

## **PROJECT MILESTONE 3**

### **DATA PRESENTATION AND VISUALIZATION**

#### **DATA SCIENCE II**

#### **COSC 4337**

##### **Members**

Luong Thuy Ngo (2048659)

Syaneth Khatt (2114829)

Karla Castello (2138671)

## **DATASET INTRODUCTION**

Dataset: Customer Personality Analysis

- The selected dataset is the Customer Personality Analysis dataset, sourced from the Kaggle repository. With this data, we want to see common patterns of customers to see what type of customers is more likely to buy more from this company and determine certain strategies to target these customers in future campaign

Dataset link:

<https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis?resource=download>

## **DATA ANALYSIS**

Data analysis is a process of inspecting, cleansing, transforming, and modeling data so that we can derive some useful information from the data and use it for future predictions.

To seek predictions for this company, we will use the dataset “Customer Personality Analysis”. In this dataset, we have a total of 2240 customers with different types of information pertaining to them and their relationship with the company. Using the characteristics of each customer (i.e. age, marital status, education, number of children, income, education) and their involvement with the company (i.e. products bought, average spending, campaigns accepted), we will predict what type of personalities are most likely to contribute in sales within the company and how their marketing strategies can change in order to target a variety of different customers.

## ANALYSIS OF CUSTOMERS AMOUNT SPENT ON PRODUCTS BY PEOPLE



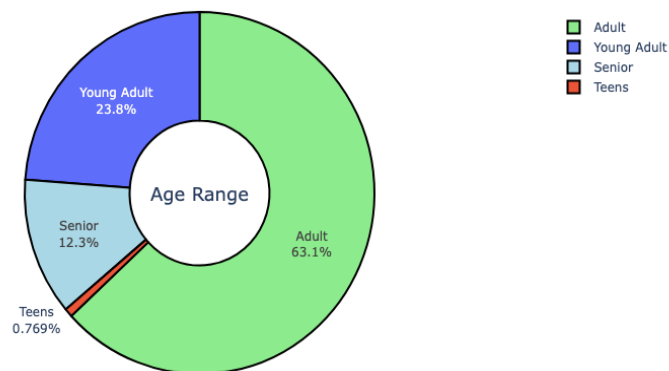
In the graph above, we are comparing each product with different types of customer information. In the upper left graph, we can see that the age group that spent the least amount on every product are teens, which are customers with the age of 13 to 19. Senior Adults, customers over the age of 60, are the customers that are most likely to spend more on every type of product. To target a wider range of ages, the company should promote its items in a way where younger customers are most likely to see them. This could mean an increase in social media presence.

In the upper right graph, we can predict that people that have an income between \$75,000 and \$100,000 are more likely to buy wine. Then, we can see that people who make over \$150,000 are most likely to buy meat from this company. Lastly, people who make between \$100,000 to \$150,000 are more likely to buy every other product. In the lower left graph, we can see that people with no children are more likely to spend than people with children. Lastly, in the lower right graph we can see that single people spend slightly more on each product than people with partners.

From these predictions, the type of customers that this company has yielded are single, older customers with no children that make over \$75,000. Almost half of the company's revenue

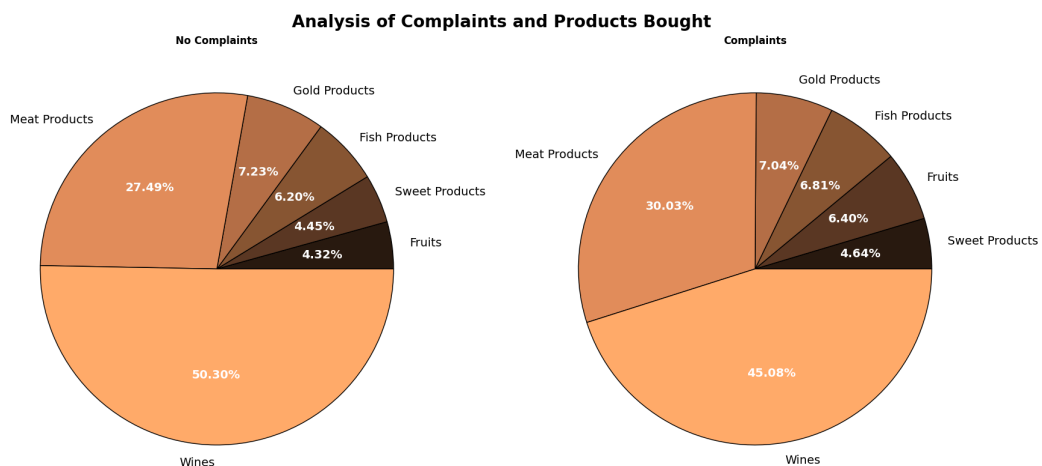
came from wine sales, while over a quarter of the revenue came from meat sales. Since most customers make over \$75,000 and the items that are bought more are meats and wines, this company does not sell average-priced products that people making less than \$75,000 can buy. This company may need to lower its prices on regular products that middle-class customers and families are able to buy comfortably.

## ANALYSIS OF THE MOST AGE RANGE BUYING FROM OUR COMPANY



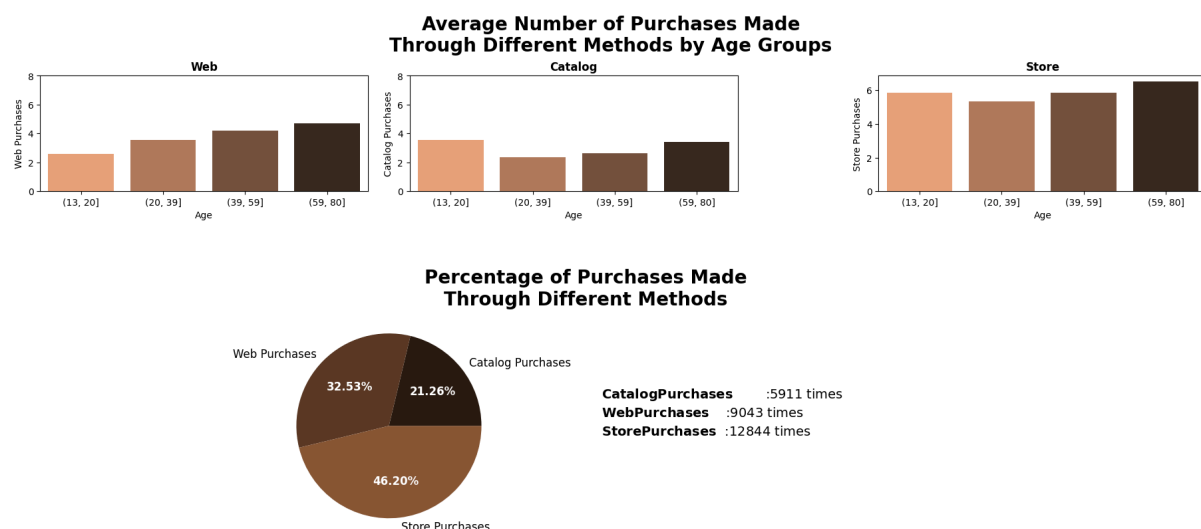
According to the above graph, our target customers are people from the age of 36 to 59, 63.1% of all revenue. This may be because the products that the company sells are only products that adults buy for their homes. From this information, the company should direct its market strategy to this group of people. If they want to expand their clientele, the company should start selling products that teens and seniors would want to buy.

## ANALYSIS OF COMPLAINTS AND PRODUCTS BOUGHT



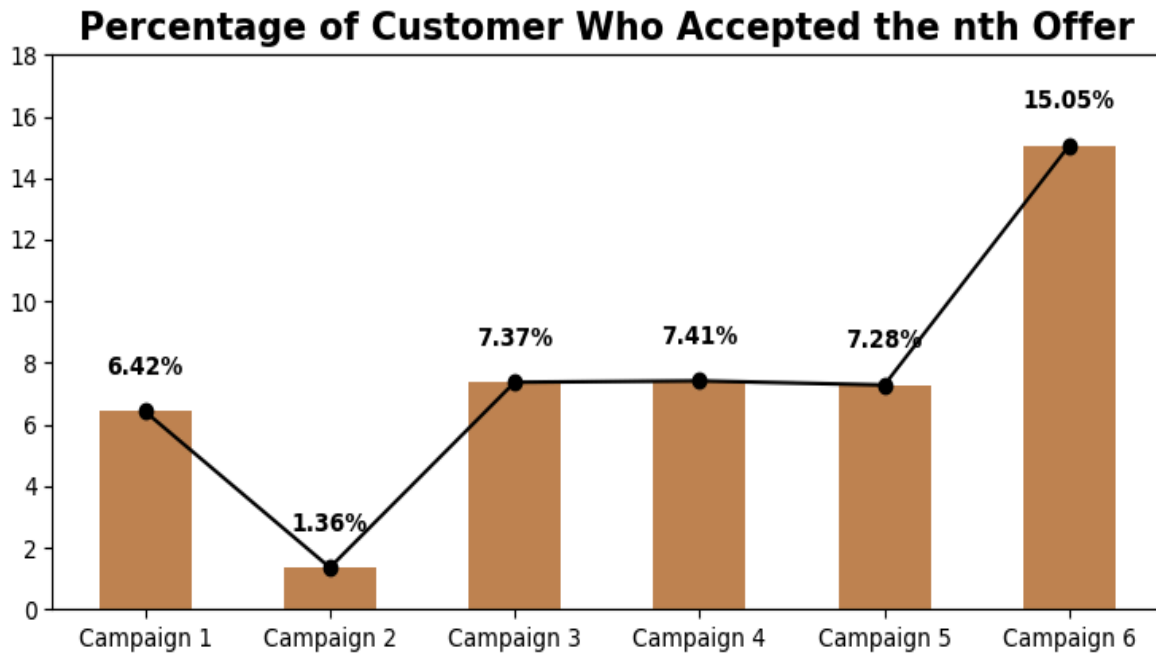
We now analyze each of the company's products with the complaints people made to the company. People who bought did not file a complaint bought less meat than people who did file a complaint, with 27.49% compared to 30.03%. Also, people who did not file a complaint bought fewer fruits than people who did file a complaint, again with 4.32% compared to 6.4%. This could mean that people most likely complained due to the meats and the fruits. Although the meat products are a top-seller in the company, it could just mean that some people did not like them. On the other hand, fruits are not a top-seller therefore this could mean that this product is not as good as it is marketed within the company. A solution is to source these fruit products from farms that have higher-quality fruits.

## ANALYSIS OF PURCHASES MADE THROUGH DIFFERENT METHODS BY AGE GROUP



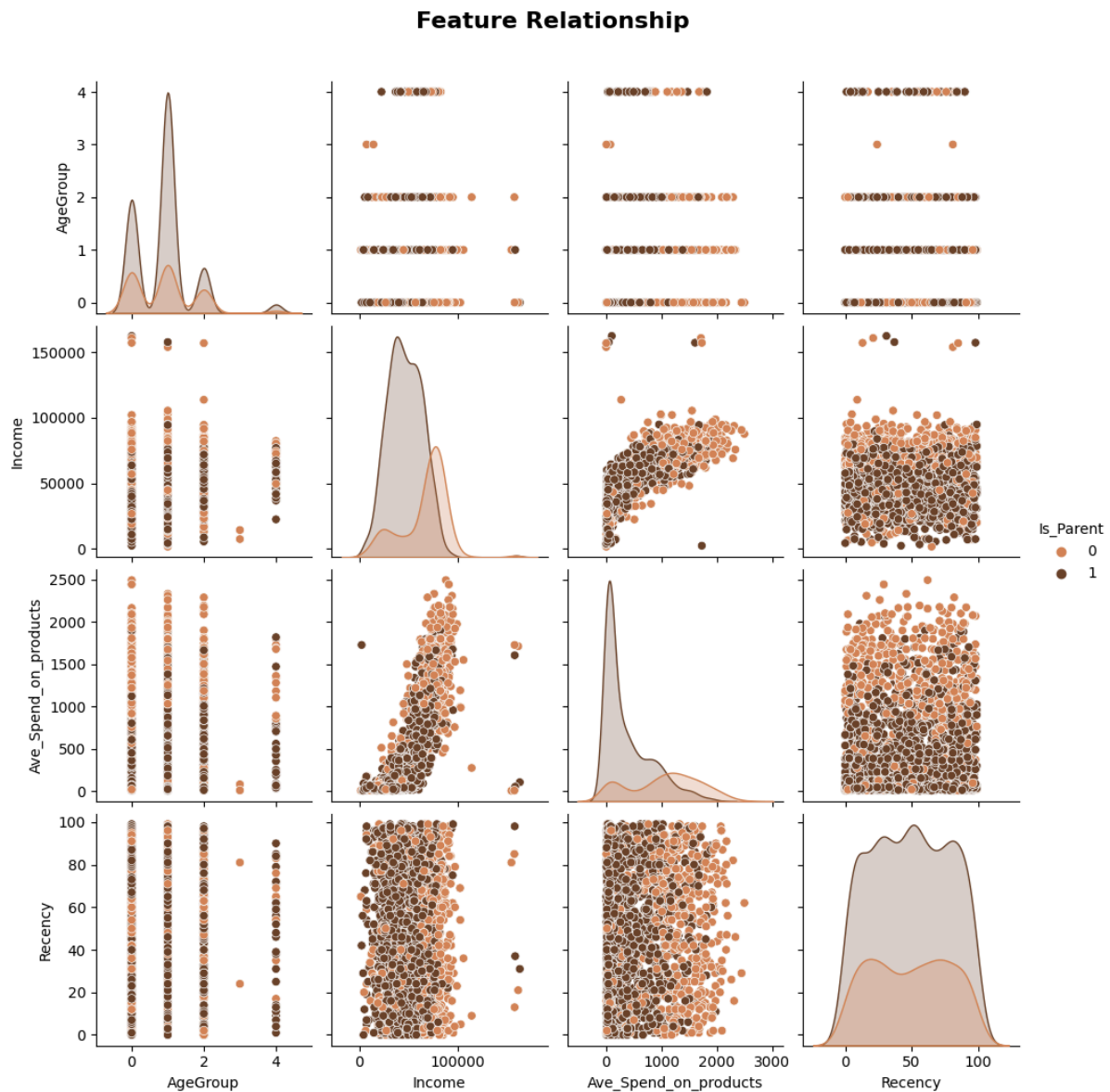
In the graphs above, we can see that most of the purchases made from this company are in-store purchases, at 46.2%, while web purchases are second at 32.53%. People between the ages of 59 and 80 are most likely to buy from this company in-store or through their website. Since in-store purchases are most popular within this company, their store marketing is very effective. Only around 21% of purchases are catalog purchases, therefore it would be better for this company to focus on improving their website to attract more customers and less on the catalog. This company also seems to attract more older customers, therefore a marketing strategy towards young adults may be needed to increase sales overall.

## ANALYSIS OF THE PERCENTAGE OF CUSTOMERS WHO ACCEPTED THE Nth OFFER



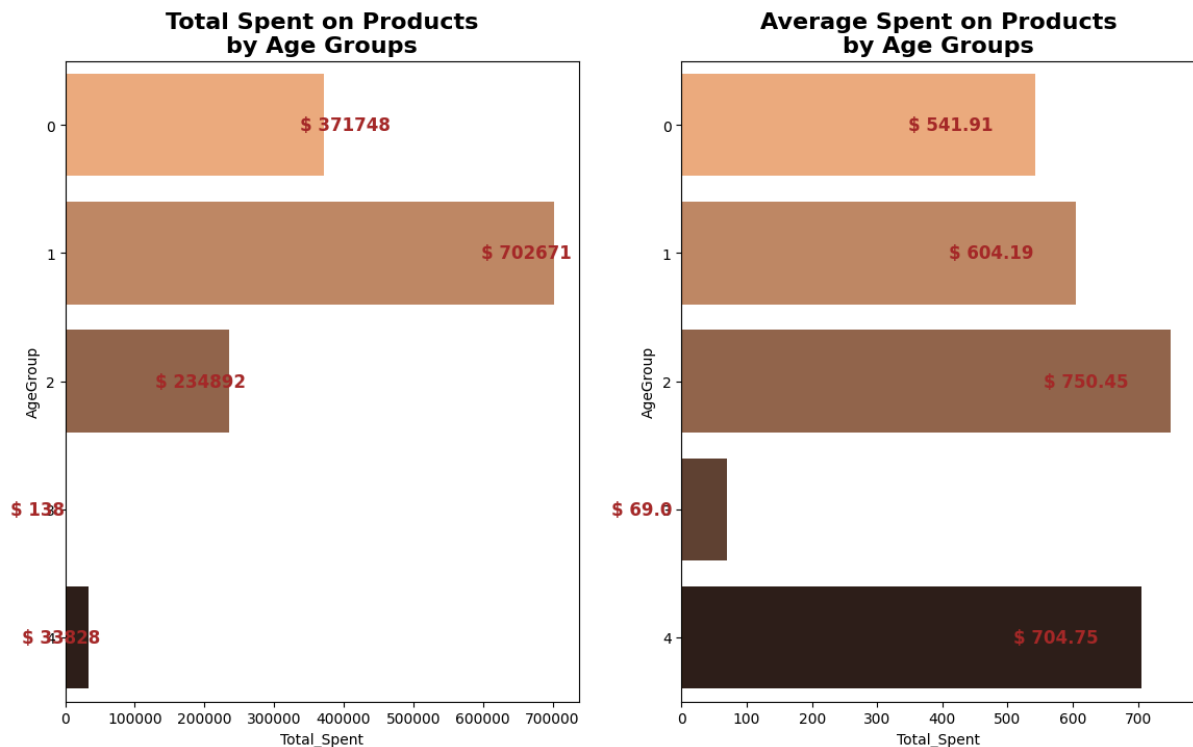
This graph shows the percentage of our customers who accepted offers from the company. The percentage of people who directly accepted the offer in the first campaign was just 6.42%. The number became even smaller in the second campaign with just 1.36% of the customers. The last campaign successfully drew the most customers compared to the previous campaigns with 15.05% percent of customers. The strategies of the first, third, fourth, and fifth campaigns yielded the same percentage of customers therefore there was no improvement. However, in the last campaign, there were more customers that participated in the campaign, therefore improving sales. This means that this strategy in campaigning worked for most customers than in previous campaigns.

## FEATURE RELATIONSHIP



From the relationships above, we observe that parents spent less money on company products on average than single people. However, we see an opposite pattern in the relationship between recency and marital status. The highest number of days since parents last bought products was around 40 days while single people's was up to 100 days. In addition, people who have yet to marry have higher incomes than married people. This company attracts both married and single people, but it differs in different categories. Difference relationships with customers' marital status and other attributes show that parents contributed to the company's sales more than single people. The company should reinforce its campaigns or development strategies to attract more single people in every purchase method.

## ANALYSIS OF TOTAL SPENT ON PRODUCTS BY AGE GROUPS



Age Group:

- 0 is Teen (Age  $\leq 19$ )
- 1 is Adult ( $20 \leq \text{Age} \leq 39$ )
- 2 is Middle Age ( $40 \leq \text{Age} \leq 59$ )
- 3 is Seinor (Age  $> 60$ )

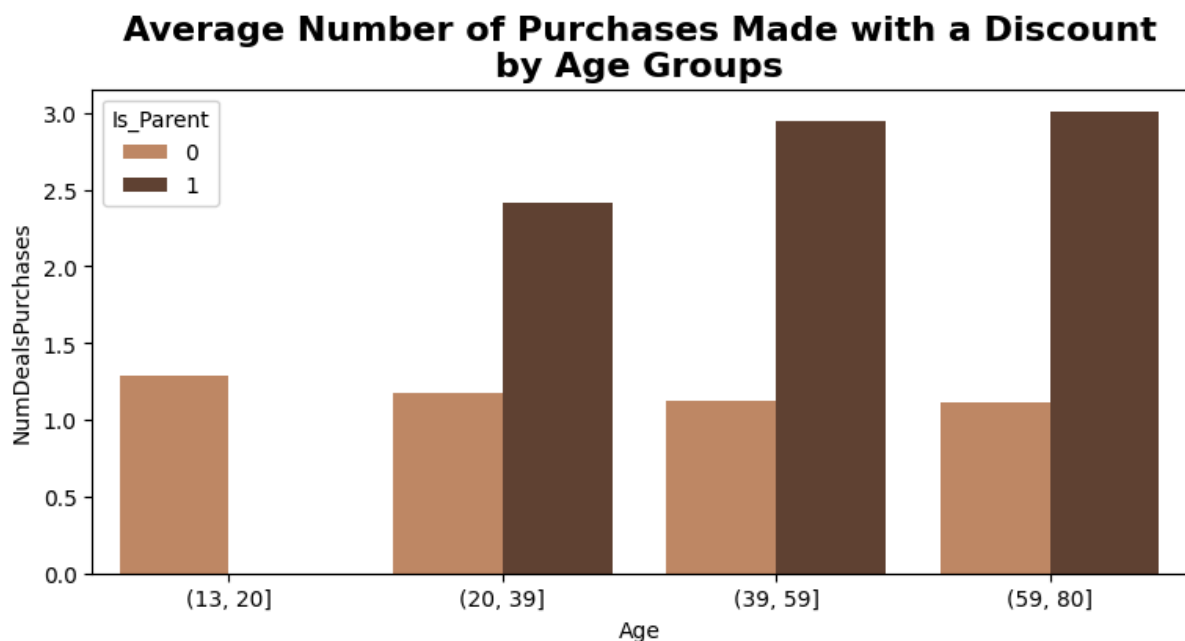
We now want to focus on the distribution across all age groups with their average spending and total spending. From the bar charts shown, the total spending of age group 1 (between 20 and 39) is the highest among other age groups, thus their average spending is also as high as \$604.19. Interestingly, middle aged-group has the opposite distribution between their average and total spending. The total spent on products for middle-aged adults came in third with \$234,829 but average spending per person in this unit accumulated up to \$750.45 which is noticeably high among all groups. Having said this, the company has most sales based on adults and middle-aged groups, thus, the company should mainly focus on marketing towards these two groups and teens to maximize their profits.

Now, we want to show the significant point of the relationship between age groups and their expenses, and purchase methods. If we look at these two relationships from side to side, we can definitely say that the company should improve its sales on all types of purchases,



especially towards teens, middle-aged adults, and older adults. As we learn that these age groups are those spent the most in total spent and as well as average spent.

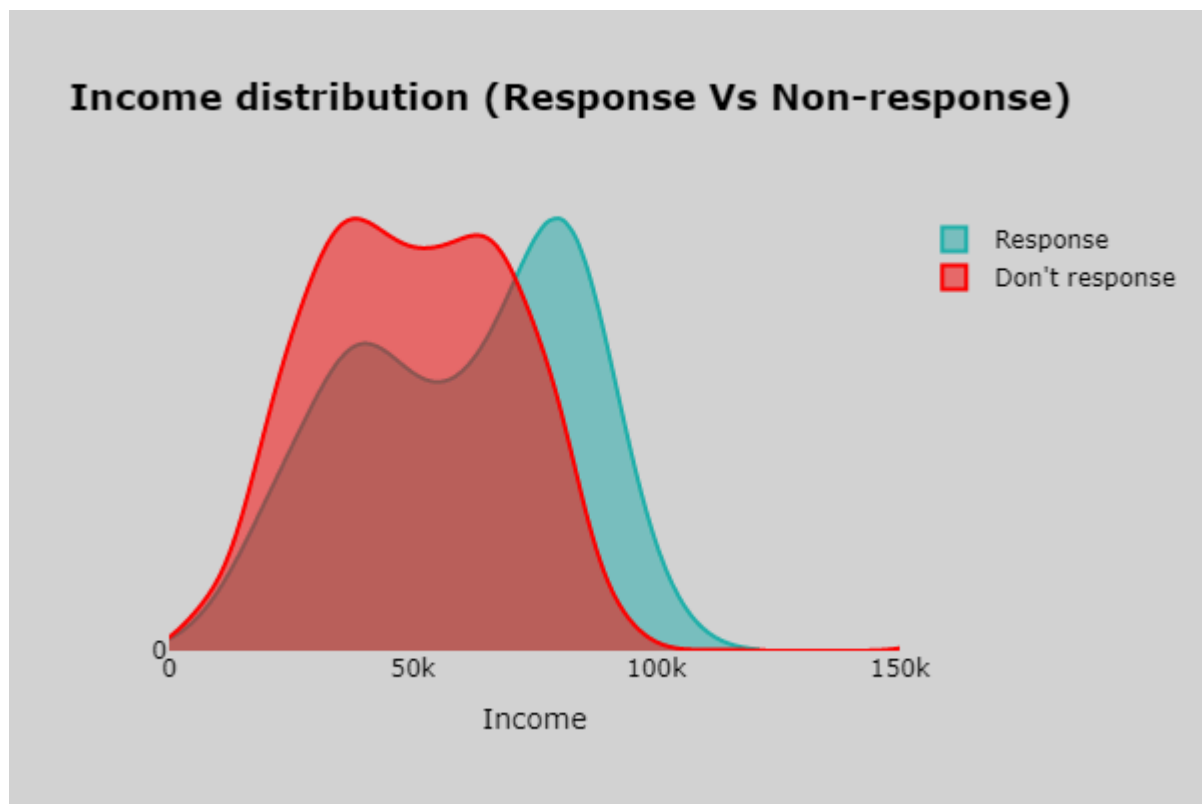
### ANALYSIS OF THE AVERAGE NUMBER OF PURCHASES MADE WITH A DISCOUNT BY AGE GROUPS



(0 represents single people, 1 represents parents)

The company attracted a lot of parents aged between 20 and 80 with a total number of deals purchases. Single people across all age groups had an almost similar number of deals purchases and lower than customers who classified as parents. We observe that no teens are parents, therefore, no result is shown on this portion except for single teens. Parents bought discounted products at an average of 2 to 3 and accumulated higher sales in these categories. The efficient way for the company to increase sales based on deals purchased is to advertise more deals and create more deals for customers such as giving out coupons or free samples so that it can gather more single people or reinforce the contribution of parents customers to purchase more items or try out new things from this company.

## ANALYSIS OF WHETHER INCOME HAS A BIG EFFECT ON THE RESPONSE RATE OR IT IS OUR PRODUCTS DID NOT ATTRACT THE CUSTOMERS



This distribution indicates that customers' income does not have a huge impact on responses which is a good sign. This could also mean that the price of products is reasonable and customers are satisfied with the company's decision. We can see that there is a difference in response and non-response when customers' income is around 30k to 70k, however, this should not be an issue. It could be because the average income customers would not want to purchase some products whose price is a little higher than others which led to more of non-responses. Starting from an income of around 70k and higher, we observe both response and non-response dropped gradually as income proceed to go higher. This could also show that customers are satisfied with the products, and other involving categories of the company. Even though the response does not have an impact on sales but without this category will not give feedback from different types of customers. Receiving comments or responses from customers is a good evaluation for the further development of the company.

## CONCLUSION

After thorough observations in different categories, we have obtained significant assumptions and efficient ways that the company should pay attention to. The company attracted mostly adults, senior adults, and middle-aged adults except for teens who barely see a contribution to sales, especially since these age groups have a higher average amount spent on every product. The company should reinforce strong advertisements towards these groups as they do now and perhaps add more teenage products to attract more teens. It can also be seen that single people and parents with 0 and 1 child spent money in this store among all other different types of customers. In addition to this, most customers with an average income of 60k and higher contributed to the company's sales and most of the sales came from wines and meats followed by other products. From this perspective, the company should have more inventory of wine products and meat products to meet the demand of customers. At the same time, this company should bring more quality products at reasonable prices to increase sales of other products. Most importantly, customers like going in-store to purchase products of over 50% in total, than web purchases and catalog purchases. By observing this characteristic of customers, the company should invest in improving its website such as design and technical alterations so it is more convenient to place orders through the website or application. They might also focus on the delivery method which is the most ideal method for customers these days. Catalog purchase is more of an outdated style, but it still works for this company, therefore, it can be improved by design or make it easy to access to customers. Regarding marketing campaigns, they will need to launch more campaigns with free samples or give out coupons to different types of customers. Since Campaign 6 was very effective and appealed to a lot of customers, they could recycle or develop that idea to make it even more intriguing. Overall, this company is pretty well performing, and it needs to focus on increasing sales of different products on top of wines and meat products. The company should also create more surveys to get more feedback and keep up with the customer's demands.

## REFERENCES

PennState: Statistics Online Courses, “11.6 - Example: Places Rated after

Standardization: STAT 505”

<https://online.stat.psu.edu/stat505/lesson/11/11.6>

## INSTRUCTIONS TO RUN THE CODE

1. Download the dataset from this link:

<https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis/download?datasetVersionNumber=1>

2. Extract the zip folder to get the file named 'marketing\_campaign.csv'
3. Place the dataset into the same folder as the file named 'notebook\_thirdDeliverable.ipynb'
4. Open the file 'notebook\_thirdDeliverable.ipynb' with Jupyter Notebook
5. Run the code (will need to download the necessary libraries if it is not pre-installed on your computer)