Vigilada Mineducación

Estimating a relationship between tombstone cost and longevity using cemetery data in Medellín city in Colombia

ASTRID LIZETH MARTÍNEZ RODRÍGUEZ

Tesis

Asesor, docente

Uriel Zapata Múnera

UNIVERSIDAD EAFIT

ESCUELA DE INGENIERÍAS

MAESTRÍA EN CIENCIAS DE LOS DATOS Y LA ANALÍTICA

MEDELLÍN

2023

Estimating a relationship between tombstone cost and longevity using cemetery data in Medellín City in Colombia

Astrid Martínez,[1] Gary O'Donovan (ORCID: 0000-0001-6748-6218),[2,3] Santiago Flórez,[4] Ana Isabel Cadavid Castrillón,[5] and Uriel Zapata (ORCID: 0000-0002-9309-7571)[3]

[1]School of Applied Sciences and Engineering, EAFIT University, Medellín, Colombia.
[2]Latin American Brain Health Institute (BrainLat), Universidad Adolfo Ibáñez, Santiago, Chile.
[3]Facultad de Medicina, Universidad de los Andes, Bogotá, Colombia.
[4]Fundación Universitaria Bellas Artes, Medellín, Colombia.
[5]Independent Researcher, Medellín, Colombia.

Acknowledgements

We thank Valentina Mobil Sandoval for her help in developing the mobile phone app used to facilitate data collection. We also thank Luciana Zapata for her help with data collection.

Abstract

Recent studies in the West suggested that tombstone cost is associated with longevity. The main goal of this study was to investigate the association between tombstone cost and life expectancy in a large cemetery in Latin America. Age at death was obtained from 2,273 consecutive death certificates held at the San Pedro Cemetery Museum in Medellín, Colombia. Subjects died between 2020 and 2022. Tombs are arranged in galleries in the cemetery, and tombstone cost is based on the material from which the tombstone was made its position in the gallery, and its ornamentation. Approximately 76% of the tombstones were low cost, 16% medium cost, and 8% high cost. Analysis of variance was used, and the assumption of equal variance was not violated. Because the data did not show a normal distribution, it was necessary to apply non-parametric techniques to assess statistical differences. The Kruskal-Wallis's test was employed for this purpose.

Longevity was similar in the low-cost group and medium-cost group: 61.5±23.9 versus 61.6±24.6 years [estimated mean (95% confidence interval)]. Longevity was lower in the high-cost group: 55.8±25.8 years. The inverse association between tombstone cost and longevity would suggest that people in Medellín are inclined to spend more on tombstones when commemorating the tragic death of a young person.

Keywords

Burial; Cemeteries; Life expectancy; Death; Latin America.

1. Introduction

Life expectancy is an important synthetic indicator for evaluating the economic and social development of a country or region [1]. One of the most influential factors on life expectancy that has been studied is socioeconomic status. For instance, attempts have been made to relate the increase in indicators such as the per capita gross domestic product (GDP) of nations to the increase in life expectancy at birth [2]. An interesting perspective that has not yet been widely studied is the relationship between socioeconomic status reflected in the gravestones awarded to the deceased and their longevity. Therefore, it is worth investigating the connections that may exist between the cost of the funeral monument and life expectancy.

Intuitively, one could conjecture that a high-cost funeral monument would be associated with individuals who enjoyed a high socioeconomic status. However, the peculiarities associated with local culture cannot be disregarded. Currently, knowledge of the relationship between socioeconomic status and life expectancy is still scarce. International researchers have suggested that the cost of funeral monuments would be related to socioeconomic status and life expectancy. It was expected that the burial cost of a deceased person would be influenced by aspects such as the prestige of the funeral gallery and the location of the vault within the gallery.

Latin American cemeteries represent significant sources of unexplored social data, possibly because the prevalence of their social, sanitary, and historical functions has detracted from the importance of digitizing information and constructing databases, which has been a necessity for businesses and institutions for years. This hinders access to information and poses challenges for its study and analysis. Data science and analytics, as interdisciplinary fields of knowledge, offer solutions that range from data recording to communication, processing, and analysis, justifying their presence in research across various disciplines.

For the case of the present problem to be addressed, it is pertinent to perform an initial analysis supported by proper data management and an objective quantitative study using statistical tools. This will be crucial to set up the foundation for future research designs with the aim of replicating the study in more cemeteries across Latin America because funeral rituals are similar across Central and South America.

With the context clarified, the research hypothesis is as follows: Does the overall cost of gravestones have any relationship with the longevity of the deceased person? Is it possible that the more expensive the gravestone, the longer the person buried lived compared to someone who was given a less expensive gravestone?

To conduct an objective study in this research, information provided by the San Pedro Cemetery Museum in the city of Medellín, Colombia, was available. Since this institution lacks a digitally

formatted database containing information about the burials conducted there, one of the critical aspects to be resolved is the collection and storage of data that will be subsequently analyzed. Additionally, the fact that the cemetery does not have a classification system that categorizes gravestones according to their cost adds further complexity.

The main purpose of this work is to investigate, from an interdisciplinary perspective, the existence of a relationship between the cost of funeral monuments and life expectancy using data from the San Pedro Cemetery Museum in Medellín, Colombia. Specifically, the aim is to establish appropriate criteria that allow for the categorization of vault gravestones in the San Pedro Cemetery Museum into three groups (high, medium, and low) based on their overall cost. Additionally, an effective mechanism will be developed to enable the digital recording and storage of variables that will determine the classification of gravestones according to their overall cost, as well as the statistical comparison of the longevity of individuals belonging to the categories that group gravestones according to their cost (high, medium, and low).

1.1. Literature review

Life expectancy is an estimate of the average number of years a person of a certain age can expect to live. Life expectancy is a hypothetical measure that assumes the specific age-specific mortality rates for the given year will apply throughout the lives of individuals born in that year. This measure differs by sex, age, race, and geographic location. Therefore, life expectancy is usually given for specific categories rather than the general population [3].

One of the factors often associated with life expectancy is income, suggesting that higher incomes imply greater longevity [4]. Studies conducted in Nordic countries, known for their significant social investment, show that premature deaths are not a general phenomenon and are not evident in low-income population groups. This expands the understanding of the dynamics of mortality inequalities related to economic resources, which can provide a basis for more specific public policy responses [5].

The influence of socioeconomic differences on mortality is a phenomenon that has already been studied. In the 19th century, pioneers of social epidemiology began to link poverty and deprivation with longevity. Villermé, Chadwick, and Virchow were among the first researchers in Europe to document the importance of socioeconomic conditions for life expectancy [6]. In the second half of the 20th century, the Whitehall Study drew attention to the social determinants of health by linking employment status with mortality in British civil servants using modern analytical methods [7]. Following this work,

several studies have confirmed that income and other measures of social position are strongly correlated with life expectancy [8]. Today, subjects with higher incomes can expect to live 10 to 15 years longer than those with lower incomes [4]. These studies indirectly suggest that subjects with higher incomes may have greater access to preventive and interventional healthcare systems.

Social sciences have few tools that allow them to retrospectively address the relationship between socioeconomic status and life expectancy. One source of information that can reveal interesting findings is cemeteries. However, it is necessary for these places to have detailed and reliable information about the people who use their services.

Cemeteries primarily serve the function of housing the mortal remains of the deceased. However, they also indirectly expose other facets of social life. The aesthetic differences between tombstones can reflect differences in social classes, customs/traditions, and historical or religious trends [9]. The cemetery is a socially constructed space that symbolizes the relationship between life and death. Cemeteries are usual places intended for burial and exhumation activities that respond to the sanitation needs or religious expressions of a particular society [10].

Around funeral rituals, research has been conducted to describe the before and after social life of communities, influenced by the presence of the Catholic Church, which affects burial customs, leading to the construction of initially isolated cemeteries that were eventually absorbed by urban development. It was common to see in central cemeteries of several cities in the country how hierarchies were established, allowing the bodies of prominent individuals to be housed in prominent locations within the burial grounds and adorned with expensive funerary monuments. The power dynamics of the deceased occupants were clearly reflected in the allocation of graves and mausoleums within a cemetery, perpetuating the social castes of the living [11].

The few existing studies suggest that tombstone cost is associated with longevity [12, 13]. Davey Smith and colleagues assumed that taller obelisks cost more than shorter obelisks and investigated the association between obelisk height and longevity in eight cemeteries in Glasgow, Scotland [12]. Data were obtained from obelisks built from 1801 to 1920, and it was found that obelisk height was correlated with age at death [12]. Kang-Auger and colleagues assumed that larger tombstones cost more than smaller tombstones and investigated the association between tombstone size and longevity in a cemetery in Quebec, Canada [13]. Data were obtained from tombstones built from 1820 to 1992, and it was found that tombstone height and volume were associated with age at death after adjusting for sex, marital status, and year of death [13]. These studies were inspired by the need to find alternative ways of investigating associations between socioeconomic status and longevity due to the lack of decent quality data in Scotland and Canada in the nineteenth and early twentieth centuries [12, 13]. Little is known about associations between socioeconomic status and longevity in Latin America due to the lack of available data in the region [14].

In the Colombian context, it is interesting to study characteristics such as the degree of ornamentation presented by the tombstones installed in cemeteries since the relationship between the living and the dead can be inferred through decorative elements. These elements provide access to the universe of meaning and historicity of the funeral behavior of the population, focusing on individual characteristics that address sociodemographic aspects and the cult of the deceased [15], displaying a personalization that recalls their identity.

Colombian research has shown that marble and stone tombstones were materials that only wealthy individuals could afford in the early decades of the 20th century [16]. On the other hand, materials such as wood and other simple materials were used by less affluent social sectors [17], suggesting that the economic capacity of the population decides the quality of the material of the tombstone installed in a burial. Considering that over time the materials used in the production of tombstones vary according to trends of the era, it is pertinent to investigate the variety of materials used today, as well as the different manifestations of ornamentation that can give greater ostentation to the tombstones.

Previous studies have sought to identify the relationship between the size of the funeral monument and life expectancy, using age at death as the dependent variable through linear regression, where the outcome was continuous. Additionally, sources of bias such as sex, marital status, and year of death were adjusted [13].

2. Materials and Methods

To reach the proposed objectives, the implementation of the CRISP-DM methodology is proposed, which is widely used in data mining processes. In general terms, the CRISP-DM method consists of 6 phases that follow the general methodological model of analysis, design, construction, and implementation, with a clear focus on obtaining, processing, and analyzing data to derive analytical models [22]. The six phases of this method are: business understanding, data understanding, data preparation, modeling, evaluation, and deployment, as shown in Figure 1. The following sections will detail the phases of this method within the framework of research development:

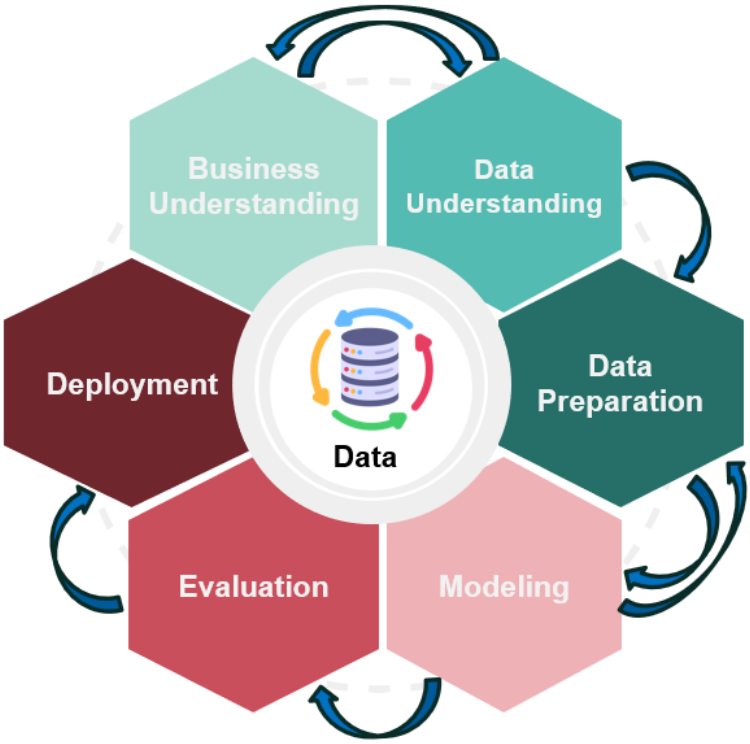


Figure 1. Illustrative diagram of the phases of the CRISP-DM methodology.

2.1. Business Understanding: This phase encompasses the task of understanding the project's goals and requirements from a business perspective to translate them into technical aims and a project plan. Within the context in which the project will be developed, it is important to consider the data source space, which is the San Pedro Cemetery Museum in Medellín, Colombia. The cemetery was founded in 1842 and was recognized as a museum in 1998 and as a national heritage site in 1999 [23].

Figure 2. Tombs are arranged in galleries in the San Pedro Cemetery Museum in Medellín in Colombia. The figure shows Gallery 87, San Gabriel. Photographs by Santiago Flórez (@santi.ph6).

Considering that the purpose of the research is to study the relationship between the cost of a funeral monument (tombstone) and longevity, it is necessary to understand the factors that influence this cost. These factors include: (1) the material of the tombstone, (2) the height at which it is located within a gallery, and (3) its level of ornamentation. Understanding these factors requires knowledge of the diverse options of tombstones that can be found in this cemetery, and a classification must be defined to categorize the cost as high, medium, or low.

2.2. Classification according to material

Seven stonemasons near the San Pedro Cemetery Museum were visited to determine the costs of the various materials used to make tombstones currently. After consolidating the information gathered during visits to the marble shops, 8 types of materials were identified. In consensus with the research area of the Cemetery, it was decided to establish a scoring system ranging from 1 to 3, where 1 corresponds to a low-cost material, 2 to a medium-cost material, and 3 to a high-cost material.

Table 1 shows the tombstone cost according to material. The high-cost category included red quartz, green Ubatuba granite from Brazil, and black marble. The high-cost category was given a score of three points. The medium-cost category included cream marble and grey marble. The medium-cost category was given a score of two points. The low-cost category included laminated stone and the plain stones provided by the cemetery free of charge. The low-cost was given a score of one point. Costs are shown in Colombian pesos (COP).

Table 1. Classification of tombstone cost according to material.

Material	Example	Price (COP)	Cost category	Cost score
Red quartz		650,000	High	3
Green Ubatuba granite		450,000	High	3

Black marble		400,000	High	3
Cream marble		300,000	Medium	2
Grey marble		300,000	Medium	T2

Table 1 (continued). Classification of tombstone cost according to material.

Material	Example	Price (COP)	Cost category	Cost score
Laminated stone		280,000	Low	1
Cemetery tombstone		Free	Low	1
Cemetery tombstone with laminate		Free	Low	1

2.3. Classification according to position in the gallery

Tombs are arranged in galleries in the San Pedro Cemetery Museum. The commercial department staff of the cemetery was interviewed to understand how the cost varies based on the height at which the tombstone is installed. It is indicated that the cost of the location is influenced by the accessibility to the vault; therefore, extreme locations are less accessible and consequently less expensive. According to the information received, a scoring system ranging from 1 to 3 is set up, where one corresponds to a low-cost material, 2 to a medium-cost material, and 3 to a high-cost material. Table 2 shows the tombstone's cost according to its position in the gallery. The high-cost category included tombs around the head´s height that were easier to reach. The high-cost type category was given a score of three points. The medium-cost category included tombs just above head height. The medium-cost class was given a score of two points. The low-cost category included tombs at the top of the gallery that were harder to reach and tombs at the bottom that were harder to view. The low-cost category was given a score of one point. Tombs can be bought, but most are rented. The cost of renting a tomb at head height is around COP 11,950,000, including the possibility of making a down payment of COP 5,975,000 and 24 monthly payments of COP 298,297. The cost of renting a tomb in a less desirable position is around COP 10,500,000, including the choice of making a down payment of COP 5,250,000 and 24 monthly payments of COP 262,102 (costs do not sum because interest is charged on monthly payments).

Table 2. Classification of tombstone cost according to its position in the gallery.

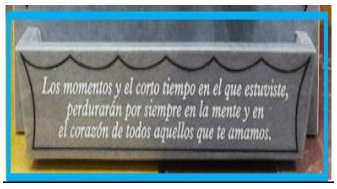
Position in gallery	Cost category	Cost score
Row 6 (highest)	Low	1
Row 5	Low	1
Row 4	Medium	2
Row 3 (head height)	High	3
Row 2	High	3
Row 1 (lowest)	Low	1

2.4. Classification according to ornamentation

The cemetery was toured, and it was observed that the tombstones exhibit a certain degree of personalization. Various forms of expression were evident, including religious images, vases, engraved images, and other accessories. Eight types of accessories commonly used today were identified. Subsequently, visits were made to 7 marble workshops near the San Pedro Cemetery Museum to inquire about the prices of these accessories. Table 3 shows some of the ornamentations used in the present day. Costs varied from as little as COP 12,000 for a flowerpot to COP 150,000 for an engraving. The costs of

the ornamentations on a tombstone were summed to create cost categories. The high-cost category included ornamentation to the sum of greater than COP 200,000. The medium-cost category included ornamentation in the sum of COP 101,000-200,000. The low-cost category included ornamentation to the sum of COP 0-100,000. The high-cost category was given a score of three points, the medium-cost category two points, and the low-cost category one point.

Table 3. Ornamentations that may be added to a tombstone.

Ornament	Example	Cost (COP)
Flowerpot		12,000-25,000
Photo of the deceased		35,000-45,000
Engraving attached to the tombstone.		45,000
Flower box		45,000-60,000

Balcony		50,000
Figurine		50,000
Engraving within the tombstone		120,000-150,000
Roof		150,000-200,000

2.5. Weighting of tombstone characteristics

Once the cost of each aspect of interest (material, height, and accessories) was determined, various possibilities were examined, it was found that the material composition is the most influential factor, as it can range from COP 0 to 650,000. On the other hand, height where the tombstone is installed was considered the second factor contributing to the overall price, although to a lesser extent, as the cost of

purchasing the vault can vary from COP 8,850,000 to 11,950,000. Lastly, the number of accessories that can be installed in the vault has a minor influence. These findings are discussed with the interdisciplinary group, which includes the historian leading the research area of the cemetery, and the weights of these characteristics are decided as follows: Tombstone material was given a weighting of 0.7, tombstone position was given a weighting of 0.1, and tombstone ornamentation was given a weighting of 0.2.

Table 4. weights distributed according to variable.

Variable	Weight
Cost Material	0.7
Cost Location	0.1
Cost Accessories	0.2
Total	1

2.6. Final classification

Tombstone characteristics were multiplied by the respective weightings and summed to create the final score. For example, consider the tombstone in Figure 3. The tombstone material is cream marble, giving a score of 2 x 0.7 = 1.4. The tombstone position is head height, giving a score of 3 x 0.1 = 0.3. The tombstone has a flower box and an engraving, and the price of these ornamentations is greater than COP 200,000, giving a score of 3 x 0.2 = 0.6. The final score is 1.4 + 0.3 + 0.6 = 2.3. Final scores of 2.4-3.0 were used to classify tombstones as high cost. Final scores of 1.7-2.3 were used to classify tombstones as medium cost. Final scores of 1.0-1.6 were used to classify tombstones as low cost. Therefore, the tombstone in Figure 3 is classified as medium cost.

Figure 3. A tombstone in the San Pedro Cemetery Museum in Medellín in Colombia.

2.7. Data Understanding: This phase includes aspects such as data collection, preliminary analysis, and showing data quality. It is considered a critical phase for the research since, lacking a digital database, the burial records archived physically were used, and a database was constructed to record information about the deceased.

Due to the cemetery only kept physical files, photographs were taken of 2550 burial certificates to extract information about the deceased, including the burial vault location within the cemetery, name, age at the time of death, gender, marital status, cause of death, occupation, date of death, and city of residence. Considering that these certificates are manually filled out by cemetery personnel, difficulties in reading the information or fields left incomplete are met, affecting data quality and causing the exclusion of records that introduce significant uncertainty. Additionally, it should be noted that the cemetery receives deceased individuals from municipalities outside of Medellín, resulting in non-uniformity in the formats. Consequently, a dataset consisting of 2318 records was constructed. Figure 4 is an example of a burial record. It can be observed that several fields are manually filled out. There are also instances where fields are incomplete. These are some of the challenges when dealing with physical information.

Figure 4. Burial records format for deceased individuals from municipalities other than Medellin.

Subjects in the present study were people who died in 2022, 2021, or 2020. This study was conducted in collaboration with the San Pedro Cemetery Museum.

Subsequently, the tombstones corresponding to the subjects whose data had been recorded were visited, and characteristics such as their level in the gallery, ornamentation, and material were seen. This information was recorded through an application built using the Power Apps tool, which stores the data in a spreadsheet saved in the Microsoft online service for the next statistical analysis.

Each death certificate held at the San Pedro Cemetery Museum had a code that showed the location of the tomb in the cemetery. A mobile phone app was created to match up death certificates and tombstones characteristics. The app was created using Microsoft Power Apps. Figure 5 shows the process followed by the app.

One of the most important steps that the developed application follows is to save a photograph of all the tombstones visited. This is done to ensure that the information of the deceased is correctly associated with the characteristics of their tombstone. The saved image is used to verify the name and date of death of the deceased. Additionally, the image serves to confirm that the characteristics were recorded accurately. When there was no concordance between the information of the deceased and the image of the tombstone, the record was discarded.

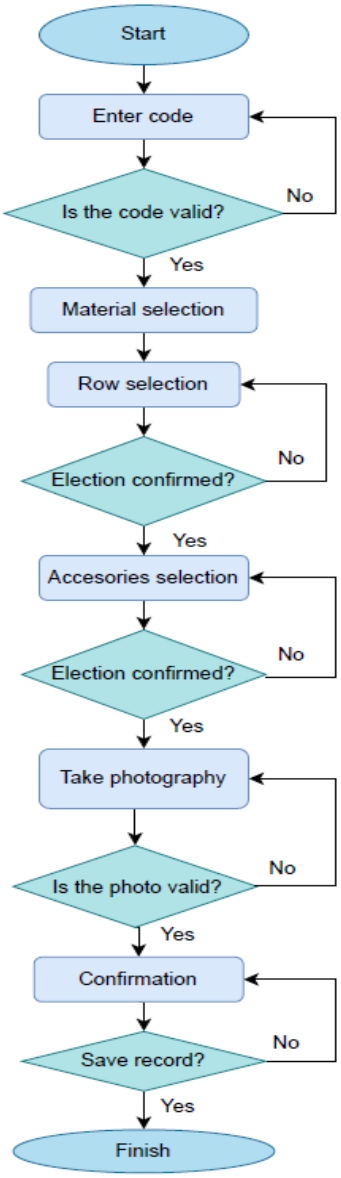


Figure 5. Flowchart of the process followed by the application.

Figure 6 shows the user interface, consisting of six steps from entering the tombstone's code to be evaluated, selecting the material from which this tombstone is made, followed by choosing the height of the vault in which the tombstone is installed, then selecting the type of accessories and their quantity, and finally taking a photograph of the tombstone and confirming the record. The application includes the functionality to return to the previous step if changes to the selections are needed.

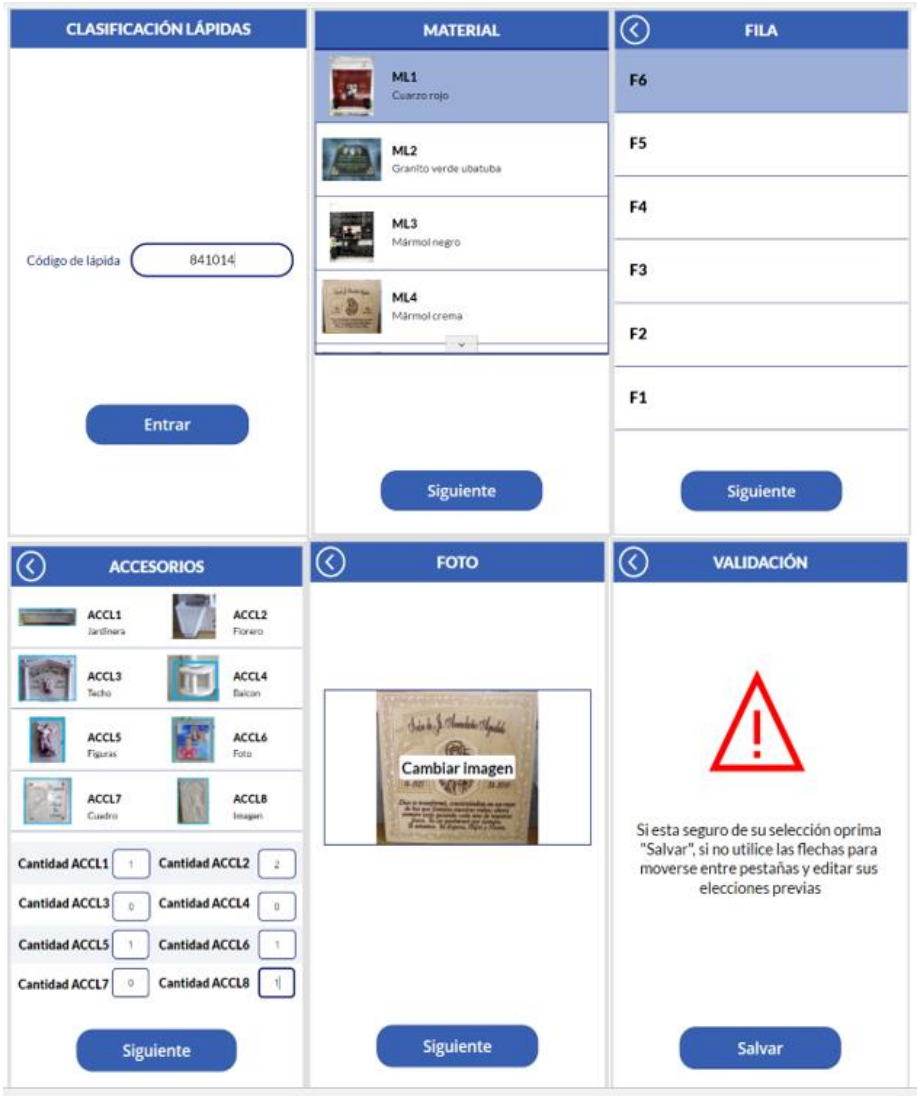


Figure 6. User Interfaces of the developed application.

2.8. Data Preparation: Once the data has been collected, it is necessary to verify if it is amenable to ETL (extract, transform, load) processes, where data cleansing is performed, data formats are corrected when necessary, and a subset of data having the required variables for the desired analysis is selected. The following variables were considered: Tombstone Code, Date of Death, Age at Death, Gender, Occupation, Marital Status, Cause of Death, Material Score, Height Score, Accessories Score, Weighted Score, and Classification.

The information recorded in the application needs to be transformed, as the choices of material, height, and accessories with quantities must adhere to the classification described in the business understanding section. Taking advantage of the fact that the application stores records in a cloud based .xlsx file, the file is

downloaded, and formulas are used to place the choices into one of the classification groups. Additionally, the data quality of the dataset constructed from the burial records is reviewed. Records that exhibit inconsistencies in death dates, names, and cases where the tombstone location code was not found in the cemetery are removed.

Since information was previously extracted from burial records, it is now necessary to relate that information with the records obtained through the application, using the tombstone code as the key. Figure 7 shows the script developed in the KNIME Analytics Platform, where the data is consolidated and a file with 2,273 records is obtained for the next statistical analysis.

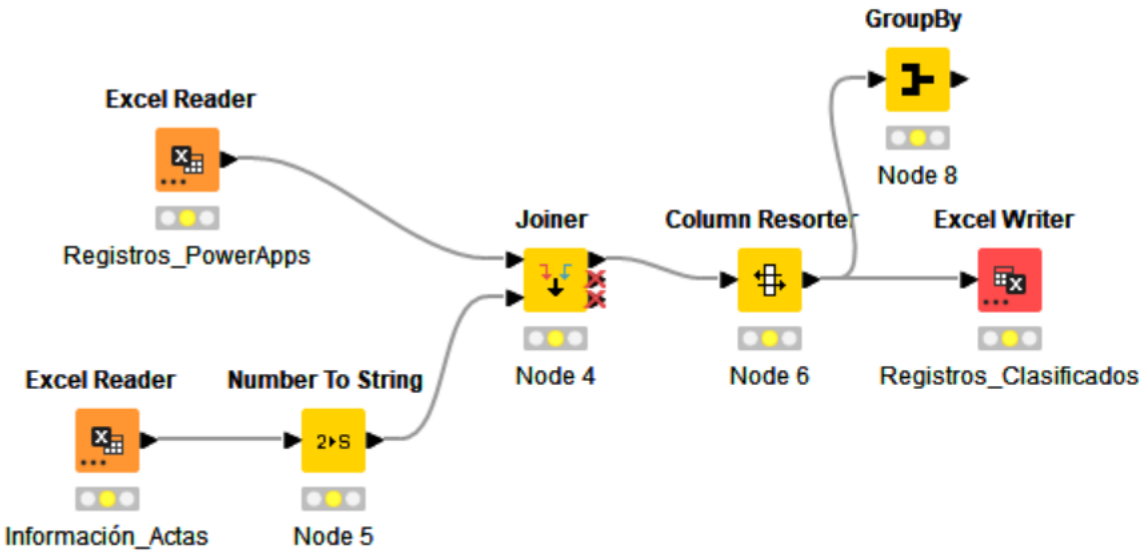


Figure 7: KNIME script developed for the consolidation of information taken from burial records and the records made in the Power Apps application.

2.9. Modeling:

The descriptive analysis of the data in this research has revealed that the variable 'Longevity' does not follow a normal distribution. As a result, parametric techniques such as ANOVA (analysis of variance) and ANCOVA cannot be employed to assess statistically significant differences, as one of the assumptions is violated [18]. Non-parametric techniques, such as the Kruskal-Wallis's test, are therefore considered, which is used as an equivalent to the one-way ANOVA test [19].

The test as proposed by Kruskal and Wallis [20] evaluates whether two or more samples are from the same distribution. The null hypothesis is that all the samples come from the same distribution.

Assuming there are k samples, each with its respective set of values. To apply the Kruskal-Wallis's test, it is initially necessary to assign a rank to all values regardless of which sample they belong to, followed by calculating the sum of all rank values within each sample, if there is no tie in all the values, the test statistic is:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1) \quad (1)$$

where N is the total number of values in all samples; n_i is the number of values contained in the ith sample, and R_i is the sum of ranks in ith sample [21].

Data were obtained from 2,273 consecutive death certificates, starting with the most recently available.

2.9.1. Variables

The dependent variable was longevity, defined as age at death. The independent variable was tombstone cost. Tombs are arranged in galleries in the San Pedro Cemetery Museum, as shown in Figure 2. There is no single measure of tombstone cost, so we created a scoring scheme based on the material from which the tombstone was made, the position of the tomb in the gallery, and the ornamentation added to the tombstone. For example, quartz, granite, and marble tombstones were scored higher than laminated tombstones. Tombstones at head height were scored higher than tombstones above or below head height. Tombstones with engravings and embellishments were scored higher than plain tombstones. Tombstones were classified as being of low cost, medium cost, or high cost as described in detail in the online supplement. Each death certificate had a code that showed the location of the tomb in the cemetery. Others independent variables were also evaluated: gender, cause of death, Covid association with the death of the subjects, variables were selected because they had better data quality, with fewer missing values.

2.9.2. Descriptive Statistical analysis

Descriptive statistical analyses were conducted, reporting percentages, means, and standard deviations, for this purpose, the 'scipy.stats' and 'pingouin' libraries in Python were utilized. Graphs were created to provide visual support for analyzing the relationship between variables. Data normality was assessed to determine which techniques should be employed to verify the presence of statistically significant differences to compare longevity the three classified groups of

tombstones (high, medium, and low), while considering relevant covariates such as gender, marital status, and violent death.

2.9.3. Statistical analysis

There is no evidence of any studies of tombstone cost and longevity in Latin America. Therefore, this study had to use data from elsewhere to estimate the sample size. We used data from Kang-Auger and colleagues' study of 276 people buried between 1820 and 1992 in a cemetery in Quebec, Canada [13]. We assumed that the standard deviation in longevity was 20 years [14]. This study also assumed that mean life expectancy was 60 years in those buried in tombs of low cost, 63 years in those buried in tombs of medium cost, and 70 years in those buried in tombs of high cost [13]. These assumptions are broadly in keeping with life expectancy in Colombia during the period of investigation [24]. For example, life expectancy at birth in Colombia was around 60 years in 1960 [24]. We used the *twomeans* command in Stata SE version 17.0 for Mac (Stata Corp, Texas, USA) to estimate sample size, where power was 0.8 and alpha was 0.05. The estimated sample size was 699 tombstones of low cost, 699 of medium cost, and 130 of high cost. Analysis of variance was used to investigate differences in longevity according to tombstone cost. Longevity was a continuous variable (age at death), and tombstone cost was a categorical variable (low versus medium versus high).

2.9.4. Evaluation: The Shapiro-Wilk normality test was performed to assess the normal distribution of the data. Its outcome finds the technique used to evaluate the difference in longevity between the studied groups. Since the data does not show a normal distribution, the Kruskal-Wallis's test is used to verify differences between groups, followed by the post hoc test Dunn's test was performed to find which groups display statistically significant differences. The Bonferroni correction was applied to adjust for multiple comparisons.

2.9.5. Deployment: It is important to highlight the absence of research studying the relationship between the cost of funeral monuments and life expectancy in Latin America. Therefore, this research is a first approach to exploring cemeteries as potential sources of socioeconomic information. While the results cannot be generalized, they could serve as references for further investigations in this line of research in other cemeteries, cities, and countries.

3. Results

Once the dataset was consolidated, descriptive statistics were started, revealing that the 'Longevity' field held missing values. Therefore, in tests involving this variable, records with missing values were discarded, resulting in 2247 valid records.

Table 5. Subject characteristics*

Characteristic	Value
Age at death, mean±SD (n)	61.1±24.2 (2,247)
Tombstone cost	
Low, n (%)	1,731 (76.15)
Medium, n (%)	361 (15.89)
High, n (%)	181 (7.96)
Missing, n (%)	0 (0)
Total, n (%)	2,273 (100)
Sex	
Male, n (%)	1,274 (56.05)
Female, n (%)	999 (43.95)
Missing, n (%)	0 (0)
Total, n (%)	2,273 (100)
Civil status	
Not married or with partner, n (%)	1,092 (48.04)
Married or with partner, n (%)	667 (29.34)
Missing, n (%)	514 (22.61)
Total, n (%)	2,273 (100)
Violent death	
Not violent, n (%)	1,531 (67.36)
Violent or under investigation, n (%)	725 (31.90)
Missing, n (%)	17 (0.75)
Total, n (%)	2,273 (100)
Year of death	
2022, n (%)	480 (21.12)
2021, n (%)	1,154 (50.77)
2020, n (%)	639 (28.11)
Missing, n (%)	0 (0)
Total, n (%)	2,273 (100)

*Values are from 2,273 consecutive death certificates, starting with the most recently available. SD is standard deviation. N is number.

Box-and-whisker plots were generated to visualize the median longevity achieved by subjects belonging to the three groups classifying the overall cost, material cost, location cost, and accessory cost. Table 6 presents the means and standard deviations of the 'Longevity' variable for all classification groups. It can be at once seen that the mean is lower for the 'high-cost' group, except in the case of the 'cost_location' variable, where the longevity mean is higher in the 'high-cost' group than in the other groups.

Table 6. Longevity means and Standard Deviations of Classified Groups.

	Mean	Std
Tombstone cost		
low-cost	61.49	23.99
medium-cost	61.56	24.67
high-cost	55.81	25.81
Material cost		
low-cost	61.46	23.99
medium-cost	61.11	25.00
high-cost	45.02	22.94
Location cost		
low-cost	59.61	23.52
medium-cost	60.02	22.83
high-cost	62.58	25.30
Accessories cost		
low-cost	61.47	24.12
medium-cost	62.01	25.45
high-cost	58.11	24.88

Figure 8 illustrates the comparison of longevity data distribution by quartiles for the three groups classifying the overall tombstone cost, material cost, height, and accessories cost. It can be seen that the median values of the 'low-cost' and 'medium-cost' groups are quite similar across all variables, while the median value of the 'high-cost' group is lower in variables related to the overall cost, accessory cost, and material cost, with the most pronounced difference seen in the latter. Conversely, it is noteworthy that the median for the 'high-cost' group in the 'cost location' variable is higher compared to the other groups.

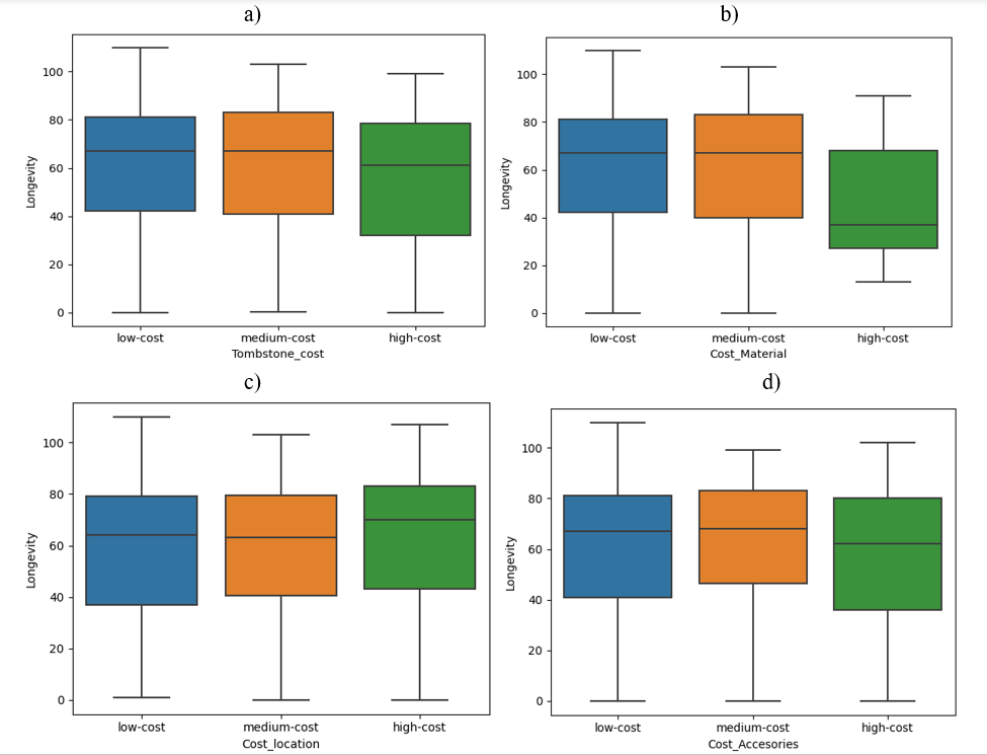


Figure 8: Box and Wisker plots of Longevity for each group. a) Tombstone cost, b) Cost Material, c) Cost location, d) Cost accessories.

Given the observed differences, it is necessary to assess whether these differences are statistically significant. To achieve this, tests such as ANOVA can be employed; however, their use is contingent upon meeting assumptions such as independence of observations, homoscedasticity, and normal distribution. Figure 9 illustrates the probability density function of the dependent variable 'longevity' for the various groups classified by the overall cost of a tombstone. In all three cases, a typical normal distribution curve is not clear; instead, bimodal distribution curves are seen. To corroborate the findings depicted visually, a normality check of the data is conducted using the Shapiro-Wilk test.

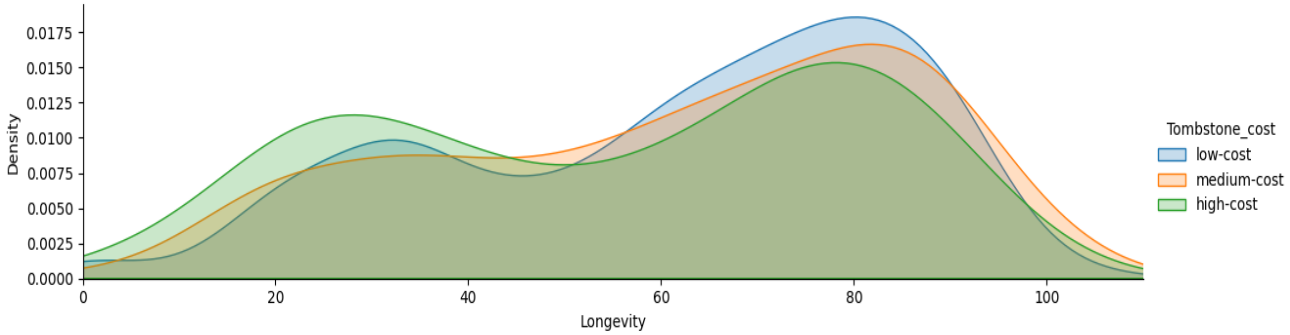


Figure 9: Longevity continuous probability density curve in variable Tombstone cost.

The results from Table 7, obtained when assessing data normality, display p-values < 0.05, showing that the null hypothesis that the data follows a normal distribution must be rejected.

Table 7. Results of assessing data normality using the Shapiro-Wilk test.

	W	P-val	Normal
Tombstone_cost			
low-cost	0.945189	1.025126e-24	False
medium-cost	0.945485	3.268062e-10	False
high-cost	0.930789	1.519773e-07	False

The use of a one-way ANOVA test is ruled out for confirming statistically significant differences due to the non-normality of the data. Non-parametric techniques are considered, and the Kruskal-Wall test is chosen, which can be applied when there are two or more independent groups, qualitative independent variables, and a continuous dependent variable. The 'he 'stats.kruskal' function from the Statistical Functions module (scipy.stats) in Python was used, specifically yielding a statistic of 7.4653 and a p-value of 0.023928. This would allow inferring that there are significant differences between at least 2 groups of the variable 'Tombstone_cost' The same finding was seen for the variables 'Material_cost' and 'Location_cost.' As for the 'Accessories_cost' variable, the test did not yield significant differences among the groups. The summary of the test results is presented in Table 8:

Table 8. Results Kruskal-Wallis's test.

Variable	Statistic	p value
Tombstone cost	7.46	0.023
Material cost	18.89	7.88e-05
Location cost	13.19	0.0013
Accessories cost	4.81	0.090

In order to find which groups exhibit these significant differences, Dunn's test was applied, which is often employed for conducting multiple comparisons between groups following a non-parametric analysis of variance. The 'sp.posthoc_dunn' function from the scikit-posthocs library in Python was utilized, using Bonferroni correction for p-values adjustment. The results obtained are presented in Table 9. It is seen that the longevity of the 'high-cost' group is significantly different from the 'low-cost' and 'medium-cost' groups in the variables 'Tombstone cost' and 'Material cost.' Conversely, in the 'Location cost' variable, there is

only a significant difference between the 'high-cost' and 'low-cost' groups. Multiple comparisons were not assessed for the 'Accessories cost' variable since it did not yield significant differences following the Kruskal-Wallis's test.

Table 9. Results Dunn's test for significantly different variables.

	low-cost	medium-cost	high-cost
Tombstone cost			
low-cost	1.000000	1.000000	0.022078
medium-cost	1.000000	1.000000	0.047000
high-cost	0.022078	0.047000	1.000000
Material cost			
low-cost	1.000000	1.000000	0.000043
medium-cost	1.000000	1.000000	0.000086
high-cost	0.000043	0.000086	1.000000
Location cost			
low-cost	1.000000	1.000000	0.001970
medium-cost	1.000000	1.000000	0.058183
high-cost	0.001970	0.058183	1.000000

As mentioned in the methods section, weights were assigned to the variables that make up the global cost of the tombstone, based on the experience and knowledge of the cemetery research area and the cost ranges that could be achieved according to the values reported by the marble workshops. Logistic regression was used to assess the contribution of the Material cost, Location cost, and Accessories cost variables. Table 10 shows the variable that most significantly affects the tombstone cost is the Material cost, followed by the Accessories cost. As for the cost associated with height, it would have no influence on the global cost.

Table 10. Coefficients calculated by logistic regression for the parameters.

$R^2 = 0.981$	
Params	Coef
Material cost	0.828169
Location cost	-0.002318
Accessories cost	0.242575

In addition to the variables related to the tombstone cost, the influence of other variables on longevity is studied, namely, gender (female or male), type of death (violent or non-violent), and marital status (with a partner or without a partner).

Figure 10 compares the longevity of subjects belonging to each classification group, differentiating by gender, type of death, and marital status. In a), it can be seen that in all three groups, the longevity of female subjects was higher than that of male subjects, with the 'low-cost' group showing the least noticeable difference. In b), differentiation by type of death is shown, revealing that subjects whose cause of death was violent barely reached or surpassed 40 years of life, marking a very clear difference compared to subjects with non-violent deaths. It is noteworthy that subjects with high-cost tombstones lived a shorter life, especially if the death was violent. In c), differentiation by marital status is displayed, with subjects who have a partner having, on average, a longer lifespan than those without a partner.

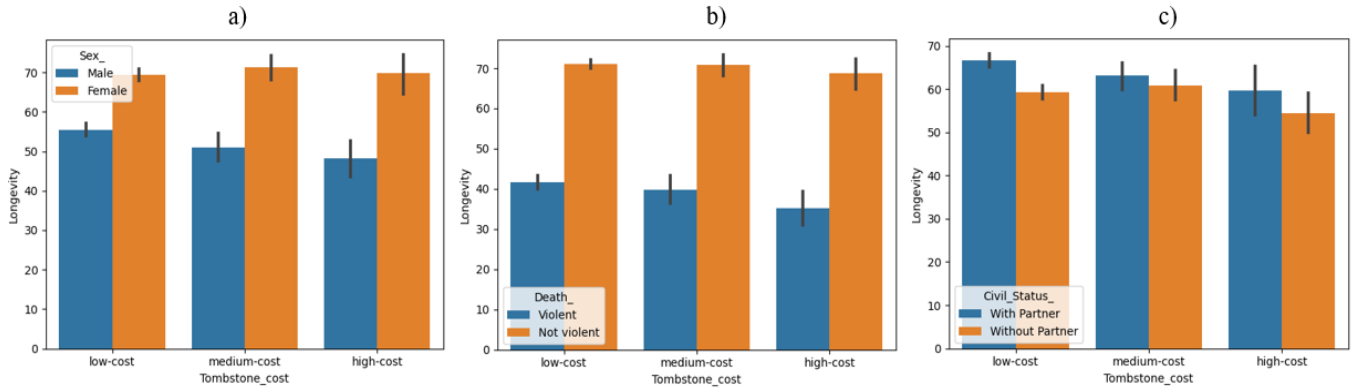


Figure 10: Longevity in the classification groups of global cost differentiated by variables Gender, Death, and Civil Status.

With the visual support of the graphs, differences were found, the statistical significance of which was evaluated using the Kruskal-Wallis's test. The results shown in Table 11 display a p-value < 0.05 for all evaluated variables. Therefore, the longevity achieved by the subjects differs depending on their gender, the type of death, and marital status.

Table 11. Results Kruskal-Wallis's test for variables Sex, Death and Civil Status.

Variable	Statistic	p value
Sex (male, female)	233.47	1.0391e-52
Death (violent, not violent)	736.54	3.38645e-162
Civil status (with partner, without partner)	17.73	2.54521e-05

After finding significant differences in longevity based on the variables Gender, type of death, and marital status, it is of interest to examine the proportion of subjects in the dataset by combining the variables Gender and type of death (Figure 11), Gender and marital status (Figure 12), and type of death and marital status (Figure 13). In Figure 11, it is seen that in female subjects, violent deaths occur in a much smaller proportion compared to non-violent deaths, while in male subjects, the number of violent deaths is similar to non-violent deaths. It is also clear that in the group with high-cost tombstones, more than 60% of male deaths occur.

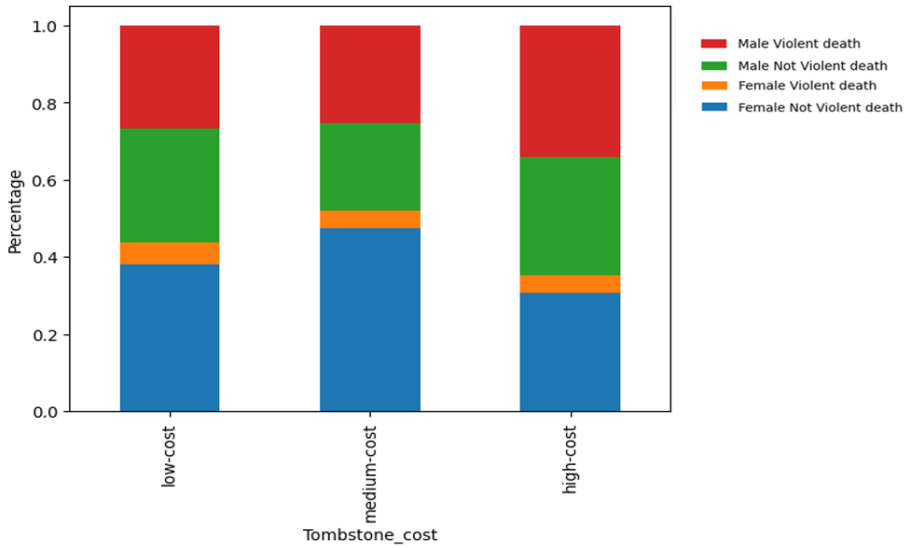


Figure 11: Percentage count by combination of Sex and Death variables.

Figure 12 illustrates the proportion of subjects in the groups classifying the overall tombstone cost by combining the variables Gender and Marital Status. It is evident that the number of subjects without a partner is greater than the number of subjects with a partner in all classification groups, regardless of gender. In the high-cost group, it is observed that it is predominantly composed of male subjects without a partner. It is also noted that the number of female subjects with a partner is lower than the number of male subjects in the low-cost and high-cost groups.

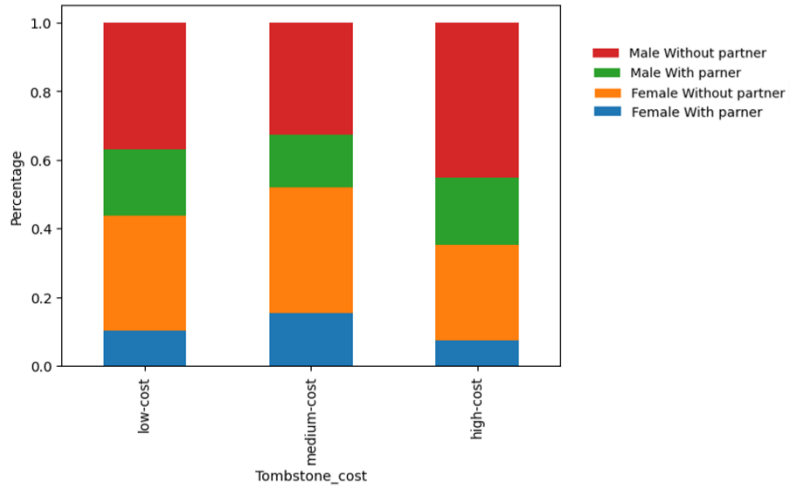


Figure 12: Percentage count by combination of Sex and Civil Status variables.

Figure 13 displays the combination of the variables type of death and marital status. In the graph, it can be seen that the proportion of subjects with non-violent deaths who had a partner is similar in all classification groups. The significant difference that stands out pertains to subjects with violent deaths who did not have a partner, as this part is greater in the high-cost group.

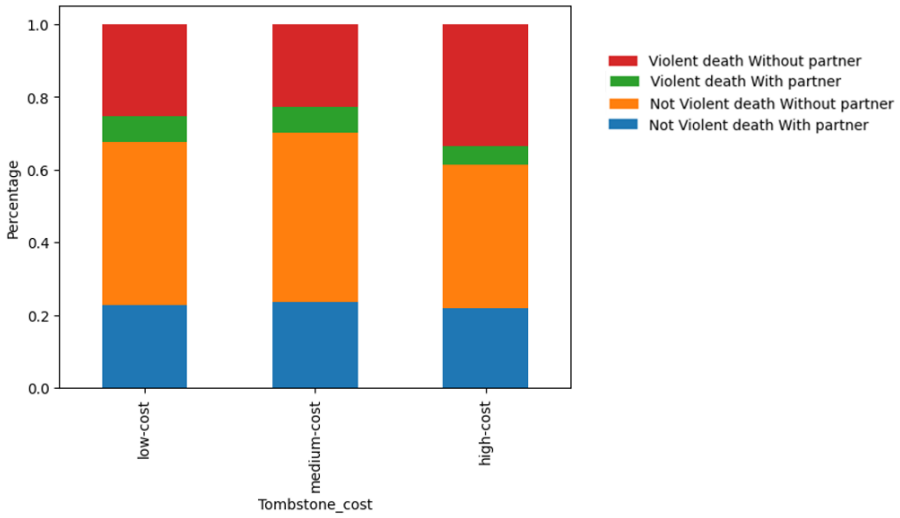


Figure 13: Percentage count by combination of Death and Civil Status variables.

Since the period of death certificate issuance spanning the years 2020, 2021, and 2022 coincides with the COVID-19 pandemic, 47 cases were identified where the cause of death was COVID-19. Figure 14 presents a comparison of longevity between subjects whose cause of death was COVID-19 and those who died from

other causes. It can be observed that subjects who died from COVID-19 had a longer life, which may be related to the fact that this disease primarily affected an older population.

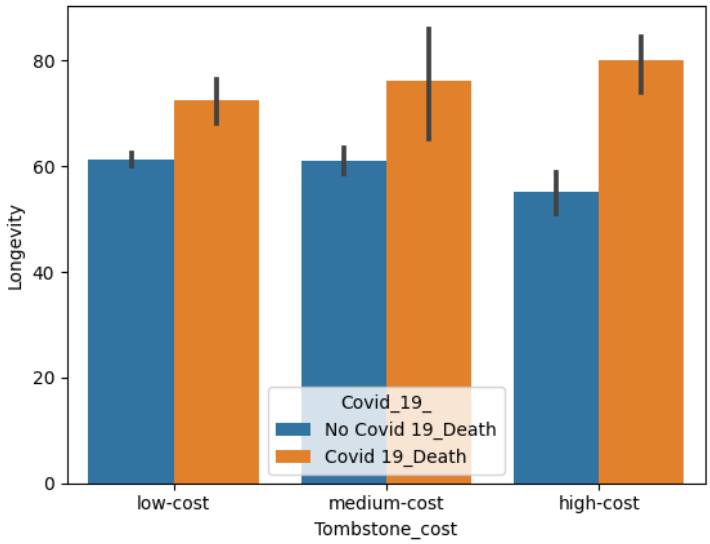


Figure 14: Comparison of Longevity Between Subjects with COVID-19-Related Deaths and Other Causes of Death

4. Discussion

According to the results, individuals whose tombstones were classified in the high-cost group had, on average, a significantly shorter lifespan than those with tombstones classified in the low or medium cost groups, with respect to the overall cost. Considering that the current life expectancy in Colombia is 74 years [25], individuals in this group lived on average 20 years less, suggesting that these subjects experienced sudden or unexpected deaths.

When examining the distribution of longevity among the groups classifying the global cost of tombstones based on the variables Gender, type of death, and marital status, aspects are found that expand the understanding of the results. The intersection of these variables reveals that the lower average longevity in subjects of the 'high-cost' group is due to the fact that most cases belong to violent deaths, which would suggest that the violent death of a young subject leads to the acquisition of a more expensive tombstone. The results revealed a higher proportion of violent deaths in men than in women. The Colombian Institute of Legal Medicine reported that in 2022, out of the total number of violent deaths, 86% were men and 14% were women [26]. In this study, the proportion was 84% for men and 16% for women, closely aligning with the figures reported by official authorities.

The significant differences found according to the subject's gender, where female subjects have a longer life, align with the information disseminated by the Presidential Council for Women's Equality in Colombia [27], which indicates a difference of 6.4 years in life expectancy at birth.

Additionally, it is found that longevity varies according to marital status. This research shows that subjects with a partner had a longer life. This aligns with previously reported studies that suggest this phenomenon is supported by the so-called 'protective effect of marriage' associated with lower morbidity [28].

The aim of this study was to investigate the association between tombstone cost and longevity in a large cemetery in Medellín in Colombia. The results did not support the hypothesis that tombstone cost was associated with longevity. On the contrary, the most expensive tombstones were those of the young. The inverse association between tombstone cost and longevity would suggest that people in Medellín are inclined to spend more on tombstones when commemorating the tragic death of a young person.

Studies in the West support the hypothesis that tombstone cost is associated with longevity [12,13].The height of 843 obelisks from the nineteenth and early twentieth centuries was associated with longevity in a study of eight cemeteries in Glasgow in Scotland [12]. The size of 165 tombstones from the nineteenth and twentieth centuries was also associated with longevity in a study of a large cemetery in Quebec in Canada [13]. This study in Latin America supports the alternative hypothesis that tombstone cost is inversely associated with longevity. We investigated 2,273 tombstones from 2020 to 2022 in the San Pedro Cemetery Museum in Medellín in Colombia. We found that the most expensive tombstones were those of the young. Socioeconomic status was not assessed in the study in Glasgow or the study in Quebec, but it was assumed that affluent families spent more to commemorate the deaths of wealthy individuals [13,14]. Socioeconomic status was not assessed in the present study either and we can only assume that ordinary families spend more to commemorate the deaths of young individuals.

It is quite normal to find tombstones of the young in the San Pedro Cemetery Museum in Medellín. Many of the young will have lost their lives in violent deaths, including gun-related deaths, knife-related deaths, traffic accidents, and suicides. The threat of death is so great in the poorest and most marginalized neighborhoods that it is said colloquially that “they are born with a tombstone on their back”. Drug dealing and violence have been commonplace in Medellín since the 1980’s [29]. Many young people in marginalized communities do not have father figures and they think of drug dealing and violence as their only ways of becoming successful [29]. Gangs run through many neighborhoods and there is an eye-for-

eye philosophy that results in an endless cycle of violence [29]. Violence and violent deaths are also part of the folklore of many communities in Medellín [29, 30]. While many of the young subjects in the present study will not have been gang members, the glorification of violence may help explain why the most expensive tombstones were those of the young.

This study has strengths and limitations. To the best of our knowledge, this is the largest study of the association between tombstone cost and longevity and the only such study in Latin America. Age at death was obtained from death certificates, tombstone cost was carefully estimated, and the analyses were adjusted for potential confounders. The number of subjects in the high-cost group was lower than planned, but the narrow confidence intervals in each group would suggest that the study was large enough to give precise results. Consistency is an important consideration in epidemiology [31] and more research should be conducted in other cemeteries to investigate whether tombstone cost is positively or negatively associated with longevity.

In conclusion, the present study does not support the hypothesis that tombstone cost is associated with longevity. Rather, the present study supports the alternative hypothesis that tombstone cost is inversely associated with longevity. This is the largest study of its kind, and the results would suggest that people in the city of Medellín in Colombia are inclined to spend more on tombstones when commemorating the tragic death of a young person.

5. References

1. Bilas, V., Franc, S., & Bošnjak, M. (2014). Determinant factors of life expectancy. Collegium Antropologicum, 38(1), 1–9.
2. Miladinov, G. Socioeconomic development and life expectancy relationship: evidence from the EU accession candidate countries. Genus, 2020.76(1). https://doi.org/10.1186/s41118-019-0071-0
3. Bezy, J. M. Life expectancy. Encyclopedia Britannica. 2021.
4. Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., & Cutler, D. The association between income and life expectancy in the United States, 2001-2014. JAMA: The Journal of the American Medical Association, 2016. 315(16), 1750–1766.
5. Brønnum-Hansen, H., Östergren, O., Tarkiainen, L., Hermansen, Å., Martikainen, P., van der Wel, K. A., & Lundberg, O. Changes in life expectancy and lifespan variability by income quartiles in four Nordic countries: a study based on nationwide register data. BMJ 2021. Open, 11(6), e048192. https://doi.org/10.1136/bmjopen-2020-048192
6. Bynum, B. Road maps to health. Lancet, 372(9650), 2008. 1626–1627.
7. Marmot, M. G., & Shipley, M. J. Do. Socioeconomic differences in mortality persist after retirement? 25 years follow up of civil servants from the first Whitehall study. BMJ 1996 (Clinical Research Ed.), 313(7066), 1177–1180. https://doi.org/10.1136/bmj.313.7066.1177
8. Shaw, M., Tunstall, H., & Smith, G. D. Seeing social position: visualizing class in life and death. International Journal of Epidemiology, 2003. 32(3), 332–335. https://doi.org/10.1093/ije/dyg17
9. Quintero Cardona, P. S., Gómez Sepúlveda, C. L., & Giedelmann Reyes, M. J. Y yo…, ¿a quién lloro? Manifestaciones de duelo en el Jardín Cementerio Universal de Medellín, Colombia. Antipoda, 2022, 46, 204–234. https://doi.org/10.7440/antipoda46.2022.09
10. Criado, J. C. H. Análisis histórico del proceso de inserción del cementerio central a la dinámica urbana de Bogotá. 2015. Universidad distrital Francisco José de Caldas.
11. Colón, L. Espacios para los muertos y ritos para la memoria. En IDRD, Las ciudades y los muertos. Cementerios de América Latina-Patrimonio Cultural. 2004. (págs. 37-39). Bogotá: Alcaldía Mayor de Bogotá. Obtenido de patrimoniocultural.gov.
12. Davey Smith G, Carroll D, Rankin S, Rowan D. Socioeconomic differentials in mortality: evidence from Glasgow graveyards. BMJ 1992;305(6868):1554-7.
13. Kang-Auger S, Kang-Auger M, Kang-Auger G, Ayoub A, Auger N. Tombstone size and life expectancy: a cross-sectional analysis of cemetery data before the turn of the century. Eur J Epidemiol 2021;36(12):1219-23.
14. Hambleton IR, Howitt C, Jeyaseelan S, Murphy MM, Hennis AJ, Wilks R, Harris EN, MacLeish M, Sullivan L, Group USCAfHDR. Trends in Longevity in the Americas: Disparities in Life Expectancy in Women and Men, 1965-2010. PLoS One 2015;10(6):e0129778.
15. Giedelmann, M. J., & Alvarado, L. J. (. Losas sepulcrales como documentos históricos en el Cementerio Universal de Bucaramanga, Colombia. Anuario de historia regional y de las fronteras, 2013, 18(1), 193–218. https://revistas.uis.edu.co/index.php/anuariohistoria/article/view/3418
16. Pinzón, L. R. P. Historiar la muerte. Representaciones historiográficas sobre la muerte en el nororiente de Colombia. 2010. División de Publicaciones UIS

17. Herrera, J. C. S. Del catafalco al ataúd y a la urna. Cambios significativos en las prácticas funerarias, 2007. Bogotá 1910-2007. Uniandes.
18. Cayuela, L. Modelos lineales: Regresión, ANOVA y ANCOVA. Area de Biodiversidad y Conservación, Universidad Rey Juan Carlos, Marid-España, 2014
19. Dalgaard, P. Analysis of variance and the Kruskal-Wallis's test. Introductory Statistics with R (2002): 111-127.
20. W.H. Kruskal, W.A. Wallis Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association, 47 (260) (1952), pp. 583-621
21. S. Guo, S. Zhong, y A. Zhang, "Privacy-preserving Kruskal-Walli's test", Comput. Methods Programs Biomed., vol. 112, núm. 1, pp. 135–145, 2013.
22. Cortina, V. G. Aplicación de la metodología crisp-DM a un proyecto de minería de datos en el entorno universitario. 2015. Universidad Carlos III de Madrid
23. San Pedro Cementerio Museo. San Pedro Cementerio Museo. Internet: https://cementeriosanpedro.org.co (accessed January 2023)
24. World Bank. Life expectancy at birth, total (years) - Colombia. Internet: https://data.worldbank.org/indicator/SP.DYN.LE00.IN?end=2019&locations=CO&start=1960
25. DANE - DANE. (s/f). Gov.co. Recuperado el 2 de octubre de 2023, de https://www.dane.gov.co/index.php?option=com_content&view=article&id=853&Itemid=28&phpMyAdmin=3om27vamm65hhkhrtgc8rrn2g4
26. Gov.co. [En línea]. Disponible en: https://www.medicinalegal.gov.co/documents/20143/742818/Boletin_Enero_2022.pdf. [Consultado: 02-oct-2023].
27. C. Generales, "Características poblacionales de las mujeres en Colombia", Gov.co. [En línea]. Disponible en: https://observatoriomujeres.gov.co/archivos/publicaciones/Publicacion_258.pdf. [Consultado: 02-oct-2023]
28. Jia H, Lubetkin EI. Life expectancy and active life expectancy by marital status among older U.S. adults: Results from the U.S. Medicare Health Outcome Survey (HOS). SSM Popul Health. 2020.
29. Duzán MJ. Sicarios. Edtion ed. Crónicas que matan. Bogotá, Colombia: Penguin Random House, 2012.
30. Riaño-Alcalá P. Jóvenes, memoria y violencia en Medellín: una antropología del recuerdo y el olvido. Medellín, Colombia: Editorial Universidad de Antioquia, 2006
31. VanderWeele TJ, Lash TL, Rothman KJ. Causal Inference and Scientific Reasoning. Edtion ed. In: Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ, eds. Modern Epidemiology. Philadelphia: Wolters Kluwer, 2021