# **SALES ANALYSIS PROJECT**

**By Abhijna Kulal**

**Contents**

[**SALES ANALYSIS PROJECT** 1](#_Toc175304197)

[**1.Introduction** 3](#_Toc175304198)

[**2. Description of the Sales Data** 4](#_Toc175304199)

[**2.1 Overview of the data** 4](#_Toc175304200)

[**2.2 Exploratory data analysis** 5](#_Toc175304201)

[**2.3 Table structure and table relationships** 5](#_Toc175304202)

[**2.3 Data issues** 6](#_Toc175304203)

[**2.4 Drawbacks of the current data model** 7](#_Toc175304204)

[**3. Project Implementation** 8](#_Toc175304205)

[**3.1 Project overview** 8](#_Toc175304206)

[**3.2 Analytics on the sales data** 9](#_Toc175304207)

[**3.3 Project future scope** 10](#_Toc175304208)

# **1.Introduction**

Analytics team of PEI group to build an analytics solution for the sales team. Analytics solution would help the sales team to track their key metrics and provide actionable insights which would help the sales team in strategic decision making.

# **2. Description of the Sales Data**

**2.1 Overview of the data:** There are total 3 dataset provided by the sales team. They are

a. Customer data

b. Order data

c. Shipping data

**Customer data** has the details about the customer which includes name, age, gender and country.

Below is the sample data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Customer\_ID | First | Last | Age | Country |
| 1 | Joseph | Rice | 43 | USA |
| 2 | Gary | Moore | 71 | USA |

**Order data** has the details about the order placed by the customer which includes order id, order amount and item.

Below is the sample data

|  |  |  |  |
| --- | --- | --- | --- |
| Order\_ID | Item | Amount | Customer\_ID |
| 1 | Keyboard | 400 | 139 |
| 2 | Mouse | 300 | 250 |

**Shipping data** has the details about the shipment of the order.

Which includes the shipping id and shipping status of the order.

|  |  |  |
| --- | --- | --- |
| Shipping\_ID | Status | Customer\_ID |
| 1 | Pending | 173 |
| 2 | Pending | 155 |

## **2.2 Exploratory data analysis**

We have done the EDA on all the three datasets.

* We didn’t find any null values in 3 tables
* We didn’t find any duplicate entries in all the tables
* Only order table had numerical column i.e amount which had any outliers in the data
* As there was only one metric, we didn’t do any correlation analysis
* As the data was mostly clean, we renamed the columns and added derived columns based on the existing columns (full name using first and last)

We have attached the python notebook file as well which consists of the source code of our analysis.

## **2.3 Table structure and table relationships**

Below is the attached data model of the 3 datasets provided

A diagram of a product

Description automatically generated with medium confidence

The data model clearly explains the relationship between the table.

Customer and orders tables have one to many relationships between them and the key is customer\_id.

Customer and Shipping tables have one to many relationships between them and the key is customer\_id.

Orders and Shipping table don’t have direct relationship but can be joined via customer table.

## **2.3 Data issues**

There were few gaps in the data sets after joining them.

Below are the key findings:

a. Customer master table has many customer\_ids(customer numbers) who neither 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 |

b. 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 |

## **2.4 Drawbacks of the current data model**

* Current Data model is not optimized because of less scope for addition of few more data sets.
* Customer table acts as the main table but it’s just a lookup table
* There is no correlation between the order table and shipping table and shipment details of the order can be tracked only using the customer table
* There is no date column in dataset which helps in analysing the trends.
* Data lineage and source system should be identified properly

# **3. Project Implementation**

## **3.1 Project overview**

Following the Data description step, we developed the high-level project design outlined below.

A computer screen shot of a diagram

Description automatically generated

**3.1.1 ETL:**  The data need to be extracted, transformed and loaded via ETL step.

This is mostly data ingestion from the source system. We have used python for the ETL step.

We can use many licensed low code platforms like alteryx, talend or prep or any cloud platforms like GCP, AWS and Azure.

**3.1.2 Storage:** Transformed data should be stored in database. We can use warehouses in cloud platforms like google Big query, Amazon redshift or any other databases like postgres or SQL server.

Databases will have two data layers.

1. Raw data layer: Actual source system data.
2. Aggregated data layer: aggregated source system data mainly for analytics purpose.

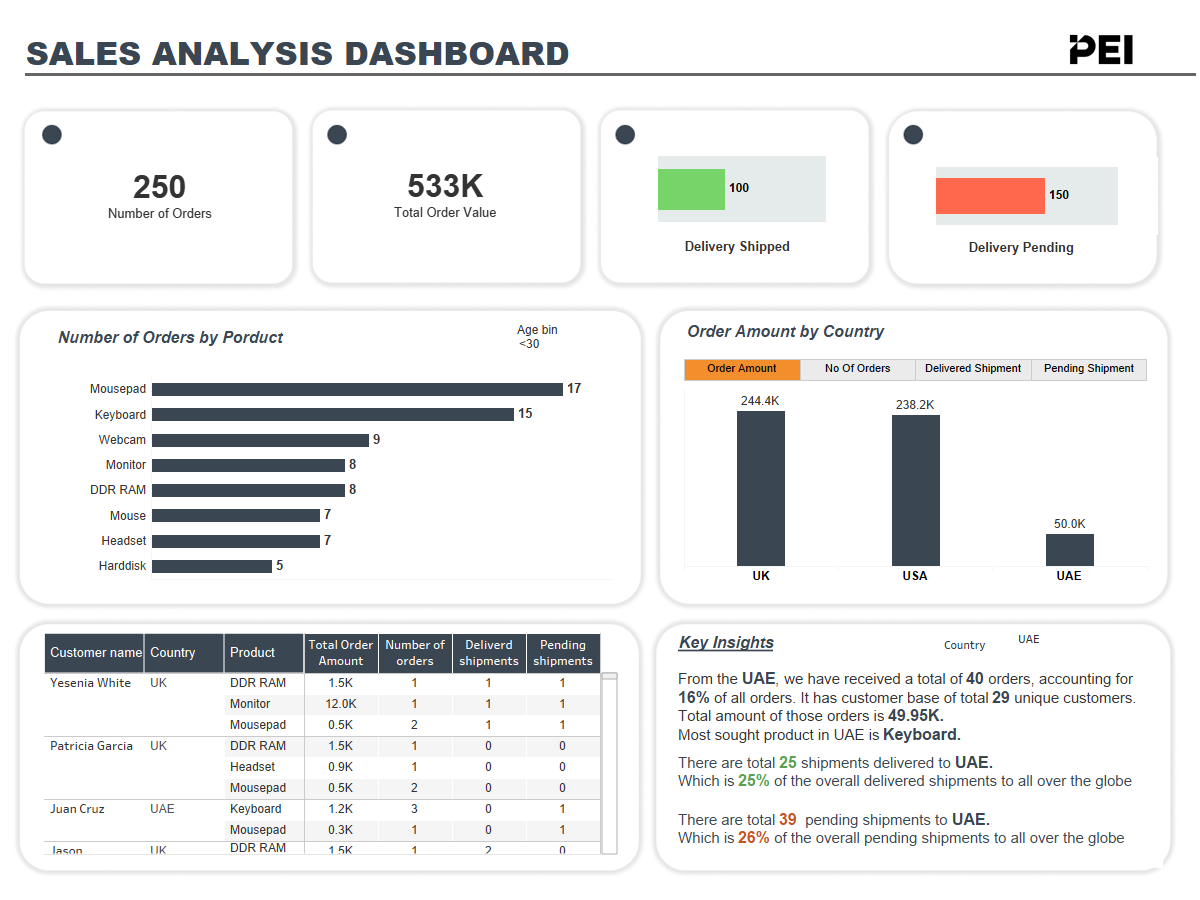
**3.1.3 Reporting:** Aggregated data layer can be visualized based on the business requirements. Here we have used tableau for creating dashboard for the source data.

## **3.2 Analytics on the sales data**

We have created below tableau dashboard. Please do interact with same dashboard using the below link.

<https://public.tableau.com/app/profile/abhijna8270/viz/Salesdashboard_17243901864020/SalesDashboard>

Sample image of the dashboard.



There are few key takeaways from the analysis:

* UK has the highest sales and highest number of orders.
* There are 250 orders have been placed and have an order value of 0.5 million.
* UAE has the less customer base but still it contributes to delayed shipments
* USA is second in terms of customer base and sales has better shipment completion compared to UK and UAE.
* Most sought products are mouse pad and keyboard

## **3.3 Project future scope**

We have designed the project such a way that it’s fully automated and requires minimal support.

But the project can be extended with addition more datasets and many insights can be derived such as customer cohort analysis, repeat customer ratio, sales forecast.

We have created separate documents for Data Engineering team where we have proposed additional project scope.