Customer Segmentation

Maddie Bauer
DSC 680-T301 – Summer 2021
https://madelinebauer.github.io/MBauer/

Which Domain?
Any company who sells products or services – physically or virtually – can benefit by segmenting their customer base. Knowing your audience is the first step in helping to sell your products and/or services successfully. Segmentation is an organizational strategy that companies use to break down their audience into smaller, more manageable groups. Customizing these groups based on similarities – such as by demographics, location, behaviors or by interests - helps to create an overall better customer experience that leads to loyalty, brand recognition and ultimately higher profits. This project will specifically be focused on customer segmentation within the retail industry.

References

https://www.cmgconsulting.com/post/market-segmentation-strategy - Understanding the benefits of customer segmentation

https://www.kdnuggets.com/2019/11/customer-segmentation-using-k-means-clustering.html - Customer segmentation with Python (k-means) example

https://seifip.medium.com/starbucks-offers-advanced-customer-segmentation-with-python-737f22e245a4 - real world company (Starbucks) segmentation example

https://www.mktr.ai/applications-and-methods-in-data-science-customer-segmentation/ - Customer segmentation background knowledge, pros/cons, and Python examples

https://www.kaggle.com/uciml/german-credit - Similar example but within banking industry

https://bdtechtalks.com/2020/12/28/machine-learning-customer-segmentation/ - More information about k-means and how to apply it

https://en.wikipedia.org/wiki/K-means_clustering - Learning more about K-means clustering

https://www.analyticsvidhya.com/blog/2020/10/a-simple-explanation-of-k-means-clustering/ - Learning more about K-means clustering

Which Data?
Data Link: https://www.kaggle.com/vjchoudhary7/customer-segmentation-tutorial-in-python#

This data set is relatively small consisting of 200 customers. It was created for a Kaggle competition to help learn about customer segmentation. The idea is that basic data was collected for mall shoppers through the use of their membership cards. The goal is to analyze the variables (Gender, Age, Annual Income and Spending Score) to determine marketing strategies moving forward. Below is a little bit more information of each variable.

Gender – Only male or female are possible values in this dataset

Age – Customer's age (numeric)

Annual Income – Numeric value in thousands (ex. 15 = 15,000)

Spending Score – Range of 0-100 that is assigned by the mall based on the customer's behavior and previous spending nature

Research Questions? Benefits? Why analyze these data?
1. Are there customers with similar spending patterns?

2. Does age, gender and/or income affect a customer's spending score?

3. Can we determine which cluster/group of customers are not so valuable, valuable, and very valuable? (rank)

4. Based on the clusters found, what are the next steps for targeting each group?

5. Can a customer shift to a different cluster if their spending score changes?

What Method?
I plan to start with EDA to dive into each variable. I will look to see how spending score is possibly affected by age, gender, and income. I then plan to take a clustering approach with this data. Specifically, I will use K-Means clustering. I will research the elbow method and silhouette method to determine which one is the best approach for me in determining the right k value.

Potential Issues?
I have only implemented K-Means clustering once before in a prior class. I will need to spend time researching and implementing different methods of determining how many clusters to create. I have come across the elbow method, gap statistic and silhouette method so far in my research for determining the correct k value. Another issue might be if there are any outliers. Outliers can significantly disrupt the work of K-means, so I'll need to research about how to deal with outliers if my dataset does in fact contain any.

Concluding Remarks
All retail companies, large and small, can benefit by grouping their customers into smaller subgroups for the benefits of providing targeted marketing, advertisements, coupons, etc. Customer segmentation is a tool that companies can utilize to organize their audience into meaningful groups to customize the overall consumer experience. By tailoring a customer's experience, you make them feel welcomed, appreciated, heard, etc. which leads to brand loyalty, great consumer retention and higher profits for the company.