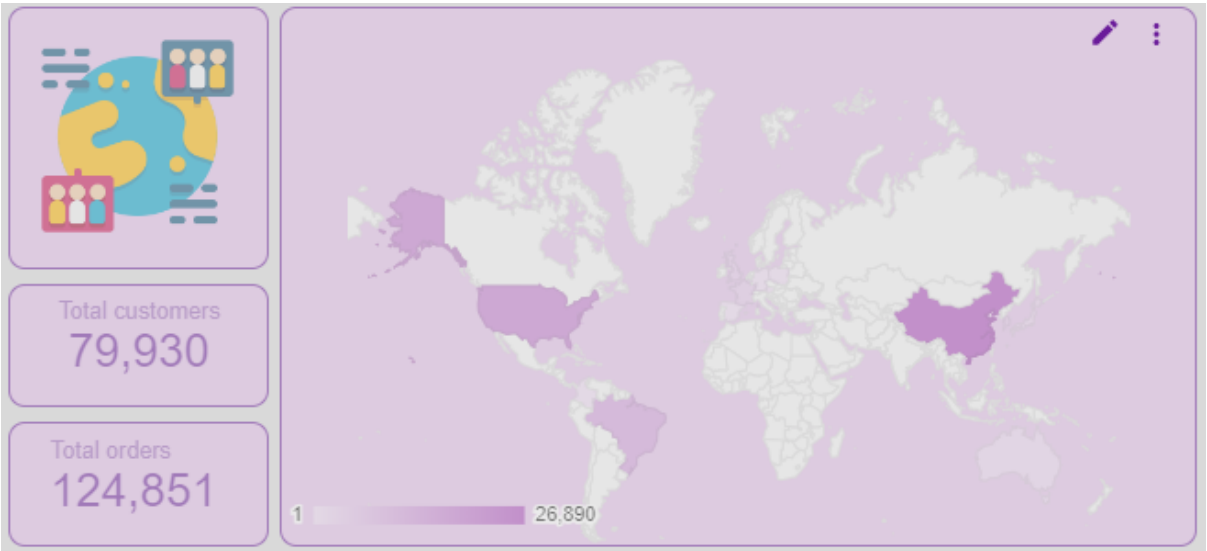
THE LOOK ECOMMERCE

Sales Analysis

-

Demographic approach



December 202

TABLE OF CONTENTS

1. THE LOOK ECOMMERCE PROJECT DESCRIPTION

In this section, it is explained the description of the project, which can be observed in table 1.1.

Table 1.1 - Project description using SMART Method.

Situation:	Described in detail in section 3, the analyzed data set was collected to better understand the distribution of customers considering their nationality, gender, age group or city. The data was extracted from Google Big Data and didn't have relevant noise, so the only change performed was to create a new column to classify each customer according to age, splitting them by 6 distinct age groups: teens, twenties, thirties, forties, fifties and elders.
Task:	Section 2 gives more details about the key questions to be analyzed and answered, which pretend to identify possible trends and patterns related with customers' behavior and their gender or nationality, among other characteristics. The main task was performed by asking 14 questions mentioned along the report.
Action:	In order to achieve the final results, and implement my analysis, the data was cleaned using python and to perform the visualizations, looker studio was used. Section 4 shows the main findings of the project.
Result:	The results are discussed in section 5 with more detail, but, generally speaking, it was found that, considering gender, in general, men tend to spend more money and buy more expensive products. The top 3 countries with more money spent are China, United States and Brazil, but France and the United Kingdom are the countries with more money spent by person, for men and women respectively.

2. PROBLEM DEFINITION

1. What is the distribution of nationality, age and gender among our customer base?
2. Which countries and age groups spend more money on ecommerce?
3. Which countries and age groups spend more money per person on ecommerce?
4. Which countries and age groups have more orders on ecommerce, for male and female individuals?
5. How are the product's categories distributed among gender and age groups?
6. How much money was spent on each product's categories, among male and female customers considering the most popular countries?
7. Considering the 4 most popular categories among both genders, how much money was spent on such products?
8. What are the most popular brands among the 3 most popular countries?
9. Can we find any specific pattern among the 1000 customers who spent more money in total?
10. Can we find any specific pattern among the 1000 most ancient customers?
11. How the number of orders created varies along the time for the most 5 popular countries?
12. How the number of customers created varies along the time for the most 5 popular countries?
13. What is the distribution of citizens, age and gender among our customer base?
14. With this data is it possible to create a model that can perform predictions, like number of orders or total spent money, considering the population's features?
15. What happens to the total retail price if we double the sales of Calvin Klein, Carhartt, Diesel, True Religion and 7 for All Mankind, or do we do the same for most popular products among men and women?

3. DATA DESIGN

For this specific topic, the data design started with the extraction of all tables from the BigQuery database, namely: distribution_centers, events, inventory_items, order_items, orders, products and users.
The extraction was performed using bigquery and SQL and saved on .csv files, splitted according to the total file size permitted.
After having the .csv files, the dataset was analyzed using python and in this case, it wasn't needed to perform the cleaning of missing/corrupted data. It was just created new columns used to facilitate the data analysis file "age_group", according to the customer age. In the end, we used 2 main datasets: one that joined users' tables with products' table and another which joined users' tables with order_items table.
At the end, specific datasets were created and saved in .csv files to create visualizations in Looker.

4. FINDINGS

4.1 What is the distribution of nationality, age and gender among our customer base?

To answer this question, some graphs were created to visualize the distributions among the different age groups, as well as, genders and countries. For that last one, only the 8 countries with the higher number of users were considered.

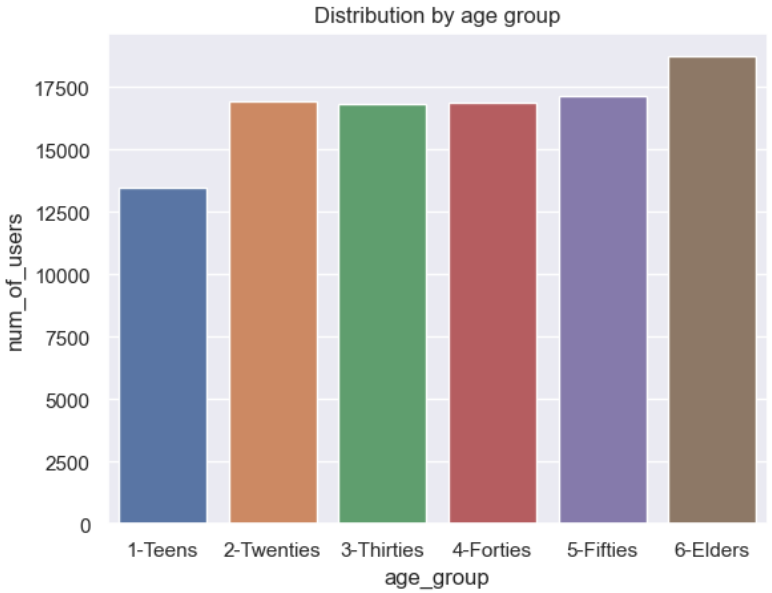


Fig. 4.1 - Distribution of users among all age groups.

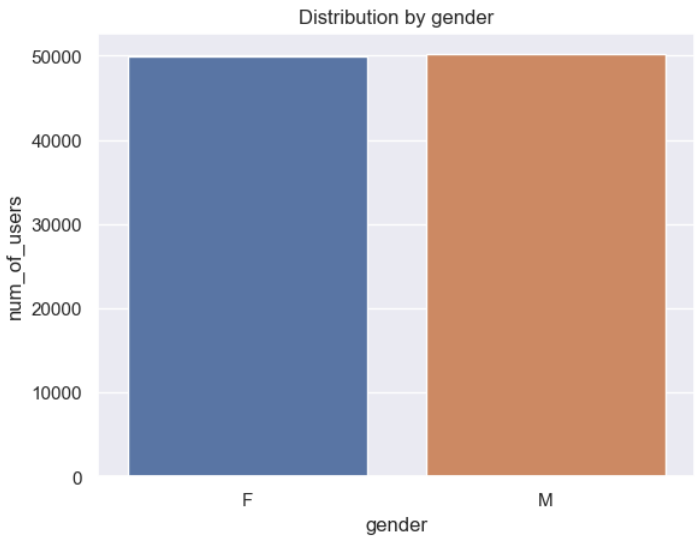


Fig. 4.2 - Distribution of users among both genders.

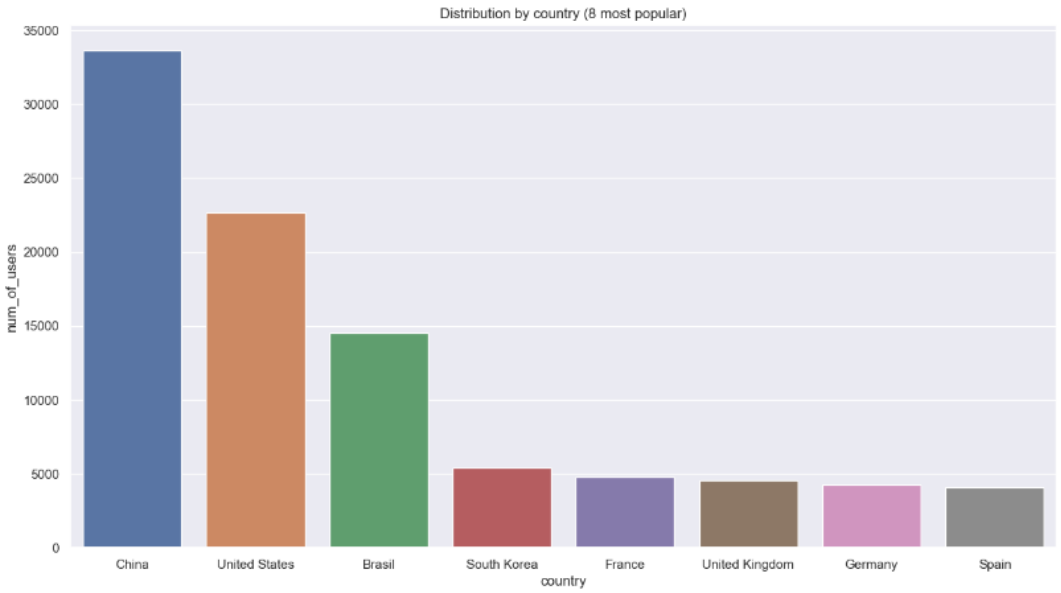


Fig. 4.3 - Distribution of users among the top 8 countries.

Another metric important to take into account is the distribution of genders in the different age groups, for each country (shown in percentage of total by country). Below we can see that distribution for the top 3 countries.

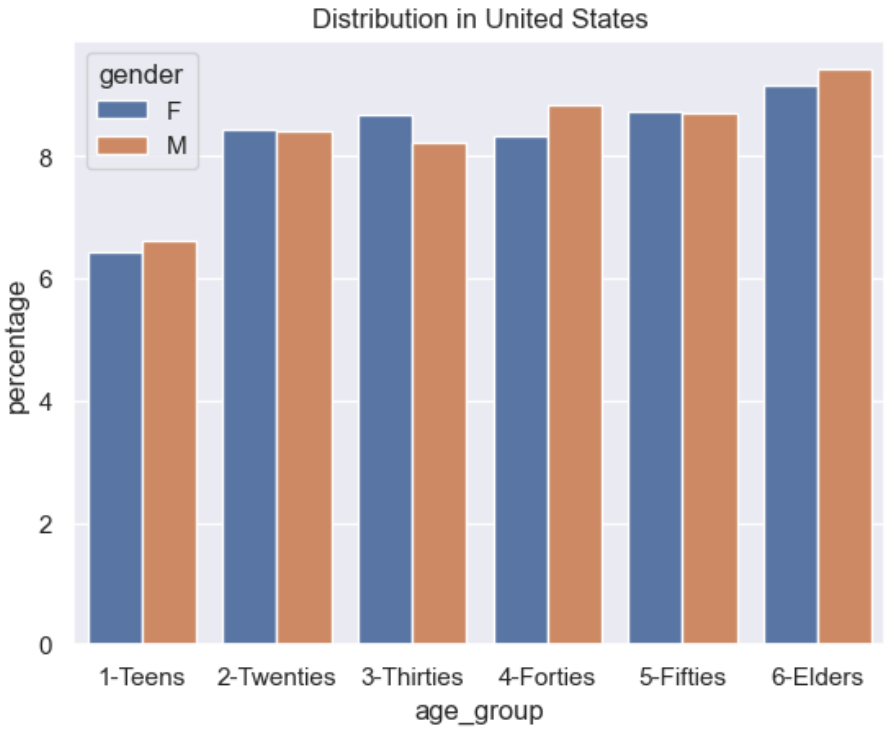


Fig. 4.4 - Distribution of users in the United States considering genders and age groups.

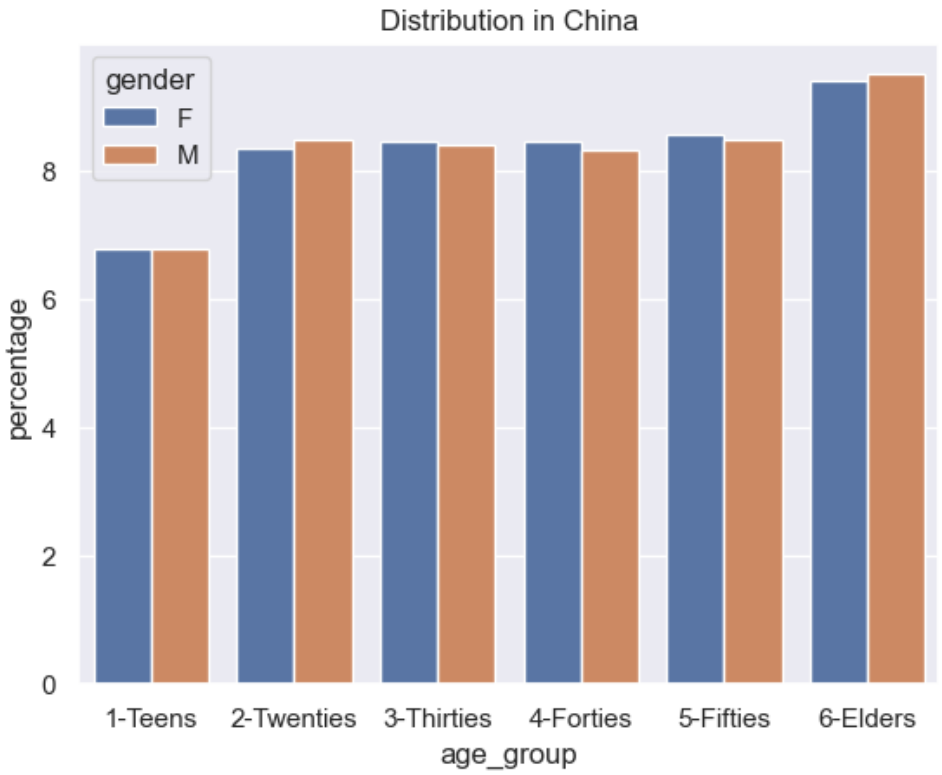


Fig. 4.5 - Distribution of users in China considering genders and age groups.

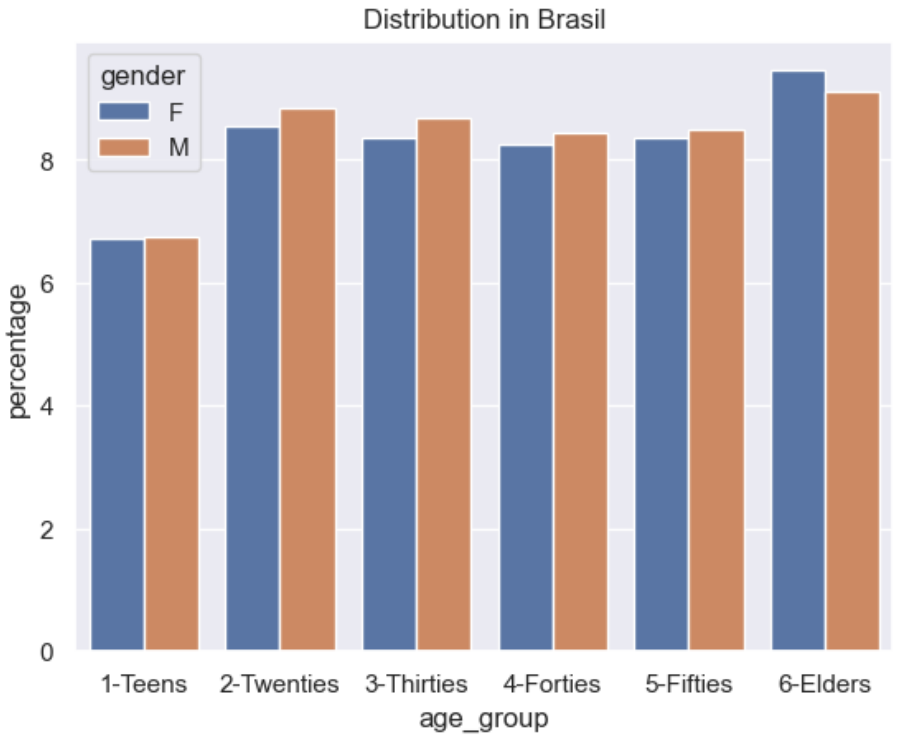


Fig. 4.6 - Distribution of users in Brazil considering genders and age groups.

Besides the age groups' distribution being quite homogenous, one can see that the elderly people have strictly more users, so below we show for that group age the distribution on the main states for China, USA and Brazil.

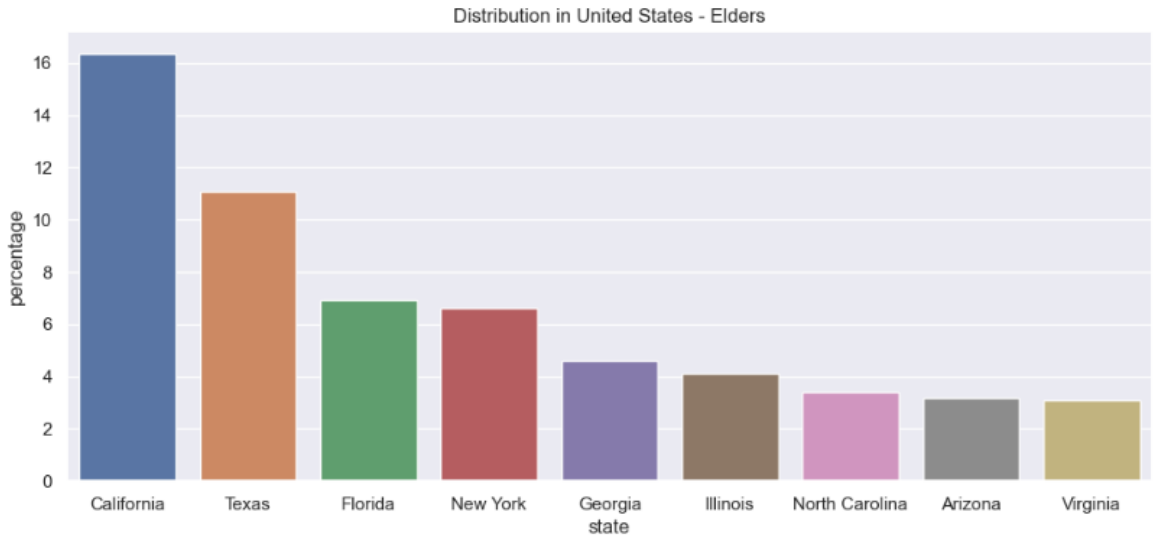


Fig. 4.7 - Distribution of users by state in the United States.

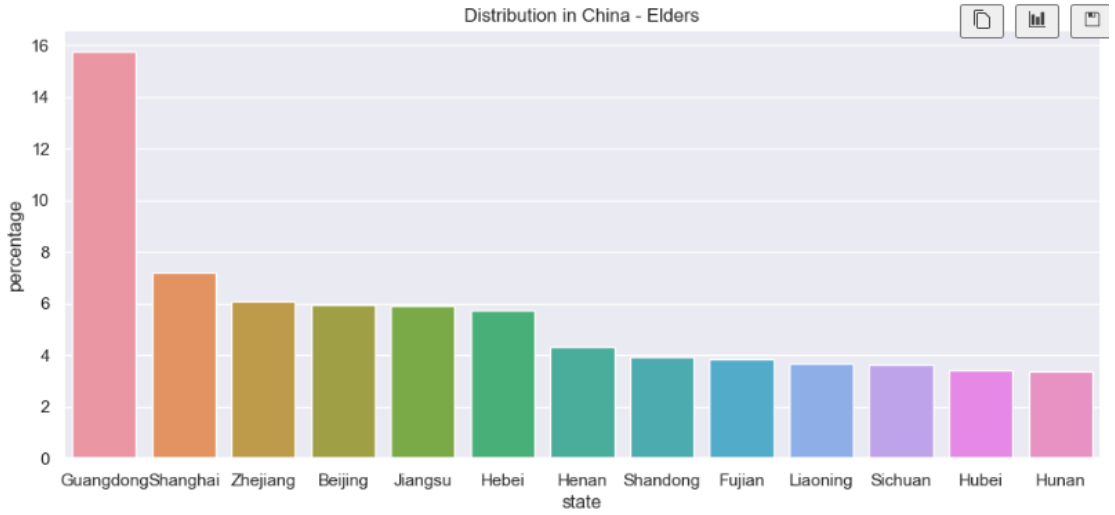


Fig. 4.8 - Distribution of users by state in China.

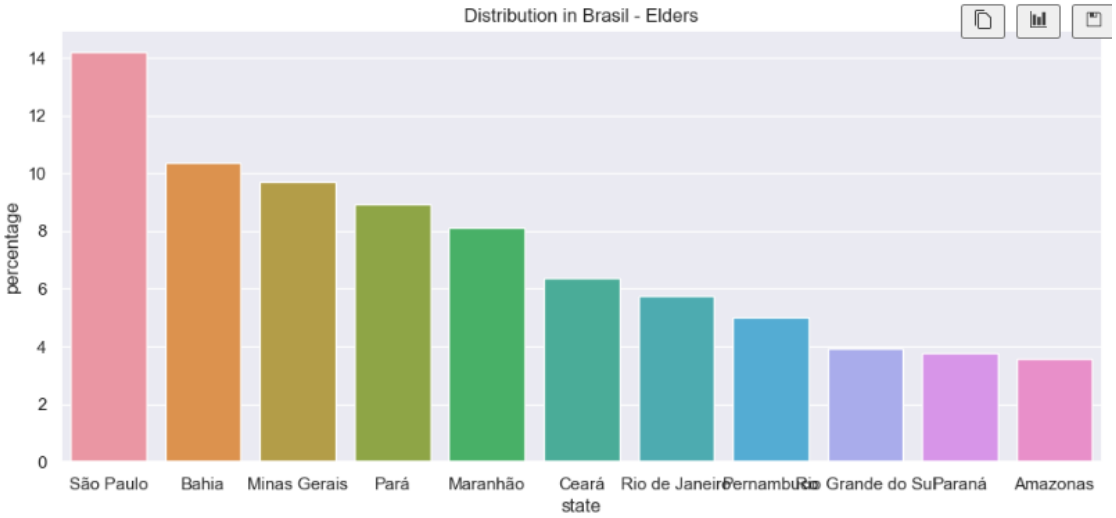


Fig. 4.9 - Distribution of users by state in Brazil.

By the data, we can see that the distribution among genders and age groups is homogeneous, having a strictly higher elder users. This is valid in general and for the top countries. Regarding the distribution by country, with no surprise, China is the country with more users, followed by the United States and Brazil. If we analyze the distribution of elderly people in each of those three countries, we easily see that the states of California, Guangdong and São Paulo have more users subscribed.

4.2 Which countries and age groups spent more money on ecommerce?

In the figures below, one can see the distribution of total spent money among countries and the same metric, but also splitting by age groups, respectively.

Fig. 4.10 - Distribution of the total spent money by country..

Fig. 4.11 - Distribution of the total spent money by country and age group..

By the data we can see China, USA and Brazil are the countries with higher total retail prices for all age groups, but especially for elderly people.

4.3 Which countries and age groups spend more money per person on ecommerce?

Another interesting metric we can analyze is the average money spent by each customer. In that sense, the bar graphs shown below were created, splitting the data into male and female users.

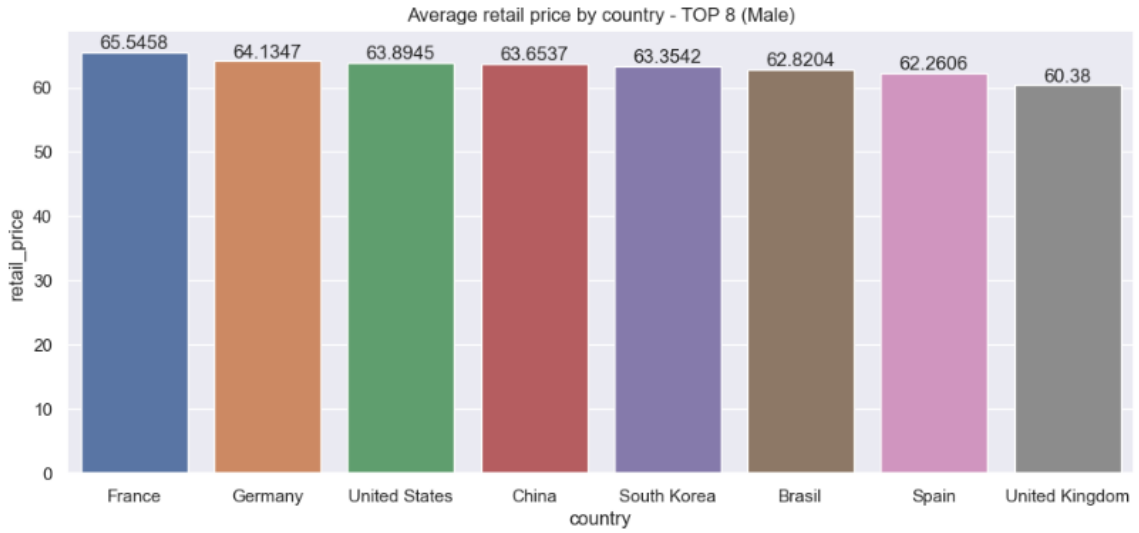


Fig. 4.12 - Distribution of average retail price by country for the top 8 countries and male users.

Fig. 4.13 - Distribution of average retail price by country for the top 8 countries and female users.

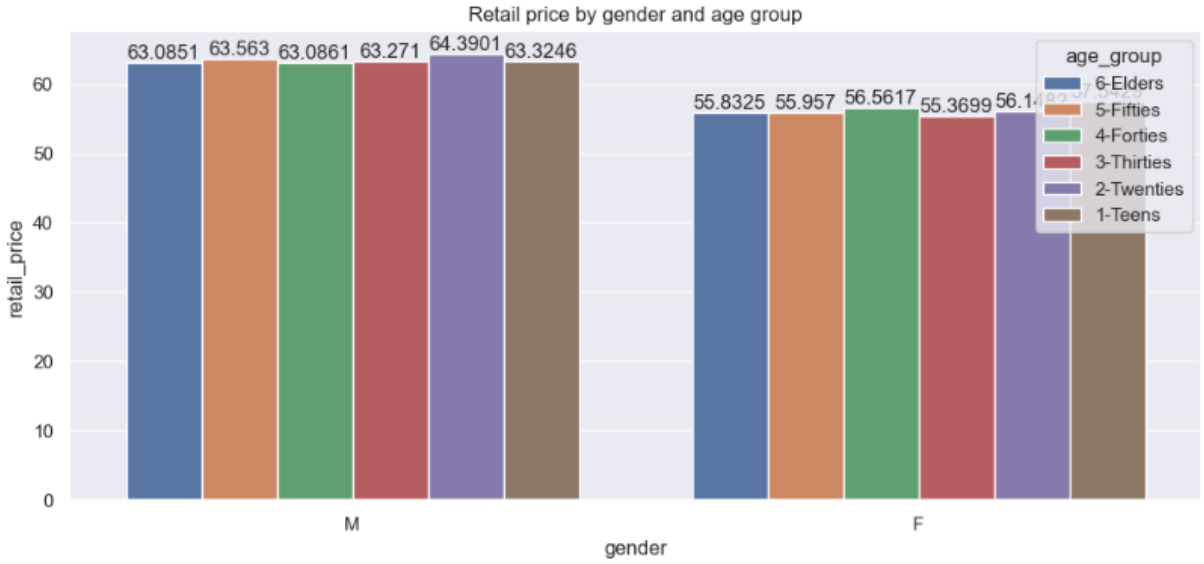


Fig. 4.14 - Distribution of average retail price by age group and gender.

Contralily of what we've seen in the previous questions, if we talk in terms of price spent by each user, USA, China and Brazil are not the top 3 countries.
Besides having a small difference, we can easily conclude that France and Germany have the higher values for male individuals, and the United Kingdom and Germany for female individuals. Another thing we can observe is that the values presented by male individuals are strictly higher than the female's.
Generally speaking the trend is similar to the one observed on figure 4.12 and 4.13.

4.4 Which countries and age groups have more orders on ecommerce, for male and female individuals?

Below we can observe three graphs that show the distribution of orders by country for men and women, as well as, the distribution among age groups and genders.

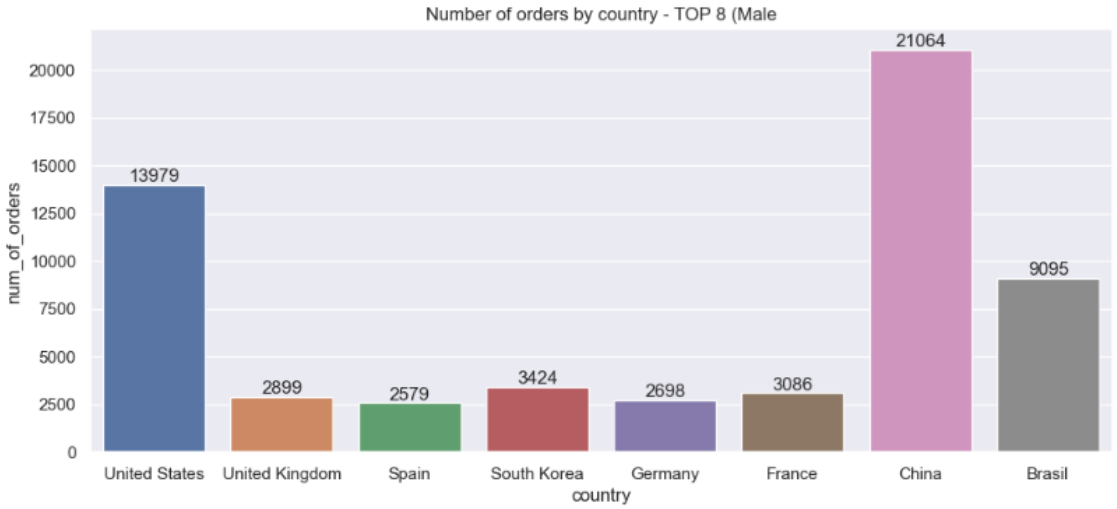


Fig. 4.15 - Distribution of average retail price by age group and gender.

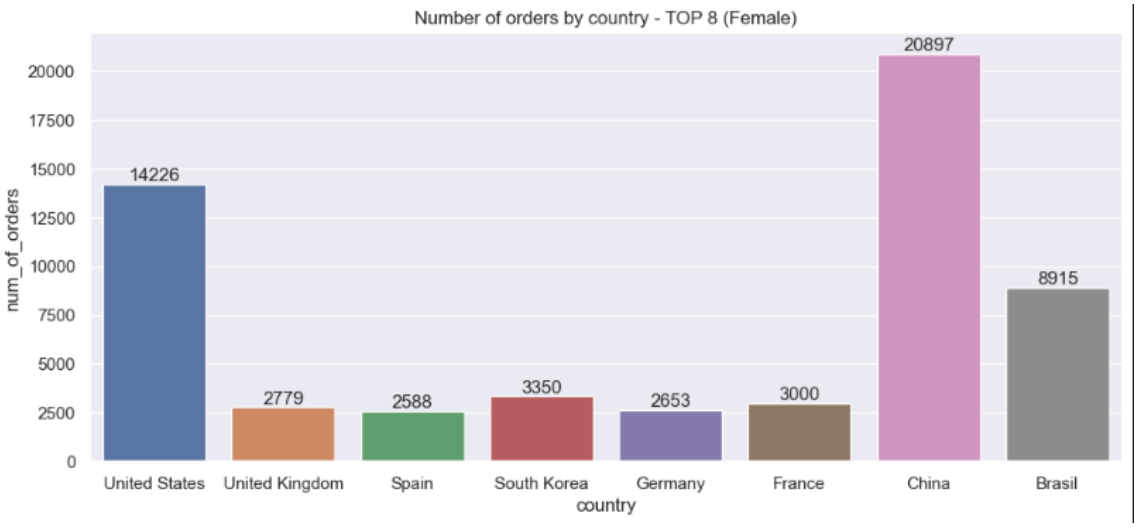


Fig. 4.16 - Distribution of average retail price by age group and gender.

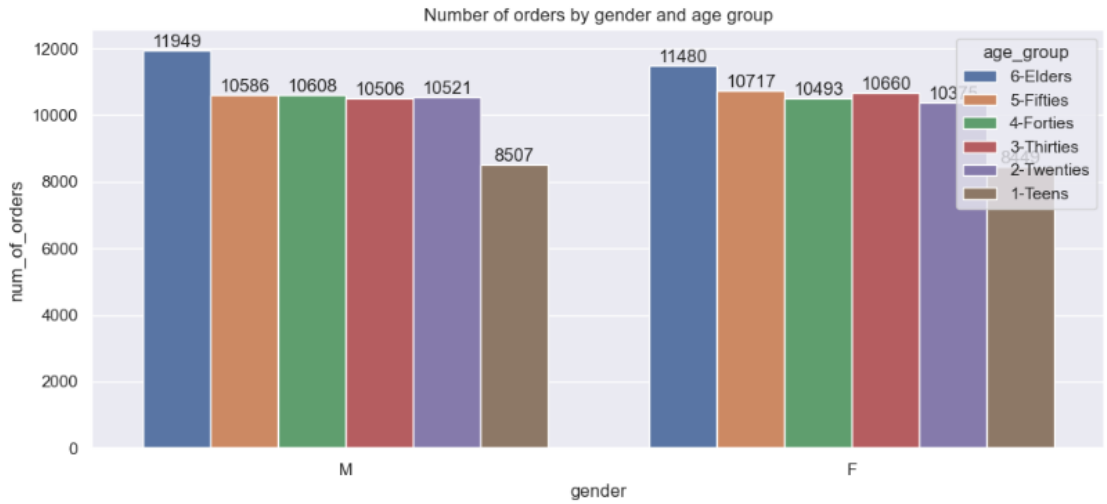


Fig. 4.17 - Distribution of average retail price by age group and gender.

By observing the figure above, again, China, USA and Brazil are the countries that sum more orders, and the distribution among genders is quite homogeneous. Regarding the ages, again elderly people tend to have a higher number of orders.

4.5 How are the product's categories distributed among gender and age groups?

The distribution of product's categories was analyzed and below we can see a bar chart that shows it.

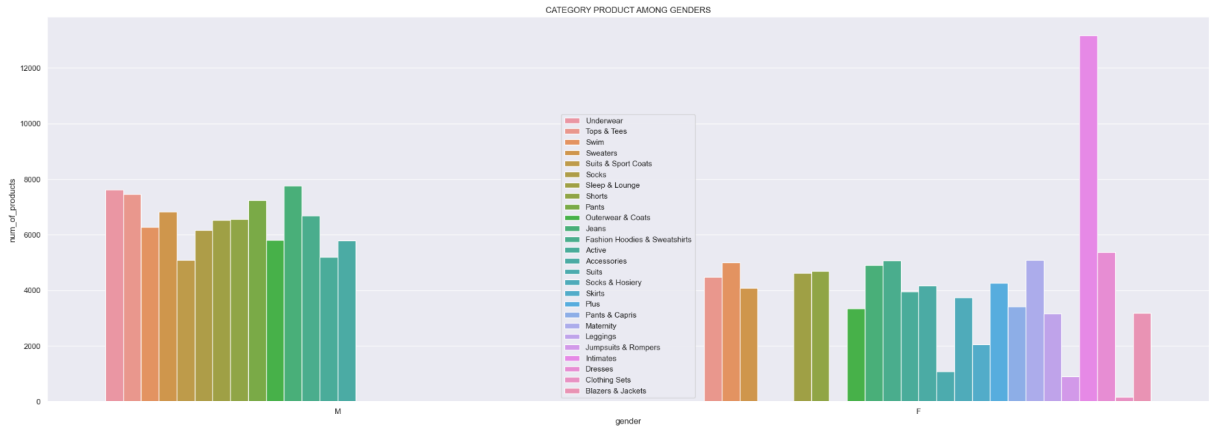


Fig. 4.18 -Distribution of products by its category, for men and women.

The graphs tell us that it's clear women have more tendency to buy intimate products and men are more homogeneous, showing not buying products women buy, like intimates, Dresses and Clothing Sets.

4.6 How much money was spent on each product's categories, among male and female customers considering the most popular countries?

If we refer not to the amount of bought products, but the price spent instead, we can see the result in the bar charts below.

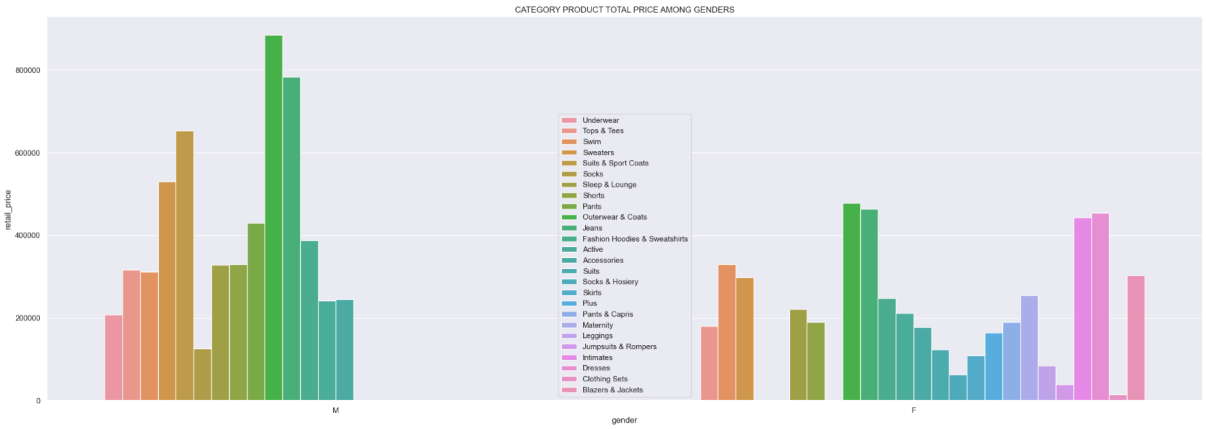


Fig. 4.19 -Distribution of total retail price by its category, for men and women.

The trend is similar for some of the most popular countries and we can see that clearly the type of products men buy tend to be more expensive, ie, even having a clear higher quantity of intimate products bought by women, the money spent is not as high as the one spent on outerwear and coats for men.

4.7 Considering the 4 most popular categories among both genders, how much money was spent on such products?

If we continue exploring the most popular products' categories , we can observe the data shown below, which is divided by the 3 most popular countries (in general matters the trend is the same) and it is clear that the 4 categories (underwear, suits and sport coats, jeans, outerwear and coats) men spend more money at are more expensive than the 4 corresponding to women (intimates, dresses, pants, outerwear and coats).

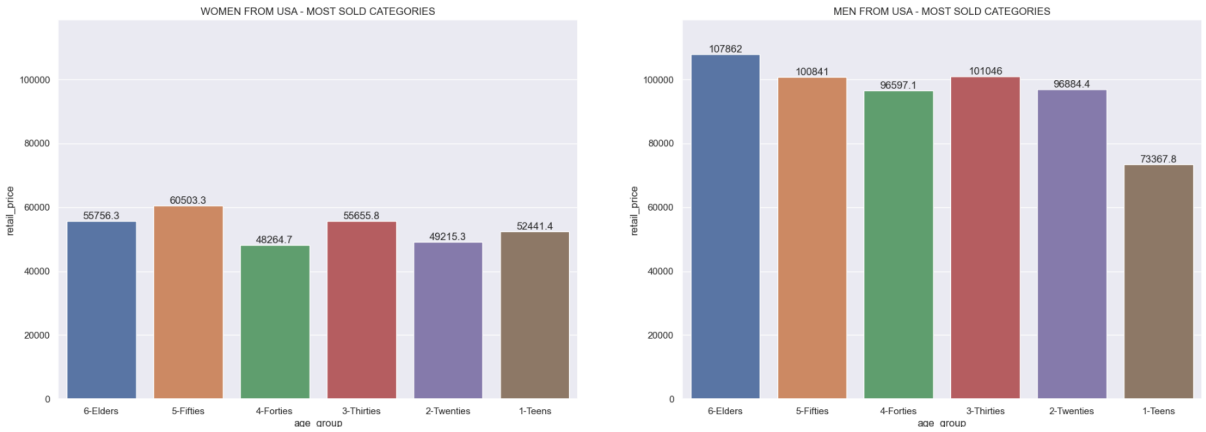
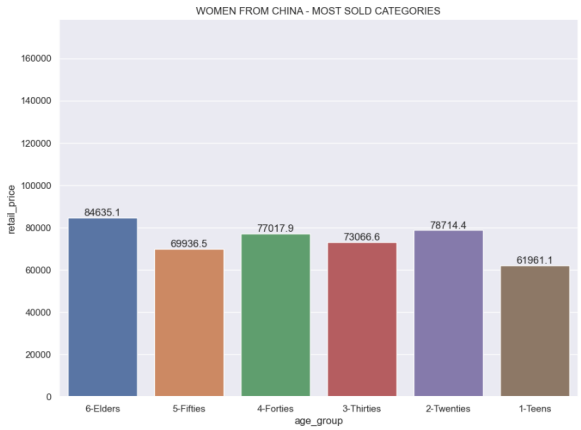


Fig. 4.20 -Distribution of total retail price by products' category and age group, considering 4 most popular products' categories among both genders, for the USA.



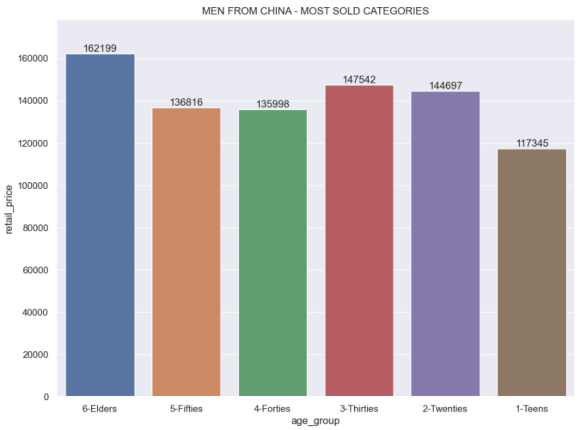
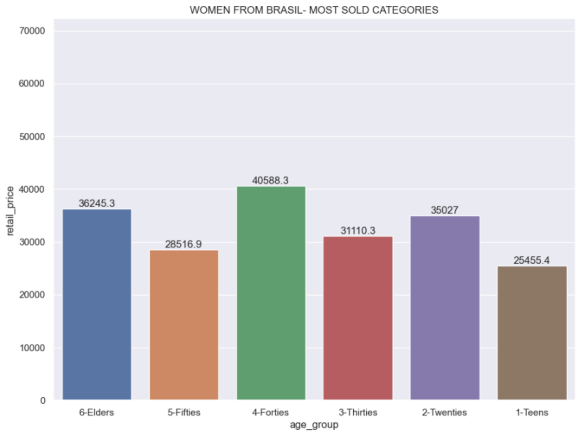


Fig. 4.21 -Distribution of total retail price by products' category and age group, considering 4 most popular products' categories among both genders, for China.



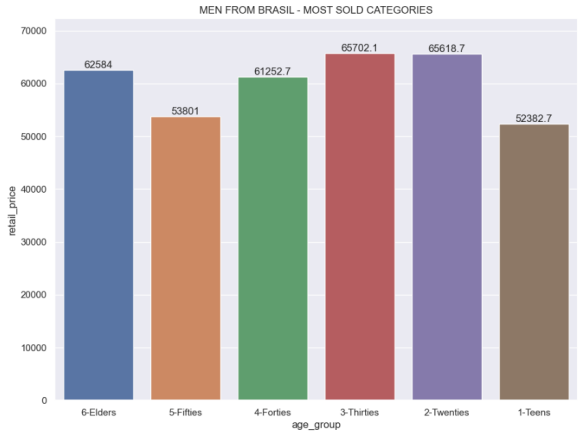


Fig. 4.22 - Distribution of total retail price by products' category and age group, considering 4 most popular products' categories among both genders, for Brazil.

Again, observing the data, generally speaking men tend to spend more money on their most popular items and the group age with more expenses is again the elders, except in the case of Brazil. In that country men in their thirties and twenties tend to spend more money compared with other ages.

4.8 What are the most popular brands among the 3 most popular countries?

Below we can see, in the table, the total retail price spent by the 3 most popular countries on the three most sold brands in each country.

Tab. 4.1 - Three most popular countries and the brands that are more sold there.

Country	Brand	Retail price ($)
United States	Calvin Klein	46106.91
	Carhartt	44481.32
	Diesel	41965.20
China	Calvin Klein	67841.21
	True Religion	65870.56
	Diesel	65027.07
Brasil	True Religion	28911.58
	Calvin Klein	27747.98
	7 For All Mankind	25725.37

If we create a bar chart which only considers the 5 brands shown above, distributed by both genders, we obtain the chart shown below.

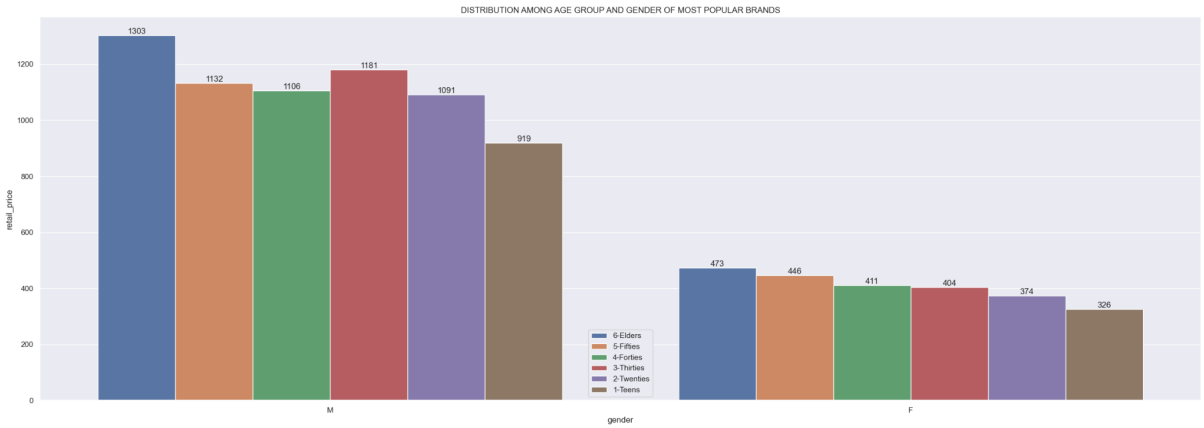


Fig. 4.23 - Distribution of total retail price by genders, considering the 5 most popular brands among the USA, China and Brazil, divided by gender.

By the chart above, once again we can conclude that men tend to spend more on expensive brands, and the elderly men are the ones who spend more money.
Other graphs for each one of the most popular countries were created and are not shown here, but the trend is similar, with a strict augmentation on the twenties and thirties age groups.

4.9 Can we find any specific pattern among the 1000 customers who spent more money in total?

By looking at the two graphics below, we can easily observe that considering the money spent in absolute terms, men spent more money than women in general, especially for countries like China and the United States.
The interesting fact is that in the second bar chart we can see that the average spent money is higher for women in several countries, and China is included in that group. It means that, besides having more Chinese men in that group of 1000 customers compared with Chinese women, each woman, individually, tends to spend more money, but their representation in the group is lower than men.

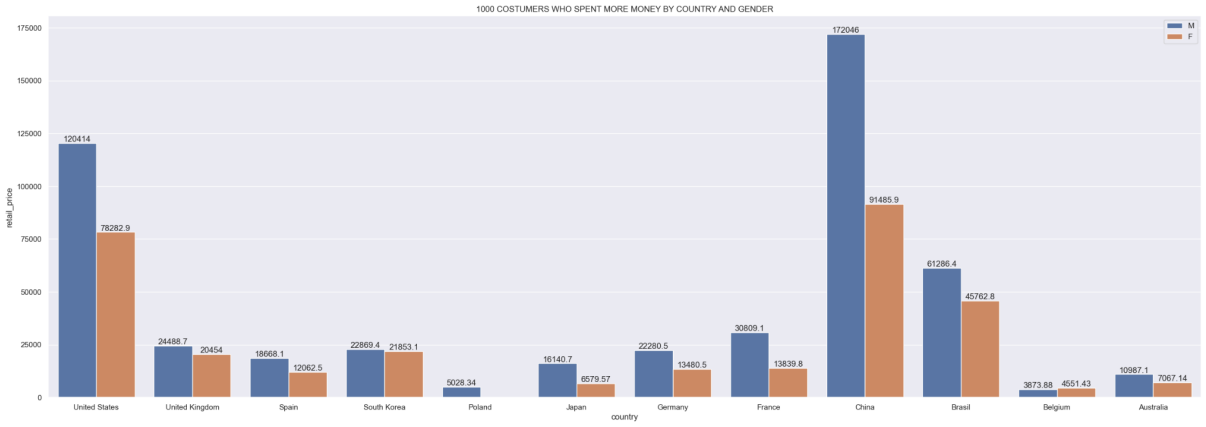


Fig. 4.24 - Distribution of total retail price by genders and countries, considering the top 1000 in terms of money spent.

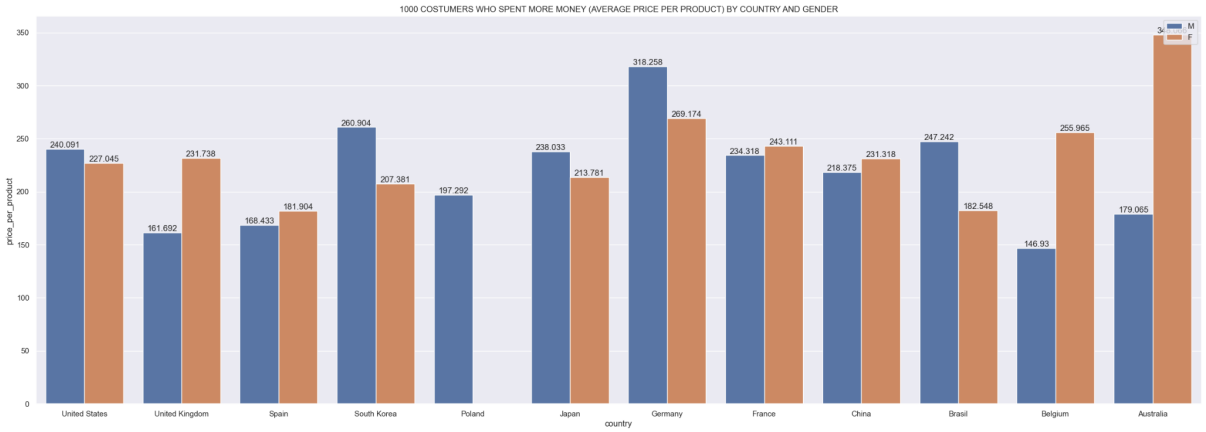


Fig. 4.25 - Distribution of total retail price by genders and countries, considering the top 1000 in terms of average money spent by person.

4.10 Can we find any specific pattern among the 1000 most ancient customers?

Among the 1000 customers that are subscribed for more time, we can't conclude for sure that those are the ones who created more orders because it would be needed more information about other customers not belonging to this group, but we can see clearly, by the second graph, that the fact that the client is more ancient doesn't necessarily indicates that he would spend more money in shopping because, for example for China, the difference of total retail price between the graph 4.24 and 4.26 is huge (144850 $ and 73393 $, for men and women respectively). The other countries show the same trend.

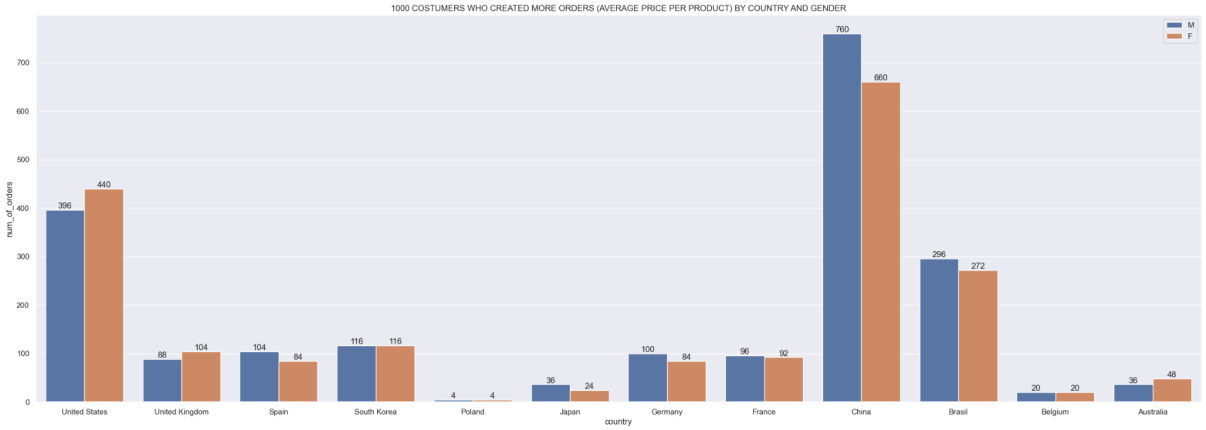


Fig. 4.26 - Distribution of total number of orders by genders and countries, considering the top 1000 in terms of subscription time.

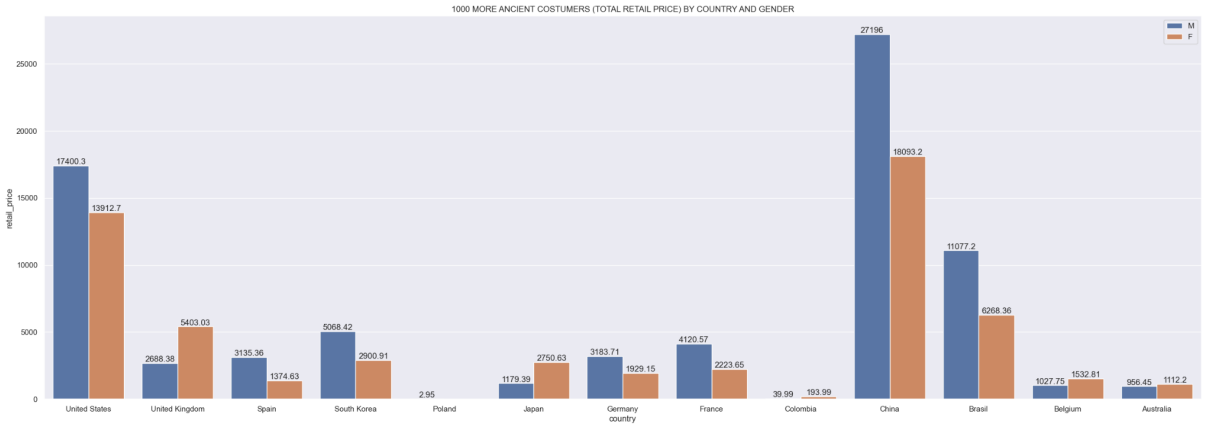


Fig. 4.27 - Distribution of total retail price by genders and countries, considering the top 1000 in terms of subscription time.

4.11 How the number of orders created varies along the time for the most 5 popular countries?

Two time series were created to show the variation of number of orders created in each month since January 2019 and we just can observe that we can't clearly find a pattern in general terms for each age group. The only conclusion from the graph 4.2.28 is that the average number of orders created in each month have been increasing, even if we see some instability.

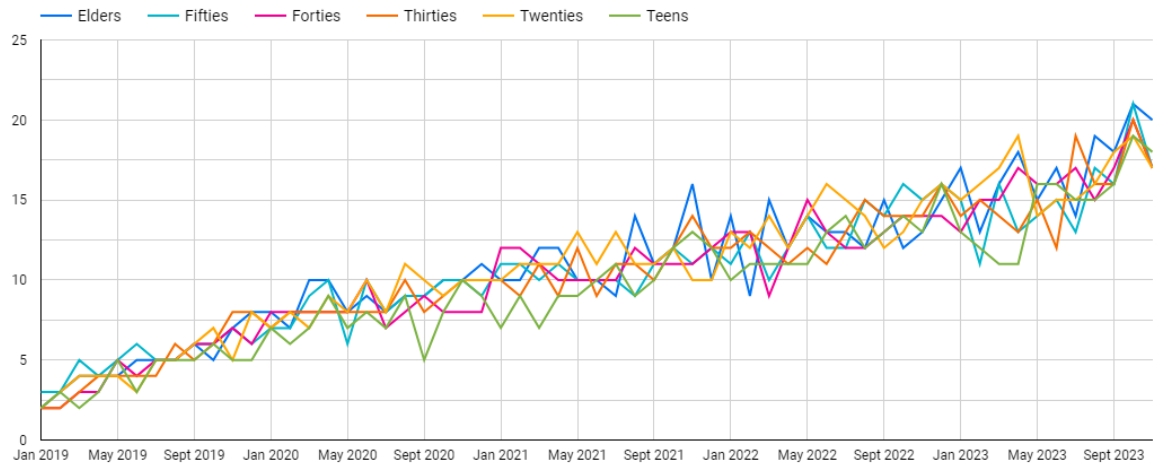


Fig. 4.28 - Variation of number of orders created along the time for each age group.

The same pattern can be observed among each country: the average number of orders is increasing with some instability, especially in the case of China.
South Korea is having a higher increment in the last few months.

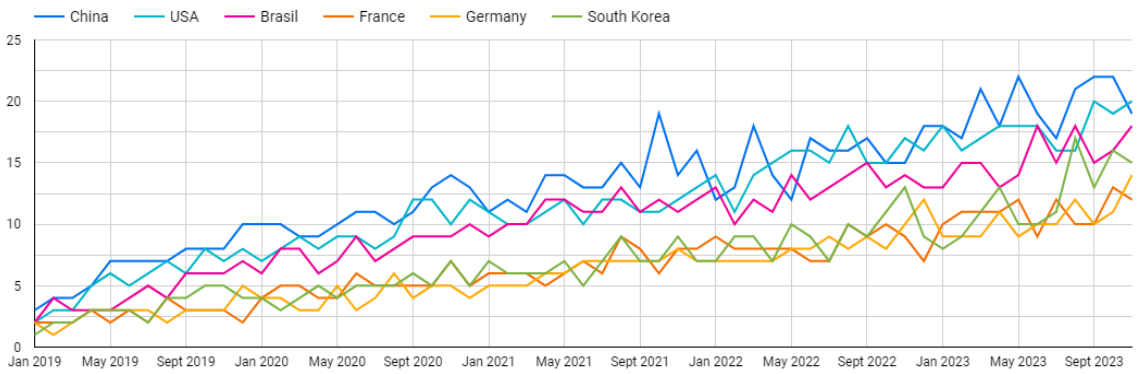


Fig. 4.29 - Variation of number of orders created along the time for each country.

4.12 How the number of customers created varies along the time for the most 5 popular countries?

Regarding the number of new subscriptions along the time, by the time series below, we can see that the values are quite estables, but all February months we see a clear decrement on the number of subscription, probably because it's an time of the year were people shop less, especially after Christmas and winter holidays.

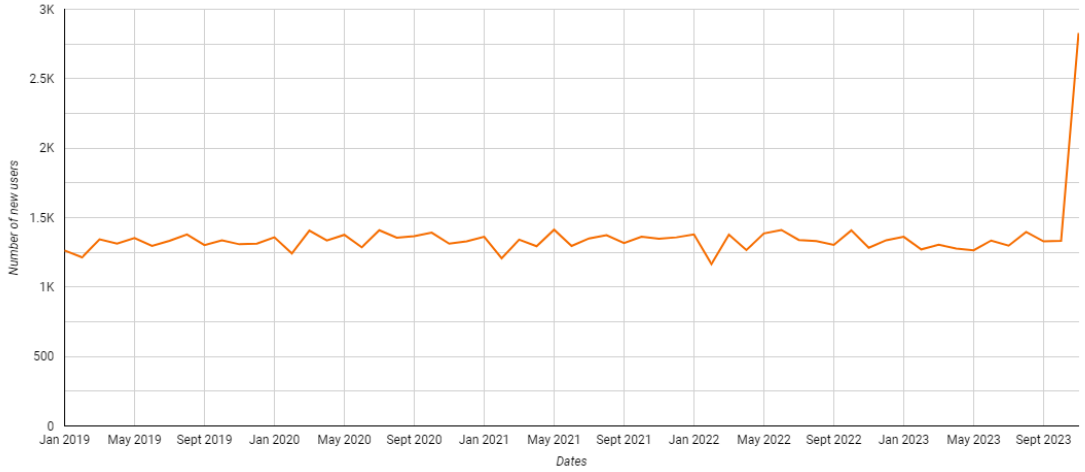


Fig. 4.30 - Variation of number of new subscriptions created along the time.

4.13 What is the distribution of citizens, age and gender among our customer base?

Below we can see the top 8 cities regarding the total retail price. Fig.4.2.31 represented the retail_price spent by customer and fig. 4.2.32 the same values in absolute terms and we can see clearly that in the first case the main cities don't belong to the most popular country, contralily the second one in which we can see that 87.5% of the 8 top cities are Chinese.

Fig. 4.31 - Distribution of average retail price per person among the top 8 cities in terms of average retail price per person.

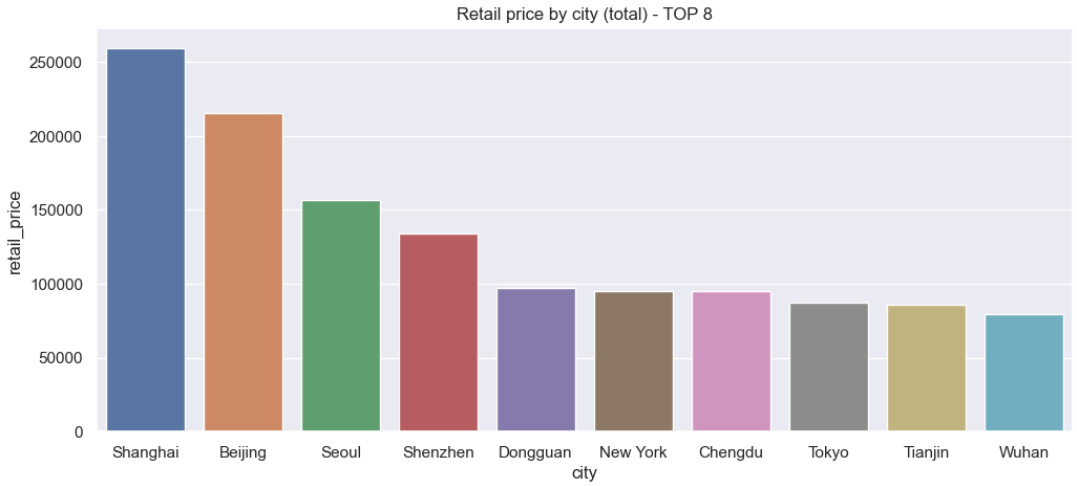


Fig. 4.32 - Distribution of retail price among the top 8 cities in terms of retail price.

More detailed representations were created to analyze the distribution among all age groups but the conclusions are not clear because for those cities, the average retail price and total retail price are similar for all age groups.
Regarding the genders, looking for the charts below it's clear that the men, on average, tend to spend more money than women among all top 6 cities.

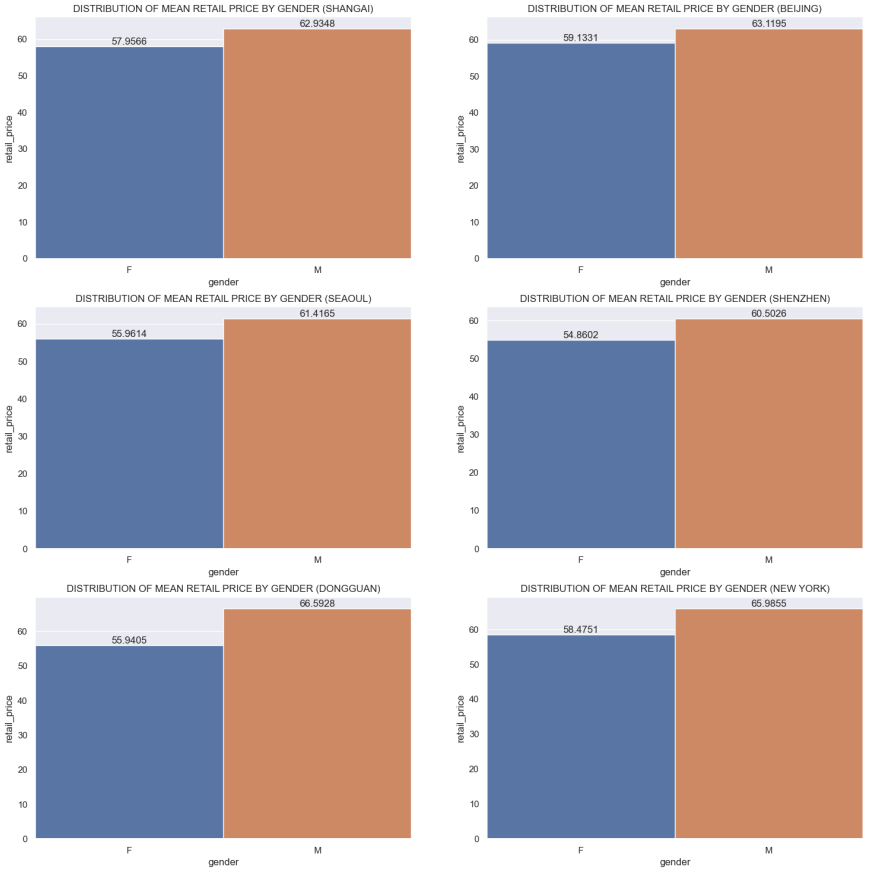


Fig. 4.33 - Distribution of average retail price per person among the top 6 cities and both genders, in terms of average retail price per person.

4.14 With this data is it possible to create a model that can perform predictions, like number of orders or total spent money, considering the population's features?

Regarding the consideration of models to perform predictions of retail prices, number of orders and other parameters, I calculated the correlation between all the numeric parameters and created 3 scatter plots, but the results were unsatisfactory. Below we can see some examples of such analysis:

Tab. 4.2 - Correlation between some of the parameters for all customers.

	Retail Price	**Customer time**	**Age**	**Gender**	**Number of products**
Retail Price	1	-0.0012	-0.0257	-0.0861	0.6745
Customer time	-0.0012	1	-0.0300	0.0397	0.0029
Age	-0.0257	-0.0300	1	0.0020	-0.0107
Gender	-0.0861	0.0397	0.0020	1	-0.0039
Number of products	0.6745	0.0029	-0.0107	-0.0039	1

Tab. 4.3 - Correlation between some of the parameters for female customers.

	Retail Price	**Customer time**	**Age**	**Gender**	**Number of products**
Retail Price	1	-0.0054	-0.0132	-	0.3836
Customer time	-0.0054	1	0.0042	-	0.0016
Age	-0.0132	0.0042	1	-	-0.0015
Gender	-	-	-	-	-
Number of products	0.3836	0.0016	-0.0015	-	1

The pattern is the same if we consider several specific customers among all the dataset, so with such weak correlations it is useless to proceed to the creation of models for predictions.

4.15 What happens to the total retail price if we double the sales of Calvin Klein, Carhartt, Diesel, True Religion and 7 for All Mankind, or do we do the same for most popular products among men and women?

In order to analyze the impact of the increment of the number of sold popular products in terms of brand and categories among women and men, we performed an estimating of the total retail price if we double the number of such products.
The following three tables show the current total retail prices by country, the estimated total retail prices considering the mentioned conditions by country and the augmentation in percentage, of the increment on the total price.
Among the main countries we can say the increase in percentage terms is around 8/9% which is significant if we think that we just considered that the sold products of most popular brands were doubled.

Fig. 4.34 - Distribution of total retail price for the top 10 countries.

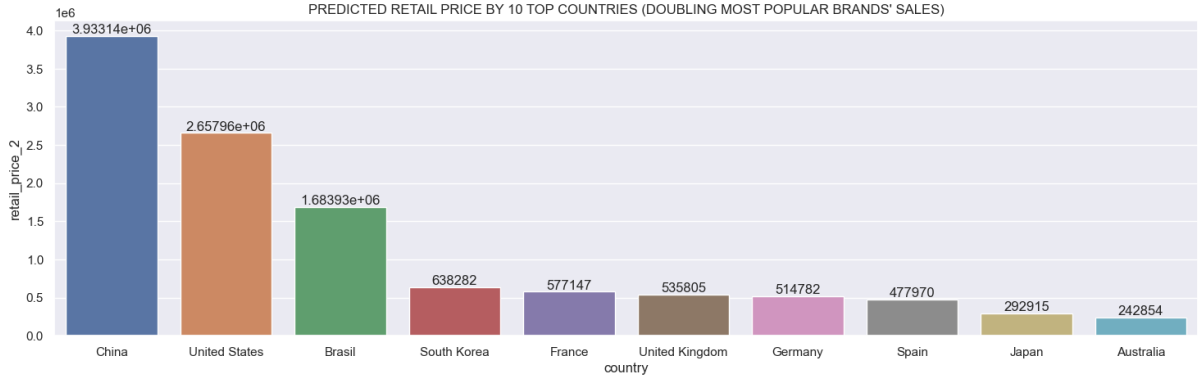


Fig. 4.35 - Distribution of total predicted retail price for the top 10 countries.

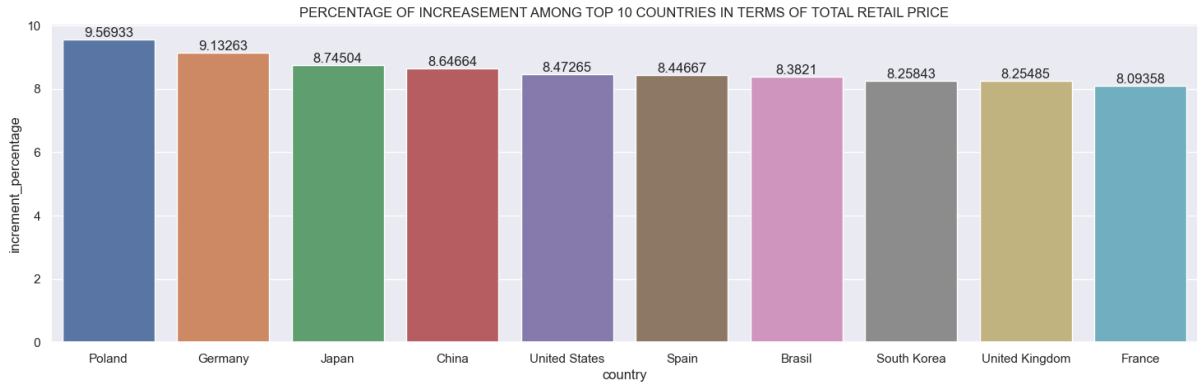


Fig. 4.36 - Distribution of increment percentage for the top 10 countries.

From the charts shown below we can easily conclude that the increment in terms of retail price is more significant for men rather than women, and it makes sense because the brands we are considering for the augmentation are more popular among male individuals.

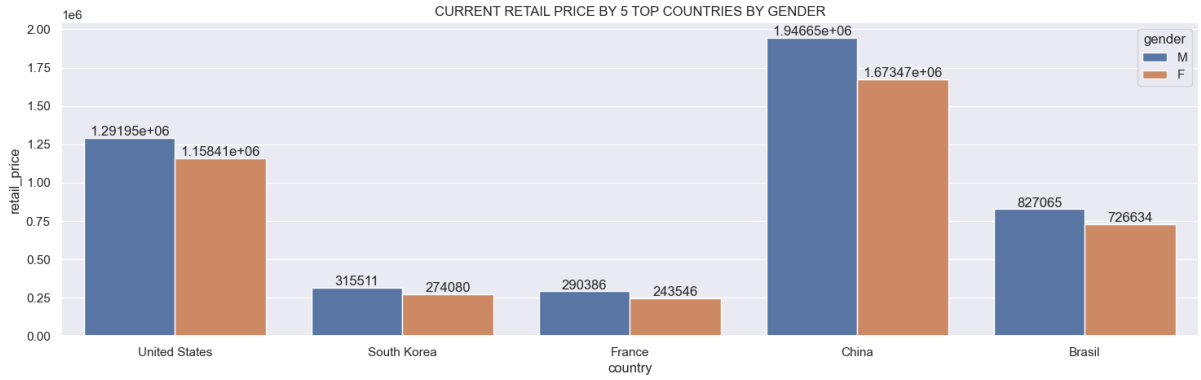


Fig. 4.37 - Distribution of total retail price for the top 10 countries, by country and gender.

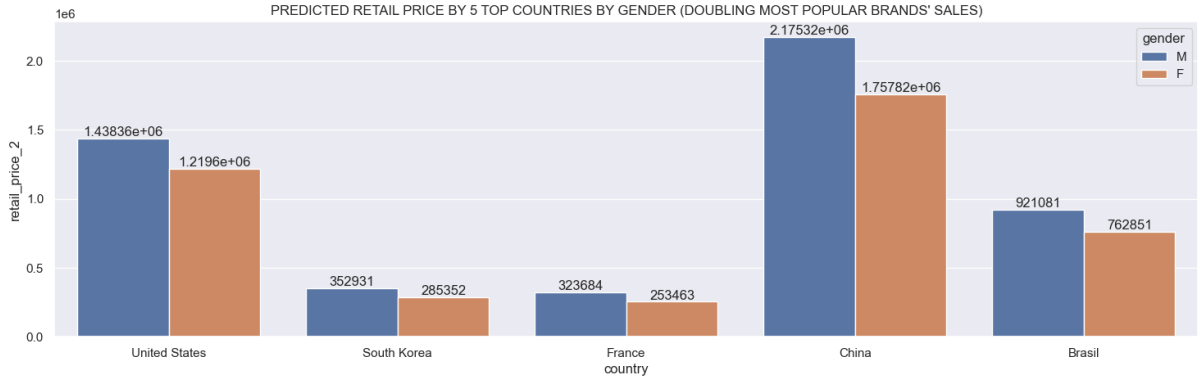


Fig. 4.38 - Distribution of total predicted retail price for the top 10 countries, by country and gender.

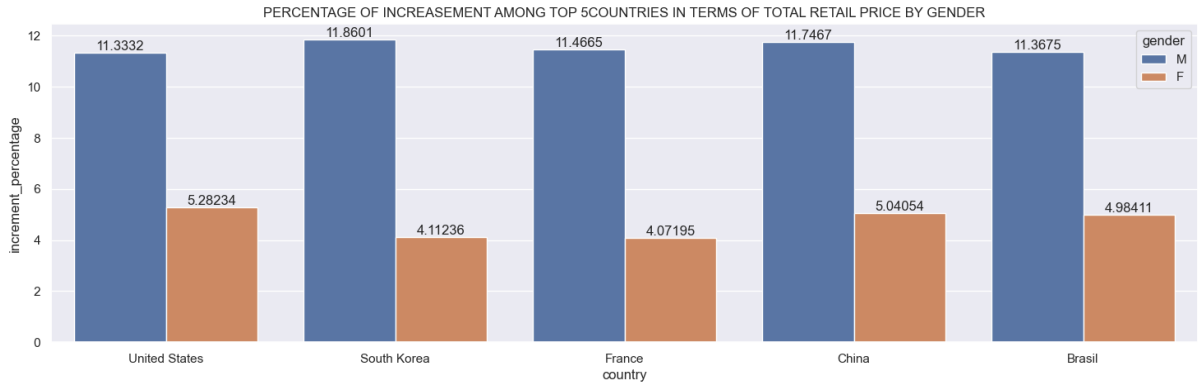


Fig. 4.39 - Distribution of increment percentage for the top 10 countries, by country and gender.

A similar study was performed, but now considering the products' category more popular among women and men, the results are quite peculiar because the percentage increment of the total retail price is more significant in countries that don't belong to the most popular, which correspond to values higher than 35%.

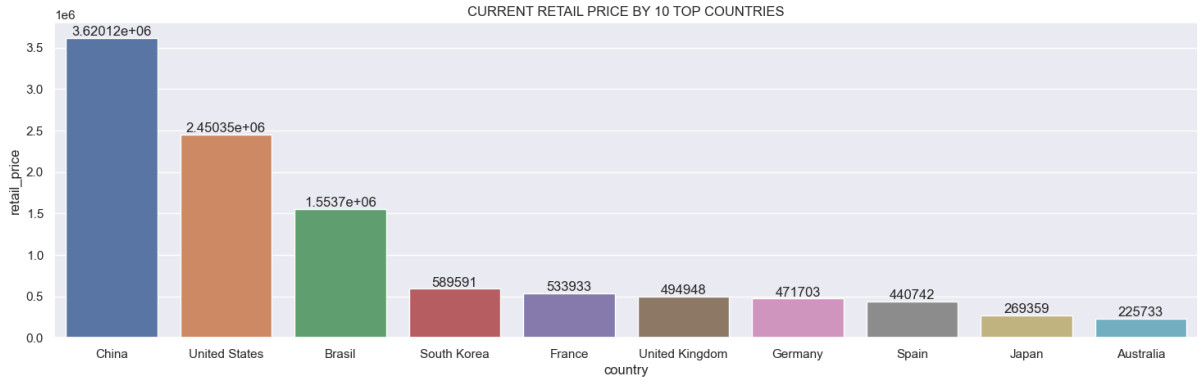


Fig. 4.40 - Distribution of total retail price for the top 10 countries, by country.

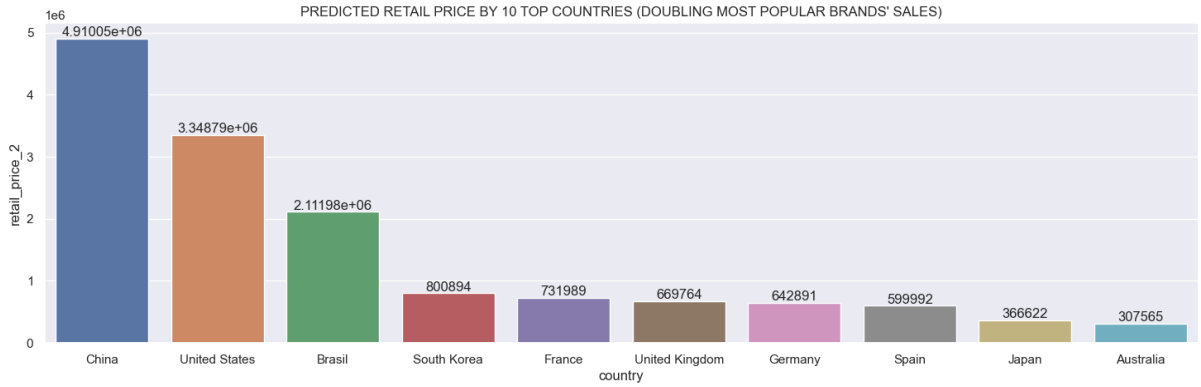


Fig. 4.41 - Distribution of predicted total retail price for the top 10 countries, by country.

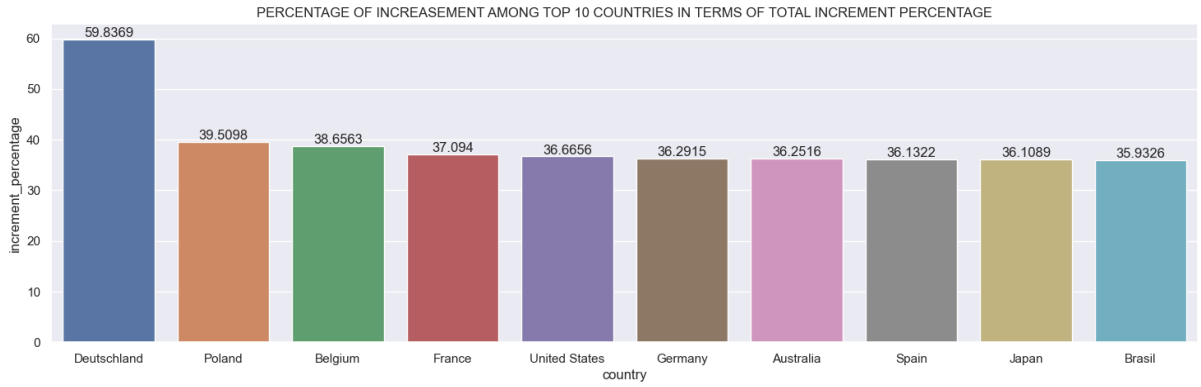


Fig. 4.42 - Distribution of increment percentage for the top 10 countries, by country.

If we look for the most popular countries, we can see that the increase in retail price in percentage is still high. We are talking about values around 43% and 27% for men and women respectively.

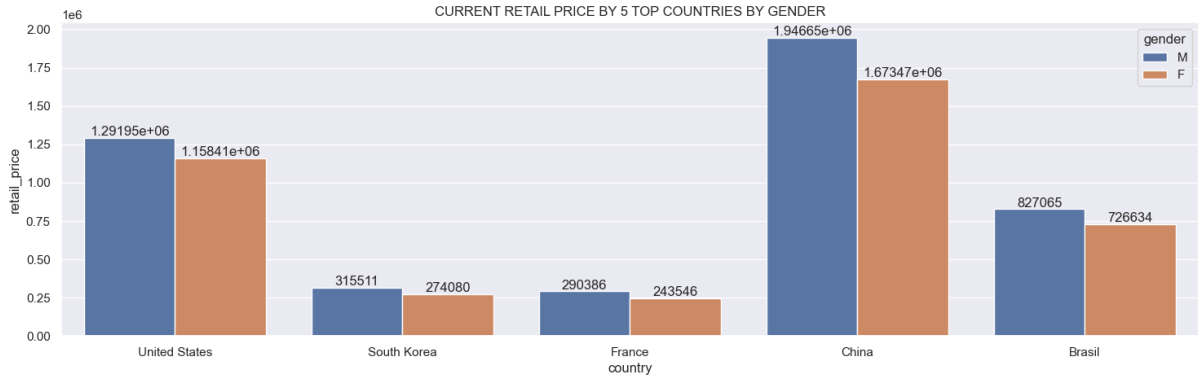


Fig. 4.3 - Distribution of total retail price for the top 5 countries, by country and genders.

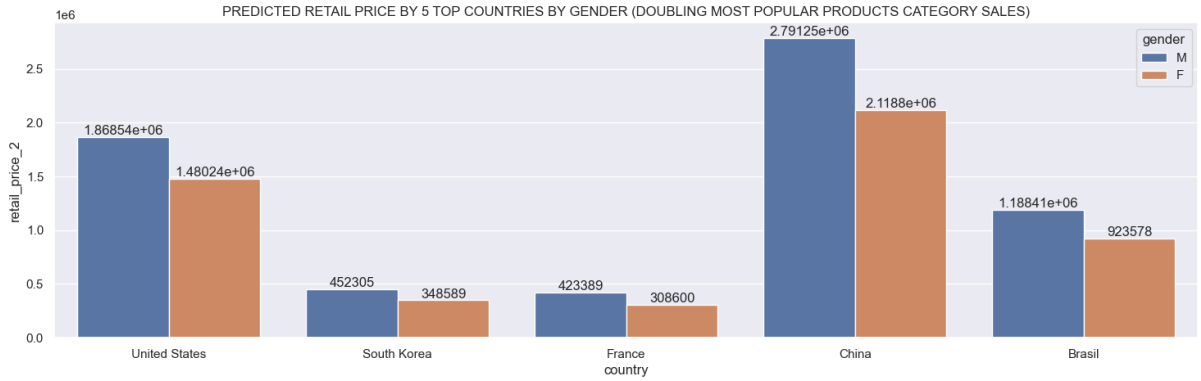


Fig. 4.44 - Distribution of predicted total retail price for the top 5 countries, by country and genders.

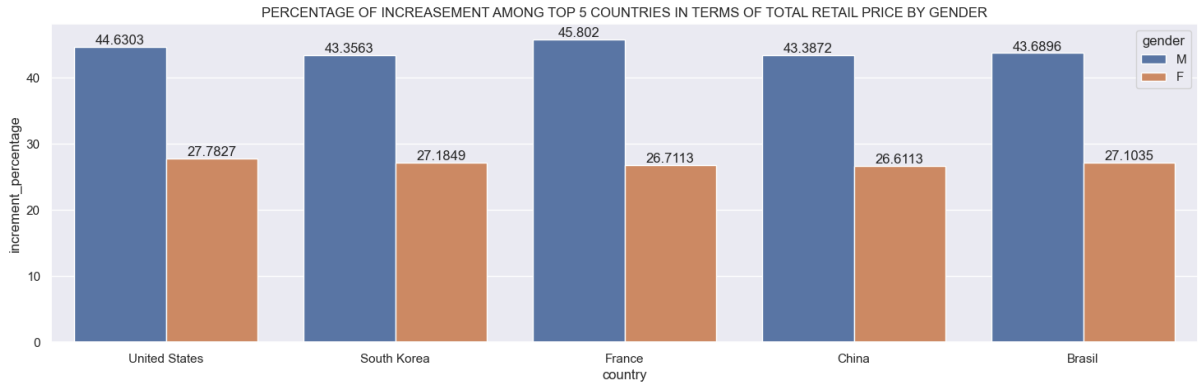


Fig. 4.45 - Distribution of increment percentage for the top 5 countries, by country and genders.

5. ANALYSIS AND CONCLUSIONS

After a intensive and deep exploration and analysis of the data, one can highlight the follow main conclusions:

- Genders are distributed almost equally
- Regarding the total retail price and number of orders, China, USA and Brazil are the 3 countries
- We have a clear difference between the behavior of men and women regarding products' categories and brands
- If we estimate the increasement of most popular sold product, the total price has a growth sound 43% for men and 27% for women
- In general the elderly men are the ones who tend to buy more expensive products

6. RECOMMENDED ACTIONS

Apart from the conclusions mentioned before, it is certain this data set has relevant information about the customers, but it would be interesting to collect more data to study for example the social class of each customer or job title.
Furthermore, some of the data could be studied in more detail and it would be interesting to perform the actions mentioned below.

- Regarding further actions, it would be interesting to study in more detail the behavior of each city population to find pattern
- The difference among different group ages could be analyze in more detail
- Regarding the insights, It is clear that if we pretend to increase the sales and the total retail price earn, we should target the increment of sales for most popular products among elderly men and in the countries were people tend to spend more expensive products