Masterarbeit

From   Lucía Murlanch Dosset

                   Matrikelnummer: 400112

Topic: Development of a business model for the utilization of customer data in supplier development

Information about customer requirements is important for manufacturing companies. Knowing these requirements can improve the products offered thus obtaining competitive advantages. Therefore, manufacturing companies try to gather customer information from numerous sources. Due to the variety of format of such data, obtaining useful knowledge is a complicated process that has not been standardized to date.

Nowadays, manufacturing companies generally no longer produce parts themselves, but assemble the parts they order from their suppliers. For this reason, suppliers have freedom in the design of the pieces, giving importance to the obtained customer requirements. Thus, it is possible to generate benefits from the transfer of the obtained knowledge from the assembling companies to the suppliers, encouraging the extraction of information from the available data. In addition, it is required to determine which kind of information is exchanged between manufacturers and suppliers, thus getting a possible lack of information.

The aim of this work is to develop a procedure of obtaining information from customer data. By this way, a structured methodology will be established in companies in order to automatize the important process of obtaining customer information. First, a reference methodology of data analytics is selected. In addition, the most commonly used sources of information are selected and analysed. Once the different sources of customer information, the respective information and references methodologies are gathered, the new methodology will be created. In order to demonstrate the functionality of it, an exemplary customer information will be analysed in the framework, showing each phase of it and the required treatment of data. On the suppliers' part, it is created a procedure to generate benefice from customer extracted knowledge.

In detail, the following subtasks are to be solved:

- Familiarization with the topics information exchanging between suppliers and manu-facturers in both general industry and automotive industry, IT industry

- Selection and analysis of useful sources of customer's information

- Determine the information exchanged between suppliers and manufacturers to know possible white spots

- Selection of a frame methodology of Data analytics

- Selection and development of the phases of the new methodology

- Validation of the new methodology

Prof. Dr.-Ing. Robert Schmitt

I   Index

II Index of figures

III Index of Tables

1 Introduction

1.1 Motivation

The automotive industry is a sector of great potential and competitiveness, which affects the global economy to a large extent. As a result, this sector has evolved rapidly in recent decades, causing a change of roles in manufacturing companies and suppliers. Such has been this change and evolution that, at present, there is a vacuum in the process of analysis and exchange of information with suppliers. Thus, this gap creates a serious problem, since the continuous improvement of the vehicles and the processes involved cannot be based on the information obtained from the consumer and increase their satisfaction. In this way, it is this vacuum in the process from which the need for research in this field arises.

1.2 Objective

The development of this project aims to generate a methodology that provides a basis for the analysis and transfer of information to suppliers.

In this way, it is intended to cover all the necessary phases from the beginning of the process until the information reaches the hands of the supplier, so that he can analyze it and take advantage of it to improve its production process.

1.3 Structure of the work

The work has three main parts; literature review, problem statement and methodology development and validation.

In the literature review is intended to provide a basis of knowledge necessary for the development and understanding of the project. It could be considered as a scientific framework of the project. The first point of this section explains the evolution of the automotive industry, naming the most remarkable changes in production. The second section deals with the strategies of manufacturing companies, with the aim of introducing the strategy of modularisation, which is very relevant for the analysis of supplier information. The following section is devoted to the evolution of suppliers, highlighting both the relations between them and the methods of communication between the two parties. The following point presents an initial classification of the methods for obtaining consumer information, considering primary, secondary research and internal sources. The consumer's behaviour is introduced in the following section, explaining in turn the factors that affect it and the market segmentation strategy, a strategy based on the consumer's behaviour. The literature review ends with an introduction to data analysis.

The second part of the project, problem relevant method, begins with the information exchanged between supplier and manufacturing companies, with the introduction to the problem of the information gap relating to consumer needs. This point continues with the sources of information most used by manufacturing firms. The study of consumer behaviour is the next topic dealt with in this section. Next, the main tools of data analysis are presented, and this second block is concluded with a small internal conclusion, in which some issues to be solved with the development of the methodology are defined.

The last block is dedicated to the development of the methodology and its validation. In this last section, a brief summary of the project and a small critical review of it have also been included.

2 Scientific foundations

2.1 Evolution of the automotive industry

The automotive industry is considered as a dynamic industry, that has changed during the last few decades. In general terms, the evolution of the automotive industry has two main points, which are a turning point in the production forms. First, at the beginning of the last century, the company Ford revolutionized the car manufacturing process with the use of the assembly line. This innovation lead to massive production process, leaving the hand-made process behind, to the increase of the production volumes and to the reduction of the unit production costs[1]. Use the "Insert Citation" button to add citations to this document.

This production system was characterized by standardised products, enough stock to solve variety demand, the application of quality control after the whole production process and to workers specialized in one task.

The second important point was after the Second World War, when the Toyota company redesigned the production process of vehicles and its parts. This innovation is known as Lean Production or Toyota Production System, and requires working teams with different types of qualification, so that each worker can develop different tasks in the production, supervision and quality control areas. In addition, it differs to the last system in the importance of the quality control, since it emphasizes in the prevention of total defects in order to eliminate costs. Until then, automobile companies were limited to the regional market, i.e. each company sold its vehicles in their countries and surroundings. With this revolution in production, Toyota managed to enter the U.S. market, causing American companies to change their strategies in order to compete, some of them will be shown in the following sections.

Throughout this evolution, the functions of the automotive companies have changed. In the beginning, vehicle production was carried out mostly in the same plant. Most of the car parts were manufactured in that plant, where they were later assembled to obtain the final car. However, with the technological development and specialization in the sector, the parts of the cars increasingly have reached a higher degree of perfection and accuracy. For this reason, during the last decade, the automotive sector has undergone a profound transformation that has led the large assembly companies to directly contribute

---

[1] **Carbajal, Y**. (Sector automotriz: reestructuración tecnológica y reconfiguración del mercado mundial), 2010. Page 26.

less and less added value to the manufacturing process of motor vehicles[2]. By this way, the manufacturing plants have been diverting the production works towards the suppliers, with a greater amount of added value. Nowadays, automotive companies are responsible for assembling the blocks of parts, known as modules, forming the final car, without participating in the production of any of its parts.

2.2 Strategies of assemblers

The changes in the production systems mentioned in the previous section reflect the dynamism that has characterized the automotive company since its inception. However, in recent years it has been characterized by a profound transformation that has led to a dynamic of high competitiveness, the main characteristic of which is that the large assembly companies contribute less and less added value to the manufacturing process of motor vehicles, and that more and more responsibility is shared in all respects with suppliers; for example, while in 1995 they contributed 40%, today they do so with 25%. Therefore, automotive companies have developed different strategies to become more competitive.

The first of the strategies consists of an adoption of the global perspective in their operations. As it has been said before, few decades ago, competition between car assemblers occurred within regions; North American companies dominate the United States market, the Japanese dominate the Asian market and the Europeans dominate their own regional market. After the entrance of Toyota in the American market, this structure changed completely. The rest of the companies changed their competitive strategies, opening new plants and relocating some in different foreign countries, leading to a greater presence of foreign competition in almost all parts of the world. As a result, assemblers plan operations with a global scale. With new investments, companies intend to repeat their supply chain, requesting suppliers to be present in the new regions they are relocated. However, this global perspective is slightly particularized in each region, adapting the models to certain social and governmental factors. For example, classified in the highest price ranges and market niches are prioritized in the markets of developed countries. Government regulations mainly affect safety factors to which cars must adapt. Thus, the technological factors of each region imply the incorporation of different electronic devices, since in some areas they can be considered basic while in other places they can remain unknown and be considered differentiating factors.

---

[2] **Jiménez, J**. (Un análisis del sector automotriz y su modelo de gestión en el suministro de las autopartes), 2006, page 13

2 Scientific foundations

The second strategy deals with the acquisition of capabilities and the company growth. However, within these strategies, companies followed strategies diametrically opposed. Some of the companies sought to acquire capabilities through alliances, mergers and acquisitions with other companies, concentrating horizontally the industry. This kind of alliances or mergers must overcome both technical aspects and personnel aspects, considering that all the workers must learn to interact with different cultural backgrounds[3]. In this way, for regroupings were significant for the global automotive industry in the nineties: the merger of Daimler-Chrysler and its entry into the capital of Mitsubishi; the takeover of Nissan, Dacia and Samsung by Renault; Ford's control over Mazda and its acquisition of Volvo; and, finally, General Motors' alliance with Fiat[4]. In the opposite, some companies sought development through internal growth strategies even though they could have taken control of other assemblers, since they had enough resources to that. That is the case of Toyota, Honda and Volkswagen.

On the other hand, automotive assemblers organized vehicle production around platforms. That organization merged because of the decline of the sales, the shortening of the life-cycle of the product and the impossibility of achieving economies of scale in design and manufacturing. In the platforms, multiple models can be designed, developed and built using a system of common parts and modules with related components; that is, they are considered fixed points or common systems that allow different vehicles to be built on the same manufacturing platform. Thus, factories are much more flexible in production and can easily change the model depending on market conditions and circumstances, labour relations or a certain strategic objective of the company[5]. This strategy involved a reduction of costs, since by using these platforms a wide range of products could be produced for a multitude of tastes and preferences, thus ensuring sufficient differentiation to neutralize the proliferation of competing brands and maintain an efficient scale of production.

The fourth strategy includes the operations and efforts that companies have developed in order to increase customer satisfaction after the car sale. Since assemblers have reduced their manufacturing operations, they have focussed their resources on getting better sales assistance, after-sales services and maintenance of the vehicle. Thus, automotive companies have created different financing system, among others measures.

---

[3] **Hahn, D., Hartley, J.,** (Supply-Chain Synchronization: Lessons from Hyundai Motor Company), 2000

[4] **Álvarez, M. L.,** (Cambios en la industria automotriz frente a la globalización: el sector de autopartes en México), 2002. Page 30-32

[5] **Jiménez, J**. (Un análisis del sector automotriz y su modelo de gestión en el suministro de las autopartes), 2006, page 27

The modifications included in these strategies sought to reduce asset intensity to increase share value while improving responsiveness and quality.

The last strategy is one of the most relevant point of the automotive industry, since it means a new model of production, identified as "modular". The establishment of this new model appeared because of the changes in the production system. This new model consisted in dividing the car into several modules, and each supplier is responsible for designing, innovating, producing and placing its module on the assembly line.[6] Given the importance of this last strategy for the development of suppliers, and the objective of this work, it will be further explored in the following section.

2.2.1 **Modularization**

The concept of modularization is a broad concept that varies according to the region in which it is applied. Thus, there are three main types of modularization;

- Modularization in product architecture (modularisation in design).
- Modularization in production
- Modularization in inter-firm systems

The latter one, modularization in inter-firm systems, is developed in the European and American automotive industry, while the second one, modularization in production, is carried out in Asian countries, such as Japan. None of the large automotive companies involves in its process the first type of modularization, so it will not be considered. On the other hand, considering that this work is carried out in Europe, modularization in inter-firm systems will be studied in greater depth.

Modularization in inter-firm systems consists of the externalization of subsystems in large amounts of units. The ratio of outside production can vary from a low degree, in which small modules are supplied by different suppliers, and they are incorporated in later modules, before being introduced in the car. In the opposite, modularization in European assemblers is based on a highly modular supply system, in which large modules are assembled by outside suppliers on their subassembly lines and are delivered and assembled into finished products on the main line of the automaker[7]. Thus, the modules

---

[6] **Álvarez, M. L.,** (Cambios en la industria automotriz frente a la globalización: el sector de autopartes en México), 2002. Page 30-32

[7] **Takeishi, A., Fujimoto, T. ,** (Modularisation in the auto industry: interlinked multiple hierarchies of product, production and supplier systems), 2001. Page 384

2 Scientific foundations

are completely assembled in the reception area of the final assembly plants, ready to be incorporated into the vehicles.

The main feature of modularization is the reduction of the scope of internal operations, deriving responsibilities to suppliers and thus freeing company resources to invest in customer satisfaction. This diversification of responsibilities has led to the development of suppliers, increasingly consolidated and with greater potential. In addition, modularisation implies the consolidation of the supplier network, even the relocation of supply plants. Since larger modules are more difficult and expensive to move, the adoption of the modular assembly process is linked to the establishment of supplier plants next to assembler plants. Thus, the delivery of parts is synchronized with the demand and adjusts to the sequence of the cars in the assembly line. For this reason, automobile companies choose a supplier, after examining and comparing it, to provide the corresponding module for a longer period, as will be explained in later sections.

Another characteristic of the modularization system deals with the customer satisfaction. Modules can be combined in different versions, reaching different versions of a car model and thus covering a wider range of consumers.

In the following diagram, the main modules with their parts are shown:

PART	MODULE	SYSTEM
Fabric Foam Seat manufacture	Seat	
Indoor Ceiling Interior Panels Terminations	Indoor	
Instrument panels Meters Gearshift lever Handlebar	Conductor module	**Interior system**
Motor Control Sensors Cables, spark plugs, distributor Alternator	Ignition	
Suspension/transmisión mechanisms Cable harness Anti-lock brakes	Chassis Electronics	

Sound Lights Air conditioning Navigation Air Bag	Interior electronics	**Electrical and electronic system**
Engine Axles Transmission	Motor Train	**Chassis systems**
Suspension Brakes Wheels and tires Shock absorbers Terminations Radiator Ventilator Lights	Mobile chassis Front and rear corner modules	

Table 1. Strategy of modularization

2.3 Suppliers evolution

The evolution of the automotive industry and the transformation of automotive companies from manufacturers to assemblers has led to drastic changes in all supplier aspects. As a result, the supply chain has become complex and extensive, and both the roles of suppliers and the relationships between suppliers and assemblers have undergone considerable changes, as presented in the following sections.

2.3.1 Supply chain

The dynamism and evolution of the automotive industry has led to the evolution of the supply chain, making it extensive and complex. The size of the industry is quite large and diversified, and above all, the relationships between customers and suppliers are intensely complex, which complicates the analysis of the supply chain. As an example, the supply chain of Hyundai has approximately 400 direct suppliers, reaching more than 2500 suppliers in the second rank and an unknown number of third rank suppliers. In this way, the supply chain can be identified as a network of companies involved in a sequence of production and services, from the supply of raw materials to the delivery of the final product, considering it as the vehicle. Thus, the network is defined by the distance between the supplier and the final assembly (vertical dimension), by the number of suppliers and customer at each level (horizontal dimension) and by the position of a company in the chain.

The vertical dimension refers to the number of levels in the supply chain, depending on the complexity of the supplied module. In some cases, the supply chain for some of the modules is excessively large, since the initial fabricants are quite far from the final assembly. Nowadays, automotive assemblers try their supply chain to be the shortest as

2 Scientific foundations

possible, even they induce their direct suppliers to locate their plants as much near as possible, as said before.

The horizontal dimension is formed by the number of suppliers or customers at each level, or tier, which depends on the technological level of the companies or type of product they manufacture. In the automotive sector, assembly companies prefer to deal with as few suppliers as possible at the highest levels by implementing the strategies mentioned in the previous section. This makes the assemblers the central company, and the most influential link in the chain.

The following is a diagram for visualizing the supply chain:

Horizontal dimension: Number of suppliers of the same element (Supplier A, Supplier B)

Supplier A.1 - Supplier B.1

Supplier A.2 - Supplier B.2

Manufacturing company

Vertical dimension: Distance between first supplier and market (Supplier 1, Supplier 2)
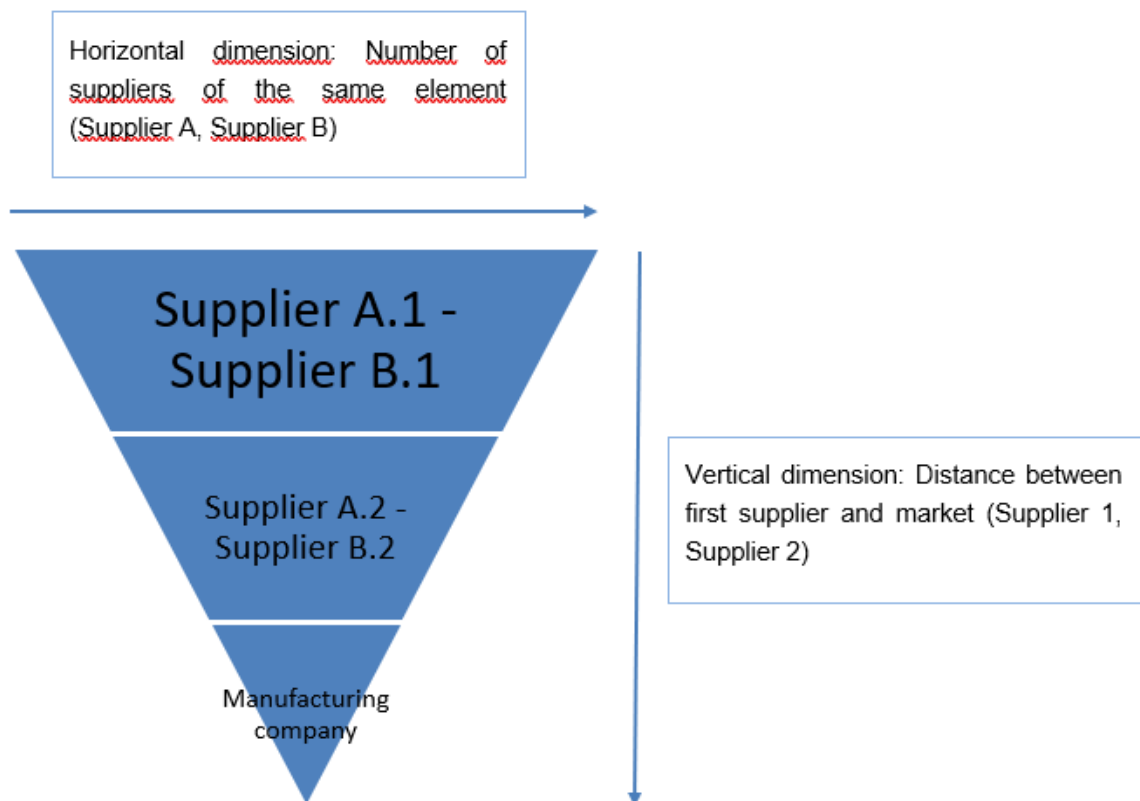
Figure 1. Supply chain

The distance of a supplier from the final assembler is determined from the location in the chain with respect to the main market. In that sense, assemblers form the central company of the chain, and for this study, they will form the end of the production chain of vehicles.

Apart from that, the recognition of the member of the supply chain is very important to facilitate the flow of information and, by this way, the improvement and the correct production system. The members of the chain are identified by their role and participation in the production process. Besides, to classify different members, the way they interact with car assemblers is analysed, which can be directly or indirectly.

2.3.2   **Relationship between suppliers and assemblers**

The relationship between suppliers and assemblers has also undergone an evolution. The increase in the dependence of the assemblers on suppliers in order to successfully achieve their production has led to a strengthening of the relations between them, most of which have been established by contract.

At the beginning of the industry, assemblers used to contact some suppliers in order to obtain some parts of the future vehicle. So that, they bought the required quantity of the desired product and, when the purchase was done, the relation between both was finished. However, the required increase in quality and the new zero-failure strategies led to an increase in the necessary quality controls, before to the choice of suppliers. Thus, the agreements between companies were established according to price, quality and delivery time. By this way, the relationships between suppliers and assemblers were predominant economic-commercial relations. In the face of the new forms of productive and territorial restructuring, generated after the relocation of suppliers and after the transfer of functions from the assemblers to the suppliers, it was necessary to add other technological, organizational and cultural conditioning factors. The relation between suppliers and assemblers became inter-organizational relations, since a lot of them share technological investments and cooperate to overcome competitive pressure and market changes.

In general terms, it is difficult to describe relationships between assemblers and Tier1-suppliers, because they vary according to the size of the company, the available resources and the product exchange. However, in general terms, most of them have some common characteristics, included in the next table:

Characteristic	Description
Long term relationship	Contracts for around 5 years in order to create a stable and efficient relationship.
Multidivisional personnel	Supplier personnel multidivisional
Quality requirements	Annual evaluations to suppliers
Just-in-time strategies	Production on exact date, quantity and quality

Table 2. Characteristics of supplier-manufacturer relationships

2 Scientific foundations

In short, automotive industry suppliers have a relationship with their customers based on quality, just-in-time, production and logistics strategies using information systems, where there is an active policy of creating and developing new products to address the technical aspects of market evolution and result in competitiveness and a long-term customer relationship. Quality, productivity and learning capacity are the formula for success in highly competitive markets such as the automotive industry.

Apart from that, in addition to the dependence and the quality, there are many reasons that explain the cooperation between assemblers and suppliers, such as resource complementation, especially in technology transfer, access to innovations and knowledge, penetration into new markets, the joining of competencies and political factors. Besides, the fact of establishing long-term relationships allows the division of risks and costs, also reducing the responsibilities of the assemblers.

2.3.3 Methods of communication

Communication between suppliers and manufacturing companies, with an effective information exchange, leads to reduction of errors and to increase efficiency of the work processes, even more so in the case of an exchange by technological means. When one company uses the information of other companies in the supply chain, the negative effects of uncertainty can be mitigated in theory[8]. In practice, however, the exchange of information between companies is not as easy as it seems, and it is not as frequent and free-flowing as it should be. In the next diagram, the most common methods of communication are shown[9].

---

[8] **S. Boyson** et al. Transportation Research Part E 39 (2003) p.175–192

[9] **Maximo, R**. (Tres razones por las que implementar un portal de proveedores). 2014

Instantaneous

- Standing out: telephone calls, e-mails
- Fast communication and not need to fix a date previously
- Problems: lack of information, misinterpreted data.

Meetings

- During the establishment of the relationship: representatives of both parts meet in order to set up coditions of the relationship (deadlines, quality, guarantees).
- Long- term relationship: frequently visits from assemblers to suppliers to support them, update conditions of the relationships, exchange possible problems experimented, among others.
- These meetings used to be regularly, even routinely, confirming the evolution from a traditional relationship to a more associative system.

Supplier portal

- New method of communication
- A supplier portal is a software solution that streamlines communication and management between a company and those who provide services, furniture, raw materials or other items necessary for their daily operations.
- Transparency: Continuous available access and visibility over the internal information (ERM system).
- Increment of the amount of data that can be exchanged.
- Due to its relevance to this work, its main characteristics are shown below.

The most basic supplier portals allow the exchange of information of an order, related to the number of products, dates and prices. In some cases, it includes quality characteristics and some details about the product. Some newer supplier portals, in addition to providing an efficient way to place orders or receive complaints from suppliers, allows them to check the status of their invoices, due dates and payments, digitally download vouchers and discount invoices. By this way, a supplier portal opens a communication and self-management channel that drastically reduces operating costs and allows both suppliers and manufactures to focus on higher value-added activities. In addition, a supplier portal integrates suppliers into the manufacture company process, giving them a common space in which both share part of their daily operations and strengthening the relationship. Finally, there are a few ones that allow suppliers to enter invoices, to check the status of procedures and to make claims with the sole help of the platform.

2.4 Customer information

Obtaining relevant information regarding customer requirements is a great challenge for the improvement of the product, and by this way, the development and continuous improvement of a company. In addition, if this information reaches who produce the product, in the case of this study the producers would be the suppliers, they will also be able to improve those aspects. This challenge is because the customer rarely if ever formulates his requirements sufficiently precisely and unambiguously, but the non-fulfilment of them can result in massive loss of customer satisfaction and trust. For this reason, almost companies, not only automotive companies, dedicate a great part of its investments in getting information of its clients to improve their products.

The process of obtaining information from the consumer can be divided into primary research and secondary research; although in a general way, the information obtained in the second one is used as a starting point for primary research.

2.4.1 Primary research

Primary research is defined as new research specially designed and compiled for the purposes of a current research problem[10]. In this case, it would be the creation of a consumer research. In addition, the aim of a primary research is to obtain original data, since the information is collected directly from a source. Within this group, there are two categories of research; qualitative techniques and quantitative techniques. The division reflects the purpose and orientation of the research and the orientation. If the purpose of the division is to generate new ideas, for example to position or reposition a product, a qualitative study is suggested. On the contrary, if descriptive and quantitative information is required, then it is best to carry out a quantitative study.

2.4.2 Secondary research (field and market)

In this research, secondary data is analysed and evaluated under the specific aspects of the problem, understanding secondary data as existing information that was originally collected for research purposes. This method is cheaper and fast than primary methods, since it is usually started. It serves to form an understanding of the research topic and to economise the survey work[11]. In addition to forming an initial picture of the situation, the motivation for collecting and analysing existing information is that unnecessary expenditure and resources can be avoided. This means that if the secondary information is useful

---

[10] **Schiffman, L.** (Comportamiento del consumidor), 2010, page 24

[11] **Kotler, P., et al.** (Principles of Marketing), 2018, p. 132

and provides enough information about the consumer, the primary research is simplified and may even be eliminated.

Within secondary survey three kind of sources are especially important, shown in the following image:

Field data

- Information of the real usage: feedback information (5)
- It allows the company to redesign its products in customer-oriented way
- It is transmitted to the company via customers, service and automatic recording devices.

Market data

- The market in which the product will be sold.
- Statics for industry standard sales figures, market shares of competitors, external databases and studies

Social media and online commerce

- All customer opinions available in Internet.
- Remarkable importance: higher repercussion.
- Customer participates in innovation process by expressing their positive and negative opinions online. (6)

Figure 2. Kind of sources of secondary survey

2.4.3   Internal company sources

This information consists of data previously collected within the organization, which were originally used for other purposes. They are usually obtained from customer service calls, letters with questions for consumers or warranty services, or even from previous investigations that have already been filed.[12] This group includes every document generated inside the company, which provide sales, statistics and economic information regarding

---

[12] **Schiffman, L.** (Comportamiento del consumidor), 2010, page 26

to the product involved. Documents of the cost accounting, general statistics regarding sales, or customer statistics are some examples. In addition, internal audits are a source of information about weak points, possible improvements and required changes in the product. Especially, in the case of product audits*.

Increasingly, companies are using internal secondary data to calculate customer lifetime value profiles for various customer segments. Such profiles include; the costs of customer acquisition (the resources needed to establish a relationship with the customer), the profits generated from individual sales to each customer, the cost of handling customers and their orders (some may place more complex and variable orders that cost more to serve), and the expected duration of that relationship[13]. In this way, the company also archives current consumer information that can be used for further segmentation.

2.5 Customer behaviour

A final or private consumer is one who acquires a product or service that he considers he needs, in order to give it a particular use. On the other hand, industrial consumers are the factories or industries that produce goods and services that are used to manufacture other goods and institutional consumers are those bodies (schools, museums, town halls, etc.) that buy goods to use them and carry out their activities.

Consumer behaviour refers to the set of activities carried out by a person or organisation from the time they have a need until the time they make the purchase and then use the product. The study of such behaviour also includes the analysis of all the factors that influence the actions carried out. The activities of searching for, buying, using and disposing of goods to satisfy needs and desires include mental and emotional processes as well as physical actions.

The study of final consumer behaviour includes the following aspects:

---

[13] **Schiffman, L.** (Comportamiento del consumidor), 2010, page 26

- Buying or acquiring behaviour. It covers all the activities carried out to obtain a good or service, including payment for it.

- The behaviour of final use or consumption of the goods and services acquired by oneself or by third parties.

- The factors internal and external to the individual that influence the purchasing process and the use or consumption of the purchased product.

Thus, consumer behaviour can be analysed with three different approaches:

- The economic approach. This approach is based on the concept of the "economic man", who directs his behaviour towards maximising his usefulness. It is considered that the behaviour of the final consumer always implies a choice, based mainly on economic reasons. This approach, which does not follow the purchase process, bases on the limitation of the resources that each consumer have, the limitation of his needs and the variety of offers to cover consumer satisfaction.

- The psychological approach. Considers primarily the human being as a social animal adapted to the rules and forms the subcultures and groupings to which life is subject. Their desires and behaviour are mainly shaped by their current affiliations to different groups and the influence of them. It includes a field of variables that affect consumer behaviour larger than the previous approach, since, in addition to economic factors, it analyses psychological factors, which are internal factors, and social factors, which are external factors. Psychological variables collect the internal characteristics of the person, their needs and desires; external variables specify the influence exerted by the environment.

- The motivational approach. It intends to explain behaviours from the causes that produce them. According to this approach, the human being acts stimulated by needs. These can be defined as sensations of lack of something, which predispose to act in such a way that they can be alleviated. Thus, motivation is considered the driving force that pushes people into action, and this driving force is caused by a state of tension

due to an unsatisfied need. Within this scope of needs, Maslow established a hierarchy of needs, represented in the figure below.
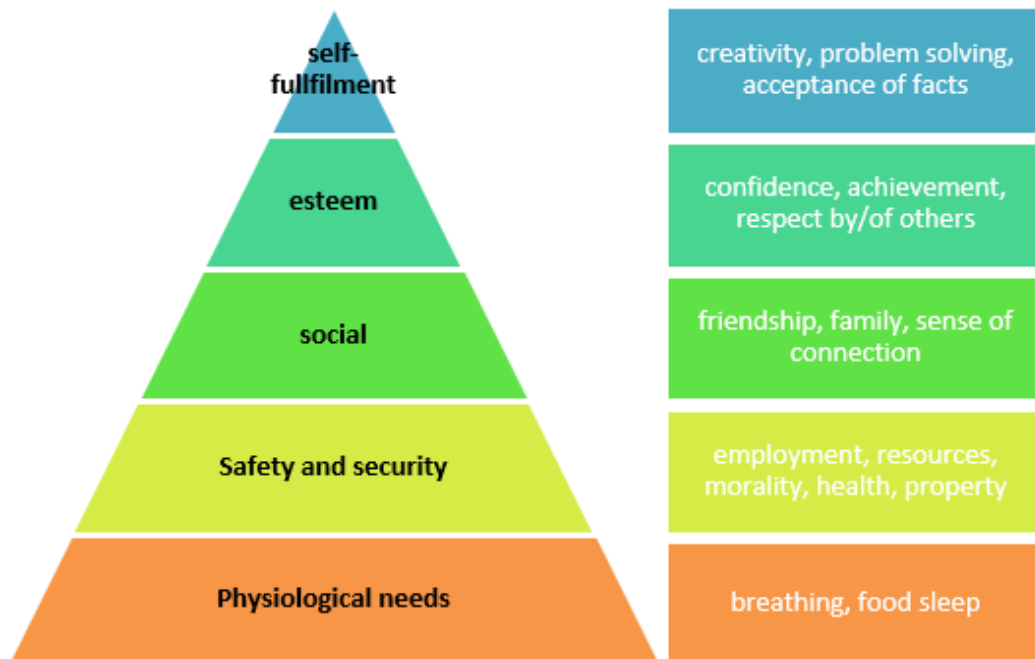
Figure 3. Maslow phyramide

Thus, he asserts that every human being has five basic levels of needs (physiological, safety, social, esteem and self-fulfilment) ranging from low-level or biogenic to high-level or psychogenic. This hierarchy implies the correlative satisfaction of needs from the lowest to the highest level. Thus, it can be considered that the human being moves by unsatisfied needs; when a need is satisfied, it ceases to be a need and loses importance in the dynamics of the human being. Those needs that are permanently gratified cease to exist as active determinants of human behaviour. On the other hand, Freud tries to explain people's behaviour based on his psychoanalytic theory. This theory affirms that people's personality is shaped throughout their lives and develops in three major areas or levels; the subconscious, where the strongest motivations are generated, the conscious, where the actions of the individual are generated to respond to the impulses of the subconscious and the self, which tries to direct impulses and actions towards socially admitted behaviours. In its application to the explanation of consumer behaviour, Freud's model makes it clear that buyers are not only influenced by economic variables, but also by the symbolic aspects of the product. In this way, the importance of product design and the symbols with which it is associated in order to influence consumer behaviour can be explained. In general, it should be borne in mind that human behaviour is very complex and should not be treated from a partial perspective but needs an integrative approach (internal factors, external factors, decision making process and consumer response). Motivational research aims to discover what

he real motivations of human actions are, knowing that human beings are often not fully aware of the reasons that drive our behaviours, our purchasing behaviours.

These three approaches try to analyse and explain the purchase reasons or factors that affect the process of purchasing a product. This process can be divided into four consecutive phases:[14]

Recognition of the need
- Recognition of the problem or need that must be satisfied
- Physiological needs (internal stimulus) and the rest of them (external stimulus).

Information search (17)
- Passive way (being receptive to advertising) or active way (consulting friends, professionals or websites).
- Intensity of the searchs depends on: complexity of the purchase (house / loaf of bread), experience of the costumer and type of purchase

Evaluation of alternatives
- Balance of benefits of different brands

Purchase decision
- Buy one of the evaluated alternatives, look for an additional one or postpone the purchase
- Postpone: more information, gather money.

Figure 4. Phases in the purchase process

These four phases are not always followed by the consumer. For example, in impulse buying the first three phases are not carried out. However, in the field of vehicle buying,

---

[14] **University of Jaen**, (El Comportamiento Del Consumidor Y De Las Organizaciones). 2002. Page 15

consumers follow the purchase process at each of its stages, as it is considered an important product and usually takes a long decision period.

It should be noted that post-purchase behaviour is very important for manufacturing companies. This is because if the consumer has been satisfied with the product, it will most likely repeat the purchase or repeat with the company on its next purchase. Thus, if you publish your opinion, it will be positive and the brand image will be positive for the rest of potential buyers

2.5.1   Factors affecting consumer behaviour

Factors affecting consumer behaviour are divided into two main groups: internal factors and external factors. The first group includes those variables of the consumer's environment that, without depending on the consumer, exert an influence on his behaviour (external). Economic, political, cultural, technological, familiar, and social are some of these variables. The second group includes all variables related to personal characteristics and consumer experience. Motivation, learning and age are some of the examples of these factors.

Factors affecting the consumer behaviour are classified in the next diagram in order to provide a visual idea about what will be explained in the following sections.

external	**internal**
•Cultures and subcultures •Social class •Social factors •Others	•Motivation •Learning •Perception •Behavours •Age and state in the life cycle •Education, occupation and income •Life style •Personality
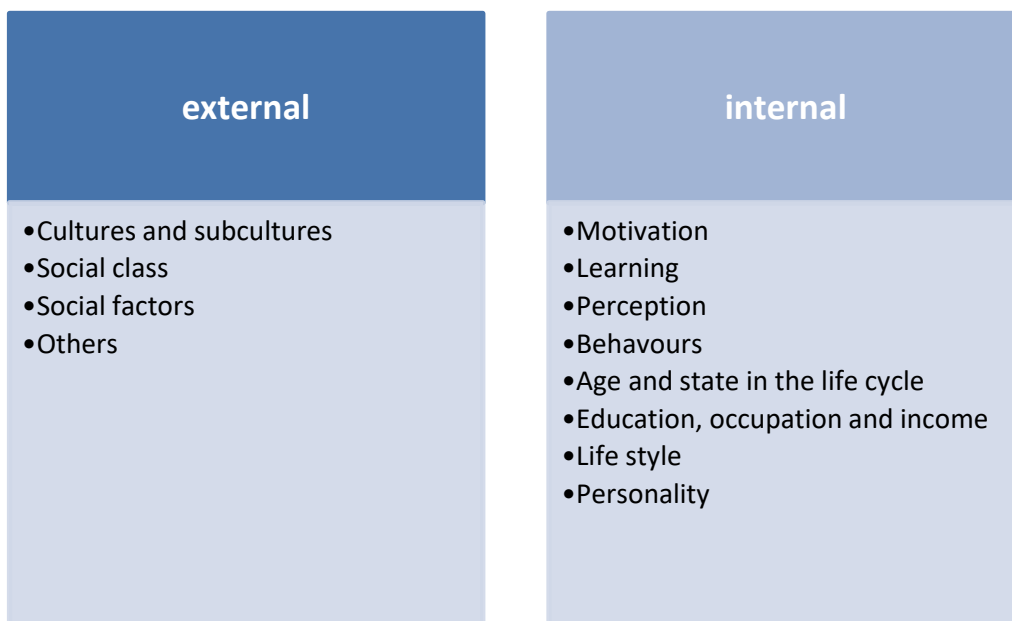
Figure 5. Factors affecting customer behaviour

2.5.1.1 External factors

Cultures and subcultures

Culture is the set of norms, beliefs and customs that are learned by society and lead to common patterns of behaviour. Each culture includes subcultures, or smaller cultures, identified by groups that share the same values. These subcultures can be delimitated by geographic, religious, or ethnic factors.

In addition, the culture is one of the most decisive factors in the human behaviour, since it underlies in the human decision-making process and it determines the life form of each society.

Moreover, human values are significant elements for the compression of the culture of a society. Consumer will act in their behaviours and conduct so that they can achieve certain socially established values as correct. Nowadays, some of these values, in the field of consumers, are materialism, jealousy and liberty, among others.

Social class

Social stratification is the process by which a society is divided into strata, given by their degree of prestige, kind of job, ownership or power. Members of the same social class share certain values and attitudes because they have similar socioeconomic characteristics; however, they are not formally grouped, nor do they meet or communicate as such a group.

The most important factors in order to distinguish social classes are income level, occupation, prestige, social and economic power and class consciousness[15]. All of them are directly proportional to the social class, so that, for example, the more prestige you have, the more social class you belong to. It should be noted that factors such as income or training have declined in union with the social class.

Among the aspects that are influenced by belonging to a social stratum is the attitude towards consumption. In general, a high social class is linked to a higher level of income, so consumption, both in quantity and price, is higher. In addition, the place where the purchase is carried out used to be different places than in low social classes.

---

[15] **University of Jaen**, (El Comportamiento Del Consumidor Y De Las Organizaciones). 2002. Page 15

2 Scientific foundations

Social factors

Social groups, family and the status and role in each one also determine the consumer behaviour. Social groups are reference groups with which individuals identify and which influence the formation of their beliefs, attitudes and behaviours[16]. This influence can be informative, comparative and normative. Within each group, each member produces a different influence due to the role they play in the group, the status they have in the group or the knowledge they have.

There are different factors that favour the influence of the reference group, which may be related to the products and others to the characteristics of the consumers. In the first group, one would find the visibility of consumption and the perceived risk in the purchase. As a characteristic of the consumer, self-confidence influences, when the consumer seeks the approval of the group, the action of the group is reinforced.

Within social groups, the primary group is the family, which influences personality, behaviour and motivations of the individual to a high degree. The family is the basic group that provides the person, from childhood, learning about the products to be acquired and patterns of consumption behaviour. As in the rest of the social groups, each member of the family has a role in the group and exerts a different influence on the other members of the group. These roles depend on each family and have varied greatly over time.

Other factors

Personal influence is an external factor of great power, since the credibility given to someone the consumer trusts can make his/her information prevails over information from the media, for example.

The situation in which the purchase is made also affects consumer behaviour. By way of example, consumers do not buy the same way when it is for themselves as when the purchase is intended for someone else.

2.5.1.2 Internal factors: personal and psychological factors

Motivation

The purchase decision process begins with the recognition of a need and motivation is necessary to that recognition and evaluation. In addition, the motivation is the ultimate

---

[16] **University of Jaén**. (El comportamiento del consumidor y las organizaciones). Page 10

cause of all conduct or behaviour of a person, since it is the driving force that pushes him to action. Motivation directly affects the specific criteria for assessing a need.

Motivation can be defined as a general predisposition that directs behaviour toward obtaining what is desired. Thus, as an individual reaches his goals, new ones are considered.

Learning

The consumer behaviour and interests are modified as consumers learn, since they acquire experience. By this way, experience is the result of the learning process. New knowledge and experience will feed back continuously to the individual, who will guide his future behaviour in the light of new experience.[17] This learning and experience are related to the satisfaction, or in-satisfaction, that the purchase produces in the consumer, as it has been previously mentioned.

Learning can lead to habit and brand loyalty. As purchases are repeated and if the experience is satisfactory, the time spent searching for and evaluating information is reduced. In addition, as the theories of learning hold, the behaviour that is reinforced.

Perception

It refers to the personal way of interpreting and giving meaning to the stimuli to which one is exposed. The reality of a person is only the perception that he has of his immediate environment; people act and react based on their perceptions of reality and it is normally different from the realty of the rest of people. Even two friends perceive the reality in a different way.

By this way, not all the information is perceived by the consumer. Of the large amount of data and stimuli that, simultaneously, a person receives from his environment, he will only pay attention to those which are for his/her interest and reject the rest. For this reason, it is said that perception is selective.

In short, consumer perception is a very important field for the marketing department. The image of the brand, the positioning of the product, the image of the company are the result of the way in which the potential consumer perceives and interprets the information that the company launches. So, yes, in order to reach consumers and influence them effectively, marketers must understand the effects that perceptions have on their purchasing decisions.

---

[17] **University of Jaen**, (El Comportamiento Del Consumidor Y De Las Organizaciones). 2002. Page 15

2 Scientific foundations

Behaviours

With learning and experience, consumers become predisposed, favourable or unfavourable, to buying. They are constituted by the knowledge acquired about the object, the value of the person against the object and the tendency to act. In the same way as with perception, the study of a person's attitudes is of great interest in the part of marketing, since through their knowledge is expected to induce the purchasing behavior of the consumer in question, and thereby make a forecast of the demand for the product or brand considered. However, it is necessary to recognise that the identification of an attitude does not imply an exact prediction of behaviour, as the relationship between attitude and behaviour is difficult to establish in a satisfactory manner.

Age and state in the life cycle

Throughout a person's life, their tastes, habits and the need to purchase different products change, so they go change the consumer behaviour. Moreover, the need to purchase one product or another also varies with age. Even the way consumers buy varies with age. By way of example, young people tend to be more cautious about investing. In addition, age is often linked to the difficulty of deciding to buy.

Education, occupation and income

The kind of job that a consumer has exerts a great influence in the purchase decision. The time the buyer spends on his working day, the money he earns at his job, the relationships he establishes in it, are some of the aspects that affect consumer behaviour. Education, occupation and income are often closely related; high-level occupations that produce high incomes generally require advanced academic training and are more prestigious than occupations that require less education[18].

Life style

It includes every fact related to the life form of a person. His interest, opinions and activities influence in the behaviour at the consumption time. Interests and opinions can be measured through surveys, although they are not evidence-based. A study of each person's lifestyle, also known as psychographic style, is based on a battery of selected questions, accompanied by a scale on which the individual answers. Some of the features examined are related to personality. Questions of purchase motives, attitudes, values and beliefs are also often included. It is also said that, while the demographic data

---

[18] **Schiffman, L.** (Comportamiento del consumidor), 2010, page 32

of each individual (age, location, among others) determine the needs of consumers, psychographic data explain purchasing decisions and final purchasing choices.

Personality

Everyone has a different personality. Autonomy, self-confidence or sociability are some of personal characteristics that makes person face the life in a different way, and by this way, the consumer behaviour changes.

2.5.2   **Market segmentation**

The process of market segmentation and the search for target consumers consists of identifying segments within a given market or population, evaluating the marketing potential of each segment, choosing the segments to be targeted, and creating a marketing mix (i.e., product, price, marketplace, and promotion) for each selected target segment. By this way, the market segmentation process identifies groups of consumers that are like each other in one or more ways, and then designs marketing strategies that attract one or more groups. Consumers can be segmented by considering different factors such as demographic indicators, lifestyles (including hobbies) and usage habits associated with a product or other factors. After segmenting a consumer market and choosing one or more target markets, it is necessary to position the product or service.[19]

Market segmentation, or consumer segmentation, becomes necessary due to the variety of consumers, their tastes and needs. Because of this variety, mass marketing ceases to be a logical strategy, since by trying to sell the same product to all potential customers with the same advertising campaign, the product ends up not completely attracting almost any potential buyer. For that reason, it is necessary to establish groups to which the product will be directed and to focus the marketing and the propaganda to that group, to achieve the success of the same one.

Segmentation, target market selection and positioning allow producers to avoid face to face competition in the market by differentiating their offerings based on characteristics such as price, style, packaging, promotional appeal, distribution system and service level. Thus, it was discovered in the companies that the investment in market research, product, advertising and distribution that required segmentation instead of resorting to mass marketing, was profitable, since clearly differentiated products were offered that covered the needs of the consumers to whom it was directed.

---

[19] **Schiffman, L.** (Comportamiento del consumidor), 2010, page 55

2 Scientific foundations

In addition, market segmentation also identifies the most effective means of communication to place advertising messages, prices appropriate to the product and many other characteristics that must meet the product.

It should be mentioned that for a market segment to be effective, it must be identifiable, enough, stable or growing, accessible and congruent with the company's objectives and resources.

A segmentation strategy begins with the selection of the base which represents the relevant attributes of a group of potential or existent clients. The characteristics used to segment consumers coincide to a large extent with the factors that affect their behaviour, presented in previous sections. In turn, they are usually classified into four different groups, which are the result of following two criteria in that classification. These two criteria are as follows: 1. facts, which are determined using direct questions and are classified with a single objective measure as opposed to cognitions, which are abstract and can only be determined by complex questions and where most of the concepts measured do not have unique and universal definitions, 2. Ingrained attributes to the consumer that arise from the consumer's physical, social and psychological characteristics, against the consumer's specific usage behaviours (i.e. facts) or attitudes and preferences towards specific products or purchasing situations (i.e. cognitions). The four groups of characteristics are shown below:

	Ingrained attributes to the consumer	Consumption specifications
Facts	*Empirical personal characteristics* Demographic data: age, age groups, gender, marital status, family life cycle, income, education, occupation, social class Geographical location, address and geodemographic data	*Usage and purchase behaviours* Rate of use Situation/occasion of use Brand loyalty (the behavioural component) Psychographic data: real behaviours (such as hobbies)
Cognitions	*Personality, lifestyles and socio-cultural values* Personality traits Lifestyles, psychographic data and Values and Lifestyle Sociocultural values and beliefs	*Product attitudes and preferences* Required benefits Level of involvement Awareness of product alternatives Brand loyalty: perceived commitment and relationship level

Table 3. Groups of characteristics

However, most segmentation strategies mix characteristics from facts and cognition groups; it means that they are hybrids.

Demographic segmentation and geodemographic segmentation are among the segmentation strategies according to consumer attributes. The first one is widely used because the classification of consumers based on their demographic data is simple and logical, the location of the segmented group according to such data is efficient and because many habits and tastes are linked to this type of data.

Geodemographic segmentation is mainly based on the consumers' place of residence. The reason for this is that the place where a person lives determines some aspects of their consumer behaviour. For example, the climate determines the type of clothing most people wear, while fashions and styles in large cities are often very different from those in non-urban areas. Local customers, unlike tourists, are also a different segment. In addition, this strategy is based on the premise that people living close to each other are likely to have similar financial means, tastes, preferences, lifestyles and consumer habits.

The segmentation based on consumption specifications includes facts related to the real usage of a product (segmentation by rate of use and by situation of use) and cognitions that the consumers have about the product, in the way of behaviours and preferences (by benefits and by brand loyalty).

The segmentation by rate of use arises from the differences between frequent, average and occasional users and non-users of a specific product, service or brand. An advanced approach to usage rate involves identifying factors that directly affect usage behaviour. The main drawback of this strategy is that, being a very common strategy, all companies go to the same customers, so conquering those consumers requires a lot of expensive advertising.

Benefit segmentation is based on the benefits consumers look for in the products and services you purchase. These benefits represent unmet consumer needs. However, this segmentation is a little more abstract than the previous ones, since consumers do not objectively recognize the benefits obtained from a product.

2.6 Fundamentals of Data Analytics

In the age of Internet of things, a larger quantity of data is generated and consumed every day. For example, machines, financial transactions and sensors such as cameras, are increasingly gathering data from very diverse and widespread sources. That is the reason why the volume and variety of computerized information in digital databases as in other sources has grown spontaneously in recent decades. Much of this information is historical, that is, it represents situations that have already occurred. Most companies, organizations and institutions rely on this kind of information, about past experiences drawn from a wide variety of sources[20]. These data provide a rich source of information that can be transformed into new, useful, valid and human-understandable knowledge. Thus, there is a growing interest in exploring these data to extract this knowledge, using it to support decision making in a wide variety of fields. By this way, several companies are interesting in the potential of these data to support their work, reduce waste and dangerous and tedious work activities, and, in general, to reduce costs and increase profits[21]. For this reason, "Data Analytics" appeared, defined as a science of inspecting, cleaning, and transforming raw data to highlight useful information, suggest conclusions, and support decision-making[22]. Essentially, data analytics methods are applied to help understand hidden patterns and relationships in large and complex datasets. This chapter discusses some relevant aspects and concepts in data analytics to lay the foundation for using data analytics as a tool for the resolution of the problem.

2.6.1 Data Analytics

Data analysis is the science created by the need to analyse information and obtain knowledge from it. This process of analysis is usually considered as a cycle, since it is common to learn something at the end of a phase that cause the comeback to last phases and redefine the work already done based on new insights[23]. This lifecycle consists of six phases briefly described below[24].

---

[20] **Hernández, J., Ramírez, MªJ., Ferri, C**., "Introducción a la minería de datos" 2004, pag 4

[21] **Moreira, J.M**. "A General Introduction to Data Analytics" (2018) p. 3

[22] **Moreira, J.M**. "A General Introduction to Data Analytics" (2018), p. 4

[23] **Mishra, S.** "Applied Statistical Modeling and Data Analytics" 2017

[24] **Riquelme, J.C., Ruiz, R., Gilbert, K.,** "Minería de Datos: Conceptos y Tendencias", (2006), pag. 13
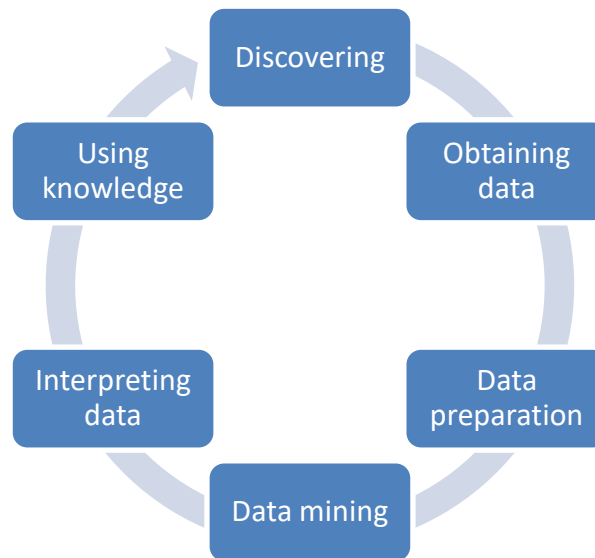
- Phase 1- Discovering and understanding the application domain. The previous knowledge about the matter, the goals of the work and the availability of resources are learnt, in terms of people, data, time and technology. In addition, this phase aims to frame the business problem as an analytical challenge and formulate initial hypotheses that will be tested later with data analysis[25].

- Phase 2- Obtaining of data. This phase includes the steps to gather, explore, evaluate and condition data to become familiar with them. The data are collected from secondary sources, previously commented on, or would be obtained in research processes created for the process, such as through surveys, questionnaires, etc., which will be commented on in later sections.

- Phase 3- Data preparation. In this phase a data cleaning, transformation, integration and reduction are carried out, with the aim to streamline future phases. It is an important phase of the process as it forms the basis on which to work and investigate. Poor data preparation will lead to poor analysis and poor research findings.

- Phase 4- Data Mining. In this phase knowledge is obtained from data. This phase, which includes the planification and creation of the model, includes many tools for analysis the data, such as classifying, regression, clustering, summarising, extraction of rules, etc. It is the central phase of the process, as it concentrates the final objective of the whole. For this reason, it will be explained in greater detail in later sections.

---

[25] [25] **Moreira, J.M**. "A General Introduction to Data Analytics" (2018) p. 3

[25]

- Phase 5. Interpreting the extracted data. It aims to explain obtained data and represent them in a visual form. Moreover, it is critical to frame the results of the work in a manner that demonstrates clear value. At this stage it is usual to share the results with stakeholders to check that the results have been successful, based on the objectives set in the first stage. In addition, team should identify key findings, quantify the business value, and develop a narrative to summarize and convey findings to stakeholders[26].

- Phase 6. Use, communicate the knowledge and establish the whoole process. In this phase final reports are presented and transmitted to the rest of the workers.

These steps, in an iterative and interactive way applied, ensure a useful extraction of data. In addition, the methods and tools used in data analysis can be divided into four categories:

- Descriptive Analytics. Its goal is to identify behaviours and describe the shape and patters of existing data. Classical descriptive methods include segmentation, clustering and classification.

- Diagnostic Analytics. This category includes those analytical activities that examine correlations in the system under consideration that are generally more complex and promise the collection of higher-value information. However, they are based on a limited number of data.

- Predictive Analytics. Based on historical data, it intends to establish prognostics and predict how the analysed topic will evolve.

- Prescriptive Analytics. Based on what should be done, it derives potential actions in order to increase the probability of the occurrence of a desired future development.

Given the importance of obtaining knowledge from already available information, the science of data analysis has been formalized in a process of discovering knowledge from data, KDD, which will be explained in future sections.

In addition, there are different data analysis methodologies, depending on the type of information that will be analysed. Some of the aspects that influence the selection of these methodologies are the structure of the data and the scale on which they are represented, presented in the next section.

---

[26] **Dietrich D**. "Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data" (2015). Pag 30..

2.6.1.1 Data structures

The first aspect to be considered during the choice of a methodology is the structure of data, since data can be shown with a great variety of structures and formats. In the figure, it is represented a visual classification of the different format of data.
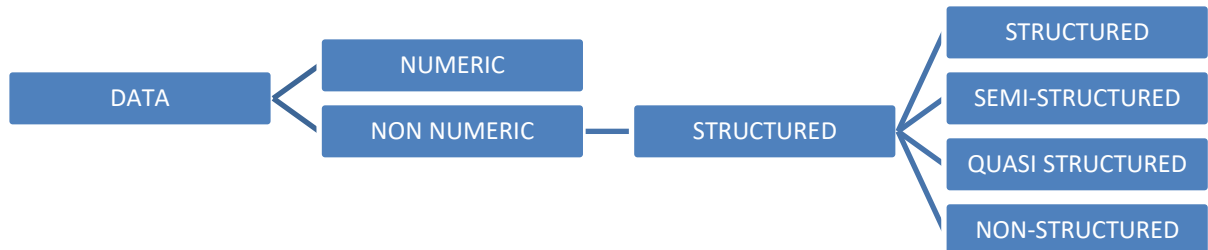
Figure 6. Different format of data.

As it can be seen, the first difference is to be numeric or non-numeric. This second group refers to structured information, dividing it according to its degree of structuration, from completely structured to unstructured data. Although analysing structured data tends to be the most familiar and simple technique, different techniques are required to meet the challenges to analyse the rest of "structured data"[27].

Within the group of structured information, the most structured data contain a defined data type, format and structure. Some examples of them are CSV files and simple spreadsheets. Semi-structured data are, for example, text files with a recognizable pattern, such as Ex-tensive Mark-up Language (XML) files, which allow parsing. Parsing is the splitting of textual data into a format that is suitable for further processing of the information. Quasi-structured data are text files with an irregular format, which can be formatted with the necessary effort for further processing. Data that have no inherent structure, such as text documents, PDFs, image and video files, are unstructured data[28].

---

[27] **Dietrich D**. "Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data" (2015). Pag 6.
[28] **Dietrich D**. "Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data" (2015). Pag 6.

2.6.1.2 Types of scales

Data can be measured and presented with a great variety of scales. To facilitate its processing and analysing, scales can be classified by qualitative scales and quantitative scales. Qualitative scales are also divided into ordinal and nominative scales, for which only equality or inequality is defined. Nominal data cannot be ordered according to how big or small a certain characteristic is, but ordinal data can[29]. Typical examples for nominative scales are object data such customer names or product names. In the ordinal data, in addition to equality or inequality, an order relation is also defined. In the second group, quantitative scale type, it is also differentiated between relative, or interval scales, and absolute scales.

In fact, we can order the four scales types according to the amount of information than each one provides. From most informative to less informative, the classification will be: absolute scale, relative scale, ordinal and nominal. In addition, the more informative is a scale, the more operations are possible. That means that it is possible to convert a more informative scale into a less one, but not in the opposite way, since it would lead to a loss of information[30].

2.6.1.3 Data mining

As I have mentioned before, Data mining is an integral part of Data Analytics, which is the overall process of converting raw data into useful information. Thus, there are different definitions for data mining, two of which are presented below.

First, data mining is defined as the process of extracting useful and understandable knowledge, previously unknown, from large amounts of data stored in different formats. In that sense, the fundamental task of data mining is to find intelligible models from the data collected[31].

Second, data mining has been defined as the process by which a useful model for prediction is generated. This model is built based on data found in a database, to which some algorithm has been applied precisely in order to propose a model[32].

In conclusion, it could be said that data mining is a process that integrates the data from different sources in order to subsequently extract important knowledge, that is, identify

---

[29] Vgl. **Moreira, J., et al.** (A General Introduction to Data Analytics), 2019, S. 35
[30] **Moreira, J., et al.** (A General Introduction to Data Analytics), 2019, pag. 37
[31] **Hernández,J., Ferrari, C., & Ramírez, M**. (Introducción a la minería de datos). 2004.
[32] **González, L**. (Una arquitectura para el análisis de información que integra procesamiento analítico en línea con minería de datos). (2005)

transcendent, valuable and useful information from which institutions will be able to make significant decisions[33]. By this way, the main purpose of data mining is to recognize hidden or new patterns and meaningful correlations in data.

"Data analytics" and "data mining" terms are very often confused. Data mining is a particular phase in data analytics process, which finalize with the extraction of real knowledge from data sources. Other phases in data analytics process are the preparing of the general data and selection and cleaning of them, among others, as it has been already named[34].

Just as in data analysis there is a classification of types of analysis, data mining models are classified into predictive and descriptive. In the first case, there is a variable with an unknown value, and the purpose is to determine it. This variable is called response, dependent variable or objective, while those used to make the prediction are the predictors or independent variables.

At the same time, predictive models need to be trained, using a set of training data whose target variable value is known. The idea is that the model delivers results based on learning, in other words, that adjusts to the known reality. These types of models are also known as supervised learning models, because the estimated or calculated values are compared with the known results and therefore there is a clear measure of the success or failure of the prediction. Some algorithms used in these models are classification and regressions.

On the other hand, there are descriptive models, in which there is no known result to guide the algorithms. Therefore they are known as unsupervised learning models, where the model is adjusted according to the observations or data provided, and heuristic arguments are often used to evaluate the quality of the results. Some algorithms used in these models are clustering and association rules. Unsupervised learning is used in cases where there is no prior knowledge of the outcome to be achieved, for example, by segmenting clients into groups that have not been previously defined. After the model is trained, a data sample independent of that used for the construction and training phase of the model is used, with the intention of evaluating the predictive capacity of the model.

Within data mining, we have to situate types of tasks, each of which can be considered as an independent problem, as it provides different information and does not have to use the same base data. As it has been commented previously, among the predictive tasks

---

[33] **Daza, A**. (Data Mining: Minería de datos).
[34] **Riquelme, J.C., Ruiz, R., Gilbert, K.,** "Minería de Datos: Conceptos y Tendencias", (2006), pag. 12

we find the classification and the regression, whereas the grouping, the rules of association, both sequential and correlations are descriptive tasks. They are described in greater detail below:

- Classification. It is probably the most used task in data mining. It classifies a data within one of the predefined categorical classes. By this way, the purpose of classification is to assign a label to each data observed. In this technique, the total set of classes, or number of groups, is determined. In addition, most classification techniques are supervised, i.e., they begin with a set of already labelled observations to learn the probability that the attributes of these observations may contribute to the classification of future unlabelled observations. For example, existing vehicle purchase data can be used to develop a "buy" or "do not buy" label for potential new customers.

- Regression. Is also a predictive task, which is used to establish relationships between one or more independent variables and a dependent variables. In turn, there are two types of regression, linear regression, used to model the relationship between several input variables and a continuous outcome variable, and logical regression, which can be used to predict the likelihood of an outcome based on the input variable.

- Clustering. It is the descriptive task par excellence and consists of obtaining "natural" groups from the data. We are talking about groups and not classes in this case, because, unlike classification, instead of analysing data tagged in a class, it analyses them to generate this label. Data is grouped based on the principle of maximizing similarity between the elements of a group by minimizing similarity between the different groups. That is to say, groups are formed such that the objects of the same group are very similar to each other, and at the same time, very different to the objects of other groups. Grouping is also often referred to as segmentation, since it divides or segments data into groups that may or may not be disjointed.

- Summarizing. This model provides a compact description of a data subset. In some cases, it is considered a type of grouping in which each group generated is considered a summary of the elements that make it up in order to more concisely describe the data.

- Correlations. They are a descriptive task used to examine the degree of similarity of the values of two numeric variables. Correlation analysis can be very useful for establishing rules of correlated items.

Data mining applications

The integration of data mining techniques into day-to-day activities becomes commonplace. Distribution businesses and targeted advertising have traditionally been the areas in which mining methods have been used the most, as they have allowed or increased the receptivity of offers. But these are not the only areas in which they apply. In fact, we can find examples in all kinds of applications: financial, education, industrial processes, etc.

Below is a list to illustrate in which areas mining can be used, with respect to some of the areas mentioned above.

Financial applications and banking: calculation of correlations between financial indicators, identification of stock market rules from historical data, credit risk analysis.

Applications in education: selection and recruitment of students, detection of dropouts and failures, estimation of length of stay in institutions.

Applications in industrial processes: extraction of models on compound behaviour, detection of parts with problems (quality models), prediction of failures and accidents, extraction of cost models, extraction of production models.

In addition, in relation to this work, data analysis through data mining is performed for analysis of market, distribution, and general trade. In this aspect, data mining techniques are applied to evaluate advertising campaigns, analyse customer loyalty, for customer segmentation and estimation of stocks, sales, among others.

2.7 Interim conclussion

The automotive industry has undergone greatly in recent decades. The two major changes that characterize the dynamism of this sector were the creation of in-line production and the emergence of Lean manufacturing. In addition, this evolution has reached the point where manufacturing companies have delegated the production of the vehicle to suppliers, thus reducing production responsibilities and enhancing their work in contact with the customer, such as marketing and growth strategies.

Thus, automobile companies have developed numerous strategies to increase their competitiveness, due to the strong growth and potential of this sector. One of these strategies is the modularisation strategy, whereby vehicle production is divided into modules manufactured by different suppliers. In this way, each supplier will manufacture one of these modules, being able to improve and deepen in technical details.

The evolution of this industry sector has led to the evolution of the supplier chain, which has become complex and extensive. To facilitate their knowledge, the supplier chain is defined through the vertical dimension, formed by the suppliers acting at consecutive levels in the production of the vehicles, and the horizontal dimension, occupied by the suppliers who perform the same functions. Within the vertical chain, it is important to know the consecutive levels in order to facilitate the flow of information and to be able to improve the vehicles based on the needs of the customers.

Finally, communication between suppliers and manufacturers is the basis for the flow of information between them, and with it, for the improvement of products based on customer satisfaction. In addition to the traditional means of communication (telephone, meetings, mail and fax), in recent years companies have developed a new way of communication that facilitates instant exchange between both parties, suppliers and manufacturers, the supplier portal.

2 Scientific foundations

As a result of this evolution, manufacturing companies have improved their relationship with customers, strengthening the activities dedicated to the orientation of the product towards the consumer. Among other measures, companies have expanded techniques for obtaining information from customers, to improve products around customers and to be able to establish consumer patterns. However, it is concluded that the consumer profile is complex to establish due to the wide variety of factors, both external and internal to the consumer, of the variety of approaches with which it can be analysed.

Also, due to the wide variety of consumers that exist and the involvement of companies in customer-focused processes, manufacturing companies have developed numerous strategies of product to supplier approach to increase their sales. The most developed is market segmentation, of which, in turn, there are numerous aspects, the most common being demographic segmentation.

Finally, data analysis is a science aimed at obtaining knowledge from information. It is a very broad science, considered as a cycle of successive phases, in which the most important phase, for this work, would be the phase of data mining. Thus, within this last part, there are numerous tools that allow the processing of information in order to obtain knowledge with the highest possible quality.

3   Problem relevant method

3.1 Supplier exchange information

Over the last few decades, the relationship between suppliers and assembly companies has evolved, becoming more intense. One of the factors reflecting the change is the amount of information currently exchanged between the two members, which is presented in the following sections.

In the next image, the main information exchanged is included, in order to give a visual idea of it, before being explained.
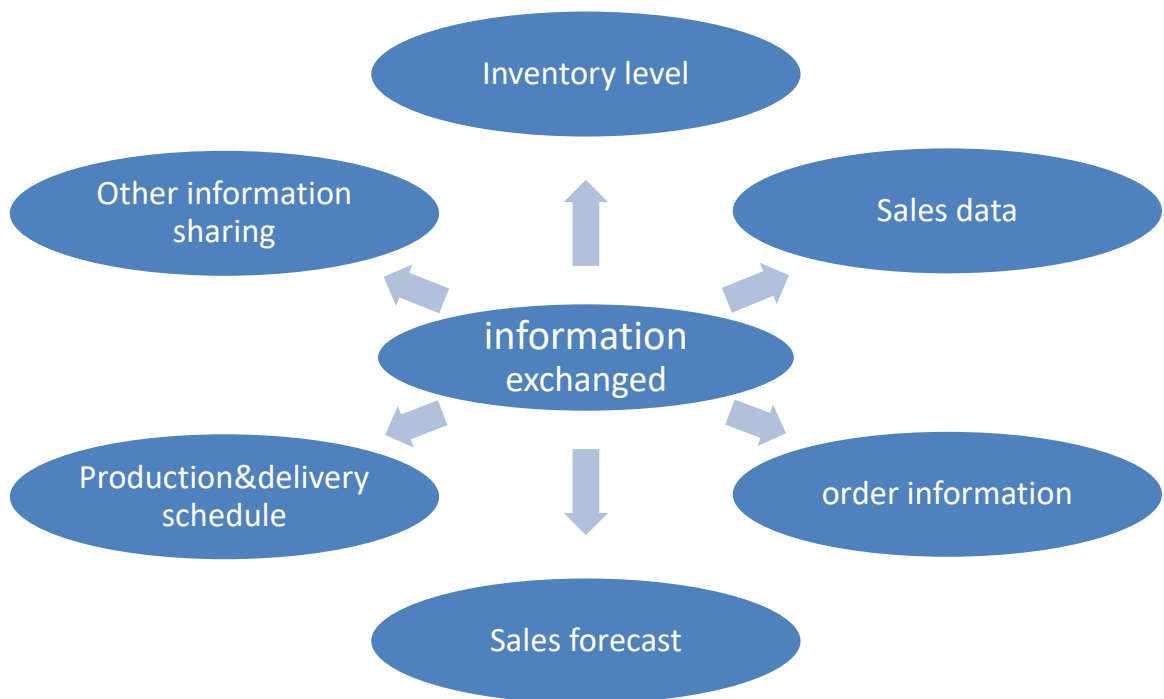
Figure 7. Information exchanged between manufacturers and suppliers

3 Problem relevant method

3.1.1 Inventory level

Information relating inventory level includes the information exchange regard to the inventory of different actors within the interfaces[35]. Between an assembler and its supplier, information about their inventory levels of the product provided is exchanged. In addition, when the product provided is not selling well is also included in this type of information.

Inventory level is one of the most common data shared between supply chain partners. Access to supply chain inventory status can contribute to lowering the total inventory level in the supply chain[36]. Knowing the inventory level of a supplier allows one of them to reduce safety inventory, since if one of them has already safety inventory, the other can preclude to have it. In addition, develop a vendor-managed inventory, suppliers make replenishment decisions based on customer's specific inventory. Thus, supplier need to be clear about both their own stocking point and the customer's receiving point[37].

3.1.2 Sales data

Sales data more refers to the data directly collected from the purchasing activity of downstream member's customer, sometimes even including end customers' gender or estimated ages[38]. Some of the data exchanged as sales data is the quantify of product sold during a previous period.

Not sharing sales data leads to erratic order patterns of the suppliers, because, in that case, sales data are usually replaced with orders from downstream, which are almost never the same amount of sales. So, orders from downstream cannot replace sales data.

3.1.3 Order Information

Inside order information, a distinction between order details and order status for tracing and tracking can be made. The first one includes specifics in terms of the type and quantity of the product, which is basic information exchanged in every purchase. The latter

---

[35] **Burmeister, C., Liang, Y**., "Information exchange between a retailer and its supplier". (2016) pag. 8

[36] **Lee, H.L. and Whang, S**. (2000) 'Information sharing in a supply chain', Int. J. Manufacturing Technology and Management, Vol. 1, No. 1, pp. 79–93

[37] **Angulo, A., Nachtmann, H., & Waller, M. A**. "Supply chain information sharing in a vendor managed inventory partnership. Journal of business logistics", 25(1), (2004). pag. 101–120.

[38] **Burmeister, C., Liang, Y.**, "Information exchange between a retailer and its supplier". (2016) pag. 7

one, status for tracing and tracking, further includes delivery time and place, delivery performance in terms of on-time delivery[39].

The exchanging of this information lead to some benefits relating to manufacture relationships, since the quality of manufacture service is improved. Moreover, savings in the labour cost of manual operation appears, because automated service is more productive and faster.

3.1.4 Sales forecast

Sales forecast represents the data that shows future demand of the marketplace, made or calculated by companies. The period for the prediction demand is defined by each company, varying from months to seasons. Occasionally, this information includes the date for big up-comings campaigns.

Suppliers uses sales forecast as the basis of their production planning. By this way, the more accuracy in the predictions, the best production plan for the supplier. To get the most accuracy sales forecast, it is required the collaboration of both parts, manufactures and suppliers.

3.1.5 Production/delivery schedule

Product information contains all information regard to the product such as production, schedule, producing ability and exploitation information of new products. The availability of production schedule and producing ability information allow suppliers to ensure a reliable delivery and to improve their own production schedule. In addition, information about input/job availability helps buyers to expand the planning horizon of their own production schedule[40]. To launch new products to the market, exploitation information of new products is required.

The most commonly exchanged information of delivery schedule is delivery date, place and time. The accuracy of the information depends on each company. Many companies include shipment tracing and tracking as well as notifications about delays or urgent orders.

---

[39] **Burmeister, C., Liang, Y**., "Information exchange between a retailer and its supplier". (2016) pag. 8

[40] **Lee, H.L. and Whang, S.** (2000) 'Information sharing in a supply chain', Int. J. Manufacturing Technology and Management, Vol. 1, No. 1, pp. 79–93 [pag

3.1.6 **Other information sharing**

Other information sharing between suppliers and manufacturers are capacity information and performance metrics, among others. The second one includes quality data, lead times, service performance and queuing delays at workstations. This information provides suppliers to improve their supply chain, by recognizing bottlenecks, for example. Moreover, suppliers and manufacturers can jointly develop a performance metric that measures the performance of the entire supply chain. This metric provides a valuable information to improve the delivery time performance of the overall chain[41]. Capacity information of manufacturers to suppliers allows suppliers to prepare themselves against possible shortages of manufactures.

In the following table, all the information exchanged is included in order to give a general idea.

3.1.7 **Interim conclusion**

According to last sections, information exchanged between a supplier and a manufacturing company refers to inventories, sales, orders and production process. This information comes either from the company's own activity or from the predictions it makes. Hence, it could be considered as information generated within the company. However, customer's information is not included in the exchanging of information. That fact involves a problem in the product development, because most of assemblers no longer produce pieces themselves, but assemble pieces they order from their suppliers. In this way, suppliers are responsible for the design and production of the different modules of the vehicle. Therefore, suppliers are responsible for a product being developed according to the requirements of the consumer, not the assembly company. Then, to improve the quality of products by satisfying customers' expectations, customers information must reach suppliers.

In this way, it is required a process to canalize customer useful information from assemblers to suppliers, by creating a methodology to obtain customer information and analyse it, to facilitate its transfer to the suppliers. In the following sections, sources of customer information are presented and analyse to start with the creation of the methodology.

---

[41] **Lee, H.L. and Whang, S**. (Information sharing in a supply chain), Int. J. Manufacturing Technology and Management, Vol. 1, No. 1, (2000) pp. 79–93 . page 8

3.2 Sources of customer information

This section presents the different methods of obtaining information from manufacturing companies, represented in the following diagram:
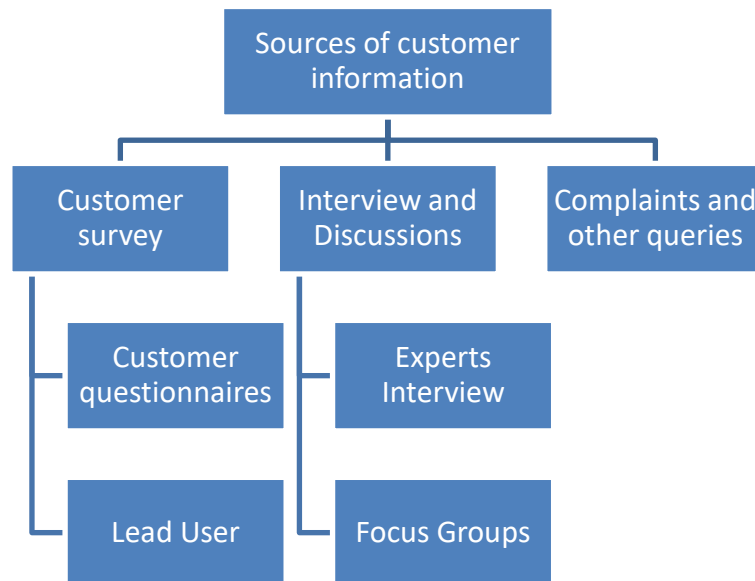
Figure 8. Methods of obtaining customer information

3.2.1 Customer survey

In this section, process in which the information comes directly from the customer are analysed. It means that customers are the sources of information for the product development. Inside customers, it does not only consider direct buyers, but buyers of competing products. The process of obtaining customer information by this method is based on consulting special groups of customers to anticipate future trends and requirements about the handling product. In this evaluation, they judge how satisfied they are with a product, how they assess the product quality and whether a follow-up purchase is possible. This method also includes the analysis of the reasons for the decision for the competitive product, since it can also provide information about customer requirements. There are two types of customer survey, developed in the following image[42].

---

[42] **Jiménez y Villegas MC y cols**. (Satisfacción del usuario como indicador de calidad Revista Mexicana de Enfermería Cardiológica) 2003;11 (2): 58-65.

3 Problem relevant method

┌───┐
│ **Customer questionnaire**   │

- Customer surveys and questionnaires are the most direct sources of information, providing information about both possible and needed improvements product. It is required the preparation of the document, which will be answered.
- To motivate customers to participate, questionnaires must be interesting, objective, unambiguous, easy to fill out and not overloaded.
- To improve analysis and facilitate the classification, questionnaires must have substantive and personal questions.
- Questions can be open-ended, generating more valuable information but more difficult to analyze, or closed-ended, limiting the answers to pre-developed alternatives.
- Three types of questionnaires: qualitative, quantitative and mix questionnaires.
- Results of this source are a set of questions of numerical and short answers.

└───┘

┌───┐
│ **Lead User**   │

- Lead user approach assumes that information on future needs and conceivable solutions is only held by a few leading users, the so-called lead users.
- . Lead users are users of a product or service that currently experience needs still unknown to the public and some of them have already developed or implemented their own ideas to improve the products they use, showing an unusual creativity in their special field.
- Workshops are defined by a working meeting in which some lead users work on a specific topic, with the aim to identify measures for the future. Workshops offer great potential for working together with various interest groups to develop measures to solve specific problems and are also suitable for developing product ideas.
- As a special form of the classic workshop, the Lead User Workshop serves to systematically record customer wishes and requirements. In particular, this approach can partly anticipate future market requirements for products. This procedure mainly consists of four steps. Firstly, important market trends are identified, in order to get the aspect of the product in which it can work. Secondly, lead users are identified from relevant customer groups, who have the above-mentioned lead user characteristics. Next, the created lead-users group examines product requirements, within the framework of a needs-analysis and with a special focus on modification the unusual benefits of the product. Finally, the reached information is validating if the "non-lead users" customers from the developed product concepts achieve similar satisfaction of needs as lead users. Once that these customers have approved the new ideas of the lead user workshops, they are collected and detailly explained in a report, thus generating the results of the methodology.

└───┘

3.2.2 Interviews and Discussions

3.2.2.1 Experts interviews

In this group of sources, the customer does not play an important role, since the information is provided by experts. It could be said that these methods rely on the know-how of experts or other persons with a special role in product development. In addition, an expert is usually regarded as representing a certain group.

The interviews can be classified according to their degree of structuring. In the one side there are the openness and non-structured interview against the fully-structured and completely planned interviews. In this scale, expert interview is in the middle in data

collection[43]. The reason of that is the use of a guideline as a basis, which provides a structure in the course of the interview, but at the same time allows sufficient flexibility for an open discussion. This serves above all to orient the interviewer regarding content, so that all questions previously considered important will be addressed and a certain comparability of the interview results can be ensured when several expert interviews are conducted[44]. In turn, interview questions can be of four main types: short questions, answered with yes or no, numerically answered questions, one-interval answered questions, and descriptive questions. By this way, results of an expert interview are a set of questions with different format, which must be treated differently.

The focus of the expert interview is on his or her experience and knowledge in connection with the research topic, in contrast to other target groups of a survey in which the focus is on the interviewee. The comments of the expert are examined from the outset against the background of his role in the field of investigation.

The function of the expert, which is usually coupled with a responsible position and privileged access to information on groups of persons, organisational processes and decision-making processes, provides an insight into specialist knowledge and background knowledge as well as into the specific experiences the expert has made in his field of activity[45].

Finally, to obtain reliable and as undistorted information as possible, it is important to ensure that the survey situation for the interviewee corresponds as closely as possible to an everyday situation.

3.2.2.2 Focus Group – Customer group

Like lead user workshops, focus group sessions are also based on customer integration into the development or innovation process. However, the customers involved in a focus group discussion do not have a focus on the innovative lead user characteristics, but on a defined task of the product. For this reason, the focus group is less suitable for identifying future customer requirements than current ones, but it is a qualitative technique of studying the opinions or attitudes of these customers, as it will be explained in the following.

A focus group is an artificial group, since it does not exist before or after the conversation session, in which the appropriate number of participants ranges from 6 to 12 people[i], to

---

[43] **Lamnek, S**. (Qualitative Forschung: Grundlagen und Methoden), 2002, p.173
[44] **Lamnek, S**. (Qualitative Forschung: Grundlagen und Methoden), 2002, p.173
[45] **Albers, S., Klapper, D., Konradt, U., Walter, A., & Wolf, J.** (Eds.). (2009). Methodik der empirischen Forschung

make it an effective group. An essential participant here is the moderator, in charge of asking questions and leading the discussion. Its task is to channel the discussion so that it does not move away from the topic of study. To sum up, focus group meetings contain moderated, relatively open discussions that address a defined problem. To capture usage problems, focus groups can also be observed when using existing products in artificial or real usage situations[46].

As I have said before, the special result quality of the focus group method is based on the reflection of the exchange and discussion, not only on the individual opinions of the participants. As a rule, focus groups are used where questions that are difficult to structure hardly permit a standardised survey and only qualitative results can be achieved. Therefore, it is particularly suitable for the survey of attitudes, feelings, ideas and conceptions of certain customer groups[47].

The obtained answers during the discussion are analysed by the moderator, who makes a report collecting all of them. Thus, the result of a focus group is a report, which describe facts to improve in the product or points to be worked.

3.2.3   Complaints and other queries

Complaints and other queries made by users are a great opportunity to measure the level of user satisfaction with the brand or with a specific product or service. In addition, consumer's complaints are for a company an opportunity to identify and rectify specific problems with their current product or service[48].

Increasingly companies are recognizing the value of a customer complaint in that it is feedback on their experience, and an opportunity to not only resolve a problem for that customer but perhaps also for a much larger number of customers. In addition, a complaint includes inherent defects of the product, that encourage the improvement of the product. Moreover, the analysis of the customer complaints in a company lead, not only to an improvement of the product, but to a creation of new ideas about new products.

Because of its importance and its usefulness, complaints can be presented in both spoken and written form. Spoken form was more commonly used in last years, where there were not any complaints management. It is also used with small problems, which can be

---

[46] **Albers, S., et al.** (Handbuch Technologie- und Innovationsmanagement), 2005, p. 272

[47] **Schmitt, R., et al.** (Qualitätsmanagement), 2015, p 155

[48] **Chugani, S., Govinda, K., & Ramasubbareddy, S.** (Data Analysis of Consumer Complaints in Banking Industry using Hybrid Clustering.) 2018. Page 1.

solved instantaneously. However, if the complaint is related to the product, it must be presented in the written form, since written form allows users to document the problem in detail and provides to the company time and capacity to solve it.

Within the written form, there are different presentation structures. Thus, most companies have a fixed complaint format, in order to speed up the process of analysing them. This format can be divided into three parts. The first part would be the personal-data part, in which all the data needed to contact the user are included. It must be said that the more quantity of data, more efficiently the process of the complaint and the feedback with her or him. The second part consist of a little formulary, that helps the following classification of the complaint. In this part it is selected the service, product or part of the product which the complaint is about. Finally, the written form includes a section in which the customer explains in detail the problem. In some cases, customers propose a solution to the problem. In terms of analysis customer data, the most useful part of the document is the description of the failure. So, it can be said that complaints are presented as a text.

The central problem with this type of information source seems to be that most people who have a reason to file a complaint do not. However, this trend seems to be changing in recent years.

3.2.4 Social Media and online reviews

The rapid growth of the Web in the last decade makes it the largest publicly accessible data source in the world, since it contains a wealth of opinions about product, services, and more. which are expressed in newsgroup posts, review sites, and elsewhere[49]. For example, an abundance of customer opinions is available in Internet forums free of charge and up-to-date. This great quantity of available data counters the lack of information provided by complaints. Moreover, because of this phenomenon, the "face-to-face" communication no longer occupies the first position, but customers prefer to express their opinion and dissatisfaction by Internet. Here, customers reveal personal experiences, make recommendations or warn against purchasing certain products. Due to the almost unlimited circle of addressees, social media represent a powerful instrument with regard to the dissemination and reception of customer statements[50].

Customer complaints that are not addressed to the company are on average brought to the attention of at least ten other persons from the personal environment by the complainant carrier. Social networks such as Facebook, Twitter and Google+ increase this

[49] **B. Liu,** (Web Data Mining: Exploring Hyperlinks, Contents and Usage Data), Springer, BerlinHeidelberg, 2007.
[50] **Schmitt, R., et al.** (Die Meinung immer dabei), 2014, p. 20-21

negative snowball effect. Given the large amount of opinions about products and services, social networks can be considered as the mirror of the customer information.

Through these comments and opinions, the consumer participates in the improvement of the product, if this information is treated and analysed in a convenient way. An adequate analysis adapted to the type of information is reached by knowing the format with reviews are presented on the Internet. After the observation of some automotive forums, such as *Motor-talk, autoextrem* and *carpassion*, among others, it can be seen that opinions and reviews are written in a text form, without following a specific structure which would facilitate the classification and analysis of them.

Thus, due to the unlimited amount of field data available on the Web there are numerous text mining tools, which extract evaluations of social media according to a given objective. The object of the analysis is the general perception of the product by the customer in the field and the derivation of concrete deficiencies, desires and requirements by the customer[51].

3.2.5 Interim conclusion

There is a wide variety of sources of consumer information. This wide range of sources so different from each other implies that the information obtained from them is also very varied, in terms of format and quality of the information, among other information criteria. This makes it difficult to establish a single analysis procedure. For this reason, this methodology will consider how to analyse customer data in order to obtain the most complete analysis possible and using all possible sources of information; avoiding that the variety of formats hinders the analysis and obtaining of knowledge.

3.3 Customer behaviour study

The field of customer behaviour includes many questions, from the simple purchase of a small object or food until the purchase of a factory. By this way, the study of the consumer behaviour is not easy because of the peculiarities it presents. The activities that consumers carry out and the process that they develop are simple manifestations of behaviours as human beings. This behaviour is essentially complex, since it is influenced by many factors, both internal and external, and about which there is little and fragmentary knowledge. The main characteristics of the consumer behaviour that makes its study difficult are:

---

[51] **Schmitt, R., et al.** (Qualitätsmanagement), 2015, p. 161

complex	• As it has been said, the behaviour of a customer is influenced by many both internal and external factors. In addition, the market response to the stimuli tends to be non-linear; it tends to have threshold effects, which are minimun required levels of stimuli to produce an answer
Changeable with the life cycle of the product.	As the product consolidates in the market, throughout its life cycle, and reaches maturity, consumers undergo a learning process and gain greater experience of the characteristics and benefits of the product. It leads to an evolution of the characteristics of the product, in which thoose characteristics that gave diferentiation to the product and produced satisfaction in the consumer become required characteristics and consequentily, thoose that were required become basic characteristics, according to the Kano Model. Thus, the consumer behaviour changes with the life cycle of the product.
Changeable with the kind of product	• Consumers do not have the same interest in all the products, nor their purchase represent the same risk for the consumer. For example, the consumer will show a high level of implication in the purschase if both it is important for him, although its price is not high, and the associated risk, because of the consequences for a wrong purchase decision, is also high. Otherwise, the pruchase will be of low implication and can be acted by inertia

Figure 9. Characteristics of the consumer behaviour.

Despite the difficulty of analysis, knowing consumer behavior has benefits for consumers as well as for the company and even in a macroeconomic environment. Related to consumers, the study of the consumer behaviour facilitates the adaptaption of the products to consumer, from the characteristics of the product until the price it is launched. By this way, facilitates, guides and makes the purchase and consumption of products more satisfactory. If the purchasing decision process is also known and the factors that influence it, the products can be distributed and promoted in such a way that the decision is easier and more pleasant for the buyer.

In the field of businesses, the study of the consumer behaviour translates into a better understanding of consumer needs and market opportunities for companies. This knowledge leads to the development of a commercial strategy more adapted to the consumer, that will increase the demand for the products offered, increasing the participation in the market and the benefits of the product. In addition, knowing consumer behaviour is essential for any strategic marketing approach, such the previous mentionated market segmentation. For example, criteria such as desired benefits, brand loyalty, occasion, preferences for types of packaging, etc. are subjective segmentation criteria based on consumer behaviour. Thus, in order to position a product it is necessary

3 Problem relevant method

to know which attributes are taken into account and valued by consumers. Marketing mix decisions, related to product, price, distribution and communication also require knowledge of price perceptions, assessment of product attributes, media exposure habits, purchasing habits and preferences, which are involved in consumer behaviour. Finally, The study of consumer behaviour also helps to explain the successes and failures of marketing strategies. For example, the success of rechargeable batteries is largely due to new consumer values related to environmental conservation[52].

Finally, from a macroeconomic point of view, knowledge of consumer behaviour is important, because it helps to understand how a society satisfies its needs, such as communication, food, transportation, etc. Besides, the Economic policy measures will be more effective if politicians know how the consumer spends time and money.

On the other hand, there are numerous studies in this field, which try to draw conclusions about the establishment of a consumer profile. Most of them do not establish a single methodology or establish a generic pattern, but they do pose a series of premises that the consumer profile must meet in order to be useful for the company[53].

The defined profile must be a generic profile, which includes a group of profiles with common characteristics generating a purchase pattern.

- The consumer profile must refer to general characteristics of the product to be acquired; it cannot be based on very specific criteria, since it would make it impossible to establish a generic pattern.

- The consumer profile should be defined on the basis of the personal characteristics of consumers, without incurring in privacy policies.

- Consumer behaviour is not determined; it depends on social, educational, etc. factors, as explained in previous sections.

---

[52] **Alonso, J., Grande, I**., (Comportamiento del consumidor. Decisiones y estrategia de marketing).

[53] **Molero, G**.. Una arquitectura para el análisis de mercado: inteligencia de negocios y gestión de ventas. (2008)

3.4 Tools in data analytics

As previously mentioned, the science of data analysis is a very important science for obtaining knowledge from large amounts of data. That is why in the last decades certain procedures have been established to guide analysis, to make it a routine task in companies. In this way, the process of discovering knowledge from the data, KDD, and some of the methodologies of this science will be explained below, in order to provide a basis on which to develop the methodology of our study.

3.4.1 Knowledge Discovery from Data

The extraction of knowledge from raw data is an important task to improve products and production processes, both for assemblers and suppliers. The science of data analysis is formalized in a process known as Knowledge discovery from Databases, KDD. This process consists of an iterative and interactive sequence of stages or phases, as has been proposed with the science of data analysis. It is iterative because the output of some of the phases can make return to previous steps and because often several interactions are necessary to extract high quality knowledge. It is interactive because the expert or the one developing the project must help in the preparation of the data, validation of the extracted knowledge, etc. During the next sections, a detailed explanation of these phases is developed.

3.4.1.1 Integration and compilation phase

Generally, the data needed to carry out a KDD process belongs to different organizations, different departments within the same company, and has very different origins. In many cases, data will have to be acquired from outside the company. This makes the start of the process a challenge, as each data source will come with different record formats, different degrees of data aggregation, and so on. Therefore, the first thing that is done is the integration of all these data. The general idea would be to create a database that encompasses all these documents, thus ensuring future accessibility and visibility. It should be noted that the creation of databases is not essential if the volume of data is not excessive.

3.4.1.2 Selection, cleaning and transformation phase

The quality of the knowledge discovered depends not only on the tools used in the subsequent phase of data mining, but also on the quality of prepared data. Therefore, after collection, the next step in the KDD process is to select and prepare the subset of data to be worked on. This step is essential because some data collected in the previous stage are irrelevant or unnecessary for the mining task to be performed.

But in addition to irrelevance, there are other problems that affect the quality of the data. One of these problems is the presence of values that do not adjust to the general

behaviour of the data. These anomalous data can represent errors in the data or they can be correct values that are simply different from the others. Some techniques in the later phase of data mining ignore these data, but others are very sensitive, so the result is adversely affected by it. However, it is not always advisable to eliminate them, since in some occasions, such as the detection of fraudulent purchases, it can be more interesting than normal data. In these cases, it is necessary the interpretation of the one who carries out the study, evaluating the maintenance of the data or its elimination. Another problem to assess would be the presence of missing data, which can lead to inaccurate results. The selection of relevant attributes is an important procedure to avoid the generation of erroneous conclusions.

On the other hand, the type of data can also be modified to facilitate the use of techniques that require specific types of data. Thus, some attributes can be digitized, which reduces space and allows the use of numerical techniques or comparisons. The inverse process would be to discretize continuous attributes, i.e. to transform numerical values into discrete or nominal attributes.

3.4.1.3 Data mining phase

The data mining phase is the most characteristic of the process, and therefore, as previously mentioned, it is often confused with the entire process. The objective of this phase is to produce new knowledge that can be used by the user. This is done by building a model based on data collected for this purpose. The model consists of a description of the patterns and relationships between the data that can be used to make predictions, better understand the data or to explain past situations. It is necessary to make a series of decisions before starting the process:

- Determining which type of mining task is most appropriate, considering that some of the most relevant task have been presented before.

- Determining a model using the data mining techniques presented below.

- Choosing a mining algorithm that solves the task and gets the type of model that is desired.

3.4.1.3.1 Common techniques in data mining

Among the most common techniques in data analysis, we find the statistical techniques, used both for regression and for classification or grouping tasks. The most common method is Fisher's linear discriminant analysis.

On the other hand, decision trees are a group of conditions organized hierarchically, like a tree, giving its name. They are very useful to find structures in high dimensional spaces and in problems that mix categorical and numerical data. This technique is used in

classification, clustering and regression tasks. Decision trees used to predict categorical variables are referred to as classification trees, as they divide instances into classes. When decision trees are used to predict continuous variables they are known as regression variables. In the next image, there is a little example about buying or not a bicycle.
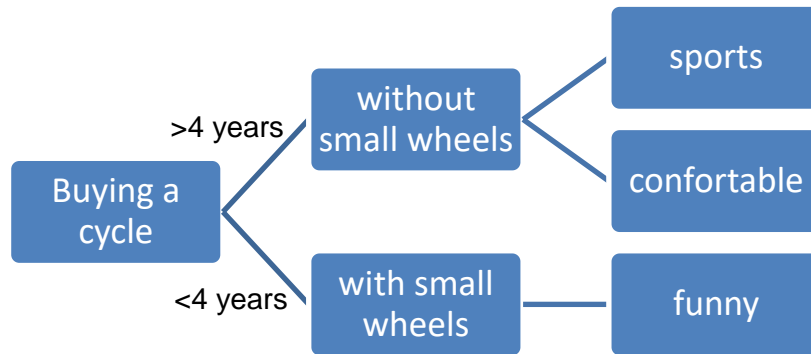
Figure 10. example of decision tree

3.4.1.4 Interpreting phase

In this phase the evaluation of the obtained data is considered, with the objective of converting them into knowledge. For this reason, the results obtained by applying the previously commented techniques are analysed and conclusions are drawn. In addition, it is in this phase that the expectations of the analysis proposed in the first phase must be met. By way of example, if in the first phase a prototype client has been considered, in this phase the results obtained will be studied in order to define it.

Likewise, the results obtained, and the quality of the established conclusions will be evaluated. In the most specific cases, the evaluation of the method is carried out through complex algorithms that relate test data with real data and establish an evaluation of the study.

3.4.1.5 Using and communicating phase.

Once the data have been analysed and the knowledge has been extracted, it is necessary for this knowledge to be extended among the interest groups for later use. For this reason, the results obtained are prepared so that they can be explained visually and their dissemination is as extensive as possible. In general, this knowledge should reach the entire analysis group involved and those interested in that knowledge. For example, in our case it must be guaranteed that the suppliers receive this information and that they can handle it and use it on future occasions.

3 Problem relevant method

In many cases, these data are monitored, that is, they are entered into the company's database so that they can be used in subsequent studies.

3.4.2 CRISP-DM

CRISP-DM provides a framework for carrying out data analysis projects, by defining a process model, and it aims to define knowledge discovery methodologies. Since it was conceived in 1996, most of the data analysis projects have followed this standard process model, because its large amount of advantages[54]. One of them is that it is independent of both the industry sector and the technology used, since it was designed to provide a generic model that could specialize according to the needs of any particular company or sector, allowing analysts to have a reasonable assurance that their efforts will be useful and valid. In addition, it allows to carry out data mining project faster, less costly, reliably and more manageable[55]. CRISP-DM consists of four levels, organized hierarchically into tasks ranging from the most general level to the most specific cases. At the most general level, the process is organized in six phases, which will be detailed below. In turn, each phase consists of general tasks, which describe the actions to be developed for each specific situation[56].
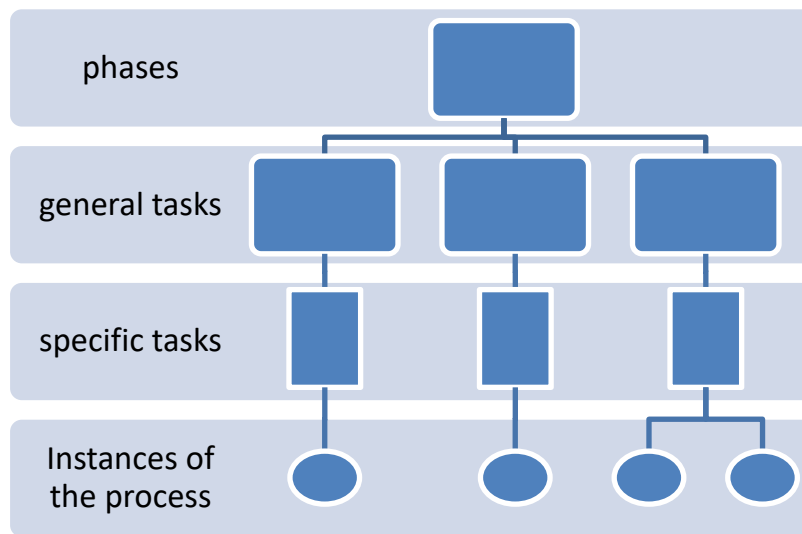

Figure 11. Ranking phases

---

[54] **Nuva, L., Schmitta, R., Briele, K.,** (Methodical data-driven integration of perceived quality into the product development process).

[55] **Wirdth, R**. (CRISP-DM: towards a standard process model for data mining)

[56] **Daza, A.** (Data mining. Minería de datos). Marcombo, s.a. 2017.

For example, if the general task is data cleaning, the third level will specify tasks such as: numerical data cleaning and/or categorical data cleaning. Finally, the fourth level determines the set of actions and decisions to be carried out to solve the tasks established in the previous level.

Similar to data mining, CRISP-DM methodology divides data projects in six phases, briefly outlined in the following[57]:

- Business understanding. Before starting with the data project, it is required to knowing the business perspective of the company and the project objectives. By this way, it is possible to work more efficiently and accordingly to the company principles.

- Data understanding. In this step, it is carried out the collection and familiarization with data, trying to discover quality problems, first insights into the data or detect hidden information.

- Data Preparation. This step intends to prepare the data that will be used in the modelling tools. Tasks include table, record, data cleaning, construction of new attributes, transformation, among others. The activities are not carried out in a prescribed order and they are usually performed multiple times.

- Modelling. Data mining techniques are selected and applied. Normally, this step requires the repetition of the last phase, since some techniques require specific data formats.

- Evaluation. In this step, the created model, with its steps, is evaluated to ensure the achievements of the business objectives and its quality.

- Deployment. The last step consists of the preparation of the knowledge gained so that they are understood by people outside the project, such as customers.

On the other hand, CRISP-DM is described in terms of a hierarchical process model, in which four levels of abstraction are defined, from general, covering all possible data mining situations, to specific, in which the data process is a record of actions, decisions and results.

---

[57] **Wirdth, R**. (CRISP-DM: towards a standard process model for data mining) pag 5-7

3.4.3 **SEMMA**

The SEMMA process corresponds to the acronym of Sample, Explore, Modify, Model, Asses, and refers to the process of conducting a data mining project[58]. It was created by the SAS Institute, who defined this methodology as process of selecting, exploring, and modelling large amounts of data to uncover unknown business patterns[59].

The phases involved in the SEMMA process can be described as:

- Sample. This stage intends to select the amount of analysed data to be analysed, remaining a sample big enough to contain a significant information, but yet small enough to manipulate quickly.

- Explore. In this stage unanticipated trends and anomalies are searched, to gain understanding and ideas. In addition, first associations and selections of variables are carried out.

- Modify. In this stage the main transformations of the data are developed, by creating, selecting and transforming variables to focus the model selection process. Filters to the data out of range and clusters are also relevant actions.

- Model. This stage models the data with the aid of statistical methods, such as regression, decision trees, etc. Thus, the software can search automatically for a combination of data that reliably predicts desired outcome.

- Assess. In this stage the usefulness and reliability of the findings from the data mining process is evaluated, to establish measures and its performance.

The main disadvantage of SEMMA is its linked to the SAS software. So that, it pretends to guide the user on the implementations of DM applications, although it is independent from Data mining chosen tool[60].

To sum up, SEMMA allows to automatize data mining processes, by creating and establishing a structure for their conception.

---

[58] **Azevedo, A., Santos, M.F**. (KDD, SEMMA and CRISP-DM: A parallel overview) page 2
[59] **León, E**. (Módulo Minería de Datos)
[60] **Azevedo, A., Santos, M.F.** (KDD, SEMMA and CRISP-DM: A parallel overview) page 3

3.5 Interim conclusion/reflection

In this section, four important aspects have been studied. First, the information exchanged between suppliers and manufacturing companies was studied. From this small analysis it can be concluded that suppliers do not reach the information regarding consumers. Considering the evolution of the automotive industry, where suppliers have become the new manufacturers, this information is indispensable for the improvement of vehicles aimed at consumers. It is therefore necessary to develop a methodology to channel the flow of information.

Secondly, the different sources of information used by manufacturing companies to obtain information from their clients have been defined. From this study, it can be observed that most of them are sources or processes that prepare them. Therefore, it can be concluded that these sources of information will be focused on obtaining knowledge for manufacturing companies and will not be so useful for suppliers. Two questions arise from this reflection that will be tried to be resolved in the development of the methodology: Would more sources of information about consumers be necessary? ¿How do customer data differ, if you do not produce the parts yourself?

The second important aspect studied in this section is consumer behaviour, defined as complex and changing. In addition, it has been observed that the definition of a consumer profile is very important for the establishment of competitive and marketing strategies of companies. Therefore, it would relevant to the assemblers to correctly analyse customer data in order to establish patterns of behaviour. That is why the answer to the question "How do I analyse customer data in order to fully define customer behaviour?" could be important to find.

Finally, the most common models or methodologies within data analysis science have been defined. This is due to the need to analyse customer information in order to obtain consumer information and convert it into vehicle improvement. However, the analysis should be focused on the use of the data by suppliers. Therefore, a new question arises: ¿How do you have to analyse the data if you want to use them for your supplier?

4 Methodological development

4.1 Justification of the development

As it has been seen in the development of this paper, there is a gap in the exchange of consumer information between suppliers and manufacturing companies in the automotive industry. However, it is not because such information is not obtained, but because the flow of information between the two parties is not adequate. In addition, the interruption in the flow means that, in line with the evolution of the automobile industry in which manufacturing companies have delegated production functions to their suppliers, consumer information does not reach the real automobile manufacturers. Therefore, the process of improvement based on customer needs is slowed down and is not completely focused on those needs. For this reason, the development of a methodology for analysing consumer data, encompassing all the phases of the process, from the time the information is collected until it is available to the interested supplier, is proposed below.

On the other hand, the rapid evolution of the industry has prevented studies focused on this field from developing at the same speed. Thus, in addition to the information flow vacuum, studies focused on the analysis of information in this sector have not been promoted, since, although there are some of them, they tend to be either too specific (focused on a specific company) or too general, without specifying the type of industry to which they refer. For this reason, there is a need to study the analysis of consumer data in the automobile industry, as proposed in the methodology to be presented.

First, the methodology focuses on both manufacturing companies and suppliers. This involves analysing and classifying the data so that both can obtain useful information from the consumer for their tasks.

In general, the methodology will consist in the creation of a data analysis process based on the previously mentioned techniques (KDD, CRISP-DM and SESSMA), and the analysis and interpretation of data will be deepened, with the objective of converting the data obtained from the consumer to a suitable and useful format, both for assemblers and suppliers. The aim is to achieve two different objectives. Firstly, it is intended that manufacturing companies take advantage of the information obtained from consumers to reinforce and validate their market strategies relating to consumer segmentation. In this way, they will be able to check whether their strategic orientation is correct or whether they must take corrective measures in this area. The second objective of this methodology is to facilitate the flow of information to suppliers by classifying the information obtained according to the suppliers to whom it is addressed, which are those who are really interested in the different parts of the vehicle. In order to consider these interests, it Will

be evaluated the strategy of modularization, previously commented, by which the manufacturing process of a vehicle is divided into different parts and then assembled in the manufacturing company.

4.2 Methodology choice

The first step in the development is the choice of a methodology from those presented above, or a mixture of them, which provides a basis on which to build the methodology itself. The processes presented in previous sections have been: KDD, CRISP-DM and SESSMA. The SESSMA methodology, although it could form a good basis for the process, lacks applications in today's industry. This fact results in a lack of information and knowledge of it, which makes it difficult to study. Thus, the fact that it is not used in the industry would make our methodology less attractive, due to the problems involved in its establishment, for example, the training from scratch of future analysts and those involved in the processes. Therefore, the SESSMA methodology will be discarded as the basis for our methodology.

The two remaining processes, KDD and CRISP-DM, are very similar, as both are closely related to the general science of data analysis. However, both lack some important phase for our methodology, so neither would completely cover the needs of our methodology. KDD lacks a previous phase of knowledge of the business, which leads to a start of the analysis without stated objectives and, in general, without a complete idea of the future development of the analysis and the company, a fact that is considered necessary in the proposed methodology. CRISP-DM is a more specific methodology of the data mining process. This fact can be observed because the last phase of the process is the analysis and preparation of the results, without posing a phase of presentation of the same or establishment of the general process.

Therefore, it is considered that the methodology to be developed will mix both processes; KDD and CRISP-DM, covering all the fields and objectives proposed. The "mix" methodology will include an initial phase of knowledge of the company and the process and a final phase of presentation of results and establishment of the process. In this way, a global methodology will be achieved that ranges from the knowledge phase of the process until the information is in the hands of the supplier, as desired

4.3 Methodology development

The methodology to be developed is proposed as a mixture of the processes of KDD and CRISP-DM, to cover all the phases that have been discussed previously, and thus meet the objectives of the work. In addition, the methodology will be iterative and interactive, following the characteristics of the KDD process, although the cyclical character will be reinforced in the Data mining phase, a concept explained in the following sections. In this way, it will include the following phases:
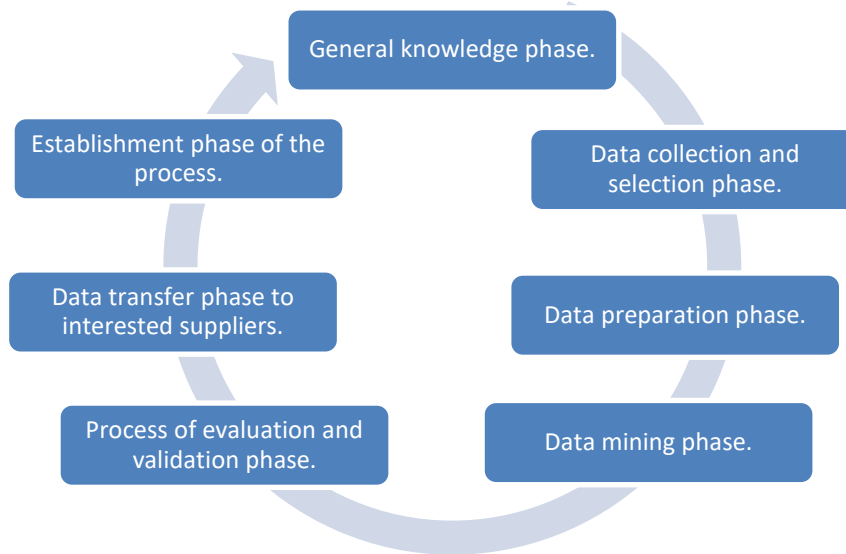
Figure 12. Phases of the methodology

4.3.1 Knowledge phase

This phase intends that the members of the group involved in the project have the necessary knowledge for its correct development. This involves training participants at three different levels, ranging from knowledge of the automotive industry to full knowledge of the project to be developed, through a training stage about the company itself. In this way, in a process of general knowledge, the members must know the evolution of the automobile industry, by which the manufacture of vehicles is carried out by the old suppliers. In a second level, the members must know in detail the operation of the company in which they are going to develop the project. With it, they will have to be informed about their suppliers, the methods of communication that they maintain, people involved in the processes of obtaining information of the consumers and of the people dedicated to the strategies of market, among other aspects. The last step in this phase would be the formation of the group directly related to the project. This is intended to know or establish main objectives, available resources, delivery dates, and all aspects necessary for completion. This will focus and guide the work to effectively meet these objectives, avoiding overwork or work that does not add value to the project.

4.3.2 Data collection and selection phase.

The final objective of this phase is the collection of consumer information, as much as possible to achieve the highest accuracy of the results. Assessing the sources of consumer information raised in previous sections, it will be decided which are carried out or from which sources it is decided to obtain the information, according to the objectives,

resources and dates raised in the previous phase. For example, if the time for the development of the project is reduced and personal resources are scarce, the search for information in forums and obtaining information from the consumer through questionnaires and/or simple tools will be promoted. On the opposite, if the time available is high and the human resources also allow it, it is proposed to carry out a working group.

4.3.3   Data preparation phase.

The data preparation phase is considered a tedious and important step in the process, since in this phase all the information is channelled with its different formats and possible invalid documents, and from it comes the information prepared to be studied in a simple way. The preparation of the obtained documents will be carried out, with several objectives. In the first place, the data will be cleaned, with which obsolete documents will be eliminated and information obtained with incomplete or erroneous data, among other deleted documents. Secondly, a process of homogenization of the data will be carried out, a phase that differs from the rest of the methodologies proposed to date, as will be seen below.

In this way, once the cleaning of the information has been completed, the homogenisation of the information will be carried out. According to the previous sections, it can be observed that there are many sources of consumer information. At the same time, these sources provide information in different formats and structures. This breadth of sources and formats makes it difficult to analyse the information, as no only one analysis technique can be applied to simplify the process. So that, as the first phase of the homogenization process, the grouping of information sources according to their characteristics is considered, for example, according to their level of structuring. In order to achieve this grouping, some of the most relevant characteristics of the information sources are presented below, in order to subsequently establish the values of each source for each characteristic and thus be able to classify and/or compare them.

- Format. It refers to the way with documents are presented.
- Degree of structuring. It depends on the format. A structured format would be a questionnaire with short answers, and, in the opposite, an unstructured format would correspond a text. The more structured, the easiest to extract useful information and the more automatized the process.
- Degree of specification. It refers to the level of detail with which the information is given.
- Scalability. It refers to the ability of extracted data to be added in a scale type.
- Costs. It refers to the economic efforts for the company that each source involves.

4 Methodological development

Sources of information	Format	Degree of structuring	Degree of specification	Scalability	Cost
Lead User workshop	Report	0%	50%	0	75%
Customer survey	numerical and short answers	100%	25%	50%	50%
Focus group	Report	0%	50%	0	50%
Expert interviews	Set of questions with different format	50%	75%	50%	50%
Complaints	Text	0%	0	0	50%
Social media and online reviews	Text	0	0	0	50%

Table 4. Characteristics and score for each source

 Firstly, Lead user workshop, complaints and social media and online reviews are presented by a text format, so the degree of structuring is null. In the opposite, customer survey has the total degree of structuring since it is formed by a set of questions answered by short or numerical answers, making its information easy to classify and analyze. Finally, results of an expert interviews are presented by a set of questions, which are usually a mix of open-ended questions and short-answered questions. By this way, it is in the middle of the structure's scale.

Secondly, degree of specification is also related with the usefulness of the given information. It means that a customer can explain its requirements in a detailed way but not explained in technological terms, which can be directly applied in the improvement process of the product. For this reason, the degree of specification usually depends on the person who elaborates the final document with the information. The more expert in the matter the person is, the more degree of specification reaches the final information. By this way the most specified degree corresponds to expert interview, since the whole process involves experts. Following it, Lead User workshop and focus group are in the middle, because the final report of the analysis is carried out by an expert, who applies his or her knowledge to interpret and reproduce the customer's requirements. Customer survey has a low degree of specification, since the preparation and elaboration of the questionnaire is developed by an expert, guiding the customer to specific answers. Finally, the degree of specification of complaints and social media and online reviews is zero, because of already explained reasons.

The classification according to scalability is easy to develop since only number-answered questions can be categorized in this group. So, complaints, social media and online reviews, lead user workshop and focus group have a null degree of scalability. On the other

hand, the scalability of both customer survey and expert interviews depends on the number of number-answered questions they have. Although that number can vary greatly, they usually present one part of its questions with this format.

The cost of each methodology is a factor to consider choosing one source or another. However, it is not such a relevant factor for the development of a consumer information analysis methodology, so it will not be given much importance in the following sections.

As it can be seen in the table, complaints and social and online reviews have the same score, so they will be treated as a single source, called complaints and reviews. Likewise, focus groups and Lead User Workshops have the same values, except for the cost that has already been commented as not relevant. Therefore, this pair of sources will also be treated as a single source, to simplify subsequent processes. In this way, when this phase begins, the first task to perform will be the grouping of these sources according to the following table:

Sources of information	Format	Degree of structuring	Degree of specification	Scalability	Cost
Customer groups	Report	0%	50%	0	75%
Customer survey	numerical and short answers	100%	25%	50%	50%
Focus group	Report	0%	50%	0	50%
Expert interviews	Set of questions with different format	50%	75%	50%	50%
Complaints and reviews	Text	0%	0	0	50%

Table 5. Sources grouped.

Once the sources are grouped, the homogenization process continues. This is where the proposed methodology differs greatly from the rest of the methodologies proposed to date, since the methodologies developed in this area involved a parallel analysis of the sources, it means that the information from each source was analysed separately and, once converted into a common format, they were added together for subsequent study. In this work, a cyclical analysis of the information is proposed, considering its degree of structuring, from less structuring to greater. The aim is to reduce the probability of overlooking relevant data, to prevent erroneous data not previously detected from continuing in the process and to establish relationships between components that might not otherwise be perceived, among others. In short, it is proposed as a cycle to achieve in-depth analysis. In this way, the analysis process would begin with the sources of information in which the data are generated with the least degree of structuring. In this

4 Methodological development

case, the methodology will begin with customer groups and complaints and reviews. After applying the appropriate tools, such as summarising or classifying, the information analysed will have the same format as the next less structured group, i.e. expert reviews. In this way, all complaints and reviews already analysed and expert reviews will be subjected to the corresponding analysis techniques, which for this group will be more specific. The result of the analysis will have a format like a questionnaire. Finally, the analysed data, which comes out of this second stage of analysis, will be introduced again together with the last group not analysed so far, the customer survey. The result of this last phase will compose the data that will be analysed to establish conclusions by the assemblers and suppliers. In this way, the result of the analysis will consist of a table of adjectives or characteristics of the vehicle, with a pointed numerical score. Thus, the personal data of each result will continue to be linked to these scores, without being submitted to any analysis tool, in order to be able to use these data for market segmentation.

In order to provide a more intuitive idea of the process, an outline of what each phase would be like with the sources that would pass each one and the results of this one is attached below.

Step in the cycle	
Group of documents to be analysed	Group analysed
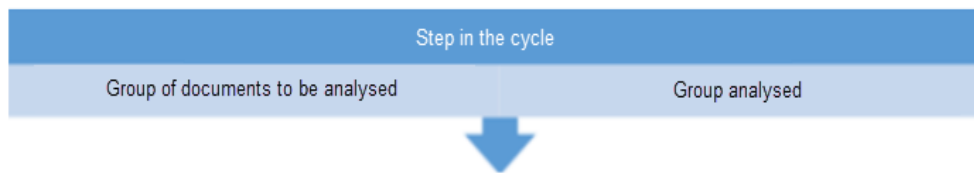
Figure 13. Structure of each phase

Step 1. From text to semi structured questions.	
Group 1: Complaints and reviews + Customer groups	Group 1 modified.

Step 2. From semi structured questions to structured questions	
Group 2: Group 1 modified + Expert interviews	Group 2 modified.

Step 3. From structured questions to unique format	
Group 2 modified + customer survey	Data to be analysed
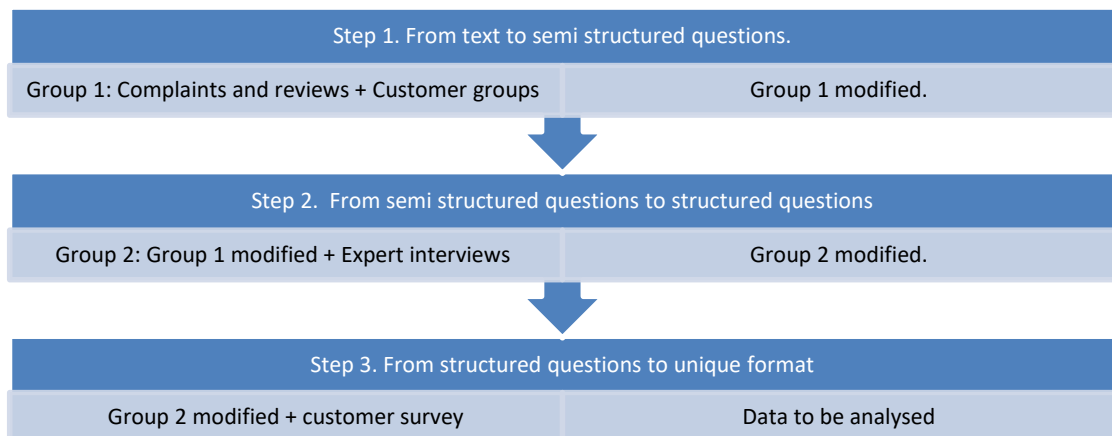
Figure 14. Scheme of each phase

Once that the information has been homogenised in a format, it is proposed its introduction in Excel, to facilitate the later representation of the results, by taking advantage of its tools. The template used will have two differentiated parts; personal data of the consumer and consumer needs. One of the aims to separate information is that personal information will be used by the assembly company only, while consumer preferences or needs will be useful to suppliers. Therefore, it is a way to facilitate the analysis carried out in the next phase. In addition, It is common that official questionnaires present that format.

4.3.4   Data mining phase

Once the data is unified in a single format, the data is analyzed. To do this, the analysis is divided into two parts: analysis for assemblers and analysis for suppliers.

4.3.4.1 Analysis for assemblers

In the first place, a data analysis will be carried out oriented to the manufacturing company involved in the study. As mentioned in previous sections, manufacturing companies carry out segmentation strategies according to various factors such as location, family life or age, among others. The objective of this segmentation is to improve advertising campaigns, achieving a better orientation of their products and thus achieving higher sales volume. For this reason, in the first analysis the personal data of the clients are analyzed to establish conclusions that serve as a guide for the company's own strategies. In general, the aim is to establish relationships between the personal data of consumers and the general characteristics of vehicles, which can form a purchasing pattern. This pattern, related to what has been said in the previous sections, is defined as the profile of the consumer, characterised by being complex and changing. This character makes it difficult to establish a consumer profile, which facilitates the process of improving vehicle models, aimed at established consumer profiles. In this way, this phase will try to establish, or re-establish, different patterns of consumers, which will serve as a basis for segmentation strategies. For example, in an age-based segmentation strategy it would try to visualise whether young people tend to buy cars of a more striking colour, or whether older people (50-60 years) prefer more neutral colours, such as matt grey or black. With this decision, the youth-oriented propaganda of the models would offer such a model in the color studied.

In order to carry out this analysis, the data mining tasks mentioned in previous sections and some of their models, such as decision trees, will be used. Thus, according to the premises commented on in previous sections, the consumer profile must be generic, must refer to general characteristics of the vehicle, must be linked to the personal data of the person surveyed and depends on various factors alien to the individual surveyed. In this way, it is concluded that, in order to know a consumer profile, personal data related

to the most general data of the vehicle must be analysed. It is possible to establish intervals that include ranges of values or criteria in order to include similar criteria in the same profile and thus achieve a generic profile. The analyses will always be carried out without considering data relating to the exact name or location of the dwelling, thus avoiding the publication of confidential data that could endanger the consumers surveyed. Finally, in order to assess the factors that influence consumer behaviour, a list of criteria will be added to analyse the social factors of each person; their professional situation and their personal situation.

In this way, certain characteristics of the vehicle will be graphically represented with respect to age and location, in order to be able to draw future conclusions. In this project these factors have been chosen to provide an idea of the methodology and the visualisation of the results, but the choice of these may vary according to the company's strategies.

The decision tree model, a tool presented in previous sections, will be used to establish the consumer profile.

4.3.4.2 Analysis for suppliers.

The second analysis will focus on suppliers, with the objective of analysing and transforming the data into useful information for them. As mentioned in previous sections, the manufacturing process of a vehicle is divided into modules, so each supplier is responsible for manufacturing and assembling each module and transporting it to the assembly company. In this way, the suppliers' interest is centred on the information relative to the module they manufacture. For this reason, it will be a question of classifying the data obtained by modules in order to be able to visualise the preferences of the consumers in each module. However, there may be cases in which the data are related, especially in cases of the electronics module, as there are multiple possibilities that raise the use of electronics in different modules: thermal sensors in the seats to modify the temperature, loudspeakers in the doors, and so on. For that reason, it will be proposed in later phases its presentation of joint form, all the modules in a same file, in the final phase.

4.3.5 **Data interpretation phase.**

Once the data have been analysed and represented, the experts in the project will carry out the analysis and interpretation of the results. Thus, the fulfilment of objectives proposed in the first phase of the process will be assessed, in terms of dates and in terms of quantity and quality of the information obtained. If this is not enough or if the quality is poor, it is necessary to analyse the phase in which the problem arose and repeat the process from there. Some of the problems to be assessed will be the possibility that the sources of information have been scarce, that in the data preparation phase they have

leaked badly and that erroneous or obsolete documents, etc. have continued in the process. If the origin phase of the problem is not found, the general process will be repeated from the first of the phases.

In addition, the analysis of the results is carried out with a practical approach. Since data analysis is required to validate the methodology, the necessary conclusions are drawn for the strategies. For this reason, in this phase the specialists in the departments of marketing, production, among others, value the results obtained in order to establish conclusions in the company. Therefore, modifications in market segmentation strategies and even in the company's general strategies will be evaluated.

This phase is carried out before the transfer of the results to the suppliers in order to guarantee the success of the project. Once the data is received by the suppliers, they will interpret the results in order to take advantage of the information obtained and to be able to take the appropriate measures in the production process. The reason for deriving its analysis is the greater use of information. This is because suppliers know their manufacturing process, their products and their capabilities. Therefore, they will be able to obtain more knowledge from the information from which the manufacturing company could draw.

4.3.6   **Data transfer phase.**

This phase is an important phase in the work carried out, as the flow of information with suppliers must be formalised. This project aims to ensure not only the transfer of information but also a continuous and updated transfer of information. The aim is to ensure that each time a consumer study is carried out or information is obtained that could be beneficial for the supplier, it can be channelled quickly and easily. In other words, it is proposed to create a channel of communication between manufacturing companies and suppliers that allows manufacturing companies to include the information obtained in consumer studies almost instantaneously. On the supplier side, this route should offer easy access. In addition, it must guarantee the correct visualization of the results obtained.

In this way, the communication channel must:

- Have easy access.
- Enable fast and efficient updating.
- Allow results to be displayed.

In order to achieve a channel with these characteristics, the communication channels proposed in the development of the work will be evaluated, in order to choose the most appropriate one for the objectives.

4.3.7   **Establishment of the methodology.**

As we have seen in previous sections, the profile of the consumer is not permanent, but changing with the life cycle of the products and changing with the type of product. From this characteristic of consumer behaviour arises the need for periodic consumer research. In this way, the establishment of this methodology in the company is considered necessary. Thus, the process would be streamlined and could be repeated in more occasions.

For this reason, the formation of a team that delves into this methodology in order to be able to study it and give it up is proposed.

5   Validation of the methodology

In order to validate the methodology, an example of a consumer study will be carried out in the automotive industry, deepening the phases of data mining and information transfer, as these are the objective phases of the work. In this way, all the phases of the methodology that would be carried out by a manufacturing company in the automotive industry will be worked on.

The first phase to be developed in the process will be the knowledge phase. In order to do this, the team involved in the consumer study will be informed about the automotive industry and the company in which the research is carried out. In addition, they will work on establishing the objectives, dates and resources available for the Project. In the following figure, the minimum information that those involved in the project should know at the end of this stage is attached.

General knowledge	•Evolution of the automotive industry •Modularization •General methods of communication with suppliers.
Knowledge of the company	•Current Suppliers •Methods of communication with them •Segmentation strategies •Consumer surveys conducted to date.
Process knowledge	•Establishment of objectives. •Personnel involved and the functions of each one •Available material resources

Figure 15. Minimum knowledge at the end of the first phase

Once the researchers know the automotive industry, the company and the objectives of the project, the information gathering phase begins. This phase would consist of searching as many sources as possible to obtain very precise results. However, due to the lack of available resources, for the validation of the methodology it is proposed to work on a single source of information; questionnaires. This choice is justified by its accessibility and reduced privacy. Thus, the participation of experts is unnecessary. In this case, a

5 Validation of the methodology

questionnaire template has been used, obtained from an investigation of the automobile market in Spain[61].

According to the proposed methodology, the data analysis process would begin with the least structured sources, so that they would be worked on until they became completely structured sources. It would be a question of applying the tools of summary, tokenization and grouping successively until having the correct format. In the second phase of this stage, the results of the first one would enter together with the source of information of a slightly higher degree of structuring. The tools of tokenization, grouping and classification would be applied again, until obtaining a document formed by a set of short questions. Again, this document would enter the last phase of the data preparation cycle, together with the previously commented questionnaire. In this last phase the scalability tools are applied, which would transform the nominal responses to numerical ones, on the same general scale. By way of example, the nominal questions with several options are broken down into three questions with the answer given in the statement. The following is a specific case of this transformation:

Which of the following options is best suited for your car seats?	
Leather ☐	Fabric ☒

Do you prefer the seats of your car made by fabric?	10
Do you prefer the seats of your car made by leather?	0
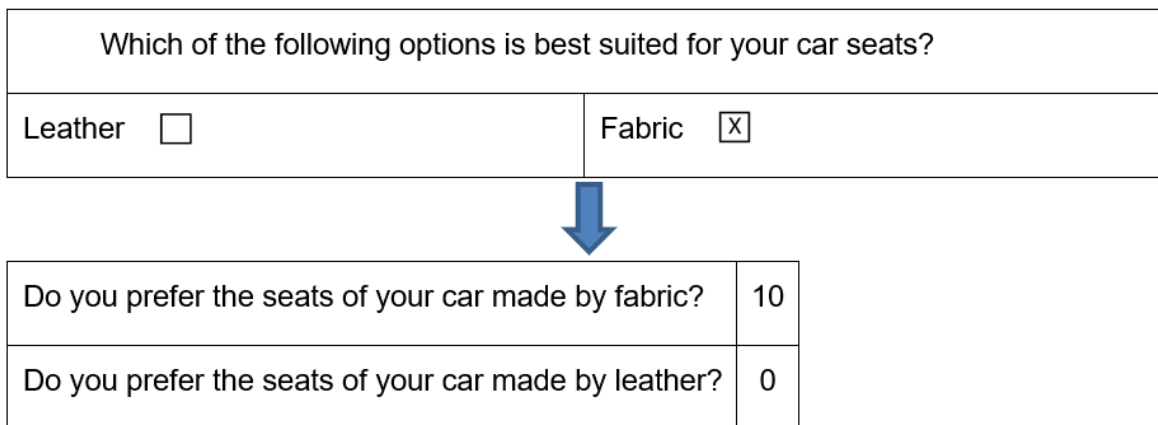
Figure 16. Example of transformation of questions

In this way, the data preparation process is completed with a final file that will compose a set of documents with a single format and a single scale. It should be noted that the personal information of the consumer interviewed will remain linked to the preferences or needs of the consumer in order to carry out the study of the consumer. If they were separated, the definition of the consumer profile would be impossible to study.

---

[61] **Rivero, L.**. (Factores de fidelización de clientes en la industria automovilística en España). Thesis de maestria. (2003), page 23

PERSONAL INFORMATION			
Gender	M	Age	34
Location			
Parking:	~~Yes~~ / No	City:	Zaragoza
Country:	Spain		
Personal Status			
Marital Status	single	Professional Status	Engineering
Nº of children	0		
HAVING A VEHICLE:			
Brand:	Opel Astra	Age of purhase	2016
Period to change:	5 years	Reasons to change:	Age of the car
Frecuency of use:	Each day		
Characteristics of actual car:			
Model:		Launch year	
Electronic devices		General information	
GPS	~~Yes~~ / No	Color	
Video	Yes / ~~No~~	Seat fabric	
Movil	Yes / ~~No~~	Automatic handbrake	Yes / ~~No~~
Radio	~~Yes~~ / No	Winter/summer wheels	~~Yes~~ / No
Speed monitoring	Yes / ~~No~~	Roof window	Yes / ~~No~~

CAR DESIRES			
Seat of Leather	10	Seats of Fabric:	0
número of seats:			
2 seats	0	4 seats	0
5 seats	10	>5 seats	0
Adjusting seat in:			
height	10	slope	10
Speakers included	10	no Speakers included	0
Heat control seat	10	No heat control seat:	0
Sound control by			
Voice	0	Bottons	10
Roof window	10		
Control Panel by			
Voice	10	Tactile	0
Normal bottons	0	No panel	0
Gearshift lever characterized by:			
Ergonomics		Appeareance	
Reverse gear			
By botton		By moving to a side	
Adjustable lights:	0	No Adjustable lights	10
Air conditioning	10	No air conditioning	0

Figure 17. Questionnaire

5 Validation of the methodology

As it can be seen, the personal data are visually separated from the customer's needs, but remain in the same file, for the reasons discussed above.

Once the data have been collected in this format, the data analysis phase begins. First, the analysis for the manufacturing company begins. The objective of this analysis will therefore be to draw conclusions regarding segmentation strategies and the establishment of the consumer profile. Based on the above, the analysis will provide a generic consumer profile, without personal data and involving some of the factors that affect consumer behaviour. The first tool used for this will be the classification, applied to vehicle models. In this way, the different commented characteristics will be classified and represented according to the vehicle model that the respondents possess. For this classification, an excel tool has been developed, which facilitates the tasks to be carried out. This tool consists of a table programmed so that the entered data are classified according to the desired criteria. In the following figure is attached an image of this tool together with an example of data classification for an age range between 18 and 39 years:

PERSONAL DATA		LOCATION			PERSONAL STATUS			VEHICLE					
Age	Gender	Parking	City	Country	Marital Status	Nº of children	Professional Status	Model	Frequency of use	Purchase year	Electronic devices	color	Seat fabric
55	F	Yes	Teruel	Spain	Married	3	High	High Range	Every day	2014	Medium	White	Leather
19	M	Yes	Zaragoza	Spain	Married	3	High	Ergonomic	Every day	2012	Medium	Black	Leather
38	M	Yes	Zaragoza	Spain	Married	0	High	Other	Every day	2015	High	Red	Fabric
29	F	No	Teruel	Spain	Married	0	Medium	Other	Once per month	2008	Basic	Silver	Fabric
22	M	No	Teruel	Spain	Single	0	Student	Other	Once per week	2010	Basic	White	Fabric
24	M	No	Huesca	Spain	Single	0	Student	Other	Frequently	2015	Basic	Silver	Fabric
46	F	No	Zaragoza	Spain	Married	2	Medium	Ergonomic	Every day	2016	Medium	Silver	Leather
58	M	Yes	Zaragoza	Spain	Single	0	High	High Range	Frequently	2017	High	White	Leather
28	F	Yes	Madrid	Spain	Single	0	Medium	High Range	Once per week	2018	High	Red	Leather

PERSONAL DATA		LOCATION			PERSONAL STATUS			VEHICLE					
Age	Gender	Parking	City	Country	Marital Status	Nº of children	Professional Status	Model	Frequency of use	Purchase year	Electronic devices	color	Seat fabric
19	M	Yes	Zaragoza	Spain	Married	3	High	Ergonomic	Every day	2012	Medium	Black	Leather
38	M	Yes	Zaragoza	Spain	Married	0	High	Other	Every day	2015	High	Red	Fabric
29	F	No	Teruel	Spain	Married	0	Medium	Other	Once per month	2008	Basic	Silver	Fabric
22	M	No	Teruel	Spain	Single	0	Student	Other	Once per week	2010	Basic	White	Fabric
24	M	No	Huesca	Spain	Single	0	Student	Other	Frequently	2015	Basic	Silver	Fabric
28	F	Yes	Madrid	Spain	Single	0	Medium	High Range	Once per week	2018	High	Red	Leather

Figure 18. Table excel tool

As it can be seen, this tool facilitates and simplifies the analysis process; once the data are collected and processed to the correct format, researchers will only have to enter them in the annotated table to view the results.

In order to continue with the analysis process, the previously commented data representation is now performed. To do this, a small program has been created, based on the table shown in the previous image, which allows to visualize the relationship between two criteria of the same one. As an example, the representation of the data of the consumers of high-end cars according to age and professional status is included below:

High Range - Age	
18-39	1
40-55	1
>55	2

High Range - Professional Status	
High	2
Medium	1
Low	0
Student	0

High Range - Age

■ 18-39 ■ 40-55 ■ >55

High Range - Professional Status

■ High ■ Medium ■ Low ■ Student
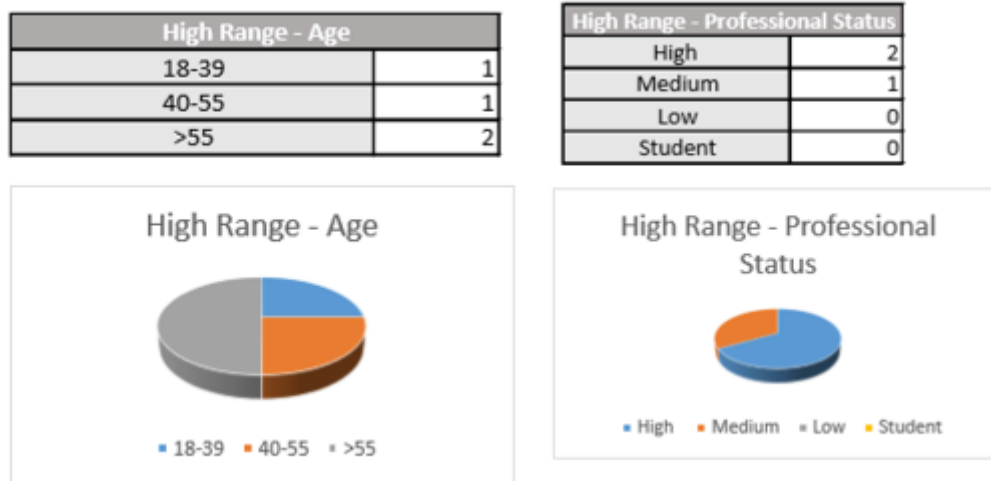
Figure 19. Data representation

Regarding the above representation, two aspects should be commented upon. First, this age range has been chosen according to a study at the University of Navarra, Spain, which created this general classification as the basis for beginning consumer studies. Secondly, the professional status has been classified according to high, medium, low and student, which would correspond to the salary associated with the professions of each consumer researched. Once again, this classification has been carried out to facilitate the analysis process and to include possible consumer profiles according to a shared criterion, the professional status.

As far as programming is concerned, this is a generic tool that can be easily modified to increase the range of analysis possibilities and thus include infinite variations that companies wish to observe. In order to modify it, it is necessary to consider three aspects:

- The number of clients surveyed added, since it will be necessary to expand the range of representation.
- The factors to be represented.
- The intervals for each factor to be evaluated in the representation.

Once the data have been represented, the decision tree, the data mining tool chosen for the validation of this methodology, is constructed. This tool is constructed taking as reference the previous representations, which will serve to define the main branches and the discarded options. This means that, if in the previous representation it is observed that a classification does not have significant representation, it can be eliminated from the options of the tree, in order to build a simpler model. In the previous representation it can be observed that top models only have consumers of medium and high social status, so that students and consumers of low social status will not be included in the creation of the consumer profile definition for top vehicles. In addition to the simplifications, the probabilities of following a branch of the tree according to the proportion represented are added.

5 Validation of the methodology

Again, in the previous representation, the probability that a consumer of high professional status will buy a high-end car is 66.6%.

To represent the tree in this methodology, the first box will begin with the "buy" option, as the main decision to be made will be whether to buy a vehicle. It should be mentioned that in this case it has been only considered the option "buy a high range car", in order to simplify the example, since otherwise the possibilities would be too many. Next, the first of the criteria to be studied is placed, which in our case corresponds to the age of the respondent, accompanied by the probability of occurrence of each criterion. The third criterion considered has been the professional status, just to show how the path is modified if the probability is cero. In our case, there was any customer between 40-55 years old with a low and Student status. So, both paths have been deleted. In general, this process is repeated as many times as criteria to be assessed. Once the tree is built, the path with the highest cumulative probability is studied, which will determine different consumer profiles. Next, the decision tree for the purchase of a high-end vehicle is represented, in order to facilitate the validation of the methodology.
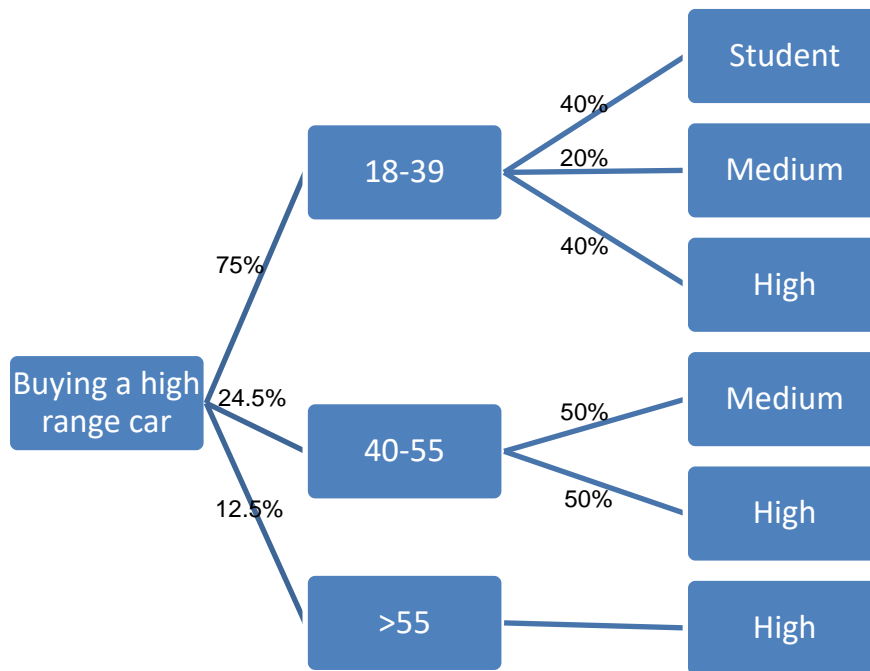

Figure 20. Decision tree for buying a high range car

.

From our data, the conclusion of the consumer profile for high-end vehicles would be young people, with a high social status or student. It can be observed that it does not fit particularly well with the general idea that was envisaged. This fact can be alluded to the scarcity of data for validation.

Once the analysis oriented to the manufacturing company has been carried out, the analysis oriented to the suppliers is carried out. As mentioned in previous sections, the

analysis will be based on the modularisation strategy, whereby suppliers are classified according to the module they manufacture. For this methodology, the first phase to be carried out is the reorganisation and classification of the data based on modularisation previously mentioned. To this end, a table like the previous one is constructed, but which includes the data relating to the needs of the consumer for each module discussed in previous sections, for which reason as many tables will be created as production modules. The initial idea was to include all this information in a single table, but the extension of the that table makes the work difficult instead of facilitating it. For this reason, it has been decided to build as many tables as there are suppliers. As an example, the table corresponding to the seating module is shown below:

SEATS OF		Nº OF SEATS				ADJUSTING SEATS IN		SPEAKERS INCLUDED		HEAT CONTROL SEAT	
LEATHER	FABRIC	2	4	5	>5	SLO	HEIGH	YE	N	YES	NO
10	0	0	10	0	0	10	10	10	0	10	0
10	0	0	0	10	0	10	10	10	0	10	0
0	10	0	0	10	0	10	10	10	0	0	10
10	0	0	10	0	0	10	10	0	10	0	10
10	0	0	0	0	10	10	10	10	0	0	10
0	10	0	10	0	0	10	10	10	0	10	0

Figure 21. Information classified for suppliers

As explained in the section on data preparation, only two figures appear; 10 and 0. A score of 10 means that the respondent prefers this option to the others. In this way, suppliers can observe consumer preferences and use them to modify their production process. Thus, taking advantage of the filters applied in the table, suppliers will be able to sort the table according to a preference of consumers. For example, they will be able to rearrange the table according to the consumers who prefer leather instead of fabric for seats. Thus, the programming tool used in the analysis of the manufacturing company is provided so that suppliers can carry out the same study as these; representation of data and realization of needs tree, but it does not go deeper into this area since it is not the object of the project. In addition, the joint representation of data allows suppliers to observe the existence of possible relationships with other modules, as discussed above.

Once the data mining stage is over, the results are interpreted. In this phase, the researchers assess the reliability of the results in order to decide whether to repeat the process or to transfer the data to the suppliers. In our validation, the data have led to consistent results in the consumer analysis part. Furthermore, no data gaps or exceptional cases indicating that the methodology has not been successful have been detected. Therefore, it is concluded that the research is valid and the data analysis by the manufacturing company is closed.

Finally, in previous sections it has discussed the most common methods of communication between assembly companies and their suppliers; via telephone, e-mail, face-to-face meetings and the supplier portal. The first two communication channels, although they

stand out for their practicality and speed, do not facilitate a constant flow of information. Moreover, the presentation of a data analysis is not feasible via telephone. For this reason, neither telephone nor e-mail will be evaluated in this methodology. Meetings are a practical method for presenting results and for their analysis and continuous improvement of the process. However formal meetings usually take place punctually and with suppliers considered important, such as suppliers of modules with a high productive load, for example. In addition, moving experts to suppliers would be time consuming and costly. Therefore, if only face-to-face meetings were valued for the flow of information, the improvement process would be greatly slowed down, and this information would never even reach the suppliers. Therefore, for this methodology, the supplier portal will be valued.

As already mentioned in previous sections, a supplier portal is an internal website, which can only be accessed by users of the supplier company and the manufacturing company. So the study of a supplier portal is complex. However, it is known that information relating to consumer needs, and therefore all the information studied in this project, is not shared in supplier portals. Therefore, a new information flow channel must be provided.

In order to achieve a channel with these characteristics, it is proposed to introduce a section on the Web intended solely for updating consumer information. It is valued the possibility of including it in the Web of each manufacturing company, so that a section appears in the Web of the company that carries out the study, where all the results of the studies are uploaded. However, access to this section must be restricted, as the information must only be visible to suppliers, researchers and personnel involved in the business of the manufacturing company, but not to the public or general clients. For this reason, placing this section on the generic Web of the companies would entail an extra job of generating users and passwords for all those involved. Therefore, it is considered easier to publish it in the Suppliers Portal. In this "miniWeb" both companies have access at present, so this inconvenience would not be a big problem. In addition, all users know and manage this portal, thus eliminating training times in this area. As a consequence, the insertion of a section dedicated to the publication of the results of the consumer study in the supplier portal is proposed in this methodology.

This section will also contain all the consumer studies carried out, all classified by date of completion.

5.1 Reflection of the methodology

5.1.1 Summary

Throughout this project several topics related to the production chain in the automotive industry have been dealt with, which could be classified in three main sections: supplier-manufacturer relationship, consumer behaviour and data analysis. In this way, after presenting the introduction, objectives and structure of the project, in the second point, a review of the existing literature has been presented in order to present the commented topics. This point started with the evolution of the automotive industry, to give a general idea of the situation. Next, the strategies of the manufacturing companies have been commented, deepening in the productive strategy of the modularization, given that it is the one developed now and around which the production is organized and, with it, the suppliers. Next, the evolution of suppliers and consumer behaviour has been commented on, given their importance for the establishment of market strategies for manufacturing companies. The last section of this literature review has been devoted to presenting the fundamentals of data analysis science, as they were necessary to understand the development of the methodology and the different tools used.

In the third point of the project, it has begun with the approach of the different problems to be solved with the development of the methodology. In this way, the void in the exchange of information between suppliers and manufacturing companies about the needs of the consumer has been reflected. Thus, the different sources of information most used by manufacturing companies to obtain consumer information have been described. The next point raised was consumer behaviour, explaining its importance and main characteristics. Finally, the main useful tools for data analysis were presented.

The last two points of the work have been the development and validation of the methodology, for which the different phases chosen for its creation have been created and explained. In the validation, the methodology developed has been applied to a small example of the automotive industry, to observe the possibilities and limitations of it.

To sum up, this project raises the situation in the automobile industry, from its evolution to the relations between suppliers and manufacturers. During the second section, the problem in the treatment of information is raised, which would justify the need for the prepared methodology. The work ends with this methodology developed and validated through an example.

5.1.2 Benefits of the methodology

The methodology presented in this project differs from the methodologies presented to date in several respects.

5 Validation of the methodology

Firstly, the number of phases that make up the general cycle of the methodology is greater. The aim was to offer a more complete cycle from the creation of the research team to the transfer of information. This last phase characterises the proposed methodology, as it is one of the least addressed phases to date. In addition, this transfer route has been chosen to guarantee the fluidity and permanence of the information exchanged.

On the other hand, the analysis of the data also makes a difference to the proposed methodology. As previously mentioned, a data preparation phase has been used, in which the information sources enter the process in an order established according to their degree of structuring. This fact allows the deepening of the analysed data and facilitates the detection of possible erroneous data, avoiding reaching erroneous conclusions.

Continuing in the phase of data analysis, it is worth mentioning the use of the excel tools and the decision tree. The first is characterized by providing clarity and easy visualization of the data to the methodology. Thus, a small programming has been carried out for the representation of the results, slightly explaining its modification to enlarge the range of study. It is this second factor of the tool that makes the proposed methodology stand out, since to date the published methodologies do not go so deep into the analysis of the data, but rather name possible examples that could be used, but do not detail or develop them.

On the other hand, another outstanding aspect of the methodology is the definition of a consumer profile. To date, the study of the consumer has not been deepened due to its difficulty and the changing aspect of it. For this reason, this methodology delves into this area, offering the possibility of greater depth.

For all these reasons, the aspects for which the methodology would stand out most are its easy understanding and application, its deepening in the field of data analysis and the transfer of information to suppliers in the automobile industry.

5.1.3 Objectives fulfilment

In order to verify that the methodology has been successful in terms of the proposed objectives, the questions raised in previous sections are commented on and answered.

The first one referred to the way of analysing the data in order to fully define the consumer's behaviour. As mentioned above, the analysis should focus on the consumer's personal data, relating them to generic data from the consumer's vehicle. Thus, in order to define the profile of the consumer, these relationships must be plotted on a decision tree, generating the route of greatest probability, which will be the profile of the consumer sought. The consumer profile can be created around any desired characteristic of the

consumer, depending on the market strategies of the automobile company. Thus, the process will have to be repeated periodically due to its complex and changing character.

Secondly, the question has been raised as to how the data differ if it is not produced by itself, i.e. if the analysis is carried out by the manufacturing companies. As can be seen, manufacturing companies are interested in information on general vehicle characteristics and personal information, as this is necessary for the establishment of a consumer profile. However, they are not involved in the search for more detailed information in each module of the vehicle, for example. Therefore, it is concluded that the data differ if it is produced or not in terms of depth of the information obtained, level of detail and technical specifications.

Another of the questions raised is the need to include more sources of information. The number of existing sources and methods of obtaining information means that the information obtained can be very broad and, in turn, very detailed. I believe that it is not necessary to include more sources than those valued in this project. However, I would raise the introduction of suppliers in the development of sources and methods, in order to obtain the data they need to improve their production process. Therefore, I conclude that it is not necessary to introduce new sources of information but the introduction of suppliers in the development of sources.

Finally, I consider that the question of how to analyse the data to give them to the suppliers has been resolved in the creation of the methodology, since it has been created for this purpose, among others. As a summary, data analysis must filter out erroneous data and classify and sort them according to the module to which they refer.

5.1.4   Critical reflection

The work carried out covers a wide range of aspects, from the automotive industry to data analysis. This made it difficult to compile the information, as the bibliography had to be consulted very differently.

On the other hand, it is a general process, which allows for a brief explanation of each proposed phase but prevents further study. For this reason, it is proposed as future work the deepening in some of these phases, proposing the main phase of data mining.

One limitation of the proposed methodology is that, when using simple tools, the amount of data to be worked on should not be too high. In the case of covering a very high quantity, the workload would be excessive, and computer tools designed for this purpose should be used.

Thus, after reviewing the bibliography, it can be concluded that both the study of consumer behaviour and the information exchanged by suppliers and manufacturing companies are two sectors with little research. For this reason, the study and research of these two factors is also proposed as future work.

5 Validation of the methodology

By way of conclusion, a general methodology has been developed for the processing of information, with the aim of converting it into useful knowledge and applying it to the improvement of the productive process of the automobile industry. This has implied the consideration of the evolution of this industry, to justify the difficulty of this process.

6   Bibliography

Hernández, J. (Introducción a la minería de datos). 2013

Daza, A., (Data mining: minería de datos). 2010

Velandia, A., López, W., (Investigación cualitativa y psicología del consumidor: alternativas de aplicación Avances en Psicología Latinoamericana), vol. 26, núm. 2, julio-diciembre, 2008, pp. 290-303

Álvarez, M.L., (Cambios en la industria automotriz frente a la globalización: El sector de autopartes en México". Revista Contaduría y Administración, No. 206, julio-septiembre). 2002

Hahn, C.,Duplaga, D., Hartley, J. (Supply-Chain Synchronization: Lessons from Hyundai Motor Company). 2000