```
/* Load Gender Submission */
proc import datafile="C:\Users\stoneleiker.AUTH\Desktop\AI Project\SAS (Titanic Machine Learning from
    out=gender
    dbms=csv
    replace;
    getnames=yes;
run;

/* Load Test Data */
proc import datafile="C:\Users\stoneleiker.AUTH\Desktop\AI Project\SAS (Titanic Machine Learning from
    out=test
    dbms=csv
    replace;
    getnames=yes;
run;

/* Load Train Data */
proc import datafile="C:\Users\stoneleiker.AUTH\Desktop\AI Project\SAS (Titanic Machine Learning from
    out=train
    dbms=csv
    replace;
    getnames=yes;
run;

/* Merge test data with gender_submission to add Survived */
proc sort data=test; by PassengerId; run;
proc sort data=gender; by PassengerId; run;

data test_full;
    merge test (in=a) gender (in=b);
    by PassengerId;
    if a;
run;

/* Combine train and test_full into one dataset */
data combined_all;
    set train test_full;
run;

/* Clean Combined Data */
data clean_all;
    set combined_all(keep=PassengerId Survived Name Sex Age SibSp Parch Embarked Pclass);
    sex = lowcase(strip(sex));
    if age >= 0 and age <= 100;
    FamilySize = SibSp + Parch + 1;
    if FamilySize = 1 then IsAlone = 1;
    else IsAlone = 0;
run;

/* Frequency table: Pclass by Embarked controlling for Survived */
proc freq data=clean_all;
    tables Pclass * Embarked * Survived / norow nocol nopercent;
run;

/* Frequency: FamilySize by Survived */
proc freq data=clean_all;
```

```
    tables FamilySize * Survived / norow nocol nopercent;
run;

/* Frequency: IsAlone by Survived */
proc freq data=clean_all;
    tables IsAlone * Survived / norow nocol nopercent;
run;

/* Logistic Regression Model */
proc logistic data=clean_all descending;
    class sex embarked pclass / param=ref;
    model Survived = sex age pclass embarked FamilySize IsAlone;
run;

/* Random Forest using HPFOREST */
proc hpforest data=clean_all maxtrees=100 seed=12345;
    target Survived;
    input sex age pclass embarked FamilySize IsAlone / level=nominal;
    id PassengerId;
run;

/* Survival by Sex */
proc sgplot data=clean_all;
    vbar sex / group=Survived groupdisplay=cluster stat=percent;
    title "Survival by Sex";
run;

/* Survival by Pclass */
proc sgplot data=clean_all;
    vbar pclass / group=Survived groupdisplay=cluster stat=percent;
    title "Survival by Pclass";
run;

/* Survival by Family Size */
proc sgplot data=clean_all;
    vbar FamilySize / group=Survived groupdisplay=cluster stat=percent;
    title "Survival by Family Size";
run;

/* FamilySize breakdown by Pclass */
proc freq data=clean_all;
    tables FamilySize * Pclass / norow nocol nopercent;
run;
```