



Data scientist assignment - Raven data

Objective

Conduct customer segmentation to identify distinct groups within the customer base for retail C-Levels, Mainly CEO and CMO. Your work should encompass data exploration, preprocessing and feature engineering, as well as model training and evaluation.

Data Source

[UCI Machine Learning Repository - Online Retail dataset.](#)

Task

- Conduct EDA to understand customer purchase behavior.
- Data cleaning and validation.
- Create features that distinguish between customer behavior clusters.
- Apply clustering algorithms to segment customers.
- Visualize and interpret the segments.
- Provide a report with business recommendations based on the segments identified.

Solution Outline

- EDA: According to the retailer's main KPIs.
- Preprocessing: Calculate clusters values, standardize features.
- Clustering: Determine the optimal number of clusters using the elbow method or silhouette score, apply clustering algorithms.
- Visualization: Use PCA or t-SNE for visualization of clusters.
- Reporting: Your suggestion for main KPIs, describe each segment, suggest targeted marketing strategies, and potential value of each segment.

Submission

- Documentation of your process with a (short) explanation of each step: You can put your work in a document or in your code notebook.
- Code file

Notes

The requirement is a simple, straightforward model with a valid logic, the actual result does not matter. If there are steps and processes you would have included but the implementation is too time consuming add them as a comment in the pipeline or as future steps at the end of your documentation.

Good luck!