The following is a summary of my data analysis:

**Background:** The dataset used for my analysis contains data from real transactions that took place between January 12, 2010 and September 12, 2011 at a UK based online retail store. The online store sells gifts for any occasion and its customers are mostly wholesalers. The data can be found at <https://archive.ics.uci.edu/ml/datasets/online+retail>.

**Objective/Problem:** The company’s marketing team wanted to understand the behavior of their customers so that they could determine the customers to target. The problem was that they did not have a way of grouping similar customers together so that they could make that determination.

**Approach:** My analysis involved helping the marketing team with learning their target customers by using a machine learning technique to group customers based on their behaviors. The machine learning technique I used to group customers was k-means clustering. K-means clustering is an algorithm that places data into a specified number of clusters based on certain attributes. The attributes I used for the algorithm were frequency of purchases and total amount spent between January 12, 2010 and September 12, 2011. The frequency was calculated based on a count of the invoices of each customer. The amount spent was calculated by taking the total quantity and multiplying it by the unit price and then summing all the results to get the aggregate amount. I determined that the optimal number of clusters for the algorithm was four based on my use of the Elbow method. Then, the data was put through the k-means algorithm and returned the four specified clusters (See **Figure 1** in **Appendix**).

**Conclusion:** After analyzing the four clusters produced by the k-means algorithm, I found cluster 4 customers to be the highest spenders with low purchase frequency; they make very large purchases sporadically. Cluster 2 customers are the lowest spenders with the lowest purchase frequency; they make small purchases very sporadically. Cluster 3 customers have the highest frequency and a high amount spent; they purchase very regularly and spend a large amount of money. Cluster 1 customers have a low frequency of purchases and a low amount spent; they make small purchases sporadically. Based on the characterization of each cluster, cluster 3 consists of the company’s target customers. The marketing team should focus more attention on cluster 3 since these customers purchase most often and spend a good amount of money on the company’s products. The least attention should be given to cluster 2 customers because these customers are not the company’s target.

**Appendix**

**Chart, scatter chart

Description automatically generatedFigure 1**