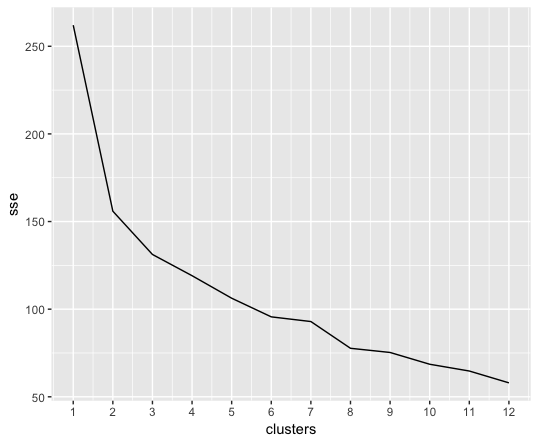
Assignment 3

Sam Robinson

Graph of SSE clusters based on number of clusters selected. I ended up selecting 6 clusters because of the diminishing returns after 6.



Question 1:

In the first test (Home => Search => Prod\_B) there was a 47.83% correlation with also viewing product A. However, that did not end up leading to a sale.

In the first test if we also assume that this person inevitably purchased then it still makes sense for the person to be shown product A.

In the second test (Home => Prod\_C) there was an 80% correlation with also viewing product A and only a 20% correlation with viewing product B

Like above if we assume that this person also ended up making a purchase than there is a 50% correlation with viewing product A in both scenarios.

Question 2:

We can achieve this goal by adding more weight to people who only use the search function and we can alternatively add a column for not using the search function and adding weight to that column as well. This would let us get clusters that are more based on not using search or using search. Alternatively we can add more weight to people viewing multiple columns by creating a column that sums viewing A/B/C but ends up as a zero if they used the search function I.E. (sum(Prod\_A, Prod\_B, Prod\_C) \* search)

When looking at the data generated from creating a window shopper feature there is not a strong correlation between window shopping and purchasing.

Window shopping score is the mean of the window shopping column

|  |  |  |
| --- | --- | --- |
| Cluster | Purchase Percent | Window Score |
| 1 | 42% | 1.79 |
| 2 | 53% | 1.32 |
| 3 | 50% | 1.62 |
| 4 | 50% | 0 |
| 5 | 25% | 2.38 |
| 6 | 6% | 0 |

Truth be told this is not how I would approach this problem, which was also mentioned in class.

Option Question:

In my 6 clusters, I did not see any two products correlate strongly. I find this more interesting than products having strong relationships. This means to me that the 3 products are very distinct from each other and people looking for one of them may not be looking for the other ones.

Among clusters with strong purchasing habits there was a strong correlation with avoiding the home screen and not viewing other products as well. This leads me to believe that A must be the sites strongest product and that some people visit this site specifically to find prod A.

Here are the centroids for my clusters.

| **cluster** | **home** | **products** | | **search** | **prod\_A** | **prod\_b** | **prod\_c** | **cart** | **purchase** | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | |  |  |  |  |  |  |
| **1** | 1 | 0.86 | 0.79 | | 1.00 | 0.57 | 0.00 | 0.71 | 0.93 | 0.50 |
| **2** | 2 | 0.90 | 1.00 | | 0.00 | 0.90 | 0.00 | 0.30 | 1.00 | 0.60 |
| **3** | 3 | 1.00 | 0.94 | | 0.44 | 0.72 | 1.00 | 0.44 | 0.67 | 0.33 |
| **4** | 4 | 0.50 | 0.00 | | 0.50 | 0.83 | 0.33 | 0.17 | 1.00 | 0.92 |
| **5** | 5 | 0.00 | 1.00 | | 0.31 | 0.03 | 1.00 | 0.66 | 0.48 | 0.31 |
| **6** | 6 | 0.88 | 0.29 | | 0.35 | 0.71 | 0.24 | 0.18 | 0.00 | 0.00 |