Market Segmentation Analysis for McDonalds dataset

Subhadip Chakraborty

Summary

The market segmentation analysis aims to identify distinct market segments within the fast food industry to facilitate targeted marketing strategies. This report presents the findings of the analysis conducted for a fast food company using Python. The analysis involves 10 key steps, including data collection and preprocessing, exploratory data analysis, clustering, profiling segments, describing segments, selecting target segments, customizing the marketing mix, and evaluation and monitoring.

Introduction (step-1)

Background and Objectives: This section provides an overview of the fast food industry and the specific objectives of the market segmentation analysis. It highlights the importance of understanding consumer preferences and behaviors to develop effective marketing strategies.

Methodology Overview: The methodology section describes the approach used in the analysis, including data collection, preprocessing, and the application of clustering techniques. It also outlines the steps involved in profiling and describing market segments, selecting target segments, and customizing the marketing mix.

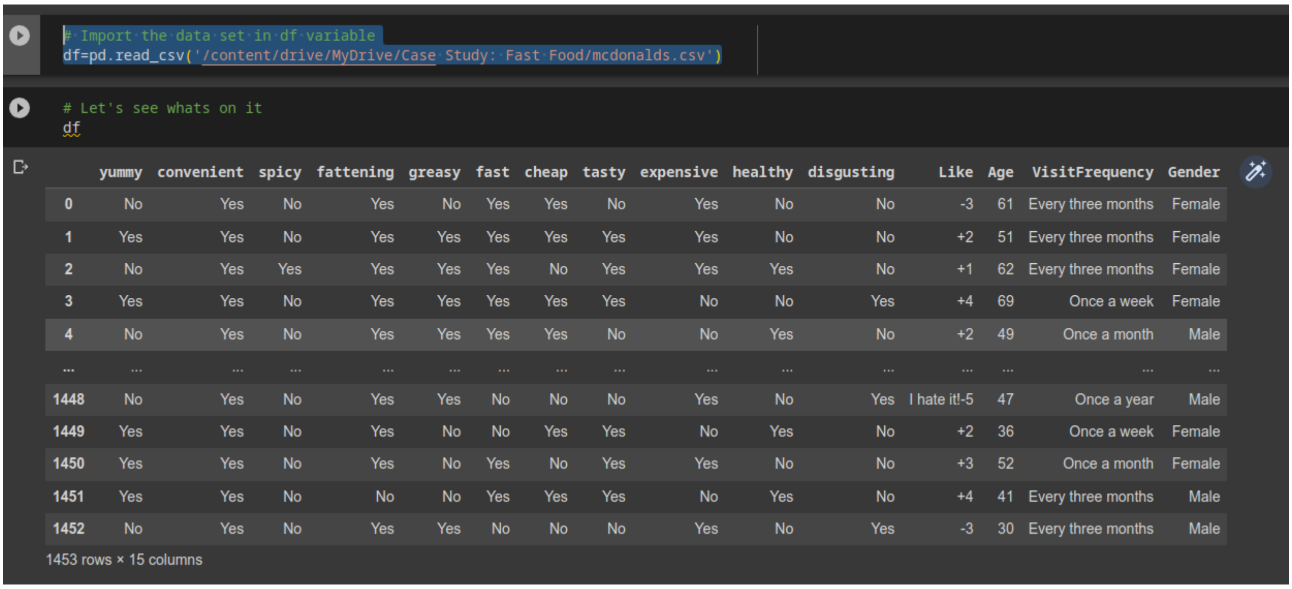
Data Collection and Preprocessing (step-2)

Data Sources and Variables: This section explains the sources of data used in the analysis, such as customer surveys or transactional data. It lists the variables considered for segmentation, including demographic information, preferences, and purchasing behavior.

Data Cleaning and Transformation: Here, the data cleaning process is detailed, including handling missing values, removing duplicates, and transforming variables as needed. Any necessary data transformations, such as scaling or normalization, are performed to ensure compatibility with the clustering algorithms.

Exploratory Data Analysis (step-3)

Data Visualization and Insights: This section explores the data through visualizations and descriptive statistics. It includes charts, histograms, and summary statistics to gain insights into consumer behavior, preferences, and trends within the fast food industry.



```
# Import the data set in df variable
df=pd.read_csv('/content/drive/MyDrive/Case Study: Fast Food/mcdonalds.csv')
```

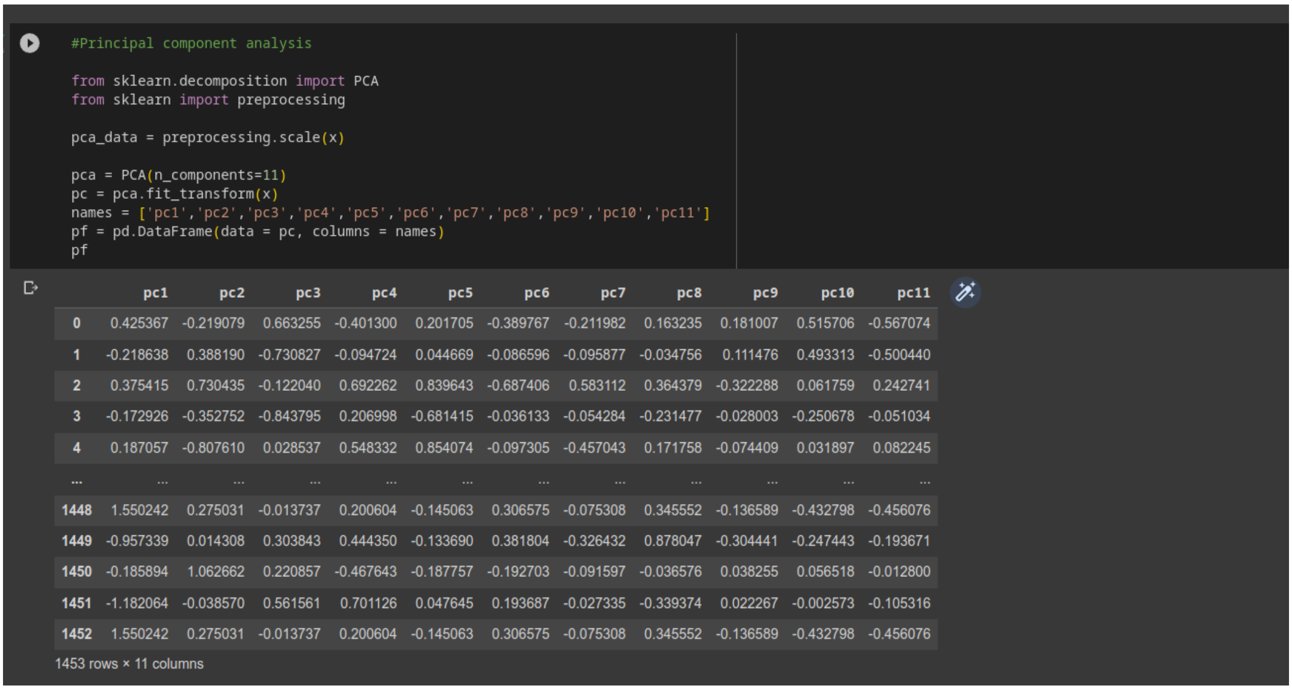
```
# Let's see whats on it
df
```

```
      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
...     ...        ...   ...       ...    ...  ...   ...   ...       ...     ...        ...        ... ...                ...    ...
1448     No        Yes    No       Yes    Yes   No    No    No       Yes      No        Yes I hate it!-5  47        Once a year   Male
1449    Yes        Yes    No       Yes     No   No   Yes   Yes        No     Yes         No         +2  36        Once a week Female
1450    Yes        Yes    No       Yes     No  Yes    No   Yes       Yes      No         No         +3  52       Once a month Female
1451    Yes        Yes    No        No     No  Yes   Yes   Yes        No     Yes         No         +4  41 Every three months   Male
1452     No        Yes    No       Yes    Yes   No    No    No       Yes      No        Yes         -3  30 Every three months   Male
1453 rows × 15 columns
```

Data Transformation and Standardization(step-4)

Scaling and Normalization of Variables: In this step, numerical variables are scaled and normalized to ensure they have a similar range and distribution. This process is crucial for clustering algorithms that rely on distance-based calculations.

Code Snippet: Scaling and Normalization in Python:



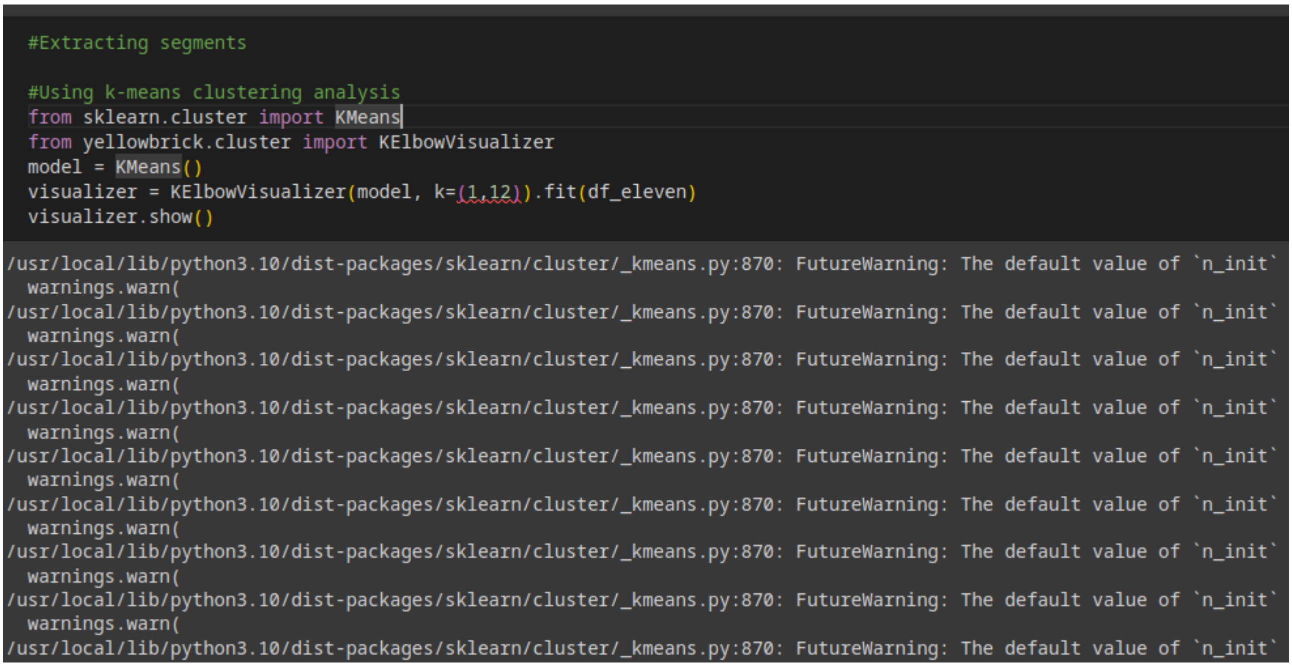
```
#Principal component analysis

from sklearn.decomposition import PCA
from sklearn import preprocessing

pca_data = preprocessing.scale(x)

pca = PCA(n_components=11)
pc = pca.fit_transform(x)
names = ['pc1','pc2','pc3','pc4','pc5','pc6','pc7','pc8','pc9','pc10','pc11']
pf = pd.DataFrame(data = pc, columns = names)
pf
```

```
           pc1       pc2       pc3       pc4       pc5       pc6       pc7       pc8       pc9      pc10      pc11
0     0.425367 -0.219079  0.663255 -0.401300  0.201705 -0.389767 -0.211982  0.163235  0.181007  0.515706 -0.567074
1    -0.218638  0.388190 -0.730827 -0.094724  0.044669 -0.086596 -0.095877 -0.034756  0.111476  0.493313 -0.500440
2     0.375415  0.730435 -0.122040  0.692262  0.839643 -0.687406  0.583112  0.364379 -0.322288  0.061759  0.242741
3    -0.172926 -0.352752 -0.843795  0.206998 -0.681415 -0.036133 -0.054284 -0.231477 -0.028003 -0.250678 -0.051034
4     0.187057 -0.807610  0.028537  0.548332  0.854074 -0.097305 -0.457043  0.171758 -0.074409  0.031897  0.082245
...        ...       ...       ...       ...       ...       ...       ...       ...       ...       ...       ...
1448  1.550242  0.275031 -0.013737  0.200604 -0.145063  0.306575 -0.075308  0.345552 -0.136589 -0.432798 -0.456076
1449 -0.957339  0.014308  0.303843  0.444350 -0.133690  0.381804 -0.326432  0.878047 -0.304441 -0.247443 -0.193671
1450 -0.185894  1.062662  0.220857 -0.467643 -0.187757 -0.192703 -0.091597 -0.036576  0.038255  0.056518 -0.012800
1451 -1.182064 -0.038570  0.561561  0.701126  0.047645  0.193687 -0.027335 -0.339374  0.022267 -0.002573 -0.105316
1452  1.550242  0.275031 -0.013737  0.200604 -0.145063  0.306575 -0.075308  0.345552 -0.136589 -0.432798 -0.456076
1453 rows × 11 columns
```

Extracting Segments(step-5)

K-means Clustering Algorithm: This section introduces the K-means clustering algorithm, which partitions the data into distinct clusters based on similarities. The optimal number of clusters is determined using techniques such as the elbow method or silhouette analysis.

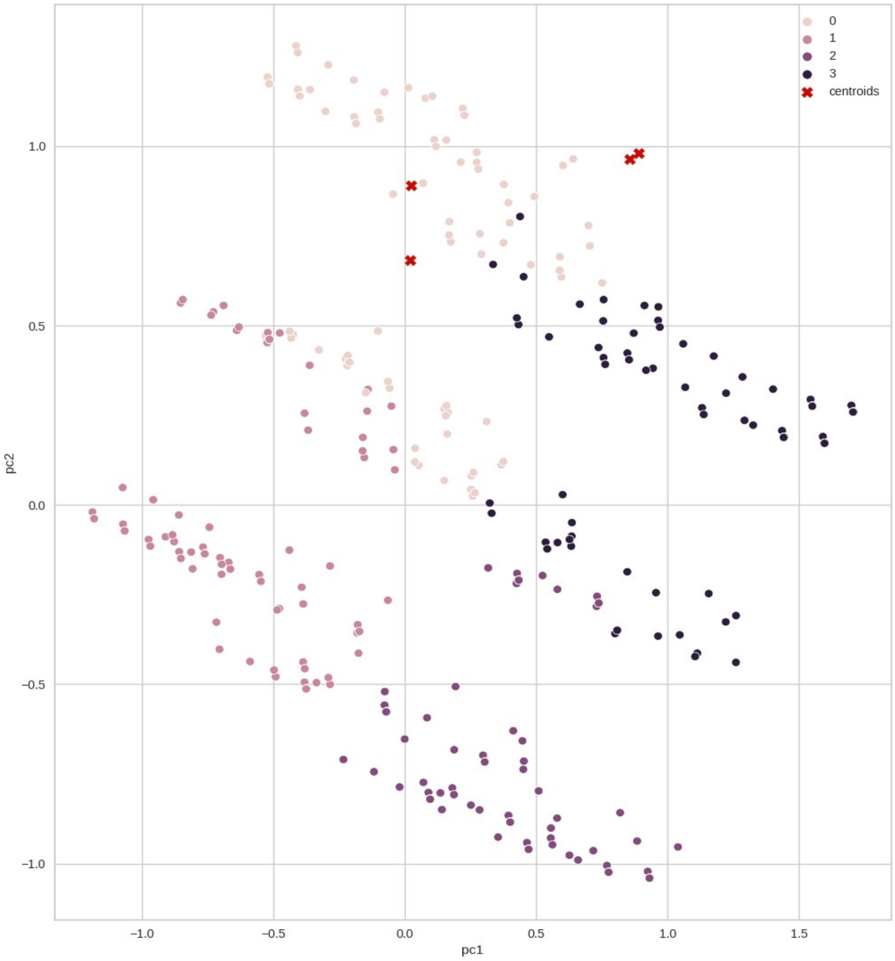
Code Snippet: K-means Clustering in Python:



```
#Extracting segments

#Using k-means clustering analysis
from sklearn.cluster import KMeans
from yellowbrick.cluster import KElbowVisualizer
model = KMeans()
visualizer = KElbowVisualizer(model, k=(1,12)).fit(df_eleven)
visualizer.show()
```

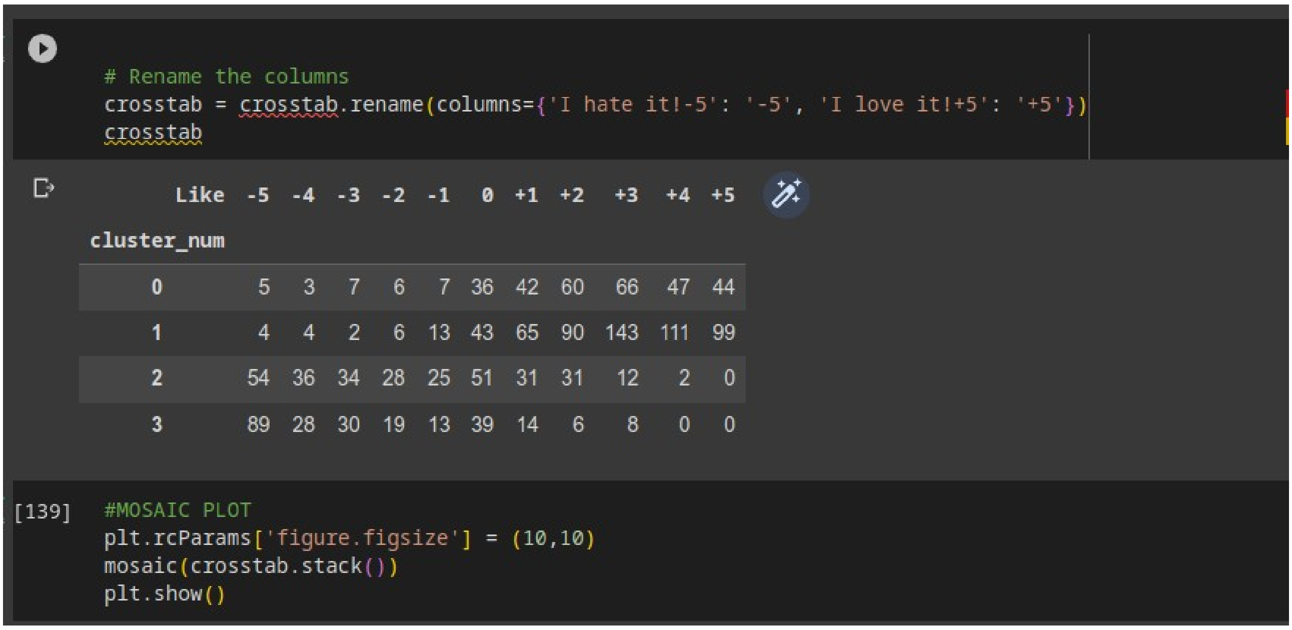
```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init`
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init`
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init`
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init`
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init`
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init`
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init`
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init`
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init`
```

Result:



Descriptor Variables Analysis: This section analyzes additional descriptor variables, such as gender and age, to gain a deeper understanding of market segments. Mosaic plots, box plots, and statistical tests are utilized to examine the relationships between segment membership and descriptor variables.

Code Snippet: Mosaic Plot in Python:

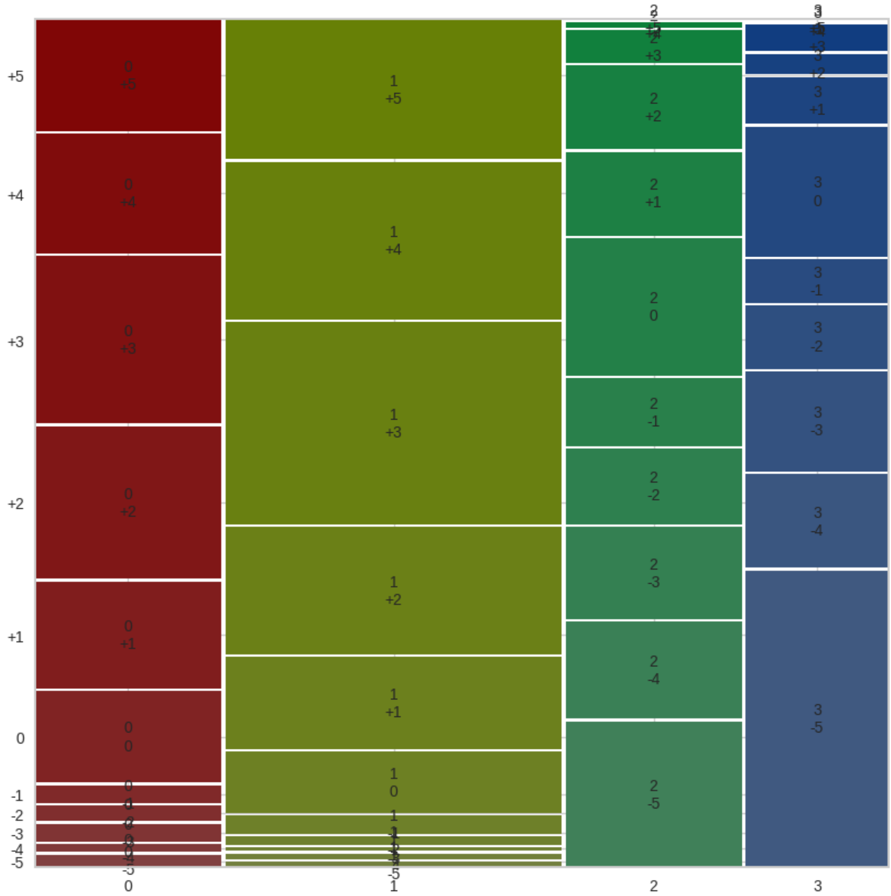


```
# Rename the columns
crosstab = crosstab.rename(columns={'I hate it!-5': '-5', 'I love it!+5': '+5'})
crosstab
```

```
        Like  -5  -4  -3  -2  -1   0  +1  +2   +3   +4  +5
cluster_num
     0         5   3   7   6   7  36  42  60   66   47  44
     1         4   4   2   6  13  43  65  90  143  111  99
     2        54  36  34  28  25  51  31  31   12    2   0
     3        89  28  30  19  13  39  14   6    8    0   0
```

```
[139] #MOSAIC PLOT
      plt.rcParams['figure.figsize'] = (10,10)
      mosaic(crosstab.stack())
      plt.show()
```

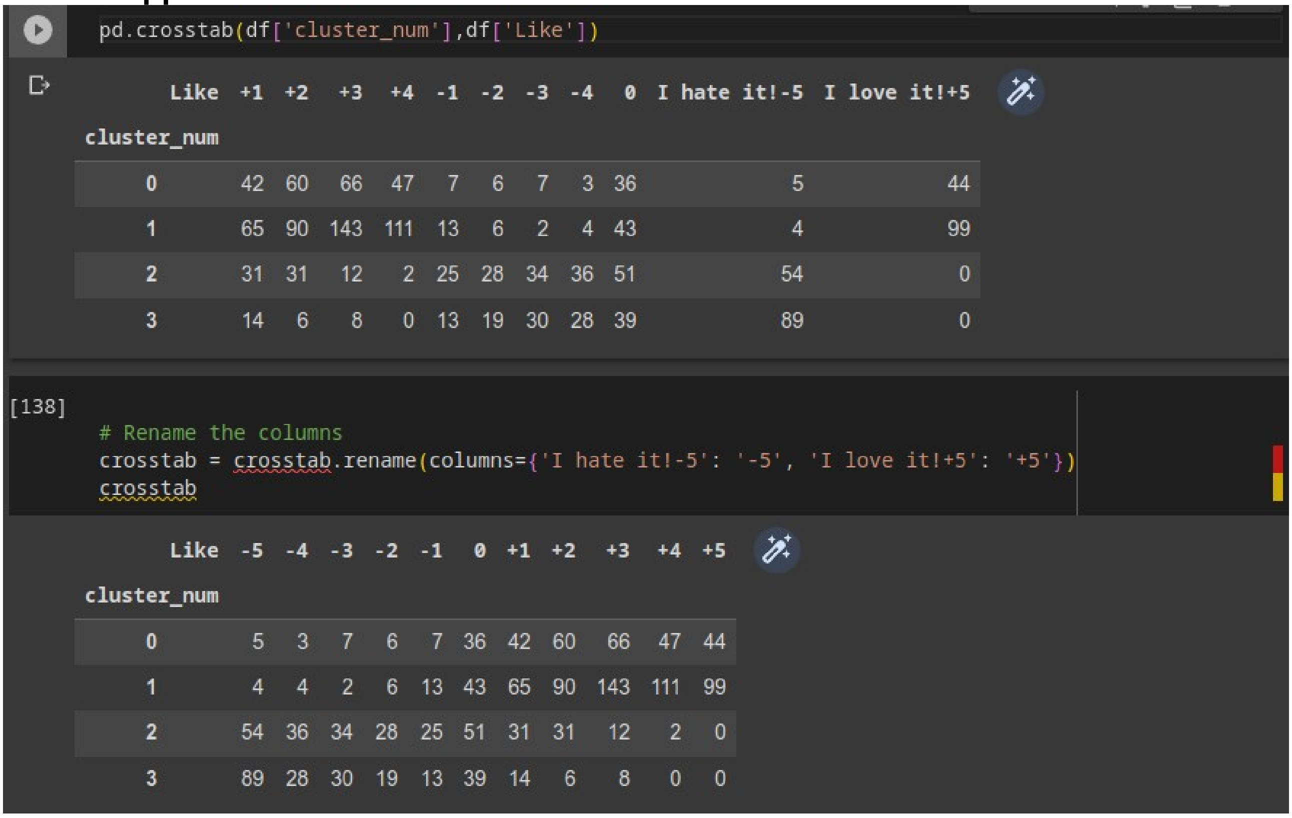
Out Put:



Describing Segments (step-7)

The fast food data set is not typical for data collected for market segmentation analysis because it contains very few descriptor variables. Descriptor variables – additional pieces of information about consumers – are critically important to gaining a good understanding of market segments

Code Snippet:



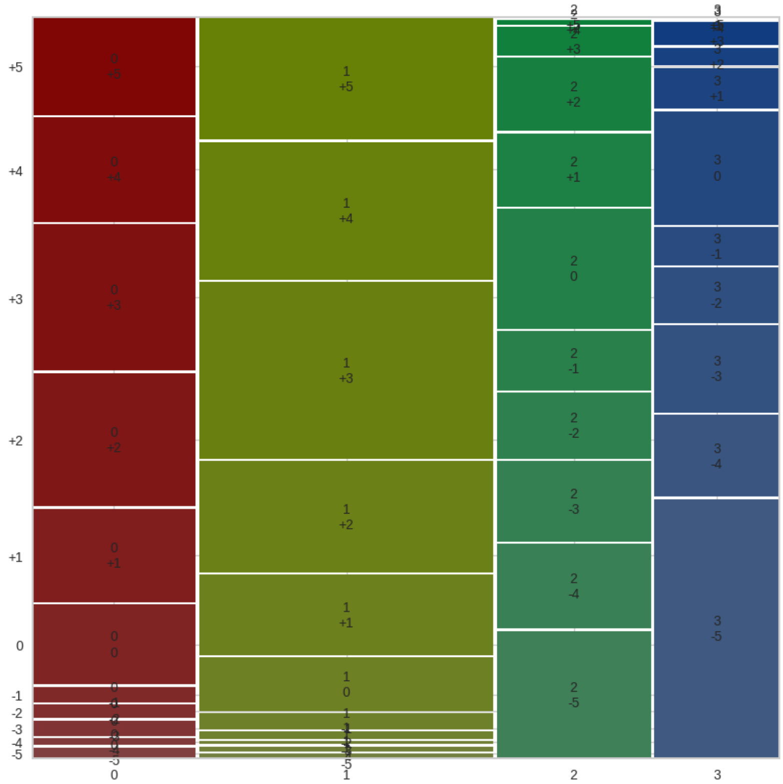
```
pd.crosstab(df['cluster_num'],df['Like'])
```

```
Like         +1   +2   +3   +4   -1   -2   -3   -4   0   I hate it!-5   I love it!+5
cluster_num
     0       42   60   66   47    7    6    7    3  36              5             44
     1       65   90  143  111   13    6    2    4  43              4             99
     2       31   31   12    2   25   28   34   36  51             54              0
     3       14    6    8    0   13   19   30   28  39             89              0
```

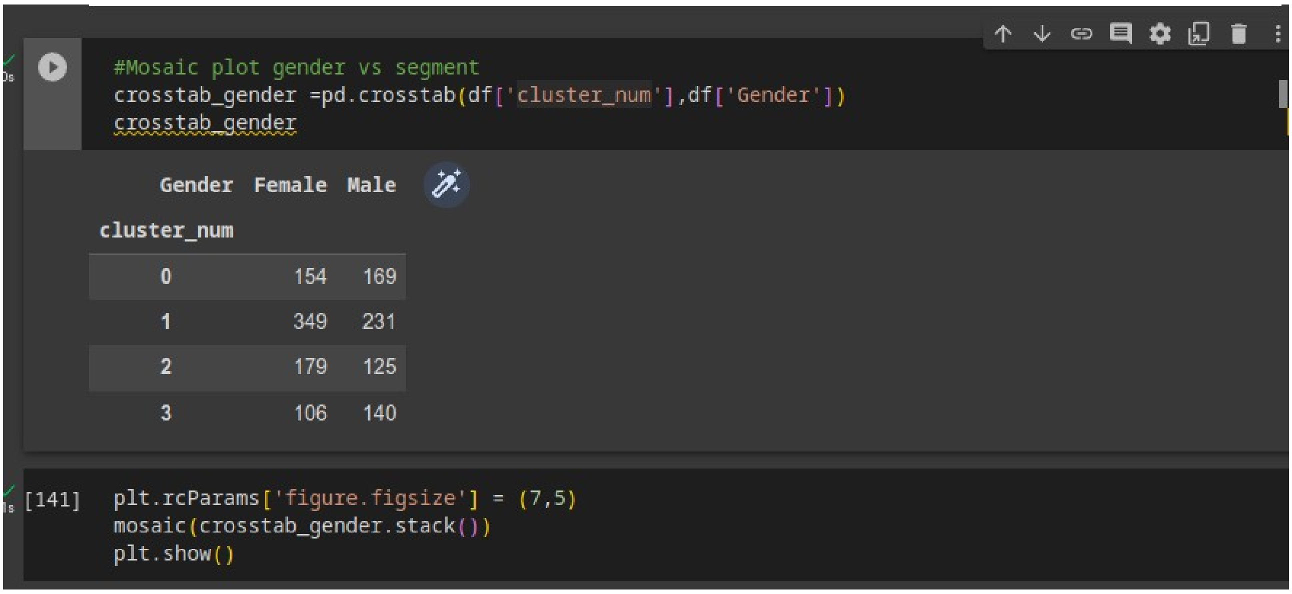
```
[138]
# Rename the columns
crosstab = crosstab.rename(columns={'I hate it!-5': '-5', 'I love it!+5': '+5'})
crosstab
```

```
Like         -5   -4   -3   -2   -1    0   +1   +2   +3   +4   +5
cluster_num
     0        5    3    7    6    7   36   42   60   66   47   44
     1        4    4    2    6   13   43   65   90  143  111   99
     2       54   36   34   28   25   51   31   31   12    2    0
     3       89   28   30   19   13   39   14    6    8    0    0
```

Output:



Mosaic plot gender vs segment:

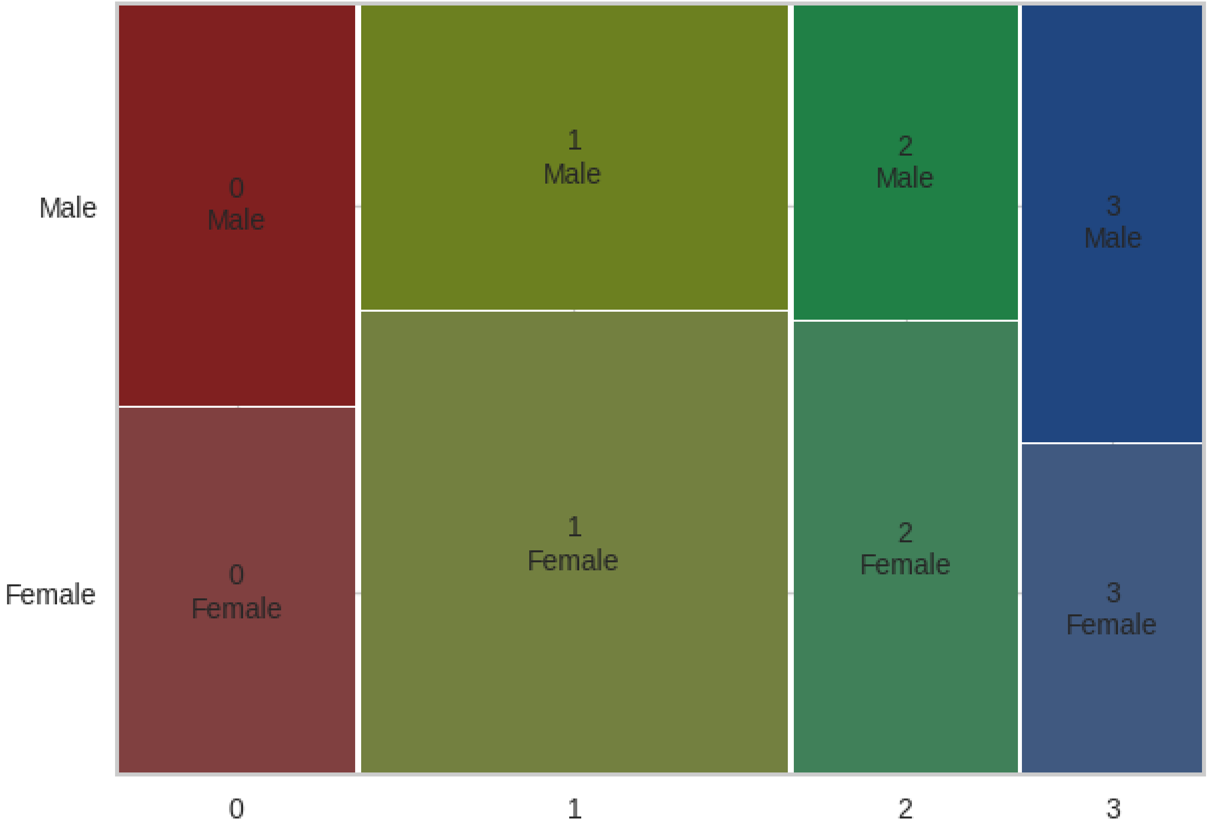


```
#Mosaic plot gender vs segment
crosstab_gender =pd.crosstab(df['cluster_num'],df['Gender'])
crosstab_gender
```

```
     Gender   Female   Male
cluster_num
     0           154    169
     1           349    231
     2           179    125
     3           106    140
```

```
plt.rcParams['figure.figsize'] = (7,5)
mosaic(crosstab_gender.stack())
plt.show()
```

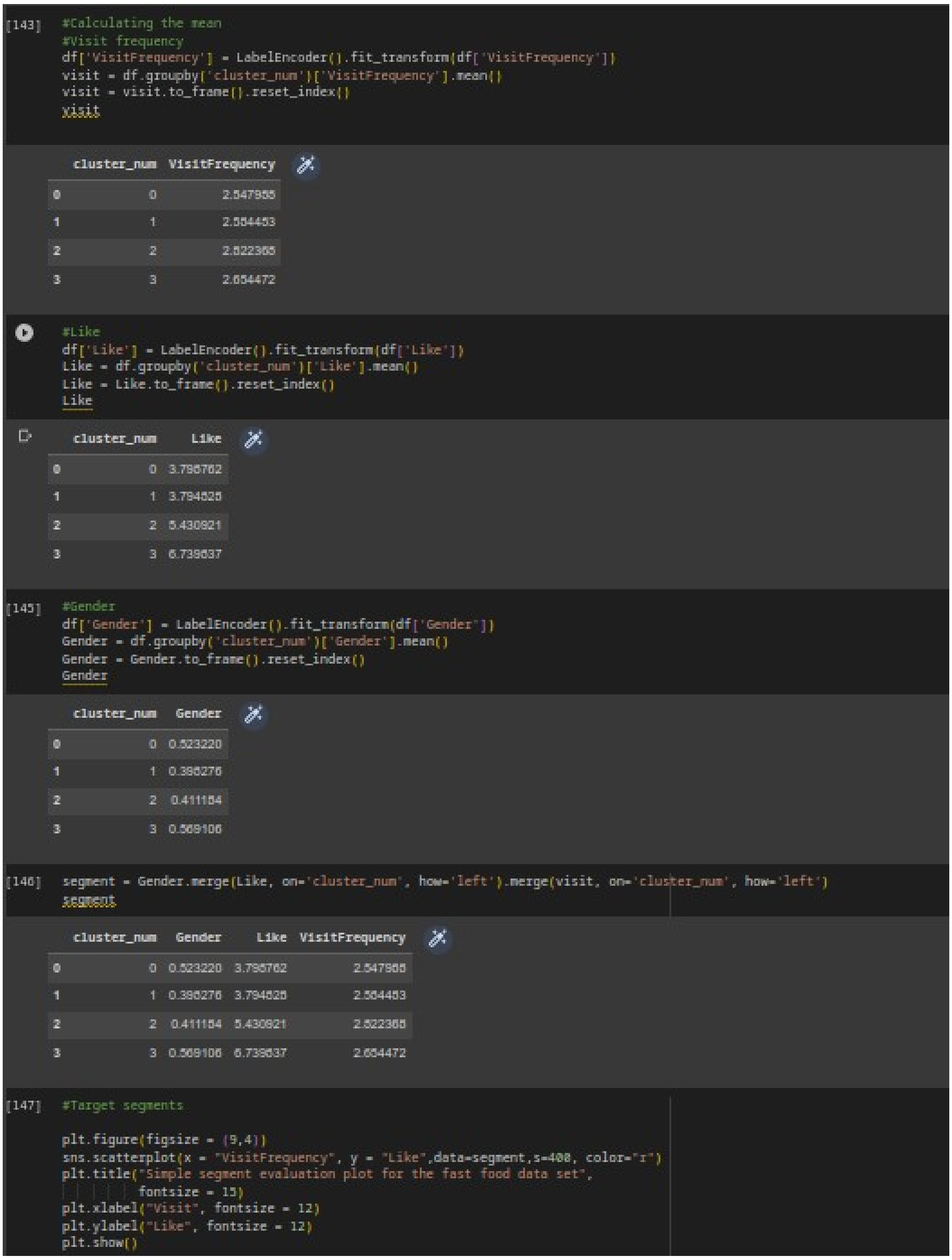
Output:



Selecting target segment (step-8)

The selection of a target segment is a crucial step in the market segmentation process as it determines where a company will focus its marketing efforts and allocate resources. By identifying a specific target segment, companies can tailor their strategies and offerings to better meet the needs and preferences of their most valuable customers. In this section, we will discuss the process of selecting the target segment based on various criteria and considerations.

Code Snippet: Selecting target segment



```
#Calculating the mean
#Visit frequency
df['VisitFrequency'] = LabelEncoder().fit_transform(df['VisitFrequency'])
visit = df.groupby('cluster_num')['VisitFrequency'].mean()
visit = visit.to_frame().reset_index()
visit
```

```
   cluster_num  VisitFrequency
0            0        2.547955
1            1        2.584453
2            2        2.822368
3            3        2.654472
```

```
#Like
df['Like'] = LabelEncoder().fit_transform(df['Like'])
Like = df.groupby('cluster_num')['Like'].mean()
Like = Like.to_frame().reset_index()
Like
```

```
   cluster_num      Like
0            0  3.795762
1            1  3.794528
2            2  5.430921
3            3  6.739637
```

```
#Gender
df['Gender'] = LabelEncoder().fit_transform(df['Gender'])
Gender = df.groupby('cluster_num')['Gender'].mean()
Gender = Gender.to_frame().reset_index()
Gender
```

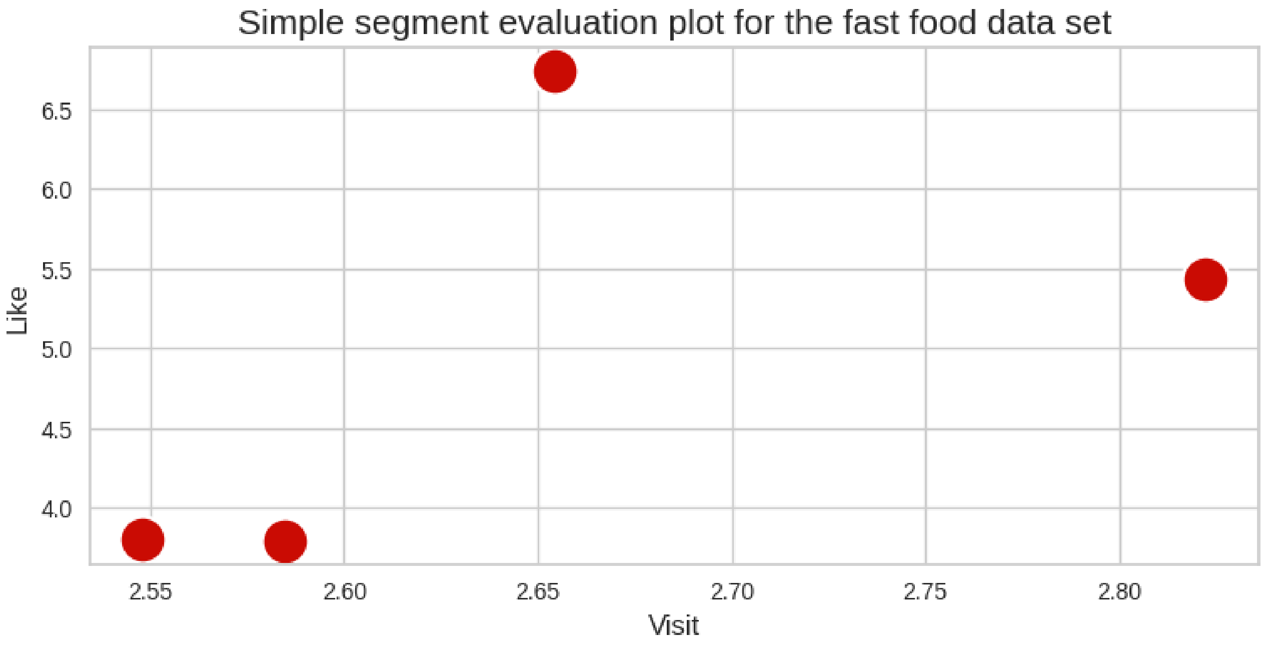
```
   cluster_num    Gender
0            0  0.523220
1            1  0.386276
2            2  0.411154
3            3  0.569106
```

```
segment = Gender.merge(Like, on='cluster_num', how='left').merge(visit, on='cluster_num', how='left')
segment
```

```
   cluster_num    Gender      Like  VisitFrequency
0            0  0.523220  3.795762        2.547955
1            1  0.386276  3.794528        2.584453
2            2  0.411154  5.430921        2.822368
3            3  0.569106  6.739637        2.654472
```

```
#Target segments

plt.figure(figsize = (9,4))
sns.scatterplot(x = "VisitFrequency", y = "Like",data=segment,s=400, color="r")
plt.title("Simple segment evaluation plot for the fast food data set",
          fontsize = 15)
plt.xlabel("Visit", fontsize = 12)
plt.ylabel("Like", fontsize = 12)
plt.show()
```

Output:



Customizing the Marketing Mix (step-9)

In Step 9, the marketing mix is tailored to the selected target segment. For example, if McDonald's focuses on Segment 3, which consists of young customers who like the food but find it expensive, they could introduce a MCSUPERBUDGET line. This line caters to the price expectations of Segment 3 and aims to develop customer loyalty. Distinct product features, effective communication channels, and strategic distribution channels are essential in designing the marketing mix. Continuous evaluation and monitoring ensure the marketing strategy remains effective in meeting the needs of the target segment.

Evaluation and Monitoring(step-10)

Assessing Market Segmentation Strategy Success: The success of the market segmentation strategy is evaluated based on predetermined metrics and key performance indicators. This section discusses the evaluation process and the metrics used to measure the effectiveness of the strategy.

Monitoring Market Changes: The market is continuously monitored to identify any changes in consumer preferences, competitor activities, or market dynamics. This section highlights the importance of ongoing monitoring to ensure the company remains adaptable to changing market conditions.

Conclusion

The market segmentation analysis provides valuable insights into the fast food industry, enabling the fast food company to identify distinct market segments and tailor their marketing strategies accordingly. By continuously evaluating and monitoring the market, the company can adapt its strategies to meet changing consumer needs and market dynamics, ensuring long-term success and competitiveness.

Colab link- https://colab.research.google.com/drive/10_qkVo-l1p6O73uxpVj2KIUG5WAmZshK?usp=sharing

Github link-https://github.com/Subhadip023/market-segmet-analysis