Project Documentation

ABC Banking Optimization

24.11.2017

Natalia Kaźmierczak
Ashmitha Bonaventure
Keerthana Kokatam
Sonia Grzywacz
Aytac Macit
Baraa Hjih
Carlos Calatrava

Table of contents

1. Problem Description

The ABC is a banking group which serves n = 1000 customers and staffed by 10 employees. For the expansion of the business, it has become aware of the need to gain in-depth knowledge of the bank's customers, so as to offer products and services that best meet their needs and also ABC wants to study the behaviour of the customer because it can help us develop more accurate, predictive forecasts for our business.

2. Segments

To delineate behavioural segments (Platinum,Gold,Silver,Bronze..) of ABC's customers, with a threshold for each segment. Each customer belongs to some segment. Each segments contain different properties.

- **BRONZE Segment**

  This is the largest group. Generally they have only one current account, make few transactions above average value, make traditional, prudent choices (only Deposit Certificates), and make little use of credit cards for payment of goods.

- **SILVER Segment**

  Those forming this segment are mainly young nuclear families characterised by high spending, meagre resources and high use of credit cards.

- **GOLD Segment**

  Those belonging to this segment have large investments in shares and Certificates of Credit, ample liquidity in current accounts and scant need of services. Their accounts show high liquidity and movement of large sums, but in very small numbers.

- **PLATINUM Segment**

  Rich and progressive, very attached to the institute, whose services they use widely. They have a high average balance, but low liquidity and frequent transactions, and they hold two or three accounts. They are few but great customers.

A general rule of classification should allow us to classify all the Bank's customers in the identified behavioural targets.The implementation of the general rule of classification, should allow us to classify each customer to the most segment to which he could belong.

3. Thresholds of features

We need to analyse the features describing each customer. Chosen features:

- X - total amount of money used in transactions per month
- Y - number of transactions per month
- Z - standard deviation of the transactions over a day

More sample features :

- Customer demographics(retail/corporate client)
- Product information (number and types of products, loyalty)
- Transactional behavior :
  - » Number, amount and types of transactions per defined time period
- Balances :
  - » Average, opening and closing balances per defined time period.
- Utilisation of distribution channels :
  - » Channel usage number (ATM, online, branch) per given time period.
- Risk profile :
  - » Delinquency/claim (number and amount per defined period)

3.1. Thresholds values for each segment

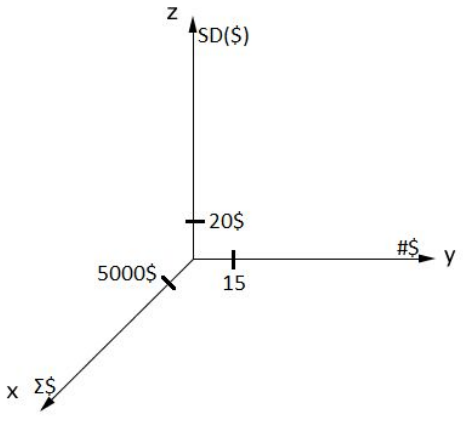
X - total amount of money used in transactions per month
Y - number of transactions per month
Z - standard deviation of the transactions over a day

Assumptions:

We assume the threshold values for the different segments/classes of customers as follows:

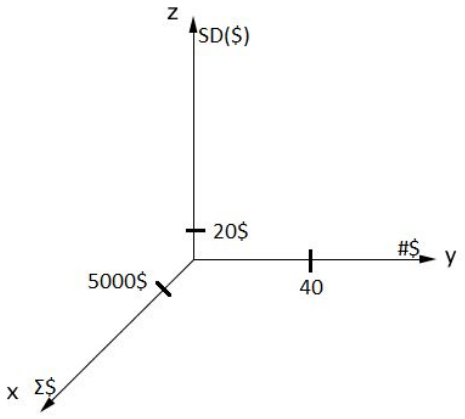
- **BRONZE Segment:**

Feature	Class/Segment	Threshold Value
X - total amount of money used in transactions per month (in $)	BRONZE	5000
Y - number of transactions per month	BRONZE	15
Z - standard deviation of the transactions over a day(in $)	BRONZE	20



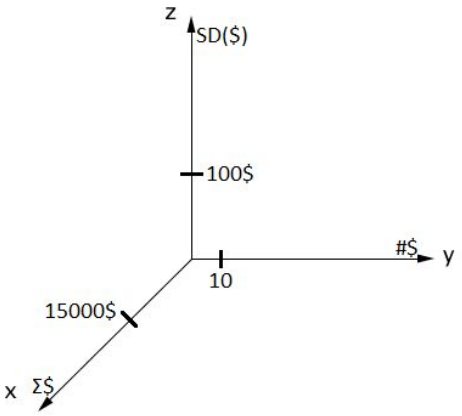
- **SILVER Segment**

Feature	Class/Segment	Threshold Value
X - total amount of money used in transactions per month (in $)	SILVER	5000
Y - number of transactions per month	SILVER	40
Z - standard deviation of the transactions over a day(in $)	SILVER	20



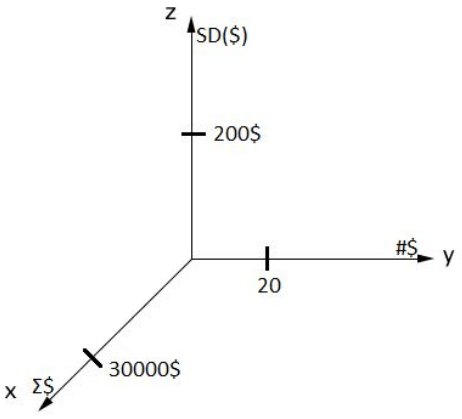
- **GOLD Segment**

Feature	Class/Segment	Threshold Value
X - total amount of money used in transactions per month (in $)	GOLD	15000
Y - number of transactions per month	GOLD	10
Z - standard deviation of the transactions over a day(in $)	GOLD	100



- **PLATINUM Segment**

Feature	Class/Segment	Threshold Value
X - total amount of money used in transactions per month (in $)	PLATINUM	30000
Y - number of transactions per month	PLATINUM	20
Z - standard deviation of the transactions over a day(in $)	PLATINUM	200



4. Decision rule for grouping A, B in each segment

We need to design a decision rule for the threshold of each segment and the features of each customer.

$t_1, ..., t_m$ - threshold of segment

$f_1, ..., f_m$ - feature of each customer

To belong to the group A the value of all features of the customer should be greater than or equal to each of the given threshold value of each segment. Otherwise, customer belongs to B group.

$$\bigcap_{i=1...m} f_i \geq t_i$$

Group A - all features exceeded threshold in each segment
Group B - all features did not exceed threshold in each segment

5. Complete group A, B

$$A_{(Platinum)} \cup A_{(Gold)} \cup A_{(Silver)} \cup A_{(Bronze)} = A$$

The new set A is a Union of all the customers of all the segments who satisfy the decision rule.

$$B_{(Platinum)} \cup B_{(Gold)} \cup B_{(Silver)} \cup B_{(Bronze)} = B$$

The new set B is a Union of all the customers of all the segments who does not satisfy the decision rule.

6. Rule for grouping Green, Red

The new sets are sent for analysis as to check who fits in the GREEN band and the RED band. Data analyst would be specifying this grouping. Each test on one customer is paid for the analyst. For testing purposes grouping will be done with random distribution of selecting RED customers in the dataset in each segment.

RED group - customers which do not pay in time or have some wrong properties
GREEN group - customers which data analysts classify as good customers

$\#A_{red}$ - number of red customers in A group

$\#B_{red}$ - number of red customers in B group

Initial random distribution parameter of number of chosen RED in group whole dataset will be set to 20%. Since the amount of customers in group B will be bigger, then by

probability we can say that number of RED customer will be bigger than 10% in group B, which is the acceptance rule described in the next chapter.

We will need to split the dataset of customers in each segment of group B by taking sample of whole group B set to 10% of all number in B group by sampling with random seed. This will minimize the expense of checking each customer in real life.

$\#\widehat{B}_{red}$ - number of red customers in sample of B group

$$\#B_{red} = \#\widehat{B}_{red} * 10$$

where 10 is the the size of sample in percentages.

7. Acceptance rule

The most important for the ABC company is to have the least amount of RED customers in the B group for each segment. Company pays also for each customer for checking by the data analyst. We need to check the percentage of number of RED customers in B group divided by all red customers in each segment. This percentage must be less than 10%.

Err - error (the percentage of red customers of overall customers)

$$Err = \frac{\#B_{red}}{\#A_{red} + \#B_{red}} * 100\%$$

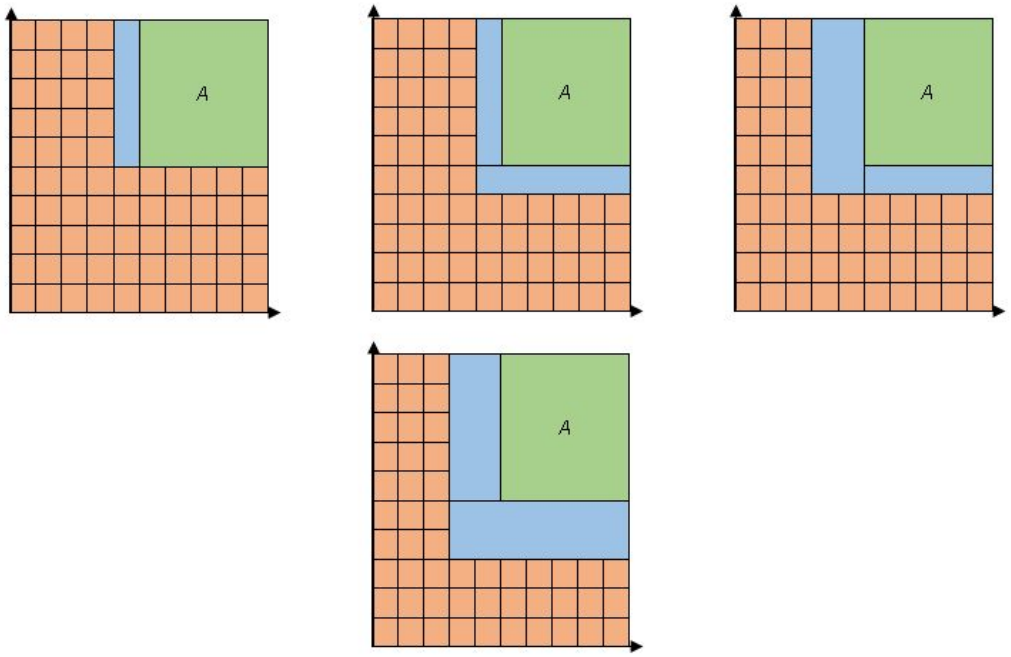
$$\boxed{Err < 10\%}$$

8. Solution

suggested solution:
Let we say we have three dimensions or variables to determine our data X,Y,Z.
we will generate temporary data set contain splits subsets from our original data with values of dimensions and check number and percentage of Good and Bad customer and after that will choose best one depending on our requirements to redefine threshold.

pesucode:

```
do i=Xto 0 by - %5;
        do j=Y to 0 by - %5;
                do k=Z to 0 by - %5;
                        proc sql;
                        select count(customer),count(red),count(green) from
                        ((select * from DATA where X<i and Y<j and Z<k)
Minus
                        (select * from DATA where X>=i and Y>=j and Z>=k))
                        quit;
                        red=count(red);
                        green=count(green);
                        percent= count(red)/(count(red)+count(green));
                end;
        end;
end;
run;
```



X index	Y index	Z index	Number of Green	Number of Red	Percentage of Green depending on subset Data	Percentage of Green depending on whole Data
X	Y	Z	…	…	%	%
X-1	Y-1	Z-1	…	…	%	%
…….						
0		0	…	…	%	%