**BUSINESS INTELLIGENCE**

**MID TERM**

**Name**: Shaheer Ahmed Farooqui

**Erp**: 19722

**Dataset**: Ecommerce dataset

PROBLEM STATEMENT: What are those factors that are responsible for people taking interest in buying from their website?

**ABOUT DATA**

**KPI:**

Status

MV (market value)

Grand total

**Dimensions:**

Product Category

Payment Method

Order Status

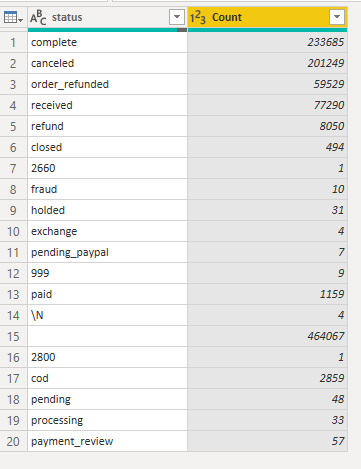
Working date

**DATA TRANSFORMATION**

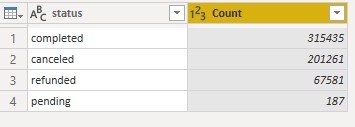
1. **Status Column**

This column had multiple dirty values like if a order was canceled then there many different entries were found in this column with different names that had the same meaning as order canceled so in order to tackle that all the dirty values were removed and only 4 values were kept in the column which were COMPLETED, CANCELED, REFUNDED and PENDING.

**The data before cleaning**

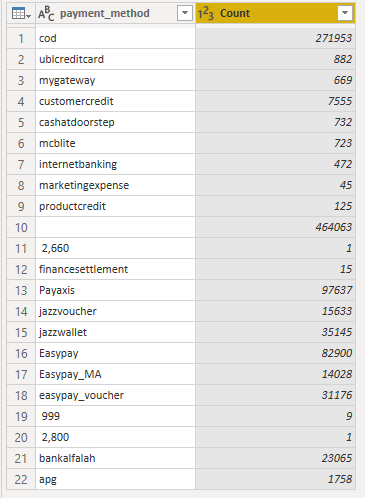


**Data after cleaning**

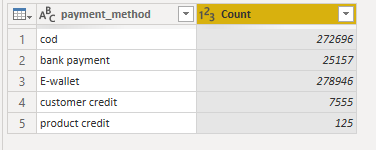


1. **Payment Method Column**

**Data before Cleaning**

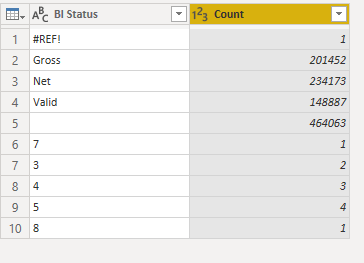


**Data after Cleaning**

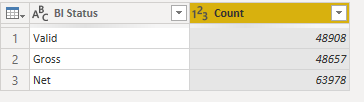


1. **BI status column**

**Data before cleaning**

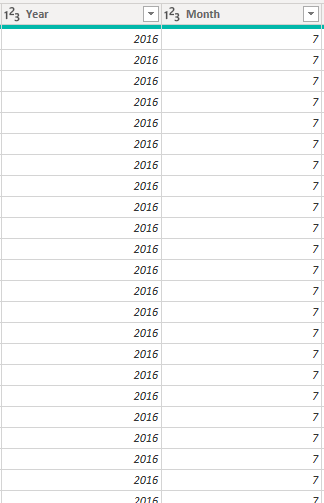


**Data after Cleaning**



1. **Year and Month Column**

Year and Month column were removed from the dataset as a separate date column was already present in the dataset so the Year and Month as two separate columns were unnecessary for the dataset



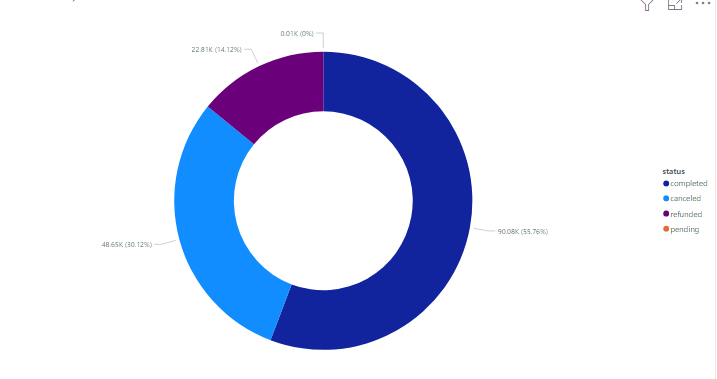
**OTHER POINTS**

* There were 460k null values in each column, so in order to tackle that all the null values of a column were removed by using the REMOVE EMPTY option of powerBI. Removing null of one column removed the null values of all other columns as all the columns has 460k+ null values in common.
* Incremental\_id column was removed as it was not necessary for the dataset.
* Date columns like working\_date has some errors like the format error which was tackled by using the REMOVE ERRORS function in rows.
* The name of price column was changed to price/unit
* Total column was removed and a new column was added with the name of grand total and a formula was used to calculate the values of that column. The formula was

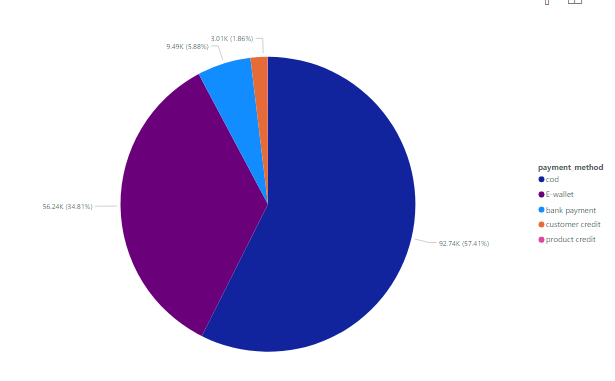
Grand total = (price/unit \* quantity) – discount

* There were some negative values in the grand total column which is because of when discounted amount was greater than the MV value so that negative value was changed to zero.

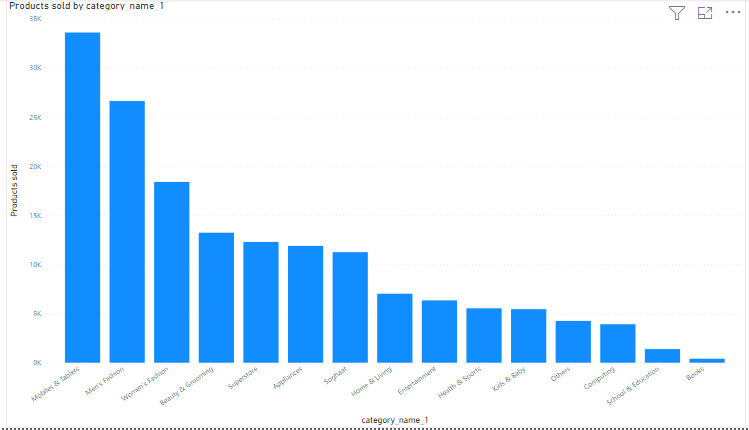
**CHARTS**



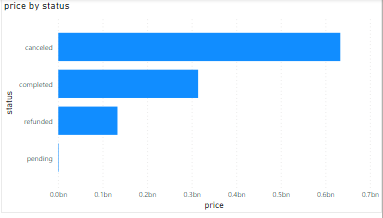
This Donut Chart is indicating the number of orders that were completed, canceled, refunded or pending to get an idea about the progress of the company



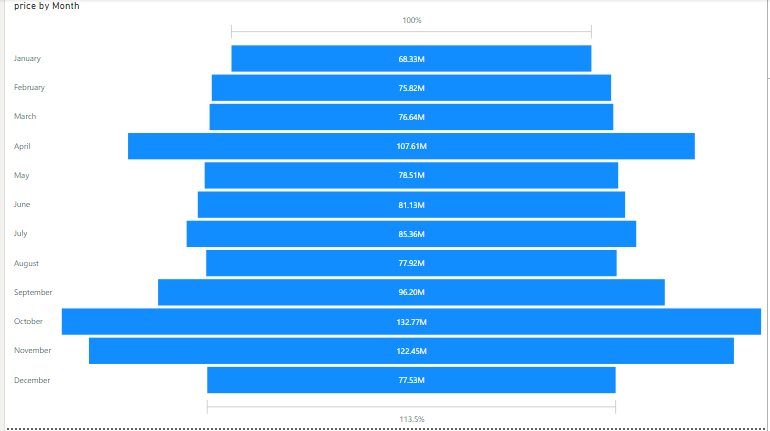
This Pie chart is showing the payment method of the orders received like whether it was cash on delivery, E-wallet , bank-payment etc. It can be used to predict the preference of the customers that how majority of them would like to pay for their orders.



This chart is showing the number of products sold in each category of the Ecommerce website. It will help the company to get an idea that which category is sold the highest number of times so that they might introduce new items in that category to boost their sales.



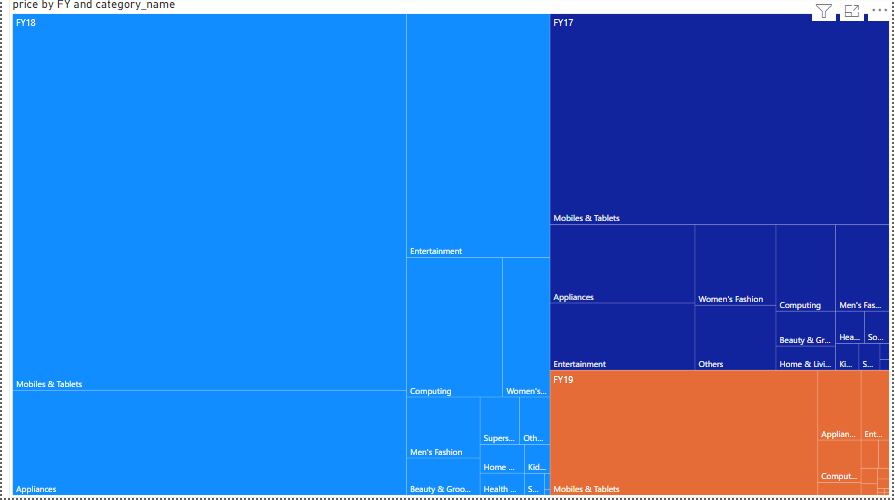
This chart is indicating the price value of the orders that were completed, canceled, refunded or pending. It shows that the value of orders canceled was more than the total value of completed, refunded and pending status.



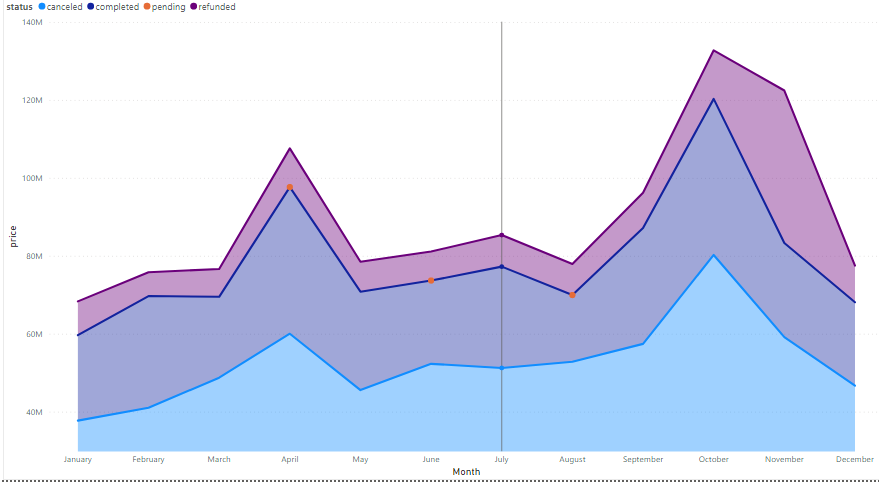
This chart is showing the revenue generated by each month. It can be used to check the progress whether the revenue generated each month is increasing or not.



This tree map is showing the revenue generated in each category. It can be clearly seen that the highest revenue generated is from the mobiles and tablets category.



It is showing the products sold in each category by fiscal year 2017, 2018 and 2019.



This stacked area chart has Month on its x axis and price on y axis and the value field has status column. This is showing the revenue generated from each status each month like how much the revenue was generated In January on completed orders.