```sas
libname c 'C:\Users\pietro.marino4\Desktop\BI';
run;
data c.datamobility;
set datamobility;
run;

data c.datamobilityclean;
set c.datamobility;
label
d1='Gender'
d2='Age'
d3='Occupation'
d4='Origin'
d5='Population'
d6='Offsite'
b1='Preference'
b2='Frequency'
b3='Spending'
c1='Car'
c2='Bike'
c3='Scooter'
c4='Motorbike'
c5='Carpooling'
p1='Enjoy'
p2='Corrente'
p3='Car2go'
p4='Drivenow'
p5='Zipcar'
p6='Mobike'
p7='Jump'
p8='Bikemi'
p9='Lime'
p10='Bird'
p11='Helbiz'
p12='ECooltra'
p13='Zigzag'
p14='Mimoto'
p15='Blablacar'
e1='Comfort'
e2='Payment'
r1='Money'
r2='Environment'
r3='Time&Energy'
r4='Parking'
r5='Availability'
r6='Safety'
r7='Comfy'
r8='Risk'
r9='Ztl'
r10='Variety'
id=id;
run;
proc contents data=c.datamobilityclean;
run;

proc freq data=c.datamobilityclean;
table d1--d6;
run;

proc freq data=c.datamobilityclean;
table b1--b3;
run;
```

```sas
proc freq data=c.datamobilityclean;
table e1--e2;

proc freq data=c.datamobilityclean;
table c1--c5;
run;

proc freq data=c.datamobilityclean;
table p1--p15;
run;
proc means data=c.datamobilityclean;
var r1--r10;
run;
proc means data=c.datamobilityclean;
var r1--r10;
class d1;
run;
proc means data=c.datamobilityclean;
var r1--r10;
class d2;
run;
proc means data=c.datamobilityclean;
var r1--r10;
class d3;
run;
proc means data=c.datamobilityclean;
var r1--r10;
class d4;
run;
proc means data=c.datamobilityclean;
var r1--r10;
class d5;
run;
proc means data=c.datamobilityclean;
var r1--r10;
class d6;
run;
proc means data=c.datamobilityclean;
var r1--r10;
class b1;
run;
proc means data=c.datamobilityclean;
var r1--r10;
class b2;
run;
proc means data=c.datamobilityclean;
var r1--r10;
class b3;
run;
proc means data=c.datamobilityclean;
var r1--r10;
class e1;
run;
proc means data=c.datamobilityclean;
var r1--r10;
class e2;
run;
/*PRINCIPAL COMPONENT ANALYSIS*/
proc princomp data=c.datamobilityclean out=c.datamobility2;
var r1--r10;
run;
```

```sas
data c.datamobilitysentinel3; set c.datamobility2;
avg_i=.;
avg_i=mean(of r1--r10);
run;

proc corr data=c.datamobilitysentinel3;
var avg_i prin1;
run;

/*REMOVAL OF SIZE EFFECT*/
data c.datamobilityclean3;
set c.datamobilityclean;
Minimum=min (of r1--r10);
Maximum=max (of r1--r10);
Mean=mean (of r1--r10);
Array x1 r1--r10;
Array x2 new_1-new_10;
Do over x2;
X2=.;
If x1<mean then x2=(x1-mean)/(mean-minimum);
If x1>mean then x2=(x1-mean)/(maximum-mean);
If x1=mean then x2=0;
If x1=minimum then x2=-1;
If x1=maximum then x2=1;
If x1=. then x2=0;
End;
Run;

data c.datamobilityclean3;
set c.datamobilityclean3;
label
new_1='Money se'
new_2='Environment se'
new_3='Time&Energy se'
new_4='Parking se'
new_5='Availability se'
new_6='Safety se'
new_7='Comfy se'
new_8='Risk se'
new_9='Ztl se'
new_10='Variety se';
run;

proc means data=c.datamobilityclean3;
var new_1--new_10;
run;
proc princomp data=c.datamobilityclean3 out=c.datamobilityclean4;
var new_1--new_10;
run;


/*CLUSTERING*/
proc cluster data=c.datamobilityclean4 method=single;
var prin1-prin5;
id id;
run;
proc tree;
run;

proc cluster data=c.datamobilityclean4 method=ward;
var prin1-prin5;
id id;
run;
```

```sas
proc tree;
run;

proc tree ncl=4 out=c.datamobilityclean_cluster;
id id;
run;

proc freq data=c.datamobilityclean_cluster;
table cluster;
run;

proc sort data=c.datamobilityclean_cluster;
by id;
run;
proc sort data=c.datamobilityclean4;
by id;
run;

data c.datamobilityclean5; merge c.datamobilityclean4
c.datamobilityclean_cluster;
by id;
run;


proc means data=c.datamobilityclean5;
var new_1-new_10;
run;
proc means data=c.datamobilityclean5;
var new_1-new_10;
class cluster;
run;


/*T-TEST*/
data c.datamobilityclean5_fake;
set c.datamobilityclean5;
cluster=5;
run;


data c.datamobilityclean5_ttest;
set c.datamobilityclean5 c.datamobilityclean5_fake;
run;


%macro do_k_cluster;
%do k=1 %to 4;
proc ttest data=c.datamobilityclean5_ttest;
where cluster=&k or cluster=5;
class cluster;
var new:;
ods output ttests=c.datacluster_&k (where=( method='Satterthwaite')
rename=(tvalue=tvalue_&k) rename=(probt=prob_&k));
run;
%end;
%mend do_k_cluster;
%do_k_cluster;



proc sort data=c.datacluster_1;
by variable;
run;
```

```sas
proc sort data=c.datacluster_2;
by variable;
run;
proc sort data=c.datacluster_3;
by variable;
run;
proc sort data=c.datacluster_4;
by variable;
run;

data c.datacluster_all; merge c.datacluster_1 c.datacluster_2 c.datacluster_3
c.datacluster_4;
by variable;
run;

data c.cluster_all_label;
set c.datacluster_all;
if variable='new_1' then label='Save Money';
if variable='new_2' then label='Environmental Benefits';
if variable='new_3' then label='Save Time&Energies';
if variable='new_4' then label='Disposable Parkings';
if variable='new_5' then label='Number of Available Means';
if variable='new_6' then label='Safety of Means';
if variable='new_7' then label='Comfort of Means';
if variable='new_8' then label='Avoid Risk of Theft';
if variable='new_9' then label='Possibility to Circulate in ZTL';
if variable='new_10' then label='Variety of Services';
run;

proc freq data=c.datamobilityclean5;
table (d1 d2 d3 d4 d5 d6 b1 b2 b3 e1 e2)*cluster / expected chisq ;
run;


data c.datamobilitycheck; set c.datamobilityclean5;
cluster3=.;
if cluster=3 then cluster3=1;
else cluster3=2;
run;

proc freq data=c.datamobilitycheck;
table cluster3*d4 / expected chisq;
run;
```