

ACCIDENTS IN SPAIN 2013

An application of Clustering algorithms by
Jorge Hidalgo Calderón

Objective:

In the link:

https://sedepi.dgt.gob.es/WEB_IEST_CONSULTA/subcategoria.faces

I find official data about accidents in Spain during the year 2013. In this .ppt I use the online software BigMI to, via a clustering algorithm, find an answer to the following question:

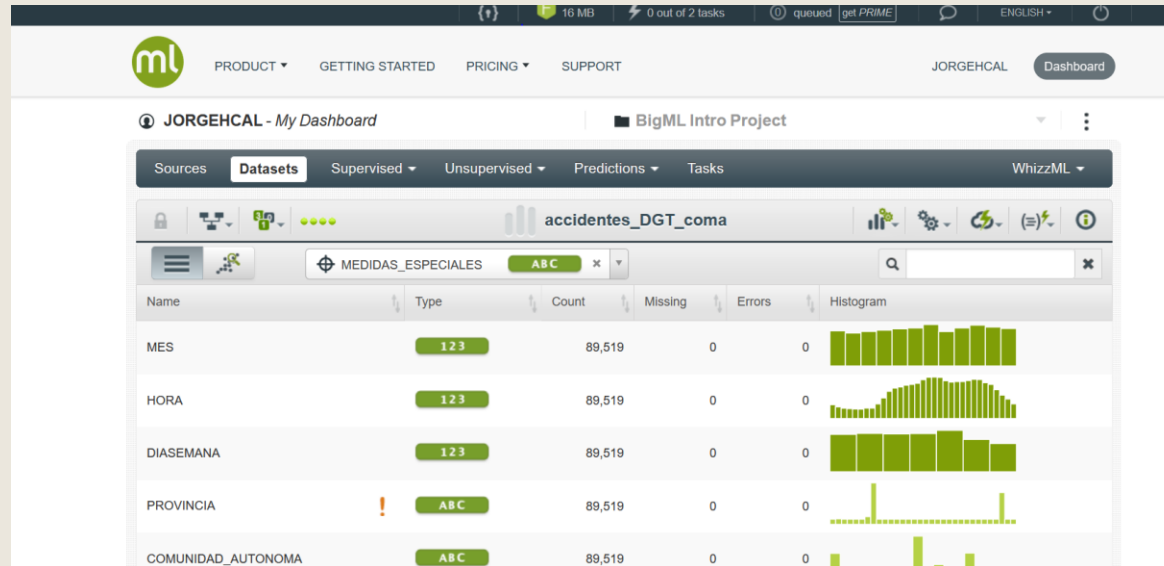
¿How is the amount of vehicles involved related to the accident's gravity?

¿How the amount of vehicles involved related to the accident's gravity?

After some attempts, I apply a clustering k-means with a number of clusters $k = 5$, using the following variables:

- TOT_MUERTOS, TOT_HERIDOS_GRAVES, TOT_HERIDOS_LEVES, TOT_VEHÍCULOS_IMPLICADOS (Deaths, Serious Injuries, Light Injuries, Amount of Vehicles Involved)

Two images, one of the dataset other of the configuration of the clustering:



The screenshot shows the 'CLUSTER CONFIGURATION' interface for the dataset 'accidentes_DGT_coma'. The 'Clustering algorithm' is set to 'K-means'. The 'Number of clusters (K)' is set to 5. The 'Default numeric value' is set to 'Mean'. The 'Model clusters' section shows a cluster icon. The 'Cluster name' is set to 'accidentes_DGT_coma'. There are buttons for 'Reset' and 'Create cluster'.

CLUSTER CONFIGURATION

Clustering algorithm: K-means

Number of clusters (K): 5

Default numeric value: Mean

Model clusters: [Cluster Icon]

Advanced configuration [Gear Icon]

Cluster name: accidentes_DGT_coma

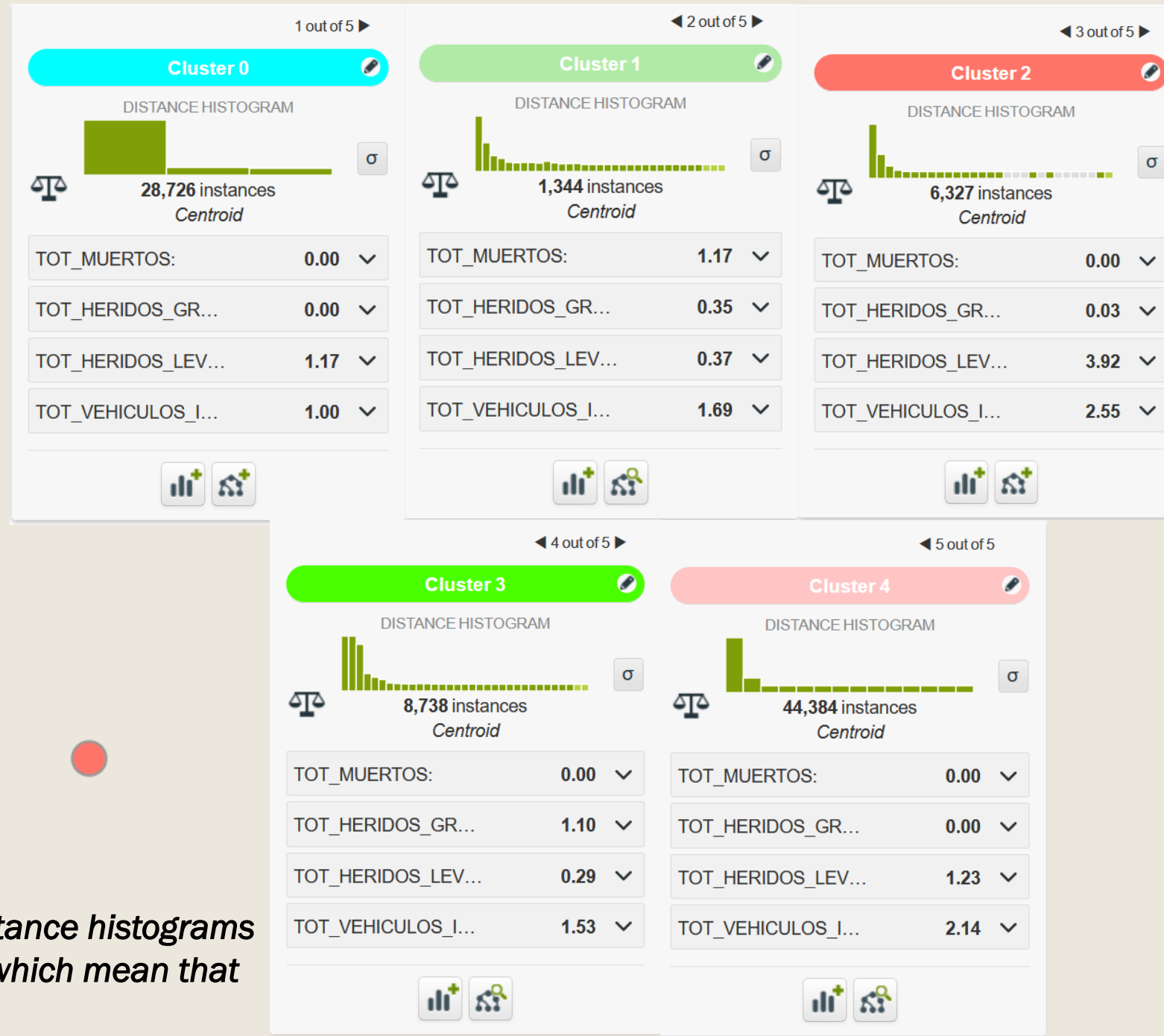
[Reset] [Create cluster]

The clustering use the K-means algorithm: this means that the algorithm computes distances among data and create groups using an iterative process which has been proved to converge to a solution (a stable collection of clusters).

The result:



This clustering is a good one: the distance histograms tend to accumulate data at the left, which mean that the approximation are good enough



Concluussions from the clustering:

- **With only one vehicle** (Cluster 0), there is only light damage, **NOT deaths NOR serious damage**. This Clustering is interesting since it is common to think that more vehicles imply more serious damage, and this seems to be false.
- In the other clusters there is more than one vehicle implied in the accident. Note that in clusters where the vehicle amount increases, deaths and serious injuries dissappear, and light injuries increase.
- The cluster which shows **deaths** also shows the others kind of injuries, and **more than one vehicle involved**.