**Attrition Analysis**

1. **Import the required dataset**

**SAS Code:**

%web\_drop\_table(WORK.IMPORT);

FILENAME REFFILE '/folders/myfolders/sasuser.v94/Project 03\_Attrition Analysis\_Datasets - Renamed.xlsx';

PROC IMPORT DATAFILE=REFFILE

DBMS=XLSX

OUT=WORK.IMPORT;

GETNAMES=YES;

RUN;

PROC CONTENTS DATA=WORK.IMPORT; RUN;

%web\_open\_table(WORK.IMPORT);

1. **Check the frequency of churn**

**SAS Code:**

/\*Check the frequency of churn \*/

proc freq data = WORK.IMPORT;

tables Retain\_Indicator;

RUN;

**SAS Results Output:**

| **Retain\_Indicator** | | | | |
| --- | --- | --- | --- | --- |
| **Retain\_Indicator** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **Employee Left** | 28 | 56.00 | 28 | 56.00 |
| **Employee Retained** | 22 | 44.00 | 50 | 100.00 |

**The FREQ Procedure**

From the above results we get the **Frequency of Churn = 56%**

**SAS Code for frequency of churn:**

ods graphics / reset width=6.4in height=4.8in imagemap;

proc sgplot data=WORK.IMPORT;

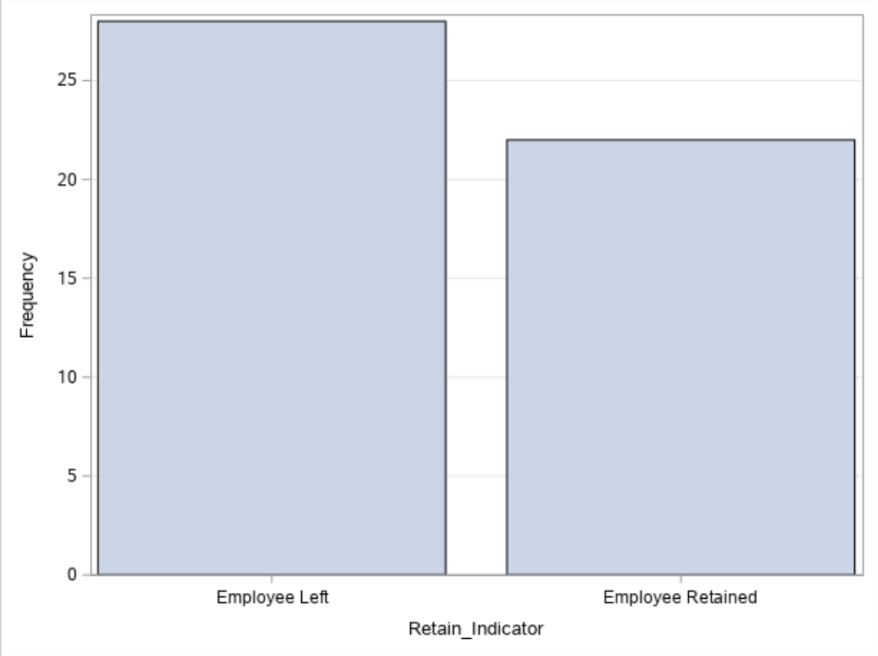
vbar Retain\_Indicator /;

yaxis grid;

run;

ods graphics / reset;

**Output:**



1. **Perform descriptive statistics for the dataset**

**SAS Code:**

/\* cross tab \_\_2 way frequency between Gender and Retain indicator\*/

proc freq data = WORK.IMPORT;;

tables Sex\_Indicator\*Retain\_Indicator/nocol norow nopercent nocum;

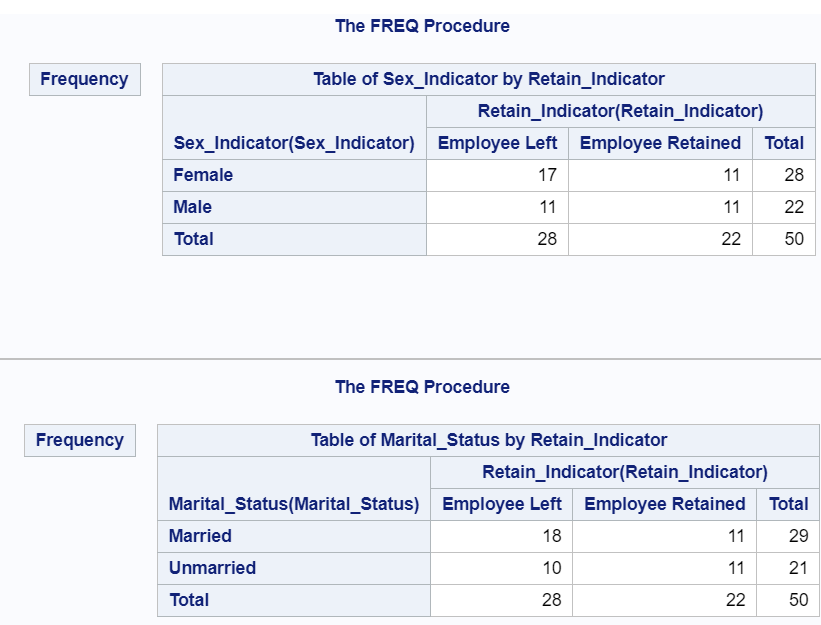
run;

/\*2 way frequency between Marital Status and Retain indicator\*/

proc freq data = WORK.IMPORT;

tables Marital\_Status\*Retain\_Indicator/nocol norow nopercent nocum;

run;

**SAS Results Output:**

**Insights:** Married employees are more likely to attrite than unmarried & we have seen that Married employees are more likely to attrite than unmarried.

1. **Perform logistic regression**

**SAS Code:**

%web\_drop\_table(WORK.IMPORT);

FILENAME REFFILE '/folders/myfolders/sasuser.v94/Project 03\_Attrition Analysis\_Datasets.xlsx';

PROC IMPORT DATAFILE=REFFILE

DBMS=XLSX

OUT=WORK.IMPORT;

GETNAMES=YES;

RUN;

PROC CONTENTS DATA=WORK.IMPORT; RUN;

%web\_open\_table(WORK.IMPORT)

proc logistic data = WORK.IMPORT;

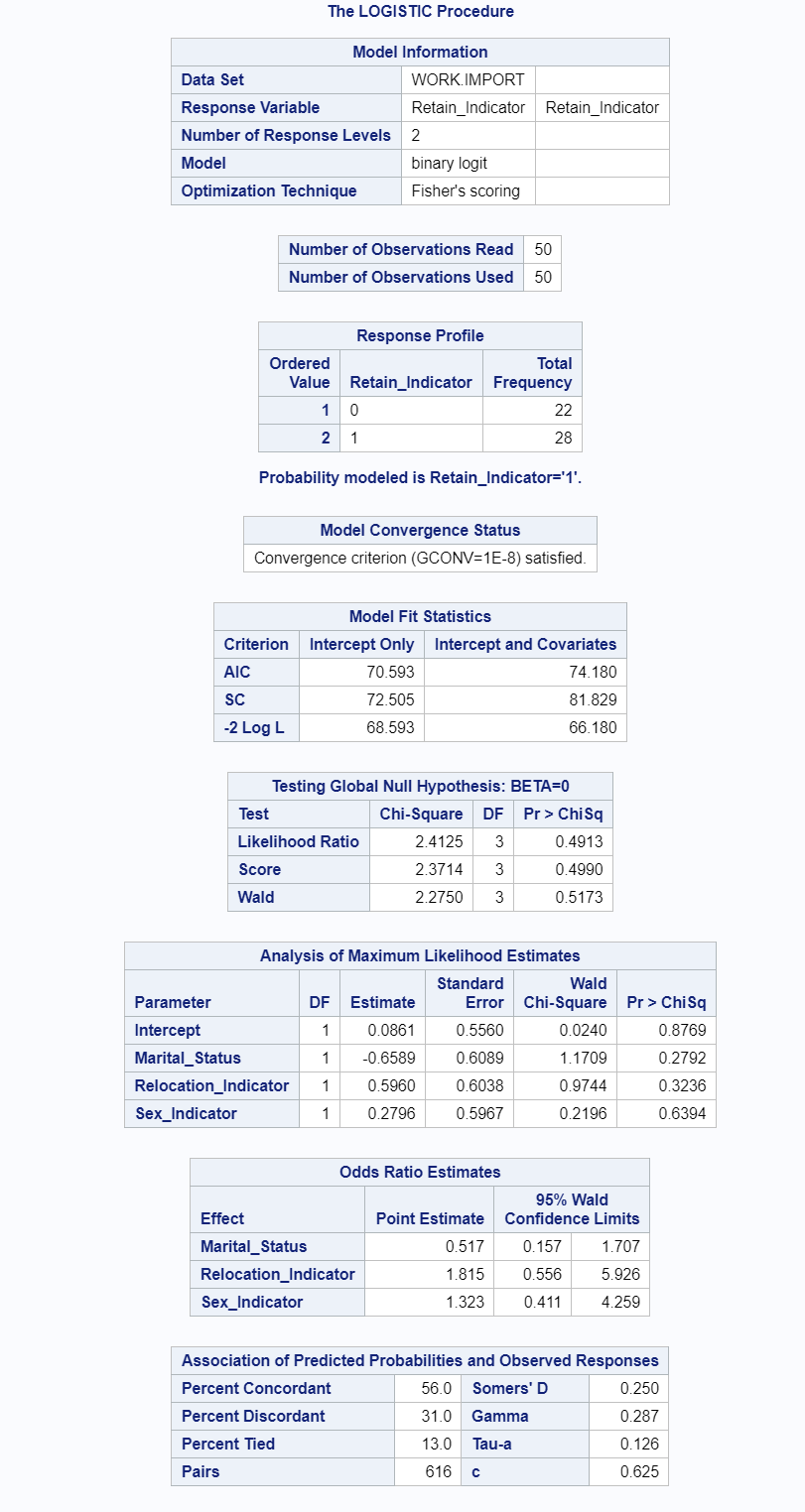
model Retain\_Indicator(event = '1') = Marital\_Status Relocation\_Indicator Sex\_Indicator/link=logit;

output out = \_PredicteDs p = \_Predicted\_prob\_attrition;

run;

**Insights:** Inthe table ‘**Analysis of Maximum Likelihood Estimates’** we have the p value for all the independent variables. Here we see that the variables ‘Marital\_Status’ & ‘Relocation\_Indicator’ have a much greater influence on the dependent variable (i.e. ‘Retain\_Indicator’) ,than the variable Relocation\_Indicator,which have the least influence on the dependent variable.

**SAS Results:**



1. **Check Max and Min values for the probability of churn**

**SAS Code:**

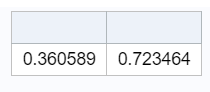
proc sql;

select min(Predicted\_prob\_attrition), max(Predicted\_prob\_attrition) from \_PredicteDs;

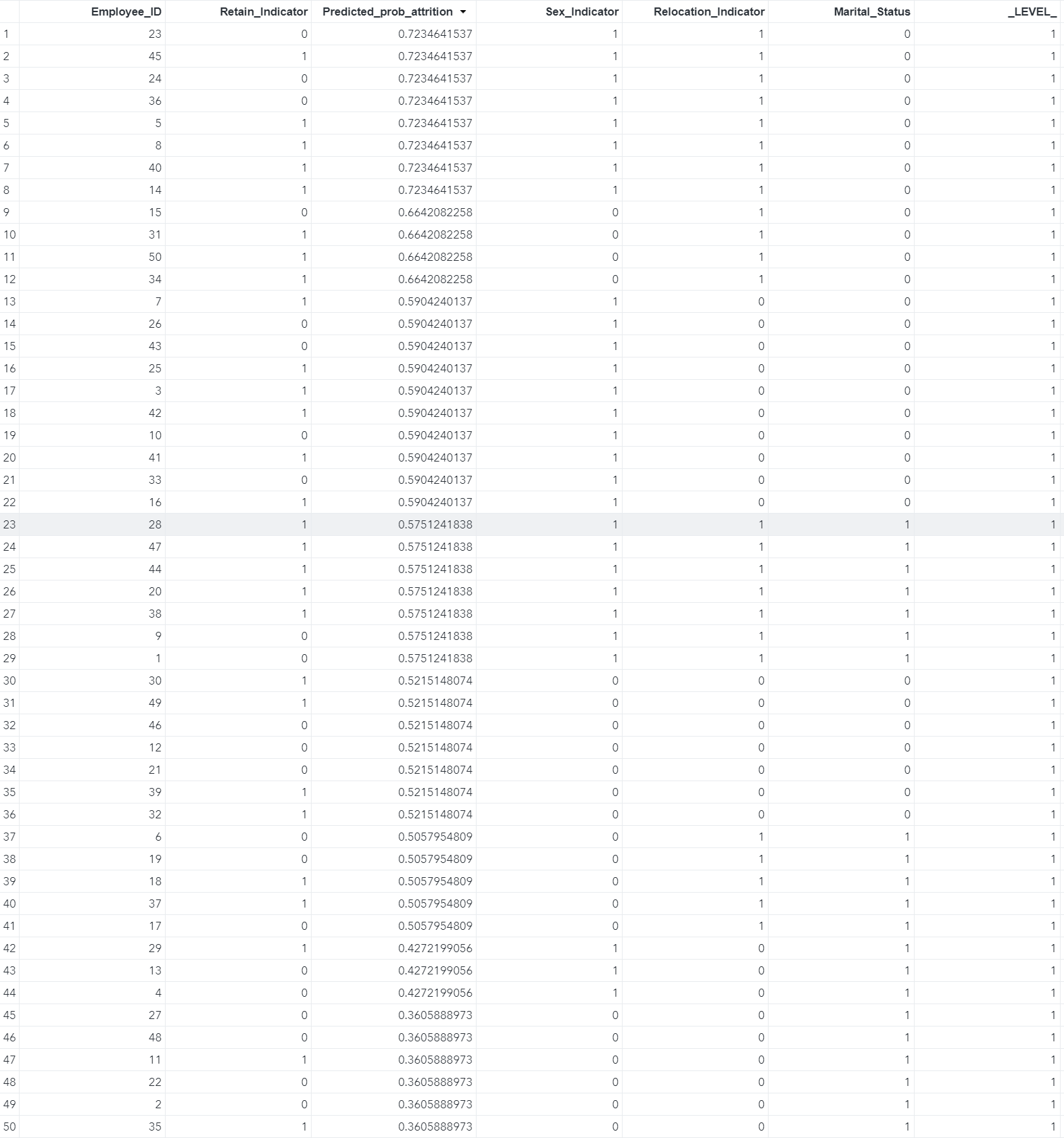
quit;

**SAS Results:**

Min value Max value



1. **Create new dataset to add all “churned” employees above the cut-off value**

**SAS Output Dataset:**

After observing the above SAS output data , we can say that 0.575 or 57.5% can be a good cut off value to predict & categorize all the employees on the basis of their churn probability.

**SAS Code:**

/\*Creating a new column with discrete predicted values using the Predicted\_prob\_attrition column and a cutoff value\*/

data predicted;

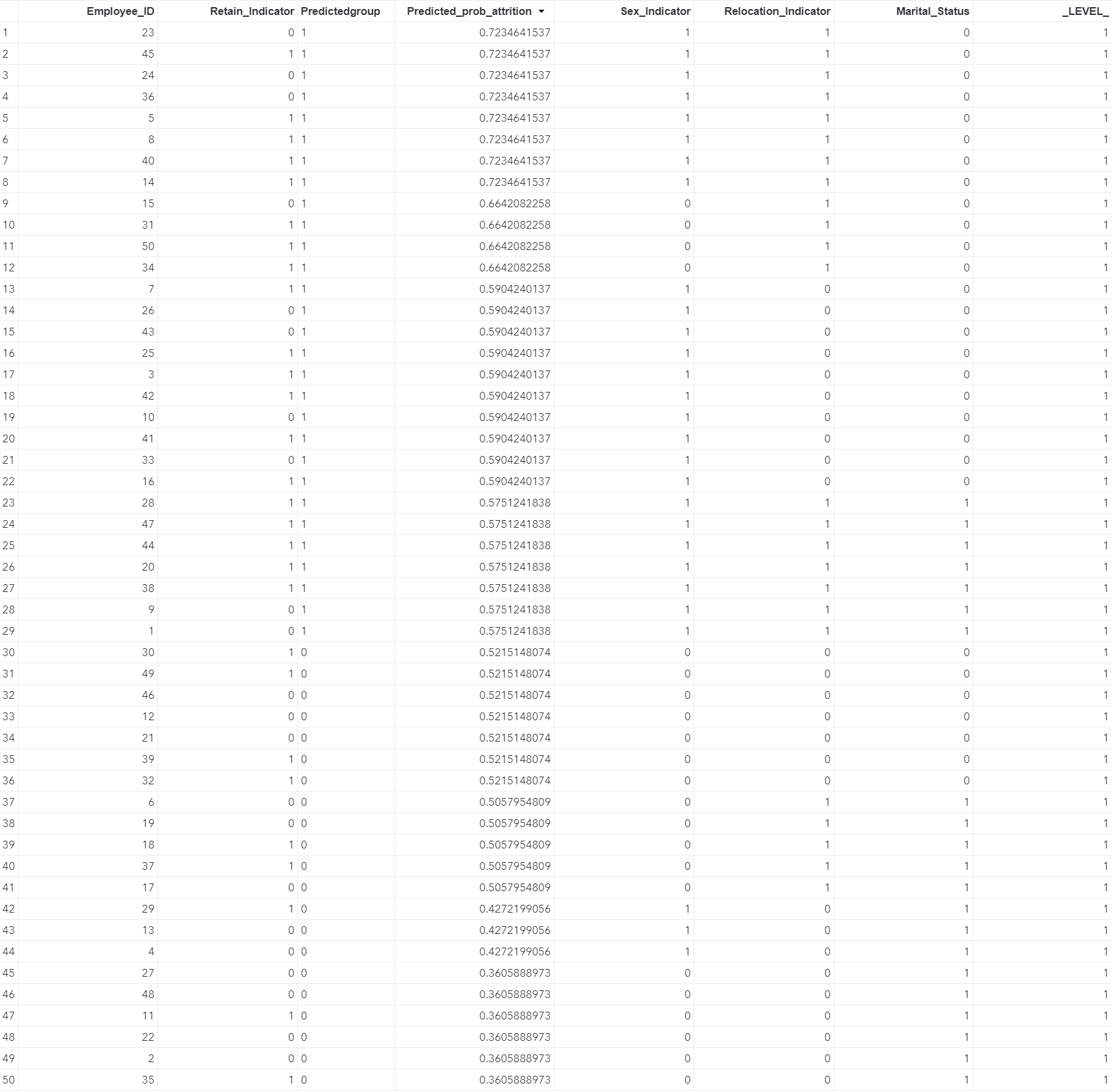
set \_PredicteDs;

if Predicted\_prob\_attrition<=.575 then Predictedgroup = "0";

if Predicted\_prob\_attrition>=.575 then Predictedgroup = "1";

run;

**SAS Output:**



**SAS Code:**

/\*Creating a new dataset to add all “churned” employees above the cut-off value\*/

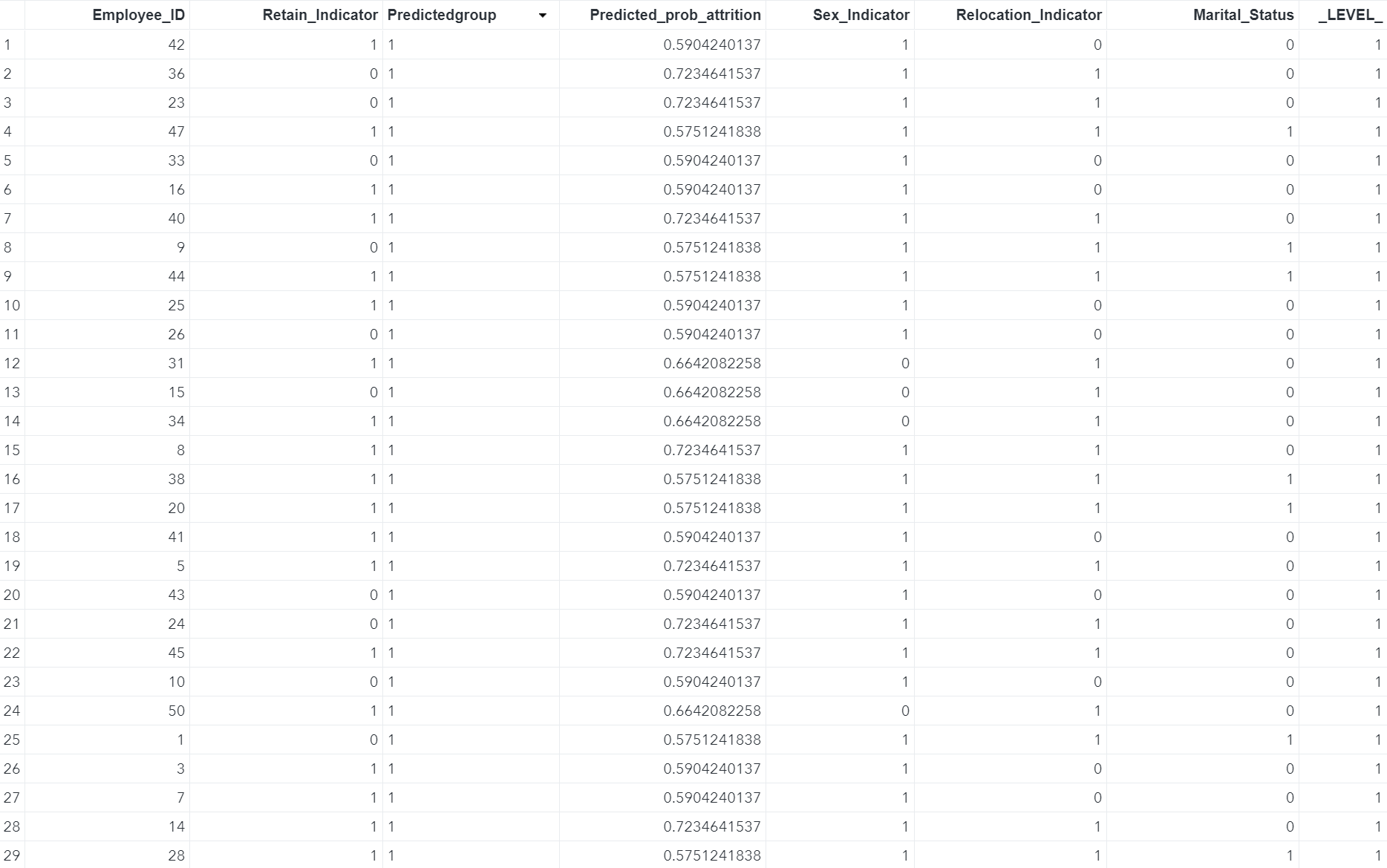
data Churned\_list;

set predicted;

if Predictedgroup = 0 then delete;

run;

**SAS Output Data:**

****