**Team- Abhik**

Project 2 Study Task

Summary Report step-4 and step-5 (Market Segmentation Analysis - McDonald case study)

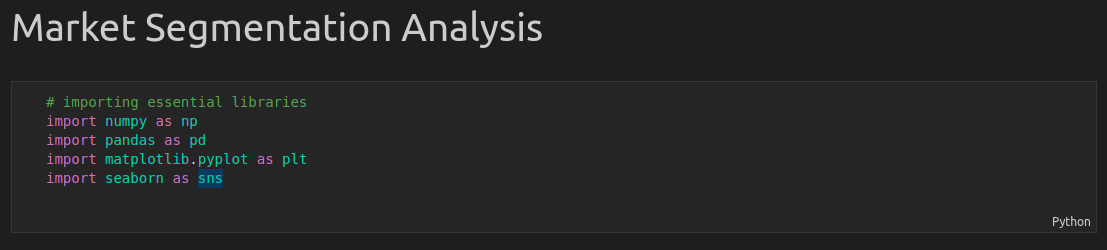
*By Yash Dawande*

Summary Report:

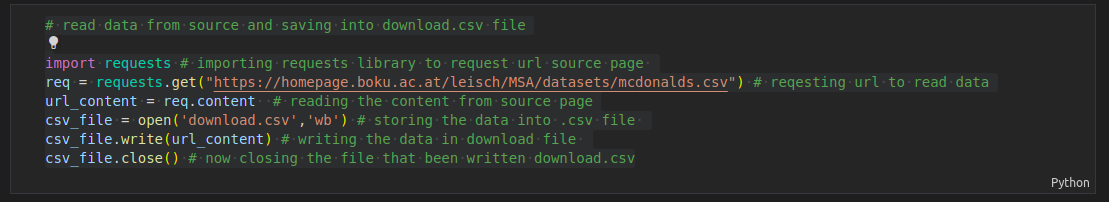
Step 4: Exploring data

Frist we explore the key characteristics of the data set by loading the data set and inspecting basic features such as the variable names, sample size, and the first three rows of the data:

Import data:



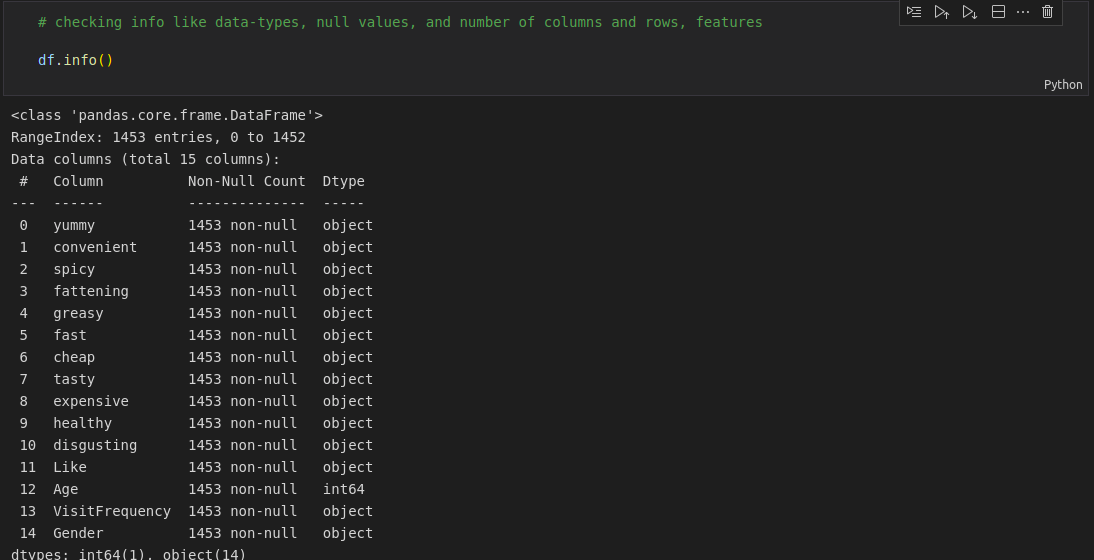
Download data from source:



Reading dataset:

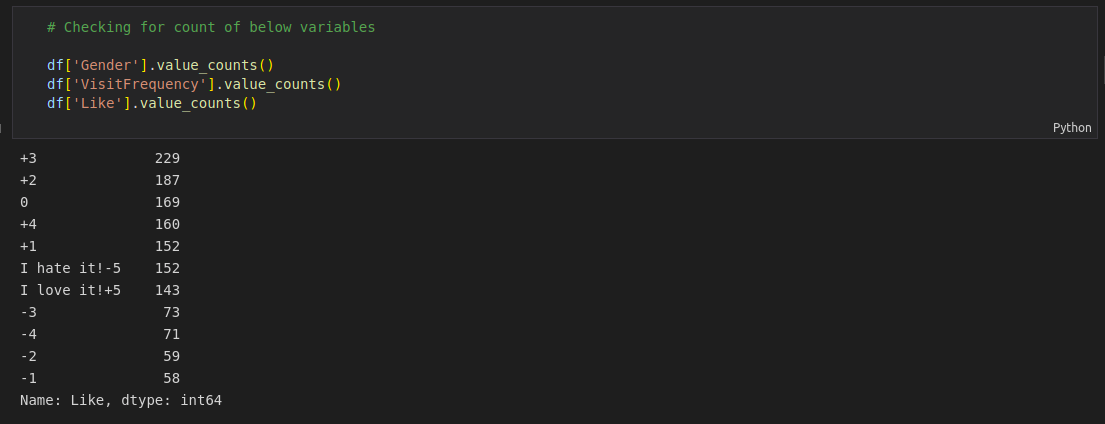


Check information about data:

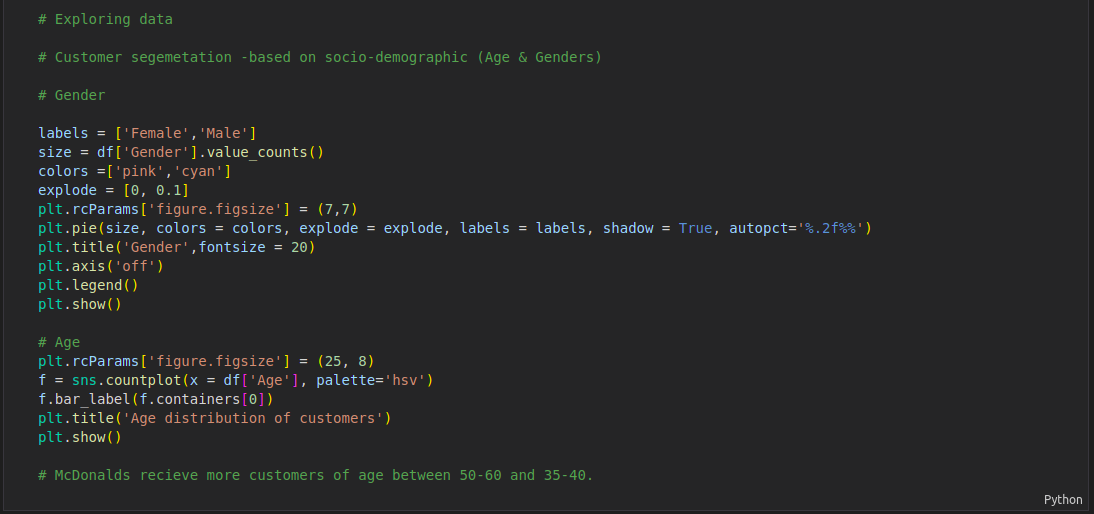


Count of variables like “ Gender ”, “ VisitFrequency “, “ Like “ :

Highest no. of likes is 229 where lowest no. of likes 58 can be seen.

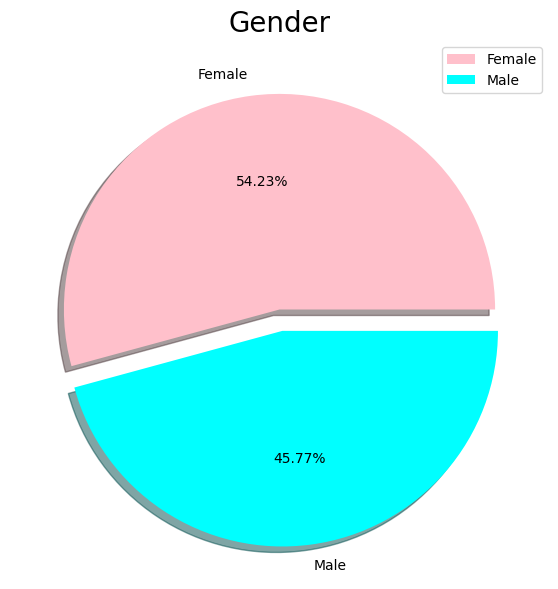


Exploring Data:

As you can see above, we are exploring data of customer segmentation based on socio-demographic (Age & Gender)

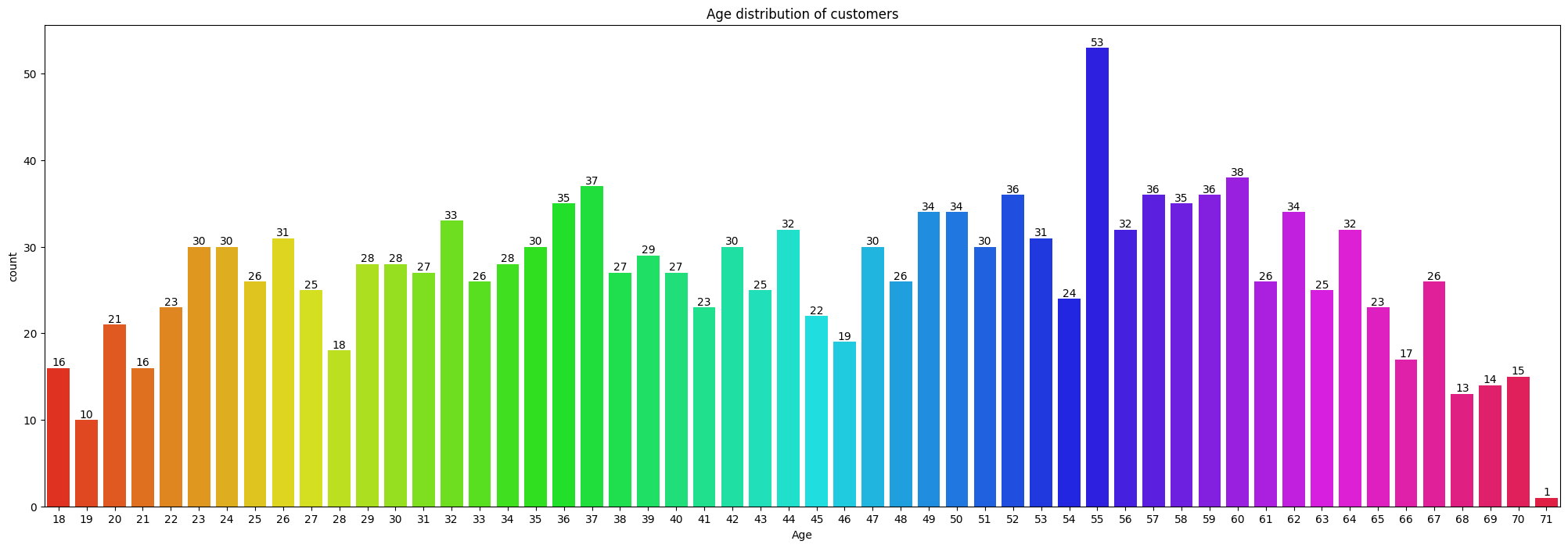
So the result is like this

Gender segmentation



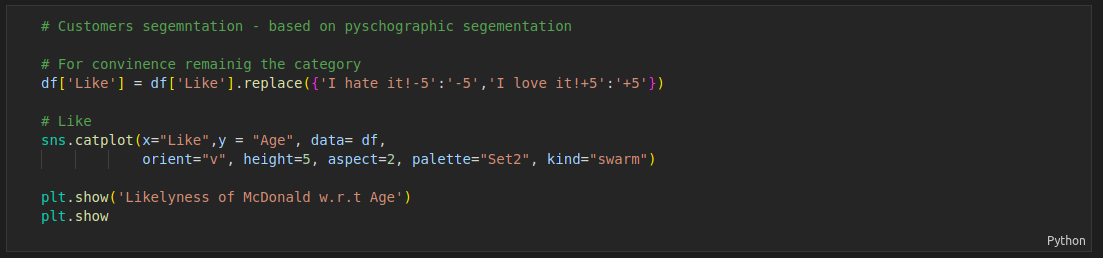
As can be seen in the above diagram that segmentation of male Is 54.23% and female is 45.77%.

Age segmentation

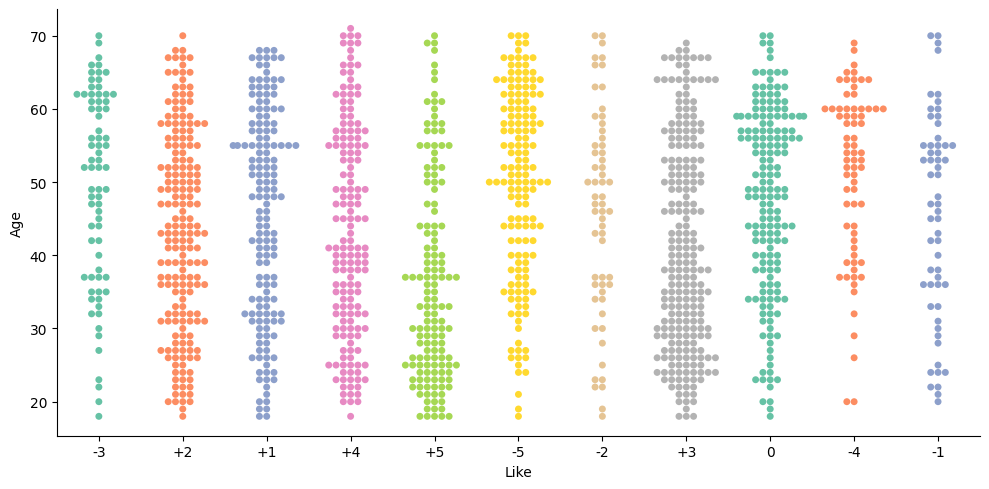


Where in this diagram we can see that the highest number of counts is 53 of age 55 and the lowest count is 2 of age 71.

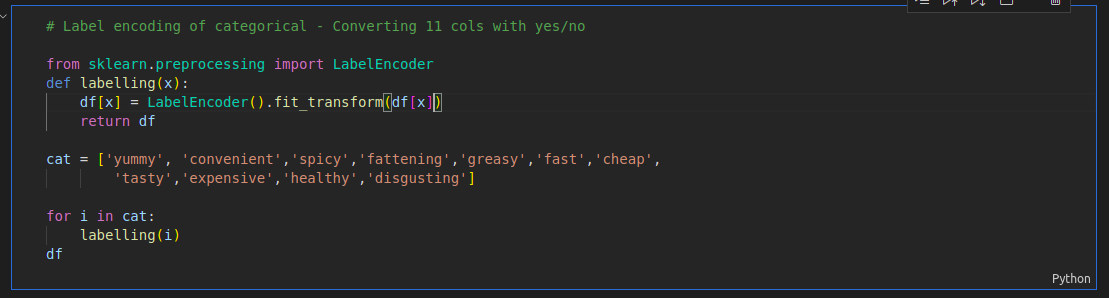
Customers segmentation based on psychographic segmentation



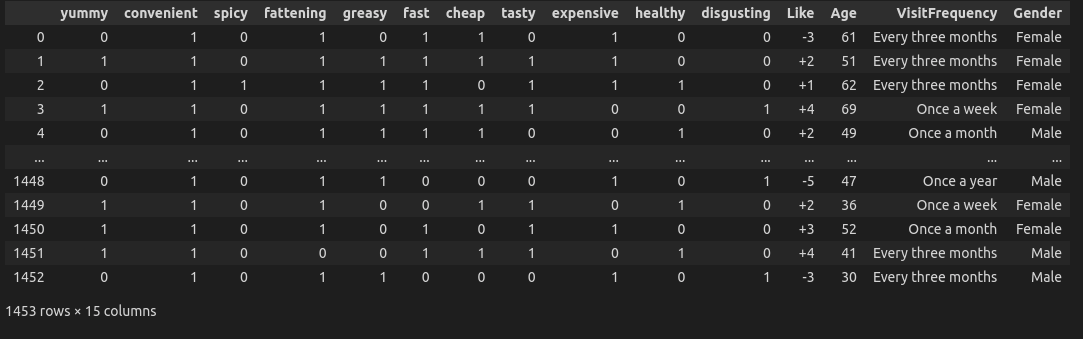
Result of the code

So here in the above code we replace ”I hate it! -5” with ” -5 ” and ” I love it! +5” with ” +5”

This quick glance at the data shows that the segmentation variables (perception of McDonald’s) are verbal, not numeric. This means that they are coded using the words YES and NO. This is not a suitable format for segment extraction. We need numbers, not words. To get numbers, we store the segmentation variables in a separate matrix and convert them from verbal YES / NO to numeric binary. First, we extract the first eleven columns from the data set because these columns contain the segmentation variables and convert the data to a matrix. Then we identify all YES entries in the matrix. This results in a logical matrix with entries TRUE and FALSE. Adding 0 to the logical matrix converts TRUE to 1, and FALSE to 0. We check that we transformed the data correctly by inspecting the average value of each transformed segementation variable.

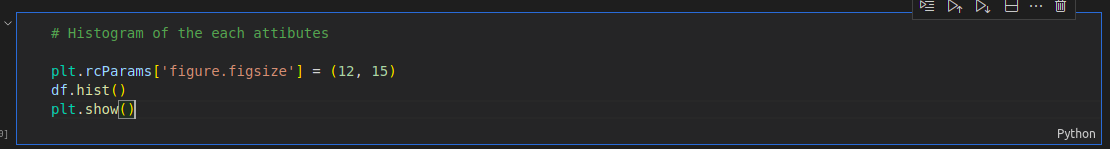


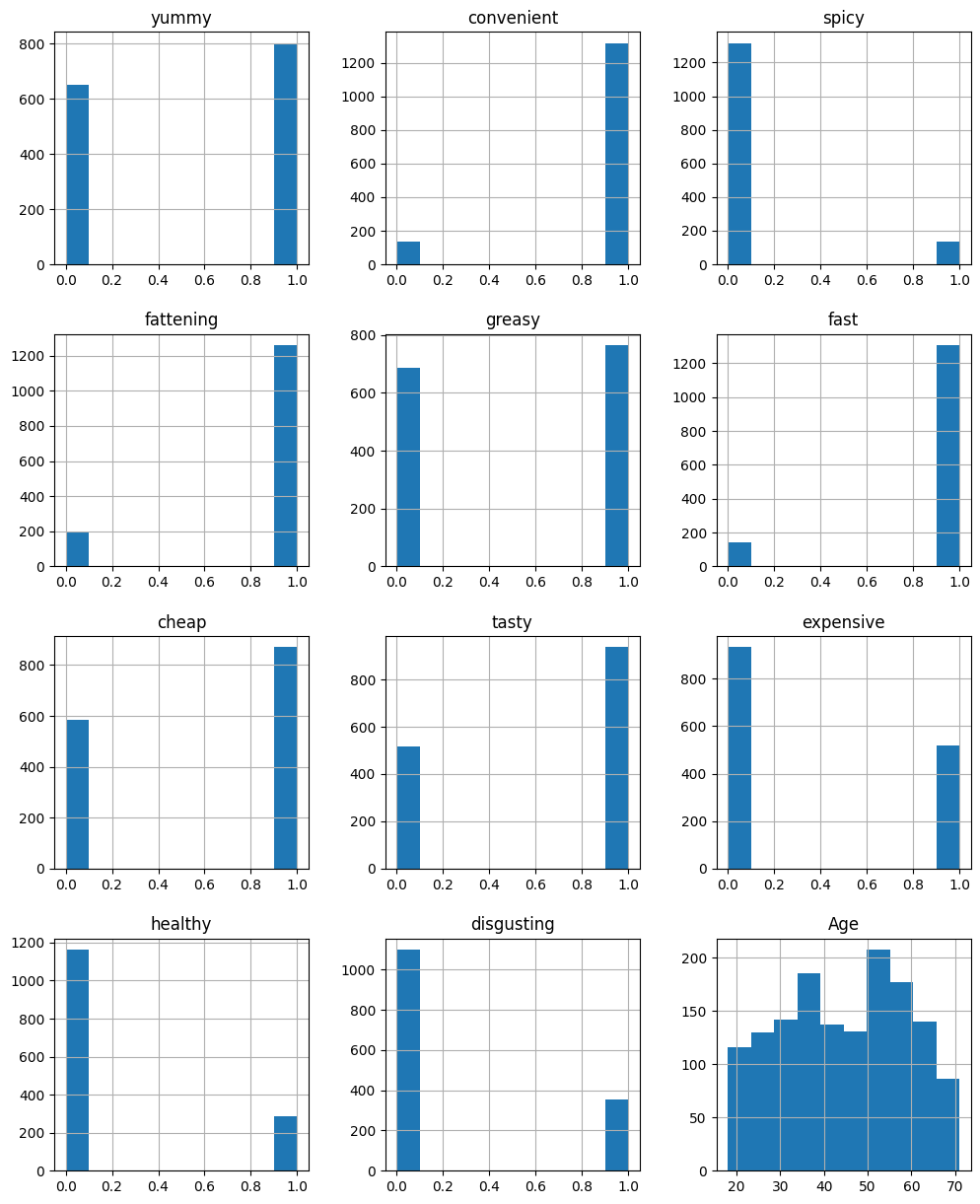
This is what the result took like



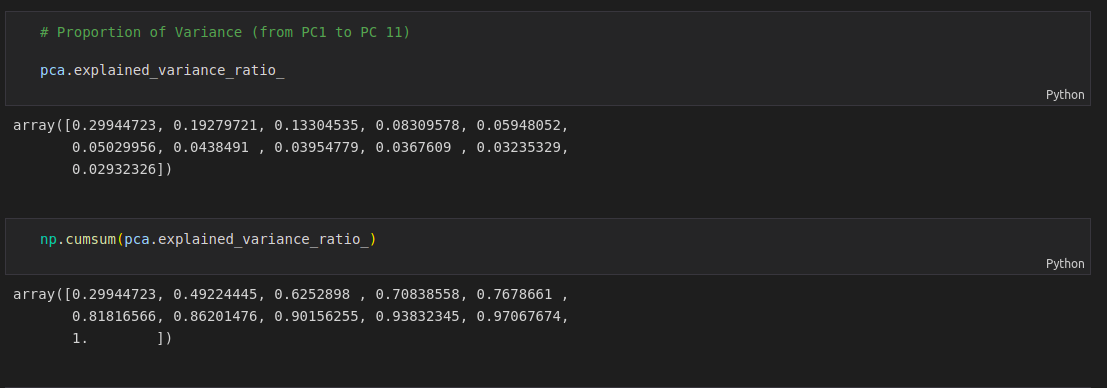
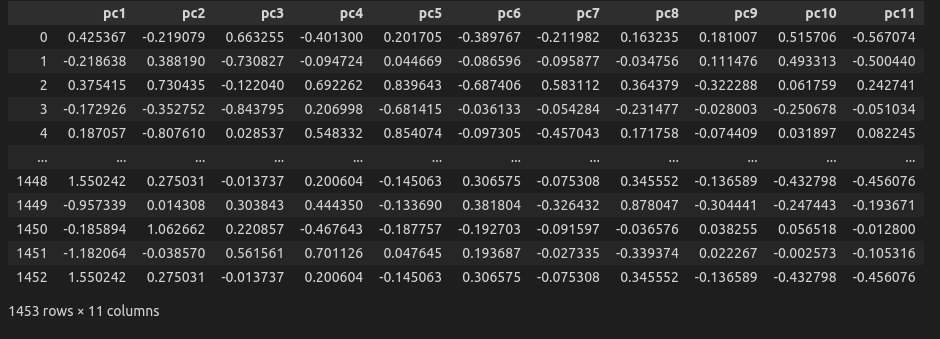
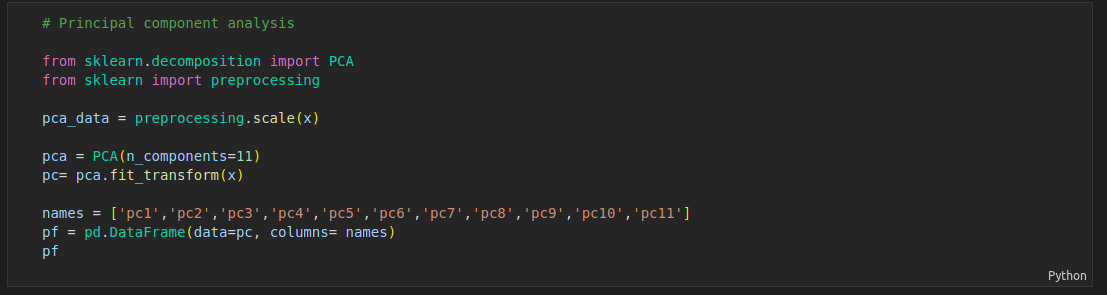
As can be seen that yes/no converted into 1 and 0.

Now taking look at histogram

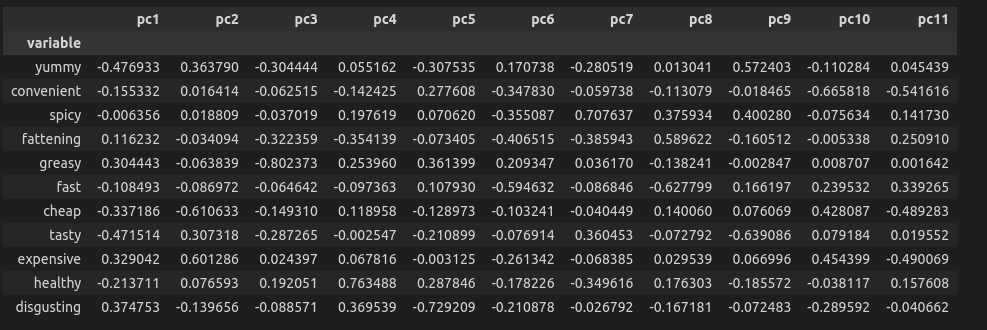
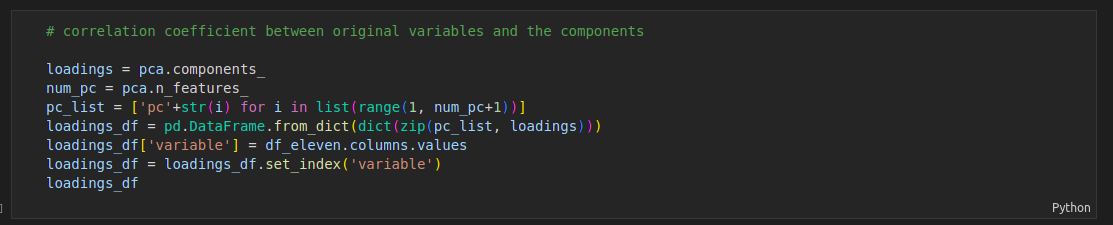


In the comparison of each graph between one and zeros values are high and low but at the Age graph is shows highest age is between 50 to 60 range.

Another way of exploring data initially is to compute a principal component analysis and create a perceptual map. A perceptual map offers initial insights into how attributes are rated by respondents and, importantly, which attributes tend to be rated in the same way. Principal components analysis is not computed to reduce the number of variables. This approach – also referred to as factor-cluster analysis – is inferior to clustering raw data in most instances (Dolnicar and Grün 2008). Here, we calculate principal components because we use the resulting components to rotate and project the data for the perceptual map. We use unstandardized data because our segmentation variables are all binary.

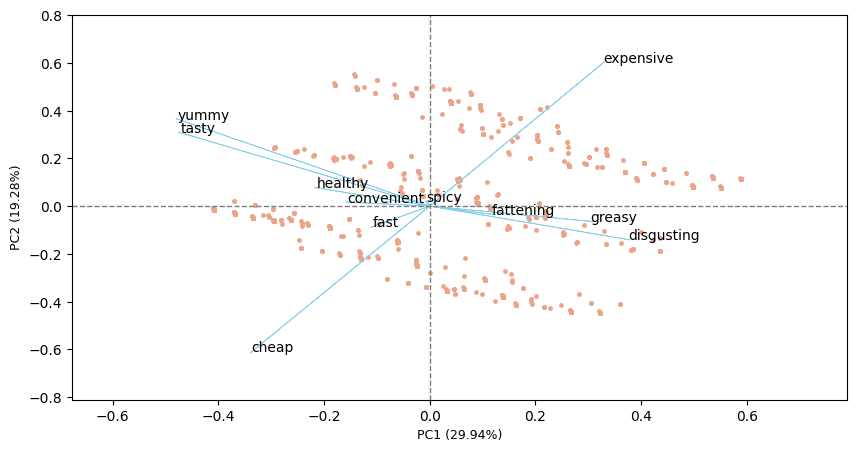


Results from principal components analysis indicate that the first two components capture about 50% of the information contained in the segmentation variables. The following command returns the factor loadings:



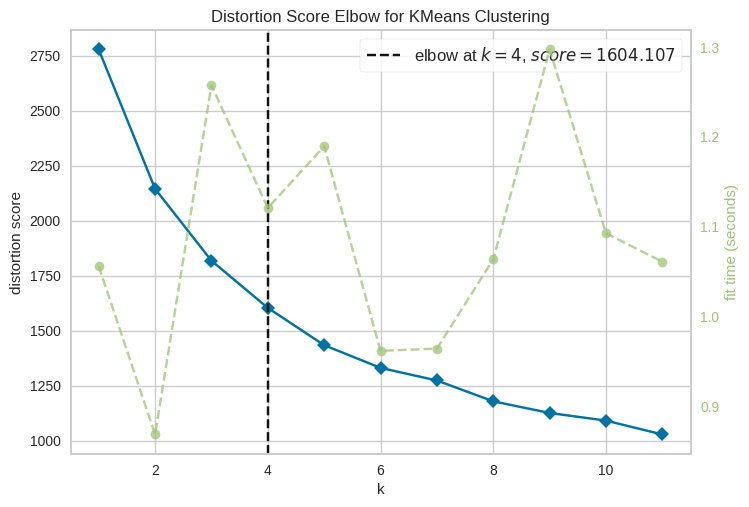
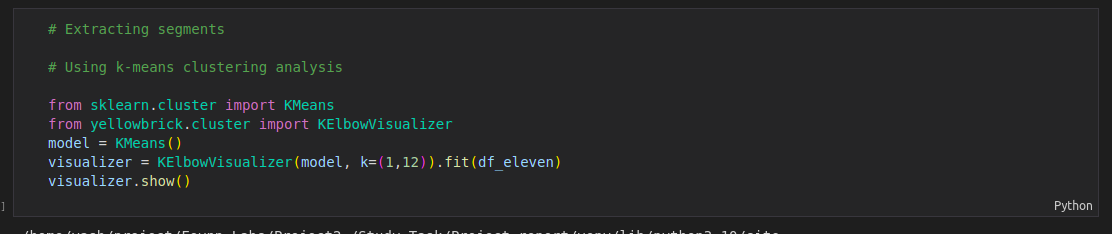
The loadings indicate how the original variables are combined to form principal components. Loadings guide the interpretation of principal components. In our example, the two segmentation variables with the highest loadings (in absolute terms) for principal component 2 are CHEAP and EXPENSIVE, indicating that this principal component captures the price dimension. We project the data into the principal component space with predict

The attributes CHEAP and EXPENSIVE play a key role in the evaluation of McDonald’s, and these two attributes are assessed quite independently of the others. The remaining attributes align with what can be interpreted as positive versus negative perceptions: FATTENING, DISGUSTING and GREASY point in the same direction in the perceptual chart, indicating that respondents who view McDonald’s as FATTENING, DISGUSTING are also, likely to view it as GREASY. In the opposite direction are the positive attributes FAST, CONVENIENT, HEALTHY, as well as TASTY and YUMMY. The observations along the EXPENSIVE versus CHEAP axis cluster around three values: a group of consumers at the top around the arrow pointing to CHEAP, a group of respondents at the bottom around the arrow pointing to EXPENSIVE, and a group of respondents in the middle. These initial exploratory insights represent valuable information for segment extraction. Results indicate that some attributes are strongly related to one another, and that the price dimension may be critical in differentiating between groups of consumers.



Step 5: Extracting Segments

**Using k-Means:** We will use standard k-means analysis.

If we calculate a range of solutions, we can compare them and choose the one which extracts segments containing similar consumers which are distinctly different from members of other segments.

the sum of distances within market segments drops slowly as the number of market segments increases. We expect the values to decrease because more market segments automatically mean that the segments are smaller and, as a consequence, that segment members are more similar to one another. But the much anticipated point where the sum of distances drops dramatically is not visible. This scree plot does not provide useful guidance on the number of market segments to extract.

