Business Insight Report from EDA task

EDA process used:

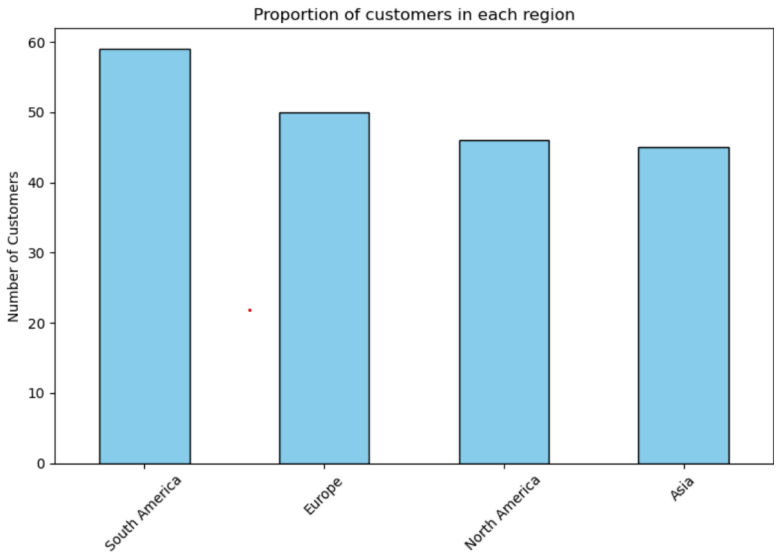
From the 3 datasets given in csv format, in order to do an accurate EDA (Exploratory Data Analysis), I also needed to preprocess the data to check for different important aspects of the dataset, such as null values present in the dataset, and duplicated values present, and the different categories present in certain features of the datasets. For the customers table the basic preprocessing was done, along with visualizations to see the proportion of customers in each region, the sign up trend over time, and a scatter plot of region vs the signup date.

For the Products table, again preprocessing was done, along with basic visualizations to see how the different features are interacting with each other, so a count plot was used to see which category of product had the most number of products, along with a box plot to see the distribution of price along with seeing if there are any outliers present, and lastly a histogram was also used to see the price distribution.

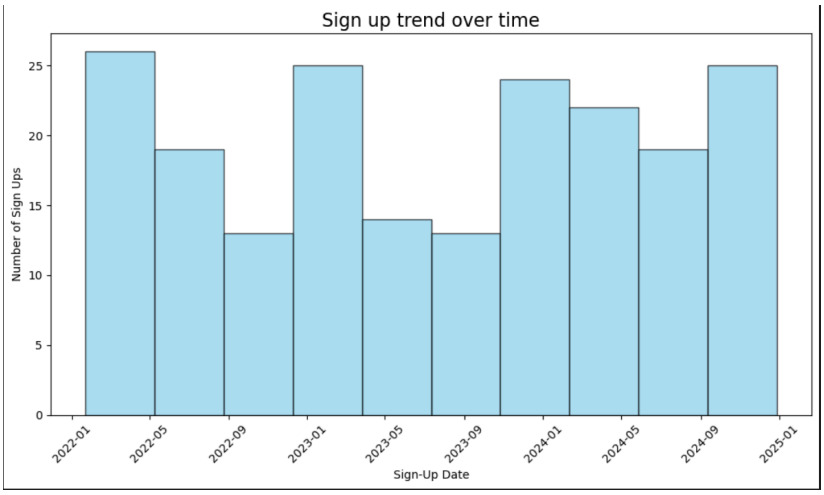
Lastly for the Transactions table, along with basic preprocessing, a histogram is used to see transaction value distribution, the distribution of the total value in transactions. The correlation heatmap of the numerical values in the dataset, to see how well they are correlated to each other.

Business Insights

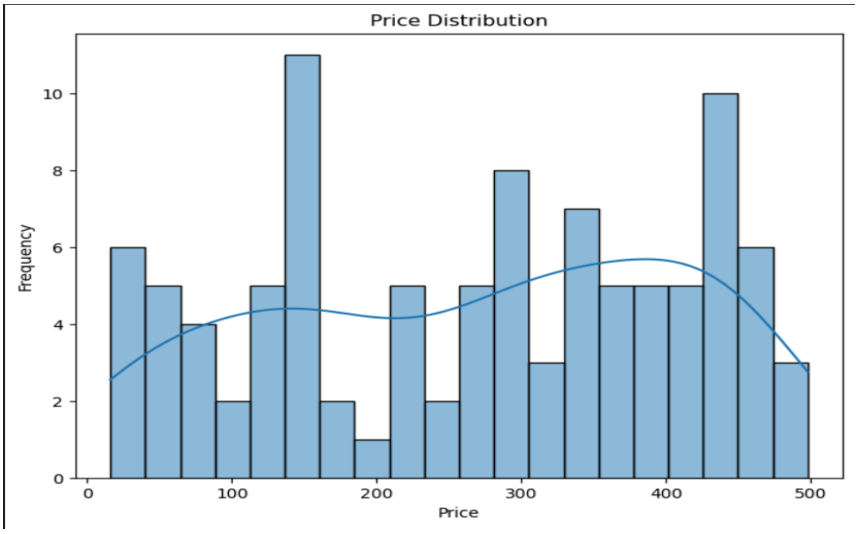
Customers table insights:



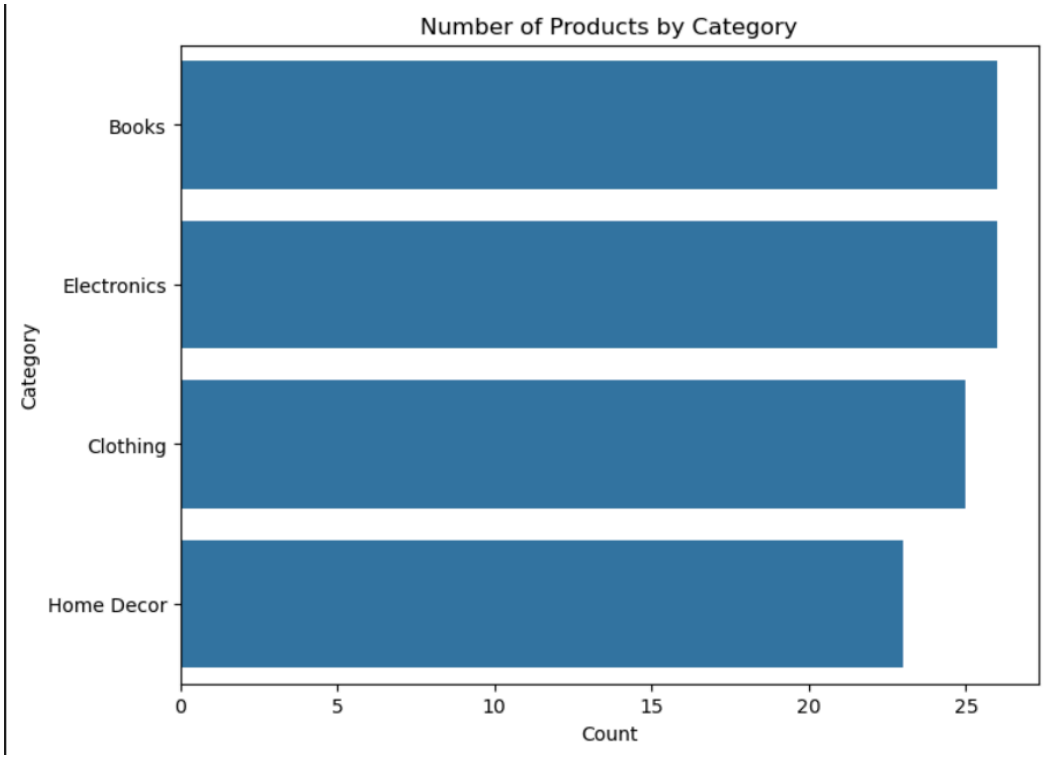
1) From this chart I am able to see a certain trend of how the different regions are influencing the number of customers according to the dataset, by having this information, We are able to see in which region we should probably focus on more as South America has more number of customers at around 60 in the count of customers, compared to other countries which have only around 47 in the count of customers, so comparatively South America has 22 % more customers compared to the other countries hinting towards a higher customer base in South America



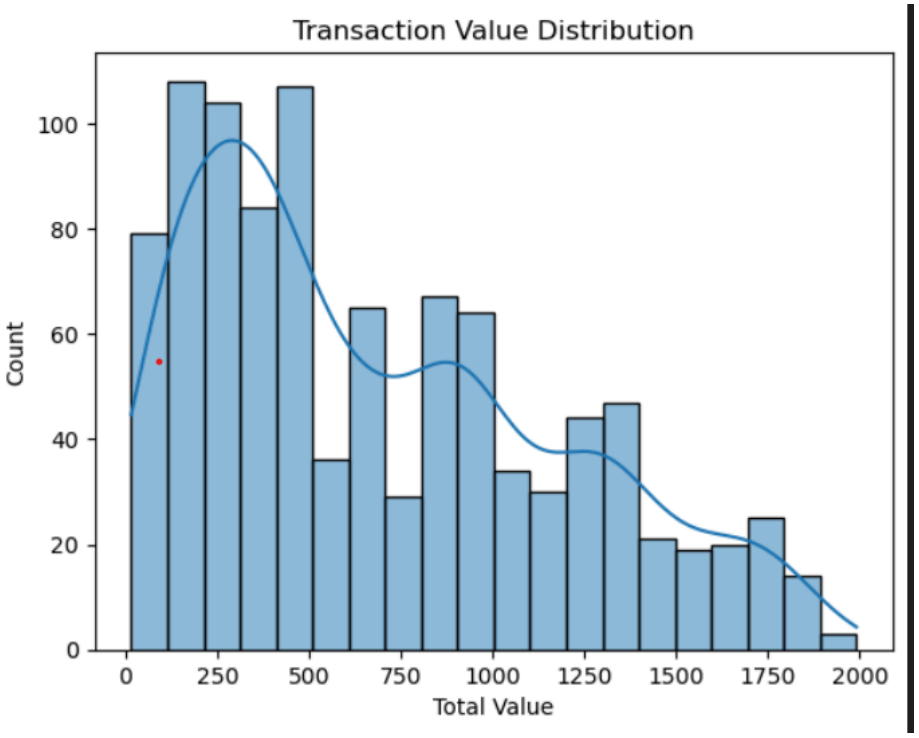
2) From this chart of the Customers table I can see that in the months of January of each year were the peaks of the number of sign-up as recorded which shows a certain trend in the data, and a pattern which could easily be used to increase the revenue during these particular months in the entire year. However, only these months are the peaks as the number of sign ups reduces from 25 to 20 and then 13 only in around 9 months time.



3) From this chart above the insight which I am able to see clearly is that around the price of 150 is the highest peak of the count of product sold at around 12 units, however another second highest peak can also be seen around the 450 price point at around 10 units, so focusing around these points would be a very beneficial decision in order to maximize sales and amount of products being sold.



4) From the above chart I am able to see some interesting insights such as how many products are seen in the data with basic appliances like electronics and books having the highest number of entries in the dataset at around 25 units, with clothing and home decor having around 20 units. Which then also shows where most of the inventory of the products are in which category.



5) Lastly from the above visualization of the transaction data I am able to see most of the total value which is being generated is rightly skewed, and shows where the most total value is being generated, and around the 1750 and 2000 range very less total value is being generated as seen from the dataset.