ABSTRACT:

The Main Objective The Objective Of Market Segmentation Is To Minimize Risk By Determining Which Products Have The Best Chances Of Gaining A Share Of A Target Market And Determining The Best Way To Deliver The Products To The Market .With Full Fledged Logic And Environment, Using New Trending Technologies In Market With Maximum Functionalities And Efficiency.

KEYWORDS:

* Unsupervised Learning
* K-Means Clustering
* Customer Segmentation

INTRODUCTION:

The Main Objective The Objective Of Market Segmentation Is To Minimize Risk By Determining Which Products Have The Best Chances Of Gaining A Share Of A Target Market And Determining The Best Way To Deliver The Products To The Market. With Full Fledged Logic And Environment, Using New Trending Technologies In Market With Maximum Functionalities And Efficiency.

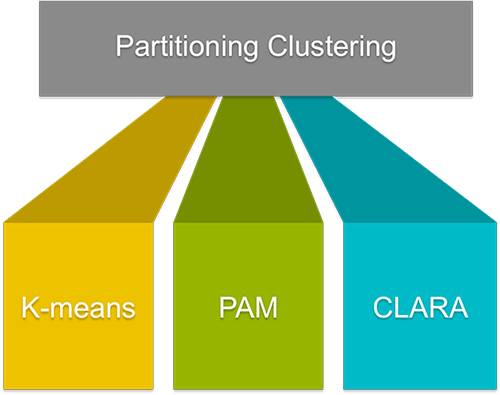
PACKAGES:

* Matplotlib
* Seaborn

LIBRARIES:

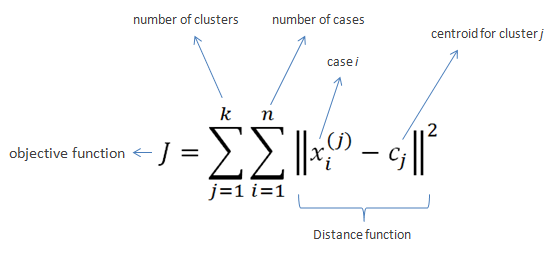
* Numpy
* Pandas

LITERATURE REVIEW:



K-MEANS CLUSTERING ALGORITHM

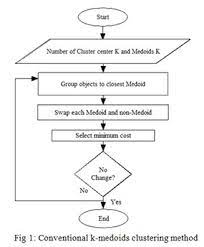
* Step-1: Select the number K to decide the number of clusters.
* Step-2: Select random K points or centroids. ...
* Step-3: Assign each data point to their closest centroid, which will form the predefined K clusters.
* Step-4: Calculate the variance and place a new centroid of each cluster.



PAM CLUSTERING ALGORITHM

The PAM algorithm searches for k representative objects .a data set (k medoids) and then assigns each object to the closest medoid in order to create clusters. Its aim is to minimize the sum of dissimilarities between the objects in a cluster and the center of the same cluster (medoid)

The Flow Chart Of The PAM Algorithm is shown in fig1



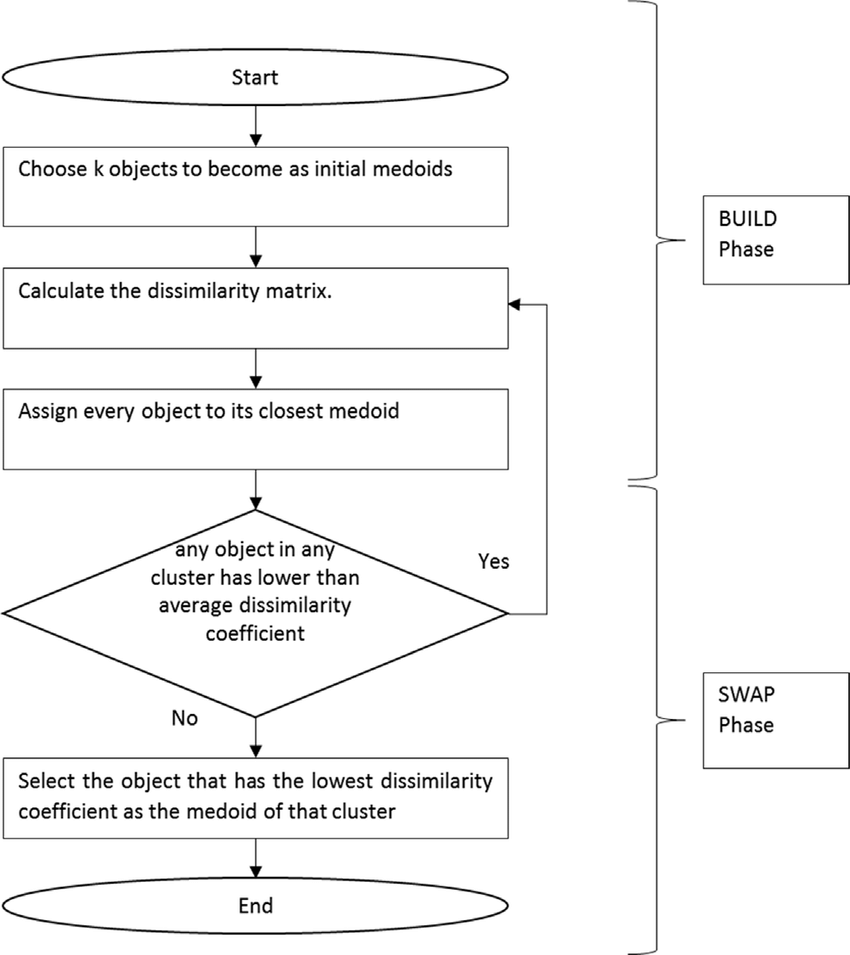
S(i)=b(i)-a(i)/larger of b(i) and a(i)

a(i)=average distance inside cluster

b(i)=average distance nearest other cluster

CLARAS CLUSTERING ALGORITHM:

1. Select ‘k’ random data points and label them as medoids for the time being.
2. Select a random point say ‘a’ from the points picked in step (1), and another point say ‘b’ which is not included in those points.
3. We would already have the sum of distances of point ‘a’ from all other points since that computation is required for selecting the points in step (1). Perform similar computation for point ‘b’.
4. If the sum of distances from all other points for point ‘b’ turns out to be less than that for point ‘a’, replace ‘a’ by ‘b’.
5. The algorithm performs such a randomized search of medoids ‘x’ times where ‘x’ denotes the number of local minima computed, i.e. number of iterations to be performed, which we specify as a parameter. The set of medoids obtained after such ‘x’ number of steps is termed as ‘local optimum’.
6. A counter is incremented every time a replacement of points is made. The process of examining the points for possible replacement is repeated till the counter does not exceed the maximum number of neighbours to be examined (specified as a parameter).
7. The set of medoids obtained when the algorithm stops is the best local optimum choice of medoids.



EXISTING WORK:

Customer segmentation is the process by which you **divide your customers up based on common characteristics** – such as demographics or behaviours, so you can market to those customers more effectively. These customer segmentation groups can also be used to begin discussions of building a marketing persona.

PROPOSED WORK:

* Here, We Test The Different Data Sets Of Various Organizations To Know The Accuracy Of The Demand And Sales Of The Particular Product.
* Based On The Result Obtained, Marketing Strategy Can Be Applied To Maximize The Sales And Profit.
* With This Approach Risk Of Business Is Reduced.
* This Approach Depends Largely On The Type Of Algorithm And The Quality And Quantity Of The Data Used.
* This Approach Can Easily Deal With Tons Of Data.

DATA COLLECTION:

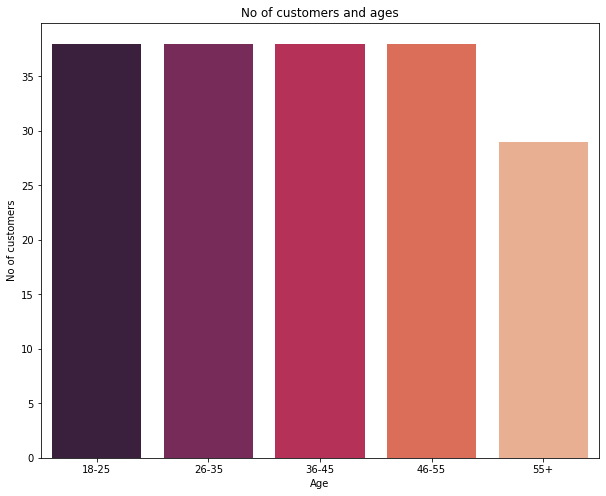
* I Have Collected The Data Set From “Kaggle” Website Which Consists The Data Of 200 Citizens, With .Csv Extension
* I Have Manually Created A Dataset Using Microsoft Excel And Saved It With .csv Extension
* EXAMPLE:

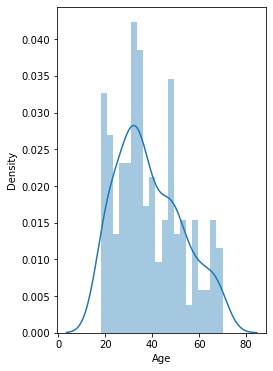
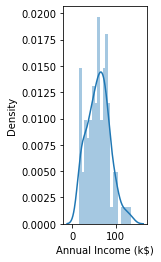
dataset.csv

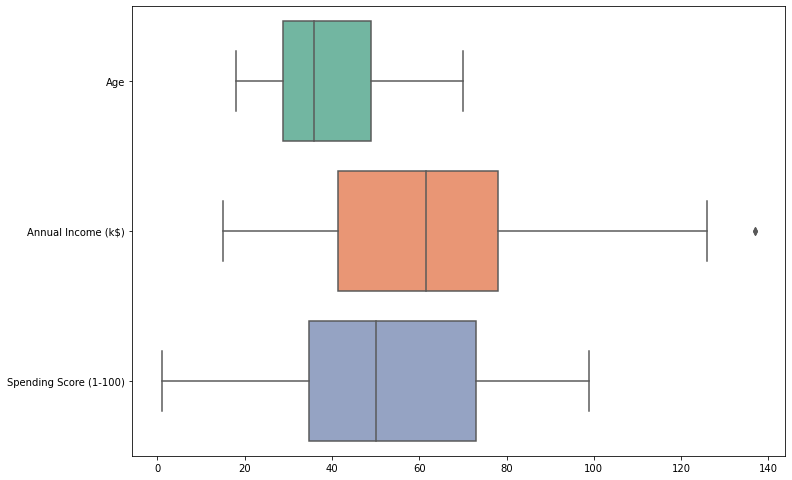
dairymilk.csv Etc.…

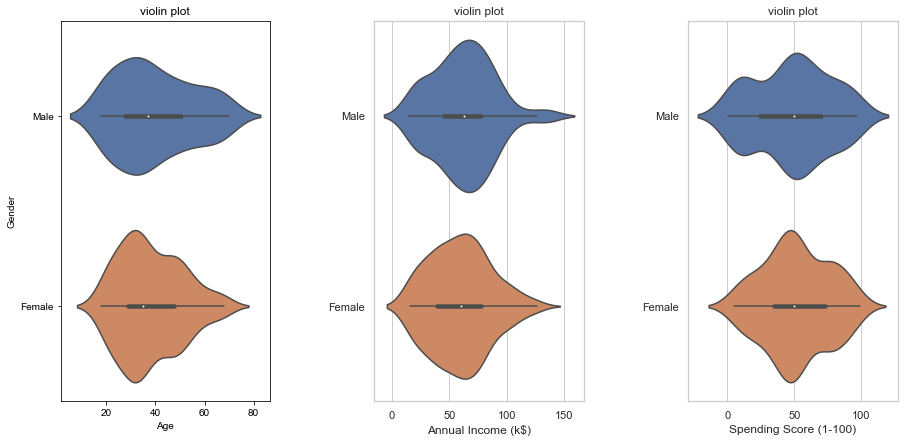
OUTPUT:

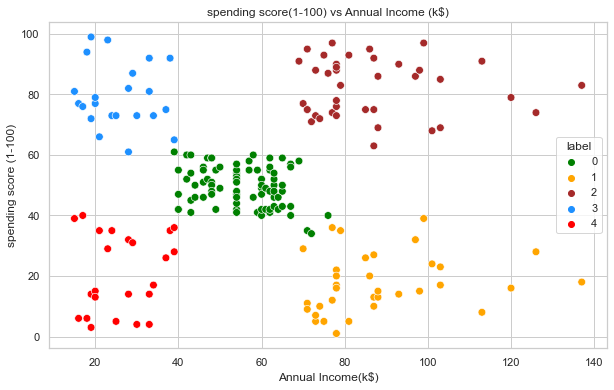




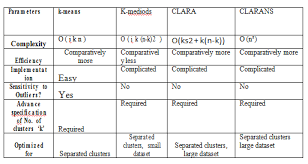
 







DISCUSSION:



CONCLUSION:

K-Means Has Been Around Since The 1970s And **Fares Better Than** Other Clustering Algorithms Like Density-Based, Expectation-Maximisation. It Is One Of The Most Robust Methods, Especially For Image Segmentation And Image Annotation Projects. According To Some Users, K-Means Is Very Simple And Easy To Implement.

REFERNCES:

[1]: EXPOSYS DATA LABS DOCUMENTATION

[2]: ONE STOP SOLUTIONS DOCUMENTATION

[3]: W3SCHOOLS DOCUMENTATION

BY:

GADDAMIDHI SAI HARSHITH REDDY

MENTOR:

DR.GLADSON BRITTO SIR