

**Title:**

# An Overview of State of the Art Clustering Techniques in R

—

## Evidence for Customer Segmentation from an Online Fashion Retailer

**Outline:**

1. Introduction/Motivation
  - a. The Idea of Clustering
  - b. Its Application in Marketing
  - c. Overview of the Paper
2. Theory of Clustering
  - a. Choice of Clustering Technique
    - i. Partitioning (K-Means)
      - Elbow Method
      - Information Criteria (Bayesian Inference Criterion for k means)
    - ii. Hierarchical
      - Agglomerative/Divisive
      - Choice of Proximity Measure (depending on variable scale)
    - iii. Density (DBSCAN)
  - b. Evaluation (Internal, External)
3. Data from an Online Retailer (Application in R)
  - a. Data Manipulation and Cleaning
  - b. Total orders per Month, Amount Spent per Month, RFM (Recency, Frequency, Monetary)
  - c. Depending on: Gender, age, state?
4. Evidence (Application in R)
  - a. Results
  - b. Graphical representation
    - i. Dendrogram
    - ii. Scatterplotters for all variables, colored by cluster
5. Discussion (of the different methods with respect to the results)
  - a. Internal/External plausibility
  - b. Problems of Data Scales
6. Conclusion

**Literature:**

- W. Härdle, L. Simar: *Applied Multivariate Statistical Analysis*. Springer, New York 2015  
→this procedure is partly replicated

**Tasks:**

- Make folder in cloud (*till end of the week*)
  - o Includes all relevant docs (literature, code)
- Make Github Folder (*till end of the week*)
- Get all submission information (*till end of the week*)
- Required Structure of Paper (as above) (*till end of the week*)
- Coding and its Style of Coding
- Slides (Preliminary Report) (*till appointment*)