SAS® GLOBAL FORUM 2018

USERS PROGRAM

A Simple Methodology for Customer Classification in Two Dimensions

Adriana Mara Guedes Barbosa, Caixa Econômica Federal, Brazil

Alan Ricardo da Silva, University of Brasília, Department of Statistics, Brazil

April 8 – 11 | Denver, CO

#SASGF

A Simple Methodology for Customer Classification in Two Dimensions

Adriana Mara Guedes Barbosa and Alan Ricardo da Silva

Caixa Econômica Federal and University of Brasilia, Brazil

ABSTRACT

This paper aims to show a simple way for customer classification in 2 dimensions. Several variables were used to create only two major characteristics (customer attractiveness and profitability) and then it was possible to identify potential customers for grant credit. This methodology uses PROC REG, PROC GPLOT, PROC GINSIDE and some data steps, allowing the visualization of the results in a simple scatterplot.

METHODS

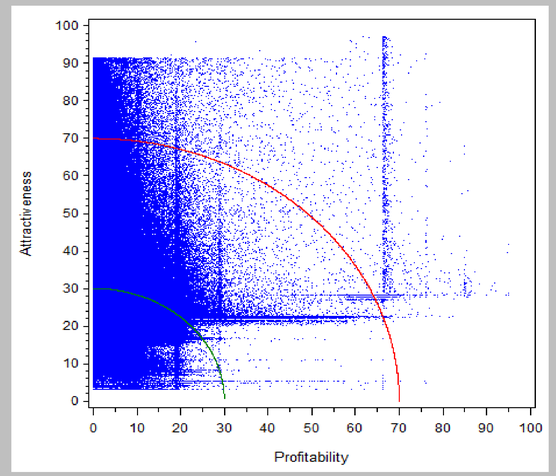
$$Rating\_Score = \alpha + \beta_1\, Market\_Indebtedness + \beta_2\, Segment\_Score + \beta_3\, Relationship\_Time\_Score + \beta_4\, Age\_Score + \epsilon \quad (1)$$

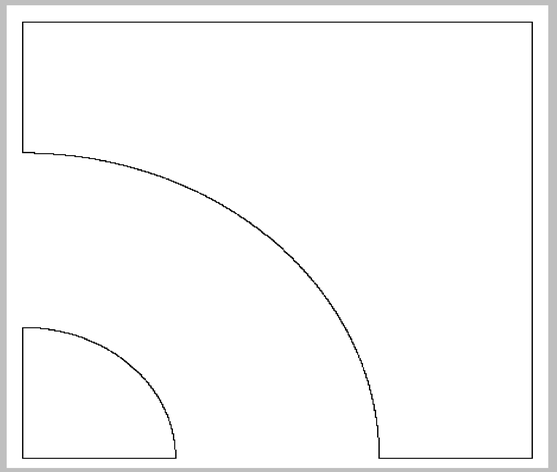
$$Month\_payment = \alpha + \beta_1\, Commercial\_Credit\_Taken + \beta_2\, Housing\_Credit\_Taken + \beta_3\, Investiments\_Volume + \beta_4\, Amount\_of\_Products + \beta_5\, Banking\_Tariff + \epsilon \quad (2)$$

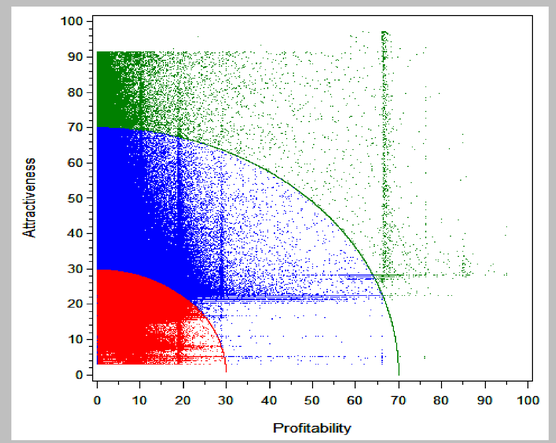
$$Attractiveness = 0.689 \times Market\_Indebtedness + 0.286 \times Segment\_Score + 0.024 \times Relationship\_Time\_Score + 0.001 \times Age\_Score$$

$$Profitability = 0.371 \times Commercial\_Credit\_Taken + 0.347 \times Housing\_Credit\_Taken + 0.248 \times Investiments\_Volume + 0.026 \times Amount\_of\_Products + 0.008 \times Banking\_Tariff$$

RESULTS





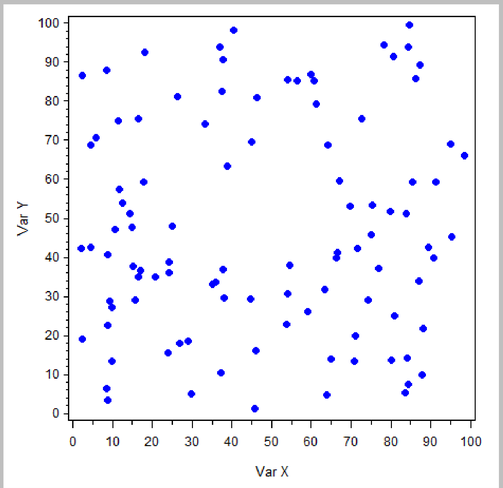


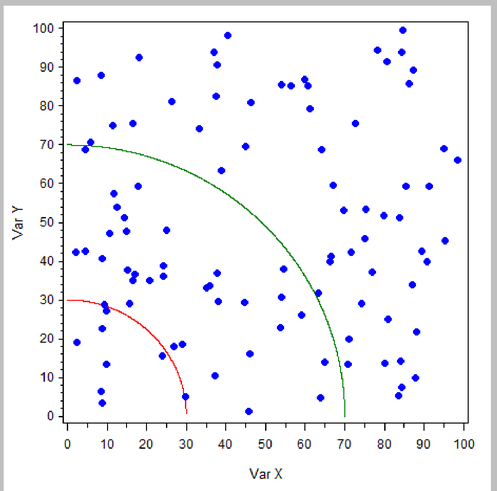
A Simple Methodology for Customer Classification in Two Dimensions

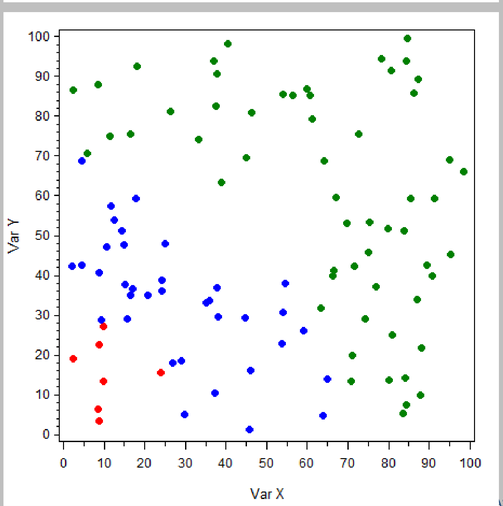
Adriana Mara Guedes Barbosa and Alan Ricardo da Silva

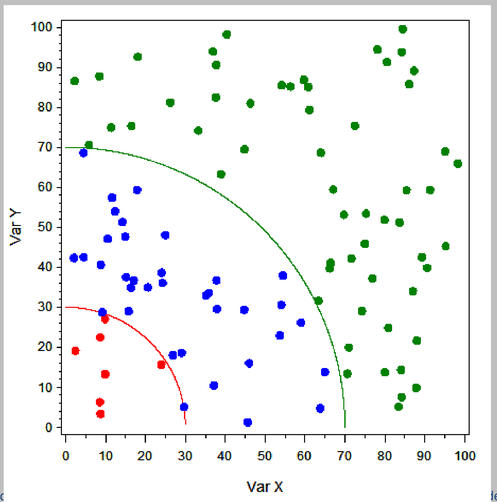
Caixa Econômica Federal and Universidade de Brasilia, Brazil

RESULTS CONTINUED









Conclusions

The main challenge of any financial institution is maximizes the likelihood of the customers pay the credit taken. This paper showed a simple methodology for customer classification using only 2 major characteristics: customer attractiveness and profitability. In summary, the steps for this classification are: 1) define weights for some candidate variables using PROC REG, with METHOD=STEPWISE; 2) define groups by using circle equation and plot it using PROC GPLOT; 3) classify the customers into the groups using PROC GINSIDE. The final result is a plot for the classification.

This simple methodology enhanced the customer classification, because it combines the classical credit risk methodology with the attractiveness profile, helping the decision maker of the institution.

SAS® GLOBAL FORUM 2018

April 8 – 11 | Denver, CO
Colorado Convention Center

Thank you!!

adriana.g.barbosa@caixa.gov.br
alansilva@unb.br

#SASGF

Paper 2721-2018

A Simple Methodology for Customer Classification in 2 Dimensions

Adriana Mara Guedes Barbosa, Caixa Econômica Federal, Brazil

Alan Ricardo da Silva, Universidade de Brasília, Departamento de Estatística, Brazil

ABSTRACT

This paper aims to show a simple way for customer classification in 2 dimensions. Several variables were used to create only two major characteristics (customer attractiveness and profitability) and then it was possible to identify potential customers for grant credit. This methodology uses PROC REG, PROC GPLOT, PROC GINSIDE and some data steps, allowing the visualization of the results in a simple scatterplot.

INTRODUCTION

Caixa Economica Federal is a Brazilian public bank handling about 82 millions of customers. The CRM (Customer Relationship Management) sector always deals with the challenge to develop constraints to grant credit in case of economic scenario showing a credit reduction. In this case, it is necessary to evaluate the customer in the credit risk perspective, as well as, in the attractiveness perspective, such as a good payer, credit in other financial institutions etc.

Faced with the economic situation that Brazil has been experiencing since 2014, another constraint beyond the credit risk was necessary to grant credit. In this way, the attractiveness dimension has been raised to assess customer profiles and consequently prioritize those customers more involved with the institution rather than anyone which request credit.

Imagine the situation in which two customers are requesting credit: one of them has a good historic of good credit payer, but has few products from the institution, such as a car insurance and a housing financing. The other one does not have a historic of credit taker, but has several products from the institution, such as an account with special credit, credit cards, car and life insurance and no debts, showing he is a good payer. In this case, which customer will be prioritized?

The main purpose of this paper is to show a simple methodology to evaluate the customer in the credit risk perspective and in other situation of financial engagement, as well as, to show these results in a scatterplot.

CUSTOMER CLASSIFICATION

Two major characteristics about customer attractiveness and profitability are essential to know the customer profile and, thus, maximize the likelihood of the credit taken be payed. Among several variables, a proxy for customer attractiveness has been created as:

$$Rating\_Score = \alpha + \beta_1\, Market\_Indebtedness + \beta_2\, Segment\_Score + \beta_3\, Relationship\_Time\_Score + \beta_4\, Age\_Score + \epsilon \quad (1)$$

And a proxy for customer profitability has been created as:

$$Month\_payment = \alpha + \beta_1\, Commercial\_Credit\_Taken + \beta_2\, Housing\_Credit\_Taken + \beta_3\, Investiments\_Volume + \beta_4\, Amount\_of\_Products + \beta_5\, Banking\_Tariff + \epsilon \quad (2)$$

These models can be easily estimated using PROC REG and the idea was to estimate a weight between the variables of each model. For that, it has been used the STEPWISE method and the PARTIAL R^2 was used to create the weights (W) for the customer attractiveness as:

$$W_{A1} = \frac{0.0682}{(0.0682+0.0283+0.0024+0.0001)} = 0.689 \quad (3)$$

$$W_{A2} = \frac{0.0283}{(0.0682+0.0283+0.0024+0.0001)} = 0.286 \quad (4)$$

$$W_{A3} = \frac{0.0024}{(0.0682+0.0283+0.0024+0.0001)} = 0.024 \quad (5)$$

$$W_{A4} = \frac{0.0001}{(0.0682+0.0283+0.0024+0.0001)} = 0.001 \quad (6)$$

$$Attractiveness = 0.689 \times Market\_Indebtedness + 0.286 \times Segment\_Score + 0.024 \times Relationship\_Time\_Score + 0.001 \times Age\_Score \quad (7)$$

In the same way, the customer profitability was created as:

$$W_{P1} = \frac{0.1437}{(0.1437+0.1342+0.0959+0.0102+0.0029)} = 0.371 \quad (8)$$

$$W_{P2} = \frac{0.1342}{(0.1437+0.1342+0.0959+0.0102+0.0029)} = 0.347 \quad (9)$$

$$W_{P3} = \frac{0.0959}{(0.1437+0.1342+0.0959+0.0102+0.0029)} = 0.248 \quad (10)$$

$$W_{P4} = \frac{0.0102}{(0.1437+0.1342+0.0959+0.0102+0.0029)} = 0.026 \quad (11)$$

$$W_{P5} = \frac{0.0029}{(0.1437+0.1342+0.0959+0.0102+0.0029)} = 0.008 \quad (12)$$

$$Profitability = 0.371 \times Commercial\_Credit\_Taken + 0.347 \times Housing\_Credit\_Taken + 0.248 \times Investiments\_Volume + 0.026 \times Amount\_of\_Products + 0.008 \times Banking\_Tariff \quad (13)$$

The Equations (7) and (13) were rescaled to be between 0 and 100, and the customers were classified in 3 groups: a group with low likelihood for grant credit; a group with middle likelihood for grant credit; and a group with high likelihood for grant credit. The cut points for these 3 categories were defined based on levels of Pearson Correlation of low correlation and high correlation, respectively: customer attractiveness and profitability less than 30, customer attractiveness and profitability between 30 and 70, and customer attractiveness and profitability higher 70.

Figure 1 shows the scatterplot of customer attractiveness and profitability and Figure 2 shows the curves for the 3 groups. Theses curves are easily created by using the circle equation defined as:

$$x^2 + y^2 = r^2 \quad (9)$$

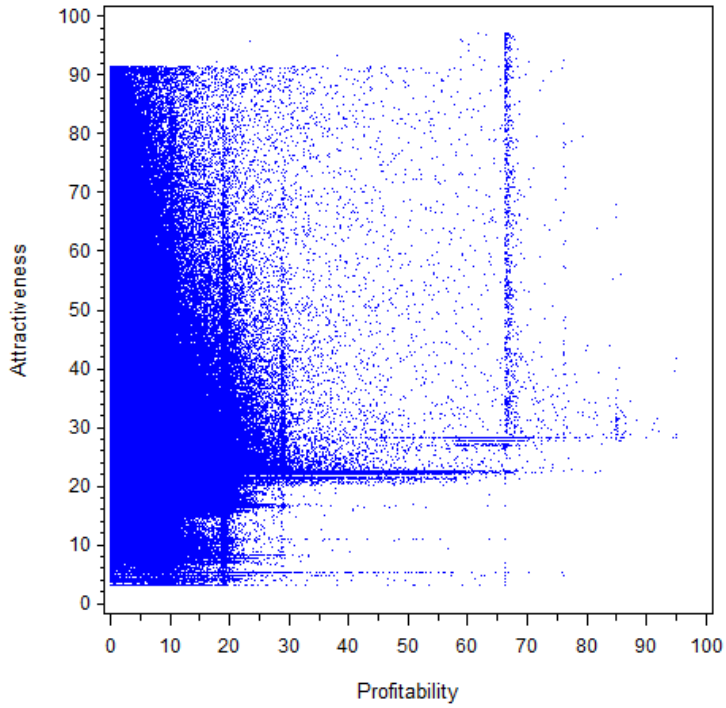


Figure 1. Scatterplot of customer attractiveness and profitability.

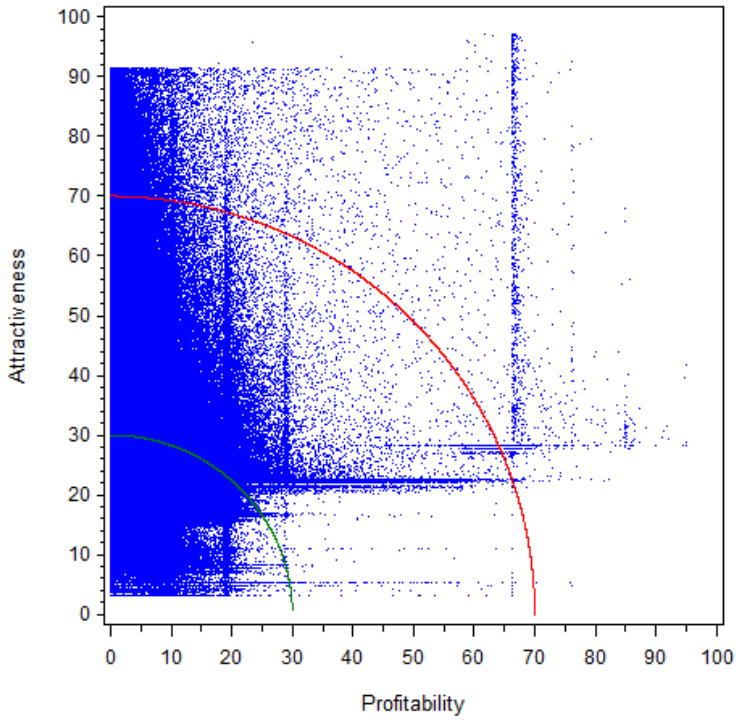


Figure 2. Scatterplot of customer attractiveness and profitability and group curves.

The next step is the classification of the customers into the 3 groups. One way is to create a shape for the first polygon defined by the radius equal to 30 and other for the third polygon defined by the radius equal to 70, and then use PROC GINSIDE to classify them. Figure 3 shows the shape and Figure 4 shows the result of the classification.

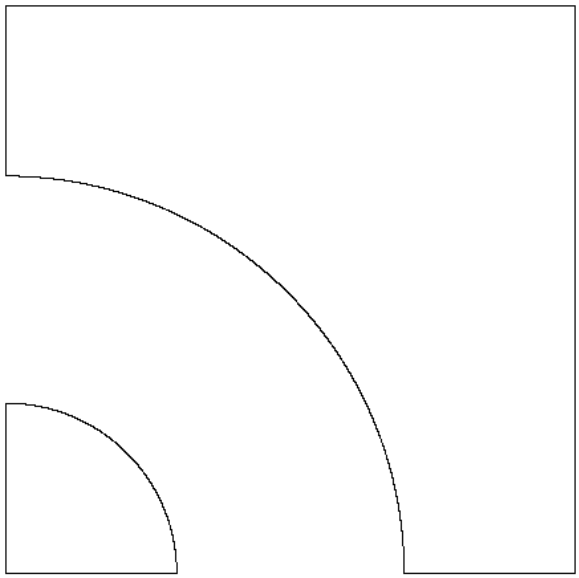


Figure 3. Polygons for the groups.

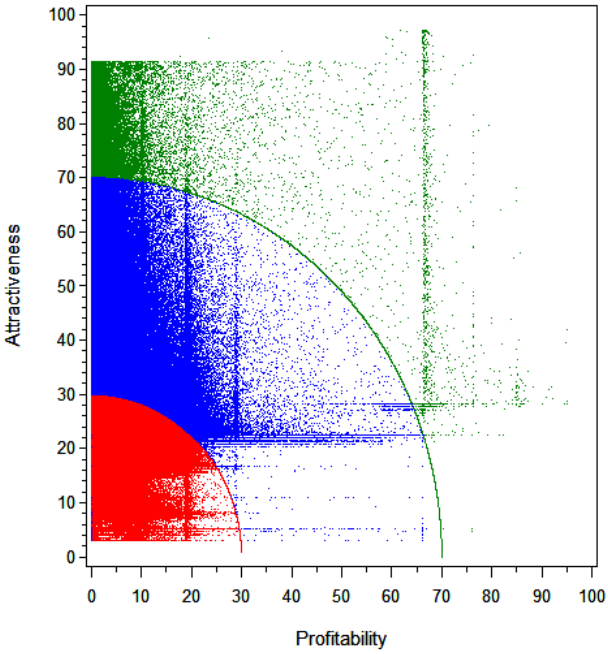
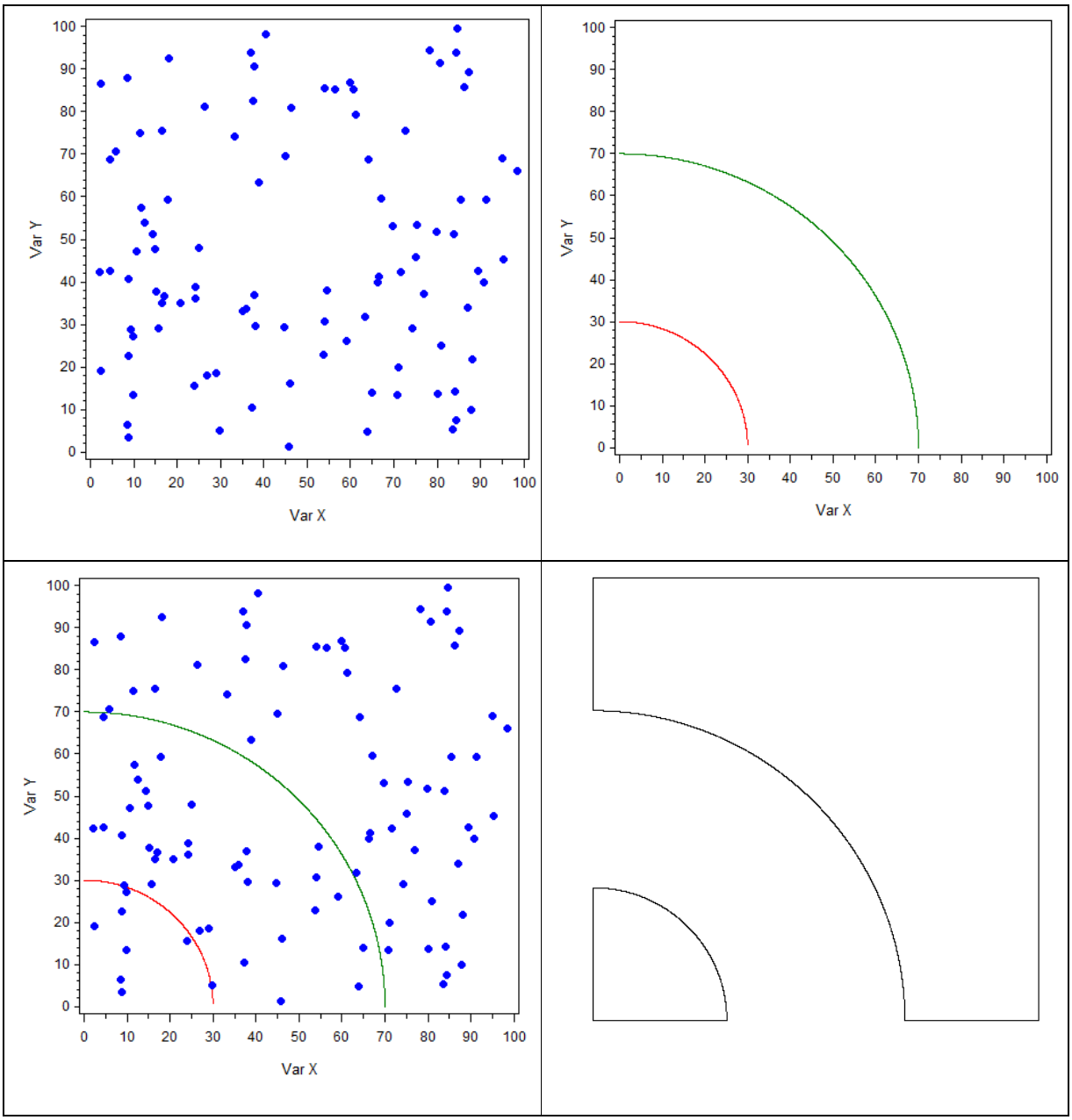


Figure 4. Final classification of the customers.

In Appendix, there is a program that illustrate the entire process. The plots of this program are in Figure 5.



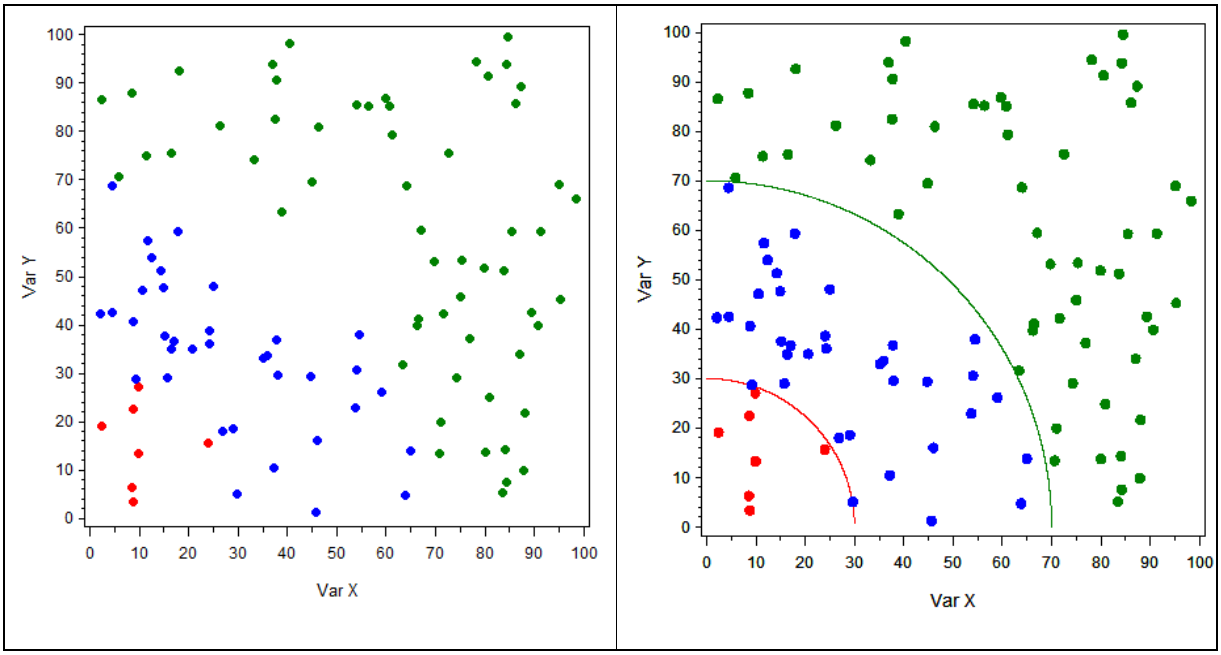


Figure 5. Example of classification process.

CONCLUSIONS

The main challenge of any financial institution is maximizes the likelihood of the customers pay the credit taken. This paper showed a simple methodology for customer classification using only 2 major characteristics: customer attractiveness and profitability. In summary, the steps for this classification are: 1) define weights for some candidate variables using PROC REG, with METHOD=STEPWISE; 2) define groups by using circle equation and plot it using PROC GPLOT; 3) classify the customers into the groups using PROC GINSIDE. The final result is a plot for the classification.

This simple methodology enhanced the customer classification, because it combines the classical credit risk methodology with the attractiveness profile, helping the decision maker of the institution.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Adriana Mara Guedes Barbosa
Enterprise: Caixa Econômica Federal
Address: Setor Bancário Sul, Quadra 4, edifício matriz, 3º andar
City, State ZIP: Brasília, DF, Brazil, 70092-900
Work Phone: +55 61 3206 6333
E-mail: adriana.g.barbosa@caixa.gov.br
Web: www.caixa.gov.br

Name: Alan Ricardo da Silva
Enterprise: Universidade de Brasília
Address: Campus Universitário Darcy Ribeiro, Departamento de Estatística, Prédio CIC/EST sala A1 35/28
City, State ZIP: Brasília, DF, Brazil, 70910-900
Work Phone: +55 61 3107 3672
E-mail: alansilva@unb.br
Web: www.est.unb.br

APPENDIX I – EXAMPLE FOR SIMULATED DATA

```
data simdata;
do i=1 to 100;
x=ranuni(2)*100;
y=ranuni(3)*100;
output;
end;
run;

proc gplot data=simdata;
plot Y*x;
symbol v=dot c=blue;
run;
quit;

axis1 order=0 to 100 by 10 length=4 in;
axis2 order=0 to 100 by 10 length=4 in;
proc gplot data=simdata;
plot y*x/ haxis=axis2 vaxis=axis1;
symbol i=none v=dot c=blue;
run;
quit;

data circle;
r1=30;
r2=70;
do x1=0 to 100 by 0.01;
y1=sqrt(r1**2-x1**2);
y2=sqrt(r2**2-x1**2);
output;
end;
run;

axis1 label=(a=90 'Var Y') order=0 to 100 by 10 length=4 in;
axis2 label=('Var X') order=0 to 100 by 10 length=4 in;
proc gplot data=circle;
plot y1*x1 y2*x1 / overlay haxis=axis2 vaxis=axis1 name='circle';
symbol1 i=join v=point c=red;
symbol2 i=join v=point c=green;
run;
quit;

data simdata1;
merge simdata circle;
run;

axis1 order=0 to 100 by 10 length=4 in;
axis2 order=0 to 100 by 10 length=4 in;
proc gplot data=simdata1;
plot y1*x1 y2*x1 y*x/ overlay haxis=axis2 vaxis=axis1;
symbol1 i=join v=point c=red;
symbol2 i=join v=point c=green;
symbol3 i=none v=dot c=blue;
run;
quit;
```

```
data circle30;
set circle(keep=x1 y1);
if x1=. or y1=. then delete;
id=30;
run;

data circle31;
x1=0;y1=0;id=30;
run;

data circle30;
set circle31 circle30;
run;

data circle70;
set circle(keep=x1 y2);
if x1=. or y2=. then delete;
id=70;
run;

data circle71;
x1=100;y2=0;id=70;output;
x1=100;y2=100;id=70;output;
x1=0;y2=100;id=70;output;
run;

data circle70;
set circle71 circle70;
run;

proc sql;
insert into circle30
values (30,0,30);
quit;

data circle3070;
set circle30(rename=(x1=x y1=y)) circle70(rename=(x1=x y2=y));
run;

data xx;id=10;v=3;run;
proc gmap data=xx map=circle3070 all;
id id;
choro v /nolegend;
run;
quit;

proc ginside data=simdata map=circle3070 out=frequency;
id id;
run;

data frequency;
set frequency;
if id=. then id=3070;
run;

axis1 label=(a=90 'Var Y') order=0 to 100 by 10 length=4 in;
axis2 label=('Var X') order=0 to 100 by 10 length=4 in;
```

```
proc gplot data=frequency;
plot y*x=id/ haxis=axis2 vaxis=axis1 name='datac' nolegend;
symbol1 i=none v=dot c=red;
symbol2 i=none v=dot c=green;
symbol3 i=none v=dot c=blue;
run;
quit;

proc greplay igout=work.gseg gout=work.gseg nofs tc=sashelp.templt
template=whole;
treplay 1:circle 1:datac name='datafin';
run;
quit;
```