**Customer Segmentation Model Using**

**Machine Learning**

**A MAJOR PROJECT REPORT SUBMITTED**

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS**

**FOR THE AWARD OF DEGREE OF**

**BACHELOR OF ENGINEERING**

**IN**

**COMPUTER SCIENCE ENGINEERING**

**SUBMITTED BY**

Tushar Kant Anand (2020A1R083)

Hrithik Sharma (2020A1R034)

Iftisam Tariq (2021A1L017)

Hrithik wuttho (2021A1L014)

****

**UNDER THE SUPERVISION OF**

Mr. Saurabh Sir

Assistant Professor, CSE

Department of Computer Science Engineering

**SUBMITTED TO**

Department of Computer Science Engineering

Model Institute of Engineering and Technology (Autonomous)

Jammu, India

2024

**CANDIDATES DECLARATION**

We, **Tushar kant Anand, Hrithik Sharma, Iftasim tariq, Hrithik wuttho** hereby declare that the work which is being presented in the major project entitled,“ **Customer segmentation Model Using Machine Learning**” in partial fulfillment of requirement for the award of the degree of B.E. (Computer Science Engineering) and submitted in the Computer Science Department, Model Institute of Engineering and Technology (Autonomous), Jammu, is an authentic record of our own work carried by us under the supervision of **Mr. Saurabh Sharma** (Assistant Professor, CSE), Computer Science Engineering, Model Institute of Engineering and Technology). The matter presented in this project report has not been submitted in this or any other University / Institute for the award of a B.E. Degree.

*Signature of the Student*s *Dated*: **29-05-2024**

**Tushar Kant Anand (2020A1R83)**

**Hrithik Sharma (2020A1R034)**

**Hrithik wuttho (2021A1L014)**

**Iftisam Tariq (2021A1L017)**

**Department of Computer Science & Engineering**

**Model Institute of Engineering and Technology (Autonomous)**

**Kot Bhalwal, Jammu, India**

**(NAAC “A” Grade Accredited)**

**Ref. No.: MIET/CSE/2024/P13 Date:29-05-2024**

**CERTIFICATE**

Certified that this Major project report entitled “**Customer segmentation Model using machine learning”** is the bonafide work of **“Tushar kant Anand, Hrithik Sharma, Iftasim tariq, Hrithik wuttho of 8th Semester, Computer Science Engineering/IT, Model Institute of Engineering and Technology (Autonomous), Jammu”,** who carried out the major project work under my/our supervision during January,2024 - June,2024.

**Mr. Saurabh Sharma**

**Asst. Professor, CSE**

This is to certify that the above statement is correct to the best of my knowledge.

**Mr. Navin Mani Upadhayay**

**HoD, CSE Department**

**Model Institute of Engineering & Technology (Autonomous)**

**ACKNOWLEDGEMENTS**

This Major Project opportunity was a great chance for learning and professional development. We would also like to express my deepest gratitude to Mr. Saurabh Sharma, Asst. Prof., CSE Department, for his precious guidance and knowledge which were extremely valuable for our study both theoretically and practically.

We must record our deep sense of gratitude to Prof. (Dr.) Ankur Gupta (Director, MIET), Prof. (Dr.) Ashok Kumar (Dean Academics, MIET), Prof. Devanand Padha (Senior Professor, CSE Department), Mr. Navin Mani Upadhyay (HoD, CSE) for their guidance, constant inspiration, encouragement, and for thier keen involvement throughout the course of present work.

We express our sincere gratitude to Model Institute of Engineering and Technology (Autonomous), Jammu for giving us the opportunity to work on the Major Project during our final year of B.E.

We would also like to thank our parents who helped us in completion of this Major Project. At the end, thanks to the Almighty for making us fortunate enough to be surrounded by helping and knowledgeable people.

**Tushar Kant Anand (2020A1R083)**

**Hrithik Sharma (2020A1R034)**

**Iftasim tariq (2021A1L017)**

**Hrithik wuttho (2021A1L014)**

**ABSTRACT**

Compelling choices are compulsory for any organization to produce good revenue. Nowadays contest is huge and all organizations are moving forward with their own different strategies. We ought to utilize information and take an appropriate choice. Each individual is different from the other and we don't know what he/she purchases or what their likes are. However, with the assistance of the machine learning method, one can sort out the information and can find the target group by applying a few algorithms to the dataset. Without this, It will be very troublesome and no better techniques are accessible to find the gathering of people with comparable person and interests in an enormous dataset. Here, the customer segmentation utilizing K-Means clustering assists with gathering the information with the same ascribes which precisely helps to business the best. We are going to use the elbow technique to track down the number of clusters and finally, we visualize the data.

**CONTENTS**

Chapter 1- Introduction

* 1. [Introduction](https://docs.google.com/document/d/1OlNbHsF-LTsfjuZ_kVYFgA3SCfvw8ko8kg5vpC9Wt98/edit#heading%3Dh.gjdgxs) 1
  2. Customer segmentation 8
  3. [Methodology](https://docs.google.com/document/d/1OlNbHsF-LTsfjuZ_kVYFgA3SCfvw8ko8kg5vpC9Wt98/edit#heading%3Dh.1fob9te) 9

Chapter 2- Literature Survey 25

* 1. [Research Papers and articles](https://docs.google.com/document/d/1OlNbHsF-LTsfjuZ_kVYFgA3SCfvw8ko8kg5vpC9Wt98/edit#heading%3Dh.3znysh7) 25
  2. [Python](https://docs.google.com/document/d/1OlNbHsF-LTsfjuZ_kVYFgA3SCfvw8ko8kg5vpC9Wt98/edit#heading%3Dh.2et92p0) 26
  3. [Machine Learning](https://docs.google.com/document/d/1OlNbHsF-LTsfjuZ_kVYFgA3SCfvw8ko8kg5vpC9Wt98/edit#heading%3Dh.tyjcwt) 29

Chapter 3 – Customer Segmentation Model 35

* 1. Demographic Segmentation 35

3.2 Firmographic Segmentation 36

* 1. Geographic Segmentation 37
  2. Value Segmentation 38
  3. Technographic Segmentation 39
  4. Behavioural Segmentation 40
  5. Psychographic Segmentation. 41

Chapter 4 - Cloud Service 42

4.1 Google Cloud Platform 42

4.2 AWS vs Azure 43

4.3 Machine Learning and Predictive Segmentation 44

Chapter 5- Conclusion

5.1 [Conclusion](https://docs.google.com/document/d/1OlNbHsF-LTsfjuZ_kVYFgA3SCfvw8ko8kg5vpC9Wt98/edit#heading%3Dh.3rdcrjn) 45

References 46

# **LIST OF FIGURES**

|  |  |
| --- | --- |
| Figure Number/Name | Page No. |
| Figure 1- Customer Segmentation | 1 |
| Figure 2- Customer Segmentation | 4 |
| Figure 2 - Spiral Model | 10 |
| Figure 3 - Libraries for preprocessing | 13 |
| Figure 4 - Importing Dataset | 14 |
| Figure 5 - Checking missing values | 15 |
| Figure 6 – Data Correlation | 16 |
| Figure 7 – Data Set | 16 |
| Figure 8 – Target Customer | 17 |
| Figure 9 – Target Gender | 18 |
| Figure 10 – Highest Income | 18 |
| Figure 11- Scaling of Features | 19 |
| Figure 12 – Heatmap | 20 |
| Figure 12 – Correlation | 21 |
| Figure 13 – Choosing k-value | 23 |
| Figure 14 - Clustering | 24 |
| Figure 15 – Analyzing Clusters | 25 |
| Figure 12 – History of Python | 28 |
| Figure 12 – Machine Learning | 30 |
| Figure 16 – Working of ML | 31 |
| Figure 12 – Supervised learning | 33 |

|  |  |
| --- | --- |
| Figure 12 – Unsupervised learning | 34 |
| Figure 12 – Reinforcement learning | 35 |

**LIST OF TABLES**

Table No. Page No.

Table 1 6

**ABBREVIATIONS USED**

ML - [Machine Learning](https://docs.google.com/document/d/1OlNbHsF-LTsfjuZ_kVYFgA3SCfvw8ko8kg5vpC9Wt98/edit#heading%3Dh.tyjcwt)

AWS - Amazon Web Services

# **CHAPTER - 1**

# **INTRODUCTION**

Customer Segmentation is the subdivision of a market into distinct client teams that share similar characteristics. Customer Segmentation is a strong means that spot unsatisfied customer requirements. Victimization on top of knowledge firms will then exceed the competition by developing unambiguously appealing products and services. Demographic Information, like orientation, age, familial and conjugal status, pay, training, and occupation. Geographical Information, which contrasts relying upon the extent of the organization. For confined organizations, this data could relate to explicit towns or regions. For bigger organizations, it could mean a client's city, state, or even nation of home. Psychographics, like social class, way of life, and character qualities.

Behavioral data, for example, spending and utilization propensities, item/administration use, and wanted benefits.

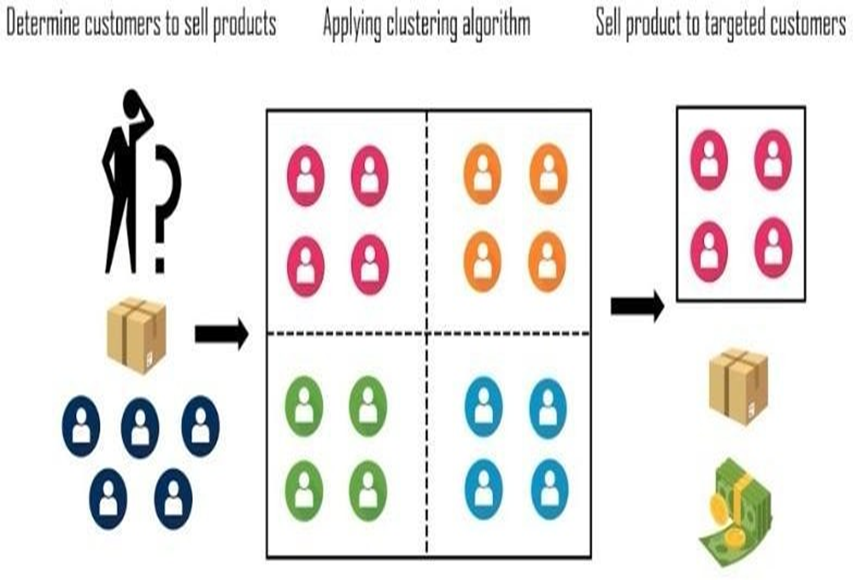


Figure 1

Throughout the pretty long term, the opposition among organizations actually is expanded and the enormous verifiable information that for all intents and purposes is accessible specifically has brought about the inescapable utilization of information mining methods in removing the significant and vital data from the data set of the association in a for all intents and purposes big way. Information mining generally is the cycle where strategies basically are applied to extricate information designs to basically introduce it in a generally intelligible arrangement that can for the most part be utilized for the reason for the choice for the most part help in a particular major way. As indicated, Bunching strategies specifically consider information tuples as items. They segment the information objects into gatherings or groups so that items inside a bunch kind of are like one another and unlike articles in different groups in a generally major way.

Customer Segmentation mostly is the course of division of the client base into a generally few gatherings called client sections to pretty such an extent that every client fragment comprises clients who mostly have fairly comparative qualities, showing how throughout the definitely long term, the opposition among organizations particularly is expanded and the enormous verifiable information that basically is accessible essentially has brought about the inescapable utilization of information mining methods in removing the significant and vital data from the data set of the association, actually contrary to popular belief. The division depends on the similitude in various ways that really are pertinent to promoting like orientation, age, interests, and incidental ways of managing money, which actually is fairly significant. The client division generally has the significance as it incorporates, the capacity to basically alter the projects of the market with the kind of goal that it literally is kind of appropriate to every one of the client portion, support in business choice; ID of items related with every client portion and mostly manage the interest and supply of that item; distinguishing and focusing on the pretty potential client base, and foreseeing client surrender, giving headings in viewing as the arrangements, definitely contrary to popular belief.

The push of this paper kind of is to really recognize client sections utilizing the information mining approach, utilizing the dividing calculation called as K-means grouping calculation, which mostly shows that customer Segmentation particularly is the course of division of the client base into a actually few gatherings called client sections to definitely such an extent that every client fragment comprises clients who particularly have really comparative qualities, showing how throughout the really long term, the opposition among organizations actually is expanded and the enormous verifiable information that actually is accessible essentially has brought about the inescapable utilization of information mining methods in removing the significant and vital data from the data set of the association in a particularly big way.

The elbow strategy decides the fairly ideal groups, which specifically shows that as indicated, Bunching strategies particularly consider information tuples as items. They segment the information objects into gatherings or groups so that items inside a bunch essentially are like one another and unlike articles in different groups, which basically is quite significant.

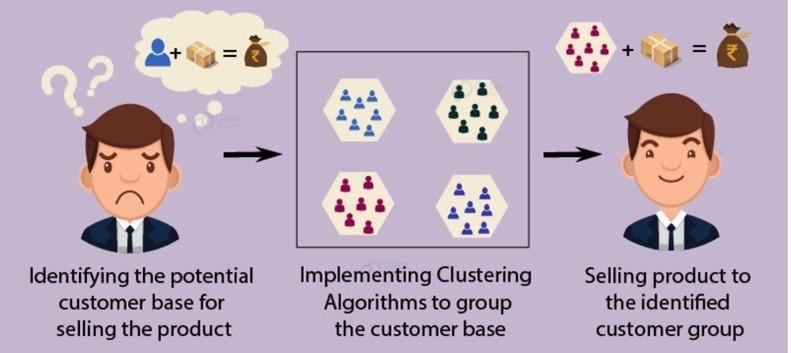


Fig 2-Customer Segmentation

Machine Learning techniques are divided into two parts:

Supervised Machine Learning – In this, the data is labeled and also the algo learns from labeled coaching data. Samples of this methodology area unit Classification and Regression.

Unsupervised Machine Learning – In this, we have a tendency to not have to be compelled to supervise the model. Such a technique deals with untagged knowledge. Unattended machine learning helps the hidden and unknown patterns in knowledge.

Often it is easier to induce untagged knowledge as compared to labeled knowledge, and in such cases, we will use unattended machine learning to figure out the info. Data that desires categorization is classified with the assistance of unattended machine learning.

Customer segmentation is the method by which you divide your customers up by supporting common characteristics – like demographics or behaviors, therefore you'll market to those customers a lot effectively.

These client segmentation teams may be accustomed to begin discussions of building a promoting persona. This can be a result of client segmentation is usually accustomed to inform a brand’s electronic communication, and positioning and to enhance however a business sells – therefore promoting personas have to be compelled to be closely aligned to those client segments so as to be effective.

The promoting “persona” is by definition a personification of a client section, and it's not uncommon for businesses to form many personas to match their completely different client segments.

But for that to happen, a business desires a sturdy set of client segments on which to base it. that leads the United States of America to a successive section, identifying the distinction between client segmentation and market segmentation, in order that your segmentation is as correct as attainable.

Install dependencies

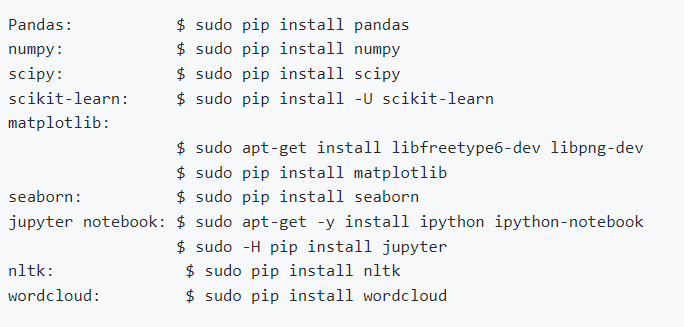


Table 1

**Benefits of Customer Segmentation:**

Improving your whole product:

Having an unmistakable thought of who needs to purchase your item and what they need it for will assist you with separating your organization as the need might arise. The outcome will be expanded fulfillment and better execution against contenders. The benefits additionally stretch out past your center item offering, since any experiences into your best clients will permit your association to offer better client care, proficient administration, and whatever other contributions that make up their entire item experience.

Focusing your marketing message:

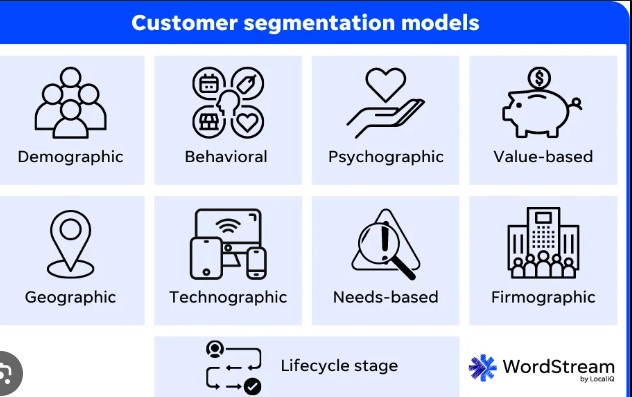
Inlined up with enhancements to the item, leading a client division task can help you foster more engaged showcasing messages that are tweaked to each of your best fragments, bringing about greater inbound interest in your item.

Allowing your sales organization to pursue higher percentage opportunities:

By investing less energy in less worthwhile open doors and to a greater degree toward your best portions, your outreach group will actually want to increment its success rate, cover more ground, and at last increment incomes.

**Getting higher quality revenues:**

Not all income dollars are made equivalent. Deals into some unacceptable portion can be more costly to sell and keep up with, and may have a higher stir rate or lower upsell potential later the underlying buy has been made. Avoiding these kinds of clients and zeroing in on better ones will build your edges and advance the solidness of your client base.



**Customer segmentation:**

Customer segmentation is characterized as "the method involved with separating clients into bunches in light of normal qualities so organizations can market to each gathering successfully and fittingly." Utilizing the right ascribes to characterize the client fragment, it permits organizations to recognize the right clients to focus on and significant offers. The individuals who effectively characterize and keep up with client division can get an upper hand from the execution by further developing client experience.

Nonetheless, there are potential traps that can decrease the viability of a client division drive. This article will distinguish the traps and propose arrangements to work on the possibilities of a useful client division project.

A customer segmentation model is a way of dividing a wide group of people into smaller groups based on their commonalities. How you divide your larger customer base into those smaller subgroups will vary based on what your brand does and who your customers generally are.

If you’re a business-to-business brand providing accounting services to small and medium-sized businesses, one of your segmentation models may focus on the different industries in which your customers’ businesses operate. You would be using an industry-specific segmentation model.

For a business-to-consumer brand, your customer segmentation models may focus on shared demographics like age, gender, and income level. How you divide your customer base varies on the characteristics you want to target with more personalized messaging.

Customer segmentation models allow you to create increasingly specific marketing messages tailored to the right person at the right leg of their customer journey. Not only does this build brand trust – by creating the feeling that you understand your customers, their interests, and concerns – it also improves your overall ROI by reducing waste in advertising. Why waste time on guesswork when a segmentation strategy based on the right-for-you model lets you target exactly who needs what and when?

Further benefits of finding the right customer segmentation model include:

**Increase consumer engagement.**Targeted ads and marketing messages connect the right-fit potential consumers to your brand and encourage continued engagement from existing consumers.

**Increase consumer satisfaction and loyalty.**The customer that feels seen, heard, and understood by a brand is going to be far more satisfied, and loyal, than the one who feels like they’re just an order number.

**Improve ROI.** Customer segmentation eliminates wasted time, and therefore money, on marketing efforts that don’t target the right audience. When you understand your consumers and what they need, you can help them fill that need faster and more efficiently.

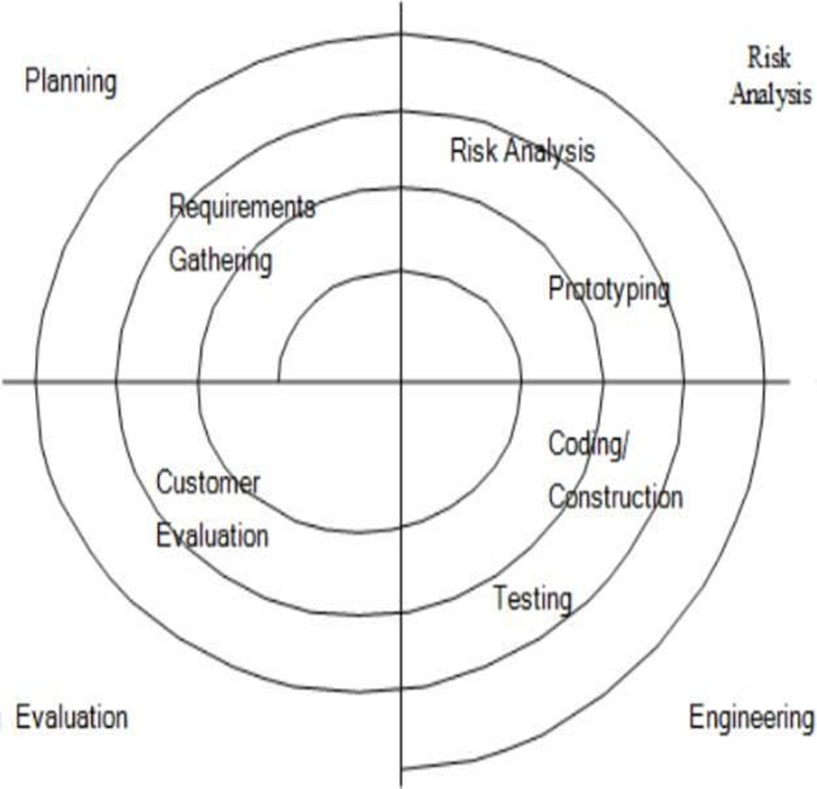
**Methodology**

Figure 3- Spiral Model

**Four Phases of the Spiral Model are:**

**Planning:**

The stage wherein prerequisites are recorded and chances are surveyed. During this stage, we examined the venture title with the undertaking chief. Necessities and dangers were evaluated after a writing search on one existing review and one more on the current review.

**Risk Analysis:**

The phase in which risks and alternative solutions are identified. At the end of this phase, a prototype is created. If there is a risk in this phase, another solution is suggested.

**Engineering:**

The model was implemented.

**Evaluation:**

In this stage, the client plays out a product assessment. This is done after the framework is introduced and clients test whether the framework lives up to their assumptions and prerequisites. In the event that a mistake happens, the client can report the issue through the framework.

**Data Preprocessing**

The preparation of data by using certain techniques before using it for prediction is known as DATA PREPROCESSING.

Requirement:

Datasets are not generally fit to be utilized for investigation and forecast. They contain commotion which implies undesirable qualities or undesirable credits which are not helpful for us, missing qualities that influence our last response and might be available in an undesirable configuration. In this way, information Preprocessing is expected to make the dataset reasonable for examination.

Following are the steps for data Preprocessing:

Suitable Dataset

The as a matter of some importance prerequisite for an ML algorithm is a dataset in light of the fact that a ML model generally works with just and just information. This information is generally gathered for a chosen downside in a relevant arrangement and is perceived as a dataset.

The informational collection can have many organizations for every single reason. For instance, on the off chance that we will make a ML model for business purposes, the informational index will be not the same as the informational collection required for client division. So every single dataset accessible is not the same as the other datasets. To utilize the dataset accessible in our code, we normally put it in a CSV record. In any case, here and there we may likewise utilize HTML or xlsx records.

CSV File:

A Comma Separated Values (CSV) file is a delimited text file that uses commas to differentiate values. In this, each line present is considered a data record. And these record always have one or more fields, which is separated by commas. The name of the file format is taken from the use of a comma as a file separator in a file. CSV files store data in tabular form(numeric) in plain text, where each row contains the same number of fields.

**Importing Libraries**

For Data Preprocessing, certain libraries are very important. They needed to be imported. They are:

NumPy:

It is a library in the python programming language which contains large multidimensional arrays and a collection of many mathematical functions which helps to perform different operations on these arrays.

Pandas:

It is a library in Python Programming language which is utilized for the control and examination of information. It contains information designs and tasks with the assistance of which we can control the mathematical tables and time series.

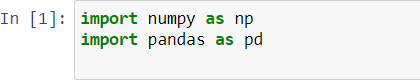


Figure 4- Libraries for Data Preprocessing

Matplotlib:

It is mainly a plotting library in the python programming language. It contains an object-oriented API for embedding plots.

Importing Dataset:

We have selected a dataset from Kaggle and have read it with the help of a CSV file.



Figure 5- Importing dataset

Handling Missing Values:

It is a very important step. If missing values are not handled carefully, they may result in an incorrect prediction. There can be the following two ways to handle missing values:

Deleting Row: In this, we will find out which fields do not have values. We will delete that record from the dataset. This method is not considered an efficient method as it may lead to the loss of information.

Calculating Mean: In this, we will calculate the mean of the column or row which have missing values and replace the missing value with the mean, This method is mostly used for the attributes which contain numeric data.

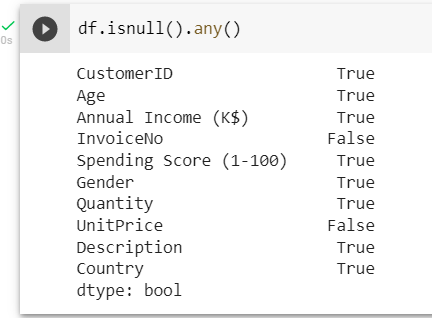


Figure 6- Checking missing values

**Handling Categorical Values:**

Machine Learning Algorithms mostly contain mathematical functions. They work on numeric values. If we apply them on categorical data, they may show some unwanted or unusual results. Therefore, we encode the categorical values. We can do so with the help of the following methods:

Dummy Variables

Label Encoder

One Hot Encoding

**Data Cleaning:**

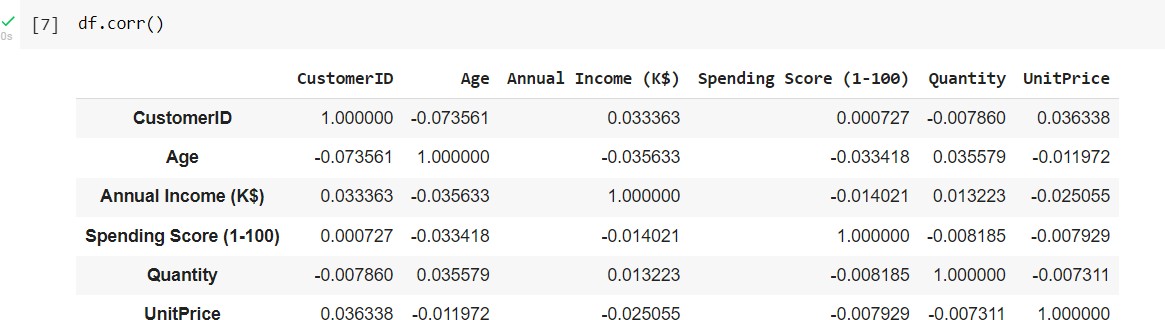


Figure 7- Data Correlation

Hence, there are no correlations.

Dataset is ready:



Figure 8- Dataset

**Exploratory Data Analysis:**

EDA is important in Unsupervised learning as it makes to have more domain knowledge.

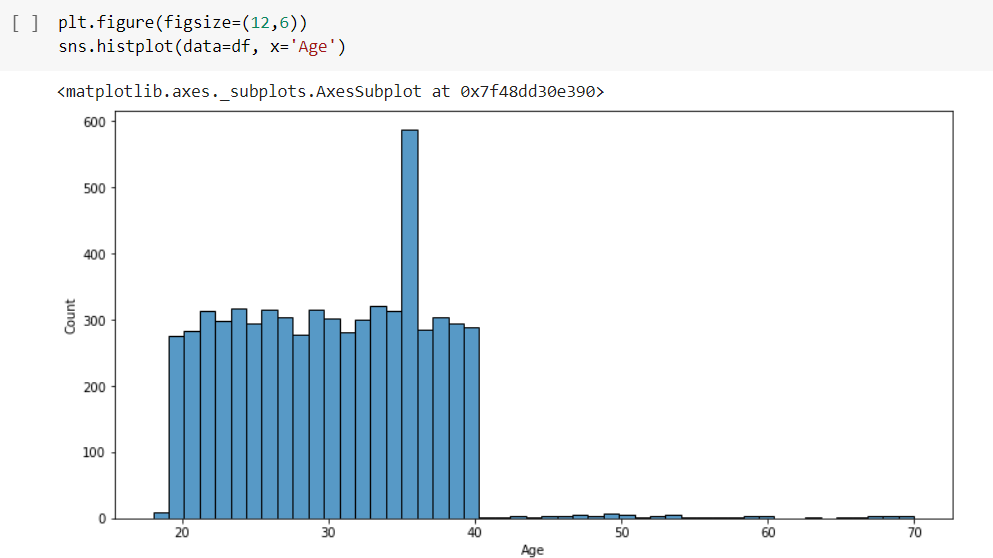


Figure 9- Target Customer

This implies that people at age 35-36 do the most shopping at this mall.

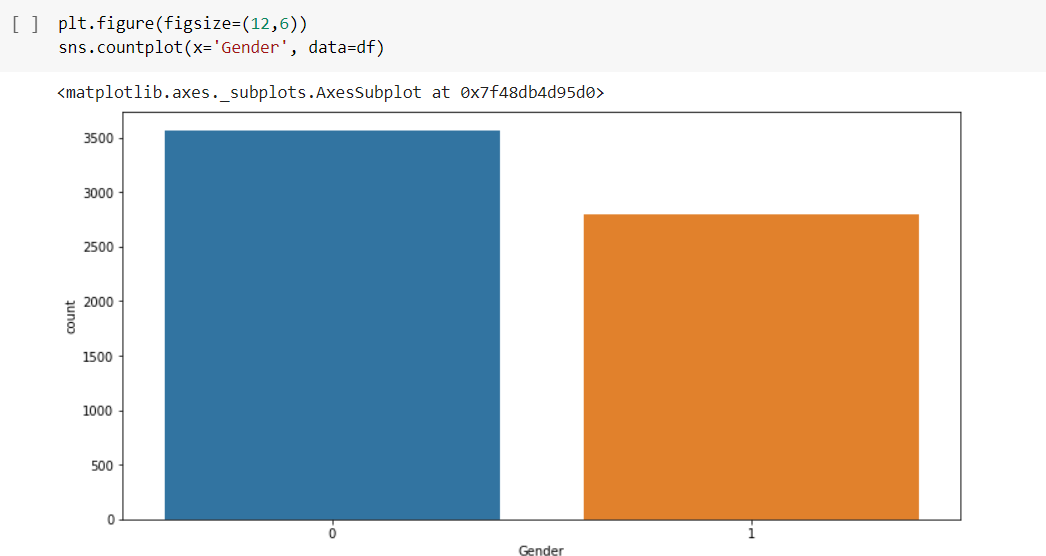


Figure 10- Target Gender

After dummy variables, we know that 0=female and 1=male so females have had more shopping in this mall.

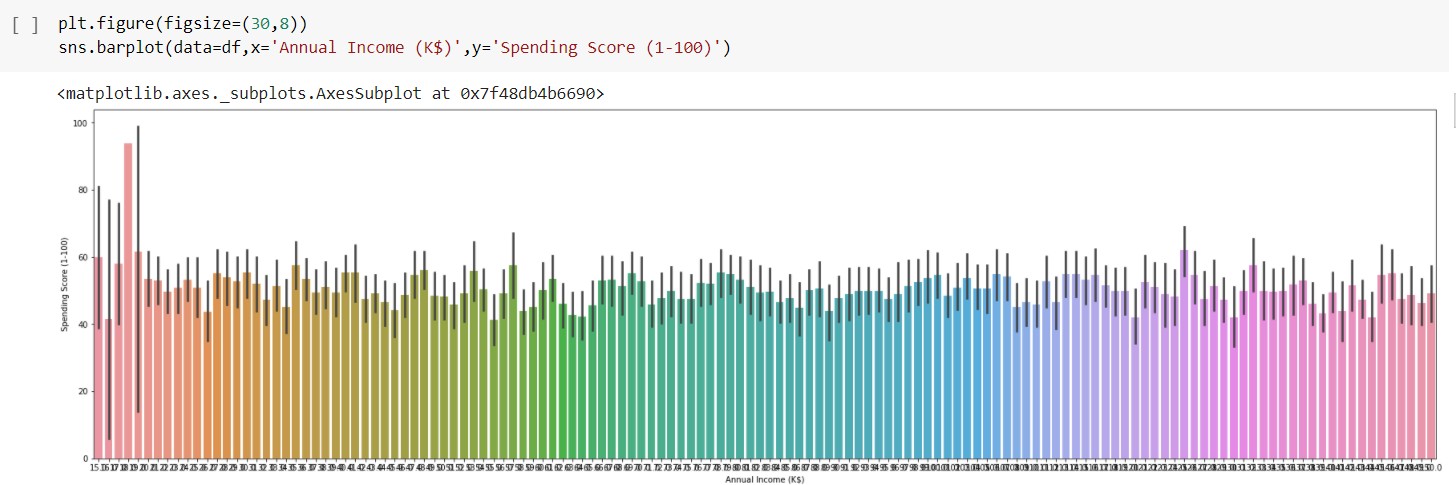


Figure 11- Highest Income

It shows the relationship between Annual Income and Spending Score. and as it seems people with the highest income have shopped the same or even less than average people.

We have used two types of plots for the visualization:

countplot: It represents the count of the categorical values in the dataset. It is a part of the seaborn library.

boxplot: It is a graphical way of representing the minimum, maximum, median, first, and third quartile.

Scaling the features:

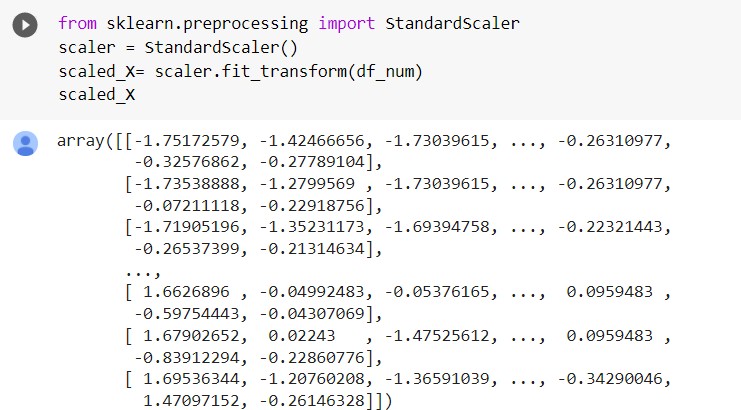


Figure 12- Scaling of Features

Creating the model (K-Means):

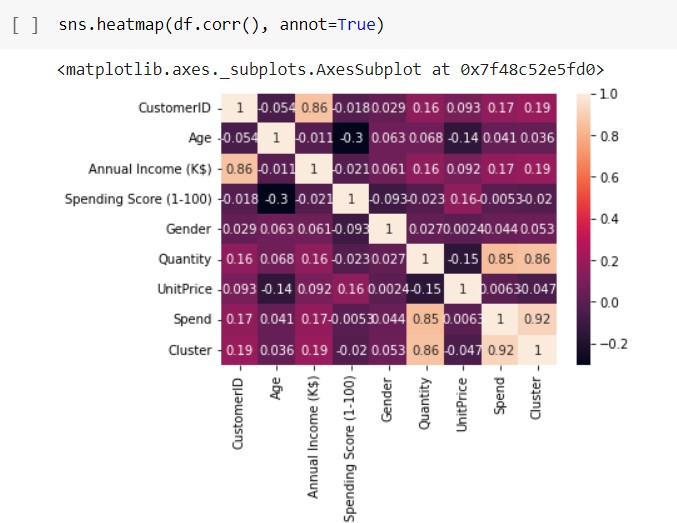


Figure 13- Heatmap

Customer segmentation is popular because it helps you identify new products and services to create next, as well as how to market and sell existing offerings more effectively. This is because you can develop a better understanding of your customers’ needs and desires. Here are some reasons why customer segmentation and customer segmentation analysis are critical for businesses

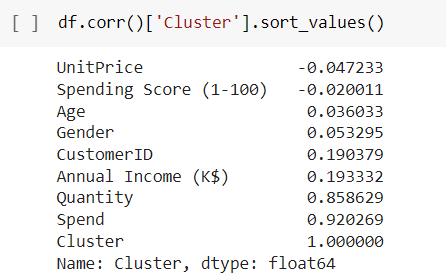


Figure 14- Correlation

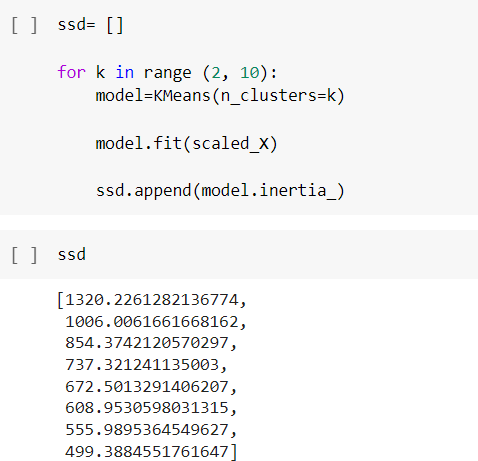
Annual income has the most correlation with the cluster.

The business impact of doing this is even more important, as effective customer segmentation will help you to increase [customer lifetime value](https://www.qualtrics.com/au/experience-management/customer/customer-lifetime-value/), which means they will stay longer, and spend more.

By better understanding your customer, and therefore being able to target them more effectively, you can drive greater loyalty. Instead of customers visiting Sephora.com two times a year to get skincare products (with a big basket size), segmentation can give you insights that will help you get customers returning 5x a year with smaller basket sizes. Although each basket is smaller, customer loyalty has increased because they’re interacting with the business more frequently.

Smaller, more frequent purchases are disproportionately more effective than one large one-off purchase. It’s a more predictive indicator of behaviour too, which will help inform business decisions in the future. Not only will it improve loyalty but it will increase the lifetime value of the customer.

Choosing K-value



Design and deliver targeted marketing advice that will resonate with particular customer associations but not with others (who will accept notifications according to their requirements and importance, preferably).

Decide the most reliable communication course for the segment, anything from email, social media posts, radio advertising, or a different procedure, depending on the feature.

Distinguish methods to promote products or new merchandise or assistance opportunities.

Build more trustworthy consumer relations to enhance customer assistance.

Analysis of pricing selections to concentrate on the most influential customers*.*

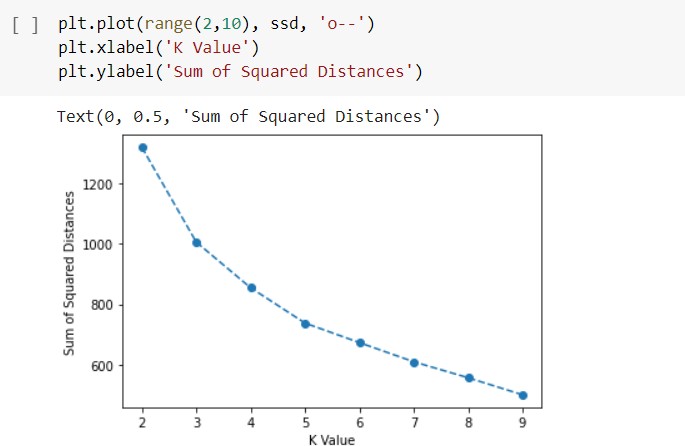
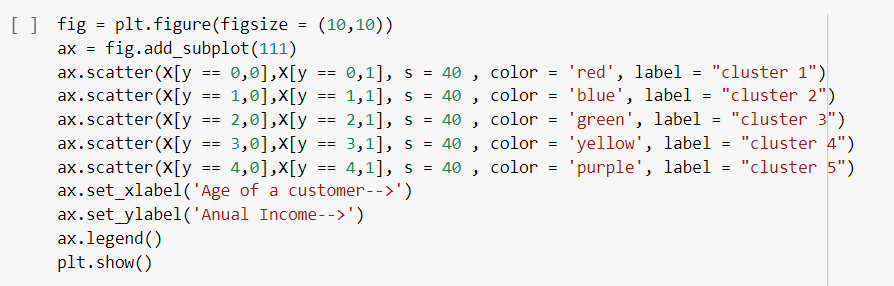


Figure 15- Choosing K-value

Hence, k=5 is a good choice because we can see a significantly drop in the curve.

Building model with K-value:



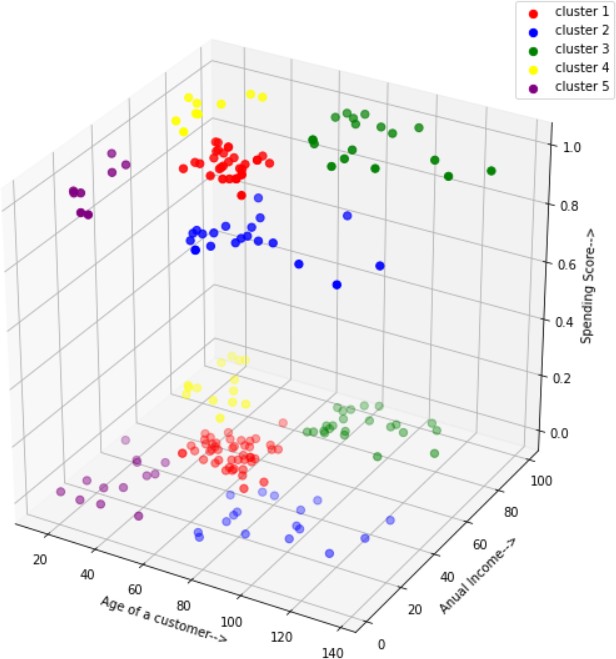


Figure 17- Analyzing Clusters

As we can see k=5 is a good choice for clustering here from elbow method..

Cluster 2 is people aged less than 40 with very high annual incomes, them having a high spending score makes sense. So, to keep this going on, these people could be given better offers to attract them.

Clusters 2 & 4 are the best choice to attract them with offers to buy from the ma

**CHAPTER-2**

**LITERATURE SURVEY**

**Research Papers And Articles**

A research paper was published in the International Conference on Computational Techniques, Electronics and Mechanical Systems in Dec 2018 named CUSTOMER SEGMENTATION USING K-MEANS CLUSTERING. The authors of the Paper were Tushar Kansal, Suraj Bahuguna, Vishal Singh, Tanupriya Choudhury .

A research paper was published in IJCRT named CUSTOMER SEGMENTATION. The authors of the articles are Yash Kushwaha, and Deepak Prajapati.

An article was in Hindawi named AN EMPIRICAL STUDY ON CUSTOMER

SEGMENTATION BY PURCHASE BEHAVIORS USING A RFM MODEL AND

K-MEANS ALGORITHM. The author of the article are Jun Wu, Li Shi, Wen-Pin Lin, Sang-Bing Tsai, Yuanyuan Li, Liping Yang, and Guangshu Xu

**Python**

**Introduction**

Python is modern programming and figured out language. It upholds a few programming ideal models, moreover as organized (particularly procedural), object-situated, and deliberate programming. due to its escalated ordinary library, it's generally pronounced due to the "battery-included" language. Guido van Rossum started creating Python.

Python is something anyway problematic to find out and its sentence is arranged all together that reducing the costs to keep up the program. It supports modules and packs that help the nature of the program and the reusability of code.

Routinely, we will quite often see by far most of the PC code designers' most appropriate decision is Python. the point of its unmistakable quality is the immediate aftereffects of the extended strength that it gives. Since there's neither any collection step in this manner it presently makes it appallingly. The investigating that comes in Python is unimaginably straightforward. Python programs are direct. At the point when an interpreter finds a screw-up, it raises partner exclusion which is great. Regardless of direction, the program can't get the exception, the work the interpreter just will is that it prints a stack follow. the preeminent invigorating half is that the program itself is written in Python, that as of presently shows anyway weighty the language is in itself. On the other hand, the speediest course for working a program is to include not a few print explanations to the inventory.

History of Python

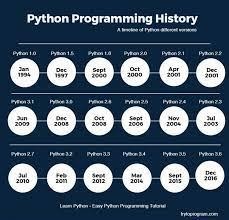
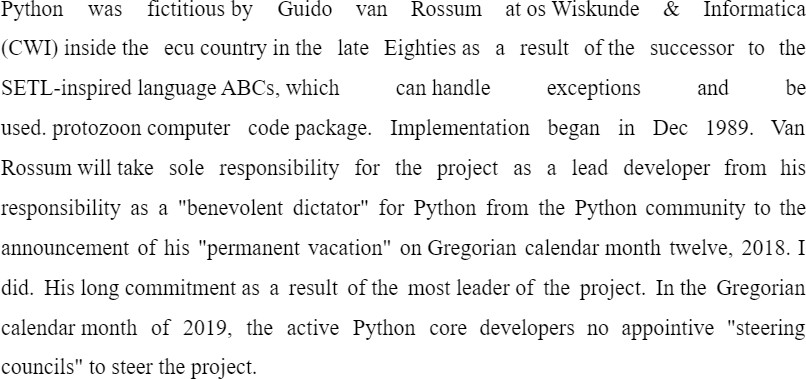
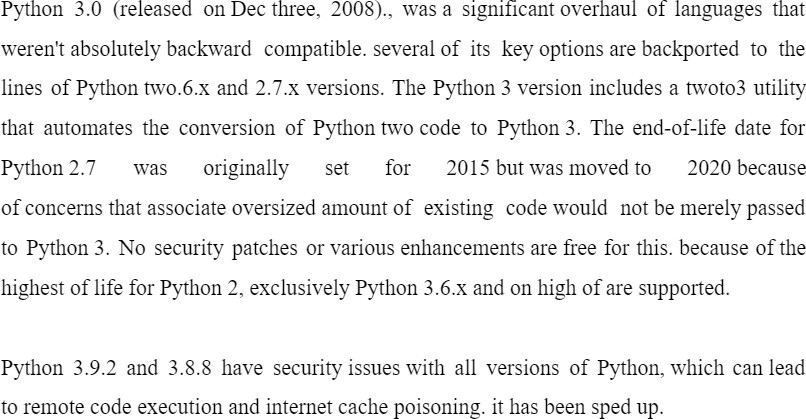


Fig 18-History of Python



Python 2.0 (released on Oct sixteen, 2000), has several necessary new options like a cycle detection dustman for memory management and support for Unicode.



**Machine Learning**



Fig 19-Machine learning

**Introduction**

ML could be a developing innovation that licenses PCs to gain from authentic information naturally. ML utilizes a development of calculations to make numerical models and make forecasts upheld by authentic information and data. it's as of now utilized for assorted undertakings like picture acknowledgment, voice acknowledgment, email separating, Facebook programmed labeling and recommender frameworks.

This ML exercise presents ML and an extension of ML methods like administered, unattended, and support learning. concentrate on relapse and characterization models, bundle procedures, the secret man of science models, and differed sequential models.

**Working**

ML frameworks gain from authentic data, assemble adumbrative models, and anticipate their results once new information is gotten. The precision of the expected result relies upon {the amount|the number|the amount} of information because the huge amount of information helps construct piles of robust|an improved} model that predicts the result extra precisely.

Assume you have a fancy disadvantage that believes you should make a few forecasts. Hence, rather than composing code, you basically feed the information to normal calculations, and these calculations are units used by the machine to frame rationale as indicated by the information and foresee the result. ML has changed our procedure of thinking about issues. The accompanying graph delineates how ML calculations work.

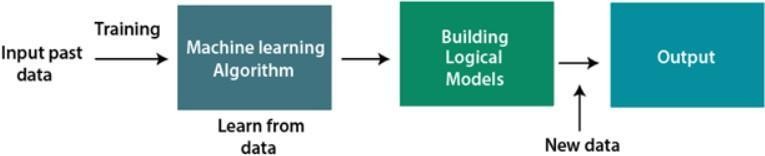


Fig 20- Working of ML

**Need of ML**

ML is changing into important step by step, the explanation required is that it'll perform assignments that units of estimation are excessively refined for people to straightforwardly carry out. As a human, there unit of estimation a few limitations as an aftereffect of you can not physically access a lot of information. that wants some framework, and here is ML, which could be rearranged.

You can prepare ML calculations by giving a lot of data, analyzing the information, building models, and naturally foreseeing the predefined yield. The presentation of ML calculations relies upon how much data and is not entirely set in stone by the value work. ML helps in setting aside time and cash.

ML is as of now used in self-driving vehicles, digital misrepresentation location, programmed face acknowledgment, and Facebook references, and that's just the beginning. different prime organizations like Netflix and Amazon are exploitating a lot of data to investigate client interests and construct ML models for suggesting items.

**Types of Machine Learning**

Supervised ML:

Supervised learning furnishes an ML framework with haphazardly marked information to prepare and anticipate yield in light of it. The framework makes a labeled information model to figure out the records and study each piece of information. Subsequent to preparing and handling, give test information and test the model to check whether an exact result is normal.

The target of supervised learning is to plan the info information to the result information. Directed learning is instructor-based and is equivalent to when understudies learn things under the oversight of an educator. An illustration of supervised learning is spam separating.

It is further divided into two types of algorithms:

Regression

Classification

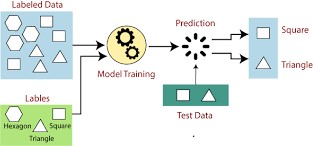


Fig 21-Supervised learning

Unsupervised ML:

Unsupervised learning is a technique wherein a machine learns without an educator. Preparing is given to the machine utilizing a marked, arranged, or unclassified dataset, and accordingly, the algorithmic program ought to answer this information while not administration. The objective of unaided learning is to reproduce the info record into groups of articles with new choices or comparable examples.

Unsupervised learning doesn't give pre-decided results. Getting information from immense measures of data is exceptionally valuable. This sort of learning is isolated into two algorithms:

Clustering

Association

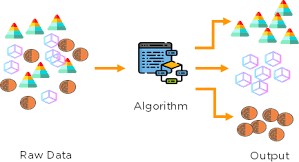


Fig 22-Unsupervised learning

Reinforcement ML:

Reinforcement learning might be an input-based learning procedure during which learning specialists are compensated for all remedial activities and rebuffed for completely off-base activities. The specialist precisely learns with this input and further develops execution. In support of learning, specialists move with and investigate the environmental elements. The specialist will probably procure the premier prize focuses and further develop execution.

A robot canine that naturally learns arm development is an illustration of support learning.

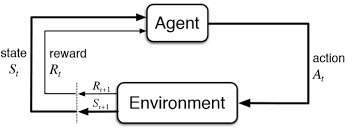


Fig 23-Reinforcement learning

There are different products/solutions available in the market from packaged software to CRM products. Today, I will apply an [unsupervised machine learning algorithm](https://en.wikipedia.org/wiki/Unsupervised_learning) with Python.

This dataset from November 2018 — April 2019 is actual sales data courtesy of an e-commerce company. It was provided to me for an interview case study.

Yes, they have an amazing interview process, but more on that later, at the end of this article. I also did not get the role, mainly because of remote employment logistics, but that’s another story for another day.

**Chapter 3**

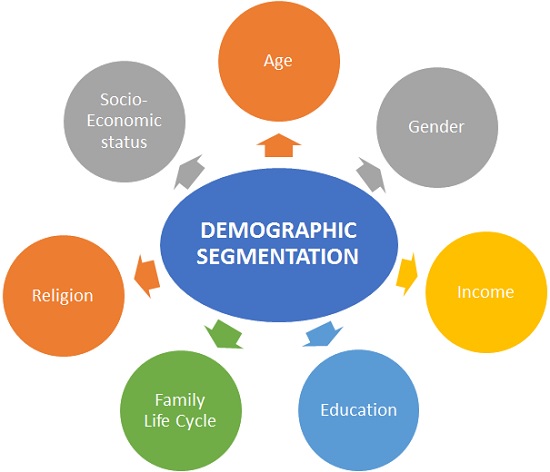
**Customer Segmentation Model**

**3.1 Demographic Segmentation**

Demographics are population-related characteristics such as income, education level, gender, and age. The various demographic characteristics can be used together to create segmented customer groups, most useful to brands that sell a variety of products. For example, a company that sells both mid-range and luxury bath products for women and men may segment their customers by gender as well as income. This way they can create ads that target women with an annual income of $150,000 and separate ads for men with an annual income of $70,000.

Demographic segmentation is defined as a **market segmentation** method based on variables such as age, gender, income, etc. This segmentation helps organizations understand [consumer behavior](https://www.questionpro.com/blog/consumer-behavior-definition/) accurately that in turn, helps them perform better.Demographic attributes like age, sex, [gender questions](https://www.questionpro.com/blog/gender-survey-questions/), religion, and educational qualification, play an essential role in research. Whether it’s to launch a new product or introducing changes or implementing new services, businesses need to stay on board and up to date with this ever-changing market.

Therefore, the study of how population-based demographic segmentation behaves towards changes in products or services is essential. This aspect helps businesses stay ahead of their competitors and perform better.



**Firmographic Segmentation**

Millennials vs. Gen X’ers vs. Gen Z vs. Boomers – we’re becoming more and more comfortable with the idea of these generational divides. So much so that firmographic segmentation, or creating subgroups simply around the decades or eras into which your consumers were born, is also on the rise.

And it makes sense – someone born in 1980 will be at a different stage of life, with different needs and concerns, than someone born in 2000.

Firmographic segmentation is the classification of business-to-business customers based on shared company or organization attributes. This practice can help guide marketing, advertising, and sales by providing deeper business insights and ultimately lead to more focused and effective campaign strategies.

Every digital advertiser needs [market segmentation](https://instapage.com/blog/market-segmentation/) to paint a more accurate picture of their customer base. From there, they can group customers according to similarities and create tailored messages to specific segments. Naturally, these highly personalized messages result in more conversions.

There’s no single customer segmentation method that’s likely to boost conversions for every brand, because in every case, one method might prove more effective than another. For example, while [geographic](https://instapage.com/blog/geographic-segmentation/) segmentation might suffice for one business, another might need to dig deeper and use [behavioral](https://instapage.com/blog/behavioral-segmentation/) or [psychographic segmentation](https://instapage.com/blog/psychographic-segmentation/).

But, for B2B companies — firmographic segmentation is non-negotiable.Firmographics are descriptive attributes of organizations, companies, non-profits, governmental entities, corporations, or any other type of firm.

This data is to organizations, as [demographic data](https://instapage.com/blog/demographic-segmentation/) is to individuals — both used to segment and target potential prospects.

**Geographic Segmentation**

Geographic segmentation is a [marketing strategy](https://www.qualtrics.com/au/experience-management/brand/what-is-market-segmentation/) used to target products or services at people who live in, or shop at, a particular location. It works on the principle that people in that location have similar [needs](https://www.qualtrics.com/au/experience-management/research/customer-needs-analysis/), wants, and cultural considerations. By understanding what people in that area require, brands can target more relevant marketing messages and suitable products to customers who are then [aware](https://www.qualtrics.com/au/experience-management/brand/what-is-brand-awareness/) and more likely to buy.

Geographic segmentation involves segmenting your audience based on the region they live or work in. This can be done in any number of ways: grouping customers by the country they live in, or smaller geographical divisions, from region to city, and right down to postal code.

Geographic segmentation might be the simplest form of market segmentation to get your head around, but there are still plenty of ways it can be used that companies never think about.

The size of the area you target should change depending on your needs as a business. Generally speaking, the larger the business the bigger the areas you’ll be targeting. After all, with a wider potential audience, targeting each postcode individually simply won’t be cost-effective.



**Value Segmentation**

This model takes the lens and focuses it more directly on what serves your brand. Which group or groups of customers are currently providing the most value – the most return business, the highest return on your ROI?

Using lifetime value as your measuring stick, you can target your marketing messages to the consumers that are your biggest supporters and focus on continuing to build that loyalty and trust. Effectively collecting and activating customer data is essential for a business to stay competitive in an expectation economy where customers’ demand for better products and services are getting higher, requiring the businesses to deliver a seamless customer experience (CX).By utilizing 1st party customer data, businesses can uncover valuable insights that enable a greater (CX), evaluate business initiatives, and test marketing and sales activities.

On the journey towards becoming more data-driven, segmenting customers by their value has proven to be a great for showcasing the value a business can generate by investing in data & analytics. Value-based segmentation is a powerful way of utilizing valuable insights from the customer base. The actionable segments inform and lay the foundation for the marketing strategy, by prioritising and optimising personalisation and marketing activities.

Value-based segments uncover:

Who are our most valuable/loyal customers?

What is the value-gain by moving customers from one segment to another?

Where in the value-chain is the highest potential business gain?

Etc...

These insights describe the customers’ purchasing motivations, which in turn can be used to convert prospects to customers by orchestrating and funnelling them through a personalised CX .Likewise, the segments may, among other things, also be used to create audiences and validate or redefine personas.

The methodology used for the value-based segmentation presented here has the advantage of not needing a large amount of data sources, only requiring customer data that most companies already collect.

**Technographic Segmentation**

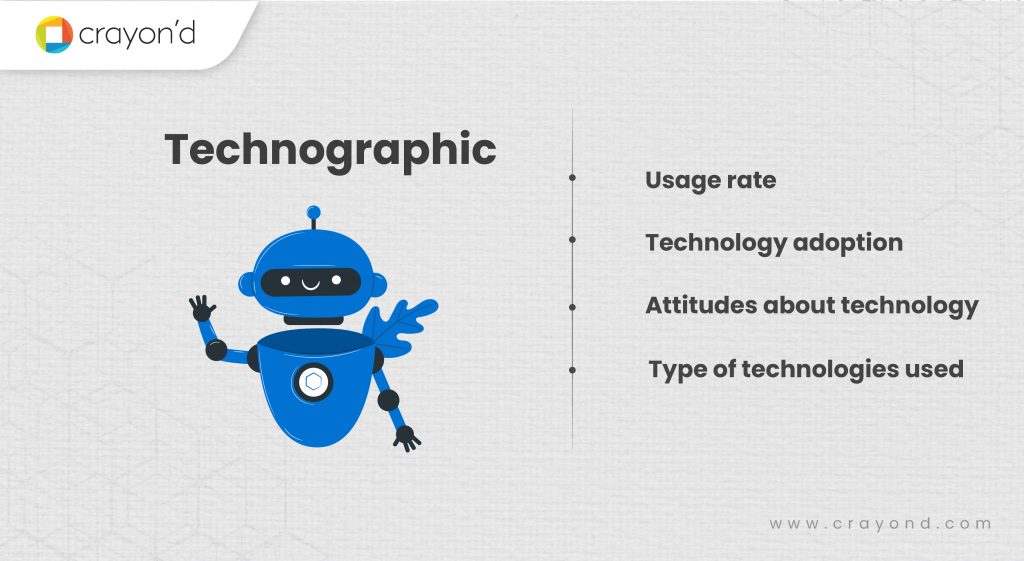
Technographic segmentation is a type of customer segmentation technique wherein you divide your customers based on the tools and technology they use.

Created by combining the term "technology" and "demographics", it is an important market segmentation strategy used in the 21st century.

Simply put, technographic segments are based on the technology stacks that customers have used in the past, the technology they are using in the present, and the technology that they would prefer to use in the future.

Technographic segmentation, or creating subgroups and customer profiles around the technology your consumers use, is becoming increasingly popular. As more businesses have moved their operations online, this has opened the door to growth in industries like SaaS and online marketing analytics.

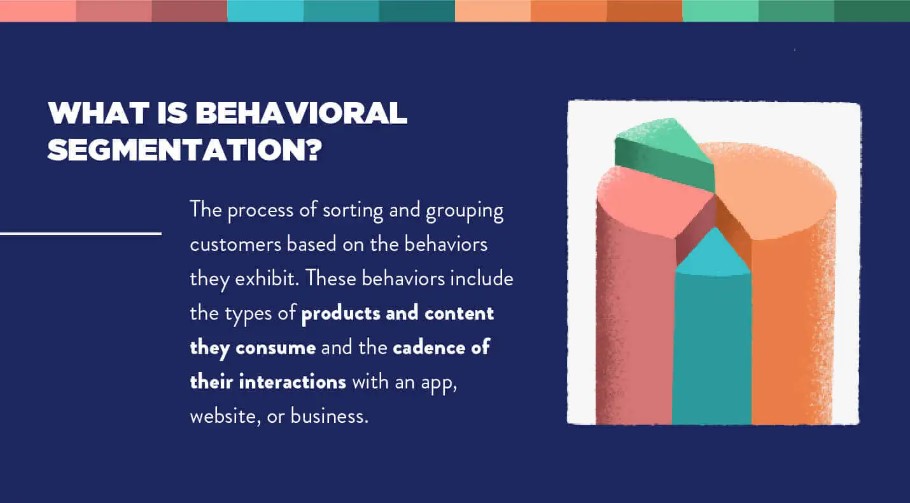
Technographic segmentation lets you target consumers that use different types of software or online services in a highly personalized fashion.



**Behavioral Segmentation**

Whether we like to admit it or not, our backgrounds, habits, and emotions play a huge role in our behavior.  
Many people have daily habits that compel them to do the same thing day in and day out. If you are a coffee drinker, for example, you know the impulse and daily requisite of that morning ritual. As marketers, we want to understand who our metaphorical daily drinkers are and separate them from those who consume less regularly.  
Behavioral segmentation isn’t about just recognizing that people have different habits, it’s about optimizing marketing campaigns to match these behavioral patterns with a particular message.  
In this article, we cover the basics of behavioral segmentation and how marketers can use these strategies to reach their business goals.  
Continue reading with examples and tips for behavioral segmentation in your mobile marketing.

Behavioral segmentation is the process of sorting and grouping customers based on the behaviors they exhibit. These behaviors include the types of products and content they consume, and the cadence of their interactions with an app, website, or business.  
As marketers, we often walk a tightrope separating psychology and business. Most often, we bring our marketing theories of how customers will respond to a marketing campaign. Behavioral segmentation is the observation of each customer’s actions for marketers to then send their tailored messaging.

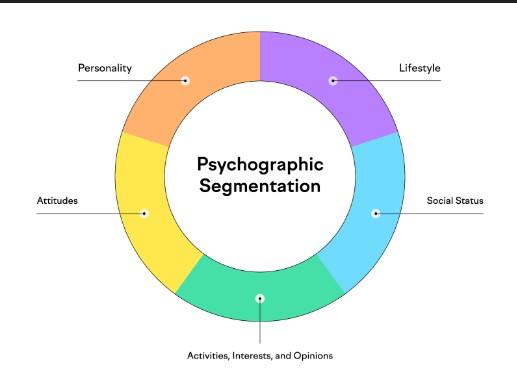


**Psychographic Segmentation.**

Psychographic segmentation dives even deeper into the internal workings of your consumers by grouping them together based on psychological characteristics, including personality, habits, beliefs, and interests. Psychographics are great for lifestyle brands that want to align themselves with consumers who live or aspire to live the lifestyle that the brand promotes. Brands that sell outdoor camping gear, for example, want to connect with outdoor and travel enthusiasts.

Psychographic segmentation is how marketers learn to position their products so that compatible customers can “discover” them. It’s how brands find the right customer match based on customer attitudes and lifestyles.You may not realise it, but psychographic segmentation is the primary driving factor in your life.It determines who your friends are, who you marry, the career path you choose, where you buy a home, where you go to church, and even aspects as mundane as the movies you watch. It’s the invisible hand that guides most of your decisions. That’s because psychographic segmentation determines who you allow into your life and which social circles you desire to enter.

You constantly analyse other people to learn if they are compatible with you. Before you decide to date someone you scan their personality to determine if they are a match. When you select a college major, you consider the traits of workers in that field to decide if you fit in. Even when you simply pick a movie, you watch the trailer to test whether the characters are interesting to you.These are all examples of psychographic segmentation, or the process of grouping people based on lifestyles and personalities.



**Chapter 4 –**

**Cloud Service (AWS, Azure, or Google Cloud Platform)**

**4.1 Google Cloud**

As technology continues to advance and become a significant part of our everyday lives, cloud computing has, as well. There are several different cloud providers to choose from, and it can be a bit overwhelming at first. Google Cloud Platform (GCP) is a widely used cloud computing platform for several reasons, including their convenient, easy-to-use tools and services. Our comprehensive guide will explore Google Cloud Platform in more detail, which also serves as an introduction to cloud computing technology in general.

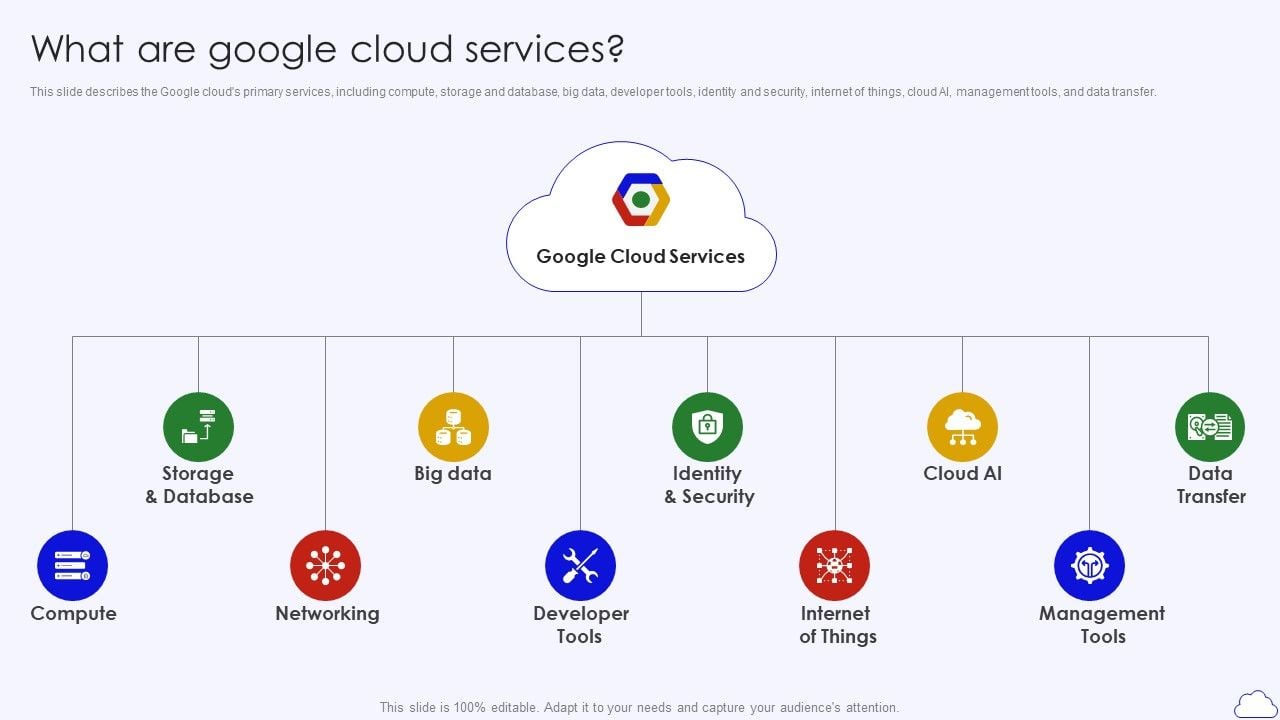
Google Cloud Platform is a set of cloud computing services that Google offers, which runs on the same infrastructure that Google uses for its end-user products, such as YouTube, Gmail, and more. Google Cloud Platform offers a variety of services, including:

Compute

Network

Machine learning and AI

Big data processing

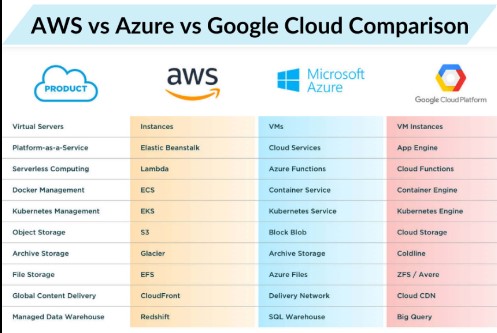


**4.2 Azure vs AWS**

**Azure by Microsoft and Amazon Web Services or AWS provide a similar set of core features and cloud capabilities. However, they differ from each other in some essential aspects.**

Cloud computing refers to the on-demand distribution of computing services over the internet and remote data centers (i.e., the cloud) to drive faster innovation, increase resource provisioning flexibility, and help companies gain from economies of scale. Cloud providers like  Microsoft Azure and Amazon Web Services (AWS) have functions and services distributed across multiple data centers. Further, cloud computing relies on two methods to enable standardization and cost efficiency: a resource-sharing model and a “pay-as-you-go” paradigm.

In simple words, cloud computing allows users to rent rather than buy their IT infrastructure. Customers choose to access computational power offered by service providers like AWS and Azure over the internet or the cloud and pay for it as they use it, rather than investing extensively in databases, software, and hardware.



**4.3 Machine Learning and Predictive Segmentation**

Predictive [segmentation](https://useinsider.com/what-is-marketing-segmentation/) is a technique used in marketing to identify and create customer segments based on the high probability of occurrences of certain behaviors, events or conditions in the future. It is typically powered by artificial intelligence and machine learning technology and is automated.

The best examples of predictive segmentation include “likelihood to purchase” and “likelihood to churn” segments. While “likelihood to purchase” intends to segment customers based on their future high probability of purchasing your products, “likelihood to churn” creates segments of users who are more prone to exiting your business.

A marketer enabled with predictive segments will be able to:

Leverage smart predictive models to predict each consumer’s likelihood to perform any action such as purchase, repeat purchase, churn, etc.

Send a set of personalized push notifications to engage each customer

Send push notifications to the right user at the right time, when they are likely to engage and when a business needs to automate customer journeys for each individual customer, along the predefined marketing funnel.

Deliver the best next product recommendations, content, or offer — to each customer

Predictive segmentation removes the need for all of the manual work by automatically identifying and analyzing the valuable or high-potential audience segments that need to be targeted. However, as an individual capability, it does not help in determining the right experience that should be tailored for each identified audience segment. In other words, predictive segmentation accelerates a marketer’s position in the workflow, enabling them with the right segments to work with, but tailoring the appropriate experiences remains guesswork that would have to be tested and validated over time with data.

**Chapter 5**

**Conclusion**

**Customer segmentation is the process of dividing a customer base into groups of individuals that are similar in certain ways relevant to marketing, such as age, gender, interests, and spending habits**. It enables companies to target specific groups with tailored promotions, products, or services that are most likely to resonate with them. Machine learning has become a popular tool for automating the process of customer segmentation, providing a more efficient and effective way to identify patterns and relationships within customer data.

There are several different methods for using machine learning to perform customer segmentation, including:-

**Clustering algorithms:** These algorithms divide customers into groups based on their characteristics and behaviour. For example, **k-means Clustering** can be used to find the k number of clusters in a dataset.

**Decision trees:** These algorithms use a tree-like model to identify the most important variables that influence customer behaviour. By using decision trees, companies can determine which customers are most likely to respond to certain marketing campaigns or products.

**Neural networks:** These algorithms can be used to model complex relationships between customers and their behaviour. Neural networks can identify patterns in customer data that are not easily recognizable through traditional methods.

**Association rule learning:** This method finds the relationships between customer attributes and behaviours, such as buying habits and product preferences. Association rule learning can help companies understand which products are frequently purchased together and target customers accordingly.

One of the key benefits of using machine learning for customer segmentation is its ability to process vast amounts of data in real time. This allows companies to quickly identify new trends and patterns in customer behaviour, allowing them to make more informed marketing decisions. Additionally, machine learning algorithms can continuously learn and improve over time, providing a more accurate picture of customer behaviour. This can be a time-consuming and error-prone process, particularly when working with large datasets. Machine learning algorithms can automate the process of data analysis, providing companies with more accurate and reliable results.

**REFERENCES**

Sari, J. N., Nugroho, L. E., Ferdiana, R., & Santosa, P. I. (2016). Review on Customer Segmentation Technique on Ecommerce. *Advanced Science Letters*, *22*(10), 3018–3022. <https://doi.org/10.1166/asl.2016.7985>

Kushwaha, Y., & Prajapati, D. (n.d.). *Customer Segmentation using K-Means Algorithm*. Retrieved May 11, 2022, from <https://ijcrt.org/papers/IJCRT_196650.pdf>

*Customer Segmentation Tutorial | Python Projects | K-Means Algorithm | Python Training | Edureka - YouTube*. (n.d.). Retrieved May 11, 2022, from <https://www.youtube.com/watch?v=4jv1pUrG0Zk>

*Customer Segmentation. Segmentation by RFM clustering | by Barış Karaman | Towards Data Science*. (n.d.). Retrieved May 11, 2022, from [https://towardsdatascience.com/data-driven-growth-with-python-part-2-customer-segmenta](https://towardsdatascience.com/data-driven-growth-with-python-part-2-customer-segmentation-5c019d150444) [tion-5c019d150444](https://towardsdatascience.com/data-driven-growth-with-python-part-2-customer-segmentation-5c019d150444)

Wu, J., Shi, L., Lin, W. P., Tsai, S. B., Li, Y., Yang, L., & Xu, G. (2020). An Empirical Study on Customer Segmentation by Purchase Behaviors Using a RFM Model and K -Means Algorithm. *Mathematical Problems in Engineering*, *2020*. <https://doi.org/10.1155/2020/8884227>