



# Online Consumers Behavior

## Data Analytics Report

27/11/2020

---

María González González

# Index

|                                 |    |
|---------------------------------|----|
| .....                           | 0  |
| <i>Introduction</i> .....       | 2  |
| <i>Objectives</i> .....         | 3  |
| <i>Workflow</i> .....           | 4  |
| Dataset Structured: .....       | 4  |
| Top 10 Products Categories..... | 5  |
| Preferred Brand .....           | 6  |
| Price or Brand .....            | 7  |
| Top Brands Smartphone.....      | 8  |
| Busiest Hour .....              | 9  |
| Days With More Activity .....   | 10 |
| Linear Regression .....         | 11 |
| <i>Conclusion</i> .....         | 12 |
| <i>References</i> .....         | 12 |



## Introduction

The Internet is playing a crucial role in our daily life, such as; it is able for us to search information, to be communicated worldwide, to be entertained, and even to buy things online.

Online shopping has been widely accepted since it provides consumers more choices and information, as it allows us to compare prices and reviews. Another advantage of online purchases is that it is easier to find any product and most of the times, is the fastest way to get it. All of this proves that, Internet shopping is providing consumers satisfaction, at the time of seeking their needs. On the other hand, there are some factors that may have a negative impact at the time of purchasing online, consumers would like to try the product or to examine it, or even the trust of providing their personal data.

## Objectives

During this project it is going to be answered the following questions:

1. Top 10 categories of products
2. Do consumers have a preferred brand at the time of selecting their products?
3. Is the price or the brand affecting their decisions?
4. Top brands chosen by customers to buy their smartphones
5. There is a preferred hour of the day customers search more online, add the product to cart or buy
6. Days of the month that customers buy more and less
7. There is a Lineal Regression between add the product to cart and purchase it?

## Workflow

In order to answer the objectives of this project it is necessary to work with Big Data, a total amount of 55 million of data is needed, in order to analyze the people behavior from February and March of 2020.

Since the platform Tableau does not allow to work with such an amount capacity of data, it was needed to get a random sample of the entire dataset for both months in order to create some of the plots beneficial to answer the objectives of this project.

At the same time, the data was obtained from a large multi-category online store, where each row in the dataset represents an event. All events are related to the products and users.

### Dataset Structured:

- ❑ Event\_time: from February and March 2020 and the event happened at UTC time
- ❑ Event\_type: can be:
  - ❑ View - a user viewed a product
  - ❑ Cart - a user added a product to shopping cart
  - ❑ Purchase - a user purchased a product

Commun event: view => cart => purchase

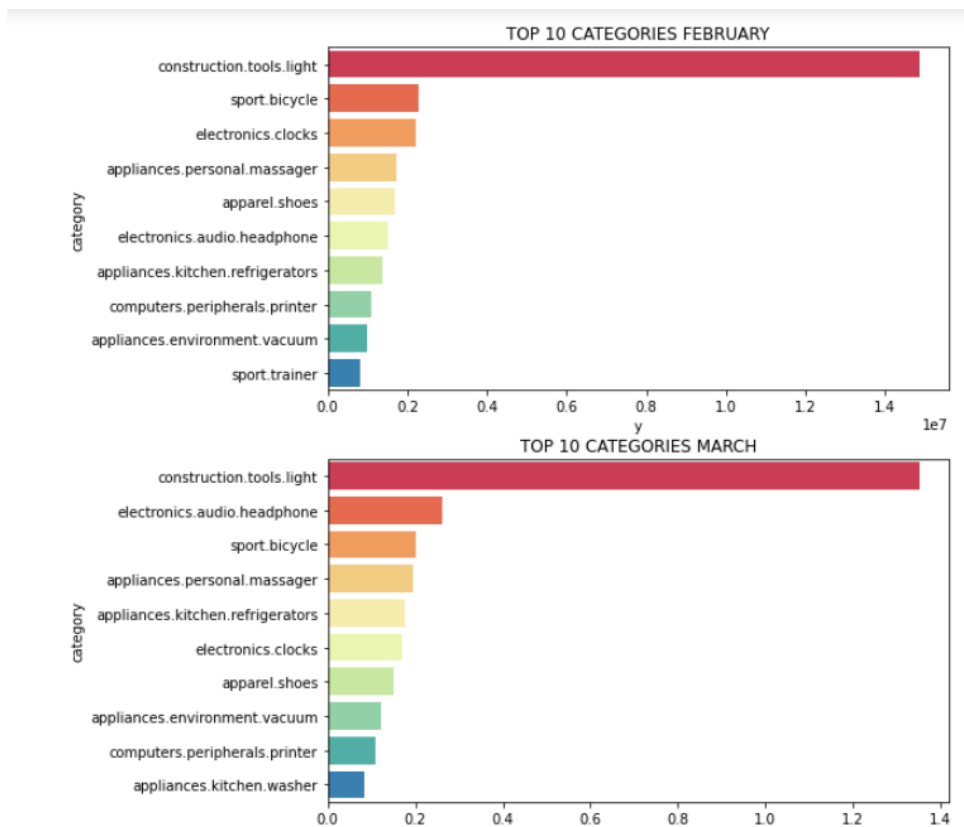
- ❑ Product\_id - ID of a product
- ❑ Category\_id - product's category ID
- ❑ Brand - brand of the product
- ❑ Price - float price of the product
- ❑ User\_id - permanent user id
- ❑ User\_session - Temporary user's session ID. Same for each user's session. Is changed every time user come back to online store from a long pause

## Top 10 Products Categories

Both February and March there is a drastic difference between the first category and the others. The first category for February will be Construction Tools Light, followed by sport bicycle and by electronic clocks, the same situation, regarding the Construction category, happens in March. However, the second position in March will be for the electronics audio headphone and the third for the sport bicycle.

From this result it can be concluded that during the start of the worldwide pandemic situation may have affected the online shopping decision as the needs of the population has changed.

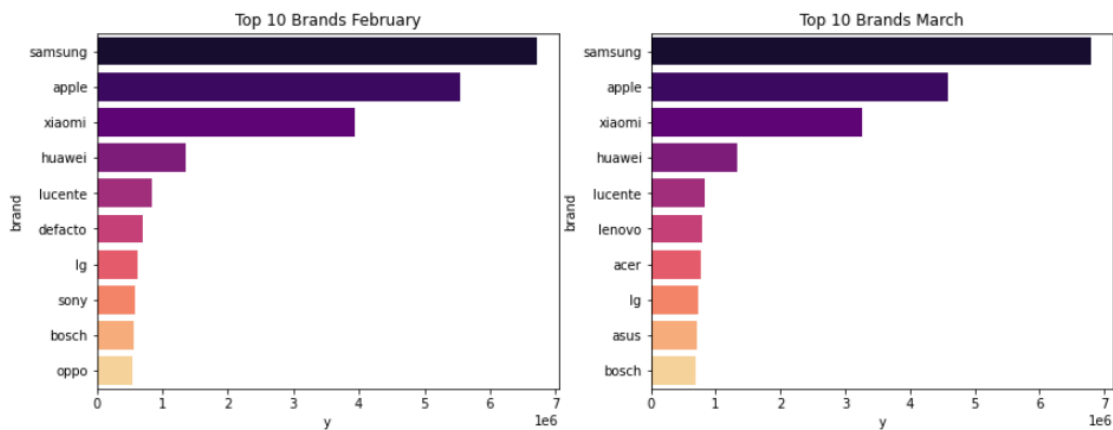
As it shows the plot below:



## Preferred Brand

In contemplation of the second objective: Do consumers have a preferred brand at the time of selecting their products? It is crucial to be aware that the most popular brands are the ones from the technology industry. The reasons could be that customers already know the products, as well as, these brands have a wide variety of products and that there are the most famous brands worldwide.

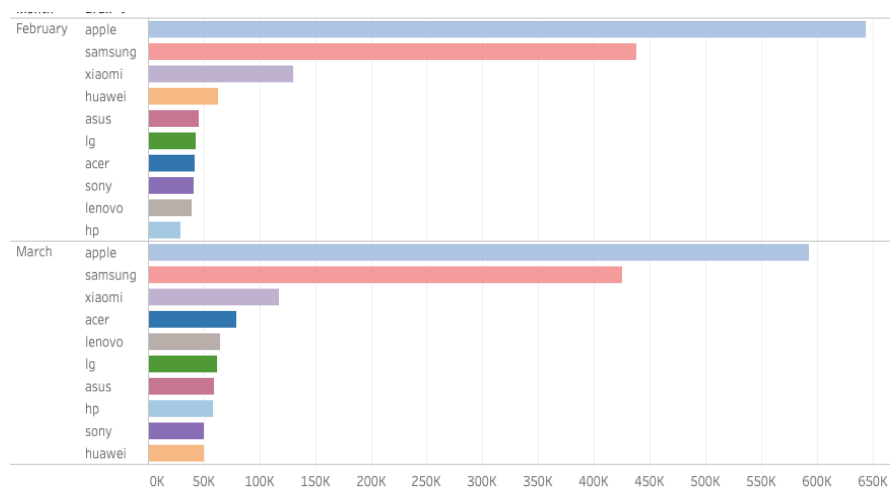
In the below plot it is acknowledged that most of these brands have a strong online position since in both months they are categorized as the top 10 brands, as well that under the fourth position there is a balance between all of the following brands. As the top 3; Samsung, Apple and Xiaomi are in lead, far away from the other brands.



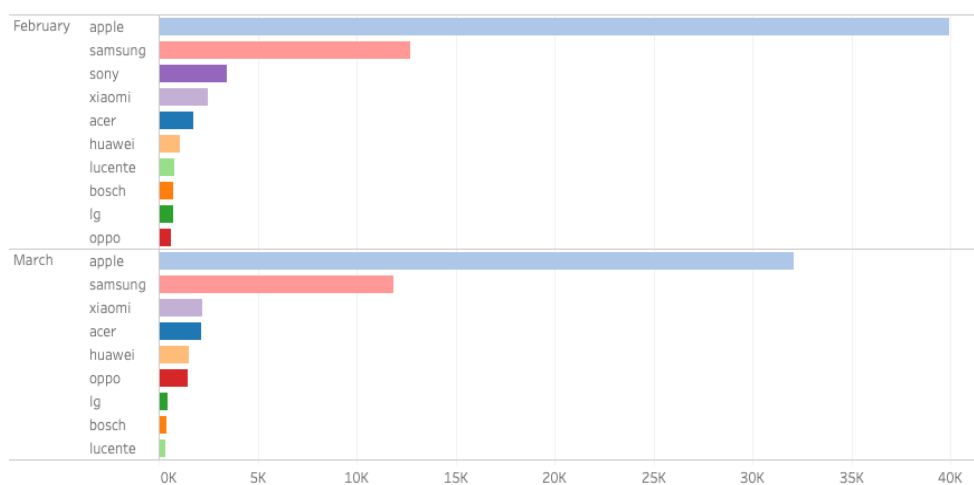
## Price or Brand

Is the price or the brand affecting consumers decisions? In order to answer this question, it is needed to compare which type of brands customers are searching online and which are their final decisions.

In the first plot it samples which brands are receiving the most online visits, it is crucial to remark the huge difference from the others that Apple has, below it is appreciated:



With the below plot it indicate that Apple is the most representative brand and the one that consumers choose, followed by Samsung, so it can be concluded that the online shopping decision is affected by the brand of the product.

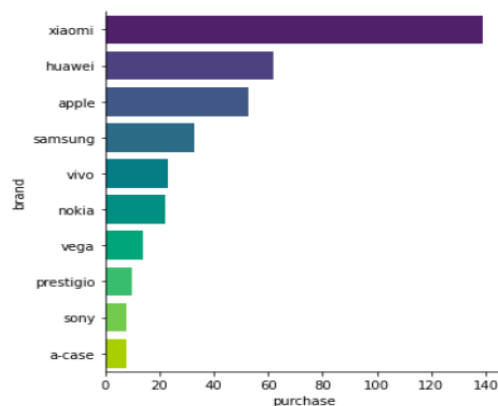




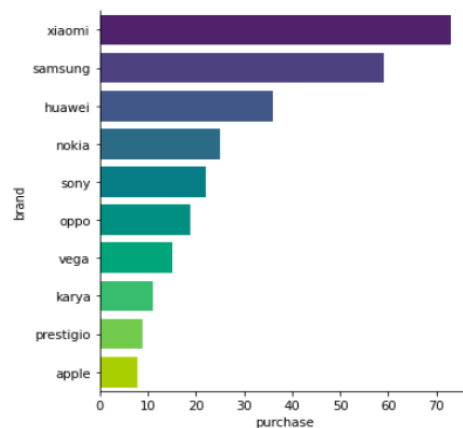
## Top Brands Smartphone

Previously, it was concluded that Apple followed by Samsung are the top chosen brands at the time of Online shopping, however looking into a specific product such as smartphones, it is observed that cheaper brands are the one leading the sales.

In order to answer Which are the top smartphone brands? This graphic would be helpful, as it shows that Apple and Samsung are still at the third and fourth position and the leaders are Xiaomi and Huawei, concluding that consumers decide to buy their smartphone depending on the price.

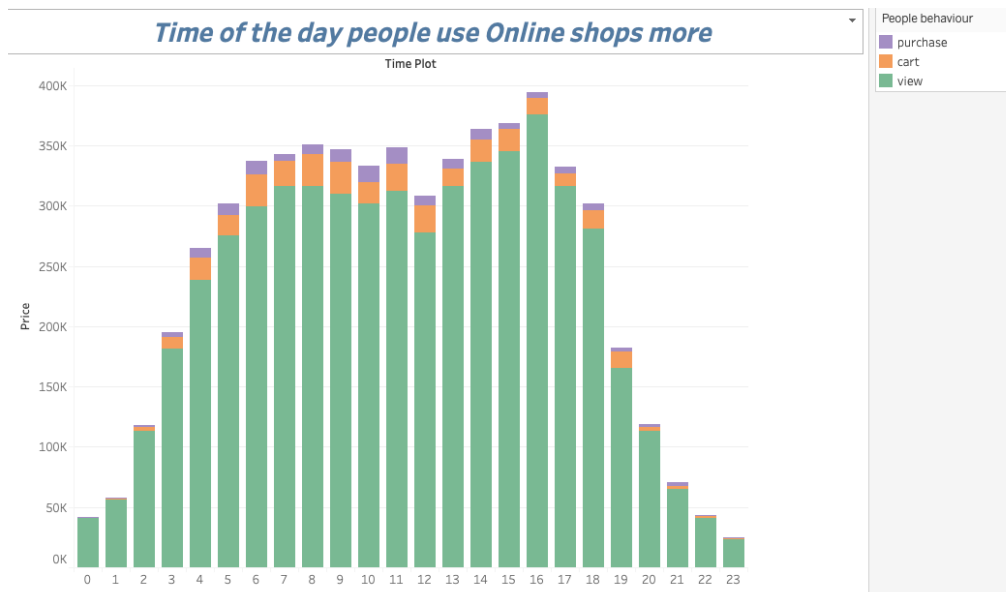


Even so, in March Apple took the last position and Xiaomi, Samsung and Huawei are the leaders, this result could be affected by the beginning of the worldwide pandemic.



## Busiest Hour

It is wondering if there is a preferred hour of the day customers search more online, add the product to cart or buy it? As it is notice below there is more activity from the 06:00 until the 18:00, and the time when people buy the highest peak where more purchases occur will be during the 09:00 and the 10:00.

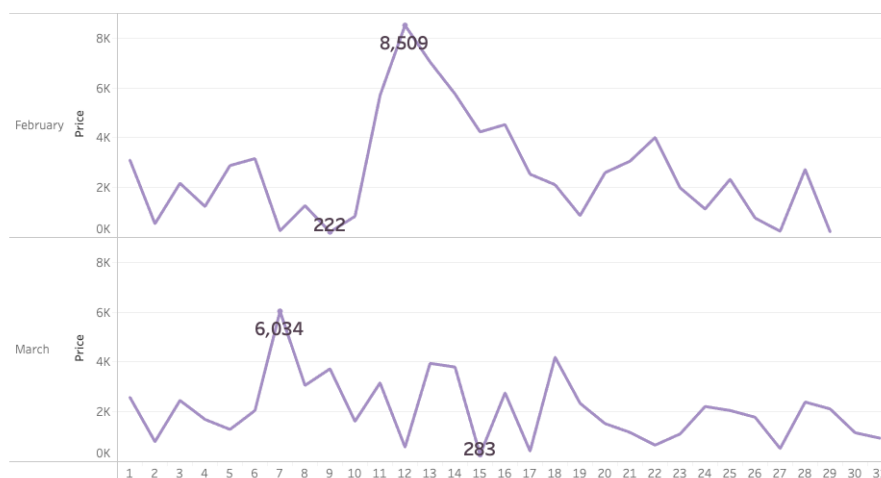


## Days With More Activity

In response of the days of the month that customers buy more or less. There is a big impact during the first 15 days of the month and the last ones, which means that people buy more online during the first 15 days. These results could be due to wages are entered at the beginning of each month.

On the other side, during the weekends time is where the lowest peak occurs for both months February and March. Moreover, the highest peaks happen during the week days.

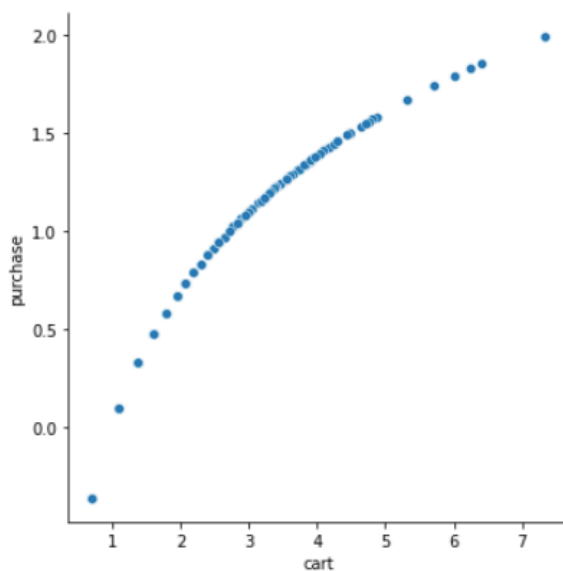
As it highlights in the graphic, there is not any balance between the days, since there is drops and downs during both months.



## Linear Regression

Linear regression models are used to show or predict the relationship between two variables or factors. Each observation consists of two values. One value is for the dependent variable, in this case the cart factor and one value is for the independent variable, the purchase variable.

Which reference to the last objective if there is a lineal regression between add the product to cart and purchase it. As the line chart shows there is a linear regression between purchase and cart variable.



## Conclusion

Online shopping is growing more and more nowadays, it is crucial to be aware that consumers take into consideration the brand of the product, the review and the price at the time of taking their decision. On the other hand, the industry of technologies is the chosen by the consumers at the time of buying online.

For the next step will be analyzing this situation before, during and after the worldwide pandemic in order to conclude if this extreme situation has change customers online behavior.

## References

- <https://www.kaggle.com/mkechinov/ecommerce-behavior-data-from-multi-category-store>
- [https://github.com/MariaGonzalezGonzalez/Maria\\_Gonzalez\\_Project5](https://github.com/MariaGonzalezGonzalez/Maria_Gonzalez_Project5)