Customer Segmentation with k-Means for Demo Online Shop

Instructions

Assignment

Task Description and Requirements

The results of the practical application of the Marketing Data Science project should be processed in a "Management Briefing," oriented towards the most important phases of the CRISP-DM model. Present the key findings concisely and precisely.

1. Business Understanding & Marketing Data Science Canvas

A Marketing Data Science Canvas should be created for the clustering project "Customer Segmentation" of the test demo online shop "sustainable-fit.com". This includes the analysis of the business case, the use case, the presentation of the opportunities this project offers for practice, how success can be measured, which framework conditions are crucial for the project, and how it can be implemented in the company.

It can be assumed that the team behind the online shop startup is still very small and, in addition to the founder, consists of product suppliers and some voluntary supporters.

For the online shop, it can be assumed that it was launched a few months ago, and although there are already many customers, sales are only developing tentatively. The offerings still have much room for expansion, and the question of in which direction the product range should be further developed also arises.

According to the funnel logic and the selected use case, different phases (upper funnel phase, middle funnel phase, or lower funnel phase) can be tested and selected, whereby ultimately one phase should be chosen with a corresponding justification. It is up to you which features you assign to the respective phase in the context of feature selection, as long as there is a plausible explanation for it.

2. Dataset & Cleaning, Preparation

For data protection reasons, a fictitious dataset is provided as a CSV file, which is already largely cleaned. Alternatively, another dataset, such as a self-created dataset, can also be used.

As part of the exploratory data analysis, a data screening with interpretations should then be performed, e.g., using box plots, correlation matrices, etc. - the dataset should thus be meaningfully preprocessed in this process, e.g., filtering out extreme values or cleaning the data.

3. Modeling

With the finally cleaned dataset, a clustering model should be tested using the k-means algorithm. Visualization should be done with an appropriate tool, whereby the analysis within the framework of an iterative process may require adapting the feature selection or the number of clusters until a meaningfully explainable model is obtained.

4. Evaluation & Deployment

An assessment of the results compared to the previous situation should be discussed. It can be assumed that no clustering using ML algorithms has taken place before.

The most important findings should be presented, for example:

• What is the model quality?

• What are the limitations of the clustering model, and how can the results be further utilized in practice?

• What are the recommendations for further action?

• How can stakeholders be involved in the use case - how can acceptance for the project be created?

Of course, graphics for visualizing model concepts or other elements for better understanding can also be used in the management briefing.

The utilized workflow should also be included in the management briefing.

Appendix: CSV dataset "dataset.csv"

Length: 10 pages of pure body text, plus illustrations.