**ABSTRACT**

Customer segmentation and pattern extraction are one of the key aspects of business decision support system. Today's business runs based on such innovation has the ability to enteral the customers with the products, but with such a large raft of products leave the customers confounded, what to buy and what to not and also the companies are nonplussed about what section of customers to target to sell their products.

In order to grow the business intelligently in a competitive market, identification of potential customers should be done timely. This is where machine learning comes into play, various algorithms are applied for unraveling the hidden patterns in the data for better decision making for the future. This elude concept of which segment to target is made unequivocal by applying segmentation. The process of segmenting the customers with similar behaviors into the same segment and with different patterns into different segments is called customer segmentation [1]. In this paper K-means clustering algorithm has been implemented to segment the customers and finally compare the results of the clusters obtained from the algorithms.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Contents** | **Page No.** |
| **1** | **INTRODUCTION** | 1 |
|  | 1.1 Problem Statement | 2 |
| **2** | **EXISTING METHODS** | 5 |
| **3** | **PROPOSED SYSTEM** | 10 |
|  | 3.1 Architecture | 11 |
|  | 3.2 KMeans Algorithm | 11 |
|  | 3.3 Elbow Method | 14 |
|  | 3.4 Tools and Technologies used | 16 |
| **4** | **METHODOLOGY** | 17 |
| **5** | **IMPLEMENTATION** | 27 |
| **6** | **CONCLUSION** | 33 |

**CHAPTER 1**

**INTRODUCTION**

* 1. **INTRODUCTION**

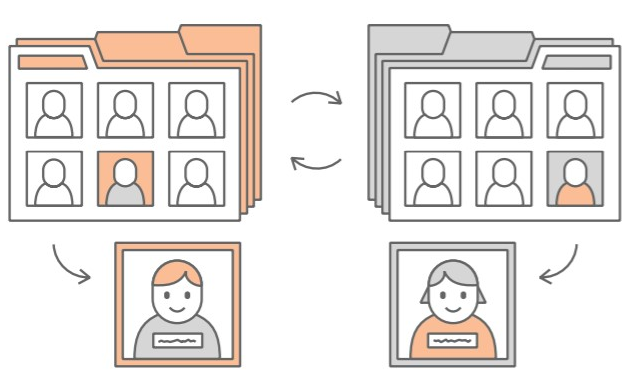
**1.1 PROBLEM STATEMENT**

The importance of treating customers as the principal asset of an organization is increasing in value. Organizations are rapidly investing in developing strategies for better customer acquisition, maintenance and development. The concept of business intelligence has a crucial role to play in making it possible for organizations to use technical expertise for acquiring better customer insight for outreach programs.

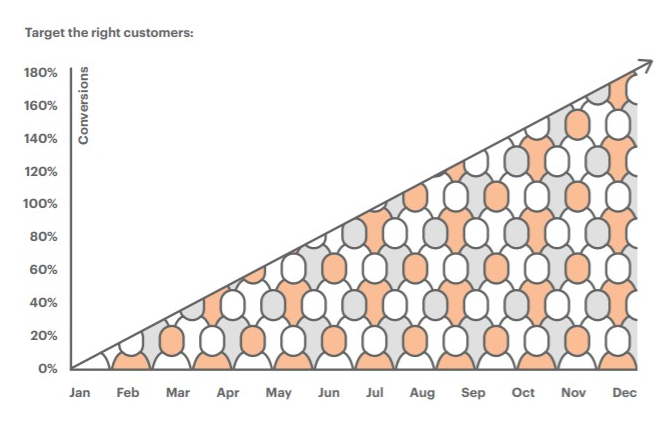
One of the most useful techniques in business analytics for the analysis of consumer behavior and categorization is customer segmentation. By using clustering techniques, customers with similar means, end and behavior are grouped together into homogeneous clusters. Customer Segmentation helps organizations in identifying or revealing distinct groups of customers who think and function differently and follow varied approaches in their spending and purchasing habits. customer segmentation is a method of analyzing a client base and grouping customers into categories or segments which share particular attributes. Key differentials in segmentation include age, gender, education, location, spending patterns and socio-economic group. Relevant differentials are those which are expected to influence customer behavior in relation to a business. The selected criteria are used to create customer segments with similar values, needs and wants.

Clustering techniques reveal internally homogeneous and externally heterogeneous groups. Customers vary in terms of behavior, needs, wants and characteristics and the main goal of clustering techniques is to identify different customer types and segment the customer base into clusters of similar profiles so that the process of target marketing can be executed more efficiently. This study aims to explore the avenues of using customer segmentation, as a business intelligence tool within the CRM framework as well as the use of clustering techniques for helping organizations redeem a clearer picture of the valuable customer base [2].

Think of an online pet supply store as an example: Its customers are very diverse, with very different needs and preferences. What they buy is highly dependent on the type of pets they own, their pets’ ages and dietary needs, their lifestyles and income, and how they interact with their pets. From a marketing perspective, it wouldn’t make sense for a pet supply store to communicate to its entire customer base in the same way. But customer segmentation is about more than matching customers with appropriate product offers. It’s also about changing the way you communicate with your customers based on what you know about them. It’s about identifying your most profitable customers and tailoring your products and services to meet their specific needs. Ultimately, customer segmentation is about creating relevant shopping experiences that build brand loyalty.



Customers are becoming more sophisticated in how they navigate their shopping choices, and online retailers are discovering that one-size-fits-all marketing approaches aren’t so effective any more. In fact, targeting the wrong customers can cost you—not just in wasted marketing dollars, but in higher operational costs associated with processing product returns, handling customer service calls, and responding to lackluster customer reviews. Conversely, targeting the right customers (with the right messages at the right time) can pay off in terms of higher conversion rates, higher average order values, and increased profits. Targeting the right customers can also lead to brand advocacy and word-of-mouth advertising, valuable product insights and greater overall customer satisfaction[3].



**CHAPTER 2**

**EXISTING METHODS**

**2. EXISTING METHODS**

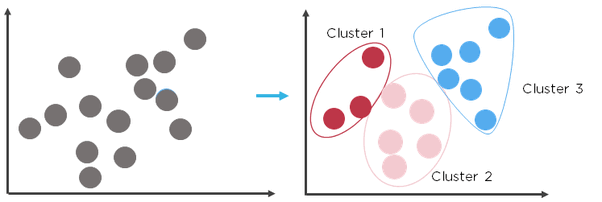
Customer segmentation divides your lists into groups based on common features that tend to predict buying habits, such as demographics or interests, in order to better serve the customer. This is called a priori segmentation– a priori is Latin for from the former, and basically means that you’ve deducted these segments based on anecdotal knowledge or observed trends in your marketing efforts. Here are some a priori segments that’s been used: Since the task of clustering is subjective, the means that can be used for achieving this goal are plenty. Every methodology follows a different set of rules for defining the ‘*similarity’* among data points. In fact, there are more than 100 clustering algorithms known. But few of the algorithms are used popularly, let’s look at them in detail:

* **Connectivity models:** As the name suggests, these models are based on the notion that the data points closer in data space exhibit more similarity to each other than the data points lying farther away. These models can follow two approaches. In the first approach, they start with classifying all data points into separate clusters & then aggregating them as the distance decreases. In the second approach, all data points are classified as a single cluster and then partitioned as the distance increases. Also, the choice of distance function is subjective. These models are very easy to interpret but lacks scalability for handling big datasets. Examples of these models are hierarchical clustering algorithm and its variants.
* **Centroid models:** These are iterative clustering algorithms in which the notion of similarity is derived by the closeness of a data point to the centroid of the clusters. K-Means clustering algorithm is a popular algorithm that falls into this category. In these models, the no. of clusters required at the end have to be mentioned beforehand, which makes it important to have prior knowledge of the dataset. These models run iteratively to find the local optima.
* **Distribution models:** These clustering models are based on the notion of how probable is it that all data points in the cluster belong to the same distribution (For example: Normal, Gaussian). These models often suffer from overfitting. A popular example of these models is Expectation-maximization algorithm which uses multivariate normal distributions.
* **Density Models:**These models search the data space for areas of varied density of data points in the data space. It isolates various different density regions and assign the data points within these regions in the same cluster. Popular examples of density models are DBSCAN and OPTICS.

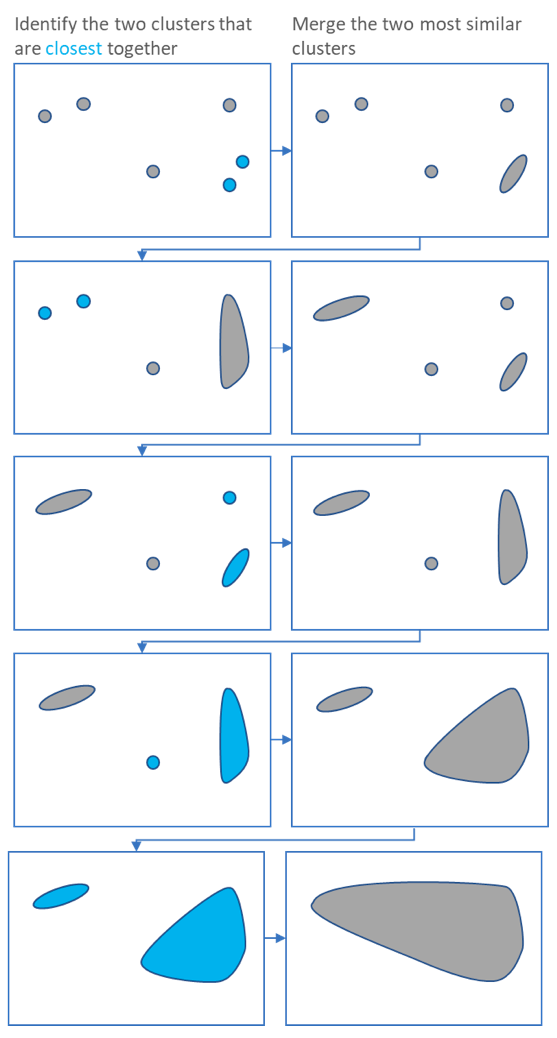
**Techniques Used: -**

Which algorithm to be use depends on which type of clustering we want – Flat or hierarchical? Even within the flat category the algorithm depends on whether we want the clustering to be hard or soft. (Hard meaning an item belongs to exactly one cluster) [12].

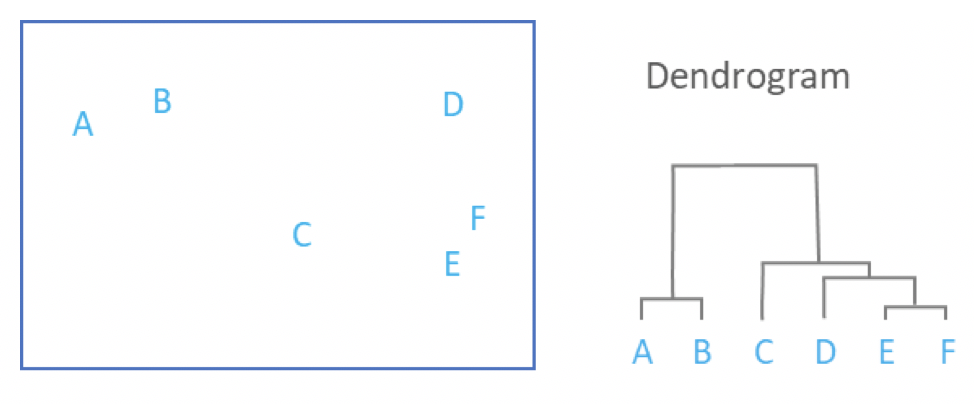
**Flat Clustering**: In Flat clustering, we have to tell the machine how many categories to cluster the data into. Here we tell the data must be clustered into 3 categories.



Algorithms suited to hard clustering include k-means clustering, dbscan, connecting components clustering and other. Algorithms for soft clustering are often softened version of the hard-clustering ones. One such is soft – also called probabilistic – k-means clustering. This is based on powerful algorithm call the EM algorithm of which (hard) k-means clustering is a special case.

**Hierarchical Clustering:** This is an algorithm that groups similar objects into groups called *clusters*. The endpoint is a set of clusters*,*where each cluster is distinct from each other cluster, and the objects within each cluster are broadly similar to each other. 

The main output of Hierarchical Clustering is a **dendrogram***,*which shows the hierarchical relationship between the clusters:



Popular hierarchical clustering algorithms include agglomerative (bottom -up) clustering or divisive(top-down) clustering. In the former, items are first clustered into tight clusters, then clusters are clustered into coarser clusters. This process is repeated until we have a tree of cluster. Its leaves are the items. The tree’s root is a single cluster representing the entire data set. Divisive clustering works in the opposite direction. It starts from the cluster representing the entire data set; a cluster is then split into two. The resulting clusters become the children of the former. Thus, the tree of clusters is grown top- down.[6]

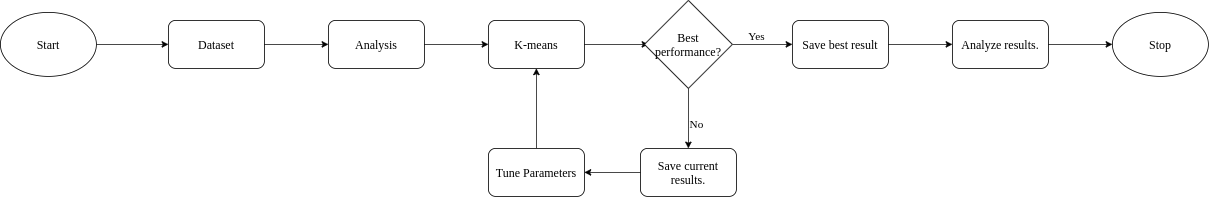
Although, hierarchical clustering algorithm isn’t that useful for some scenarios, [8] have used it for intelligent customer segmentation and [9] have made use of it for applying clustering algorithms on the transaction data from a supermarket.

**CHAPTER 3**

**PROPOSED SYSTEM**

**3. PROPOSED SYSTEM**

**3.1 Architecture**

****

The figure shown above portrays the flow of our system. Initially, the dataset was collected from the given link. This step was followed by loading the dataset into the system and analysing it. The entire system was developed using python programming language. The dataset has 200 datapoints and about 5 columns, we removed some irrelevant or undesired columns and then conducted a detailed analysis on the dataset. The proceeding step involves a loop of applying K-means, evaluating the performance, and tuning the hyperparameters, this process is known as Elbow method. Once the optimal score is achieved, we analyse all the results including the results with suboptimal performance. After analysis, we fine-tuned our verdicts as per the final clustering results and presented the results. The subsequent sections describe the general overview of the K-means clustering algorithm followed by the Elbow method.[7]

**3.2 K-Means**

K-means[10] is a centroid-based algorithm, or a distance-based algorithm, where we calculate the distances to assign a point to a cluster. In K-Means, each cluster is associated with a centroid. The main objective of the K-Means algorithm is to minimize the sum of distances between the points and their respective cluster centroid[4].

Let’s now take an example to understand how K-Means actually works:

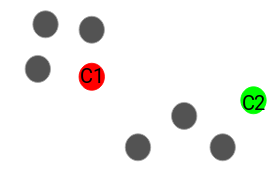
[](https://cdn.analyticsvidhya.com/wp-content/uploads/2019/08/Screenshot-from-2019-08-09-12-21-43.png)  
We have these 8 points and we want to apply k-means to create clusters for these points. Here’s how we can do it.

**Step 1: Choose the number of clusters *k***

The first step in k-means is to pick the number of clusters, k.

**Step 2: Select k random points from the data as centroids**

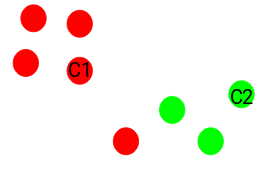
Next, we randomly select the centroid for each cluster. Let’s say we want to have 2 clusters, so k is equal to 2 here. We then randomly select the centroid:

[](https://cdn.analyticsvidhya.com/wp-content/uploads/2019/08/Screenshot-from-2019-08-09-12-23-55.png)

Here, the red and green circles represent the centroid for these clusters.

**Step 3: Assign all the points to the closest cluster centroid**

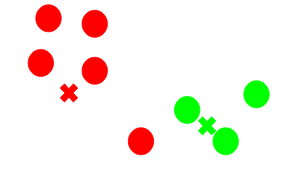
Once we have initialized the centroids, we assign each point to the closest cluster centroid:

[](https://cdn.analyticsvidhya.com/wp-content/uploads/2019/08/Screenshot-from-2019-08-09-12-24-35.png)

Here you can see that the points which are closer to the red point are assigned to the red cluster whereas the points which are closer to the green point are assigned to the green cluster.

**Step 4: Recompute the centroids of newly formed clusters**

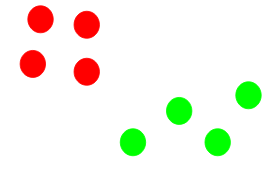
Now, once we have assigned all of the points to either cluster, the next step is to compute the centroids of newly formed clusters:

[](https://cdn.analyticsvidhya.com/wp-content/uploads/2019/08/Screenshot-from-2019-08-09-12-26-59.png)

Here, the red and green crosses are the new centroids.

**Step 5: Repeat steps 3 and 4**

We then repeat steps 3 and 4:

[](https://cdn.analyticsvidhya.com/wp-content/uploads/2019/08/Screenshot-from-2019-08-09-12-28-14.png)

The step of computing the centroid and assigning all the points to the cluster based on their distance from the centroid is a single iteration.

**Stopping Criteria for K-Means Clustering**

There are essentially three stopping criteria that can be adopted to stop the K-means algorithm:

1. Centroids of newly formed clusters do not change
2. Points remain in the same cluster
3. Maximum number of iterations are reached

We can stop the algorithm if the centroids of newly formed clusters are not changing. Even after multiple iterations, if we are getting the same centroids for all the clusters, we can say that the algorithm is not learning any new pattern and it is a sign to stop the training.

Another clear sign that we should stop the training process if the points remain in the same cluster even after training the algorithm for multiple iterations.

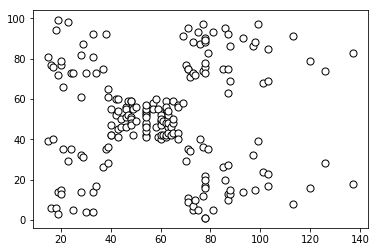
Finally, we can stop the training if the maximum number of iterations is reached. Suppose if we have set the number of iterations as 100. The process will repeat for 100 iterations before stopping.[5]

**3.3 Elbow method**

**3.3.1 Choosing the number of clusters**

A fundamental step for any unsupervised algorithm is to determine the optimal number of clusters into which the data may be clustered. The Elbow Method is one of the most popular methods to determine this optimal value of k.

Consider the scatter plot like these consisting of Annual Income On x-axis and Spending score (1 to 100) on y-axis:



From the above visualization, we can see that the optimal number of clusters should be around 5. But visualizing the data alone cannot always give the right answer. Hence, we demonstrate the following steps.

We now define the following: -

1. **Distortion:** It is calculated as the average of the squared distances from the cluster centres of the respective clusters. Typically, the Euclidean distance metric is used.
2. **Inertia:** It is the sum of squared distances of samples to their closest cluster centre.

**3.4 Tools and Technologies used**

Tools Used:

* Jupyter Notebook - It is used to create and share documents that contain live code, equations, visualizations etc.
* Python Programming Language

Libraries Used:

* Pandas (0.23.0) - It is used for data manipulation and analysis
* Matplotlib (2.2.2) – It is plotting library usually used for data visualization.
* NumPy (1.14.3) – It is used for performing number of mathematical operations on arrays. Such as statistical, algebraic etc.
* Sklearn (0.19.1)– It is used for accessing various types of machine learning algorithms. E.g. SVM, KNN, Kmeans etc.
* Seaborn (0.8.1) – It is use same as matplotlib for data visualization.

**CHAPTER 4**

**METHODOLOGY**

**4. METHODOLOGY**

In this section we’ll describe the analysis done using the flow as specified in previous section.

**Dataset:**

We tested KMeans algorithm on the dataset of customers which is received from the Mall. The dataset consists of 5 features namely Customer ID, Gender, Age, Annual Income, Spending Scores and total of 200 tuples. In Data Preprocessing stage we removed out Customer ID Column from the testing dataset since that column isn’t adding any valuable contribution to the formation of clusters later. Later Gender column consist of two string datatype values ‘Male’ and ‘Female’. Since numeric computations couldn’t been performed on string data types we convert the columns into binary values representing Male as 1 and Female as 0.

**Data Analysis:**

We describe some common mathematical features of our given dataset, given in the following figure.

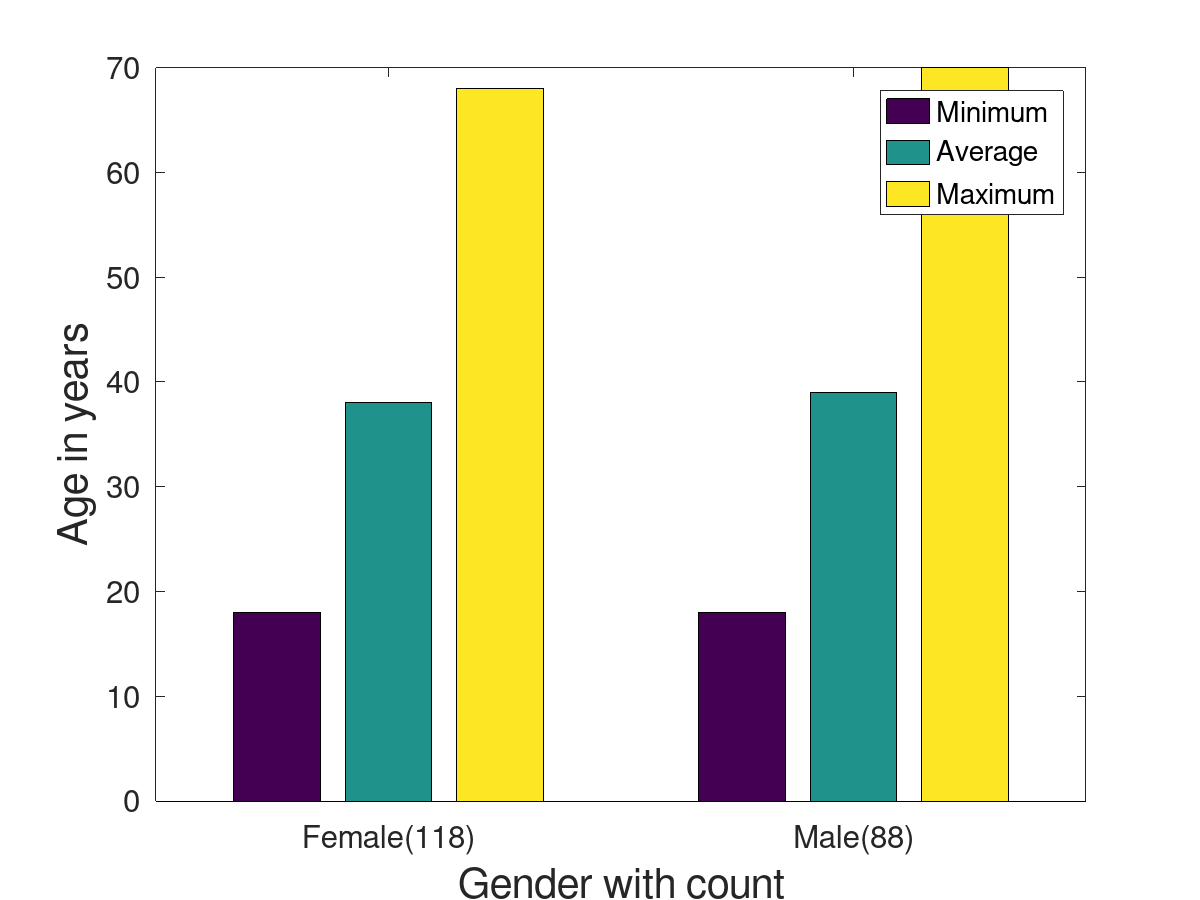
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Gender** | **Age** | **Annual Income (k$)** | **Spending Score (1-100)** | **Clusters** |
| count | 200 | 200 | 200 | 200 | 200 |
| mean | 0.44 | 38.85 | 60.56 | 50.2 | 1.78 |
| std | 0.497632586 | 13.96900733 | 26.26472117 | 25.82352167 | 1.195300681 |
| min | 0 | 18 | 15 | 1 | 0 |
| max | 1 | 70 | 137 | 99 | 4 |

Here for Gender column, mean value 0.44 represents that number of Female are more as compared to number of males. For Age column mean value is 38.85 representing that average age of people shopping at mall is around mid-age.

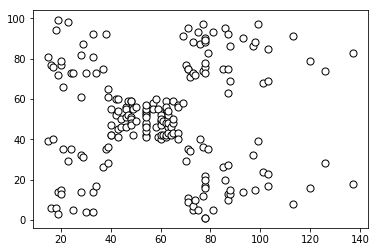
**Plots of various features:**

a) Visualizing Gender(x-axis) and Age(y-axis):





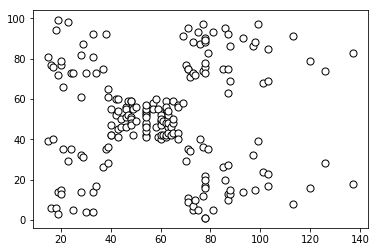
b) Visualizing Annual Income(x-axis) and Spending Scores(y-axis):



**Analyzing the Optimal number of clusters using Elbow Method:**

The Elbow Method is one of the most popular methods to determine this optimal value of k [11].

Consider the scatter plot like these consisting of Annual Income On x-axis and Spending score (1 to 100) on y-axis:



We iterate the values of k from 1 to 9 and calculate the values of distortions for each value of k and calculate the distortion and inertia for each value of k in the given range.

**After Tabulating and Visualizing results for Distortions:**

1: 35.58782757352547

2: 30.122051138558877

3: 24.020507648685765

4: 20.47612410788973

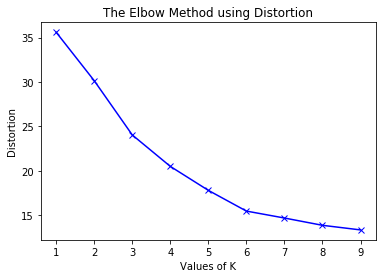
5: 17.773099154184546

6: 15.442648884708749

7: 14.573101197758696

8: 13.850513224645463

9: 13.281042253254888



**After Tabulating and Visualizing results for Inertia:**

1: 308862.06

2: 212889.44245524297

3: 143391.59236035682

4: 104414.67534220166

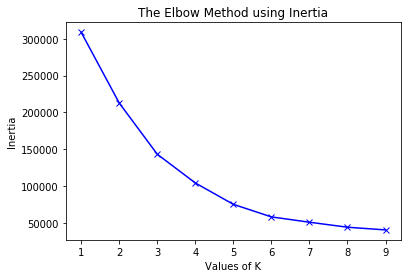
5: 75399.61541401483

6: 58350.65449462819

7: 51132.703212576904

8: 44359.634641148325

9: 40693.47096922685

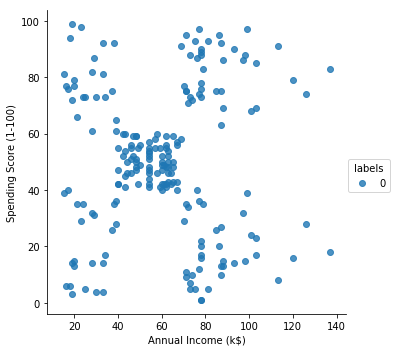


To determine the optimal number of clusters, we have to select the value of k at the “elbow” i.e. the point after which the distortion/inertia start decreasing in a linear fashion. Thus, for the given data, we conclude that the optimal number of clusters for the data is **5.**

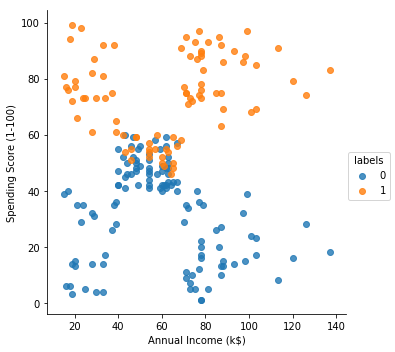
**Cluster Analysis:**

**The clustered data points for different value of k: -**

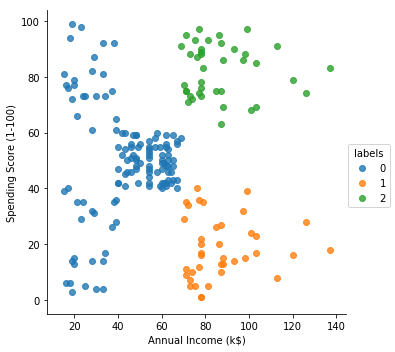
1. k = 1



2. k = 2



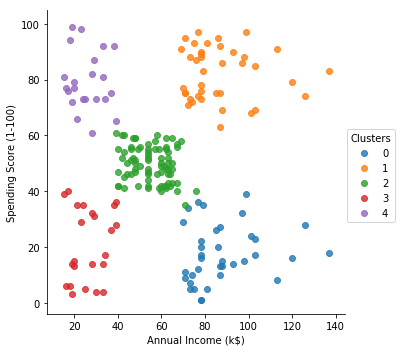
3. k = 3

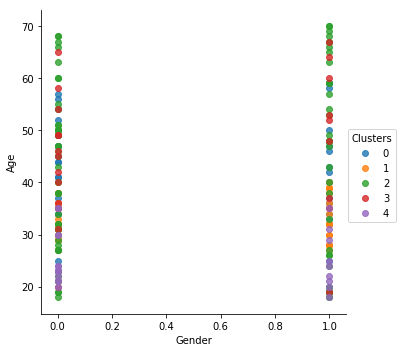


4. k = 4



5. k = 5





Here each cluster in **Annual Income Vs Spending score** scatterplot depicts some kind of information/pattern

**1) Cluster 0:** This cluster represents the customer having Annual income between 80 – 140 (certainly very high income) but their spending score is very less (0 – 40).

**2) Cluster 1:** This cluster represents the customer having Annual income between 80 – 140 and also their spending score is very high (60 – 100). This represents that this cluster consist of customers who are very frequent buyer from the Mall.

**3) Cluster 2:** This cluster represents the customer having average Annual income (40 – 60) and average Spending Scores (40 – 60).

**4) Cluster 3:** This cluster represents the customer having low Annual income (0 – 40) as well as low Spending scores (0 – 40).

**5) Cluster 4:** This cluster represets the customer having low Annual income(0 -40) but certainly very high spending scores(60 – 100).

**CHAPTER 5**

**IMPLEMENTATION**

**5. IMPLEMENTATION**

Code:

# Importing the libraries

from pandas import read\_csv as rc

from matplotlib import pyplot as plt

import numpy as np

from sklearn.cluster import KMeans

# Importing the dataset

data\_set = rc(r'C:\Users\shrey\OneDrive\Desktop\Exposys Data Science Project\customer-segmentation-dataset\Mall\_Customers - Usable.csv')

data\_set.head(15)

# Converting the gender values to 1(Male) and 0(Female)

gender = {'Male' : 1, 'Female': 0}

data\_set.Gender = [gender[item] for item in data\_set.Gender]

data\_set.head(15)

# Since customer ID is unusable for us we'll remove it

data\_set = data\_set.iloc[:, 1:]

data\_set.head(15)

# Describing some common features of our dataset

data\_set.describe()

# Visualization of Gender and Age

# Filtering out Gender and Age columns into gender\_age variable

gender\_age = data\_set.iloc[:,0:2]

gender\_age.head(10)

# Describing some common features of Gender and Age

gender\_age.describe()

# Histogram distribution of age

gender\_age.iloc[:,1].plot.hist(alpha= 0.5, stacked = True, bins = 25)

# Histogram distribution of Gender

gender\_age.iloc[:,0].plot.hist(alpha= 0.5, stacked = True, bins = 100)

# Describing some common features of Gender 'Male'

male\_gender = gender\_age[gender\_age['Gender'] == 1]

male\_gender.describe()

# Describing some common features of Gender 'Female'

female\_gender = gender\_age[gender\_age['Gender'] == 0]

female\_gender.describe()

# Histogram distribution of Age with Male as a gender

male\_gender.hist(column = "Age", bins = 25)

# Histogram distribution of Age with Female as a gender

female\_gender.hist(column = "Age", bins = 25)

# Hexbin of Gender and Age

data\_set.plot(kind = 'hexbin',x='Gender', y = 'Age',gridsize = 20,sharex=False)

# Scatterplot of Gender and Age

plt.scatter(data\_set.iloc[:, 0], data\_set.iloc[:, 1], c = 'white', marker = 'o', edgecolor='black',s=50)

plt.xlabel('Gender')

plt.ylabel('Age')

plt.title("Gender Age Scatterplot")

plt.show()

# Scatterplot of Annual Income(x - axis) and Spending Score(y-axis)

plt.scatter(data\_set.iloc[:, 2], data\_set.iloc[:, 3], c = 'white', marker = 'o', edgecolor='black',s=50)

plt.xlabel('Annual Income')

plt.ylabel('Spending Score')

plt.title('Annual Income Spending Score Scatterplot')

plt.show()

# Choosing the optimal number of clusters

# Checking the values of Inertia and Distortion using Elbow method

from scipy.spatial.distance import cdist

distortions = []

inertias = []

mapping1 = {}

mapping2 = {}

K = range(1,10)

for k in K:

#Building and fitting the model

kmeanModel = KMeans(n\_clusters=k).fit(data\_set)

kmeanModel.fit(data\_set)

distortions.append(sum(np.min(cdist(data\_set, kmeanModel.cluster\_centers\_,

'euclidean'),axis=1)) / data\_set.shape[0])

inertias.append(kmeanModel.inertia\_)

mapping1[k] = sum(np.min(cdist(data\_set, kmeanModel.cluster\_centers\_,

'euclidean'),axis=1)) / data\_set.shape[0]

mapping2[k] = kmeanModel.inertia\_

for key,val in mapping1.items():

print(str(key)+' : '+str(val))

plt.plot(K, distortions, 'bx-')

plt.xlabel('Values of K')

plt.ylabel('Distortion')

plt.title('The Elbow Method using Distortion')

plt.show()

for key,val in mapping2.items():

print(str(key)+' : '+str(val))

plt.plot(K, inertias, 'bx-')

plt.xlabel('Values of K')

plt.ylabel('Inertia')

plt.title('The Elbow Method using Inertia')

plt.show()

# Applying KMeans clustering algorithm to the dataset

kmeans = KMeans(n\_clusters=5, init='random',n\_init=10,max\_iter=300,tol=1e-04, random\_state=0)

clustered\_data = kmeans.fit\_predict(data\_set)

print(clustered\_data)

# Concatenating this cluster label to the original data\_set

from pandas import DataFrame as df

from pandas import concat as cc

clusters = df(kmeans.labels\_)

data\_set = cc((data\_set, clusters), axis = 1)

data\_set = data\_set.rename({0:'Clusters'}, axis = 1)

data\_set.head()

# Importing the seaborn library to visualize the graphs

import seaborn as sns

# Visualizing the Gender vs Age clusters

sns.lmplot(x = "Gender", y = "Age", data = data\_set, hue='Clusters',fit\_reg=False)

# Visualizing the Annual Income vs Spending Scores clusters

sns.lmplot(x='Annual Income (k$)', y ='Spending Score (1-100)',data=data\_set,hue='Clusters',fit\_reg=False)

# Plotting pairplot to get pairwise relationships in a dataset

sns.pairplot(data\_set,hue='Clusters')

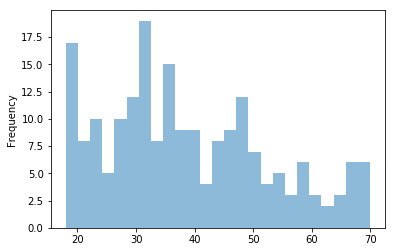
**CHAPTER 6**

**CONCLUSION**

**6. CONCLUSION**

1) Gender (x-axis) and Age (y-axis) Visualization

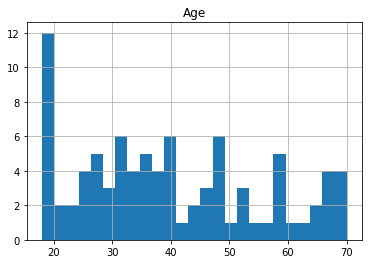
1.1 Histogram distribution of Age



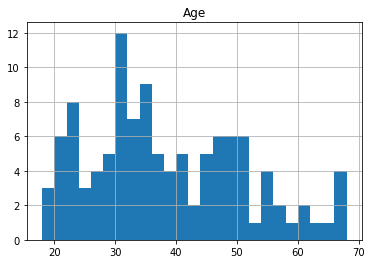
1.2 Histogram distribution of Gender



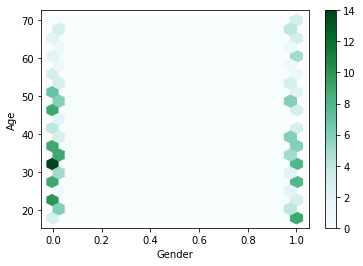
1.3 Histogram distribution of Age with Male as a Gender



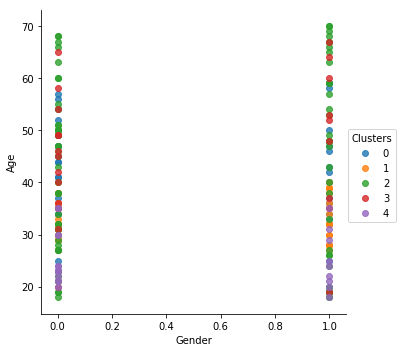
1.4 Histogram distribution of Age with Female as a Gender



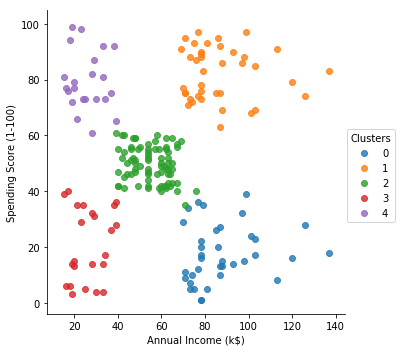
1.5 Hexbin of Gender and Age



1.6 Scatterplot of Gender and Age



2) Analyzing Annual Income (x – axis) and Spending Score (y – axis)



**Business Strategies that can be applied for this clusters:**

* **Cluster 0:** This cluster is very important to target to digital marketing since this cluster have very high annual income but they are not spending much. Maybe we’re missing out on lot of customers who don’t seem to take much action in the mall even though having good annual income.
* **Cluster 1:** We can target these customers whenever new deals arises because indeed these are with customers that you’ll get the highest chance to sell your products. This cluster have high potential to buy your products.
* **Cluster 2:** There is not much of digital marketing should be done here. Since they aren’t doing much of spending in mall. More advertising could result into making that cluster similar to cluster 3.
* **Cluster 3:** There is nothing to do to this cluster. Since their annual income is low and their spending score is also less.
* **Cluster 4:** These customers shouldn’t be targeted as much as cluster 1 as they have pretty low annual income but their spending is quite high. Because too much new deals and new irresistible offer that these customers don’t need.

**References**

[1] GHZANFARI, MEHDI, et al. "Customer segmentation in clothing exports based on clustering algorithm." (2010): 59-86.

[2] J. Lee and S. Park, “Intelligent profitable customers segmentation system based on business intelligence tools”, Expert Systems with Applications, vol. 29, no. 1, pp. 145-152, 2005.

[3] D. A. Kandeil, A. A. Saad and S. M. Youssef, “A two-phase clustering analysis for B2B customer segmentation”, in International Conference on Intelligent Networking and Collaborative Systems, Salerno, 2014, pp. 221-228.

[4] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman and A. Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 881-892, July 2002, doi: 10.1109/TPAMI.2002.1017616.

[5] D. MacKay, “An example inference task: Clustering”, in Information theory, inference, and learning algorithms. Cambridge, UK: Cambridge University Press, 2003, pp. 284-292.

[6] O. Maimon and L. Rokach, “Clustering methods”, in Data Mining and Knowledge Discovery Handbook. Boston: Springer US, 2005, pp. 321-352.

[7] G. Lefait and T. Kechadi, “Customer segmentation architecture based on clustering techniques”, in Fourth International Conference on Digital Society, Sint Maarten, 2010, pp. 243-248.

[8] M. Namvar, M. Gholamian and S. KhakAbi, “A two-phase clustering method for intelligent customer segmentation”, in International Conference on Intelligent Systems, Modelling and Simulation, Liverpool, 2010, pp. 215-219.

[9] Gaur, Deepti, and Seema Gaur. "Comprehensive analysis of data clustering algorithms." *Future Information Communication Technology and Applications*. Springer, Dordrecht, 2013. 753-762.

[10] J. MacQueen, “Some methods for classification and analysis of multivariate observations”, in Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, 1967, pp. 281-297.

[11] T. Kodinariya and P. Makwana, “Review on determining number of cluster in K-means clustering”, International Journal of Advance Research in Computer Science and Management Studies, vol. 1, no. 6, pp. 90-95, 2013.

[12] Tripathi, Shreya, Aditya Bhardwaj, and E. Poovammal. "Approaches to clustering in customer segmentation." *International Journal of Engineering & Technology* 7.3.12 (2018): 802-807.