Market Segmentation Analysis for Agri-Tech Start-up

By:-
Team T1

1. Rahul Chawla
2. Joydeep Baidya
3. Priyakshi Goswami
4. Naman Chib

Abstract

The purpose of this project is analysing the respective market in India using segmentation analysis for an Agri-Tech startup launching a Bio-Pesticide as its first product. Segmentation analysis is an important step before we embark on a marketing plan. It is important to learn how to analyze your audience and market.

Market Overview

The Indian pesticides market reached a value of INR 232 Billion in the year 2020. Pesticides are substances or a mixture of substances intended for preventing, destroying, repelling or mitigating any pest. Pesticides represent the last input in an agricultural operation and are applied for preventing the spoilage of crops from pests such as insects, fungi, weeds, etc., thereby increasing agricultural productivity.

The significance of pesticides has been rising over the last few decades catalyzed by the requirement to enhance the overall agricultural production and the need to safeguard adequate food availability for the continuously growing population in the country. In India, pests and diseases, on an average eat away around 20-25% of the total food produced. Looking forward, it is expected of the Indian pesticides market to exhibit moderate growth during the next five years.

The total available arable land per capita has been reducing in recent years as a result of increasing urbanization levels and is expected to reduce further in the coming years. Driven by rising population levels, food demand is expected to continue increasing in the coming years. We expect pesticides to play a key role in increasing the average crop yields per hectare.

Based on the product type the market has been segmented as synthetic pesticides and biopesticides. Synthetic pesticides currently dominate the market, holding the largest share. The global biopesticides market is projected to grow at a CAGR of 14.7% from an estimated value of USD 4.3 billion in 2020 to reach USD 8.5 billion by 2025. The usage of synthetic chemicals can lead to pollution and contamination of the soil, as well as have harmful effects on the food chain. With this concern, there has been an increase in awareness of residue-free food, due to which high importance is given for biological products.

Due to the factors, such as the ban on chemical pesticides in the major countries attributed to the deteriorating soil conditions, extensive farm practices, and growing concern regarding residue levels in food products, there is high scope and demand for biopesticide products. Biopesticide products are mainly used for fruits & vegetables, cereals & grains, and oilseeds & pulses.

Agricultural Marketing in India

India is an agricultural country and one third population depends on the agricultural sector directly or indirectly. Agriculture remains as the mainstay of the Indian economy since times immemorial. Indian agriculture contribution to the national gross domestic product (GDP) is about 25 per cent. With food being the crowning need of mankind, much emphasis has been on commercialising agricultural production. For this reason,bio-pesticides production and even distribution of food has of late become a high priority global concern.

Agricultural marketing is mainly the buying and selling of agricultural products. In earlier days when the village economy was more or less self-sufficient the marketing of agricultural products presented no difficulty as the farmer sold his produce to the consumer on a cash or barter basis.

There are three marketing functions involved in this, i.e., assembling, preparation for consumption and distribution. Selling any agricultural produce depends on some couple of factors like the demand of the product at that time, availability of storage etc. The products may be sold directly in the market or it may be stored locally for the time being. Moreover, it may be sold as it is gathered from the field or it may be cleaned, graded and processed by the farmer or the merchant of the village. Sometimes processing is done because consumers want it, or sometimes to conserve the quality of that product. The task of the distribution system is to match the supply with the existing demand by wholesaling and retailing in various points of different markets like primary, secondary or terminal markets.

In India, there are several central government organisations, who are involved in agricultural marketing like, Commission of Agricultural Costs and Prices, Food Corporation of India, Cotton Corporation of India, Jute Corporation of India, etc. There are also specialised marketing bodies for rubber, tea, coffee, tobacco, spices and vegetables.

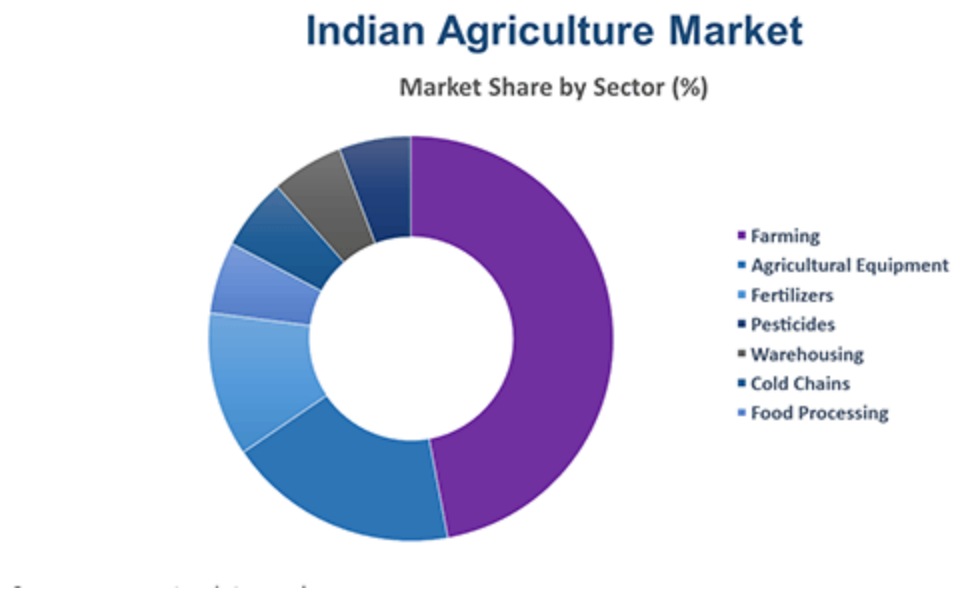
Under the Agricultural Produce (grading and marketing) Act of 1937, more than forty primary commodities are compulsorily graded for export and voluntarily graded for internal consumption. Although the regulation of commodity markets is a function of state government, the directorate of marketing and inspection

provides marketing and inspection services and financial aid down to the village level to help set up commodity grading centers in selected markets.

As we have a tradition of agricultural production, marketing and allied commercial activities, now it is the time for us to brainstorm and come out with new ideas of value added services. These value added services will give the existing agricultural engine a new dimension. The next logical step could be food-processing which not only could be another revenue generating area but also can provide lots of full-time employment to our youths. With the changing agricultural scenario and global competition, there is a need to exploit the available resources at maximum level.

There was a survey undertaken by the directorate of marketing and inspection in the ministry of agriculture in 1970-71 and 1971-72, of five hundred regulated markets was, with a view to assessing the adequacy and efficiency of the existing regulated markets and highlighting their drawbacks and deficiencies and suggesting measures to develop them. One of the most important drawbacks has been the inadequate financial resources of some of the market committees. During the fourth plan, a central sector scheme was drawn up by the ministry of agriculture to provide a grant at 20 percent of the cost of market development of the market, subject to a maximum of Rs. 2 lakhs. The balance will have to be provided by the commercial banks.

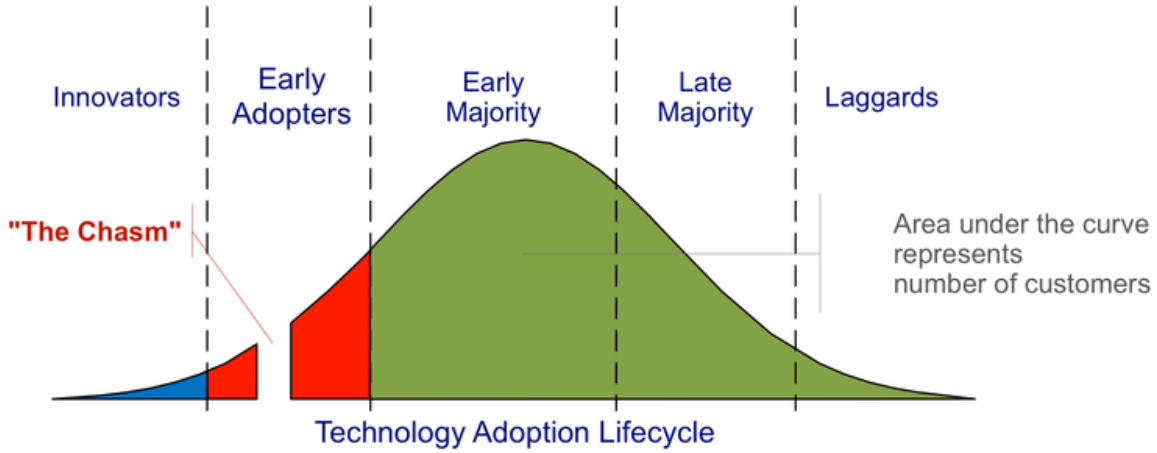
Another important development in the field of regulated markets is the keen interest taken by the International Development Agency (IDA) in the development of the infrastructure in regulated markets. The IDA is financing the development of infrastructure in 50 markets of Bihar.



Adaptation Innovation Life Cycle

The technology adoption lifecycle is a sociological model that describes the adoption or acceptance of a new product or innovation, according to the demographic and psychological characteristics of defined adopter groups. The process of adoption over time is typically illustrated as a classical normal distribution or "bell curve". The model indicates that the first group of people to use a new product is called "innovators", followed by "early adopters". Next come the early majority and late majority, and the last group to eventually adopt a product are called "Laggards" or "phobics." For example, a phobic may only use a cloud service when it is the only remaining method of performing a required task, but the phobic may not have an in-depth technical knowledge of how to use the service.

The demographic and psychological (or "psychographic") profiles of each adoption group were originally specified by the North Central Rural Sociology Committee (Subcommittee for the Study of the Diffusion of Farm Practices) by agricultural researchers Beal and Bohlen in 1957.



The report summarized the categories as:

• **Innovators**– had larger farms, were more educated, more prosperous and more risk-oriented

• **Early adopters** – younger, more educated, tended to be community leaders, less prosperous

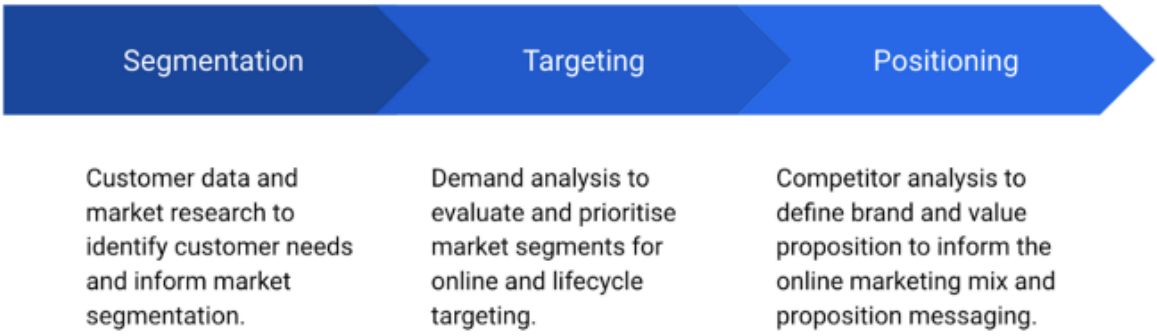
• **Early majority** – more conservative but open to new ideas, active in community and influence to neighbors

• **Late majority** – older, less educated, fairly conservative and less socially active

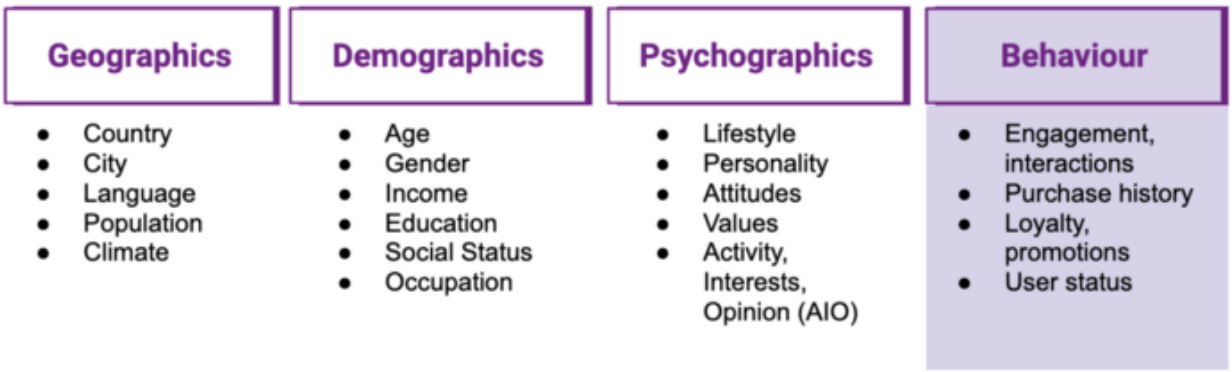
• **Laggards** – very conservative, had small farms and capital, oldest and least educated

Market Segmentation

Segmentation, Targeting and Positioning (STP) is a common strategic model in today's marketing approach. It reflects the increasing popularity of customer centric marketing strategies over product differentiation strategies. The audience focused approach in marketing e.g. helps to deliver more relevant and targeted communications to commercially relevant segments. STP therefore goes hand in hand with marketing personas.



The popularity of segmentation in the strategy derives on the one hand from past limitations of CRM and ad-tech systems as well as a dependency on human decision making in the STP process on the other hand. Before the rise of data driven personalisation most systems could only handle a smaller number of segments and consequently targeting was limited to your general audience segmentation.



Target Market

The target market of Agri-Tech startups can be categorized in terms of demographic, behavioural and psychographic.

With an increasing number of people being productively employed in lucrative industries, they have increased their living standards and made their wallets fatter too. The target group are mainly the middle income segment who have a higher disposable income. Agri-Tech Company targets the population in urban and semi

urban locations because the city dwellers have a very busy life. The customers need to keep up with the hectic schedule and they need immediate, reliable and assistance for their day-to-day tasks.

Why we need Biopesticides in Agriculture

There are many elements that the farming industry has to battle with to grow healthy crops, from irrigation and bad weather to ever changing markets and demands. So the three reasons to biopesticides in India are:

1. *Ease of Use:*

   Chemical pesticides are toxic, and there is always a recommended period for which you are not allowed back in sprayed fields which can be up to 24 hours.

   Biopesticides have been developed to be applied in a similar way as chemical pesticides, which is no longer a barrier to using biological applications. There is no waiting time between spraying a field with a biological product, which means more productive time.

2. *Residue:*

   Spraying crops with pesticides, or using pesticides in the soil, can leave residue on produce, and are potentially toxic to humans depending on the level of exposure.

   Biopesticides are not harmful to humans or the crops, and there is no damage or residue left on the produce. By matching the correct natural enemy to the pest, nature takes its course in destroying the pest and nothing else.

3. *Resistance:*

   Over time and a multitude of applications, pests build up genetic resistance to widely used pesticides. They have adapted over a period to become resistant to the pesticides, so the chemicals become less and less effective. This is a huge problem for growers, as they are in danger of losing existing methods to manage their pest problems.

   Biopesticides are a dependable solution in this instance, as pests don’t develop ‘chemical resistance’ to biopesticides so these products don’t become less effective over time.

With the dissolution of barriers around practicalities of using biopesticides, this method of pest management is seeing growth in the agricultural sector. With biopesticides being non-toxic to humans and not harmful to crops or neighbouring species, we should see this over taking chemical pesticides as the pest management tool of choice.

Analysis and Approaches used in the Project

In the project we have broken down our analysis approach into two parts, in the first part we worked on the dataset of crop production in India and in second part we have again taken the different agriculture dataset and analysed it using K-means clustering.

PART 1

Link to the Dataset:
https://www.kaggle.com/abhinand05/crop-production-in-india

Dataset Description: The dataset consists of textual as well as numerical attributes.

```
df = pd.read_csv('crop_production.csv')
```

```
df.head()
```

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254.0	2000.0
1	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2.0	1.0
2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102.0	321.0
3	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176.0	641.0
4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720.0	165.0

The dataset comprises 246091 rows and 7 attributes and out of which 3730 instances are missing in the *Production* attribute.

```
df.shape
```

```
(246091, 7)
```

```
df.isnull().sum()
```

```
State_Name          0
District_Name       0
Crop_Year           0
Season              0
Crop                0
Area                0
Production       3730
dtype: int64
```

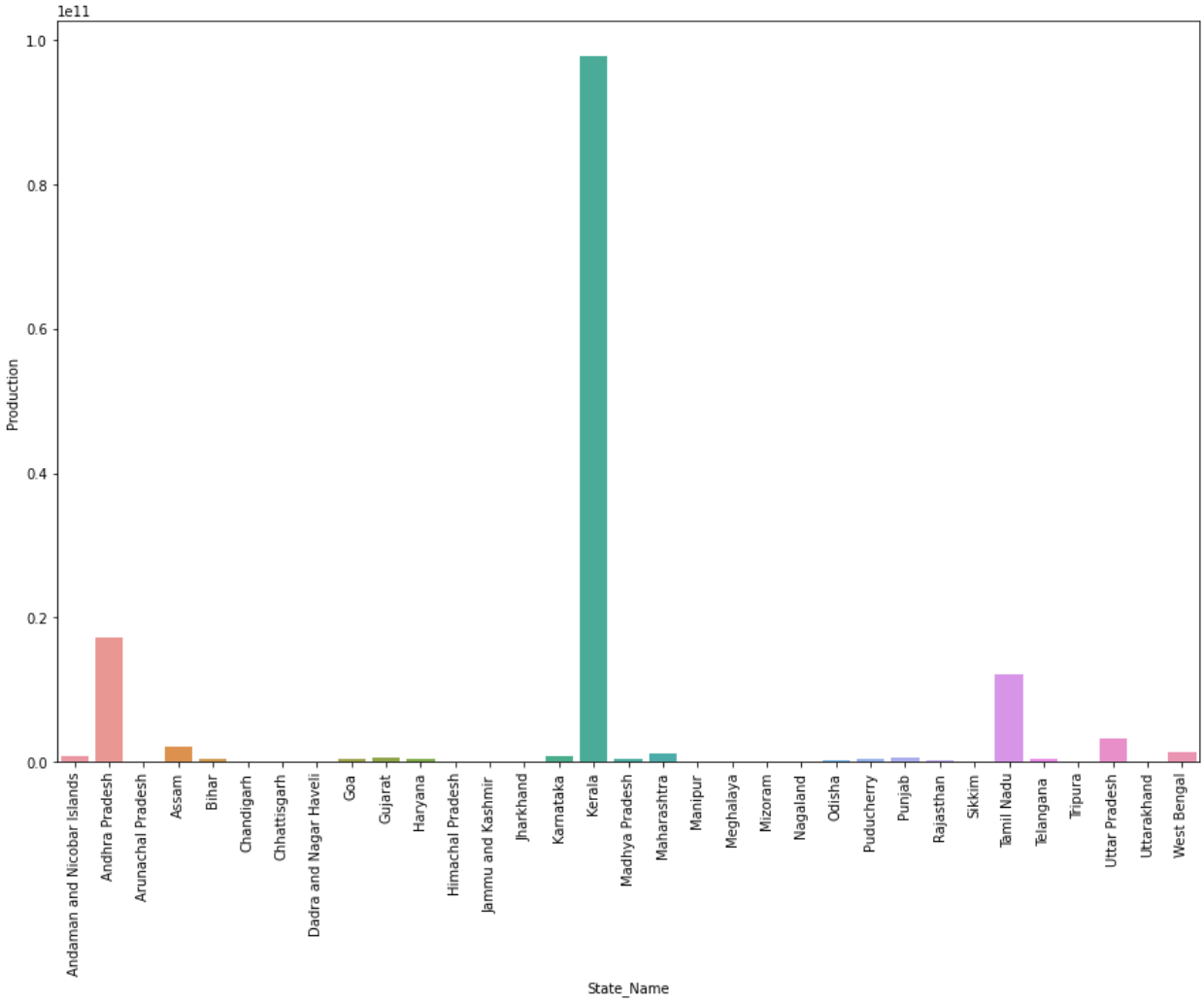
Exploratory data analysis

From the EDA, it has been found that the top 5 states with highest production were Kerala, Andhra Pradesh, Tamil Nadu, Uttar Pradesh and Assam with their corresponding production values from the year 1997-2014 as shown below,

```
state.nlargest(5, columns = 'Production', keep = 'first')
```

	State_Name	Production
15	Kerala	9.788005e+10
1	Andhra Pradesh	1.732459e+10
27	Tamil Nadu	1.207644e+10
30	Uttar Pradesh	3.234493e+09
3	Assam	2.111752e+09

and the barplot for the same is shown below,



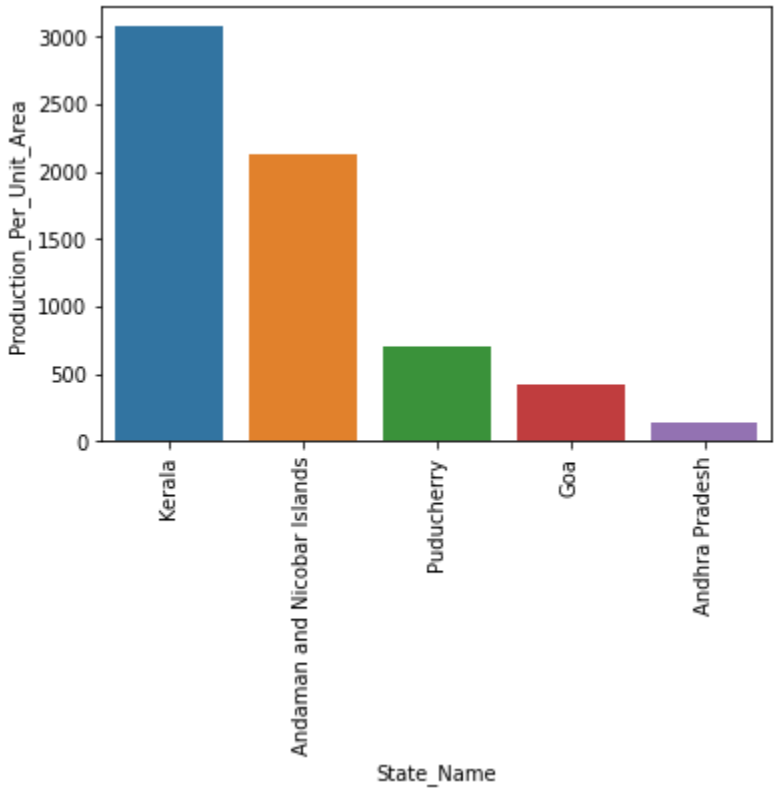
The production per unit area of different states have been analysed and it has been found that,

Kerala is the state where production per unit area is the highest and after that Andaman and Nicobar Islands being the second in terms of production per unit

area.

State_Name	Area	Production	Production_Per_Unit_Area
Kerala	3.180225e+07	9.788005e+10	3077.770880
Andaman and Nicobar Islands	3.370834e+05	7.182232e+08	2130.698931
Puducherry	5.487360e+05	3.847245e+08	701.110374
Goa	1.205678e+06	5.057558e+08	419.478126
Andhra Pradesh	1.315073e+08	1.732459e+10	131.738602

The barplot of the same is shown below,



PART 2

Link to Dataset: [All Agriculture related Datasets for India | Kaggle]

Dataset Description: The dataset consists of numerical as well as textual data.

Methodology Used in Performing Market Segmentation for Agri-Tech Startup

Clustering: Clustering is one of the most common exploratory data analysis techniques used to get an intuition about the structure of the data. It can be defined as the task of identifying subgroups in the data such that data points in the same subgroup (cluster) are very similar while data points in different clusters are very different. In other words, we try to find homogeneous subgroups within the data such that data points in each cluster are as similar as possible according to a similarity measure such as euclidean-based distance or correlation-based distance. The decision of which similarity measure to use is application-specific. Clustering analysis can be done on the basis of features where we try to find subgroups of samples based on features or on the basis of samples where we try to find subgroups of features based on samples.

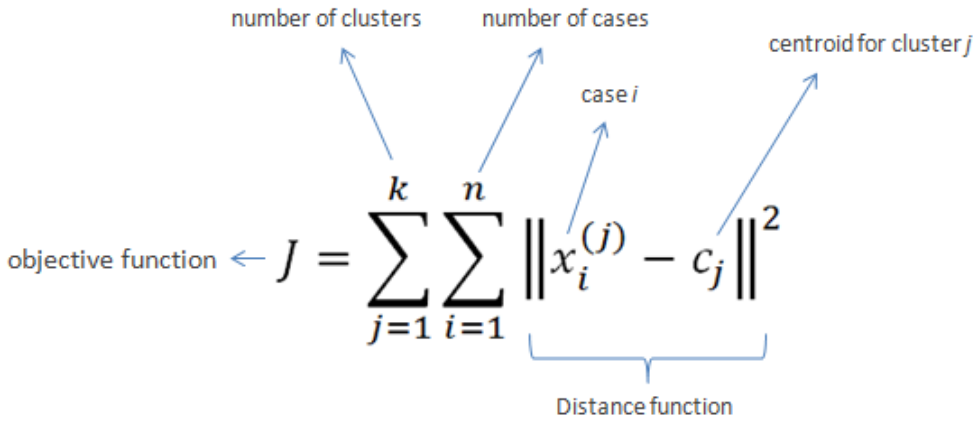
K-Means Algorithm: It is an iterative algorithm that tries to partition the dataset into pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid (arithmetic mean of all the data points that belong to that cluster) is at the minimum. The less variation we have within clusters, the more homogeneous (similar) the data points are within the same cluster.

The way k means algorithm works is as follows:

1. Specify number of clusters K.
2. Initialize centroids by first shuffling the dataset and then randomly selecting K data points for the centroids without replacement.
3. Keep iterating until there is no change to the centroids. i.e assignment of data points to clusters isn't changing.

The approach k-means follows to solve the problem is expectation maximization The E-step is assigning the data points to the closest cluster. The M-step is computing the centroid of each cluster. Below is a break down of how we can solve it mathematically

The objective function is:



Basic Libraries used in Analysis

1. Import numpy as np
2. Import pandas as pd
3. Import matplotlib.pyplot as plt
4. Import seaborn as sns

Numpy has been imported for performing mathematical operations. Pandas library has been used for dealing with dataframe. Matplotlib and seaborn have been used for basic visualisations.

Results and Discussion

Loading data file

```
In [2]: df=pd.read_csv('data.csv',index_col=0)
        df.head()
```

Out[2]:

	state	district	market	commodity	variety	arrival_date	min_price	max_price	modal_price
0	Andaman and Nicobar	South Andaman	Port Blair	Amaranthus	Amaranthus	04/03/2019	6000	8000	7000
1	Andaman and Nicobar	South Andaman	Port Blair	Banana - Green	Banana - Green	04/03/2019	4500	5500	5000
2	Andaman and Nicobar	South Andaman	Port Blair	Bhindi(Ladies Finger)	Bhindi	04/03/2019	6000	8000	7000
3	Andaman and Nicobar	South Andaman	Port Blair	Bitter gourd	Other	04/03/2019	6000	8000	7000
4	Andaman and Nicobar	South Andaman	Port Blair	Black pepper	Other	04/03/2019	110000	130000	120000

The given data has been read and the first five rows are displayed.

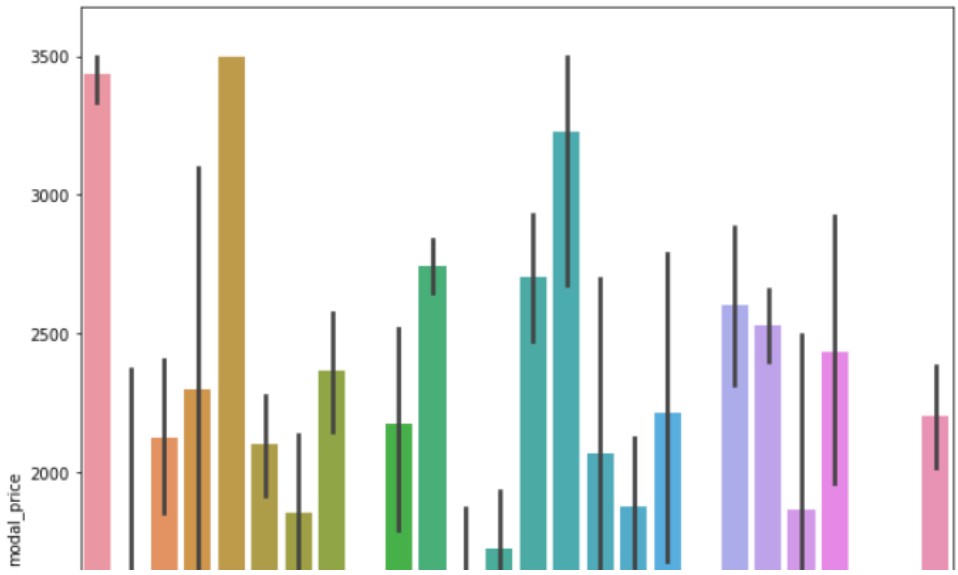
```
In [10]: print(df.min_price.quantile(0.75))
         print(df.min_price.quantile(0.25))

         3200.0
         800.0

In [11]: df.loc[df['min_price']>=3200,'min_price']=3200

In [12]: print(df.max_price.quantile(0.75))
         print(df.max_price.quantile(0.25))

         3700.0
         1000.0

In [13]: df.loc[df['max_price']>=3700,'max_price']=3200

In [14]: print(df.modal_price.quantile(0.75))
         print(df.modal_price.quantile(0.25))

         3500.0
         950.0

In [15]: df.loc[df['modal_price']>=3700,'modal_price']=3500
```

Using the quantile method the lowest and highest price has been calculated and the values lesser than the lowest are set to the lowest and the values greater than the greatest are set to the greatest to remove outliers.

```
In [17]: df['state'].value_counts()

Out[17]: Uttar Pradesh          574
         Kerala                 274
         Tamil Nadu             183
         Punjab                 180
         Maharashtra            162
         West Bengal            162
         Gujarat                113
         Himachal Pradesh       102
         Odisha                  84
         Haryana                 65
         Assam                   58
         Manipur                 52
         Madhya Pradesh          44
         Rajasthan               41
         Karnataka               36
         Uttrakhand              21
         Tripura                 18
         Andaman and Nicobar     15
         Telangana               15
         Andhra Pradesh           7
         Pondicherry              7
         Meghalaya                6
         Nagaland                 6
         Chattisgarh              5
         Jammu and Kashmir        3
```

States with most agriculture in decreasing order.

```
In [18]: plt.figure(figsize=(10,12))
         sns.barplot(x='state',y='modal_price',data=df)

Out[18]: <AxesSubplot:xlabel='state', ylabel='modal_price'>
```



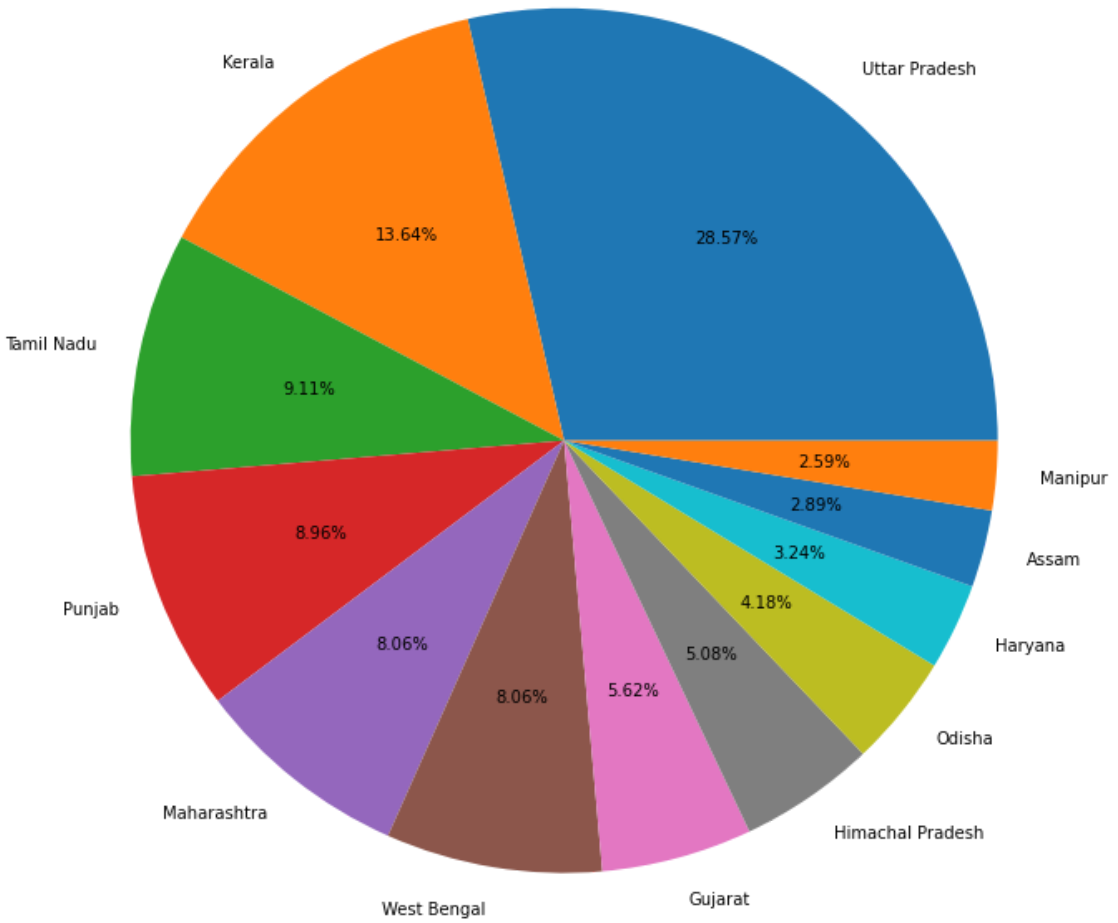
Price capped for crops in each state.

In [21]:
```
df['arrival_date']=pd.to_datetime(df['arrival_date'])
```

In [22]:
```
df.info()
```

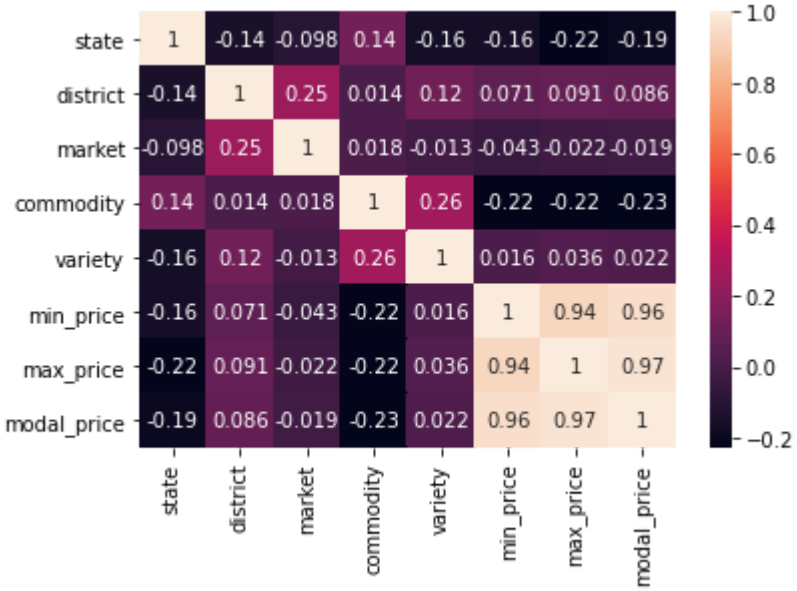
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2236 entries, 0 to 2237
Data columns (total 9 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   state         2236 non-null   object
 1   district      2236 non-null   object
 2   market        2236 non-null   object
 3   commodity     2236 non-null   object
 4   variety       2236 non-null   object
 5   arrival_date  2236 non-null   datetime64[ns]
 6   min_price     2236 non-null   int64
 7   max_price     2236 non-null   int64
 8   modal_price   2236 non-null   int64
dtypes: datetime64[ns](1), int64(3), object(5)
memory usage: 239.2+ KB
```

In [23]:
```
plt.figure(figsize=(8,8))
plt.pie(df.state.value_counts().to_list()[:12],labels=df.state.value_counts().index[:12],radius=1.5,autopct="%0.2f%%")
```

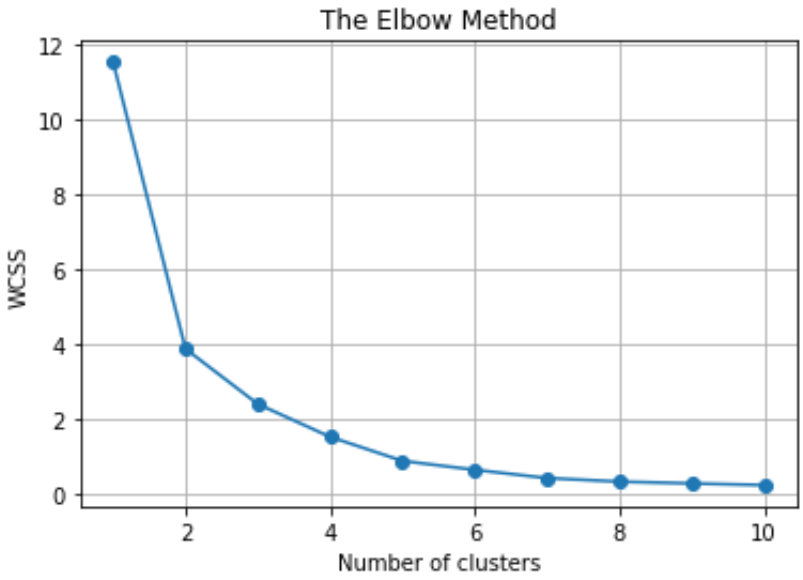


Top 12 states where most of the agriculture takes place in our country with the percentage of agriculture carried out in the respective state.

```
In [26]: obj=[c for c in df.columns if df[c].dtype=='object']
         obj

Out[26]: ['state', 'district', 'market', 'commodity', 'variety']

In [27]: from sklearn.preprocessing import LabelEncoder
         le=LabelEncoder()

In [28]: for i in obj:
             df[i]=le.fit_transform(df[i])

In [29]: sns.heatmap(df.corr(),annot=True)
```



In [36]:
```
X = x.iloc[:,[0,7]].values
from sklearn.cluster import KMeans
wcss = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters = i, init = 'k-means++',max_iter = 1500, n_init = 20)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)
plt.plot(range(1, 11), wcss, marker = 'o')
plt.title('The Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.grid()
plt.show()
```



The Elbow method has been used to find the most optimum clusters for the data. 2 clusters are the optimum clusters.

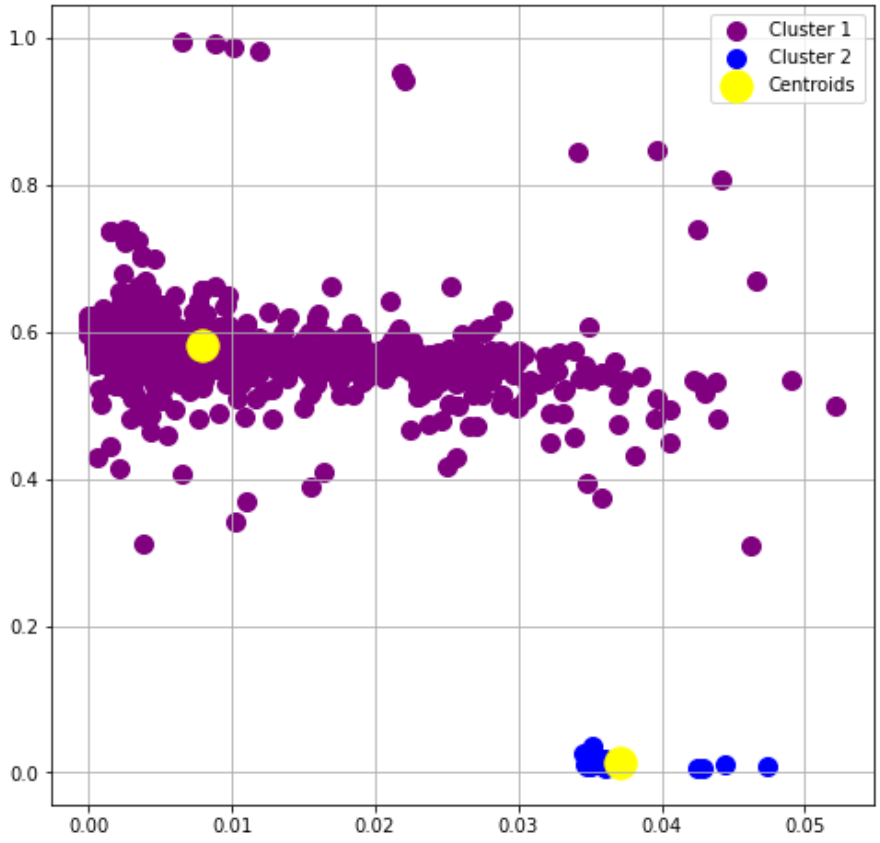
```
In [37]: kmeans = KMeans(n_clusters =2, init = 'k-means++', random_state = 69)
         y_kmeans = kmeans.fit_predict(X)
```

```
In [38]: from sklearn.metrics import silhouette_score
         score = silhouette_score(X, kmeans.labels_, metric='euclidean')
         print(score)
```

```
0.9325561631131797
```

```
In [39]: plt.figure(figsize=(8,8))
         plt.scatter(X[y_kmeans == 0, 0], X[y_kmeans == 0, 1],s = 100, c = 'purple', label = 'Cluster 1')
         plt.scatter(X[y_kmeans == 1, 0], X[y_kmeans == 1, 1], s = 100, c = 'blue', label = 'Cluster 2')
         plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1], s = 300, c = 'yellow', label = 'Centroids')
         plt.legend()
         plt.grid()
```

Silhouette score has been calculated to predict how well clusters are fitted over the given data.



Respective clusters and their cluster centers have been plotted. The whole data has been divided into two clusters with most of the customers falling under one cluster

and the rest into the remaining one. Therefore the conditions favour most of the customers.

Conclusion

It is virtually impossible to satisfy all customers, so it is up to the company to select the specific parts of the market which they can best serve. Therefore, businesses could identify market segments, select a few profitable segments, and develop products and marketing mixes that are aimed at particular customers. In this project we have done segmentation of the agriculture market in India.

References

https://www.domain-b.com/economy/agriculture/20040713_marketing.html

https://en.wikipedia.org/wiki/Technology_adoption_life_cycle

https://towardsdatascience.com/data-science-powered-segmentation-models-ae89f9bd405f?gi=1ec64af6e25

https://bionema.com/5-reasons-to-use-biopesticides-in-agriculture