

Factors Contributing to School Dropout at the University of Cauca

Maria Falla¹

Rubiel Vargas-Cañas¹

¹Universidad Del Cauca (Unicauca), Cauca – Colombia

ABSTRACT – Factors Contributing to School Dropout at the University of Cauca. University dropout is a major problem in Latin America and the Caribbean; in Colombia, around 50% of students who enter university complete their studies. This phenomenon is multi-causal; therefore, it is necessary to perform an analysis in different scenarios, in order to create strategies that help permanence and graduation. This paper presents an alternative to identify some factors that lead a student to drop out of the Universidad del Cauca, based on data science methodologies. It was found that desertion is influenced by academic factors, such as areas of knowledge, semester and average, social context such as stratum and place of origin, and individual factors such as age and gender.

Keywords: College Dropout. Dropout Factors. Data Science.

RESUMEN – Factores que contribuyen a la Deserción Escolar en la Universidad del Cauca. La deserción universitaria constituye un importante problema en América Latina y el Caribe; en Colombia alrededor del 50% de los estudiantes que entran a la universidad, culminan sus estudios. Este fenómeno es multicausal, por tanto, es necesario realizar un análisis en diferentes escenarios, para crear estrategias que ayuden a la permanencia y graduación. Este trabajo presenta una alternativa para identificar algunos factores que llevan a que un estudiante deserte en la Universidad del Cauca, basada en metodologías de ciencia de datos. Se encontró que en la deserción influyen factores académicos, como las áreas de conocimiento, semestre y promedio, de contexto social como estrato y lugar de procedencia e individual como edad y género.

Palabras-clave: Deserción Universitaria. Factores de Deserción. Ciencia de Datos.

Introduction

In 2007, the Organization for Economic Cooperation and Development (OECD) stipulated that a well-educated and skilled population is essential for the social and economic well-being of both countries and individuals (OECD, 2007). Education plays a central role in providing individuals with the knowledge, skills, and competencies necessary to effectively participate in society and the economy. Therefore, it is crucial to focus on increasing retention and graduation rates at all educational levels. In this context, studies on student dropout rates in Latin America and the Caribbean highlight the significant problem faced in this field, particularly within public education systems (Pineda; Pedraza, 2011).

A study conducted by the International Institute for Higher Education in Latin America and the Caribbean (IESALC), a specialized institute within the United Nations Educational, Scientific and Cultural Organization (UNESCO), revealed that the average student dropout rate in Latin American and Caribbean universities was 50% in 2004. Specifically for Colombia, the dropout rate was 49% (IESALC, 2006).

Higher education in Colombia has been reserved for small segments of the population. These include those who manage to meet the conditions to secure one of the limited spots in public universities (state institutions primarily funded by the government) or those with the financial resources to enroll in private universities (Gómez Campo; Celis Giraldo, 2009). Since the 1990s, student dropout rates have gradually and consistently increased, with no significant differences between public and private universities. Consequently, dropout has become a structural characteristic of the Colombian university system and a focal point of concern for the state (Barragán Díaz; Patiño Garzón, 2016).

Based on these results, the Colombian national government set a goal to achieve a more equal and supportive society, focusing on four fundamental aspects: increasing coverage, quality, relevance, and financing within the education system. Since 2003, the Ministry of National Education (MEN) – the ministry responsible for formulating national education policies – has promoted actions to address the phenomenon of dropout. Notable initiatives include: System for Prevention and Analysis of Dropout in Higher Education (SPADIES) - This system consolidates and organizes information, allowing for tracking of the academic and socio-economic conditions of students who have entered higher education in the country (Guzmán; Durán; Franco; Castaño; Gallón; Gómez; Vásquez, 2009) - Creation of Calls for Proposals, these calls support the development of projects related to dropout prevention, and Implementation of Dropout Rates as Quality Indicators, Dropout rates serve as indicators of effective management and quality within higher education institutions (IES).

In recent years, the Ministry of National Education (MEN) and the State University System of Colombia (SUE) – an organization

comprising thirty-four public universities that regulates higher education in Colombia – have directed their efforts toward studying the phenomenon of university student dropout. This focus comes in response to state guidelines that aimed to increase coverage and the number of university students. However, this expansion did not translate into equivalent improvements in student retention and graduation rates. According to research by Gvirtz and Oría (2016) and Hurtado-Tobón, García-González, Galvis-Soto, and Méndez-Parra (2022), approximately 50% of students successfully complete their studies. These studies confirm that graduation and degree attainment rates remain low, and the trend in dropout rates has persisted at around 50% over the past two decades.

Taking the above into account, Colombian higher education institutions (IES) have designed and developed various proposals to address the issue of student dropout. Several of these initiatives receive support from the Ministry of National Education (MEN) (Patiño Garzón; Cardona Pérez, 2012). For instance, research has been conducted to quantify the phenomenon and identify the determinants of student dropout. However, the dispersion of results makes it challenging to identify clear trends (Rojas Betancur; González, 2008). Additionally, there is not enough evidence or databases available to make a comparison regarding the factors and size of student desertion (Gartner Isaza; Dussán Lubert; Montoya, 2016).

The absence of general rules for mitigating student attrition has compelled IES to construct their own analytical sources, emphasizing the importance of institution-specific databases for proper contextualization. Meanwhile, the Colombian government aims to standardize dropout indicators at the national level (Higuera-Martínez, 2017).

Since 2010, the Ministry of National Education (MEN) has required universities to identify and monitor factors associated with student dropout and strategies aimed at reducing it (Colombia, 2013). At the University of Cauca, following the ministry's guidelines, a process has been established to create various programs and tools for analyzing, tracking, and managing dropout within the institution. As part of this effort, a diagnostic process was conducted to identify the different causes or factors related to dropout among students who formally initiated the cancellation process. This involved the completion of forms and psychological consultations. It's important to note that the number of students who go through the cancellation process is very low compared to the total number of dropouts. Therefore, there may be additional information not considered, which could provide a more accurate understanding of the dropout phenomenon within the institution.

The diagnostic process initiated in 2010 was repeated in 2014 to analyze dropout rates for the period 2010-2013. Subsequently, in 2016, data from the 2015-2016 period was included (University of Cauca, 2016). Starting from 2017, the plan is to conduct this diagnostic process every three years, ensuring a continuous characterization of stu-

dents. This involves updating institutional information systems, such as the Integrated Enrollment and Academic Control System (SIMCA), which handles academic and financial enrollment processes, among other services. Additionally, national systems like SPADIES are utilized (University of Cauca, 2017). Each process involves a semester-based analysis, allowing visibility into students who cease enrollment. These students are grouped by faculty, program, study level, schedule, campus, age, gender, and other relevant factors. This information facilitates a general tracking of dropout rates and serves as the primary data source for conducting more in-depth studies on the determinants of institutional dropout. Consequently, efforts should be directed toward creating and enhancing information systems and analytical models, such as the one developed in this study.

Based on these diagnostics, programs have been created to monitor dropout rates and the key factors related to this phenomenon. These initiatives aim to establish a support network that can improve student retention and graduation rates. Currently, the University of Cauca has a program specifically dedicated to student retention called “Permanecer”. Additionally, other programs like “Tramados”, which focuses on preventing and addressing substance abuse and other addictions, significantly contribute to student persistence. Since 2017, an early warning system has been implemented, following a periodic referral protocol. This system not only detects students at risk of academic dropout but also enables timely interventions by various stakeholders, including program coordinators and parents.

As seen so far, there is currently no in-depth and specific mechanism for analyzing the factors related to student dropout at the University of Cauca beyond the academic realm. The existing data primarily comes from registration, enrollment, and academic record tracking processes managed by the Division of Admissions, Registration, and Academic Control (DARCA) through the SIMCA system.

In this document, the results obtained from an exploratory data analysis process is presented, focusing mainly on academic and social factors. These factors include faculty, program, birthplace, age, city of origin, city of residence, gender, and participation in Black communities, among others. By applying data mining techniques, it can be identified some factors that exhibit a strong relationship with student dropouts at the University of Cauca during the period from the first semester of 2018 to the first semester of 2019.

The work is carried out at the University of Cauca, a public institution whose funding is provided by the national government and territorial entities. It also receives resources from stamps (taxes on acts involving territorial entities, created by the departmental assembly). Additionally, the university generates its own resources through activities related to education (tuition fees, other services), research, and extension (Bayona-Rodríguez; Rueda-Delgado; Ome, 2022). The University of Cauca was founded in 1827 (196 years ago) and currently offers a wide range of academic programs, including forty-three under-

graduate programs and forty-eight graduate programs. It is located in the municipality of Popayán, the capital of the Cauca Department in southwestern Colombia.

The Cauca Department, where the university is located, primarily relies on agricultural and livestock production, forestry exploitation, fishing, and commerce (Beltrán CH; De David, 2015). Additionally, it boasts significant ethnic and cultural diversity, with 62% of the population being of indigenous and Afro-Colombian descent. However, this diversity, combined with the presence of armed groups controlling territories and engaging in illegal economies, as well as political and economic elites controlling productive land, has made Cauca one of the departments most affected by armed conflict and conflicts related to land concentration, violence, poverty, and economic underdevelopment (Tose Vergara; Ortiz Ruiz, 2019).

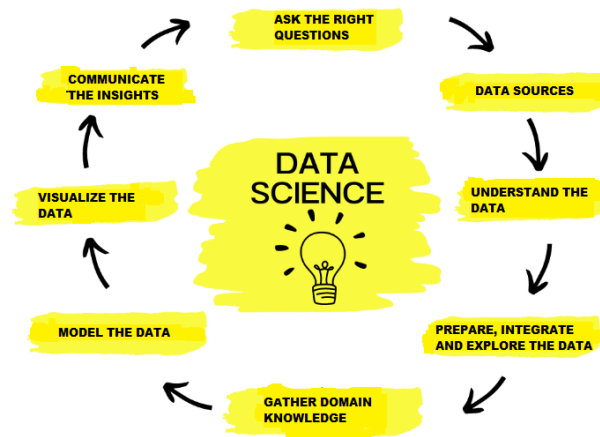
El Cauca shares borders with the departments of Valle del Cauca and Tolima to the north, Huila and Caquetá to the east, Nariño and Putumayo to the south, and the Pacific Ocean to the west. Most of the students at the University of Cauca come from these regions. Despite the socio-economic conditions and violence, they recognize this institution as one of the best opportunities for higher education in the region and the country.

This work is primarily developed through the methodology, analysis, and results sections, followed by the conclusions. The methodological process consists of seven phases: Formulation of Questions, Identification of Data Sources, Understanding the Information, Data Integration, Collection and Verification of Expert Knowledge, and Data Modeling and Exploratory Analysis. The analysis and results section is further divided into the following subsections: Understanding and Integrating Information, Exploratory Data Analysis and Visualization, Dropouts by Program and Faculty, Dropout, Semester, Age, and Gender, Dropout and Academic Performance, Correlation of Data for Students Abandoning Their Professional Career, and Dropout and Geographic Conditions.

Metodology

The development of this work followed a methodological process for data exploration divided into eight phases (Figure 1): formulation of questions, identification of data sources, understanding the information, data integration, collection and verification of expert knowledge, data modeling, visualization, and communication of discoveries. This methodology, known as the Data Science Wheel (Martínez; Viles; Olaizola, 2021), serves as a general guide for applying methods, processes, and arguments to achieve expected results. In this work, the modeling stage is narrowed down to an exploratory data analysis, which reveals patterns in the data that will later serve as the basis for hypotheses to be tested in future studies.

Figure 1 – Methodological Process for Data Exploration



Source: Generated by the authors.

Question Formulation – Correct Questions

Within the university’s school dropout prevention programs, students voluntarily attend these. However, with the information provided so far by the Division of Admissions, Registration, and Academic Control (DARCA), an initial follow-up of all students can be done. By applying Data Science techniques based mainly on descriptive statistics, data grouping and ordering, already known at many areas of knowledge (Morales Rodríguez; Caro Silva; Contreras Bravo, 2022), a closer analysis of the university’s reality can be made and finally identify the main factors affecting student retention. This project was developed around the question, “What are the main factors that lead a student to drop out of their studies at the University of Cauca?” However, to get an accurate answer and in line with the problem within the university, different questions that arise as the exploration and analysis of the Institution’s data progresses must be taken into account, such as: What is the total number of dropout students in the first semester of 2018 at the University of Cauca? How is dropout shown in the different Programs and Faculties of the University of Cauca? Are academic factors such as career average and semester average decisive in the process of abandoning the program initially enrolled by students? Can personal factors such as age and gender be related to dropout? What is the relationship between dropout and social factors such as stratum and place of origin?

Identification of Data Sources

The MEN, committed to providing more accurate data that allows tracking the phenomenon of dropout, through the System for the Prevention and Analysis of Dropout in Higher Education Institutions

(SPADIES), has consolidated mainly academic and socioeconomic information from different sources since 1998. This tool is enriched with information reported by higher education institutions to the National Information System of Higher Education Institutions (SNIES) and from two external sources. The first is the Colombian Institute for Educational Credit and Technical Studies Abroad (ICETEX) - a financial entity linked to the MEN. The second is the Colombian Institute for the Evaluation of Education (ICFES) - an entity specialized in evaluation (Hurtado-Tobón; García-González; Galvis-Soto; Méndez-Parra, 2022). This source of information is interesting, however, it only contains data from first-year students, and it only allows specific queries by variables such as sex, age, stratum, etc., and the comma-separated values files (csv) show consolidated information mainly in percentages, meaning that there is no access to raw data. Then, the Office of Planning and Institutional Development of the University of Cauca is consulted. This is a technical advisory dependency attached to the rectorate, which assists in academic-administrative management, and among its functions is to detect students who do not enroll semester by semester (with information provided by SIMCA). However, dropout students are not detected, according to the definition given by the MEN.

Considering the above, an initial approach to the study of student dropout at the University of Cauca was developed by exploring the data provided by the Division of Admissions, Registration, and Academic Control, a department attached to the Vice-Rector for Academic Affairs and responsible for carrying out the processes of registration, admission, enrollment, archiving, custody of the academic records of graduates and carding (through the SIMCA system), among others. Through these processes, the raw data was obtained which finally allowed the creation of the working database.

Understanding the Information

One of the main objectives of this work is to identify all students who do not enroll for two consecutive semesters (without registering graduation or without being withdrawn for disciplinary reasons) and who are therefore considered dropouts, according to the definition given by the Ministry of Education. The data worked here are mainly academic and some social, which are collected in the process of registration, admission, enrollment, and follow-up of the academic records of the students, such as: faculty, program, level, schedule, modality, headquarters, status, new status, grade, identification type, identification, code, birth, age, city of origin, city of residence, place of issuance of ID, gender, black communities, indigenous peoples, stratum, disability, career average, semester average. These data correspond to the students enrolled from the first semester of 2018 to the first semester of 2019.

Data Integration

In this part, the preparation, integration, and exploration of the data is carried out by applying functions that allow changes of varia-

bles and operate them later, changes in the name of the attributes to identify them more easily in the different files and to be able to relate them. In this stage, a standardization of the date format, handling of duplicate and lost data (López Gómez, 2018) was initially carried out. As a result, three comma-separated files (csv) were obtained, one file per semester, each with 24 features, and with these files, a “raw data” dataset was created that allows exploring many important characteristics in the group of students identified as dropouts.

In a new dataset, seven features are preserved (code, identification, status, new status, grade, entry period, and age), and the information is limited to students enrolled at the professional level and at the main headquarters. With the new dataset, a process of identifying students who did not enroll for two consecutive semesters and who are known as dropouts is developed. For this process, each student enrolled in the first semester of 2018 was identified and searched in the following two semesters (second semester of 2018 and first of 2019), for this the group of students graduated in the first semester of 2018 was not considered. After recognizing the target group, a column was added to the initial file where dropout and non-dropout students can be identified.

Collection and Corroboration of Expert Knowledge

The data initially provided by DARCA have several features closely related to the performance and academic process of the students, these attributes are: status, new status, career average grade, and semester average grade. To correctly interpret the statuses, it was necessary to consult the person in charge of managing this database, which allowed understanding that all statuses are current or last (upon leaving the university), the status per semester is not saved, which would be appropriate for this study. After making the corresponding consultation, the “new status” attribute was created, which indicates the graduation date, in order to group the graduated students each semester and correctly identify the dropout students. In this process, the meaning of each of the attributes in question was also understood.

According to the person in charge, the career average grade attribute corresponds to the cumulative average up to the last semester that the student has taken at the university, and the semester average is a created attribute, which shows the cumulative average per semester, the status attribute shows the options of Active or Inactive depending on the academic or financial condition in which the student is, for example, there are options: Inactive Financial Enrollment (has not paid his enrollment), Inactive two Programs, Inactive Cancels Semester and Inactive Previous Debts.

Within this attribute, there are also three options related to academic performance (balanced 2 subjects, balanced R3 and balanced first semester), a graduated status, and another deceased. A student who loses his place due to poor academic performance is known as balanced.

Data Modeling - Exploratory Analysis

After executing the data integration process, descriptive statistics techniques and functions, variable grouping, and data ordering are applied, as different IES have done (Hernandez Gonzalez; Melendez Armenta; Morales Rosales; García Barrientos; TecpanecatI Xihuitl; Algreto, 2016; Himmel, 2002) to obtain information from each variable, sets of variables, as well as the entire dataset (Ahuja; Kankane, 2017). In this part, work was mainly done with the numerical variables of age, stratum, career average, semester average, and the categorical variables program, faculty, city of origin, both of the dropout students, as well as the students who continue with their studies.

Data Visualization

To discover behaviors, patterns, and trends of the different variables; find atypical variables and analyze correlations, graphical and statistical functions and tools were applied for the visualization and analysis of the data, which could later be converted into factors related to student dropout. The results obtained from the exploration, visualization, and analysis will allow the formulation of hypotheses related to the possible causes that lead to student dropout at the University of Cauca.

Results and Analysis

Understanding and Integrating Information and Consulting Experts - Raw Data Dataset

With the csv files, the raw data dataset (Table 1) was structured, from this a smaller dataset was built, taking as reference the guidance of the person in charge of data management at DARCA.

Table 1 – Dataset Structure: corresponds to information of all enrolled students

Name	Description	Type
ENTRY_PERIOD	Period in which the student enters the Institution	Number
FACULTY	Faculty to which the student belongs	Text
PROGRAM	Program in which the student enrolls	Text
LEVEL	This can be Professional or technological	Text
SCHEDULE	Schedule of the program to which the student belongs, can be Daytime or Nighttime	Text
MODALITY	This is purely face-to-face in the studied semesters	Text
HEADQUARTERS	Headquarters of the University of Cauca, these are located in Popayán, Santander de Quilichao, Miranda and Silvia	Text

Factors Contributing to School Dropout at the University of Cauca

STATUS	Current status of the student or status at the time of leaving the University, these appear as ACTIVE or INACTIVE, according to the academic and/or financial condition within the University	Number
NEW_STATUS	Graduation date of the student	Number
GRADE	Current grade of the student or at the time of leaving the University, these appear as REPETITIVE, BALANCED, LOW PERFORMANCE, REGULAR, according to the academic condition of the student	Text
IDENTIFICATION_TYPE	Type of identification of the current student	Number
IDENTIFICATION	Number of the type of identification	Number
CODE	Code of the student given in the Institution	Number
AGE	Age of the enrolled student	Number
CITY_OF_ORIGIN	City of origin of the enrolled student	Text
CITY_OF_RESIDENCE	City of residence of the enrolled student	Text
PLACE_OF_ISSUANCE_ID	Place of issuance of the identification document	Text
GENDER	Gender of the enrolled student	Text
BLACK_COMMUNITIES	Student who belongs to black communities	Text
INDIGENOUS_PEOPLES	Student who belongs to Indigenous communities	Text
STRATUM	Stratum of the student who enrolls	Number
DISABILITY	Disability of the student who enrolls	Text
CAREER_AVERAGE	Cumulative average of the career of the enrolled student	Decimal
SEMESTER_AVERAGE	Cumulative average of the semester in which the student enrolls	Decimal

Source: Generated by the authors.

With the defined dataset, it was continued with a refinement of it to improve the data quality, avoiding mainly coherence problems when applying the different functions. Within the process, mainly the cleaning of empty data, the correction of data to handle them correctly, as well as the handling of dates is carried out.

Exploratory Analysis and Data Visualization

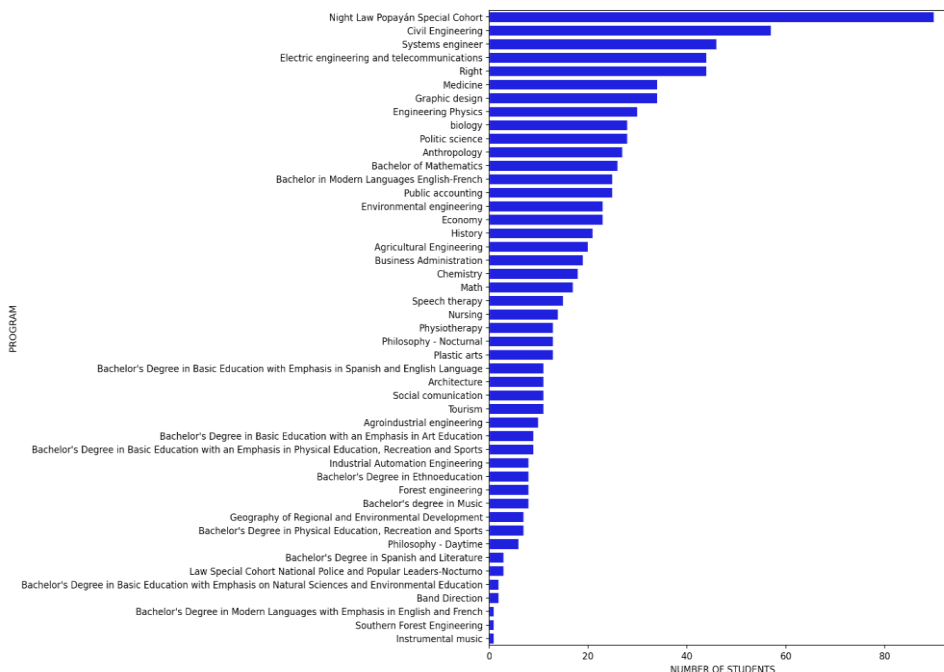
The number of students who appear as enrolled in the first semester of 2018, at the Popayán headquarters at the professional level is 14,265 students, after removing the graduated students (643), there are 13,622 students left, these are searched in the following two semesters and it was found that 886 of these students are not reported as enrolled in those semesters, that is, all these students have voluntarily or involuntarily abandoned their career temporarily or permanently and therefore meet the dropout criterion proposed by the MEN. The factors related to involuntary dropout are those associated with the institution's regulations, that is, academic aspects (non-disciplinary), which are analyzed using data shown in the dataset. One of the greatest advantages of this analysis is that it takes into account the information of all students who drop out and not just those who carry out the formal cancellation process.

Dropouts by Program and Faculty

Once the group of students considered dropouts and non-dropouts was found, an exploration of the variables or possible factors that most influence a student to abandon their studies was carried out, thus some cases of dataset analysis are proposed. As an initial measure, dropout by program (Figure 2) was detected. According to the results obtained, the programs with the highest number of dropout students belong, in their order, to social sciences, engineering, basic sciences, and health sciences. The ten programs with the highest dropout rate are: Special cohort night law, Civil Engineering, Systems Engineering, Law, Electronic and Telecommunications Engineering, Medicine, Graphic Design, Physics Engineering, Biology, and Political Science. However, when looking at the percentage of students who drop out in relation to the total number of students enrolled by program, the figures vary significantly.

The highest dropout percentage is between 14% and 18% and is presented in the programs of: History, Bachelor's Degree in Physical Education, Architecture, Philosophy - Night, Bachelor's Degree in Languages, and Graphic Design; the remaining programs show a dropout rate between 8 and 9%. When comparing the previous results, it is found that both the number of students and the percentage of dropouts are important in the programs related to social and basic sciences, however, the same is not shown in engineering (8%) and health sciences (4.5%), although the number of dropout students is significant, the percentage is not as much as would be expected.

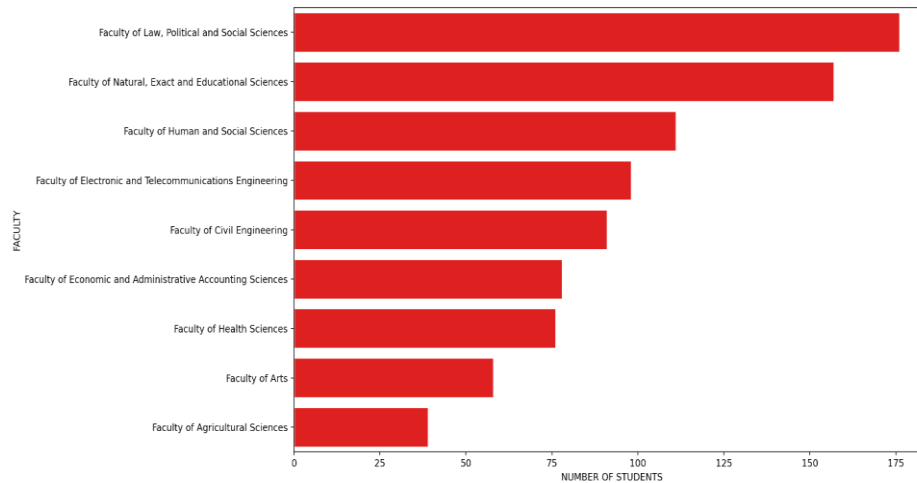
Figure 2 – Number of Students who drop out by Program



Source: Generated by the authors with data from the University of Cauca.

According to Figure 3, the number of students who drop out of their study program is higher in the faculties of Social Sciences and Exact Sciences, followed by Engineering, Economic Sciences, Health Sciences, and finally the faculties of Arts and Agricultural Sciences. When comparing the percentage of dropout students with respect to the total number of students enrolled by faculty, the order changes as the faculty of arts shows a higher percentage close to 8%, then there is the faculty of Exact Sciences with 7.3%, then the faculties of Social Sciences and Engineering are interspersed with percentages between 6.8% and 4.9%, finally there are the faculties of Economic Sciences, Health and Agricultural, with a range of 4.8% - 3.4%. The order found of the percentages is similar to the number of dropout students by faculty shown initially, however, it is found that, although the number of dropout students in the faculty of Arts is small compared to the other faculties, the percentage of dropout students is higher. On the other hand, the faculty of law, political sciences and social sciences, although it has the highest number of students who drop out of their career, the percentage of dropout students is around 5%, ranking fifth among all faculties.

Figure 3 – Number of Students who Drop Out by Faculty



Source: Generated by the authors with data from the University of Cauca.

In general, it is noted that both the programs and the faculties have a higher dropout rate in Social Sciences, Exact Sciences, and Engineering, and for the faculties, they are followed by Economic Sciences, Health Sciences, and finally the faculties of Arts and Agricultural Sciences.

This result partially agrees with the dropout rates in Colombia by areas of knowledge, where there is a higher percentage in engineering, followed by fine arts, economics and related, agronomy, education sciences, basic sciences, social sciences, and with a lower percentage in health sciences. The dropout in engineering it seems similar in Cauca University and the rest of the country and although social sciences and basic sciences are in another order, their percentage is also significant in most higher education institutions in the country (Guzmán; Durán; Franco; Castaño; Gallón; Gómez; Vásquez, 2009).

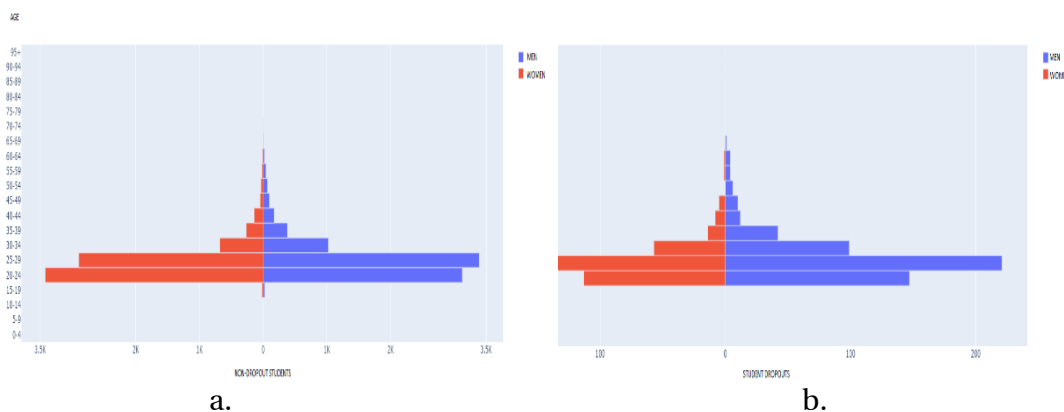
The dropout by areas of knowledge can be related to the degree of difficulty of the program, its duration, relationships with peers, teachers and administrators, problems in the transition and coupling process (Pascarella; Terenzin, 2005), in vocational guidance (Tinto, 1989), deficiencies in the cognitive and human capacities and skills important in the program (Colombia, 2010), etc., these are situations that can be analyzed in greater depth, when looking for information for example from the curriculum of the career and the tutorials that students normally take in the permanence and graduation programs (in this program support is offered in literacy, mathematics, psychosocial and socioeconomic support).

Dropout, Semester, Age, and Gender

Several studies related to dropout have estimated that the majority of students who enter university drop out at an early age and in the first semesters of training (Colombia, 2010), which has led the MEN to seek strategies to study and decrease this phenomenon. The ages of the students who enrolled in the first semester of 2018 at the Popayán headquarters at the professional level are mainly in the range of 20 to 29 years (Figure 4a), within the group of dropout students, the age range mainly oscillates between 21 and 35 years (Figure 4b). When looking for first-year students within the total group of dropout students, it is found that they correspond to 24.7%, that is, they are approximately a quarter of the group. When looking for the number of dropouts by semester, it is found that students who drop out while in their second semester are 7.2%, being in the third semester 5.6%, in the fourth semester 5.7%, fifth semester 6%, sixth semester 4.4%, in general, it is observed that the percentages decrease as higher semesters are reached, therefore, it is evident that students are more likely to abandon their studies in the first semesters of their career.

In relation to gender, within the group of enrolled students, men represent 52.9% and women 47.1%, within the group of students who do not enroll for two consecutive semesters, it is found that 61.8% are men and 38.2% are women, it is clear that in the first semester of 2018 the majority of dropout students are men (Figure 4b), it can also be seen that this difference is preserved in all age ranges, even in the range where there is a higher number of enrolled women (range of 20-24 in Figure 4a).

Figure 4 – Age Pyramids of Comparison of Age and Sex of the Student Population a. non-Dropouts and b. Dropouts, Belonging to the First Semester of 2018

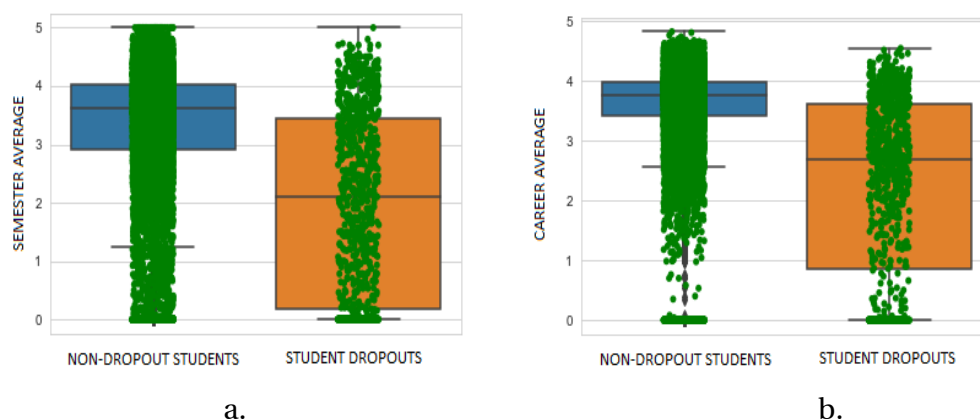


Source: Generated by the authors.

Dropout and Academic Performance

Many of the cases of dropout are related to academic conditions such as low performance due to the impact on their academic average. By applying standardized statistical methods, a group of numerical data such as semester and career averages can be graphically represented, allowing to highlight aspects of the frequency distribution, through quartiles and their outliers. When comparing the box plots (Figure 5), it is noted that the distributions in the 50% of the total sample of dropout and non-dropout students in the two averages are asymmetric with negative biases, meaning that most of the data are distributed in the smaller values of the range; in addition, the median is closer to the higher academic averages, which indicates that in the upper part of the box there is a higher concentration of averages in a smaller range, compared to the lower part of the box, however, dropout students show semester averages with a positive bias and the median is closer to the lower averages.

Figure 5 – Box Plots of the Averages of a. Semester in which they Drop Out and b. Accumulated of the Career



Source: Generated by the authors.

Academic performance is a factor that can clearly affect students' determination to abandon their educational project, coupled with the academic situation experienced in the semester in which the decision is made. The ranges and medians of the semester and career averages (Table 2) are lower in the semester average, this may be related to the student's decision to abandon their studies, as in the semester in which the decision is made, their average is even more affected. If we now move on to observe the values of the ranges, it is clear that academic performance is lower in the group of dropout students, where the lower ranges for them are below three (minimum passing grade) and for the group of non-dropouts they are above 2.9.

It can also be seen that a large group of students who are dropouts have good academic performance (see upper ranges - 25% of the sample), therefore, for this group the decision to abandon their studies must be affected by other causes different from academic ones.

Table 2 – Medians and Ranges of Averages for Dropout and Non-Dropout Students at the University of Cauca in the First Semester of 2018

	Semester Average		Career Average	
	Dropouts	Non-Dropouts	Dropouts	Non-Dropouts
Median	1.99	3.11	2.29	3.58
Upper Range	2.10 - 3.50 (1.40)	3.60 - 4.00 (0.4)	2.69 - 3.62 (0.93)	3.76 - 3.99 (0.23)
Lower Range	0.18 - 2.10 (1.92)	2.90 - 3.60 (0.7)	0.86 - 2.69 (1.83)	3.42 - 3.76 (0.34)

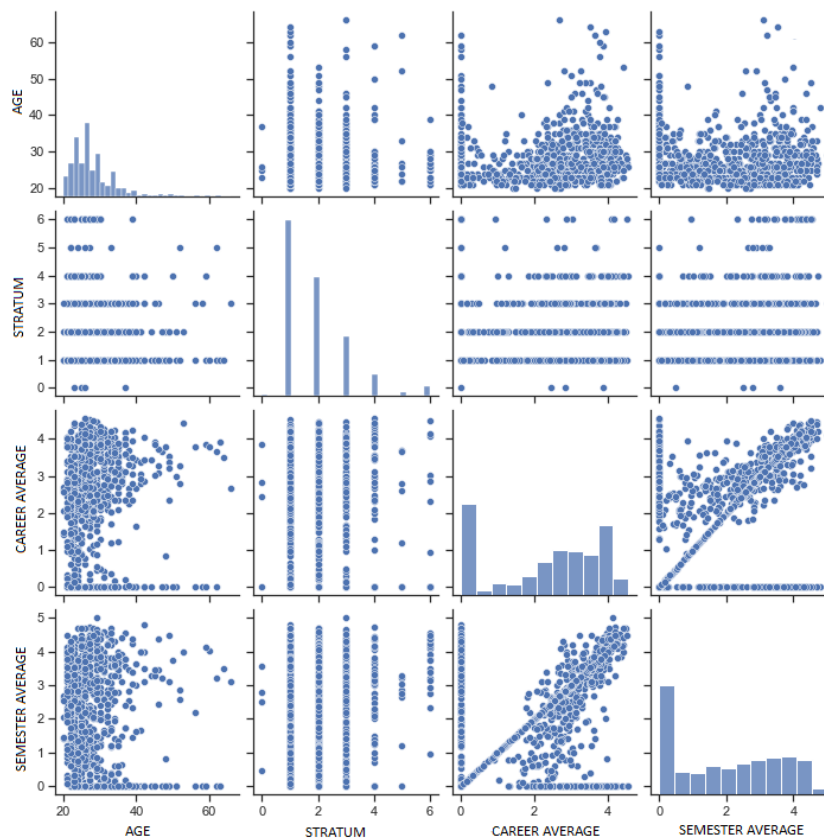
Source: Generated by the authors.

On the other hand, the diagrams show that for non-dropouts there is 25% of the sample below the lower range shown in the table, in which students have semester averages below 2.9, and career averages below 3.42, that is, there are also students with low performance who continue their studies.

Correlation of Data of Students who Abandon their Professional Career

Initially, a set of graphs was obtained with bivariate distributions of age, stratum, career average, and semester average of the students considered dropouts (Figure 6).

Figure 6 – Bivariate Distributions of Age, Stratum, Career Average, and Semester Data of Dropout Students from the First Semester of 2018



Source: Generated by the authors.

The main relationships found are as follows: Age and Age. According to the graph, the majority of students are under 40 years old, it seems that there is a larger number of students aged between 20 and 30 years, this is related to the results shown previously in which it is indicated that the majority of dropout students are aged in the range of 21 to 35 years.

The relationship between Age and Stratum is inverse, that is, older students come from lower strata, it is also observed that there is a large number of dropout students of all ages in lower strata, it can be highlighted that the number of students in higher strata is much lower, as well as their ages, this may be related to young students who belong to families with a stable economic income. The other group, on the contrary, may have economic difficulties or of another origin. The strata that present the highest dropout are 1, 2, and 3, therefore, the majority of students who drop out of their studies belong to low-low, low, and medium-low strata.

The relationship of Age and Career Average is positive, as age increases the career average increases. There is a concentration of data in the grade range of 2 to 4, and ages between 20 and 40 years, in the group of students with grades less than 2 there is a decrease in the age range, which can be between 20 and 30 years. The relationship of Age and Semester Average is quite similar to the previous one, although the relationship is positive, the slope tends to be lower, this coincides with the previous results where it is shown that the averages in general are lower in the semester in relation to the career averages. In the two named relationships, there is a significant group of students, who have averages greater than three, even greater than four.

This result is interesting because it shows that it is important to focus attention on younger students, and according to previous results, who are in their first semesters, to look for the factors that influence their academic performance, these can be of different nature, as indicated by the findings found by the Ministry of Education that shows that the main determining factor of dropping out of studies in Colombia is located in the academic dimension, that is, it is associated with the potential or cultural and academic capital with which students enter higher education, other factors found are financial and socioeconomic, followed by institutional and vocational and professional orientation.

In the relationship between Stratum and Career Average, as well as in the relationship Stratum – Semester Average, in strata 1, 2, 3 the range of grades seems to be distributed uniformly in each stratum. As the stratum grows, groups of students concentrate in higher grade ranges, it can be seen that the highest concentration of students is in the grade range of 2 to 4.

The stratum and average relationship shows that many of the students with fewer economic resources (strata 1, 2, 3) have averages below 3 and another smaller group above 3, there is also a group that is visualized in stratum 4 (medium) where clearly the majority of averages are above 2, indicating that they do not have very low averages and their academic performance is better compared to the lower strata. For now, it can be corroborated that dropout students mainly belong to lower strata and that performance tends to improve in higher strata, especially in the middle stratum. There is not much information about the higher strata.

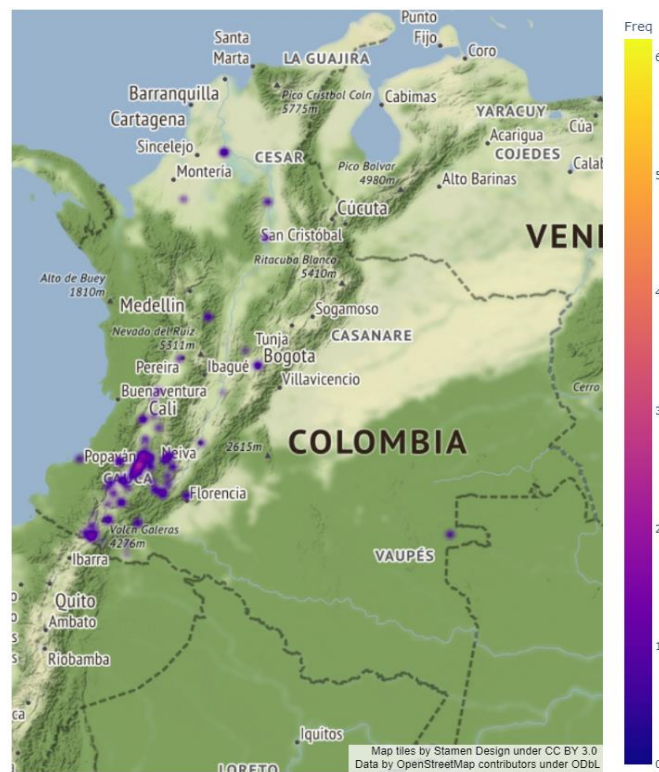
Career Average and Semester Average: there is a large concentration of averages close to zero, which correspond to the students who drop out in the first semester, which according to previous results represent approximately a quarter of the dropout students. A linear relationship is also observed in two groups, one with very low averages both semester and career (below 3), who drop out motivated by their low academic performance and another with averages above 3 that, according to previous results, it can be speculated that for these students the reason for dropping out is not related to low academic performance.

Dropout and Geographical Conditions

The origin of the students from the University of Cauca enrolled in the first semester of 2018 and that of the dropout group was analyzed using heat maps, which indicate the geographical places from where they come; it was observed that the students mainly come from the southwestern Colombian region, with a concentration in the departments of Nariño, Huila, Cauca, Valle del Cauca, and Putumayo, which correspond to the departments with which it borders. This result shows the importance of the University of Cauca in the educational development of the region. In smaller quantities, there are also students from the Andean and Caribbean regions.

The dropout students (Figure 7) also concentrate in the southwestern Colombian region, especially in Cauca, near its capital Popayán. This indicates an interesting and perhaps unexpected result, as it is generally thought that students who come from other departments could have greater difficulties than the locals, such as, for example, higher maintenance and stay expenses, separation from the family, etc., factors that could influence their performance at the university.

Figure 7 – Heat Map of Dropout Students in the First Semester of 2018



Source: CC BY 3.0 Data by OpenStreetMap.

One of the reasons for this phenomenon may be that many of the students from Popayán and its surroundings belong to schools that do not have a good educational quality, another important reason is that they belong to areas with problems of violence, such as drug trafficking, illicit crops, territorial wars, poverty etc., conditions that since the time of the colony have affected the region and that may be significantly influencing the educational process of our students. Although the University is one of the most important in the country, it is located in one of the departments with the most problems of armed conflict and territory.

Another possible reason for the result found could be that students who do not live in Popayán, but in its vicinity have difficulty commuting daily to the university facilities, in addition to the fact that food and transportation expenses increase. These possible reasons for dropout in Popayán and its surroundings can be analyzed in more detail by seeking information about the quality of secondary education received by students, as well as information related to social conditions and violence in the place of origin.

Conclusions

The most important source of information for the work has been the Division of Admissions, Registration, and Academic Control of the University of Cauca (DARCA), where academic and social data are found. The periods analyzed are those between the first semester of 2018 and the first semester of 2019. After creating the dataset, it is possible to identify the total group of students who stop enrolling for two consecutive semesters and who are considered dropouts.

Through the application of different techniques and functions of Data Science, the process of exploration and analysis of the data provided by DARCA was carried out, obtaining the following: within the areas of knowledge, it was found that both the programs and the faculties have a higher dropout rate in Social Sciences, Exact Sciences, and Engineering. The highest percentage of dropout by programs (Social Sciences and Exact Sciences) is between 14% and 18%, and for engineering around 7%, although the number of dropout students in these programs was large, their percentage was not so much. For the faculties, it was found that the range of dropout percentages is between 3% and 8%, and those that had the highest percentage, agreed with those that have the highest number of dropout students, except for the faculty of arts. These results are of great interest as they allow guiding this study in the direction of the programs and faculties to look for possible causes that can contribute to dropout, such as the degree of difficulty of the program, its duration, relationships with the academic and administrative community, problems in the transition and coupling process, in vocational guidance, etc., these are situations that can be analyzed in greater depth, when looking for related information.

According to the study made on individual and academic factors such as age, gender, and semester, it was found that within the group of dropouts, students are aged between 21 and 35 years and within this age range, the percentage of first-year students is 25.6%, that is, a little more than a quarter of the group in question. When continuing to search for the group of dropouts by semester, it was found that as the semester and age increase, the percentages decrease, indicating a lower probability of dropping out in higher semesters. These results clearly show a higher probability of dropout in younger students who are in the first semesters, this may be related to the transition period between secondary education and higher education where the student's academic and social adaptability is determined; also in the age range found, students begin their work and family experience.

In relation to gender, the percentages of enrolled men and women differ only by 3.8%, with a higher number of men enrolled, within the group of dropouts the difference is 23.6% with a higher number of men who drop out of their studies. This result must be corroborated to know if there is a significant difference in dropout between genders, which can help find the reasons why men make the decision to drop out of the program.

In Colombia, it has been found that the factors that mainly determine the abandonment of studies in higher education are academic. This phenomenon was analyzed through the semester and career averages of dropout and non-dropout students. It was found that when comparing the averages of the two groups, the academic performance is clearly lower in the group of dropout students, where the lower ranges for them are below 3 and for the group of non-dropouts, above 2.9. The results found when comparing career and semester averages indicate that in general the latter are lower, a factor that could influence the decision to drop out, as the semester in which they stop enrolling, the averages are lower. Within the group of dropout students, it was found that in the upper 25% of the sample the averages are above three, therefore, this group of students have good academic performance, this indicates that the decision to abandon their studies is not related to academic factors, for that particular group.

Through the application of data visualization functions, statistical graphs are obtained that allow seeing the relationship between different factors mainly age – stratum, average – stratum. The results show that the students who drop out of their studies at the University of Cauca mainly belong to the low – low (1), low (2) and medium – low (3) strata, in these strata the ages of the students are distributed throughout the age range, however, when looking at the other strata (4, 5, 6) the number of students is much lower, as well as the ages. It is clear that economic conditions are very much related to student dropout and the possible situations can be diverse such as lack of resources, change of work situation, family economic difficulties, etc.

When relating the stratum and the average, it was found that the majority of dropout students who have low strata (1, 2, 3) have aver-

ages less than three, when moving to the middle stratum (4) it is clearly visualized that the averages are above two, that is, in general, there is better performance. Once again it is corroborated that the majority of dropout students belong to the lower strata and it is also found that as the stratum increases, academic performance improves. In this way, we see that the social and academic factors of stratum and average are very related and are determinants in the phenomenon of dropout at the University of Cauca.

Heat maps allow visualizing the origin of all the students of the University of Cauca enrolled in the first semester of 2018, which is mainly from the South West of Colombia, color concentrations are found in the departments of Nariño, Cauca, Huila, and Valle del Cauca. Comparing this result with that obtained for dropout students, it was found that the origin is mainly from Popayán and its surroundings, an interesting result that leads to propose possible reasons for dropout, such as the academic quality of the schools to which the students belong, the socioeconomic conditions of the students living in Popayán and the nearby areas, the conditions of violence and family situation, etc¹.

Received on March 4, 2023

Approved on August 30, 2023

Note

¹ *Acknowledgments:* Special thanks to the Division of Admissions, Registration and Academic Control (DARCA) and Division of Planning belonging to the Vicerector's Office for Culture and Welfare that have made it possible to carry out this study.

References

AHUJA, Ravinder; KANKANE, Yash. Predicting the probability of student's degree completion by using different data mining techniques. In: INTERNATIONAL CONFERENCE ON IMAGE INFORMATION PROCESSING, 4., 2017, Shimla, India. *Annals [...]*. Shimla, India: IEEE, 2017. p. 1-4.

BARRAGÁN DIAZ, Diego; PATIÑO GARZÓN, Luceli. Elementos para la comprensión del fenómeno de la deserción universitaria en Colombia. Más allá de las mediciones. *Cuadernos Latinoamericanos de Administración*, Colombia, v. 9, n. 16, p. 55-66, 2016.

BAYONA-RODRÍGUEZ, Hernando; RUEDA-DELGADO, Gabriel; OME, Luis Germán. ¿Cómo cobran la matrícula las universidades públicas en Colombia? Evidencia 2009-2019. *Nómadas*, Bogotá, v. 56, p. 67-91, 2022. Disponible en: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0121-75502022000100067&nrm=iso. Acceso el: 2 ene. 2022.

BELTRÁN CH, Jesica Wendy; DE DAVID, Cesar. Dimensión espacial del desarrollo en el departamento del Cauca, Colombia. La vía Panamericana como eje de aglomeración. *Pampa*, Santa Fe, v. 11, n. 11, p. 39-62, 2015. Disponible en: <https://bibliotecavirtual.unl.edu.ar/publicaciones/index.php/Pampa/article/view/5190>. Acceso em: 5 ago. 2023.

COLOMBIA. Ministerio de Educación Nacional. *Educación Superior: Boletín Informativo* nº 14. Bogotá: Panamericana Formas e Impresos S.A., 2010. Dis-

ponible en: <https://www.mineducacion.gov.co/1621/article-92779.html>. Acceso el: 3 mar. 2022.

COLOMBIA. Ministerio de Educación Nacional. **Lineamientos Política de Educación Superior Inclusiva**. Bogotá: MinEducación; Dirección de Fomento para la Educación Superior, 2013. Disponible en: https://www.mineducacion.gov.co/1759/articulos-357277_recurso.pdf. Acceso el: 3 feb. 2022.

GARTNER ISAZA, Lorena; DUSSÁN LUBERT, Carmen; MONTOYA, Diana Marcela. Caracterización de la deserción estudiantil en la universidad de Caldas el período 2009-2013. Análisis a partir del sistema para la prevención de la deserción de la educación superior – Spadies. **Latinoamérica de Estudios Educativos**, Manizales, v. 12, n. 1, p. 132-158, 2016. Disponible en: <https://www.redalyc.org/pdf/1341/134149742008.pdf>. Acceso el: 2 ene. 2022.

GÓMEZ CAMPO, Víctor Manuel; CELIS GIRALDO, Jorge Enrique. Crédito educativo, acciones afirmativas y equidad social en la educación superior en Colombia. **Revista de Estudios Sociales**, Bogotá, n. 33, p. 106-117, 2009. Disponible en: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-885X2009000200010&nrm=iso. Acceso el: 2 mar. 2022.

GUZMÁN, Carolina; DURÁN, Diana; FRANCO, Jorge; CASTAÑO, Elkin; GALLÓN, Santiago; GÓMEZ, Karoll; VÁSQUEZ, Johanna. **Deserción estudiantil en la educación Superior Colombiana**. Bogotá: Imprenta Nacional de Colombia, 2009.

GVIRTZ, Silvina; ORÍA, Ángela. Evaluar el Rendimiento Interno y Académico: un Desafío para la Macro y la Micro Política. Lecciones a Partir de un Estudio de Caso. **Revista Iberoamericana de Evaluación Educativa**, Madrid, v. 3, n. 2, 2016. Disponible en: <https://revistas.uam.es/riee/article/view/4494>. Acceso el: 2 mar. 2022.

HERNANDEZ GONZALEZ, Arnulfo Gamaliel; MELENDEZ ARMENTA, Roberto Angel; MORALES ROSALES, Luis Alberto; GARCIA BARRIENTOS, Abel; TECPA-NECATL XIHUITL, Jose Luis; ALGREGO, Ignacio. Comparative Study of Algorithms to Predict the Desertion in the Students at the ITSM-Mexico. **IEEE Latin America Transactions**, v. 14, n. 11, p. 4573-4578, 2016. Disponible en: <https://ieeexplore.ieee.org/document/7795831>. Acceso el: 2 ene. 2022.

HIGUERA-MARTÍNEZ, Oscar Ivan. Deserción estudiantil en Colombia y los programas de ingeniería de la Uptc seccional Sogamoso. **Ingeniería Investigación y Desarrollo**, Boyacá, v. 17, n. 1, p. 70-76, 2017. Disponible en: <https://doi.org/10.19053/1900771X.v17.n1.2017.5593>. Acceso el: 5 mar. 2022.

HIMMEL, Erika. Modelo de análisis de la deserción estudiantil en la educación superior. **Calidad en la Educación**, Santiago de Chile, n. 17, p. 91-108, 2002. Disponible en: <http://dx.doi.org/10.31619/caledu.n17.409>. Acceso: 5 mar. 2022.

HURTADO-TOBÓN, Luís Hernando; GARCÍA-GONZÁLEZ, María Dolly; GALVIS-SOTO, Diana Milena; MÉNDEZ-PARRA, Rosa María. Modelo para el seguimiento y control de la deserción en la población universitaria. **Revista De Investigaciones Universidad Del Quindío**, Quindío, v. 34, n. S4, p. 12-17, 2022. Disponible en: <https://doi.org/10.33975/riuq.vol34nS4.1061>. Acceso el: 2 ene. 2022.

IESALC. Instituto Internacional de la UNESCO para la Educación Superior en América latina y el Caribe. **Informe sobre la educación superior en América Latina y el Caribe 2000-2005: la metamorfosis de la educación superior**. Caracas: IESALC, 2006.

LÓPEZ GÓMEZ, Juan Gerardo. Comparación en el porcentaje de deserción entre alumnos de preparatorias públicas y privadas que ingresaron en el ciclo escolar 2015-2016 a la UASLP. **Revista Acta Educativa**, v. 6, n. 1, p. 1-14, 2018.

MARTÍNEZ, Iñigo; VILES, Elisabeth; OLAIZOLA, Igor. Metodologías de ciencia de datos: desafíos actuales y enfoques futuros. **Big Data Research**, v. 24, 2021. Disponible en: <https://doi.org/10.1016/j.bdr.2020.100183>. Acceso el: 2 feb. 2022.

MORALES RODRÍGUEZ, Danna Lorena; CARO SILVA, Joan Alejandro; CONTRERAS BRAVO, Leonardo Emiro. A review on the prediction of students' academic performance using ensemble methods. **Ingeniería Solidaria**, Bogotá, v. 18, n. 2, p. 1-28, 2022. Disponible en: <https://revistas.ucc.edu.co/index.php/in/article/view/4312>. Acceso el: 30 ene. 2022.

OECD. Organización Para La Cooperación y el Desarrollo Económicos. **Reporte anual 2007**. Paris: OECD Publishing, 2007. Disponible en: <https://doi.org/10.1787/annrep-2007-en>.

PASCARELLA, Ernest; TERENCEZIN, Patrick. **How college affects students: a third decade of research**. San Francisco: Jossey-Bass, 2005.

PATIÑO GARZÓN, Luceli; CARDONA PÉREZ, Angélica María. Revisión de algunos estudios sobre la deserción estudiantil universitaria en Colombia y Latinoamérica. **Theoria**, Chile, v. 21, n. 1, p. 9-20, 2012. Disponible en: <https://www.redalyc.org/articulo.oa?id=29931769002>. Acceso el: 3 mar. 2022.

PINEDA, Clelia; PEDRAZA, Alexandra. **Persistencia y Graduación. Hacia un modelo de retención estudiantil para Instituciones de Educación Superior**. Bogotá: Universidad de la Sabana, 2011.

ROJAS BETANCUR, Mauricio; GONZÁLEZ, Diana Carolina. Deserción estudiantil en la Universidad de Ibagué, Colombia: una lectura histórica en perspectiva cuantitativa. **Zona Próxima**, Colombia, n. 9, p. 70-83, 2008. Disponible en: <https://www.redalyc.org/pdf/853/85312286006.pdf>. Acceso el: 3 feb. 2022.

TINTO, Vincent. Definir la deserción: una cuestión de perspectiva. **Revista de Educación Superior**, v. 18, n. 71, 1989.

TOSE VERGARA, Paula Andrea; ORTIZ RUIZ, Nicolás. Análisis de política pública centrado en actores: violencia por conflicto armado y construcción de paz en el Cauca (2012-2014). **Revista Mexicana de Ciencias Políticas y Sociales**, Ciudad de México, v. 64, n. 237, p. 341-375, 2019. Disponible en: <https://www.redalyc.org/articulo.oa?id=42164494014>. Acceso el: 2 ene. 2022.

UNIVERSIDAD DEL CAUCA. **Inicio**: Programa para atender la deserción, permanencia y graduación estudiantil en la Universidad del Cauca. Popayán: Universidad del Cauca, 2016. Disponible en: <http://facultades.unicauca.edu.co/vicecultura/programa-para-atender-la-desercion-permanencia-y-graduacion-estudiantil-en-la-universidad-del-cauca>. Acceso el: 2 ene. 2023.

UNIVERSIDAD DEL CAUCA. **Modelo de Permanencia y Graduación Estudiantil**. Popayán: Universidad del Cauca, 2017. Disponible en: https://vicecultura.unicauca.edu.co/viceculturav2/sites/documentos/programas/politica_de_permanencia_y_graduacion_estudiantil-21-12-17.pdf. Acceso el: 2 ene. 2023.

Maria Falla received B.S in physics engineering from University of Cauca, Popayán, in 2008, M.Sc. degree in physics science from the University of Puerto Rico, Mayaguez campus, Puerto Rico, in 2013. From 2009 to 2010, she was professor in the university Fundación Universitaria de Popayán, in Popayán. From 2014 to the present, she is professor with the Physics Department, University of Cauca, Popayán.

ORCID: <https://orcid.org/0000-0002-4603-6987>

E-mail: mfalla@unicauca.edu.co

Rubiel Vargas-Cañas received the B.S. in computer sciences engineering from the Universidad Industrial de Santander, Bucaramanga, in 2000, M.Sc. degree in engineering from the Universidad del Valle, in 2010 and the Ph.D. degree in biomedical engineering from City University, London, UK, in 2012. From 2001 to 2002, he was the coordinator of the academic Databases for the library of Universidad Industrial de Santander. Since 2002, he has been a Professor in the Physics Department of the Universidad del Cauca.

ORCID: <https://orcid.org/0000-0003-1548-942X>

E-mail: rubiel@unicauca.edu.co

Availability of research data: the dataset supporting the results of this study is published in this article.

Editor in charge: Lodenir Karnopp

