BIO-TECH STARTUP MARKET SEGMENTATION

T-CORVUS

Parth Rajendra Salunke

A Chingkheinganba Singha

Prachi Nikhare

Akash Gupta

30/12/2021

Github Link: https://github.com/prachinikhare/BIO-TECH-STARTUP-MARKET-SEGMENTATION

Abstract :

In the given report We are mainly focused on innovation in health technology and development processes in healthcare industries. We Analyze the Medical Market in India with respect to Geographic, Demographic, Psychographic, Behavioral pattern and finally we come up with a feasible strategy to enter in the biotech market.



Problem Statement

Our team is working under a Biotech Startup which is going to launch its Home Checkup Service with Online Booking offering the following initial services.

1. Full Body Checkup with a Bio-Tech Device based on Blood Samples

2. Online Health Techs offering i. Diabetes checkup device, ii. Blood Pressure checkup device, iii. Vitamins deficiency checkup device.

so , we need to analyse the Medical Market in India with respect to the given problem statement using Segmentation analysis and come up with a feasible strategy to enter the market, targeting the segments most likely to use their product in terms of Geographic, Demographic, Psychographic, Behavioral pattern .

Data Collection

1. **Some of the different data collecting sources:**

- Collecting new data from internet and other sources
- Using the previously collected and stored data
- Reusing someone else's data
- Purchasing data

2. The data collection methods depend on the following:

- The research problem under study
- The research design
- The information gathered about the variable

Data Preprocessing:

Step 1. Loading the data set

Importing libraries:

The absolutely first thing we need to do is to import libraries for data preprocessing. There are lots of libraries available, but the most popular and important Python libraries for working on data are Numpy, Matplotlib, and Pandas. Numpy is the library used for all mathematical things. Pandas is the best tool available for importing and managing datasets. Matplotlib
(Matplotlib.pyplot) is the library to make charts.

Step 2. Loading data into pandas

we downloaded your data set and named it as a .csv file, we need to load it into a pandas DataFrame to explore it and perform some basic cleaning tasks removing information you don't need that will make data processing slower.

Step 3. Exploring the data set

Deciding on a target column

With a filtered data set explored, we need to create a matrix of dependent variables and a vector of independent variables. First, you should decide on the appropriate column to use as a target column for modeling based on the question we want to answer. For example, if you're going to predict the development of cancer, or the chance the credit will be approved, we need to find a column with the status of the disease or loan granting and use it as the target column.

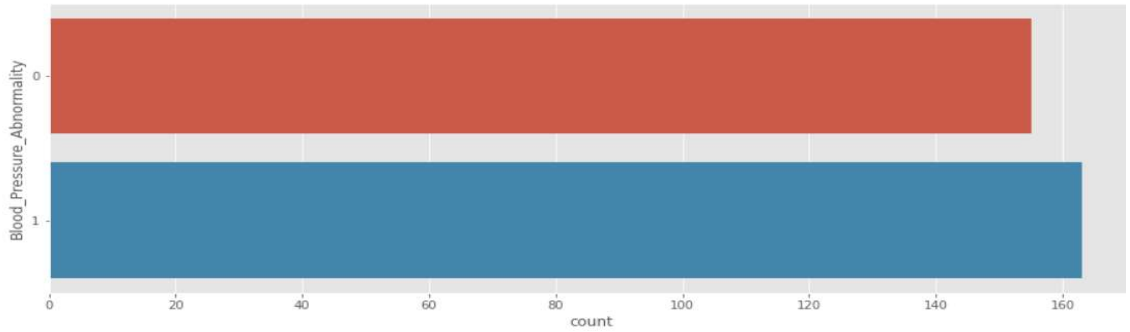
Step 4. Preparing the Features for Machine Learning

Finally, it's time to do the preparatory work to feed the features for ML algorithms. To clean the data set, we need to handle missing values and categorical features, because the mathematics underlying most machine learning models assumes that the data is numerical and contains no missing values. Moreover, the scikit-learn library returns an error if we try to train a model like K-Means Clustering and Hierarchical Clustering. using data that contain missing or non-numeric values. Clustering algorithm is a technique that assists customer segmentation which is a process of classifying similar customers into the same segment. Clustering algorithms helps to better understand customers, in terms of both static demographics and dynamic behaviors.

Data Visualization:

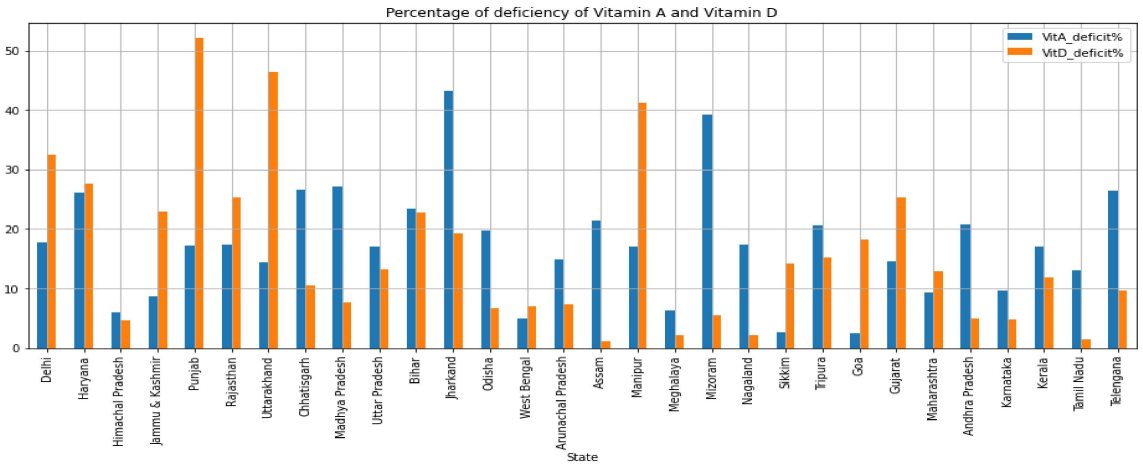
1. Blood Pressure

Blood pressure data visualization for the management of hypertension: designing for patient and physician information needs.

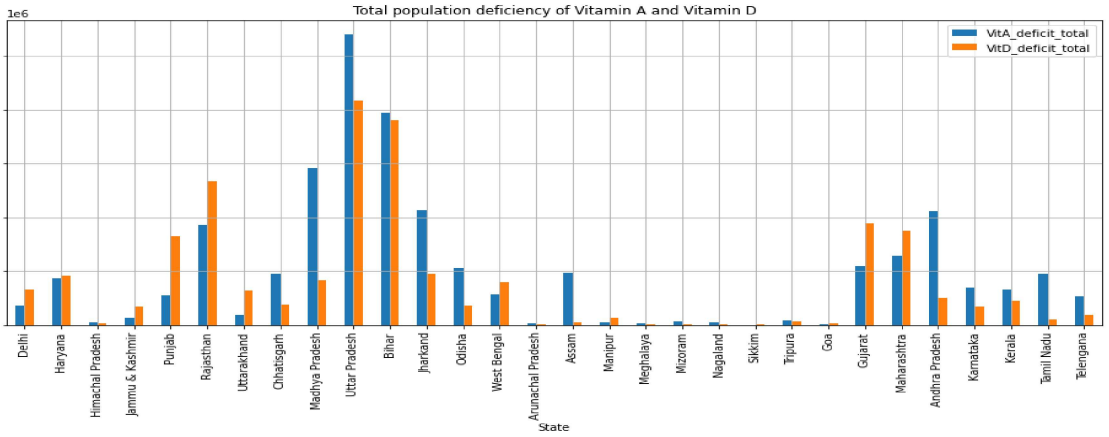


2. Vitamins

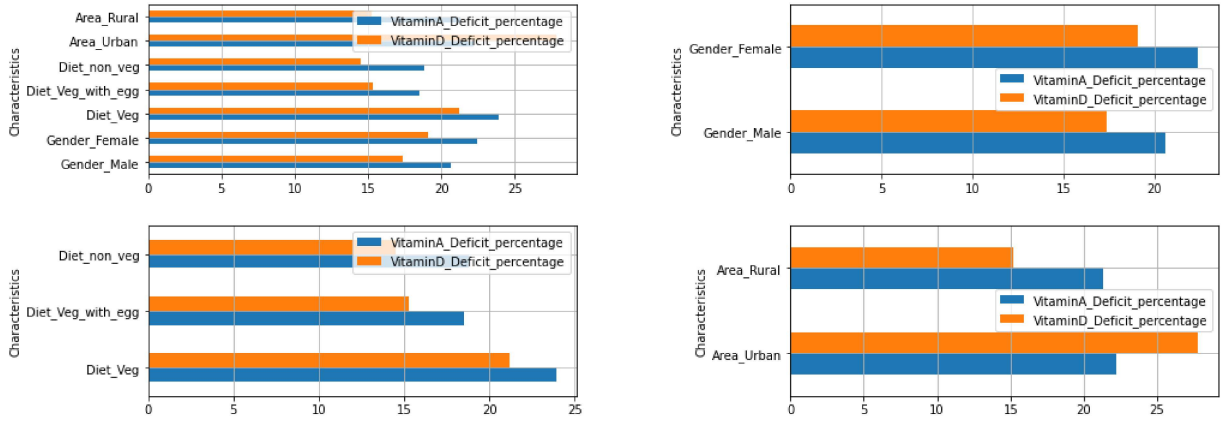
The data used in the report has been collected from a survey during the year 2016 – 2018. It includes the data about the deficiency of Vitamin A and Vitamin D among children aged between 0 – 6 in various states of India.



In the above plot, we observe that Jharkhand and Mizoram have higher percentage of Vitamin A deficiency whereas Punjab, Uttarakhand and Manipur have higher percentage of Vitamin D deficiency.



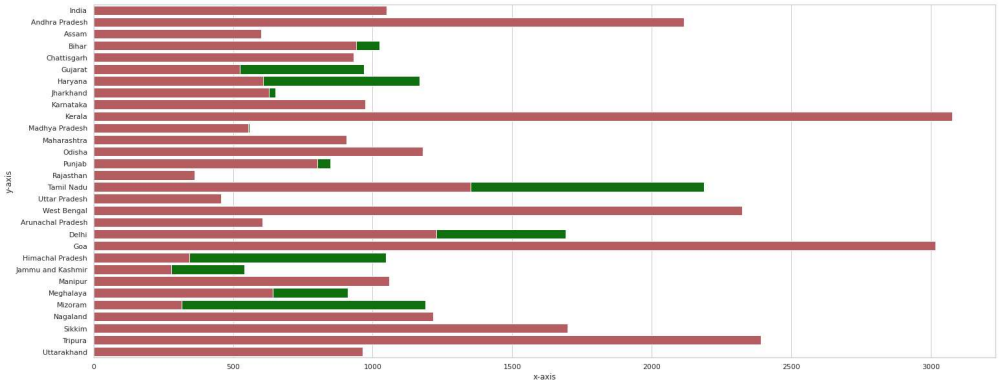
In the above plot, we observe that Uttar Pradesh has the highest population of Vitamin A deficiency and Uttar Pradesh and Bihar have higher population of Vitamin D deficiency.



Here, we observe that female children have higher percentage of deficiency both in Vitamin A and D. Also, vegetarians have comparatively higher percentage of deficiency both in Vitamin A and D. Moreover, the percentage of deficiency both in Vitamin A and D is higher in Urban areas.

3. Diabetes

Diabetes is a chronic (long-lasting) health condition that affects how your body turns food into energy. Most of the food you eat is broken down into sugar (also called glucose) and released into your bloodstream. When your blood sugar goes up, it signals your pancreas to release insulin.Although not everyone with type 2 diabetes is overweight, obesity and an inactive lifestyle are two of the most common causes of type 2 diabetes.

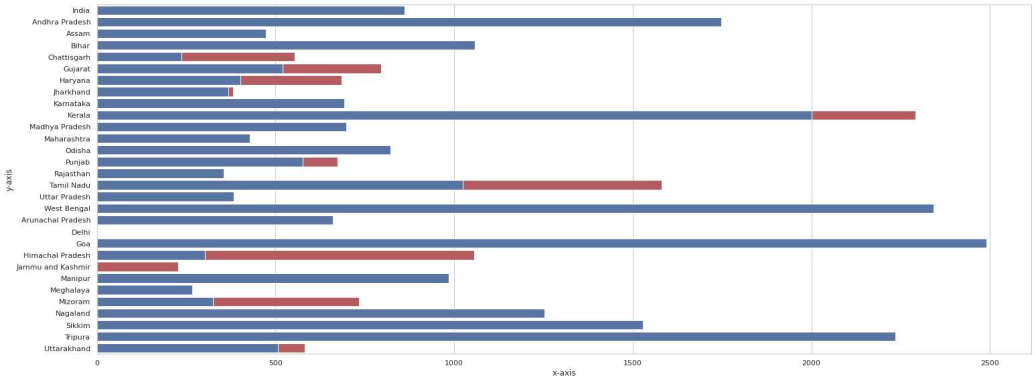


More number of diabetes patients in Goa ,Kerala and Andhra Pradesh. Lowest number of diabetes patients in Rajasthan , Uttar Pradesh and Jammu & Kashmir.

In Tamilnadu I have noticed that there are more number of diabetes patients female than male.

Same thing noticed in Delhi Himachal Pradesh Jammu and Kashmir meghalaya Mizoram Haryana Gujarat.

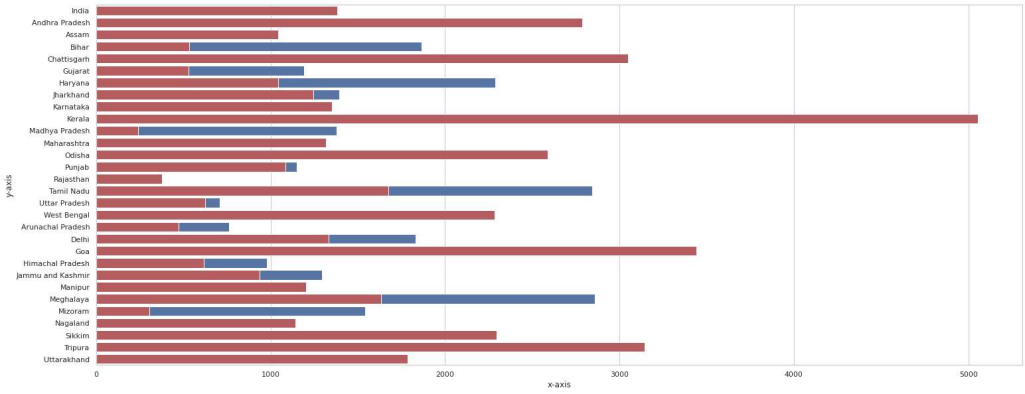
A **Rural Area** is an open swath of land that has few homes or other buildings, and not very many people. A rural area's population density is very low. Many people live in a city, or urban area. Agriculture is the primary industry in most rural areas.



Agriculture is primary in most of the rural areas. So when we analyse the above plot , we notice that Kerala, Goa , Andhra Pradesh, Tripura have large rural areas so they also have high no. Of diabetes patients.In plot i notice that Delhi don't have a single

patient in rural area because Delhi is a proper City which don't have mainly rural area.Kerala Tamil Nadu Himachal Pradesh Jammu Kashmir Mizoram have high number of female diabetes patient .West Bengal Goa and Andhra Pradesh Bihar have high number of diabetes patients.

Urban Areas are very developed, meaning there is a density of human structures such as houses, commercial buildings, roads, bridges, and railways. "Urban area" can refer to towns, cities, and suburbs.



Urban areas are very developed areas like town cities and suburbs. There are more number of diabetes patients in Kerala in rural areas and also urban area.Goa also has a high number of diseases in rural areas and also urban TamilNadu meghalaya Haryana Bihar have high number of female diabetes patients. There were over 2,000 patients in 2005 Andhra Pradesh Kerala Odisha Tamilnadu Goa Tripura.

Model Building

Finding the optimal clusters using:

i) K-Means

ii) Hierarchical Clustering (Agglomerative)

iii) Plotting Elbow Method

iv) Silhouette Score

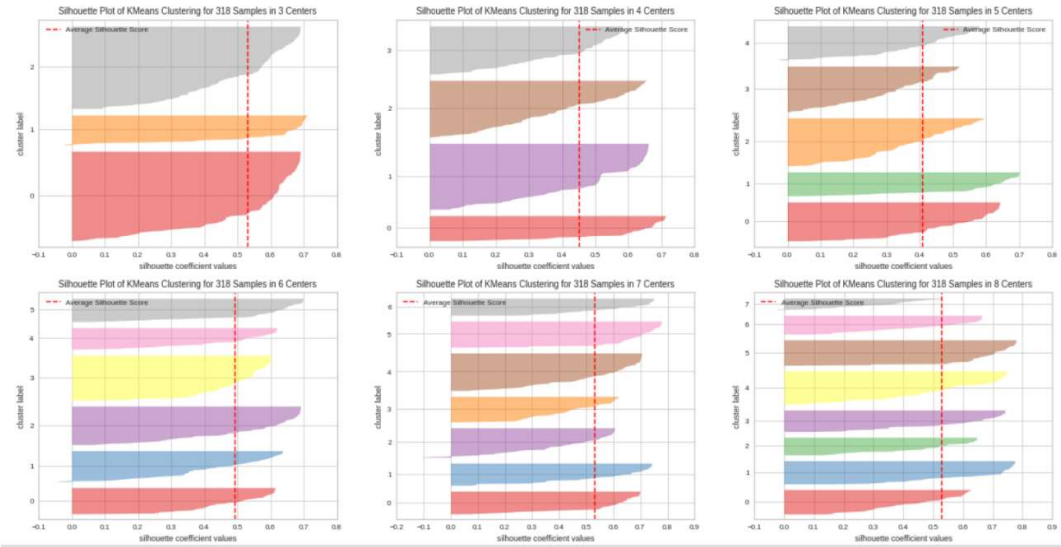
v) Finding the Segment Distribution for each cluster

vi) Performing Profiling

1. Blood Pressure

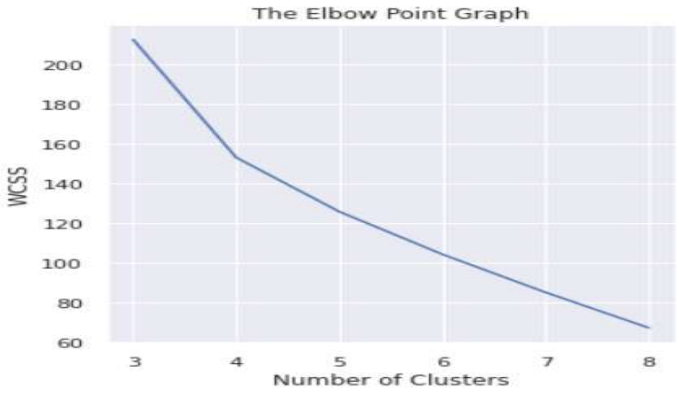
- Step 1: Finding the Cophenetic Distance Correlation Coefficient for different Linkages
  Conclusion: For the Agglomerative Clustering, the Cophenet distance correlation coefficient helps in deciding which linkage method is the most beneficial for the dataset. The higher the coefficient, the better it is. From the above, the best method is Average computed using City Block distance.

- Step 2:The Optimal clusters using KMeans, Silhouette Coefficient Score for both K-Means and Agglomerative Clustering

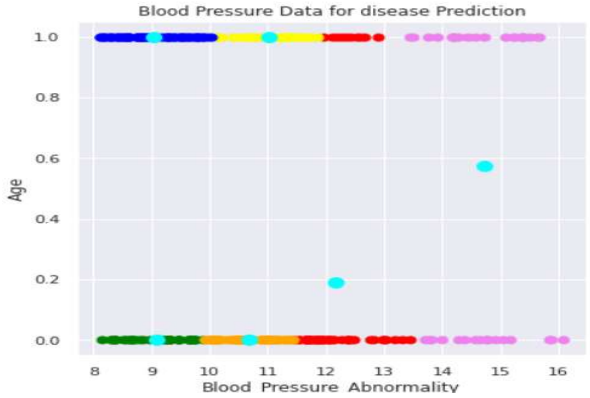


- Step 3: Elbow graph

  Inertia or Sum of Squared Errors within the Clusters is also known as the Cluster Errors. Cluster error will decrease after some Clusters

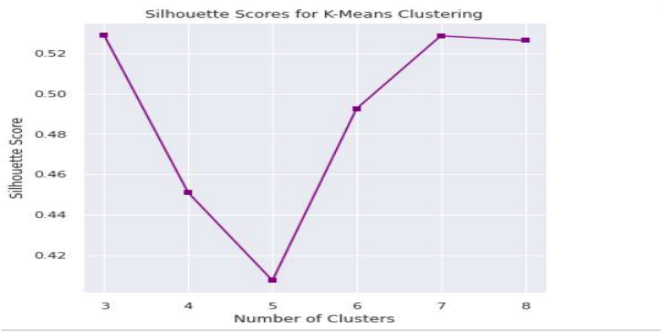


- Step 4: Plotting All The Clusters And Their Centroids

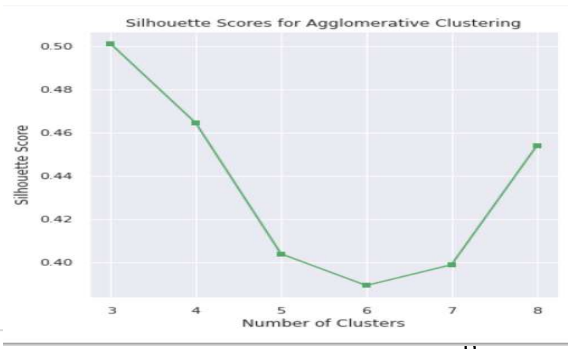


- Step 5: Plotting the Silhouette Score for the clusters found from K-Means and Agglomerative Clustering

sil_kmeans



sil_agc



- Step 6 : Segment Distribution

  Segment distribution gives the percentage of values in each Segment

  i) Making K-Means Cluster and Labels for finding out the distribution of Segments and then performing Profiling
  1)K = 3, 2)K = 4, 3)K = 5, 4)K = 6, 5)K = 7, 6)K = 8

  ii) The Segment Distribution for cluster 1)K = 3, 2)K = 4, 3)K = 5, 4)K = 6, 5)K = 7, 6)K = 8

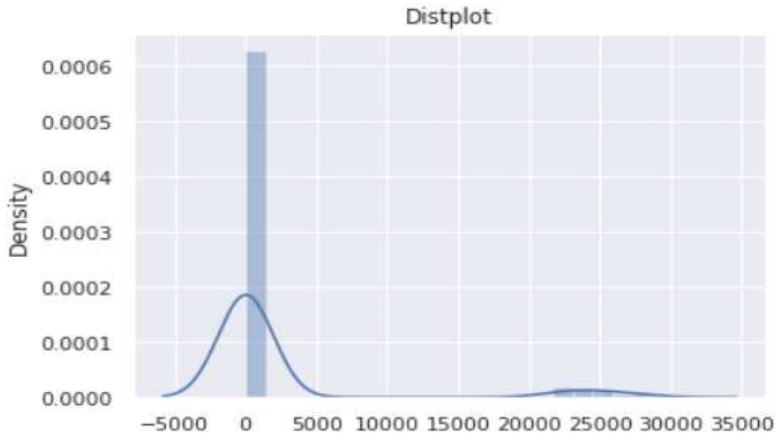
- Step 7 : Profiling And Describing Potential Segments

  The goal of Profiling is to be able to differentiate between the groups with the proper distinction that means each group has some unique characteristics about it and then that cluster is the ideal solution to segment the values.

  Steps to do Profiling are: 1) Find the count of each segment (value_counts) ----> that gives how many observations or records are present in each segment 2) For each of the variable: find the overall average and the individual segment-wise average 3) Perform the above two steps for each K value to find the best profiling

  Step 1 of Profiling :- Finding how many observations or records are present in each segment

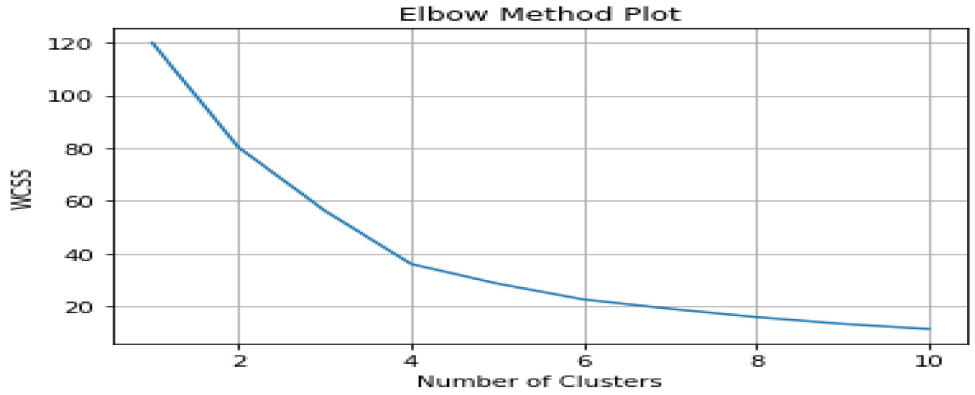
  Step 2 of Profiling :- For each of the variables : find the overall average and the individual segment wise average.
  A good indication of the distribution of data is the mean value, hence will find the average value for each variable and for each cluster.



2. Vitamin

i) Elbow graph

Inertia or Sum of Squared Errors within the Clusters is also Known as the Cluster Errors. Cluster error will decrease after some Clusters.



ii) Segment Distribution

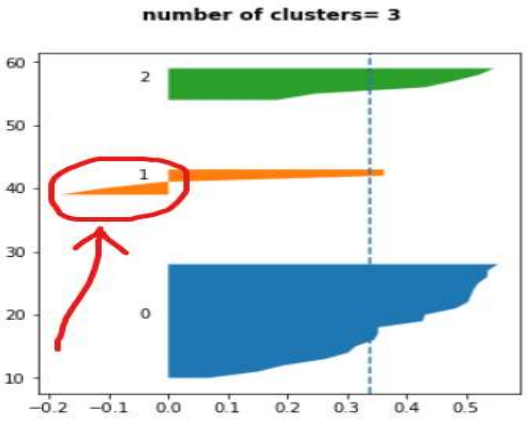
Segment distribution gives the percentage of values in each Segment.

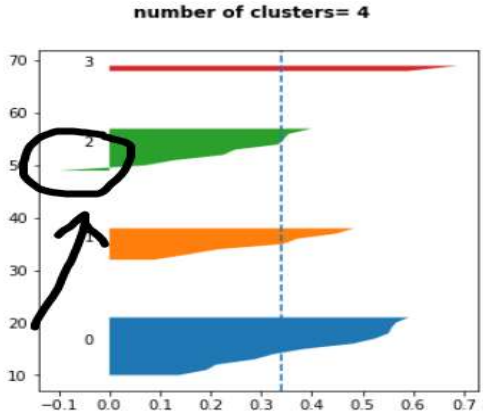
Making K-Means Cluster and Labels for finding out the distribution of Segments
1) K = 3, 2) K = 4, 3) K = 5, 4) K = 6

The Segment Distribution for cluster 1) K = 3, 2) K = 4, 3) K = 5, 4) K = 6

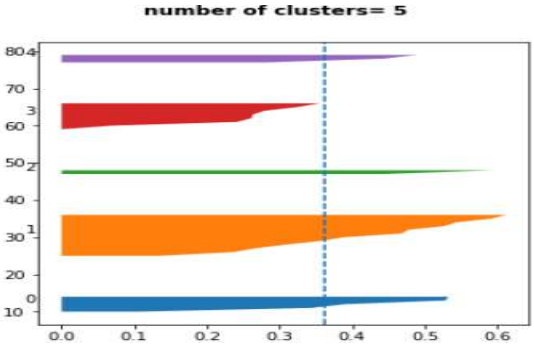
Plotting the Silhouette Score for the clusters found from K-Means.

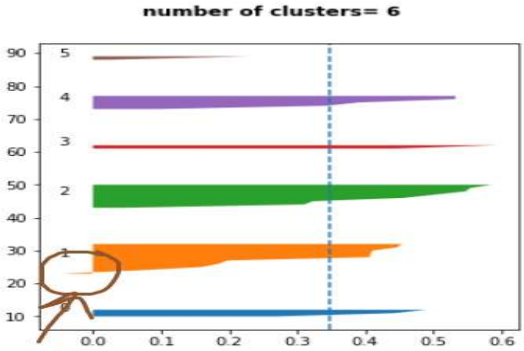
```
The silhouette score for no. of cluster k =  3  is :-  0.3390144574166799
The silhouette score for no. of cluster k =  4  is :-  0.3396809882591862
The silhouette score for no. of cluster k =  5  is :-  0.36173462627882186
The silhouette score for no. of cluster k =  6  is :-  0.3463507399737702
```





From above, we observe that the Silhouette





Score = **0.36173462627882186** is the highest for number of clusters = 5

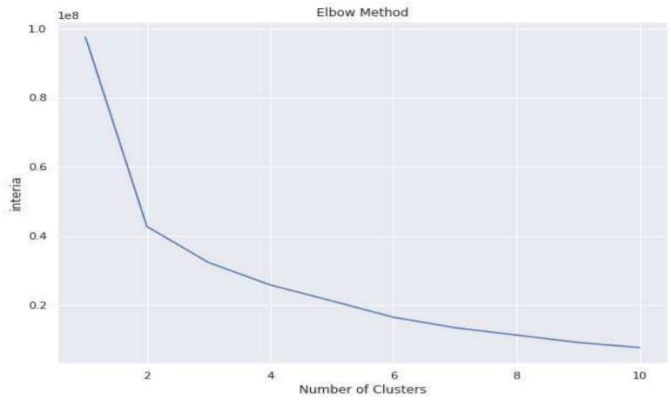
iii) States in Cluster

- States in Cluster 1 are ['Delhi' 'Haryana' 'Punjab' 'Uttarakhand' 'Manipur']
- States in Cluster 2 are ['Himachal Pradesh' 'Jammu & Kashmir' 'West Bengal' 'Arunachal Pradesh' 'Meghalaya' 'Nagaland' 'Sikkim' 'Tripura' 'Goa' 'Karnataka' 'Kerala' 'Tamil Nadu']
- States in Cluster 3 are ['Uttar Pradesh' 'Bihar']
- States in Cluster 4 are ['Chhattisgarh' 'Madhya Pradesh' 'Jharkhand' 'Odisha' 'Assam' 'Mizoram' 'Andhra Pradesh' 'Telangana']
- States in Cluster 5 are ['Rajasthan' 'Gujarat' 'Maharashtra']

3.Diabetes

i) Elbow graph

Inertia or Sum of Squared Errors within the Clusters is also Known as the Cluster Errors. Cluster error will decrease after some Clusters.

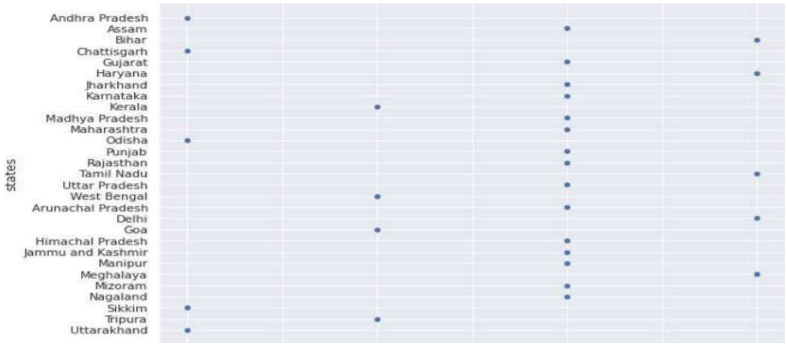


ii) States in Cluster

States wise visualization with Clusters

Making K-Means Cluster and Labels for finding out the distribution of Segments

1) K = 2, 2) K = 4, 3) K = 6, 4) K = 8, 5) K = 10



Selection of Target Segment

Target marketing in healthcare is the practice of marketing your healthcare center to your target audience by using strategies like audience research, segmentation, and more. With healthcare target marketing, you can better reach your ideal client or patient.

❖ **Four ways to define your target audience**

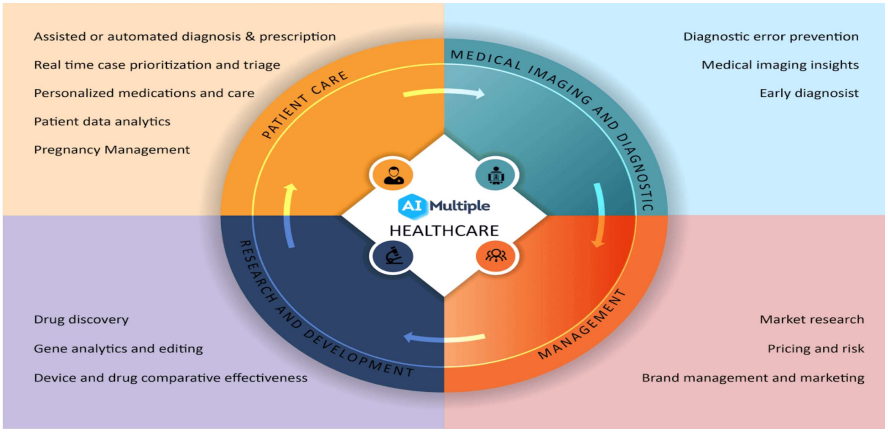
There are few points that will help you find the right target audience for your marketing initiatives. However, the goal should be to specific about the target audience you select so that your **healthcare marketing** campaigns provide expected results. Let us explore the details here.

- **Geo-location**: When researching your target audience, you will have to decide about the location that you will focus on in your marketing campaigns. This decision will be based upon the physical location of your establishment and the distance from which you expect patients to visit.
- **Demographics**: Marketing campaigns are to be created on the basis of chosen demographics of your audience. For instance, you will have to analyze the gender, age, education, size of family, occupation, household income, and similar other details while deciding about demographics to target.
- **Psychographics**: As part of psychographic analysis, few of the points to look into would be lifestyle, behavior, and personality of your audience base. In addition, you need to examine whether the audience is responsive to new innovations.
- **Behavior**: analyze needs of our audience base as well as evaluate their knowledge level. Additionally, we need to check the response they provide to specific health care services and products.

- **OUR TARGET:-** For Vitamins checkup, we can target female children of States in Cluster 3 ['Uttar Pradesh' ,'Bihar'] and States in Cluster 5 ['Rajasthan', 'Gujarat' ,'Maharashtra'] who residing in Urban areas and being a vegetarian.

  For Diabetes checkup, we can target men of Chhattisgarh, Kerala and West Bengal who reside in urban areas.Although we get different target states, we can still recommend our customers with both the checkup services as it is better to have good health.

A marketing programme uniquely designed for a particular customer. In international marketing the term describes a marketing programme designed uniquely for a particular country.



A marketing mix includes multiple areas of focus as part of a comprehensive marketing plan. The term often refers to a common classification that began as the four Ps: **product, price, placement, and promotion.** Effective marketing touches on a broad range of areas as opposed to fixating on one message.

i) Product

Initially, we will be providing Vitamins and Diabetes checkup services. Later, we can expand to other services, routine, cholesterol check ups etc.

ii) Price

The sale price of the product reflects what consumers are willing to pay for it. Marketing professionals need to consider costs related to research and development, manufacturing, marketing, and distribution—otherwise known as cost-based pricing. Pricing based primarily on consumers' perceived quality or value is known as value-based pricing.

iii) Placement

The type of product sold is important to consider when determining areas of distribution. Basic consumer products, such as paper goods, often are readily available in many stores. Premium consumer products, however, typically are available only in select stores. Another consideration is whether to place a product in a physical store, online, or both.

iv) Promotion

Activities might include advertising, sales promotion, personal selling, and public relations. A key consideration should be for the budget assigned to the marketing mix. Marketing professionals carefully construct a message that often incorporates details from the other three Ps when trying to reach their target audience. Promotions can also be made via social sites like Facebook, Instagram, Youtube etc.

Conclusion

- The primary aim of AI in healthcare is to analyze relationships between treatment techniques and patient outcomes.

- AI can achieve fast and accurate Diagnostics.

- It will be very helpful to reduce the human errors as well as the cost of treatment. Artificial Intelligence is definitely improving the healthcare industry. From predictive medical care and more accurate diagnosis to motivating the patients to take care of their health, AI will certainly continue enhancing the patient experience and healthcare expertise in general.

Reference

1. Miller D.D., Brown E.W. Artificial intelligence in medical practice: the question to the answer? *Am J Med.* 2018;131(2):129–133. [PubMed] [Google Scholar]

2. Kirch D.G., Petelle K. Addressing the physician shortage: the peril of ignoring demography. *JAMA.* 2017;317(19):1947–1948. [PubMed] [Google Scholar]

3. Combi C., Pozzani G., Pozzi G. Telemedicine for developing countries. *Appl Clin Inform.* 2016;07(04):1025–1050. [PMC free article] [PubMed] [Google Scholar]

4. Bresnick J. Artificial intelligence in healthcare market to see 40% CAGR surge; 2017.

5. Lee K.-F. *AI superpowers: China, Silicon Valley, and the new world order.* 1st ed. Houghton Mifflin Harcourt; 2019. [Google Scholar]

6. King D, DeepMind's health team joins Google Health.

7. Hoyt R.E., Snider D., Thompson C., Mantravadi S. IBM Watson Analytics: automating visualization, descriptive, and predictive statistics. *JMIR Public Health Surveill.* 2016;2(2):e157. [PMC free article] [PubMed] [Google Scholar]

8. Marr B. How is AI used in healthcare—5 powerful real-world examples that show the latest advances. Forbes; 2018.

9. Kalis B, Collier M, Fu R. 10 promising AI applications in health care. Harvard Business Review; 2018.

10. Singhal S, Carlton S. The era of exponential improvement in healthcare? McKinsey Co Rev.; 2019.

11. Konieczny L, Roterman I. Personalized precision medicine. Bio-Algorithms Med-Syst 2019; 15.

12. Love-Koh J. The future of precision medicine: potential impacts for health technology assessment. *Pharmacoeconomics.* 2018;36(12):1439–1451. [PMC free article] [PubMed] [Google Scholar]

13. Kulski JK. Next-generation sequencing—an overview of the history, tools, and 'omic' applications; 2020.

14. Hughes J.P., Rees S., Kalindjian S.B., Philpott K.L. Principles of early drug discovery. *Br J Pharmacol.* 2011;162(6):1239–1249. [PMC free article] [PubMed] [Google Scholar]

15. Ekins S. Exploiting machine learning for end-to-end drug discovery and development. *Nat Mater.* 2019;18(5):435–441. [PMC free article] [PubMed] [Google Scholar]

16. Zhang L., Tan J., Han D., Zhu H. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug Discov Today.* 2017;22(11):1680–1685. [PubMed] [Google Scholar]