Profiling application users

Presentation and visualization of results

Mateusz Tchórzewski

GOAL

Characterizing the group of potential users to whom ads can be targeted

Data analysis process

Exploratory Analysis

Output: 551 subjects

Tool: Sweetviz, correlation matrix

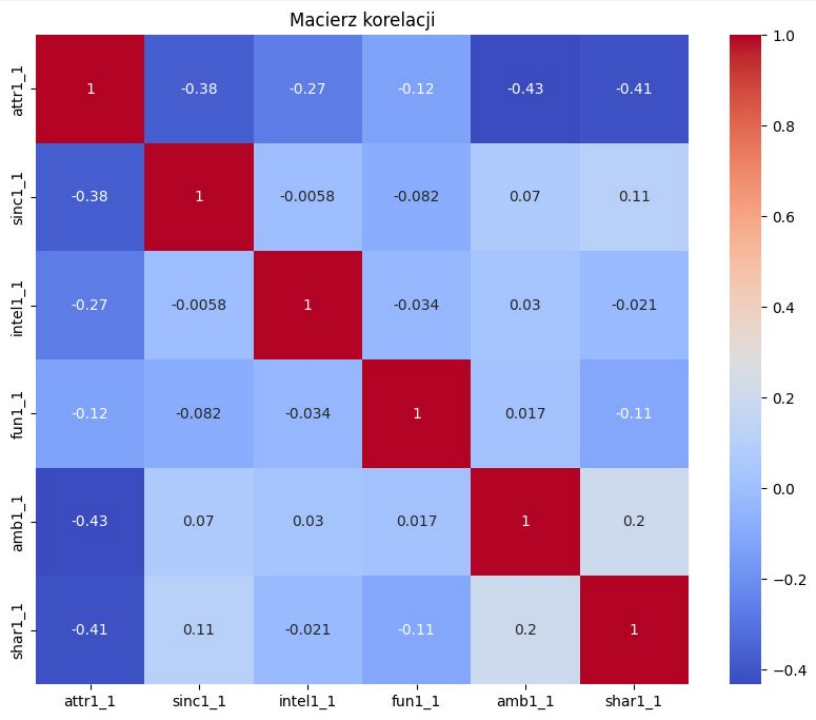
Purpose: Understanding data distribution, identifying missing values, basic descriptive statistics.

Data Preparation

Removing duplicates and filling missing values (filling gaps with zeros).

Data standardization with StandardScaler

Data selection



Fun Feature

Clustering Method

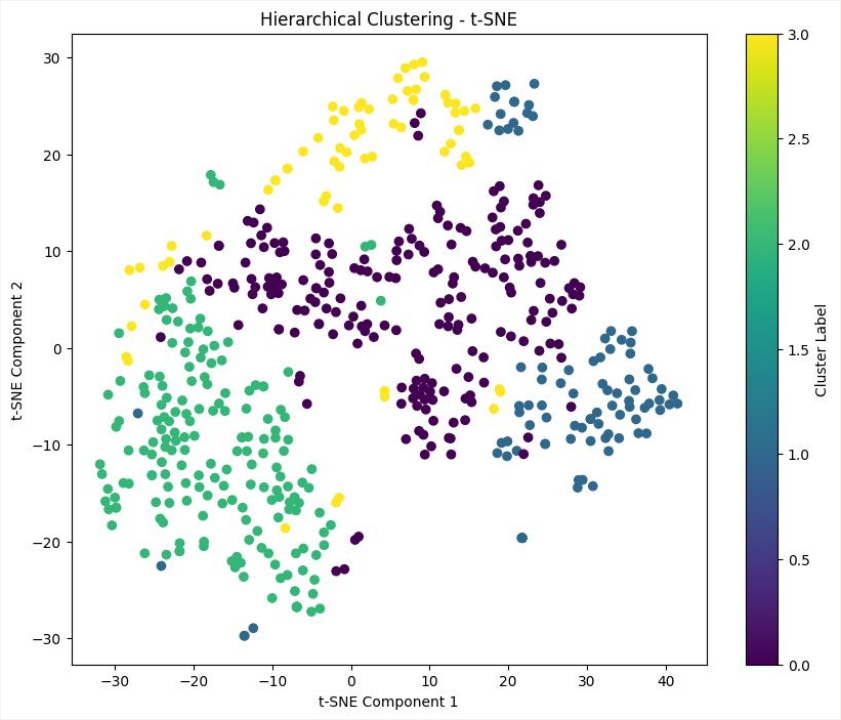
Algorithm: Hierarchical Clustering (Agglomerative Clustering).

Number of clusters: 4.

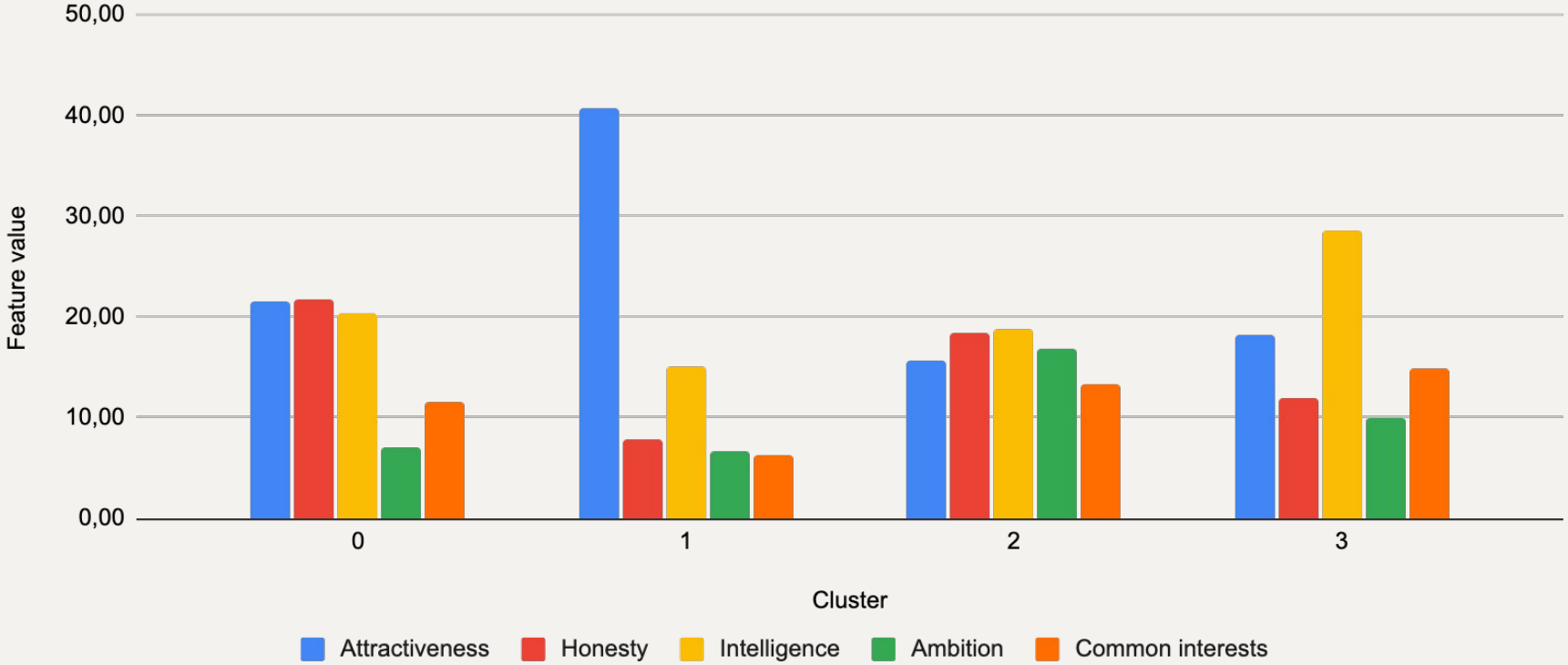
Distance measure: Euclidean.

Connection method: Ward

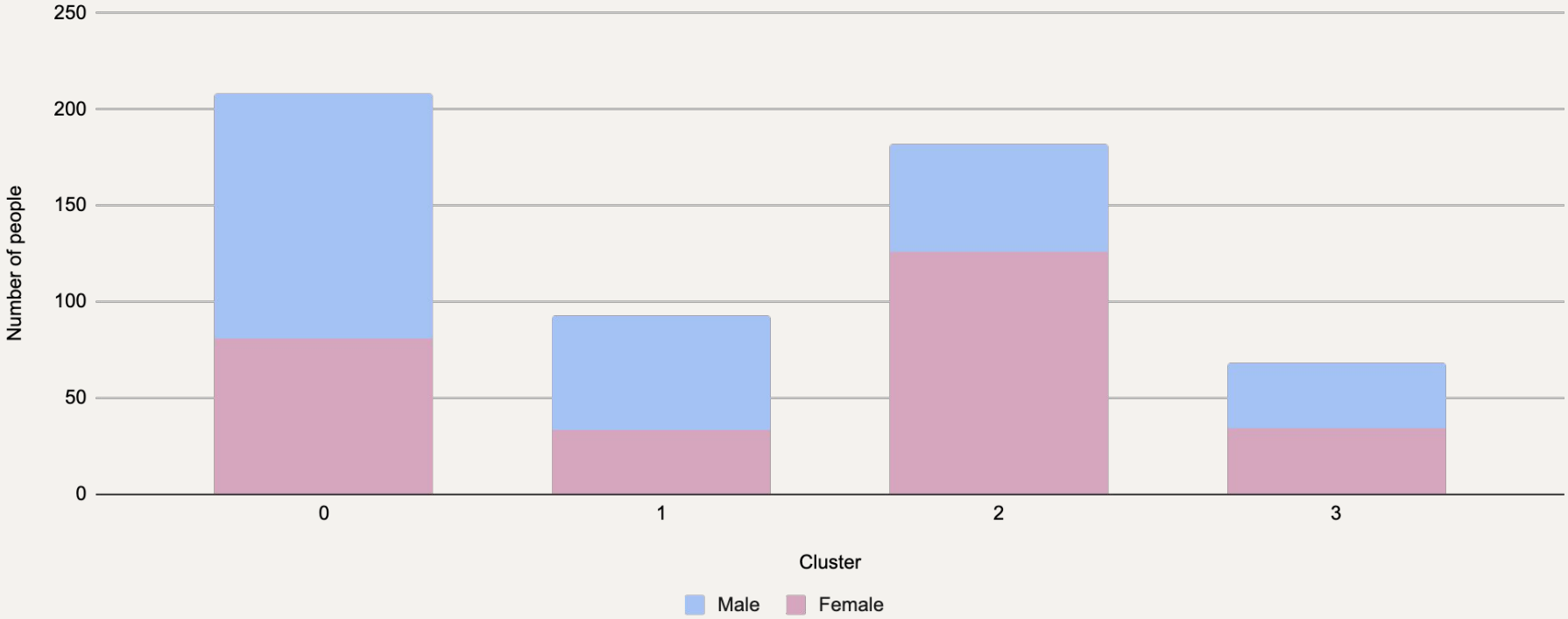
Evaluation: Silhouette method



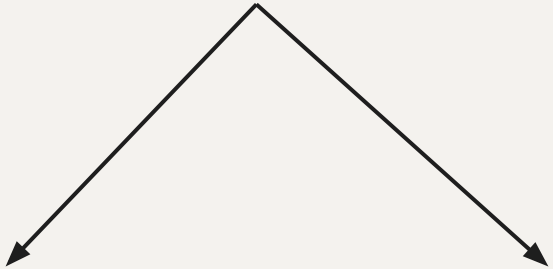
Distribution of desired feature values in clusters



Gender distribution in clusters



Data Split



Women

cluster - no. of people

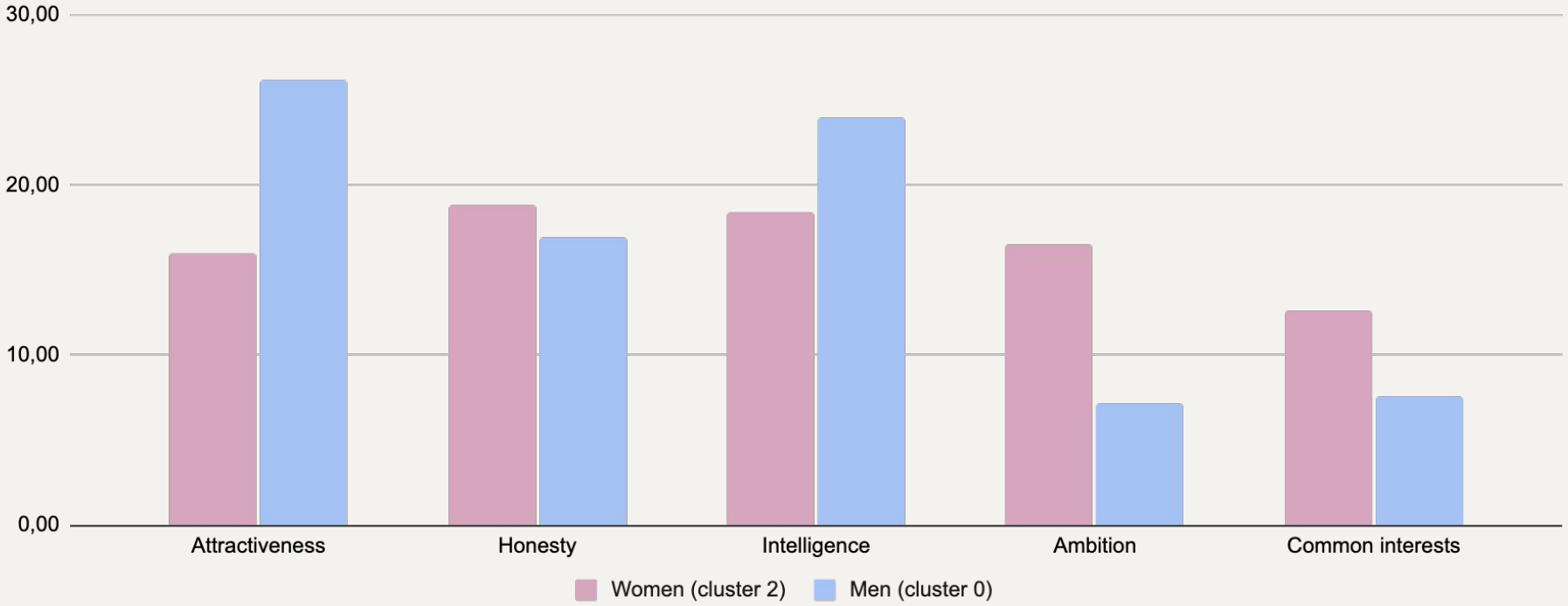
- **2 140**
- 0 111
- 1 18
- 3 5

Men

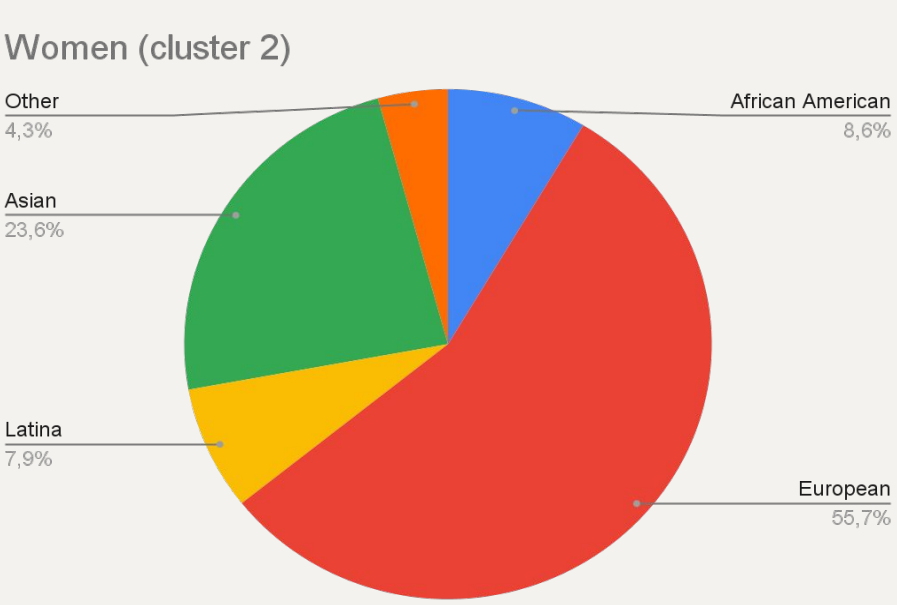
cluster - no. of people

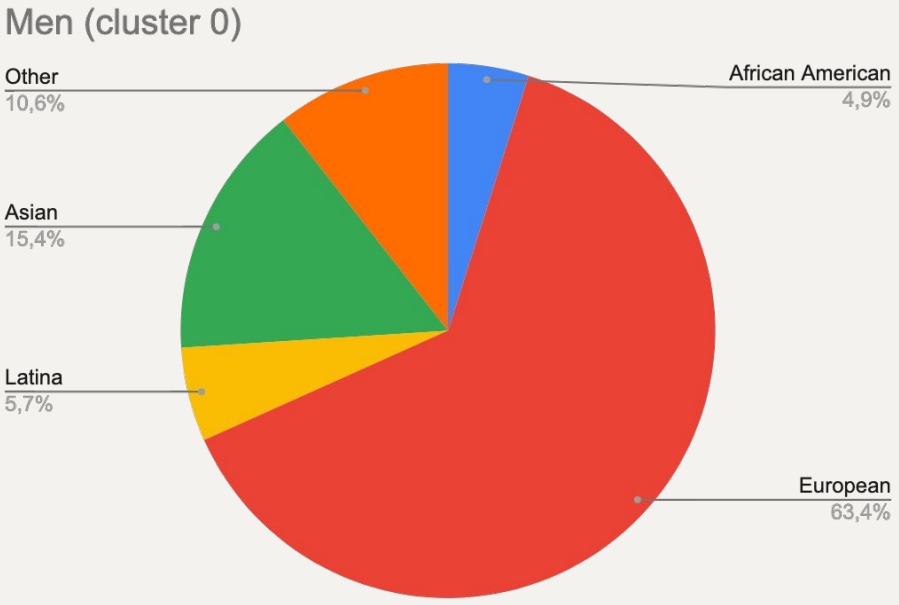
- **0 123**
- 2 93
- 3 32
- 1 29

Distribution of desired features



Ethnic group

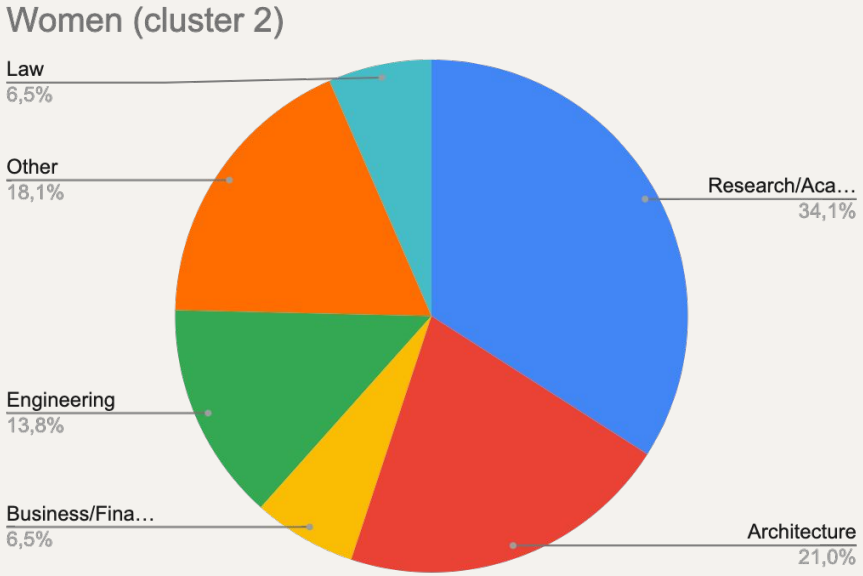


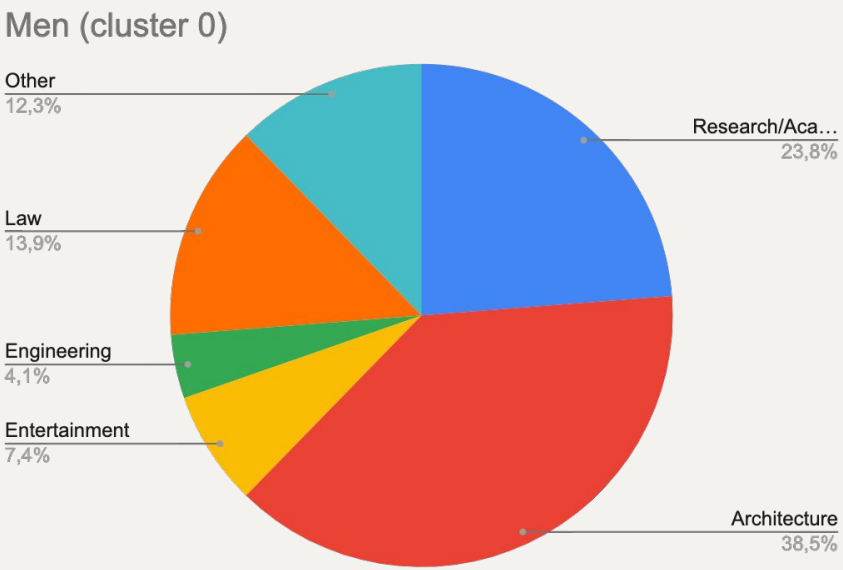


Interests


10,00
8,00
6,00
4,00
2,00
0,00
Food
Museums
Art
Reading
Theater
Movies
Music
Concerts
Women (cluster 2)
Men (cluster 0)

Professional field





Conclusions and recommendations

- Advertisements should be targeted to the European ethnic group, which can be translated into the geography of the advertisements.
- Common interests are not highly valued (as a desirable feature), so advertisements should not feature couples with common interests.
- Good areas for advertisements include primarily films, but also museums, art, reading, theatre and music.
- The median age for the groups in both sexes is 26.5 and 25 years, respectively. Therefore, the form and language of the advertisement should be age-appropriate.
- Employment area should not be taken into account due to the source of the data