Validation
Product Clustering

Karolina Mączka, Tymoteusz Urban

Preliminary remarks

- Absence of introduction, the columns in the dataset have not been explained
- Incorrect scale on plots
- Unjustified number of clusters
- Imprecise function cluster_names. It returns unmeaningful labels
- Referring to an external variable within a function

Second Checkpoint

- Introduction has been written
- Flawed functions have been corrected
- Plots have been corrected
- Still, there are no comments on how the number of clusters has been chosen
- Poor choice of metrics (lack of methods such as average distance between clusters etc.)
- Lack of explanation

Base Word Cloud

Unique clusters: 180

Validation Word Cloud

Silhouette: 0.04694754348168872
Calinski-Harabasz: 192973.75059555445
Davies-Bouldin: 1.334166307922311

Results

—

- The clusters are reasonably spaced.
- The most important words are meaningful
- The validation set confirms the results.
- There were not many algorithms tested so there is still a potential for achieving better results.