```
/* Read Table and cleaning the data*/
/* read the table */

FILENAME REFFILE '/folders/myfolders/data/noshow.csv';

PROC IMPORT DATAFILE=REFFILE DBMS=CSV OUT=noshow replace;
    GETNAMES=YES;


RUN;

/* drop the columns */

data noshow;
set noshow;
drop  PatientID AppointmentId ScheduledDay AppointmentDay neighbourhood;
run;

/* format No_show column */

data noshow;
format No_show $20.;
set noshow;
run;

/* change lebel */
data noshow;
    set noshow;

    if No_show='No' then
        No_show='Showed up';
    else
        No_show='not showed up';
run;

/* Check the data*/
PROC PRINT DATA=noshow (firstobs=1 obs=5);
RUN;

/* check structure of the data */
Proc contents data=noshow;
run;

/* change column name of the data */
data noshow;
    set noshow;
    rename Hipertension=Hypertension Handcap=Handicap;
run;

/* check missing value  */
/* numeric value column */
proc means data=noshow n nmiss;
    var _numeric_;
run;


/* character value column*/
```

```
proc freq data=noshow;
    tables No_show Gender ;
run;

/* other way to check missing value */
proc format;
    value $missfmt ' '='Missing' other='Not Missing';
    value missfmt .='Missing' other='Not Missing';
run;

proc freq data=noshow;
    format _CHARACTER_ $missfmt.;
    tables _CHARACTER_ / missing missprint nocum nopercent;
    format _NUMERIC_ missfmt.;
    tables _NUMERIC_ / missing missprint nocum nopercent;

    /* summary statistics */
proc means data=noshow;
    var age;
run;

/* categorical variable */
proc freq data=noshow;
    table gender;
run;

proc univariate data=noshow;
    var age;
run;

proc freq data=noshow;
    table gender hypertension diabetes No_show;
run;

/* summary stattistics of all numeric variable */
proc means data=noshow;
run;

/* drop age less than 0 value */
/* find the observation */
proc print data=noshow;
    where age < 0;
run;

/* delete the observation less than 0 */
data noshow;
    set noshow;

    if age<0 then
        delete;
run;

/* check agian the mean */
proc means data=noshow;
    var age;
run;
```

```
/* EXPLORATORY ANALYSIS */


/* age vs no show histogram */
proc sgplot data=noshow;
    histogram age / group=No_show transparency=0.4;

run;

/* age vs noshow box plot */
PROC SGPLOT DATA=noshow;
    VBOX Age / category=No_show;

run;

/* bar diagram of age vs no show */
proc SGPLOT DATA=noshow;
    vbar age / group=No_show;

run;

/* t test */
proc ttest data=noshow sides=2 h0=0 plots(showh0);
    class No_show;
    var Age;
run;

/* gender vs no show */
/* bar diagram */
proc SGPLOT DATA=noshow;
    vbar Gender;

run;

/* stacked bar diagram */
proc SGPLOT DATA=noshow;
    vbar Gender / group=No_show;

run;

/* other way */
proc freq data=noshow order=freq;
    tables No_show*Gender / plots=freqplot(twoway=stacked);
run;

/* proportion */
proc freq data=noshow order=freq;
    tables No_show*Gender / plots=freqplot(twoway=stacked scale=grouppct);
run;

/* chi square test */
/* with plot */
proc freq data=noshow;
    tables (gender)*(No_show) / chisq plots=(freqplot(twoway=grouphorizontal
```

```
        scale=percent));
run;

/* simple way */
proc freq data=noshow;
    tables (gender)*(No_show) / chisq;
run;

/* no show vs scholarship */
proc freq data=noshow order=freq;
    tables No_show*scholarship/ plots=freqplot(twoway=stacked);
run;

proc freq data=noshow;
    tables (scholarship)*(No_show) / chisq plots=(freqplot(twoway=grouphorizontal
        scale=percent));
run;

/* no show vs Hypertension*/
proc freq data=noshow order=freq;
    tables No_show*hypertension/ plots=freqplot(twoway=stacked);
run;

proc freq data=noshow;
    tables (hypertension)*(No_show) / chisq plots=(freqplot(twoway=grouphorizontal
        scale=percent));
run;

/* no show vs diabetes*/
proc freq data=noshow order=freq;
    tables No_show*diabetes/ plots=freqplot(twoway=stacked);
run;

proc freq data=noshow;
    tables (diabetes)*(No_show) / chisq plots=(freqplot(twoway=grouphorizontal
        scale=percent));
run;

/* no show vs alcoholism*/
proc freq data=noshow order=freq;
    tables No_show*alcoholism/ plots=freqplot(twoway=stacked);
run;

proc freq data=noshow;
    tables (alcoholism)*(No_show) / chisq plots=(freqplot(twoway=grouphorizontal
        scale=percent));
run;

/* no show vs handicap*/
proc freq data=noshow order=freq;
    tables No_show*handicap/ plots=freqplot(twoway=stacked);
run;

proc freq data=noshow;
    tables (handicap)*(No_show) / chisq plots=(freqplot(twoway=grouphorizontal
        scale=percent));
run;
```

```
/* no show vs sms received*/
proc freq data=noshow order=freq;
    tables No_show*sms_received/ plots=freqplot(twoway=stacked);
run;

proc freq data=noshow;
    tables (sms_received)*(No_show) / chisq plots=(freqplot(twoway=grouphorizontal
        scale=percent));
run;

/* appointment days vs No show */
/* Neighborhood vs No show */
/* logistic regression */
proc logistic data=noshow;
    class gender Scholarship Hypertension Diabetes Alcoholism Handicap
        SMS_received / param=glm;
    model No_show(event='Showed up')=gender Scholarship Hypertension Diabetes
        Alcoholism Handicap SMS_received Age / link=logit selection=backward
        slstay=0.05 hierarchy=single technique=fisher;
```