

```

proc import out = cbb
datafile = "/home/u63824887/cbb.csv"
dbms = csv replace;
guessingrows = max;
run;

/* Check of the different results of postseason so I can make Bernoulli */

proc freq data = cbb;
table POSTSEASON;
run;

/* Creation of Bernoulli Postseason variable */
data cbb;
set cbb;
postseason_b = postseason;
if postseason in ("N/A","NA") then postseason_b = 0;
else postseason_b = 1;
run;

/* There are too many levels in team, but Conference can be used. Since there are a lot
of conferences as well we will divide them between Power 5, Mid-Major, & low-major */
proc freq data = cbb;
table conf;
run;

data cbb;
set cbb;
conf_group = conf;
/* Even though there are 6 power conferences, the P12 no longer exists, and it is now generally
known as the Power 5, so that is how I will designate it*/
if conf in ("ACC","B10","B12","SEC","BE","P12") then conf_group = "P5";
else if conf in ("A10","CUSA","Amer","MWC","MVC","WCC","MAC","SB") then conf_group = "MM";
else conf_group = "LM";
run;
/* I determined the classifications by 2 factors:
1. Generally accepted groupings across NCAA MBB
2. Conference NET rankings, a metric used for a power index
*/

/* We will use logistic regression using the following variables:
G: Games played
W: Games won
ADJOE: Adjusted Offensive Efficiency
ADJDE: Adjusted Defensive Efficiency
BARTHAG: Power rating (chance of beating an average D1 team)
EFG_O: effective field goal %
EFG_D: effective field goal % allowed
Adj_T: Adjusted tempo
*/

proc logistic data = cbb;
class conf_group / param = reference;
model postseason_b(event = '1') = G W ADJOE ADJDE BARTHAG EFG_O EFG_D Adj_T conf_group/clparm = both;
run;

proc freq data = cbb;
tables conf_group * postseason_b;
run;

proc logistic data = cbb;

```

```
class conf_group;  
model postseason_b(event = '1') = conf_group;  
run;
```