ISSN-e: 2244-8276

USE OF INDICATORS AS A BUSINESS INTELLIGENCE SOLUTION FOR A CUSTOMER SERVICE LOGISTICS PROCESS OF CERTAIN VENEZUELAN PRIVATE BANKS.

USE OF INDICATORS AS A BUSINESS INTELLIGENCE SOLUTION FOR A LOGISTICS PROCESS OF CUSTOMER SERVICE OF CERTAIN VENEZUELAN PRIVATE BANKS.

SIRO TAGLIAFERRO[1]
Universidad Metropolitana de Caracas, Venezuela
stagliaferro@unimmet.edu.ve
0000-0001-7501-3568

JOSÉ VALENTÍN SALAZAR[2]
Universidad Metropolitana de Caracas, Venezuela
jvalentin@unimmet.edu.ve

Summary

Business Intelligence (BI) tools allow ordering and analyzing data from various sources to obtain knowledge that facilitates the interpretation and understanding of information, improving decision making in a company and achieving a competitive advantage. This article seeks to propose BI solutions for customer service, using two anonymous banking institutions as a database and source of information. Through the use of Rstudio©, as a digital tool to work with Data Mining, processes and statistical analysis were performed to extract all relevant information for the development of management indicators. Through these indicators and the Benchmarking methodology, the results between both banks were compared, where it was determined that there were two serious problems: customer segmentation and excess of products offered. Finally, a series of solutions based on business intelligence were proposed to optimize resources and improve customer service, thus increasing the level of competitiveness of both banks.

Keywords: Business Intelligence, business intelligence, management indicators, customer service, cluster analysis, Rstudio©.

1 Siro German Tagliaferro Isturiz: Production Engineer (2012) UNIMET, MBA (2016) IESA, Data Science (2022) UCV, full time professor with more than 5 years of experience.

2 José Valentín Salazar: Production Engineer (2021) UNIMET

Summary

Business Intelligence (BI) tools allow ordering and analyzing data from various sources, to obtain knowledge that facilitates the interpretation and understanding of information, improving decision-making in a company and achieving a competitive advantage. This article seeks to propose BI solutions for customer service, using two anonymous banking institutions as a database and source of information. Through the use of Rstudio©, as a digital tool to work with Data Mining, statistical processes and analyzes were carried out to extract all the pertinent information for the development of management indicators. Through these indicators and the Benchmarking methodology, the results between both banks were compared, where it was determined that there were two serious problems: customer segmentation and excess products offered. Finally, a series of solutions based on business intelligence were proposed to optimize resources and improve customer service, thus increasing the level of competitiveness of both banks.

Keywords: Business Intelligence, business intelligence, business intelligence, management indicators, customer service, cluster analysis, Rstudio©.

RECEIVED: 09-07-2024 / ACCEPTED: 13-09-2024 / PUBLISHED: 22-12-2024

How to quote: Tagliaferro S. & Salazar J. (2024). Use of Indicators as a Business Intelligence solution for a customer service logistics process of certain Venezuelan Private Banks. *Almanaque*, 44, 59 - 84. https://doi.org/10.58479/almanaque.2024.9

ISSN-e: 2244-8276

INDEX

Summary	59
Summary	60
Introduction	63
Research Phases	63
Obtaining information	63
Data cleansing	63
Data Mining	66
Evaluation and comparison	68
Construction of indicators	71
Results of the study	74
Current values of indicators	74
Product efficiency	75
Best selling products	76
Sales by State	77
Dashboard generation	78
Conclusions and recommendations	79

Conclusions 79

Recommendations: 80

Acknowledgments 81

References 81

Introduction

Today, one of the key factors for decision making in a company is the speed with which it has access to accurate information. In any decision-making process, if one party fails, all parties fail. On the other hand, if the data is not presented in a way that facilitates its interpretation, reliable information cannot be obtained. Indicators make it possible to establish measurement parameters and understand the state of a business, in addition to facilitating evidence-based decision making. Business intelligence allows to extract quality information from the collection and analysis of any amount of data from different sources, especially indicators, a fundamental part for a Business Intelligence solution to work, on the other hand, data analysis supported by this solution allows the use of Data Mining tools, which can detect patterns or actions that are hidden and can not be obtained empirically, which in the end can become a competitive advantage for companies.

Another factor of utmost importance for companies is customer service. The cost of acquiring a customer is much higher than the cost of keeping one, and customers are increasingly demanding, so a bad experience becomes a bad reputation for the company and loss of customers.

From the above, this research proposal arises, which proposes to implement business intelligence solutions for customer service through indicators, built from information provided by two banking institutions, and the Benchmarking methodology applied between these two companies.

Research Phases

Obtaining information

The information was obtained through banks A and B to obtain the following study variables such as: types of accounts, products, records, sex, nationality, age, region, last entry in the account, among others. This information was stored in a csv file.

With this data, data mining was performed to develop indicators and detect patterns that were of interest for the study.

Data cleansing

The database provided by the banks consisted of the following: clients (rows) and variables (columns), where the most important variables are: Date of account opening, client's age, gender, products purchased, if he/she is a main client, client's location, nationality, opening funds, income, if he/she has a mortgage, among others. On the other hand, there are others that were useful such as: Client Code, account number, among others since they are data that are transactions that do not contribute in the study, later in the study another cleaning will be made for the level of effect between the variables, by means of Data Mining tools. On the other hand, there were data that could not be taken for the study due to lack of information or errors in the values.

The data were archived in Excel documents, which were supplied by the banks from their information systems. Table 1 below shows an example of the above.

Table 1. Bank data

Source: Own elaboration with banks

The variables were then coded to facilitate the calculation in the Rstudio© tool. This coding is used to be able to introduce qualitative data in the software, given that the software does not recognize the *character* variables in its Data Mining algorithms.

Table 2 below shows how the coding was done to be recognized by the Rstudio© software, only the qualitative data is transformed, everything else that was in quantitative data was kept the same as it was delivered. This transformation is done in an orderly manner and with consecutive numbers according to the number of values of the variable, in some cases a Boolean is used in its value Yes or No (0 or 1) for the coding.

Table 2. List of coded qualitative variables

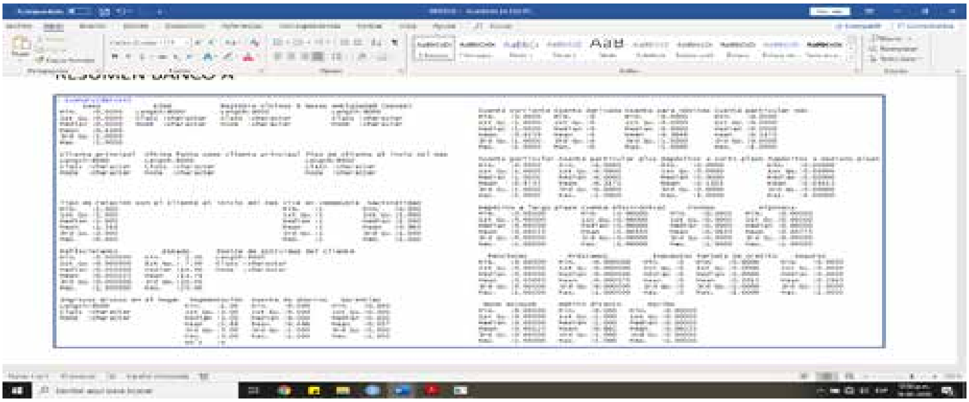
Variable	Valores originales	Valores nuevos
Sexo	**H** (Masculino)	1
	M (Femenino)	0
Vive en Venezuela	Si	1
	No	0
Nacionalidad Venezolana	Si	1
	No	0
Fallecimiento	Si	1
	No	0
Nombre del estado	Amazonas	1
	Anzoátegui	2
	Apure	3
	Aragua	4
	Barinas	5
	Bolívar	6
	Carabobo	7
	Cojedes	8
	Delta Amacuro	9
	Dependencias Federales	10
	Distrito Capital	11
	Falcón	12
	Guárico	13
	Lara	14
	Mérida	15
	Miranda	16
	Monagas	17
	Nueva Esparta	18
	Portuguesa	19
	Sucre	20
	Táchira	21
	Trujillo	22
	Vargas	23
	Yaracuy	24
	Zulia	25

Source: Own elaboration

Next, the data already coded in the Rstudio© software were extracted, and the *summary* function was applied to check if the program was reading the data correctly and generating descriptive statistics of the data, which should be done to see what the software is detecting and to have an empirical view of the behavior of the data. Table 3 shows the values of each variable: maximum and minimum, its average, variance, quartiles and the type of data it contains. Being able to verify what data the software detects is of utmost importance so that the Data Mining algorithms can work correctly and continue with the study.

It is important to mention that this article has been applied and analyzed in bank A, since the methodological steps and algorithms applied are the same for bank B.

Table 3. Summary of descriptive statistics for Bank A



It can be seen that the results of this table are measures of central tendency of the variables, and it can also be seen that there was no problem in reading the data.

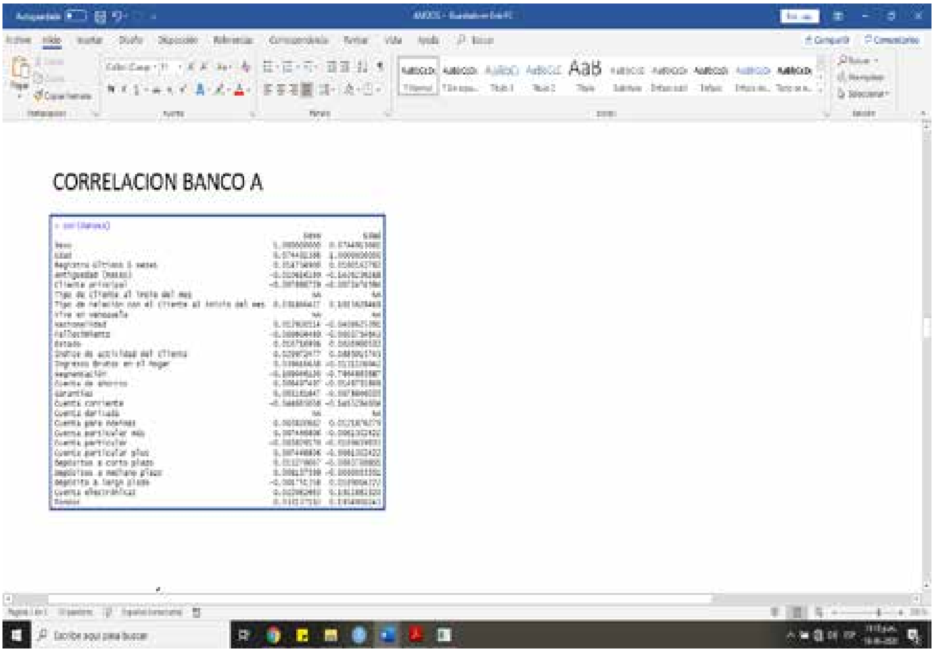
Source: Own elaboration

Data Mining

This stage was divided into two phases, the first one of data exploration and transformation, and the second one of selection and application of the data mining method that best suited the research.

The data exploration and transformation phase is composed of a superficial exploratory analysis of the data. Through the *cor* and *summary(cor)* functions, specific aspects such as the degree of correlation and variance for each variable, maximum, minimum and median values, among others, were observed. This was done in order to pinpoint the independent and/or duplicated variables, which were eliminated using the *null* function. This analysis can be seen in Table 4.

Table 4. Correlation analysis by *Cor* function in Rstudio © Table 4.



Source: Own elaboration

This correlation analysis was then generated using the *corrplot* function to graphically observe the degree of correlation between the filtered variables and to facilitate the location of the most important variables for future analysis.

A data transformation process was carried out where all the variables were scaled using the *scale* function; this is nothing more than subtracting the mean and dividing by the standard deviation, so that the resulting variables have mean 0 and standard deviation 1. This was done in order to prevent some variables from having a greater influence than others on the variance and, therefore, to prevent the selected grouping algorithm from generating erroneous results.

For the phase of selection and application of data mining methods, different types of hierarchical clusters based on Euclidean distances were used, since each one uses a different function (equation) and representation, so that various groupings (models) can be obtained to explain the largest possible amount of data through an optimal number of patterns or, at least, to indicate dependencies in the data. In order to choose each of the models to be used, three criteria of preference were established: that it should be an optimal method for the amount of data, that it should fit the problem and that it should work with Euclidean distances. Finally, all the models obtained were compared with each other to decide which of them best described the information analyzed.

Evaluation and comparison

At this stage, the information obtained was evaluated and deciphered according to the clustering method used. In addition, the number of clusters that best described the data was selected and the basis for the design of the indicators was established.

Table 5 shows the different algorithms used for the *clusters*. This table summarizes how the clustering process is established, using the mathematical distances between the data obtained and grouping them according to the algorithms used.

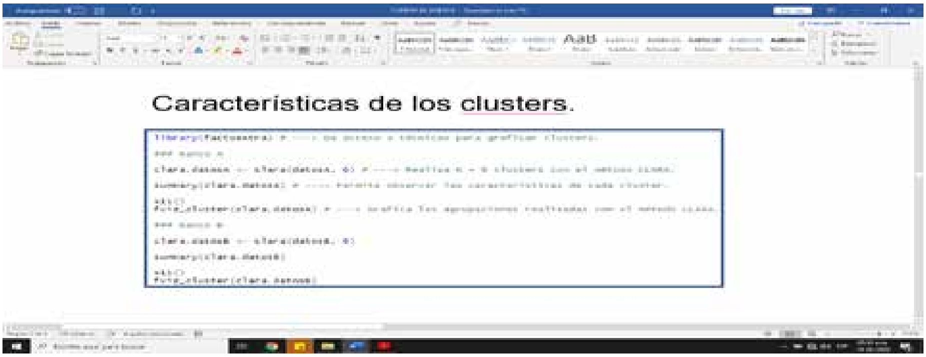
Table 5. Types of distances applied in the *Clusters*

Method	Description
Minimum distances *(single linkage)*	Algorithm that relates clusters according to the minimum distance between the closest entities.
Maximum distance *(complete linkage)*	Algorithm that relates clusters according to the maximum distance between the nearest entities.
Ward (minimum variance)	Algorithm that attempts to reduce the variance by considering the union of each pair of clusters and choosing to combine those that generate a smaller increase in the sum of the squares of the deviations when joined.
Ward.D2	Algorithm similar to the previous one. The difference is that the dissimilarities are squared before updating the cluster.
Centroids	Algorithm that groups individuals according to the geometric center of the cluster (centroid).
CLARA *(Clustering Large Applications)*	Algorithm that allows to perform as many groupings as indicated by the K parameter. It is used when there are large amounts of data because it reduces RAM consumption.

Source: Own elaboration

Figure 1 shows how to program the algorithm in the Rstudio© software; this procedure is the same for each cluster algorithm, only the one that was really useful for the study is presented, which was the CLARA algorithm. This algorithm is used because it was the one that generated more accuracy in the distribution of the groups and fulfilled the objective of achieving a simpler segmentation for the banks.

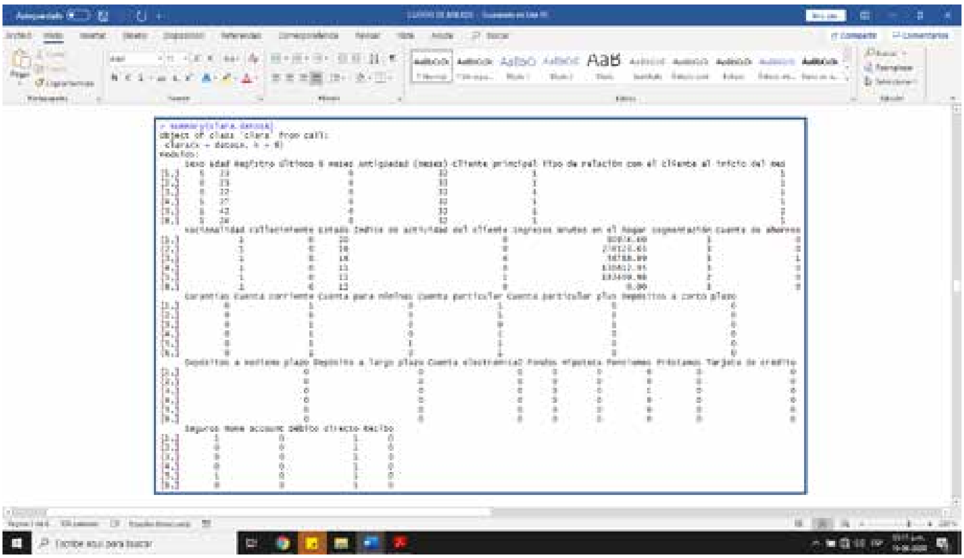
Figure 1. Algorithm *script* used to perform the *cluster* analysis.



Source: Own elaboration

Figure 2 shows the grouping performed by the algorithm and the characteristics of each of these groups. The objective of this characterization is to generate types of customers with their qualities, which will be used to generate a key indicator for the study and detect patterns that the banks did not know about their customers.

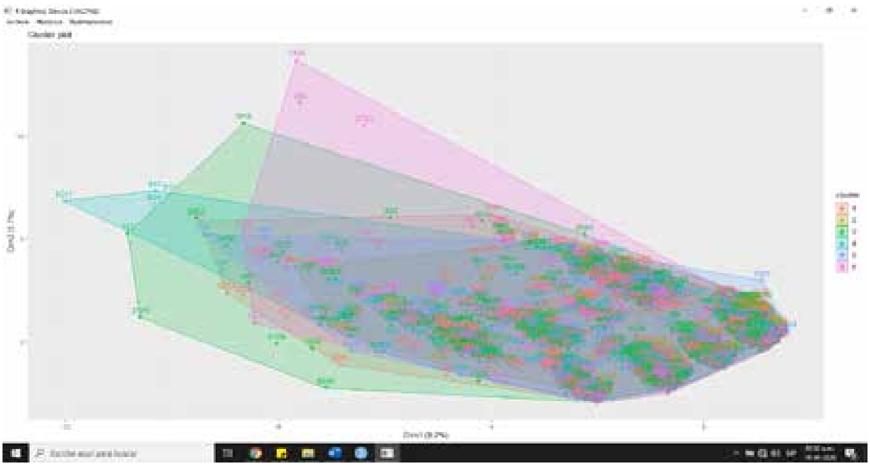
Figure 2. Results obtained from the grouping and characteristics of each cluster.



Source: Own elaboration

Figure 3 shows graphically how the banks' clients were grouped, although it is difficult to see each of the points, the most important is represented in Figure 2.

Figure 3. Graphical representation of the clusters



Source: Own elaboration

Table 6 shows the different algorithms with the optimum number of groups detected by the algorithm and then the number recommended by the researchers.

Table 6. Grouping by each clustering algorithm

Cluster method	Optimal number of groupings	Recommendation index
Minimum distances	6	9
Maximum distances	7	11
Ward.D	6	7
Ward.D2	6	7
Centroids	5	11

Table 7 shows the characteristics of each cluster translated into the original values, which allow us to identify patterns and knowledge to generate indicators, on the other hand it can be seen that in the relevant products there is an acronym, which indicates the type of account held by customers, for example: CA (Savings Account), CP (Main Account), CC (Checking Account) and DD (Direct Debit). Finally, there are the number of customers belonging to a cluster.

On the other hand, there are variables that are not contributing, as was the case with the customer segmentation that the banks had carried out.

Table 7. Client characteristics by *cluster*

Cluster	Segmentation	Sex	Age	Nationality	State	Revenues	Relevant products	Size
1	3	Male	23	Venezuelan	Sucre	90.974,40	CC, CP, Insurance and DD.	1848
2	3	Female	23	Venezuelan	Miranda	274.123,65	CP and DD.	366
3	3	Female	22	Venezuelan	Miranda	56.788,89	CA, CC, CP+, Pension and DD.	2459
4	3	Male	27	Venezuelan	Capital Dept.	130.612,95	CC, CP AND DD.	1077
5	2	Male	42	Venezuelan	Capital Dept.	183.409,98	CC, CN, CP, Insurance and DD.	577
6	3	Male	24	Venezuelan	Falcon	-	CC, CP AND DD.	1669

Source: Own elaboration

Construction of indicators

In this stage, management indicators were constructed based on the information obtained from the *cluster* algorithms and statistics generated by the Rstudio® software, where 4 indicators were created with their respective indicator sheets. The following tables show the elements that make up the indicators' data sheet: name of the indicator, definition, calculation method, sources, person responsible for the calculation, unit of measurement, periodicity and observations.

After developing the indicators, they were applied to real data to obtain information that could be used in decision making.

Table 8. Output Efficiency Indicator Sheet

INDICATOR DATA SHEET (1)		
NAME OF THE INDICATOR:	Product efficiency.	
DEFINITION:	It represents the percentage of success (sales) that a product has in the market.	
CALCULATION METHOD:	Ratio of the quantity of product X sold to the total number of customers.	**Numerator:** quantity of product X sold.
		Source of numerator: company.
		Denominator: total number of customers.
		Denominator source: company.
RESPONSIBLE:	Sales manager of the company.	
UNIT OF MEASUREMENT:	Percentage.	
PERIODICITY:	Monthly.	
REMARKS:		

Source: Own elaboration

Table 9. Company Coverage Indicator Sheet

INDICATOR DATA SHEET (2)		
NAME OF THE INDICATOR:	Coverage of the company.	
DEFINITION:	This is the percentage of products adapted to reality.	
CALCULATION METHOD:	Quotient of the sum of the efficiency of all products divided by the total amount of products.	Numerator: Sum of the efficiency of each product.
		Source of numerator: company.
		Denominator: total number of products offered by the company.
		Denominator source: company.
RESPONSIBLE FOR THE CALCULATION:	Sales manager of the company.	
UNIT OF MEASUREMENT:	Percentage.	
PERIODICITY:	Monthly.	
REMARKS:		

Source: Own elaboration

Table 10. Most Sold Product Indicator Sheet

INDICATOR DATA SHEET (3)		
NAME OF THE INDICATOR:	Best selling products.	
DEFINITION:	Percentage of revenues represented by the five best-selling products.	
CALCULATION METHOD:	Ratio of the number of best-selling products to the total number of products sold.	**Numerator:** number of best-selling products.
		Source of numerator: company.
		Denominator: total number of products sold.
		Denominator source: company.
RESPONSIBLE FOR THE CALCULATION:	Sales manager of the company.	
UNIT OF MEASUREMENT:	Percentage.	
PERIODICITY:	Monthly.	
REMARKS:		

Source: Own elaboration

Table 11. Consumption Index Indicator Sheet

INDICATOR DATA SHEET (4)		
NAME OF THE INDICATOR:	Consumption index.	
DEFINITION:	Average number of products purchased per customer.	
CALCULATION METHOD:	Ratio between the total amount of products sold and the total amount of customers.	**Numerator:** number of products sold.
		Source of numerator: company.
		Denominator: total number of customers.
		Denominator source: company.
RESPONSIBLE FOR THE CALCULATION:	Sales manager of the company.	
UNIT OF MEASUREMENT:	For each customer.	
PERIODICITY:	Monthly.	
REMARKS:		

Source: Own elaboration

Table 12. Sales by State Indicator Sheet

<table>
<tr><th colspan="3">INDICATOR DATA SHEET (5)</th></tr>
<tr><td>NAME OF THE INDICATOR:</td><td colspan="2">Sales by state.</td></tr>
<tr><td>DEFINITION:</td><td colspan="2">Percentage represented by each state in terms of sales.</td></tr>
<tr><td rowspan="4">CALCULATION METHOD:</td><td rowspan="4">Quotient between the quantity of products sold in state X and the total quantity of products sold.</td><td>Numerator: number of products sold in state X.</td></tr>
<tr><td>Source of numerator: company.</td></tr>
<tr><td>Denominator: Total number of products sold.</td></tr>
<tr><td>Denominator source: company.</td></tr>
<tr><td>RESPONSIBLE FOR THE CALCULATION:</td><td colspan="2">Sales manager of the company.</td></tr>
<tr><td>UNIT OF MEASUREMENT:</td><td colspan="2">Percentage.</td></tr>
<tr><td>PERIODICITY:</td><td colspan="2">Monthly.</td></tr>
<tr><td>REMARKS:</td><td colspan="2"></td></tr>
</table>

Source: Own elaboration

Results of the study

Both banks are present in these results in order to achieve a *benchmark* with respect to the indicators described above.

Current values of indicators

After the indicators were constructed, they were applied to determine the current situation of each bank, as well as to demonstrate the usefulness and purpose of the indicators. The table below shows the meaning of each color in the indicator graphs.

Table 13. Color index in the graphs.

Color	Score
Green.	Excellent.
Purple.	Adequate.
Red.	Deficient.

Source: own elaboration.

Product efficiency

The results obtained by applying the indicator defined as "Product Efficiency" to each banking institution are shown in the graphs below:

Efficiency is measured by the percentage of success (sales) that a product has in the market.

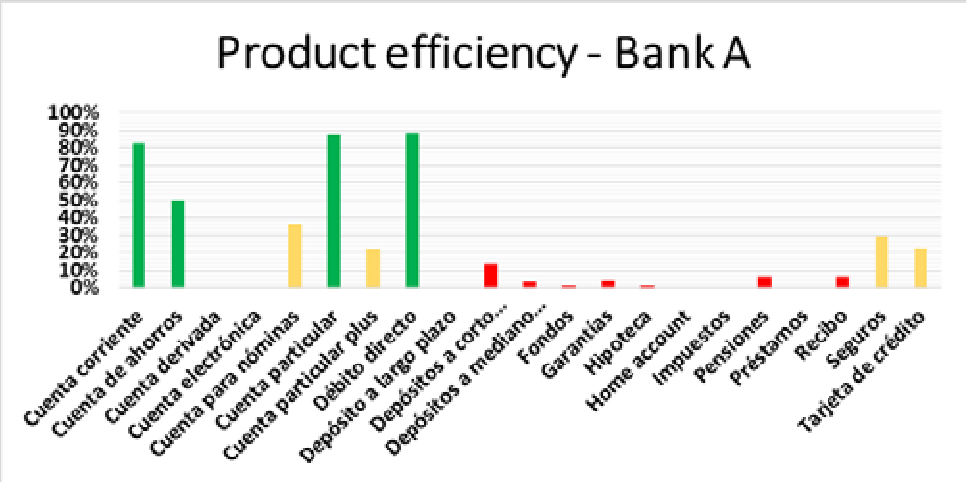


Figure 4. Product efficiency, Bank A.

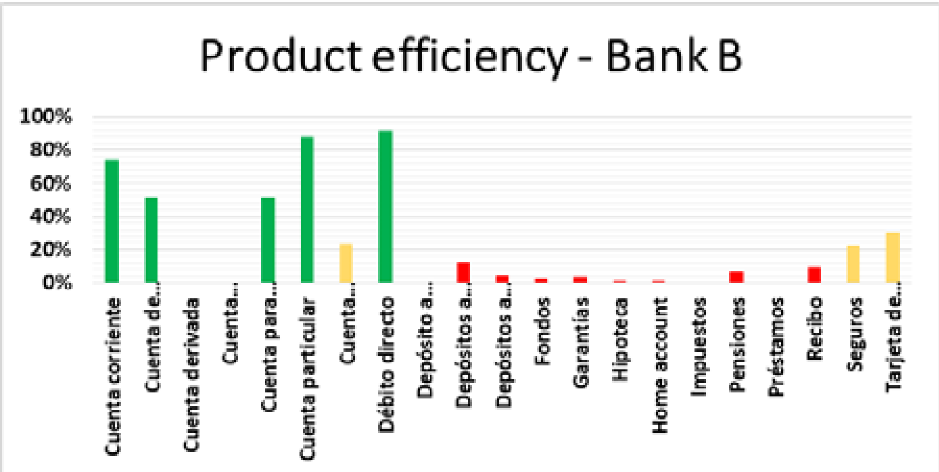


Figure 5. Product efficiency, Bank B.

It can be seen that there are many products that are not used by the customers of both banks.

Best selling products

The results obtained by applying the indicator defined as "Best-selling products" to each banking institution are shown in the graphs below:

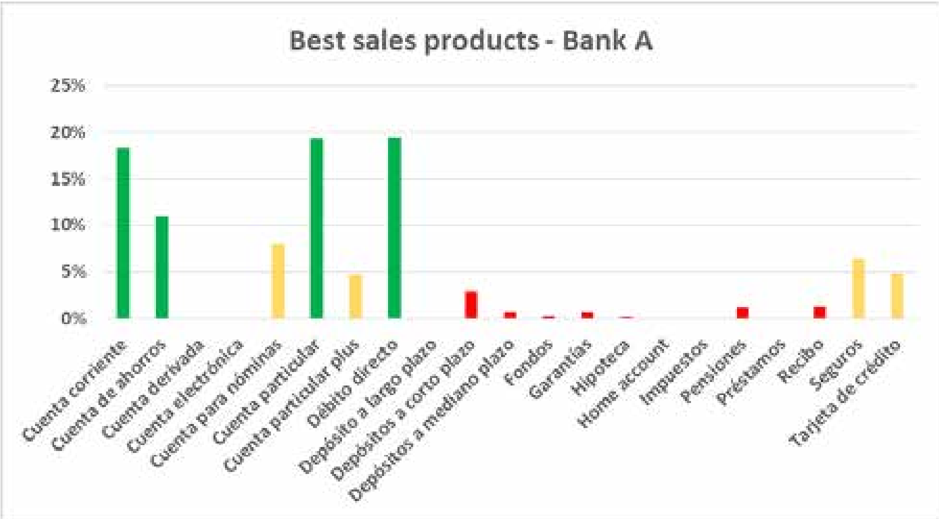


Figure 6. Percentage of total sales, Bank A.

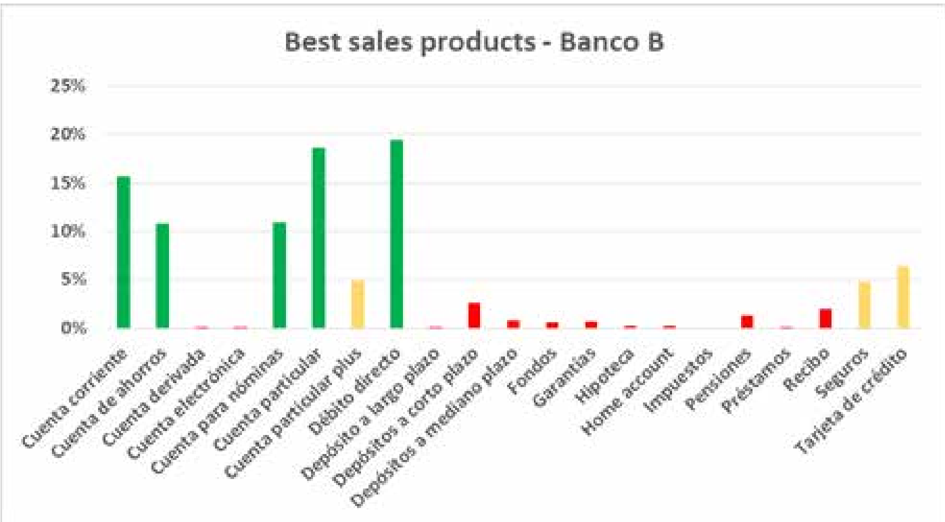


Figure 7. Percentage of total sales, Bank B.

Where it can be seen that the banks have too many products that should be reviewed in view of the current market.

Sales by State

The results obtained by applying the indicator defined as "Sales by State" to each banking institution are shown in the graphs below:

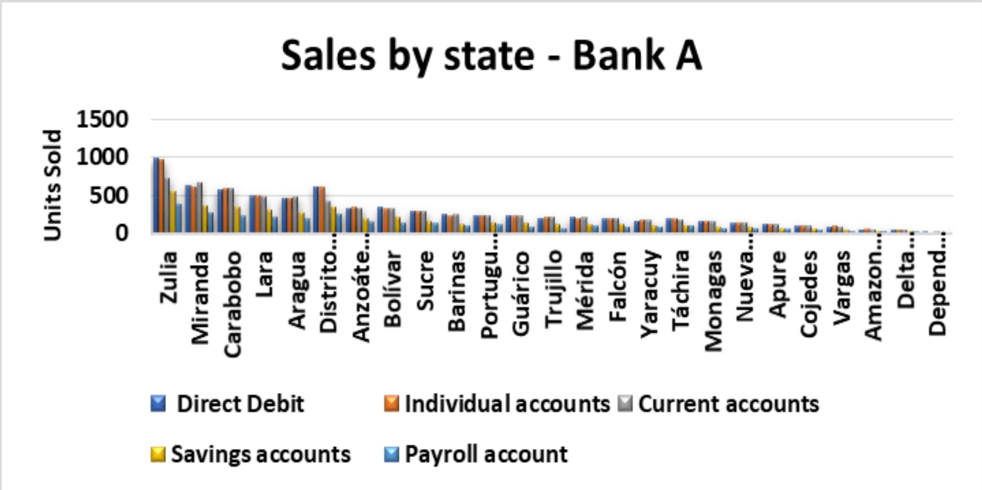


Figure 8. Sales by State, Bank A.

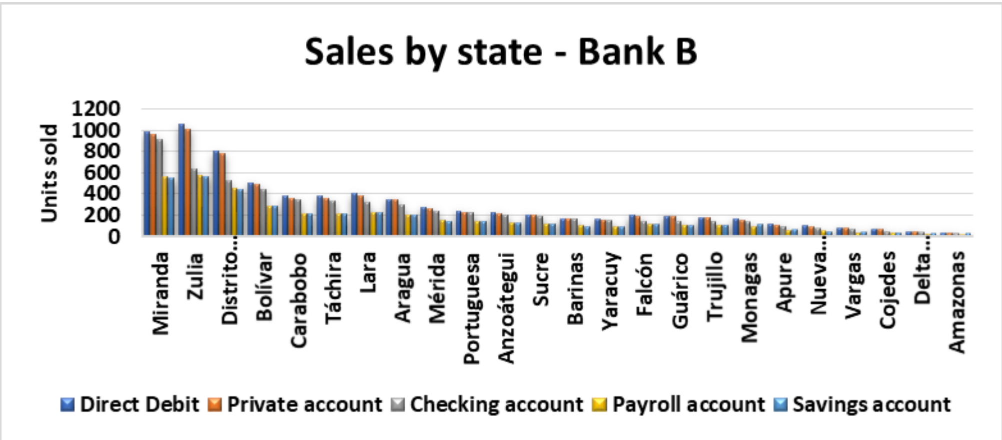


Figure 9. Sales by State, Bank B.

Dashboard generation

To facilitate the reading of these graphs, a global *dashboard* of each of the indicators was generated using pivot tables and the result of each of the proposed indicators

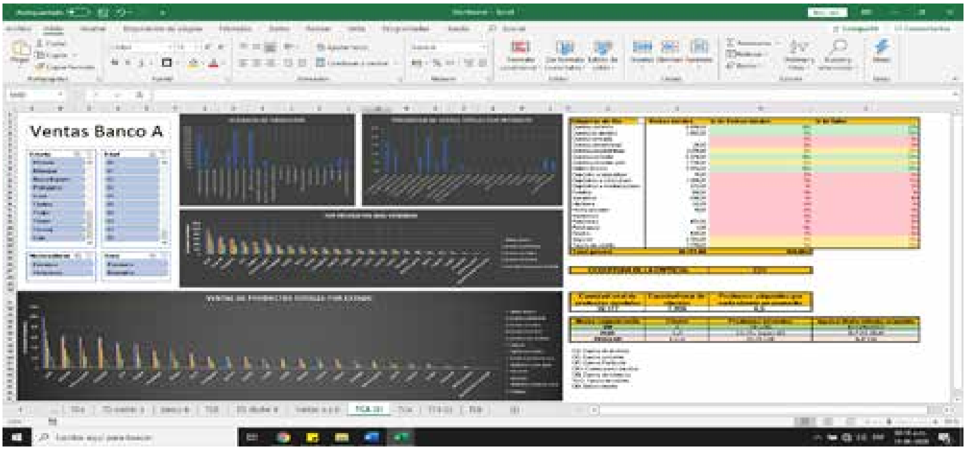


Figure 10. Dashboard of Bank A

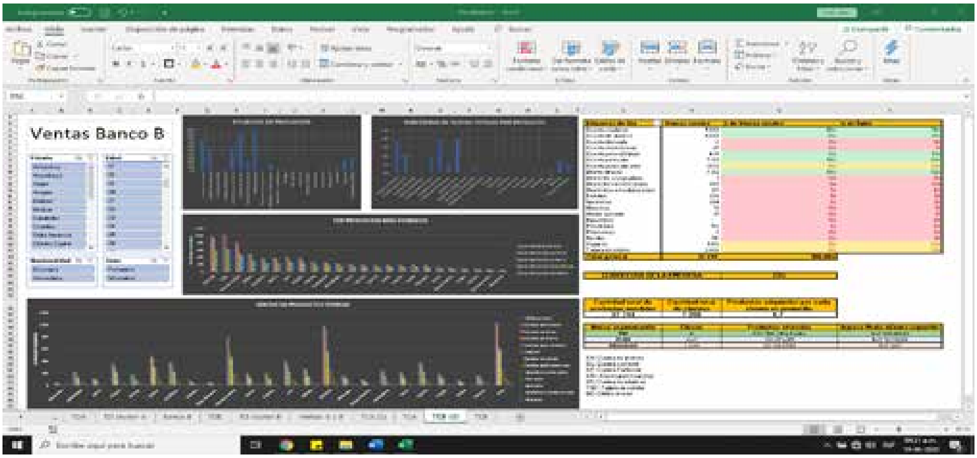


Figure 11. Dashboard of Bank B

Conclusions and recommendations

Finally, the conclusions of the research project, the new objectives and recommendations to be taken into account to improve the customer service process and the future situation of each bank in general are presented.

Conclusions

- The format of the data provided is never ideal. A well-executed data selection, cleaning and data transformation process is the basis for designing a strategy to generate desired and safe results because if the input data is not reliable, neither will be the information obtained (Garbage In, Garbage Out). This methodology prevents obtaining erroneous information by ignoring non-influential variables, reducing noise, eliminating inconsistencies and validating that the information is adequate for the data mining algorithm.
- Although *data mining* processes are methodical, it is very common to perform several iterations throughout the project varying several parameters, such as the mining algorithm, objectives and even the data. This is done in order to obtain accurate information of the highest possible quality to facilitate decision making.
- Both companies have certain deficiencies at the time of collecting information and/or recording the data, since the following errors were found: one variable with all its corresponding values empty (Last date as main client) and they did not reflect the date of obtaining the data but the date of recording the entire set. In addition, the existence of empty values in the gross income of the household may pose a legal problem, since Venezuelan regulations make it mandatory to declare income.
- Two major problems were identified in each banking institution: customer segmentation and the excess of products offered. This arises due to a deficiency in customer service, since the companies did not know how to pinpoint the characteristics and needs of their clients. Therefore, in addition to classifying them in an incongruent hierarchy, they created a large number of products that are not generating benefits for the company.
- The critical variables associated with the segmentation process were gross household income, cluster size and the following products: savings accounts, checking accounts, individual accounts, individual plus accounts, payroll accounts, credit cards and direct debit.
- The following indicators were designed: product efficiency, company coverage, best-selling products, consumption index and sales by state, to measure the current situation of each company in terms of sales and to identify products that do not have added value.

The results obtained by applying the company's Coverage indicator are well below the suggested acceptable value; however, they are not definitive due to the lack of information on revenues for each product. Even so, taking into account that this is the most important indicator, the results obtained from the application of the Coverage indicator are well below the suggested acceptable value.

- Through statistical tools and procedures, reliable information was obtained to determine the current situation of a company in one or several specific aspects, which, in turn, facilitated the benchmarking process by comparing only the most influential details for the analysis.

Recommendations:

- The database should be improved or an automated database should be created to collect and record the information. This will facilitate the application of the indicators and data mining process by not having to invest so much time and resources in data cleaning.
- It is essential to constantly evaluate the results of the indicators collected, according to the frequency established for each one of them, in order to take preventive and/or corrective measures in the event of finding any value outside the range established as optimal by the company.
- Compare the results obtained with those of the previous period to verify that sales are in line with the objectives and that the corrections made, if any, are effective.
- Products that score 20% or less on the Product Efficiency indicator should be subjected to a thorough analysis, taking into account the income generated by such products, to decide whether they should be eliminated, modified or left as they are. It is very likely that in most cases products with low efficiency will be eliminated if they happen to be interest-bearing assets, such as loans or mortgages, since in an inflationary situation money loses value over time if the contract is not anchored to a strong currency such as the dollar. However, if the relevant studies are carried out as indicated above and it turns out that customers need this type of product, the company must make the corresponding decisions in order to meet those needs.
- Implement a new segmentation that works for each bank, so that they can provide personalized customer service and offer products related to the characteristics and/or requirements of the customer or consumer.

Acknowledgments

José Valentín Salazar

I would like to thank my parents first of all, for giving me the opportunity to study at this prestigious university, for all the support throughout my career and for letting me know that I am capable of achieving anything I set my mind to.

I wish to express my gratitude to the authorities of the Universidad Metropolitana for keeping this university functioning at the highest level and, especially, to my tutor Siro Tagliaferro for guiding me throughout this research work.

I thank my brother and my friends, who knew how to help me in difficult moments.

Siro Tagliaferro

I would like to thank my school, Universidad Metropolitana, and all the professors who gave me their support to achieve this research.

References

American Productivity And Quality Center (1993). *The benchmarking management guide*. Productivity Press, Cambridge, Massachusetts. Retrieved October 25, 2019, from: http://mautic.agenciaf2b.com.br/The-Benchmarking-Management-Guide-New_Release_2019.pdf

Balagueró, T. (2018). What are the business intelligence components in *Big Data*. Deusto Formación. Retrieved October 25, 2019 from: https://www.deustoformacion.com/blog/gestion-empresas/cuales-son-componentes-business-intelligence-big-data

Bisquerra, R. (1989). *Educational research methods: A practical guide*. Retrieved October 11, 2019, from: http://dip.una.edu.ve/mead/metodologia1/Lecturas/bisquerra2.pdf

Botella, P., Alacreu, M. and Martínez, M. (n.d.). *Installation and introduction to R statistical software and the R-Commander library. Descriptive statistics*. Retrieved November 4, 2019, from: https://www.uv.es/~mamtnez/IRCED.pdf

Camargo-Vega, J, Camargo-Ortega, J and Aguilar, L. (2014). *Getting to know Big Data*. Revista Facultad de Ingeniería. Issue 38. Retrieved on September 28, 2019 from: http://www.scielo.org.co/pdf/rfing/v24n38/v24n38a06.pdf

Cebotarean, E. (n.d.). *Business intelligence*. Retrieved October 25, 2019, from: http://www.scientificpapers.org/wp-content/files/1102_Business_intelligence.pdf

Gallardo, J. (n.d.). *Introduction to Cluster Analysis. General considerations.* Retrieved January 19, 2020 from: https://www.ugr.es/~gallardo/pdf/cluster-1.pdf

Guerrero, F. and Rodríguez, J. (2013. *Design and development of a guide for the implementation of a Big Data environment at the Catholic University of Colombia*. Catholic University of Colombia. Retrieved October 25, 2019, from: https://repository.ucatolica.edu.co/handle/10983/1320

Gurutze, M. and Ochoa, C. (2005). *A THEORETICAL REVIEW OF THE BENCHMARKING TOOL*. Revista de Dirección y Administración de Empresas. Number 12, Spain. Retrieved October 25, 2019 from: https://www.ehu.eus/documents/2069587/2113623/12_6.pdf

Hernández, R, Fernández, C and Baptista, P. (2004). *Research methodology.* Retrieved October 11, 2019 from: http://sistemas.unicesar.edu.co/documentossistemas/sampieri.pdf

Hare, J. (2016). *Gartner Survey Reveals Investment in Big Data Is Up but Fewer Organizations Plan to Invest.* Stamford, EU. gartner. Retrieved October 11, 2019 from: https://www.gartner.com/en/newsroom/press-releases/2016-10-04-gartner-survey-reveals-investment-in-big-data-is-up-but-fewer-organizations-plan-to-invest

IDATHA (2014). *BIG DATA -General Concepts.* Retrieved October 25, 2019, from: http://idatha.com/whitepapers/Whitepaper-bigdata.pdf

López, P. (2004). *SAMPLE POPULATION AND SAMPLING*. Point Zero, 09(08), 69-74. Retrieved November 4, 2019, from: http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S1815-02762004000100012&lng=es&tlng=es

Luhn, H. (1958). *Business Intelligence System*. IBM Journal. Retrieved October 25, 2019, from: http://altaplana.com/ibm-luhn58-BusinessIntelligence.pdf

Macassi, S. and Mata, M. (1997). *How to develop samples for audience surveys*. Cuadernos de investigación No 5. ALER, Quito.

McAfee, A. and Brynjolfsson, E. (2012). *Big Data*: The Management Revolution. Harvard Business Review. Retrieved November 14, 2019, from: http://tarjomefa.com/wp-content/uploads/2017/04/6539-English-TarjomeFa-1.pdf

Molina, L. (2002). *Data mining: torturing data until they confess*. Retrieved October 25, 2019 from: https://www.businessintelligence.info/resources/assets/dss/molina-torturando-datos.pdf

Mondragón, A. (2002), What are indicators? National Institute of Statistics, Geography and Informatics. Journal of information and analysis. Number 22. Retrieved January 19, 2020 from: https://www.orion2020.org/archivo/sistema_mec/10

Muñiz, L. (2018). *Business Intelligence key tool to improve business management*. Retrieved October 25, 2019 from: http://www.sistemacontrolgestion.com/Portals/1/Ebook_Mejorar_gestion_BI_SCG_Estrategia_v18.pdf

Niño, M and Illarramendi, A. (2015). *UNDERSTANDING BIG DATA: BACKGROUND, ORIGIN AND FURTHER DEVELOPMENT.* University of the Basque Country. Retrieved October 25, 2019 from: https://www.dyna-newtech.com/busqueda-NT/entendiendo-big-data-antecedentes-origen-y-desarrollo-posterior

Paz, R. (2005). Customer service. Vigo, Spain: Ideaspropias Editorial.

Robles, R. (2016). *The advantages of Big Data.* Retrieved September 28, 2019 from: https://www.icemd.com/digital-knowledge/articulos/las-ventajas-del-big-data/

Raffino, M. (2018). Concept of customer service. Argentina. Concept of. Retrieved November 6, 2019, from: https://concepto.de/servicio-al-cliente/

Ruiz, B., Batanero, C. and Arteaga, P. (2011). *Linking the Random Variable and Statistics in the Realization of Informal Inferences by Future Teachers*. Boletim de Educação Matemática, 24(39). Retrieved November 4, 2019, from: https://www.redalyc.org/pdf/2912/291222099006.pdf

SAS (n.d.). *Customer intelligence in the era of data-driven marketing*. Retrieved October 25, 2019 from: https://www.sas.com/content/dam/SAS/es_mx/doc/infographic/inteligencia-de-cliente-en-la-era-del-marketing-basado-en-datos-spanish_ebook.pdf

Sinnexus (n.d.)*. Business Intelligence solutions for your enterprise*. Retrieved October 25, 2019, from: https://www.sinnexus.com/downloads/SnxFolletoComercial.pdf

Tascón, M. (2013). *Introduction to Big Data. Past, Present and Future*. Telos Magazine. Journal of Thought on Communication, Technology and Society. Issue 95. Retrieved September 28, 2019 from: https://telos.fundaciontelefonica.com/archivo/numero095/pasado-presente-y-futuro/

Timarán, S., Hernández, I., Caicedo, S., Hidalgo, A. and Alvarado, J. (2016). *The knowledge discovery process in databases.* In Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional (pp. 63-86). Bogotá: Ediciones Universidad Cooperativa de Colombia. Retrieved October 25, 2019, from http://ediciones.ucc.edu.co/index.php/ucc/catalog/download/36/40/230-1?inline=1

Universidad Santo Tomás, (n.d.). *The Statistical Method.* Retrieved November 4, 2019, from: http://soda.ustadistancia.edu.co/enlinea/Segunda%20unidad%20Cuanti/el_mtodo_estadstico.html

Valcárcel, V. (2004). *Data Mining and the Discovery of Knowledge*. Journal of the Faculty of Industrial Engineering Vol. (7) 2. Retrieved October 25, 2019 from: https://www.researchgate.net/publication/307181857_DATA_MINING_Y_EL_DESCUBRIMIENTO_DEL_CONOCIMIENTO

Vanegas, E and Guerra, L. (2013). *Business intelligence system to support the decision making process*. INGENIERÍA UC Magazine, 20(3). Retrieved September 30, 2019 from: https://www.redalyc.org/pdf/707/70732641004.pdf

Vicente, J. (n.d.). *Introduction to cluster analysis.* University of Salamanca. Retrieved June 15, 2020 from: https://www.yumpu.com/es/document/read/14514099/analisis-cluster-estadistica-universidad-de-salamanca