

Market_Segmentation

Haden Loveridge, Alex Parson, Biagio Alessandrello

2024-08-13

To create market segments we first scaled all variables around 0. We then iterated through 3 clusters to 8 and compared which clusters had significantly high values (greater than or less than 1, indicating that it is at least one standard deviation away from the average content).

For each cluster if no column is specified, then no attributes showed significance. The numbers in parenthesis show the value in standard deviations for that cluster. Next to the number of the cluster we can also see the number of followers in that cluster. For example, 'Cluster 3 (538):' means that cluster 3 has 538 followers.

3 Clusters

cluster 1:

cluster 2:

cluster 3: sports_fandom (1.9), food (1.79), family (1.42), religion (2.2), parenting(2.05), school (1.61)

5 Clusters

cluster 1: high travel (1.85), high politics (2.433), high news (1.93), high computers (1.64), high automotive (1.07)
cluster 2: cluster 3: high health_nutrition (2.157), high outdoors (1.66), high personal_fitness (2.12),
cluster 4: cluster 5: high sports_fandom (2.05), high food (1.82), high family (1.5), high religion (2.26), high parenting (2.122), high school (1.658)

7 Clusters

cluster 1 (805): health_nutrition (2.19), outdoors (1.72), personal_fitness (2.16)

cluster 2 (3583):

cluster 3 (517): photo_sharing (1.2), cooking (2.77), beauty (2.55), fashion (2.65)

cluster 4 (706): sports_fandom (2.08), food (1.84), family (1.5), religion (2.65), parenting (2.16), school (1.68)

cluster 5 (1280): chatter (1.2), shopping (1.09)

cluster 6 (620): travel (1.91), politics (2.49), news(1.99), computers (1.68), automotive (1.05)

cluster 7 (371): online_gaming (3.5), college_uni (3.2), sports_playing (2.17)

8 Clusters

cluster 1 (617): travel (1.9), politics (2.49), news (2.00), computers (1.67), automotive (1.05)

cluster 2 (513): photo_sharing(1.21), cooking (2.78), beauty (2.56), fashion (2.65)

cluster 3 (1270): chatter (1.21), shopping (1.12)

cluster 4 (801): health_nutrition (2.19), outdoors (1.72), personal_fitness (2.16)

cluster 5 (703): sports_fandom (2.01), food (1.84), family (1.5), religion (2.25), parenting (2.14), school

(1.67)

cluster 6 (49): spam (12.4), adult (3.75)

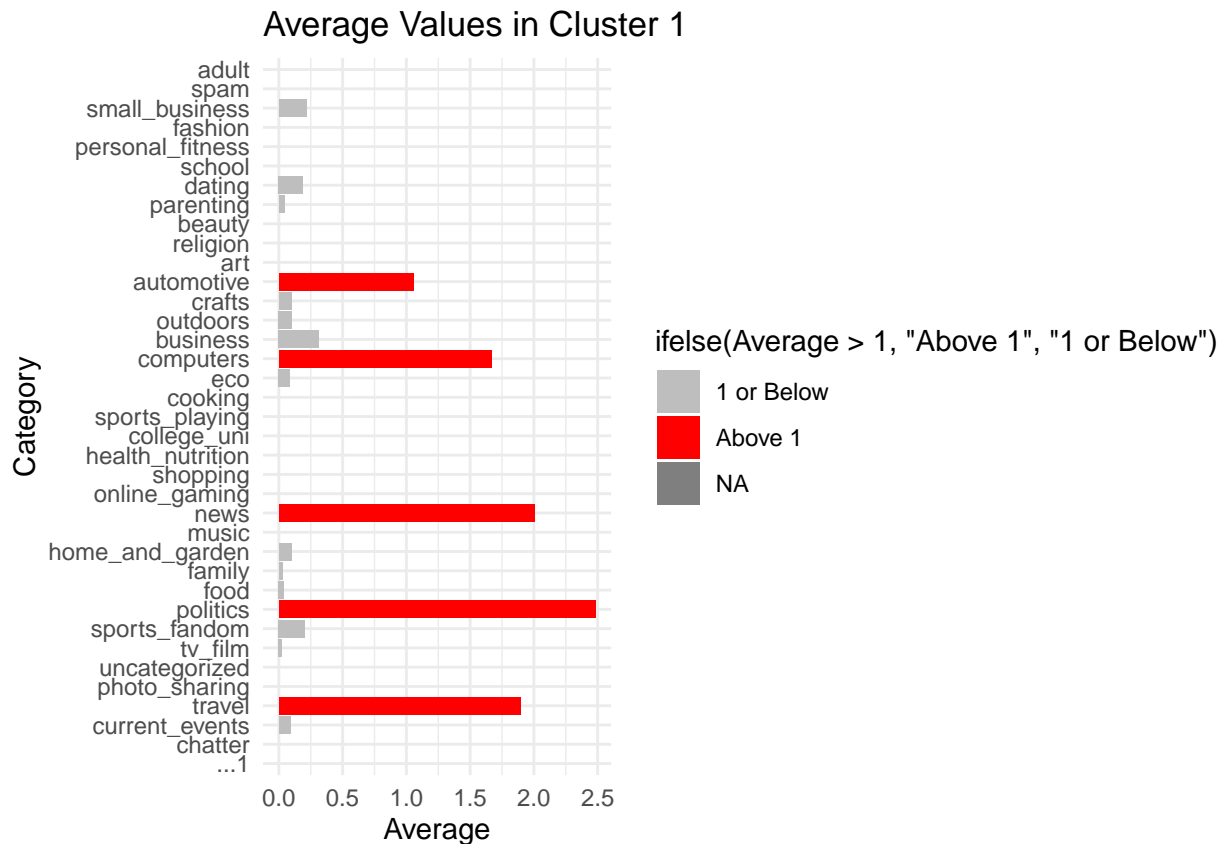
cluster 7 (3561):

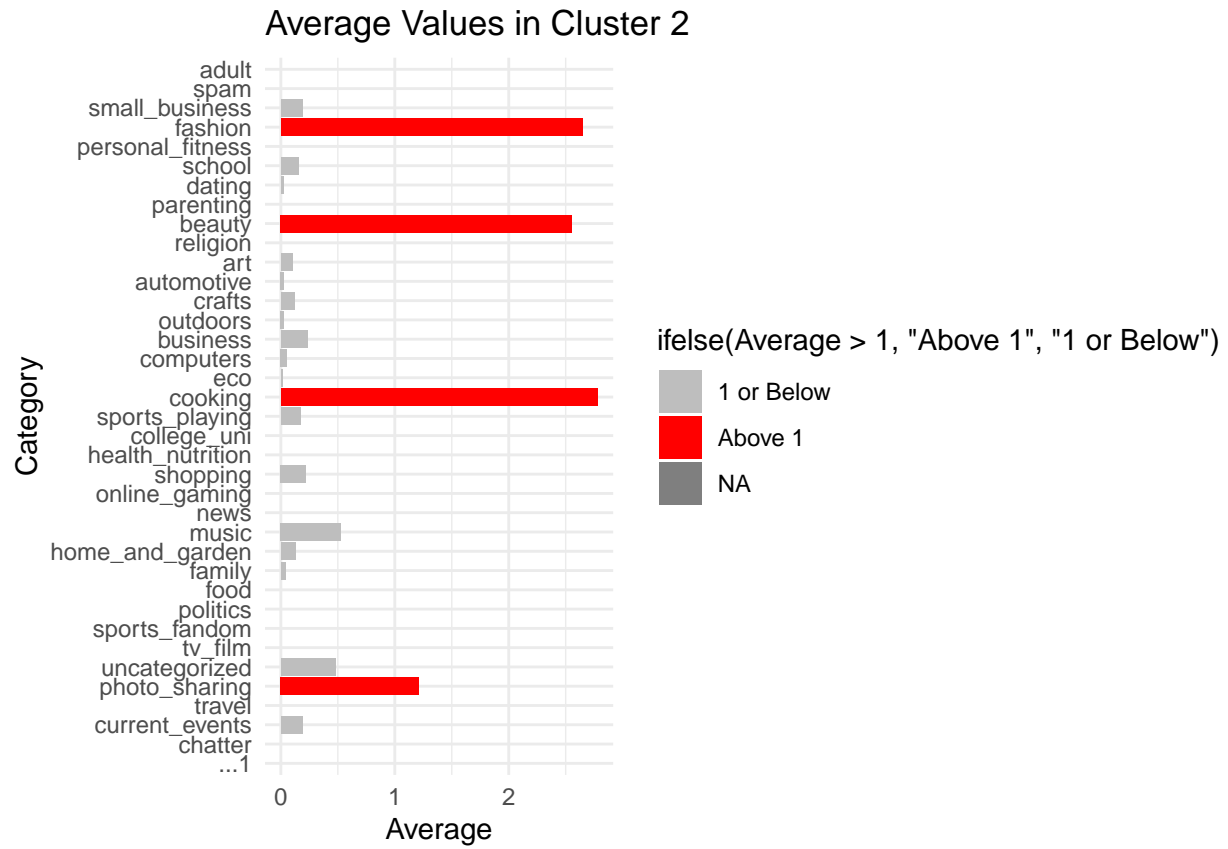
cluster 8 (368): online_gaming(3.5), college_uni (3.28), sports_playing (2.18)

Irelevant Columns: current_events, uncategorized, tv_film, home_and_garden, eco, business, crafts, art, dating, small_business

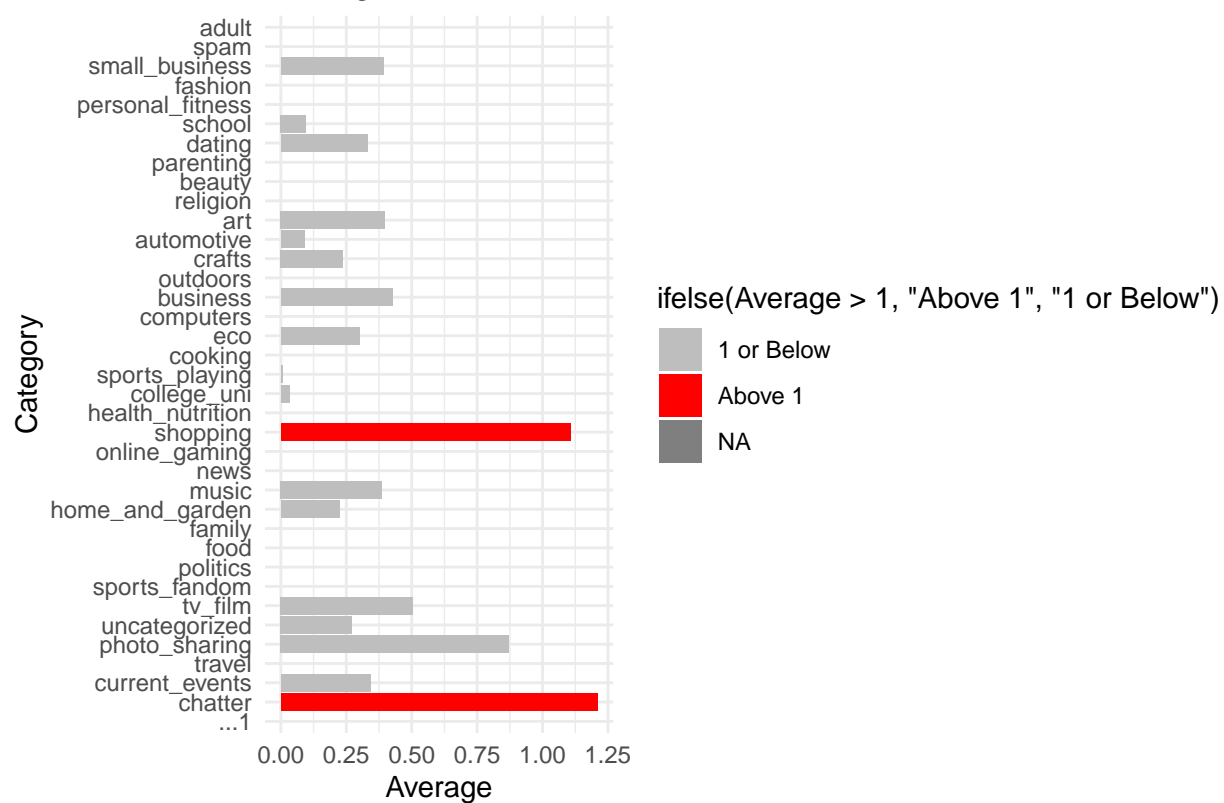
Final Findings:

We can see that of the 8 clusters, cluster 7 seems to have no significance, but it is the bulk of the followers. Cluster 6 is the spam accounts that don't provide insights. The other 6 clusters are true customer groups. For example, cluster 8 shows that we have a distinct group of followers that are into online gaming, college university, and playing sports. This group could be college students. Cluster 2 has 617 people that are very active about travel, politics, news, computers, and automotive. This group could be grandparents. Each group could also be classified as cluster 2 - Moms, 3- women, 4, healthy people, 5- dads, 6- junk / robots, 7- noise, 8- college students

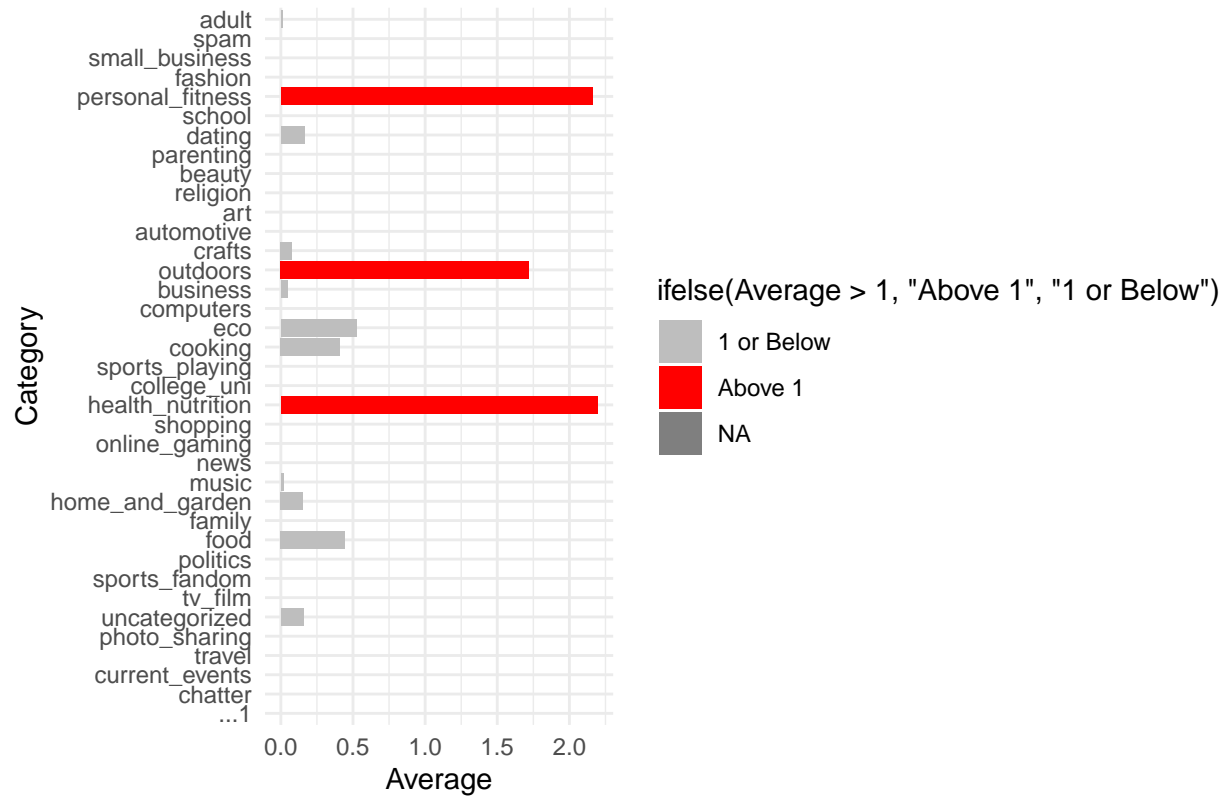




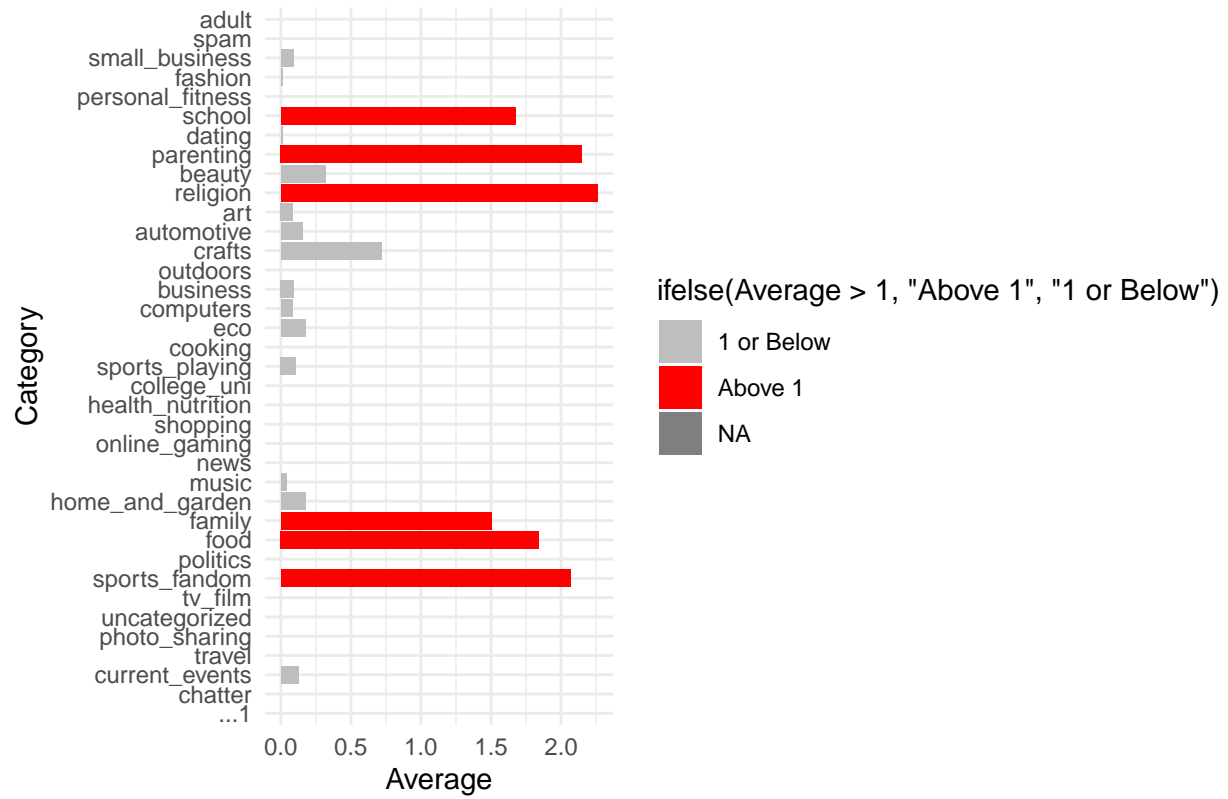
Average Values in Cluster 3



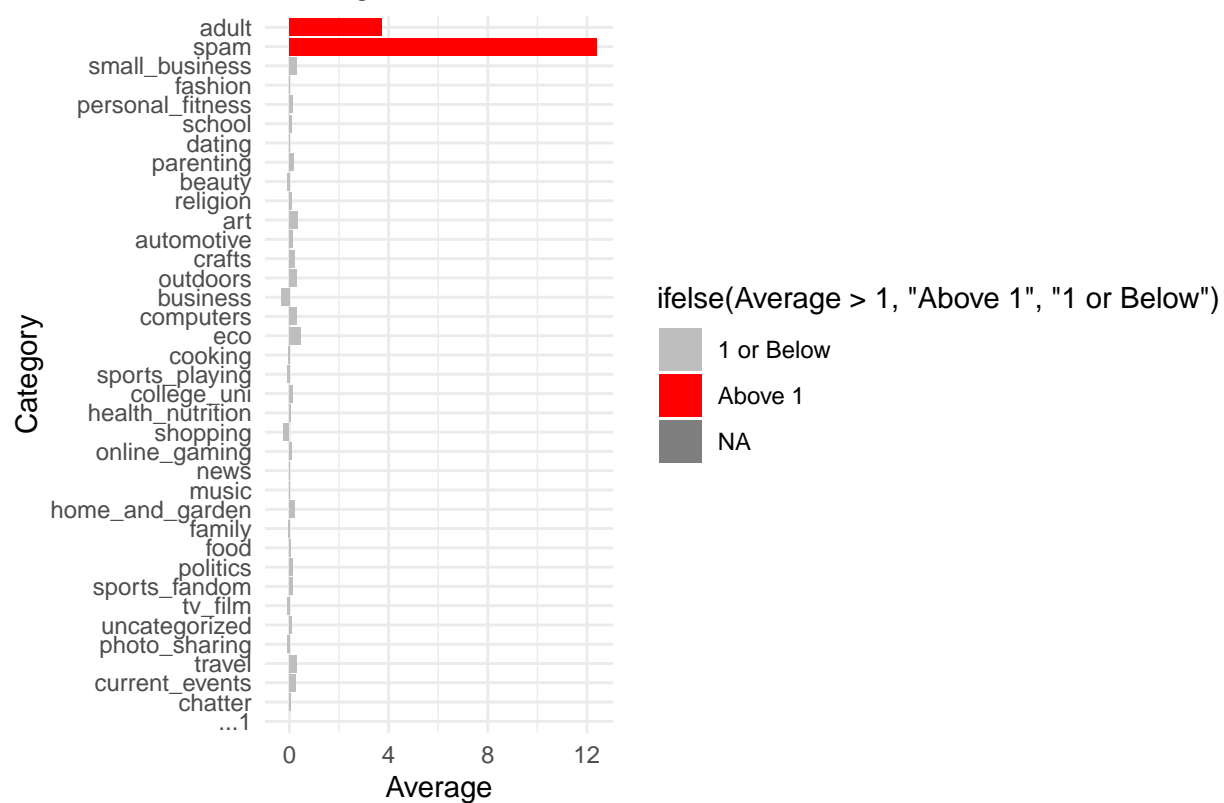
Average Values in Cluster 4

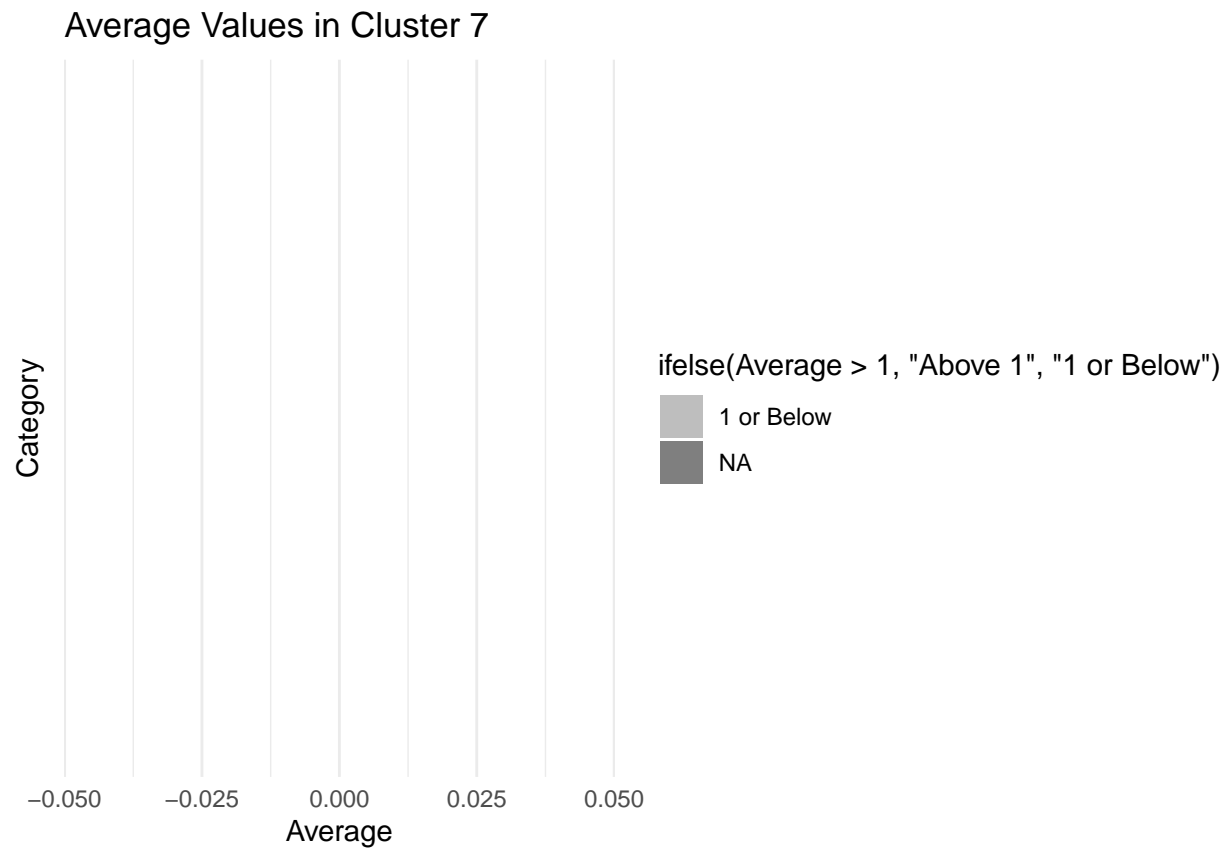


Average Values in Cluster 5



Average Values in Cluster 6





Average Values in Cluster 8

