Mini Project: Clustering

Scenario

A Shopping Mall tries to better understand its customers in order to be able to form clusters of customers aiming at a better individual approach to the customer experience. To that extent, data has been collected from 200 customers recording their *gender, age, yearly income* (in k$) as well as a *spending score* (between 1 and 100). Based on this data, an initial clustering shall be performed and the characteristics of the clusters shall be identified.

Tasks

1. As a preparation, use the free e-book Python Machine Learning von Sebastian Raschka (https://www.packtpub.com/free-ebook/python-machine-learning) to read up on the topics

   - encoding and scaling from chapter 4 (and as an additional source: https://scikit-learn.org/stable/auto_examples/preprocessing/plot_scaling_importance.html)

   - Principal Component Analysis (PCA) from chapter 5

   Also, you should research, what the *CRISP-DM-process* is and think about what this might have to do with the project at hand.

2. Then, read the data from the file `data/Kunden_EKZ.csv` into a *pandas data frame* in a *Jupyter notebook*. When doing so, think about how to deal with the column `Customer_ID`.

3. Perform a profiling of the data.

4. Subsequently, transform the data in such a way that you can work with two explanatory variables which are able to explain that main part of the variance.

5. Next, perform a clustering using the optimal number of clusters. Also, visualize your results in an appropriate fashion.

6. Afterwards, interpret your result by applying the clustering you found to the original data and characterizing them based on typical values of the recorded data.

7. Finally, match the steps you took to solve the mini project to the phases of the CRISP-DM process. Which phases might not yet have been gone through an how might these look for the case at hand?

As you solve the tasks, be sure to explain what you are doing in an adequate fashion (by making use of the markdown possibilities provided by the Jupyter notebooks).