Mall Customer Segmentation

Wong Zhao Wu

AIML
CA2

PROBLEM OVERVIEW & EDA

Modelling Objective and Data Exploration

DATA PREPROCESSING

Preprocessing Technique before Cluster Modelling

CLUSTERING ALGORITHMS

Explore Different Clustering Algorithms

UNSUPERVISED FEATURE SELECTION

Identify Best Subset of Features for Clustering Algorithms

CLUSTERS INTERPRETATION

Interpret Clusters Formed to Identify Target Customer

CONCLUSION : THE MARKETING STRATEGY

Final Marketing Strategy for Target Customer

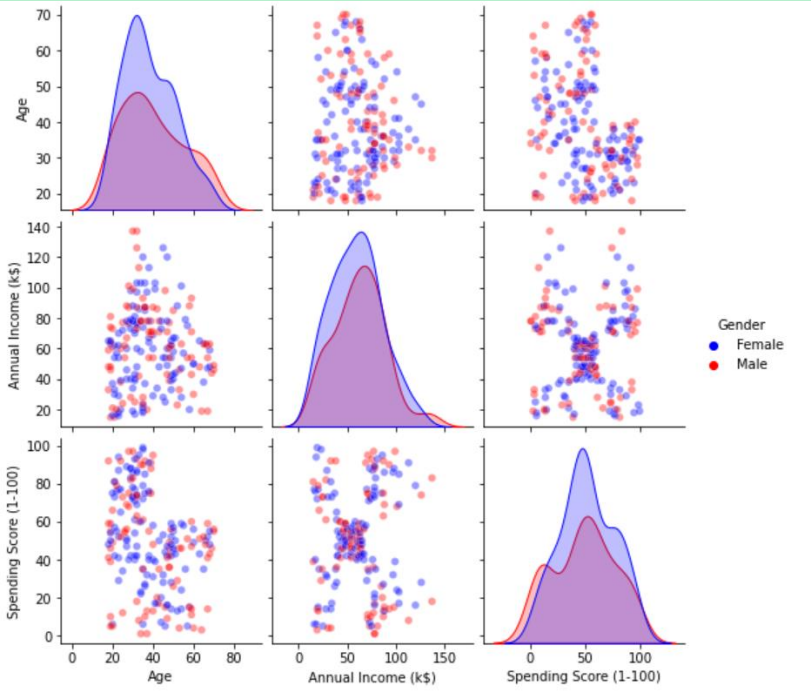
PROBLEM OVERVIEW

As the owner of the mall and you want to **understand your customers** so that **appropriate directions can be given to marketing team** for them plan their strategy accordingly.

Modelling Objective

- Achieve customer segmentation using unsupervised ML algorithms
- Identify target customers and devise a marketing strategy to boost the growth of mall supermarket.

EDA – Pairplot



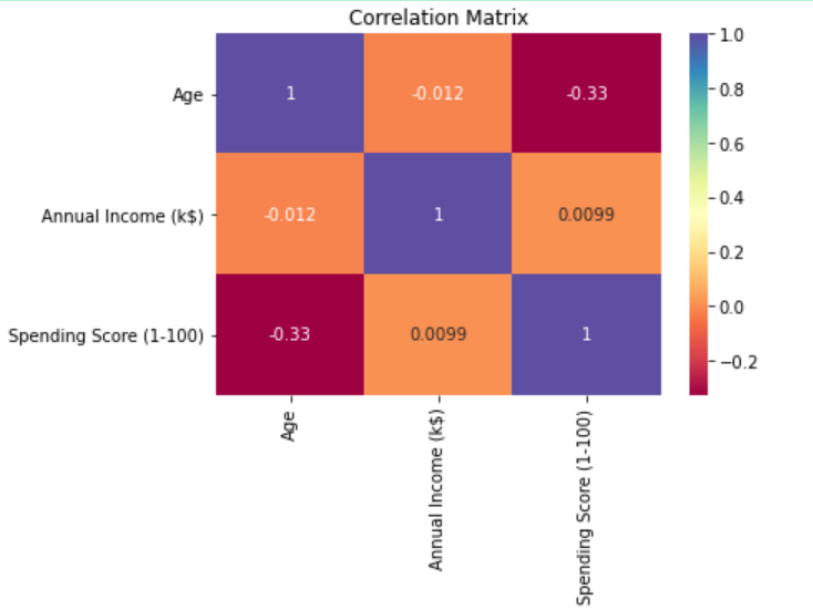
Observations:

1. By just comparing `Annual Income (k$)` and `Spending Score (1-100)`, there **seems to be 5 identifiable clusters**.

2. Distribution of male and female datapoints seems to be **randomly scattered across different measures**. Hence, it is arguable that there *might not be a significant different of spending habits across Male and Female.*

3. There are **no identifiable clusters wrt to Age attribute**, which might lead to ambiguity for the clustering algorithm. However, further analysis is required to unveil the relationship of Age with other attributes.

EDA – Correlation Plot



Observations:

1. **Week Linear Relationship** between Age and Spending Score (1-100), implying there might be some relationship for Age that we have yet to understand. Hence, we will keep the Age feature for the Clustering Algorithm.

DATA PREPROCESSING

Dropping CustomerID

Drop CustomerID column since it does not review any useful information.

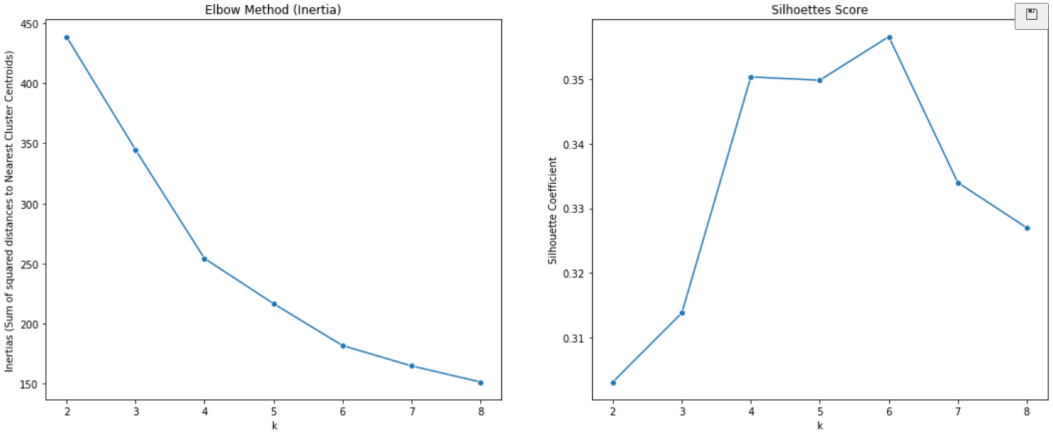
Standard Scaling

Scale datapoints to ensure all features are in same scale.

Dichotomous Variable for Gender

Dummy encode Gender column to represent the category value.

CLUSTERING ALGORITHMS – K-Means Clustering



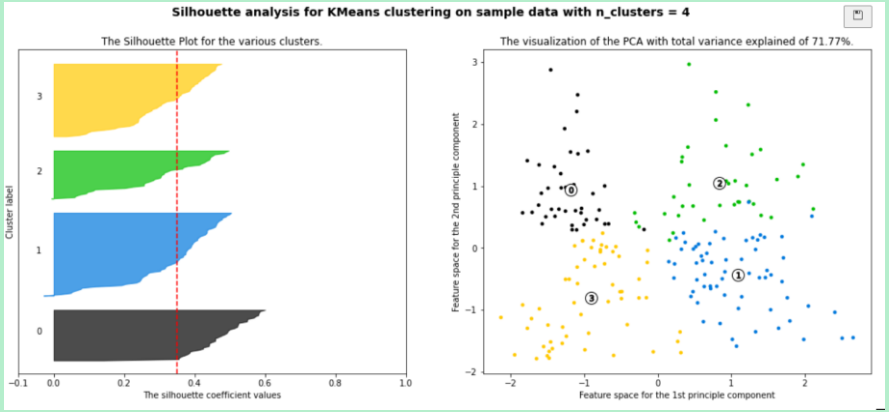
Choosing Number of k

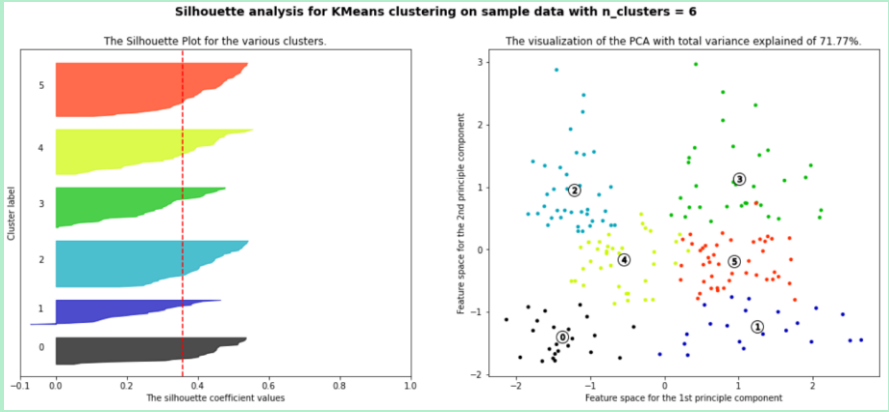
1. **Elbow Method (Inertia)** choosing the number of k is **ambiguous** as the Inertia seems to decrease uniformly when k increase.

2. **Silhouettes Score** max out at $k \in [4,5,6]$ with **peak at $k = 6$**.

Hence, $k \in [4,5,6]$ seems to be a good candidates for the number of clusters generated. Let us perform more comprehensive **Silhouettes Analysis** and **Visualisation of clusters through PCA**.

CLUSTERING ALGORITHMS – Silhouette Analysis



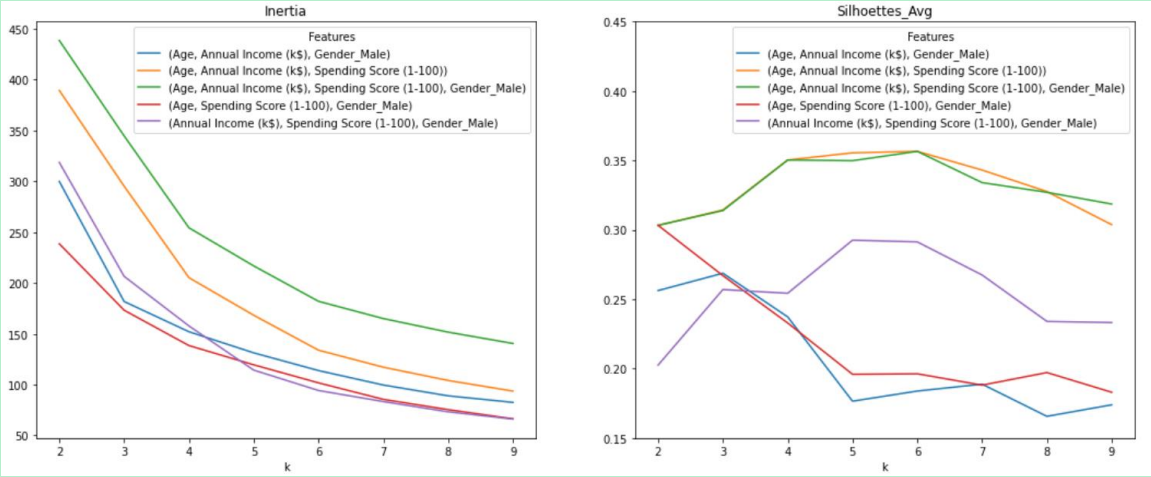


Choosing Number of k

1. From the observations gathered, the plausible values of k is between 4 or 6 when we perform clustering with all 4 features. Before we proceed to the next algorithm let us explore further on how Feature Selection can help to improve the quality of clustering.

UNSUPERVISED FEATURE SELECTION

To examine which combinations of features can generate more quality clustering, one method is to **perform clustering with different subset of features**. The aim of this is to **maximise the Silhouette Score** and generate high quality and interpretable clusters.



Observations:

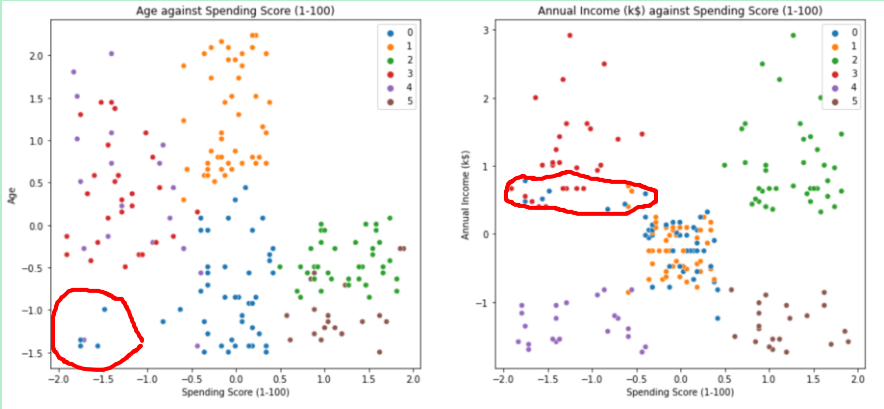
- Removing Gender_Male as clustering feature has close to no effect to the clusters formed as shown by the Silhouettes_Avg of having similar shape.
- The finding is corroborated with the ambiguous datapoints as shown in Pairplot earlier.

For the sake of easier visualization, I will **Remove Gender_Male** from my Feature Set.

CLUSTERING ALGORITHMS

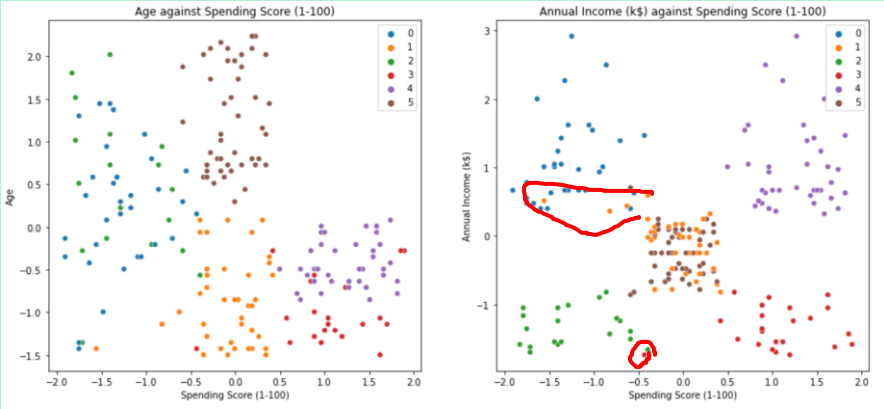
Aggrolomative Clustering

Form Cluster based on hierarchical relationship.



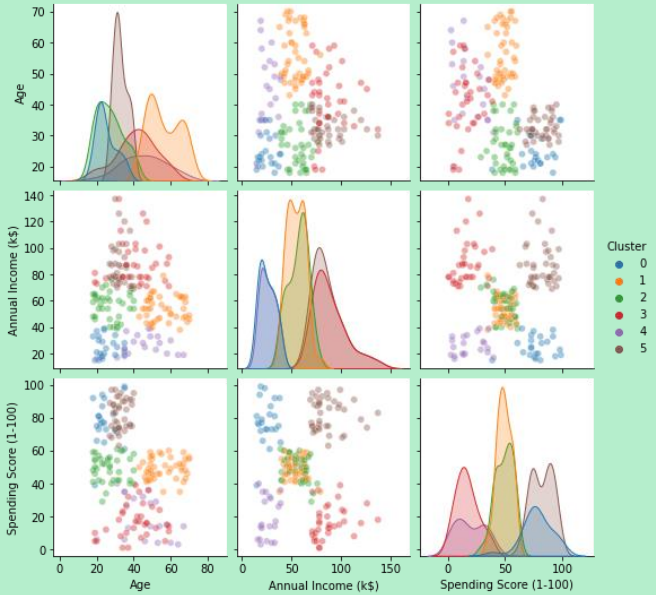
Spectral Clustering

Make no assumptions on the shape of clusters.

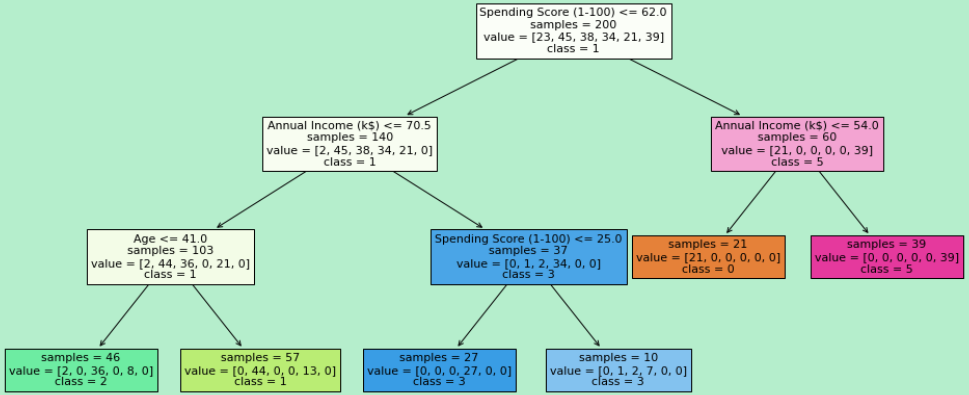


CLUSTERING INTERPRETATION – K-Means Clustering

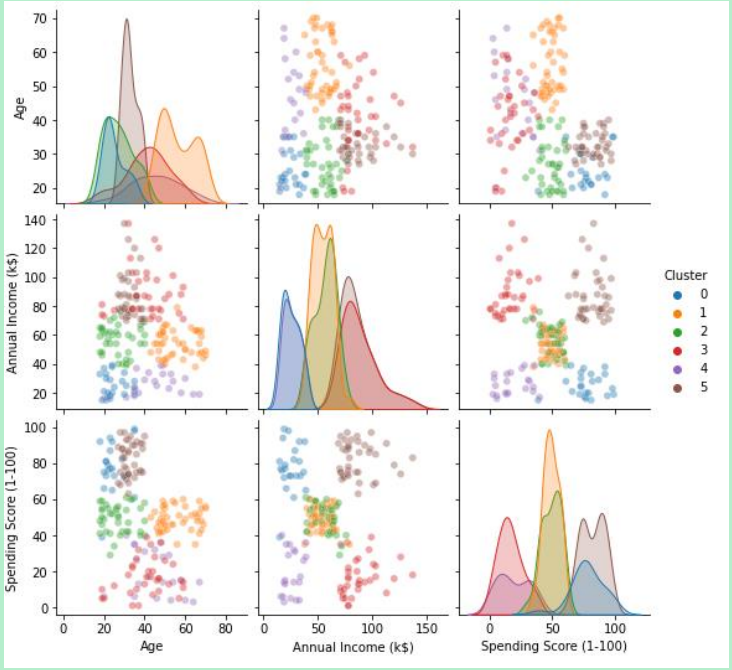
Colored Pairplot



Decision Tree



CLUSTERING INTERPRETATION – K-Means Clustering



00

Overspending Youngsters a.k.a. Trendy Ah Beng

Youngsters that can spend more and earn less are most likely getting their incentives from their parents, and the marketing strategy to devise is probably ***to sell items that can resonate with youths*** **better** like sport equipments or idols endorsed products.

01

Senior Moderate Buyers a.k.a. Housewife Aunty

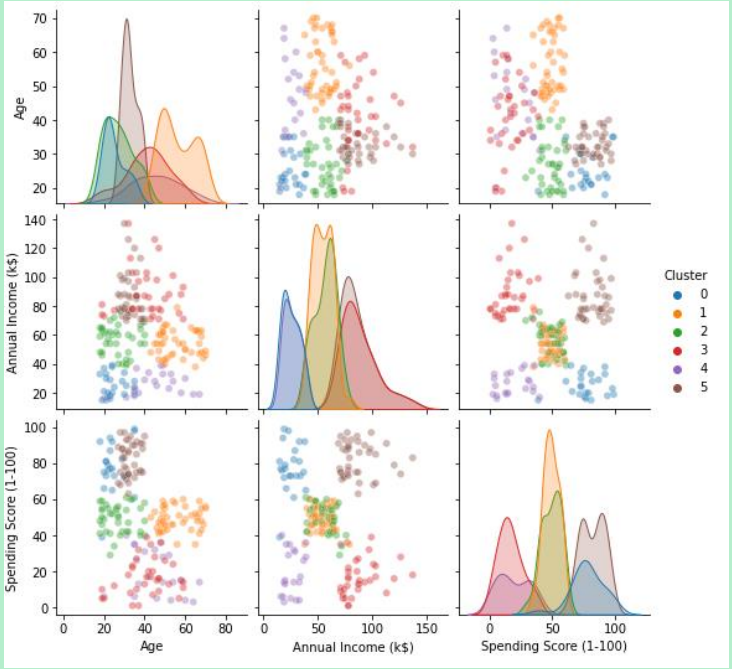
Being the most loyal customer to the supermarket, strategy like campaigns would not actually make much effects to their spending behaviours. The subtle strategy that could be adopted is through ***occasional offerings*** like Bundle Sale or Buy 1 Get 1 Free to boost their spending behaviour for staple goods.

02

Junior Moderate Buyers a.k.a. Employed Alan

They are youth who are just starting to work with relative moderate money to spend on groceries. Perhaps some ***campaigns or games that allows them to exchange vouchers*** or winning lucky draws can inspire them to spend more and earn more points for the campaign.

CLUSTERING INTERPRETATION – K-Means Clustering



03

The Stingy Buyers a.k.a. Kiamsiap Uncle John

They should be ***main focus of marketing strategy*** as they has the ***capabilities to spend more than what they are spending now***. Hence, further analysis should be done to study closely their spending behaviours in order to devise more specific strategy like maybe expanding the luxury item section etc.

04

Low Income Buyers a.k.a. Needy Jenny

They are the one that ***needs the most help*** from the supermarket although it might be a blunt to just directly provide offers and incentives for them, one strategy is to organise campaigns that ***allow them exchange staple goods through the green act*** of maybe collecting reusable materials or returning plastics bags to boost the companies' image of being environmental aware and to promote supermarket to their peers.

05

Young Generous Buyers a.k.a. Entrepreneur Ben

Since this customer group is highly populated by youth, stategy mention in Class 1 like importing items that can ***resonate with youths*** and expanding the luxurious items sections like wines could catch their attentions.

THE MARKETING STRATEGY

Appealing to the Millennials and Gen-Z

As we have two major customer cluster(Class 0 & Class 5) that is made up of **Millennials(Age 24-40) and Gen-Z(Age 6-24)**, the mall should make sure relevant action is taken to **instill some element resonance** to appeal these two groups of young customer. As millennials are *experience seeker, eager for self-expression and identity and tech-savvy,* the following are the approach proposed:

- Online Purchasing Platform with Delivery Service
- Diversify Product Sold to Include Youth-Appealing Items *(e.g. Fashionables, Entertainment Systems and Mobile Devices, Sports Peripherals)*

Engage the Mature Customer

From Class 1 and Class 3 we noticed a huge proportional of our customer are made up of **mature buyer** be it those with higher spending score or those with moderate spending score. Their **spending behaviour are relatively consistent** and hence the approach taken should not be more subtle and not too drastic to improve their spending behaviour. The following are the approach proposed:

- Bundle Sales and Rotational Discount
- Coupon Based Campaign/Lucky Draw to encourage Spending and Bulk Purchase by the customer

THANKS!

Personal Learning Journey

I do enjoy the entire process of Unsupervised Learning : Clustering as I found it thrilling to **Discover the Hidden Chest** in the dataset. Although I am aware that this example is in an extremely ideal state, it serves as a good example to demonstrate the clustering process. I also appreciate the multidisciplinary nature of data science that allows me to gain more exposure on devising inclusive and viable business strategy for the Mall Supermarket.