Felipe Guth - 14210231

P1

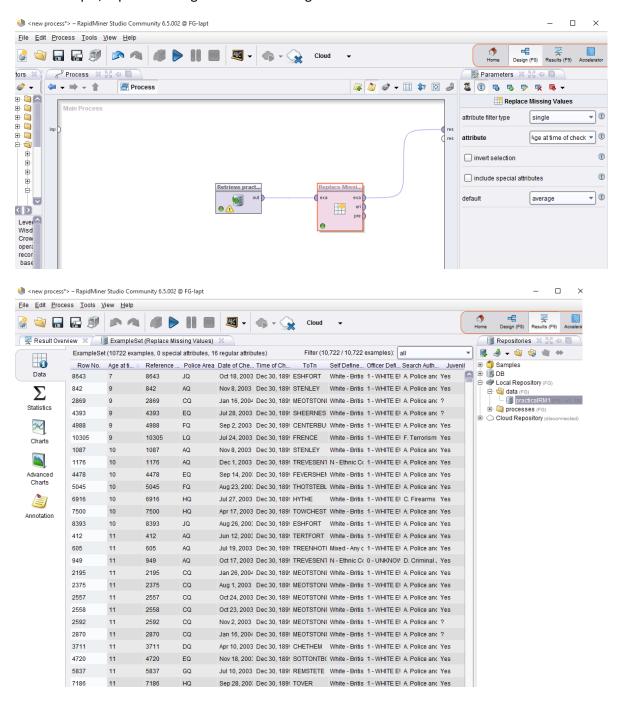
Missing values are identified with an "?".

Percentage of missing values: 379/10722. 3.53%

The percentage is not very important because it represents a small fraction of the data.

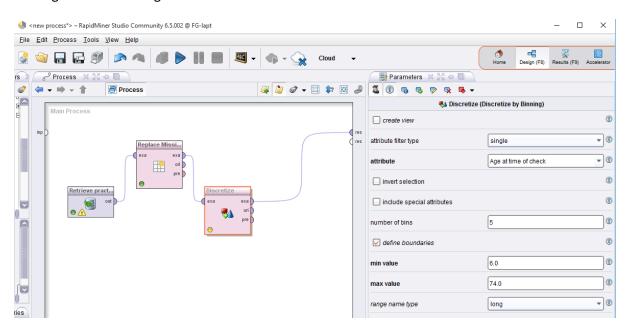
The best way would be filling the age as 17 if the column juvenile contain the value yes and fill with the mean value in case of not juvenile or in case of missing this second information as well.

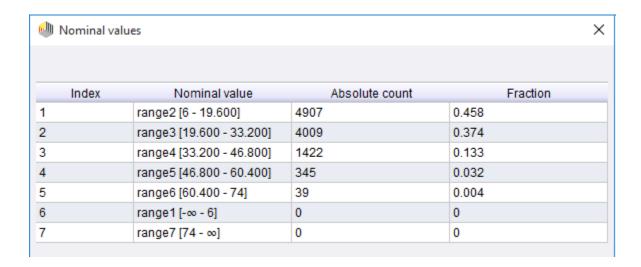
RM technique, replace missing values with average of attribute.

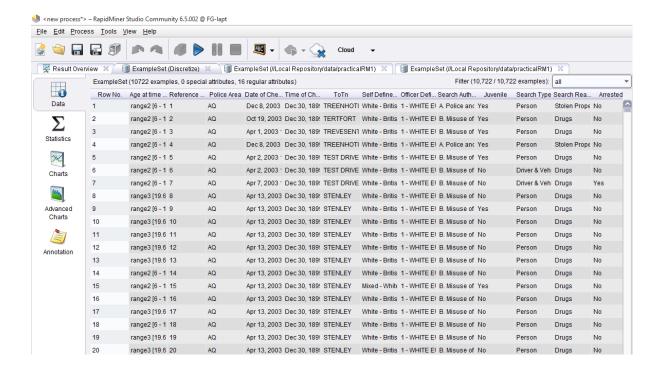


The noise in the case is the variance error due to data entry problems. The officers in the moment of inputting the data may have had problems with the equipment or performed a wrong operation leading to erroneous inserting of data.

Binning method to categorize the data into 5 classes.







Р3

It is not possible to normalize the attributes Juvenile, Search Reason and Police area, once they are string, non-numerical values.

The only attributes possible of normalization are age at time of check and reference number (which do not make sense).

