Contents

[**1. Introduction:** 1](#_Toc83075967)

[**1.1** **Reasons for selecting the subject area AND DATA** 1](#_Toc83075968)

[**1.2** **Data Description** 1](#_Toc83075969)

[**1.3** **Vision and Goals** 2](#_Toc83075970)

[**1.4** **Key Stakeholders** 3](#_Toc83075971)

[**1.5** **Business requirements** 3](#_Toc83075972)

[**2** **Schema** 3](#_Toc83075973)

[**2.1** **Snowflake Schema:** 3](#_Toc83075974)

[**2.2** **Fact Constellation Schema or Galaxy Schema:** 4](#_Toc83075975)

[**3 ETL / SQL 6**](#_Toc83075976)

[**4 Visualizations and Reports 7**](#_Toc83075977)

[**5 Conclusion 11**](#_Toc83075978)

[**6 Reference 12**](#_Toc83075979)

## **1.Introduction:**

The aim of this assessment finds a best matching dataset which is related with the finance and the market situation of sales and revenue and to carry on the analysis and recommendations. This data is sourced from the Kaggle , E-Commerce website real-time dataset, this dataset describes about the total sales on different categories of the store products. From October 2013 to December 2013 of a Famous E-Commerce company in India. As published by a user in the Kaggle Dataset Repository.

## **Reasons for selecting the subject area AND DATA**

The reason for selection this dataset is that it’s the live real-time dataset of the popular E-Commerce Website in India and this sales information shows us the good data visualisation techniques and the Database modelling techniques can be understood in the real word heavy dataset.

## **Data Description**

The dataset consists of E-Commerce checkout data The fields given in the dataset are as follows:

* **Attempt\_key** (Unique key to each transaction),
* **Customer\_id** (Unique ID to every customer),
* **Date** (Date of the transaction),
* **Product** (Product added for checkout),
* **Gender** (Gender of the customer),
* **Device Type** (Web or Mobile device used for the transaction),
* **City** (City of the customer),
* **State**( State of the customer),
* **Country**(Country of the customer),
* **Category** (Product Category). ,
* **Customer\_Login\_type** ( Type of the customer-Member or Guest),
* **Delivery\_Type** (Type of Delivery chosen for the transaction(Normal or 1 Day Delivery),
* **Quantity** (number of products added in the checkout) ,
* **STARTS** (Boolean : 1 for all transaction -This is to indicate every transaction),
* **DONES** (Boolean : 0 OR 1 -This indicates whether a transaction is successful (1) or failure (0) ),
* **Amount** (value of the Quantity for checkout items),
* **Month** (Month of the transaction date)

The dataset has Total of 65536 number of rows for each above columns respectively on total it has 1,179,648 Records of data available, since this is the real time data of a famous E-Commerce company in India, Name of the company has been hidden out and we can assume the veracity of the data and we can say that this might be of Top E-Commerce like Amazon, Flipkart who started their business during the time of 2013 respectively. And the data has the time bound of 3 months from October to January 2014.

## **Vision and Goals**

The Vision and Goals is to find the Total Number of sales respective of the category of the Products and other several items, Visualizing and creation of the database in the SQL Server and creating tables and other Manipulation Techniques incorporated in this dataset. The Ultimate goal from this dataset is to understand how the sales of the E-Commerce institution has been improved in several location and how sales are lacking in the areas which we can identify from this analysis and visualise them clearly and show the results, so that the E-Commerce company can improve in the future and save form loss, where the people who are not interested in purchasing products or items and which item they prefer more of the choice based on the final check out dataset provided. From this Dataset we can find the growth of the company and give the Constructive recommendation for the future development of the institution and its scalability financially.

## **Key Stakeholders**

From this Dataset we can understand that the Final End user of this product are the main stake Holders, Sellers are the Next important stakeholders, Middle-Man (Such as the Delivery and other Logistics) we can tell them as stake holders if this three get interrupted then there is a positive impact on the sales and the company revenue.

## **Business requirements**

Business Requirements of this Project defines the Good Recommendation to improve the sales of the E-Commerce company and the Growth of its sales in future and Profitability, Identifying the customer centric approach on the specific products sold by the company and the most likelihood of choice of items, based on the quantity and quality of the individual item measured in means of the sales result Visualization,

# **Schema**

A schema is a logical description of a database's contents. It contains the name and description of all record kinds, as well as all associated data items and aggregates. A data warehouse, like a database, necessitates the upkeep of a schema. The relational model is used in databases, while the Star, Snowflake, and Fact Constellation structure are used in data warehouses.

## **Snowflake Schema:**

A snowflake schema is a logical arrangement of tables in a multidimensional database that resembles a snowflake form on the entity relationship diagram. When all the dimension tables are normalized, the resultant structure resembles a snowflake, with the fact table in the centre.

Likewise in our database we have also made the schema with Fact table at the centre, and which contains the Customer\_Dimension, Item\_Dimension, Time\_Log, Location\_Dimension and Payment\_Dimension, this are all the main fact table of our database which contains all information, then they are subbranches to other dimensions tables with other several attributes of the database records.

Diagram

Description automatically generated

## **Fact Constellation Schema or Galaxy Schema:**

A galaxy Schema or Fact Constellation schema is used for displaying a multidimensional model, It's a set of multiple fact tables with some shared dimension tables. It is also known as Galaxy schema since it is a compilation of multiple star schemas. It is a significantly more complicated schema than the star and snowflake schemas and is one of the most often utilized for data warehouse construction. We need fact constellations for complicated systems.

Here in our E-Commerce Dataset, we have brought the complex form of the entire facts available in the database and its more in depth than the star and snowflake schemas, which we can understand easily, similarly centre, is the same fact table with all dimensions and then it extends to further dimensional tables and furthermore it expands to the last available fact of the data.

Sales Fact table has the Following Attributes,

* Customer\_Dimension
* Items\_Dimension
* Location\_Dimension
* Time\_Log
* Payment\_Dimension

**Customer\_Dimension:**

The Customer\_Dimension has the entire details about the Customer Details like

* Customer\_Personal:

Which, further has the table of Customer\_id and Customer\_Gender

* Device\_Type:

Which, further has the table of Mobile and Web

* Customer\_Login\_Type

Which, further has the table of Member and Guest

* Delivery\_Info

Which, further has the table of Standard or Normal Delivery and One \_Day\_Delivery

**Items\_Dimension**

Item Dimension is another schema table, and which contains multiple table in them , such as the Categories, Quantity, on further Quantity is divided into Five Attributes and they all are segregated in a separate table ,

each Category such as:

* Clothing
* Fashion
* Electronics
* Stationeries
* Wearables
* Households
* Accessories
* Vehicle

Again, this all Categories has their own individual Products in them for clothing they are Jeans and Shirt, they are mentioned as a separate Table , like wise Fashion has its own Table and it contains spectacles, Facial Cream and Bag. Electronics category contains watch, Stationaries category contain Book and Finally Wearables category contains Shoes as their product.

A picture containing text, indoor, screenshot

Description automatically generated

The Above Galaxy Schema explains all in detail about the database modelling.

# ETL / SQL

First dataset is loaded into the SQL – Microsoft SQL Server Management tool Lit by creating the new database and then import the existing dataset in the Dashboard to edit the data, view specific columns , top 300 etc.

A screenshot of a computer

Description automatically generated

# **Visualizations and Reports**

**Gender Based :**

Below we have made agender based visualisation on tableau for the Overall sales from the September 13 till January 2014. Darker the colour, more the number of Sales in the time

Table

Description automatically generated

**Sales Revenue Based on each geography location**

Below Represent show sus the Category based upon their geographic loaction to find the optimal sales revenue, from the below result we can say that Chennai is still under development and on other side Hydrabad gainining momnentum as it tops the revenue on compared with other geopgraphic loactions.

Chart, line chart

Description automatically generated

**Unsuccessful Transactuions:**

Un successful transactrion is deckined on the sepcial request and we can find that the total sales irrespective

Chart, line chart

Description automatically generated

Device-Type : which tells aboutt the Total Number of customers who have used to purchadese

Either in Web based Application or Mobile based application. From this we can understabnd aboutt the taste of the customers.

Chart, pie chart

Description automatically generated

Like wise same for thr dvice-type failure occurs.(Web and Mobile based applications)

Chart, pie chart

Description automatically generated

**Delivery Information**

Now we must think how the delivery of the product can reduce the sales.As we can see the One-day delivery has the high impact rather han the normal and standard delivery, Though there is a steady fight between them we can take a constructive feedback that customer like the products to buy from the Store regulary irrespective of Delivery type.

Chart

Description automatically generated with medium confidence

Starts:

Starts tells about the Transaction start (1 is for all transaction) and in this visualisation we are taking the geographical locations vs the Starts. Below Shows the extent of the Infromation.

Application

Description automatically generated with low confidence

Chart, line chart

Description automatically generated

Survey : (Research on the Preference of the Customer )

A survey has been created to understand the right constructive feedback among the most users online, so we have planned to create a google forms survey, though it was a short period we got a good number of response to validate all the requested information for this constructive Analysis of E-Commerce to increase the sales Revenue.

Chart, pie chart

Description automatically generated

We covered all the key aspects of the Stakeholder point of view as we early stated , one is the Seller/ Retailer, Next is the Customer and the Middle person is the delivery and Logistics. Thought customer pays the very high level in this,

* On the Age Group Classification most people were from 21 to 30 Years Interested in the Online shopping a maximum point of view,
* on the Question of which Interface is the most preferable one , for that we received a response of 74.7 % mobile Application,
* For the Category of the Shopping , most people prefer Clothing, Fashion, Electronics to purchase online.
* For the Delivery Preference most people prefer for the less time delivery and in short period: 18.4 % for less than 1-day delivery time, 32% for 2-day delivery time and 20% like to set their priority based on the order items.
* For the Sign-Up or Guest question Most prefer for Sign-Up -67.15% and rest for Guest Option.

From this analysis we can understand the Mood of the end-Point Customer who face this problems during Purchasing online. More Interactive Mobile Application can support the Sales with all security features, Based on Targeting the specific age group can increase in revenue , as they tend to purchase more Fashion, Clothing, Electronics, and other next percentage of items on list, we must have specific criteria for marketing and additional discounts offer can booster the sales and increase revenue, with less delivery time and good software application as well.

# Conclusion

From this Overall Analysis we can tell that the deep understanding of the data has been studied , Schematic Diagram of the Snowflake and Galaxy has been performed and then the loading of the dataset in the SQL Server Management, and then the In-Depth Visualization using the Tableau has been carried out to showcase the best and constructive feedback and the state of the E-Commerce company at the current dataset provided date position and the Future improvement Recommendations has been provided. And more over this was a good assignment to learn deep and understand the Financial Situation and its implied co- conditioning ( Sales/Products/ Customer) state of E-Commerce Company in India from September 2013 to January 2014.

# Reference

[1] <https://www.kaggle.com/roopeshbharatwajkr/ecommerce-dataset/settings>

[2] <https://www.tutorialspoint.com/dwh/dwh_schemas.htm>

[3] <https://public.tableau.com/en-us/s/>

[4] <https://rigorousthemes.com/blog/15-best-tableau-dashboard-examples/>

[5] <https://www.sqlshack.com/import-data-excel-file-sql-server-database/>

[6] <https://www.sqlservertutorial.net/sql-server-basics/sql-server-create-table/>