figure 3,4,5 and 6 shows the raw data from multiple data sources for developing and implementing a data warehouse. From this raw data tables, data is transformed, cleaned, primary keys are set and the data is populated in the respective dimension tables. Below are the screenshots of the data which is loaded in the SQL server (SSMS) by the ETL process performed in SSIS. In the below screenshots we can see the preview of data populated in the respective dimension table from the multiple sources. The screenshots below are of only few dimension tables in our data warehouse project. The dimension tables were created using the create SQL query and were populated using the insert SQL query.

Raw data for product:



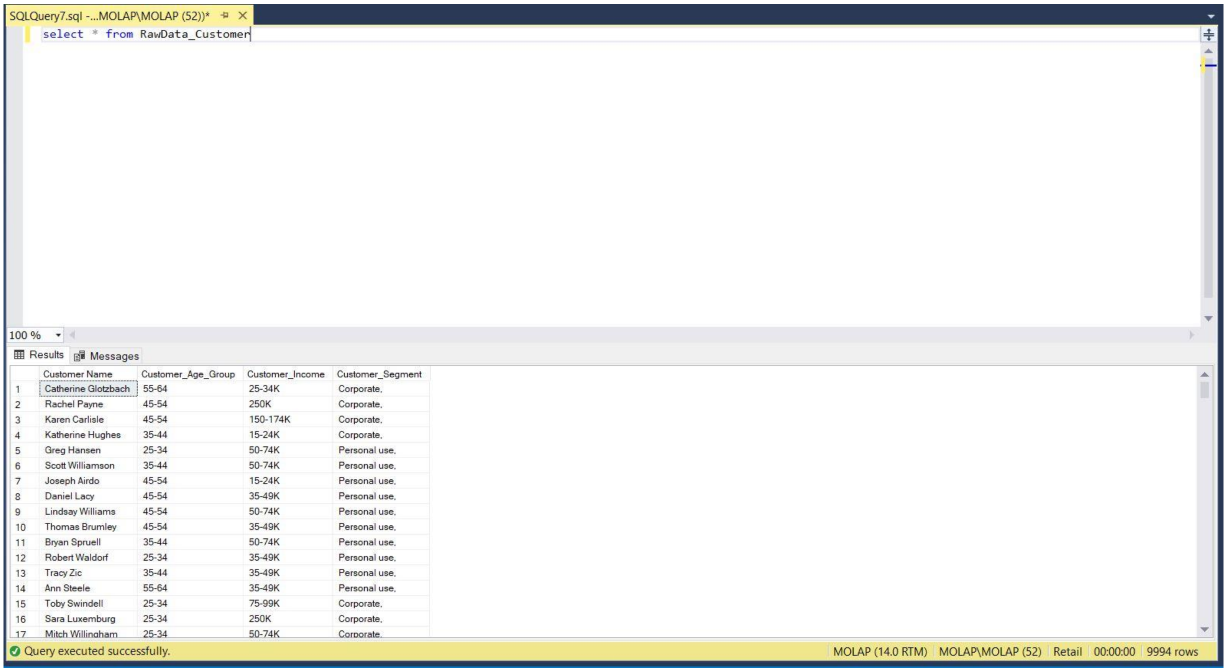
```
select * from RawData_Product
```

	Category	Sub-Category	Product Name
1	Furniture	Bookcases	Bush Somerset Collection Bookcase
2	Furniture	Chairs	Hon Deluxe Fabric Upholstered Stacking Chairs Rounded Back
3	Office Supplies	Labels	Self-Adhesive Address Labels for Typewriters by Universal
4	Furniture	Tables	Bretford CR4500 Series Slim Rectangular Table
5	Office Supplies	Storage	Eldon Fold 'N Roll Cart System
6	Furniture	Furnishings	Eldon Expressions Wood and Plastic Desk Accessories Cherry Wood
7	Office Supplies	Art	Newell 322
8	Technology	Phones	Mitel 5320 IP Phone VoIP phone
9	Office Supplies	Binders	DXL Angle-View Binders with Locking Rings by Samsill
10	Office Supplies	Appliances	Belkin F5C206VTEL 6 Outlet Surge
11	Furniture	Tables	Chromcraft Rectangular Conference Tables
12	Technology	Phones	Konftel 250 Conference phone - Charcoal black
13	Office Supplies	Paper	Xerox 1967
14	Office Supplies	Binders	Fellowes PB200 Plastic Comb Binding Machine
15	Office Supplies	Appliances	Holmes Replacement Filter for HEPA Air Cleaner Very Large Room HEPA Filter
16	Office Supplies	Binders	Storex DuraTech Recycled Plastic Frosted Binders
17	Office Supplies	Storage	Stur-D-Stor Shelving Vertical 5-Shelf 72 H x 36 W x 18 1 2 D

```
Query executed successfully.    MOLAP (14.0 RTM) | MOLAP\MOLAP (52) | Retail | 00:00:01 | 109934 rows
```

Fig. 2 Raw data from product data source

Raw data for customers:



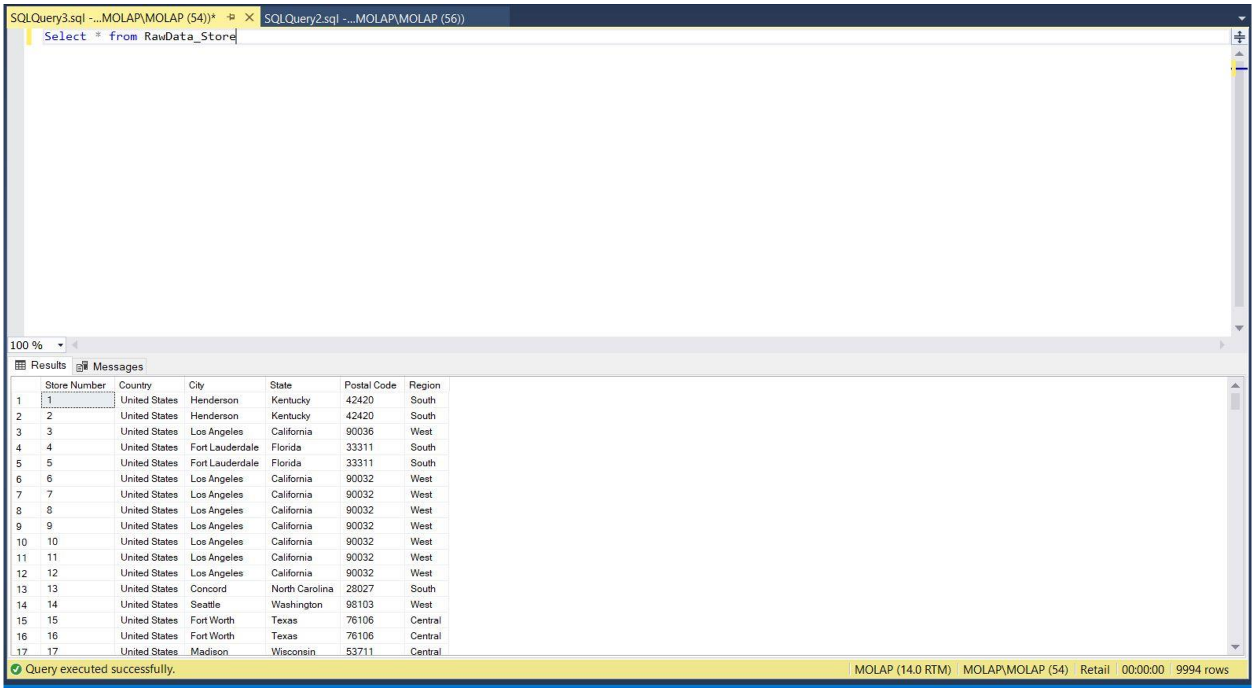
```
select * from RawData_Customer
```

	Customer Name	Customer_Age_Group	Customer_Income	Customer_Segment
1	Catherine Glotzbach	55-64	25-34K	Corporate,
2	Rachel Payne	45-54	250K	Corporate,
3	Karen Carlisle	45-54	150-174K	Corporate,
4	Katherine Hughes	35-44	15-24K	Corporate,
5	Greg Hansen	25-34	50-74K	Personal use,
6	Scott Williamson	35-44	50-74K	Personal use,
7	Joseph Airdo	45-54	15-24K	Personal use,
8	Daniel Lacy	45-54	35-49K	Personal use,
9	Lindsay Williams	45-54	50-74K	Personal use,
10	Thomas Brumley	45-54	35-49K	Personal use,
11	Bryan Spruell	35-44	50-74K	Personal use,
12	Robert Waldorf	25-34	35-49K	Personal use,
13	Tracy Zic	35-44	35-49K	Personal use,
14	Ann Steele	55-64	35-49K	Personal use,
15	Toby Swindell	25-34	75-99K	Corporate,
16	Sara Luxemburg	25-34	250K	Corporate,
17	Mitch Willingham	25-34	50-74K	Corporate,

Query executed successfully. | MOLAP (14.0 RTM) | MOLAP\MOLAP (52) | Retail | 00:00:00 | 9994 rows

Fig. 3 Raw data from customers data source

Raw data for store:


SQLQuery3.sql -...MOLAP\MOLAP (54))*
SQLQuery2.sql -...MOLAP\MOLAP (56))
100 %
Results
Messages

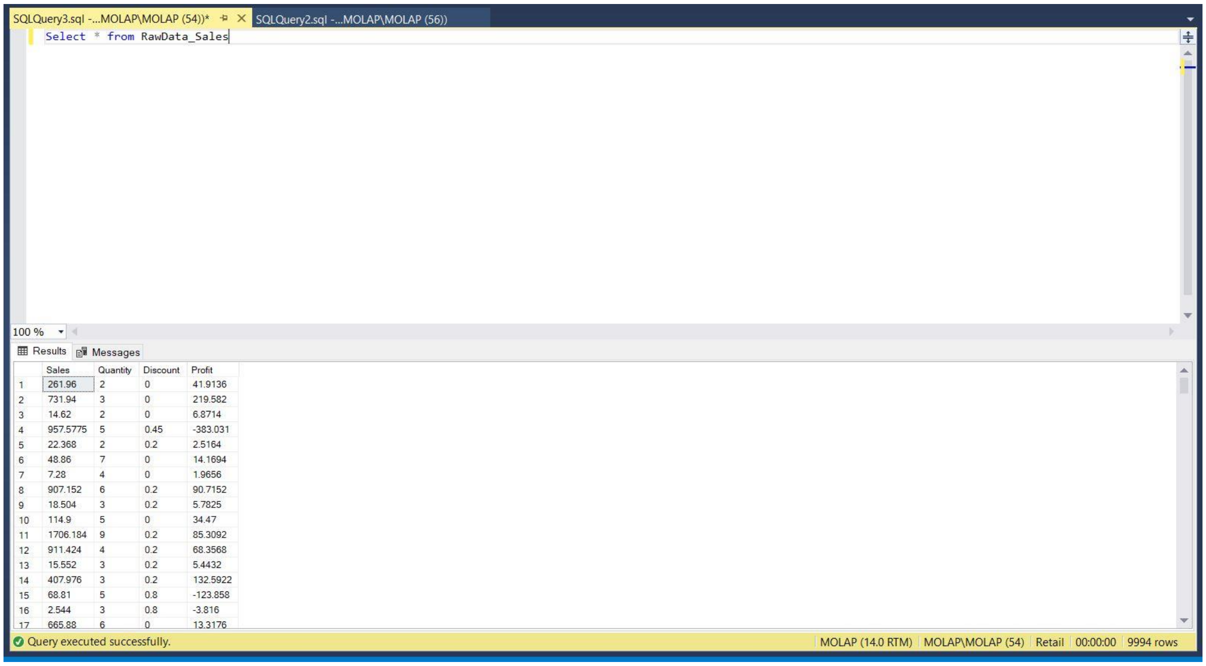
```
Select * from RawData_Store
```

	Store Number	Country	City	State	Postal Code	Region
1	1	United States	Henderson	Kentucky	42420	South
2	2	United States	Henderson	Kentucky	42420	South
3	3	United States	Los Angeles	California	90036	West
4	4	United States	Fort Lauderdale	Florida	33311	South
5	5	United States	Fort Lauderdale	Florida	33311	South
6	6	United States	Los Angeles	California	90032	West
7	7	United States	Los Angeles	California	90032	West
8	8	United States	Los Angeles	California	90032	West
9	9	United States	Los Angeles	California	90032	West
10	10	United States	Los Angeles	California	90032	West
11	11	United States	Los Angeles	California	90032	West
12	12	United States	Los Angeles	California	90032	West
13	13	United States	Concord	North Carolina	28027	South
14	14	United States	Seattle	Washington	98103	West
15	15	United States	Fort Worth	Texas	76106	Central
16	16	United States	Fort Worth	Texas	76106	Central
17	17	United States	Madison	Wisconsin	53711	Central

Query executed successfully. | MOLAP (14.0 RTM) | MOLAP\MOLAP (54) | Retail | 00:00:00 | 9994 rows

Fig. 4 Raw data from store data source.

Raw data for sales:


SQLQuery3.sql -...MOLAP\MOLAP (54))*
SQLQuery2.sql -...MOLAP\MOLAP (56))
100 %
Results
Messages

```
Select * from RawData_Sales
```

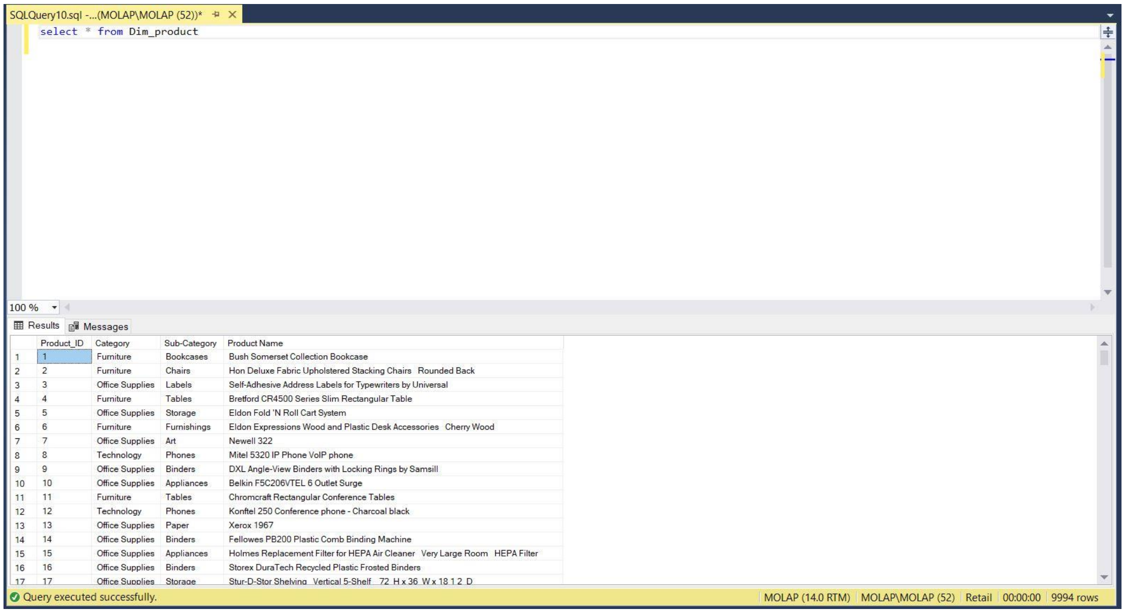
	Sales	Quantity	Discount	Profit
1	261.96	2	0	41.9136
2	731.94	3	0	219.582
3	14.62	2	0	6.8714
4	957.5775	5	0.45	-383.031
5	22.368	2	0.2	2.5164
6	48.86	7	0	14.1694
7	7.28	4	0	1.9656
8	907.152	6	0.2	90.7152
9	18.504	3	0.2	5.7825
10	114.9	5	0	34.47
11	1706.184	9	0.2	85.3092
12	911.424	4	0.2	68.3568
13	15.552	3	0.2	5.4432
14	407.976	3	0.2	132.5922
15	68.81	5	0.8	-123.858
16	2.544	3	0.8	-3.816
17	665.88	6	0	13.3176

Query executed successfully. MOLAP (14.0 RTM) | MOLAP\MOLAP (54) | Retail | 00:00:00 | 9994 rows

Fig. 5 Raw data for sales.

Dimension table for product:

The dimension table for product contains the product_id which is the primary key for the dimension product. It contains the category and sub category of products and the proper names of the different products sold.



```
select * from Dim_product
```

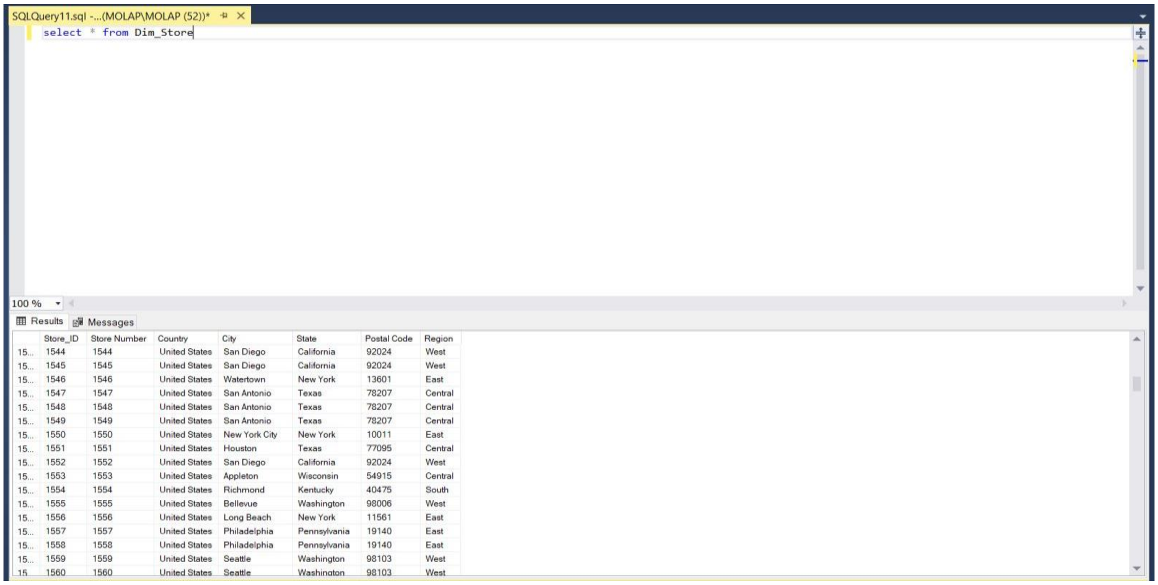
	Product_ID	Category	Sub-Category	Product Name
1	1	Furniture	Bookcases	Bush Somerset Collection Bookcase
2	2	Furniture	Chairs	Hon Deluxe Fabric Upholstered Stacking Chairs Rounded Back
3	3	Office Supplies	Labels	Self-Adhesive Address Labels for Typewriters by Universal
4	4	Furniture	Tables	Bretford CR4500 Series Slim Rectangular Table
5	5	Office Supplies	Storage	Eldon Fold 'N Roll Cart System
6	6	Furniture	Furnishings	Eldon Expressions Wood and Plastic Desk Accessories Cherry Wood
7	7	Office Supplies	Art	Newell 322
8	8	Technology	Phones	Mitel 5320 IP Phone VoIP phone
9	9	Office Supplies	Binders	DXL Angle-View Binders with Locking Rings by Samsill
10	10	Office Supplies	Appliances	Belkin F5C206VTEL 6 Outlet Surge
11	11	Furniture	Tables	Chromcraft Rectangular Conference Tables
12	12	Technology	Phones	Konftel 250 Conference phone - Charcoal black
13	13	Office Supplies	Paper	Xerox 1967
14	14	Office Supplies	Binders	Fellowes PB200 Plastic Comb Binding Machine
15	15	Office Supplies	Appliances	Holmes Replacement Filter for HEPA Air Cleaner Very Large Room HEPA Filter
16	16	Office Supplies	Binders	Storex DuraTech Recycled Plastic Frosted Binders
17	17	Office Supplies	Storage	Stur-D-Stor Shelving Vertical 5-Shelf 72 H x 36 W x 18 1 2 D

Query executed successfully. MOLAP (14.0 RTM) | MOLAP\MOLAP (52) | Retail | 00:00:00 | 9994 rows

Fig. 6 Dimension product preview in SSMS

Dimension table for store:

The dimension store contains various attributes of the stores, store_id is set as the primary key and it contains the store locations from country, region, state, city and postal code.



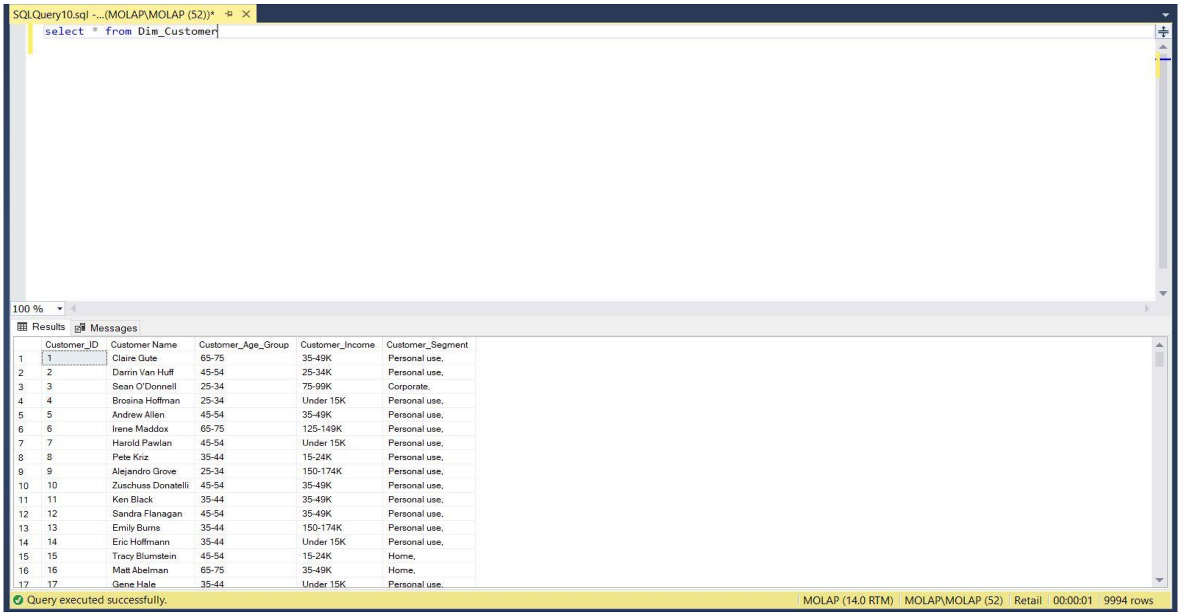
```
select * from Dim_Store
```

	Store_ID	Store Number	Country	City	State	Postal Code	Region
15...	1544	1544	United States	San Diego	California	92024	West
15...	1545	1545	United States	San Diego	California	92024	West
15...	1546	1546	United States	Watertown	New York	13601	East
15...	1547	1547	United States	San Antonio	Texas	78207	Central
15...	1548	1548	United States	San Antonio	Texas	78207	Central
15...	1549	1549	United States	San Antonio	Texas	78207	Central
15...	1550	1550	United States	New York City	New York	10011	East
15...	1551	1551	United States	Houston	Texas	77095	Central
15...	1552	1552	United States	San Diego	California	92024	West
15...	1553	1553	United States	Appleton	Wisconsin	54915	Central
15...	1554	1554	United States	Richmond	Kentucky	40475	South
15...	1555	1555	United States	Bellevue	Washington	98006	West
15...	1556	1556	United States	Long Beach	New York	11561	East
15...	1557	1557	United States	Philadelphia	Pennsylvania	19140	East
15...	1558	1558	United States	Philadelphia	Pennsylvania	19140	East
15...	1559	1559	United States	Seattle	Washington	98103	West
15	1560	1560	United States	Seattle	Washington	98103	West

Fig. 7 Dimension store preview in SSMS.

Dimension table for customer:

Dimension customer is having all the attributes of customers. Customer_id is the primary key in the dimension. Attributes of customers are their age group, Customer income level and customer segment which we are using in our business intelligence query.



```
select * from Dim_Customer
```

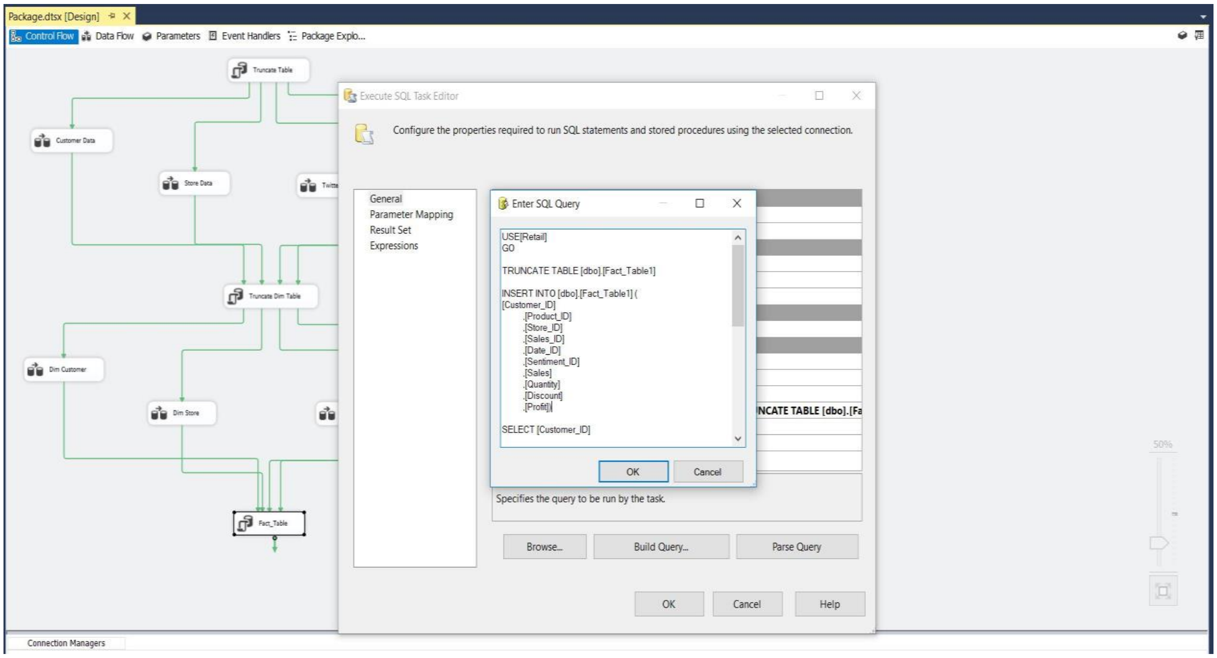
	Customer_ID	Customer Name	Customer_Age_Group	Customer_Income	Customer_Segment
1	1	Claire Gute	65-75	35-49K	Personal use,
2	2	Darrin Van Huff	45-54	25-34K	Personal use,
3	3	Sean O'Donnell	25-34	75-99K	Corporate,
4	4	Brosina Hoffman	25-34	Under 15K	Personal use,
5	5	Andrew Allen	45-54	35-49K	Personal use,
6	6	Irene Maddox	65-75	125-149K	Personal use,
7	7	Harold Pawlan	45-54	Under 15K	Personal use,
8	8	Pete Kriz	35-44	15-24K	Personal use,
9	9	Alejandro Grove	25-34	150-174K	Personal use,
10	10	Zuschuss Donatelli	45-54	35-49K	Personal use,
11	11	Ken Black	35-44	35-49K	Personal use,
12	12	Sandra Flanagan	45-54	35-49K	Personal use,
13	13	Emily Burns	35-44	150-174K	Personal use,
14	14	Eric Hoffmann	35-44	Under 15K	Personal use,
15	15	Tracy Blumstein	45-54	15-24K	Home,
16	16	Matt Abelman	65-75	35-49K	Home,
17	17	Gene Hale	35-44	Under 15K	Personal use,

```
Query executed successfully.    MOLAP (14.0 RTM) | MOLAP\MOLAP (52) | Retail | 00:00:01 | 9994 rows
```

Fig. 8 Dimension customer preview in SSMS.

Fact Table:

Fact table for this project was created using the execute SQL task in SSIS during the ETL process. Fact table has its own unique primary key known as fact_id in our data warehouse project. Fact table consists all the foreign keys of our dimension tables and the measures values such as sales in dollars, profit earned, discount and quantity of product sold. Fact table is populated using the insert SQL commands and all the primary key of the fact table is joined with the foreign keys of all the dimension tables. Figure below shows the SQL query for populating the fact table.



Configure the properties required to run SQL statements and stored procedures using the selected connection.

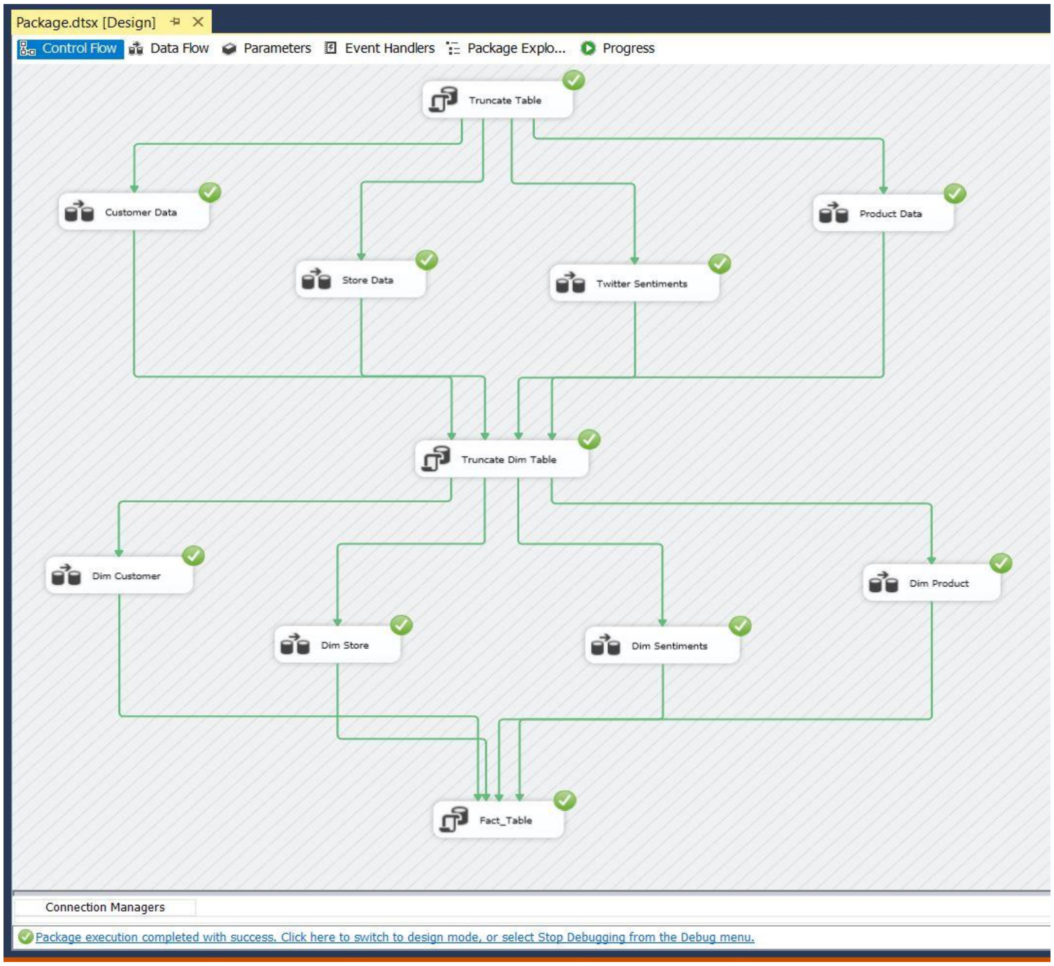
```
USE[Retail]
GO

TRUNCATE TABLE [dbo].[Fact_Table1]

INSERT INTO [dbo].[Fact_Table1] (
[Customer_ID]
      ,[Product_ID]
      ,[Store_ID]
      ,[Sales_ID]
      ,[Date_ID]
      ,[Sentiment_ID]
      ,[Sales]
      ,[Quantity]
      ,[Discount]
      ,[Profit])

SELECT [Customer_ID]
```

Specifies the query to be run by the task.

Fig. 9 Fact table insert SQL query

ETL strategy:

ETL stands for extraction, transformation and loading the extraction process is used to extract the data from different data sources. In, transformation process the raw data undergoes certain transformations such that the data is cleaned, converted into logical data and then data is loaded in the data warehouse server in a well presentable manner. ETL operations are processed in SSIS. The area where ETL process is carried out in SSIS is also called as staging area. The figure below illustrates the ETL process for our project and the data is stored in the SQL server which can be accessed using SSMS.



Package execution completed with success. Click here to switch to design mode, or select Stop Debugging from the Debug menu.

Fig. 10 successful execution of the ETL process.

In our ETL process shown above in figure 2, we first have a truncate table which deletes the data which is already present in the table, so that there are no duplicate values and our ETL is a re – runnable ETL. In the second step we have imported the raw data files using the import flat file function and stored to its destination table in the server. Raw data is extracted from the CSV file formats downloaded from multiple sources mentioned above. Here the data can go under certain transformation, cleaning and then the dimension tables are created for our data warehouse project. These dimension tables are then populated and the data is stored in the SQL server. In this project we have five-dimension tables and a fact table.

Steps performed in ETL process:

1. Extraction:

We have downloaded structured data from multiple sources discussed above and the data is stored in CSV file format which is then extracted in the staging area. The extraction of file is done using the data flow task in SSIS in which we use flat file source to OLE DB destination. Then this data from raw tables is populated to the respective dimension tables. The unstructured data is extracted from the Twitter tweets using the Twitter API and is fetched and converted into CSV file using R code.

2. Cleaning:

Huge amount of structured data was downloaded for this project in CSV format. These large amounts of data in CSV files were having NULL values and some special characters which were removed in the cleaning process. Unstructured data was cleaned using the R code.

3. Transformation:

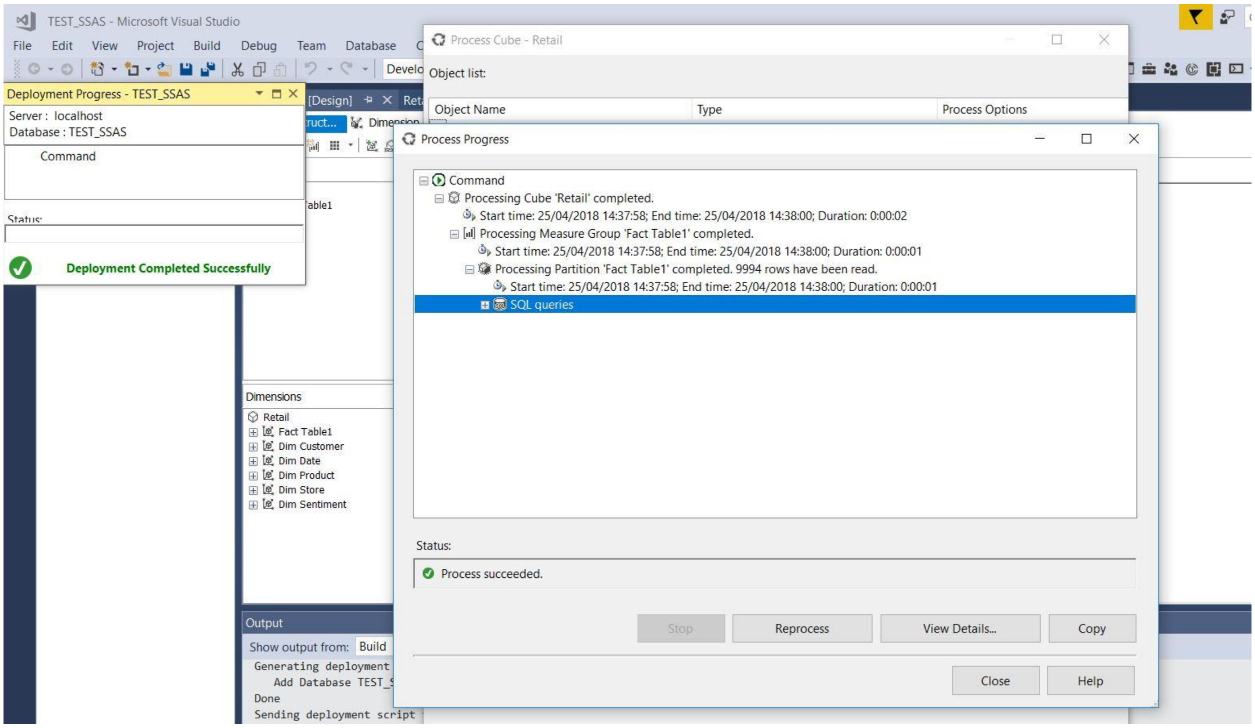
Transforming the data to its appropriate format so that it is further ready to load in the data warehouse server. In transformation process we have converted the data types to its suitable format in SQL. Transformation process of data is done using SSIS tool.

4. Loading:

Loading is the final step of the ETL process in which the data is loaded in the data warehouse server. Our data from the various sources is loaded into the destination that it the dimension tables and fact table. Data is loaded using the SQL commands. We have processed the loading in SSIS tool using the execute SQL task in which we used the INSERT SQL command to populate the data. We can see the data populated in the fact table and dimension tables in the SQL server using SSMS.

Creating and deploying the CUBE:

The CUBE is created using the fact table and dimension tables which we have created earlier. Creation and deployment of CUBE was done using SSAS. Our CUBE was created and deployed using the star schema.



Deployment Progress - TEST_SSAS

Server : localhost
Database : TEST_SSAS

Command

Status:

Deployment Completed Successfully

```
Command
  Processing Cube 'Retail' completed.
    Start time: 25/04/2018 14:37:58; End time: 25/04/2018 14:38:00; Duration: 0:00:02
    Processing Measure Group 'Fact Table1' completed.
      Start time: 25/04/2018 14:37:58; End time: 25/04/2018 14:38:00; Duration: 0:00:01
      Processing Partition 'Fact Table1' completed. 9994 rows have been read.
        Start time: 25/04/2018 14:37:58; End time: 25/04/2018 14:38:00; Duration: 0:00:01
        SQL queries
```

Status:

Process succeeded.

Show output from: Build

```
Generating deployment
   Add Database TEST_S
Done
Sending deployment script
```

Fig. 11 Successful deployment of CUBE in SSAS.

In our DWBI project we used the analytical service in which firstly, we created a data source then we moved further to the data source view and created a new data source view in which we selected our dimension tables and fact table. Then we created new CUBE selected the fact table measure and then we processed the CUBE. Once the CUBE is deployed and processed we open the browser and check whether fact table and dimension tables are populated.

Application of data warehouse:

Business intelligence queries:

Tableau is used for visualization of Business intelligence queries. Tableau provide a better understanding of data and is a good approach for visualization of BI queries.

1. Which category of product will be having maximum sales in $ in year 2020 and in which region of USA?

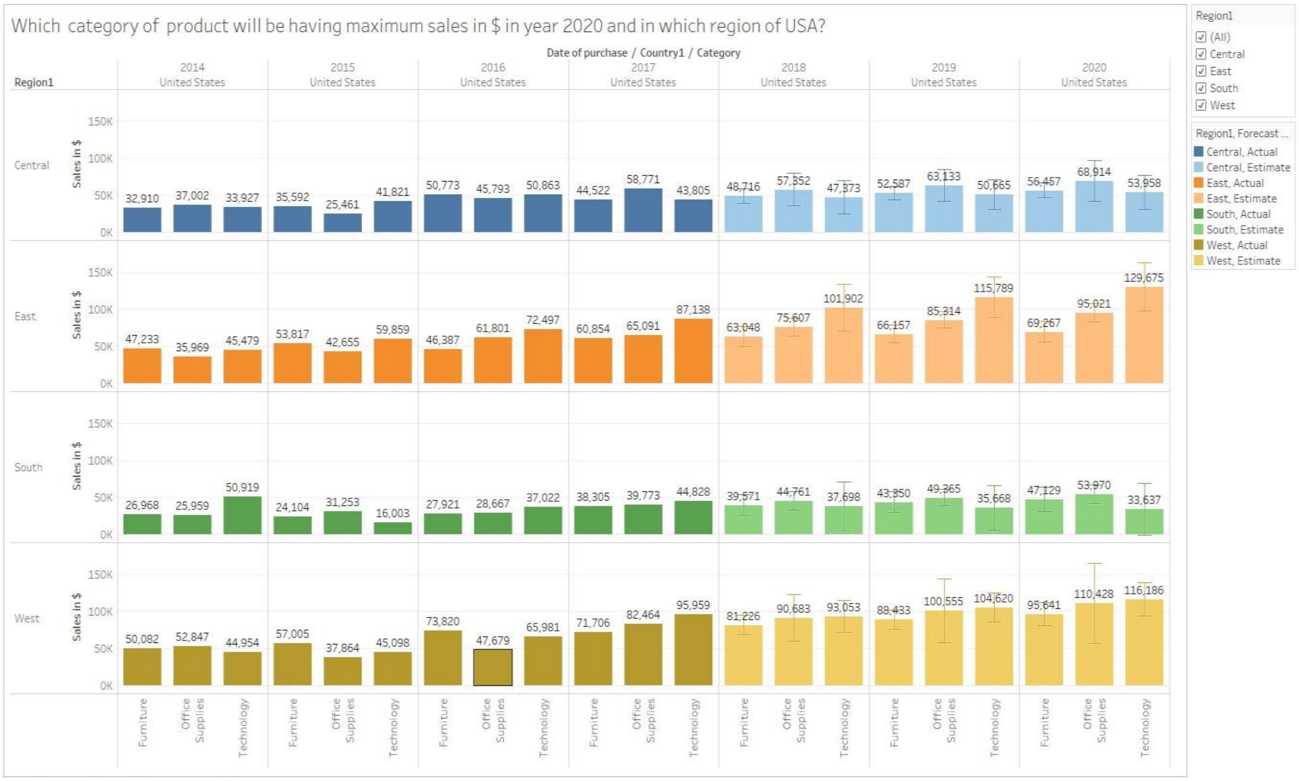
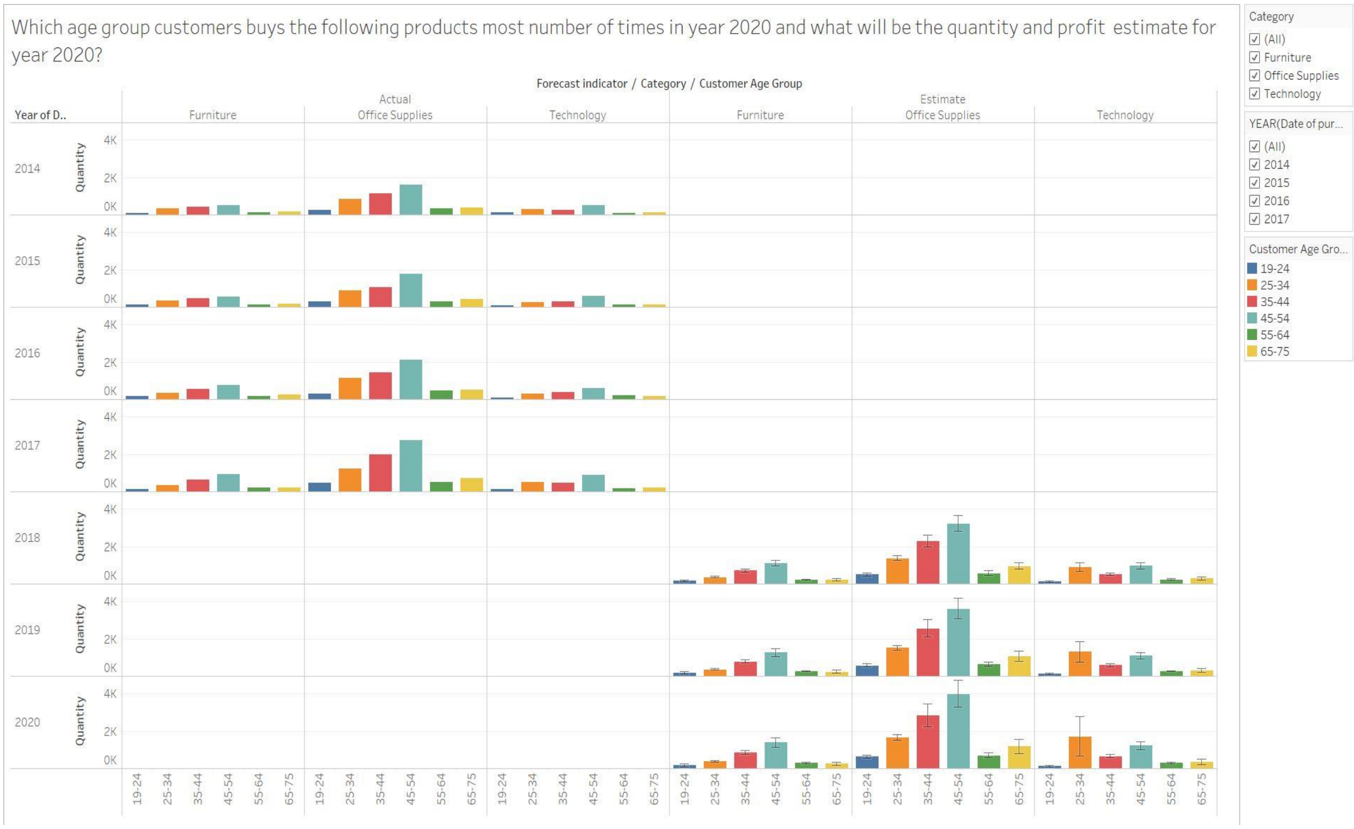


Fig.12 Business intelligence query 1 visualization in Tableau.

This BI query predicts the which product will be having maximum amount of sales for a retail superstore in the year 2020 and categorizes the product sales in all four regions of USA. We can say from the visualization that central and south region will be having maximum sales for office supplier in 2020. Whereas east and west region are more into technology products. We have differentiated the actual sales and estimates sales with different colors in Tableau.

The product categories can be drill down to product sub categories and product names. Store regions have an option to drill to its state and city. Two structured data sources were used for this business query that are product dataset and the store location dataset in addition to it a sales measure was used to add value to this business intelligence query.

2. Which age group customers buys the following product most number of times in the year 2020? What will be the estimated quantity sold and profit earned for year 2020?



Which age group customers buys the following products most number of times in year 2020 and what will be the quantity and profit estimate for year 2020?

Category
☑ (All)
☑ Furniture
☑ Office Supplies
☑ Technology

YEAR(Date of pur...
☑ (All)
☑ 2014
☑ 2015
☑ 2016
☑ 2017

Customer Age Gro...
🟦 19-24
🟧 25-34
🟥 35-44
🟦 45-54
🟩 55-64
🟨 65-75

Fig.13 Business intelligence query 2 visualization in Tableau.

This BI query tells us about which age group of customers are more interested in buying which products and what will be estimated quantity sold and profit earned by each age group, so that a business user can apply changes in the marketing strategies to attract more consumers according to their age group and product interests which ultimately enhances business. So as per visualization we estimated for year 2020 the age group 45-54 will be more common in buying the products for both categories furniture and office supplies. People in the age group 25-34 will be more attracted towards the technology products.

This BI query was implemented with the use of two structured data sources first is the customer data source and second is the product data source. Measures used were the quantity of products sold and the profit earned.

3. What are the sentiments of customer on three product categories and in each region?

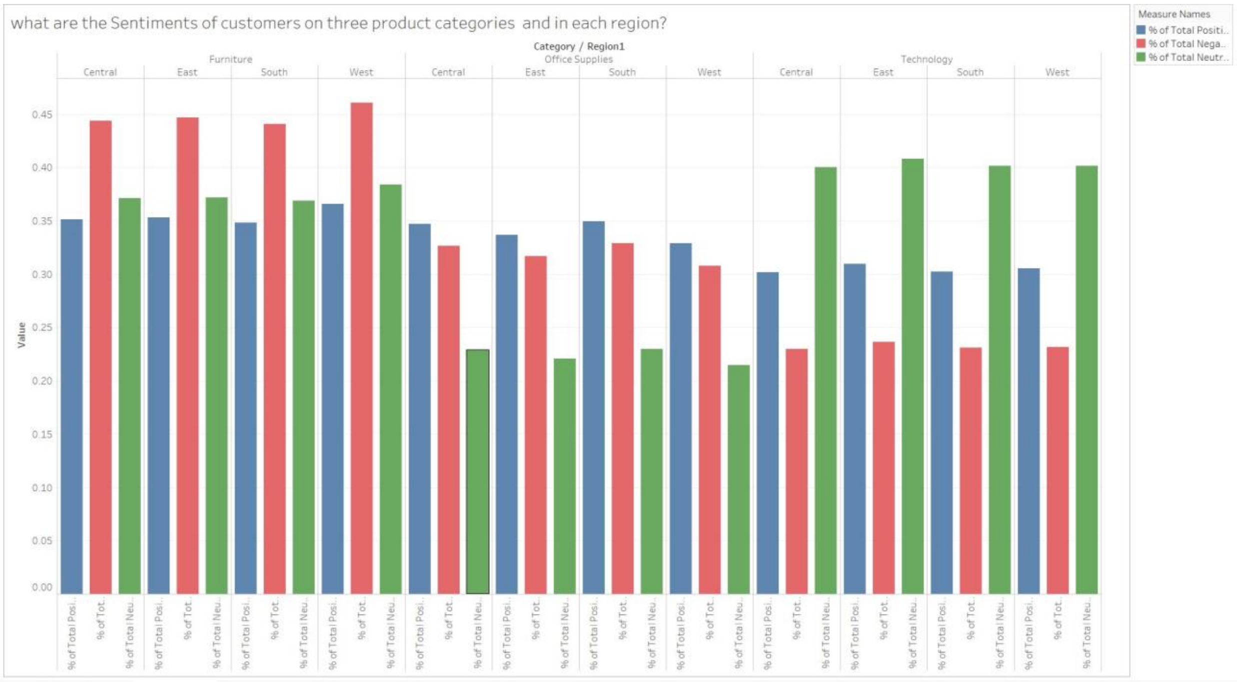


Fig.14 Business intelligence query 3 visualization in Tableau.

This BI query tells us about the customers sentiments about the three product categories in each region. Red color is for negative sentiments, blue is for positive and green is neutral. We have calculated the percentages for each product in all four regions. From the visualization we conclude that most unlikely category for consumers is the furniture with more negative and less positive sentiments. Technology product category is more neutral and less unlikely. Office supplies are little more into the positive direction.

We have used three data sources in this query one is the unstructured data from twitter, second is the product data and the third is the store location dataset.

Conclusion:

Thus, we have successfully implemented the data warehouse and business intelligence model for analysis of sales for a retail store. Any business user can analyze and forecast its sales for a retail store on various perspective such as categories of products, store locations and customers age, income. Analysis can be done using the CUBE deployed with different dimension tables and the fact table in our data warehouse project.

References

Kimball, R. & Ross, M., 2013. *The Data Warehouse Toolkit.* 3rd ed. s.l.:Wiley.