

Exercise 07

Consider the following file: <https://github.com/masterfloss/data/raw/main/realEstate2018lab09.xlsx>

The variables

- * Municipalities - municipalities (concelho) names
- * Waste total - total collected waste, in tons
- * Collected waste - total selected waste, in tons
- * Indifferent waste - non selected waste, in tons
- * Unemployment - Total of unemployment, number of people
- * Tourism - Nights spent in tourist accommodations, number per 100 inhabitants
- * Crime - Crimes registered by police, number per thousand inhabitants
- * Wage - Average monthly earnings of employees, Euros
- * Bank evaluation M2 - bank evaluation M2, Euros
- * Tax - sum o total real estate taxes paid in the region, euros per capita
- * Purchasing power - purchasing power compared to the national value, Nation average =100
- * Population - estimated total population, number of people
- * price2000 - average prices of real estate sold in 2000 by
- * price2018 - average prices of real estate sold in 2018

1. change name from "collected waste" to "selected waste".

2. Calculate the following variables:

- $\text{percentage selected waste} = \text{selected waste} / \text{waste total}$
- $\text{percentage unemployment in the total population} = \text{unemployment} / \text{population}$

3. Calculate regression using OLS. Use as target price 2018

4. Calculate regressions using other algorithms

5. Identify possible clusters

6. Identify PCA.