

DataWiz - Saksham 2021

A Business Analytics Case Study

SCOPE, SCMHRD (Pune)



OVERVIEW



Data is key in today's world, where multiple strategic decisions are made by generating insights from it.



"Customers are King"

Customer data are the prominent feature of any services. Brands and services use customers' data to get insights which further can be used to generate leads. Similarly shopping complexes and malls utilize customer data to understand the current customer behaviours, which helps in identifying potential customers in order to have more data driven strategy to target customers by proper planning the layout of the mall.

TABLE OF CONTENTS



01

INTRODUCTION

*A basic understanding
of the given dataset*

03

SEGMENTATION

*Drawing inferences
from the customer
segmentation*

ANALYSIS

*Analysing the different
data features*

02

FINAL REMARKS

Final Takeaways

04

01. INTRODUCTION

Customer Data, Articuno Mall, Kanto

About the Dataset :

The customer data represents the customer visiting the malls per day

The dataset consists of the following features

- CustomerID : Unique ID given to a customer
- Gender : Sex of the customer
- Age : Age of the customer visiting the mall
- Annual Income : It is the annual income of the customer in K\$
- Spending Score : It is a score given to a customer based on their spending trends. Value ranges between 0-100



02. ANALYSIS

Using exploratory data analysis, we look into the dataset and its features

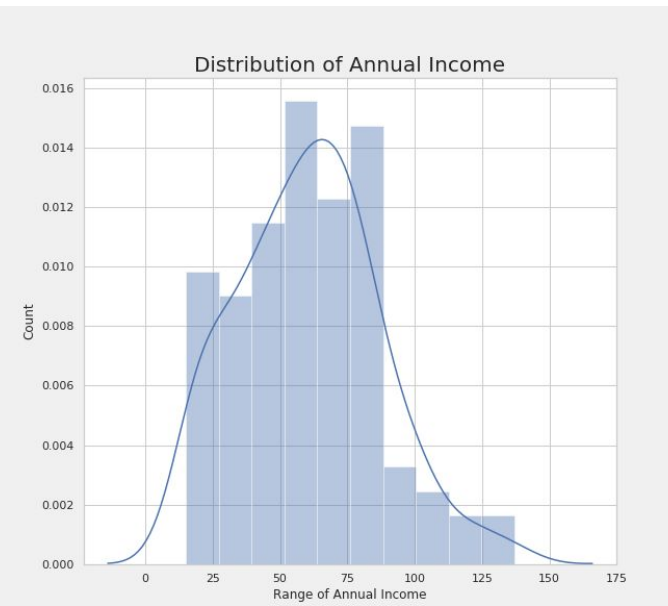
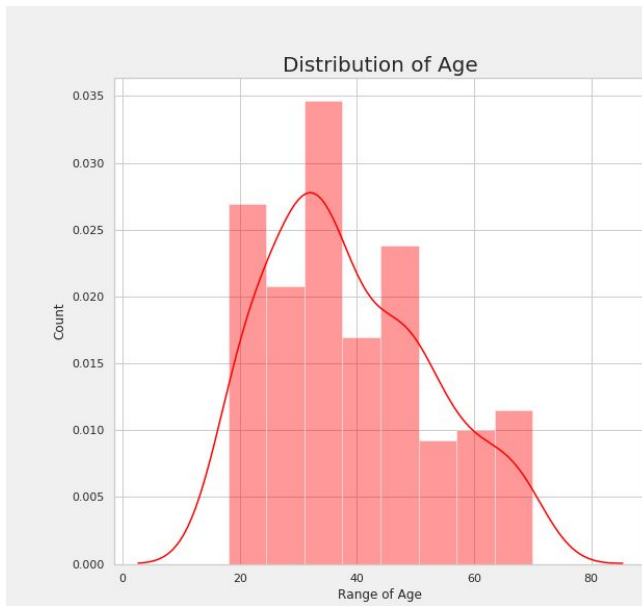
- GENDER



Distribution of Male and Female customers:

We can see that dataset has a fairly even split among male and female customers





- AGE

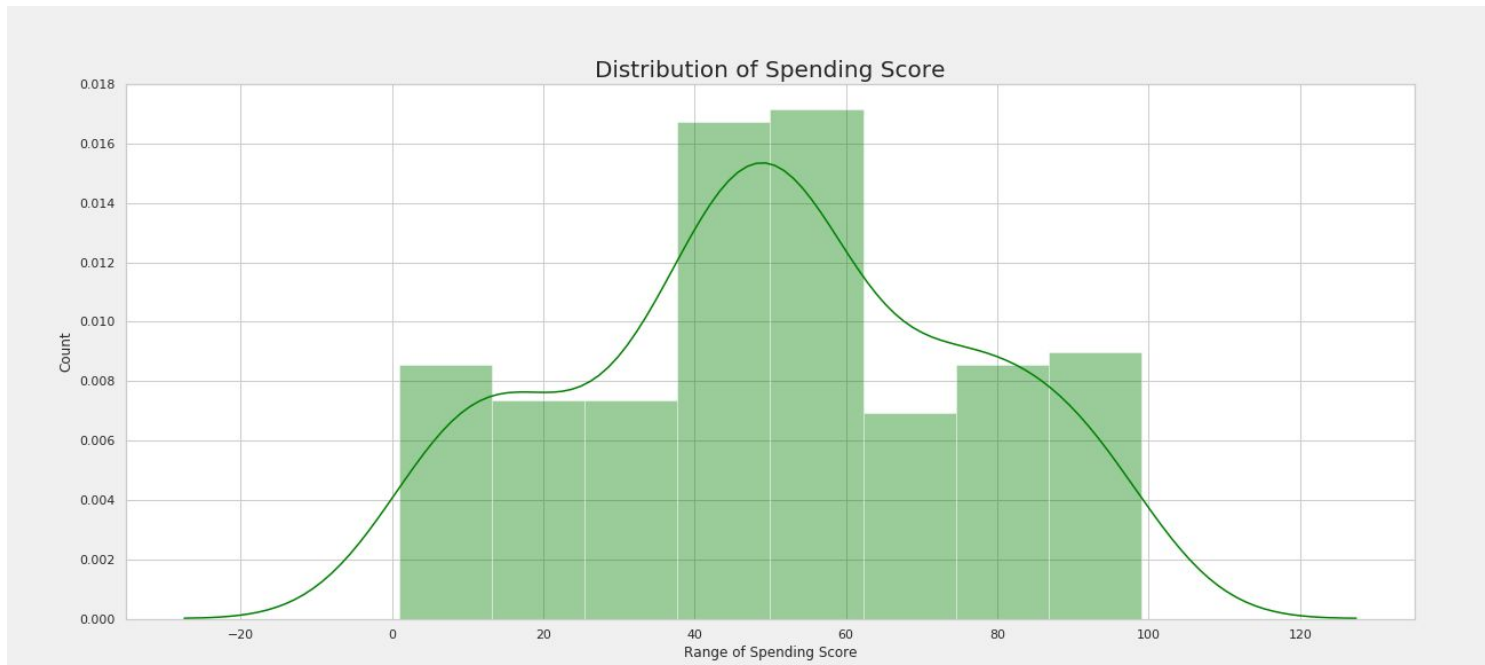
Distribution of Age:

Majority of customers are less than 50 years of age.

- INCOME

Distribution of Income:

Very few people have an annual income of more than \$ 100k



- **SPENDING SCORE**

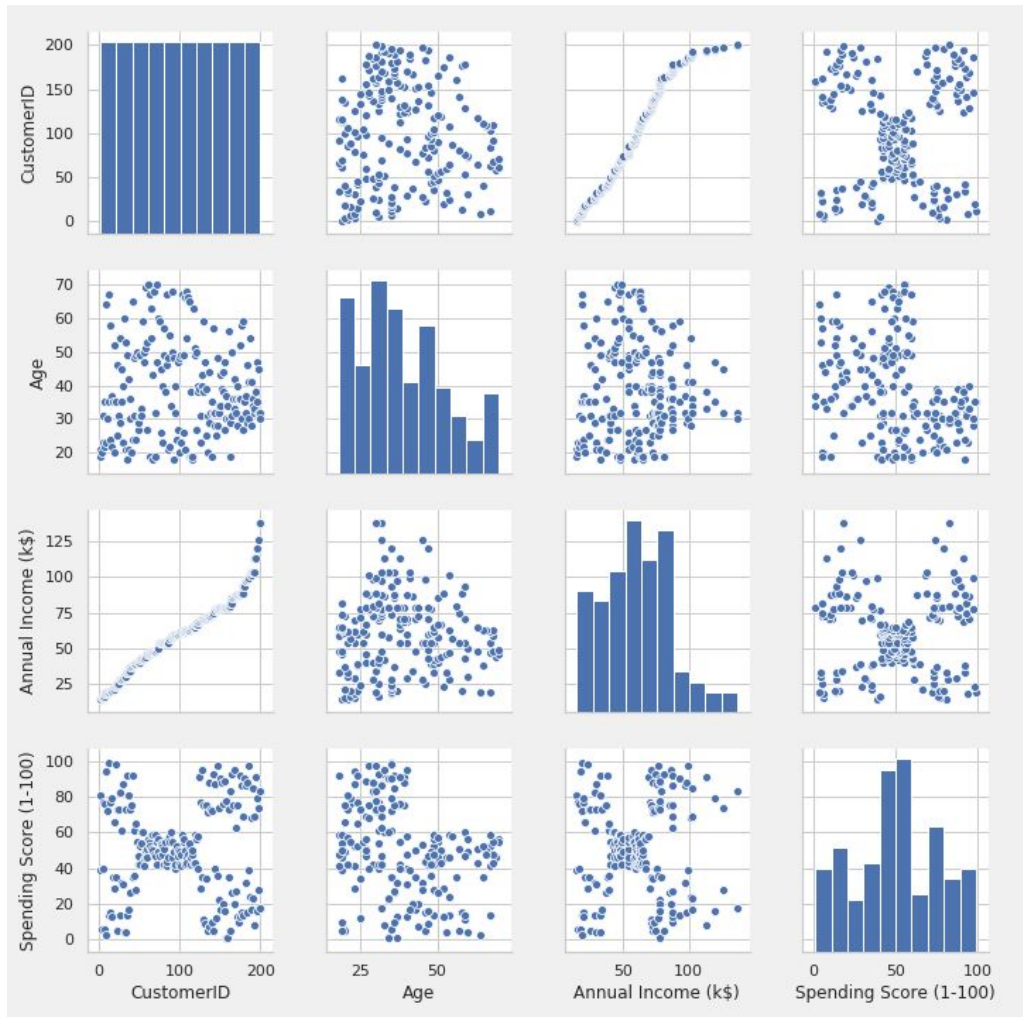
Distribution of Spending Score:

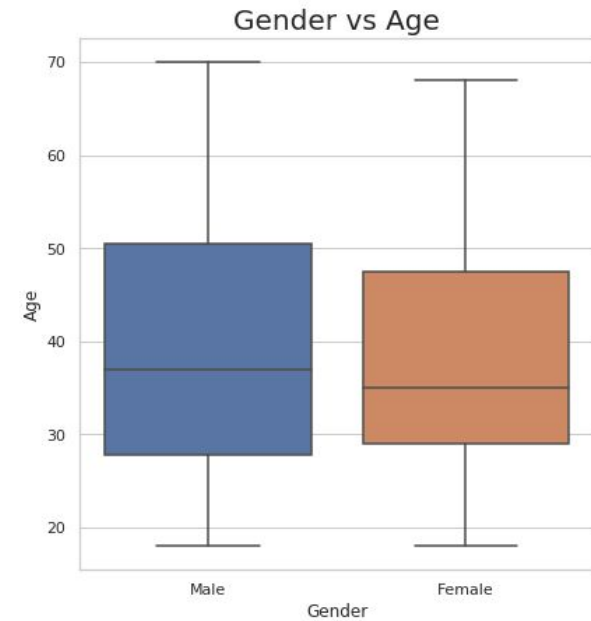
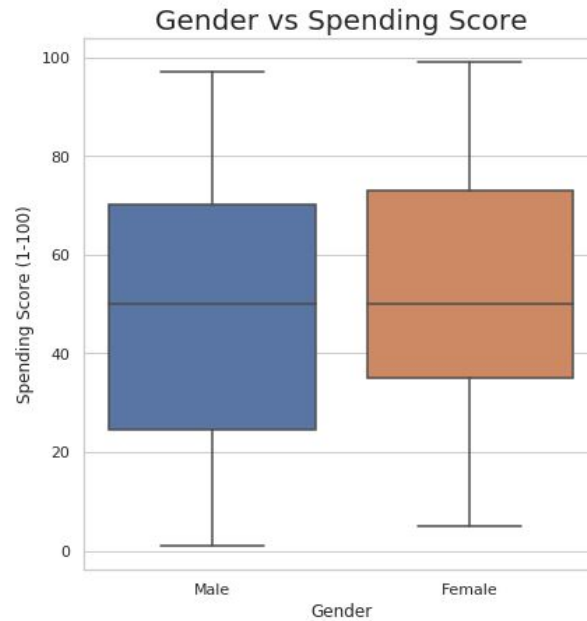
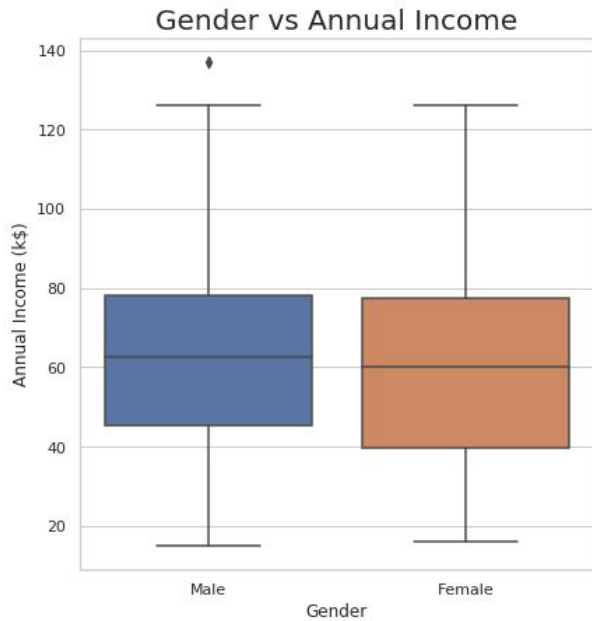
For most of the customers, the spending score centers between 40 and 60

- PAIR PLOT

The last row gives the insights of the Clusters of customer segment.

Also from the third plot of the last row, it is evident that the Spending score of a customer is related to their Annual Income.





We can see from the above that there is only slight differences between the male and female in terms of Annual Income, Spending Score and Age. So, it can be inferred that gender has little impact on the segmentation of customers.

A few points can be inferred like

- More number of females have annual income of about \$ 40k than that of male
- Spending score of female are more than that of male
- Slightly more number of male customers have age 50 than that of female.

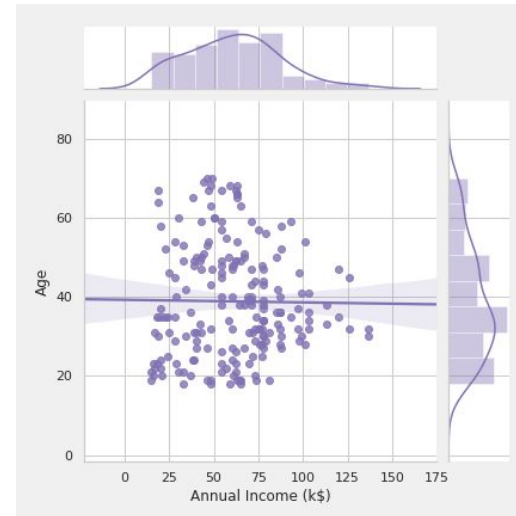
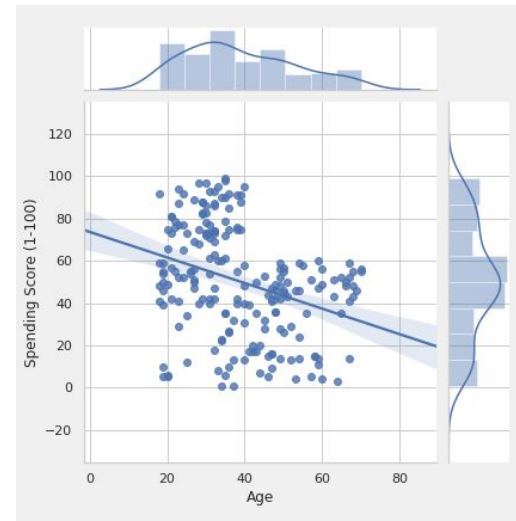
● SPENDING SCORE VS AGE

- The maximum spending score is in the category of age 20-40.
- The linear fit plot has a negative slope which suggests that the spending score is decreasing with age
- Customers with age above 40 has a spending score of less than 60

Hence, we can infer that there seems to be 2 segments of customers by age vs score.

● AGE VS ANNUAL INCOME

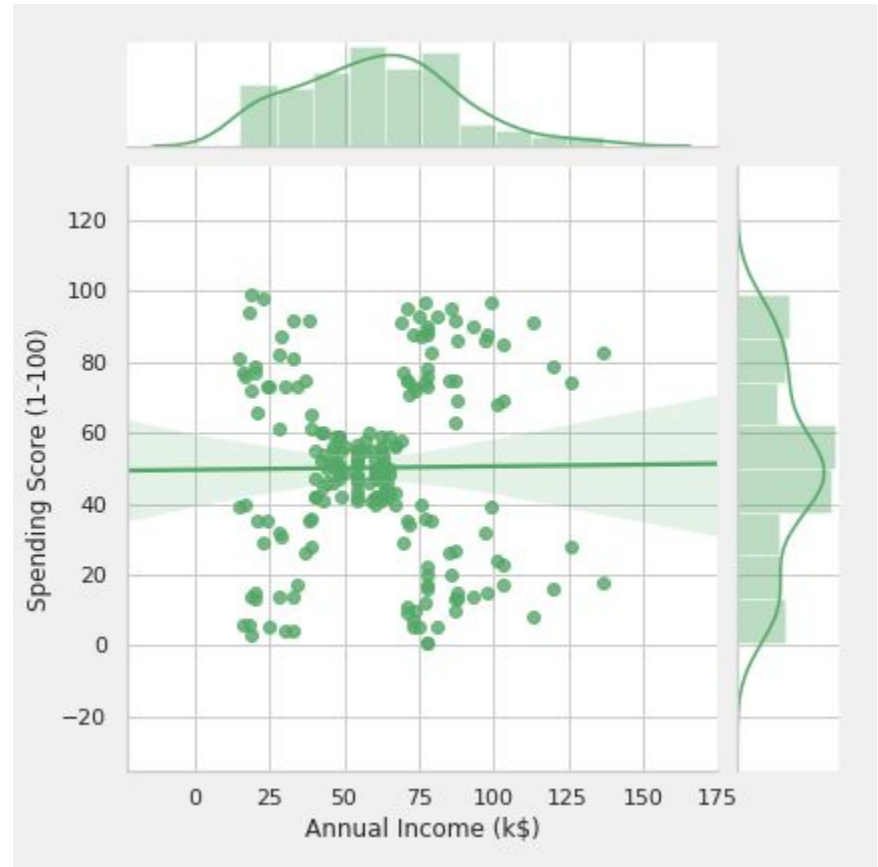
From the age vs annual income, there is not much information about segmentation of customer as only one distinct pattern can be observed.



- SPENDING SCORE VS ANNUAL INCOME

In Annual Income vs Spending Score, we can infer that it shows different patterns. It might be because of the different purchasing power of the customer and their different spending habits.

The different patterns shown are interesting and gives us the ideal way to segment the customers based on income and score!



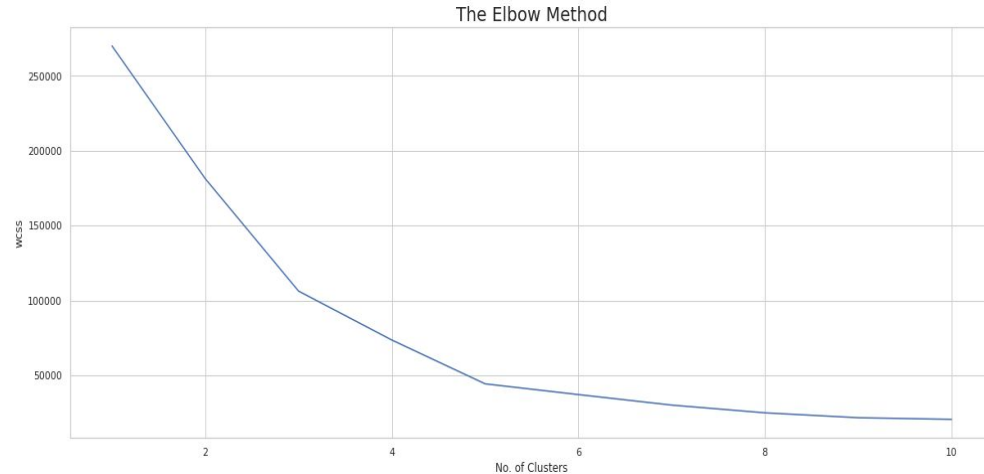
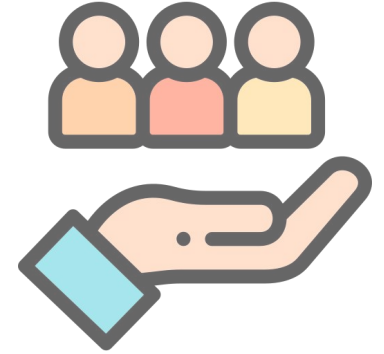
03. SEGMENTATION

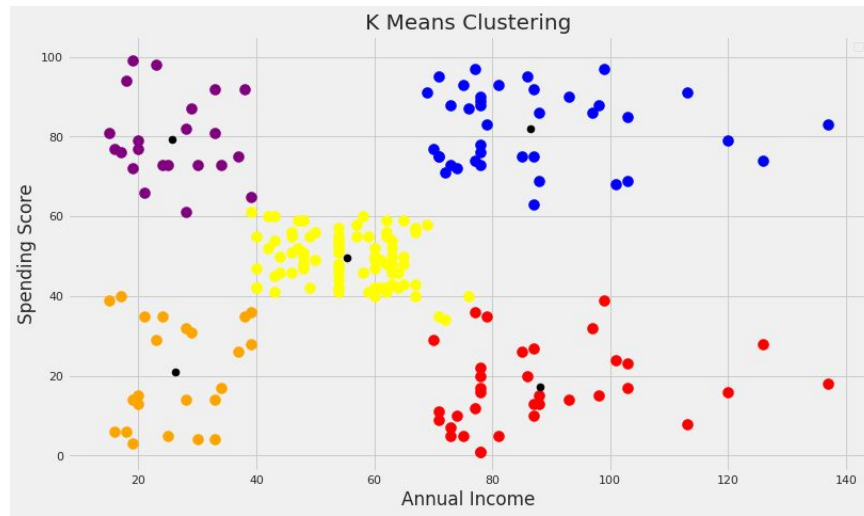


In order to segment the customers, we will be using K-means clustering algorithm.

Finding optimal number of clusters:
We will be using the Elbow Method to get the number of optimal clusters.

It is evident from the plot that the optimal number of cluster is 5, as the slope of the curve is not much steep after that.





Using K-means, we get five clusters of customer segments based on annual income and spending score

- Cluster 1(Orange): Customers low with annual income and low spending scores. These people know how to spend and save money.
- Cluster 2(Purple): Customers low with annual income and low spending scores. These are those who love to buy more often even though they have a low income and are the **satisfied customers**.
- **Cluster 3(Yellow):** Customers with average income and an average spending score. These people again will **not be the prime targets** but the mall can implement other promotional strategies to **increase their spending score**.
- Cluster 4(Blue): Customers with high income and high spending scores. These are **the regular and satisfied customers** of the mall.
- **Cluster 5(Red):** Customers high income but low spending scores. These can be **the prime targets** of the mall, as they have the **potential to spend money** but are unsatisfied by the mall's services. So, the mall authorities must implement new strategies to so attract these people and can convert them into regular ones.

04. FINAL REMARKS



So, for the mall in order to increase the profits, the authorities should target people belonging to cluster 3 (Yellow) and cluster 5 (Red).

They should also maintain its standards so that people belonging to cluster 2 and cluster 4 are happy and satisfied as they are the high spenders and the mall authorities do not want to lose them.



Thank
You!

Team-A
IIT(BHU), Varanasi
Abhinash Daulaguphu
Ajeya Krishna

Credits: The icons have been used from [Flaticon](#)