

```

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

/* Clean Train Data */
data clean_train;
    set train(keep=PassengerId Survived Name Sex Age SibSp Parch Embarked Pclass);
    sex = lowercase(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_train;
    tables Pclass * Embarked * Survived / norow nocol nopercent;
run;

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

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

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

/* Random Forest using HPFOREST */

```

```

proc hpforest data=clean_train 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_train;
  vbar sex / group=Survived groupdisplay=cluster stat=percent;
  title "Survival by Sex";
run;

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

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

/* Step 1: Create FamilySize column */
data family_size_detail;
  set clean_train;
  FamilySize = SibSp + Parch + 1;
run;

/* Step 2: Frequency Table: FamilySize (1,2,3,4,5,6,7,8, etc) by Pclass */
proc freq data=family_size_detail;
  tables FamilySize * Pclass / norow nocol nopercnt;
run;

```