Topic : Shopping Mall Clustering Analysis

Team Members:

Yerramsetty Sai Naga Sabarish- 20BCE2370

Shiva Pranav- 20BCE2507

Sai Kiran Balaga- 20BCE0023

Farhan Abdul Khader Kooliyattayil- 20BCE2530

Course Code: CSE3021

Social and Information Network

Submitted to Professor: **Dr. Ilanthenral Kandasam**

1) Introduction

India is a market that is growing in the global economy. Since progress, the country has improved at an exceptional rate. The nation has experienced improvements in the foundation, capital business sectors, banking, affirmation, and other fields thanks to modifications in every area that truly important. This breakthrough has made it possible for the nation's retail business to expand to one more region. The expansion of business levels brought by industry advancement has improved the discretionary spend of the typical customer. Consistently asks the affiliate the board to consider its systems, destinations, and friendliness. Energy shifts have the ability to quickly render the prior successful company planning guidelines obsolete. What fresh challenges are there? With increased general competition, condition deterioration, foundation neglect, poor job skills, and a variety of other financial, political, and social difficulties, associations are feeling the pinch. In actuality, there are problems, but there are also opportunities. The globalization show off hypothesises that associations may rely on a much larger market potential for their goods and initiatives. Environmental degradation offers associations endless freedoms to come up with more practical methods for cleaning up the planet.The retail sector is happy about progress being made in closely related fields, whether it be in strip malls, supermarkets, or single-brand abundance shops. These businesses have noticed the metropolises' thriving business environments and have even flocked to the Tier II and Tier III metropolitan areas. The nuclear family corporate goliaths like Reliance, KK Modi, Aditya Birla social event, and Bharti pack are also going through a period of retail development, in addition to the general players like Wal-Mart, TESCO, and Metro group. Reliance said that it will invest $3.4 billion to establish a network of 1,575 outlets and rise to the position of the nation's foremost retailer. The expansion of Uber shopping malls in India is changing opinions about the impacted retail market.

Buying habits in the country of merchants are rapidly evolving and changing. Shopping strips are quickly becoming sought-after destinations after being redirected. The greatest opportunity exists in the sale of food and groceries because almost all of it is now segregated.

Reliance, the Aditya Birla Group, and the Bharti Group, which has partnered with Walmart, the most renowned retailer in the world, are just a few of the large companies joining the filtered food retail market. This group of novice players is expected to have a 25–30% annual improvement rate.

2) Literature Review Summary Table

Authors	Title (Study)	Concept	Implementation	Dataset Analysis	Relevant Finding	Limitations
Mohammed Ismail El-Adly	Shopping Malls Attractiveness: A Segmentation Approach	Social The purpose of the paper is to identify the shopping malls in the UAE that are most appealing to consumers, and then to divide consumers into groups based on these qualities..	Due to the discovered criteria, a segmentation strategy employing K-means cluster analysis was also applied to segment mall visitors.	The variables that make a shopping mall attractive were determined using a poll of university employees and principal component factor analysis.	Influencer's number of followers does not imply the influencer's spreading capability or whether it is the most central and significantmember of the network. Centrality measures on their own are not enough to identify influencers	This study is limited in that it surveyed UAE University staff as shoppers. Thus, findings may not be representative of UAE shoppers in general.

Devinder Pal Singh	Integration of materialism with shopping motivations: motivations based profile of Indian mall shoppers	It describes accounts' performance of the top, most visited museums worldwide and next investigates their interconnection Recording and analyzing performance characteristics	In addition to utilitarian motivations, the research seeks to investigate materialism as one of the retail shopping motives in the Indian setting. It seeks to pinpoint the primary driving forces behind why people shop and how they justify the benefit of mall shopping. Additionally, it aims to characterise the motivational and demographic traits of the identified segments and create a typology of Indian mall shoppers based on their shopping motivations.	The data were collected through a mall intercept survey	Shoppers at Indian malls are driven by hedonistic, materialistic, and utilitarian goals. In general, they can be divided into four groups: balanced consumers, materialist shoppers, hedonistic shoppers, and value shoppers.-Information reaches museums of the same type or the same location more quickly	Understanding unique customer categories can enable marketers in developing marketing and promotional tactics to increase mall foot traffic. Understanding unique customer categories can enable marketers in developing marketing and promotional tactics to increase mall foot traffic.

Diana Simona Damian, José Dias Curto José Castro Pinto	The impact of anchor stores on the performance of shopping centres: the case of Sonae Sierra .	This article examines how anchor shops affect shopping centre performance and other retailer costs. It examines anchor store customer spillover on Sonae Sierra retail centres. Sonae Sierra is an international retail centre company founded in Portugal in 1989. Sonae (Portugal) and Grosvenor (UK) hold 50% each.	The anchor retailers provide roughly 41% of the overall gross lettable space, while they typically only contribute 18% of the developer's total rent revenue. The hypotheses are tested using ordinary least squares and the Kruskal-Wallis statistic (to prevent ANOVA assumption violations).	The data collection targeted 35 shopping centres in Portugal and Spain with 1,200,000 square feet (or more), for three consecutive years (2005-2007)	The empirical research reveals that anchors enhance mall sales and rentals for non-anchor retailers. The authors show that externalities are internalised by efficient allocation of space and incentives across stores, and that anchor stores increased the malls' customer drawing power, measured as the number of people who visited the mall at a given time, although lately they have had less impact on sales per person.	This research only examined Sonae Sierra retail malls, hence it can only be extrapolated using this model. The impossibility to collect data on consumer buying power in the regions around Sonae Sierra retail malls and the requirement to protect data confidentially prevented the incorporation of extra independent variables in the models.

Chae Mi Lim, Rodney Runyan Youn-Kyung Kim	Segmenting luxe-bargain shoppers using a fuzzy clustering method	This research uses a fuzzy clustering approach based on psychographic characteristics linked to luxury consumption and bargain processes to identify consumer groups among luxe-bargain buyers and profile their behavioural inclinations.	Constructing a bipartite graph that represented the relationship between visitors and the Louvre Museum Calculated centrality measures Degree centrality, betweenness centrality and eigenvector centrality	The sample consists of 500 consumers who purchased a luxury brand at a bargain	The studies consist of a confirmatory factor analysis, a fuzzy clustering analysis based on psychographic variables, and an analysis of variance (ANOVA) for profiling the segments.	Due to the sample technique, the product category of luxury brands, and the restricted number of luxury brands that were employed in the research, the results of the present study should be considered with care.
Alison Elizabeth Lloyd Ricky Y.K. Chan Leslie S.C. Yip Andrew Chan	Time buying and time saving: effects on service convenience and the shopping experience at the mall	The field of user-friendliness in service provision is yet only partially researched. The value that is put on time, however, is likely to impact the significance that is placed on convenience. This is because time permeates all elements of consuming. Previous studies have called for an assessment of convenience that goes beyond the setting of stores; shopping malls, which are a one-stop shopping destination, provide an excellent location for investigations like this one.	A conceptual model delineating the relationships between service convenience, shopping trip value, customer satisfaction and several retail outcomes is developed.	This model is empirically tested using survey data collected from 619 mall shoppers; and analyzed using structural equation modelling.	Customers who place a high economic value on their time will experience a greater increase in hedonic value as a direct result of increased service convenience. On the other hand, the impacts of service convenience are stronger on the utilitarian value for those individuals who put a low economic value on time. The effects that shopping value has on retail results may likewise be rather variable depending on the value that is put on time.	It is recommended to conduct more research making use of the entire convenience scale as well as more intricate time style measurements.

Dr. Shailesh K. Kaushal Ram Komal Prasad	Segmentation of Shopping Mall Shoppers: A Cluster Analysis Approach	This paper attempts to identify various segments of shopping mall shopper viz., Quality conscious, Price conscious and Recreational conscious.	This research study has been undertaken with the aim of identifying prominent segments for shopping mall shopper. Both primary and secondary data have been used in the study. Primary data have been collected from Lucknow, using well-structured questionnaire. Total 157 customers were selected from this city for the purpose of data collection using judgment sampling. 24 preselected statements were used to find out the prominent segment. All of these were five point Likert-type scales in which respondents were asked to indicate their level of agreement (1 = strongly disagree to 5 = strongly agree). Cluster analysis using SPSS software has been applied to determine the significant segments	Market segmentation has been accepted as a strategic marketing tool for defining markets and thereby for allocating resources. The objectives of this research were to investigate the consumer behaviour segmentation in shopping mall and to study variation in the segmentation of shopping mall shoppers across different demographic variable. This research study found three main consumers segments such as quality conscious, price conscious and recreational conscious buyer. In addition, this study shows that average Indian shoppers in our sample were not very brand conscious, but quite quality, price and recreational conscious.	Information on consumer buying behaviour segmentation will be useful for retailing targeting Indian markets. As Indian retail industries is in boom today and more and more national and international players would be interested in the emerging retail market in India, an understanding of Indian mall shopper behaviour, with particular reference to their segmentation is crucial. Profiling consumers by combining segmentation of shopping mall shoppers and demographic variables provide more meaningful ways to identify and understanding various consumers segments and target each segment with more focused marketing strategies	There are several limitations that warrant future research. This study has been conducted in Lucknow city of Uttar Pradesh, India. The results of the same, if conducted in other part of the country may vary. This difference is too significant to be ignored. The sample size consisted of 157 active mall shoppers. The small sample size is also error prone.

K. Archana Dr.K.G.Saranya	Mall Customer Segmentation Using Clustering Algorithm	It is a web application for segmentation of mall customers which is useful for the shopkeepers to market the product based on the planned strategy. The cluster which is generated by the application is stored in the image format.	Clustering is one of the most common methods used in exploring data to obtain an clear understanding of the data structure. It can be characterized as the task of finding the subgroups in the complete dataset. Similar data is clustered in the same subgroup. A cluster refers to a collection of aggregated data points due to some similarities. Clustering is used in Market basket analysis used to segment the customers based on their behaviors. K Means Clustering is the most common and simplest Machine learning algorithm and it follows an iterative approach which attempts to partition the dataset into different “k” number of predefined and non overlapping subgroups where each data point belongs to only one subgroup.	Cluster 1 represents earning high but spending less Cluster 2 represents average in terms of earning and spending Cluster 3 represents earning high and also spending high [Target customers] Cluster 4 represents earning less but spending more Cluster 5 represents earning less, spending less Stimulus: The marketing team presses the K-Means execution button. Response: The results are generated in the form of a graph and stored as image. With clusters K=5.	Consumer data is increasing exponentially due to with this large data magnitude.	Such clustering models need to have the ability to effectively process this huge information.

Hemashree Kilari Sailesh Edara Guna Ratna Sai Yarra Dileep Varma Gadhiraju	Customer Segmentation using K-Means Clustering	The average silhouette coefficient across all dataset occurrences is used to calculate the silhouette score. The silhouette coefficient, which ranges from -1 to 1, measures how close points in one cluster are to points in nearby clusters.	A shopping center store provided the dataset for clustering using the K-means algorithm. Five attributes and 200 tuples make up the data set, which represents the information of 200 consumers. The characteristics in the data collection are CustomerId, gender, age, yearly income (k$), and spending score on a scale of (1-100).	If a dataset contains null values, duplicates, or other noisy data, data cleaning must be performed. Data cleansing ensures that information is reliable, usable, and available for analysis. When we have the data, we may visualize it by comparing the annual income and spending score, which is gender-specific. According to the study, there are five different types of plots that illustrate groups of customers who engage in the following activities, as well as customer behaviors linked to yearly income and expenditure scores: 1.Score of High Income/Low Spending 2.Low Income- A high score for spending 3.A high score for spending-despite Low Income 4.Average Income-Average Spending Score 5.High Income-High Spending Score.	By analyzing the data, we can predict customer behavior based on their Annual Income and Spending Score. This cluster analysis may be applied to a number of consumer marketing methods. We'd want to keep our target clientele, who have a high income and a high spending score, because they deliver the biggest profit margin. Because of their lifestyle demands for a high income and low spending score, customers will be lured to the Mall Supermarket because of the great variety of things available. Less Income Less Spending Scores can obtain more promotions, and they will be tempted to spend by receiving offers and discounts on a frequent basis. A cluster analysis may be used to establish what kind of things clients wish to consume, allowing for the development of more targeted marketing efforts. The people in clusters 3 and 4 are the potential	Mall managers must be able to understand what customers require and, more importantly, how to meet those needs. analyze their purchasing habits, and establish frequent encounters with customers that make them feel comfortable in order to satisfy their demands.

					clients in this situation.	
M. Thirunavakarasu Kuncham Pavan Kumar Reddy G. Srinivasa Teja	Customer Segmentation In Shopping Mall Using Clustering In Machine Learning	The Data Science Methodology aims to answer basic questions in a predescribed sequence, that cover the five main aspects of data science projects. These aspects are: ● Problem to Approach ● Requirements to Collection ● Understanding to Preparation ● Modelling to Evaluation ● Deployment to Feedback	Machine learning approaches are an incredible instrument for dissecting customer information and tracking down bits of knowledge and examples. Misleadingly wise models are useful assets for chiefs. They can exactly recognize client fragments, which is a lot harder to do physically or with ordinary logical techniques. There are many machine learning algorithms, each reasonable for a particular sort of issue. One extremely normal AI calculation that is appropriate for client division issues is the k-means clustering algorithm.	Presently we will prepare the model on the preparation dataset and make expectations for the test dataset. However, would we be able to approve these forecasts? One approach to doing this is we can partition our train dataset into two sections: train and approval. We can prepare the model on this preparing part and involving that make expectations for the approval part.	By the customer segmentation method the project is evaluated successfully. The accessible informative elements in the data set is thing perspectives, preferences and discussions, but the underlying rendition of the framework assessed in this report just utilizes sees.	During the final phases of the execution a small variant issues were risen, but no critical enhancements were noticed. All things considered, the spotlight during this task was on the bunching investigation, the preprocessing phase of this undertaking could be improved by joining preferences and discussion to the appraisals computations utilizing some weighting of these three elements.

Asith Ishantha	Mall Customer Segmentation Using Clustering Algorithm	We are going to aim to cluster a data set that is about behaviour of the customers having credit card using many unsupervised algorithms.	Mini Batch K-means algorithm ‘s main idea is to use small random batches of data of a fixed size, so they can be stored in memory. Each iteration a new random sample from the dataset is obtained and used to update the clusters and this is repeated until convergence. Each mini batch updates the clusters using a convex combination of the values of the prototypes and the data, applying a learning rate that decreases with the number of iterations. This learning rate is the inverse of the number of data assigned to a cluster during the process. As the number of iterations increases, the effect of new data is reduced, so convergence can be detected when no changes in the clusters occur in several consecutive iterations.	The data set has 8950 transactions or information about account that belong to customers.	Cluster 0: Second highest purchase frequency which tend to pay in instalment, that is lower credit limit and long duration customers. Cluster 1: Pretty low balance and purchase frequency. They rarely use credit card and also, they have lower credit limit. Cluster 2: This group is having the highest number of customers and lowest usage of cards. Inactive customers, also long duration customers. Cluster 3: High tendency of payment instalment, higher purchase frequency and their tenure time is above average. Cluster 4: The highest balance amount but purchase frequency is not that good. Tend to cash in advance, higher credit limit than others. They don’t like spending money. Cluster 5: Second highest purchase frequency and also higher tendency	They use unsupervised Learning Algorithms which is intrinsically more difficult than supervised learning as it does not have corresponding output. The result of the unsupervised learning algorithm might be less accurate as input data is not labeled, and algorithms do not know the exact output in advance.

					payment in instalment. They are long duration customers.	

3) Objective of the project:

The objective of this project is to determine how committed and satisfied customers are with the retail centres. The experiment sought to determine whether there were any notable differences in the approaches used for customers of shopping centres when taking into account their age groups and sexual orientation.

4) Innovation component in the project:

Incorporate new data sources: While traditional shopping mall clustering analysis may focus on factors such as store types and demographics, there may be other data sources that could be incorporated to provide new insights. For example, social media data could be analyzed to understand how people are interacting with different malls or retailers, or data on foot traffic patterns could be used to identify which malls are most popular at different times of day or days of the week.One emerging trend in the shopping mall industry is a focus on sustainability and environmental responsibility. A clustering analysis project could incorporate factors related to sustainability, such as the use of renewable energy or sustainable building materials, to identify which malls are leading the way in this area.Traditional clustering analysis is typically based on static data sets. However, by incorporating real-time data from sources such as sensors or mobile devices, it may be possible to develop more dynamic and responsive insights about mall performance and customer behavior.

5) Work done and implementation

a) Methodology adapted:

K Means:
K-means clustering is a powerful technique for analyzing shopping malls and identifying patterns in customer behavior and preferences. By using this technique, shopping mall owners and marketers can gain valuable insights that can help improve customer engagement, increase sales, and enhance the overall shopping experience.
Affinity propagation:

Compared to K-means clustering, affinity propagation has some advantages in that it doesn't require the number of clusters to be specified in advance and can handle datasets with more complex structures. However, it can be more computationally expensive and may require more tuning of parameters. The choice of algorithm will depend on the specific goals and characteristics of the shopping mall dataset being analyzed

Hierarchical clustering

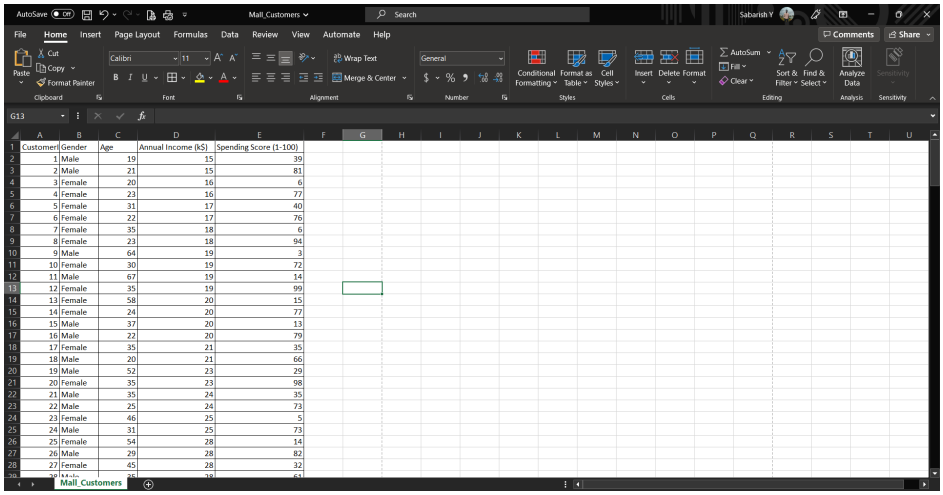
It is a type of clustering algorithm that groups data points into nested clusters using a hierarchical approach.
Compared to K-means clustering and affinity propagation, hierarchical clustering can be more interpretable, as it generates a dendrogram that visually shows the hierarchical relationships between clusters. It can also handle non-spherical cluster shapes and doesn't require the number of clusters to be specified in advance. However, it can be computationally expensive and may require more tuning of parameters. The choice of algorithm will depend on the specific goals and characteristics of the shopping mall dataset being analyzed.

Db scan (Density-Based Spatial Clustering of Applications with Noise)
Compared to K-means clustering, affinity propagation, and hierarchical clustering, DBSCAN can handle datasets with arbitrary shapes and doesn't require the number of clusters to be specified in advance. However, it can be sensitive to parameter selection and may not work well for datasets with varying densities. The choice of algorithm will depend on the specific goals and characteristics of the shopping mall dataset being analyzed.

Silhouette:
Silhouette analysis is a technique that can be used to evaluate the quality of clustering results in shopping mall clustering analysis. It provides a measure of how well each data point fits into its assigned cluster, and can be used to identify the optimal number of clusters.

Silhouette analysis can be a useful technique for evaluating clustering results in shopping mall clustering analysis. It provides a quantitative measure of the quality of the clustering results and can help to identify the optimal number of clusters. However, it is important to note that silhouette analysis should be used in conjunction with other techniques for evaluating clustering results, such as visualization and domain expertise.

b) Dataset or Data collection



Customer	Gender	Age	Annual Income (k$)	Spending Score (1-100)
1	Male	19	15	39
2	Male	21	15	81
3	Female	20	16	6
4	Female	23	16	77
5	Female	31	17	40
6	Female	22	17	76
7	Female	35	18	6
8	Female	23	18	94
9	Male	64	19	3
10	Female	30	19	72
11	Male	67	19	14
12	Female	35	19	99
13	Female	58	20	15
14	Female	24	20	77
15	Male	37	20	13
16	Male	22	20	79
17	Female	35	21	35
18	Male	20	21	66
19	Male	52	23	29
20	Female	35	23	98
21	Male	35	24	35
22	Male	25	24	73
23	Female	46	25	5
24	Male	31	25	73
25	Female	54	28	14
26	Male	29	28	82
27	Female	45	28	32

c) Tools Used:

Python modules:

- CSV (dataset)
- Pandas
- Google colab

d) Code:

```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import io

df2 = pd.read_csv('Mall_Customers.csv')

df2.head()
```

```
df2.shape


df2.info()


df2.describe()


df2.isnull().values.any()


labels = ['Female', 'Male']


size = df2['Gender'].value_counts()


colors = ['lightgreen', 'orange']


explode = [0, 0.1]


plt.rcParams['figure.figsize'] = (9, 9)


plt.pie(size, colors = colors, explode = explode, labels =
labels, shadow = True, autopct = '%.2f%%')


plt.title('Gender', fontsize = 20)


plt.axis('off')


plt.legend()


plt.show()


plt.figure(figsize=(25,10))


# Passing X axis and Y axis along with subplot position
```

```
plt.title('Age vs Annual Income', fontsize = 20)

plt.xticks(rotation=90)

sns.barplot(x = df2['Age'] , y = df2['Annual Income (k$)'] ,
palette='icefire');

plt.rcParams['figure.figsize'] = (25, 8)

sns.countplot(df2['Age'], palette = 'hsv')

plt.title('Distribution of Age', fontsize = 20)

plt.show()

sns.pairplot(df2)

plt.rcParams['figure.figsize'] = (25, 8)

plt.title('Pairplot for the Data', fontsize = 20)

plt.show()

plt.rcParams['figure.figsize'] = (15, 8)

sns.heatmap(df2.corr(), cmap = 'Wistia', annot = True)

plt.title('Heatmap for the Data', fontsize = 20)

plt.show()

ax = sns.barplot(y= "Spending Score (1-100)", x = "Age", data
= df2, palette=("Blues_d"))
```

```
sns.set(rc={'figure.figsize':(27.7,6.30)})

sns.set_context("poster")

plt.rcParams['figure.figsize'] = (35, 14)

sns.countplot(df2['Spending Score (1-100)'], palette =
'magma')

plt.title('Distribution of Spending Score', fontsize = 20)

plt.show()

plt.rcParams['figure.figsize'] = (18, 7)

sns.boxenplot(df2['Gender'], df2['Spending Score (1-100)'],
palette = 'Accent_r')

plt.title('Gender vs Spending Score', fontsize = 20)

plt.show()

plt.rcParams['figure.figsize'] = (18, 7)

sns.violinplot(df2['Gender'], df2['Annual Income (k$)'],
palette = 'gnuplot')

plt.title('Gender vs Spending Score', fontsize = 20)

plt.show()

plt.rcParams['figure.figsize'] = (35, 14)
```

```
sns.countplot(df2['Spending Score (1-100)'], palette =
'gist_rainbow')

plt.title('Distribution of Spending Score', fontsize = 20)

plt.show()

ax = sns.barplot(y= "Spending Score (1-100)", x = "Annual
Income (k$)", data = df2, palette=("Blues_d"))

sns.set(rc={'figure.figsize':(11.7,8.27)})

sns.set_context("poster")

x = df2['Annual Income (k$)']

y = df2['Age']

z = df2['Spending Score (1-100)']

sns.lineplot(x, y, color = 'blue', palette = 'Accent_r')

sns.lineplot(x, z, color = 'pink', palette = 'Accent_r')

plt.title('Annual Income vs Age and Spending Score', fontsize
= 20)

plt.show()

x = df2.iloc[:, [3, 4]].values

print(x.shape)
```

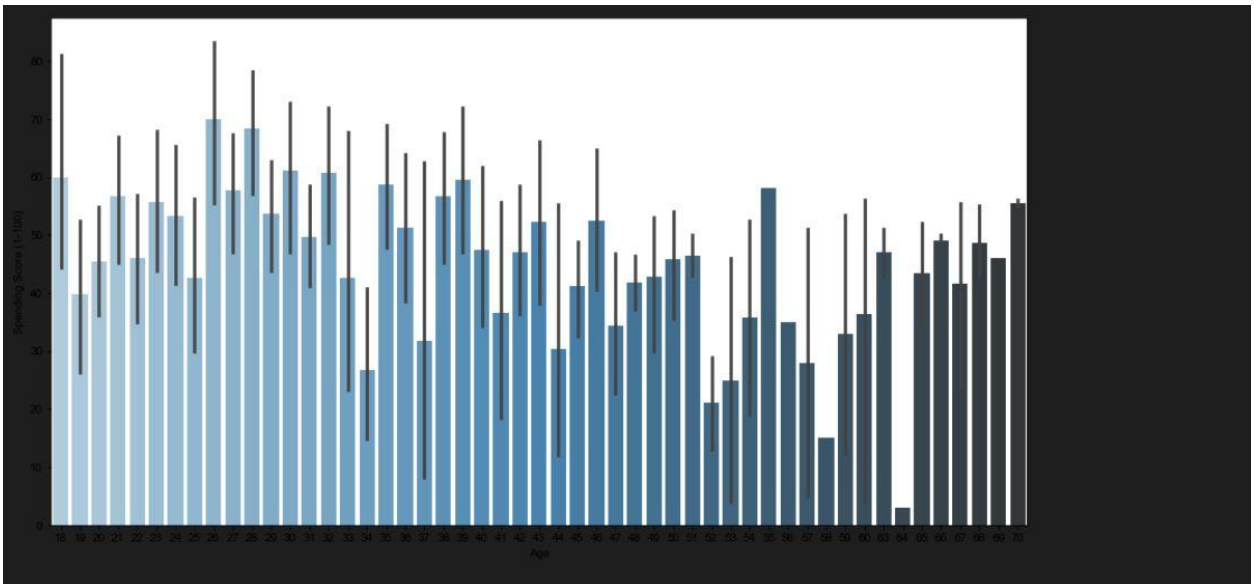
```
from sklearn.cluster import KMeans

wcss = []

for i in range(1, 11):

    km = KMeans(n_clusters = i, init = 'k-means++', max_iter
= 300, n_init = 10, random_state = 0)

    km.fit(x)

    wcss.append(km.inertia_)

plt.plot(range(1, 11), wcss)

plt.title('The Elbow Method', fontsize = 20)

plt.xlabel('No. of Clusters')

plt.ylabel('wcss')

plt.show()

km = KMeans(n_clusters = 5, init = 'k-means++', max_iter =
300, n_init = 10, random_state = 0)

y_means = km.fit_predict(x)

plt.scatter(x[y_means == 0, 0], x[y_means == 0, 1], s = 100,
c = 'pink', label = 'miser')

plt.scatter(x[y_means == 1, 0], x[y_means == 1, 1], s = 100,
c = 'yellow', label = 'general')
```

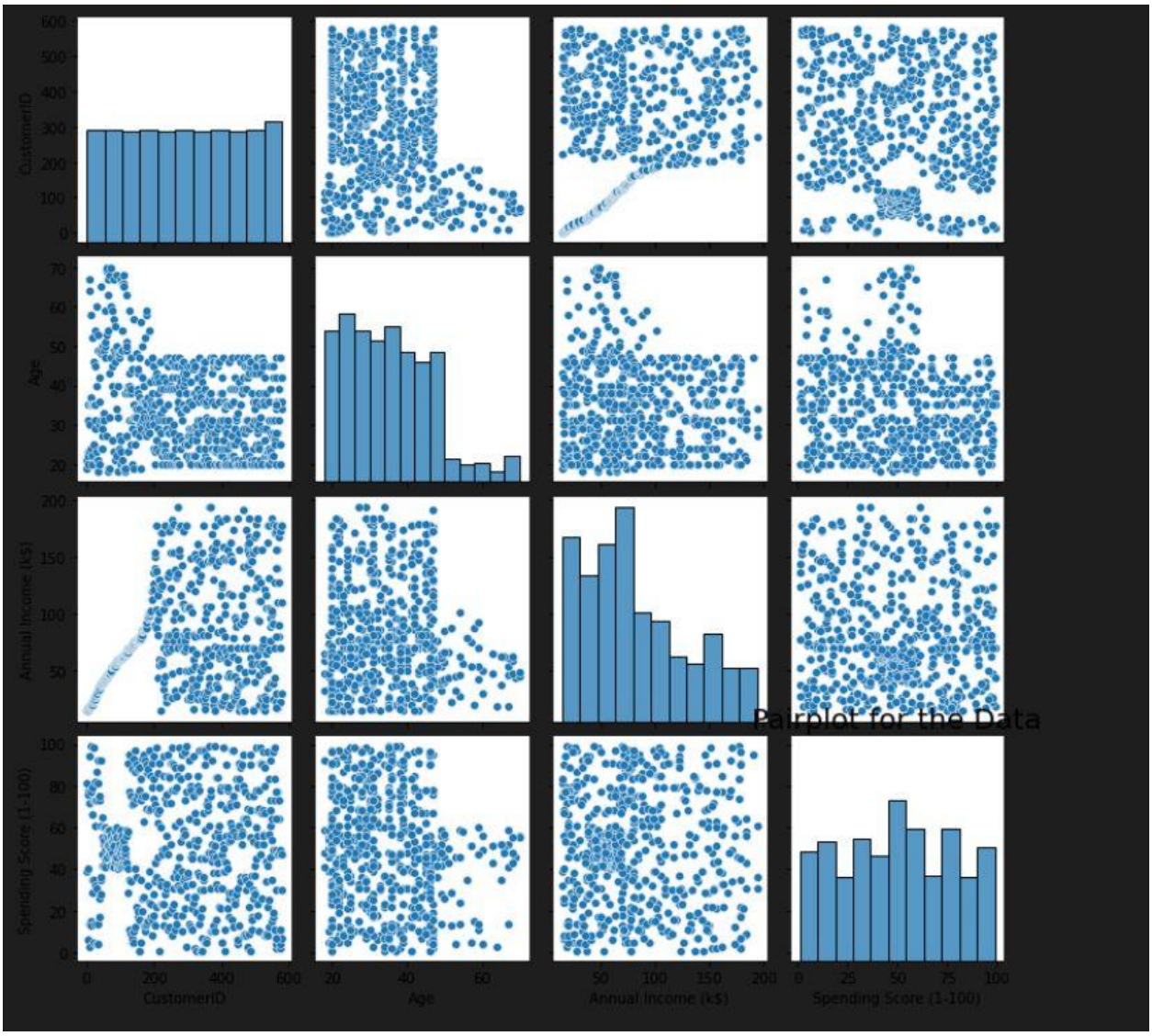
```
plt.scatter(x[y_means == 2, 0], x[y_means == 2, 1], s = 100,
c = 'cyan', label = 'target')

plt.scatter(x[y_means == 3, 0], x[y_means == 3, 1], s = 100,
c = 'magenta', label = 'spendthrift')

plt.scatter(x[y_means == 4, 0], x[y_means == 4, 1], s = 100,
c = 'orange', label = 'careful')

plt.scatter(km.cluster_centers_[:,0], km.cluster_centers_[:,
1], s = 50, c = 'black' , label = 'centeroid')

plt.style.use('fivethirtyeight')

plt.title('K Means Clustering', fontsize = 20)

plt.xlabel('Annual Income')

plt.ylabel('Spending Score')

plt.legend()

plt.grid()

plt.show()

import scipy.cluster.hierarchy as sch

dendrogram = sch.dendrogram(sch.linkage(x, method = 'ward'))

plt.title('Dendrogam', fontsize = 20)

plt.xlabel('Customers')
```

```
plt.ylabel('Ecuclidean Distance')

plt.show()

from sklearn.cluster import AgglomerativeClustering

hc = AgglomerativeClustering(n_clusters = 5, affinity =
'euclidean', linkage = 'ward')

y_hc = hc.fit_predict(x)

plt.scatter(x[y_hc == 0, 0], x[y_hc == 0, 1], s = 100, c =
'pink', label = 'miser')

plt.scatter(x[y_hc == 1, 0], x[y_hc == 1, 1], s = 100, c =
'yellow', label = 'general')

plt.scatter(x[y_hc == 2, 0], x[y_hc == 2, 1], s = 100, c =
'cyan', label = 'target')

plt.scatter(x[y_hc == 3, 0], x[y_hc == 3, 1], s = 100, c =
'magenta', label = 'spendthrift')

plt.scatter(x[y_hc == 4, 0], x[y_hc == 4, 1], s = 100, c =
'orange', label = 'careful')

plt.scatter(km.cluster_centers_[:,0], km.cluster_centers_[:,
1], s = 50, c = 'blue' , label = 'centeroid')

plt.style.use('fivethirtyeight')

plt.title('Hierarchial Clustering', fontsize = 20)

plt.xlabel('Annual Income')
```

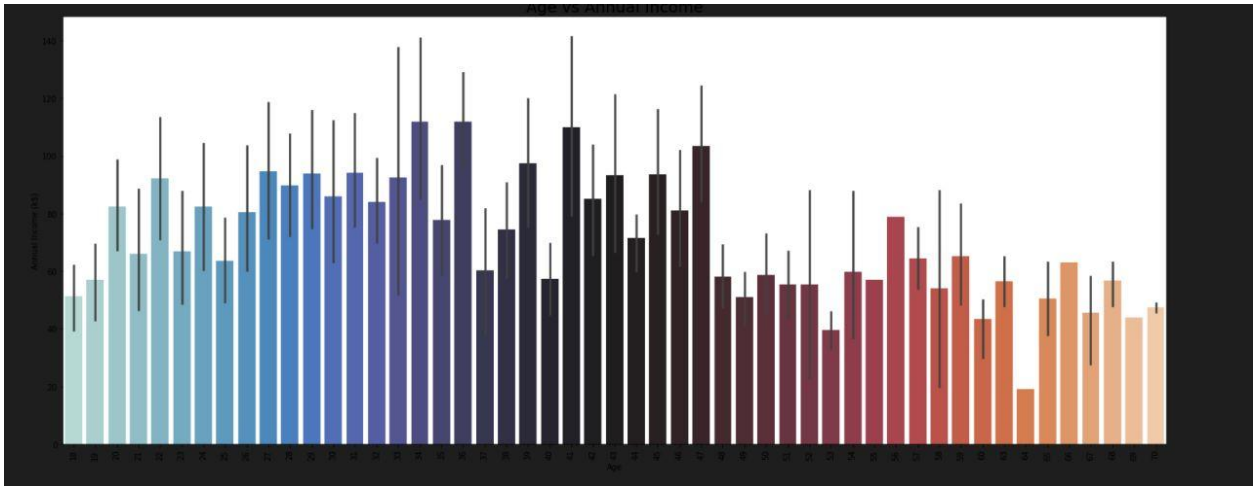
```
plt.ylabel('Spending Score')

plt.legend()

plt.grid()

plt.show()
```

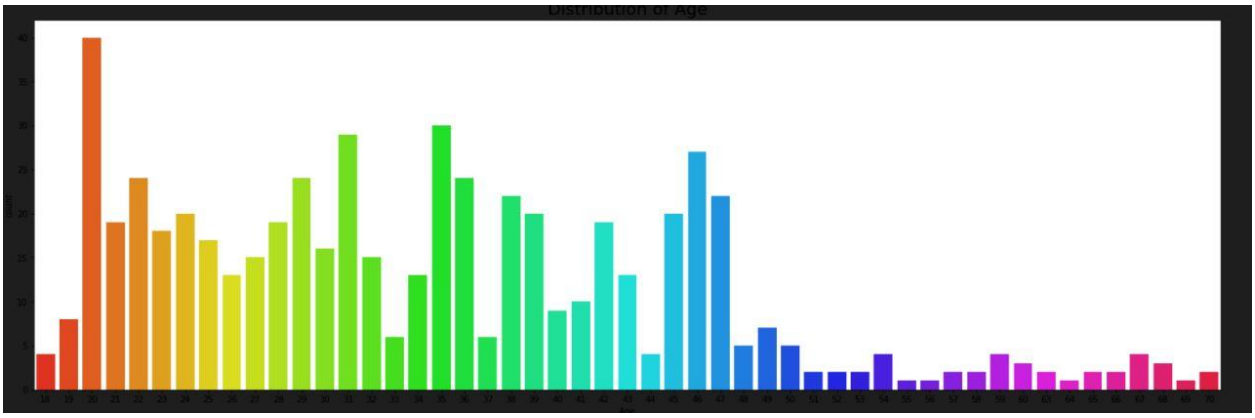
e) Output Screenshots

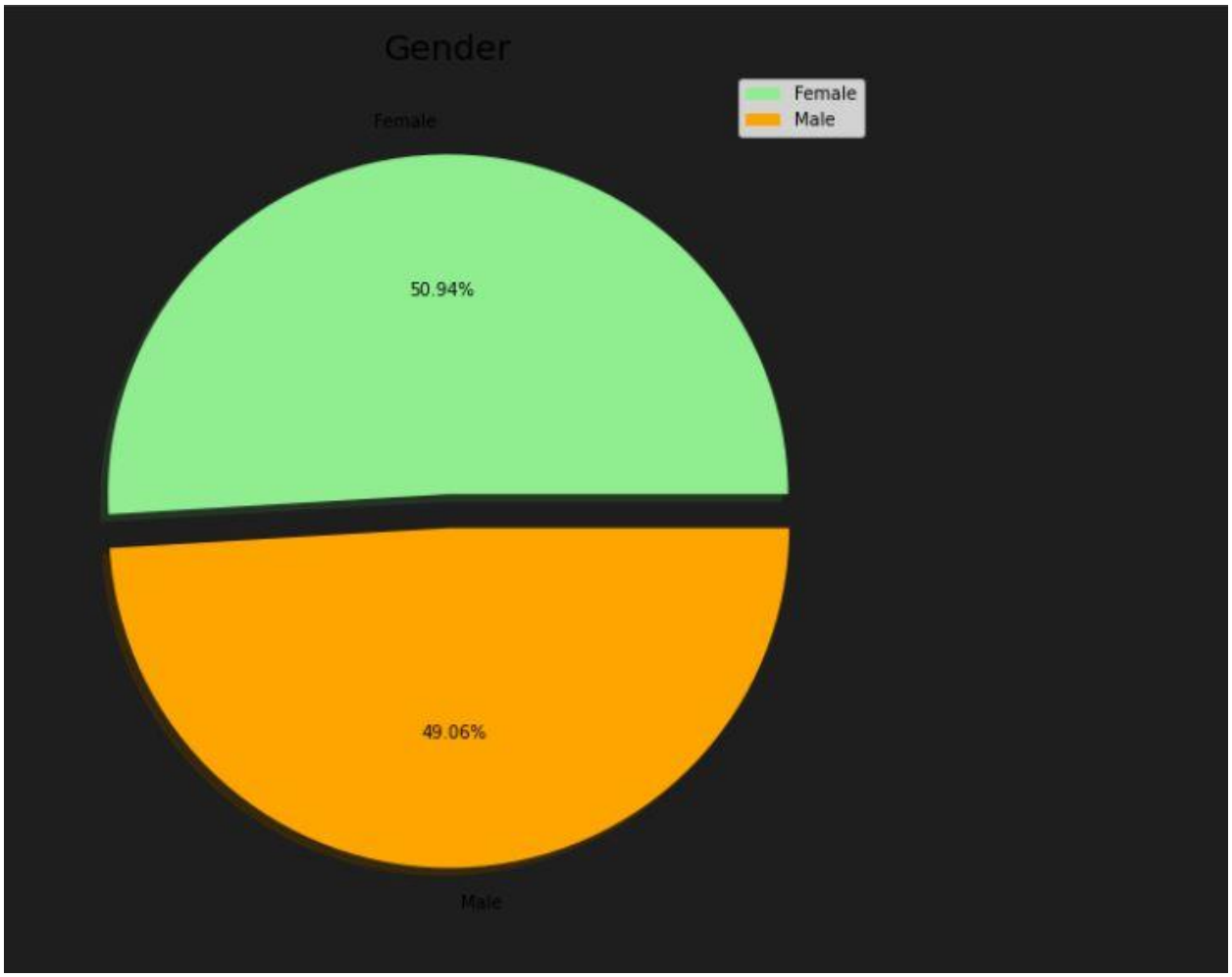


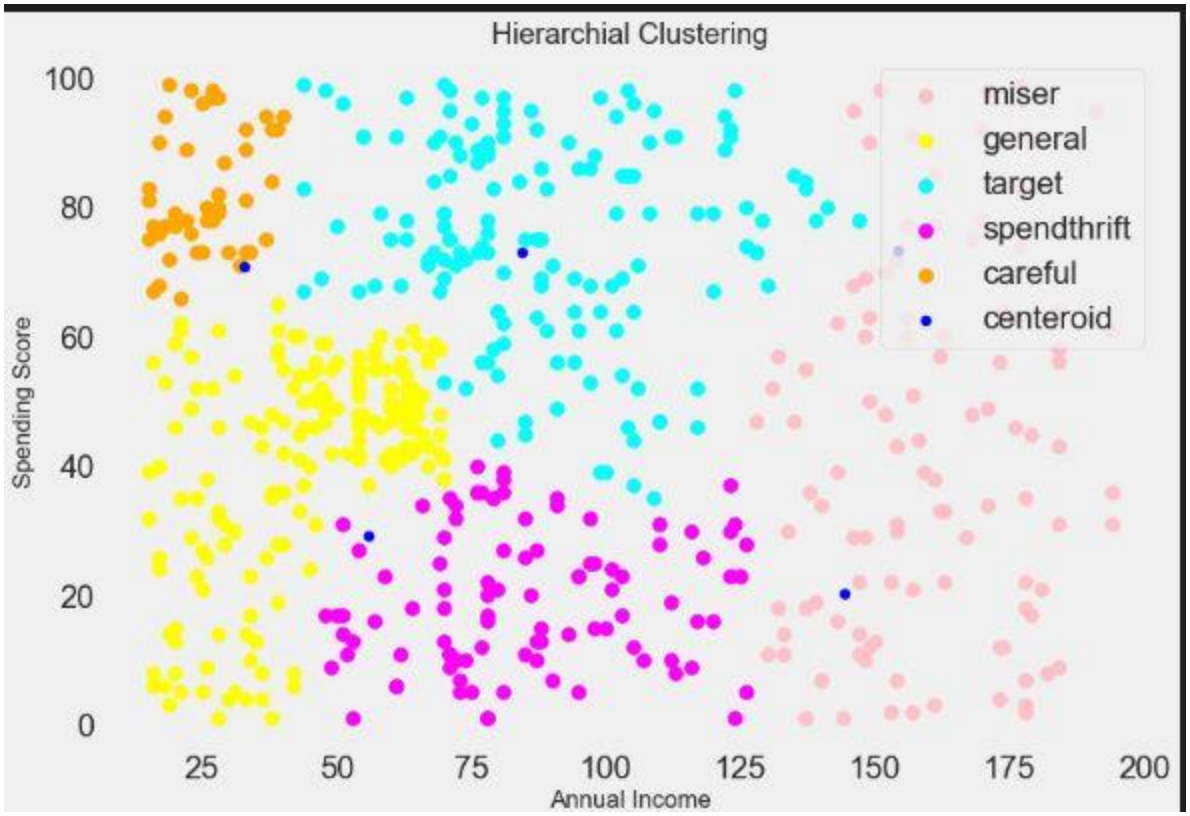


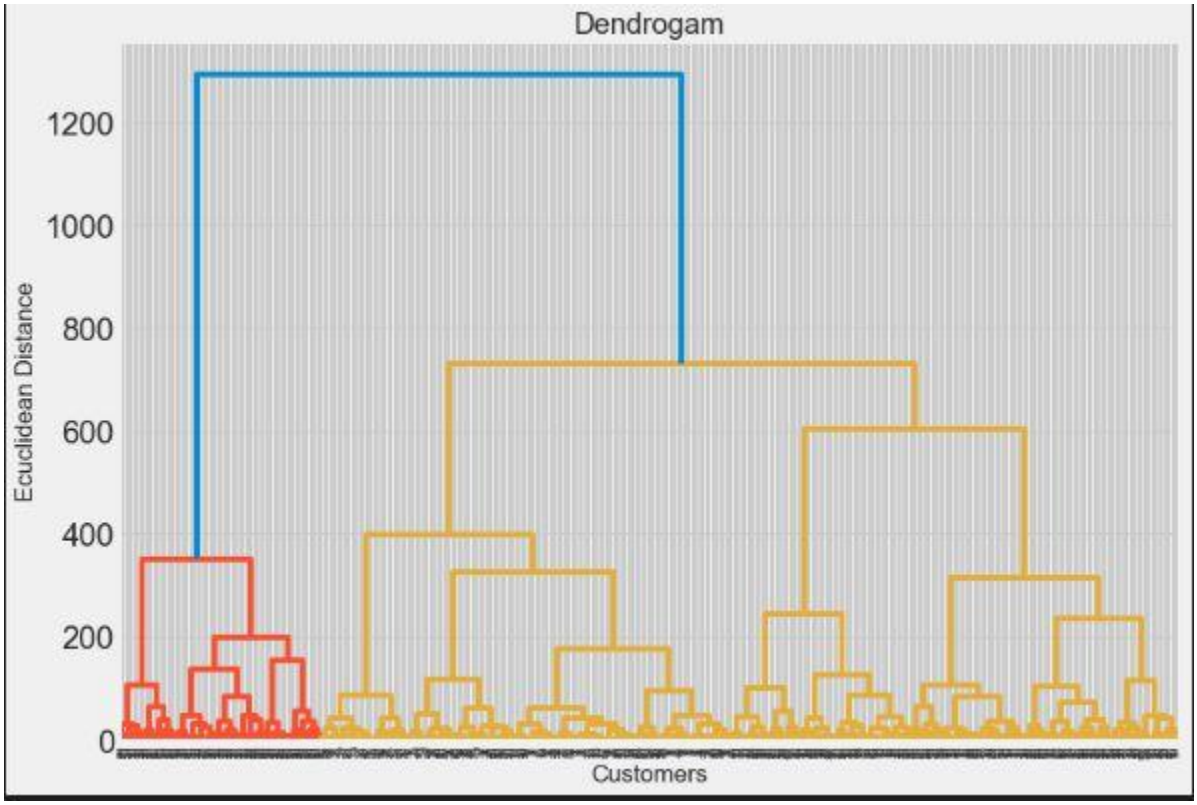
Pairplot for the Data

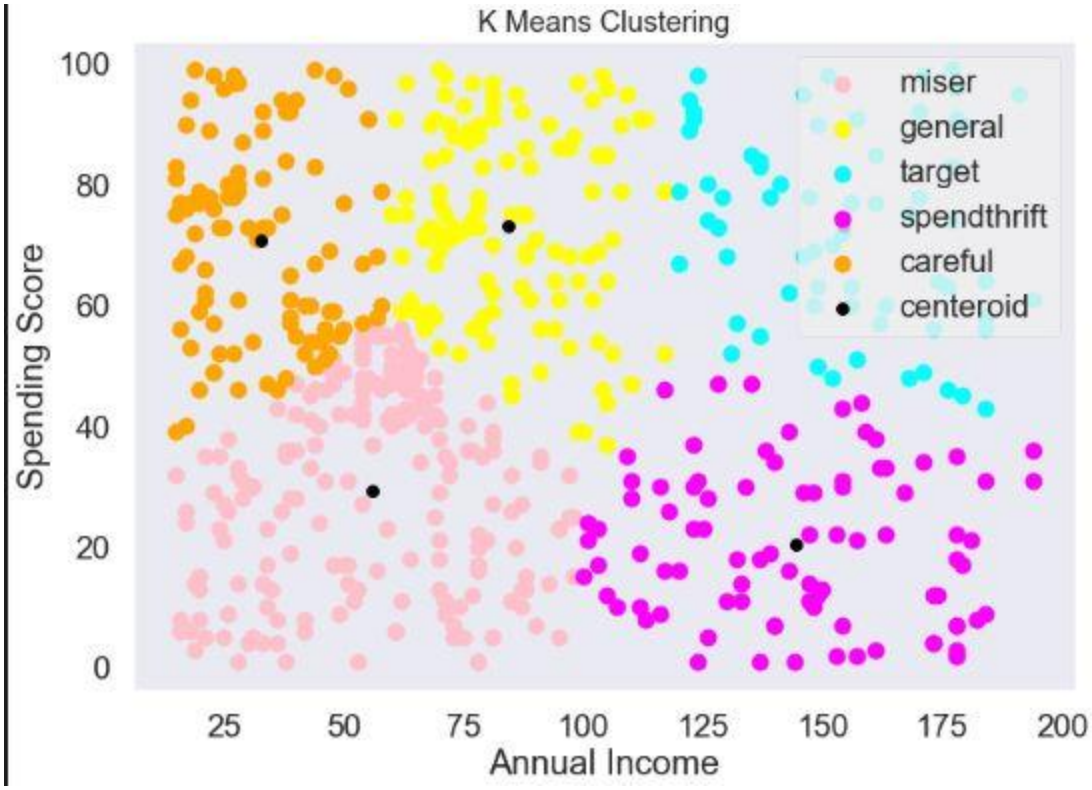


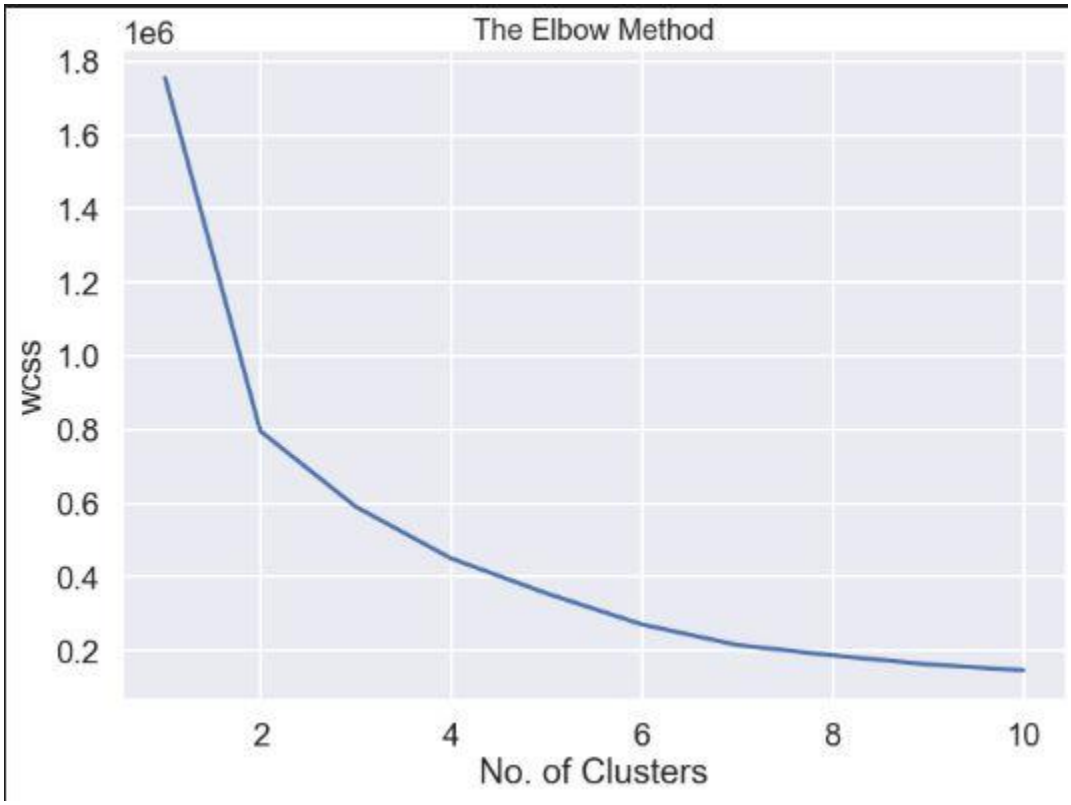

Distribution of Age

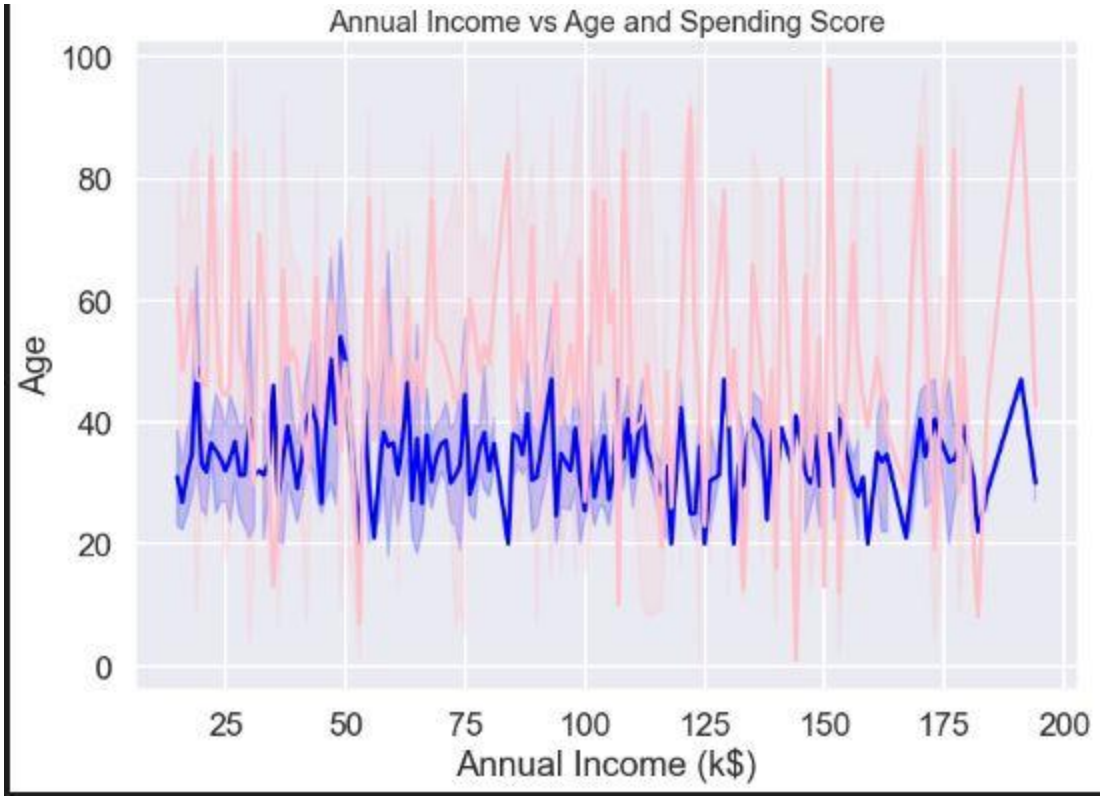

Gender
Female
Male
Female
50.94%
49.06%
Male

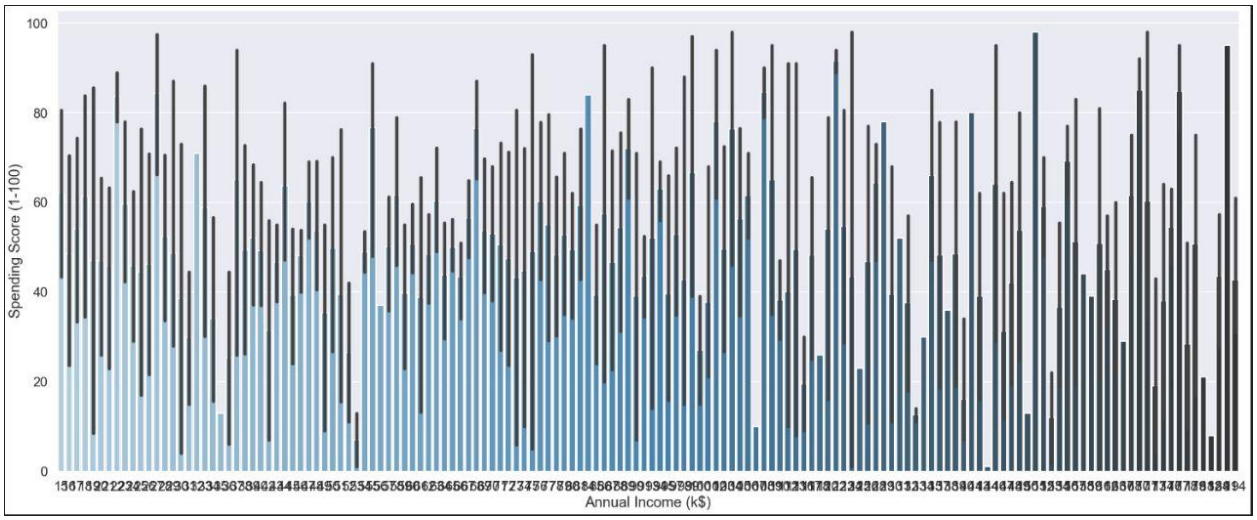

Hierarchial Clustering
miser
general
target
spendthrift
careful
centeroid
Spending Score
Annual Income
100
80
60
40
20
0
25
50
75
100
125
150
175
200

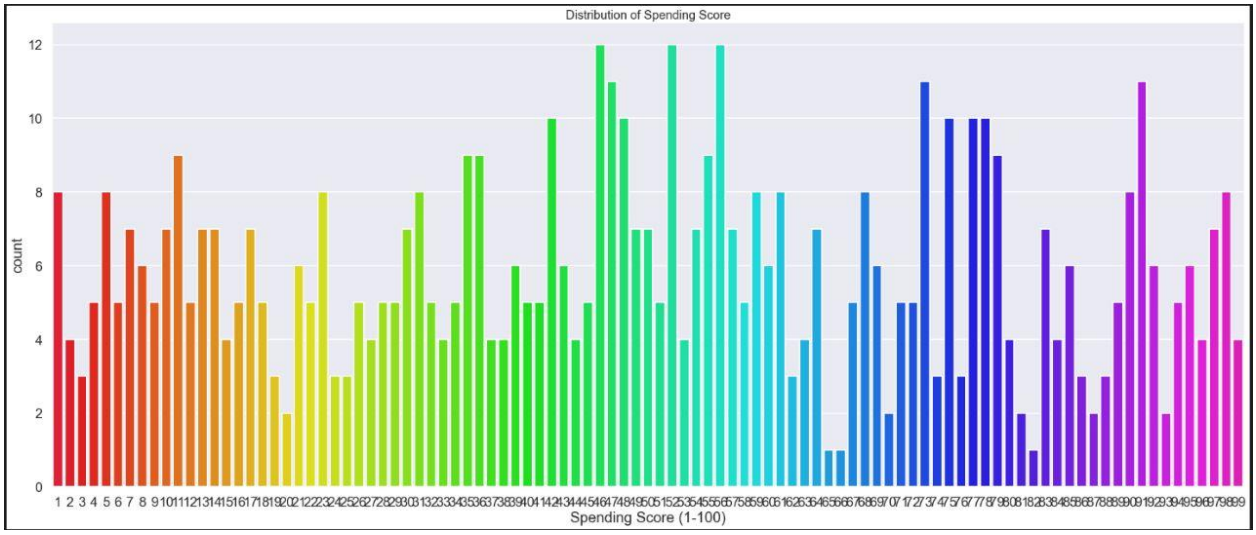

Dendrogam
Ecuclidean Distance
Customers
1200
1000
800
600
400
200
0

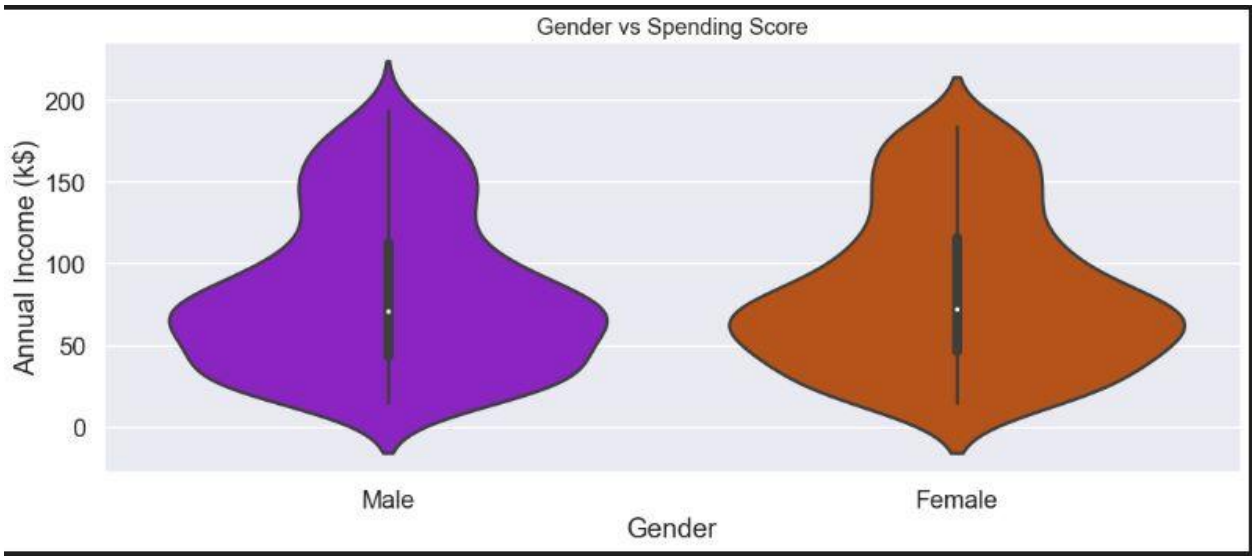

K Means Clustering
Spending Score
Annual Income
miser
general
target
spendthrift
careful
centeroid
100
80
60
40
20
0
25
50
75
100
125
150
175
200

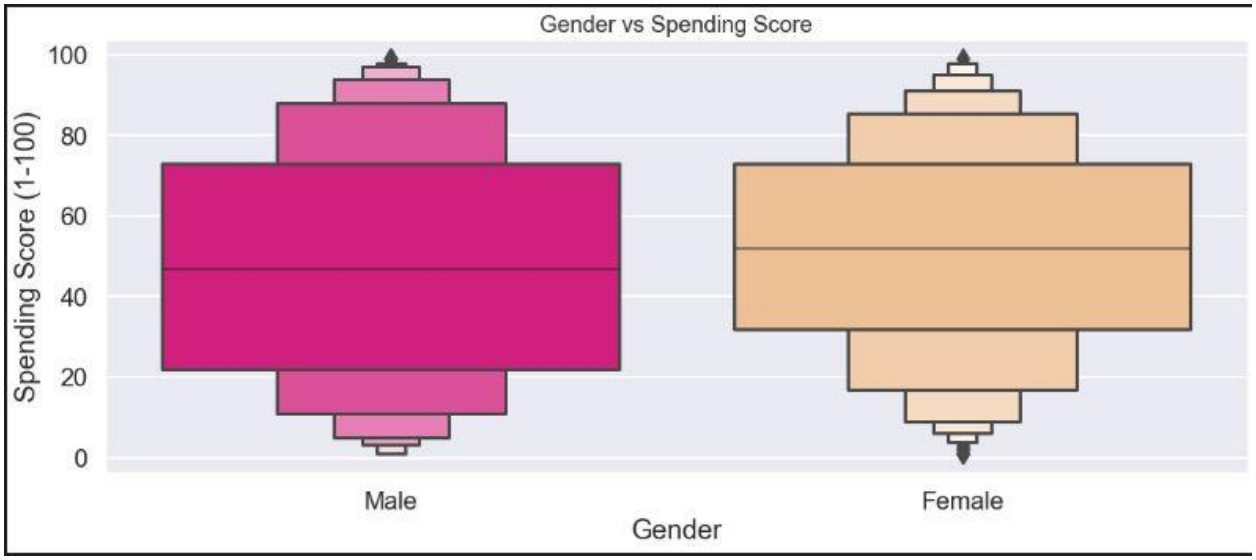

The Elbow Method
1e6
wcss
No. of Clusters
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
2
4
6
8
10

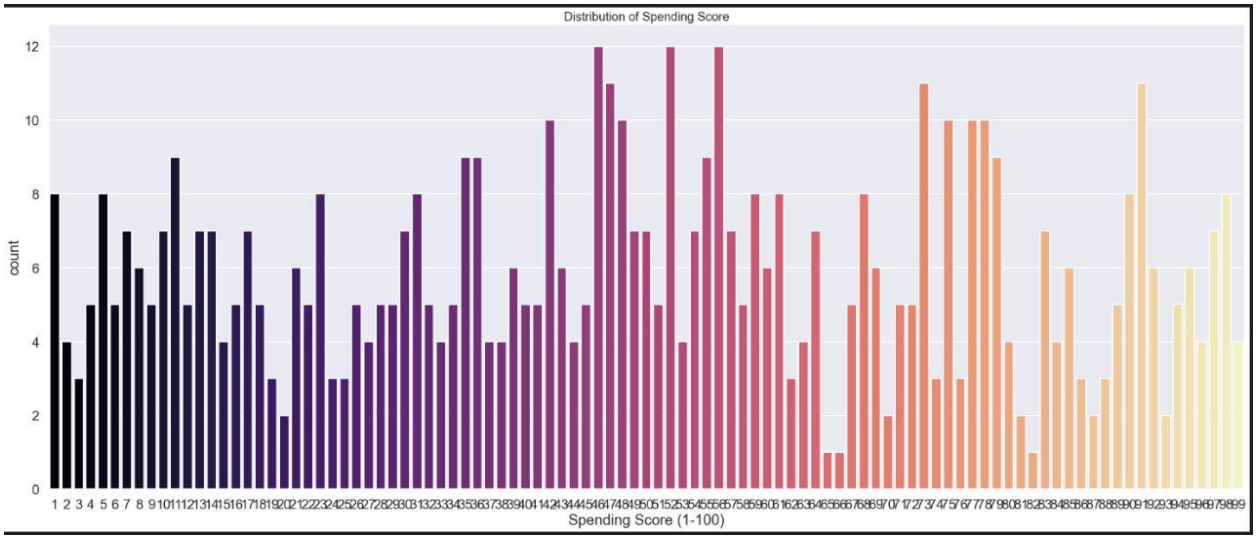

Annual Income vs Age and Spending Score
Age
Annual Income (k$)
0
20
40
60
80
100
25
50
75
100
125
150
175
200


Spending Score (1-100)
Annual Income (k$)
0
20
40
60
80
100


Distribution of Spending Score
count
0
2
4
6
8
10
12
Spending Score (1-100)


Gender vs Spending Score
Annual Income (k$)
0
50
100
150
200
Male
Female
Gender





6) **Expected Results:** The expected result of a shopping clustering analysis would be to identify different groups or segments of customers based on their shopping behavior.

   For example, clustering analysis could reveal that there are certain customers who tend to purchase high-end luxury items, while others may be more price-conscious and buy primarily discounted or sale items. Clustering analysis could also reveal groups of customers who have similar shopping habits, such as those who frequently purchase items online versus those who prefer to shop in physical stores.Ultimately, the goal of shopping clustering analysis is to better understand customer behavior and preferences, so that businesses can tailor their marketing and sales strategies to better meet the needs and expectations of different customer

segments. This can help businesses improve customer satisfaction and loyalty, while also increasing sales and revenue.

7) References

https://www.kaggle.com/datasets/vjchoudhary7/customer-segmentation-tutorial-in-python?resource=download

Ismail El-Adly, M. (2007), "Shopping malls attractiveness: a segmentation approach", *International Journal of Retail & Distribution Management*, Vol. 35 No. 11, pp. 936-950. https://doi.org/10.1108/09590550710828245

Singh, D.P. (2018), "Integration of materialism with shopping motivations: motivations based profile of Indian mall shoppers", *Journal of Asia Business Studies*, Vol. 12 No. 4, pp. 381-401. https://doi.org/10.1108/JABS-05-2016-0075

Simona Damian, D., Dias Curto, J. and Castro Pinto, J. (2011), "The impact of anchor stores on the performance of shopping centres: the case of Sonae Sierra", *International Journal of Retail & Distribution Management*, Vol. 39 No. 6, pp. 456-475. https://doi.org/10.1108/09590551111137994

Lim, C.M., Runyan, R. and Kim, Y.-K. (2013), "Segmenting luxe-bargain shoppers using a fuzzy clustering method", *International Journal of Retail & Distribution Management*, Vol. 41 No. 11/12, pp. 848-868. https://doi.org/10.1108/IJRDM-01-2013-0012

Elizabeth Lloyd, A., Y.K. Chan, R., S.C. Yip, L. and Chan, A. (2014), "Time buying and time saving: effects on service convenience and the shopping experience at the mall", *Journal of Services Marketing*, Vol. 28 No. 1, pp. 36-49. https://doi.org/10.1108/JSM-03-2012-0065

Kaushal, Shailesh & Prasad, Ram Komal. (2022). SEGMENTATION OF SHOPPING MALL SHOPPERS: A CLUSTER ANALYSIS APPROACH.

Ishantha, Asith. (2021). MALL CUSTOMER SEGMENTATION USING CLUSTERING ALGORITHM.

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIO NAL RESEARCH ISSN: 2277- 7881; IMPACT FAC TOR – 6.014; IC VALUE:5.16; ISI VALUE:2.286 VOLUME 8, ISSUE 12(6), DECEMBER 2019

Hemashree Kilari , Sailesh Edara , Guna Ratna Sai Yarra , Dileep Varma Gadhiraju, 2022, Customer Segmentation using K-Means Clustering, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 11, Issue 03 (March 2022),

M. THIRUNAVAKARASU, KUNCHAM PAVAN KUMAR REDDY, G. SRINIVASA TEJA, 2022, CUSTOMER SEGMENTATION IN SHOPPING MALL USING CLUSTERING IN MACHINE LEARNING, International Research Journal of Engineering and Technology (IRJET) Volume: 09, Issue: 03 (March 2022), e-ISSN: 2395-0056, p-ISSN: 2395-0072