

CUSTOMER SEGMENTATION

Maddie Bauer
DSC 680 Applied Data Science
Bellevue University

Table of Contents



The problem



Findings



The Data



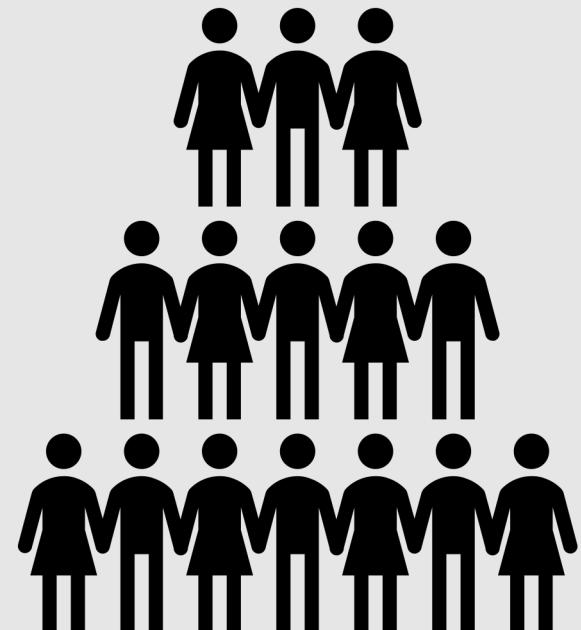
Key Takeaways

The Problem

Businesses are often:

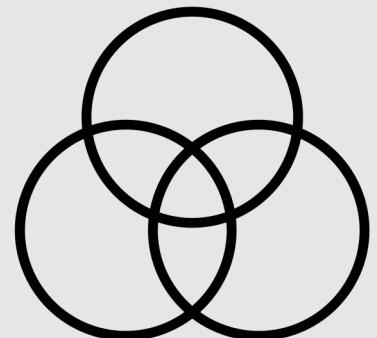
- looking for ways to build relationships & connect with their customers
- wanting to find new ways to save time & money on marketing
- use existing data to their advantage
- create consistencies for customers to better their customer retention & improve sales

Doing all of these things leads to...HAPPY CUSTOMERS!



Customer Segmentation Definition

The process of dividing a customer base into several groups where individuals within each group share similarities in different ways that are relevant to marketing such as gender, age, interests, and miscellaneous spending habits, etc.



Why Segment Your Customers?

Pros

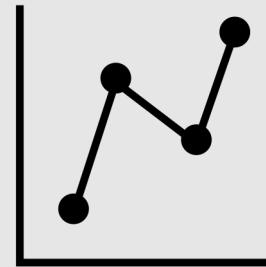
- Customizes marketing campaigns
- Helps determine appropriate product pricing
- Helps you set more specific & measurable goals
- Better customer retention
- Choose specific product features for deployment
- Prioritize new product development efforts
- Design an optimal distribution strategy

Cons

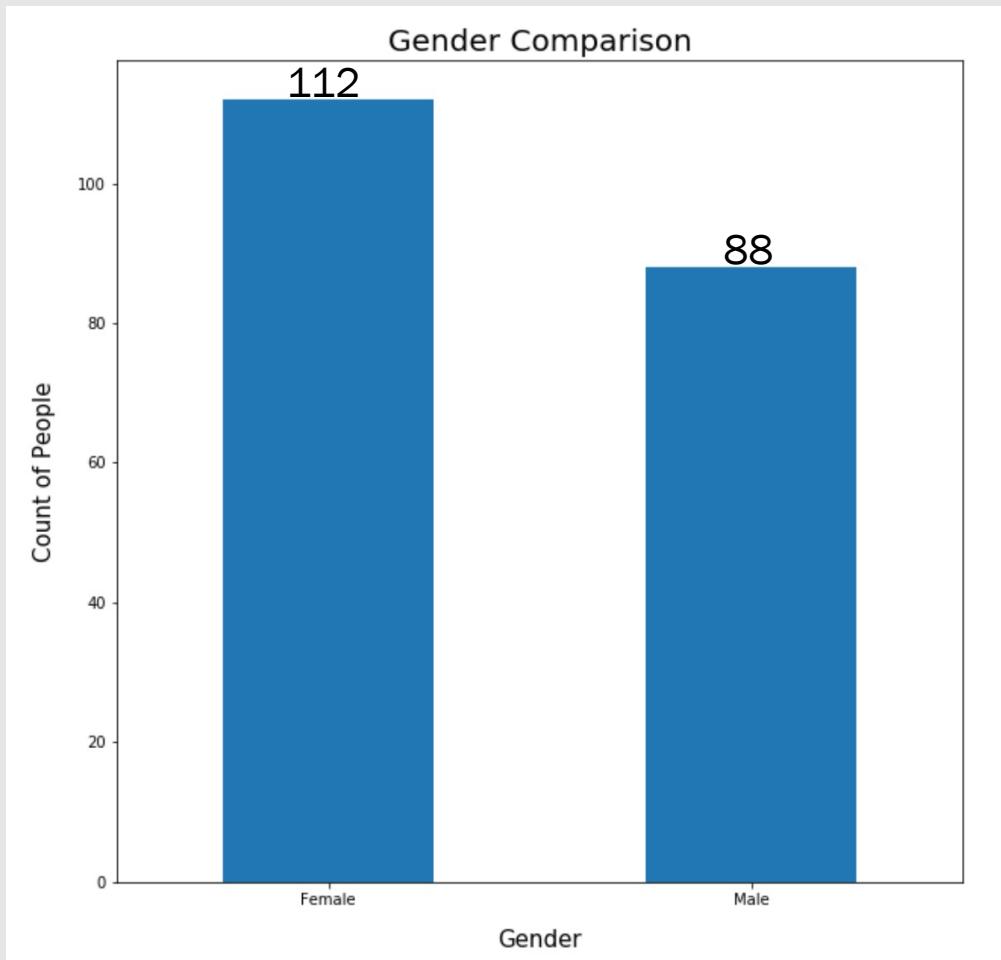
- Can be expensive (research & surveys)
- Developing customer profiles can be time-consuming
- There's a chance you might miss a customer segment – resulting in loss of sales for a specific group

The Data

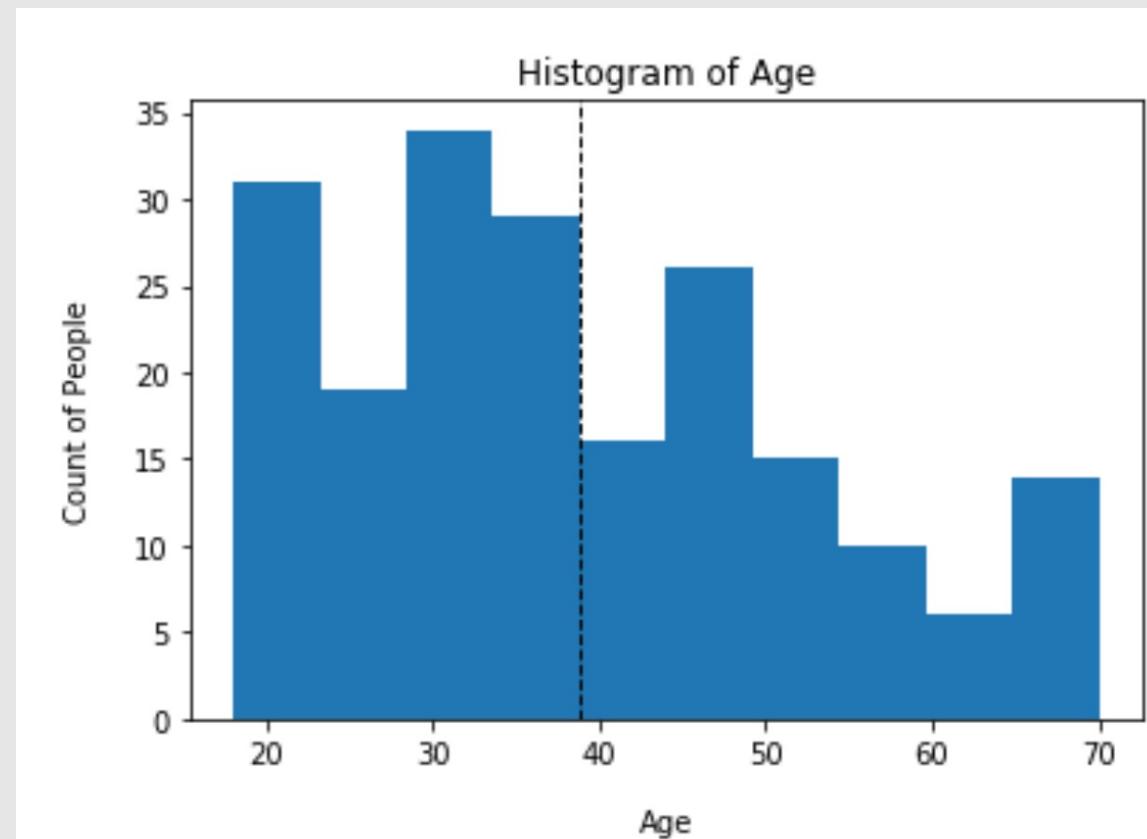
- 200 customers
- Gender – Only male or female are possible values in this dataset
- Age – Customer's age (numeric value)
- Annual Income – Numeric value in thousands (ex. 40 = \$40,000)
- Spending Score – Range of 0-100 that is assigned by the mall based on the customer's behavior and previous spending nature



Gender

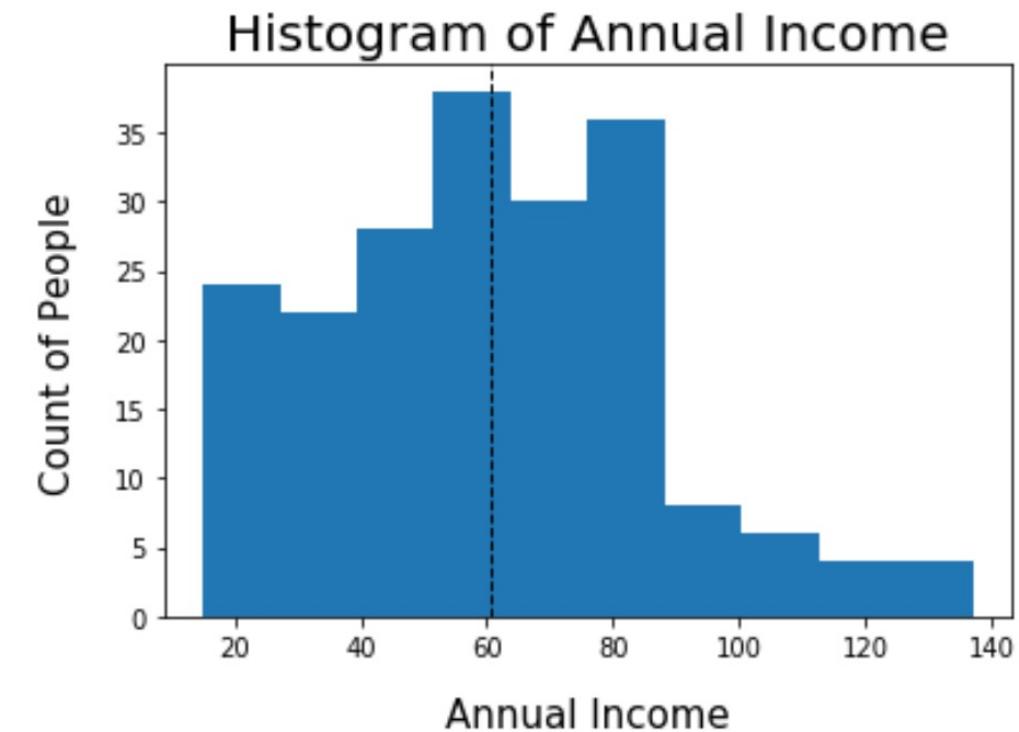


Age



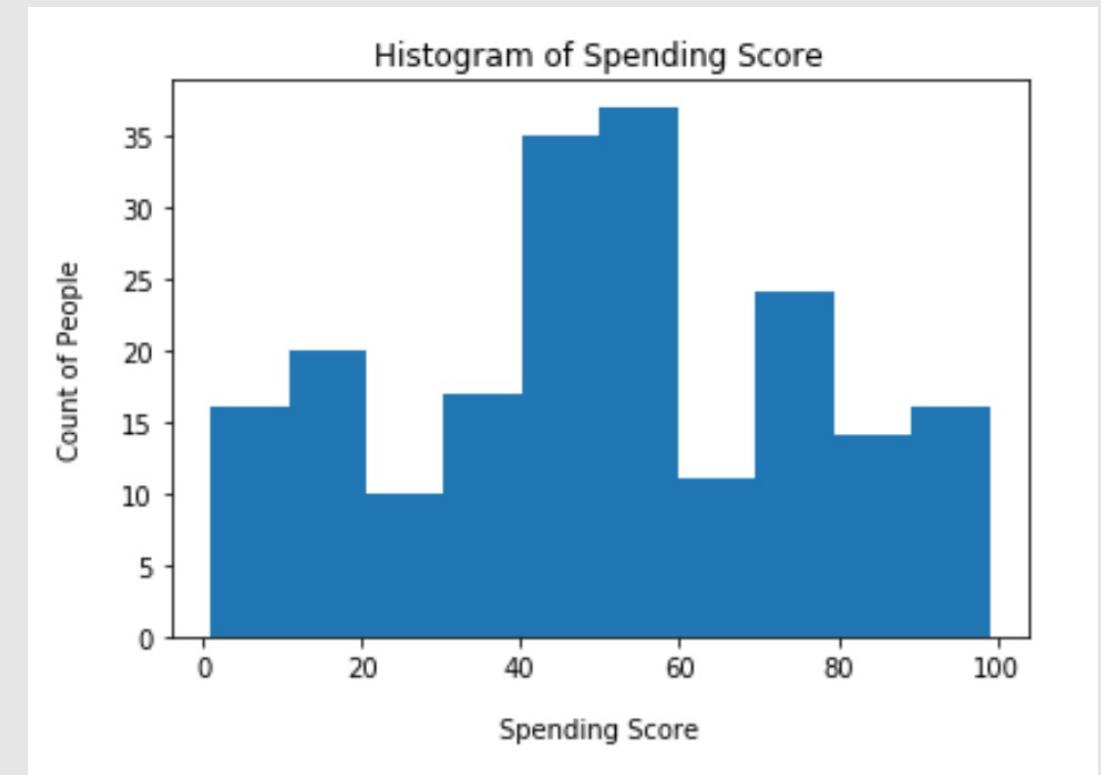
Mean Age: 38.85

Annual Income



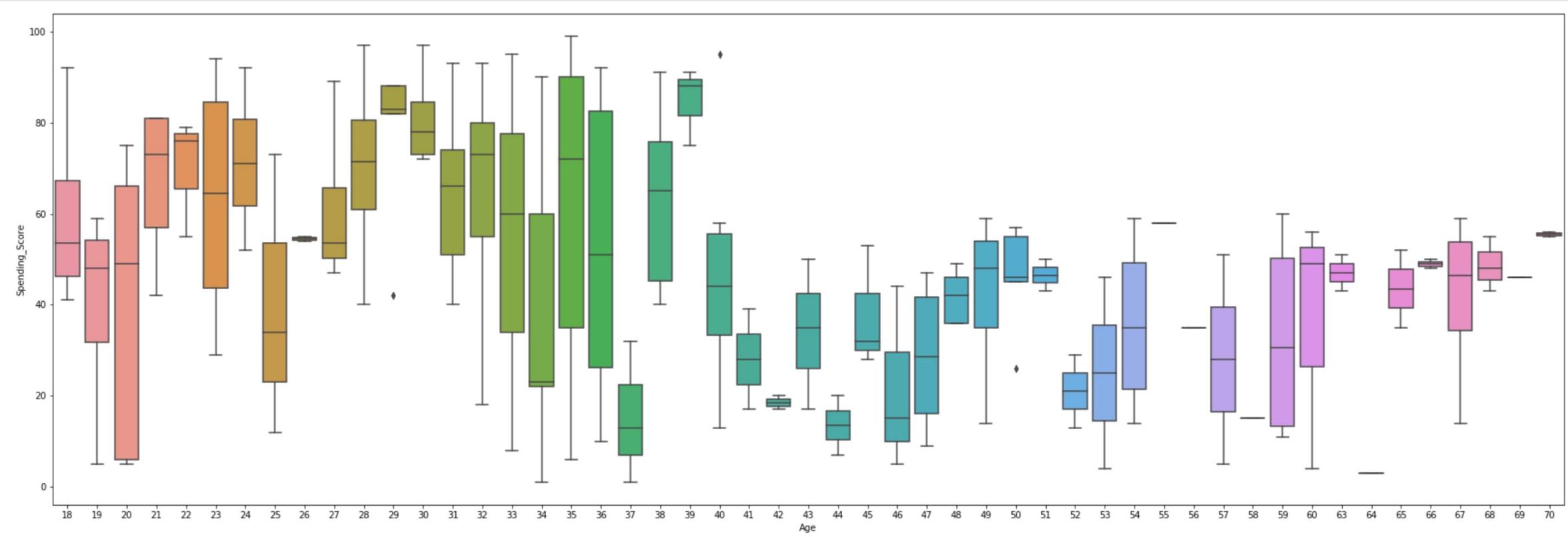
Mean Income: 60.56

Spending Score

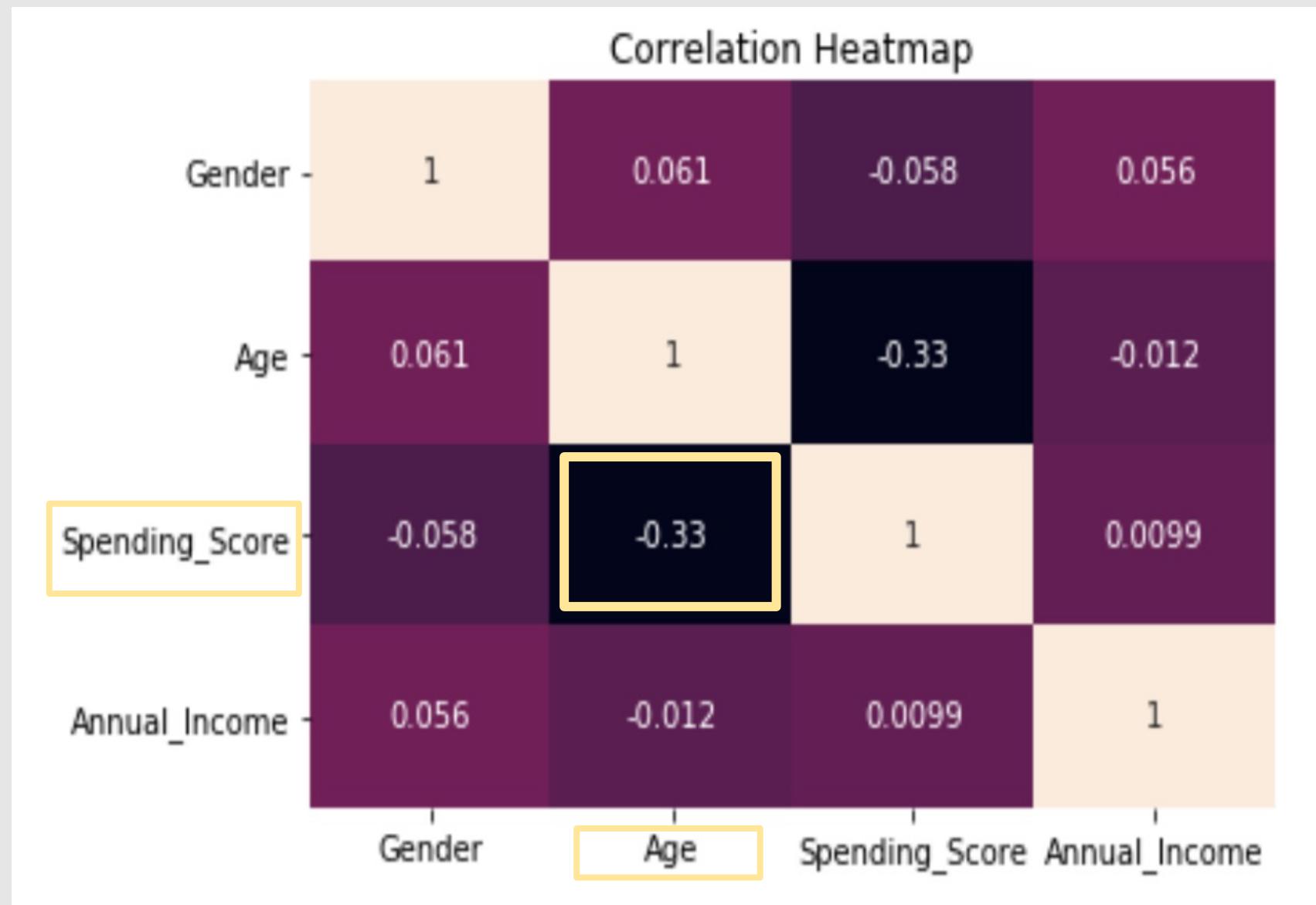


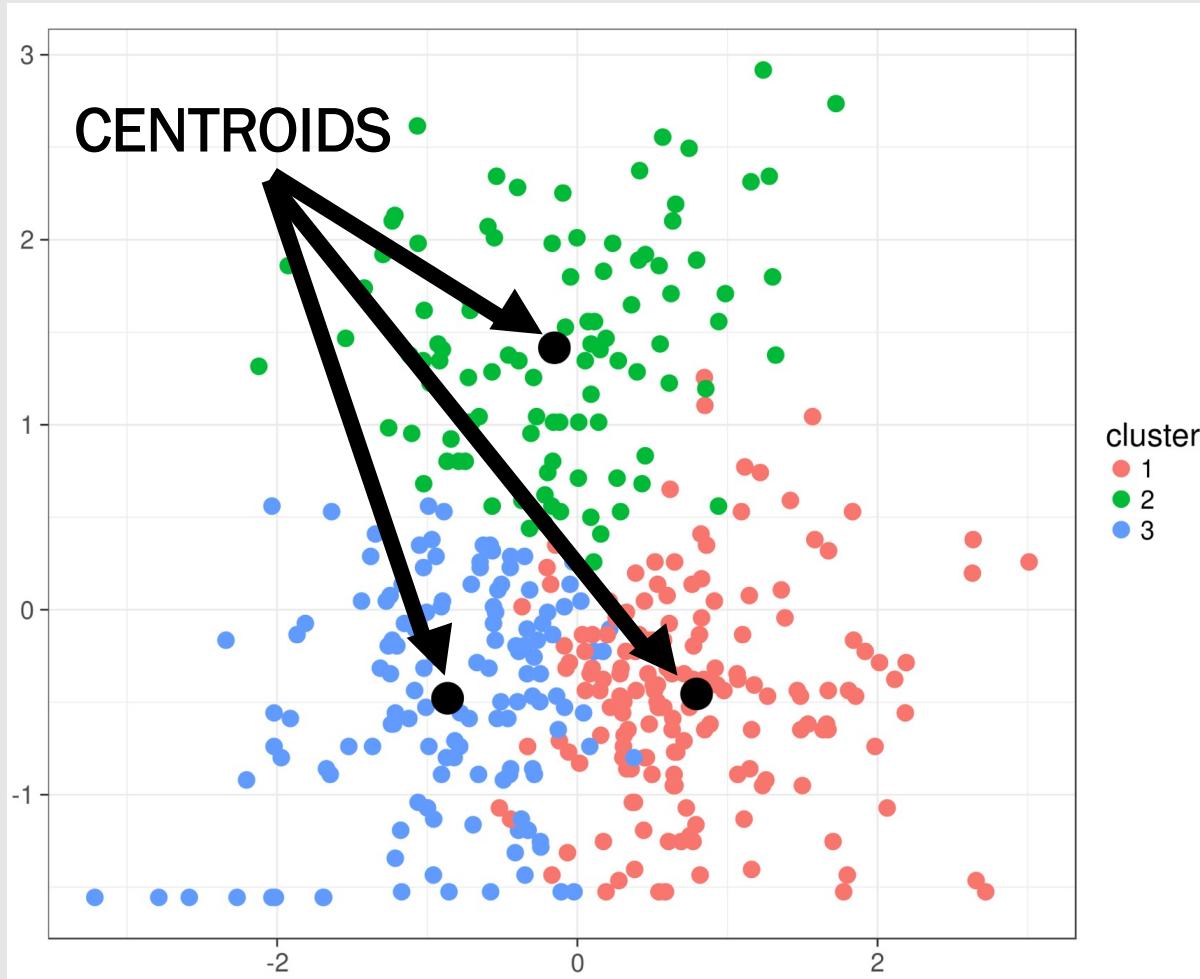
Mean Spending Score: 50.2

Spending Score Spread by Age



- Ages 34, 35 & 36 have a wide spread of spending scores.
- Ages 55, 56, 58, 64, 66, 69, 70 have almost no spread at all.
- Bulk of spending happens Age 40 and younger



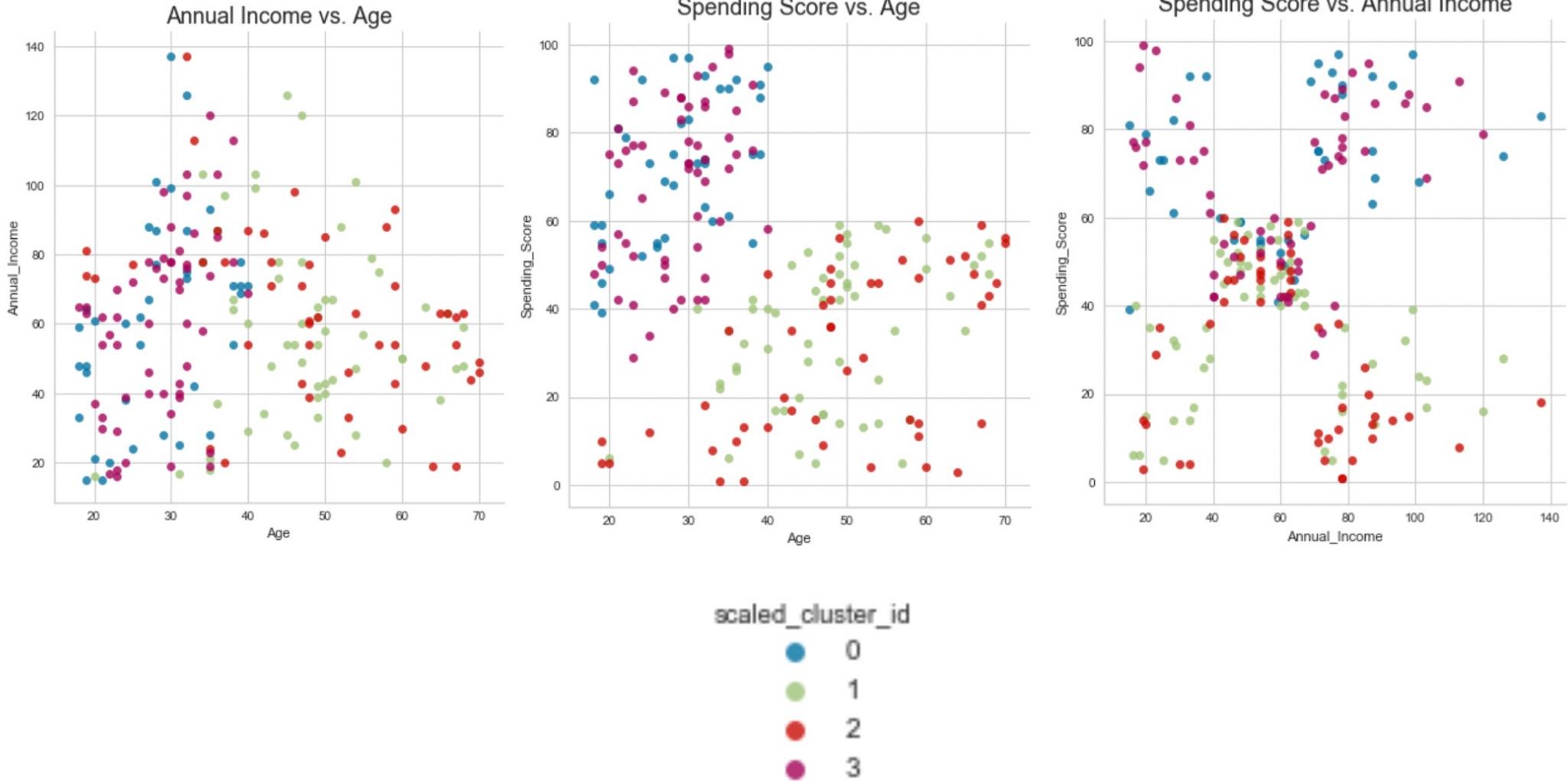


Source: <https://rpubs.com/cyobero/k-means>

K-Means Clustering

- Unsupervised machine learning technique
- Groups similar data points together to discover underlying patterns
- The number of clusters is denoted by k
- Each cluster has a center, called the centroid
- Every data point is assigned to a cluster by its relative location to the nearest centroid
- Once clusters are formed, you can analyze the characteristics of each one and determine new patterns/trends

Findings – 4 Clusters



Findings

Cluster Number	Number in the Cluster	Age (Mean)	Annual Income (Mean)	Gender (Mean)	Spending Score (Mean)
0	40	28.25	62.000	Male	71.675
1	56	47.80	58.071	Female	34.875
2	48	49.45	62.417	Male	29.208
3	56	28.34	60.429	Female	68.179

Cluster 0 – Highest spending score, younger in age, varying income levels, more males

Cluster 1 – Moderate/Low spending score, older in age, varying income levels, more females

Cluster 2 – Lowest spending score, older in age, varying income levels, more males

Cluster 3 – High spending score, younger in age, varying income levels, more females

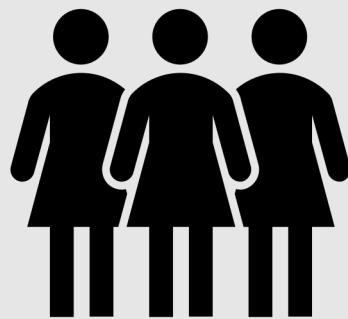
Key Takeaways

- 4 Clusters/Groups of Customers were determined
- Age & Spending Score help determine cluster more so than Gender & Annual Income
- Most Valuable to Mall – Clusters 0 & 3
- Least Valuable to Mall – Clusters 1 & 2



Conclusion & Next Steps

- Customer segmentation is a tool that companies can utilize to organize their audience into meaningful groups to customize the overall consumer experience.
- Has many benefits!
- We can now use these attributes and information to start customizing our marketing strategies for each of the 4 clusters.
- Re-run clusters quarterly to determine with any customers change clusters.
- Leads to brand loyalty, great consumer retention and higher profits for the mall.



THANK YOU!

References

- Brownlee, J. (June 30, 2020). *Why one-hot encode data in machine learning?* Retrieved June 24, 2021 from <https://machinelearningmastery.com/why-one-hot-encode-data-in-machine-learning/>
- Choudhary, V. (n.d.). *Mall customer segmentation data*. Retrieved June 12, 2021 from <https://www.kaggle.com/vjchoudhary7/customer-segmentation-tutorial-in-python#>
- Dabbura, I. (September 17, 2018). *K-means clustering: Algorithm, applications, evaluation, methods, and drawbacks*. Retrieved June 24, 2021 from <https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a>
- Dataflair.training. (n.d.). *Data science project – Customer segmentation using machine learning in R*. Retrieved June 23, 2021 from <https://data-flair.training/blogs/r-data-science-project-customer-segmentation/>
- Garbade, Dr. M. (September 12, 2018). *Understanding K-means clustering in machine learning*. Retrieved June 24, 2021 from <https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1>
- K-Means Clustering. (June 22, 2021). In Wikipedia. Retrieved June 24, 2021 from https://en.wikipedia.org/wiki/K-means_clustering
- Lintern, M. (2013). To segment or not to segment: We weigh the pros and cons. Retrieved June 28, 2021 from <https://blogs.oracle.com/marketingcloud/post/to-segment-or-not-to-segment-we-weigh-the-pros-and-cons>
- Sagar, A. (n.d.). *Customer segmentation using k means clustering*. Retrieved June 23, 2021 from <https://www.kdnuggets.com/2019/11/customer-segmentation-using-k-means-clustering.html>