Heliyon 5 (2019) e02195

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.heliyon.com

Data maturity analysis and business performance. A Colombian case study

Didier Grimaldi [a,*], Javier Diaz [b], Hugo Arboleda [b], Vicenc Fernandez [c]

[a] Department of Strategy, Entrepreneurship and Innovation, La Salle Campus, Ramon Llull University, Sant Joan de La Salle, 42, 08022, Barcelona, Spain
[b] Universidad Icesi, Calle 18 No. 122-135, Pance, Cali, Colombia
[c] Department of Management, Universitat Politècnica de Catalunya BarcelonaTech, ETSEIAT, Colom 11, TR6, 3.17, 08222, Terrassa, Spain

ARTICLE INFO

Keywords:
Business
Computer science
Information science
Economics
Colombia
Data science
Competitive analytics
Qualitative comparative analysis
Business performance

ABSTRACT

Context: Colombia over the last decade has experienced a historic economic boom and Information Technology (IT) has been emerging as a tool to enable the competitiveness of companies. The last government (2014–2018) took different actions to explain how the use of data science and open data improve the business activity. The question to identify if there is a relationship between IT capacities, the organizational structure and the performance of the companies remains unresolved and is certainly an urgent issue for new government of Iván Duque.
Purpose: Our study analyses the relationship between data structure and business performance measured through the efficiency of customer experience and provider operations processes.
Methodology: our methodology is novel compared to previous researches which develop linear regression. It is based on the use of a fuzzy-set qualitative comparative analysis (fsQCA).
Originality/value: our method allows to reveal multiple and complementary paths to achieve possible correlations between data and business performance.
Findings: Our results show that data consistency, data usage and data protection are the three more frequent conditions to a better customer experience and provider operations efficiency. Surprisingly, data-driven profile is a necessary but not sufficient condition.
Practical implications: our conclusions allow practitioners to uncover the strength of the data to orientate their digital strategy. Our recommendations could be used for the new governmental program of digital revival for Small and Medium Enterprises.

1. Introduction

Porter (1998) argues that companies can better adapt themselves to the environment changes through the acquisition of data on the markets, customers and competitors. He adds their competitive position can increase if they are able to be more informed or able to faster adapt to unexpected changes. Many scholars have followed Porter's argument developing theories in the Strategic Management field. As an illustration, models such as RBV, strategic factor markets (Mintzberg, 1978), blue ocean (Randall, 2015) are born and permit to build methodologies of data collection and data strategy. Their main objectives are first to analyse the industry, the strengths and weaknesses of the studied company along with the bargaining power of the different stakeholders engaged in the marketplace. Secondly, they look for how the company by developing specific resources may successfully compete and win in its environment (Mintzberg, 1978). This double analysis (external and internal) passes through a process to convert amounts of data into insightful inputs for the management of the firm (Mata et al., 1995).

McAfee and Brynjolfsson (2012) have defined the data science concept as the upcoming era of innovation and productivity raise. Consequently, many business executives engage their organizations into data science projects in the hopes of better monitoring and managing their organizations (Evans, 2013). Nevertheless, a lot of them are still wondering how to obtain the expected value that many consulting firms relayed by global media companies have been announced for many years (Forbes, 2014). This study aims at understanding the relationship between the use of data science and the firm performance. It participates to the topical academic discussion on how data science is extending the business strategy toolset (Grimaldi et al., 2019).

Colombia, over the last decade, has experienced a historic economic boom. In 1990, Colombia had a GDP per capital of only US$ 1,500 while in 2015 it increased to over US$14,000. By 2015, Colombia became the 4th largest in Latin America, and the world's 31st largest. Modern industries like shipbuilding, electronics, automobile, tourism, construction,

* Corresponding author.
E-mail address: didierg@salleurl.edu (D. Grimaldi).

https://doi.org/10.1016/j.heliyon.2019.e02195
Received 21 January 2019; Received in revised form 25 April 2019; Accepted 29 July 2019

and mining, grew dramatically during the 2000s and 2010s, even if most of Colombia's exports are still commodity-based (Colombia, 2018). Colombia is Latin America's 2nd-largest producer of domestically-made electronics and appliances only behind Mexico (World Bank, 2018). Due to these overwhelming results, we decided to focus our present study to the Colombian business activity considering all the industries of the economy (see demographic survey details in section 3.1).

Nevertheless, our method of analysis is different from those used in past researches (Bonabeau, 2003; Evans, 2013). Historically, scholars have conducted linear regression examining causality between independent variables and a dependent variable considered as the outcome (Davenport and Harris, 2007; Harris and Craig, 2011). Instead of statistics based in linear algebra, our study uses qualitative comparative analysis (QCA) method in order to describe the correlation between the data structure and the performance results of the firm. Our scientific approach moves from the observation of a single effect of a data characteristic to the analysis of a combination of multiple variables associated with high business performance (Fiss, 2007). The structure of the remainder is as follows. Section 2 is the literature review while section 3 shows the methodology used. Section 4 develops and analyses the results which are discussed in the section 5. In this section also, we present a conclusion and suggest future lines of research.

2. Related work

2.1. Colombian context

Information technologies (IT) are emerging in Colombia as a tool to enable the competitiveness of companies. Different digital transformation programs were launched in 2016 by the State to improve the productivity of Small and Medium Enterprises (SMEs) namely: MiPyme, Vive Digital or Digital Entrepreneur among others (Mintic, 2016). The Colombian economy is based at 99% on SMEs as Mintic (2016) recently reported. This type of company offers 81% of the employment and for this reason, the Colombian government carries out studies to fortify them. Maza et al. (2016) examine how IT contributes to the performance of Colombian organizations for the specific textile sector. For this case, they intend to identify if there is a relationship between IT capacities, the organizational structure and the performance of Bogota-based SMEs.

Rojas et al. (2014) argue that both data science and open data can transform business, government, and society and a combination of the two is especially potent. They observe the last government of Colombia (2014–2018) took small steps and it still has a long way to go to make use of data science and open data for the benefits of the economy. Even if the first results are promising and accessible for Entrepreneurs as a result of a pragmatic public policy, they are not reflected in observable and sustainable benefits for the SMEs (Fedesarrollo, 2013). They highlight also that the important question is to examine in detail how data science can impact Colombian businesses.

2.2. Correlation between business performance and analytics

Lavalle et al. (2011) determine the utilisation of data affects positively the performance of an organization. Indeed, they interview 500+ companies all around the world, collecting data that they analyse using classic Bayesian analytics model. Their findings are based on opinions (e.g. if the CEO considers his company as top performing against his competitors) and their conclusions are those who use data are twice more efficient than those who don't do it. For our paper, we decided to use a different method and to move from the analysis of single effects competition to the study of a combination of variables which all together provide the same outcome, i.e., each of them covering one portion of the final model (Fiss, 2007; Ragin and Strand, 2008).

However, we consider two different issues. On one side, we concentrate our efforts on how analytics improve the design of the front-end applications by integrating more smoothly the different online and offline steps of the customer journey between them and improving finally the Client shopping experience (Gobble, 2013; Grimaldi et al., 2019; Tan et al., 2017). On the other side, we focus our attention on how the new capabilities of the data science assist the back-end applications which allow customers and suppliers to share visibility of their inventory and needs, in order to achieve a minimum level of stock and to decrease logistics and warehouse management costs (Addo-Tenkorang and Helo, 2016; Forza and Salvador, 2001; Gunasekaran and Ngai, 2004; Hofmann, 2017). In summary, we decided to look for the possible casual conditions that lead to a firm performance improvement building separately two different models, the first for the processes dedicated to the improvement of the customer experience, the second for those dealing with a better efficiency of the provider operations.

2.3. Data maturity model

Spruit and Pietzka (2014) propose a framework to assess the data maturity of an organization and uncover the barriers due to their structure of the data. They show that one of the main impediment is the incapacity for companies to correctly manage huge amounts of data. Davenport and Prusak (2000) add the difficulties increase while the data are unstructured originated from sensors or other connected devices. Wegener (2008) observes that data require a maturity level to provide value-added insights. In transport services, the companies need accurate data to comply with hard procedures for the recording of the events. The same issue exists with the food and consumer products industry and the pharmaceutical sector while the security of the citizens requires accountability in order to satisfy rigorous controls of public health. Ecommerce, finance and trading companies who are by definition intense consumers and producers of data, develop large efforts to build and maintain precise datasets to monitor their business and decide future lines of growth. Even B2B business like L'Oreal, New Balance, Dove, … have recently modified their go-to-market strategy to directly address their final consumers' needs and collect information through web, mobile application or social media tools about their online behaviour.

Becker et al. (2009) suggest a maturity data model (MDM) as an artefact to better manage the data of an organization. After a deep analysis of the literature about MDMs, they highlight four areas to cover all the main aspects of data management which are: data consistency, data completeness, data usage (who uses the data in which systems) and data protection (confidentiality). This present study aims at exploring the different combinations of these four causal variables to the efficiency of the customer experience and the provider operations processes.

2.4. Data-driven profile

Gobble (2013) states that companies which permanently use business analytics to outperform the competitors can be considered as data driven. Davenport and Harris (2007) precise a data-driven company looks beyond basic statistics and use predictive modelling for internal and external aspects of their business. For external aspects they refer to either the management of customers e.g. understand better the purchase behaviour or suppliers e.g. simulate alternatives of route shipments to minimize the impact of unexpected malfunctions (Bonabeau, 2003). Internal aspects mainly deal with the management of the finance, risk or human resource business processes (McAfee and Brynjolfsson, 2012). In this context, Davenport and Prusak (2000) highlight the case of Healthways, a global insurance company able to gather clients' personal data with their agreement and combined with hospital procedures to predict which patients could be at highest risk for bigger future medical expenses over the next years.

Woerner and Wixom (2015) argue even if the data are easier and easier to collect, most of the CEOs that they have interviewed are still reluctant to appreciate the outputs of a big data analysis and to take their decision based on a fact-based decision-making process. They prefer instead to use their business experience or what they usually call feeling.

Consequently, we define the variable "data-driven" from now as the attribute of a company which prefers to use (big) data instead of instinct or experience to manage a business issue. Finally, our research question consists on examining how the consistency, the completeness, the usage and the protection of the data may affect the customer experience and the efficiency of the provider operations and if the data-driven variable plays a moderating role in this possible relationship.

3. Methodology

Qualitative Comparative Analysis (QCA) is a research methodology to conceptualize and analyse causality. It differs from statistics, which estimates the separate contribution of independent variable in explaining variation of a dependent variable using linear algebra. The possible variation indeed determines a correlation between them with a level of probability of finding estimated by the p-value calculation. On the other hand, QCA is a appropriate for small N (e.g. between 10 and 50 cases) (Ho et al., 2016) that incorporates Boolean logic for a comparison of principles. QCA aims at determining the conditions of necessity and sufficiency between the causes and the outcome (Ragin and Strand, 2008).

Two different modes exist in the application of QCA method. The crisp-set is usually used for variables with binary values; for example the value '1' stands for the presence of a condition and the value '0' its negation. The fuzzy-set is usually used to describe continuous variables as it is the case for our study. The computation is realised using the package QCA of R software (Dusa, 2019). The advantages of QCA in comparison with correlational techniques are double: (a) asymmetry, meaning the presence and the absence of the outcome, respectively may lead to different explanations; (b) equifinality, which means that different paths can lead to the same outcome.

Our method follows two phases that can be repeated according to the results achieved. The first phase called calibration of the values aims at determining three thresholds for the values: full membership (or full inclusion), full non-membership and cross-over. The second phase is the analysis of the data which follows in its turn, three steps. The first step consists on building a table gathering all the information received through our survey. This table has 32 rows. By adjusting the inclusion parameter, we modify the number of possible combinations. The minimum inclusion (also called consistency) should be, according to (Ragin

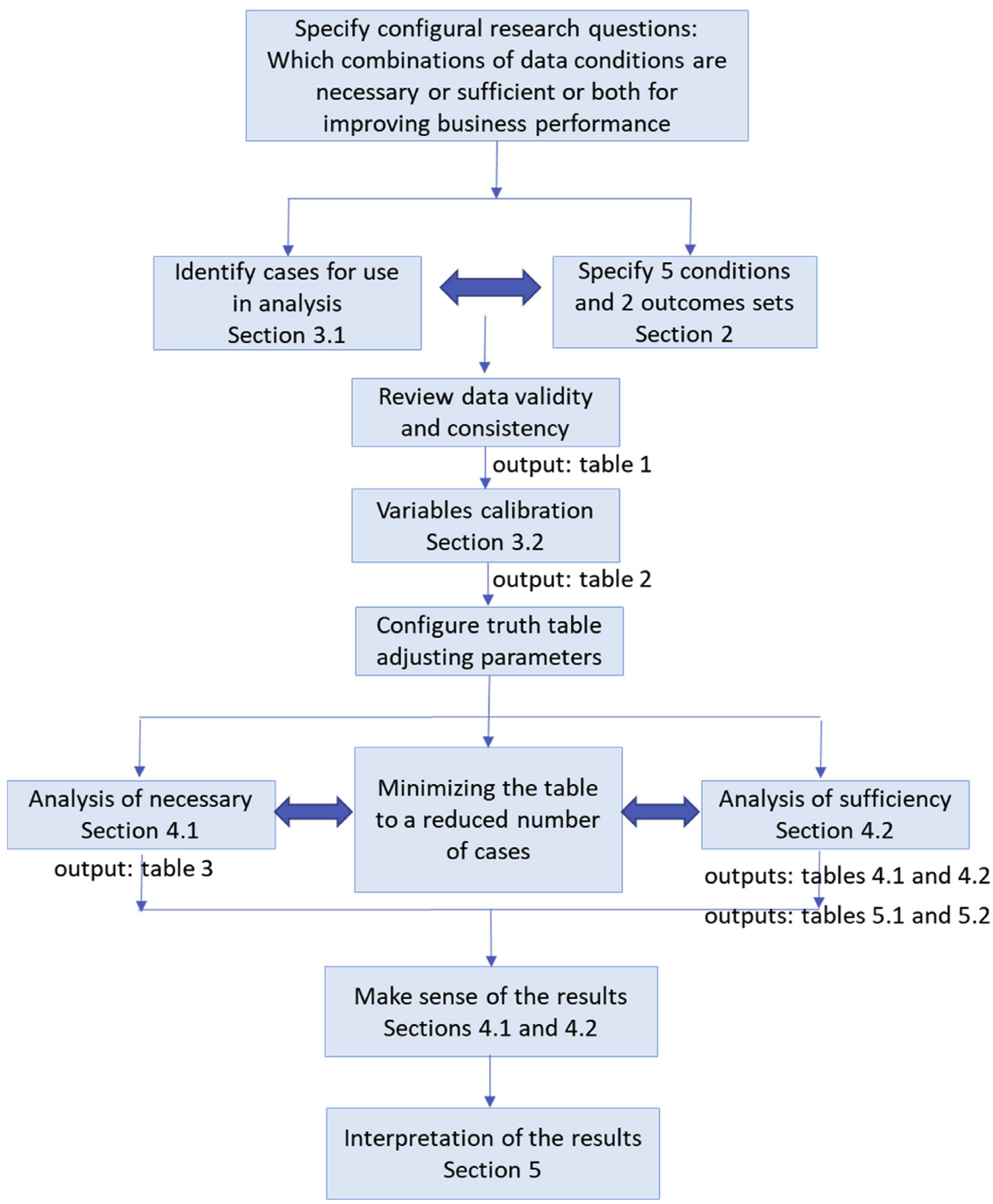

Fig. 1. Data Analysis procedure and sequence.

and Strand, 2008), equal or superior to 0.75. In our case it is 0.9.

The second step aims at minimizing the table to a reduced number of cases required to obtain the outcome. The parameter to be considered is the frequency or number of equivalent cases. The third and last step deals with transforming through a Boolean algorithm the truth table into the combinations of variables that produce the outcome. The analysis procedure and the sequence involved in the research is therefore summarised in the Fig. 1. We show in the Fig. 1 when the different steps of the analysis generate the tables and linking them with the different sections of this manuscript, we show how the data in the tables are derived.

3.1. Data collection and demographic results

Our study uses a questionnaire to collect data from companies operating in Colombia. This questionnaire is addressed to Director of Technology or Director of Information Systems from companies based in Colombia. We were helped by the yearly report issued by the Mintic institution *("Ministerio de Tecnologías de la Información y las Comunicaciones")* which describes the Colombian economic activities and includes an organizational structure overview (Mintic, 2010). We sort it out an initial list of 150 Senior Executives who were our target for our survey. We got 42 answers for a 28% response rate. After the removal of missing data and outliers, a sample of 36 responses were kept for QCA analysis. The size of the companies that answered the questionnaires are showed in Fig. 2.1 while their countries of operation are described in Fig. 2.2 and the company industry in Fig. 2.3.

We address the 10 major industries in Colombia: Retail, Academia, Consumer & Industrial Products, Energy & Resources, Financial/Insurance, Life Science & Health care, Manufacturing, Government, Transportation/Logistics and finally Technology, Media and Telecommunications even if the latter is the most representative one with 7 out of 36 (Fig. 2.3). The public sector is also represented with 4 answers. In terms of company size, half of the sample are medium companies between 250 and 499 employees (Fig. 2.1) which is aligned with the statistics published by Mintic (2010). Almost all the companies however operate in their mother country: Colombia (Fig. 2.2).

The Appendix shows the 31 measurement items of our survey. Those items are scaled according to a 5-point Likert rule where 5 corresponds to "completely agree" and 1 to "completely disagree". Our survey has a google format and three sections. First, the professionals answer how much they feel comfortable against different statements which evaluate the maturity of the company data according to 4 dimensions as described in the appendix. Then, the survey raises questions in order to evaluate the data-driven propensity of the company (i.e. to take decision based on data instead of instinct/experience). In the second section, the respondents indicate how much their data provide benefits in the management of their Customers and Providers. In the last section, we collect demographic information as control variables for our study.

For each construct of the sections 1 and 2, we calculate the Cronbach alpha, the composite reliability (CR) and the Average Variance Extracted (AVE) to check the validity and coherence of our data (see Table 1). Our results show that our dataset satisfies the conventional thresholds of 0.5 for AVE, 0.7 for either Cronbach alpha or composite reliability (Díaz-Díaz and de Saá-Pérez, 2014).

Finally, the survey concludes in section 3, with questions about demographic information about the respondent such as the name of the organization, the age, the number of employees, etc. (see Appendix).

3.2. Variables calibration

We calibrate our results using three thresholds: full membership, full non-membership and cross-over. In descendent order, the full membership is equivalent to a response "completely agree". A full non-membership is equivalent to either "completely disagree" or "disagree". Finally, the cross-over corresponds to the option "neither agree nor disagree". For the remaining, we apply this following linear regression function (1):

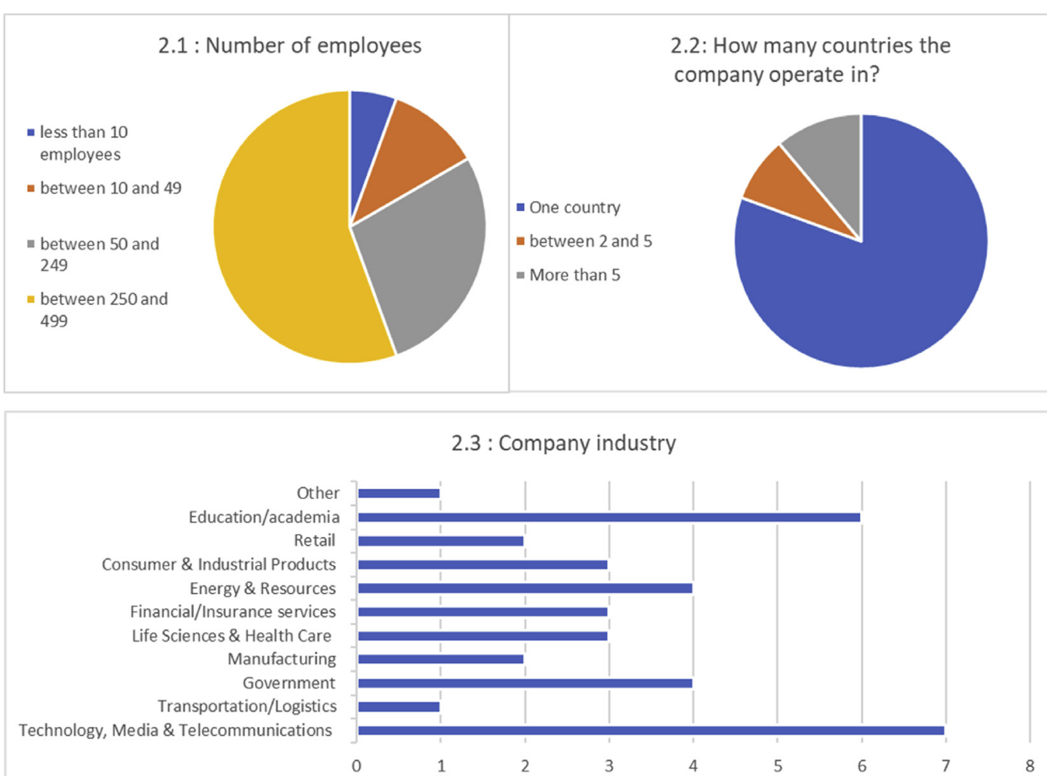

Fig. 2. Demographic results.

Table 1
Description of the conditions and the outcome.

Condition	Condition abbreviation	Description	Factor analysis
Data consistency	Dat_con	Expressing the data in the same way	ICR = .68 CR = .804 AVE = .510
Data completeness	Dat_cmp	No data are missing	ICR = .69 CR = .859 AVE = .753
Data usage	Dat_usg	Access of data is defined	ICR = .83 CR = .923 AVE = .857
Data protection	Dat_pro	Access of data is managed and controlled	Not applicable
Data-driven profile	Dat_drv	Preference to use data to take decision	ICR = .76 CR = .851 AVE = .536
Customer experience (OUTCOME)	Custom	Increase of the Customer experience	ICR = .94 CR = .955 AVE = .843
Provider operations efficiency (OUTCOME)	Prom	Increase of Provider operations efficiency	ICR = 0,91 CR = .947 AVE = .855

$$\text{Calibration} = (0{,}5\ /\ 2n) * \sum_{1\le k\le n} sum(Answers(k)) - 0.25 \quad (1)$$

Where n is the number of questions that aggregate the variable and, Answer (k) the result given by the respondent based on Likert-scale of the statement k.

Table 2 shows the calibration values.

4. Results

We analyse the conditions that lead to an increase or decrease of the customer experience and provider operations efficiency. Then, we present an analysis of necessity and sufficiency conditions.

4.1. Analysis of necessary conditions

Table 3 examines the relationship between the five conditions and the two outcomes: Customer and Provider Management. The abbreviation “**Cons.**” refers to Consistency as introduced by (Ragin and Strand, 2008). This measure represents the degree to which one set is included by another, producing evidence that is either consistent with the underlying hypothesis, inconsistent or mixed. The abbreviation “**Cov.**” refers to necessity Coverage as introduced by (Ragin and Strand, 2008) and allows an assessment of the frequency with which one set occurs relative to another one. High coverage values are synonym of greater empirical relevance (Ragin and Strand, 2008) but Consistency prevails always before Coverage in the analysis of the results.

Table 2
Calibration of the variables.

Variables	Full membership	Cross over	Full Non-membership
	0.9	0.5	0.1
Data consistency	20	12	4
Data completeness	10	6	2
Data usage	10	6	2
Data protection	5	3	1
Data-driven profile	25	15	5
Customer experience (OUTCOME)	20	12	4
Provider operations efficiency (OUTCOME)	15	9	3

The abbreviation (~) is the logical operator to describe the absence of the condition. Software R is an ideal environment within a QCA package (Dusa, 2019) which provides algorithms to calculate easily coverage and consistency parameters and generates the Table 3.

Our study analyses both the presence of the condition (Custom and Prom) and its absence (~Custom and ~Prom). The analysis in Table 3 shows that none of the conditions are either necessary for improving the customer experience/provider operations efficiency (consistency lower than 0.9) or for decreasing them (Ragin and Strand, 2008). Thus, the increase or decrease of these outcomes leads therefore to a combination of different conditions.

4.2. Analysis of sufficiency

We analyse now which aggregation of conditions is sufficient to achieve the outcome (Ragin and Strand, 2008).

4.2.1. QCA method to interpret the data

QCA is a two-step method. First of all, a truth table recompiles all the necessary configurations that lead to the two outcomes (Custom and Prom) or the absence of them (~Custom and ~Prom). The second step consists on inserting additional conditions or exclusions (frequency, accuracy, etc.) to reduce contradictions, i.e. combinations where the truth table reveals cases with the same configurations of conditions but showing divergent outcomes. During this phase, the objective is to reduce the complexity of the table and to come to a consistent explanatory model. In our case, we use the heuristic implemented in the software package released by Dusa (2019). We set as condition of reduction a consistency higher than 0.9 (Ho et al., 2016). The two-step method is followed in the next two sub-sections to analyse respectively the conditions of better Customer experience and better Provider operations efficiency.

4.2.2. Customer experience (Custom)

The model that gives conditions to increase the customer experience (Custom) presents three causal configurations (see Table 4.1). These three patterns show a consistency over 0.9, which is sufficient to produce the outcome (Fiss, 2011). The black circle ● describes the presence of the condition while the white circle ○ its absence. Reading and analysing the whole Table 4.1 reveal the following findings:

- Data protection is present in all the configurations.
- The other conditions are present at least once in each configuration.

The three casual configurations are:

- Configuration 1 (dat_con * dat_pro * ~dat_drv) being (*) the logical operator AND and (~) the logical operator NO or ABSENCE, shows the presence of data with high consistency, high protection all together lead to an increase of the customer experience.
- Configuration 2 shows a different set of combinations or paths (dat_cmp * dat_usg * dat_pro) but reflects also a situation that leads to an increase of the Customer experience.
- Configuration 3 (~dat_con * ~dat_cmp * ~dat_usg * dat_pro * dat_drv) indicates that a high data protection with a high data-driven profile both together increase the Customer experience even if the rest of conditions are absent or ambiguous or negative.

However, the model presented in the Table 4.2 that analyses the reduction of the customer experience shows 2 causal configurations This model shows the diversity of existing paths leading to the outcome (~Custom). The analysis of the Table 4.2 highlights that data completeness, protection and data-driven are pertinent in all the configurations, showing as in the previous model (Table 4.1) that data protection is a relevant condition.

Table 3
Analysis of necessary conditions.

Conditions tested	Custom		~ Custom		Prom		~ Prom	
	Cons.	Cov.	Cons.	Cov.	Cons.	Cov.	Cons.	Cov.
Data consistency	0.72	0.873	0.465	0.532	0.672	0.903	0.451	0.457
~Data consistency	0.614	0.549	0.889	0.750	0.596	0.590	0.905	0.675
Data completeness	0.508	0.848	0.337	0.531	0.504	0.932	0.291	0.405
~ Data completeness	0.719	0.535	0.903	0.634	0.291	0.405	0.952	0.592
Data usage	0.764	0.801	0.529	0.524	0.756	0.880	0.484	0.425
~ Data usage	0.547	0.551	0.800	0.762	0.505	0.565	0.863	0.728
Data protection	0.870	0.712	0.686	0.461	0.951	0.750	0.694	0.413
~ Data protection	0.243	0.450	0.571	0.870	0.256	0.525	0.580	0.9
Data-driven profile	0.573	0.918	0.325	0.491	0.498	0.883	0.310	0.415
~ Data-driven profile	0.682	0.517	0.945	0.677	0.670	0.563	0.913	0.578

Table 4.1
Analysis of sufficiency conditions (Custom).

Conf.	Conditions					Coverage		Consistency
	Data consistency	Data completeness	Data usage	Data protection	Data-driven	Raw	Unique	
1	●			●	○	0.512	0.173	0.949
2		●	●	●		0.481	0.174	0.890
3	○	○	○	●	●	0.112	0.012	0.903

Solution coverage: 0.810. Solution consistency: 0.900.
Frequency threshold = 1. Consistency threshold = 0.9.

Table 4.2
Analysis of sufficiency conditions (~Custom).

Conf.	Conditions					Coverage		Consistency
	Data consistency	Data completeness	Data usage	Data protection	Data-driven	Raw	Unique	
1	○	○		○	○	0.518	0.257	1.0
2		○	●	○	○	0.290	0.029	1.0

Solution coverage: 0.643. Solution consistency: 0.744.
Frequency threshold = 1. Consistency threshold = 0.9.
The black circles indicate the presence of antecedent conditions while the white circles show the absence or negation of antecedent conditions. The blank cells represent ambiguous conditions.

Table 5.1
Analysis of sufficiency conditions (Prom).

Conf.	Conditions					Coverage		Consistency
	Data consistency	Data completeness	Data usage	Data protection	Data-driven	Raw	Unique	
1	●		●	●		0.575	0.087	0.953
2		●	●	●		0.456	0.109	0.933
3	●	○	●		○	0.343	0.031	0.940
4	●	●		●	○	0.257	0.024	1.000

Solution coverage: 0.739. Solution consistency: 0.924.
Frequency threshold = 1. Consistency threshold = 0.9.

4.2.3. Provider operations efficiency (Prom)

The model that gives conditions to increase the provider operations efficiency (Prom) presents four causal configurations (see Table 5.1). These three patterns show a consistency over 0.924, which is sufficient to produce the outcome (Ho et al., 2016). First of all, we highlight that data usage and data protection are two conditions present in three of all configurations.

According to (Ragin and Strand, 2008), we analyse the 3 causal configurations with have the highest unique and raw coverages. They are configurations 1, 2 and 3. Configuration 4 is then discarded of our analysis:

- Configuration 1 (dat_con * dat_usg * dat_pro) and configuration 2 (dat_cmp * dat_usg * dat_pro) show that data needs three of the four conditions to lead to an increase of the Provider efficiency.
- Configuration 3 (dat_con * ~dat_cmp * dat_usg * ~dat_drv) indicates that a high data consistency and data usage improve both together the provider efficiency.

The model that analyses the reduction of the provider operations efficiency (see Table 5.2) shows only one causal configuration (~Prom). This model stresses the data-driven variable combined with three attributes of the data structure is a sufficient condition to reduce provider operations efficiency. It is an important output, especially, if we remind that the previous model did not find that a high value of data-driven condition was sufficient to increase the provider efficiency.

5. Discussion and conclusion

Our paper aims at understanding the configurations associated to

Table 5.2
Analysis of sufficiency conditions (~Prom).

Conf.	Conditions					Coverage		Consistency
	Data consistency	Data completeness	Data usage	Data protection	Data-driven	Raw	Unique	
1	○	○		○	○	0.573	-	0.979

Solution coverage: 0.979. Solution consistency: 0.573.
Frequency threshold = 1. Consistency threshold = 0.9.
The black circles indicate the presence of antecedent conditions while the white circles show the absence or negation of antecedent conditions. The blank cells represent ambiguous conditions.

better performance of a company and we decide to analyse the processes related to customer management (customer experience) and provider management (operations efficiency). We surprisingly find that an analytical competitor (Grimaldi et al., 2019) who is oriented to take decision based on the use of data is not a sufficient condition to improve the customer experience and supplier operations efficiency in four out of five main configurations that have the greater empirical relevance. It shows that the data structure and data quality play the most important role than an organisation culture oriented to the data. Moreover, we observe also that decision makers with lower data-driven profile lead to decrease customer experience and provider operations while this condition is combined also with at least two bad conditions of data quality. We definitely reach a first conclusion here that more than a favourable organization oriented to the take of decisions based on data, Colombian firms need appropriate data to contribute to their performance (Ji-fan Ren et al., 2017).

Data consistency, data usage and data protection are the three more frequent conditions amongst the four solutions analysed to improve the provider efficiency. Data protection is the unique condition 'common' along the three different configurations that lead to increase customer experience. Furthermore, due to data privacy and data protection reasons, data have to be distributed to appropriate users and not to be made available for users without access rights. It becomes a challenge when at the same time the company has to ensure the data availability at all times (Anderson and Moore, 1990). Our findings confirm the relationship between actions of data cleaning and data maturity and the company performance in terms of customer and provider management processes and show that all the paths followed by Colombian firms go through the protection and the privacy of the data as main condition that drives to the performance improvement.

Moreover, we suggest that future studies give a step further and investigate if the improvement of the customer and provider management processes have a direct impact on the financial Key Performance Indicators (KPIs) of a company e.g. the profitability, the market response, the market position value, and the new product success rate. These results could be collected extending the current survey prepared for the companies and the study could have the objective to determine if analytical competitor have better financial results than the laggards to use the data science (Davenport and Harris, 2007).

The conclusion of our paper is different combinations of casual conditions (related to three items of the data maturity) drive to a better customer experience and provider operations efficiency. Surprisingly, data-driven condition is necessary and not sufficient. The results show one solution does not fill all and further studies could improve to understand better the equifinality of the outcomes. This study has some limitation. It measures conditions based on answers reported by professionals with a technological profile (CTO and CIO). To minimize errors and ensure more accurate results, we suggest to include representative of sales, marketing, production and procurement departments and to gather and consolidate all these results to have a 360° view of the company reality.

Declarations

Author contribution statement

Didier Grimaldi, Javier Diaz, Hugo Arboleda, Vicenc Fernandez: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interest statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Appendix

The respondent evaluates each of the following statements in a Likert-scale between 1-5:

First section – Data Maturity adopted from (Butler, 2011).

Block	items	Statements	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Data-driven behaviour	M1	Our company takes into account the data analytics skills in the hiring process					
	M2	A relevant number of our employees have a science or technology academic background					
	M3	Our company delivers training classes related to data analytics & visualization					
Data consistency	M4	A common definition of the data sources (templates, order forms) allows to share data inside the organization					

(*continued on next page*)

(continued)

Block	items	Statements	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
	M5	All data (ordering details, material inventory, etc.) are managed in the same way throughout the organization					
	M6	An automatic method of maintaining data consistency is being used					
	M7	There is no input error in all the data (e.g. information manually entered incorrect, machine calibration outlier, etc.)					
Data completeness	M8	All sources (data) have been inputted by our company with no omissions. New variables are included if required					
	M9	All sources (data) have been inputted by our suppliers and clients with no omissions. New variables are included if required					
	M10	Problems due to incomplete data are hard to be found					
Data Usage	M11	The employees of functional departments (business users) find the source and have access the data they need					
	M12	There is a dialogue between the IT department and the functional departments that permits the perfect exploitation of the data					
Data Protection	M13	Data access (especially sensitive data) is restricted according to the rules defined by the user profile					

Second section – Perceived benefits.

Business Processes	items	Statements	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1. Front-office/ Customer perceived benefits	A1	Our data analytics allow our organization to improve its marketing campaigns					
	A2	Our data analytics allow to improve our brand image					
	A3	Our data analytics allow the organisation to have a 360° vision of our client and develop specific actions by customer segment (Coupons, loyalty program, special offer…)					
	A4	Our data analytics allow to make the best offer for the client in response to market conditions or competition behaviour					
2. Back office/ Provider perceived benefits		Our data analytics allow the organization to					
	A6	a. Cooperate better with suppliers/distributors, predicting demand					
	A7	b. Improve the traceability of the supply chain, manufacturing and logistics operations					
	A8	c. Decrease operation and management costs for material procurement and sales/distribution (for instance: inventory reduction)					
	A9	d. Prevent from Fraud detection and Cyber intelligence					
	A10	e. Prevent from financial risks (increase of materials price, monetary risks, etc.)					
	A11	e. Analyse the risks related to compliance, ethics and corporate responsibility					

Third section: Demographic information.

C0	Company name:				
C1	Position of the respondent in the company				
C2	Email of the person responding to this questionnaire				
C3	Where are the headquarters of the company?				
C4	In how many countries does the company have operations	1. one country 2. between 2 and 5 3. more than 5			
C5	In which sector does the company operates?				
C8	What is the size of the company?	1. Fewer than 10 employees 2. 10 to 49 employees 3. 50 to 249 employees 4. 250 to 499 employees 5. More than 500 employees			
C9	Age of the company (#years)	below 3 years	3–5 years	5–10 years	more than 10 years

References

Addo-Tenkorang, R., Helo, P.T., 2016. Big data applications in operations/supply-chain management: a literature review. Comput. Ind. Eng. 101, 528–543.

Anderson, B.D.O., Moore, J.B., 1990. Optimal Control: Linear Quadratic Methods. Prentice-Hall, Inc, Upper Saddle River, NJ, USA.

Becker, J., Knackstedt, R., Pöppelbuß, J., 2009. Developing maturity models for IT management. Bus. Inf. Syst. Eng. 1 (3), 213–222.

Bonabeau, E., 2003. In: Don't Trust Your Gut. Harvard Business Review, vol. 81. Graduate School of Business Administration, Harvard University. Retrieved from. http://cat.inist.fr/?aModele=afficheN&cpsidt=14791484.

Butler, D., 2011. Oracle Master Data Management Overview. Retrieved from. http://www.oracle.com/us/products/applications/master-data-management/mdm-overview-wp-103258.pdf.

Colombia, U. nacional, 2018. Innovar. Retrieved from. https://revistas.unal.edu.co/index.php/innovar/issue/download/4938/1081.

Davenport, T.H., Harris, J.G., 2007. Competing on Analytics: the New Science of Winning. Harvard Bu).

Davenport, T.H., Prusak, L., 2000. Working Knowledge: How Organizations Manage what They Know. Harvard Business School Press, Boston, MA, USA.

Díaz-Díaz, N.L., de Saá-Pérez, P., 2014. The interaction between external and internal knowledge sources: an open innovation view. J. Knowl. Manag. 18 (2), 430–446.

Dusa, A., 2019. QCA with R. A Comprehensive Resource. Springer International Publishing, New York.

Evans, P., 2013. How Data Will Transform Business available on-line at. http://www.ted.com/talks/philip_evans_how_data_will_transform_business.

Fedesarrollo, 2013. El papel de las tic en el desarrollo de la pequeña empresa: reflexiones de política a la luz del caso colombia. Retrieved from. http://www.fedesarrollo.org.co/wp-content/uploads/2013/12/TIC_diciembre_2013.pdf. (Accessed 31 October 2018).

Fiss, P., 2011. Building better causal theories: a fuzzy set approach to typologies in organization research. Acad. Manag. J. 54.

Fiss, P.C., 2007. A set-theoretic approach to organizational configurations. Acad. Manag. Rev. 32 (4), 1180–1198.

Forbes, 2014. Making Analytics Accountable: 56% of Executives Expect Analytics to Contribute to 10% or More Growth in 2014. Retrieved from. https://www.forbes.com/sites/louiscolumbus/2014/12/10/making-analytics-accountable-56-of-executives-expect-analytics-to-contribute-to-10-or-more-growth-in-2014/.

Forza, C., Salvador, F., 2001. Information flows for high-performance manufacturing. Int. J. Prod. Econ. 70 (1), 21–36.

Gobble, M.M., 2013. Big data: the next big thing in innovation. Res. Technol. Manag. 56 (1), 64–67.

Grimaldi, D., Fernandez, V., Carrasco, C., 2019. Exploring data conditions to improve business performance. J. Oper. Res. Soc.

Gunasekaran, A., Ngai, E.W., 2004. Information sytems in supply chain integration and management. Eur. J. Oper. Res. 159, 269–295.

Harris, J.G., Craig, E., 2011. Developing analytical leadership. Strateg. HR Rev. 11 (1), 25–30.

Ho, J., Plewa, C., Lu, V.N., 2016. Examining strategic orientation complementarity using multiple regression analysis and fuzzy set QCA. J. Bus. Res. 69 (6), 2199–2205.

Hofmann, E., 2017. Big data and supply chain decisions: the impact of volume, variety and velocity properties on the bullwhip effect. Int. J. Prod. Res. 55 (17), 5108–5126.

Ji-fan Ren, S., Fosso Wamba, S., Akter, S., Dubey, R., Childe, S.J., 2017. Modelling quality dynamics, business value and firm performance in a big data analytics environment. Int. J. Prod. Res. 55 (17), 5011–5026.

Lavalle, S., Lesser, E., Shockley, R., Hopkins, M.S., Kruschwitz, N., 2011. Big data, analytics and the path from insights to value. MIT Sloan Manag. Rev. 52 (2), 21–32.

Mata, F.J., Fuerst, W.L., Barney, J.B., 1995. Information Technology and sustained competitive advantage: a resource-based analysis. MIS Q. 19 (4), 487–505.

Maza, M.C., Fedriani, E.M., Antonio, J., Sanz, O., 2016. Fatores relevantes para otimizar os servços públicos de apoio a empreendedores e a taxa de sobrevivência das empresas. INNOVAR, Contabilidad e Finanzas, pp. 9–24.

McAfee, A., Brynjolfsson, E., 2012. Big data's management revolution. Harward Business Review.

Ministerio de Tecnologías de la Información y las Comunicaciones, 2010. El Plan Vive Digital: Diagnóstico y Punto de Partida de Colombia. Retrieved from. http://www.mintic.gov.co/portal/vivedigital/612/w3-article-1518.html. (Accessed 31 October 2018).

Ministerio de Tecnologías de la Información y las Comunicaciones, 2016. Iniciativas Plan Vive Digital.

Mintzberg, H., 1978. Patterns in strategy formation. Manag. Sci. 24 (9), 934–948. Retrieved from. https://econpapers.repec.org/RePEc:inm:ormnsc:v:24:y:1978:i:9:p:934-948.

Porter, M.E., 1998. The Adam Smith address : location, clusters, and the "new" microeconomics of competition. Bus. Econ. 33 (1), 7-.

Ragin, C.C., Strand, S.I., 2008. Using qualitative comparative analysis to study causal order: comment on caren and Panofsky (2005). Sociol. Methods Res. 36 (4), 431–441.

Randall, R.M., 2015. W. Chan Kim and renee mauborgne dispel blue ocean myths. Strateg. Leadersh. 43 (2), 11–14.

Rojas, L.A.R., Bermúdez, G.M.T., Lovelle, J.M.C., 2014. Open data and big data: a perspective from Colombia. In: Lecture Notes in Business Information Processing, vol. 185. Springer Verlag, pp. 35–41. LNBIP.

Spruit, M., Pietzka, K., 2014. MD3M: the master data management maturity model. Comput. Hum. Behav.

Tan, K.H., Ji, G., Lim, C.P., Tseng, M.-L., 2017. Using big data to make better decisions in the digital economy. Int. J. Prod. Res. 55 (17), 4998–5000.

Wegener, H., 2008. Metadaten, referenzdaten, stammdaten. In: Integrierte Informationslogistik. Springer, pp. 189–209.

Woerner, S.L., Wixom, B.H., 2015. Big data: extending the business strategy toolbox. J. Inf. Technol. 30 (1), 60–62.

World Bank, 2018. World Bank Forecasts for Colombia. Retrieved from. https://openknowledge.worldbank.org/bitstream/handle/10986/29801/9781464812576.pdf. Retrieved 11 Sept 2018.