



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Escola Superior d'Enginyeries Industrial,
Aeroespacial i Audiovisual de Terrassa

MUEI, MUEA, and MASE

Advanced Engineering Data Analysis (AEDA) Clustering

Lab Practice

Prof. Daniel Fernández

daniel.fernandez.martinez@upc.edu

Wholesale customers Data Set

Description: The data set refers to clients of a wholesale distributor. It includes the annual spending in monetary units (m.u.) on diverse product categories

Location:

The data is in the file Wholesale customers data.csv

- 1) FRESH: annual spending (m.u.) on fresh products (Continuous);
- 2) MILK: annual spending (m.u.) on milk products (Continuous);
- 3) GROCERY: annual spending (m.u.) on grocery products (Continuous);
- 4) FROZEN: annual spending (m.u.) on frozen products (Continuous)
- 5) DETERGENTS_PAPER: annual spending (m.u.) on detergents and paper products (Continuous)
- 6) DELICATESSEN: annual spending (m.u.) on and delicatessen products (Continuous);
- 7) CHANNEL: customers' Channel - Horeca (Hotel/Restaurant/Café) or Retail channel (Nominal)
- 8) REGION: customers' Region - Lisbon, Oporto or Other (Nominal)

Activity: Apply a **k-means** and assess result as we have seen in class.

Solution proposal: wholesale_solution.pdf

Edgar Anderson's Iris Data Set

Description: The famous (Fisher's or Anderson's) iris data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. The species are Iris setosa, versicolor, and virginica.



Location:

```
library(datasets)  
iris
```

Activity: Apply a **hierarchical clustering** and **model-based_clustering** and assess result as we have seen in class.

Solution proposal: A comprehensive solution including enhancing the results with the dendextend package is in iris_hc_solution.pdf and in iris_mbc_solution.pdf