



DEPARTMENT of COMPUTER SCIENCE & ENGINEERING
SOUTHEAST UNIVERSITY

CSE4000: Research Methodology

“CUSTOMER SEGMENTATION AND MARKET BASED ANALYSIS USING MACHINE LEARNING”

A dissertation submitted to the Southeast University in partial fulfillment of the requirements for the degree of B. Sc. in Computer Science & Engineering

Submitted by

Md. Sabbir Ahmed
ID: 2017100000016

Md. Arafat Hossain
ID: 2017100000017

Md. Muntasirur Rahman
ID: 2017100000140

Supervised by

Maksuda Rabeya

Lecturer

**Department of CSE
Southeast University**

Copyright©2021 Southeast University
October, 2021

Letter of Transmittal

October 25, 2021

The Chairman,
Department of Computer science and Engineering
Southeast University,
Banani, Dhaka.

Through: Supervisor, Maksuda Rabeya

Subject: Submission of Research Paper Report.

With due respect and humble submission, we the students of CSE are submitting our research paper report on “Customer Segmentation and Market Based Analysis Using Machine Learning” It gives us immense pleasure to inform you that we have completed our research paper under kind-hearted supervision.

For the better understanding and reliability ,it is being divided into different chapters and it will provide a extensive source information.

Now, we have to placed our research paper report before you for kind approval. We hope that our research paper will satisfy you.

Thank you.
Sincerely yours.

Submitted by

Md. Sabbir Ahmed
ID: 2017100000016
Md. Arafat Hossain
ID: 2017100000017
Md. Muntasirur Rahman
ID: 2017100000140

Supervised by

Maksuda Rabeya
Lecturer
Department of CSE,SEU

Certificate

This is to certify that the report “**Customer Segmentation and Market Based Analysis Using Machine Learning**” is the record of research work done by Md. Sabbir Ahmed (ID-2017100000016), Md. Arafat Hossain (ID- 2017100000017) and Md. Muntasirur Rahman (2017100000140) for the partial fulfillment of BSc requirement in Computer Science and Engineering (CSE) from Southeast University.

This research work has been satisfactorily completed under my supervision.

Author's:

Md. Sabbir Ahmed
ID: 2017100000016
Batch-46 ,CSE

Md. Arafat Hossain
ID: 2017100000017
Batch-46 ,CSE

Md. Muntasirur Rahman
ID: 2017100000140
Batch-46 ,CSE

Approved by the Supervisor:

Maksuda Rabeya
Lecturer
Department of CSE,SEU

Abstract

Due to the emergence of numerous business competitors, there is intense competition among competing businesses in terms of acquiring new customers and retaining existing ones. As a result of the foregoing, regardless of the size of the business, the need for exceptional customer service becomes critical.

Furthermore, any company's ability to understand the needs of each of its customers will give it a competitive advantage in providing targeted customer services and developing customized marketing programs for customers. This understanding is attained through systematic customer segmentation.

Customer in each segment have market characteristics that are similar. We have gone through many researcher's paper who applied different model for this recommendation system which are Density Based Clustering (DBSCAN), Affinity Propagation Algorithm (AP), hold-out validation. In this paper we performed K-means Algorithm which is best for the segment the customers. This is a k-means clustering algorithm that work for identifying the real customers. We collected a dataset which name is: "Test" from Kaggle and applied k-means algorithm in this paper.

Keywords- *Machine Learning, K-means, Affinity Propagation Algorithm, Elbow method*

Acknowledgements

At first, we would like to Thank Almighty Allah for giving us opportunity in writing in this paper. We would also like to pay our thanks to our respected teacher **Maksuda Rabeya** ma'am who has chosen us for this research topic and also provides help with our knowledge, inspiration and information.

Her useful advice and co-operative behavior for this whole work is sincerely acknowledged. It would not have been possible for us to complete the research without her sincere and loving help.

We want to express our debt to our parents as well as to our family members whose blessing and support have always helped us face the challenges ahead.

Table of Contents

Abstract.....	4
Acknowledgments.....	5
Content List	6
Chapter 1- Introduction.....	7
1.1 Objective.....	8
Chapter 2- Literature Review.....	9
Chapter 3- Research Methodology.....	10
Chapter 4- Definition of Selected Dataset.....	11
Chapter 5- Implementation.....	12
5.1 Data set.....	12-13
5.2 Data Convert.....	14
5.3 Graph of Dataset.....	15-23
Chapter 6- Result and Evaluation.....	24
Chapter 7- Future Plan.....	25
Conclusion.....	25
References.....	26

Chapter 1

Introduction

Over the years, increasing competition in the business and the availability of large-scale historical information have led to the widespread use of data mining techniques to find critical and strategic information hidden in the organization's information. Customer segmentation is the process of different group of customer into group based on common characteristics so that company can market to each group appropriately. The importance of customer segmentation includes, respectively, the ability to customize a business's market plan that will be appropriate for each segment of its customers. Support the business decisions based on risk environment such as credit of relationship with us customer.

Machine Learning is based on learning. The more users use it, the more machines will learn. That's why sometimes it won't give the right result but according to the users it will be right.

There are lot of techniques to make a model like Density Based Clustering (DBSCAN), Affinity Propagation Algorithm (AP), hold-out validation for customer segmentation. Getting accurate results we performed K-means algorithm which is comparatively best for customer segmentation among the existing model. We looking forward for our dataset and gaining to know real buyers so we chose K-means Algorithm for our research paper.

1.1 Objective

This research model helps market management who want to increase their sell of their products. We are intending to launch this model and help the management who find difficulties to know the customer's movement in the market.

1. To help the market management and analysis the sales.
2. To predict based on customer by previous customer's data-set.

The scope of our research is to develop a model where the market management/shopkeepers can easily predict based on their customer's and to predict the possibility to reach their needs which is actually real or nightmare. This research basically on predict customers who really want to buy from shop and who are not. So we hope this system or model will help to understand their running situation and predict their future customer's based on attributes.

Chapter 2

Literature Review

Craft, Stephen(2004) show in his article that was generally, customer's are ready to pay for their specific's product which is needs everyday. If market management analysis who who successfully segment the market and adapt their products to needs smaller segment stand to gain in term of increased sales and profit margins and reduce competitor pressure.New businesses ,in particular , may find market segmentation to be a key in enabling them to complete with larger firms. Many management consulting firms offer with market segmentation to small businesses. But the potential gains offered by7 market segmentation must be measured as a costs, which in addition to the market research as segment in a market with increased production and market expenses

Customers and markets are not similar for every time(Smith,1956; Claycamp & Massy,1968).Smith(1956) suggested that segmentation, ther division of market into a team of customers who share certain characteristics or keyword toward a product or services, it will be an effective way for an organization to manage diversity within a market might be an effectively broken down into actionable customer segments.

(Smith 1956) Every market have two type of product policy, one is product differentiation and another is market segmentation .(Frank,Massy & Wind, 1972, Wedel & Kamakura 2000) market segmentation refers to making products decisions after studying and key-wording the diversity of needs in market place , while product differentiation refers to products consumers from the recent population and sometime even customers with similar within a group/team and different across-group response.

Chapter 3

Research Methodology

Customer segmentation is the practice of dividing a company's customer into groups that reflect similarity among customer to each group. The goal of segmentation customer is to determine how to relate to customer in each segment in order to maximum the value of each customer in the business. The process requires a thought-out strategy understanding how to manage and group or customer and which data you will used to do this.

- 1.Find the number of cluster using Elbow method.
- 2.Compare with attributes we have nine attributes
- 3.Each two compare between themselves and find graph.
- 4.Finally show the diagram which is find from each diagram.
- 5.And get the result of customers

Chapter 4

Definition of the Selected Dataset

We are using the “Test” dataset for performing this project. This dataset contains only one file. The file is Test.csv

The Test.csv file contains 2628 rows and 8 columns. Describing the dataset below-

Table : Customer Dataset Description

Attribute	Description	Datatypes
ID	A unique id for each customer	int64
Gender	Customer’s Gender	object
Marital Status	Customer’s marital status	object
Age	Customer Age	int32
Graduated	Educational Qualification of Customer	object
Profession	Customer profession	object
Spending Score	Previous Buying History	int32
Family Size	Family size Of customer	int32

Chapter 5

Implementation

Now we are going to implement our research utilizing the Python programming language. We will be utilizing anaconda-3 and the execution will be finished utilizing Jupyter Notebook.

Jupyter can be an organization that allows anyone to use all Python library in their code without any restrictions.

We have to implement it to prove the reality of our research scenario.

5.1 Dataset

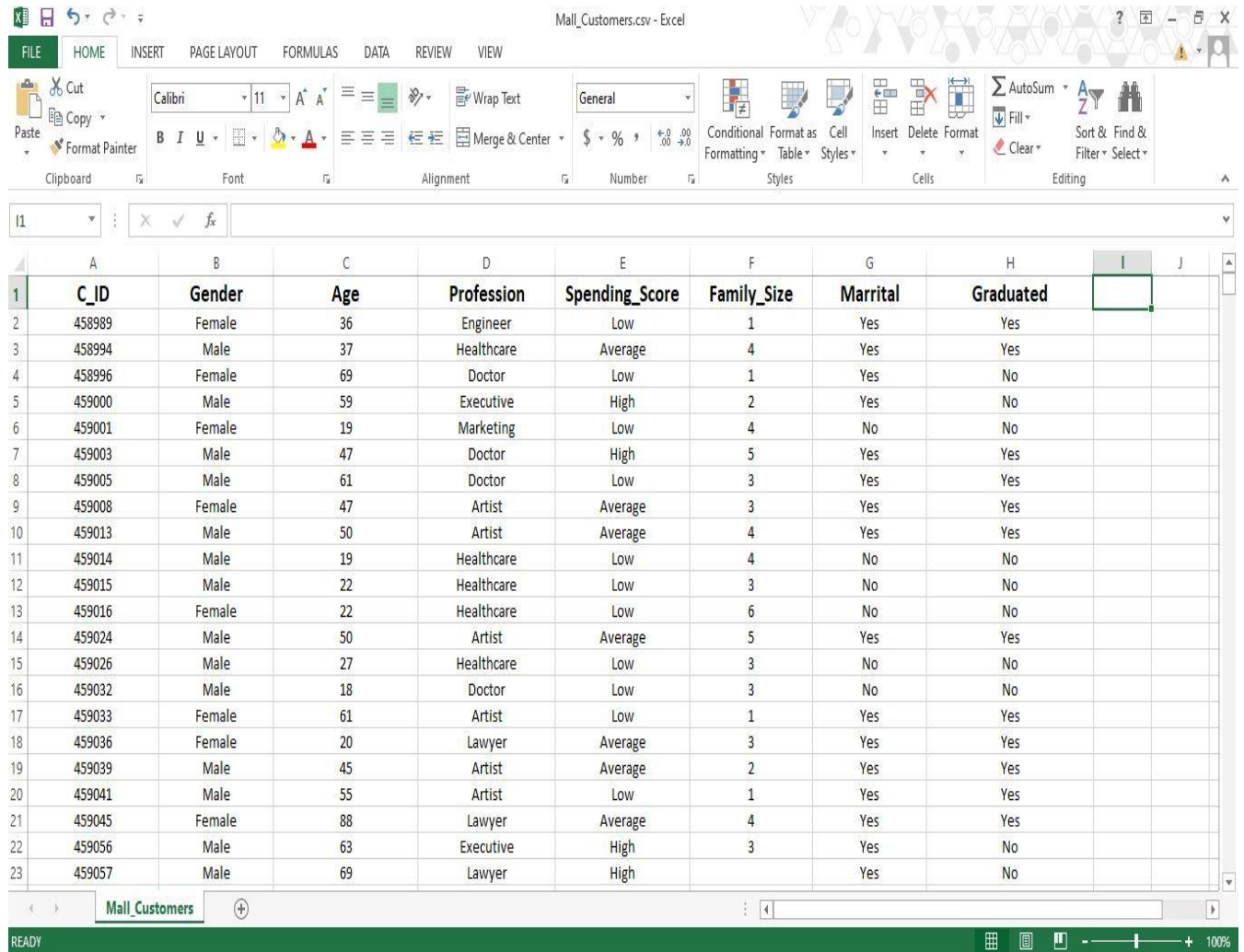
We use Test.csv file. In Dataset we are collect form

<https://www.kaggle.com/vetrirah/customer?select=Test.csv>

our dataset into 2628 data row and 9 attribute column

Here we will implement our K-means cluster algorithm on our dataset.

Our Dataset



	A	B	C	D	E	F	G	H	I	J
1	C_ID	Gender	Age	Profession	Spending_Score	Family_Size	Marrital	Graduated		
2	458989	Female	36	Engineer	Low	1	Yes	Yes		
3	458994	Male	37	Healthcare	Average	4	Yes	Yes		
4	458996	Female	69	Doctor	Low	1	Yes	No		
5	459000	Male	59	Executive	High	2	Yes	No		
6	459001	Female	19	Marketing	Low	4	No	No		
7	459003	Male	47	Doctor	High	5	Yes	Yes		
8	459005	Male	61	Doctor	Low	3	Yes	Yes		
9	459008	Female	47	Artist	Average	3	Yes	Yes		
10	459013	Male	50	Artist	Average	4	Yes	Yes		
11	459014	Male	19	Healthcare	Low	4	No	No		
12	459015	Male	22	Healthcare	Low	3	No	No		
13	459016	Female	22	Healthcare	Low	6	No	No		
14	459024	Male	50	Artist	Average	5	Yes	Yes		
15	459026	Male	27	Healthcare	Low	3	No	No		
16	459032	Male	18	Doctor	Low	3	No	No		
17	459033	Female	61	Artist	Low	1	Yes	Yes		
18	459036	Female	20	Lawyer	Average	3	Yes	Yes		
19	459039	Male	45	Artist	Average	2	Yes	Yes		
20	459041	Male	55	Artist	Low	1	Yes	Yes		
21	459045	Female	88	Lawyer	Average	4	Yes	Yes		
22	459056	Male	63	Executive	High	3	Yes	No		
23	459057	Male	69	Lawyer	High		Yes	No		

Figure: Dataset

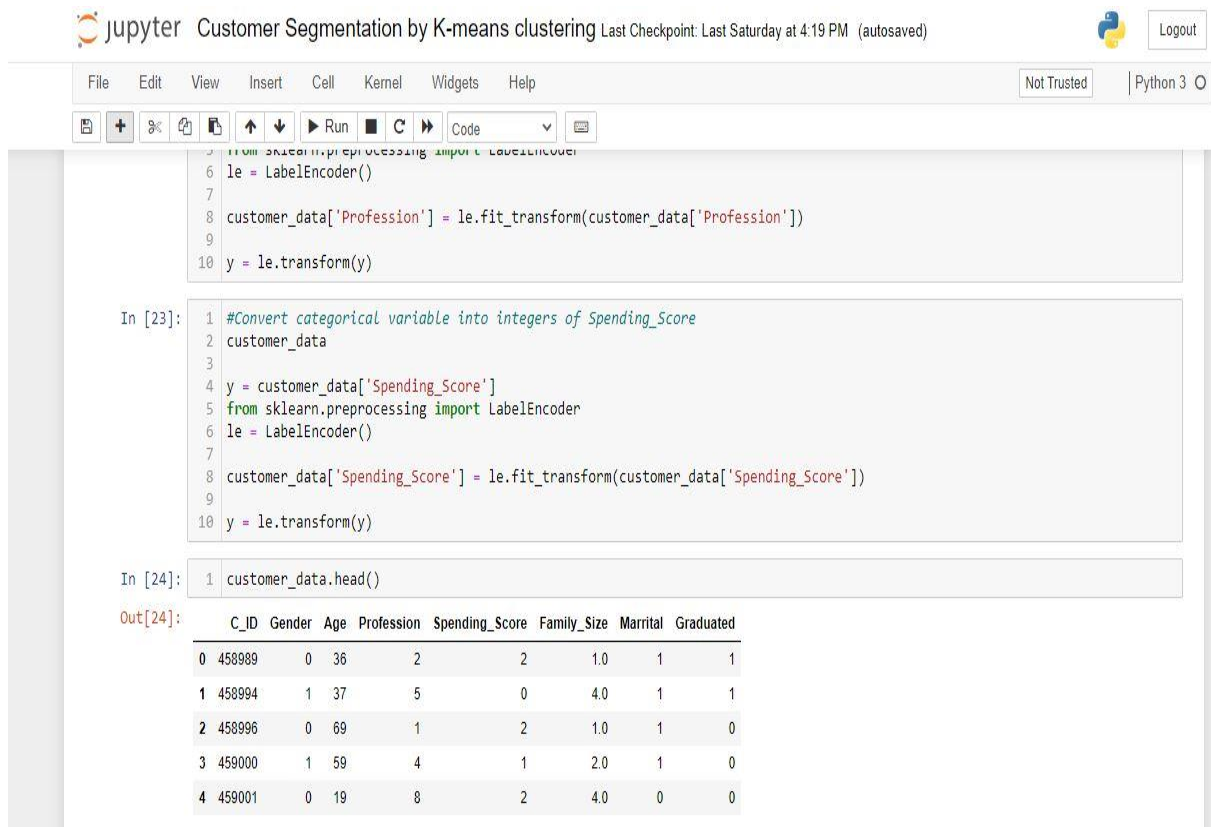
5.2 Dataset-convert

Our dataset attribute value is classified and value is string. but we can't implement string value. Then we convert the string value to float. Gender column attribute Male=1 and Female=0.

Profession column Engineer=2, Healthcare=5,

Doctor=1 Etc. Spending_score column Low=2, Average=0

High=1. Marrital status Yes=1 and No =0 similar of Graduated.



The image shows a Jupyter Notebook interface with the title "Customer Segmentation by K-means clustering". The notebook contains three code cells. The first cell imports LabelEncoder from sklearn.preprocessing and uses it to convert the 'Profession' column of 'customer_data' into integers. The second cell, labeled "In [23]:", does the same for the 'Spending_Score' column. The third cell, labeled "In [24]:", displays the first five rows of the 'customer_data' dataset using the .head() method. The output shows columns: C_ID, Gender, Age, Profession, Spending_Score, Family_Size, Marrital, and Graduated.

```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
customer_data['Profession'] = le.fit_transform(customer_data['Profession'])
y = le.transform(y)
```

```
In [23]: 1 #Convert categorical variable into integers of Spending_Score
2 customer_data
3
4 y = customer_data['Spending_Score']
5 from sklearn.preprocessing import LabelEncoder
6 le = LabelEncoder()
7
8 customer_data['Spending_Score'] = le.fit_transform(customer_data['Spending_Score'])
9
10 y = le.transform(y)
```

```
In [24]: 1 customer_data.head()
```

```
Out[24]:
```

	C_ID	Gender	Age	Profession	Spending_Score	Family_Size	Marrital	Graduated
0	458989	0	36	2	2	1.0	1	1
1	458994	1	37	5	0	4.0	1	1
2	458996	0	69	1	2	1.0	1	0
3	459000	1	59	4	1	2.0	1	0
4	459001	0	19	8	2	4.0	0	0

Figure: Dataset-convert

5.3 Graph of Dataset

Know we can show our dataset graph x=axis by y-axis. Data set attribute
C_ID,Gender,Age,Profession,Spending_Score,Family_Size,Marrital Graduated

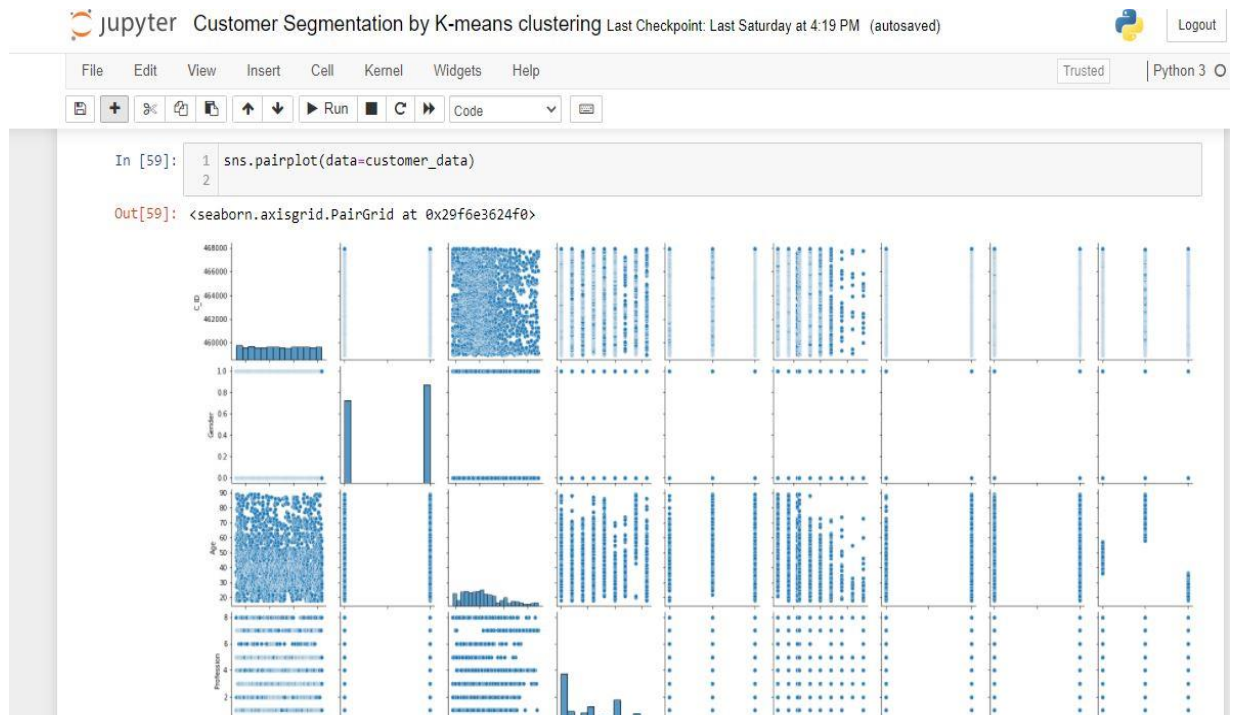


Figure: Graph

Gender graph

This is graph of dataset Gender attribute. our dataset 54% of Female person and 46% of Male person.

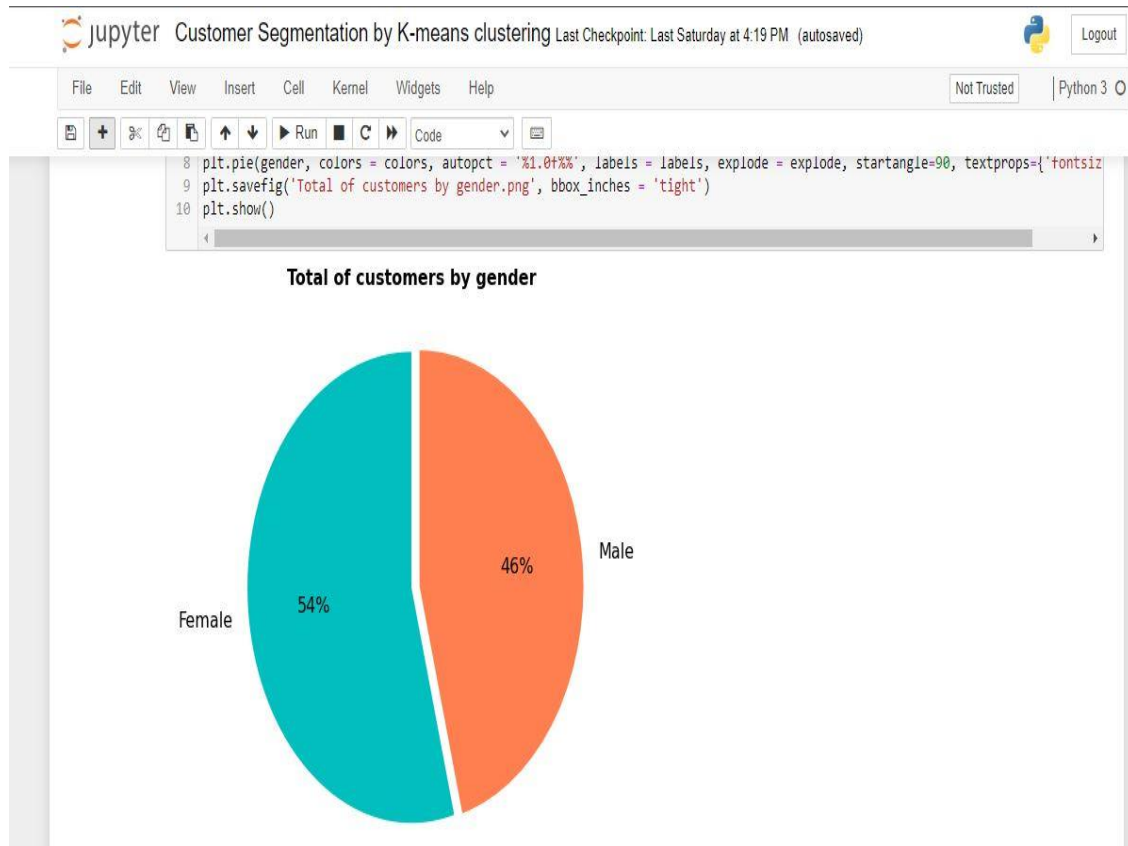


Figure: Gender

Age graph

Distribution of Age graph the 30 years old customer is high and 80 years old customer is very low.

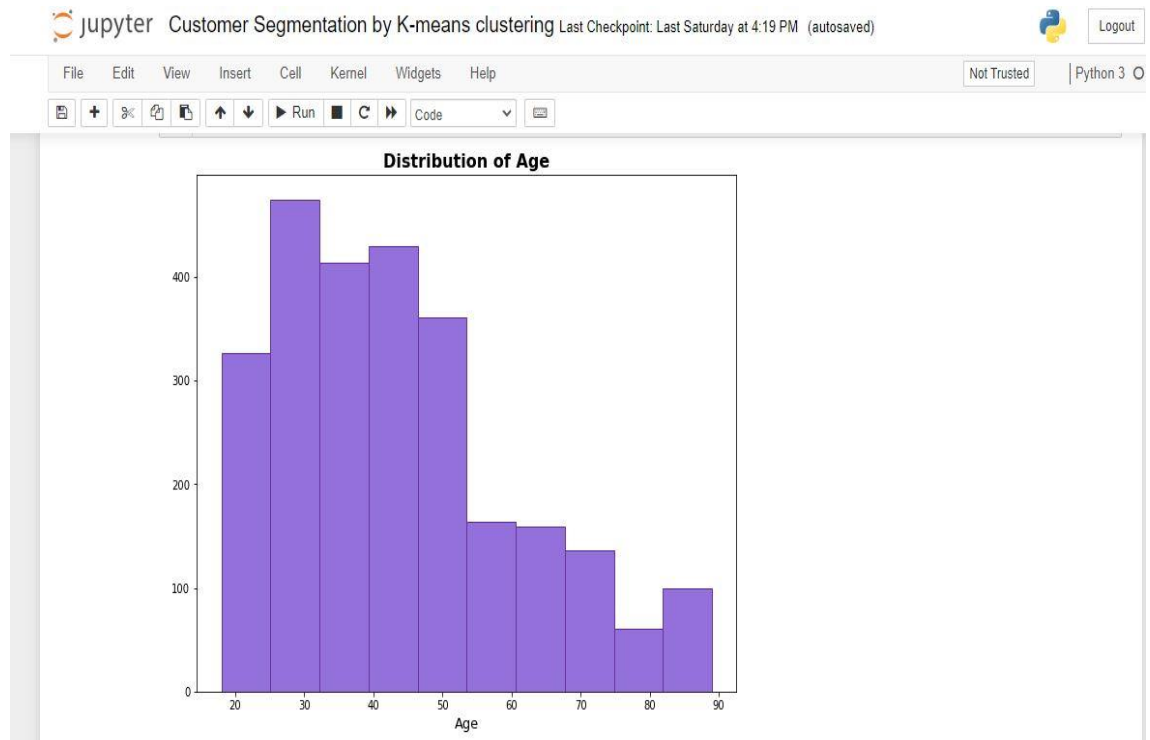


Figure: Age

Graph of Profession

We know, 0=healthcare so analysis of the graph healthcare profession customer is very high.

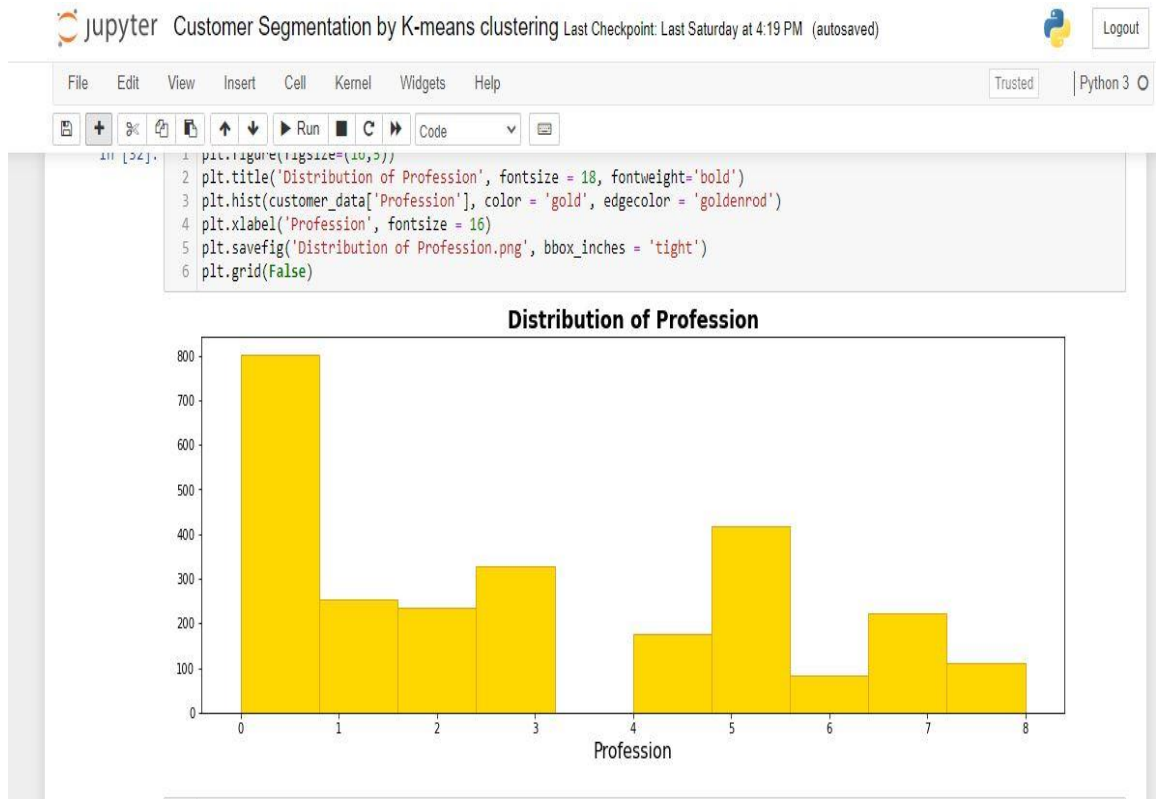


Figure: Profession

Average-Age

We are analysis this graph average age by Gender, so total customer of dataset is Female customer average age 43.46 and Male customer average age is 43.81



Figure: Avg-age

Average-Spending score

We are analysis this graph average spending score by gender, so total customer of dataset there are Female customer score is high and Male customer score is low.

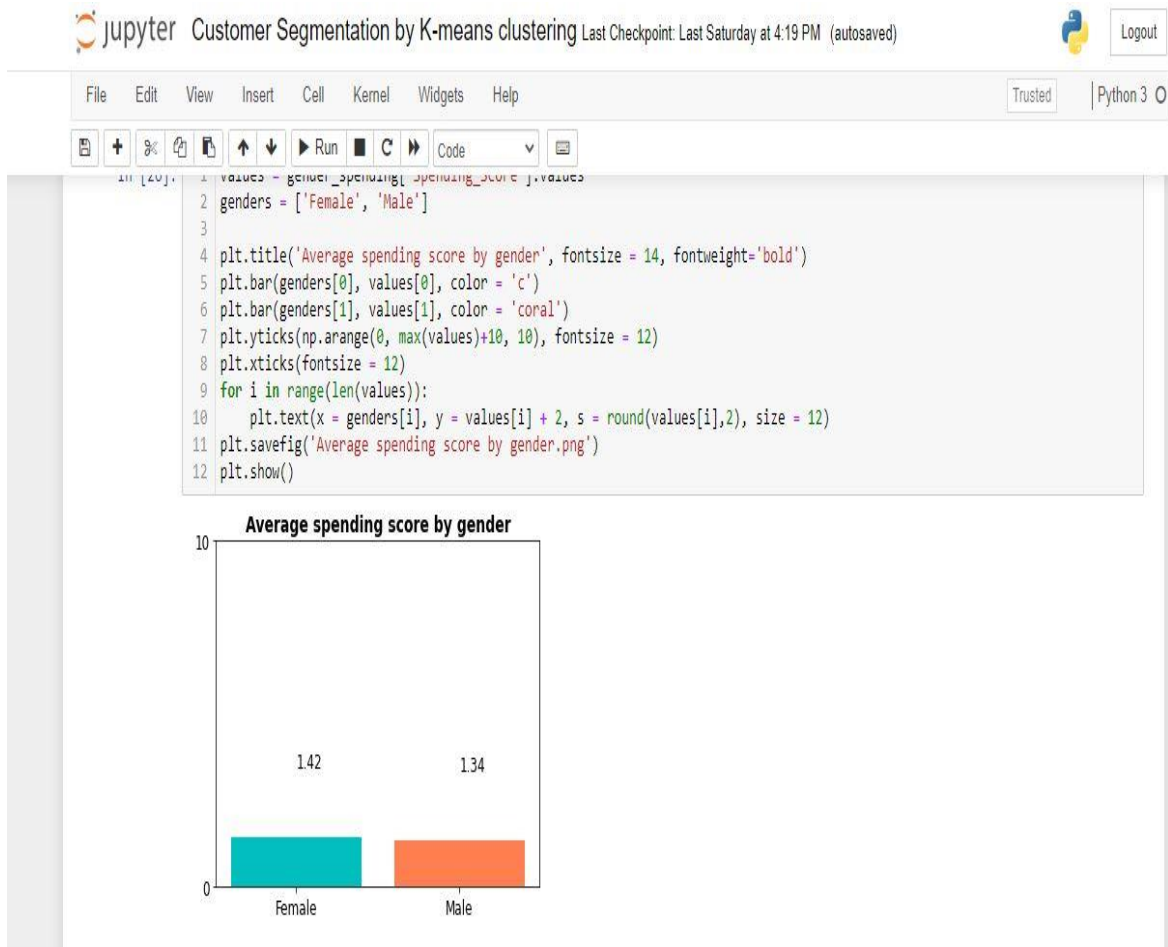


Figure: Avg-score

Average-Family size

We are analysis this graph average Family size by Gender, so total customer of dataset there are Female customer Family size is smaller than Male customer.



Figure: Avg-F-Size

Elbow-method

In cluster analysis, the elbow method is a heuristic that is used to determine the number of clusters in a data set.

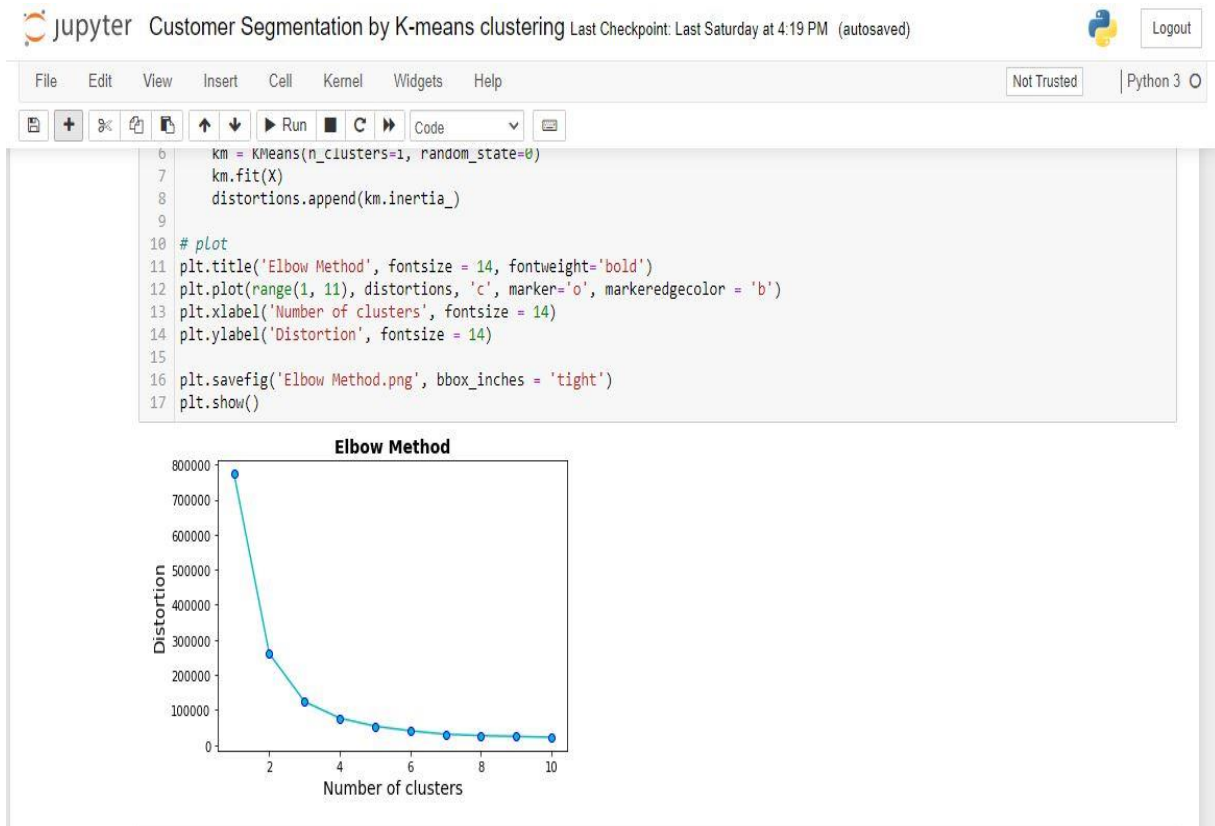


Figure: Elbow-method

Cluster of K-means

Now we are cluster of the dataset in customer age and customer profession ,we are used the k-means algorithm. We compare two attributes age and profession .Comparing these two attributes we see that there are more engineer in cluster 1 and they are 40 to 50 years old.

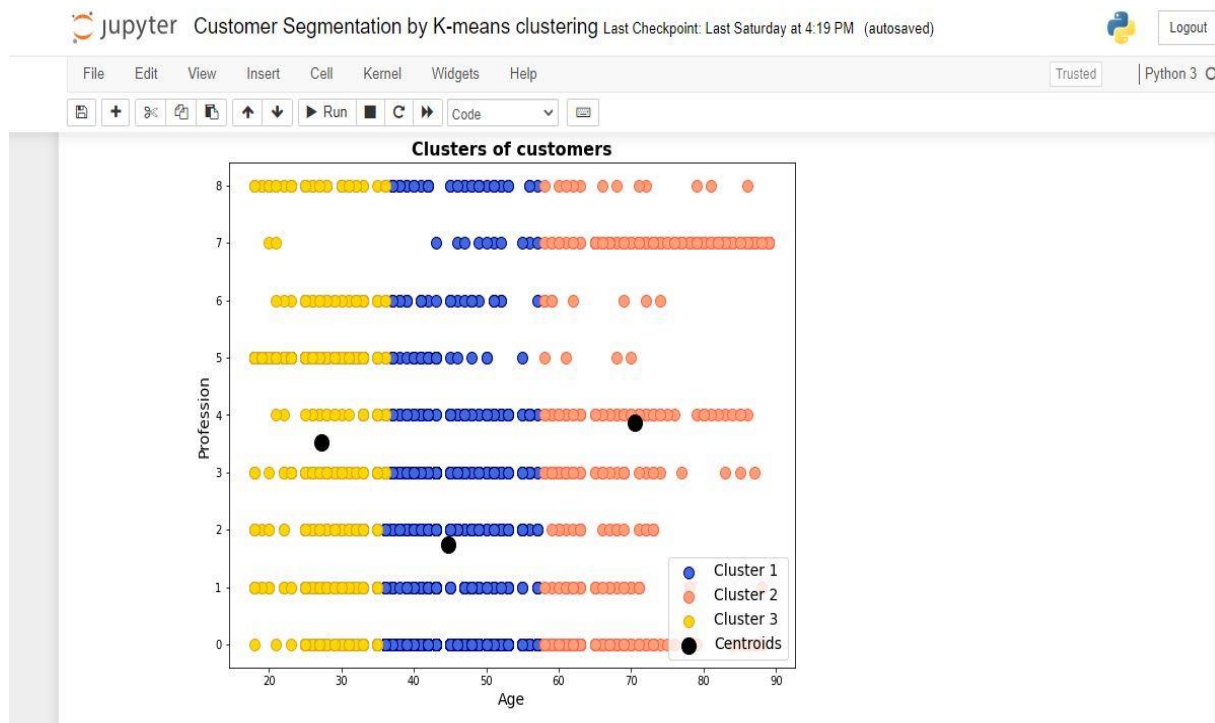


Figure: k-means

Chapter-6

Result and Evaluation

We are works with Age and profession, we got three cluster using Elbow method and k-means algorithm. We see that cluster 1(blue) has the most customers and the most engineering profession.

Their Age ratio is around 40 to 50 .and there are more Male than Female among them. If we think of cluster 2, we will see that the ratio of customer in the executive profession and their age ratio is 70. Cluster 3 has more customer between the Age of 20 to 30 and more Male to Female.

Therefore, we can say that Male customer come more in shopping mall. If we keep more Male product then our sales will be good and customer will be able to buy product very easily.

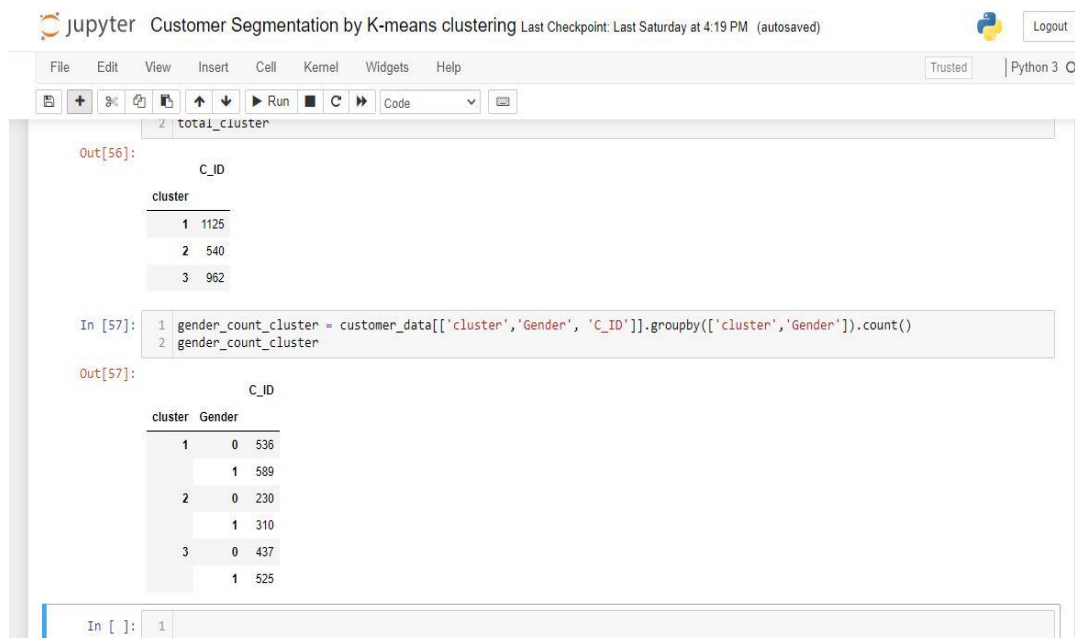


Figure: result

Chapter 7

Future Plan

In this paper we show who is the real buyer according to previous dataset. But this paper is not satisfied completely error free lack of consumers lot of things. Near future we looking forward to this topic.

Conclusion

In recent years, we have used machine-learning algorithms to predict which customer coming to really shopping mall to buying products. Customer segmentation can have a positive impact on business if it is done properly. Many companies understand the benefits of defining and implementing customer segmentation and can create a competitive advantage. Successful customer segmentation projects are the ones that avoid the pitfalls mentioned. Master data management including data steward, well-defined processes and technology can help support customer segmentation and the value it brings to a company.

References

- [1] <https://www.datacamp.com/community/tutorials/k-means-clustering-python>
- [2] <https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning>
- [3] <https://www.youtube.com/watch?v=AaUMof6f5o8&t=826s>
- [4] Blanchard, Tommy. Bhatnagar, Pranshu. Behera, Trash. (2019). Marketing Analytics Scientific Data: Achieve your marketing objectives with Python's data analytics capabilities. S.I: Packt printing is limited.
- [5] Griva, A., Bardaki, C., Pramadari, K., Papakiriakopoulos, D. (2018). Sales business analysis: Customer categories use market basket data. Systems Expert Systems, 100, 1-16.
- [6] Hong, T., Kim, E. (2011). It separates consumers from online stores based on factors that affect the customer's intention to purchase. Expert System Applications, 39 (2), 2127-2131.
- [7] Hwang, Y. H. (2019). Hands-on Advertising Science Data: Develop your machine learning marketing strategies... using python and r. S.I: Packt printing is limited
- [8] Puwanenthiren Premkanth, - Market Classification and Its Impact on Customer Satisfaction and Special Reference to the Commercial Bank of Ceylon PLC. Global Journal of Management and Business Publisher Research: Global Magazenals Inc. (USA). 2012. Print ISSN: 0975-5853. Volume 12 Issue 1.
- [9] Puwanenthiren Premkanth, - Market Classification and Its Impact on Customer Satisfaction and Special Reference to the Commercial Bank of Ceylon PLC. Global Journal of Management and Business Publisher Research: Global Magazenals Inc. (USA). 2012. Print ISSN: 0975-5853. Volume 12 Issue 1.
- [10] Sulekha Goyat. "The basis of market segmentation: a critical review of the literature. European Journal of Business and Management www.iiste.org. 2011. ISSN 2222-1905 (Paper) ISSN 2222-2839 (Online). Vol 3, No.9, 2011
- [11] T.Nelson Gnanaraj, Dr.K.Ramesh Kumar N.Monica. AnuManufactured cluster analysis using a new algorithm from structured and unstructured data. International Journal of Advances in Computer Science and Technology. 2007. Volume 3, No.2.