Project Report:

Market Segmentation for Biotech Startup,

Shadab Haque, Praneeth Kolloju, Shreyash Kharat,

24 November 2021.

Report

24 Nov 2021

Introduction

In India, the biotechnology sector is a fast-growing industry. The companies under these biotech startups are known for providing better diagnosis, high-quality drugs, and making medicines and Med-Tech products available at a reasonable price. They aim at getting the best outcome for developing health awareness and upgrading diagnosis facilities for betterment. With the developing technologies, the health care sector is also increasing their quality with the help of these startups in India.

Amid the pandemic, one of the industries that made a huge growth has to be the biotech industry without a doubt. Apart from the prescribed medication, people were also flocking to get in touch with daily vitamins and other possible nutrients too. This has enabled the biotech industry in India and all over the world take off and flourish. So, let's get to know what are these biotechnology startups in India in a little bit more detail.

Overview

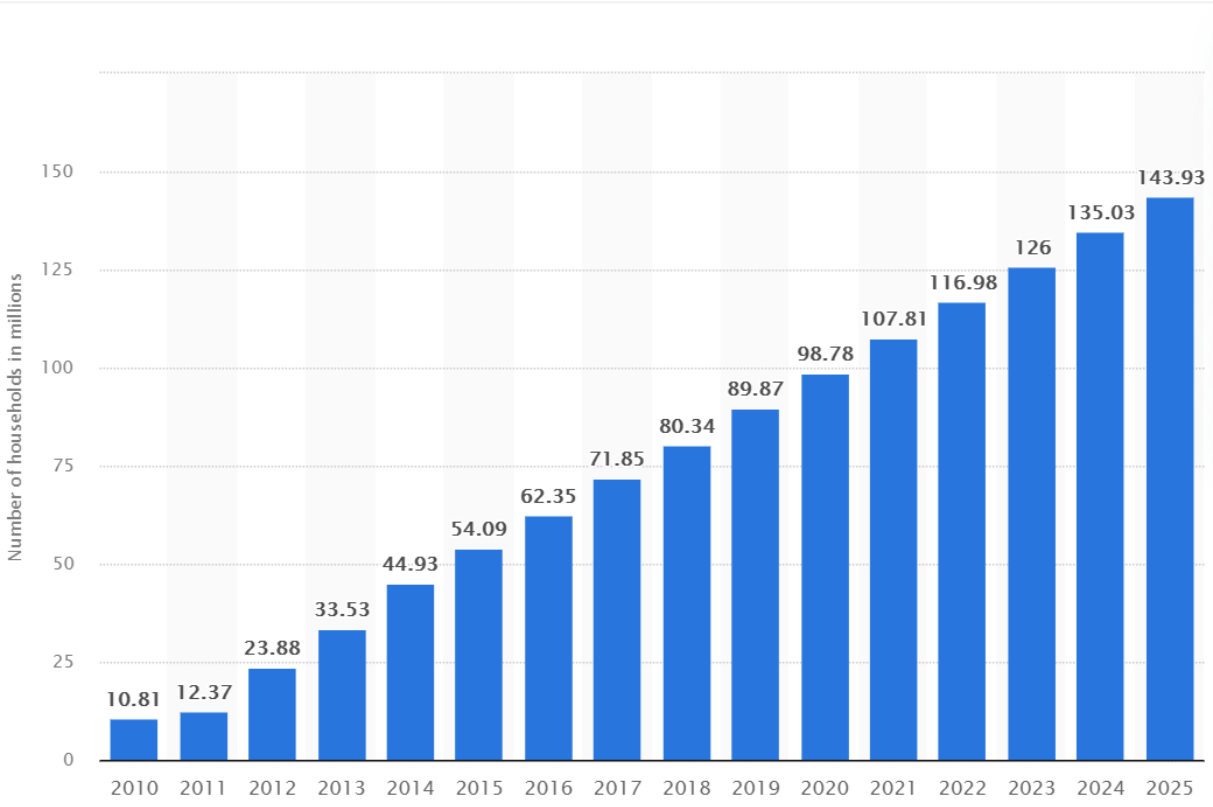
The Indian biotechnology industry, a highly knowledge-intensive industry, is on a strong growth trajectory and is expected to play a significant role as an innovative manufacturing hub. It is expected to grow at an average growth rate of around 30% a year and is poised to reach a USD 100 billion market size by 2025 (Make in India statistics). This is attributed to the growing demand for healthcare services, intensive R&D activities, strong government initiatives, fast-developing clinical capabilities with the country becoming a popular destination for global companies for clinical trials, contract research and manufacturing activities. As per the report, "India's Biotech Start-Ups Ecosystem" by the Association of Biotechnology Led Enterprises (ABLE), the bioscience sector has witnessed a constant dynamic situation with about 1022 startups being formed in this sector during a five-year period from 2012 to 2016 with investments worth Rs 18,700 crore from various sources.

The success of Biotechnology start-up depends on a number of factors. Annual health checkup is one basis with which we can gauge the success of a recently started bio-tech startup. This depends on Health literacy. Health literacy is the capacity to obtain, process and understand healthcare information and services needed to make appropriate decisions in the areas of wellness and patient care. It is one of the factors that lead to the acquisition of knowledge, positive attitude, self-efficacy, positive health behavior, and better survival. It gives a sense of control and confidence to individuals to gain access to and understand aspects of health promotion for themselves, their families and communities. It further helps in mobilizing communities to address the social, economic and environmental determinants of health. Low health literacy, on the other hand, has been found to be associated with lesser use of preventive services, and excessive use of emergency services, with high costs and dismal outcomes. .For instance, a person exposed to information related to health dangers may be more likely to undergo annual body checkups. Therefore, he/she may be more inclined to use health services provided by start-up companies.

When it comes to chronic diseases, diabetes and hypertension is at the forefront. In India, more than 77 million adults are living with diabetes. Researchers predict that this will increase to 134 million by 2045. Overall, females have a higher risk of developing diabetes than males, but as both groups get older, this risk decreases. Although diabetes figures are high, researchers estimate that 57% of cases remain undiagnosed. This is particularly concerning, as the risk of serious complications increases when people do not take medication to control their blood sugar.

For hypertension, research was conducted and the following were noted. Overall prevalence for hypertension in India was 29.8%. Significant differences in hypertension prevalence were noted between rural and urban parts [27.6% and 33.8%]. Regional estimates for the prevalence of hypertension were as follows: 14.5%, 31.7%, 18.1%, and 21.1% for rural north, east, west, and south India; and 28.8%, 34.5%, 35.8%, and 31.8% for urban north, east, west, and south India, respectively. Overall estimates for the prevalence of awareness, treatment, and control of BP were 25.3%, 25.1%, and 10.7% for rural Indians; and 42.0%, 37.6%, and 20.2% for urban Indians.

Internet accessibility is one of the most important factor for the success and profit of online biotechnology companies. The more widespread the internet access, the more exposure the firm gets to showcase its service to the customers. If the product matches the customers preference, they are more likely to encourage other people to use it too.

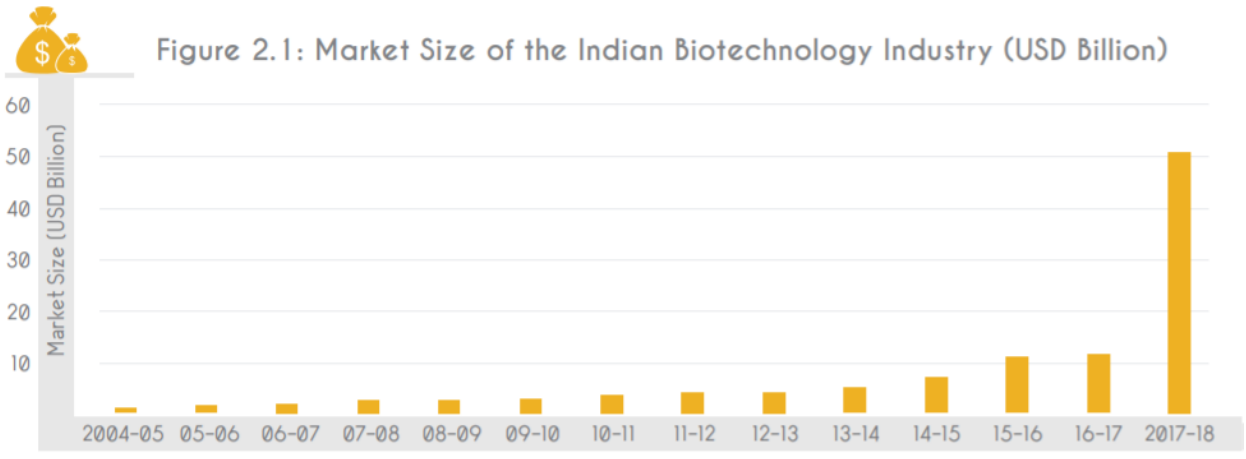


Average income also plays a part in the usage of products. Basically, the more income a person earns, the more money they are going to spend on themselves. This is true, since people in urban areas spend a lot on their health compared to people in rural areas.

Cities	Average Salary
Ahmadabad	18,849 INR
Bangalore	18,101 INR
Chennai	17,691 INR
Delhi	19,094 INR
Hyderabad	19,573 INR
Indore	16,702 INR
Jaipur	18,564 INR
Kolkata	18,328 INR
Mumbai	19,172 INR
Pune	17,796 INR

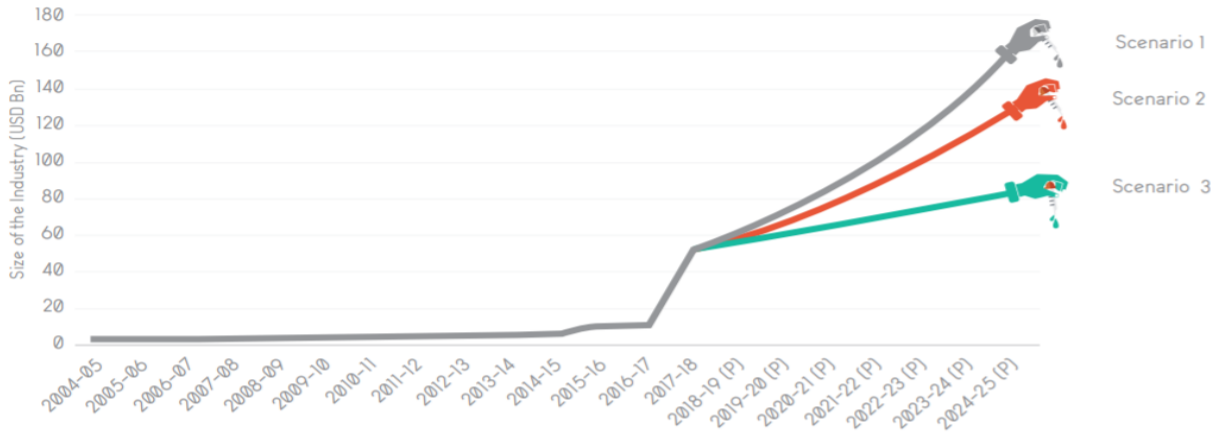
Market Overview

The biotechnology industry can be traced back to1980s when the Government created the Department of Biotechnology, India. From merely $1.1 billion in 2003, it has grown exponentially in size to a $51 billion industry as of 2018. The year-on-year growth rate of the industry is 14.7%. The growth of the industry is depicted in Figure 2.1. As of 2017, India accounted for approximately 3 percent share in the global biotechnology industry and stood among the top 12 destinations for biotechnology in the world.



The Department of Biotechnology has set an ambitious target for the industry i.e. to make a $100 billion industry by the year 2025. The Indian biotechnology industry has also become a prominent player in the global market. In the year 2017, it was around 3 percent of the global market. The Indian market, if it keeps on growing at the same growth rate of 14 percent, it would reach $128 billion by 2025. In 2017, the global market was valued at $390 billion and is projected to grow at 7.4 percent yearly. Assuming the same rate of growth, the value of the global market would be $670.9 billion by 2025.

Figure 2.2: Projecting the Growth of the Industry



Market Segmentation

Market segmentation is a decision-making tool for the marketing manager in the crucial task of selecting a target market for a given product and designing an appropriate marketing mix. Market segmentation is one of the key building blocks of strategic marketing. Market segmentation is essential for marketing success: the most successful firms drive their businesses based on segmentation. Market segmentation lies at the heart of successful marketing, tools such as segmentation have the largest impact on marketing decisions.

The segmentation criterion can be one single consumer characteristic, such as age, gender, country of origin, or stage in the family life cycle. Alternatively, it can contain a larger set of consumer characteristics, such as a number of benefits sought when purchasing a product, a number of activities undertaken when on vacation, values held with respect to the environment, or an expenditure pattern.

Market segmentation has a number of benefits. At the most general level, market segmentation forces organizations to take stock of where they stand, and where they want to be in future. In doing so, it forces organizations to reflect on what they are particularly good at compared to competitors, and make an effort to gain insights into what consumers want. Market segmentation offers an opportunity to think and rethink, and leads to critical new insights and perspectives. When implemented well, market segmentation also leads to tangible benefits, including a better understanding of differences between consumers, which improves the match of organizational strengths and consumer needs. Such an improved match can, in turn, form the basis of a long-term competitive advantage in the selected target segments.

Types of Market Segmentation

With segmentation and targeting, you want to understand how your market will respond in a given situation, like purchasing your products. In many cases, a predictive model may be incorporated into the study so that you can group individuals within identified segments based on specific answers to survey questions.
Geographic Segmentation:

Geographic information is seen as the original segmentation criterion used for the purpose of market segmentation. Typically – when geographic segmentation is used – the consumer's location of residence serves as the only criterion to form market segments. While simple, the geographic segmentation approach is often the most appropriate. For example: if the national tourism organization of Austria wants to attract tourists from neighboring countries, it needs to use a number of different languages: Italian, German, Slovenian, Hungarian, Czech. Language differences across countries represent a very pragmatic reason for treating tourists from different neighboring countries as different 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

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).

As is the case with geographic segmentation, socio-demographic segmentation criteria have the advantage that segment membership can easily be determined for every consumer. In some instances, the socio-demographic criterion may also offer an explanation for specific product preferences (having children, for example, is the actual reason that families choose a family vacation village where previously, as a couple, their vacation choice may have been entirely different). But in many instances, the socio-demographic criterion is not the cause for product preferences, thus not providing sufficient market insight for optimal segmentation decisions.

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 sociodemographic 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.

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. In a comparison of different segmentation criteria used as segmentation variables, behaviors reported by tourists emerged as superior to geographic variables. The key advantage of behavioral approaches is that – if based on actual behavior rather than stated behavior or stated intended behavior – the very behavior of interest is used as the basis of segment extraction. As such, behavioral segmentation groups people by the similarity which matters most.

Extracting Segments

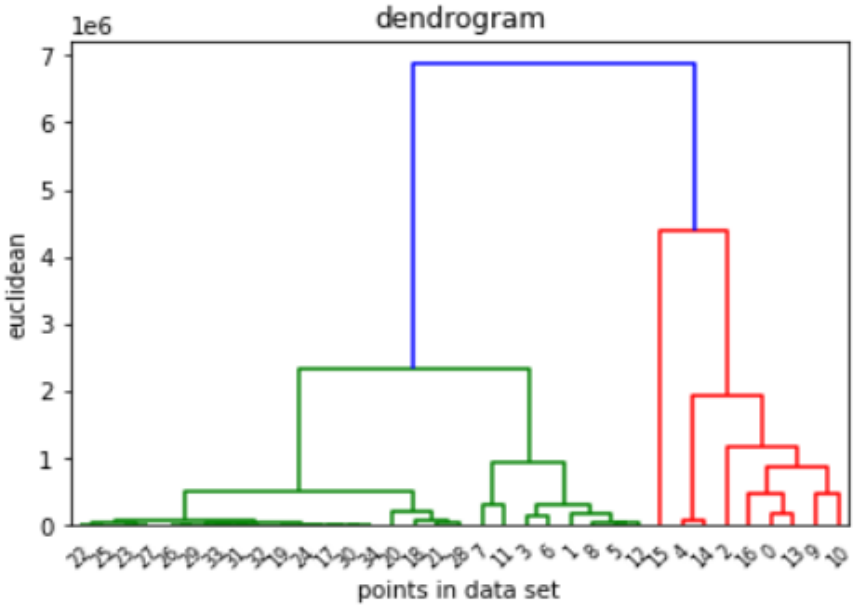
Hierarchical Clustering

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). If the aim is to have one large market segment (k = 1), the only possible solution is one big market segment containing all consumers in data X. At the other extreme, if the aim is to have as many market segments as there are consumers in the data set (k = n), the number of market segments has to be n, with each segment containing exactly one consumer. Each consumer represents their own cluster. Market segmentation analysis occurs between those two extremes. In this we will use Agglomerative hierarchical clustering which approaches the task from the other end. The starting point is each consumer representing their own market segment (n singleton clusters).

```
[330] x=df.iloc[:,14:].values
```

```
import scipy.cluster.hierarchy as hie
den=hie.dendrogram(hie.linkage(x,method='ward'))
plt.title('dendrogram')
plt.xlabel('points in data set')
plt.ylabel('euclidean')
```

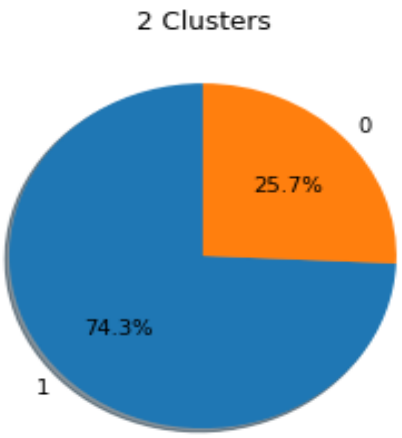
```
Text(0, 0.5, 'euclidean')
```



```
[332] from sklearn.cluster import AgglomerativeClustering
      ac=AgglomerativeClustering(n_clusters=2)
      pred=ac.fit_predict(x)
```

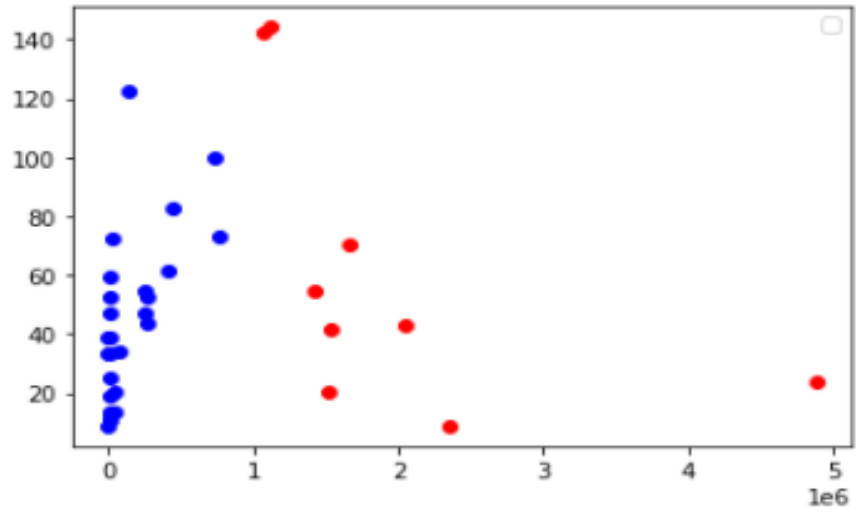
```
[333] df['clusters'] = pred
      df1 = df[df.clusters==0]
      df2 = df[df.clusters==1]
```

```
[334] c4=df['clusters'].value_counts()
      g = plt.pie(c4, labels=c4.index,
                  autopct='%1.1f%%', shadow=True, startangle=90)
      plt.title('2 Clusters')
      plt.show()
```



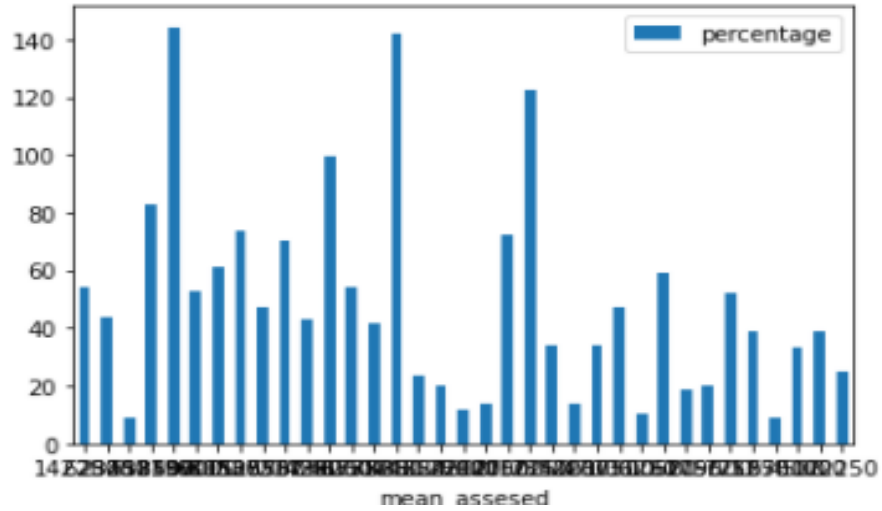
```
plt.scatter(df1['mean_assesed'],df1['percentage'],c='red')
plt.scatter(df2['mean_achieved'],df2['percentage'],c='blue')
plt.legend()
```

```
No handles with labels found to put in legend.
<matplotlib.legend.Legend at 0x7f62e86a2710>
```



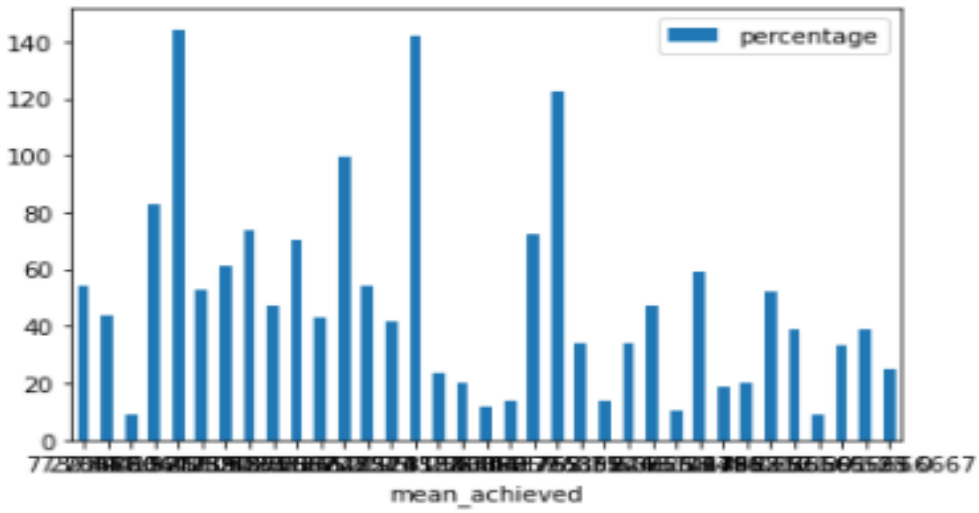

```
336] df.plot.bar(x='mean_assesed', y='percentage', rot=0)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f62e8724410>
```



```
[337] df.plot.bar(x='mean_achieved', y='percentage', rot=0)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f62e866cf50>
```



K-Means Clustering Algorithm

K-Means Clustering is an Unsupervised Machine learning Algorithm, which groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process, as if K=2, there will be two clusters, and for K=3, there will be three clusters, and so on.

It is an iterative algorithm that divides the unlabeled dataset into k different clusters in such a way that each dataset belongs only one group that has similar properties.

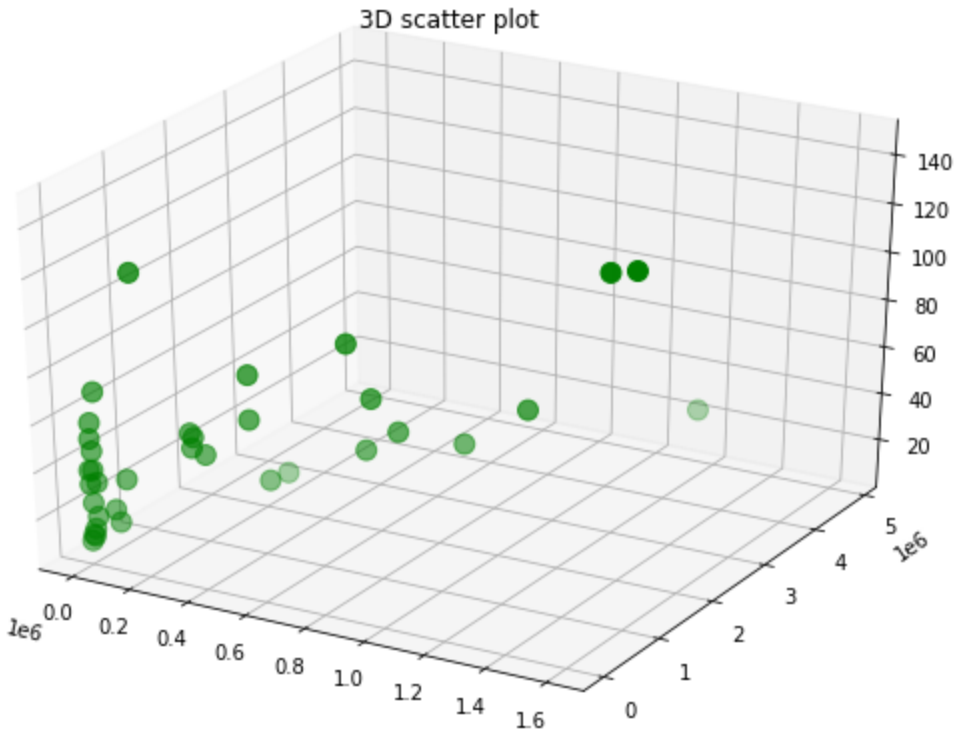
It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabeled dataset on its own without the need for any training.

It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters.

The algorithm takes the unlabeled dataset as input, divides the dataset into k-number of clusters, and repeats the process until it does not find the best clusters. The value of k should be predetermined in this algorithm.

The k-means clustering algorithm mainly performs two tasks:

- Determines the best value for K center points or centroids by an iterative process.
- Assigns each data point to its closest k-center. Those data points which are near to the particular k-center, create a cluster.

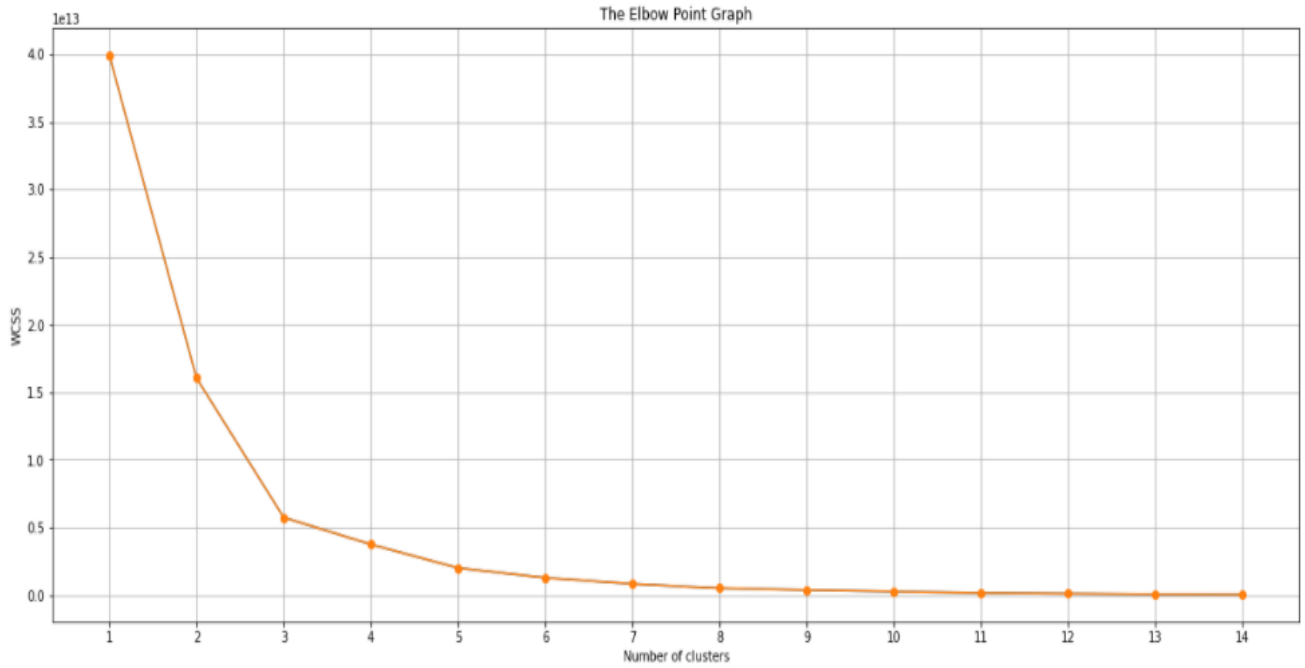


ELBOW METHOD:

In the **Elbow method**, we are actually varying the number of clusters (K) from 1 – 10. For each value of K, we are calculating WCSS (Within-Cluster Sum of Square). WCSS is the sum of squared distance between each point and the centroid in a cluster. When we plot the WCSS with the K value, the plot looks like an Elbow. As the number of clusters increases, the WCSS value will start to decrease. WCSS value is largest when K = 1. When we analyze the graph, we can see that the graph will rapidly change at a point and thus creating an elbow shape. From this point, the graph starts to move almost parallel to the X-axis. The K value corresponding to this point is the optimal K value or an optimal number of clusters.

The elbow method uses the sum of squared distance (SSE) to choose an ideal value of k based on the distance between the data points and their assigned clusters. We would choose a value of k where the SSE begins to flatten out and we see an inflection point. When visualized this graph would look somewhat like an elbow, hence the name of the method.

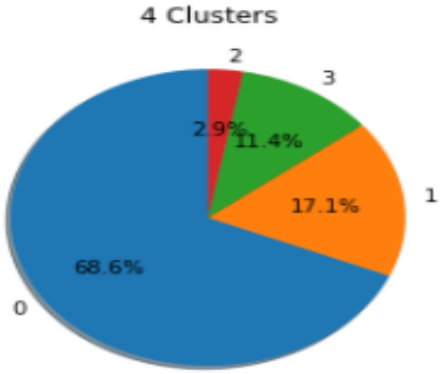
```
from sklearn.cluster import KMeans
wcss = []  # Within-Cluster-Sum-of-Squares
for i in range(1, 15):
    kmeans = KMeans(n_clusters=i, init='k-means++', random_state=10)
    kmeans.fit(x)
    wcss.append(kmeans.inertia_)

plt.figure(figsize=(20, 8))
plt.plot(range(1, 15), wcss)
plt.title('The Elbow Point Graph')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.plot(range(1, 15), wcss, "-o")
plt.xticks(range(1, 15))
plt.grid(True)
plt.show()
```

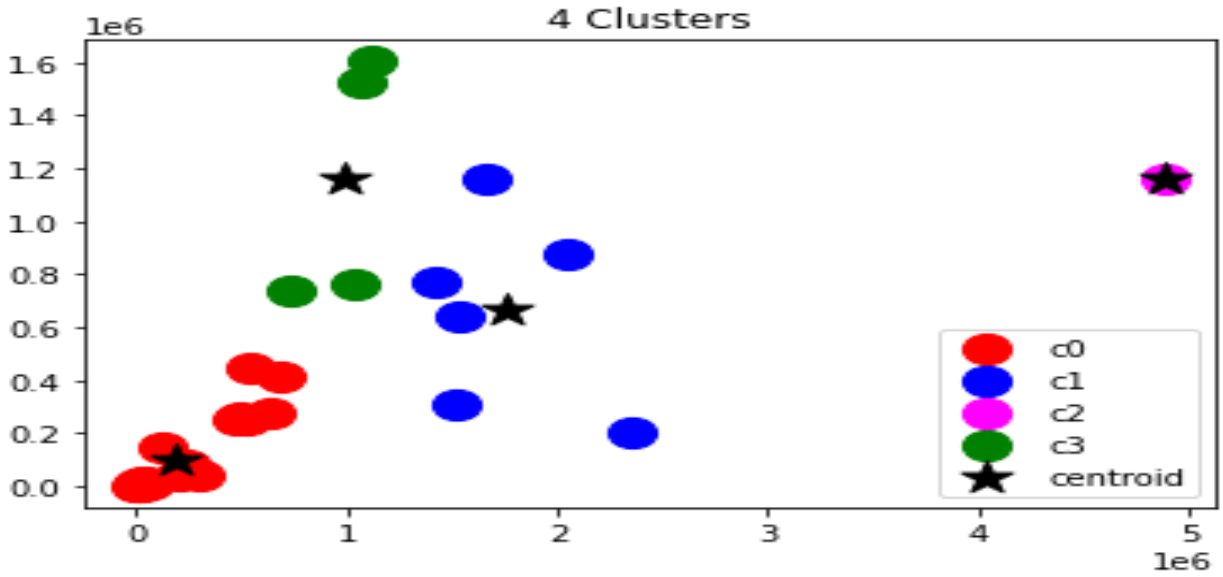


For 4 Clusters:

```
kmeans=KMeans(init='k-means++',n_clusters=4,random_state=0)
kmeans.fit(x)
pred=kmeans.predict(x)
```

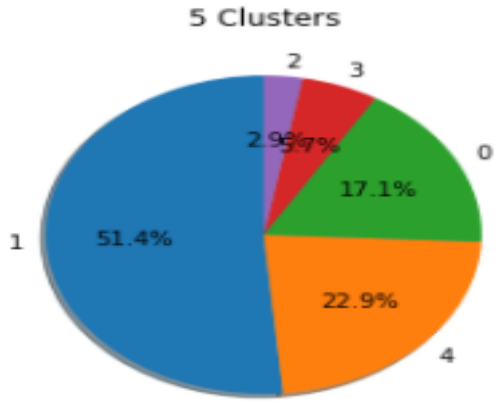
```
df['clusters4']=pred
c4=df['clusters4'].value_counts()
g = plt.pie(c4, labels=c4.index,
            autopct='%1.1f%%', shadow=True, startangle=90)
plt.title('4 Clusters')
plt.show()
```

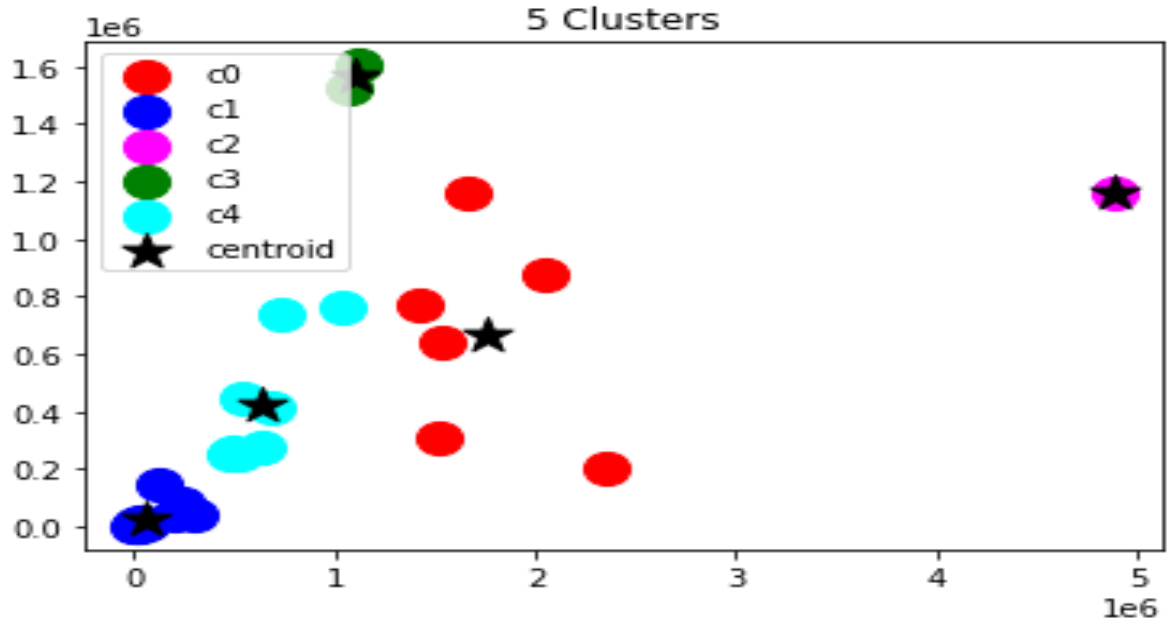




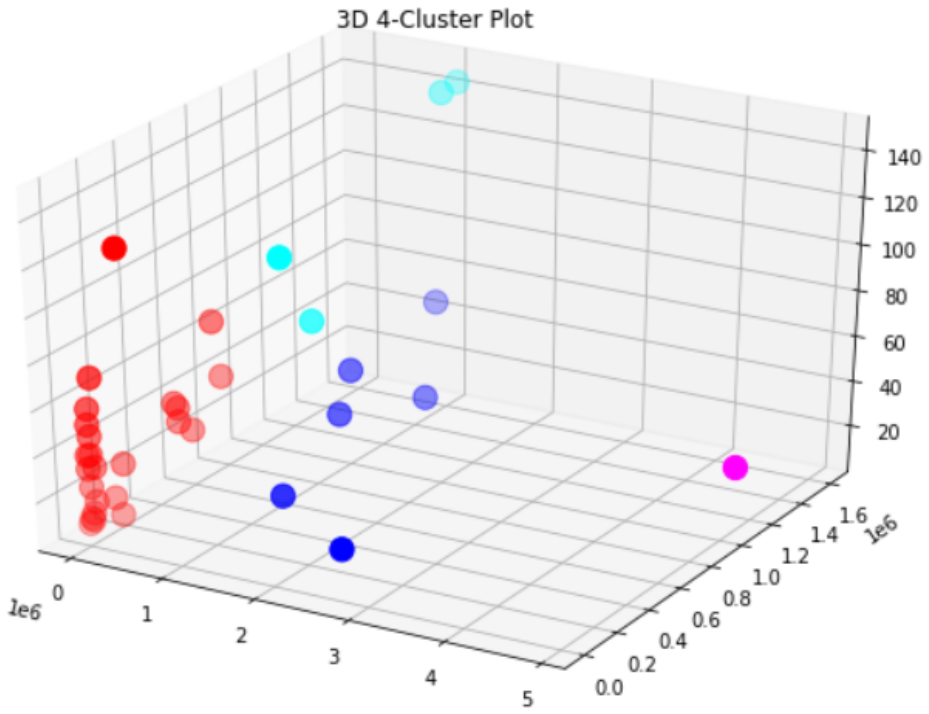
For 5 Cluster:

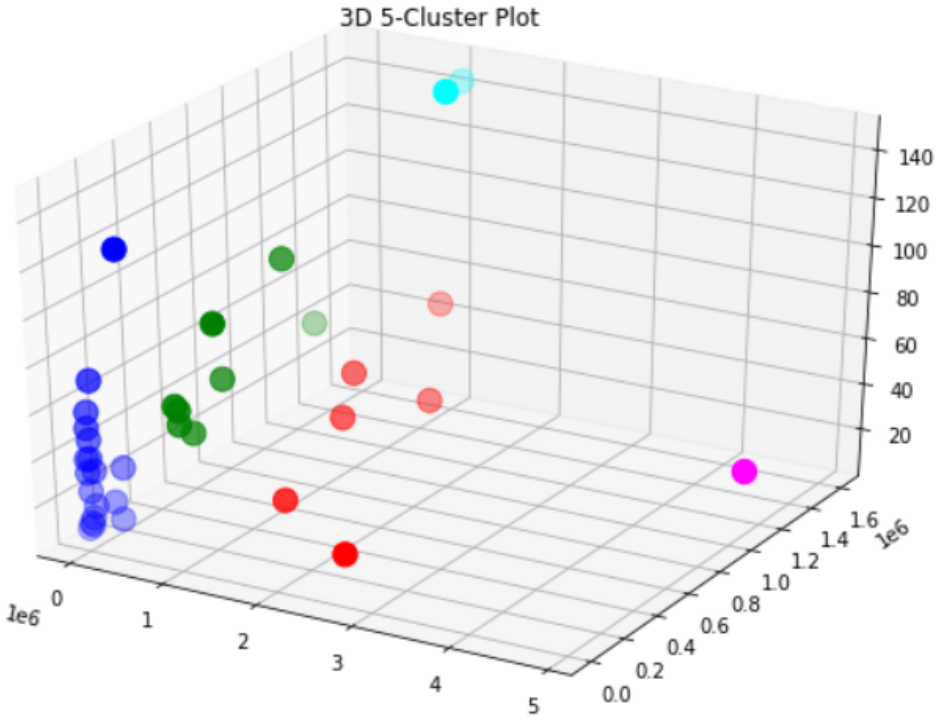
```
kmeans=KMeans(init='k-means++',n_clusters=5,random_state=0)
kmeans.fit(x)
pred1=kmeans.predict(x)
df['clusters5']=pred1
c5=df['clusters5'].value_counts()
g = plt.pie(c5, labels=c5.index,
            autopct='%1.1f%%', shadow=True, startangle=90)
plt.title('5 Clusters')
plt.show()
```





Following are 3D cluster plots for 4 clusters and 5 clusters respectively.





Marketing Mix

Setting the price of the product is both art and science. Most importantly, you must know the cost of production and understand how each parameter affects the production cost, it is the basis for marketing mix. Next you have to tune the parameter affecting your production cost according to the variables in the data collected from the previous sales like which product characteristic are prefer in a segment of the market, specific pricing strategy to attract more buyers, sensitivity or distribution of customers and their preference with respect to prices, customer values, etc. The product price is a huge deciding factor for the consumer to predict the quantity or the quality of the product depending on the product itself. It is recommended to match the quantity or the quality of the product with the overall market outlet to retain customers. To have a complete knowledge of cost of production and understand the relation between production parameters and production cost, it is recommended that keeping complete accurate records for each step of production, i.e. manufacturing, packaging, storage, advertisement, transportation, distribution and sales, will help in optimizing the production parameters for maximum profit.

Basically, there are 4 P's that are widely studied in marketing mix.

1. Product: Market Segmentation will help in finding optimum product and the quantity of it, for maximal profits in a certain geographic segment.

2. Price: Market Analysis and business knowledge in the field of bio-tech industries will help in strategically deciding the price for each product.

3. Place: Market Segmentation will help in figuring out the optimal location or the other channel for distribution (mostly e-commerce sites like 1mg, pharmeasy, etc.) for each product requirement.

4. Promotion: Market Segmentation and Analysis will figure out the segments for which we need to advertise and business intelligence in advertisement industry will help with the method preferred for promotion.

Target Segments

1. According to the 4-cluster model, the segment-colored Red has a low `mean_assesed` and a low `mean_achieved`, implies the requirement in this segment of market is low and the supply is also low, hence the market is not so active in that segment, therefor it's not recommended to enter this segment of market.

2. According to the 4-cluster model, the segment-colored Magenta has a high `mean_assesed` and a high `mean_achieved`, which implies the market there is quite active and is also has a good supply, so the new manufacturer would have to provide the products at a lower price than the present manufacturer and should do a lot of promotion, so if the firm has a good funding, then entering this market segment is recommended.

3. According to the 4-cluster model, the segment-colored Cyan has a low `mean_assesed` and a high `mean_acheived`, implies the supply is much more than the demand, hence it's not recommended to enter this segment of market.

4. According to the 4-cluster model, the ideal market segment to enter is the Blue colored segment. This segment has moderately high `mean_assesed`, and a low `mean_achieved`, implies the demand is moderately high but the supply is low. This is the ideal condition that we would wish for. Hence, any firm is recommended to enter this market segment.
5. So far as the 5-cluster model is concerned, the Blue segment is 4-Cluster model, is same as Red segment in 5-cluster model, implies the recommendations for ideal market segment would remain the same.

Reference

https://github.com/PraneethKolloju/Bio-Tech-Startup-Clustering.git