```
/*PROJECT: A study was conducted to determine whether a ‖DRUG A= helps to reduce
fasting sugar level. There are 3400 patients were enrolled for this study and
before and after fasting sugar level were taken at specific interval. */

*Loading permanent data;
libname Proj "C:\Users\Kanna\Documents\Rutgers Grad\Fall 2022\BINF5210_SAS";
filename Data1 "C:\Users\Kanna\Documents\Rutgers Grad\Fall 2022\BINF5210_SAS\Initial_Study.txt";
filename Data2 "C:\Users\Kanna\Documents\Rutgers Grad\Fall 2022\BINF5210_SAS\Secondary_Study.txt";

*Create and Print Data for First Study;
data Proj.First_Study;
    infile Data1 dlm=',' dsd missover;
    input Patient_ID
          Age
          State $
          LOS
          Total_Charge
          Initial_Sugar ;
    format Initial_Sugar 10.5;
run;

title "First Study";
proc print data=Proj.First_Study;
run;

*Create and Print Data for Second Study;
data Proj.Second_Study;
    infile Data2 dlm=',' dsd missover;
    input Patient_ID
          After_Sugar ;
    format After_Sugar 10.5;
run;

title "Second Study";
proc print data=Proj.Second_Study;
run;

*Sort Study one and Study two;
proc sort data=Proj.First_Study out=FirstStudy_Sorted;
    by Patient_ID State;
run;

proc sort data=Proj.Second_Study out=SecondStudy_Sorted;
    by Patient_ID;
run;

*Merge Files;
data Proj.Study;
    merge FirstStudy_Sorted (in=a)
          SecondStudy_Sorted (in=b);

        by Patient_ID;

        if a=1 and b=1;
run;
```

```
title "Merged Study Data";
proc print data=Proj.Study;
run;

*Eliminating extreme values within the data - will not influece statistical method (40 - 600);
*Format age to age_group;
data Proj.Study;
set Proj.Study;
    if Age <= 21 then AgeGroup = 'Young';
    else if Age > 21 AND Age <= 60 then AgeGroup = 'Adult';
    else if Age > 60 then AgeGroup = 'Old';

    if Initial_Sugar >= 40 AND After_Sugar >= 40
        AND Initial_Sugar <= 600 AND After_Sugar <= 600
        then output;
run;

title "Study Data";
proc print data=Proj.Study;
run;

*Surveyselect;
*For All data;
title "Study Subset";
proc surveyselect data=Proj.Study
                  out = Study_subset
                  method = srs
                  sampsize = 1000;
            id _all_;
run;

proc print data=Study_subset;
run;

*Final Data Contents;
proc contents data = Study_subset;
run;

*For Study_subset;
*Descriptive Stats;
title "Study Subset";
title2 "Means Procedure";
proc means data=Study_subset n mean min max;
    var Initial_Sugar After_Sugar;
run;

proc means data=Study_subset;
    class State AgeGroup;
    var Initial_Sugar After_Sugar;
run;

/*Cleaning up Blood sugar to levels cutoff at above 600 and below 40.
These numbers can create a dangerous situations for patients
- above 600 can cause diabetic coma which is hyperosmolar. Anything between 600-800 is abnormal
- below 40 is hypoglycemia, creating severe dehydration which can also leave to coma and/or death. Bet
```

```
title "Study Subset";
title2 "Frequency Procedure";
proc freq data=Study_subset;
    tables State*AgeGroup / nocum;
run;

*Univariate;
title "Study Subset";
title2 "Univariate";
proc univariate data=Proj.Study;
    var Initial_Sugar After_Sugar;
run;

*Running Paired t-test to analyze Before and After Drug A;
title "Study Subset";
title2 "Paired t-test";
proc ttest data=Study_subset;
paired Initial_Sugar * After_Sugar;
run;

*Running ANOVA to analyze age effect;
title "Study Subset";
title2 "One-Way ANOVA on AgeGroup";
proc ANOVA data=Study_subset;
    Class AgeGroup;
    Model After_Sugar = AgeGroup;
    Means AgeGroup / TUKEY;
run;

*Running ANOVA to analyze age and state interaction effect;
title "Study Subset";
title2 "Two-Way ANOVA: Interaction Effect of Age and State";
proc ANOVA data=Study_subset;
    Class AgeGroup State;
    Model After_Sugar = AgeGroup State
                    AgeGroup * State;
run;

*Correlation between LOS, initial_sugar, after_sugar;
Proc corr data = Study_subset;
        Var Initial_Sugar After_Sugar;
        With LOS;
Run;
```