ONLINE E-COMMERCE DATA ANALYSIS REPORT

Made By: Yaseen Saeed

Contents

1. Describe the dataset

This data is about E-commerce and customers' behavior, imported from Kaggle website

2. Wrangle The dataset

2.1. Data Assessing

2.1.1. Data Cleanliness

Invalid Dtypes:

The data has 14 column x 5110 rows. Some data are of wrong dtypes, such as:

Order number: This is a ID. It should be of type Object

Order Date: This is a date. It should be of dtype DateTime64

Completeness:

There are 15 missing rows

Uniqueness:

There are 14 duplicate rows

Accuracy:

No illogical data found

Consistency:

Motherboard is repeated twice in this column

2.1.2. Data Tidiness

All fields are committed to the basic Tidiness rules

2.2. Data cleaning

2.2.1. Fix Data types problems

Fields are changed to their appropriate dtypes

2.2.2. Fix Completeness

15 Empty rows are dropped

2.2.3. Fix Duplicates

14 duplicates are dropped

2.2.4. Consistency

Motherboard category is standardized

3. Getting the data ready

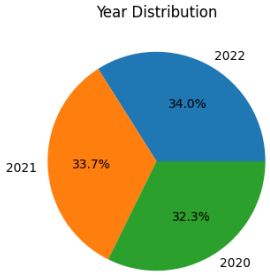
I have got the data ready for the analysis through:

1. Add the profits column
2. Make two columns for year and month

4. Data Analysis and Visualization

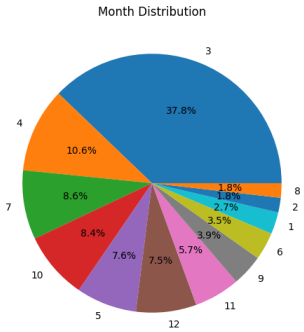
Question 1: Which year has the most orders?

2022 was the most year having orders with a percentage of 34%



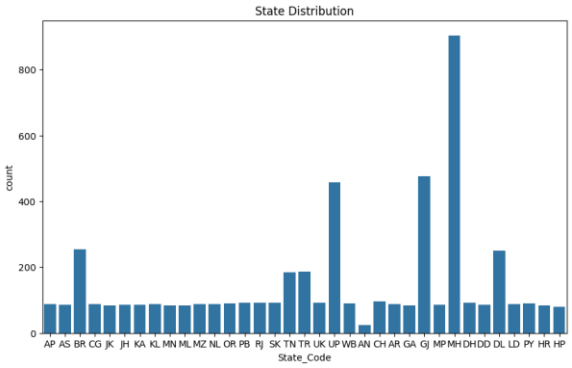
Question 2: Which month has the most orders?

The graph shows that March has the highest number of orders, with a percentage of 37.8%, while August and February have the least number of orders, with a percentage of 1.8%



Question 3: Which state has the highest percentage of orders over the three years?

The state MH has the highest number of orders, while the state AN has the least number of orders

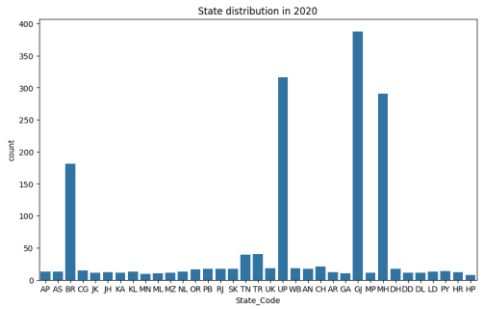


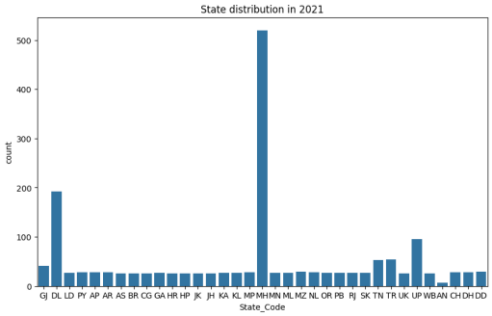
Question 4: Which state has the highest percentage of orders in each year?

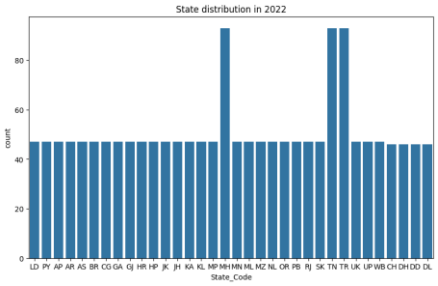
In 2020, GJ state had the highest orders

In 2021, MH state had the highest orders

In 2022, MH, TN, TR had the highest orders

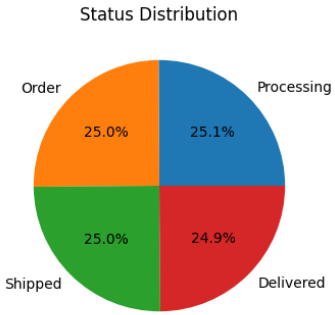






Question 5: Which status is the most common over the three years?

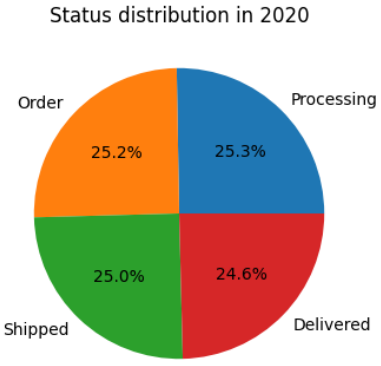
All statuses (Order, shipped, delivered, and processing) are almost equally common among the orders' data

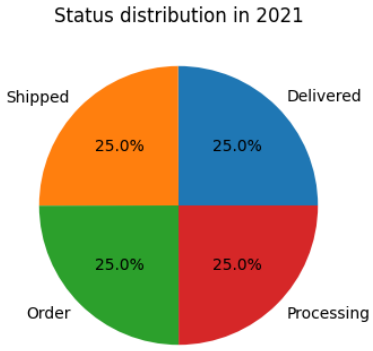


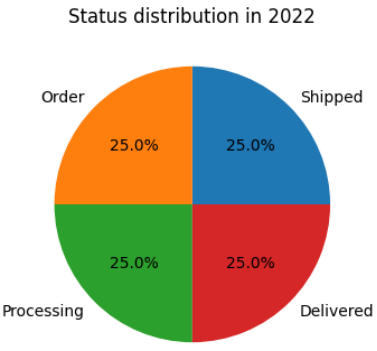
Question 6: Which status is common in each year?

In 2020, the Processing status was slightly higher than the other statuses

In both 2021 and 2022, the four statuses are equally the same

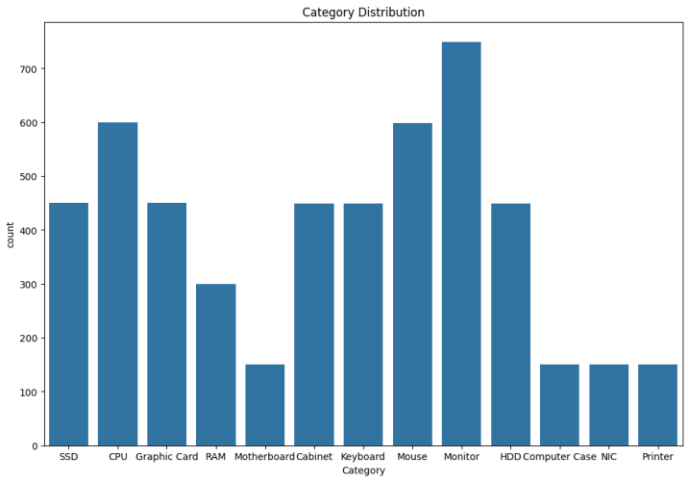






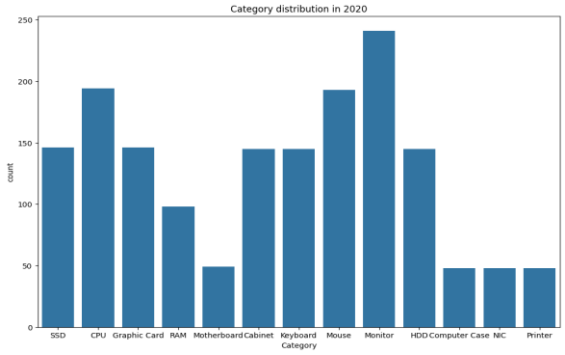
Question 7: Which category is most sold over the three years?

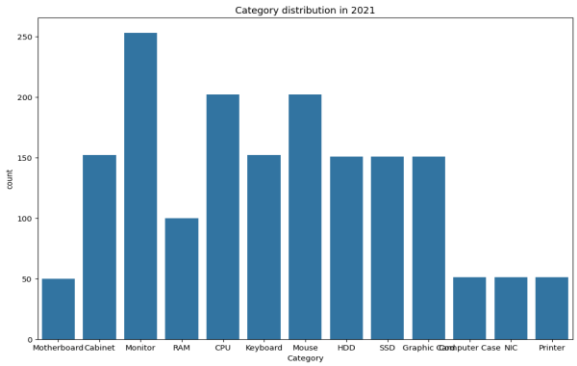
Monitors are the most common category over the three years

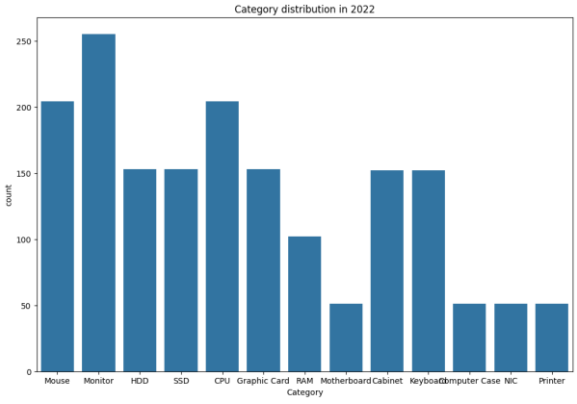


Question 8: Which category is most sold in each year?

In each of the three years, Monitors are the highest sold category

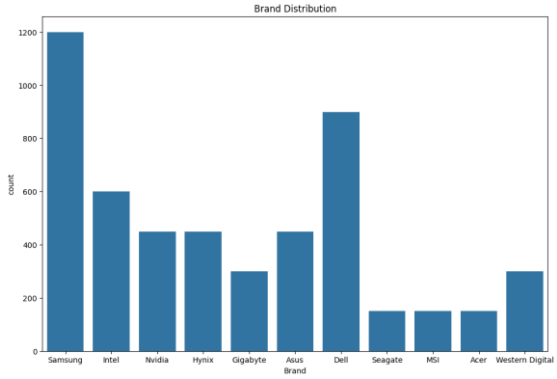






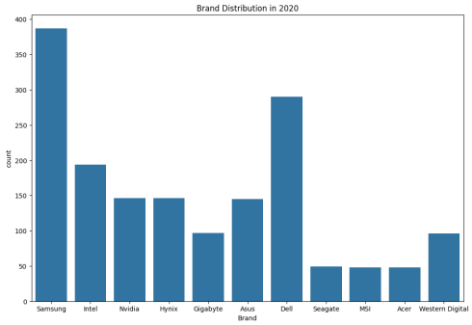
Question 9: Which Brand is most sold over the three years?

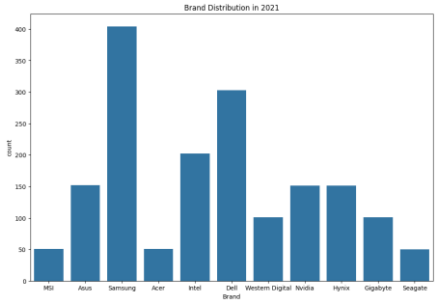
Samsung is the most sold brand over the three years

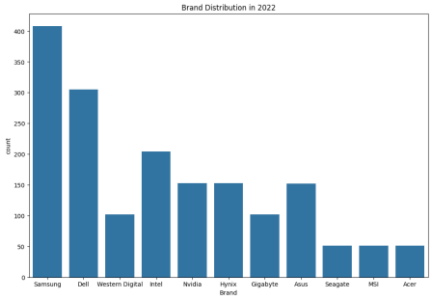


Question 10: Which Brand is most sold in each year?

In each of the three years, Samsung was the most sold brand

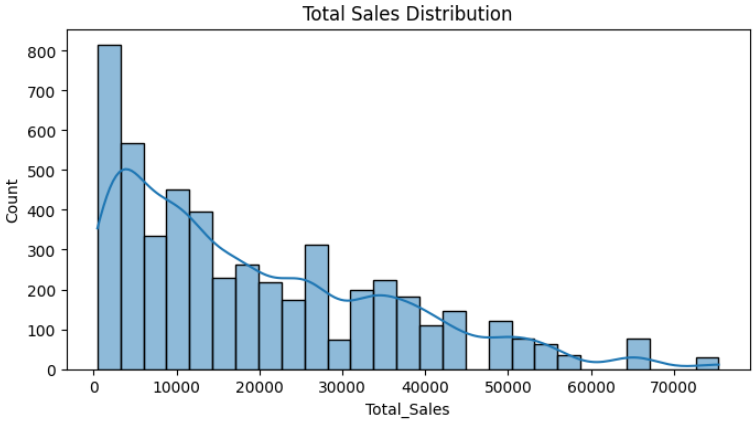






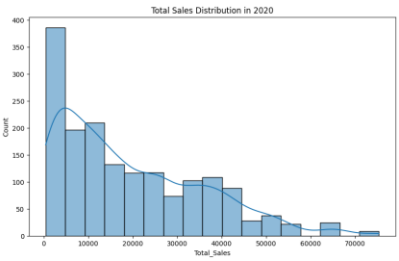
Question 11: What is the average total sales over the three years?

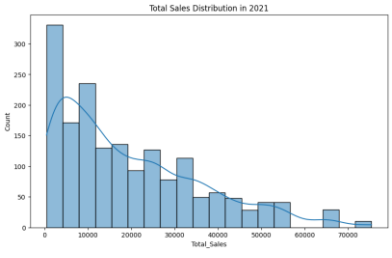
The data is right-skewed, with many outliers beyond 50000 dollars. The average of sales over the three years is almost 20000 dollars

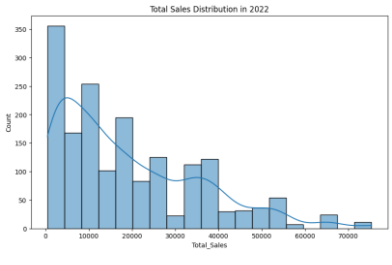


Question 12: What is the average total sales in each year?

In each of the three years, the data is right-skewed, having outliers beyond 60000, with an average of about 20000 dollars. 2020 sales were the highest in their means with only slight difference between the other two years

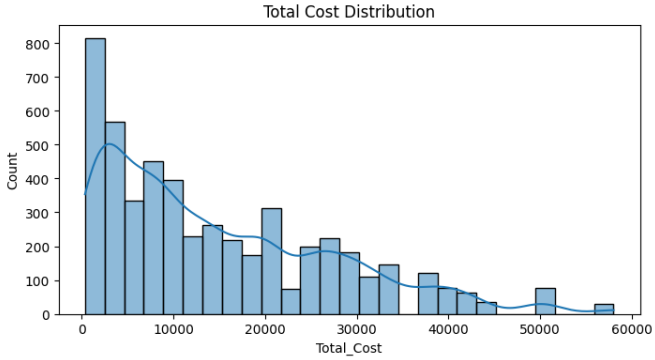






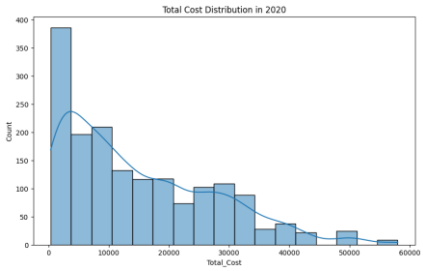
Question 13: What is the average total costs over the three years?

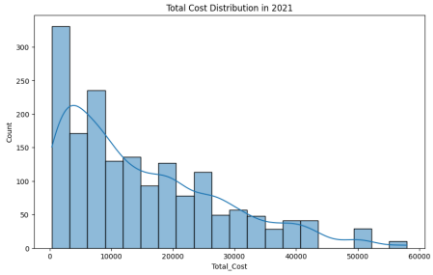
The data is right-skewed, with outliers beyond 45000 dollars, with an average of about 15000 dollars

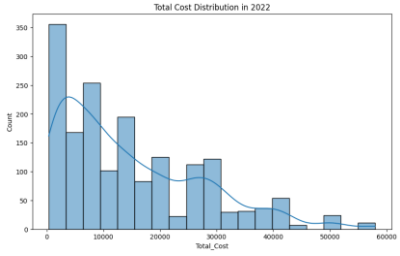


Question 14: What is the average total costs in each year?

In each of the three years, the data is very right-skewed. There are many outliers beyond 40000 dollars, with an average of about 15000 dollars

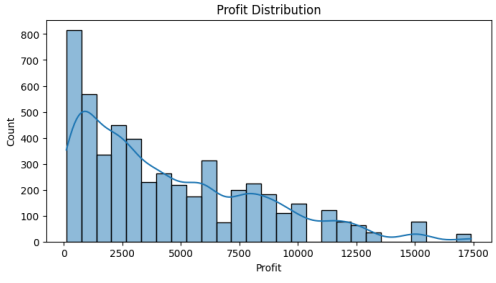






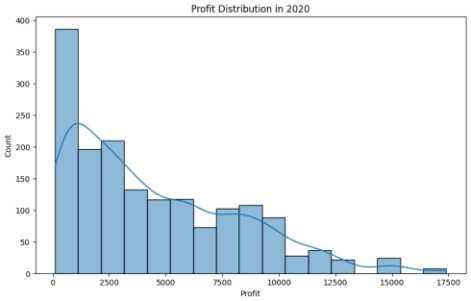
Question 15: What is the average profit (or loss) over the three years?

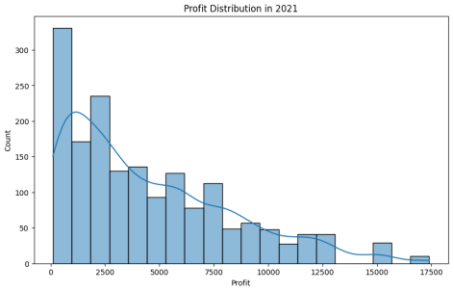
The data is very right-skewed. There are many outliers beyond 10000 dollars of profit, with an average of about 4500 dollars of profit

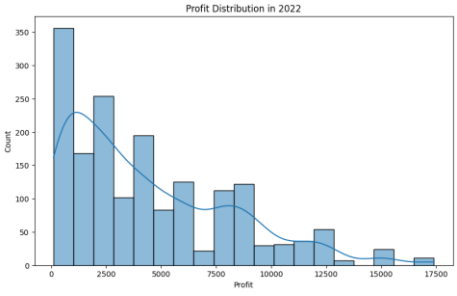


Question 16: What is the average profit (or loss) in each year?

In each of the three years, the data is very right-skewed, with outliers beyond 10000 dollars of profit, and average of about 4500 dollars of profit

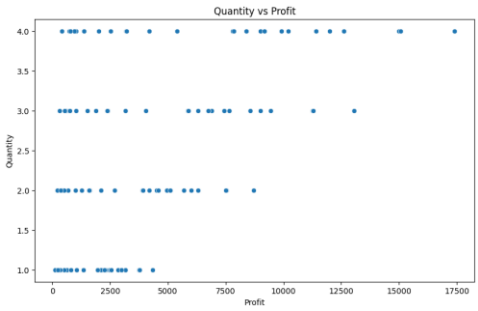






Question 17: What is the relation between quantity sold and the profit (or loss)?

There is moderate correlation between Quantity sold and profit



Question 18: What is the relation between total sales and the profit (or loss)?

There is very strong positive correlation between total sales and profit

Question 19: What is the relation between total costs and the profit (or loss)?

There is a very strong positive correlation between total costs and profits