**SALES DATA ANALYSIS**

PIE Group analytics team worked on the datasets provided by the DE team related to the sales. There were gaps in the data model provided by the DE team related to Sales table. We have provided our findings in related to sales tables in this comprehensive document.

We have given our analysis to existing data using a tableau dashboard.

Three key components of our finding related to the saes data are:

1. **Table Structures and EDA**
2. **Actual Data Model**
3. **Proposed Data Model**
4. **Analysis of the data**

**1.Table Structures and EDA:**

There was total 3 tables provided by the DE team. And they are Customer Master Table, Orders Table and Shipping table.

*Customer* Master table: It has major details of the customer demographics

*Orders table:* It has the information about the order details and can be treated as our fact table or base table.

*Shipping table:* it has the details of shipping information about the customer.

Data discrepancy related to customer\_id:

1a. Customer master table has many customer\_ids(customer numbers) who either placed an order or their orders are shipped.

The number of such customers is approximately 35. Please find the below examples.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Customer\_ID | First | Last | Age | Country |
| 7 | David | Davis | 59 | USA |
| 16 | David | Benson | 61 | USA |
| 19 | James | Johnson | 67 | UK |
| 25 | Raymond | Taylor | 36 | UAE |

There might be the possibility of historic data in the master table, if we are able to trace the data then we will be able to get important KPIs like customer cohort and repeat customer over the timeline.

1b. Data discrepancy related to order and shipping: There are few customers whose details are missing in orders table, they are appearing in shipping table. Below examples for shipping table

|  |  |  |
| --- | --- | --- |
| Shipping\_ID | Status | Customer\_ID |
| 1 | Pending | 173 |
| 2 | Pending | 155 |
| 5 | Delivered | 72 |
| 9 | Pending | 199 |

Similarly, we found cases where there were details about the customers in orders table but not in shipping table and pasted few examples of such cases in orders table below

|  |  |  |  |
| --- | --- | --- | --- |
| Order\_ID | Item | Amount | Customer\_ID |
| 1 | Keyboard | 400 | 139 |
| 2 | Mouse | 300 | 250 |
| 3 | Monitor | 12000 | 239 |
| 4 | Keyboard | 400 | 153 |

**2. Actual Data Model:**

We have created the data model as per the data given to us.It’s one to many relationships from customer to orders and shipping table. A diagram of a product

Description automatically generated with medium confidence

Limitations of the above model:

* We are unable to find the relationship between shipping and orders table. We can’t treat orders table as our fact table. Here base table will be customer table which is a dimensional table.
* Unable to track the load time in all the tables
* We have added many data points to track in the suggested model.

**3. Proposed Data Model:**

We have proposed the below data model where we are keeping the orders table as main table and rest of the two tables as dimensional tables.

The schema of the below model is star schema.

We have included load\_at column for each table which gives good idea about when the data have been inserted.

And few relatable data columns such as order\_date and ship\_date which will be helpful in trend analysis and calculating the Year over year growth.

The changes in the datatypes of each column can be changed accordingly.

Shipment ID might be same for multiple orders or may not be.

**A screenshot of a computer

Description automatically generated**

**4.Analysis of the data:**

We have built a tableau dashboard by using the existing data. There are multiple key takeaways which we have mentioned in the dashboard itself as insights sheet.

Below is the image of the dashboard.

A screenshot of a data analysis dashboard

Description automatically generated