McDonald's market segmentation

Data Mining Project

2023

POLYTECH LYON

Context

With over 40,000 restaurants, 23 billion in revenues and 200,000 employees[1], McDonald's is the market leader in fast food.

Considering the ever more frequent consumption of these establishments (83% of Americans claim to go to a fast-food restaurant at least once a week, and fast-food restaurants are seeing a 2.2% increase in their clientele per year[2]), it is legitimate to study public opinion concerning the brand with the Ronald clown.

So, using our knowledge and the tools offered by data mining, we'll be looking at the opinions of a sample of 1,453 testimonials, trying to extract as much information as possible and make the data in this report speak for itself.

Dataset

The dataset used comes from a study carried out by Austrian professor Friedrich LEISCH, and is available free of charge at the following address:

https://homepage.boku.ac.at/leisch/MSA/datasets/mcdonalds.csv

This dataset was built in 2021, which does not make it obsolete. It contains 15 columns and 1453 rows. Each line corresponds to the testimony of one person.

The first 11 columns give the person's opinions on adjectives attributed to McDonalds, including: [yummy, convenient, spicy, fattening, greasy, fast, cheap, tasty, expensive, healthy, disgusting]. The witness answers yes or no if he thinks the adjective corresponds to the sign.

The 12th column is a score, ranging from -5 to +5.

The last 3 columns contain information about the witness, including age, gender and frequency of restaurant visits (ranging from “Never” to “More than once a week”).

	yummy	convenient	spicy	fattening	greasy	fast	cheap	tasty	expensive	healthy	disgusting	like	Age	VisitFrequency	Gender
0	No	Yes	No	Yes	No	Yes	Yes	No	Yes	No	No	-3	61	Every three months	Female
1	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	No	+2	51	Every three months	Female
2	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	+1	62	Every three months	Female
3	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No	Yes	+4	69	Once a week	Female
4	No	Yes	No	Yes	Yes	Yes	Yes	No	No	Yes	No	+2	49	Once a month	Male

Image 1. Dataset overview “mcdonalds.csv”

[1] « McDonald’s - Faits et chiffres ».
[2] « Fast Food Statistics | July 2023 | The Barbecue Lab ».

Study

As we said in our introduction, our aim is to find out what the public thinks of McDonald's, and to go into more detail about whether there is a certain target audience for the company, what opinions this target audience is expressing, and how these opinions are evolving within the population.

We decided to pose the following problem: How do the consumption trends and habits of McDonald's customers evolve according to age, gender and other defined criteria?

Population

It's normal for any study to begin by knowing the population being surveyed, without which it will be impossible to proceed. We will study the population on the basis of the information we have, i.e. gender and age.

The population of this study is very well distributed on these two criteria, we observe on appendix 1 (see appendices) that there are slightly more women surveyed than men with 54% of women for approximately 46% of men. It's not a very big difference, and it doesn't tell us anything about the consumption habits of each gender, but it's important to note this difference, which could have an impact when comparing M/F opinions.

We can make exactly the same observation about the age of those surveyed, which is fairly well distributed across a population ranging from 18 to 70 (see appendix 2). There is a “peak” of 55-year-olds, which we note for the rest of the study, but which should not have a major impact.

The population surveyed is therefore very representative of a 'lambda' population in terms of age and gender, which means we can continue our study confident that the results will be representative of reality.

Visitor Frequency

We now know the population of our study and we want to answer our problem. The first piece of information we need to ascertain about our customers' consumption habits is how often they visit the restaurant.

The graph in appendix 3 (see appendices) shows us the frequency with which people surveyed frequent the restaurant, according to age and gender. The first thing we notice is a diagonal running from top left to bottom right.

This diagonal is very interesting, as it shows that attendance seems to decrease with age: very few under 30s never go, and very few over 50s go more than once a week. Similarly, the middle-aged population seems to concentrate around one visit every 2 months. This population is, however, more evenly divided between “once a

week” and ‘once a year’ compared to the under-30s and over-55s. Although we have a correlation coefficient of -0.2898 between age and frequentation (due to the fact that there are far fewer values on the edges of the graph than in the center in general), we can say that age has an impact on fast-food chain frequentation.

As far as gender is concerned, the distribution is very even (see Appendix 3), showing that gender has no influence on chain restaurant patronage. Even if this seems to be the case for certain categories, we can't say that women have a higher patronage than men, since our control set is 54% female.

We have seen that age has an impact on attendance, but we can also show that attendance has an influence on other parameters such as score. This influence can be seen more clearly here (see appendix 4), with a diagonal running from the bottom left to the top right of the graph. The density is around this diagonal, with very few outliers. It's worth noting, however, that some people who never go give a score of 5, which is not the case for the reverse.

Unsurprisingly, it's obvious that the more people surveyed frequent McDonald's, the more they like the food offered by Mr. Yellow.

Influences on scores

So, does age have an influence on attendance and attendance has an influence on grade? We'll see that it does, but not exactly as expected.
In fact, we can see from the graph in appendix 5 (see appendices) that we have a sort of division of the people surveyed from grade -1 upwards. Whereas in previous graphs we could observe something progressive, here we have a clear horizontal break between two parts of the graph (and a third part which is line
-5). This separation shows us that, overall, the people surveyed give a positive or very negative rating to the fast-food chain.
Nevertheless, we still have a diagonal line running from the top left towards the like=0 age=[52,57] point, then descending at an even greater angle towards the bottom right of the graph. What we can conclude from this is that overall age does have an influence on the rating: as age increases, the rating gradually decreases until the age of 55, when opinions change, and the rating drops drastically. It's important to note that this analysis is true for a 2021 study, but that this information is bound to change over time. This break occurs at the age of 55, as the brand has been known for some fifty years.

Links between criteria and visit frequency

So far, we've answered part of our question by looking at consumer opinions according to age and gender, but we can now go further and understand why the rating drops with age, and what criteria are behind this drop. So we're going to look at the most closely related criteria to try and understand.

The graph in appendix 6 (see appendices) shows the extent to which the criteria are interrelated, and the extent to which the frequency of visits is linked to the rating criteria. The first thing we can observe is a sort of separation of the criteria into 3 categories:

- Criteria almost never cited: spicy, healthy, disgusting, expensive

- Criteria almost always cited: fast, fattening, convenient, tasty

- moderately cited criteria: yummy, cheap, greasy

While everyone agrees on the first three and, similarly, on the last 3 (the highest percentages always coming from very large numbers of customers), we note that among these categories there are points of divergence. For example, “healthy” is cited on average, but only by customers who frequent the restaurant the most. Or “disgusting”, which is cited by 74% of those who never go to the restaurant.Cela nous donne un indice de quels sont les critères les plus liés entre eux et effectivement, lorsque nous regardons en détails nous pouvons calculer quels sont les critères les plus connectés et les séparer en clusters.

If we look for the criteria closest to each other, we see 4 pairs of criteria: fast / fattening, tasty / yummy, spicy / healthy and greasy / cheap. Where some pairings don't really give us any information (spicy / healthy, for example), others are much more telling, such as the fast / fattening / convenient group, which illustrates McDonald's role as a fast food company. In this way, we can see the policy pursued by McDonald's, which aims to be simple and fast, but plays its appetizing trump card through fattening.

Another piece of information we can draw from these comparisons is the relevance of the data. This can easily be seen with the criteria “tasty” and “yummy”. We can see that these two criteria are very closely linked, as they both represent the notion of taste, and although tasty and yummy don't have exactly the same definition, it's important to ask whether these criteria aren't duplicating each other and whether it wouldn't be relevant to use just one. You can't see it in this graph, but the same applies to cheap and expensive, since they are defined in opposite ways.

As far as frequentation is concerned, “once a month” and “once a week” are the most closely related in terms of survey results (not surprisingly, with “more than once a week” in 3rd place, being the closest frequentations in terms of time intervals).

Finally, we can see that those who never frequent McDonald's are the ones most likely to highlight the fast-food chain's clearly negative criteria, and conversely for those who frequent the restaurant most who tend to highlight its positive points.

We can, however, question the relevance of the opinions of people who never frequent McDonald's, wondering whether they might skew the study. Indeed, if they never go there, is their opinion representative of reality?

Impact of age and gender on opinion

We have seen that :

- Age has an impact on attendance

- Attendance has an impact on score

- Age has an impact on grade

- Criteria are linked to attendance

So we ask ourselves the question: “Does age influence ratings? Indeed, if age influences frequentation and frequentation is linked to reviews, then there should be a link between age and reviews.

Appendix 7 (see appendices) shows us the percentage of people in each age category who did or did not set a criterion. In most cases, age doesn't really influence whether a criterion is present or not. This is particularly true of the so-called “objective” criteria (fast, fattening, convenient, healthy).

On the other hand, this graph highlights the reasons why each age group consumes this type of food. For younger people, “yummy”, “tasty” and “greasy” are clearly adjectives that focus on taste: we go because we like it.

For the older generation (especially those over 50), it's “cheap”, “fast” and “convenient” that come to the fore: we go there because it's convenient.

We can clearly see the change in the reasons for going depending on age, which is totally in line with the analyses made in the graph “McDonald's attendance by gender and age” (see appendix 3). When you're older, eating out isn't seen as an everyday option, but rather as a way to tide you over when you're short of time, money or whatever, whereas for younger people, fast food is a habit.

We also tried to see if gender had an impact on opinion (see appendix 8), but once again, we found no difference between men and women.

Further information

If we wanted to delve a little deeper, we could look at common attributes and interests between users and define profiles for customer preferences (see appendix 9).

However, we only have an analytical task ahead of us, so we won't be doing this, but it's important to note that it is possible.

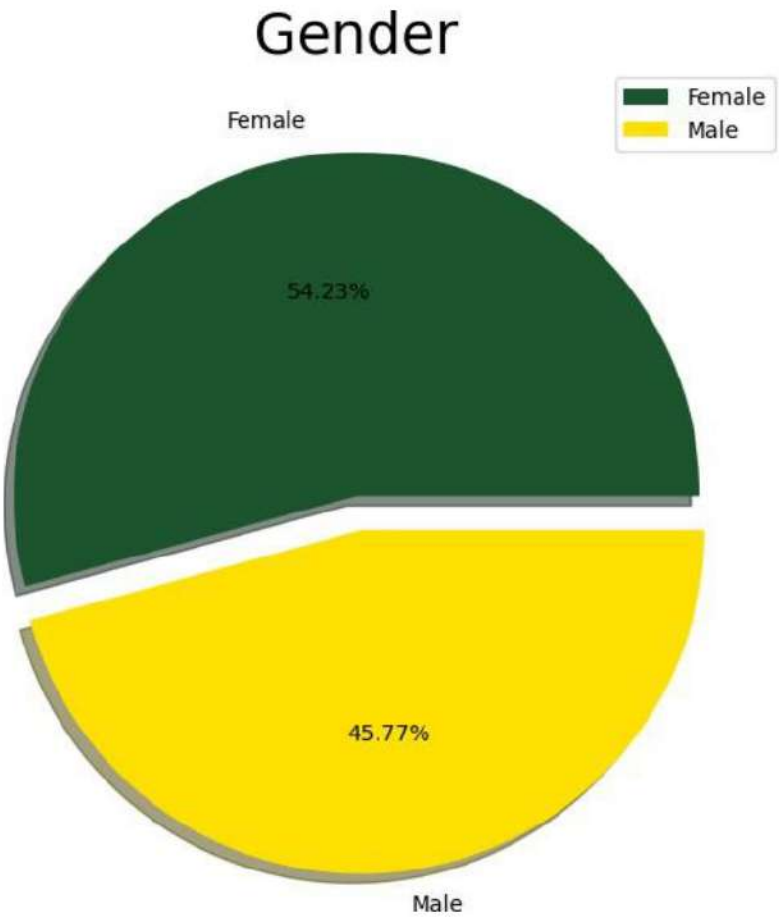
Conclusion

Our study focuses on the following question: How do the consumption trends and habits of McDonald's customers evolve according to age, gender and other defined criteria?
After this study, we can say that age does have an influence on the consumption habits of McDonald's fast-food chain. Although everyone is affected by the Ronald McDonald's phenomenon, younger people are more likely to consume its products than older people, for a variety of reasons, including taste, convenience and price.

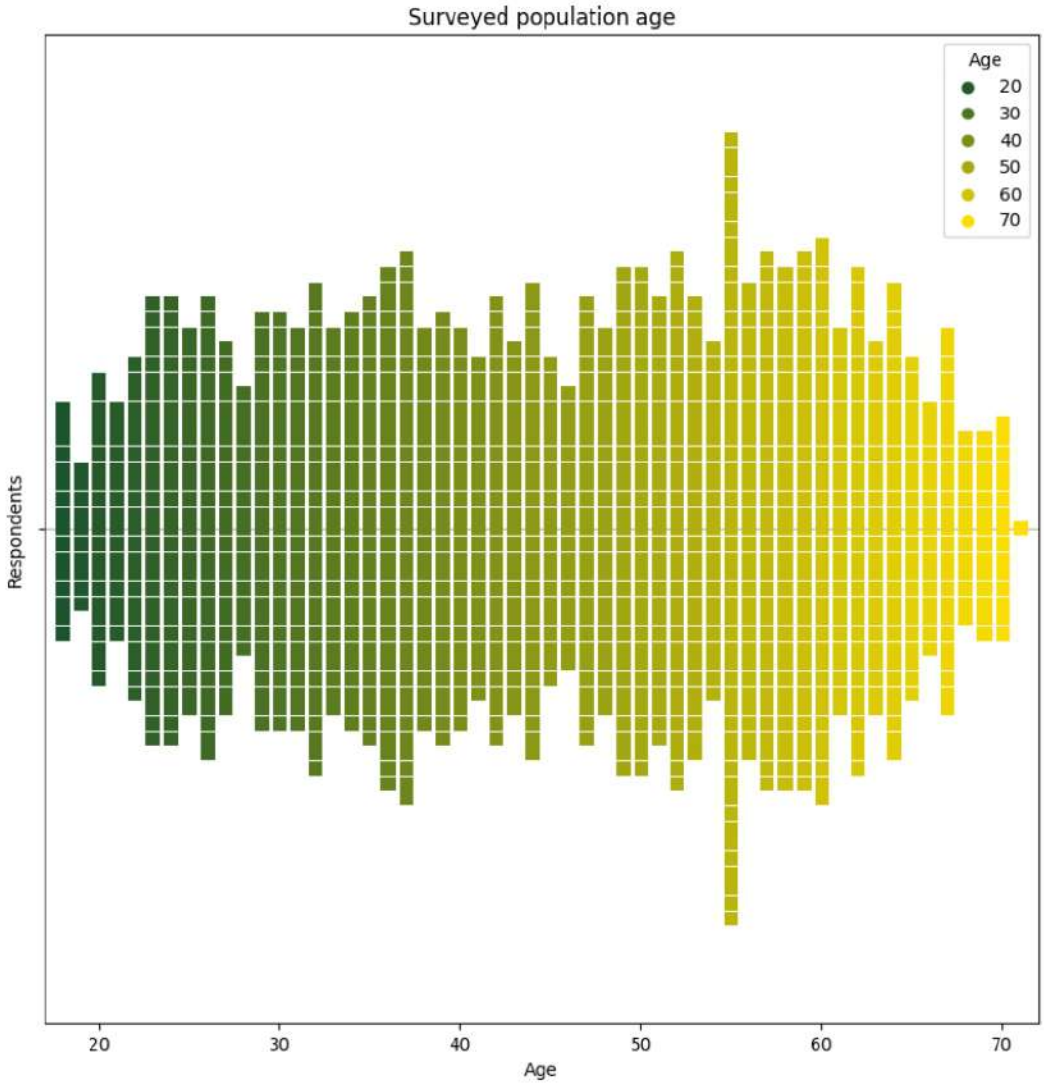
Nevertheless, we found no evidence that gender has any significant impact on McDonald's consumption.

Once again, we would like to point out that this study is based on a 2021 study carried out in Austria. The conclusions drawn from this study can therefore only be taken into account in this context and cannot be transposed to the situation in France in 2023.

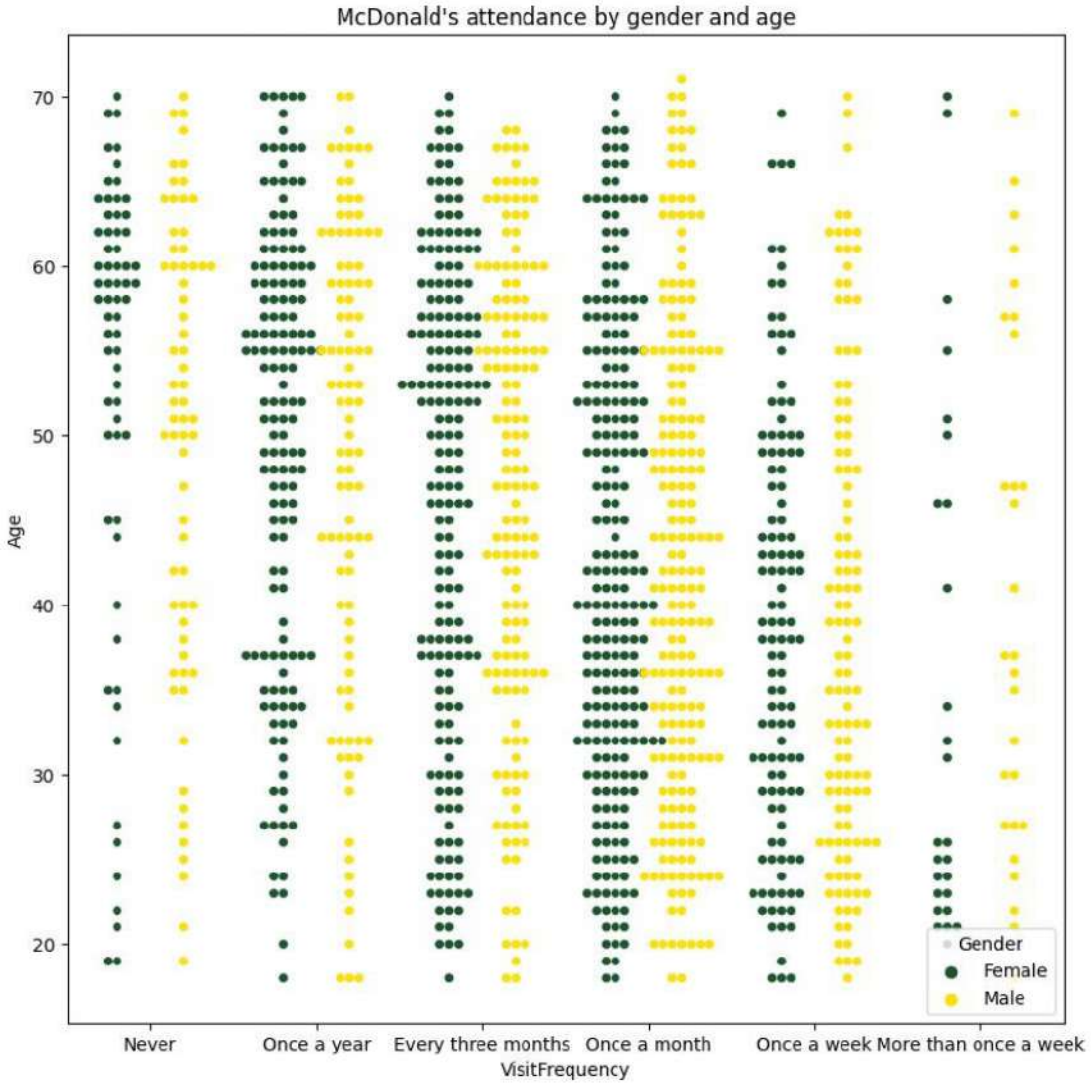
Appendices



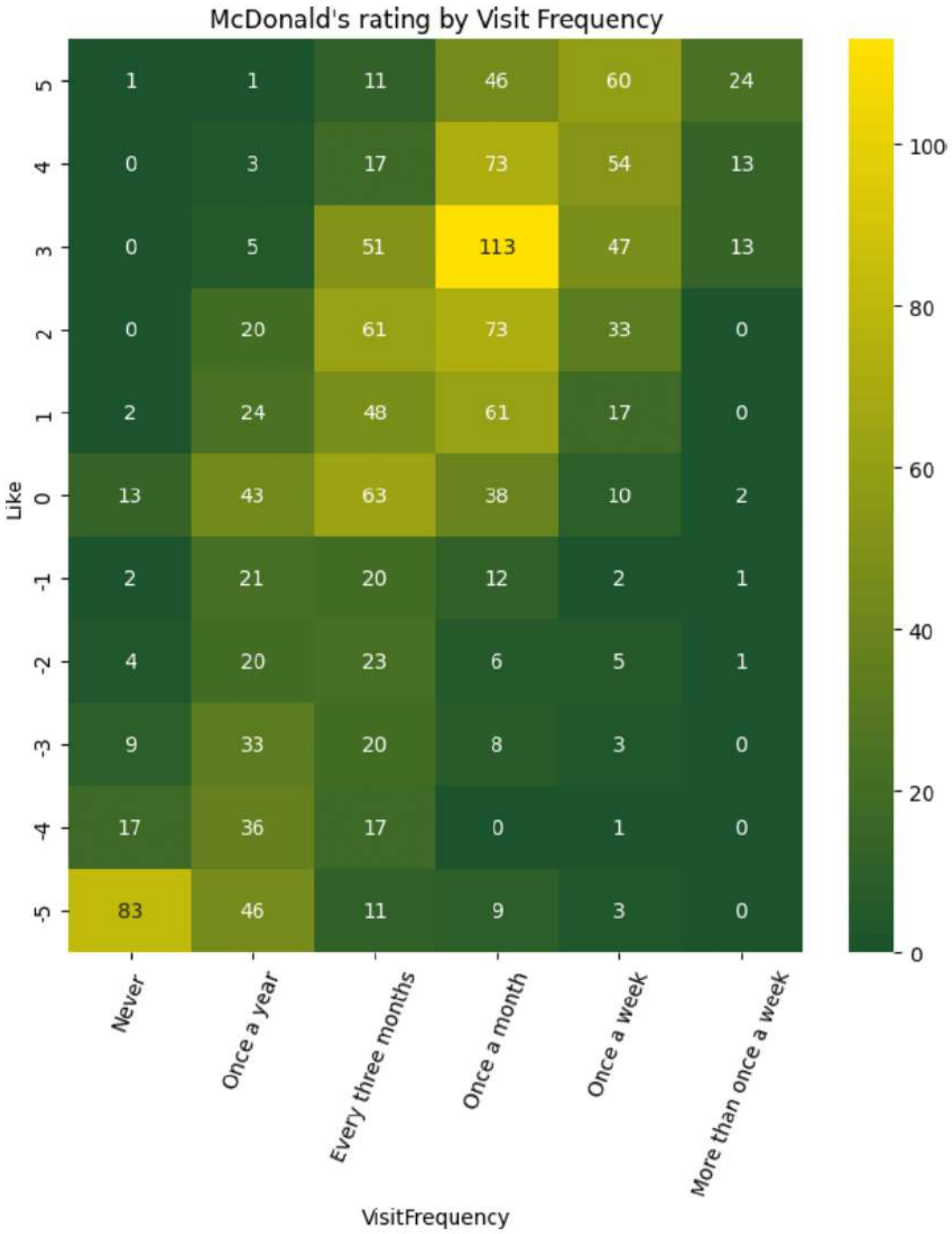
Appendix 1: Percentage of men and women in the survey population



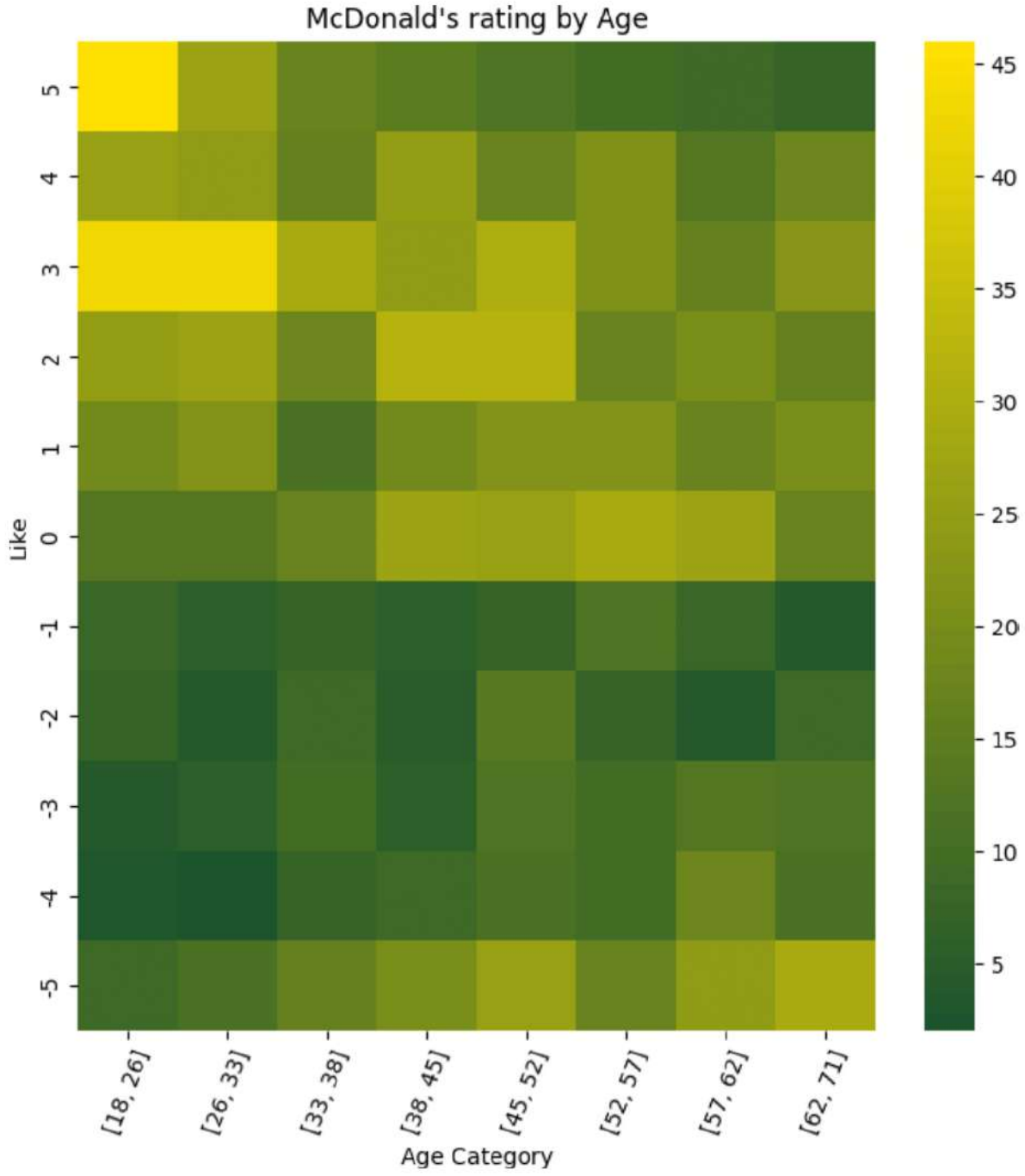
Appendix 2: Age distribution of survey population



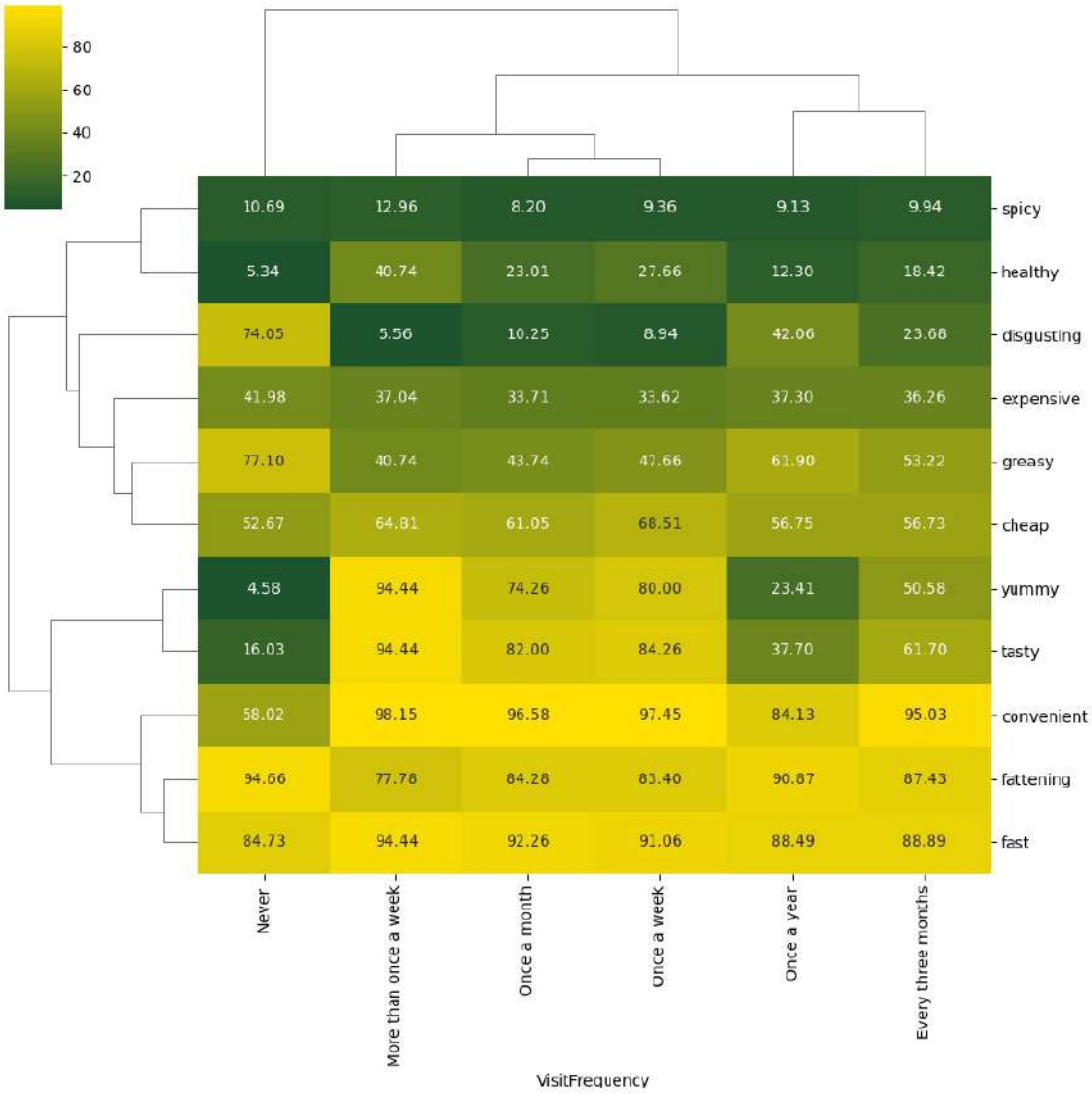
Appendix 3: Attendance by age and gender



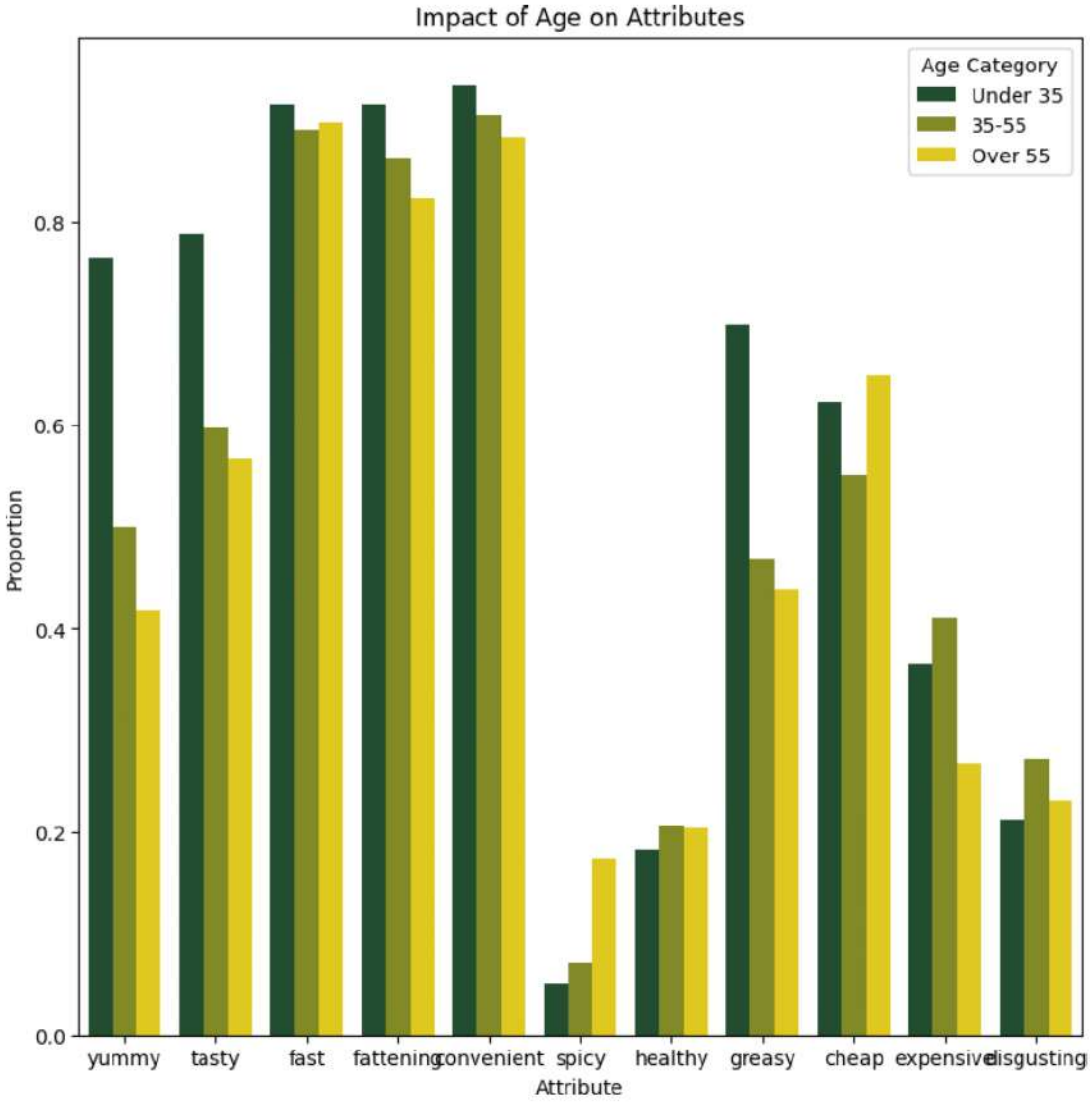
Appendix 4: Rating according to visit frequency



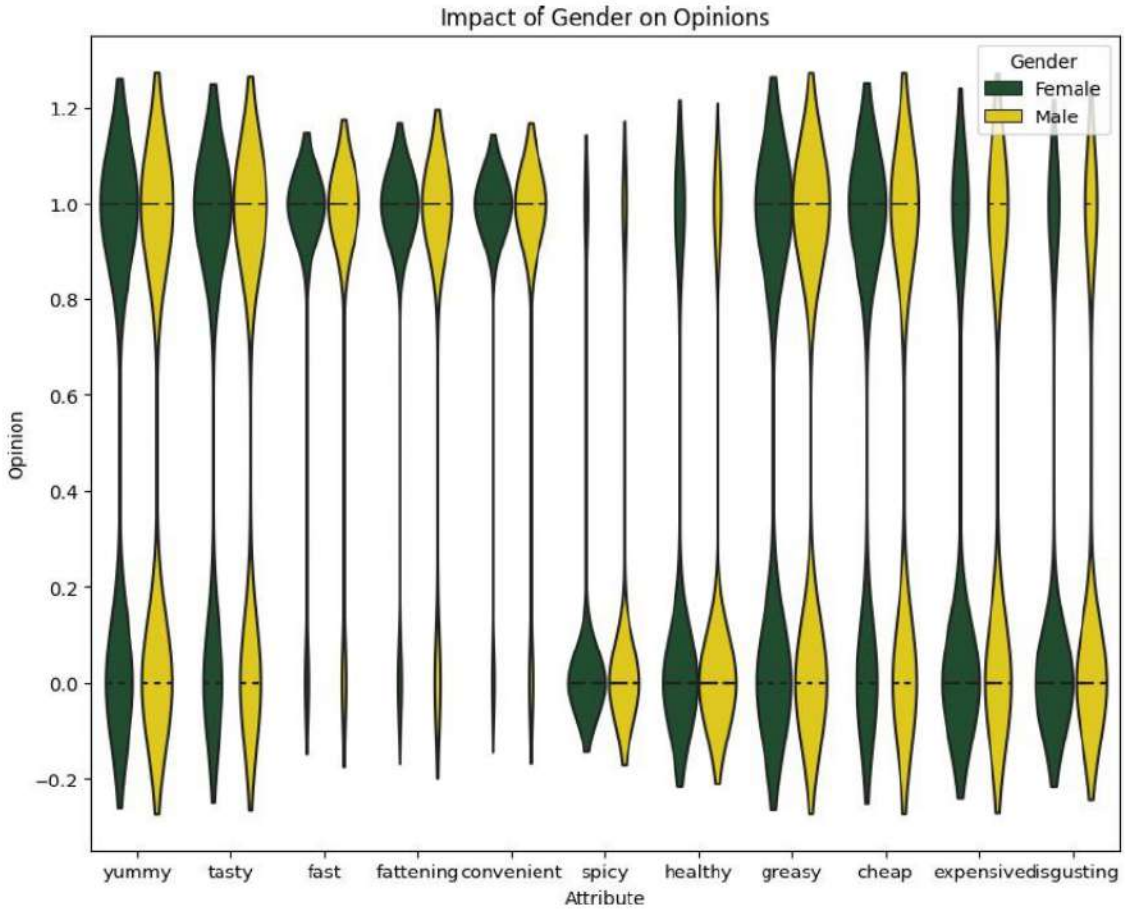
Appendix 5: Age rating



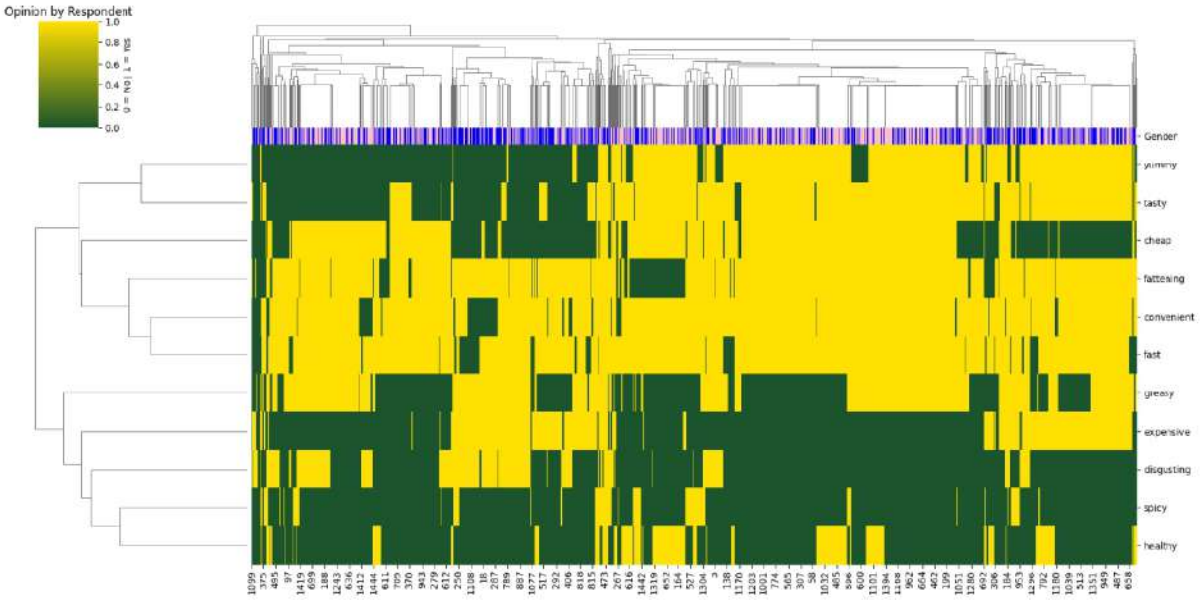
Appendix 6: Opinion according to frequentation and relationship between opinions



Appendix 7: Impact of age on opinion



Appendix 8: Impact of gender on opinions



Appendix 9: Opinion by profile