```sas
options validvarname=v7; /*change the Variable names of the tables to Valid SAS Name*/
proc import out=Cancer datafile="/home/u61150141/sasuser.v94/Data_Science/Project/Breast_Cancer.csv"
dbms=csv replace;
guessingrows=max; /*Column Grade has values 1,2,3 and anaplastic; Grade IV which was giving error in import guessingrows=max ∨
run;


proc freq data=Cancer;
    table grade;
run;

/*Data Cleaning*/
data Cancer;
    set Cancer;
    if Grade='anaplastic; Grade IV' then Grade=4;

run;

/*Checking count after updating the column */
proc freq data=Cancer;
    table grade;
run;



/*Creating Dummy Variable*/

data Cancer;
    set Cancer;
    /*Dummy Variable for Race with Race White as base value*/
    if Race='Black' then Race_Blk=1; else Race_Blk=0;
    if Race='Other' then Race_Oth=1; else Race_Oth=0;

    /*Dummy Variable for Marital_Status with Marital_Status Married as base value*/
    if Marital_Status='Divorced' then MS_DRC=1; else MS_DRC=0;
    if Marital_Status='Separated' then MS_Sep=1; else MS_Sep=0;
    if Marital_Status='Single' then MS_Sin=1; else MS_Sin=0;
    if Marital_Status='Widowed' then MS_Wid=1; else MS_Wid=0;

    /*Dummy Variable for T_Stage with T_Stage T1 as base value*/
    if T_Stage='T2' then TS_T2=1; else TS_T2=0;
    if T_Stage='T3' then TS_T3=1; else TS_T3=0;
    if T_Stage='T4' then TS_T4=1; else TS_T4=0;

    /*Dummy Variable for N_Stage with N_Stage N1 as base value*/
    if N_Stage='N2' then NS_N2=1; else NS_N2=0;
    if N_Stage='N3' then NS_N3=1; else NS_N3=0;

    /*Dummy Variable for _6th_Stage with _6th_Stage IIA as base value*/
    if _6th_Stage='IIB' then SS_IIB=1; else SS_IIB=0;
    if _6th_Stage='IIIA' then SS_IIIA=1; else SS_IIIA=0;
    if _6th_Stage='IIIB' then SS_IIIB=1; else SS_IIIB=0;
    if _6th_Stage='IIIC' then SS_IIIC=1; else SS_IIIC=0;

    /*Dummy Variable for Differentiate with Differentiate 'Moderately differentiated' as base value*/
    if Differentiate='Poorly differentiated' then Diff_PD=1; else Diff_PD=0;
    if Differentiate='Undifferentiated' then Diff_UND=1; else Diff_UND=0;
    if Differentiate='Well differentiated' then Diff_WD=1; else Diff_WD=0;

    /*Dummy Variable for Grade with Grade '1' as base value*/
    if Grade=2 then Grd2=1; else Grd2=0;
    if Grade=3 then Grd3=1; else Grd3=0;
    if Grade=4 then Grd4=1; else Grd4=0;

    /*Dummy Variable for A_Stage with A_Stage Regional as base value*/
    if A_Stage='Distant' then ADistant=1; else ADistant=0;

    /*Dummy Variable for Estrogen_Status with Estrogen_Status Positive as base value*/
    if Estrogen_Status='Positive' then EstrogenN=1; else EstrogenN=0;

    /*Dummy Variable for Progesterone_Status with Progesterone_Status Positive as base value*/
    if Progesterone_Status='Positive' then ProgesteroneN=1; else ProgesteroneN=0;

    /*Dummy Variable for Status with Status Alive as base value*/
    if Status='Dead' then SDead=1; else SDead=0;

run;
```

```sas
proc freq data=Cancer;
    table Race Race_Blk Race_Oth Marital_Status MS_DRC MS_Sep MS_Sin MS_Wid T_Stage TS_T2 TS_T3 TS_T4 N_Stage NS_N2 NS_N3
    _6th_Stage SS_IIB SS_IIIA SS_IIIB SS_IIIC Differentiate Diff_PD Diff_UND Diff_WD Grade Grd2 Grd3 Grd4 A_Stage ADistant
    Estrogen_Status EstrogenN Progesterone_Status ProgesteroneN Status SDead;
run;



/*Splitting the data in 70:30*/

Proc surveyselect data=Cancer samprate=.7 method=SRS out=Cancer_Split outall seed=12345; /*Samprate - size of the trainset*/
/*Proc surveyselect is to split the data samprate is for sample rate, dividing the data in 70:30
, method SRS is for simple random sample seed all using the same numbers to split the data */
run;


data Cancer_train Cancer_Valid;
    set Cancer_Split;
    if selected=1 then output Cancer_train; else output Cancer_Valid;
run;


/*Checking for the distribution of Biased variable*/
proc freq data=Cancer;
    table Status;
run;

proc freq data=Cancer_train;
    table Status;
run;

proc freq data=Cancer_Valid;
    table Status;
run;



/*EDA*/

proc freq data=Cancer;
    table Race Marital_Status T_Stage N_Stage _6th_Stage Differentiate A_Stage Estrogen_Status Progesterone_Status;
run;

proc means data=Cancer n mean median min max q1 q3 stddev var maxdec=2;
    var Age;  /*class is to add categorical division to the statistics*/
run;

/*Age Distribution*/
Proc sgplot data=Cancer;
    histogram Age/nbins=10 scale=count;
    title "Age Count";
    xaxis label="'Age'";
run;


/*Age by Status*/
proc means data=Cancer n mean median min max q1 q3 stddev var maxdec=2;
    var Age; class Status; /* class is to add categorical division to the statistics*/
run;

Proc sgplot data=Cancer;
    vbox Age/category=Status;
    Title"Age by Status";
run;


/*Age by Sruvival_Months*/
proc means data=Cancer n mean median min max q1 q3 stddev var maxdec=2;
    var Survival_Months; class Status;
run;

Proc sgplot data=Cancer; /*sgplot is to create scatter plot*/
    scatter x=Age y=Survival_Months;
    title "Scatterplot of Age by Survival Months";
    xaxis label="'Age'";
```

```sas
    yaxis label="Survival Month";
run;


/*Survival_Months by Status*/
Proc sgplot data=Cancer;
    vbox Survival_Months/category=Status;
    Title"Survival Month by Status";
run;

/*Age by T_Stage*/
Proc sgplot data=Cancer;
    vbox Age/category=T_Stage;
    title "Age by T-Stage";
    xaxis label="'Age'";
    yaxis label="T-Stage";
run;

/*Bar Chart --Used for Categorical variables to show percent*/
Proc sgplot data=Cancer;
    vbar Estrogen_Status/stat=pct group=Status GROUPDISPLAY = CLUSTER;/*stat is to change frequency to %*/
    yaxis values=(0 to 1 by 0.1) label="Percent";/*changing scaling for y axis*/
    title "Estrogen_Status by Status";
run;

Proc sgplot data=Cancer;
    vbar Progesterone_Status/stat=pct group=Status GROUPDISPLAY = CLUSTER;/*stat is to change frequency to %*/
    yaxis values=(0 to 1 by 0.1) label="Percent";/*changing scaling for y axis*/
    title "Progesterone_Status by Status";
run;

/*Regional_Node_Positive by Status*/
proc sgplot data=Cancer;
    vbox reginol_node_positive/category=Status;
    Title"Reginol Node Positive by status";
run;


/* Correlation matrix */
proc corr data=Cancer;
    var age tumor_size regional_node_examined reginol_node_positive survival_months;
run;



/*-----------------------------------------------------SURVIVAL_MONTH-----------------------------------------------------
                 /*----------------------------Multiple Linear Regression------------------------------------*/


/*Multiple Linear Regression to predict Survival Months using all variables except Tumor_Size and by Proc Reg*/
Proc Reg data=Cancer_train; /*Reg = Regression*/
    Model Survival_Months=Age Race_Blk Race_Oth MS_DRC MS_Sep MS_Sin MS_Wid TS_T2 TS_T3 TS_T4 NS_N2 NS_N3
    SS_IIB SS_IIIA SS_IIIB SS_IIIC Diff_PD Diff_UND Diff_WD Grd2 Grd3 Grd4 ADistant
    EstrogenN ProgesteroneN SDead Regional_Node_Examined Reginol_Node_Positive/clb corrb vif ;
run;

/*Multiple Linear Regression to predict Survival Months using all variables except T_Stage and by Proc Reg*/
Proc Reg data=Cancer_train; /*Reg = Regression*/
    Model Survival_Months=Age Race_Blk Race_Oth MS_DRC MS_Sep MS_Sin MS_Wid Tumor_Size NS_N2 NS_N3
    SS_IIB SS_IIIA SS_IIIB SS_IIIC Diff_PD Diff_UND Diff_WD Grd2 Grd3 Grd4 ADistant
    EstrogenN ProgesteroneN SDead Regional_Node_Examined Reginol_Node_Positive/clb corrb vif ;
run;

/*Multiple Linear Regression to predict Survival Months using all variables except Tumor_Size and by Proc HpReg*/
proc hpreg data=Cancer seed=12345;
    partition fraction(validate=0.3);
    model Survival_Months=Age Race_Blk Race_Oth MS_DRC MS_Sep MS_Sin MS_Wid TS_T2 TS_T3 TS_T4 NS_N2 NS_N3
    SS_IIB SS_IIIA SS_IIIB SS_IIIC Diff_PD Diff_UND Diff_WD Grd2 Grd3 Grd4 ADistant
    EstrogenN ProgesteroneN SDead Regional_Node_Examined Reginol_Node_Positive;
run;

/*Multiple Linear Regression to predict Survival Months using all variables except T_Stage and by Proc HpReg*/
proc hpreg data=Cancer seed=12345;
    partition fraction(validate=0.3);
    model Survival_Months=Age Race_Blk Race_Oth MS_DRC MS_Sep MS_Sin MS_Wid TS_T2 TS_T3 TS_T4 NS_N2 NS_N3
    SS_IIB SS_IIIA SS_IIIB SS_IIIC Diff_PD Diff_UND Diff_WD Grd2 Grd3 Grd4 ADistant
    EstrogenN ProgesteroneN SDead Regional_Node_Examined Reginol_Node_Positive;
```

```sas
    selection method=stepwise(choose=validate);
run;



/*Variables selection*/

/*Multiple Linear Regression to predict Survival Months with Forward Selection*/
Proc Reg data=Cancer_train; /*Reg = Regression*/
    Model Survival_Months=Age Race_Blk Race_Oth MS_DRC MS_Sep MS_Sin MS_Wid TS_T2 TS_T3 TS_T4 NS_N2 NS_N3
    SS_IIB SS_IIIA SS_IIIB SS_IIIC Diff_PD Diff_UND Diff_WD Grd2 Grd3 Grd4 ADistant
    EstrogenN ProgesteroneN SDead Regional_Node_Examined Reginol_Node_Positive/selection=forward;
run;

/*Multiple Linear Regression to predict Survival Months with backward Selection*/
Proc Reg data=Cancer_train; /*Reg = Regression*/
    Model Survival_Months=Age Race_Blk Race_Oth MS_DRC MS_Sep MS_Sin MS_Wid TS_T2 TS_T3 TS_T4 NS_N2 NS_N3
    SS_IIB SS_IIIA SS_IIIB SS_IIIC Diff_PD Diff_UND Diff_WD Grd2 Grd3 Grd4 ADistant
    EstrogenN ProgesteroneN SDead Regional_Node_Examined Reginol_Node_Positive/selection=backward ;
run;

/*Multiple Linear Regression to predict Survival Months with Stepwise Selection with T_Stage*/
Proc Reg data=Cancer_train; /*Reg = Regression*/
    Model Survival_Months=Age Race_Blk Race_Oth MS_DRC MS_Sep MS_Sin MS_Wid TS_T2 TS_T3 TS_T4 NS_N2 NS_N3
    SS_IIB SS_IIIA SS_IIIB SS_IIIC Diff_PD Diff_UND Diff_WD Grd2 Grd3 Grd4 ADistant
    EstrogenN ProgesteroneN SDead Regional_Node_Examined Reginol_Node_Positive/selection=Stepwise ;
run;

/*Multiple Linear Regression to predict Survival Months with Stepwise Selection with Tumor_Size*/
Proc Reg data=Cancer_train; /*Reg = Regression*/
    Model Survival_Months=Age Race_Blk Race_Oth MS_DRC MS_Sep MS_Sin MS_Wid Tumor_Size NS_N2 NS_N3
    SS_IIB SS_IIIA SS_IIIB SS_IIIC Diff_PD Diff_UND Diff_WD Grd2 Grd3 Grd4 ADistant
    EstrogenN ProgesteroneN SDead Regional_Node_Examined Reginol_Node_Positive/selection=Stepwise ;
run;



/*Hpreg with t_Stage and all variables with stepwise selection*/
proc hpreg data=Cancer seed=12345;
    partition fraction(validate=0.3);
    model Survival_Months=Age Race_Blk Race_Oth MS_DRC MS_Sep MS_Sin MS_Wid TS_T2 TS_T3 TS_T4 NS_N2 NS_N3
    SS_IIB SS_IIIA SS_IIIB SS_IIIC Diff_PD Diff_UND Diff_WD Grd2 Grd3 Grd4 ADistant
    EstrogenN ProgesteroneN SDead Regional_Node_Examined Reginol_Node_Positive;
    selection method=stepwise(choose=validate);
run;

/*Hpreg with Tumor_size and all variables with stepwise selection*/
proc hpreg data=Cancer seed=12345;
    partition fraction(validate=0.3);
    model Survival_Months=Age Race_Blk Race_Oth MS_DRC MS_Sep MS_Sin MS_Wid Tumor_Size NS_N2 NS_N3
    SS_IIB SS_IIIA SS_IIIB SS_IIIC Diff_PD Diff_UND Diff_WD Grd2 Grd3 Grd4 ADistant
    EstrogenN ProgesteroneN SDead Regional_Node_Examined Reginol_Node_Positive;
    selection method=stepwise(choose=validate);
run;



/*with SDead, EstrogenN, Tumor_Size and SS_IIB, SS_IIIA, SS_IIIB, SS_II*/
proc hpreg data=Cancer seed=12345;
    partition fraction(validate=0.3);
    model Survival_Months=Tumor_size SS_IIB SS_IIIA SS_IIIB SS_IIIC EstrogenN SDead;
    selection method=stepwise(choose=validate);
run;

/*with SDead, EstrogenN, Tumor_Size */
proc hpreg data=Cancer seed=12345;
    partition fraction(validate=0.3);
    model Survival_Months=Tumor_size EstrogenN SDead;
    selection method=stepwise(choose=validate);
run;

/*with SDead, EstrogenN, TS_T2 TS_T3 TS_T4 */
proc hpreg data=Cancer seed=12345;
    partition fraction(validate=0.3);
    model Survival_Months=TS_T2 TS_T3 TS_T4 EstrogenN SDead;
    selection method=stepwise(choose=validate);
run;
```

```sas
/*with SDead, EstrogenN */
proc hpreg data=Cancer seed=12345;
    partition fraction(validate=0.3);
    model Survival_Months= EstrogenN SDead;
    selection method=stepwise(choose=validate);
run;

/*Final model with SDead, EstrogenN */

                /*---------------------------CART-------------------------------------*/

/* CART for Survival months*/
/* with T stage */
proc hpsplit data=cancer nodes=detail;
    class Race Marital_Status T_Stage N_Stage _6th_Stage Differentiate A_Stage Grade Estrogen_Status Progesterone_Status statu
    model survival_months=age Race Marital_Status T_Stage N_Stage _6th_Stage Differentiate A_Stage Grade Estrogen_Status Proge
 regional_node_examined reginol_node_positive status;
    grow rss;
    prune cc;
    partition fraction(validate=0.3 seed=12345);
Run;

/* with tumor_size */
proc hpsplit data=cancer nodes=detail;
    class Race Marital_Status N_Stage _6th_Stage Differentiate A_Stage Grade Estrogen_Status Progesterone_Status status;
    model survival_months=age Race Marital_Status N_Stage _6th_Stage Differentiate A_Stage Grade Estrogen_Status Progesterone_
    tumor_size regional_node_examined reginol_node_positive status;
    grow rss;
    prune cc;
    partition fraction(validate=0.3 seed=12345);
run;

proc means data=cancer n min max mean median stddev var;
    var survival_months;
run;

/* with only important variables  from CART -*/
proc hpsplit data=cancer nodes=detail;
    class Estrogen_Status status;
    model survival_months=Estrogen_Status status;
    grow rss;
    prune cc;
    partition fraction(validate=0.3 seed=12345);
run;


        /*--------------------------Neural Netwrok-----------------------------------*/

/* NN for Survival months */
/* with Tumor_size*/
proc hpneural data=cancer;
    partition fraction(validate=0.3 seed=12345);
    target survival_months/level=int;
    input age tumor_size regional_node_examined reginol_node_positive /level=int;
    input Race Marital_Status N_Stage _6th_Stage Differentiate A_Stage Grade Estrogen_Status Progesterone_Status status/level
    hidden 14;
    train maxiter=1000 numtries=5;
run;

/* with T stage */
proc hpneural data=cancer;
    partition fraction(validate=0.3 seed=12345);
    target survival_months/level=int;
    input age regional_node_examined reginol_node_positive /level=int;
    input Race Marital_Status T_Stage N_Stage _6th_Stage Differentiate A_Stage Grade Estrogen_Status Progesterone_Status statu
    hidden 14;
    train maxiter=1000 numtries=5;
run;

/* using important variables from CART tumor_size*/
proc hpneural data=cancer;
    partition fraction(validate=0.3 seed=12345);
    target survival_months/level=int;
    *input age tumor_size regional_node_examined/level=int;
    input Estrogen_Status status/level=nom;
    hidden 2;
    train maxiter=1000 numtries=5;
```

```sas
Run;




/*-----------------------------------------------Status----------------------------------------------------*/

                      /*---------------------------Logistic Regression-----------------------------------*/

/*Logistic Regression with all variables and T_Stage*/
proc logistic data=Cancer_train outmodel=Cancer_model1; /*save all the parameter estimates for future analysis */
    Model SDead(event='1')=Age Race_Blk Race_Oth MS_DRC MS_Sep MS_Sin MS_Wid TS_T2 TS_T3 TS_T4 NS_N2 NS_N3
    SS_IIB SS_IIIA SS_IIIB SS_IIIC Diff_PD Diff_UND Diff_WD Grd2 Grd3 Grd4 ADistant
    EstrogenN ProgesteroneN Survival_Months Regional_Node_Examined Reginol_Node_Positive;
run;

/*Logistic Regression with all variables and Tumor_Size*/
proc logistic data=Cancer_train outmodel=Cancer_model1; /*save all the parameter estimates for future analysis */
    Model SDead(event='1')=Age Race_Blk Race_Oth MS_DRC MS_Sep MS_Sin MS_Wid Tumor_Size NS_N2 NS_N3
    SS_IIB SS_IIIA SS_IIIB SS_IIIC Diff_PD Diff_UND Diff_WD Grd2 Grd3 Grd4 ADistant
    EstrogenN ProgesteroneN Survival_Months Regional_Node_Examined Reginol_Node_Positive;
run;
/*Tumor_Size is not statistically Significant*/

/*Logistic Regression with all variables and T_Stage with forward selection*/
proc logistic data=Cancer_train outmodel=Cancer_model1; /*save all the parameter estimates for future analysis */
    Model SDead(event='1')=Age Race_Blk Race_Oth MS_DRC MS_Sep MS_Sin MS_Wid TS_T2 TS_T3 TS_T4 NS_N2 NS_N3
    SS_IIB SS_IIIA SS_IIIB SS_IIIC Diff_PD Diff_UND Diff_WD Grd2 Grd3 Grd4 ADistant
    EstrogenN ProgesteroneN Survival_Months Regional_Node_Examined Reginol_Node_Positive/selection=forward;
run;

/*Logistic Regression with all variables and T_Stage with backward selection*/
proc logistic data=Cancer_train outmodel=Cancer_model1;
    Model SDead(event='1')=Age Race_Blk Race_Oth MS_DRC MS_Sep MS_Sin MS_Wid TS_T2 TS_T3 TS_T4 NS_N2 NS_N3
    SS_IIB SS_IIIA SS_IIIB SS_IIIC Diff_PD Diff_UND Diff_WD Grd2 Grd3 Grd4 ADistant
    EstrogenN ProgesteroneN Survival_Months Regional_Node_Examined Reginol_Node_Positive/selection=backward;
run;

/*Logistic Regression with all variables and T_Stage with Stepwise selection*/
proc logistic data=Cancer_train outmodel=Cancer_model1;
    Model SDead(event='1')=Age Race_Blk Race_Oth MS_DRC MS_Sep MS_Sin MS_Wid TS_T2 TS_T3 TS_T4 NS_N2 NS_N3
    SS_IIB SS_IIIA SS_IIIB SS_IIIC Diff_PD Diff_UND Diff_WD Grd2 Grd3 Grd4 ADistant
    EstrogenN ProgesteroneN Survival_Months Regional_Node_Examined Reginol_Node_Positive/selection=stepwise;
run;

/*Tryin once again with Tumor_Size to recheck*/
/*Logistic Regression with all variables and Tumor_Size with Stepwise selection*/
proc logistic data=Cancer_train outmodel=Cancer_model1;
    Model SDead(event='0')=Age Race_Blk Race_Oth MS_DRC MS_Sep MS_Sin MS_Wid Tumor_Size NS_N2 NS_N3
    SS_IIB SS_IIIA SS_IIIB SS_IIIC Diff_PD Diff_UND Diff_WD Grd2 Grd3 Grd4 ADistant
    EstrogenN ProgesteroneN Survival_Months Regional_Node_Examined Reginol_Node_Positive/selection=stepwise;
run;
/*Removed from stepwise selection*/


/*Logistic Regression with important variablesfrom the Stepwise selection*/
proc logistic data=Cancer_train outmodel=Cancer_model1;
    Model SDead(event='1')=Age Race_Blk Race_Oth TS_T2 TS_T3 TS_T4 NS_N2 NS_N3
    Diff_PD Diff_UND Diff_WD Grd2 Grd3 Grd4 EstrogenN Survival_Months Regional_Node_Examined Reginol_Node_Positive;
run;

/*Logistic Regression with important variables from the Stepwise selection but removing t_stage*/
proc logistic data=Cancer_train outmodel=Cancer_model1;
    Model SDead(event='1')=Age Race_Blk Race_Oth TS_T2 TS_T3 TS_T4 NS_N2 NS_N3
    Diff_PD Diff_UND Diff_WD Grd2 Grd3 Grd4 EstrogenN Survival_Months Regional_Node_Examined Reginol_Node_Positive;
run;
/*No change in AUC by removing T_Stage so exculding T_Stage */

/*Logistic Regression with important variablesfrom the Stepwise selectionremoving t_STage and N_Stage*/
proc logistic data=Cancer_train outmodel=Cancer_model1;
    Model SDead(event='1')=Age Race_Blk Race_Oth
    Diff_PD Diff_UND Diff_WD Grd2 Grd3 Grd4 EstrogenN Survival_Months Regional_Node_Examined Reginol_Node_Positive;
run;
/*Removing N_STage reduced AUC and reverting to N_Stage and removing grade*/


proc logistic data=Cancer_train outmodel=Cancer_model1;
```

```sas
    Model SDead(event='1')=Age Race_Blk Race_Oth NS_N2 NS_N3
    Diff_PD Diff_UND Diff_WD EstrogenN Survival_Months Regional_Node_Examined Reginol_Node_Positive;
run;
/*Not much difference by removing Grade will test by removing Race */

proc logistic data=Cancer_train outmodel=Cancer_model1;
    Model SDead(event='1')=Age NS_N2 NS_N3
    Diff_PD Diff_UND Diff_WD EstrogenN Survival_Months Regional_Node_Examined Reginol_Node_Positive;
run;
/*Auc reduced a bit so reverting Race and removing Differentiate*/

proc logistic data=Cancer_train outmodel=Cancer_model1;
    Model SDead(event='1')=Age Race_Blk Race_Oth NS_N2 NS_N3
    EstrogenN Survival_Months Regional_Node_Examined Reginol_Node_Positive;
run;
/*AUC reduced further so reverting Differentiate*/

/*Final model with best AUC and Parsimony principle*/
proc logistic data=Cancer_train outmodel=Cancer_model1;
    Model SDead(event='1')=Age Race_Blk Race_Oth NS_N2 NS_N3
    Diff_PD Diff_UND Diff_WD EstrogenN Survival_Months Regional_Node_Examined Reginol_Node_Positive;
run;

proc logistic data=Cancer_train outmodel=Cancer_model1;
    Model SDead(event='1')=Age NS_N2 NS_N3
    EstrogenN Survival_Months Reginol_Node_Positive ProgesteroneN tumor_size;
run;


proc hpsplit data=cancer nodes=detail;
    class Race Marital_Status N_Stage _6th_Stage Differentiate A_Stage Grade Estrogen_Status Progesterone_Status status;
    model status(event="Dead")=age Race Marital_Status N_Stage _6th_Stage Differentiate A_Stage Grade Estrogen_Status Progeste
    survival_months tumor_size regional_node_examined reginol_node_positive;
    partition fraction(validate=0.3 seed=12345);
    grow gini;
    prune cc;
run;


proc hpsplit data=cancer nodes=detail;
    class N_Stage Estrogen_Status Progesterone_Status status;
    model status(event="Dead")=age N_Stage Estrogen_Status Progesterone_Status survival_months tumor_size reginol_node_positiv
    partition fraction(validate=0.3 seed=12345);
    grow gini;
    prune cc;
run;

proc hpsplit data=cancer nodes=detail;
    class N_Stage Race Estrogen_Status status Differentiate;
    model status(event="Dead")=Age Race N_Stage Differentiate Estrogen_Status Survival_Months Regional_Node_Examined Reginol_N
    partition fraction(validate=0.3 seed=12345);
    grow gini;
    prune cc;
```