Market Segmentation

Market Segmentation through Machine Learning

INDEX

Team Information

Team-Utkarsh

Team Members:

1.Utkarsh Joshi

2.Rishabh Kumar Singh

3.Madhusmita

4.Shristi Gupta

Work Segmentation:

- Deciding not to segment and Segment Evaluation
  Utkarsh Joshi
  Rishabh Kumar Singh
  Madhusmita Rout
- Collecting Data and Exploring Data
  Madusmita Rout
- Profiling Segments and Customizing Marketing Mix
  Utkarsh Joshi
- Describing Segments and Selecting Target Segments
  Rishabh Kumar Singh

Case Study: Practice Project:

"Market Segmentation is a marriage, not a date"

(Step-1),Utkarsh Joshi

Market segmentation, a key marketing strategy needs to be executed with caution. It can't be implemented in a day or two. It needs devoted time, to draft such a strategy. Once marketing strategy is drafted and executed, the organization needs to stick to it. It should not be changed once put into effect. Segmenting a market is not free, it has cost associated with it. So, it is recommended not to segment unless expected increase in sale is sufficient to justify implementing a segmentation strategy.

One of the main areas of focus when it comes to implementing market segmentation are implication barriers. The barriers include:

1.**Senior Management** plays a key role in implementing marketing strategy. Chief executive should understand need for segmentation review, understand process and show an active interest in it, then only market segmentation can be implemented successfully.

2.**Organizational Culture** also determines to a great extent that how well the strategy drafted will be executed. If there is lack of consumer orientation, resistance for new ideas and lack of creative thinking then it is almost impossible for the organization to progress.

To sum it up, Market Segmentation is a practice of dividing your target market into approachable groups. Market segmentation creates subsets of a market based on demographics, needs, priorities and common interests.

There are significant advantages of implementing market segmentation. According to reports companies who implement market segmentation found it crucial for growing profits.

Some benefits include :

1.help defining audience charactersitcs

2.developing effective marketing strategies

3.attracting right customers

4.enhanced profits

5.product development

Step 1: Madhusmita Rout

Deciding (not) to Segment:

Implications of Committing to Market Segmentation:

Market segmentation has developed to be a key marketing strategy applied in many organisations, it is not always the best decision to pursue such a strategy. Before investing time and resources in a market segmentation analysis, it is important to understand the implications of pursuing a market segmentation strategy.

The key implication is that the organisation needs to commit to the segmentation strategy on the long term. For Segmenting a market these all things are necessary:

- Costs of performing the research,

- Fielding surveys, and focus groups,

- Designing multiple packages,

- Designing multiple advertisements

- Communication messages

It's better not to segment unless the expected increase in sales is sufficient to justify implementing a segmentation strategy, stating that One of the truisms of segmentation strategy is that using the scheme has to be more profitable than marketing without it, net of the expense of developing and using the scheme itself.

Strategic business units in charge of segments offer a suitable organisational structure to ensure ongoing focus on the (changing) needs of market segments. Because of the major implications of such a long-term organisational commitment, the decision to investigate the potential of a market segmentation strategy must be made at the highest executive level, and must be systematically and continuously communicated and reinforced at all organisational levels and across all organisational units.

Implementation Barriers:

The first group of barriers relates to senior management:

- Lack of leadership,

- pro-active championing,

- commitment and involvement in the market segmentation process by senior leadership undermines the success of market segmentation.

A second group of barriers relates to organisational culture:

- Lack of market or consumer orientation,

- resistance to change and new ideas,

- lack of creative thinking,

- bad communication

- lack of sharing of information and insights across organisational units,

- short-term thinking,

- unwillingness to make changes

- office politics have been identified as preventing the successful implementation of market segmentation.

Third group of barriers is lack of training:

- If senior management and the team tasked with segmentation do not understand the very foundations of market segmentation, or if they are unaware of the consequences of pursuing such a strategy, the attempt of introducing market segmentation is likely to fail.

- Another obstacle may be objective restrictions faced by the organisation, including lack of financial resources, or the inability to make the structural changes required.

Most of these barriers can be identified from the outset of a market segmentation study, and then proactively removed. If barriers cannot be removed, the option of abandoning the attempt of exploring market segmentation as a potential future strategy should be seriously considered.

MARKET SEGMENTATION(Step-1):

Rishabh Kumar Singh

DEFINITIONS OF MARKET SEGMENTATION

Market segmentation is a marketing term that refers to aggregating prospective buyers into groups or segments with common needs and who respond similarly to a marketing action.

KEY BENEFITS OF MARKET SEGMENTATION

Market segmentation seeks to identify targeted groups of consumers to tailor products and branding in a way that is attractive to the group.

Markets can be segmented in several ways such as geographically, demographically, or behaviorally.

Market segmentation helps companies minimize risk by determining which products are the most likely to earn a share of a target market and the best ways to market and deliver those products.

With risk minimized and clarity about the marketing and delivery of a product heightened, a company can then focus its resources on efforts likely to be the most profitable.

STRATEGIC AND TACTICAL MARKETING

Marketing strategy defines priority markets, audiences, and products. This is where Segmentation, Targeting, and Positioning fit.

Tactics will involve devising the best way to communicate these to audiences but typically won't involve a strategic review of which are the best audiences or product/market fit

THE COSTS OF MARKET SEGMENTATION

Creative and effective market segmentation can lead to the development of popular new products, but unsuccessful segmentation can cost a great deal of money and still not yield the desired results.

"User input is the key for Market segmentation analysis"(Step-2),Utkarsh Joshi

Market segmentation deals with study of customers, divided into smaller groups, to understand their specific characteristics like behaviour, age, income, and personality. It's easier for companies to advertise when they are marketing a smaller segment of customers; this way, each campaign can be highly targeted and precise to the characteristics of each group.

After having drafted the market segmentation strategy, the organization needs to invest in its analysis. Two common strategies used include: Knock Out Criteria and Attractiveness Criteria

Knock Out Criteria are used to determine if market segments resulting from market segmentation analysis qualify to be accessed using segment attractiveness criteria. The criteria that come under this category include substantiality, measurability and accessibility. Knock out criteria needs to be understood by senior management.

Attractiveness Criteria are used to evaluate the relative attractiveness of marketing segments. Attractiveness Criteria are not binary in nature. Segments need not comply with a particular attractiveness criteria. A segment can be more or less attractive with respect a particular criteria. The attractiveness across all criteria determines whether a market segment is selected as a target segment.

Segment Evaluation Plot is a structured approach for evaluating market segments. In this approach segment attractiveness is shown along one axis while Organizational Competitiveness is shown along the other. The values for segment attractiveness and organizational

competitiveness is decided by segmentation team. Thus we need to have experts to in segmentation team so as to get correct predictions.

Step 2: Madhusmita Rout

Specifying the Ideal Target Segment:

Segment Evaluation Criteria:

It 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.

The user needs to be involved in most stages such as:

- The organisation has to make a major contribution to market segmentation Analysis.

- While this contribution is conceptual in nature, it guides many of the following steps, most critically data collection and selecting one or more target segments.

- 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.

Knock-Out Criteria:

Knock-out criteria are used to determine if market segments resulting from the market segmentation analysis qualifies 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.

Knock-out criteria must be understood by senior management, the segmentation team, and the advisory committee.

Attractiveness Criteria: when deciding which attractiveness criteria are most useful to their specific situation segmentation team has to consider

- 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.

Implementing a Structured Process:

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.

In this process segment attractiveness and organizational competitiveness is necessary because there is no standard set of criteria that could be used by all organization. To achieve this there are many criteria has to be investigated so that we can decide which one is important for organization, such as:

- This task should be completed by a team of people, if a core team of two to three people is primarily in charge of market segmentation analysis, this team could propose an initial solution and report their choices to the advisory committee – which consists of representatives of all organisational units – for discussion and possible modification.

- There are at least two good reasons to include in this process representatives

from a wide range of organisational units.

- First, each organisational unit has a different perspective on the business of the organisation. As a consequence, members of these units bring different positions to the deliberations.

- Secondly, if the segmentation strategy is implemented; it will affect every single unit of the organisation. Consequently, all units are key stakeholders of market segmentation

Analysis

At the end of this step, the market segmentation team should have a list of approximately six segment attractiveness criteria. Each of these criteria should have a weight attached to it to indicate how important it is to the organisation compared to the other criteria. These allocations then have to be negotiated until agreement is reached.

MARGET SEGMENTATION ANALYSIS(Step-2):

Rishabh Kumar Singh

Market segmentation analysis is the study of customers, divided into smaller groups, to understand their specific characteristics like behavior, age, income, and personality.

Segmentation divides the data into comparatively homogenous groups, marketing efforts such as targeting, positioning, retention, and product development can be more efficiently performed.

❖ **THE THREE METHODS WE WILL CONSIDER ARE:**

- Cluster Analysis

- Neural Networks

- Mixture Models

Cluster Analysis

Cluster analysis is the traditional method used for segmentation in marketing research. This is a family of methods that subsumes many variations and can be broadly classified under two distinct groups: hierarchical and non-hierarchical (or partitioning) methods.

Neural Networks

There are two basic types of neural networks: supervised learning and unsupervised learning networks.

Supervised learning networks can be used in place of traditional methods like regression and discriminant analysis and were discussed in the previous article in this series.

Unsupervised learning networks are generally used when there are no clear distinctions between dependent and independent variables in the data and when pattern or structure recognition is required.

Mixture Models

 The basic idea linking methods in this category is that the data contain many distributions or segments which are mixed. The task of the analysis then becomes one of unmixing the distributions and for this reason, they are also called unmixing models.

7 STEPS TO MARKET SEGMENTATION

1. Step 1 – Define your market. …

2. Step 2 – Analyse existing customers. …

3. Step 3 – Create buyer persona(s) …

4. Step 4 – Compare and identify gaps, groups, and opportunities. …

5. Step 5 – Define and name segments. …

6. Step 6 – Research segments separately. …

7. Step 7 – Test and optimize.

Step 3: Collecting Data: Madhusmita Rout

Segmentation Variables:

The characteristics of people that are used to determine if the people are similar are called segmentation variables. For example, if segmenting a market is based on the age of people, then age is the segmentation variable.

When people refer to segmentation variables, they are usually referring to one of the following:

- A single variable that is used to allocate people to segments

- A set of variables where people are used to allocate people to segments based on some logical relationship

- A set of variables that are used in a predictive statistical algorithm to predict segmentation membership

- A set of variables that are used in a segmentation algorithm

- A variable in a data file or database that records segment membership

Segmentation Criteria

The term segmentation criterion is used here in a broader sense than the term segmentation variable. The term segmentation variable refers to one measured value, for example, one item in a survey, or one observed expenditure category. The term segmentation criterion relates to the nature of the information used for market segmentation.

The decision which segmentation criterion to use cannot easily be outsourced to either a consultant or a data analyst because it requires prior knowledge about the market. The most common segmentation criteria are geographic, socio-demographic, psychographic and behavioral.

Geographic Segmentation:

- Geographic information is seen as the original segmentation criterion used for the purpose of market segmentation. when geographic segmentation is used – the consumer's location of residence serves as the only criterion to form market segments.

- The key advantage of geographic segmentation is that each consumer can easily be assigned to a geographic unit. As a consequence, it is easy to target communication messages, and select communication channels (such as local newspapers, local radio and TV stations) to reach the selected geographic segments.

- The typical case is best illustrated using tourism: people from the same country of origin are likely to have a wide range of different ideal holidays, depending on whether they are single or travel as a family, whether they are into sports or culture. Despite the potential shortcomings of using geographic information as the segmentation variable, the location aspect has experienced a revival in international market segmentation studies aiming to extract market segments across geographic boundaries. Such an approach is challenging because the segmentation variable must be meaningful across all the included geographic regions, and because of the known biases that can occur if surveys are completed by respondents from different cultural backgrounds.

Socio-Demographic Segmentation:

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

- For example: luxury goods (associated with high income), cosmetics (associated with gender; even in times where men are targeted, the female and male segments are treated distinctly differently).

- Demographics explain about 5% of the variance in consumer behavior.

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.

- As a consequence, most psychographic segmentation studies use a number of segmentation variables, for example: a number of different travel motives, a number of perceived risks when going on vacation.

- The psychographic approach has the advantage that it is generally more reflective of the underlying reasons for differences in consumer behavior.

Behavioural Segmentation:

- Another approach to segment extraction is to search directly for similarities in behaviour or reported behaviour. 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 behaviour.

- The key advantage of behavioural approaches is that – if based on actual behaviour rather than stated behaviour or stated intended behaviour – the very behaviour of interest is used as the basis of segment extraction.

- But behavioural data is not always readily available, especially if the aim is to include in the segmentation analysis potential customers who have not previously purchased the product, rather than limiting oneself to the study of existing customers of the organisation.

Data from Survey Studies:

Most market segmentation analyses are based on survey data. Survey data is cheap

and easy to collect, making it a feasible approach for any organisation. But survey

data – as opposed to data obtained from observing actual behaviour – 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. A few key

aspects that need to be considered when using survey data are discussed below.

- **Choice of Variables:** Carefully selecting the variables that are included as segmentation variable in commonsense segmentation, or as segmentation variables in data-driven segmentation, is critical to the quality of the market segmentation solution.

- In data-driven segmentation, all variables relevant to the construct captured by the segmentation criterion need to be included. At the same time, unnecessary variables must be avoided.

- Including unnecessary variables can make questionnaires long and tedious for respondents, which, in turn, causes respondent fatigue.

- Unnecessary variables included as segmentation variables divert the attention of the segment extraction algorithm away from information critical to the extraction of optimal market segments. Such variables are referred to as noisy variables or masking variables and have been repeatedly shown to prevent algorithms from identifying the correct segmentation solution.

- **Response Options:** Answer options provided to respondents in surveys determine the scale of the data available for subsequent analyses. Because many data analytic techniques are based on distance measures, not all survey response options are equally suitable for segmentation analysis.

- Options allowing respondents to answer in only one of two ways, generate binary or dichotomous data. Such responses can be represented in a data set by 0s and 1s.

The distance between 0 and 1 is clearly defined and, as such, poses no difficulties for subsequent segmentation analysis.

- Options allowing respondents to select an answer from a range of unordered categories correspond to nominal variables. If asked about their occupation, respondents can select only one option from a list of unordered options. Nominal variables can be transformed into binary data by introducing a binary variable for each of the answer options.

- Options allowing respondents to indicate a number, such as age or nights stayed at a hotel, generate metric data.

- The most commonly used response option in survey research, however, is a limited number of ordered answer options larger than two.

- **Response Styles:** A wide range of response styles manifest in survey answers, including respondents' tendencies to use extreme answer options (STRONGLY AGREE, STRONGLY DISAGREE), to use the midpoint (NEITHER AGREE NOR DISAGREE), and to agree with all statements.

- Response styles affect segmentation results because commonly used segment extraction algorithms cannot differentiate between a data entry reflecting the

respondent's belief from a data entry reflecting both a respondent's belief and a response style.

- **Sample Size:** The importance of having a sample size sufficiently large to enable an algorithm to extract the correct segments (if segments naturally exist in the data). Ensure the data contains at least 100 respondents for each segmentation variable.

- The importance of collecting high-quality unbiased data as the basis for market segmentation analysis.

- It can be concluded from the body of work studying the effects of survey data quality on the quality of market segmentation results based on such data that, optimally, data used in market segmentation analyses should:

- contain all necessary items; contain no unnecessary items; contain no correlated items; contain high-quality responses; be binary or metric; be free of response styles; include responses from a suitable sample given the aim of the segmentation study; and include a sufficient sample size given the number of segmentation variables.

⚜ **Data from Internal Sources:**
- 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 behaviour of consumers.

- ### Data from Experimental Studies:

  - Experimental data can result from field or laboratory experiments. For example, they can be the result of tests how people respond to certain advertisements. The response to the advertisement could then be used as a segmentation criterion.

  - Experimental data can also result from choice experiments or conjoint analyses. The aim of such studies is to present consumers with carefully developed stimuli consisting of specific levels of specific product attributes.

Step 4: Exploring Data: Madhusmita Rout

A First Glimpse at the Data:

- After collecting the data, exploratory data analysis cleans and – if necessary – preprocesses the data.

- Data exploration helps to

  (1) identify the measurement levels of the variables;

  (2) investigate the univariate distributions of each of the variables;

  (3) assess dependency structures between variables.

- Data may need to be pre-processed and prepared so it can be used as input for different segmentation algorithms.

- Results from the data exploration stage provide insights into the suitability of different segmentation methods for extracting market segments.

Data Cleaning:

- Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

- This includes checking if all values have been recorded correctly, and if consistent labels for the levels of categorical variables have been used.

- For example, gender typically has two values in surveys: female and male.

- Data cleaning is not simply about erasing information to make space for new data, but rather finding a way to

maximize a data set's accuracy without necessarily deleting information.

🔸 **Descriptive Analysis:**

- Descriptive Analysis is the type of analysis of data that helps describe, show or summarize data points in a constructive way such that patterns might emerge that fulfill every condition of the data.

- It is one of the most important steps for conducting statistical data analysis.

- In R, we obtain a numeric summary of the data with command summary (). This command returns the range, the quartiles, and the mean for numeric variables.

- For categorical variables, the command returns frequency counts. The command also returns the number of missing values for each variable.

🔸 **Pre-Processing:**

- **Categorical Variables:** Two pre-processing procedures are often used for categorical variables. One is merging levels of categorical variables before further analysis, the other one is converting categorical variables to numeric ones, if it makes sense to do so.

- Examples of categorical variables are race, sex, age group, and educational level. While the latter two variables may also be considered in a numerical manner by using exact values for age and highest grade completed, it is often more informative to categorize such variables into a relatively small number of groups.

- **Numeric Variables**: A numerical variable is a data variable that takes on any value within a finite or infinite interval (e.g. length, test scores, etc.). the numerical variable can also be called a continuous variable because it exhibits the features of continuous data.

- The range of values of a segmentation variable affects its relative influence in distance-based methods of segment extraction.

- To balance the influence of segmentation variables on segmentation results, variables can be standardised. Standardising variables means transforming them in a way that puts them on a common scale.

- If the data contains observations located very far away from most of the data (outliers). In such situations, robust estimates for location and spread – such as the median and the inter quartile range are preferable.

- 🔹 **Principal Components Analysis:**
  - Principal components analysis (PCA) transforms a multivariate data set containing metric variables to a new data set with variables – referred to as principal

Components, which are uncorrelated and ordered by importance.

  - Principal components analysis basically keeps the data space unchanged, but looks at it from a different angle.

  - Principal components analysis works off the covariance or correlation matrix of several numeric variables.

- If all variables are measured on the same scale, and have similar data ranges, it is not important which one to use.

- If the data ranges are different, the correlation matrix should be used (which is equivalent to standardizing the data).

- In most cases, the transformation obtained from principal components analysis is

used to project high-dimensional data into lower dimensions for plotting purposes.

"Correct interpretation, is critical to making good strategic marketing decisions"(Step-6)

By: Utkarsh Joshi

Profiling is required when data driven market segmentation is used. 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 becomes important if no natural segments exist in data and either a reproducible or constructive market segmentation approach has to be taken.

A segment profile is **a descriptive summary of the size, needs, behaviors and preferences of consumers within a particular market segment**, with the goal of evaluating segment attractiveness and developing suitable marketing strategies.

Profiling Segments in Python:

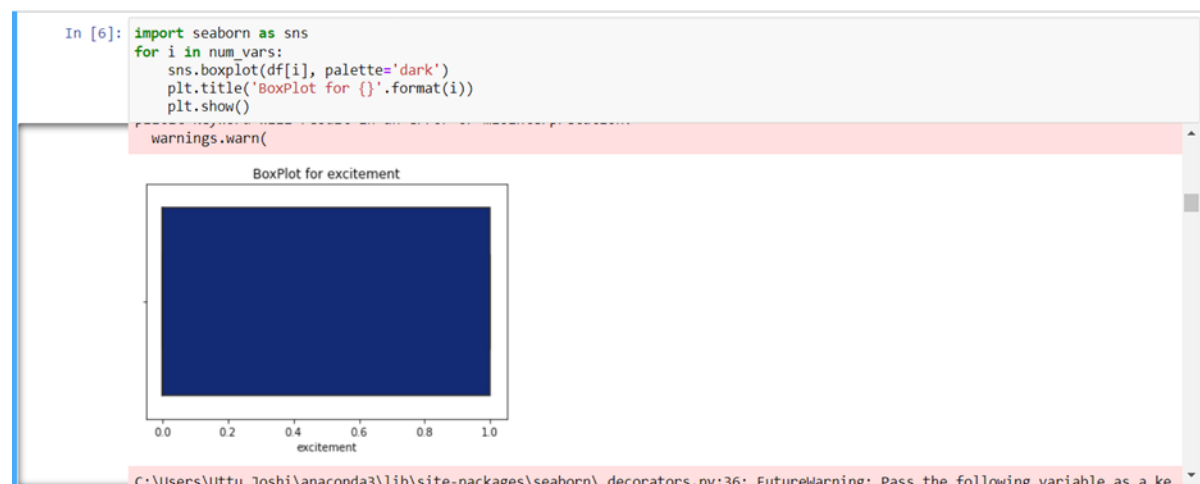1.Load The Dataset: We load the data set using pandas and display it

2. Seprate the variables: We seprate the required variables.

```
In [3]: df1 = df[['rest and relax','luxury / be spoilt','do sports' ,'excitement','not exceed planned budget','realise creativity' ,'fun
        'entertainment facilities','not care about prices','life style of the local people','intense experience of nature','cosiness/fami
        ,'cultural offers']]
        df1.head()
```

Out[3]:

	rest and relax	luxury / be spoilt	do sports	excitement	not exceed planned budget	realise creativity	fun and entertainment	good company	health and beauty	free- and- easy- going	entertainment facilities	not care about prices	life style of the local people	intense experience of nature	cosiness/familiar atmosphere	s
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	1	1	0	0	0	1	0	0	1	0	0	1	0	0	0	
2	1	0	0	0	0	0	0	0	0	0	1	1	0	1	0	
3	1	1	0	0	0	1	1	0	0	1	0	0	1	0	0	
4	1	0	0	0	0	0	0	0	1	0	0	1	0	0	1	

3.Checking for Outliers: We check for any outlier value as outliers affect our output to a good extent.

```
In [6]: import seaborn as sns
        for i in num_vars:
            sns.boxplot(df[i], palette='dark')
            plt.title('BoxPlot for {}'.format(i))
            plt.show()
```


**Our data set has got almost no outliers.

4.Performing Clustering to segment our Data: To extract segments we perform clustering.

```
In [12]: from sklearn.cluster import KMeans, AgglomerativeClustering
         from sklearn.metrics import silhouette_score
         from scipy.cluster.hierarchy import linkage, cophenet, dendrogram
         from scipy.spatial.distance import pdist
```

5. Checking Silhouette score for KMeans and Agglomerative Clustering:

```
In [12]: wcss = []
         sil_kmeans = []
         sil_agc = []
         for i in range(3,10):
             kmeans = KMeans(n_clusters=i,n_init=100,init='k-means++',random_state=0)
             kmeans.fit(df1)
             in_km = kmeans.inertia_
             wcss.append(in_km)
             sil_km = silhouette_score(df1, kmeans.labels_)
             sil_kmeans.append(sil_km)

             # Agglomerative Clusters and its Silhouette Score
             agc = AgglomerativeClustering(n_clusters = i, affinity = 'cityblock', linkage = 'average')
             agc.fit(df1)
             sil_ag = silhouette_score(df1, agc.labels_)
             sil_agc.append(sil_ag)

             print('Number of clusters:', i)
             print('KMeans Inertia', in_km)
             print('Silhouette Score for KMeans:', sil_km)
             print('Silhouette Score for AGC(HCA):', sil_ag)
             print()
```

```
Number of clusters: 3
KMeans Inertia 2782.9459324344766
Silhouette Score for KMeans: 0.08972052075790887
Silhouette Score for AGC(HCA): 0.13306580986653116

Number of clusters: 4
KMeans Inertia 2648.1980708775227
Silhouette Score for KMeans: 0.09285564163859127
Silhouette Score for AGC(HCA): 0.1121483043010524

Number of clusters: 5
KMeans Inertia 2532.697500143457
Silhouette Score for KMeans: 0.09903848906757909
Silhouette Score for AGC(HCA): 0.08577239362776054

Number of clusters: 6
KMeans Inertia 2455.3178368470726
Silhouette Score for KMeans: 0.09385892802490287
Silhouette Score for AGC(HCA): 0.07570593366553549

Number of clusters: 7
KMeans Inertia 2392.8974073293907
Silhouette Score for KMeans: 0.08145195080312345
Silhouette Score for AGC(HCA): 0.07248804687220406

Number of clusters: 8
KMeans Inertia 2338.201178821137
Silhouette Score for KMeans: 0.0855458907289171
Silhouette Score for AGC(HCA): 0.06437573864720869

Number of clusters: 9
KMeans Inertia 2297.9270564538824
Silhouette Score for KMeans: 0.08275839975080396
Silhouette Score for AGC(HCA): 0.042338288792856955
```

*Thus we see that we have got overlapping clusters, with maximum sepration between clusters when the number of clusters is 4.

6.<u>Profiling Step:</u>

Finding how many observations or records are present in each segment

```
In [39]: df1.cluster_4.size
         df1.cluster_4.value_counts()

Out[39]: 1    296
         3    257
         2    232
         0    215
         Name: cluster_4, dtype: int64

In [40]: pd.Series.sort_index(df1.cluster_4.value_counts())

Out[40]: 0    215
         1    296
         2    232
         3    257
         Name: cluster_4, dtype: int64
```

8. Grouping-by over each cluster to find the Segment wise average for each variable

```
In [47]: df1.groupby('cluster_4').apply(np.mean).T
Out[47]:
```

cluster_4	0	1	2	3
rest and relax	0.958140	0.854730	0.913793	0.887160
luxury / be spoilt	0.376744	0.206081	0.353448	0.202335
do sports	0.241860	0.070946	0.146552	0.108949
excitement	0.530233	0.094595	0.362069	0.420233
not exceed planned budget	0.195349	0.020270	0.043103	0.077821
realise creativity	0.623256	0.040541	0.939655	0.645914
fun and entertainment	0.613953	0.192568	0.504310	0.587549
good company	0.293023	0.050676	0.103448	0.089494
health and beauty	0.786047	0.233108	0.577586	0.583658
free-and-easy-going	0.241860	0.064189	0.318966	0.171206
entertainment facilities	0.209302	0.111486	0.193966	0.225681
not care about prices	0.758140	0.114865	0.021552	1.000000
life style of the local people	0.641860	0.077703	0.077586	0.140078
intense experience of nature	0.353488	0.131757	0.185345	0.124514
cosiness/familiar atmosphere	0.902326	0.158784	0.068966	0.175097
maintain unspoilt surroundings	0.227907	0.097973	0.224138	0.120623
everything organised	0.795349	0.097973	0.146552	0.112840
unspoilt nature/natural landscape	0.502326	0.060811	0.107759	0.513619
cultural offers	0.809302	0.388514	0.728448	0.832685
cluster_4	0.000000	1.000000	2.000000	3.000000
cluster_5	1.441860	1.820946	1.103448	2.735409
cluster_6	1.418605	3.148649	4.215517	1.840467

9. Concatinating the above two averages:

```
In [49]: Profiling_output = pd.concat([df1.apply(lambda x: x.mean()).T,
                  df1.groupby('cluster_4').apply(lambda x: x.mean()).T,
                  df1.groupby('cluster_5').apply(lambda x: x.mean()).T,
                  df1.groupby('cluster_6').apply(lambda x: x.mean()).T,
                  df1.groupby('cluster_7').apply(lambda x: x.mean()).T,
                  df1.groupby('cluster_8').apply(lambda x: x.mean()).T], axis=1)
```

10.Profiling Output

	0	0	1	2	3	0	1	2	3	4	...	5	6	0	1
rest and relax	0.899	0.958140	0.854730	0.913793	0.887160	0.957219	0.912664	0.843284	0.868778	0.978947	...	0.987013	0.830645	0.917431	0.842640
luxury / be spoilt	0.276	0.376744	0.206081	0.353448	0.202335	0.085561	0.323144	0.231343	0.176471	0.894737	...	0.948052	0.177419	0.229358	0.223350
do sports	0.135	0.241860	0.070946	0.146552	0.108949	0.080214	0.135371	0.074627	0.108597	0.473684	...	0.506494	0.177419	0.174312	0.086294
excitement	0.334	0.530233	0.094595	0.362069	0.420233	0.310160	0.344978	0.093284	0.429864	0.810526	...	0.792208	1.000000	0.376147	0.091371
not exceed planned budget	0.078	0.195349	0.020270	0.043103	0.077821	0.122995	0.034934	0.018657	0.072398	0.273684	...	0.311688	0.120968	0.036697	0.025381
realise creativity	0.530	0.623256	0.040541	0.939655	0.645914	0.352941	0.921397	0.048507	0.669683	0.968421	...	0.974026	0.701613	0.990826	0.172589
fun and entertainment	0.457	0.613953	0.192568	0.504310	0.587549	0.508021	0.519651	0.171642	0.570136	0.747368	...	0.766234	0.596774	0.000000	0.203046
good company	0.125	0.293023	0.050676	0.103448	0.089494	0.165775	0.082969	0.044776	0.072398	0.494737	...	0.532468	0.120968	0.055046	0.050761
health and beauty	0.522	0.786047	0.233108	0.577586	0.583658	0.796791	0.580786	0.171642	0.515837	0.842105	...	0.831169	0.588710	0.541284	0.197970
free-and-easy-going	0.189	0.241860	0.064189	0.318966	0.171206	0.058824	0.310044	0.059701	0.167421	0.568421	...	0.571429	0.209677	0.321101	0.050761
entertainment facilities	0.181	0.209302	0.111486	0.193966	0.225681	0.117647	0.196507	0.111940	0.212670	0.389474	...	0.363636	0.241935	0.000000	0.086294
not care about prices	0.459	0.758140	0.114865	0.021552	1.000000	0.705882	0.000000	0.115672	1.000000	0.789474	...	0.740260	0.943548	0.293578	0.121827
life style of the local people	0.215	0.641860	0.077703	0.077586	0.140078	0.556150	0.069869	0.055970	0.158371	0.473684	...	0.519481	0.201613	0.100917	0.055838
intense experience of nature	0.190	0.353488	0.131757	0.185345	0.124514	0.267380	0.157205	0.123134	0.117647	0.473684	...	0.571429	0.048387	0.183486	0.111675
cosiness/familiar	0.302	0.902326	0.158784	0.068966	0.175097	0.946524	0.078603	0.089552	0.090498	0.663158	...	0.701299	0.233871	0.110092	0.096447

Source Code:

Step:7 Rishabh Kumar Singh

DESCRIBING SEGMENTS

How do you develop market segmentation?

Steps in Market Segmentation

1.   Identify the target market. The first and foremost step is to identify the target market.

2.   Identify expectations of Target Audience.

3.   Create Subgroups.

4.   Review the needs of the target audience.

5.   Name your market Segment.

6.   Marketing Strategies.

7.   Review the behavior.

8.   Size of the Target Market.

The Three Visualization Tools You Can Use Are:

1.   Customer segment-specific avatar icons for each customer profile.

2.   Color coding differentiating specific segments.

3.   Adding flag labels to the segments – for example, a call-out speech bubble shape.

Predicting segmentation from descriptor variable

Predictive segmentation is a technique used in marketing to identify and create customer segments based on the high probability of

occurrences of certain behaviors, events or conditions in the future. It is typically powered by artificial intelligence and machine learning technology and is automated

Predictive segmentation removes the need for all of the manual work by automatically identifying and analyzing the valuable or high-potential audience segments that need to be targeted. However, as an individual capability, it does not help in determining the right experience that should be tailored for each identified audience segment. In other words, predictive segmentation accelerates a marketer's position in the workflow, enabling them with the right segments to work with, but tailoring the appropriate experiences remains guesswork that would have to be tested and validated over time with data.

Targeting in marketing is a strategy that breaks a large market into smaller segments to concentrate on a specific group of customers within that audience. It defines a segment of customers based on their unique characteristics and focuses solely on serving them.

There Are Typically 4 Different Types Of Market Targeting Strategy:

- **Mass marketing (undifferentiated marketing)**:

Whole marketing with one offer

- **Segmented marketing (differentiated marketing)**

Decide to target several different market segments, separates offer for each .

- **Concentrated marketing (niche marketing)**

Concentrate on one, or a few , segments or niche

- **Micromarketing.**

Local Marketing – Cities, Neighbourhoods , specific stores

Individual Marketing – 1 Pers

Step:8 Rishabh Kumar Singh

Market Segment Evaluation

THESE COSTS ARE:

1. Product costs:

To accommodate each market segment, the firm goes ahead with designing a specific product. Designing work may be in the range of mere change in the label to complete rethinking of an entirely new product. It involves a good deal of research and development activities as a product is to be separated for each market segment.

2. Production cost:

Though technological break-through has made it possible for a good many firms to reduce the number of units needed to achieve economics of scale, still there persists a problem of high production costs.

3. Promotion costs:

Every organization has to develop a promotion strategy that fits each market segment. It is but natural that multiple strategies require huge expenditure on both human and financial resources to design different ads and place them in various media. The multiplicity of ads deprives the firms of quantity discounts and concessions. This results in mounting promotion costs.

Customising the Marketing Mix(Step-9),Utkarsh Joshi

A marketing mix model is a modeling technique used to determine market attribution, the estimated impact of each marketing channel company uses. It is an analysis technique that allow marketers to measure the impact of their marketing and advertising campaigns to determine how various elements contribute to their goal.

Marketing mix modelling uses aggregate data. Marketing mix modeling allows marketers to factor in external influencers such as seasoniality, promotions, etc.

Marketing Mix Modeling:

1. Importing Libraries and dataset:

```
In [1]: import numpy as np
        import pandas as pd
        import seaborn as sns
        import matplotlib.pyplot as plt

In [2]: df = pd.read_csv('Advertising.csv')
        df.head()

Out[2]:
```

	Unnamed: 0	TV	radio	newspaper	sales
0	1	230.1	37.8	69.2	22.1
1	2	44.5	39.3	45.1	10.4
2	3	17.2	45.9	69.3	9.3
3	4	151.5	41.3	58.5	18.5
4	5	180.8	10.8	58.4	12.9

2.Plotting heatmap to show degree of correlation:

```
In [4]: plt.figure(figsize=(9,8))
        sns.heatmap(round(df.corr(),2),annot = True,cmap = 'YlGnBu',linewidths=3, fmt='.2g')
        plt.title('Correlation Matrix')
        plt.show()
```


Correlation Matrix
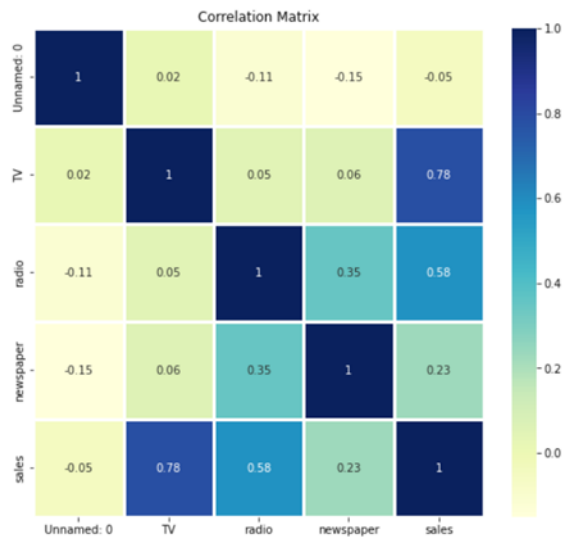
3.Feature Importance: allows us to see how important each input variable is in predicting output

```
In [6]: X = df.loc[:, df.columns != 'sales']
        y = df['sales']# Building Random Forest model
        from sklearn.ensemble import RandomForestRegressor
        from sklearn.model_selection import train_test_split
        from sklearn.metrics import mean_absolute_error as mae
        X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=.25, random_state=0)
        model = RandomForestRegressor(random_state=1)
        model.fit(X_train, y_train)
        pred = model.predict(X_test)# Visualizing Feature Importance
        feat_importances = pd.Series(model.feature_importances_, index=X.columns)
        feat_importances.nlargest(25).plot(kind='barh',figsize=(10,10))
```
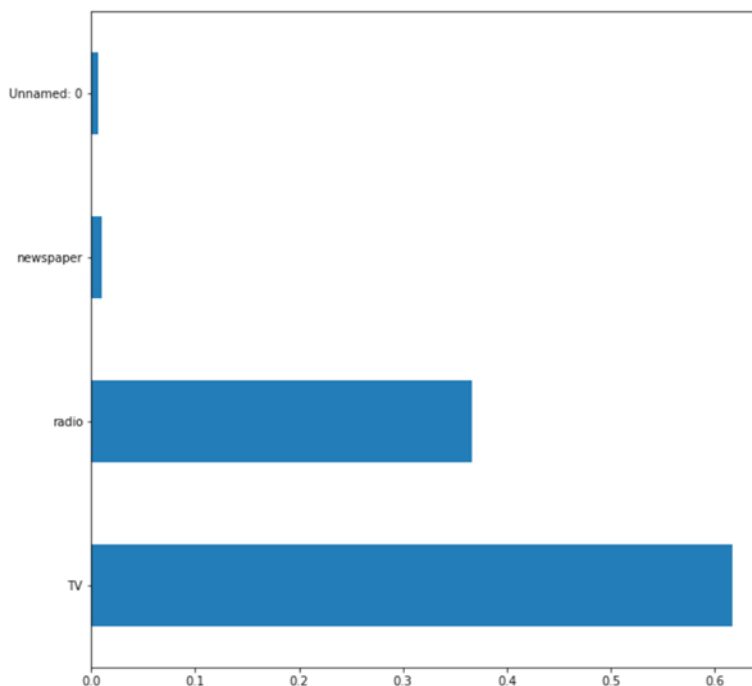

4.Building model.

```
In [7]: import statsmodels.formula.api as sm
        model = sm.ols(formula="sales~TV+radio+newspaper", data=df).fit()
        print(model.summary())
```

```
                           OLS Regression Results
==============================================================================
Dep. Variable:                 sales   R-squared:                       0.897
Model:                           OLS   Adj. R-squared:                  0.896
Method:                Least Squares   F-statistic:                     570.3
Date:               Wed, 16 Mar 2022   Prob (F-statistic):           1.58e-96
Time:                       15:48:29   Log-Likelihood:                -386.18
No. Observations:                200   AIC:                             780.4
Df Residuals:                    196   BIC:                             793.6
Df Model:                          3
Covariance Type:           nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      2.9389      0.312      9.422      0.000       2.324       3.554
TV             0.0458      0.001     32.809      0.000       0.043       0.049
radio          0.1885      0.009     21.893      0.000       0.172       0.206
newspaper     -0.0010      0.006     -0.177      0.860      -0.013       0.011
==============================================================================
Omnibus:                      60.414   Durbin-Watson:                   2.084
Prob(Omnibus):                 0.000   Jarque-Bera (JB):              151.241
Skew:                         -1.327   Prob(JB):                     1.44e-33
Kurtosis:                      6.332   Cond. No.                         454.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```
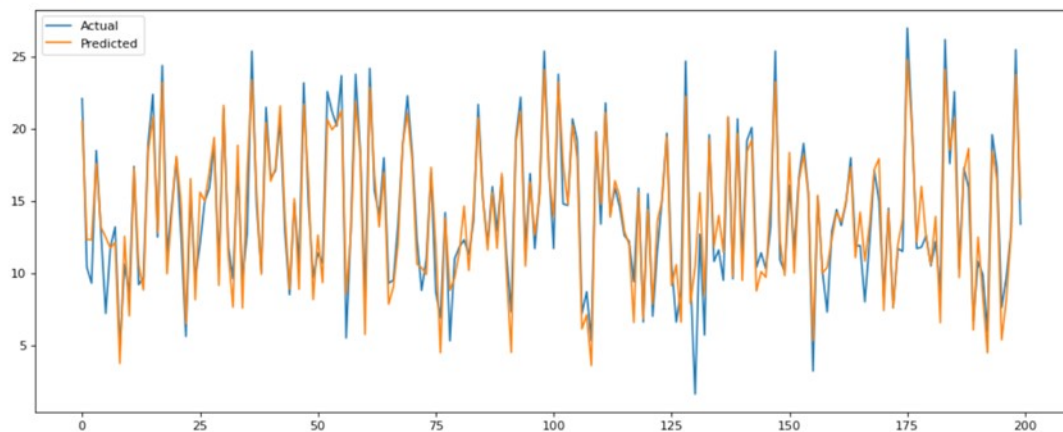
5. Plotting to see how good our model does in predicting values.

```
In [9]: from matplotlib.pyplot import figure

        y_pred = model.predict()
        labels = df['sales']
        df_temp = pd.DataFrame({'Actual': labels, 'Predicted':y_pred})
        df_temp.head()

        figure(num=None, figsize=(15, 6), dpi=80, facecolor='w', edgecolor='k')
        y1 = df_temp['Actual']
        y2 = df_temp['Predicted']
        plt.plot(y1, label = 'Actual')
        plt.plot(y2, label = 'Predicted')
        plt.legend()
        plt.show()
```


Thus our model does well in predicting values.

Source Code:
https://drive.google.com/file/d/1K8QmsnwsAmedhW0HsroBAgNW
WiARJph-/view?usp=sharing