**DATA PREPARATION**

**ASSIGNMENT ASS05**

**DONE BY:**

**SUSHMITHA MANI**

**DATA PREPARATION ASSIGNMENT ASS05**

**QUESTION 1:** *Use Median imputation on the following variables: CL\_Limit, PD\_Limit, MP\_LIMIT, CREDITSCORENUM, INSURANCEEXPERIENCEDAYSNUM, PRECREDITTIERNUM, PRIORSWITCHESCOUNT, RATEMANUALNUM, and UWTIERNUM*

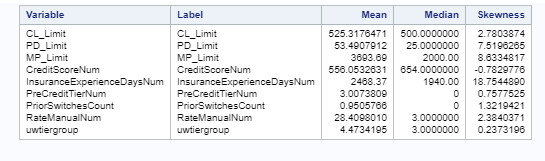
**ANSWER:**

**BEFORE TRANSFORMATION:**

proc means data=ASSG.FINALRAW chartype mean median vardef=df skewness qmethod=os;

var CL\_Limit PD\_Limit MP\_Limit CreditScoreNum InsuranceExperienceDaysNum PreCreditTierNum PriorSwitchesCount RateManualNum uwtiergroup;

run;

****

**(Table before Transformation)**

**AFTER TRANSFORMATION:**

ods noproctitle;

proc hpimpute data=work.finalraw out=work.import10;

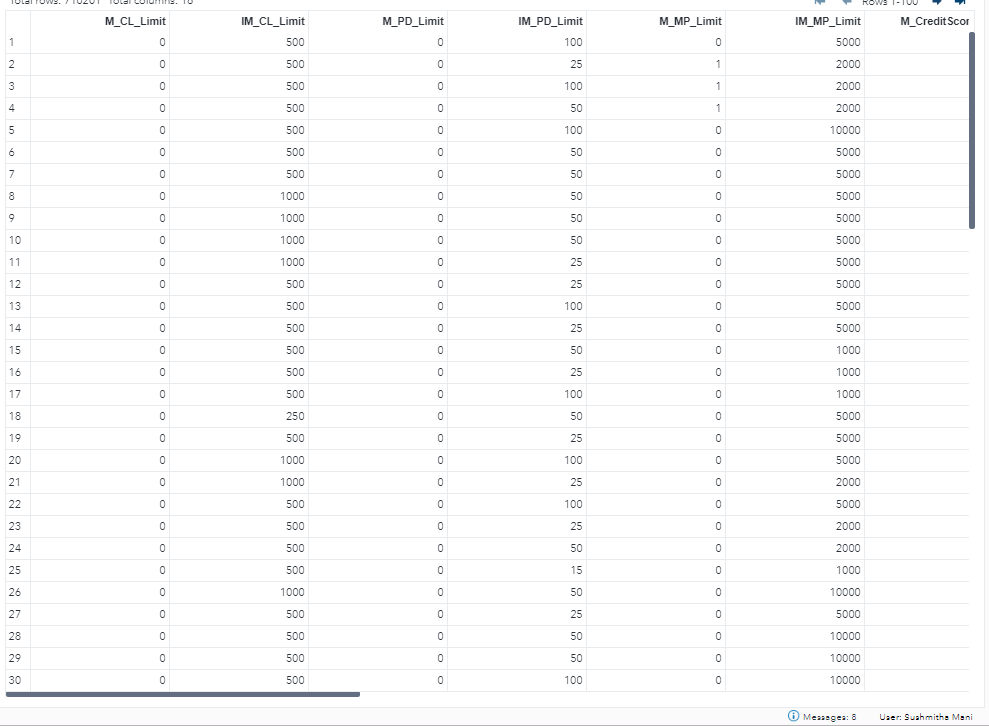
input CL\_Limit PD\_Limit MP\_Limit CreditScoreNum InsuranceExperienceDaysNum

PreCreditTierNum PriorSwitchesCount UWTierNum RateManualNum;

impute CL\_Limit PD\_Limit MP\_Limit CreditScoreNum InsuranceExperienceDaysNum

PreCreditTierNum PriorSwitchesCount UWTierNum RateManualNum / method= pmedian;

run;

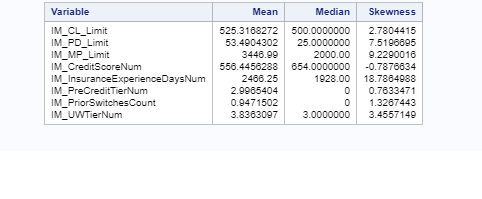


**(Output Data After Transformation)**

proc means data=ASSG.FINALRAW chartype mean median vardef=df skewness qmethod=os;

var CL\_Limit PD\_Limit MP\_Limit CreditScoreNum InsuranceExperienceDaysNum PreCreditTierNum PriorSwitchesCount RateManualNum uwtiergroup;

run;



**(Table after Transformation)**

**Question 2:** *Restrict the variable PNIAGE in the interval [16, 100].*

**ANSWER:**

LIBNAME libref 'C:\Users\97150\Documents\Assignmentss SAS';

proc sql;

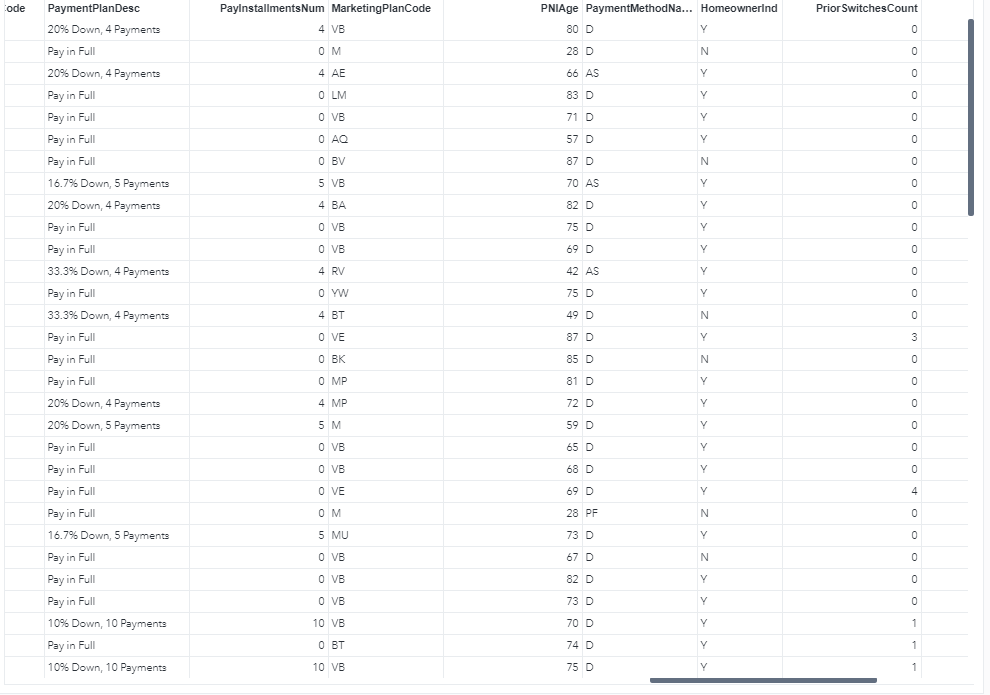
CREATE TABLE tab as

Select \*

From finalraw

where PNIAge>16 and PNIAge<100;

quit;



**(Output Data of PNIAge between interval 16 to 100)**

**Question 3:** *Perform rank transformation on the following variables RATEMANUALNUM N\_MOD\_BI\_LIMIT FIRMCODE AGEGE75\_LT30\_LT21\_POINTED PD\_LIMIT MP\_LIMIT MAXVEHVALUE PIP\_LIMIT NEXTPREMCH*

**ANSWER:**

**BEFORE TRANSFORMATION:**

proc rank data=WORK.FINALRAW out=work.Rank;

var RateManualNum agege75\_lt30\_lt21\_pointed PD\_Limit MP\_Limit maxvehvalue PIP\_Limit nextpremch FirmCode;

ranks rank\_RateManualNum rank\_agege75\_lt30\_lt21\_pointed rank\_PD\_Limit rank\_MP\_Limit rank\_maxvehvalue rank\_PIP\_Limit rank\_nextpremch rank\_FirmCode;

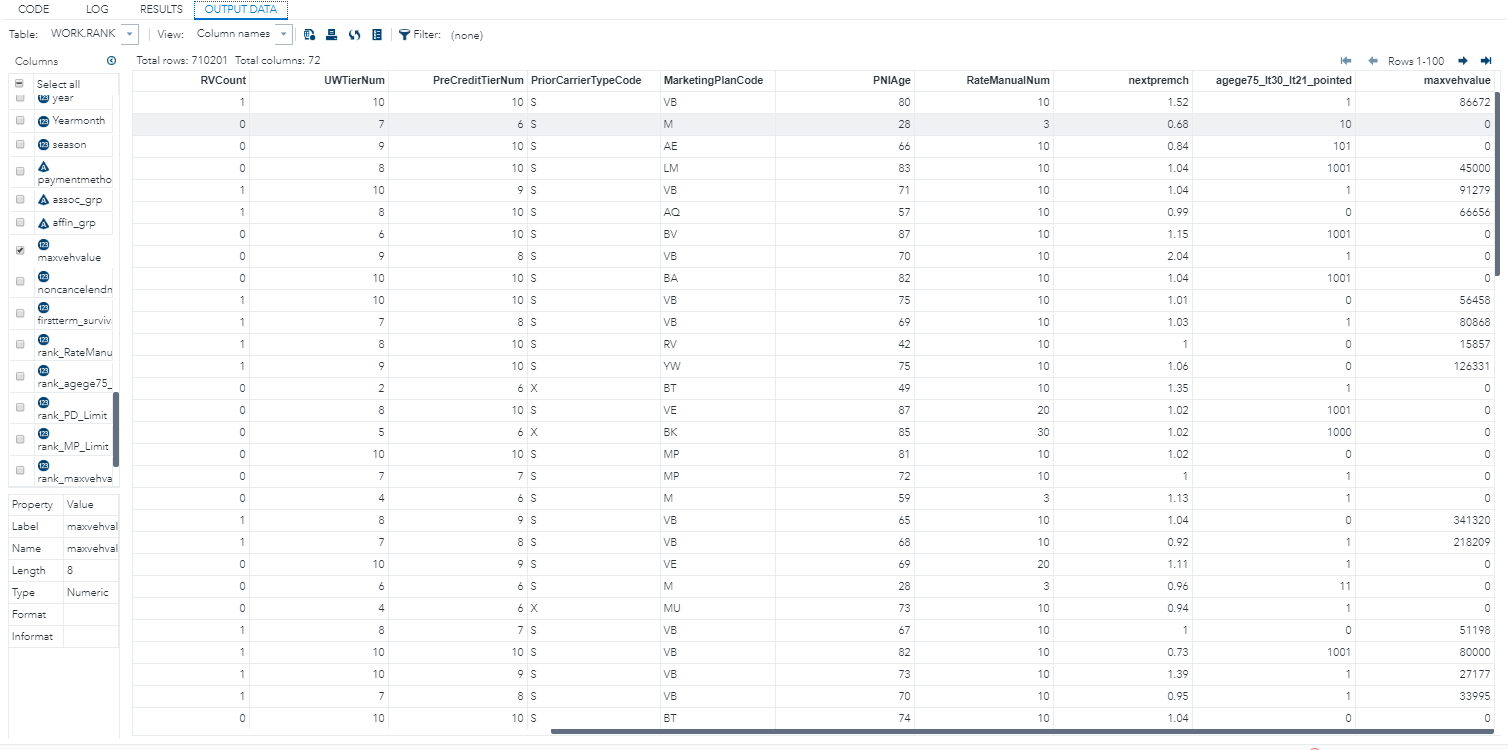
run;

proc print data=work.Rank(obs=10);

title "Subset of work.Rank";

run;

title;



**(Output Data of the variables Before Transformation)**

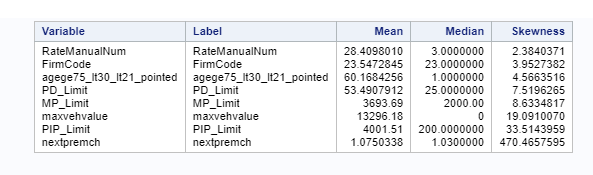
ods noproctitle;

ods graphics / imagemap=on;

proc means data=WORK.FINALRAW chartype mean median vardef=df skewness qmethod=os;

var RateManualNum FirmCode agege75\_lt30\_lt21\_pointed PD\_Limit MP\_Limit maxvehvalue PIP\_Limit nextpremch InsuranceExperienceDaysNum;

run;



**(Table Before Transformation)**

**AFTER RANK TRANSFORMATION:**

proc rank data=WORK.FINALRAW out=work.Rank;

var RateManualNum agege75\_lt30\_lt21\_pointed PD\_Limit MP\_Limit maxvehvalue PIP\_Limit nextpremch FirmCode;

ranks rank\_RateManualNum rank\_agege75\_lt30\_lt21\_pointed rank\_PD\_Limit

rank\_MP\_Limit rank\_maxvehvalue rank\_PIP\_Limit rank\_nextpremch rank\_FirmCode;

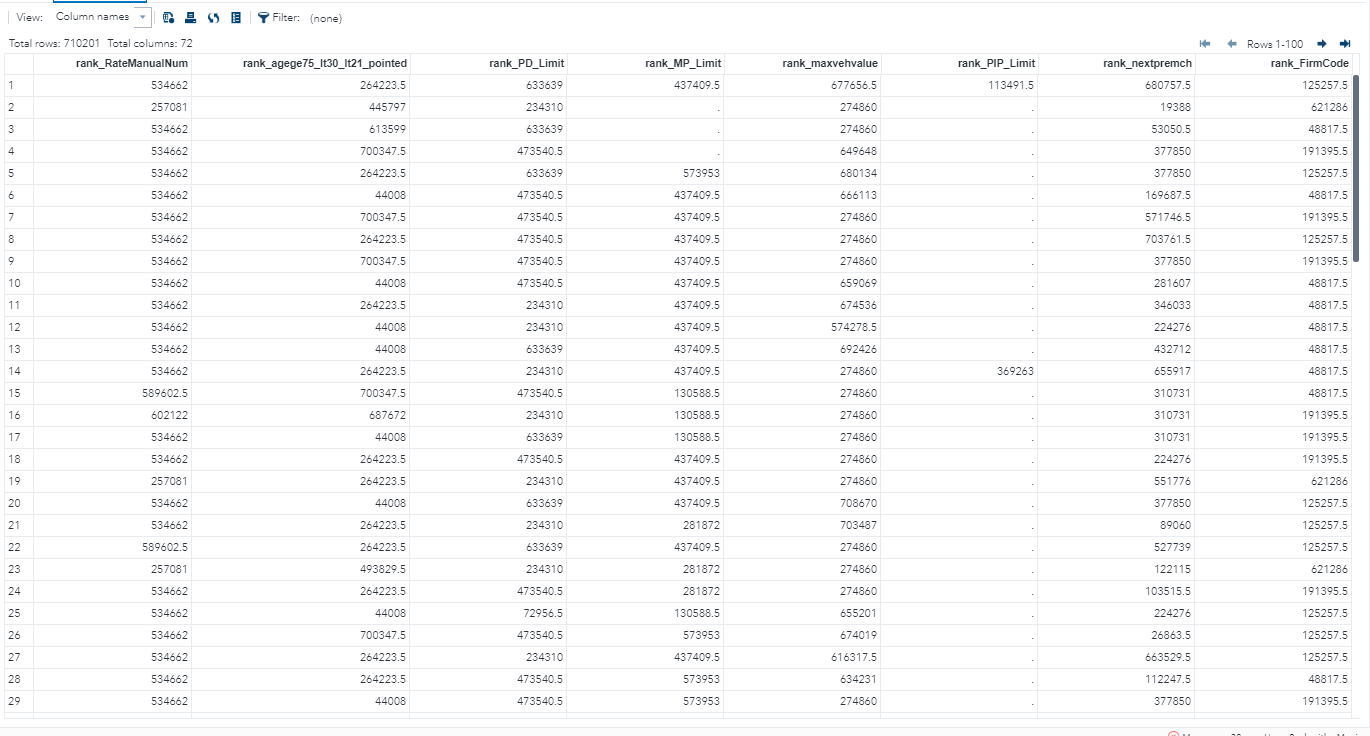
run;

proc print data=work.Rank(obs=10);

title "Subset of work.Rank";

run;

title;



**(Output Data after Transformation)**

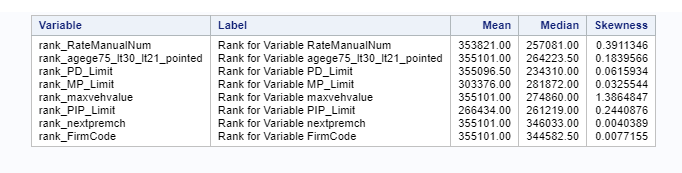
ods noproctitle;

ods graphics / imagemap=on;

proc means data=WORK.RANK chartype mean std min max n vardef=df;

var rank\_RateManualNum rank\_agege75\_lt30\_lt21\_pointed rank\_PD\_Limit rank\_MP\_Limit rank\_maxvehvalue rank\_PIP\_Limit rank\_nextpremch rank\_FirmCode;

run;



**(Table After Transformation)**

**Question 4:** *Perform log transformation on the following variables: UWTIERNUM NONCANCELENDMTS INCURRED\_LOSS.*

**ANSWER:**

**BEFORE TRANSFORMATION:**

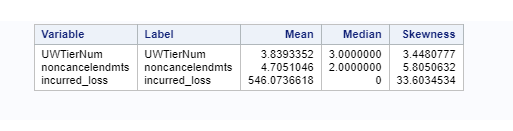
ods noproctitle;

ods graphics / imagemap=on;

proc means data=WORK.FINALRAW chartype mean median vardef=df skewness qmethod=os;

var UWTierNum noncancelendmts incurred\_loss;

run;

  
**(Table Before Transformation)**

**AFTER TRANSFORMATION:**

LIBNAME libref 'C:\Users\97150\Documents\Assignmentss SAS';

proc stdize data=\_temp0.finalraw missing=median out=work.rank1;

var UWTierNum noncancelendmts incurred\_loss;

run;

data WORK.Rank1;

set \_temp0.finalraw ;

log\_UWTierNum = log10 (UWTierNum);

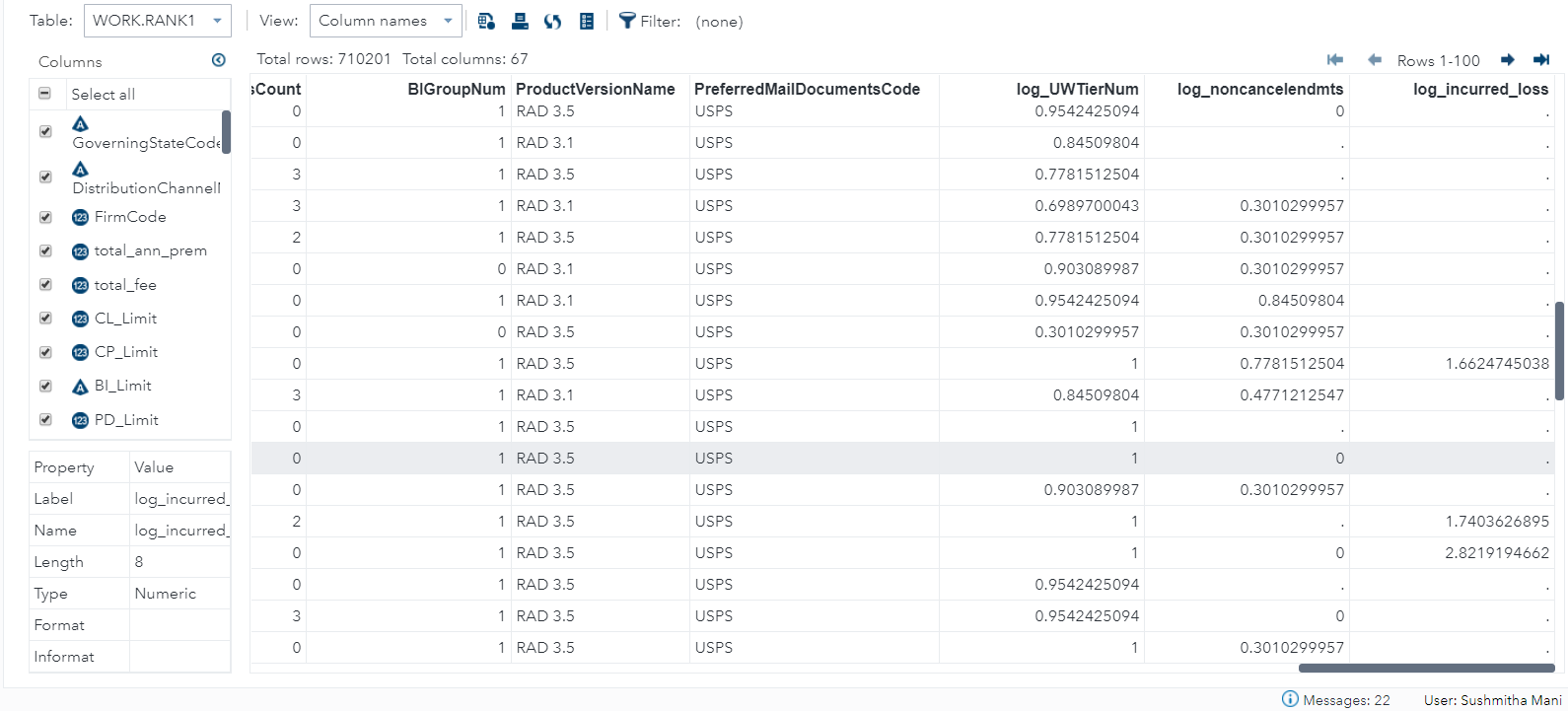
log\_noncancelendmts = log10 (noncancelendmts);

log\_incurred\_loss= log10 (incurred\_loss);

run;

proc print data= work.rank1 (obs=100);

run;



**(Output Data After Transformation)**

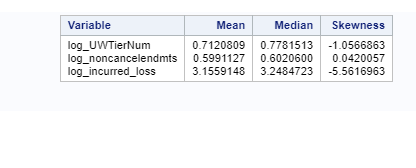
ods noproctitle;

ods graphics / imagemap=on;

proc means data=WORK.RANK1 chartype mean median vardef=df skewness qmethod=os;

var log\_UWTierNum log\_noncancelendmts log\_incurred\_loss;

run;

****

**(Table After Transformation)**

**QUESTION 5:***Perform power transformation of the following variables: TOTAL\_FEE TOTAL\_ANN\_PREM CL\_LIMIT NPCHCAT INSURANCEEXPERIENCEDAYSNUM CP\_LIMIT*

* *TOTAL\_FEE: Power 0.5*
* *TOTAL\_ANN\_PREM: Power 0.3125*
* *CL\_LIMIT: Power 0.25*
* *NPCHCAT: Power 0.125*
* *INSURANCEEXPERIENCEDAYSNUM: Power 0.375*
* *CP\_LIMIT: Power 0.4325*

**ANSWER:**

**BEFORE TRANSFORMATION:**

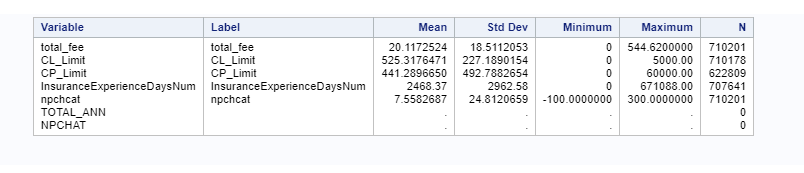
ods noproctitle;

ods graphics / imagemap=on;

proc means data=WORK.RANK1 chartype mean std min max n vardef=df;

var total\_fee CL\_Limit CP\_Limit InsuranceExperienceDaysNum npchcat TOTAL\_ANN NPCHAT;

run;



**(Table Before Transformation)**

**AFTER TRANSFORMATION:**

LIBNAME libref 'C:\Users\97150\Documents\Assignmentss SAS';

proc stdize data=\_temp0.finalraw missing=median out=work.rank1;

var UWTierNum noncancelendmts incurred\_loss;

run;

data WORK.Rank1;

set \_temp0.finalraw ;

power\_TOTAL\_FEE = TOTAL\_FEE\*\*0.5 ;

power\_TOTAL\_ANN= (TOTAL\_ANN)\*\*0.3125;

power\_CL\_LIMIT= (CL\_LIMIT)\*\*0.25;

power\_NPCHAT= (NPCHAT)\*\*0.125;

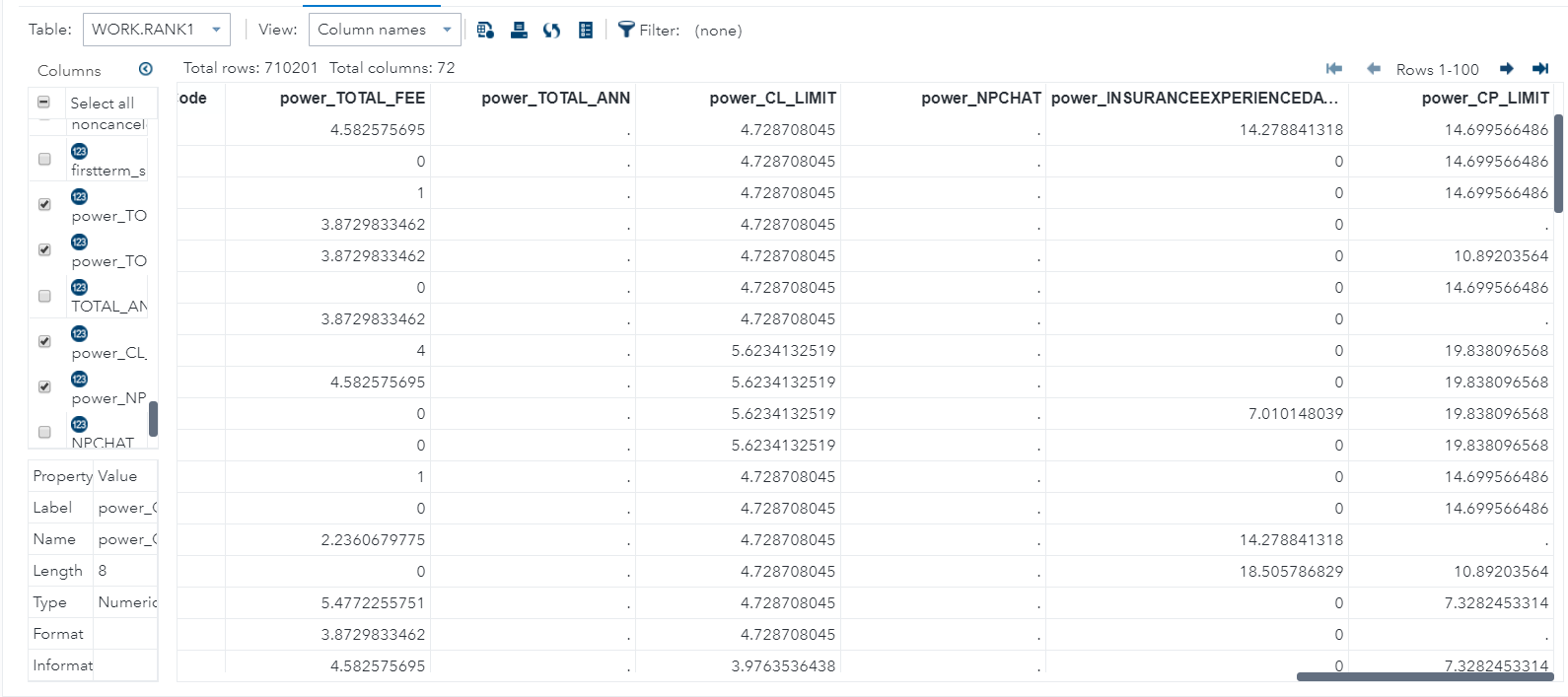
power\_INSURANCEEXPERIENCEDAYSNUM= (INSURANCEEXPERIENCEDAYSNUM)\*\*0.375;

power\_CP\_LIMIT= (CP\_LIMIT)\*\*0.4325;

run;

proc print data= work.rank1 (obs=100);

run;

****

**(Output Data After Transformation)**

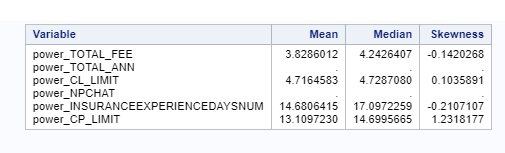
ods noproctitle;

ods graphics / imagemap=on;

proc means data=WORK.RANK1 chartype mean median vardef=df skewness qmethod=os;

var power\_TOTAL\_FEE power\_TOTAL\_ANN power\_CL\_LIMIT power\_NPCHAT power\_INSURANCEEXPERIENCEDAYSNUM power\_CP\_LIMIT;

run;

****

**(Table After Transformation)**