

Business Intelligence Lab 3

Task: Descriptive Analysis, Unsupervised Learning – IKEA Case

Objectives:

- Identify essential features of Sweden's municipalities that are relevant for IKEA stores.
- Explore the `ikea_kommun_data.txt` file and gain insights into the data.
- Transform the data into a suitable format for clustering analysis.
- Reduce the dimensionality of the data to improve the performance of the clustering algorithm.
- Apply the k-means clustering method to discover new places in Sweden that may be suitable for IKEA stores.
- Evaluate the clusters obtained and determine the most promising municipalities for IKEA expansion.
- Visualize the results to gain a better understanding of the clusters and the distribution of the data.
- Provide actionable insights and recommendations for IKEA's expansion strategy based on the findings.

This task is about finding and evaluating clusters that contain data with similar properties.

Your task: is to discover some new places here in Sweden that may be suitable for IKEA department stores. You will do this by using the *k-means* clustering method. You have a text file, `ikea_kommun_data.txt`, which contains essential features for many of Sweden's municipalities to aid you in your findings. The English term *municipality* translated to Swedish is *kommun*.

IKEA stores are already available in the following municipalities: Borlänge, Gävle, Göteborg, Haparanda, Helsingborg, Jönköping, Kalmar, Karlstad, Linköping, Malmö, Stockholm, Sundsvall, Uddevalla, Umeå, Uppsala, Västerås, Älmhult, and Örebro. Some of these municipalities are missing in the `ikea_data.txt` file. The following link shows a map of Sweden's municipalities, <https://www.scb.se/contentassets/1e02934987424259b730c5e9a82f7e74/kommunkarta09.pdf>

The general steps are data exploration, data transformation, data reduction, and the k-means clustering method.

Submission:

Upload **Yourusername_Lab3.ipynb** file, add comments to explain your analysis.

Main Python libraries to use:

- scikit-learn (a Python library that features various classification, regression, and clustering algorithms) <https://scikit-learn.org/stable/>
- pandas <https://pandas.pydata.org/docs/>
- NumPy <https://numpy.org/>
- Matplotlib <https://matplotlib.org/>
- seaborn: statistical data visualisation <https://seaborn.pydata.org/>

“You can have data without information, but you cannot have information without data.”

~ Daniel Keys Moran