```
proc import
datafile='C:\Users\gabriele.politi2\Desktop\ClamdaIM\Dataset.xlsx'
dbms=xlsx
out=project;
getnames=yes;
run;

libname rest 'C:\Users\gabriele.politi2\Desktop\ClamdaIM\';
data rest.questionario;
set project;
obs_id=_n_;
run;

proc datasets lib=rest;
modify questionario;
label s1='Frequency';
label s2='Time';
label s3='Reason';
label s4='Browsers';
label s5='Preinstalled';
label a1='imp_Privacy';
label a2='imp_Security';
label a3='imp_Feature';
label a4='imp_Search';
label a5='imp_Interface';
label a6='imp_Customization';
label a7='imp_Integration';
label a8='imp_MobileInterface';
label a9='imp_Speed';
label d1='Gender';
label d2='Age';
label d3='Region';
label d4='Occupation';
label d5='Degree';
run;
proc freq data=rest.questionario;
table d1--d5;
run;
proc means data=rest.questionario;
var a1-a9;
run;
proc princomp data=rest.questionario;
var a1-a9;
run;
proc princomp data=rest.questionario
out=rest.data_coord;
var a1-a9;
run;
proc corr data=rest.data_coord;
var prin1-prin9;
run;
data rest.data_coord_1;set rest.data_coord;
avg_i=.;
avg_i=mean(of a1-a9);
run;
proc corr data=rest.data_coord_1;
var avg_i prin1;
run;
data rest.data_adj; set rest.questionario;
avg_i=mean(of a1-a9);
```

```
min_i=min(of a1-a9);
max_i=max(of a1-a9);
array a a1-a9;
array b new1-new9;
do over b;
b=.;
if a>avg_i then b=(a-avg_i)/(max_i-avg_i);
if a<avg_i then b=(a-avg_i)/(avg_i-min_i);
if a=avg_i then b=0;
if a=. then b=0;
end;
label new1='privacy';
label new2='security';
label new3='feature';
label new4='search';
label new5='interface';
label new6='customization';
label new7='integration';
label new8='mobileinterface';
label new9='speed';
run;
proc princomp data=rest.data_adj
out=rest.data_coord_1_adj;
var new1-new9;
run;
proc cluster data=rest.data_coord_1_adj method=ward outtree=rest.data_tree;
var prin1-prin4;
id obs_id;
run;
proc template;
define statgraph dendrogram;
begingraph;
layout overlay;
dendrogram nodeID=_name parentID=_parent_clusterheight=_height_;
endlayout;
endgraph;
end;
run;
proc sgrender data=rest.data_tree template=dendrogram;
run;
proc tree data=rest.data_tree ncl=5 out=rest.data_cluster;
id obs_id;
run;
proc sort data=rest.data_coord_1_adj;
by obs_id;
run;
proc sort data=rest.data_cluster;
by obs_id;
run;
data rest.data_merged;
merge rest.data_coord_1_adj rest.data_cluster;
by obs_id;
run;
data rest.data_fake;
set rest.data_merged;
cluster=5;
run;
data rest.data_appendend;
set rest.data_merged rest.data_fake;
run;
```

```
proc ttest data=rest.data_appendend;
var new:;
class cluster;
where cluster=1 or cluster=5;
run;
proc ttest data=rest.data_appendend;
var new:;
class cluster;
where cluster=2 or cluster=5;
run;
proc ttest data=rest.data_appendend;
var new:;
class cluster;
where cluster=3 or cluster=5;
run;
proc ttest data=rest.data_appendend;
var new:;
class cluster;
where cluster=4 or cluster=5;
run;
proc freq data=rest.data_merged;
table d1*cluster/expected chisq;
run;
proc freq data=rest.data_merged;
table d2*cluster/expected chisq;
run;
proc freq data=rest.data_merged;
table d3*cluster/expected chisq;
run;
proc freq data=rest.data_merged;
table d4*cluster/expected chisq;
run;
proc freq data=rest.data_merged;
table d5*cluster/expected chisq;
run;
```