Mall Customer Segmentation and Forming Growth Strategies

medium.com/geekculture/mall-customer-segmentation-and-forming-growth-strategies-cc4130a0f4d7

> Used Python to validate the performance of K-Means, Hierarchical Clustering, and GMMs using the Silhouette score and picked the best model to segment customer information and formed growth strategies for each group

Dataset Overview

This is a mall's dataset from Kaggle, and it has some basic data about the customers such as Customer ID, age, gender, annual income, and spending score.

Problem Statement

I want to increase customer lifetime value by segmenting the customers into several groups with similar characteristics and form growth strategies for each group.

Analysis Process

> 1. Assessing Data
>
> 2. Preprocessing
>
> 3. Choosing the right K and the right model
>
> 4. Naming and plotting clustering results
>
> 5. Analyzing and forming growth strategies
>
> 6. Possible Growth Strategies Summary
>
> **1. Assessing Data**

```
CustomerID                200 non-null int64
Gender                    200 non-null object
Age                       200 non-null int64
Annual Income (k$)        200 non-null int64
Spending Score (1-100)    200 non-null int64
```

Photo by author

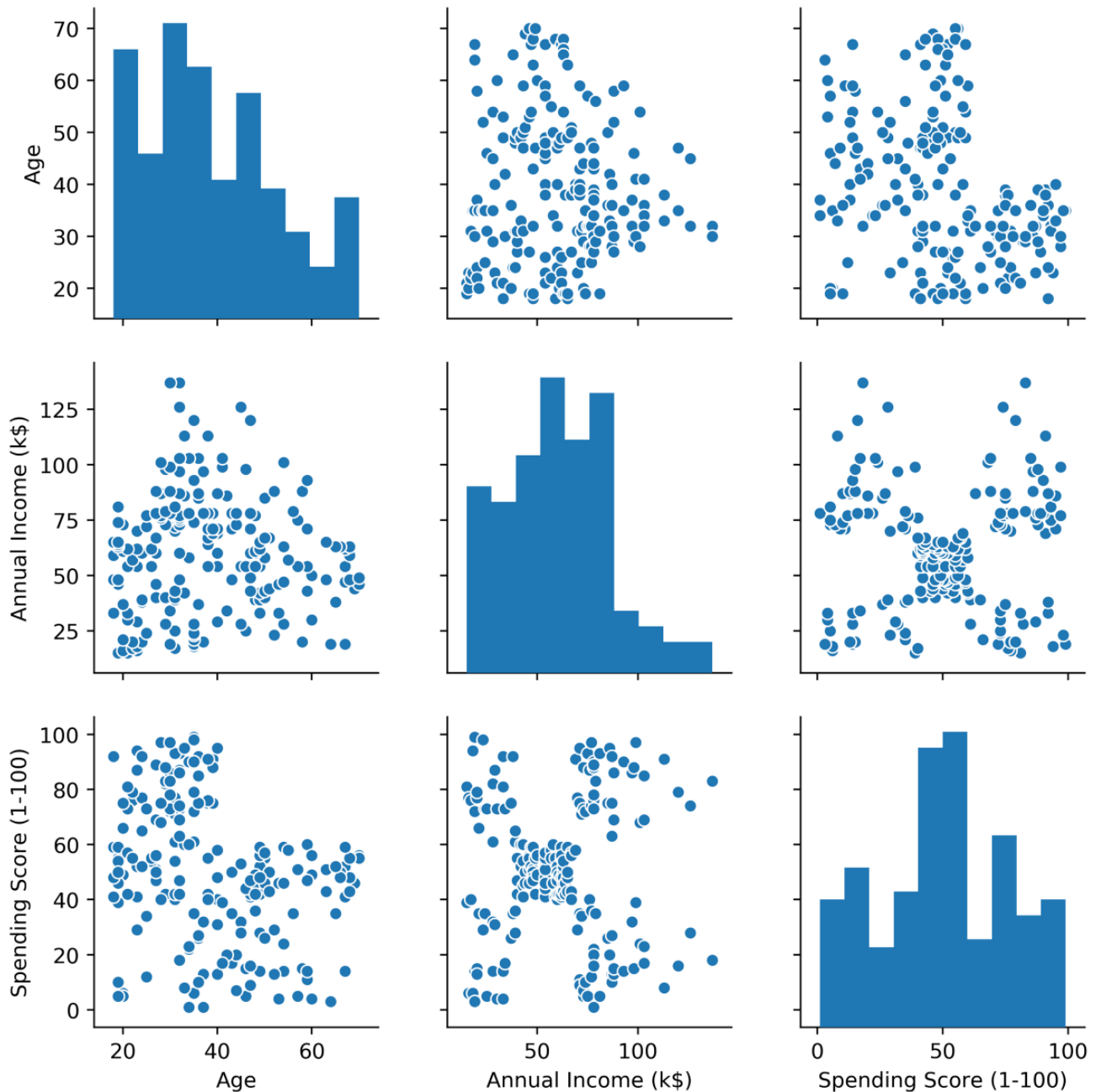This dataset has 5 columns, 200 rows, 0 duplicated rows, and 0 missing values.

Photo by author

Note 1: Age and Annual Income are slightly skewed to the right, and Spending Score is nearly normally distributed.

Note 2: For the scatter plots, there is no clear correlation between features.

Let's look at the correlation heatmap to see the correlation score.

Correlation Heatmap

Photo by author
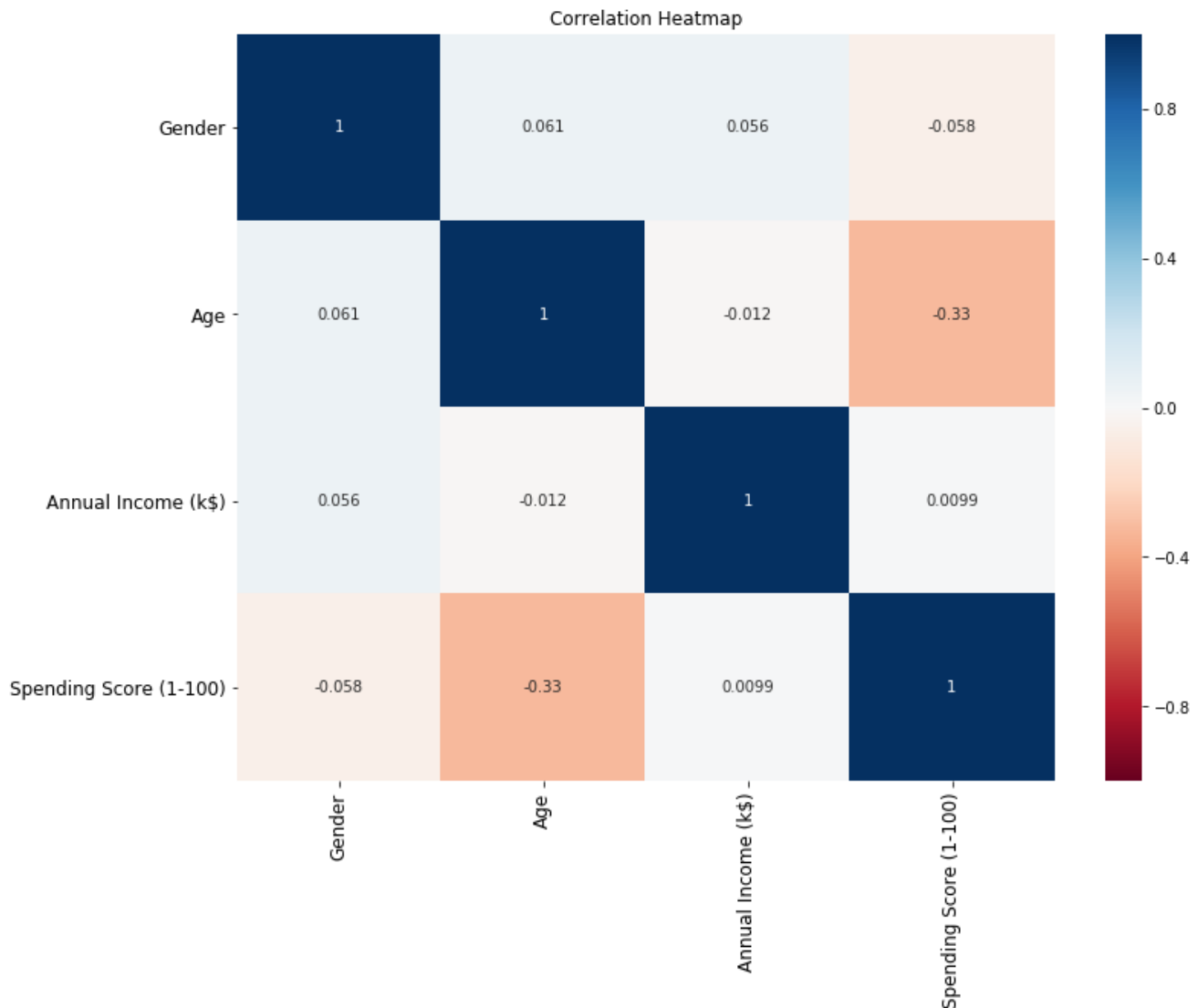
Note 1: The only slightly strong correlation is between Spending Score and Age, 0.33, but they're negatively correlated.

2. Preprocessing

Age and Annual Income are only slightly skewed to the right, so I didn't normalize them. Here, after assessing data, I will:

- **Drop gender since gender is not a continuous feature**
- **Standardize the data**

3. Choosing the right K and the right model

For choosing the right K, I used two methods: The **Elbow Method** and the **Silhouette Score.**

Elbow Method

In the elbow method, you create a plot of the number of clusters on the x-axis vs. the average distance of the cluster's center to each point on the y-axis. This plot is called a scree plot. The average distance will always decrease with each additional cluster center. And, with fewer clusters, those decreases will be more substantial. At some point, adding new clusters will no longer create a substantial decrease in the average distance. This point is known as the elbow.
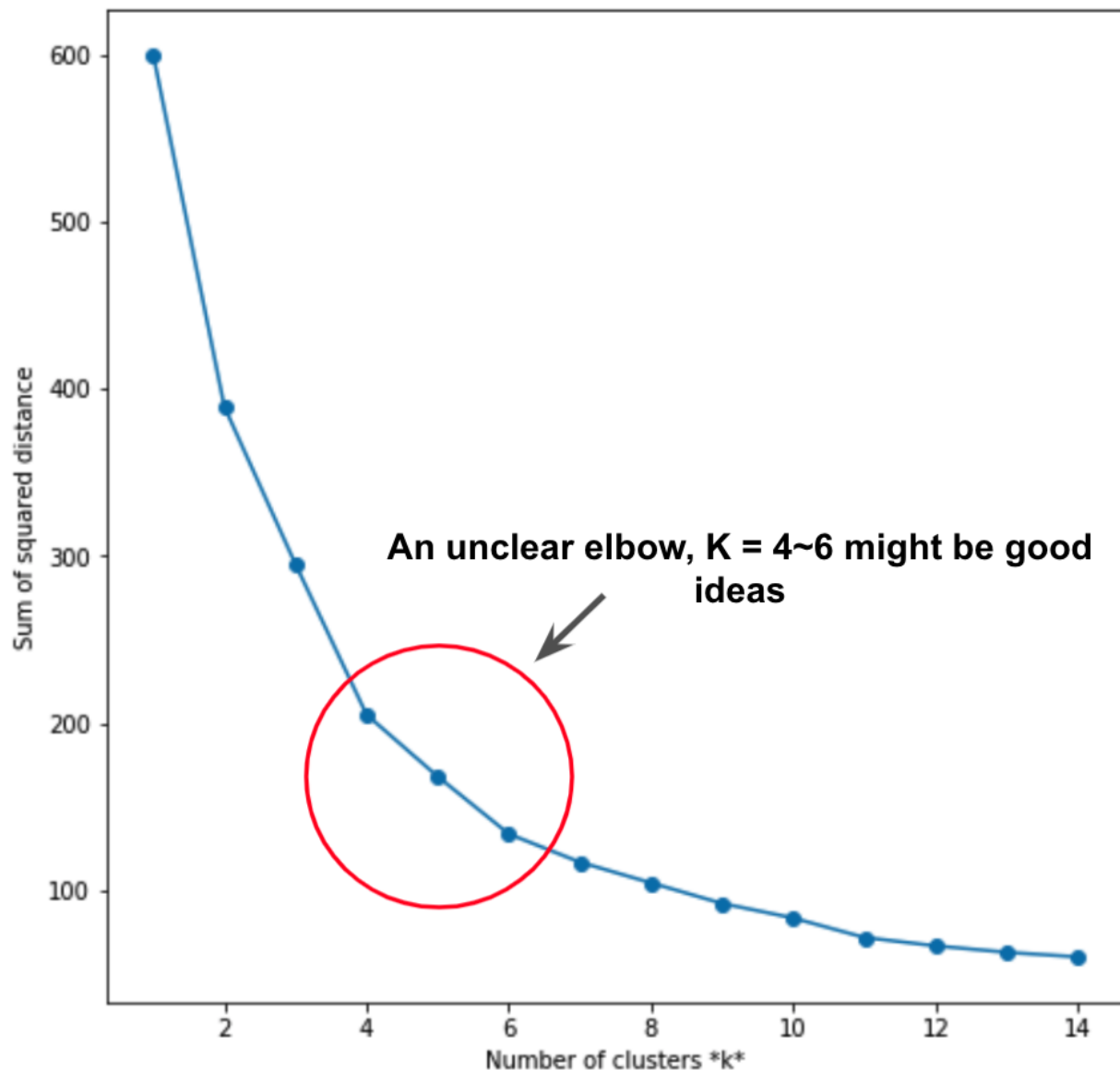
Now, we can see that the elbow method here gives us an unclear elbow, and it seems that K = 4~6 might be good ideas. Therefore, let's try another method, called, **Silhouette Score.**

Silhouette Score

Silhouette scores will compute the average distance from all data points in the same cluster, let's say A. The average distance from all data points in the closest cluster, let's say B. Compute the coefficient, (B -A) divided by the max of a or b if a is bigger than it will be the

denominator, and vice versa. And **the value will be between -1 to 1. The** higher the number, the better the k is.

Here, I want to do something more interesting. **Silhouette Score not only can be used to select the right k but it can also be used to choose the model the performs the best.** Thus, I created the function to plot **KMeans, Hierarchical Clustering, and GMMs**' Silhouette scores in different Ks(The reason I didn't use DBSCAN is that Silhouette score doesn't have the concept of noise, so even though DBSCAN performs well, it may still get a low Silhouette score). In this way, we can see which clustering performs best at which k.
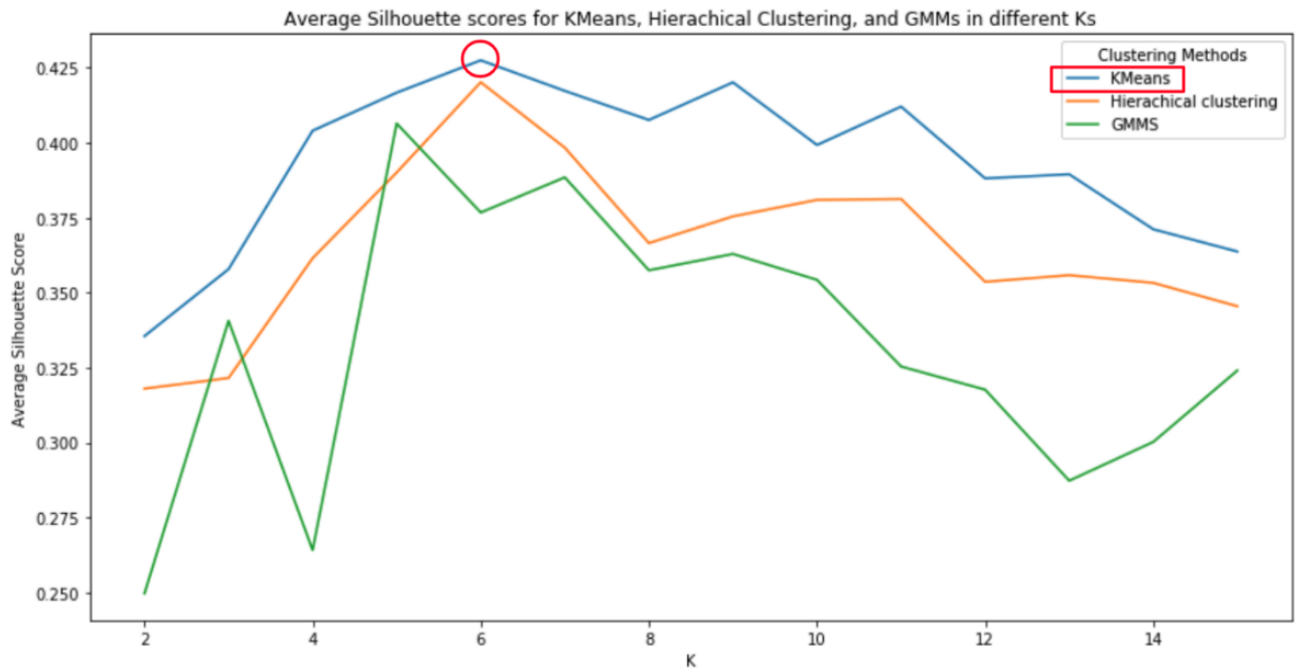

Photo by author

We can see that in this plot:

- KMeans performs best when k=6 and scores 0.428
- Hierarchical clustering performs best when k=6 and scores 0.42
- GMMs performs best when k=5 and score 0.406

Conclusion of model choosing: KMeans with k =6

4. Naming and Plotting Clustering Results

Let's run KMeans with k=6 and calculate the summary statistics of each group:

kmeans_label	Avg Age	Avg Spending Score	Median Annual Income(k)	Cnt	Male%	Female%
5	32.692308	82.128205	79.0	39	0.461538	0.538462
3	25.250000	76.916667	24.5	24	0.416667	0.583333
0	56.333333	49.066667	54.0	45	0.422222	0.577778
1	26.684211	47.789474	60.0	38	0.368421	0.631579
2	45.523810	19.380952	25.0	21	0.380952	0.619048
4	41.939394	16.969697	86.0	33	0.575758	0.424242

Photo by author

After assessing the result, I decided to name each group in this way:

- Group 5: High-income high-spending young age
- Group 3: Low-income high-spending young age
- Group 0: Middle-income mid-spending old age
- Group 1: Middle-income mid-spending young age
- Group 2: Low-income low-spending middle age
- Group 4: High-income low-spending middle age

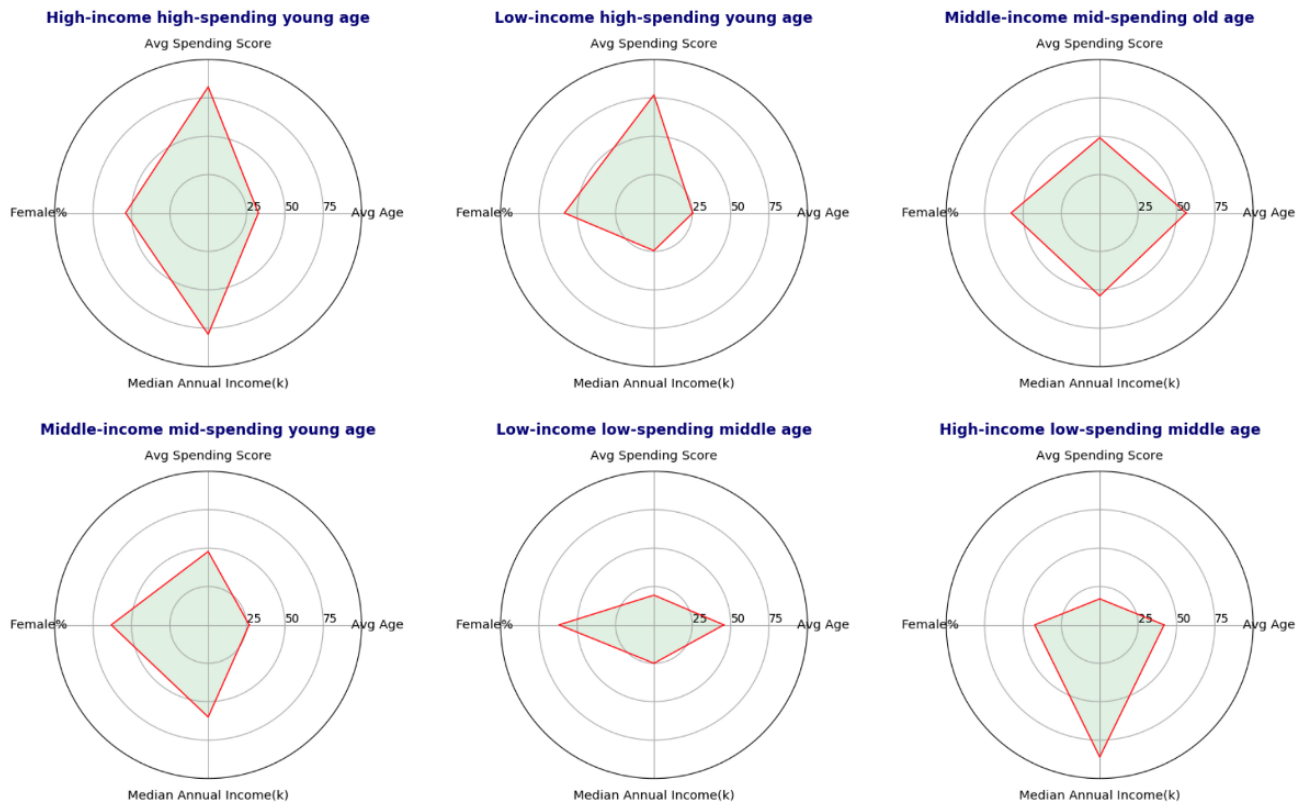Now, let's plot radar charts for each group.

Photo by author

It is clearer to see the characteristics of each group through radar charts. The next step is forming growth strategies, and I will walk you through each group one by one.

## 5. Analyzing and forming growth strategies

Firstly, I'd like to list all the common growth tactics to increase customer lifetime value that I think can be applied to a mall business:

- **New-customer programs**
- **Loyalty programs**
- **Upselling / recommending new or high-priced brands**
- **Referral programs**
- **Incentive program for ready-to-churn customers**
- **Incentive program for winning back lost customers**

And then, we can analyze each group and see what growth strategies listed above can be applied to them.

Group 1

High-income high-spending young age
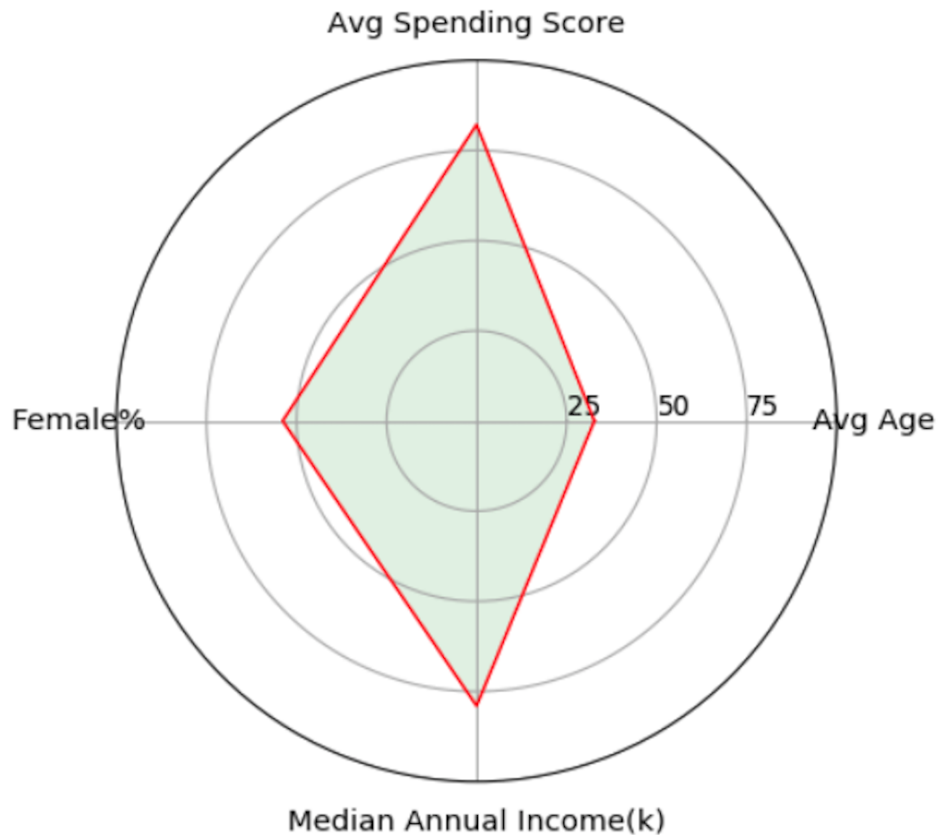
Description

This group consists of rich people spending a lot on the mall. As they are very young, their income can increase even more in the future, so keeping them loyal would be the main strategy.
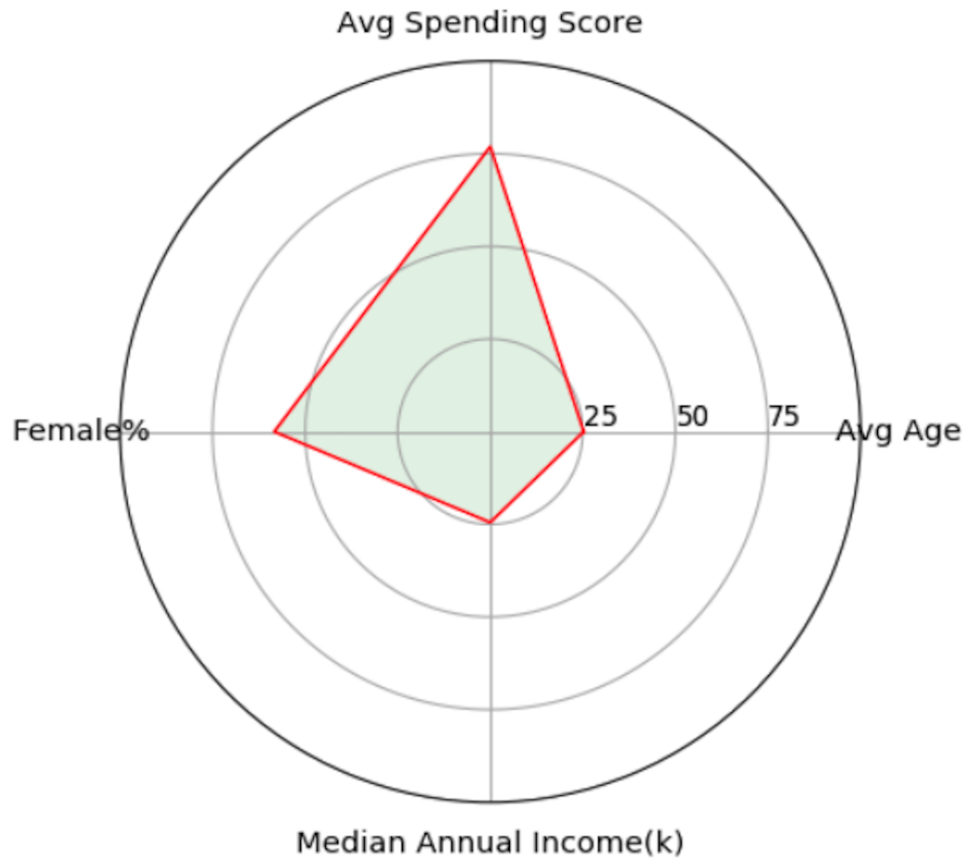
Possible Growth Strategies

1. Loyalty program
2. Up-selling
3. Referral program

Secondly, as they have the ability to spend money, we can also try to sell more high-end brands or new products that they don't know they need.

Lastly, since they had spent a lot in the mall, I assume they really enjoy their shopping experience at our mall. So, asking them to refer their friends to our mall might be a good idea. Besides, their friends might also be rich and high-spend.

Group 2

Low-income high-spending young age

Description

Group 2 consists of the customers with low income but have spent a lot. And they are the youngest among all groups.
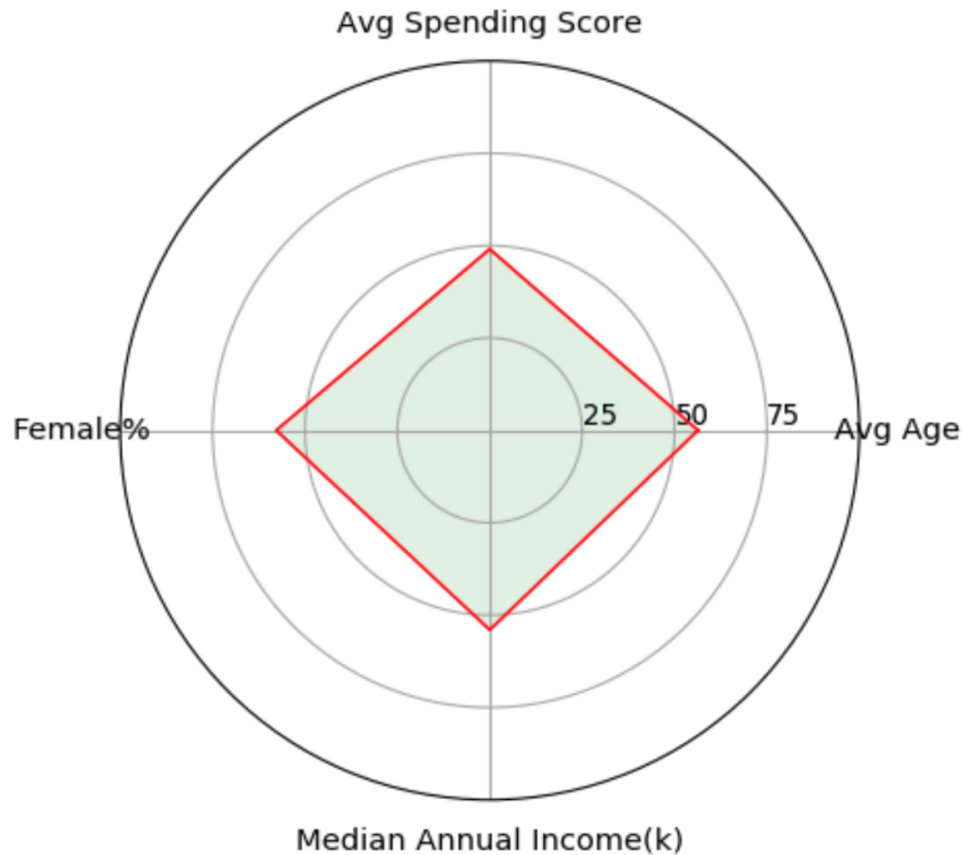
Possible Growth Strategies

1. Referral program
2. Loyalty program

The main strategy for this group is a referral program, and there are three reasons for this: First, we can tell they really love our mall as they are low-income, but spent a lot on us. Second, they might value discounts more than other groups, given their financial situation. And third, they are at a young age, meaning that they are more likely to share things on social media or invite their friends to a product.

Group 3

Middle-income mid-spending old age

Description

Group 3 consists of relatively old customers with middle income and mid-spending records.
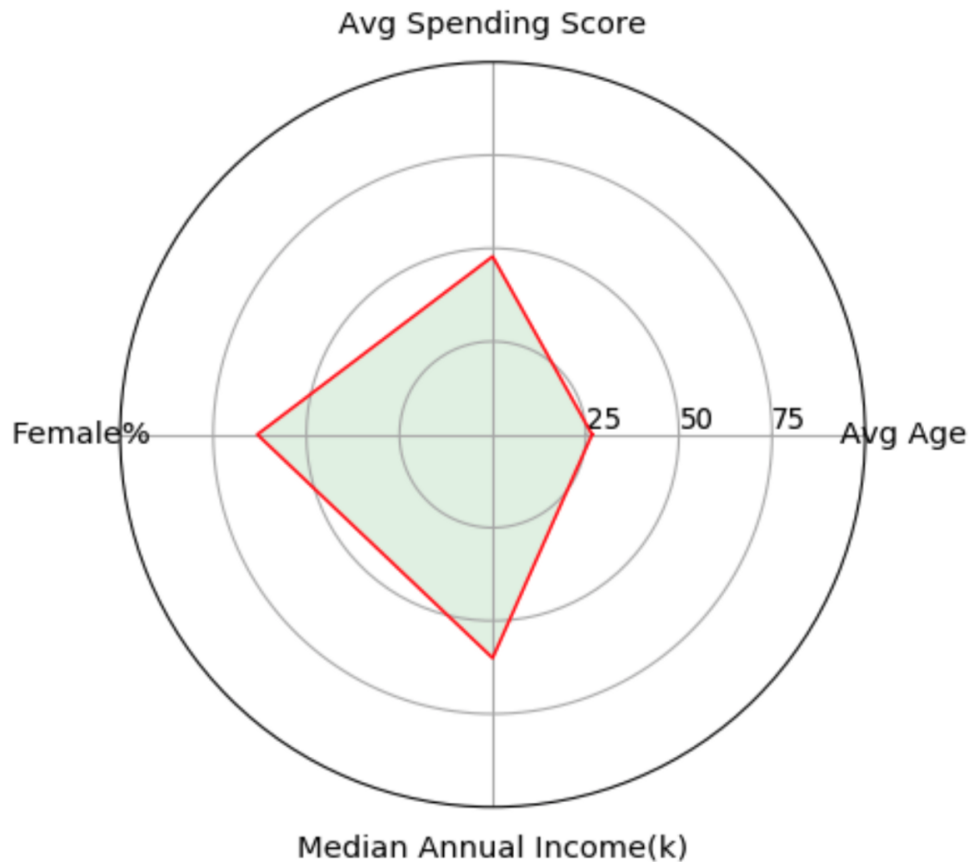
Possible Growth Strategies

1. Loyalty program

Given the limited amount of information, I assume they are long-time customers already, so their purchasing behavior won't be changed that much in the future, so keeping them loyal is this group's main strategy.

Group 4

Middle-income mid-spending young age
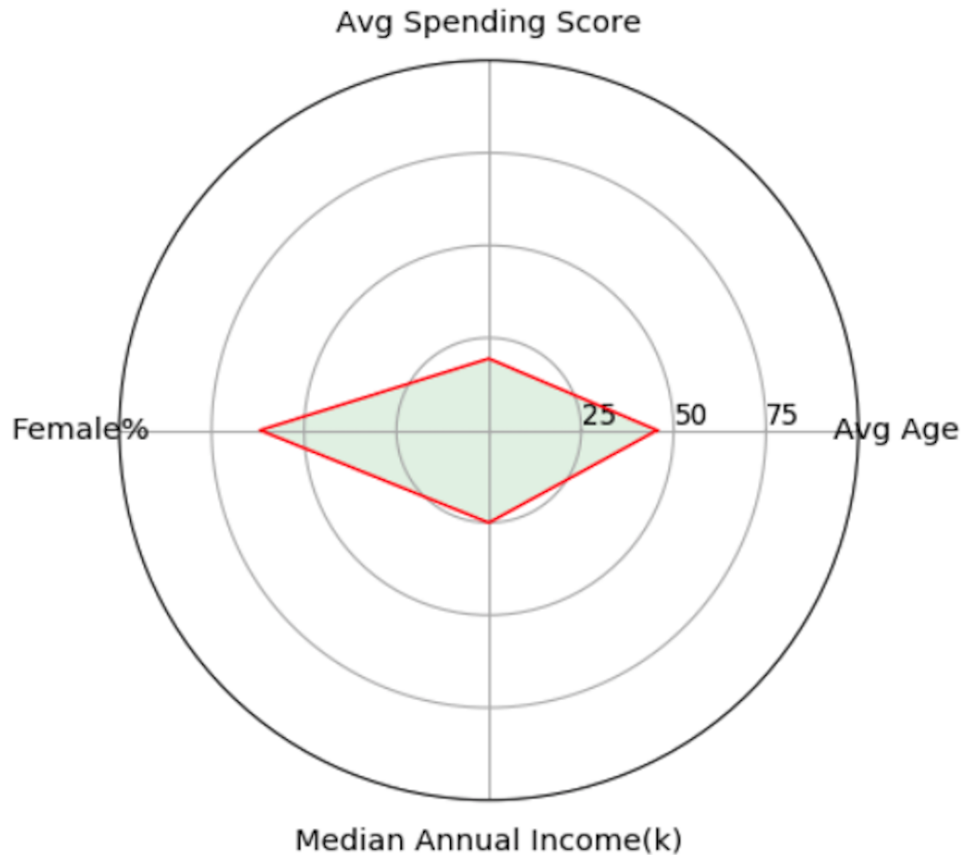
Description

Group 4 is similar to group 3, except the average age is way younger and female customers are more, accounting for 63% of customers in this group. The main strategy is also a loyalty program, but with a focus on female products.

Possible Growth Strategies

1. Loyalty program focusing on female products

Group 5

Low-income low-spending middle age

Description

Group 5 is the least desirable group as they are low-spending and low-income.
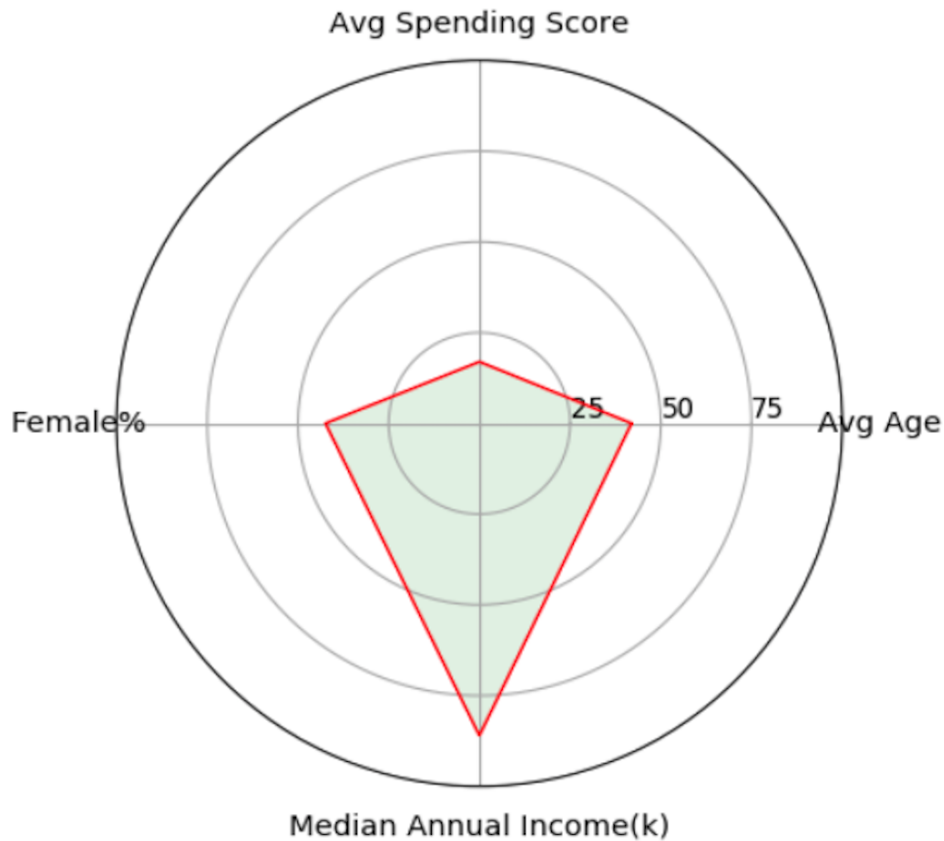
Possible Growth Strategies

None

Further Exploration Direction

Investigating why they spend so little, is it because they're low-income, staying single, or we face new competitors or other reasons?

Group 6

High-income low-spending middle age

Description

Group 6 is a high-income but low-spending group with a male ratio of 58%.

Possible Growth Strategies

1. New-customer program
2. Incentive program for winning back lost customers
3. Recommending high-pricing products

My 3 guesses of why they spend so little are: One, they are just new customers, two, they are lost customers; or three, they mostly shop at a more high-end mall. The reason is worth further investigation.

> 6. **Possible Growth Strategies Summary**
>
> **Loyalty Program: Group 1, 2, 3, 4**
>
> **Referral Program: Group 1, 2**
>
> **Upselling: Group 1, 6**
>
> **New Customer/Lost Customer Program: Group 6**

Let me go through each strategy one by one:

Loyalty Program: Group 1, 2, 3, 4

For a loyalty program, I'd recommend the mall target Group 1, 2, 3, and 4 to maintain or increase customer loyalty. There are several ways to form a loyalty program. A common way is we can build, for example, 3 to 5 levels of membership, the higher the level is, the more discounts and benefits they have, and the status is calculated by monthly or yearly spending amount.

With the incentives like this, we can anticipate the customers in these groups who want to stay in the current status or move to a higher status spend a regular amount of money on our mall, so then, we can keep their loyalty.

Referral Program: Group 1, 2

For a referral program, the mall can target Group 1 and 2, as they are the two most spending groups and are at a young age, meaning that their chances to refer a friend are higher than the older groups. To construct a referral program, I would recommend this hotspot article.

It shows a step-by-step process to build a referral program, including determining what a 'good fit' is for your company, listing possible customer referral sources, identifying channels to host your referral program, etc. Most referral program's incentives are discounts or free products.

Upselling: Group 1, 6

To upsell customers, we can target Group 1 and 6 as they are the highest-income group. For a mall business, the most common channels to communicate and upsell are emails or text messages. We can first explore what kinds of products these customers regularly buy and then run another clustering method to find smaller groups and send each group the promotion about that a more high-end brand of the same product types with a discount.

New Customer/Lost Customer Program: Group 6

Lastly, Group 6 is an unusual group, who spent very little but earned a lot. We can further investigate the true reason that they spent so little. Since they have the potential to increase our GMV, so fixing this group has a high possible return. If most of them are lost customers, we can try using emails or text messages to call them back with discounts on the products that they frequently bought when they were still our customers.

If most of them are new customers, then we can build a new customer program. And the core of new customer programs is to let them understand the value of our mall or our special product as soon as possible. To do this, we can market them our most popular items with

discounts or make them become our highest level of membership for the first month so that they can instantly experience the core benefit of being a valuable customer in our mall.

Works cited:

LaPlante-Dube, Madeleine. "*How to Build a Customer Referral Program*" Hotspot. June 14, 2021. https://brianbalfour.com/landing/p-customer-acquisition-hybrid