Thousands of people visit supermarkets every day to buy food and housing supplies. By analyzing this data, we can learn more about customer preferences regarding payment methods, visiting times, and membership in loyalty programs — based on location or gender. Then, we can adjust every single supermarket to increase customer satisfaction and profits.

**ASK**

[This dataset](https://www.kaggle.com/datasets/aungpyaeap/supermarket-sales) contains data about selling points and the customers of three branches of a supermarket company collected over three months. By analyzing this dataset, we can answer these questions:

* Which branch has the best results in the loyalty program?
* Does the membership depend on customer rating?
* Does gross income depend on the proportion of customers in the loyalty program? On payment method?
* Are there any differences in indicators between men and women?
* Which product category generates the highest income?

**Prepare**

[**This dataset**](https://www.kaggle.com/datasets/aungpyaeap/supermarket-sales) will come from a supermarket sales dataset on kaggle. It’s a historical dataset that covers 3 months between 3 branches of a supermarket at different locations. It measures what male and females bought at these locations and how this can help the supermarket make future decisions on its products. Made for predictive analytic purposes, and developing techniques for data mining, predictive modeling, and forecasting unknown events.

**Process**

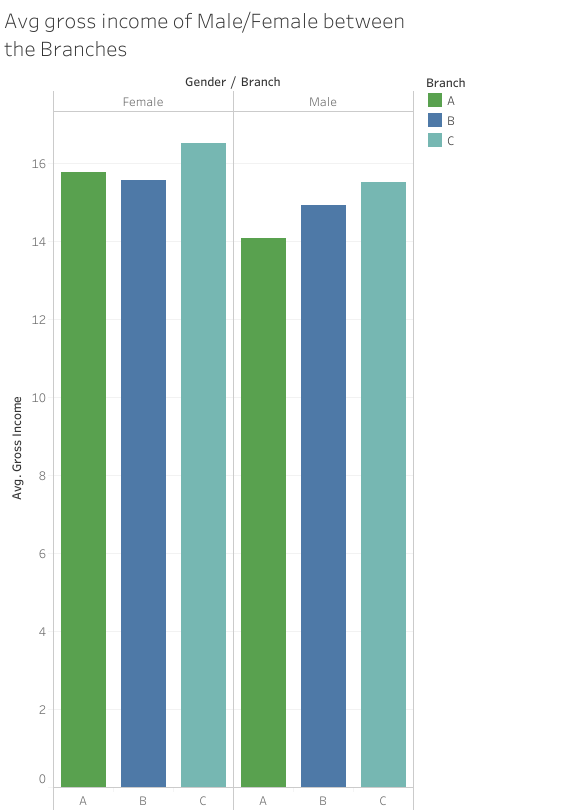
Excel, Tableau and SQL google Bigquery where used for this analysis, as they were the tools that I felt made it easiest for me to sort with. While working with the **supermarket\_sales** dataset, I chose to focus on the Branch, Customer\_type, Gender, and Product line in order to divide the customers into measurable groups and observe their purchases. I also worked with the payment column to gain an idea of customer habits on how they chose to pay for their goods, and the gross income along with the rating to see how customers viewed the products. Overall the columns I chose were able to give me insight on the customer loyalty program and how much it's benefitting the supermarket from a sales perspective, and how its 3 month data collection could help it figure out future plans that can improve its business.

Excel was used to check for spelling errors, duplicates, format issues, and ultimately the data was clean. I also used Excel to create a few pivot tables to gain a quick note of some patterns I noticed in the data, like a sum to the grand total, and how quantity correlated to the other fields. After observing this, I sent the data to SQL Bigquery in order to pull different information out of the data in order to gain a better understanding on how I could answer my questions. I chose to focus my calculations around the average of gross income even though it was similar to the tax column, because it followed the same comparison of the products as total and the fewer numbers made it easier to visualize. [Line of code](https://docs.google.com/document/u/0/d/1RrIdd8kKqwARORL4ii_TzvmCNQ1UCqevCCHLt7p0-Os/edit)

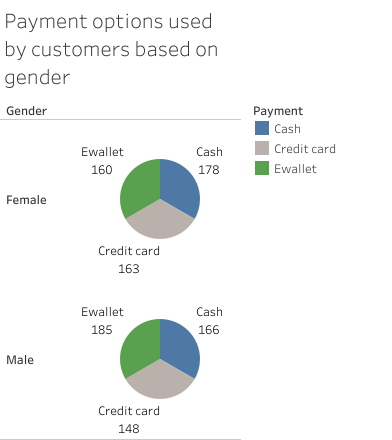
**Analyze**

Right off the bat I was drawn to the branches and the proportion of Males to Females. I knew these two sections had the information on the behavior of the customers within the different sectors, and direct insight into the mind of the consumer. Observing the buying preferences and how much each person bought can help the store determine its most popular products, and how it can improve its loyalty program. Seeing who spends the most, what's the favorite, what benefits can help the loyalty program separate itself from the normal customers, and help the store elevate to being a brand preference to the competition. However, I noticed that even though there were differences between them, there wasn't a significant difference between the loyalty program members and normal customers. There was a slightly wider comparison from females to males and overall the gaps between the measurements was small, but there is enough information to help the supermarket determine which direction they’d like to move forward with.

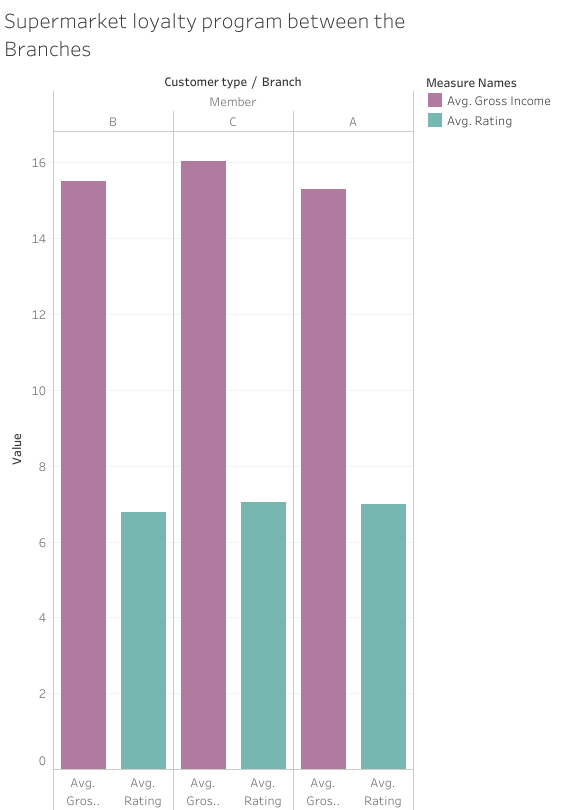
**Visualize**

****

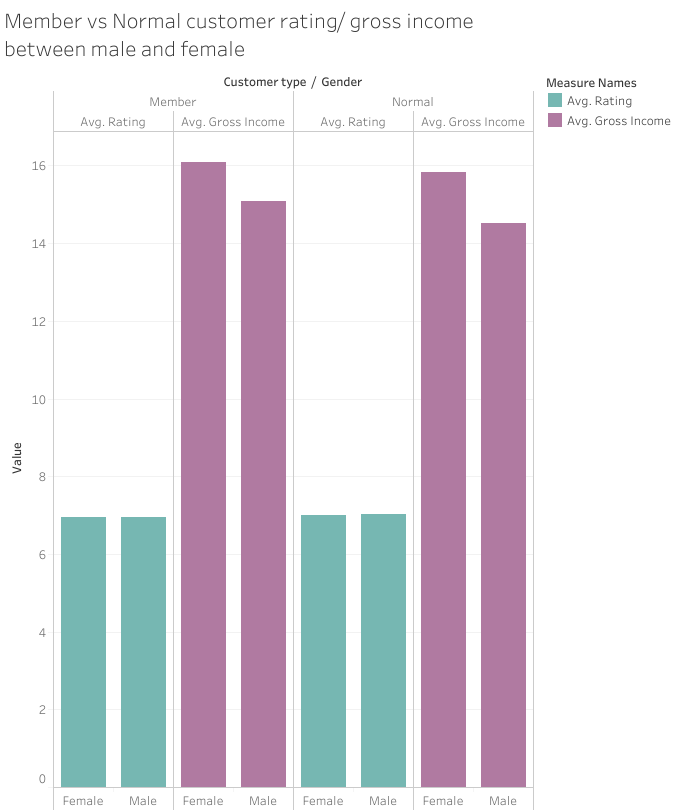
This graph shows the average gross income between males and females between the branches. For both male and female branch C had the highest gross income and now the supermarket knows how much income per male and female spend when they come in.



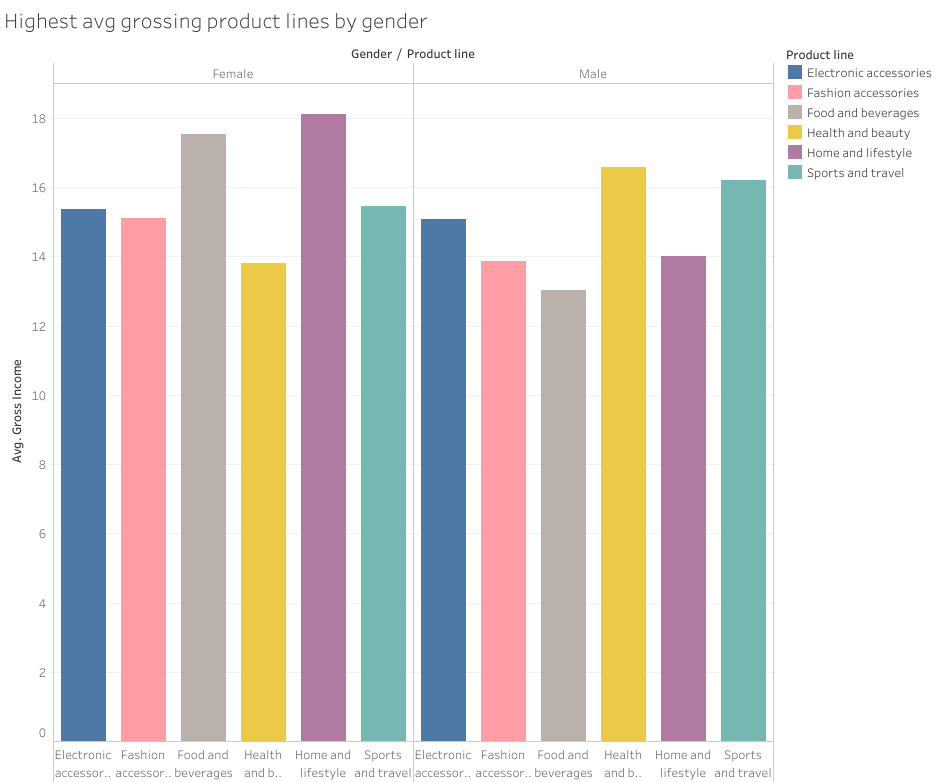
This is the divide between male and female based on payment options. These three methods of payment can help the supermarket understand how their customers prefer to pay for the items they purchase and how they can make sure these options are supported for various situations. Making sure things are simple and convenient will always be a safe move, such as a nearby atm machine, or a personal app to help out E wallets.



This is meant to provide an emphasis on branch results of the loyalty program. Here once C has the greatest average gross income of 16.027, and it also has the highest rating with 7.04. The supermarket can investigate the qualities about branch C that has given it the slight edge on branch A and B.



Now we see the comparison of gender between loyalty members and normal customers. The members have a higher gross income between genders with females at 16.082 and males at 15.095 while the normal females have 15.821 and males at 14.526. However, the normal rating for females (6.99) and males (7.019) was higher than the members rating of females (6.941) and males (6.94). Then what about the supermarket makes the normal customers rate it higher than the loyalty members, to me it seems there may need to be some improvement in the loyalty program.



Product lines that had the highest average gross income can tell the supermarket what the customers like the most. Divided between the genders helps the supermarket narrow down preferences and what people buy the most between selections. We saw earlier that females spend more than the men, and now we see what both their favorite product lines are. The supermarket can now make adjustments in supply and invest more into their top sellers.

**ACT**

Which branch has the best results in the loyalty program?

One of the main things I noticed was that the best branch in the loyalty program was branch C. The C branch had the highest gross income of 16.05 and rating of 7.04 from the customers. Even among the normal customers, Branch C had the highest results with an average gross income og 16.07 and a rating of 7.09.

Does the membership depend on customer rating?

However, in this comparison I concluded that membership didn’t necessarily depend on its rating because the normal guests (7.00) had a higher rating than the members (6.94).

* Does gross income depend on the proportion of customers in the loyalty program? On payment method?

Since the difference wasn’t significant, one could say that the gross isn’t based heavily on the loyalty program because the normal customers are outpacing the members. When it came to the payment methods, females mostly paid with cash at 178, and normal males preferred ewallet with 184. However, neither one of these options had a major impact over the other options, so I can only suggest moving forward with all three options.

* Are there any differences in indicators between men and women?

Some differences between the genders was that there were more females (261) in the loyalty program than the males(240). However, there are more normal males (259) than females (240), and between the branches the members had a higher gross income so the supermarket may need to see how they can get more normal customers to become members.The highest product line for men are health and beauty (16.576) while for females its home and lifestyle (18.105). Females also have a higher gross income average than males in both members (F 16.082) (M 15.094) and normal (F 15.820) (M 14.526).

* Which product category generates the highest income?

Home and lifestyle generates the highest income with a gross income average of 16.030.

This study was meant to give understanding to the parts the supermarket is using. The branches, loyalty program, and rating can tell the story of what the supermarket can do in the future to increase revenue. Seeing the patterns of what males and females purchased along with how they viewed the store with the rating system, helped the supermarket gaze into the mind of the customer to see what they wanted.

Some recommendations I'd make would be to do a survey, and see what customers would like to see in a loyalty program. What incentives do they value the most, and how can this be used to turn normal customers into members.

Seeing that women spend more than men, what needs and wants do women fulfill by purchasing from the supermarket? Which home and lifestyle items do they buy the most of?

Advertising sales more in the home and lifestyle for women and health and beauty sections for men.

Also gathering information on what men would like more out of the store, such as any items that are available at other stores that aren't at the supermarket that they’d want to see more of.