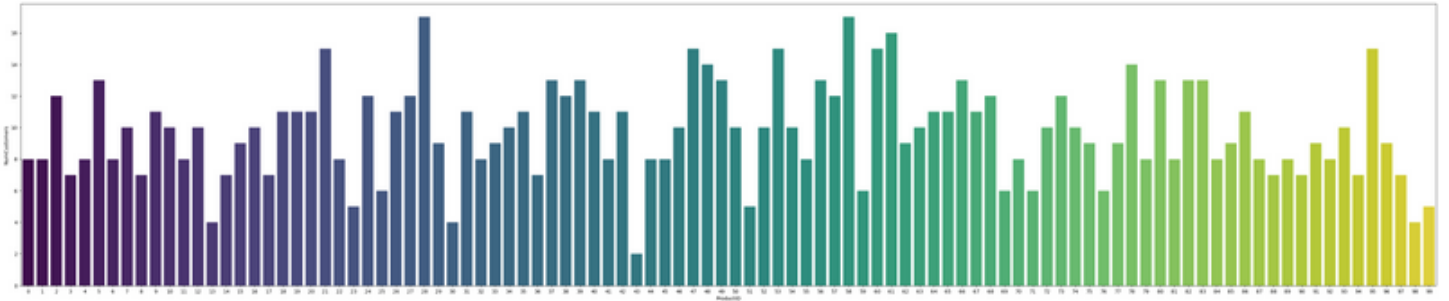# Exploratory Data Analysis (EDA)

**1.Finding number of customers will buying same product:**

on the above analysis my target is to find the no of customers will buying same product.
by observing above bar plot , i've been found that there is no only single customer will
buy only unique product which means not buying by others. the advantage of that
observation is if there is no customer will not buying unique products which is not buying
by others ,we can easily find the similar customers. all customers will have minimum
similar 2 customers and maximum 17 similar customers .

**2.Most number of transactions by each customer:**
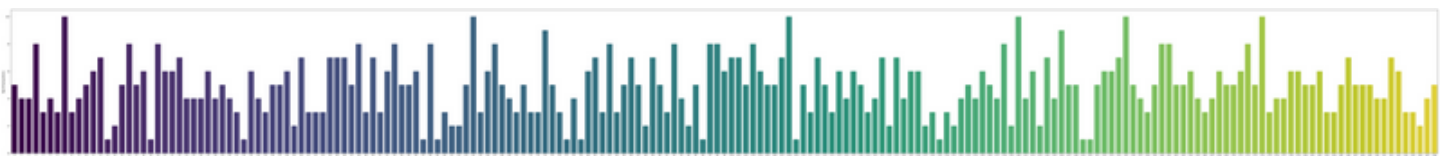
on the above analysis my target is to find the no of transactions done by each customer.
so i have been construct the bar plot ,in that on the x axis is CustomerId and on y axis
Numtransactions . by observing the above bar plot the mimimum number of transactions
done by each customer is 1 and maximum number of transactions is 10. so by this analysis
it is easy to find the similar customers by having the data of customers having more
transactions and it is hard to find the similar customers by having the data of customers who
have less transactions . but we have both type of customers who have more transactions and
less transactions. so we can get better data to build model for getting accurate result.

3.Finding Higher quantity transactions tend to have higher total value:
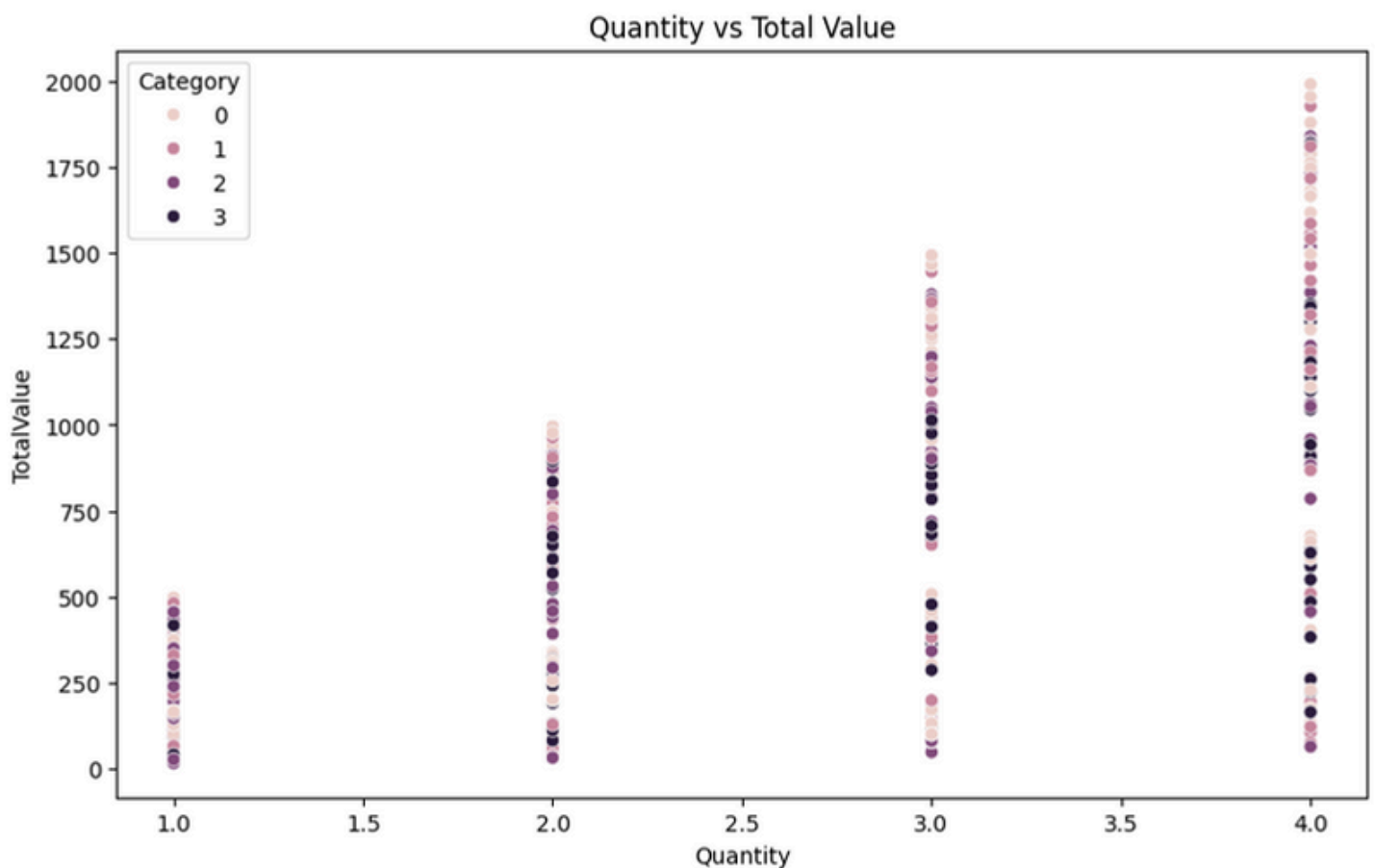
on the above analysis , there are three product categories in four colors . on the X-axis i will take totalValue and on the Y-axis i will take Quantity . on the above analysis i just conclude that category 0 wich is sandle color will have more transactions . and after that category 3 wich is black color will have more transactions . on the transactions of high quantity then the totalValue is also incresed . if we observe on the above graph if quantity is will increasing then the totalValue is also increased.so there is a relation between totalValue and Quantity. if in our data will have more relationships the we can easily build the better model .which is we can build the model wich can give more accurate result.

4.Price distribution for finding skewness:

Price Distribution
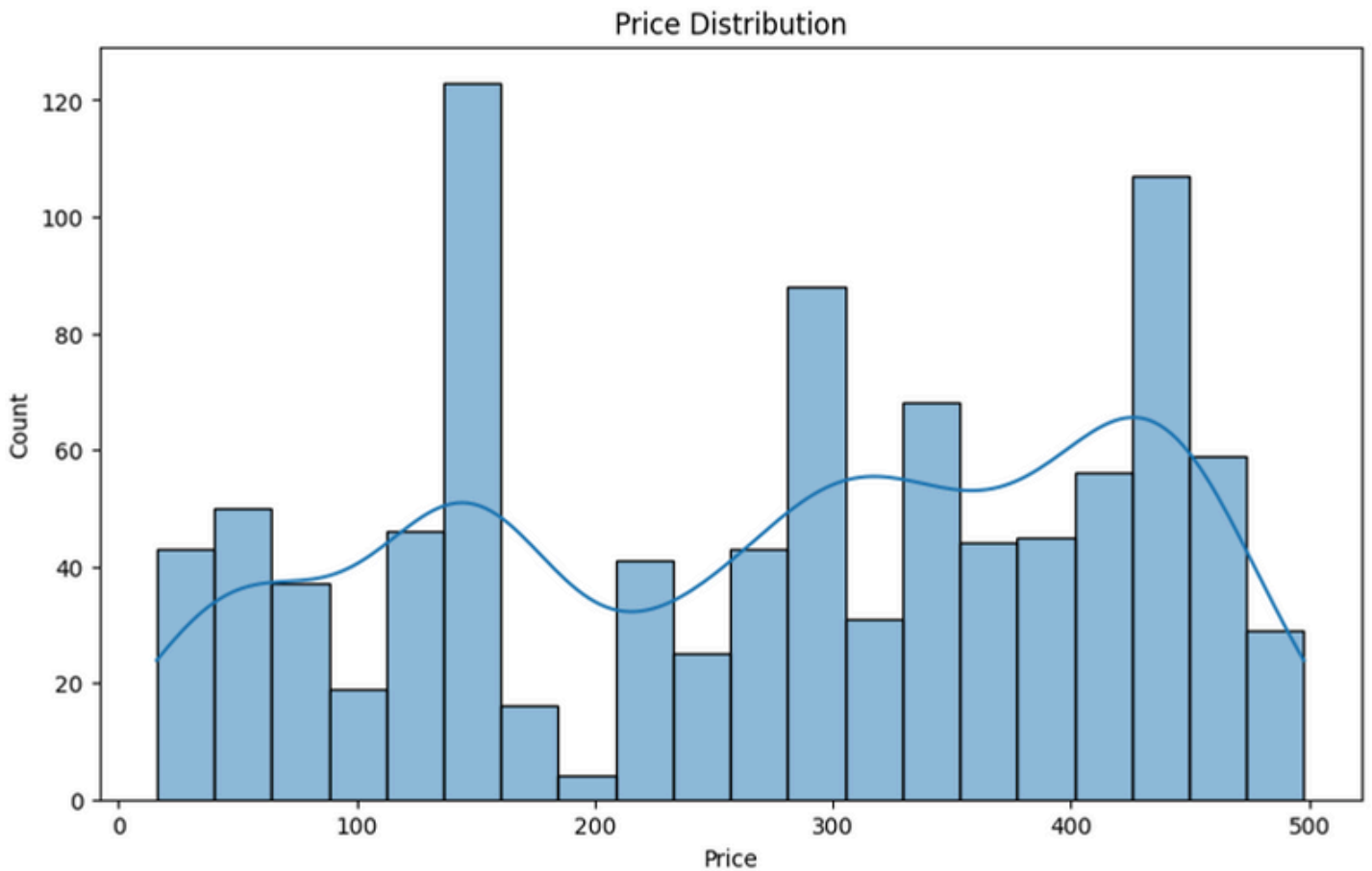
The price distribution shows a right skew, meaning a few products are highly priced.so the price distribution is positively skewed . in the positively skewed distribution the tail on the right side which are smaller values is longer than tail on the left side which are smaller values. this means majority of the data points are concentrated on the left side of the distribution and there are some extreme values on the right side . if there is no skewness in the price ditribution then we can get some accurate result .

5.clustering the customer into three categories using k-means:

Customer Clusters Based on Price and Total Value

on the above graph there are three clusters which are voilate ,green ,yellow .by analyse that graph the data points are colliding which means the three clusters are colliding which have smilarity on that clusters .clusters are overlapping, it means some customers have similar purchasing behaviors across different clusters. This suggests that strict segmentation might not be possible, and a softer approach using similarity scores or hierarchical clustering could work better.

done by

Chataparthi G V Satyanarayana

9441827299

ganeshsatyanarayana7@gmail.com