

SOCIAL INCLUSION. BIG DATA PILOT

CONTEXT AND OBJECTIVES

The Region of Madrid has developed, in cooperation with the DGIT of the European Commission and Everis, in the context of ISA Work Programme 2016, a pilot project on data analytics supporting social inclusion policies.

The fight against poverty and social exclusion is at the center of the Europe 2020 strategy for smart, sustainable and inclusive growth of society.

The aim of the pilot project was to develop a reusable platform supporting data analysis, in order to:

- Identify the **different groups** at risk of social exclusion.
- Detect the **chronological evolution** and **identifying causes and factors** that affect the rate of inclusion.
- Establish an **insertion index** to measure the potential inclusion of an individual or family unit and also detect strengths and weaknesses in social policies applied.

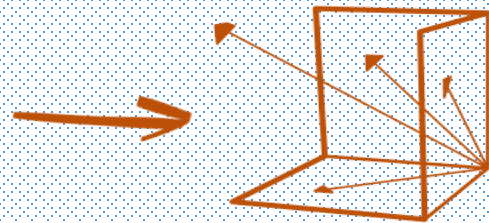
SOCIAL INCLUSION. BIG DATA PILOT METHODOLOGY

The methodology used for the treatment of the information is CRISP-DM, an analytical process model that describes commonly used approaches that experts use to tackle to guarantee the quality of the final results.

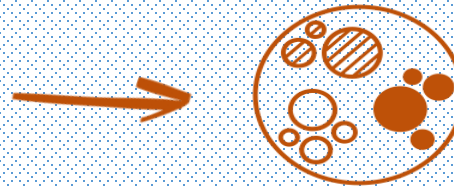
The following concrete phases and tasks were developed:



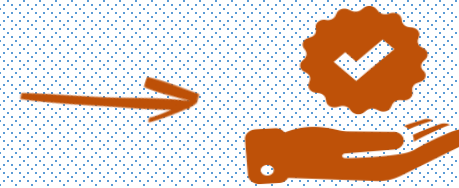
Data understanding: *in order to analyze data from an statistical point of view to obtain the initial hypothesis*



Data preparation & Axes selection: *the main citizen characteristics are chosen to find clusters*



Segmentation: *segments of citizens by previous selected characteristics are found using different analytical algorithms: k-means, hierarchical*



Validation: *business sense cluster analysis and error measuring*

SOCIAL INCLUSION. BIG DATA PILOT SEGMENTATION

The construction of the segmentation has been done applying two different groups of techniques:

1. Application of targeting statistics techniques

K-means

The k-means algorithm is a statistical model technique which consists in the creation of homogeneous groups (segments) with similar features, being the resulting segments as much heterogeneous as possible among them.

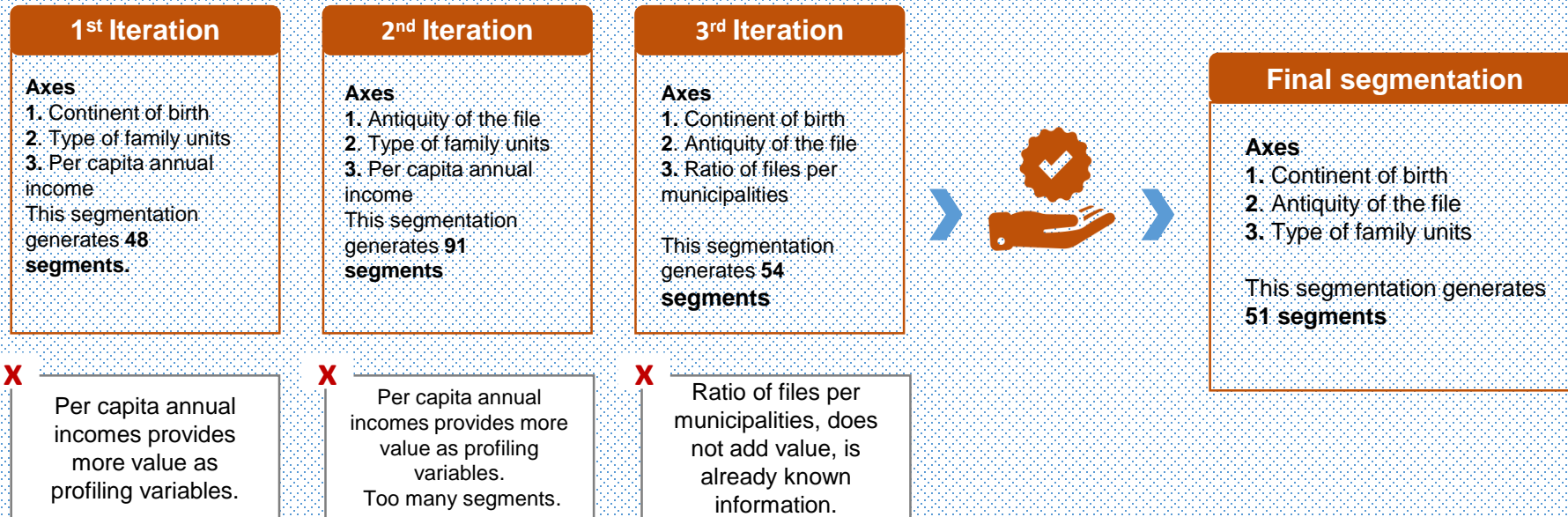
hierarchical algorithms

The hierarchical algorithms are statistical model techniques which consist in the creation of homogeneous groups (segments) from cuts at different levels of the values of the target study variables

Both algorithm has not produced homogenous groups

SOCIAL INCLUSION. BIG DATA PILOT SEGMENTATION

2. Construction of segments using business rules that allow us to classify files in homogenous groups and choose the axes of the segmentation to give us the maximum possible information.



The final segmentation produced homogenous groups and with heterogeneous characteristics between them

SOCIAL INCLUSION. BIG DATA PILOT CONCLUSIONS AND MEASURES

- **Less effective insertion policies for foreigners.** Nationals have higher chances of leaving the social exclusion.
- **Insertion handicaps in coexistence units with dependent minors.** Policies should treat this group in a personalized way to significantly increase the options of social insertion.
- **Greater chances of insertion if a minimum income is available.** If annual income is less than € 2,500 per capita, the percentage of records that are inserted is just 6%. In addition, these cases represent 66% of all existing records.
- **South of the Region have larger number of records.** these areas should be the focus where to put the available resources in order to prevent the population of those areas to request the minimum insertion income.

Efficiency of insertion policies

- To identify priority action profiles, to focus efforts.
- To customize activation plans to profiles with higher probabilities of insertion.
- To combine plans with other policies for those profiles with less probabilities of insertion

Management efficiency

- Model for process characterization to allow customization of insertion policies on each individual.
- Improvement in the application for collecting information.
- Follow-up traceability.
- Analysis and redistribution of workload of social workers.

Prevention

- Early or predictive detection of those profiles that will potentially become seekers of the benefit.
- Specific action plans in areas with the greatest volume of records, but mostly in those with higher rate of records per inhabitant