**UNIVERSIDAD CATÓLICA DEL TÁCHIRA**

**SCHOOL OF ADMINISTRATION AND PUBLIC ACCOUNTING**

**FACULTY OF ECONOMIC AND SOCIAL SCIENCES**

**BACHELOR'S DEGREE IN BUSINESS ADMINISTRATION**

**MAJOR IN MARKETING**

MARKET SEGMENTATION IN SAN CRISTÓBAL MUNICIPALITY, TÁCHIRA STATE, FOR INTERNET SERVICES (CASE STUDY: SAN JUAN BAUTISTA PARISH)

Author: Alarcón Solórzano Luigui Santiago

Educational advisor: Mariela Alvarez de Montalvo

San Cristóbal, July 2023

# Index

[Index 1](#_Toc138156449)

[Introduction 2](#_Toc138156450)

[General Objective 4](#_Toc138156451)

[Specific Objectives 4](#_Toc138156452)

[CHAPTER I 8](#_Toc138156453)

[Determining the sample design for the study to operationalize market segmentation. 8](#_Toc138156454)

[1.1 BACKGROUND 8](#_Toc138156455)

[1.2 ANALYSIS AND DISCUSSION 10](#_Toc138156456)

[CHAPTER II 16](#_Toc138156457)

[Identifying market segments based on the data obtained from the sample in the San Juan Bautista parish of the San Cristóbal municipality. 16](#_Toc138156458)

[2.1 BACKGROUND 16](#_Toc138156459)

[2.2 METHODOLOGY 18](#_Toc138156460)

[2.4 RECOMMENDATIONS 24](#_Toc138156461)

[CHAPTER III 25](#_Toc138156462)

[Specifying the obtained segments and their characteristics for internet services in the San Juan Bautista parish in the city of San Cristóbal, Táchira State. 25](#_Toc138156463)

[3.1 ANALYSIS AND DISCUSSION 26](#_Toc138156464)

[3.2 RECOMENDATIONS 31](#_Toc138156465)

[REFERENCES 33](#_Toc138156466)

[Anexos 37](#_Toc138156467)

[Anexo. N.1 Encuesta a partir del el aplicado de manera digital 37](#_Toc138156468)

# Introduction

**T**he evolution of technology throughout our history has been marked by several key points that have allowed humanity to achieve the improvement in the quality of life that exists today. It is a fact how deeply immersed we are in the use of technology, especially with the existence of a particular tool that greatly enhanced the adaptation and evolution of human beings: the internet. This creation of such magnitude was unimaginable centuries ago, let alone the impact it has had. This tool has granted us swift interconnectivity with the world, propelling new levels of economic, cultural, and social globalization. There is no doubt that it is one of the most disruptive inventions for humanity, opening many doors that are still waiting to be explored.

On the other hand, the constant change in societies has driven companies to adapt their efforts to attract more customers and consequently gain more profits. In their early stages, reaching the market was generalized, with no way to discern audiences beyond the mediums through which the product or service was advertised. Similarly, companies did not face external variables that generated a need for them to evolve.

However, with the advent of the digital era, affordable tools began to emerge for any company to improve its ability to segment the population and optimize organizational efforts. This paradigm, which changed the way we thoroughly understand the market, is called segmentation. This tool is essential in marketing fields as it provides significant advantages to those who use it. Essentially, it carries out a systematic process that categorizes the market based on certain established parameters, depending on the product or service being sold or created.

Furthermore, the existence of automated systems and work schemes that implement this concept have facilitated greater proficiency in its application despite the increasing complexity of consumer habits, motivations, and desires. This complexity can be generated by any variable present in the surrounding environment. A recent example of this is the appearance of a new virus that plagued the world, known as COVID-19. Its emergence led to unquestionable changes in various aspects of daily life for the global population, subjected to different political measures implemented by each nation.

These circumstances forced us to adopt a new way of life, where the internet acquired a new level of relevance in work, education, economy, and society. Likewise, the way we buy and consume leisure elements such as audiovisual content, video games, or podcasts was inundated by the volume of usage, altering consumption patterns worldwide.

This research aims to satisfy the lack of reliable information regarding the Venezuelan population, specifically in the city of San Cristóbal, in order to identify the segmentation structure within this population. Additionally, due to the lack of veracity and reliability in information pertaining to the Venezuelan population, there is an increased ignorance regarding the identification of consumer and client needs. Therefore, this research seeks to provide academically relevant information, serving as a precedent for future investigations and providing a space to develop topics derived from this research. From a marketing standpoint, it will assist companies that offer internet services in the city to reach this market more efficiently, ultimately increasing their profits by expanding their coverage in the municipality.

Given the nature of the problem and the difficulty of obtaining accurate and meaningful data for the research, it is expected to obtain a result close to a representative sample of the population of the Táchira state, specifically delimited in its capital. From this sample, the segments that make up the population in relation to the service, as well as the needs associated with them, will be determined. It is also possible that valuable and beneficial results for future investigations will be achieved by the end of this research.

In order to address the research questions effectively, the following objectives have been formulated:

## General Objective

To establish a market segmentation for internet services in the city of San Cristóbal, Táchira state (Case Study: San Juan Bautista Parish).

## Specific Objectives

1. Determine the study sample design for the operationalization of market segmentation.
2. Identify market segments based on the data obtained from the sample in the San Juan Bautista parish of the San Cristóbal municipality.
3. Specify the obtained segments and their characteristics for internet services in the San Juan Bautista parish in the city of San Cristóbal, Táchira State.

Given the presence of quantitative and qualitative variables, it is necessary to employ a mixed approach, which authors Hernández Sampieri and Mendoza, 2008 (cited in Sampieri, R.H., Collado, C.F., and Lucio, M.P.B., 2014)[[1]](#footnote-1), defined as "...a set of systematic, empirical, and critical research processes that involve the collection and analysis of quantitative and qualitative data, as well as their integration and joint discussion, to make inferences based on all the gathered information."

To uncover structured patterns, interrelationships, distributions, and trends of the study variables, statistical tools will be utilized due to their significant importance, as described by Barreto Villanueva, A. (2012)[[2]](#footnote-2) "Statistics is the science whose objective is to gather quantitative information concerning individuals, groups, series of facts, etc., to deduce from it, through the analysis of this data, precise meanings or forecasts for the future."

Moreover, the nature of the study is exploratory, driven by the lack of knowledge about the segments that constitute the population of the San Juan Bautista parish in the municipality of San Cristóbal regarding internet service. In this situation of uncertainty, as mentioned by Sampieri, R.H., Collado, C.F., and Lucio, M.P.B. (2014)[[3]](#footnote-3), exploration is suitable "...when the objective is to examine a little-studied research topic or problem, about which there are many doubts or has not been addressed before."

The study area will exclusively focus on the population belonging to the San Juan Bautista parish in the San Cristóbal municipality, Táchira State, Venezuela. Given the objective of achieving segmentation within this population, it is considered appropriate to use a sample due to its specific size. A quota sampling method, which Malhotra, N.K. (2008)[[4]](#footnote-4) describes as a "non-probability sampling technique that involves a restricted judgmental sampling in two stages. The first stage involves developing control categories or quotas of elements from the population. In the second stage, sample elements are selected based on convenience or judgment." Additionally, any variable that may introduce biases, such as the quantity of elements from each stratum of the sample, will be considered.

To address the geography that constitutes the San Juan Bautista parish, located in the Bolivarian Republic of Venezuela in Táchira State, in the city of San Cristóbal, the capital of both the municipality and the state, reference can be made to Moreno A. Diego & A Méndez R. Maury E. (n.d)[[5]](#footnote-5) in their urban proposal for this parish (see Figure N.1). They used the Official Gazette of Táchira State No. 444 from 1998 to explain the geographical boundaries.

TO THE NORTH: It borders Independencia Municipality along Quebrada La Laja, downstream from the bridge on the San Cristóbal-Capacho road to a site called Mata e' Guadua, until it reaches its confluence with Quebrada Zorca. It continues along the old royal road, now a branch road, until a point in front of the Church of Providencia, with UTM coordinates N. 862,950m, E. 803,215m. It then follows the old royal road from San Cristóbal to Capacho Nuevo until the confluence point of Quebrada La Parada with the Torbes River. It continues along the Torbes River upstream to the mouth of Quebrada La Machirí and follows it upstream to its head in Páramo El Oso.

TO THE SOUTH: It borders Pedro María Morantes Parish from the previous point. Starting from the head of La Machirí, it follows the watershed to the west until reaching the head of Quebrada La Parada. It then follows the downstream course of Quebrada La Parada to Carrera 11, then heads south to its intersection with Calle 9. From there, it continues westward, now bordering San Sebastián Parish, along Calle 9 and its extension until it reaches the Torbes River. From the Torbes River, it goes in a straight line to the west until reaching Hito Tononó, and then follows the ridge to the Hito de la Laguna La Linda.

TO THE WEST: It borders Independencia Municipality starting from Hito La Laguna La Linda, following a straight line in a northeast direction to the highest point of Cerro El Bolón, and then continues in another straight line in a northwest direction to the curve where the sub-branch road leading to the hamlet of Valle Verde is located. The coordinates of this point are UTM N. 862,290m, E. 802,180m. It then follows the main road of Valle, via Capacho, until its intersection with the San Cristóbal-Capacho road. It continues along this road towards Capacho until the bridge over Quebrada La Laja in Mata e' Guadua, which is the starting point.

TO THE EAST: It borders the elevation 956 of Chorro del Indio Park, from the northern boundary to the southern boundary, as previously indicated in the Mata e' Guadua sector.

On the other hand, it is worth noting that the handling of these provided data may contain sensitive content, which will adhere to the provisions stated in Article 60 of the Constitution of the Bolivarian Republic of Venezuela[[6]](#footnote-6), which states that:

Every person has the right to the protection of their honor, private life, intimacy, personal image, confidentiality, and reputation. The law shall limit the use of information technology to guarantee the honor and personal and family privacy of citizens and ensure the full exercise of their rights. (BOLIVARIAN REPUBLIC OF VENEZUELA, 1999, Art. 60)

Throughout the development of this research, compliance with the constitutional basis that generally protects information of any kind in our nation will always be ensured.

Once the nature of the investigation has been determined, instruments such as the survey of closed-ended questions defined by Sampieri, R.H., Collado, C.F., and Lucio, M.P.B. (2014) will be used.

Once the nature of the research has been determined, instruments such as closed-ended question surveys, defined by Sampieri, R.H., Collado, C.F., and Lucio, M.P.B. (2014)[[7]](#footnote-7) , will be used. They are described as "those that contain pre-defined response options. They are easier to code and analyze." These include dichotomous questions, Likert scales, or single/multiple-choice options. Open-ended questions, also defined by Sampieri, R.H., Collado, C.F., and Lucio, M.P.B. (2014)[[8]](#footnote-8), will also be used. They are described as "not delimiting the response alternatives. They are useful when there is not enough information about people's possible answers." By utilizing the advantages of both instruments, this research aims to gather fruitful information.

# CHAPTER I

### Determining the sample design for the study to operationalize market segmentation.

## BACKGROUND

As an initial background, an international study conducted by Robin, C. F., & Torres, C. A. (2001)[[9]](#footnote-9), titled "Market segmentation: seeking the correlation between psychological and demographic variables," they found difficulties in generating specific segments if analysis is not designed between demographic variables and behavioral segmentation variables. This procedure was developed with the main objective of proposing a low-cost and highly flexible methodology that allows obtaining useful segment profiles for commercial decision-making and the design of marketing mixes tailored to consumer requirements and characteristics.

As a second international background, Rodríguez, N., Rodríguez, V. A., Ramírez, E., Cediel, S., Gil, F., & Rondón, M. A. (2016)[[10]](#footnote-10) conducted an article titled "Methodological Aspects of Sample Design for the National Mental Health Survey" for the Colombian Journal of Psychiatry. In their article, they provided a summary of the methodology used in the 2015 National Mental Health Survey. They considered proportions from previous studies for determining the sample size. Similarly, aligning with the survey objectives, they divided the population into 4 age groups based on national and international institutions to obtain more detailed information.

Clearly, the methodology used for the survey provides a framework in terms of dividing the population to be studied, as well as gathering past information to support the concepts, proportions, and methods used in the study of the sample and its design for the facilitation of information gathering.

Finally, in the background of the research work for the Multiciencias Journal conducted by Andrade, N., Morales, M., & Andrade, C. (2008)[[11]](#footnote-11) titled "Market Segmentation as a Basis for the Positioning of Fast Food Franchises in the Maracaibo Municipality," the authors explain the objective of the market segmentation process as a basis for the positioning of fast food franchises in the Maracaibo municipality. This was achieved through a descriptive non-experimental research using a semi-structured interview applied to highly relevant franchises registered with the Venezuelan Franchise Chamber (Profranquicias).

This research provides a guide for the collection of relevant data, as well as the selection of the population, through national institutional databases. It also analyzes variables of different kinds to generate a more defined segmentation, such as geographic, demographic, pictographic, and behavioral variables. This will be of great utility in this present investigation to deeply understand the different dimensions of responses and analysis, leading to a better construction of a collection method for an ideal segment formation.

## ANALYSIS AND DISCUSSION

Taking into account the time and resource limitations for conducting this study, the quota sampling technique was selected (Item 2), which Malhotra, N.K. (2008) describes as a "non-probabilistic sampling technique, which is a two-stage restricted judgment sampling. The first stage involves developing control categories or quotas for elements of the population. In the second stage, sample elements are selected based on convenience or judgment."[[12]](#footnote-12)

Considering the procedure of this sampling technique, the corresponding stages were carried out. Firstly, the control category was generated, which was age. During this process of dividing the quotas, divergences in concepts regarding youth, adults, and older adults were found, creating difficulties in finding a standardized way to construct these divisions. To address this issue, the conceptual basis was taken from one of the world's most important organizations, the United Nations (UN). The concepts of youth and older adults provided a basis for generating two quotas, namely 15-24 years[[13]](#footnote-13) (referred to as youth) and 60 years or older[[14]](#footnote-14) (referred to as older adults).

Regarding the age range of 0-14 years, it was decided to discard it due to their low academic and psychological development in relation to the instrument used in the study. Lastly, the age range of 25-59 years was arbitrarily chosen due to the lack of a standard classification by the UN and other internationally recognized organizations such as ISO. Consequently, this interval was referred to as adults.

For the second stage of this sampling technique, it is necessary to know in advance the number of elements to be included in the sample. The information from the 2011 census provided by the National Institute of Statistics (INE) was used for this purpose. Despite being 11 years old, it is the most recent census available in the country. The INE census data includes population growth projections by parishes.

However, the projections are presented in five-year intervals, which makes them less useful and accurate. Therefore, the population projection for the San Cristóbal municipality in 2022 was used, as it contains the San Juan Bautista parish, which is part of the target population (Item 4). Based on this data, the population of the San Juan Bautista parish was estimated to be 292,556 individuals.[[15]](#footnote-15)

To obtain an approximation of the population of the San Juan Bautista parish in 2022, an average proportion was calculated based on the proportion of the parish's population compared to the total population of the municipality. For this calculation, data from population projections by federal entity, municipalities, and parishes from 2000 to 2050[[16]](#footnote-16) were used, considering only three years: 2000, 2005, and 2010. This calculation resulted in a proportion of approximately 34%. Multiplying this proportion by the known total number of elements in the San Cristóbal municipality, the estimated population of the San Juan Bautista parish for 2022 was determined to be 99,484 individuals. Similar calculations were performed to obtain the proportion for each age range to have a more detailed view for sample calculation.

In these circumstances, the INE divided the inhabitants of the parish into five-year intervals, which aligned with the previously generated control categories. This division allowed for obtaining proportions within the group of individuals of interest, resulting in 18.07% for 15-24 years, 49.85% for 25-59 years, and 12.32% for 60 and older.

In these circumstances, the INE divided the inhabitants of the parish[[17]](#footnote-17) into five-year intervals, which aligned with the previously generated control categories. This division allowed for obtaining proportions within the group of individuals of interest, resulting in 18.07% for 15-24 years, 49.85% for 25-59 years, and 12.32% for 60 and older.

Finally, the adapted population to the control category was calculated by multiplying the 80.24% formed by the previously presented percentages by the 99,484 individuals, resulting in a population of approximately 79,825 individuals (Item 3). This falls under the concept of a finite population, as López-Roldán, P., & Fachelli, S. (2015) state in their book: "A finite population refers to a population consisting of less than 100,000 units, while an infinite population refers to a population with 100,000 or more units."[[18]](#footnote-18)

Considering the finite population condition and the lack of knowledge about statistics such as the mean or standard deviation, the most suitable formula for calculating the sample size was chosen, which is the formula for population proportion for finite populations. López-Roldán, P., and Fachelli, S. provide conventions for the values of the formula's elements, such as a confidence level of 95% (Item 5) and proportions with a value of 50% [[19]](#footnote-19).These conventions were used for standardization purposes and to maximize data collection.

Having determined that the approximate population for the parish falls within the age range of 15 to 60 years or older, the sample size calculation for finite population with population proportion was carried out. The formula applied in this calculation is used for research where the main variable is quantitative:

Having discovered that the approximate population for the parish falls within the age range of 15 to 60 years or older, the calculation of the sample size for a finite population with population proportion was determined. The formula applied in this calculation is used for research where the main variable is quantitative, as mentioned by López-Roldán, P., & Fachelli, S. (2015)[[20]](#footnote-20) :

* N = Population size
* z = Number of standard deviation units indicating the adopted confidence level
* n = Sample size
* P = Proportion (or percentage) of individuals who have a certain characteristic.
* Q = Proportion (or percentage) of individuals who do not have a certain characteristic.
* e = Considered sampling error[[21]](#footnote-21)

When considering the availability of resources and time, conducting the survey on 383 study subjects proves to be highly complex, along with the need to obtain information and data within a limited timeframe. As a solution, the Pareto principle was adopted, discovered by the Italian economist and sociologist Vilfredo Pareto, who observed that 80% of the land or properties in Italy belonged to 20% of the population.[[22]](#footnote-22)

This main idea laid the foundations for the generation of other applications in different disciplines, as Borjas, C. M. B. (2005) did in their research, where they conceptualized this principle as "20% of the population is responsible for 80% of the problems." If we adapt this concept to this context, we can say that 20% of the sample size provides 80% of the information, resulting in a total of 76.6, which is approximately 77 respondents who theoretically will provide 80% of the information.

# CHAPTER II

### Identifying market segments based on the data obtained from the sample in the San Juan Bautista parish of the San Cristóbal municipality.

## BACKGROUND

Just like in the previous chapter, the research conducted by Robin, C. F., & Torres, C. A. (2001)[[23]](#footnote-23) used two techniques for the analysis of results to carry out homogeneous groupings, namely Hierarchical Clustering and K-means clustering, which provided the ideal number of segments to be formed in their study. Therefore, the importance of these techniques is crucial for the present research in the sample under study, applying the ones mentioned for their effectiveness in the work called " Segmentación de mercados: buscando la correlación entre variables sicológicas y demográficas".

Similarly, in his research, Raiter, O. (2021)[[24]](#footnote-24), in his article "Segmentation of Bank Consumers for Artificial Intelligence Marketing", Riter uses cluster analysis with the K-means algorithm to obtain the ideal number of segments in his segmentation, verifying the suitability of the size through calculations such as Average Silhouette Score and the Elbow method.

Additionally, he conducted reviews on the distribution structure of the study variables to observe the behavior and context they provide. Finally, he used correlation graphs to observe the level of relationship between the variables in order to identify possible characteristics that have a sufficiently significant interaction to consider when generating the segments.

Last but not least, there is the scientific research called "La Segmentación De La Demanda Turística Española" by Meiriño, R. C., Brea, J. A. F., Vila, N. A., & López, E. R. (2016)[[25]](#footnote-25). Among their objectives, they included a review of bibliography related to the object of study, detailing the methodology used in sources conducted in the Spanish nation, thus contrasting the shortcomings and deficiencies of these sources. Therefore, in the development of this work, they present the procedure and analysis for the correct identification of the different segments.

This work explains in an organized manner while presenting the procedure and results of an analysis of a real case, emphasizing the importance of analyzing the data beforehand to detect multicollinearity among variables using Cramer's V (for categorical variables), Spearman's correlation (for ordinal variables), and Eta tests (for both types of variables). Subsequently, as a complementary step, they use the statistical package SPSS to obtain the number of clusters or segments, validating this calculation with subsamples.

## METHODOLOGY

In order to collect the sample data, a survey with questions of different nature was conducted, complying with all dimensions and indicators. For this task, Typeform was used due to its easy-to-use and customizable platform, which allows creating interactive surveys with various question types such as open-ended questions, multiple-choice questions, and single-choice questions, among others. Additionally, Typeform offers an attractive and visually pleasing design, which can help improve the user experience and, therefore, increase the response rate.

Another advantage of Typeform is its ability to integrate surveys with other data analysis tools, such as Google Analytics or Excel, which provides greater ease for analysis purposes.

Once the survey was designed, the link was distributed through various channels, including email and social media platforms like WhatsApp and Instagram, to individuals who met the inclusion criteria for the sample. To ensure the quality of the collected data, assistance was provided in completing the survey in most cases, aiming to prevent misinterpretations, errors, or omissions by the respondents.

After completing the number of responses, the data was downloaded in an Excel file provided by the platform, which facilitated data cleaning. This included the removal of irrelevant columns, changing the format of some responses to a uniform format (removing any symbols or special characters that could interfere with data interpretation), and checking the consistency of the data. Records that did not meet the inclusion criteria in the sample, such as cases with missing data, were eliminated.

Following the pre-cleaning phase, the responses were encoded and further cleaned in the Python programming language to reduce any format-related biases. Following the recommendations from previous studies, the compliance of the proportions of each sample quota according to age ranges was verified. To achieve this, calculations were made based on data extracted from the National Institute of Statistics (INE) regarding the population distribution in the study area. The proportion of the population aged 15 to 24 years was found to be 18.07%, the proportion of those aged 25 to 59 years was 49.85%, and the proportion of those aged 60 years and above was 12.32%. Additionally, there was a nearly equal gender proportion of around 50% for each sex.

In this way, a representative sample of the population in the study area was obtained, and potential biases in the analysis results were avoided. However, it should be noted that the sampling technique used introduces biases since the respondents were selected based on criteria or convenience. Nevertheless, the sampling process was carried out rigorously to obtain results that are as close to reality as possible and to fulfill the exploratory nature of the present study.

To begin with the selection of segments, Principal Component Analysis (PCA) was applied to the collected data. PCA is a statistical method used to reduce the complexity of a dataset. In the case of market segmentation, the objective of applying PCA is to reduce the number of variables analyzed in the subsequent step and, thus, simplify the analysis performed by the K-means algorithm.

To start, the necessary libraries were imported in Python to perform PCA transformation. Then, categorical values were encoded into numerical variables through sequential labeling to process them with the PCA algorithm. In other cases, new columns were created to simplify the information and add more value to the analysis. These columns included:

* Price Matches: It was the combination of the question "From the following ranges, which one matches the budget you allocate for the service?" and the question "According to you, what is the ideal price range for the service?" If the responses to these questions matched, a value of “Match” was assigned. Otherwise, “Not Match” was entered.
* Best: This column was derived from the question "Rate the following characteristics of the Internet service at the time of acquisition based on their importance to you," where the characteristic in the first place was extracted as the most important.
* Best2: Using the same basis as the previous column, the second most important characteristic was extracted.
* Least: Similarly, using the question "Rate the following characteristics of the Internet service at the time of acquisition based on their importance to you," the least important characteristic for the subject was extracted.

It is worth noting that the coding system used is not optimal for the study because it implicitly gives more importance to higher values and vice versa. Additionally, the PCA transformation can assign undesired values due to scale variation.

* 1. **ANALYSIS AND DISCUSSION**

After completing these preliminary steps, PCA transformation was applied to the 40 columns of data. The result of this transformation was a series of principal components (PCs), which are linear combinations of the original variables used to represent the data in a lower-dimensional space, i.e., with fewer variables. To determine the ideal number of principal components, the cumulative sum of the percentage of variance explained by the PCs and the corresponding graph were used as tools.

In the graph, the horizontal axis represents the number of principal components, while the vertical axis represents the cumulative sum of the percentage of variance explained by the PCs. The graph shows an increasing curve, indicating that each PC explains an increasingly larger amount of the total variance. The number of principal components is selected at the point where the curve begins to flatten, and the percentage of variance explained by the PCs meets a certain threshold.

In my analysis, the graph of explained variance showed that the first 6 PCs explained most of the total variance in the data, with a cumulative percentage of 95.32%. This decision was made because selecting 19 PCs was necessary to achieve a 99% explained variance.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

After performing the PCA process, market segmentation was carried out using the K-means algorithm. This algorithm aims to group the data into k homogeneous clusters, where the distance between data points within the same cluster is minimized, and the distance between data points from different clusters is maximized.

The segmentation process was initiated by iteratively running the K-means algorithm with different values of k, and the silhouette score was used to evaluate the quality of the generated clusters. The silhouette score values were plotted against the number of clusters, which allowed identifying the maximum silhouette score value with the fewest number of clusters. In this case, it was found that the optimal number of clusters was 3, with a silhouette score of 0.3745026715440567.

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Additionally, to verify the optimal number of clusters based on the research methodology of Raiter, O., an additional review was conducted using the elbow method. This method also involves evaluating different values of k and observing the inflection point on the graph).

Gráfico, Gráfico de líneas

Descripción generada automáticamente

The inflection point represents the optimal number of clusters. In this case, it was observed that the optimal number of clusters was 5, but this was not chosen due to the loss of silhouette coefficient and the avoidance of overfitting of segments with a high number of clusters due to the low number of respondents. Therefore, 3 was chosen in this case, which coincided with the result obtained through the silhouette score, concluding the market segments for this research with this value.

## RECOMMENDATIONS

During the market research process, crafting precise and well-designed questions is a critical step for the success of the survey. It is essential for the surveyor to put themselves in the respondent's shoes and consider how the questions and potential answers could be interpreted. In some cases, creating a prototype and conducting tests with respondents can be useful to ensure that the questions are clear and effective.

Additionally, it is important to make the survey design attractive and easy to read. The presentation of the survey can influence the respondent's motivation to complete it and can impact the quality of the obtained responses. Providing support and assistance to respondents during the survey completion process is also necessary to minimize misinterpretations, errors, or omissions.

Furthermore, it is crucial to select a representative sample of the population to ensure the validity of the results. Using data from the corresponding national institute is recommended to ensure that the sample closely reflects the reality of the population.

Regarding the segmentation process, it is important to consider the use of different methods based on the available data and research objectives. In some cases, methods like PCA and K-means can be helpful and provide satisfactory results. However, it is also important to consider other methodologies that may be more reliable and accurate in obtaining the desired results, considering scaling and coding techniques in each method.

# CHAPTER III

### Specifying the obtained segments and their characteristics for internet services in the San Juan Bautista parish in the city of San Cristóbal, Táchira State.

* 1. ANALYSIS AND DISCUSSION

In the context of the constant evolution of the market and the increase in competition, having updated and accurate information about customers' needs and preferences is crucial for the survival and success of any company.

In this sense, the proposal to identify market segments for internet services in the San Juan Bautista parish in the city of San Cristóbal, Táchira state, using the K-means algorithm, represents a technological and creative opportunity for internet service organizations in the area. This would allow them to have a better understanding of their audience to design effective and personalized marketing strategies that meet the needs and desires of their potential customers. This, in turn, will enable them to offer high-quality services tailored to the demands of the local market.

Therefore, to offer these tailored services, the following specified profiles were generated for each segment obtained in this exploratory research:

In the first market segment, an age distribution is observed, where more than half of the individuals are between 25 and 59 years old, and the remaining age ranges have similar proportions. There is a slight difference in favor of male individuals in terms of their proportion. The size of their households, including the respondents, is concentrated in a total of two members, accounting for 40%, while the remaining 60% consists of three or more members. The main occupations present are related to sales, teachers, and retirees. As for the most important characteristics for this segment, speed and uninterrupted signal service are valued, while payment methods, support, and technical service are considered less relevant.

Regarding internet usage time, 86.66% of individuals use the internet for at least one hour, with 53.33% specifically using it between 1 and 3 hours. The ideal price for 73.33% of this segment is between 10 and 30 US dollars, while the rest prefer a price below 10 US dollars. It is important to highlight that 60% indicated that they would continue to use the service even if the price increased.

In terms of internet usage habits, 53.33% of individuals do not use the internet to obtain news, which is not a significant difference. Furthermore, 85.67% do not use the internet for video games, while 60% use it for educational purposes and 66.67% for work-related activities.

In reference to the way individuals gather information about internet services, 53.33% do so through the internet, and 60% use social media as a means of obtaining information about internet services. In 73.33% of cases, the budget allocated for the service coincides with the aforementioned ideal price.

The characteristics that define the second segment are based on a large portion of individuals falling within the age range of 25 to 59 years, specifically representing 72.73%. On the other hand, individuals in the age range of 15 to 24 years constitute a minority in this segment. It is worth noting that more than 60% of individuals are female. The size of their households, including the respondents, is concentrated in three or more members, accounting for 77.28%. The most common occupations in this segment are students, traders, and homemakers. The most important characteristics for this group are low and affordable price, as well as uninterrupted signal service. On the contrary, support, technical service, and payment methods are considered less relevant features.

Regarding internet usage time, 90.91% of respondents use the internet for at least one hour daily, with 54.5% of them using it between 1 and 3 hours per day. However, regarding the ideal price, there is a disparity, as 54.55% consider a range of 10 to 30 US dollars to be ideal, while 45.45% prefer a price below 10 US dollars. Interestingly, 68.18% of respondents indicated that they would continue to use the service even if there was an increase in the current price.

Regarding specific internet usage, 57.5% of individuals do not use the internet to obtain news, which is not a significant difference. Additionally, 63.64% do not use the internet for video games, 59.09% do not use it for educational purposes, and 59.09% do not use it for work-related activities. In terms of how individuals gather information about internet services, 72.73% do not do so through the internet, while 54.55% do not use social media as a source of information. In 63.64% of cases, the budget allocated for the service aligns with the ideal price.

For the last and third market segment, it is observed that more than half of the individuals fall within the age range of 25 to 59 years, followed by the age group of 15 to 24 years, while individuals aged 60 or older represent a minority in this segment. Regarding gender, there is an equitable distribution between both sexes. The size of their households, including the respondents, is concentrated in three or more members, accounting for 85%. The most prominent occupations in this segment are students, traders, and administrators. The most relevant characteristics for this segment focus on speed, uninterrupted signal service, and affordability in price. On the other hand, payment methods, support, and technical service are considered of lesser importance compared to other characteristics.

In relation to internet usage time, it is noteworthy that over 70% of individuals use the internet for 3 hours or more daily. On the contrary, only 2.5% of users use the internet for less than one hour per day. As for the ideal price, 65% of this segment considers it to be between 10 and 30 US dollars, while 30% prefer a price below 10 US dollars. It should be mentioned that 60% of respondents stated that they would not continue to use the service if there was an increase in its current price.

In respect to specific internet usage, 68.18% of individuals use the internet to obtain news, while 62.5% do not use it for video games. Furthermore, 60% use the internet for educational purposes, and 62.5% use it for work-related activities. Regarding how individuals gather information about internet services, 70% do not do so through the internet, while 65% use social media as a means to gather information about internet services. In 87.5% of cases, the budget allocated for the service aligns with the previously mentioned ideal price.

In addition to the previously discussed, general trends and commonalities were found in all market segments. Around 60% of individuals allocate an amount of 10 to 30 US dollars, while between 26% and 31% budget less than 10 US dollars. These segments also consider between 86% and 93% the needs of household members when acquiring an internet service.

In terms of internet usage habits, it is observed that between 86% and 90% of individuals use it for entertainment, and between 86% and 95% use it to obtain information. Similarly, between 85% and 93% use the internet to access social networks, and between 86% and 95% use it for information retrieval.

In relation to how individuals gather information about internet services, it is noteworthy that between 81% and 93% acquire it through acquaintances, while between 92% and 95% do not do so via television. Additionally, between 86% and 95% do not use the radio as a means to learn about these services, and between 80% and 85% do not rely on billboards, posters, or outdoor stickers. It is important to mention that all subjects in the segments affirmed that they do not gather information through newspapers or magazines regarding internet services.

With this detailed information for each segment, marketing strategies can be developed for each group, allowing internet service companies in the San Juan Bautista Parish in the city of San Cristóbal, Táchira State, to be tailored to each group in terms of product, price, promotion, and distribution approaches initially. However, this can expand to processes, personnel, and physical evidence, which will increase the effectiveness of their advertising campaigns and enable them to leverage market opportunities more effectively.

In conclusion, this proposal not only represents an opportunity to improve the effectiveness of advertising campaigns and seize market opportunities but also has the potential to foster the development of more competitive and customer-oriented businesses, generating a positive impact on the local economy.

For further information on how the results were obtained and the tools used, a repository was made available, containing the programming code used for this research, accessible through the following link: <https://github.com/SantiagoAlarconDS/Customer-Segmentation> .

## RECOMENDATIONS

In order to make the most of these opportunities, the following series of recommendations are presented for companies to improve customer satisfaction and competitiveness in the market. These recommendations cover various aspects, from improving service quality to implementing new marketing strategies, and are designed to assist companies in adapting to the needs of the local market and effectively meet the demands of potential customers.

1. Take advantage of market segmentation to design personalized marketing strategies: Companies should use the information obtained to design effective and personalized marketing strategies for each market segment. This includes adjusting their product, price, promotion, and distribution approach to meet the needs and preferences of each group.
2. Improve service quality: Companies should strive to improve the quality of the service they provide to meet their customers' expectations and foster brand loyalty. This includes offering faster connection speeds, greater connection stability, efficient customer support, and quick problem resolution.
3. Offer packages and promotions tailored to each segment: Companies should design packages and promotions tailored to each market segment, considering the specific preferences and needs of each group. For example, they can offer packages with higher connection speeds and more data for segments that require fast and stable connections, or packages with special discounts for students.
4. Monitor market trends: Companies should stay updated on market trends and new technologies in the internet services sector. This will allow them to quickly adapt to changes and offer innovative and cutting-edge services that meet their customers' needs.
5. Encourage customer feedback: Companies should encourage customer feedback through surveys and other mechanisms to gather their opinions about their services. This will enable them to continuously improve service quality and effectively adjust their marketing strategies.

In general, internet service companies in the San Juan Bautista parish in the city of San Cristóbal, Táchira state, must constantly strive for customer satisfaction and continuous improvement, leveraging the information obtained through market segmentation to offer personalized and high-quality services that meet the demands of the local market.

# REFERENCES

Barreto Villanueva, A. (2012). El progreso de la Estadística y su utilidad en la evaluación del desarrollo. Papeles de Población, p.4. https://www.redalyc.org/articulo.oa?id=11224638010

Borjas, C. M. B. (2005). Ley de Pareto aplicada a la fiabilidad. Ingeniería mecánica.

López-Roldán, P., & Fachelli, S. (2015). Metodología de la investigación social cuantitativa.

Meiriño, R. C., Brea, J. A. F., Vila, N. A., & López, E. R. (2016). Segmentación del mercado de un destino turístico de interior. El caso de A Ribeira Sacra (Ourense). PASOS Revista de Turismo y Patrimonio Cultural, 14(2), 369-383. https://doi.org/10.25145/j.pasos.2016.14.024

Moreno A. Diego & A Méndez R. Maury E. (n.d) Propuesta de red de espacios públicos Para la parroquia San Juan Bautista. San Cristóbal Edo. Táchira.

United Nations (n.d). Día Internacional de las Personas de Edad. https://www.un.org/es/observances/older-persons-day

United Nations (s. f.). Juventud. https://www.un.org/es/global-issues/youth

Pareto, V. (1896). Cours d'économie politique: professé à l'Universi̧té de Lausanne (Vol. 1). F. Rouge.

Raiter, O. (2021). Segmentation of Bank Consumers for Artificial Intelligence Marketing. International Journal of Contemporary Financial Issues, 1(1), 39-54.

BOLIVARIAN REPUBLIC OF VENEZUELA. (1999). Constitución. Gaceta Oficial Nº 36860. 30 de diciembre de 1999. Asamblea Nacional Constituyente Caracas.

Robin, C. F., & Torres, C. A. (2001). Segmentación de mercados: buscando la correlación entre variables sicológicas y demográficas.

Rodriguez, N., Rodríguez, V., Ramírez, E., Cediel, S. A. D., Gil, F., & Rondón, M. (2016). Aspectos metodológicos del diseño de muestra para la Encuesta Nacional de Salud Mental 2015. *Revista Colombiana de Psiquiatría*. <https://doi.org/10.1016/j.rcp.2016.08.009>

Sampieri R.H., Collado C.F. y Lucio, M.P.B. (2014). Metodología de la investigación sexta edición. McGraw-Hill.

Suárez Ibujés, M. O. (2011). Cálculo del tamaño de la muestra.

VENEZUELA. INSTITUTO NACIONAL DE ESTADÍSTICA. (INE) Proyecciones de población con base al Censo 2011, Proyección de la población al 30 de junio, según entidad federal y municipios (año calendario. http://www.ine.gov.ve/documentos/Demografia/SituacionDinamica/Proyecciones/xls/Entidades/Resumen\_Municipios.xls

VENEZUELA. INSTITUTO NACIONAL DE ESTADÍSTICA. (INE) Proyecciones de población con base al Censo 2011, Proyección de la población al 30 de junio, según entidad federal, municipios y parroquias (quinquenal). http://www.ine.gov.ve/documentos/Demografia/SituacionDinamica/Proyecciones/xls/Parroquias.xls

VENEZUELA. INSTITUTO NACIONAL DE ESTADÍSTICA. (INE) Cuadros de resumen del Censo 2011, Población empadronada por grupos de edad, según entidad federal, municipio y parroquia. http://www.ine.gov.ve/documentos/Demografia/CensodePoblacionyVivienda/xls/CuadrosResumenCenso2011/MunicipiosParroquias/Poblacion\_Grupos\_Edad\_Censo\_2011\_Mu\_Pa.xlsx

1. Sampieri R.H., Collado C.F. y Lucio, M.P.B. (2014). Metodología de la investigación sexta edición. McGraw-Hill, p.4. [↑](#footnote-ref-1)
2. Barreto Villanueva, A. (2012). El progreso de la Estadística y su utilidad en la evaluación del desarrollo. Papeles de Poblacion, p.4. [↑](#footnote-ref-2)
3. Sampieri R.H., Collado C.F. y Lucio, M.P.B. (2014). Metodología de la investigación sexta edición. McGraw-Hill, p.4. [↑](#footnote-ref-3)
4. Malhotra, N. K. (2008). Investigación de mercados. Pearson Educación, p. 304. [↑](#footnote-ref-4)
5. Moreno A. Diego & A Méndez R. Maury E. (n.d) Propuesta de red de espacios públicos Para la parroquia San Juan Bautista. San Cristóbal Edo. Táchira. p.69-70 [↑](#footnote-ref-5)
6. BOLIVARIAN REPUBLIC OF VENEZUELA. (1999). Constitución. Gaceta Oficial Nº 36860. 30 de diciembre de 1999. Asamblea Nacional Constituyente Caracas. [↑](#footnote-ref-6)
7. Sampieri, R. H., Collado, C. F., Lucio, P. B., Valencia, S. M., & Torres, C. P. M. (2014). *Metodología de la investigación*. [↑](#footnote-ref-7)
8. *Ibidem* [↑](#footnote-ref-8)
9. Robin, C. F., & Torres, C. A. (2001). Segmentación de mercados: buscando la correlación entre variables sicológicas y demográficas. Revista colombiana de Marketing, 2(2). [↑](#footnote-ref-9)
10. Rodriguez, N., Rodríguez, V., Ramírez, E., Cediel, S. A. D., Gil, F., & Rondón, M. (2016). Aspectos metodológicos del diseño de muestra para la Encuesta Nacional de Salud Mental 2015 [↑](#footnote-ref-10)
11. Rodríguez, N., Rodríguez, V. A., Ramírez, E., Cediel, S., Gil, F., & Rondón, M. A. (2016). Aspectos metodológicos del diseño de muestra para la Encuesta Nacional de Salud Mental 2015. Revista colombiana de psiquiatría, 45, 26-30. [↑](#footnote-ref-11)
12. Malhotra, N. K. (2008). Investigación de mercados. Pearson Educación, p. 304. [↑](#footnote-ref-12)
13. United Nations. (n.d). Juventud. [↑](#footnote-ref-13)
14. United Nations. (n.d). Día Internacional de las Personas de Edad. [↑](#footnote-ref-14)
15. VENEZUELA. INSTITUTO NACIONAL DE ESTADÍSTICA. (INE) Proyecciones de población con base al Censo 2011, Proyección de la población al 30 de junio, según entidad federal y municipios (año calendario). [online database]. [↑](#footnote-ref-15)
16. VENEZUELA. INSTITUTO NACIONAL DE ESTADÍSTICA. (INE) Proyecciones de población con base al Censo 2011, Proyección de la población al 30 de junio, según entidad federal, municipios y parroquias (quinquenal). [online database]. [↑](#footnote-ref-16)
17. VENEZUELA. INSTITUTO NACIONAL DE ESTADÍSTICA. (INE) Cuadros de resumen del Censo 2011, Población empadronada por grupos de edad, según entidad federal, municipio y parroquia. [online database]. [↑](#footnote-ref-17)
18. López-Roldán, P., & Fachelli, S. (2015). Metodología de la investigación social cuantitativa. p.8 [↑](#footnote-ref-18)
19. López-Roldán, P., & Fachelli, S. (2015). Metodología de la investigación social cuantitativa. p.8 [↑](#footnote-ref-19)
20. *Ibidem* [↑](#footnote-ref-20)
21. López-Roldán, P., & Fachelli, S. (2015). Metodología de la investigación social cuantitativa. p.22 [↑](#footnote-ref-21)
22. Pareto, V. (1896). Cours d'économie politique: professé à l'Universi̧té de Lausanne (Vol. 1). F. Rouge. [↑](#footnote-ref-22)
23. Robin, C. F., & Torres, C. A. (2001). Segmentación de mercados: buscando la correlación entre variables sicológicas y demográficas. Revista colombiana de Marketing, 2(2). [↑](#footnote-ref-23)
24. Raiter, O. (2021). Segmentation of Bank Consumers for Artificial Intelligence Marketing. International Journal of Contemporary Financial Issues, 1(1), 39-54. [↑](#footnote-ref-24)
25. Meiriño, R. C., Brea, J. A. F., Vila, N. A., & López, E. R. (2016). Segmentación del mercado de un destino turístico de interior. El caso de A Ribeira Sacra (Ourense). PASOS. Revista de Turismo y Patrimonio Cultural, 14(2), 369-383. [↑](#footnote-ref-25)