

Pakistan E-Commerce Conclusions and Summary:

By: *Abdalla Salama*.

Data Source: [Pakistan Largest E-commerce Dataset](#).

Data Overview:

The data base had 1048575 rows and 25 columns, after the data cleaning process, it contained 584524 rows and 16 columns.

The important features of the cleaned data set are:

Price, Quantity Ordered, Status, Category, Payment Method, Market Value (MV), Month-Year, Customer ID, and Product ID.

Data Exploration:

Observations:

1. Price: The mean price is 6348, and the prices range from 0 to 1012625, with the 899 median.
2. Quantity Ordered: people tend to purchase orders with the average of 1.3 items, as there are orders with 1000 items.
3. Status: There are 16 different statuses, the most frequent one is the "complete" status, with almost 40% of the orders.
4. Product ID: There are 84889 products, and the most popular product is "MATSAM59DB75ADB2F80" in terms of orders not quantity ordered (another product may be ordered in higher quantity but in less frequency).
5. Category: There are 16 different categories, and the most popular one is Mobiles & Tablets.
6. Payment Method: There are 18 payment methods, and the most popular payment method is COD-Cash on delivery, with almost 46% of the cases.
7. Fiscal Year: most of the orders were placed in the fiscal year 2018 (1-Apr-2017 to 31-Mar-2018).
8. Customer ID: There are 118327 different customers, customer 85775 is the top customer with over 2520 orders (another customer may have ordered in higher quantity but in less frequency).

Conclusions:

- People tend to place orders with low items quantity.
- The most purchased items are Mobiles and Tablets.
- People pay in cash most of times.

Data Visualisation:

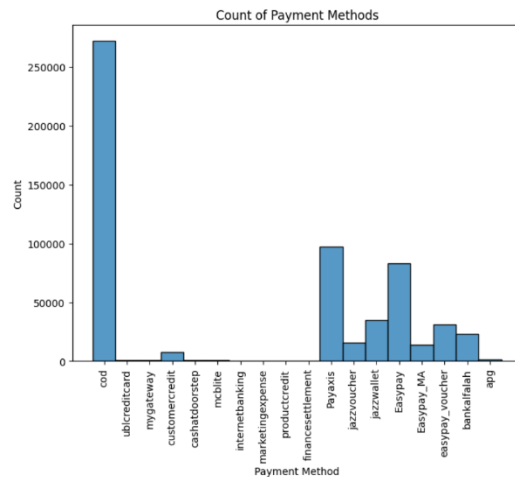


Figure 1.

People in Pakistan tend to use cash in most cases (~46% of the orders are paid in cash).

Using this information we may develop a more secured method of cash payment, such as guarantees from the client, or encourage the clients to purchase using a credit card.

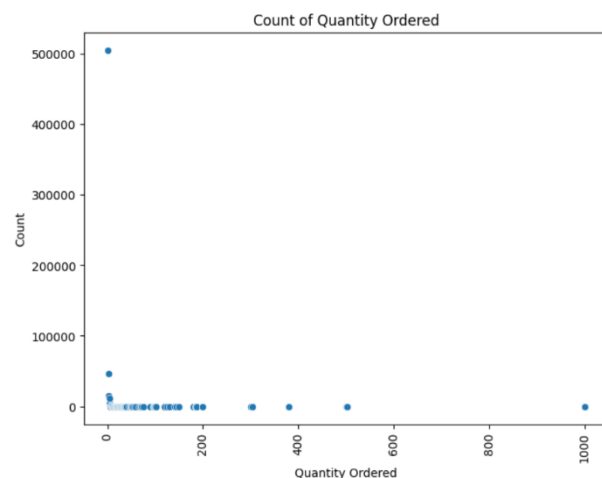


Figure 2.

Most of the orders contain low quantity of items, there is an outlier of 1000 items in a low number of orders (6 orders).

We can see that there are different types of users in the site, we can infer that there are individual buyers and large-scale buyers (traders), But the most are individuals. It would be more profitable if the client buys more than one item in a single order.

Status Distribution

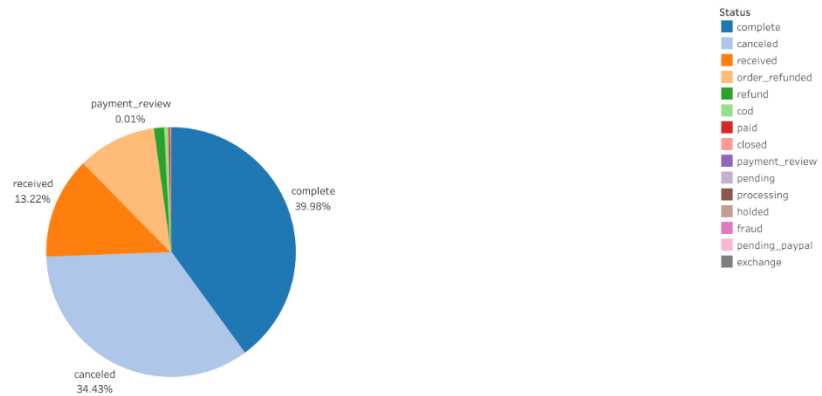


Figure 3.

Almost 40% of the orders are complete.

An interesting observation that almost 34% of the orders are cancelled! A fact to be investigated to find the causes.

Quantity Ordered by Category

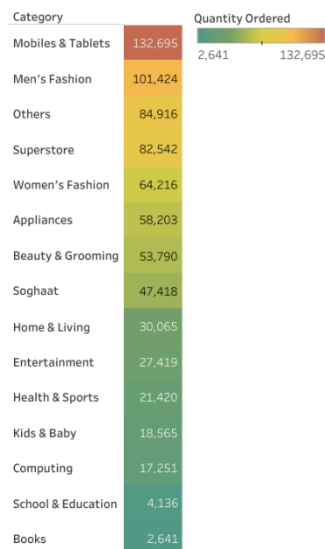


Figure 4.

The most common categories are Mobiles and Tablets, Men's Fashion, and "Others", whereas the least common are Computing, School & Education, and Books.

Category Discount Per Unit

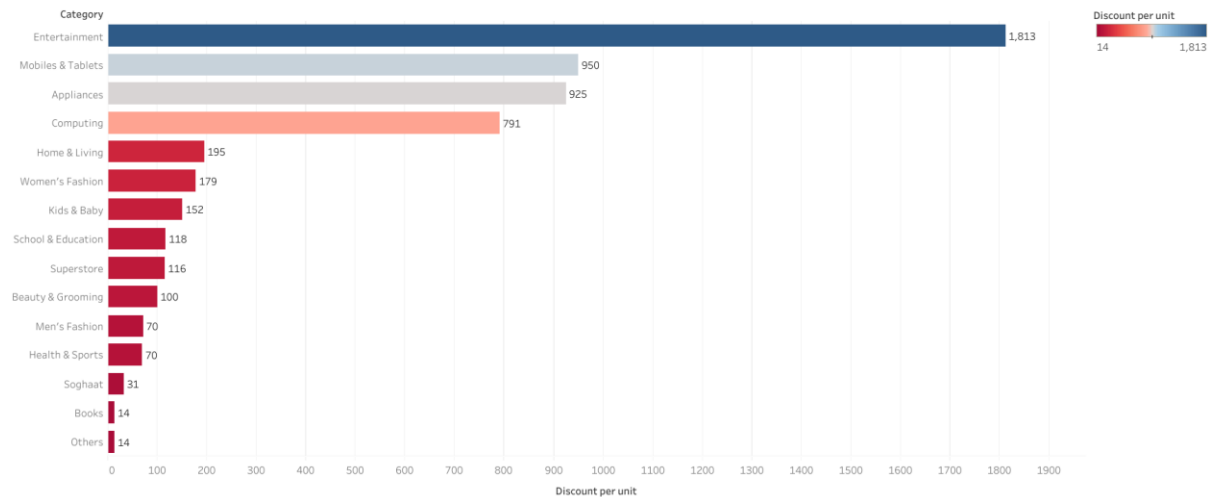


Figure 5.

Here we calculated the Discount per Unit by dividing the sum of discount by the sum of quantity ordered. We can see that the category with the most discount per unit is the Entertainment, whereas in the previous visualisation, the Entertainment category is the 10th in quantity selling. In contrast the Men's Fashion category is 2nd in quantity sold but with almost very little discount per unit.

Profit by Category

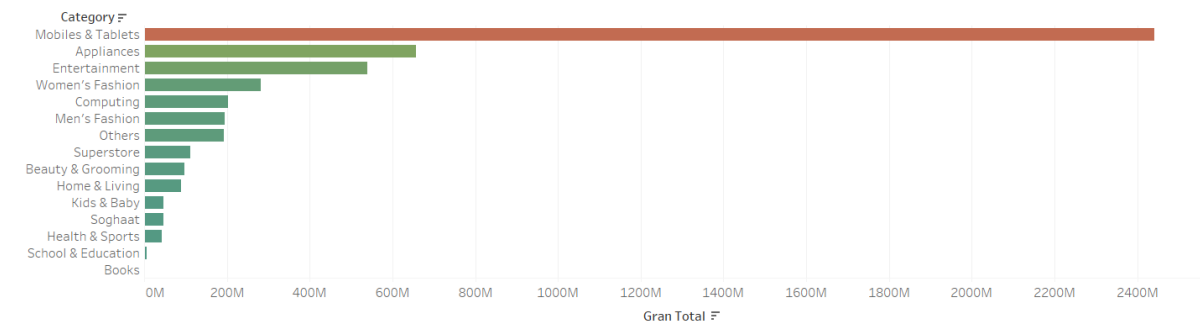


Figure 6.

Here are the most profitable categories, Mobiles & Tablets is in the lead, Entertainment is 3rd, considering it has the most discount per unit and 10th in the quantity sales. Which means the prices are very high in the entertainment category. Very similar results are observed in the Computing category.

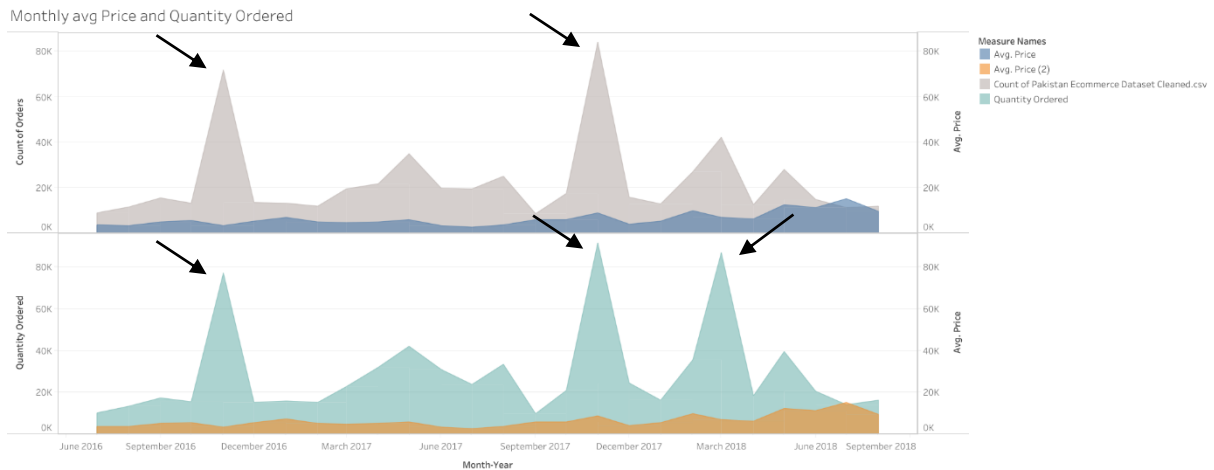


Figure 7.

In this visualisation we can see the count of orders in each month at the top, and the quantity ordered at the bottom. We note 2 distinguished peaks in both graphs in November in both years 2017 and 2018. This peak in orders may be related to the holidays period when people gift each other.

Another observation that in March 2018 there is a third peak in the quantity ordered, although the peak of the orders count is significantly lower. This indicates that there are orders with a high quantity of items in each order.

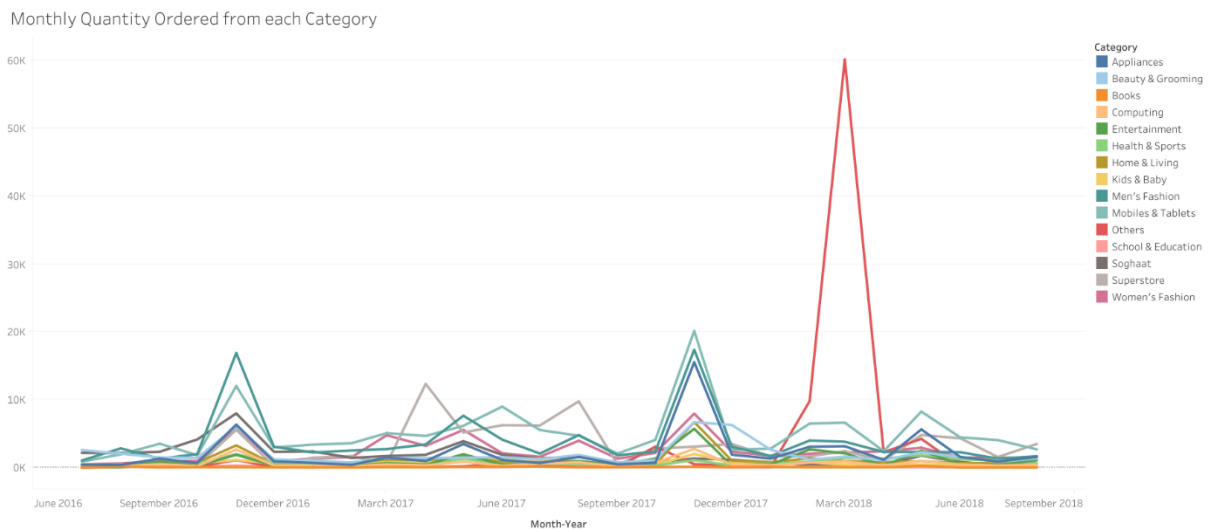


Figure 8.

To study this peak, we first look for the category responsible for the peak, if there is one. As we can see there is an abnormal peak in March 2018 in the “Others” category, almost 60K items were ordered. This peak drops vastly as soon as we enter April 2018.

"Others" Category and Payment Methods

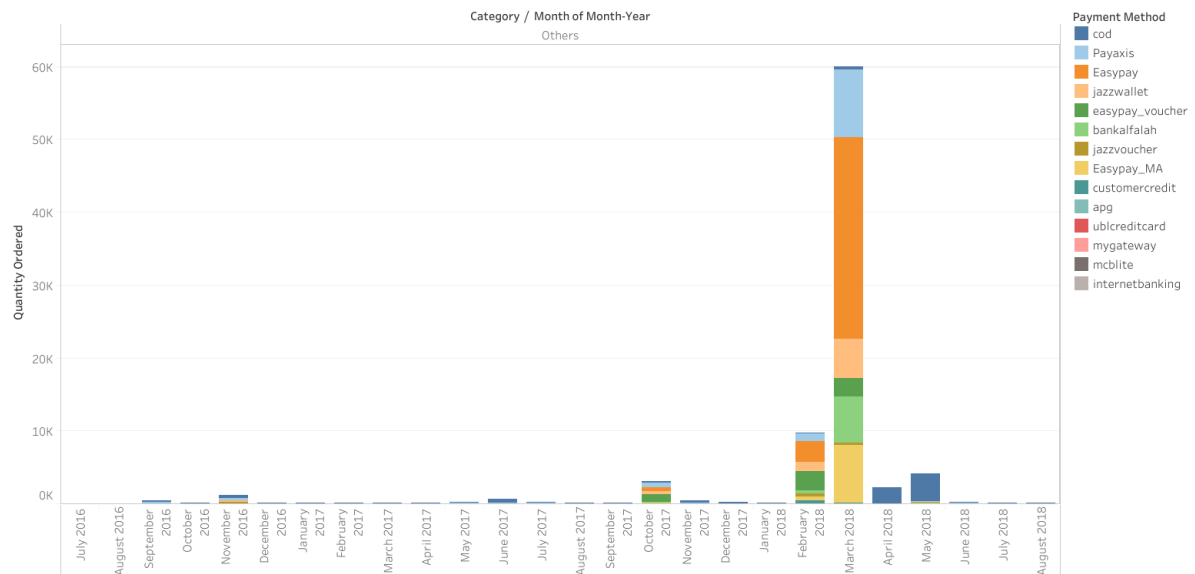


Figure 9.

If we go further with the "Others" category analysis and try to figure what Payment Method was used the most during March 2018, we discover it's Easypay.

"Others" Discount per Month

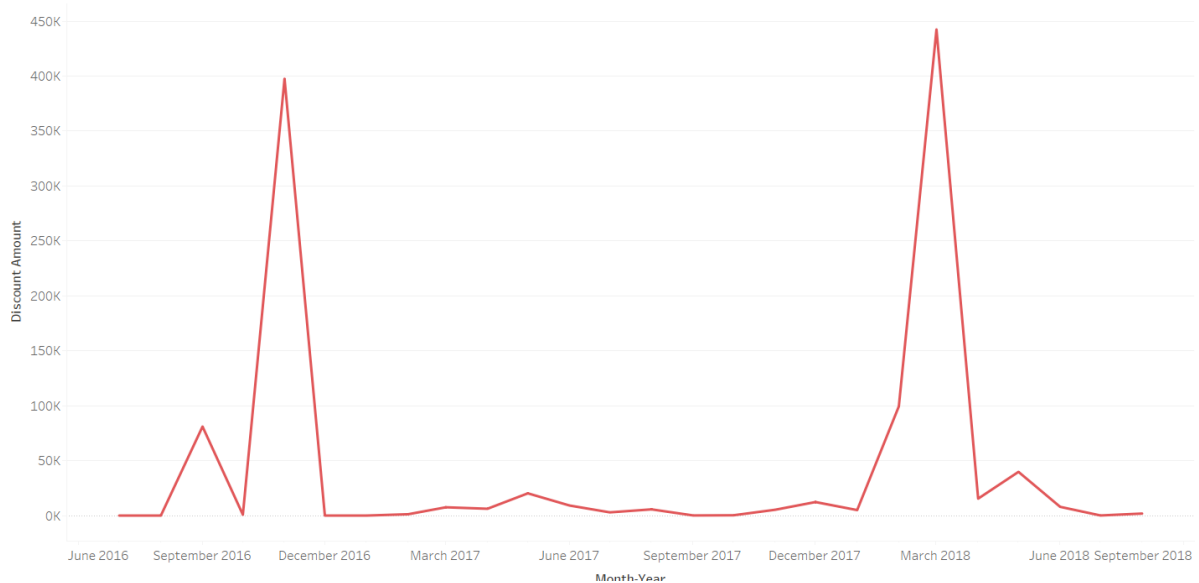


Figure 10.

The reason for the peak in the "Others" category in March 2018 may be the huge discount made in this month as shown in the graph above.

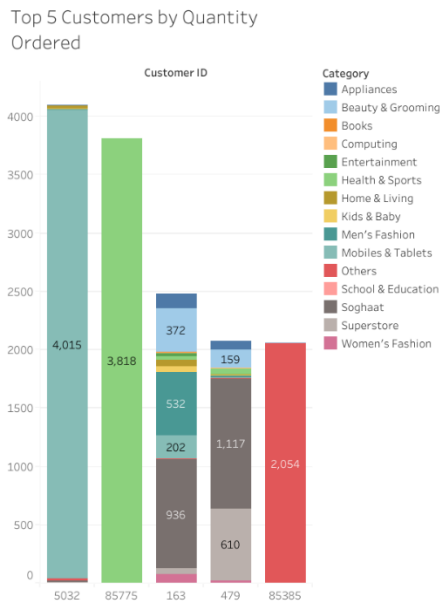


Figure 11.

Here are the top 5 customers with the most quantity ordered, 3 of which bought mainly from 1 category, customer 5032 mainly bought Mobiles & Tablets (maybe a mobile store). Customer 85775 mainly bought Health & Sports items (maybe a sports store). Customer 85385 bought only from “Others” category.

As mentioned before, we can conclude that the buyers in this site are not only individuals, but also stores. We can aim the marketing towards them also, not only toward individual buyers.

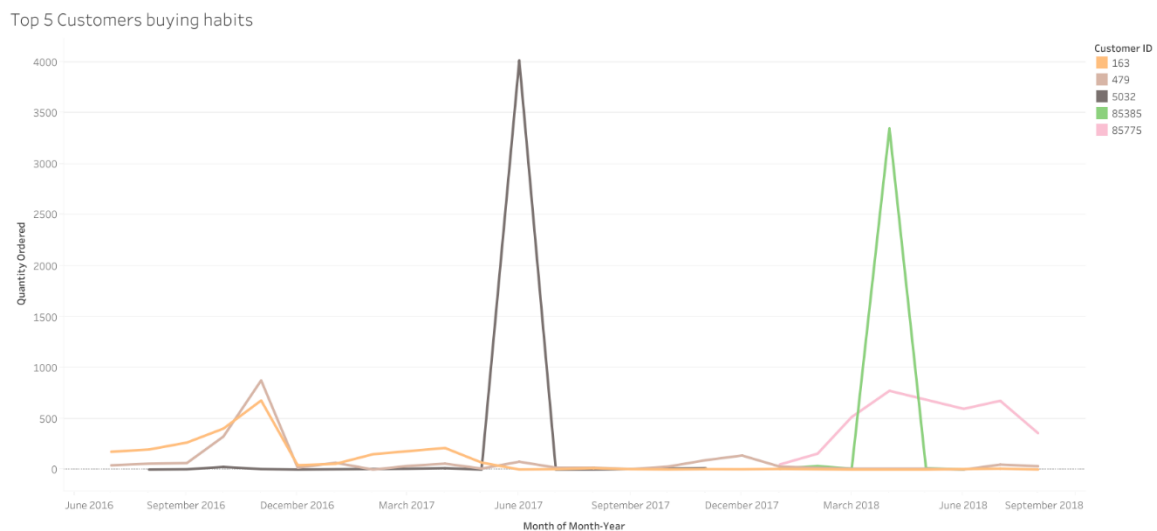


Figure 12.

The analysis of those top 5 customer purchases shows that they make a big purchase occasionally and not so frequently.

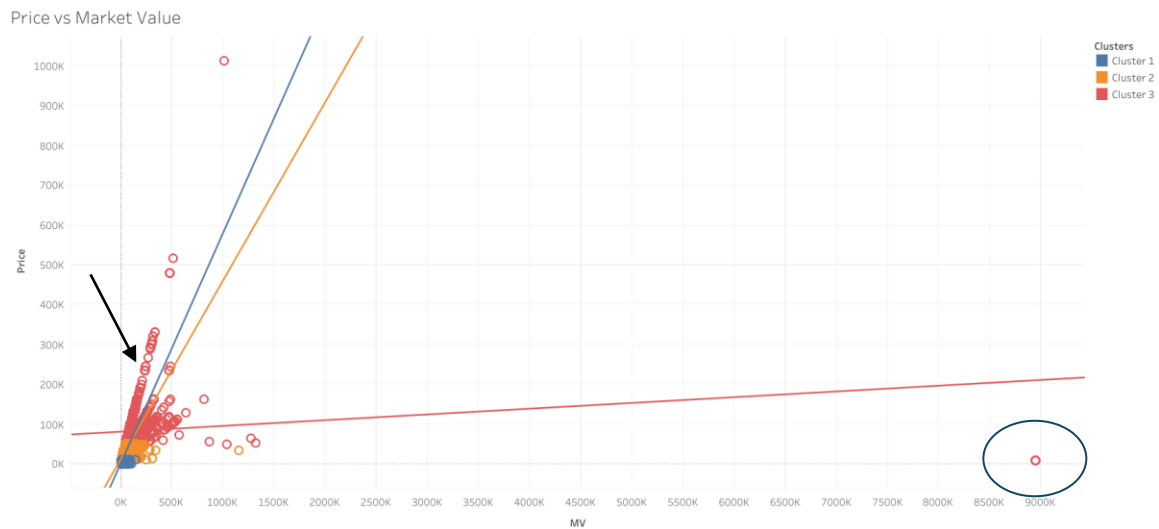


Figure 13.

In this graph we performed clustering on both Market Value and Price, it was divided into 3 clusters. In all of which the Market Value is higher than the Price. In cluster 1 (Blue), the prices are closer to the MV, in cluster 2 (Orange) they are less close to the MV, and in cluster 3 (Red) they are slightly affected by the price, this occurs despite the obvious line (pointed by the arrow) with the matching values between the price and the MV. The reason for this is the outlier in the right lower corner. If we check the values, we get:

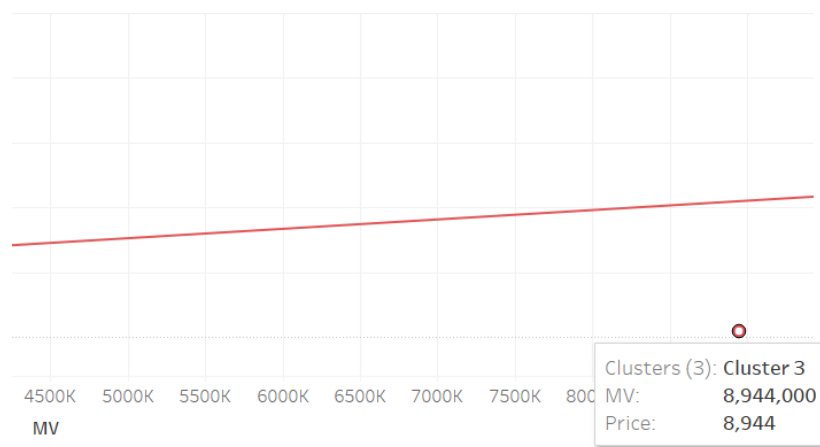


Figure 14.

This means that a typo occurred when inserting the Market Value of this item, and instead of writing 8944, the inserted value was *1000 higher.

Price vs Market Value (2)

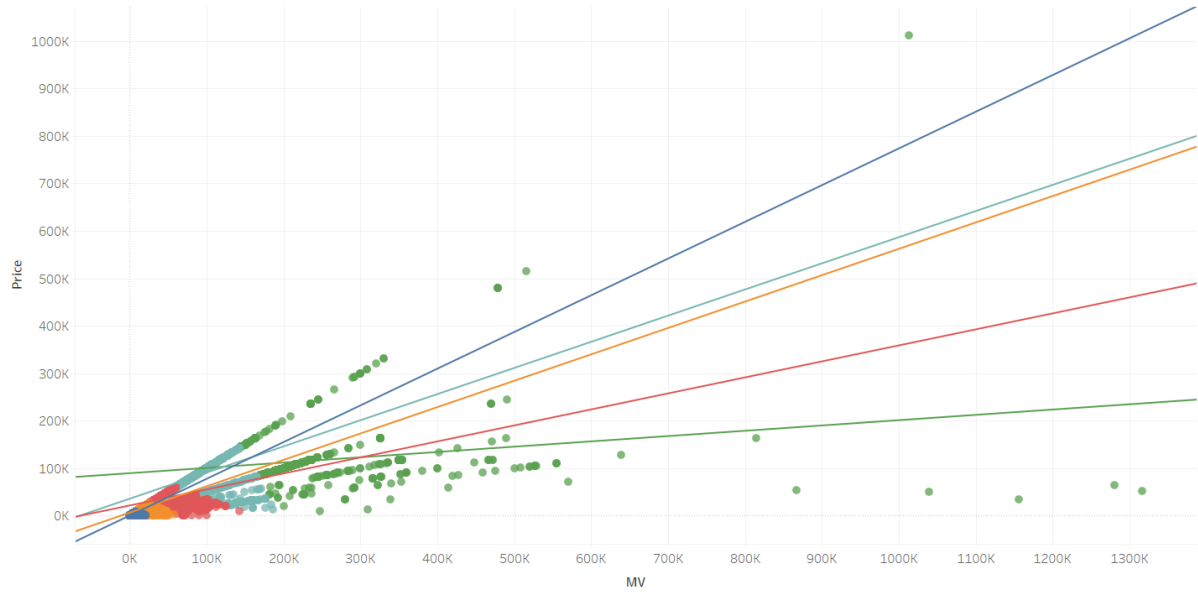


Figure 15.

After removing the outliers, the number of clusters automatically changed to 5, showing more obvious difference between the clusters.

Payment Method and Status

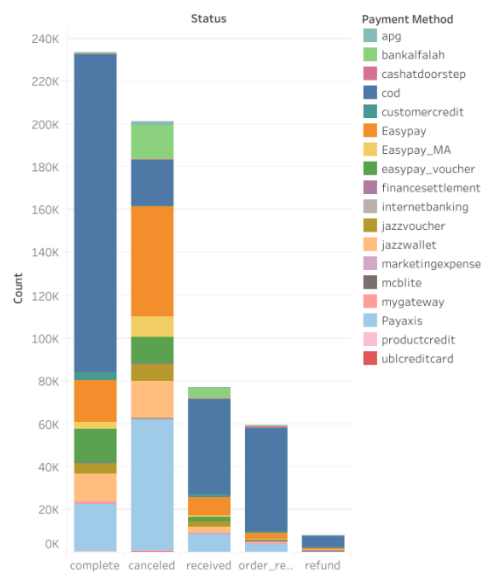


Figure 16.

From this visualisation we can determine the problematic Payment Method by looking at the cancelled status orders count, both Easypay and Payaxis are prominent in this status. Financial department may take further actions with both companies.

Summary:

- 46% of orders are cash paid. A more secured payment method may be safer ^[figure 1].
- 34% of the orders are cancelled, looking at the payment methods of the cancelled orders we can see that they are performed by Easypay and Payaxis. Further actions by the financial department may help solve this issue ^[figure 3].
- Some of the categories present high profit values coupled with low discount amounts. Which means that these categories are hot sellers or very frequently used or necessary products. Thus, to maximise the profit, big discount isn't needed ^[figures 4-6].
- In November of each year, there is a peak in sales, which may be caused by the holidays when people gift each other. Considering making discounts before the holidays to increase the sales ^[figure 7].
- A prominent peak was observed in March 2018 in the quantity ordered graph, despite the peak being much lower in the count of orders in the same timing. Which indicates the presence of orders with high quantity of items ^[figure 7].
- Further investigations show that the peak was caused by abnormal increase in "Others" category purchases. These purchases were paid mainly using Easypay, Easypay-MA, and Payaxis methods. One possible explanation for the peak being caused mainly by the "Others" category is the discount made on the category in this month ^[figures 8-10].
- 3 of the top 5 customers are large-scale buyers of a specific category. They also perform such purchases infrequently and in a one large-scale order. It's important to keep such customers because of the large orders they perform, consider discounts and rewards for such customers ^[figures 11-12].
- An as believed to be a typo was discovered in the Market Value (MV) data after grouping the Price and MV into clusters ^[figures 13-15].
- As mentioned before that both individuals and large-scale buyers (stores) use the site to place orders, a marketing and advertising plan should be directed to both types of customers. Denoting the benefits of buying from the site such as multiple payment methods and the fact that the site is widespread.