CUSTOMER SEGMENTATION

PROJECT REPORT ON "CUSTOMER SEGMENTATION"
CHETHAN B S

EXPOSYS DATA LABS

Abstract

Customer Relationship Management(CRM) has always played a crucial role as a market strategy for providing organizations with the quintessential business intelligence for building, managing and developing valuable long-term customer relationships. A number of business enterprises have come to realize the significance of CRM and the application of technical expertise to achieve competitive advantage. This study explores the importance of Customer Segmentation as a core function of CRM as well as the various models for segmenting customers using clustering techniques.

The available clustering models for customer segmentation, in general, and the major models of K-Means and Hierarchical Clustering. We use K-Means Clustering for this project.

Table of Contents

Introduction

Customer Segmentation is a method of analysing 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 behaviour in relation to a business. The selected criteria are used to create customer segments with similar values, needs and wants.

Customer segmentation is an essential tool in customer relationship management, enabling businesses to market effectively to their customers. Businesses are expected to understand their customers and demonstrate their customer insights by sending only relevant, targeted communications to their customers. Customers want to feel valued and be treated as individuals, yet for anything other than perhaps the smallest of businesses this level of customer knowledge is impossible to achieve.

Customer behaviour changes over time. Data analysis can be used to anticipate these changes in the customer lifecycle with predictive modelling. Therefore ongoing customer data gathering and analysis is essential to keep segmentation up to date and communications relevant.

Existing Method

__Demographic segmentation__ divides the market into smaller categories based on demographic factors, such as age, gender and income. Instead of reaching an entire market, a brand uses this method to focus resources into a defined group within that market.

Dividing the market into smaller segments, each with a common variable, allows companies to use their time and resources more efficiently. They can better understand the prospective market, and use advertising personalization to ensure the needs of the targeted group are fulfilled.

This customer segmentation method is one of the most commonly used because it's easy to acquire through census data, analytics software, consumer insights, and more. It's also considered by many businesses to be the cheapest way to divide a target market.

Why is demographic segmentation in marketing so important?

Companies reduce the risk of running campaigns to uninterested consumers, which quickly increases ROI. In fact, email marketers have witnessed a 760% increase in revenue by segmenting their email campaigns. Conversely, 85% of new product launches in the US fail to generate desired revenue due to poor segmentation.

In addition to better ROI, demographic segmentation allows you to:

Build long-lasting customer relationships:

Reaching your customers on a more human level with demographic-based personalized marketing creates deeper customer loyalty. It allows them to identify with your brand and feel like you are an advocate for their needs, which makes them more likely to do business with you over longer periods.

Improve your products and services:

Having loyal customer relationships encourages you to look at your products and services in a new way. When you have a deeper understanding of your target

audience, you can put yourself in their shoes to better serve them. If you develop fitness programs and know that a majority of your clients prefer the same type of program to be released, you're more likely to make that happen.

Optimize your marketing strategies:

Demographic segmentation allows you to get more specific with your marketing strategies. It helps clarify your vision, have more direction with future advertising plans, and optimize your resources, time, and budget. If 85% of your clients range from 20-35 years old, this is the segment you're going to target. Spending your time and money on seniors would be a waste.

Proposed Method with Architecture

Here, **Demographic Segmentation** method is used since the data set contains the demographic variables such as Age, Gender, Annual Income and Spending Score.
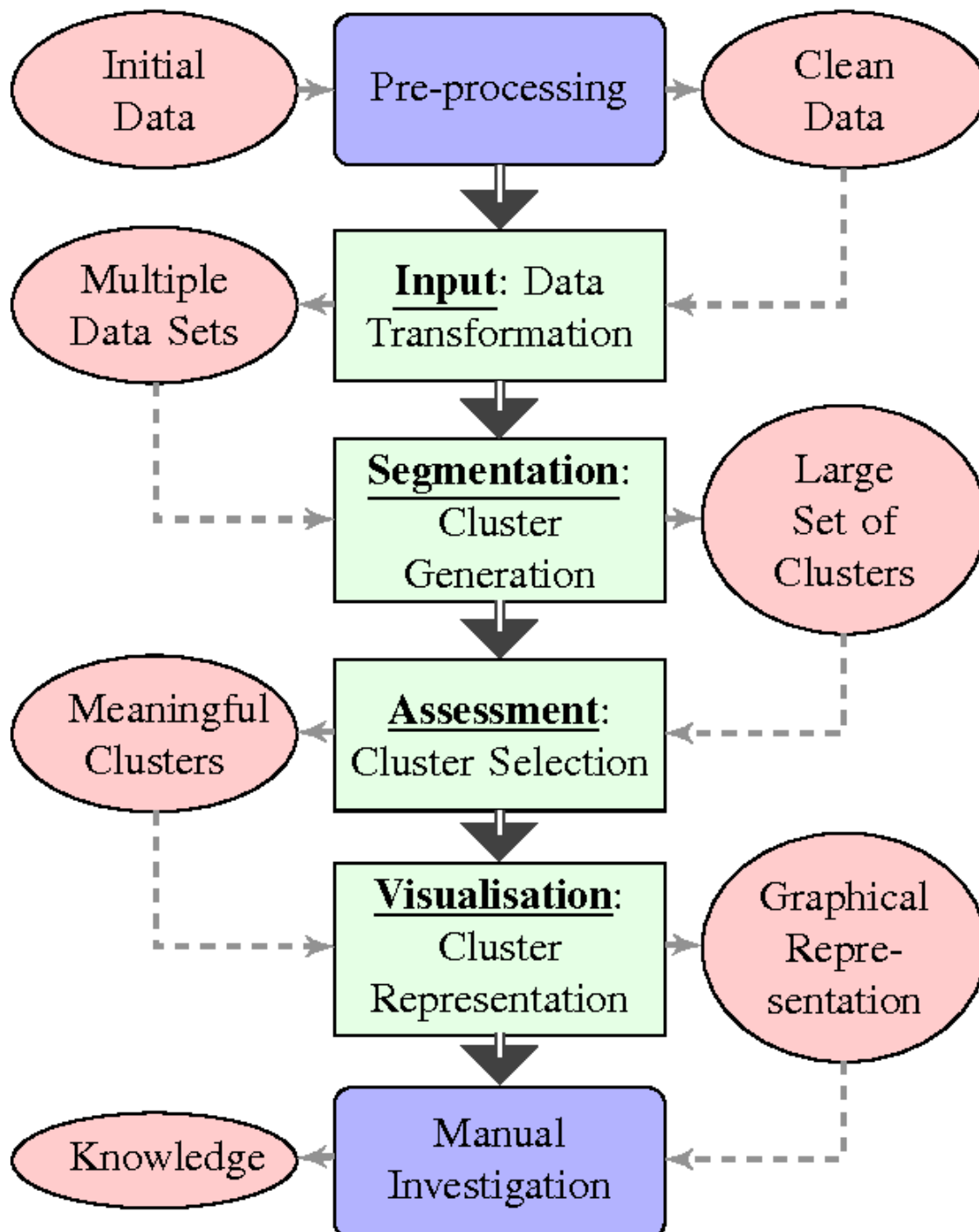
Methodology

How to get demographic data

You may not be able to get demographic information for every one of your customers or audience members. That's alright. You can get the data from enough people to build a profile of your most engaged customers. With that, you can choose the right social media platforms, create better marketing material, and grow your subscriber base more quickly.

There are a number of ways to unlock this information. We'll touch on two methods.

Analytics software

If you have a website then you're using analytics software to understand who visits and what they're interested in. Most analytics software allows you to see basic demographic information like gender and age.

Log into your analytics tool and check if it's gathering the data. If not, you can use a tool like Google Analytics (GA) for this.

Demographic segmentation variables

1. Age

Age is the most basic variable of them all, albeit the most important because consumer preferences continually change with age. Almost all marketing campaigns target age-specific audiences.

This variable can be viewed regarding specific age ranges or life cycle stages: babies, children, adolescents, adults, middle-age, and seniors. For example, many famous fashion designers have different collections to target different age groups. They aim certain clothing lines at specific age ranges, such as a chic fashion line at younger prospects, and a more formal and elegant line at older individuals.

Age segmentation is also generation-based: baby boomers, gen X, millennials, etc. Since members within each of these individual groups were born around the

same time and grew up with similar experiences, they often share similar characteristics and thought processes. Targeting baby boomers and gen X with the same offer and marketing strategy is likely to produce undesirable results because they think and act differently.

Not only do age groups and generations differ in their buying habits, but also in how they respond to advertising. They tend to have distinct ways of speaking and often spend their time on separate platforms. For example, millennials may spend most of their time on Instagram and Facebook, while seniors prefer their email inboxes.

2. Gender

Men and women generally have different likes, dislikes, needs, and thought processes. For instance, few men apply makeup, and most women don't wear boxers. Also, women typically do most of the household grocery shopping and are more likely than men to donate to charitable causes. These are all key factors to consider when creating a campaign.

They created this ad specifically for women (hence the women's swimsuits and the "for women" in the description) and purposely targeted them on Facebook for the most engagement and click-throughs.

Be careful not to assume gender stereotypes, such as considering pink a feminine color and blue a masculine color. Advertising with gender stereotypes like this could easily make your brand look sexist and cause you to miss out on or anger your target audience.

3. Income and occupation

If people can't afford your product or service, there is no point in targeting them. After all, you wouldn't promote a Mercedes or Ferrari to someone who can't afford a used vehicle with more than 100,000 miles.

Income targeting lets you measure the buying power of your audience. When you know the income range of consumers, you can usually find data to support how people spend money on both the higher and lower end of the spectrum. Many companies use this data to sell different tiers of the same product, based on income level. For instance, airlines have three classes: economy, business class, and first class.

4. Ethnicity and religion

With the tremendous increase in international business and global advertising, brings an increase in segmentation based on ethnicity, race, nationality, and religion. These groups have many individual cultures that come with conflicting interests, preferences, attitudes, and beliefs. This could impact both their response to marketing and their buying habits.

Consider Coca-Cola and Pepsi. Both companies advertise globally, but localize their campaigns for each country, too. The messages are completely different, based on local customs, religions, nationality, etc.

5. Family structure

Family makeup can be extremely useful in segmentation because when a family's dynamic changes, its needs and desires often do too. This strongly affects their buying habits and your sales process.

Single individuals tend to prioritize themselves, while newly married couples are likely prioritizing each other and their homes. Couples with several children have different needs than those who just had their first child. Large families might be more interested in low-cost household products, as compared to a couple with the same income, but without any children.

Implementation

Customer Segmentation

Importing necessary libraries

```
In [2]: import pandas as pd
        import numpy as np
```

Reading Dataset

```
In [3]: df = pd.read_csv("Mall_Customers.csv")
        df.head()
```

Out[3]:

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

Checking for null values

```
In [6]: df['Gender'].isna().sum()
```
Out[6]: 0

```
In [7]: df['Age'].isna().sum()
```
Out[7]: 0

```
In [9]: df['Annual Income (k$)'].isna().sum()
```
Out[9]: 0

```
In [12]: df.isna().sum()
```
```
Out[12]: CustomerID              0
         Gender                  0
         Age                     0
         Annual Income (k$)      0
         Spending Score (1-100)  0
         dtype: int64
```

```
In [13]: df['Gender'].unique()
```
Out[13]: array(['Male', 'Female'], dtype=object)

```
In [14]: df['Age'].unique()
```
```
Out[14]: array([19, 21, 20, 23, 31, 22, 35, 64, 30, 67, 58, 24, 37, 52, 25, 46, 54,
                29, 45, 40, 60, 53, 18, 49, 42, 36, 65, 48, 50, 27, 33, 59, 47, 51,
                69, 70, 63, 43, 68, 32, 26, 57, 38, 55, 34, 66, 39, 44, 28, 56, 41],
               dtype=int64)
```

Dropping 'Customer ID' column

```
In [17]: df.drop(columns={'CustomerID'}, axis=1, inplace=True)
```

```
In [18]: df.head()
```

Out[18]:

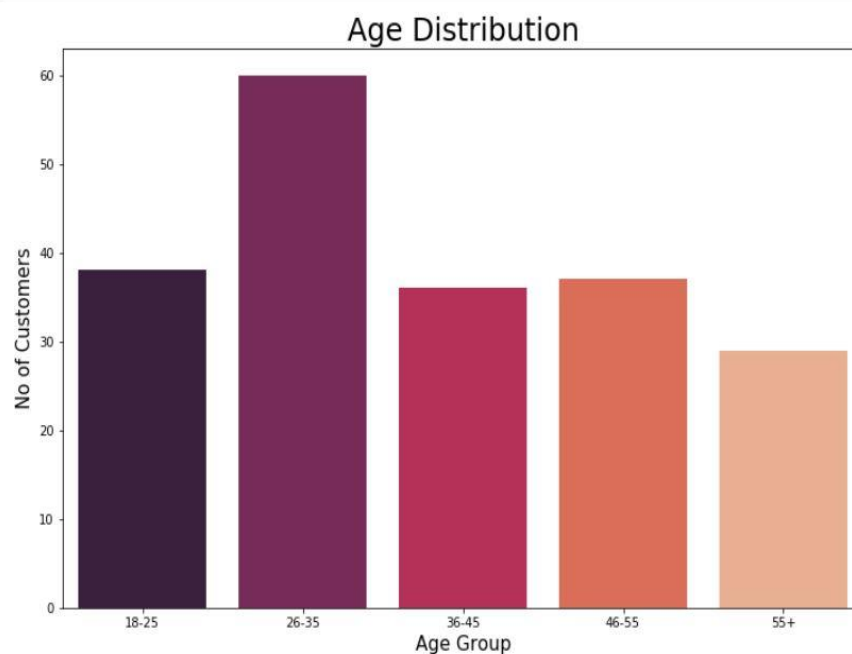|	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	Male	19	15	39
1	Male	21	15	81
2	Female	20	16	6
3	Female	23	16	77
4	Female	31	17	40

```
In [26]: import matplotlib.pyplot as plt
         import seaborn as sns
         %matplotlib inline
```

```
In [33]: plt.figure(figsize=(12,8))
         x = ['Male','Female']
         y = [88,112]
         plt.bar(x,y,color='r')
         plt.xlabel('Gender',fontsize=15)
         plt.ylabel('Count',fontsize=15)
         plt.title('Gender Distribution', fontsize=24)
         plt.show()
```

Gender Distribution
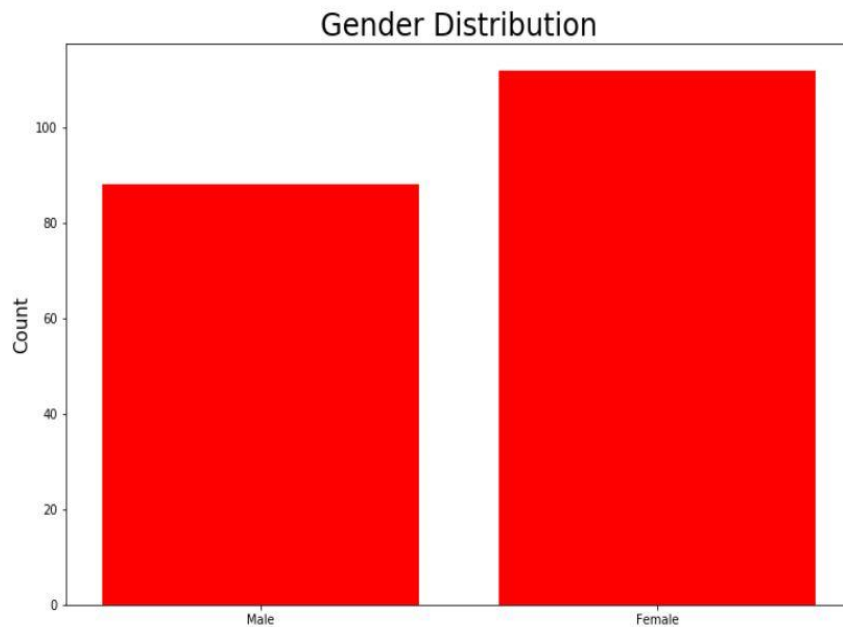
```
In [50]: x = ['18-25','26-35','36-45','46-55','55+']
         y = [len(age18_25),len(age26_35),len(age36_45),len(age46_55),len(age56_)]

         plt.figure(figsize=(12,8))
         sns.barplot(x=x,y=y,palette='rocket')
         plt.xlabel('Age Group', fontsize=15)
         plt.ylabel('No of Customers', fontsize=15)
         plt.title("Age Distribution", fontsize=24)
         plt.show()
```
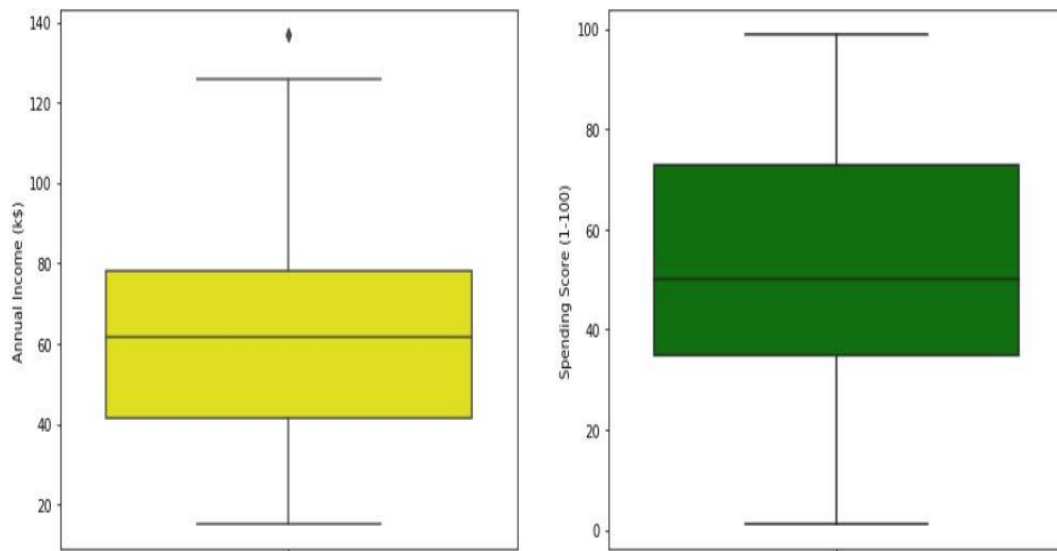
Age Distribution

Annual Income and Spending Score

```
In [38]: plt.figure(figsize=(15,6))
         plt.subplot(1,2,1)
         sns.boxplot(y=df['Annual Income (k$)'],color='yellow')

         plt.subplot(1,2,2)
         sns.boxplot(y=df['Spending Score (1-100)'],color='green')
         plt.show()
```



```
In [56]: x = ['1-20','21-40','41-60','61-80','81-100']
         y = [len(ss1_20),len(ss21_40),len(ss41_60),len(ss61_80),len(ss81_100)]

         plt.figure(figsize=(15,8))
         sns.barplot(x=x,y=y,palette='nipy_spectral_r')
         plt.xlabel('Spending Score range',fontsize=15)
         plt.ylabel('Count',fontsize=15)
         plt.title('Spending Score Distribution',fontsize=24)
         plt.show()
```
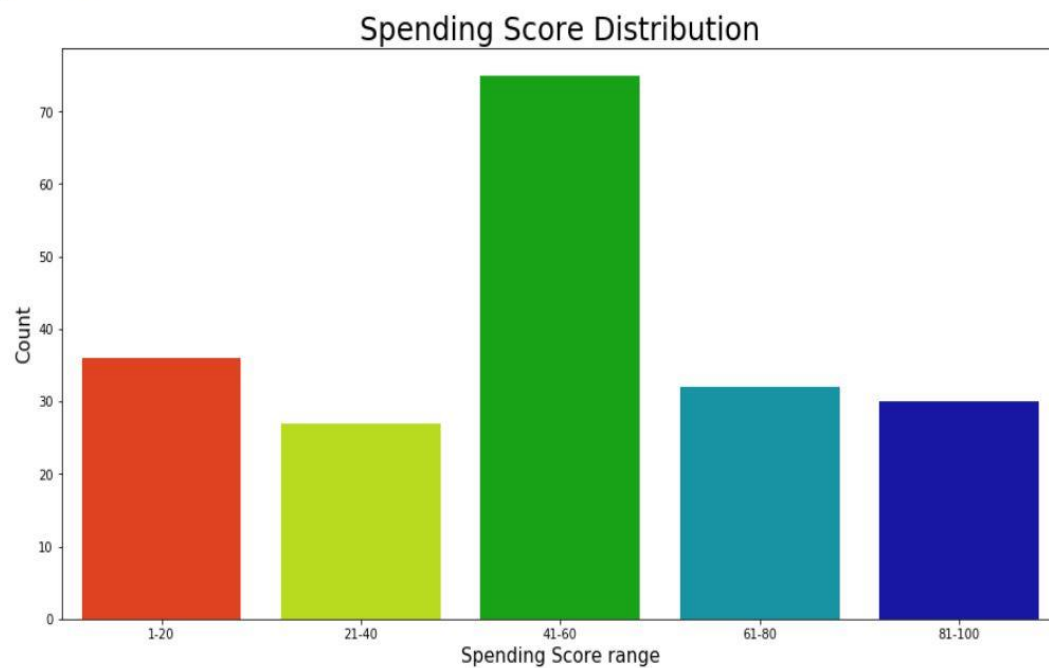
```
In [58]: from sklearn.cluster import KMeans
         wcss = []
         for k in range(1,10):
             kmeans = KMeans(n_clusters=k,init='k-means++')
             kmeans.fit(df.iloc[:,1:])
             wcss.append(kmeans.inertia_)
```

```
In [60]: plt.figure(figsize=(10,8))
         plt.plot(range(1,10),wcss,color='blue',marker="8")
         plt.xlabel('clusters',fontsize=15)
         plt.ylabel('WCSS',fontsize=15)
         plt.grid()
         plt.title('Elbow Curve',fontsize=30)
         plt.show()
```
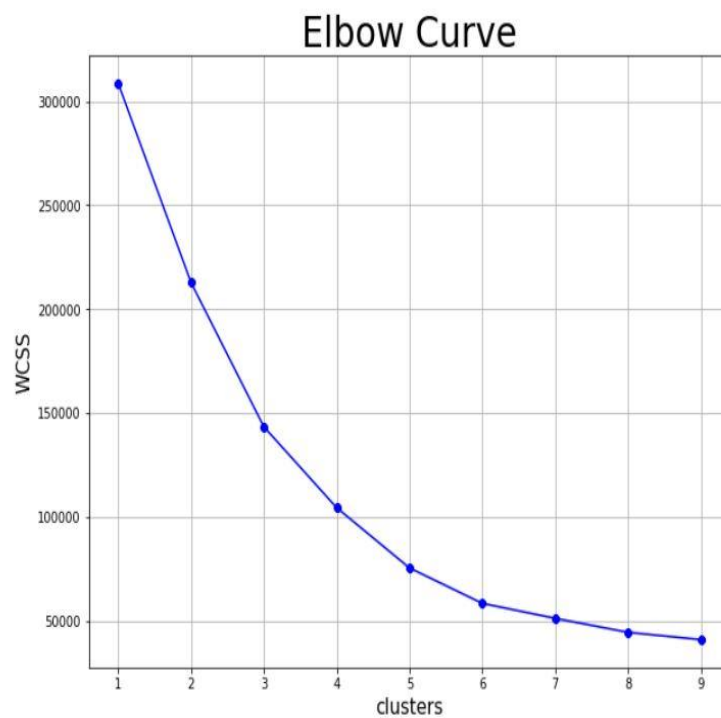
```
plt.title('3D Plot',fontsize=20)
plt.xlabel('Age',fontsize=15)
plt.ylabel('Annual Income',fontsize=15)
ax.set_zlabel('Spending Score',fontsize=15)
plt.show()
```


3D Plot
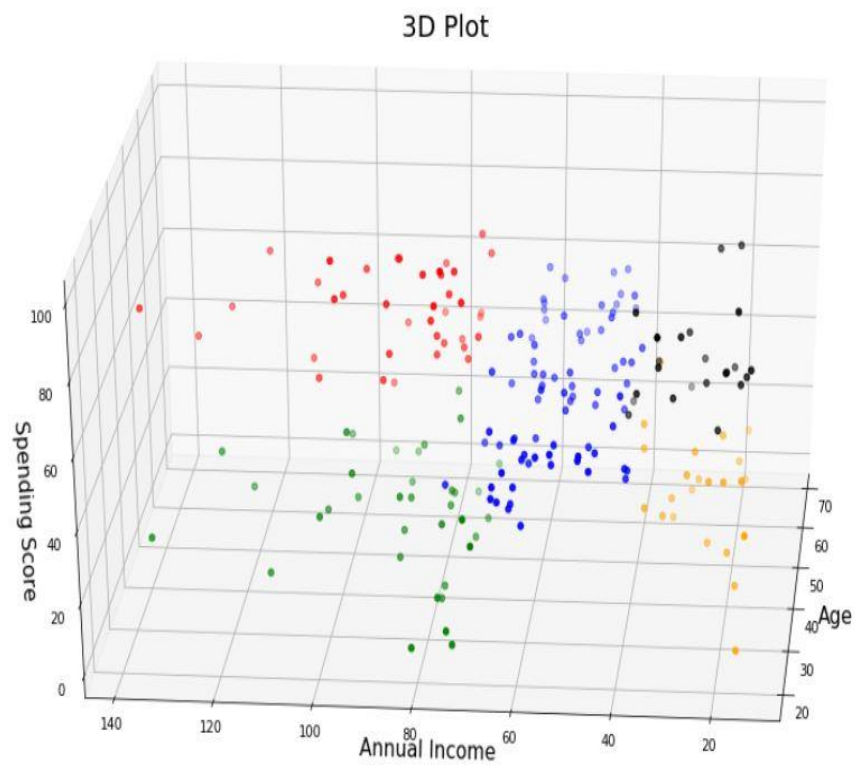
Conclusion

The first and most objective of our study is comparative study of customers buying behaviour in retail mall. The study of consumer behaviour is the most important factor for marketing of any goods and services.

The consumer behaviour suggests how individual, groups, organization select, buy, use and dispose of goods, services and ideas or experiences to satisfy their needs and wants. It also clues for improving or introducing products or services, setting price, divising channels etc.

The marketers always look for emergent trends that suggest new marketing opportunities and here in India a lot of opportunities are available.