```
/* Step 1: Import datasets */
proc import datafile="C:\Users\stoneleiker.AUTH\Desktop\AI Project\SAS (Titanic Machine Learning from
    out=train
    dbms=csv
    replace;
    getnames=yes;
run;

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

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

/* Step 2: Feature engineering for train */
data train;
    set train;
    FamilySize = SibSp + Parch + 1;
    IsAlone = (FamilySize = 1);
    Source = "Train";
run;

/* Step 3: Feature engineering for test */
data test;
    set test;
    FamilySize = SibSp + Parch + 1;
    IsAlone = (FamilySize = 1);
run;

/* Step 4: Merge test with gender_submission */
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;
    Source = "Test";
run;

/* Step 5: Combine train and test_full */
data combined_all;
    set train test_full;
run;

/* Step 6: Create clean dataset for modeling and plots */
data clean_all;
    set combined_all(keep=PassengerId Survived Name Sex Age SibSp Parch Embarked Pclass FamilySize IsA
```

```
run;

/* Step 7: Logistic regression on actual train data */
proc logistic data=train descending;
    class Sex Embarked Pclass / param=ref;
    model Survived = Sex Age Pclass Embarked FamilySize IsAlone;
run;

/* Step 8: Forest model (optional) */
proc hpforest data=train seed=12345;
    target Survived;
    input Sex Age Pclass Embarked FamilySize IsAlone / level=nominal;
    id PassengerId;
run;

/* Step 9: Bar chart - Survival rate by Passenger Class */
proc sgplot data=clean_all;
    vbar Pclass / group=Survived groupdisplay=cluster stat=percent;
    title "Survival Rate by Passenger Class";
run;

/* Step 10a: Frequency - Family Size by Class */
proc freq data=clean_all;
    tables FamilySize * Pclass / norow nocol nopercent;
    title "Family Size Distribution by Passenger Class";
run;

/* Step 10b: Frequency - Gender by Class */
proc freq data=clean_all;
    tables Pclass * Sex / norow nocol nopercent;
    title "Gender Distribution by Passenger Class";
run;

/* Step 11a: Frequency - Family Size by Survival */
proc freq data=clean_all;
    tables FamilySize * Survived / norow nocol nopercent;
    title "Survival by Family Size";
run;

/* Step 11b: Frequency - Survival by Port and Class */
proc freq data=clean_all;
    tables Embarked * Pclass * Survived / nopercent norow nocol;
    title "Survival by Embarkation Port and Passenger Class";
run;
```