Fundamentals of
Market Segmentation and McDonalds Case study

Done By – Team Vijay

Vijay Gupta(Team lead)

Anubhav Anand

Deepika Yadav

Uppari Ganesh

Arindam Mondal

Abhijeet Gajanan Pradhan

Marketing

Marketing is the set of activities a company undertakes to promote the buying or selling f a product and service. It may include advertising, promotion, delivering products to customers. Marketing plan consists of various steps, Strategic planning, tactical planning. • The strategic plan aims on long time goals of the organization. • The tactical plan aims on short time goals of the organization. SWOT analysis is done for marketing planning. S(strength) W(weakness) O(opportunities) T(threat)

Marketing Segmentation

Market Segmentation is the process of dividing a broad group of consumer or target market into sub groups of consumers based on demographics, needs, common interests, geographic locations, or psychographic behavior. Market segmentation is one of the key building blocks of strategic marketing. Market segmentation is essential for marketing success.

Benefits of Market segmentation

It helps in achieving long time goals that is market dominance in particular area which results from being best able to cater to the needs of a very specific market segment. It helps to yield a higher return on investment.

Cost of Market segmentation

Implementing market segmentation requires a large investment by the organization. A large number of people are needed for the market segmentation analysis. They have to spend a whole lot of time for this process. They have to monitor the market dynamics properly. In the worst case, if market segmentation is not implemented well, the entire exercise is a waste of resources.

Steps in Market Segment Analysis:

Step 1: Deciding (not) to segment

- In Business, market segmentation determines who is in your target market and who is not.
- The main purpose of segmenting any market is to identify different consumer needs in order to construct an appropriate marketing mix.
- When utilizing market segmentation, you look at all the people who could buy your product and decide how to break them up into groups that have similar needs, wants or demand characteristics.
- Then we have to communicate with different groups using different different messages and marketing techniques.

Customer needs according to their niche

Step 2:- Specifying the ideal target segment

Segment Evaluation Criteria:

The third layer of market segmentation analysis) depends primarily on user input. It is important to understand that – for a market segmentation analysis to produce results that are useful to an organisation – user input cannot be limited to either a briefing at the start of the process, or the development of a marketing mix at the end. Rather, the user needs to be involved in most stages, literally wrapping around the technical aspects of market segmentation analysis. After having committed to investigating the value of a segmentation strategy in Step 1, the organisation has to make a major contribution to market segmentation analysis in Step 2. While this contribution is conceptual in nature, it guides many of the following steps, most critically Step 3 (data collection) and Step 8 (selecting one or more target segments). In Step 2 the organisation must determine two sets of segment evaluation criteria. One set of evaluation criteria can be referred to as knock-out criteria. These criteria are the essential, non-negotiable features of segments that the organisation would consider targeting. The second set of evaluation criteria can be referred to as attractiveness criteria. These criteria are used to evaluate the relative attractiveness of the remaining market segments – those in compliance with the knock-out criteria. The literature does not generally distinguish between these two kinds of criteria. Instead, the literature proposes a wide array of possible segment evaluation criteria and describes them at different levels of detail.

Knock out Criteria:

- Knock-out criteria are used to determine if market segments resulting from the market segmentation analysis qualify to be assessed using segment attractiveness criteria.
- The segment must be homogeneous; members of the segment must be similar to one another.
- The segment must be distinct; members of the segment must be distinctly different from members of other segments.
- The segment must be large enough; the segment must contain enough consumers to make it worthwhile to spend extra money on customising the marketing mix for them.
- The segment must be matching the strengths of the organisation; the organisation must have the capability to satisfy segment members' needs.
- Members of the segment must be identifiable; it must be possible to spot them in the marketplace.
- The segment must be reachable; there has to be a way to get in touch with members of the segment in order to make the customised marketing mix accessible to them.

Attractiveness Criteria

Attractiveness criteria are not binary in nature. Segments are not assessed as either complying or not complying with attractiveness criteria. Rather, each market segment is rated; it can be more or less attractive with respect to a specific criterion. The attractiveness across all criteria determines whether a market segment is selected as a target segment in Step 8 of market segmentation analysis.

Implementing a Structured Process

There is general agreement in the segmentation literature, that following a structured process when assessing market segments is beneficial. The most popular structured approach for evaluating market segments in view of selecting them as target markets is the use of a segment evaluation plot showing segment attractiveness along one axis, and organisational competitiveness on the other axis. Factors which constitute both segment attractiveness and organisational competitiveness need to be negotiated and agreed upon. To achieve this, a large number of possible criteria has to be investigated before agreement is reached on which criteria are most important for the organisation.

Step 3:- Collecting Data

In common sense segmentation, the segmentation variable is typically one single characteristic of the consumers in the sample. Describing segments is critical to being able to develop an effective marketing mix targeting the segment. Typical descriptor variables include socio-demographics, but also information about media behavior. The correct description, in turn, makes it possible to develop a customized product, determine the most appropriate pricing strategy, select the best distribution channel, and the most effective communication channel for advertising and promotion.

Geographic Segmentation

Geographic information is segmentation criterion used for the purpose of market segmentation. when geographic segmentation is used – the consumer's location is considered.

Socio-Demographic Segmentation

Typical socio-demographic segmentation criteria include age, gender, income and education. Sociodemographic segments can be very useful in some industries.

Psychographic Segmentation

When people are grouped according to psychological criteria, such as their beliefs, interests, preferences, aspirations, or benefits sought when purchasing a product, the term psychographic segmentation is used. Psychographic criteria are, by nature, more complex than geographic or socio-demographic criteria because it is difficult to find a single characteristic of a person that will provide insight into the psychographic dimension of interest.

Behavioral Segmentation

Another approach to segment extraction is to search directly for similarities in behavior or reported behavior. A wide range of possible behaviors can be used for this purpose, including prior experience with the product, frequency of purchase, amount spent on purchasing the product on each occasion (or across multiple purchase occasions), and information search behavior. Survey data is cheap and easy to collect. But survey data can be contaminated by a wide range of biases. Such biases can, in turn, negatively affect the quality of solutions derived from market segmentation analysis.

Carefully selecting the variables that are included as segmentation variable in common sense segmentation, or as segmentation variables in data-driven segmentation, is critical to the quality of the market segmentation solution. Developing a good questionnaire typically requires conducting exploratory or qualitative research. Answer options provided to respondents in surveys determine the scale of the data available for subsequent analyses. Survey data is prone to capturing biases.

A response bias is a systematic tendency to respond to a range of questionnaire items on some basis other than the specific item content. Increasingly organizations have access to substantial amounts of internal data that can be harvested for the purpose of market segmentation analysis. Typical examples are scanner data available to grocery stores, booking data available through airline loyalty programs, and online purchase data. The strength of such data lies in the fact that they represent actual behavior of consumers, rather than statements of consumers about their behavior or intentions, known to be affected by imperfect memory.

The danger of using internal data is that it may be systematically biased by over-representing existing customers. What is missing is information about other consumers the organization may want to win as customers in future, which may differ systematically from current customers in their consumption patterns.

Step 4:- Exploring the data

Before analysis data is cleaned. In this all values are checked if recorded correctly, labels for levels of categorical variables. Grouping the values of some variables if necessary. Checking for unique values and ensuring that they only contain permissible values. Any other values are not permissible, and need to be corrected as part of the data cleaning procedure. Using codes to clean data can swiftly increase the speed of process. The codes written in data cleaning should be retained to ensure that every step of data cleaning, exploration, and analysis can be reproduced in future.

Descriptive Analysis

In descriptive analysis data is plotted and graphically represented to gain insights into the data. Packages of programming languages provide tools for representation. The tools can identify relation between variables and similarity and behaviour between them. Graphical methods used for numeric variables are histograms, boxplots and scatter plots, for categorical variables barplots are used, countplot and mosaic plot are used for multiple categorical variables. Using bins can provide detail information about variables. Using graphical representation outliers can easily identified. Converting binary variables into percentage using code and plotting them graphically can result in some useful insights.

Pre-processing

There are two steps to process categorical variables, one is to merge variable and other is to convert variables into numeric variables. When variables are too differentiate, merging is used. Variable can be converted into numeric variable if it can be assumed that distances between adjacent scale points on the ordinal scale are approximately equal. Another method used for multi category variables is Likert scale. Variables having multiple answer options use Likert scale.

In numeric variables to balance the influence of segmentation variables on segmentation results, variables can be standardised. Standardising means transforming variables in way that puts them into a common scale. If variables contain binary answers, then data prepossessing is not required.

Principal Components Analysis

PCA transforms multivariate dataset containing metric variables into new dataset with variables known as principal components which uncorrelated and ordered by importance. It transforms high dimensional data into lower dimensional data for plotting purposes. Sometimes it is used to reduce number of segmentation variables. It is used to identify highly correlated variables. Insights gained from such analysis can be used to remove some of the original redundant variables from the segmentation base.

After cleaning and pre-processing data in next step segments will be extracted from it.

Step 5:- Extracting Segments

Grouping Consumers:

• This illustration gives the impression that single linkage clustering is much more powerful, and should be preferred over other approaches of extracting market segments from data. • The aim of this chapter is to provide an overview of the most popular extraction methods used in market segmentation, and point out their specific tendencies of imposing structure on the extracted segments.

Distance-based methods:

Use a particular notion of similarity or distance between observations (consumers),and try to find groups of similar observations (market segments). The table contains the information needed to guide algorithm selection:

Hierarchical Methods:

Hierarchical clustering methods are the most intuitive way of grouping data because they mimic how a human would approach the task of dividing a set of n observations (consumers) into k groups (segments). • Divisive hierarchical clustering and Agglomerative hierarchical clustering approaches. • A very popular alternative hierarchical clustering method is named after Ward and based on squared Euclidean distances. • The result of hierarchical clustering is typically presented as a dendrogram .A dendrogram is a tree diagram.

Partitioning Methods:

A partitioning clustering algorithm aiming to extract five market segments, in contrast, would only have to calculate between 5 and 5000 distances at each step of the iterative or stepwise process.

k-Means and k-Centroid Clustering: ->The most popular partitioning method is k-means clustering. Within this method, a number of algorithms are available. ->These algorithms use the squared Euclidean distance. A generalisation to other distance measures, also referred to as k-centroid clustering.

*__Improved" k-Means:__

Many attempts have been made to refine and improve the k-means clustering algorithm. The simplest improvement is to initialise k-means using "smart" starting values, rather than randomly drawing k consumers from the data set and using them as starting points.

***Hard Competitive Learning:**

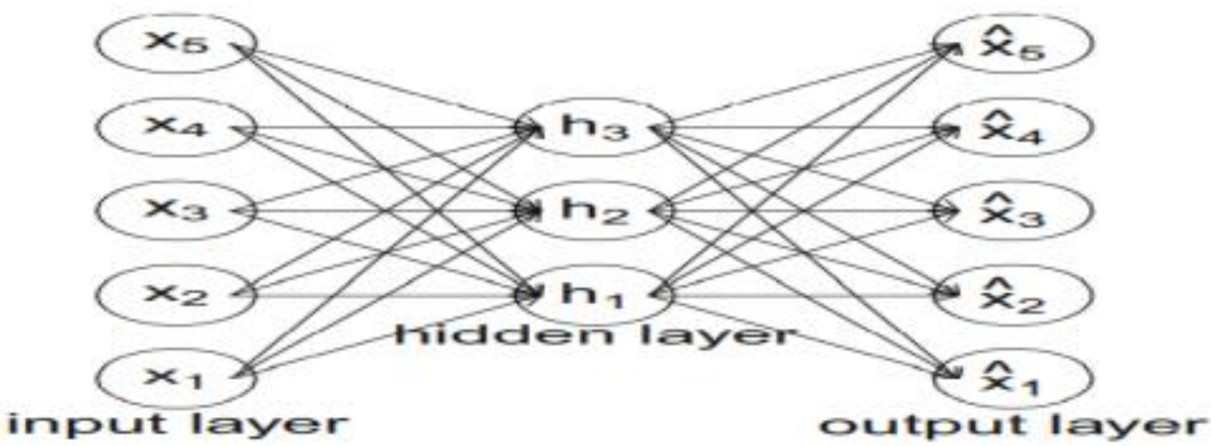
Hard competitive learning, also known as learning vector quantisation , differs from the standard k-means algorithm in how segments are extracted. *Neural Gas and Topology Representing Networks: A variation of hard competitive learning is the neural gas algorithm. Here, not only the segment representative (centroid) is moved towards the randomly selected consumer.

*** Self-Organising Maps:**

Another variation of hard competitive learning are self-organising maps. Selforganising maps position segment representatives on a regular grid, usually a rectangular or hexagonal grid.

***Neural Networks:**

Auto-encoding neural networks for cluster analysis work mathematically differently than all cluster methods presented so far. The most popular method from this family of algorithms uses a so-called single hidden layer perceptron.



Hybrid Approaches:

• The basic idea behind hybrid segmentation approaches is to first run a partitioning algorithm because it can handle data sets of any size.

*** Two-Step Clustering:**

A procedure referred to as two step clustering. The two steps consist of run a partitioning followed by a hierarchical procedure.

*Bagged Clustering:

Bagged clustering also combines hierarchical clustering algorithms and partitioning clustering algorithms, but adds bootstrapping. Bootstrapping can be implemented by random drawing from the data set with replacement. Bootstrapping has the advantage of making the final segmentation solution less dependent on the exact people contained in consumer data. Bagged clustering is suitable in the following circumstances:

• If we suspect the existence of niche markets.

• If we fear that standard algorithms might get stuck in bad local solutions.

• If we prefer hierarchical clustering, but the data set is too large.

Model-Based Methods:

Model-Based Methods is based on the assumption that the true market segmentation solution – which is unknown – has the following two general properties:

• Each market segment has a certain size.

Finite Mixtures of Distributions: →It has two main distributions to discuss:

(i).Normal Distribution: The most popular finite mixture model is a mixture of several multivariate normal distributions. The multivariate normal distribution can easily model covariance between variables; and approximate multivariate normal distributions occur in both biology and business.

(ii).Binary Distributions: For binary data, finite mixtures of binary distributions, sometimes also referred to as latent class models or latent class analysis are popular.

Finite Mixtures of Regressions:

Finite mixture of regression models assume the existence of a dependent target variable y that can be explained by a set of independent variables x. The functional relationship between the dependent and independent variables is considered different for different market segments.

Extensions and Variations:

Mixture models also allow to simultaneously include segmentation and descriptor variables. Segmentation variables are used for grouping, and are included in the segment specific model as usual. Descriptor variables are used to model differences in segment sizes, assuming that segments differ in their composition with respect to the descriptor variables.

Algorithms with Integrated Variable Selection:

These algorithms assume that each of the segmentation variables makes a contribution to determining the segmentation solution. But this is not always the case. Sometimes, segmentation variables were not carefully selected, and contain redundant or noisy variables. Pre-processing methods can identify them.

Biclustering Algorithms:

Biclustering simultaneously clusters both consumers and variables. Biclustering algorithms exist for any kind of data, including metric and binary. This section focuses on the binary case where these algorithms aim at extracting market segments containing consumers who all have a value of 1 for a group of variables. These groups of consumers and variables together then form the bicluster.

Step 6: Profiling Segments

Identifying Key Characteristics of Market Segments:

- The aim of the profiling step is to get to know the market segments resulting from the extraction step. Profiling is only required when data-driven market segmentation is used. For commonsense segmentation, the profiles of the segments are predefined. Identifying the defining characteristics of market segments with respect to the segmentation variables is the aim of profiling. Profiling consists of characterising the market segments individually, but also in comparison to the other market segments.
- At the profiling stage, we inspect a number of alternative market segmentation solutions. This is particularly important if no natural segments exist in the data, and either a reproducible or a constructive market segmentation approach has to be taken. Good profiling is the basis for correct interpretation of the resulting segments. Correct interpretation, in turn, is critical to making good strategic marketing decisions.

Segment Profiling with Visualisations:

- Neither the highly simplified, nor the very complex tabular representation typically used to present market segmentation solutions make much use of graphics, although data visualisation using graphics is an integral part of statistical data analysis.
- Graphics are particularly important in exploratory statistical analysis (like cluster analysis) because they provide insights into the complex relationships between variables. In addition, in times of big and increasingly bigger data, visualisation offers a simple way of monitoring developments over time.
- Visualisations are useful in the data-driven market segmentation process to inspect, for each segmentation solution, one or more segments in detail.

Step 7:-Describing Segments

Segment profiling is about understanding differences in segmentation variables across market segments. Segmentation variables are chosen early in the market segmentation analysis process: conceptually in Step 2 (specifying the ideal target segment), and empirically in Step 3 (collecting data). Segmentation variables form the basis for extracting market segments from empirical data.

Using graphical statistics to describe market segments has two key advantages: it simplifies the interpretation of results for both the data analyst and the user, and integrates information on the statistical significance of differences, thus avoiding the over-interpretation of insignificant differences.

Step 8:- Selecting (the) Target Segment(s)

• Market targeting means the process of evaluating the attractiveness of each market segment and selecting one or more segments to enter.

• A target segment is a set of buyers who share common needs or characteristics that the company decides to serve.

• After a firm has identified its market segment opportunities, it has to decide how many and which ones to target.

•In target method marketing, the seller divides the market into segments, chooses one or more of them, and develops product and marketing mixes most appropriate for each selected segment.

• It is impossible to appeal to all customers in the marketplace who are widely dispersed with variety of needs. Organizations that wants to succeed must identify their customers and develop marketing to satisfy their needs.

importance of Target segment

• Understanding customer’s needs.

• Suitable marketing mix.

• Differentiation.

Step 9:- Customizing the Marketing Mix

Commonly the marketing mix is understood as consisting of the 4Ps: Product, Price, Promotion and Place. Market segmentation does not stand independently as a marketing strategy. Rather, it goes hand in hand with the other areas of strategic marketing, most importantly: positioning and competition.

Product - One of the key decisions an organization needs to make when developing the product dimension of the marketing mix, is to specify the product in view of customer needs. Other marketing mix decisions that fall under the product dimension are: naming the product, packaging it, offering or not offering warranties, and after sales support services.

Price - Typical decisions an organization needs to make when developing the price dimension of the marketing mix include setting the price for a product, and deciding on discounts to be offered.

Place - The key decision relating to the place dimension of the marketing mix is how to distribute the product to the customers. This includes answering questions such as: should the product be made available for purchase online or offline only or both; should the manufacturer sell directly to customers; or should a wholesaler or a retailer or both be used.

Promotion - developing an advertising message that will resonate with the target market, and identifying the most effective way of communicating this message. Other tools in the promotion category of the marketing mix include public relations, personal selling, and sponsorship

Implementation

Data Sources :

We have gathered some datasets which are somehow related to the case. The link of the dataset is given below https://homepage.boku.ac.at/leisch/MSA/datasets/mcdonalds.csv

Packages and Tools used :

- Numpy
- Pandas
- Matplotlib
- Seaborn
- Plotly
- Sklearn
- Scipy

The dataset is based on the previous stat of the customer review and behaviour about the Online food, here we have taken the case of "Mcdonalds". Uploading packages and Tools

In [94]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly as py
import plotly.graph_objs as go
import plotly.figure_factory as ff
import plotly.graph_objects as go
import plotly.express as px
from plotly.subplots import make_subplots
```

Uploading dataset :

In [95]:
```python
df = pd.read_csv('C:\\Users\\SUSNATA BISWAS\\OneDrive\\Desktop\\mcdonalds.csv')
```

In [96]:
```python
# path_1 = "/content/drive/MyDrive/mcdonalds.csv"
# df = pd.read_csv(path_1)
df.head(10)
```

Out[96]:

	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
5	Yes	Yes	No	Yes	No	Yes	Yes	Yes	No	No	No	+2	55	Every three months	Male
6	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	No	+2	56	Every three months	Female
7	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No	No	I love it!+5	23	Once a week	Female
8	No	No	No	Yes	Yes	No	No	No	Yes	No	Yes	I hate it!-5	58	Once a year	Male
9	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	No	No	+1	32	Every three months	Female

Data Pre-processing :

● Checking is there any null values in the data set :

In [97]:
```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1453 entries, 0 to 1452
Data columns (total 15 columns):
 #   Column          Non-Null Count  Dtype 
---  ------          --------------  ----- 
 0   yummy           1453 non-null   object
 1   convenient      1453 non-null   object
 2   spicy           1453 non-null   object
 3   fattening       1453 non-null   object
 4   greasy          1453 non-null   object
 5   fast            1453 non-null   object
 6   cheap           1453 non-null   object
 7   tasty           1453 non-null   object
 8   expensive       1453 non-null   object
 9   healthy         1453 non-null   object
 10  disgusting      1453 non-null   object
 11  Like            1453 non-null   object
 12  Age             1453 non-null   int64 
 13  VisitFrequency  1453 non-null   object
 14  Gender          1453 non-null   object
dtypes: int64(1), object(14)
memory usage: 170.4+ KB
```

In our data set we have 1453 rows and 15 columns. By applying df.info (), we can see that we have no null values.

Removing duplicates :

```
In [123]: print('Duplicated value(s) on the dataset1 : ', df.duplicated().sum())
```

```
Duplicated value(s) on the dataset1 :  22
```

```
In [124]: df.drop_duplicates(keep="first", inplace=True)
```

```
In [125]: df
```

Out[125]:

	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

1431 rows × 15 columns

After this operation of the duplicacy test, we found that 22 rows were similar with each other, So, we have removed it. After removing, the total number of rows we have is (1453 - 22) = 1431.

Before removing duplicates no. of rows = 1453 After removing duplicates no. of rows = 1431

Handling categorical features :

We have manually transformed all the categorical features to all numerical values.

```
In [141]: df['Gender'] = df['Gender'].replace({'Male':1, 'Female':0})
df['yummy'] = df['yummy'].replace({'Yes':1, 'No':0})
df['convenient'] = df['convenient'].replace({'Yes':1, 'No':0})
df['spicy'] = df['spicy'].replace({'Yes':1, 'No':0})
df['fattening'] = df['fattening'].replace({'Yes':1, 'No':0})
df['greasy'] = df['greasy'].replace({'Yes':1, 'No':0})
df['fast'] = df['fast'].replace({'Yes':1, 'No':0})
df['cheap'] = df['cheap'].replace({'Yes':1, 'No':0})
df['tasty'] = df['tasty'].replace({'Yes':1, 'No':0})
df['expensive'] = df['expensive'].replace({'Yes':1, 'No':0})
df['healthy'] = df['healthy'].replace({'Yes':1, 'No':0})
df['disgusting'] = df['disgusting'].replace({'Yes':1, 'No':0})
df['VisitFrequency'] = df['VisitFrequency'].replace({'Never':0, 'Once a year':1, 'Every three months':2, 'Once a month':3, 'Once
df['Like']= df['Like'].replace({'I hate it!-5': -5, 'I love it!+5': 5,'+4': 4, '+3': 3,'+2': 2, '+1': 1})
```

After transforming, it looks like :

```
In [142]: df
```

Out[142]:

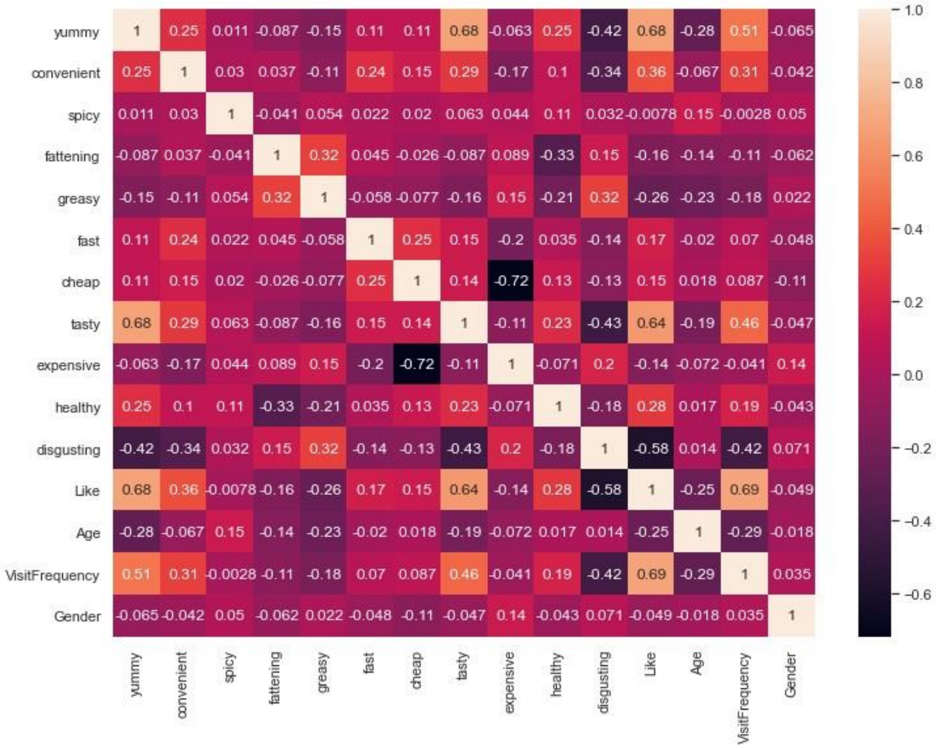
	yummy	convenient	spicy	fattening	greasy	fast	cheap	tasty	expensive	healthy	disgusting	Like	Age	VisitFrequency	Gender
0	0	1	0	1	0	1	1	0	1	0	0	-3	61	2	0
1	1	1	0	1	1	1	1	1	1	0	0	2	51	2	0
2	0	1	1	1	1	1	0	1	1	1	0	1	62	2	0
3	1	1	0	1	1	1	1	1	0	0	1	4	69	4	0
4	0	1	0	1	1	1	1	0	0	1	0	2	49	3	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1448	0	1	0	1	1	0	0	0	1	0	1	-5	47	1	1
1449	1	1	0	1	0	0	1	1	0	1	0	2	36	4	0
1450	1	1	0	1	0	1	0	1	1	0	0	3	52	3	0
1451	1	1	0	0	0	1	1	1	0	1	0	4	41	2	1
1452	0	1	0	1	1	0	0	0	1	0	1	-3	30	2	1

1431 rows × 15 columns

Analysis :

Correlation Matrix :

```
In [145]: plt.figure(figsize = (12,9))
          s=sns.heatmap (df.corr(),
                         annot = True
                         )
          #                cmap = 'Rdbu',
          #                vmain = -1,
          #                vmax = +1)
          # s.set_yticklabels(s.get_yticklabels(), rotation = 0, fontsize = 12)
          # s.set_xticklabels(s.get_xticklabels(), rotation = 90, fontsize = 12)
          # plt.title('Correlation Heatmap')
          # plt.show()
```



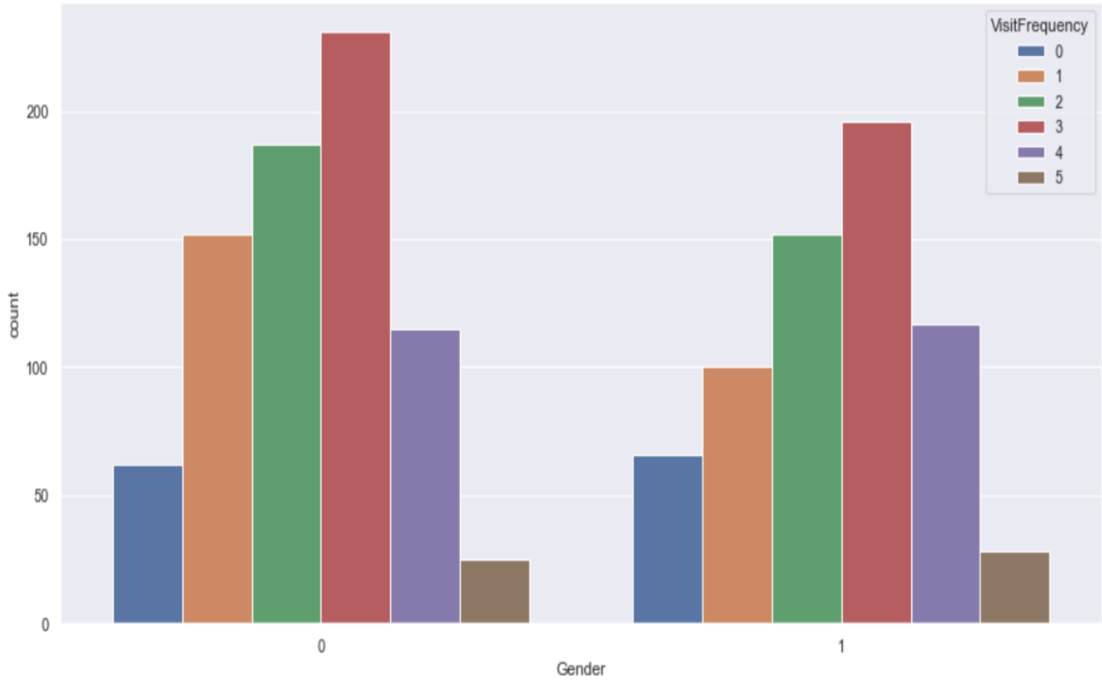
From the Correlation Matrix we see that the columns 'VisitFrequency' and 'Like' features are more likely connected as they have the correlation coefficient of 0.69.

Expensive and Cheap have correlation coefficients of -0.72, that is obvious.

Demographic Analysis :

1. **Gender**

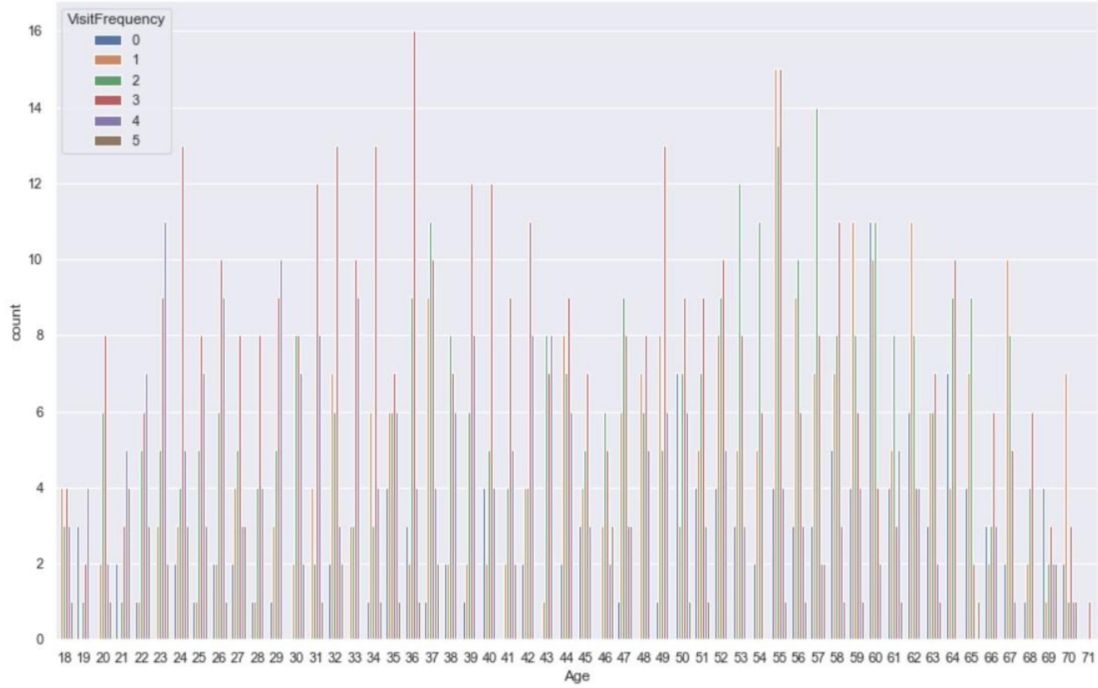
```
In [159]: plt.figure(1 , figsize = (15 , 7))
          sns.set(style="darkgrid")
          ax = sns.countplot(x= 'Gender', data=df, hue = 'VisitFrequency')
```



From the graph plot it is clear that females are more attracted to Mcdonalds. A large portion of the community, both men and women regularly visit Mcdonalds once in a month.

Age

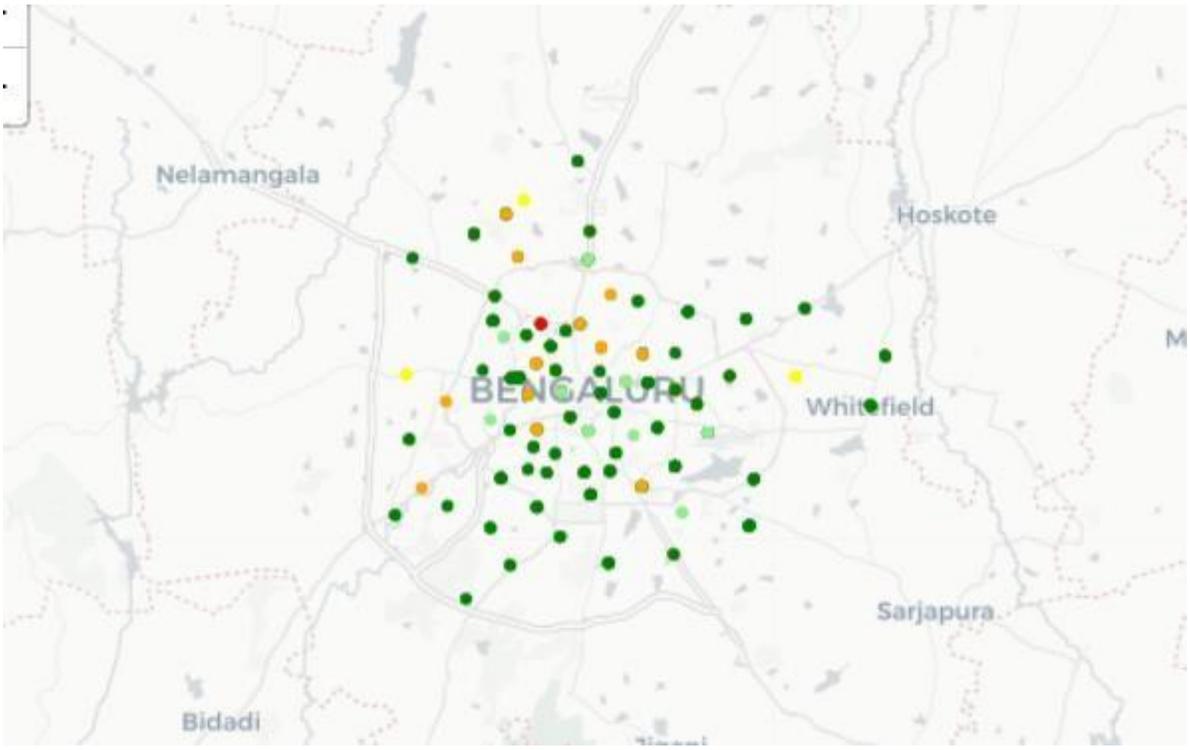
```
In [163]: plt.figure(1 , figsize = (15 , 9))
          sns.set(style="darkgrid")
          ax = sns.countplot(x= 'Age', data=df, hue = 'VisitFrequency')
```



The graph is quite messy due to the presence of 5 factors in VisitFrequency features. Though if we try to observe, we can see that aged people are likely to avoid Mcdonalds. And the optimal target age group should be 25 to 39. This is our main Target Customer.

Geographical Analysis :

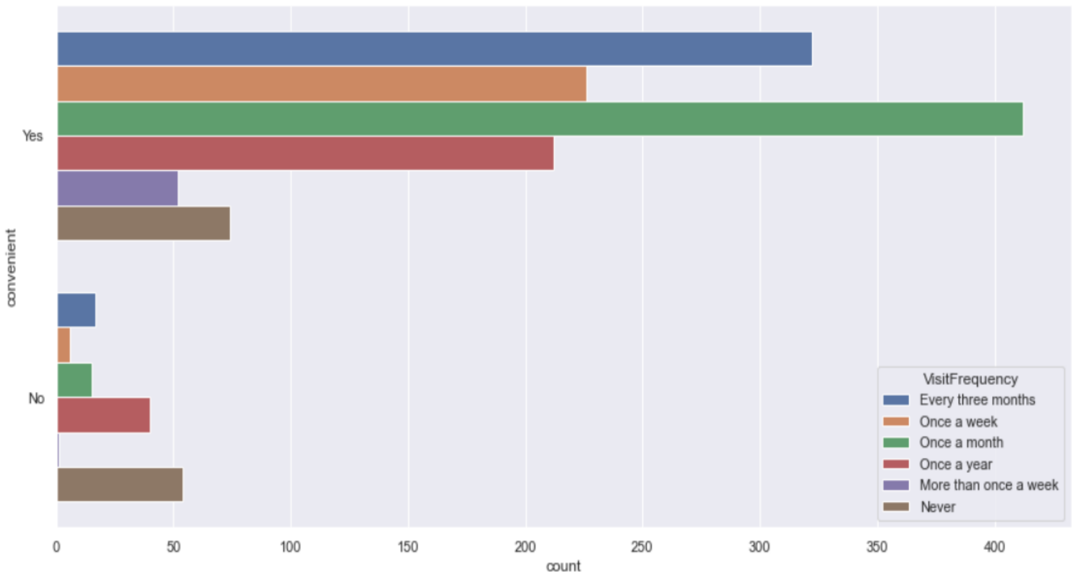
As we don’t have a good datasets to describe geographic segmentation so, we have considered the customer churn datasets where we are analysing the location with the ease and a convenience question.



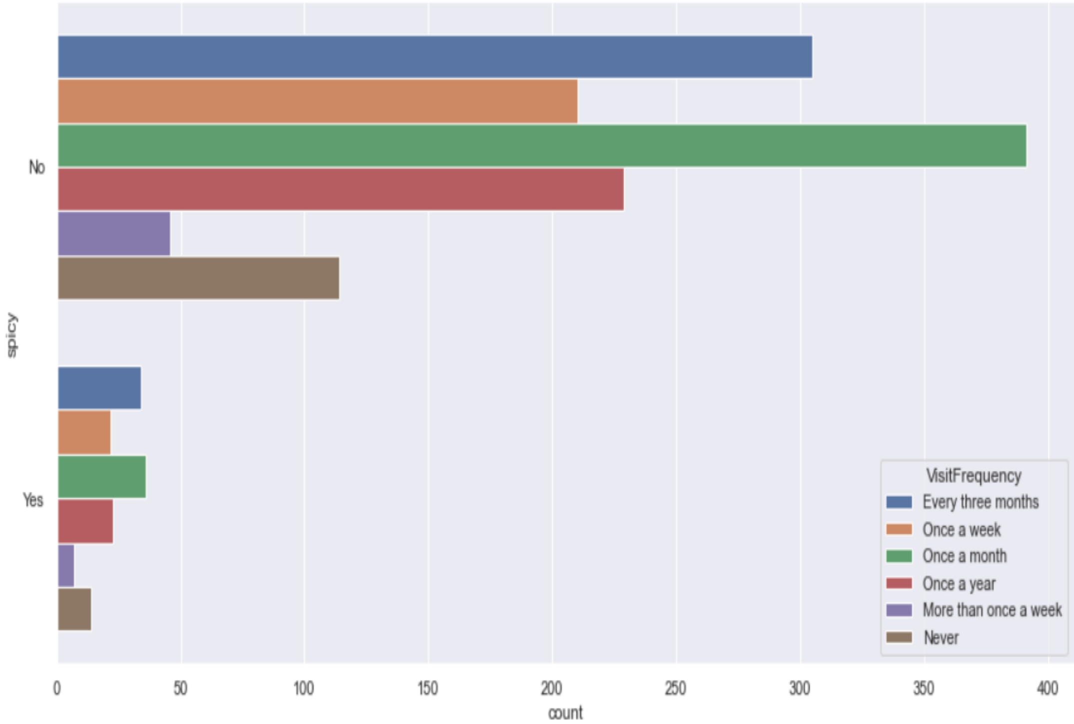
Psychographic Analysis :

Here we took a dataset and have plotted different survey questions related to Mcdonald's food survay .

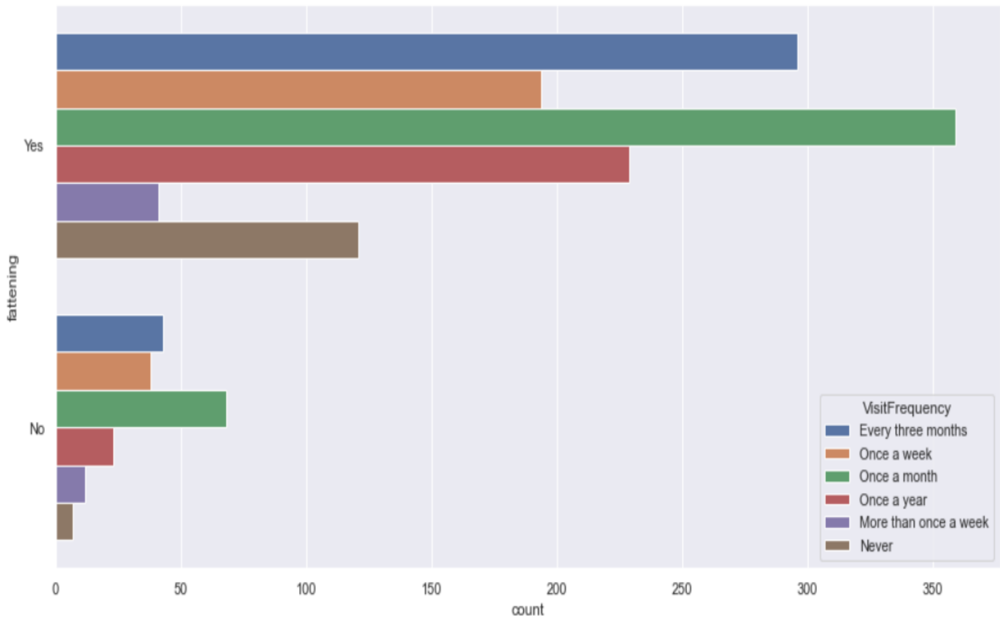
In [128]:
```
plt.figure(1 , figsize = (15 , 7))
sns.set(style="darkgrid")
ax = sns.countplot(y="convenient", data=df, hue = 'VisitFrequency' )
```



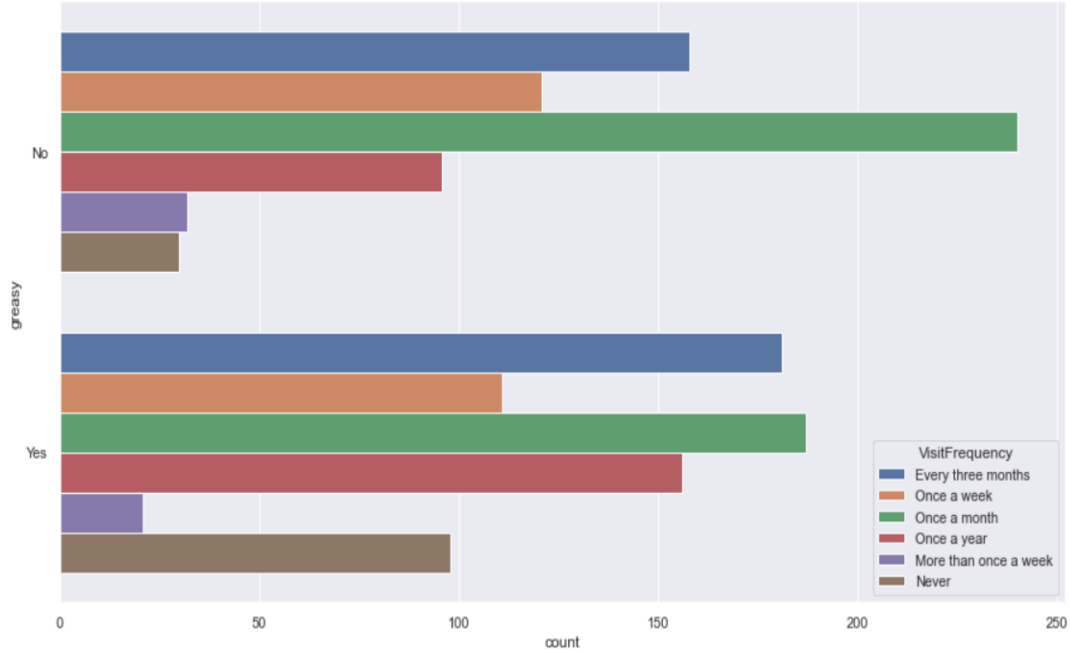
In [129]:
```
plt.figure(1 , figsize = (15 , 7))
sns.set(style="darkgrid")
ax = sns.countplot(y="spicy", data=df, hue = 'VisitFrequency')
```



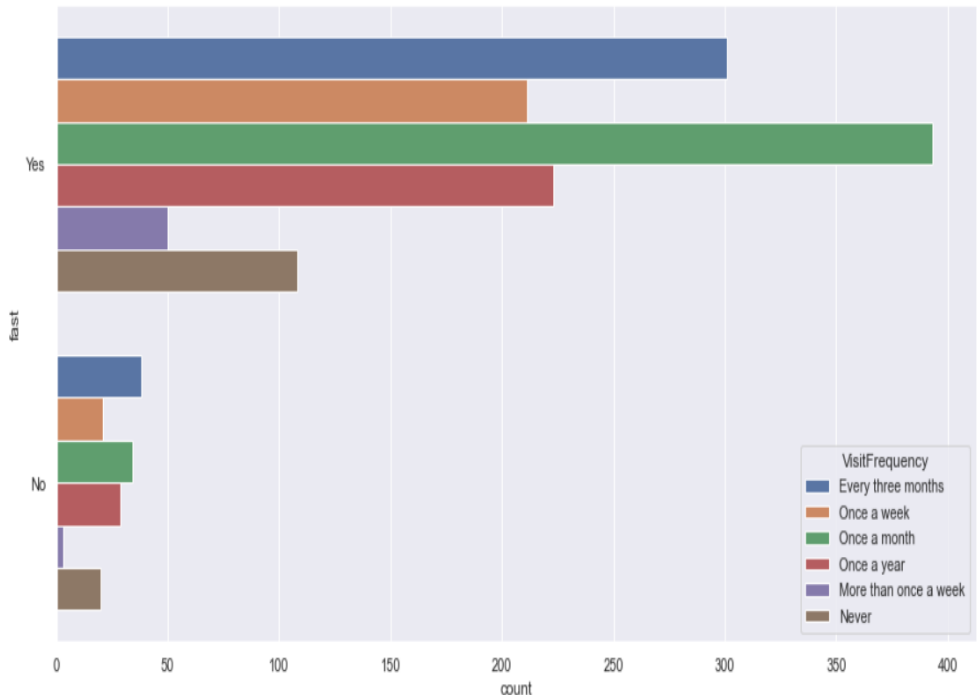
In [130]:
```
plt.figure(1 , figsize = (15 , 7))
sns.set(style="darkgrid")
ax = sns.countplot(y="fattening", data=df, hue = 'VisitFrequency' )
```



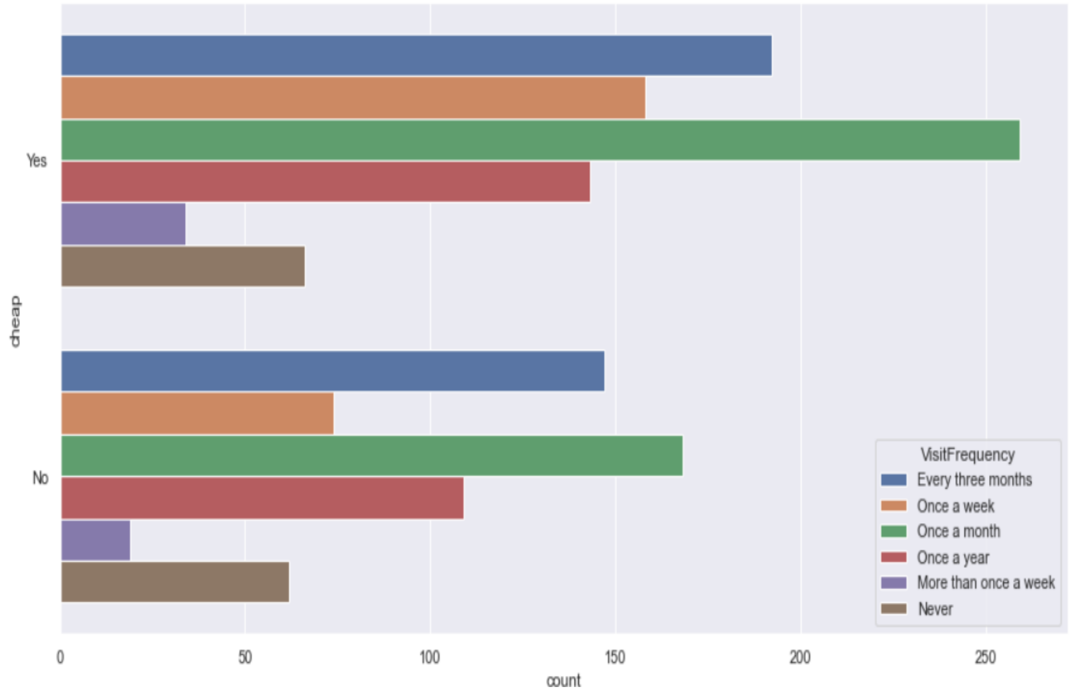
In [131]:
```
plt.figure(1 , figsize = (15 , 8))
sns.set(style="darkgrid")
ax = sns.countplot(y="greasy", data=df, hue = 'VisitFrequency' )
```



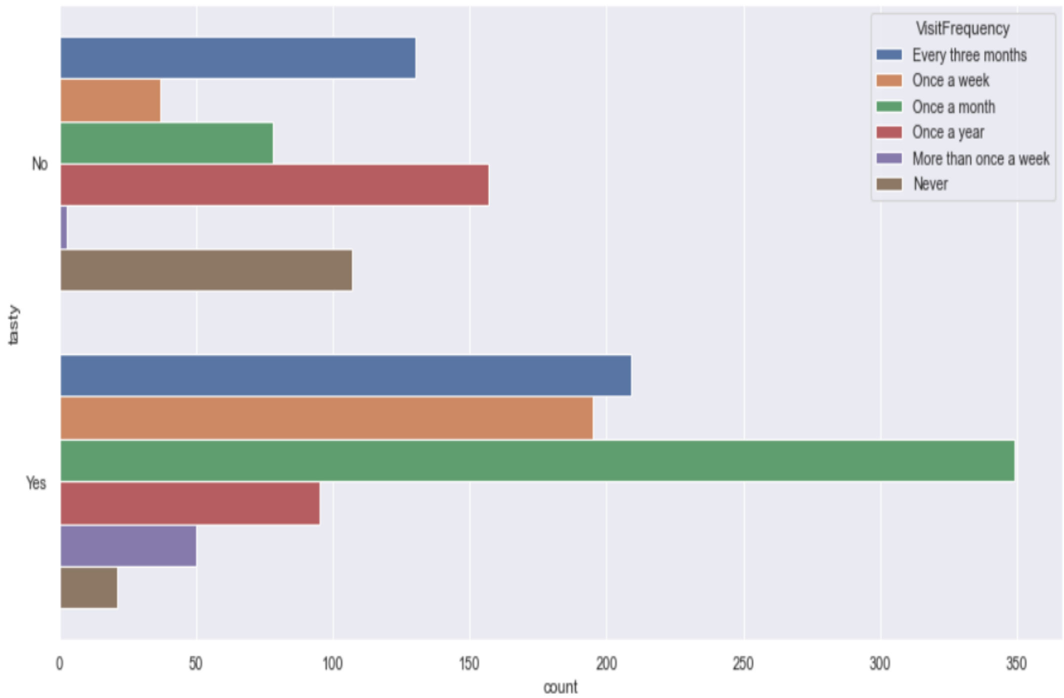
In [132]:
```
plt.figure(1 , figsize = (15 , 7))
sns.set(style="darkgrid")
ax = sns.countplot(y="fast", data=df , hue = 'VisitFrequency')
```



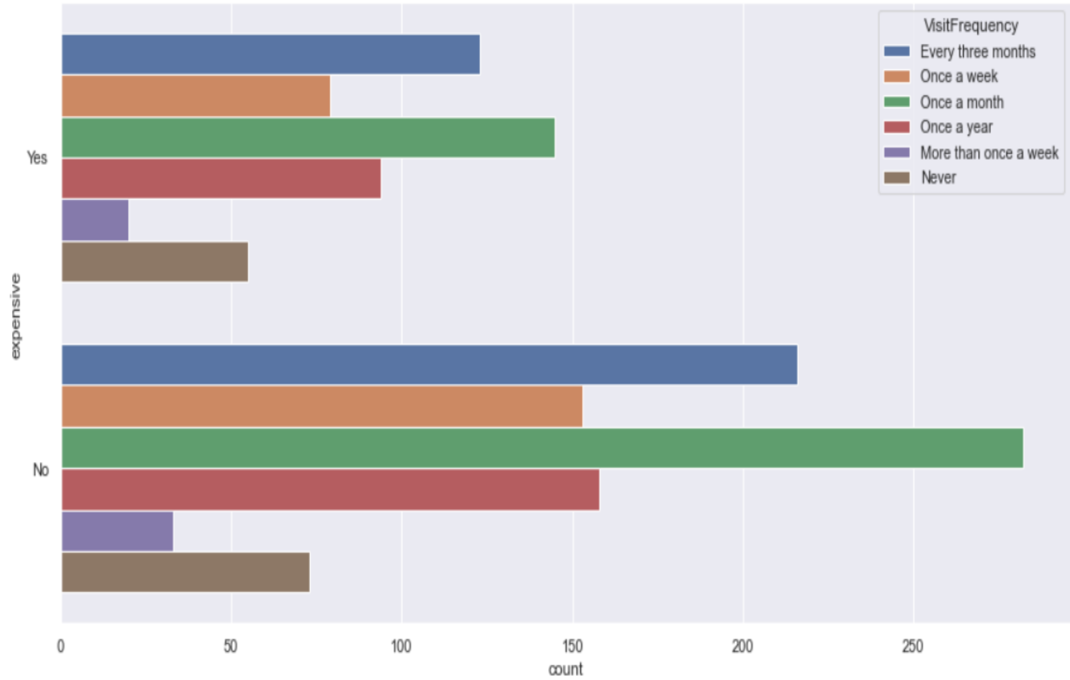
In [133]:
```
plt.figure(1 , figsize = (15 , 7))
sns.set(style="darkgrid")
ax = sns.countplot(y="cheap", data=df , hue = 'VisitFrequency')
```



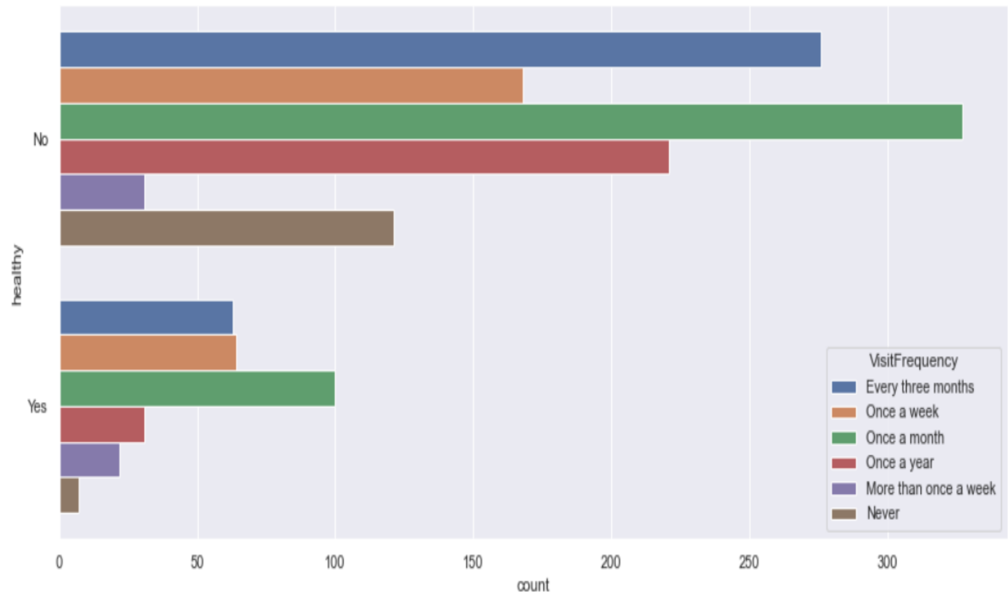
In [134]:
```
plt.figure(1 , figsize = (15 , 7))
sns.set(style="darkgrid")
ax = sns.countplot(y="tasty", data=df, hue = 'VisitFrequency' )
```



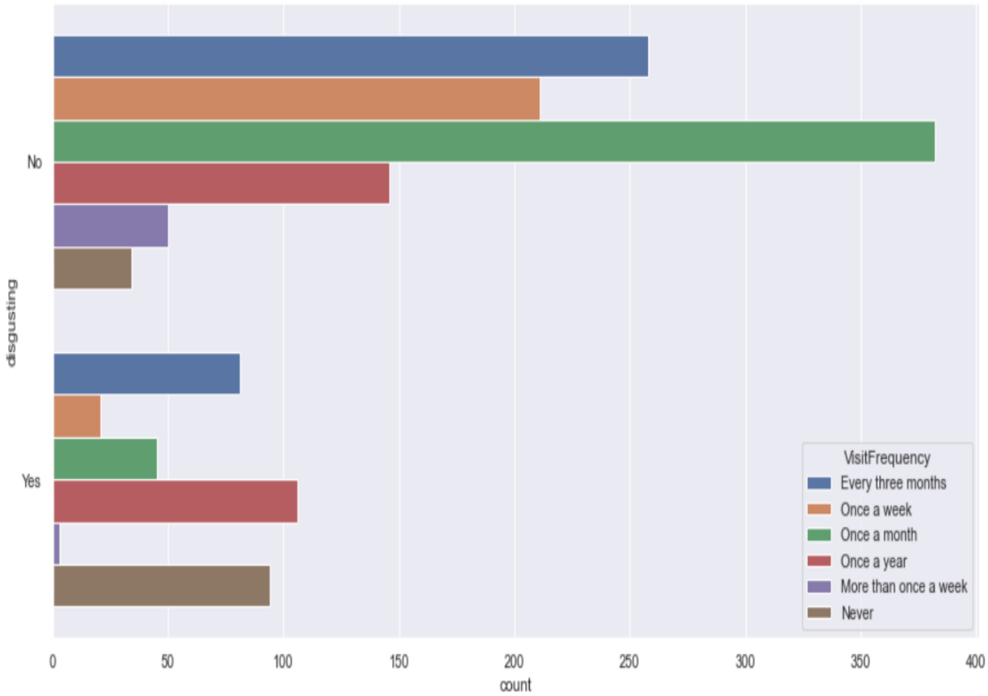
In [135]:
```
plt.figure(1 , figsize = (15 , 7))
sns.set(style="darkgrid")
ax = sns.countplot(y="expensive", data=df , hue = 'VisitFrequency')
```



In [136]:
```
plt.figure(1 , figsize = (15 , 6))
sns.set(style="darkgrid")
ax = sns.countplot(y="healthy", data=df, hue = 'VisitFrequency')
```



```
In [137]: plt.figure(1 , figsize = (15 , 7))
          sns.set(style="darkgrid")
          ax = sns.countplot(y="disgusting", data=df, hue = 'VisitFrequency' )
```



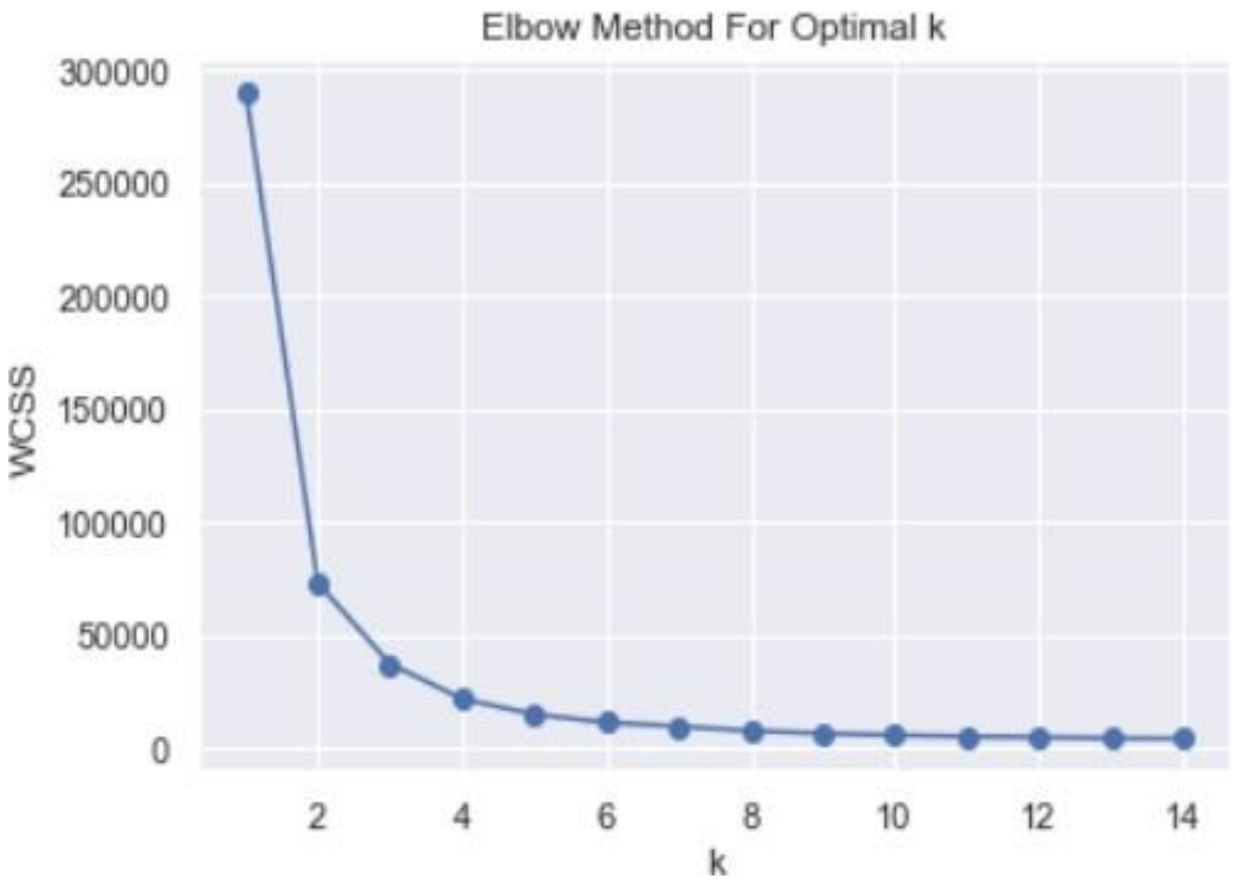
Segmentation :

Using K Means:

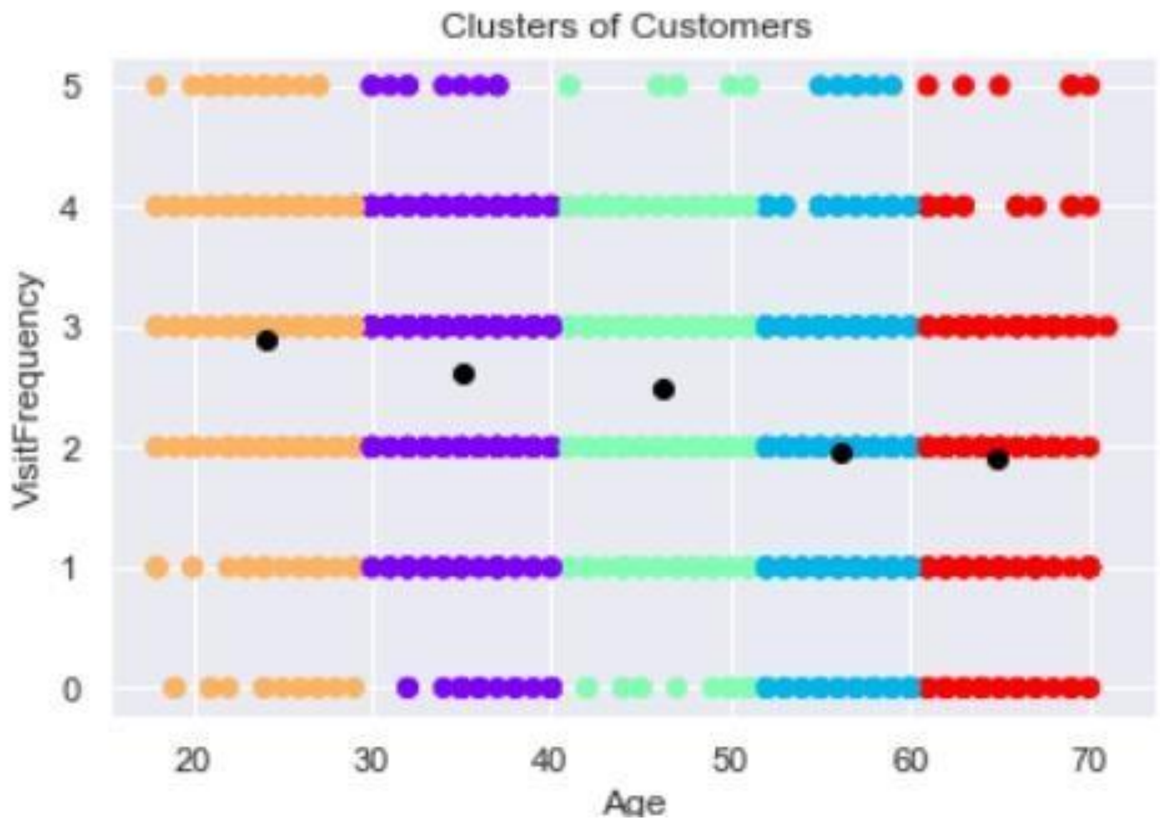
Age and Visit Frequency:

```
In [ ]: X1 = df.iloc[:, [12, 13]].values
        # selecting the columns number 12 ; Age and column number 13 ;
        # for our clustering.
```

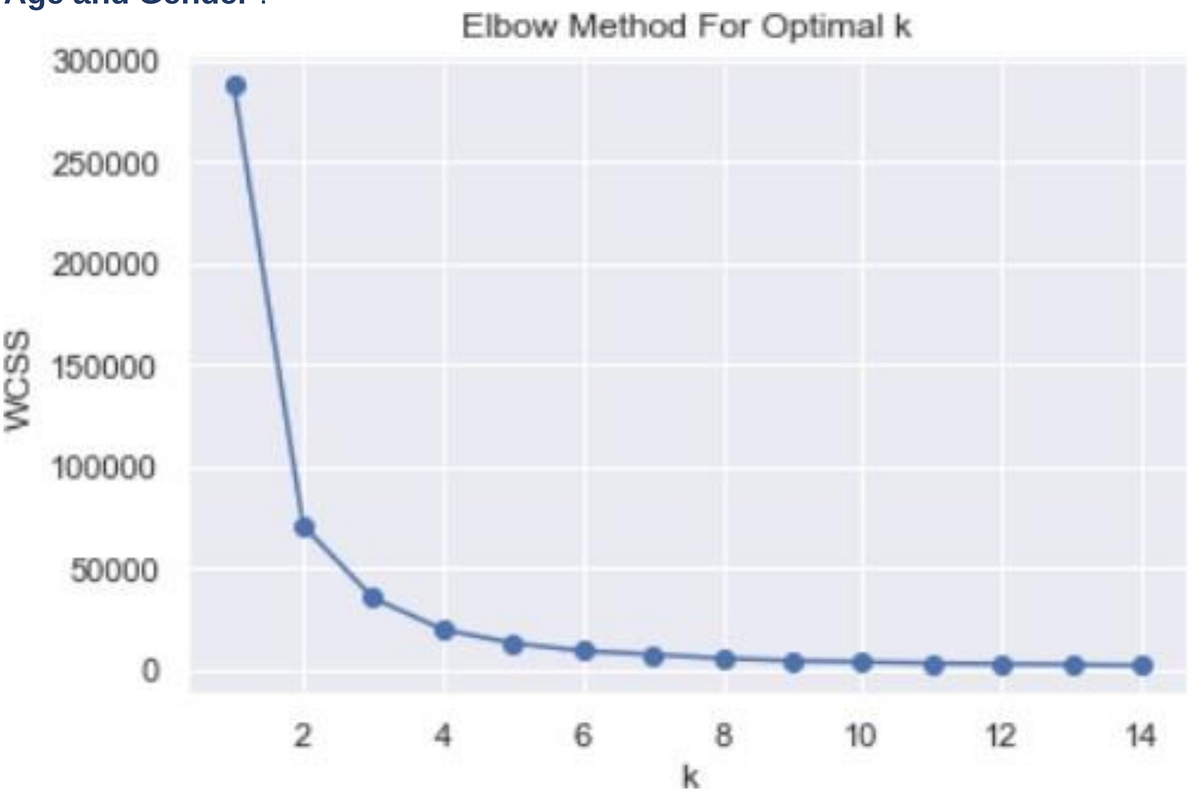
And we can use the Elbow method to find the optimum K value. For this our plot is something like this.

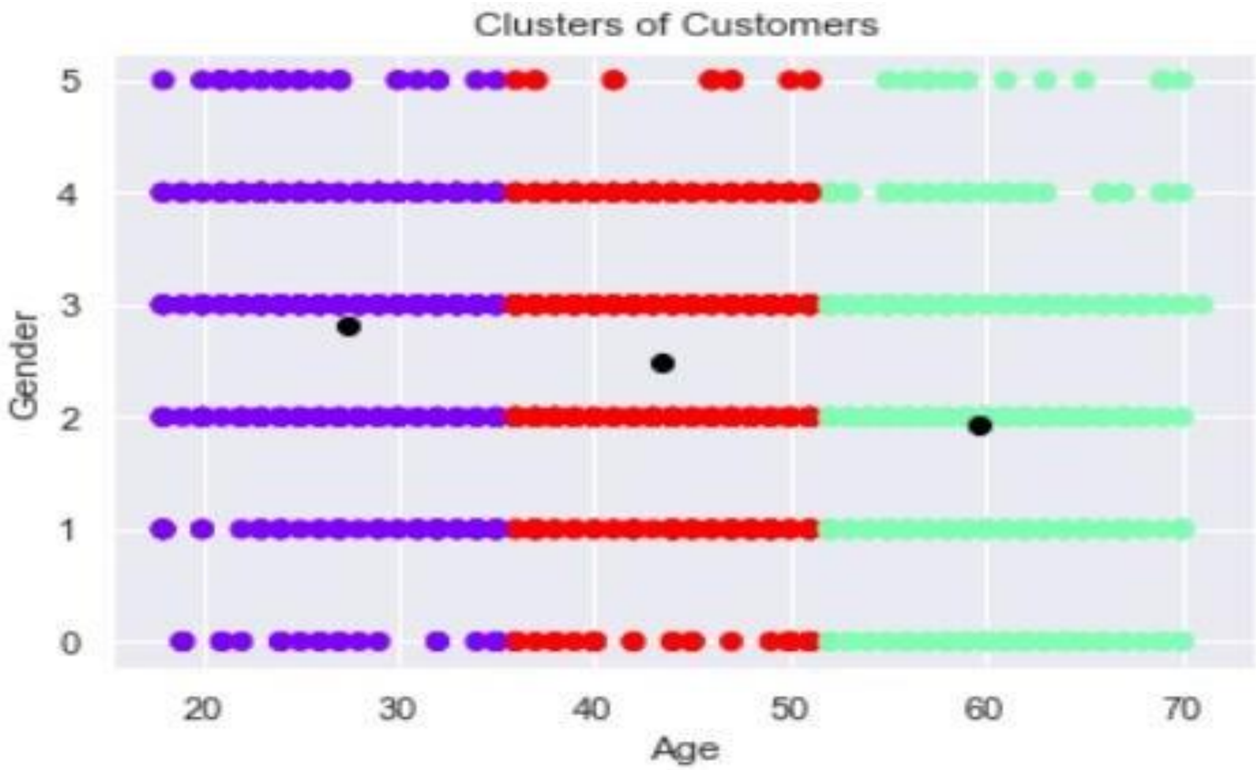


Here we can see that after the value of k is 5 the slope of the curve is increasing rapidly. So, we assume the optimal value of K is 5.



Age and Gender :

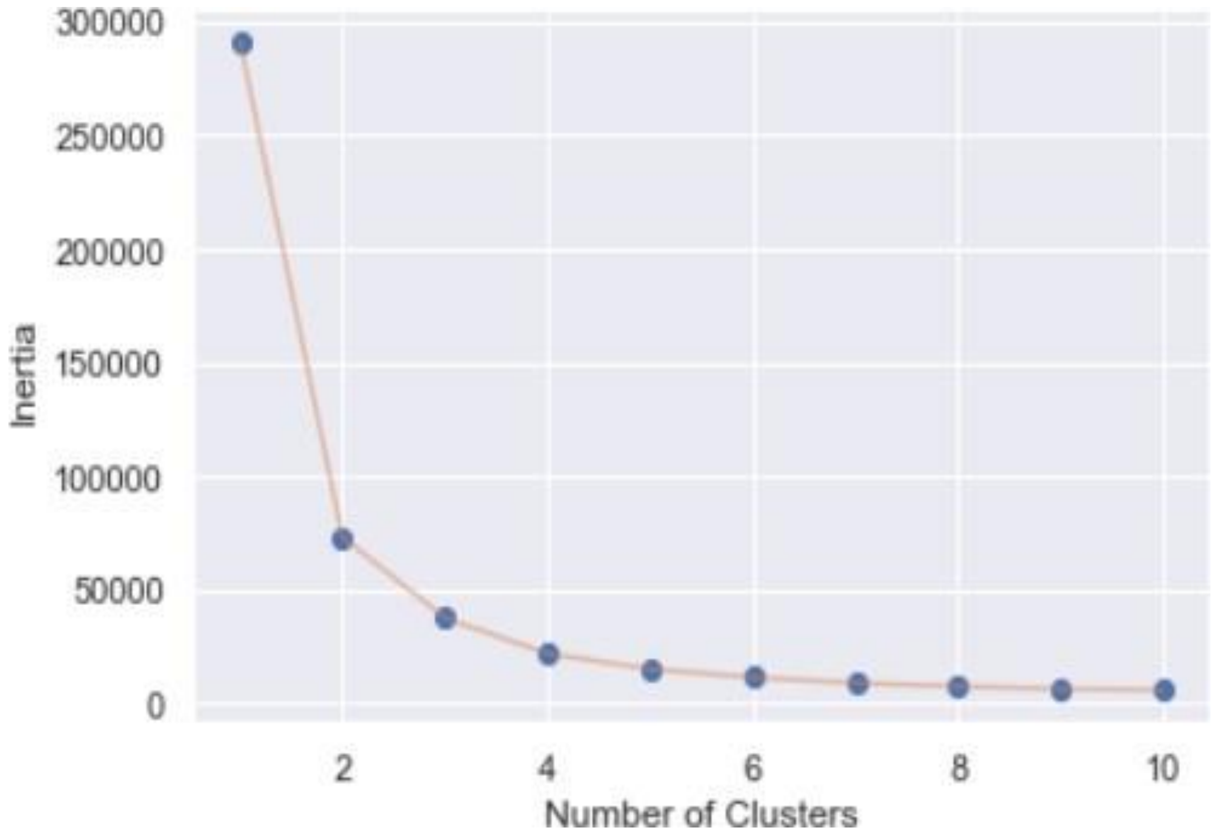




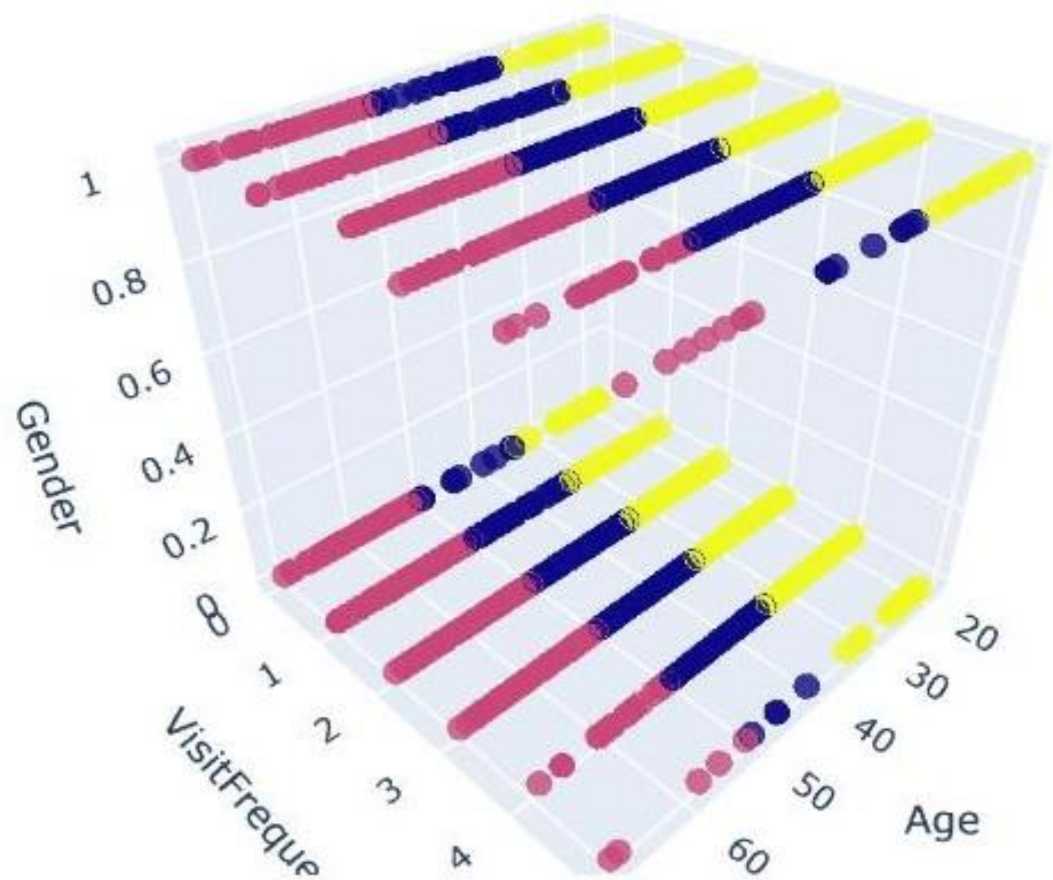
That helps a lot also.

Age, Gender and VisitFrequency :

Now considering 3 features we plot the elbow curve.



300000
250000
200000
150000
100000
50000
0
Inertia
2
4
6
8
10
Number of Clusters



Gender
1
0.8
0.6
0.4
0.2
0
VisitFreque
1
2
3
4
Age
20
30
40
50
60

Here we can see 3 clusters based on **ages, Visit frequencies and Gender** .

<u>Target segment</u> :

So from the analysis we can see that the optimum customer base or the targeted customers should be of age bracket [20-37] with visit frequency once at a time in a month.

Code:-

https://github.com/vijay2020pc/Feynn-Projects/blob/main/Task%202/Market_segmentation.ipynb