```
/*=====================================================
   IMPORTING AND VIEWING OF DATASET
=====================================================*/
proc import datafile="/home/u64153637/heart_failure_clinical_records_dataset.csv"
    out=mydata                      /*naming the dataset*/
    dbms=csv
    replace;                        /*overwrites the dataset if it exists*/
    guessingrows=max;               /* ensures SAS guessess datatypes correctly*/
    GETNAMES=YES;                   /* Use first row as variable names*/
    DATAROW=2;                      /* Start reading data from row 2 */
run;
proc contents data=mydata;          /*gives a detailed description of the dataset structure*/
run;
proc means data=mydata n nmiss min max mean std;
run;


/*=====================================================
   DATA CLEANING
=====================================================*/
      /*REMOVE DUPLICATES (IF ANY) */
proc sort data=mydata nodupkey out=mydata_nodup;
    by _all_;
run;
      /*HANDLE OUTLIERS (remove impossible ages) */
data mydata_clean;
    set mydata_nodup;
    if age < 0 or age > 120 then delete;
run;
      /*FINAL CLEANED DATASET SUMMARY*/
proc means data=mydata_clean n nmiss min max mean std;
run;

/*=====================================================
   UNADJUSTED KAPLAN-MEIER CURVE
   Survival by High Blood Pressure (no other covariates)
=====================================================*/
proc lifetest data=mydata_clean plots=survival(atrisk(maxlen=13 outside) test)notable;
    time time*DEATH_EVENT(0);               /* time variable, censoring */
    strata high_blood_pressure;             /* compare groups */
    title "Unadjusted Kaplan-Meier Survival Curves by High Blood Pressure";
run;

/*=====================================================
   ADJUSTED COX PROPORTIONAL HAZARDS MODEL
   Adjusting for other patient characteristics
=====================================================*/
proc phreg data=mydata_clean;
    class high_blood_pressure;
    model time*DEATH_EVENT(0) = high_blood_pressure
                                age
                                ejection_fraction
                                serum_creatinine
                                anaemia
                                creatinine_phosphokinase
                                serum_sodium;
    hazardratio 'Effect of High BP' high_blood_pressure;
    baseline out=cox_adj_surv survival=SurvProb / group=high_blood_pressure method=pl;
    title "Adjusted Cox Survival Curves by hign BP";
run;
```

```
      /* standardizing continuous variables */
proc standard data=mydata_clean mean=0 std=1 out=mydata_std;
    var age creatinine_phosphokinase ejection_fraction serum_creatinine serum_sodium platelets time;
run;
/*=====================================================
   UNADJUSTED KAPLAN-MEIER CURVE (Standardised Variables)
   Survival by High Blood Pressure (no other covariates)
=====================================================*/
proc lifetest data=mydata_std plots=survival(atrisk(maxlen=13 outside) test)notable;
    time time*DEATH_EVENT(0);               /* time variable, censoring */
    strata high_blood_pressure;             /* compare groups */
    title "Unadjusted Kaplan-Meier Survival Curves by High Blood Pressure (Standardised Var)";
run;

/*=====================================================
   ADJUSTED COX PROPORTIONAL HAZARDS MODEL (Standardised Variables)
   Adjusting for other patient characteristics
=====================================================*/
proc phreg data=mydata_std;
    class high_blood_pressure;
    model time*DEATH_EVENT(0) = high_blood_pressure
                                age
                                ejection_fraction
                                serum_creatinine
                                anaemia
                                creatinine_phosphokinase
                                serum_sodium;
    hazardratio 'Effect of High BP' high_blood_pressure;
    baseline out=cox_adj_surv_std survival=SurvProb / group=high_blood_pressure method=pl;
    title "Adjusted Cox Survival Curves by hign BP( Standardised Var)" ;
```