Bill_Andrews_Projects

Customer Segmentation

Business Problem

Customer segmentation is the market divided into various unique consumer groups with related characteristics. Market segmentation is a valuable tool for identifying and satisfying customer needs. Market Basket Analysis was conducted using K-Means Clustering Algorithm and Unsupervised Machine Learning Techniques. To identify the target customers that can be quickly converged, a market basket analysis is done. To enable the marketing staff to prepare a plan for promoting new products to target customers who share their interests. Many businesses must segment their customer bases to generate future projections, create various marketing plans, or implement customer-specific marketing initiatives.

Data Explanation

The data set is of 200 mall customers. It includes demographic variables such as Customer ID, Age, Gender, Annual Income, and Spending Score. The variables Customer ID and Gender were removed as they were determined not to be helpful in customer segmentation. The mean of the variables is Age 38.85, Annual Income of $60.56K, and Spending Score of 50.20. The minimum age of the customer in the data is 18 years, and the maximum age is 70. The minimum annual income of a customer is 15k$, and the maximum is 137k$. A Spending Score is assigned to the customer based on defined parameters like customer behavior and purchasing data. Here the minimum spending score assigned is one, and the maximum ranges to 99. Between 40 and 60 has the highest population of Spending Scores. The Ages between 20 and 40 have the highest Spending Scores. The Ages between 30 and 40 have the largest population. The amount between $60K and $80k has the highest population of annual income. The Ages between 30 and 50 have the highest annual income. None of the data types had to be changed, and there were no missing values to populate.

Methods

K-Means Clustering algorithm used to cluster data. To implement K-Means clustering, we need to look at the Elbow Method. The Elbow method is used within cluster analysis to determine the appropriate number of clusters in a dataset. It is clear from the figure that we should take the number of clusters equal to 5, as the slope of the curve is not steep enough after it. The data(clusters) are plotted on a spending score Vs. Annual income curve.

Analysis and Conclusion

This Clustering Analysis gives us a clear insight into the different segments of customers in the mall. There are five segments of Customers when using Annual Income and Spending Scores. It determined that they were the best variables to use to establish customer segments in a mall. Once we have the results, we can use different marketing strategies to capitalize on the

spending scores of the customer in the mall. The marketing department no longer has to use a one size fits all campaign for all of the customers. The breakdown of the clusters into various customer personas is as follows:

Cluster 1: Customers with a lower annual income and lower spending score. These customers can be identified as “The Budget-Minded Customer.” Customer Segmentation 5 Cluster 2: Customers with a higher annual income and lower spending score. They are “More Marketing Focus” customers. Cluster 3: Customers with a middle annual income and middle spending score are the “Average Joes.” Cluster 4: Customers with a lower income and higher spending score - “Carefree shoppers.” Cluster 5: Customers with a high income and high spending score - “Frequent Shoppers.”