**Superstore 12 Months Sales Analysis (EDA)**

**Context**

It is important that the retail stores or retail companies keep track on their sales data and analyse that to find customer needs, trending products, purchase behaviour and many more important attributes about their sales and business to function even better and efficient both commercially and financially.

**Content**

The data set had multiple files regarding each month of the year, so firstly I merged all the files into one to get all data in a single file, resulting in a huge data file with millions of records of sales by the Superstore over the year.

It contains not only numerical data but also geographical and Date & time data of the customers and their purchases from the superstore. Below is the list of columns from our dataset.

**Order ID –** Unique identification number of purchase/order.

**Product –** Name of the product purchased by the customer. Only one product in one cell

**Quantity Ordered –** Quantity of the product purchased.

**Price Each –** Price of one piece/quantity of the product purchased.

**Order Date –** Date of the purchase

**Purchase Address –** Address of the store the purchase has been made with postal code, city name and street name/number.

**Exploratory Data Analysis**

Exploring the dataset , checking the statistics, and treating the missing and invalid values. The data set not only consist of missing values but many invalid values as well. The presence of invalid data entries was discovered while applying mathematical formula on a column having integral entries. The column had some rows consisting of string data set which were not processing the mathematical formula and thus I could treat all the hidden invalid values from the complete data set.

After this in order to answer few business question, we needed to create some new columns by extracting data from the existing ones. For this, we had to have the columns in our dataset be in appropriate data type. As in, the column consisting of the date and time of purchase must be in date and time data type, and so on. So, I converted the data types of the features to correct one, and then extracted the required data from the already present columns applying relevant functions and using array slicing methods and lambda methods to get the desired columns through the current ones in our data set.

At this point, we now have a well structured and well elaborated data set about the sales of the superstore in last 12 months. We are now supposed to use this data set to analyse some business problems and answer some business problem statements.

Columns added during analysis:

**Sales -**  Quantity Ordered \* Price Each . Float type column

**Month –** Month in which the purchase was done out of 12 months , integer type column

**City –** City in which the purchase was done, string data type.

**Time –** Time at which the purchase was made from 24 hours in a day. Date and time data type column.

**Hour –** Hour at which the purchase was made from 24 hours in a day. Date and time data type column.

**Minute –** Minute at which the purchase was made from 60 minutes in an hour. Date and time data type column.

**Business Questions Answered and Visualizations**

Q1. What was the best month for sales? How much was earned that month?

Chart, bar chart

Description automatically generated

Q2. What city has the highest number of sales?

Chart, bar chart

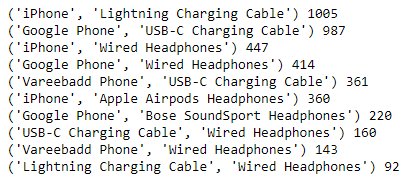
Description automatically generated

Q3. What time should we display advertisements to maximise likelihood of customer’s buying products?

Chart, line chart

Description automatically generated

Q4. What products are most often sold together?



Text

Description automatically generated

Q5. What product sold the most? Why do you think it sold the most?Chart, histogram

Description automatically generated

**Conclusion**

Following points are the conclusions or the insights from our EDA of the superstore sale data.

1. The store is selling the maximum products in the month of December, followed by October, April and November. The reason of maximum sale in December may be Christmas as so many people buy gifts for friends, family and loved ones on Christmas.
2. The superstore has made maximum sales in the city San Francisco followed by Los Angeles and New York City. The reason of high sales might be high publicity/marketing in the city or higher no. of branches of the tore in that city or longer period of service of the superstore.
3. We should display the advertisement of products or offers available in the superstore around 11:00 am and/or around 19:00 pm to attract the maximum public as these are the two times having maximum number of sales as per the data. There can be many reasons for high sales during this time namely, people go home from office around 7 pm and may be buy stuff on their way to home.
4. As per our data the 2 products sold together mostly are iPhone and Lightning Charging Cable. When we talk about 3 products sold together maximum time, we can say that Google Phone , USB-C Charging Cable, and Wired Headphones are the products sold most together.
5. The highest sold product is "AAA Batteries" and the lowest sold is "LG dryer" and "LG Washing machine". On further analysis the possible reasons for the result are found to be lower price for AAA Batteries and higher price for Dryer and the Washing machine and the secondary reason is the higher demand in public.