

A dataset containing order information as well as value of residence and age of the customer was provided. The goal of this analysis is to incentivize frequent purchasing habits within customers by providing a \$5 coupon to the 'best' customers and a \$10 coupon for the 'worst' customers. A dendrogram was initially used to guide the number of clusters set for the analysis. It was cut at a height of 190 and yielded 5 clusters with the smallest cluster having 25 members. Upon reviewing the clusters, it is evident that there is a clear "Big Spender". Cluster 3, with an average age of 34.4, has an Average Order Value of \$74.5, and also has the most number of Items in Cart, with 6.15 items on average. On the other hand, the cluster with the lowest spend is cluster 4, which has an average age of 40.3, has an Average Order Value of \$56.3, and has also has the fewest Items in Cart, with 4.30 items on average. This intuitively makes sense, as the number of items in cart is likely going to increase average order value. Interestingly, Residence Value does not seem to be highly correlated with average order value, which tells us that there are more important factors affecting Average Order Value. Cluster 3, the best cluster, could tell us a bit more about the customers. The data suggests that they could be a young working families with a new born baby, which could explain the higher number of items bought. On the other hand, the worst performing cluster could be customers who may not cook at home much. They seem to be quite well off, and may eat out at restaurants more frequently.

When using the k-means method to extract clusters, the elbow plot suggests 3-5 clusters. Given that the goal of this analysis was to find the best and worst performing clusters, I chose 5 clusters to maximize differences between the clusters. The results from the k-means generation seem to be in-line with the initial results from the dendrogram. A budget of about \$1,245 would be prudent for this promotion, as cluster 3 (best) has 87 members, and cluster 4(worst) has 81 members.

```
> segment %>% group_by(cluster) %>% summarise_all(funs(mean(.)))
# A tibble: 5 x 5
  cluster  Age AverageOrderValue ResidenceValue NoofItemsinCart
*   <int> <dbl>           <dbl>           <dbl>           <dbl>
1     1  35.9             63.6             163.             4.97
2     2  42.5             64.2             323.             5.07
3     3  34.4             74.5             467.             6.15
4     4  40.3             56.3             572.             4.30
5     5  36.4             62.2             741.             4.84
```





