

# Market Segmentation

Tyler Engalla

2025-05-12

## Market Segmentation

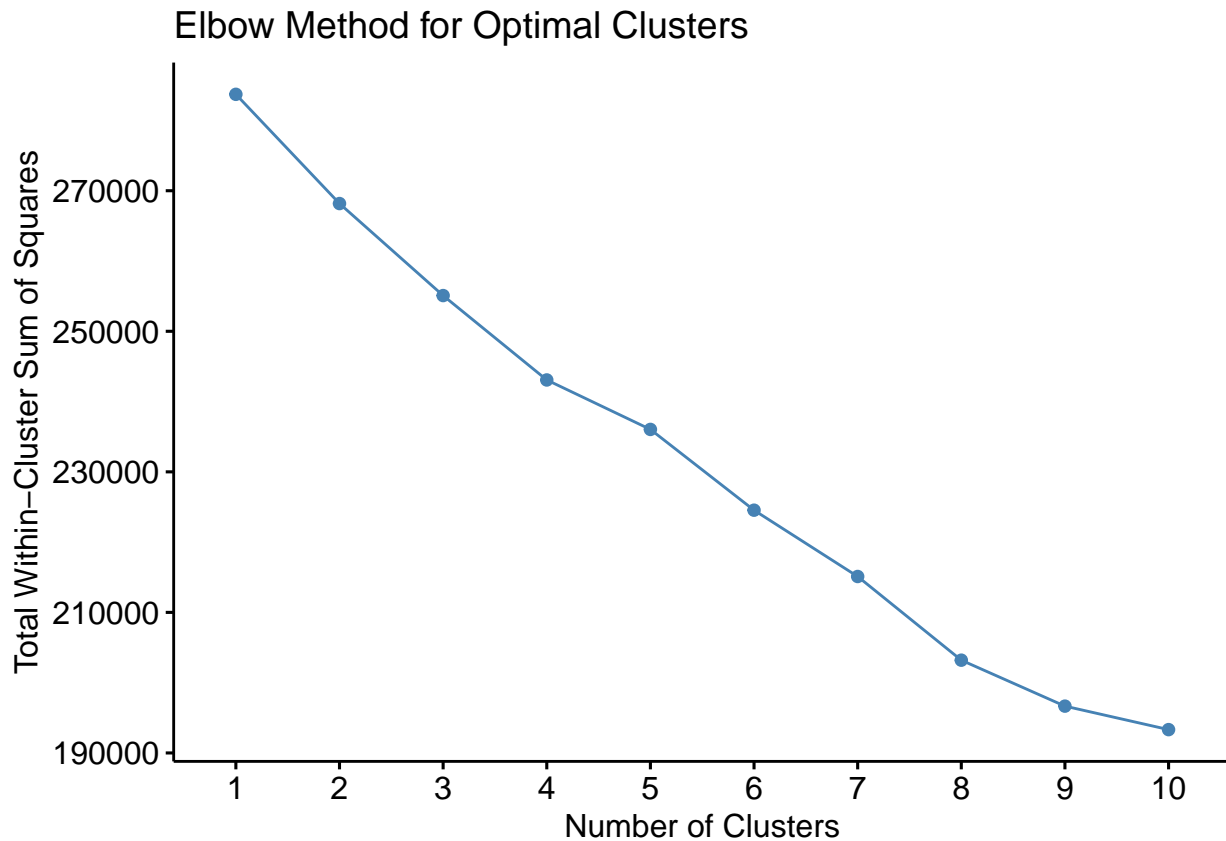
An advertising firm, NutrientH20, needs help segmenting their target audience. We have a data set of their twitter follower's activity, and using k-means clustering we'll help them segment their audience and provide insights around their follower's interest.

To begin, we start with data pre-processing - understanding the data and removing any nulls. There are 36 different categories where each tweet of each of the followers has been labeled to help us identify their user's interest.

Users may vary widely in their number of posts, so the next step taken was to normalize the data. This helps ensures that are analysis isn't skewed by a few highly active users. Next, we also scale the data so that we ensure fairness across the interest categories.

In order to break down these users into segments, we'll use K-means clustering. Using the elbow method, we can identify the optimal number of clusters to break the users into. From the chart, we see that the "elbow", or where the graph starts to flatten noticeably, appears around 4 clusters. This means our model does not substantially improve by adding any more clusters.

```
## [1] 0
```



Printing the cluster summary we can see the average proportion of posts for each of the various categories within each cluster. These proportions reflect the relative importance or frequency of each category for the users in a given cluster, thus giving us our market segments and their interests.

```
## # A tibble: 4 x 37
##   Cluster chatter current_events travel photo_sharing uncategorized tv_film
##   <fct>      <dbl>      <dbl> <dbl>      <dbl>      <dbl> <dbl>
## 1 1          0.0823      0.0378 0.0267      0.0421      0.0164 0.0200
## 2 2          0.0804      0.0348 0.0269      0.0427      0.0194 0.0174
## 3 3          0.154       0.0514 0.0354      0.0919      0.0275 0.0354
## 4 4          0.0896      0.0414 0.0860      0.0384      0.0189 0.0244
## # i 30 more variables: sports_fandom <dbl>, politics <dbl>, food <dbl>,
## #   family <dbl>, home_and_garden <dbl>, music <dbl>, news <dbl>,
## #   online_gaming <dbl>, shopping <dbl>, health_nutrition <dbl>,
## #   college_uni <dbl>, sports_playing <dbl>, cooking <dbl>, eco <dbl>,
## #   computers <dbl>, business <dbl>, outdoors <dbl>, crafts <dbl>,
## #   automotive <dbl>, art <dbl>, religion <dbl>, beauty <dbl>, parenting <dbl>,
## #   dating <dbl>, school <dbl>, personal_fitness <dbl>, fashion <dbl>, ...
```

Dim2 (7.3%)

Dim1 (8%)

cluster

- 1
- 2
- 3
- 4

**Cluster/ Market Segment 1 - Health and Nutrition Enthusiasts** - Health & Nutrition: (20.5%) - A significant portion of activity revolves around health and wellness, making this the defining interest of Cluster 1. - Cooking: (5.8%) - Reflects an interest in culinary topics, possibly tied to health and nutrition. - Personal Fitness: (11%) - Strong engagement with fitness-related content, aligning with the health-focused theme.

**Cluster/ Market Segment 3 - Politically Engaged and Outdoor Enthusiasts** - Politics: (14%) - Political discussions dominate this cluster, making it the defining feature. - Travel: (8.6%) - A secondary focus on exploration and travel-related topics. - Outdoors: (4.8%) - Engaged in outdoor and nature-related discussions.

3