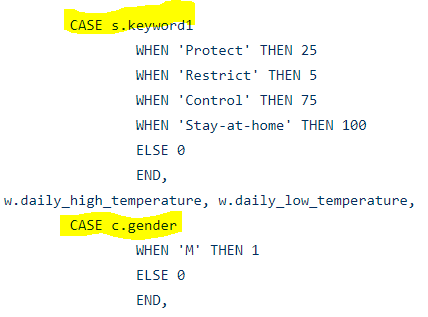
Deliverable 4 Data Mining

Data transformation:

For some age that we did not know, we will ignore the entire sample, because there are not many missing values in age measure.

For handling categorical attributes, we used conversion to ordinal data.



To normalize the numeric attributes,

For feature selection and under sampling, we remove redundant 20s samples to balance the data because compared with 60s and more, the property of 20s is too high, which would affect the attribution when we are mining data.