Journal of the Brazilian Computer Society, 2026, 32:1, doi: 10.5753/jbcs.2026.6559

An Extended Process Mining Framework for the Multi-factor Analysis of Student Trajectories in Higher Education: The Dropout Problem

Luiz F. P. Southier [**Federal University of Technology - Paraná** | ***luizsouthier@utfpr.edu.br***]
Marcelo Teixeira [**Federal University of Technology - Paraná** | ***mtex@utfpr.edu.br***]
Lovania R. Teixeira [**Federal University of Technology - Paraná** | ***lovaniateixeira@utfpr.edu.br***]
Sheila C. Freitas [**Federal Institute of Paraná** | ***sheila.freitas@ifpr.edu.br***]
Denise M. V. Sato [**Federal Institute of Paraná** | ***denise.sato@ifpr.edu.br***]
Jair J. Ferronatto [**Federal Institute of Santa Catarina** | ***jair.ferronato@ifsc.edu.br***]
Edson E. Scalabrin [**Pontifical Catholic University of Paraná** | ***edson.scalabrin@pucpr.br***]

Department of Informatics, Federal University of Technology - Paraná, Via do Conhecimento, km 01, Pato Branco, PR, 85503-390, Brazil.

Received: 12 September 2025 • **Accepted:** 16 February 2026 • **Published:** 29 April 2026

Abstract While higher education is the backbone for human capital development and economic growth, its high dropout rates remain a global concern that leads to wasted resources and unfulfilled student potential. Understanding dropout requires integrating social, economic, academic, and technical factors across students' trajectories, often interrelated in intricate, non-obvious ways. In this context, Process Mining (PM) offers a promising approach by uncovering patterns in students' interactions with academic programs and courses. However, traditional PM methods are typically established over mono perspectives of processes, which limits their ability to capture the multi-factor and correlated nature of educational trajectories. To address this gap, this paper proposes an extended PM-based approach that incorporates enriched labeling strategies that allow the simultaneous analysis of multiple dimensions of students' academic trajectories. Furthermore, the article presents a detailed application of the labeled method over real data of a Brazilian public university with 437,690 events from eight different programs, including students from the Unified Selection System (SISU). By comparing students' outcomes and paths, while considering their enrollment method, course option, and demographic information, we discovered that admission score, program, high school type, gender, and place of origin are the variables with a higher correlation to successful and less successful students. A deeper analysis of a specific program is also outlined to show how the approach can be customized for particular cases, under minor effort, while keeping standard input data.

Keywords: Data science in education, Dropout problem, Higher Education, Process Mining

1 Introduction

Human capital is crucial for world economic development [Morais *et al.*, 2024], and it has higher education as its backbone [Marques, 2020]. As the dropout rate in higher education directly impacts the qualification indexes, it has been a critical subject of discussion in the last decades [Araque *et al.*, 2009; Cerdeira *et al.*, 2018; Levy, 2007; Saqr *et al.*, 2023; Awang Long *et al.*, 2023] in several countries around the world, such as Greece [Xenos *et al.*, 2002], Spain [Ortiz-Lozano *et al.*, 2018], China [Zhang *et al.*, 2021], USA [Chen *et al.*, 2018], and Indonesia [Bäulke *et al.*, 2022]. According to the most recent report, in OECD countries, 23% of students did not complete their higher education, with the dropout rate being predominant in Latin American countries such as Brazil and Colombia [OECD, 2022]. In Brazil, for example, according to the National Institute of Educational Studies and Research Inep [2020], from 2010 to 2019, approximately 59% of higher education students dropped out, while 40% successfully completed their higher education. In 2019, 1% of students remained enrolled in the course they started a decade prior. The most recent report reveals a similar scenario: from 2,899,439 students admitted to Brazilian higher education in 2012, 1,714,956 have dropped out, and 1,153,285 have graduated [Inep, 2022].

Thus, identifying barriers that lead to dropout in higher education is crucial for helping institutions to understand factors leading students not to complete their programs [Chen *et al.*, 2018]. This also allows undergraduate course managers, entities responsible for public higher education policies, and funding agencies to target interventions. For example, knowing the causes of dropout can help institutions reallocate resources more efficiently [Bäulke *et al.*, 2022]; by identifying areas or courses where students are more likely to disengage, institutions can provide targeted support services, mentoring programs, or academic assistance to help struggling students overcome obstacles and persist in their studies [Awang Long *et al.*, 2023]; and so on.

Several approaches have been proposed to analyze different aspects of higher education processes and understand what factors may lead to dropout. Some of them [Xenos *et al.*, 2002; Awang Long *et al.*, 2023; Tayebi *et al.*, 2021] rely on data collection forms, surveys, or interviews with students. As this cannot be fully automated, the manual effort of collecting data has to be repeated whenever a new group of students is to be analyzed, undermining possible analyses. Attempts to predict dropouts [Ortiz-Lozano *et al*., 2018; Chen *et al*., 2018; Behr *et al*., 2020] or correlate dropout factors [Xenos *et al*., 2002] have also been tested, but these approaches do not consider the impact of different pathways that students can take. In most universities, as students can follow different paths to get a degree, analyzing the different paths and their impact on the students' outcomes may be crucial for accurate analysis and decision-making. Some studies [Saqr *et al*., 2023; Levy, 2007; Zhang *et al*., 2021] add the idea of analyzing students' paths, but they are applied to online learning environment processes, which differ from the purposes of this paper.

This paper proposes a PM-based approach tailored to capture aspects of higher education that are crucial for multi-factor analysis. Our approach introduces a labeling method that enables to set, carry, and simultaneously correlate multiple dimensions of students' academic paths, such as admission profiles, course preferences, and sociodemographic attributes. This favors a more holistic view of the trajectories leading to completion, delay, or dropout. Additionally, we detail the log construction process, highlighting the methodological steps required to transform raw academic records into analyzable PM event logs, an often overlooked but critical challenge when dealing with confusing university information systems. Beyond discovering process models, our approach integrates temporal, correlational, and visualization features to enhance interpretability and support decision-making by institutional managers and policymakers.

By using our approach, institutions can obtain insights into different academic paths; explore how programs, courses, and students interact; compare the trajectories of successful and less successful students; highlight critical discrepancies; identify statistically meaningful patterns linked to different outcomes; and provide actionable indicators for students, educators, and program coordinators to support informed interventions. In summary, the approach is designed to serve as a practical, interpretable tool for institutional management and policy-making.

To illustrate the proposed framework in practice, we present a detailed case study based on real data from a Brazilian public university. The case study includes both a general mapping of academic trajectories and a deep dive into a specific undergraduate program. In this focused analysis, we examine individual paths in detail and highlight how key insights can inform managerial decisions. For each analytical output, we provide concrete suggestions on how results could be translated into strategies for course offering, student advising, or program redesign. The case also reflects challenges typical of Brazilian public education, such as elective course flexibility, course prerequisites, limited availability for retaking failed courses, and high heterogeneity in student backgrounds due to nationwide admission through the Unified Selection System (SISU). This complexity reinforces the need for multi-factor, interpretable analytics, precisely what our extended PM framework aims to support.

Structurally, the background is presented in Section 2; Section 3 introduces the labeling strategy, with a methodology os use in Section 4; a case study is detailed in Section 5, while conclusions and open challenges are discussed in Section 7.

2 Background

In the context of PM, it is generally assumed that processes work connected to information systems (IS), and it is of interest to analyze their behavior. Initially, the IS is visited to extract and process its raw data, aiming to transform it into an *event log*. This can be done by processing web scraping or database queries [Zhao, 2017].

An event log contains structured data entries that represent the activities performed by the process. Each recorded event (e) corresponds to an activity (a) (e.g., a course taken) and a specific process instance (c) (e.g., a student) referred to as a *case* [Nakatumba and van der Aalst, 2009]. Events are associated with a specific moment in time (t) and can possess various *event attributes* (e.g., the obtained grade or the professor's name). This notion can be formalized as follows [Van Der Aalst, 2016].

Let A denote the set of activities in a process, C represent the set of cases, and T symbolize the time domain. E represents the set of all events, while Σ represents the set of names for data attributes that an event can assume. An event $e \in E$ is defined as a tuple (a, c, t), where $a \in A$, $c \in C$, and $t \in T$. For any event $e \in E$ and a specific attribute name $\sigma \in \Sigma$, $\#_\sigma(e)$ represents the value associated with attribute σ for event e.

Each case c (process instance) includes a *trace* (τ), which is an ordered sequence of events (e.g., the order in which courses are taken by students), and each case may have several *case attributes* (e.g., name or city of birth).

Let Γ denote the set of all names of data attributes a case can have. For any case $c \in C$ and attribute name $\gamma \in \Gamma$, $\#_\gamma(c)$ represents the value of attribute γ for case c. An *event log* is a set of cases $L \subseteq C$, such that each event appears at most once in the entire log. $\#_\sigma(L)$ is the set of values of attribute σ for all events in the log L. $\#_\gamma(L)$ represents the set of values of attribute γ for all cases in the log L.

Each case $c \in C$ contains a *trace*, represented by a finite sequence of events $\tau(c) \in E^*$, where each event appears only once. The order of events in a trace respects their timestamps. A trace $\tau(c)$ has a string representation $\hat{\tau}(c) = a_1 a_2 ... a_n$ for $a_1, a_2, a_n \in A$ that shows the activities for the events in τ and the order in which they occur.

$|L_{\hat{\tau}}|$ denotes the number of occurrences of the trace $\hat{\tau}$ on the event log L, i.e, the number of cases in which the string representations of their traces is $\hat{\tau}$. $|L_a|$, for $a \in A$, denotes the number of events $e \in E$ in the event log L that have a as activity. $|L_{a_1 \rightarrow a_2}|$, for $a_1, a_2 \in A$, denotes the number of occurrences of activity a_2 directly following activity a_1 in traces on the event log L.

To illustrate these properties, Table 1 shows an example of a short event log, where each row represents a course (activity) taken by a student (case). Course information is recorded, such as the start date (timestamp), the corresponding professor, and the final grade. Other student information, such as age and country, is also recorded.

From the structured log, it is possible to apply a PM discovery algorithm to reveal a model that represents the process

Table 1. Example of raw data converted into event log format.

Case	Student	Age	Country	Course	Start date	Professor	Grade
1	Peter	20	Brazil	Algebra	1-12-2022	James	6
				Calculus	1-02-2023	Emma	8
				Physics	1-07-2023	Thomas	10
2	Mary	21	USA	Algebra	1-12-2022	James	6
				Progr.	1-02-2023	Emma	10
				Physics	1-06-2023	Thomas	10
3	John	21	Brazil	Algebra	1-12-2022	James	6
				Calculus	1-02-2023	Emma	10
				Physics	1-07-2023	Thomas	10
4	Ruth	19	Sweden	Algebra	1-02-2022	James	1
				Algebra	1-12-2022	James	6
				Physics	1-06-2023	Thomas	10

formally and gathers its collective behavior. There are several formalisms to output the model, such as Petri nets, automata, and Directly-Follows Graphs (DFG) Chapela-Campa *et al.* [2022]. Formally, a DFG is a directed graph $G = (A_G, E_G)$ wherein each node denotes an activity $a \in A_G \subseteq A$ on the process, and each edge $a_s \rightarrow a_t \in E_G$, for $a_s, a_t \in A_G$ denotes the fact that the target activity a_t occurs immediately after the source activity a_s in traces $\tau \in E^*$ of the process. Figure 1 shows the DFG discovered for the example.

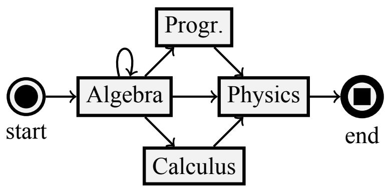

Figure 1. Example of a Directly-Follows Graph as a process model.

Remark that this model exposes all possible paths students may traverse, accounting for the sequence in which they undertake (or redo) courses.

3 Proposed context-sensitive modeling

The process maps derived from the corresponding event log (Table 1), depicted in Figure 1, visually represent the process in event log L. A closer examination of these process maps reveals valuable insights into the structural aspects of student pathways. It becomes apparent, for instance, that students attending Programming classes do not participate in Calculus. The map also highlights instances of students retaking Algebra. Therefore, it is intuitive to consider that the nodes within the process map could be labeled, color-coded, for example, based on a singular variable, such as frequency, grade, attendance, and so forth. Figure 2 shows the same example model from Figure 1. However, now the nodes are labeled according to the frequency and grades of courses.

This is the key idea for the result that follows. We calculate a number representation $\hat{a} \in \hat{A_G}$ for each $a \in A_G$ for the process map $G = (A_G, E_G)$ and associate a label, a color, for visual inspection, $c \in C$ to each $a \in A_G$ based on the number representation $\hat{a}$.

Definition 1 (Color, Color map)**.** *A color* $c \in C$ *is a triplet* (r, g, b) *where* $0 \leq r \leq 255$, $0 \leq g \leq 255$, $0 \leq b \leq 255$ *represents a color in RGB format. A color map* Cm *is a function* $Cm : \hat{A_G} \rightarrow C$ *where* $\hat{A_G}$ *is the set of number representations for each node in* A_G*, and* C *is a set of colors.*

To calculate the number representation $\hat{a} \in \hat{A_G}$ for each node $a \in A_G$ for the process map $G = (A_G, E_G)$ one of the following options (op) can be conducted based on the intended goal:

- ($op = 1$) The number representation $\hat{a}$ for a node is the number of events $e \in E$ in the event log L that have a as activity, i.e., $\hat{a} = |L_a|$. This option is recommended to understand the frequency of each activity in the process.
- ($op = 2$) The number representation $\hat{a}$ for a node is the result of some aggregation function (min, max, average, etc.) over a case attribute $\#_\gamma(c)$ for all cases c such that $a \in \hat{\tau}(c)$. This option is recommended to understand how a case attribute is distributed along the process.
- ($op = 3$) The number representation $\hat{a}$ for a node result of some aggregation function (min, max, average, etc.) over an event attribute $\#_\sigma(c)$ for all events e such that $e = (a, c, t)$, for any timestamp t. This option is recommended to understand how an event attribute is distributed along the process.

Formally, the number representations $\hat{a}$, for each node, can be calculated as in the algorithm that follows.

Algorithm 1 Number representation

Input: option op, event log L, DFG $G = (A_G, E_G)$, aggregation function agg (optional for option 1), a case attribute $\#_\gamma(c)$ (required for option 2), and an event attribute $\#_\sigma(c)$ (required for option 3).
Output: a map $f : A_G \rightarrow \hat{A_G}$ associating each node $a \in A_G$ to a number representation $\hat{a} \in \hat{A_G}$.

$f \leftarrow \emptyset$
for $a \in A_G$ **do**
  if op is 1 **then**
    $f \leftarrow f \bigcup \{a \rightarrow |L_a|\}$
  else if op is 2 **then**
    $f \leftarrow f \bigcup \{a \rightarrow agg(\#_\gamma(c))\}$, $\forall c$ such that $a \in \hat{\tau}(c)$
  else if op is 3 **then**
    $f \leftarrow f \bigcup \{a \rightarrow agg(\#_\sigma(c))\}$, $\forall e$ such that $e = (a, c, t)$, for any timestamp t
  end if
end for

To calculate the color map that is used for labeling the process model, the user has to define the starting (sc_i) and finishing (fc_i) colors for number intervals (n_i), such that all defined intervals are excluding and cover all values in $\hat{A_G}$. Formally, the user defines i number intervals $n_i \in N$ such that $\hat{A_G} \subseteq \bigcup n_i$ and $\bigcap n_i = \emptyset$.

Algorithm 2 defines how the color maps $Cm : \hat{A_G} \rightarrow C$ are calculated based on the interval definitions. The k parameter defines the number of different colors each interval can have.

The *hsl* and *rgb* functions convert colors from the RGB to HSL format and HSL to RGB format, respectively. First, for each interval i a list of colors C_i is created by adding a *step* based on the number of colors k, starting (sc_i) and finishing colors (fc_i). $\|$ denotes the concatenation of two lists. C_X is a list created by concatenating all lists of colors

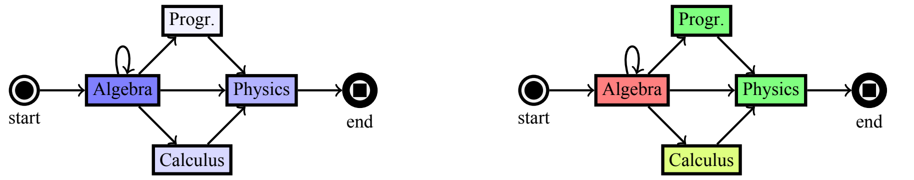

Figure 2. Example of process models colored respectively by frequency and grade. The bluer, the higher the frequency; the greener, the higher the grades.

Algorithm 2 Color map calculation

Input: activity number representation $\hat{A_G}$, number intervals $n_i \in N$ and its corresponding starting (sc_i) and finishing colors (fc_i), number of colors by interval k.
Output: a map $Cm : \hat{A_G} \to C$ associating each number representation $\hat{a} \in \hat{A_G}$ to a color C.

$Cm \leftarrow \emptyset$
for $n_i \in N$ **do**
  $step \leftarrow (hsl(fc_i) - hsl(sc_i))/k$
  $C_i \leftarrow []$ ▷ C_i is a list of colors for interval i
  for c from 0 to $k+1$ **do**
    $C_i[c] \leftarrow rgb(sc_i + k \times step)$
  end for
end for
$C_X \leftarrow C_1 \| C_2 \| \cdots \| C_i$ ▷ All list of colors are concatenated
$Cm \leftarrow \emptyset$
for $\hat{a} \in \hat{A_G}$ **do**
  $s \leftarrow min(N)$
  $c \leftarrow C_X[\lfloor \frac{\hat{a} - min(N)}{max(N) - min(N)} \times |N| \times k \rfloor]$
  $Cm \leftarrow Cm \bigcup \{\hat{a} \to c\}$
end for

$C_1 \| C_2 \| \cdots \| C_i$. To match a number representation $\hat{a}$ to a color c, we use $\hat{a}$ to calculate a certain index that matches a color in C_X.

The process maps depicted in Figure 2 are colored based on: the frequency of each activity (left); and average grade (right). Evaluation of the color-coded nodes within the process map (on the left) reveals that Algebra commands the highest occurrence rate, while Programming records the lowest frequency. Furthermore, Algebra exhibits the lowest average grades (on the right). The color-map approach presented here may also be the starting point for many other future applications of multifactor analysis embedded on PM.

4 Evaluation Methodology

We now present a methodology for performing education process analysis based on the context-sensitive approach outlined in Section 3. The steps of the approach are shown in Figure 3.

4.1 Extraction and preprocessing

First, we perform the *extraction and preprocessing* of relevant data from education systems, collecting the students' transcripts and building an *event log*. Next, four analyses are conducted to obtain relevant *charts, tables, and graphs* indicating aspects of the event log: time analysis, correlation analysis, process map analysis, and plot analysis. Finally, the *evaluation* of the analysis results can be performed by educators and other process specialists, which would suggest *improvements* for the process itself.

When *extracting* the data, the researcher must select the dataset from which the analyses are conducted. The dataset can correspond to a specific program, a set of programs offered in a specific location, or even several locations. This choice depends on what questions the research wants to answer. Also, the variables included in the dataset depend on these questions. The dataset can be obtained from education information systems. For this, the user can perform web scraping [Zhao, 2017] or database extraction.

Next, the extracted dataset has to be *preprocessed* into an event log format. An *event log* records events wherein each event information refers to an *activity* (a well-defined step in some process), executed at a certain point in time (*timestamp*), and it is related to a particular *case* (a process instance or, in other words, an occurrence of the process). Further, event logs would normally record additional information about events and cases. In our approach, at least three pieces of information are mandatory to create the event log from educational data: an enrollment or student ID (the case identifier), the courses taken by the student (the events), and when the courses were taken (the timestamps). Other information can be added at the case level (enrollment or student variables such as place of birth, gender, and ethnic group) or the event level (event variables such as grade, attendance, or the professor's name).

4.2 Time Analysis

Using the event log, our proposed approach enables four analyses. *Time analysis* enables plotting variable progress over time (timeline plots). For numerical variables, the plot shows the mean and the 95% confidence interval around the mean; for categorical variables, a kernel density estimate (KDE) histogram [Härdle *et al.*, 2004] of each variable category. An example of a KDE histogram plot is shown in Figure 4 . This example shows histograms of students by year for males and females (colored columns). The curves are the KDEs of these histograms and represent the same behavior. Since some variables have multiple categories, it is easier to see their behavior using KDE.

The approach also enables drift analysis, a technique that

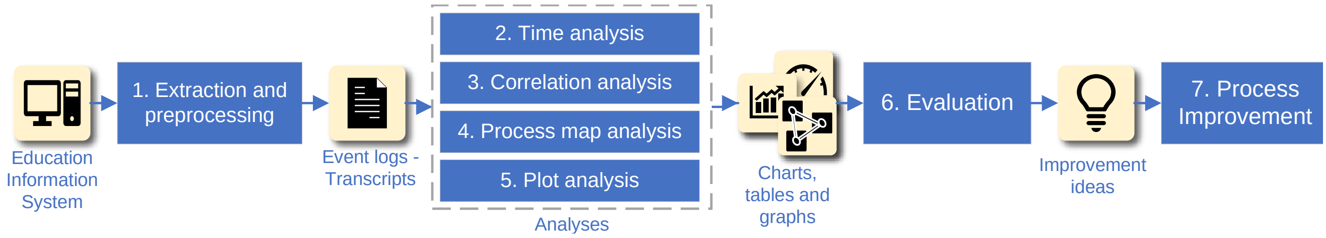

Figure 3. An overview of the PM-based approach for analyzing the paths of higher education students.

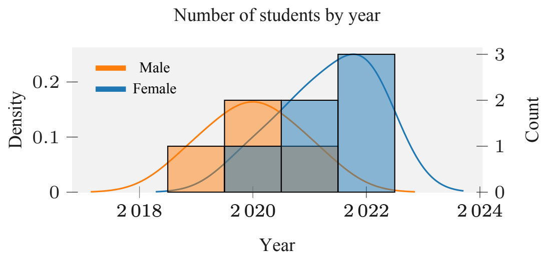

Figure 4. Examples of KDE histogram.

helps identify when a process changes in terms of business process perspectives, e.g., control flow, time, and data [Sato *et al.*, 2022]. In this paper, we use the drift framework in [Sato *et al.*, 2021] to detect drifts in a target variable of the process based on an adaptive windowing approach.

For each activity (e.g., a specific course), we derive a time series, which is a sequence of ordered data points, containing all the values for the target variable (e.g., students' grades). For every time series, we applied the ADWIN change detector, which detects drift from data sequences, providing rigorous guarantees of performance as bounds on the rates of false positives and false negatives [Bifet and Gavaldà, 2007].

For instance, we can analyze the drift in the data perspective to identify possible significant grade changes in a course. Figure 5 shows an example of average grades obtained by students in a certain course over the years and the detected drift points. Further exploration may be conducted to understand why the process has drifted, such as a professor change or other educational aspects.

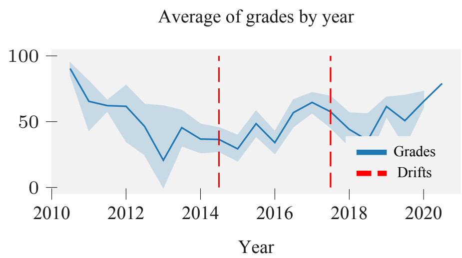

Figure 5. Examples of drift detection.

4.3 Correlation Analysis

To perform the *correlation analysis*, we calculate the correlation of certain variables to chosen outcomes. For instance, we can calculate the correlation of admission score, age, and gender concerning the student outcome: graduate or dropout. For measuring the likelihood of the null hypothesis (the observed difference is due to chance alone), we used Student's t-test [Boneau, 1960] and a chi-square test [Corder and Foreman, 2014], with significance level $\alpha = 0.01$. For the variables whose null hypothesis is false, we used the entropy test [Maimon and Rokach, 2014] to measure each variable's information gain (IG) concerning the defined outcome. We aim to identify the variables with higher IG, which would be the ones more correlated to the outcome.

4.4 Process map

Process map analysis enables using the event log to automatically calculate the process map: a model that generalizes the behavior of all cases in the log. In other words, the process map illustrates the behavior of all students in the log. From start to end, the process map models the different paths students can take, considering the order in which students take (or retake) courses. The process map can have its nodes grouped by some criteria for abstraction. Also, the process map nodes can be colored based on one variable, such as frequency, grade, attendance, etc, as described in Section 3.

By calculating the process map in a real-world educational environment, it is possible to analyze the learning paths followed by students, identifying the sequence of activities, resources, and interactions that contribute to successful learning outcomes. This analysis can help identify effective learning patterns and tailor educational content to individual student needs. The process map can be further analyzed to identify bottlenecks, inefficiencies, and opportunities for improvement. It lets people visualize activities where low grades occur, redundant activities exist, or resources are underutilized, facilitating targeted process optimization efforts.

4.5 Plot analysis

The *plot analysis* consists of creating different types of plots directly from the event log. Educators and other process specialists can use the analysis results to construct valuable insights for *process improvement*.

For example, once specialists know which scores and students' profiles are more likely to lead to dropouts, they can configure the classes, number, and paths for admitted students, who are, therefore, expected to be more likely to complete their courses; once the curriculum paths with greater retention are discovered, specialists can act specifically to treat them, either by discovering the reason of the retention (complexity or other reason), or even by trying to reduce the flow of certain students profile through those paths. Using the drift analysis of a specific attribute, e.g., students' grades, the specialists can identify significant changes in the attribute over time. Then, analysis can be performed to identify other changes that occurred in the same period when the drift was detected, e.g., professor replacement.

5 Case Study

To illustrate the proposed approach, we perform a case study where we aim to understand how programs, courses, and students interact and what statistically significant and meaningful patterns lead students to different paths and outcomes (dropouts and graduates, long vs. short graduation time, high GPA[1] vs. low GPA, etc.). Therefore, we extract the students' transcripts from the educational information system of the university. Since it is also relevant to analyze students' background information (gender, age, city of birth, etc.), we extract that information too.

5.1 Dataset composition

We process raw data collected from the information system to create an event log. Each entry in the log corresponds to a course taken by a student enrolled in a bachelor's degree program. The log contains information about 12,185 enrollments from 11,290 students in the university between 1986 and 2022. Some students have re-enrolled. The log contains students' personal information and information about each enrollment, such as the full transcript with courses, grades, and attendance.[2]

In total, the log contains 437,690 entries. The data corresponds to undergraduate students from eight bachelor's degree programs offered on one of the university's campuses. The extracted attributes are described in the following for enrollments, students, and events that occurred. For categorical variables, we present the total count of categories or the percentages of each category. We show the median (with IQR) and mean (with standard deviation) for numerical variables.

Enrollment attributes description:

Enrollment Identifier: 12,185 enrollments.

Undergraduate program Id: *Agronomy* (15.44%), *Accounting* (12.89%), *Chemistry* (9.62%), *Civil Engr.* (13.06%), *Computer Engr.* (10.52%), *Electrical Engr.* (12.74%), *Management* (12.87%), or *Mechanical Engr.* (12.87%).

Shift: *morning and afternoon* (60.92%), *afternoon and night* (13.32%), or *only night* (25.76%).

Admission type: students could apply locally at the university via an entry test (26.29%); nowadays, students can apply from anywhere in Brazil via a *Unified Selection System* [SISU, 2022] shared by most universities (61.52%); they may also be admitted via *program change* (within the university) (6.17%), *transference* (from another university) (1.99%), SISU *waiting list* (3.68%), and *other* (0.34%).

Admission score (from 0 to 1000): 613.59 (IQR 533.52-664.50), 566.20 (SD 191.69).

Admission year (1986 to 2022): 2013 (IQR 2009-2018), 2011.82 (SD 7.55).

Admission age (16 to 64): 19 (IQR 18-21), 20.5 (SD 4.43).

Admission season: students can apply twice a year at the *Fall* admission (69.74%), and the *Spring* admission (30.26%).

Admission quota group: students could be admitted when *no quota policy* was implemented (26.70%); since 2012 [Brasil, 2012], students can apply as *quota* holders (33.92%) or as *no quota* holders (39.38%); quota groups include candidates with disabilities (PwD), self-declared black, brown, or Indigenous, with *per capita* gross family income $\leq$ 1.5 minimum wages, or who have attended public high school.

Enrollment situation: *Dropout* (44.44%), *Graduate* (35.01%), and *Attendee* (20.55%).

As for the student attributes, they have the following profile:

Personal Id: 11,290 students.

Gender: Male (64.47%), Female (35.53%).

Brazilian: Yes (99.53%), or No (0.47%).

State of birth: it is the *same state* as the one where the campus is located (59.14%), or other (40.86%).

City of birth: it is the *same city* as the one where the campus is located (19.68%), or other (80.32%).

High school type: *Public* (61.82%), *Private* (32.45%), or *not informed* (5.72%).

Ethnic group: *White* (55.30%), *Brown* (11.43%), *Black* (1.36%), *Yellow* (1.06%), *Indigenous* (0.09%), or *not declared* (30.76%).

Situation: students that *have graduated* at least once (37.64%), students currently *attending* a program and who have never graduated (22.11%), and students that *have dropped out* (have never graduated and are not attending any course) (37.64%).

Finally, the event descriptions are as follows:

Identifier of an event: a student took a course on the scope of an enrollment (437,690 events).

Course identifier: 813 courses.

Class identifier: 680 classes.

Total time (hours): 60 (IQR 45-75), 64.17 (SD 24.37).

Obtained Grade (from 0 to 100): 74 (IQR 61-84), 66.91 (SD 25.45).

Attendance (from 0 to 1000): 930 (IQR 850-991), 887.36 (SD 163.51).

Course type: *mandatory* (93.28%) or *elective* (6.72%).

Professor Name: 4610 professors.

Course situation: *passed* (69.64%), *failed* (15.58%), *canceled* (3.54%), *not completed* (2.34%), had the credit *validated* (8.86%), or been *dismissed* (0.23%).

Start of the course: 2015-Jan (IQR 2010 Jul-2019 Jan) 2013-Jul.

In this particular university, the curriculum of each program is structured by semesters, with courses arranged in a recommended sequence that reflects prerequisite requirements and ensures seat availability for students who progress as expected. When a student fails a course, however, their path must be adjusted to accommodate retakes, which are

[1] In Brazil, the GPA – known as "Coeficiente de Rendimento" – is a number from 0 (worst) to 100 (best) calculated as a weighted average of course grade and course hours from all courses taken by the student.

[2] All data used in this study were fully anonymized prior to analysis, handled under strict governance procedures, and the research protocol was approved by the institution's Ethics Committee (57106622.2.0000.0177), ensuring compliance with all ethical and confidentiality standards applicable to studies involving sensitive sociodemographic information.

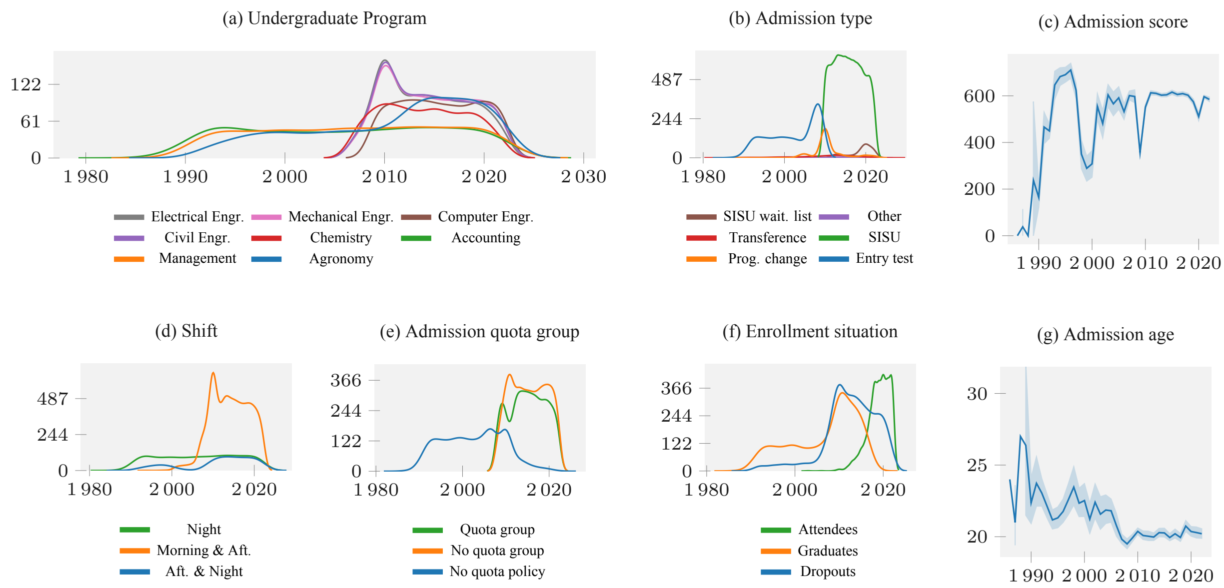

Figure 6. Enrollment variables over time.

subject to limited seats and irregular offering patterns. These curricular constraints are also represented in the event log.

Furthermore, to analyze the costs associated with this educational process, we included cost information in the log. For this, we utilized the work of Briskiewicz [2016], which defines the 2015 annual cost for each undergraduate program on campus. The annual cost is defined considering labor, security, cleanliness, depreciation, electricity, water, telephone, and other expenses. We used the Brazilian consumer price index [IBGE, 2022] to infer other years' annual cost for each program. These date is reported in the appendix, Table 5. Then, we calculated the practiced hourly cost for each program based on the annual cost and the total number of hours of courses taken by students in each program. Results are shown in the appendix, Table 6.

To illustrate the applicability of the proposed approach, we performed the two experiments in the following. The first is a time analysis that considers the data from undergraduate students from eight bachelor's degree programs offered on one university campus. For the second experiment, we selected a specific program, *Computer Engineering* (Cp), to perform further exploration. Upon processing, all the obtained results were discussed with specialists[3] from the application domain, with the purpose of better understanding the process.

5.2 Time analysis

In the following, we show the timeline plots for all programs, considering 18 different analyses, separated into 3 classes to be focused on: enrollment (Figure 6); student (Figure 8); and program (Figure 7). They were chosen to represent the most important scenarios that require intervention from managers.

In Figure 6, three programs (*Accounting*, *Management*, and *Agronomy*) have been offered since 1986, while other programs started to be offered in the middle of the 2000s. Around the same time, observe that there is a drift in the process that reflects in most variables: a significant change can be observed in "Shift" (*Morning and Afternoon* shifts started to be preeminent), "Admission type" (*SISU* started to be offered, while the *entry test* was discontinued), "Admission quota" (*Quota* policy was implemented), and "Admission Age" (students started to enroll younger).

As another effect of this drift, the proportion of *Dropouts* surpasses for the first time the *Graduates*. On the other hand, note that this drift is inversely proportional to the "Total time" (Figure 7), meaning that the actions leading to drifts (Figure 6) resulted in a reduction of the total time for completing the programs, at the price of more dropouts. This is important for managers to validate the strategies they adopted empirically, and to frame new strategies for the future of programs.

As for the student variables, in Figure 8, note that "High school type" and "Ethnic group" only started to be reported to the system in the 2000s, possibly because of the quota policy implementation. Also, the number of *Male* students surprisingly starts to be twice that of *Female* students.

In all analyses, the variables "Admission score", "Admission season", "Brazilian", "Same city", and "Course Situation" seem not to be affected by the drift in the 2000s, which may be another important information from the management perspective.

In 2020, another drift can be observed when the COVID-19 pandemic occurred. "Grades" reached the minimum in this drift, and *Not Completed* courses reached the maximum. It is interesting to notice that "Not completed" achieved a high index, and the item "Cancelled" remained similar. In this item, therefore, we can see an abandonment of the course without the request for cancellation, which may have an impact on the "Grades" index, which, in fact, happened. The abandonment choice will also impact the GPA score.

5.2.1 Practitioners: Time analysis

From the perspective of program coordinators and, to some extent, university administrators, the time analysis provides

[3] In the context of this study, discussions were conducted with program coordinators at the target university, who are formally responsible for curriculum planning, pedagogical monitoring, and academic decision-making at the program level.

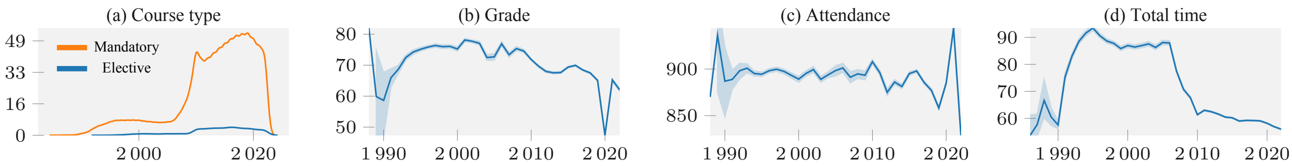

Figure 7. Program variables over time.

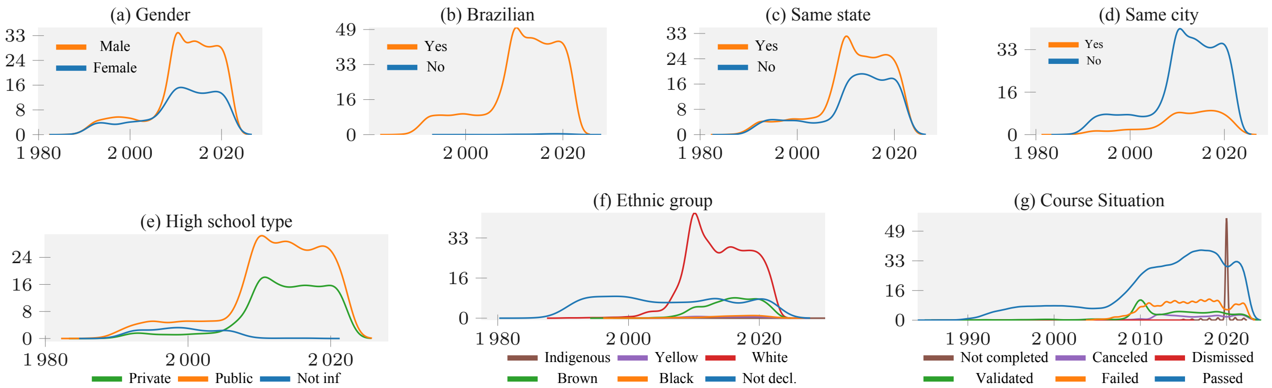

Figure 8. Student variables over time.

valuable insights to guide strategic decisions. For instance, identifying a significant proportion of students dropping out of courses or programs highlights the need for early academic support or intervention protocols. Moreover, recognizing that younger students are enrolling more frequently could inform the design of preparatory programs, marketing strategies targeting high school seniors, or even adjustments to first-semester pedagogical strategies.

Additionally, by identifying dropout rates according to the students' admission paths (e.g., SISU, entry tests, internal transfers), decision-makers may consider adjusting the preferred admission mechanisms for specific programs. If a given admission path consistently correlates with higher dropout rates or lower performance, program administrators could revise the proportion of seats allocated to each admission type or implement support programs targeted at these student groups.

University administrators or program coordinators, holding information about the large number of courses' dropouts from Figures 6, 7, and 8, could advise students to cancel courses in time, to avoid a later decrease in their GPA scores and the subsequent frustration. Also, knowing that younger students are entering university, the budget for advertising university programs can be increased, especially for the final years of elementary school, as well as high school.

5.3 Correlation for enrollment variables

Considering all undergraduate programs, the following correlations were measured for enrollment variables ("Program", "Shift", "Admission type", "Score", "Year", "Age", "Season", and "Quota group"):

- Enrollment variables with respect to "Enrollment situation" (*Dropout* and *Graduate*). We excluded the *Attendee* enrollment situation in this case.
- Enrollment variables concerning the "Duration" of the program. Only the *Graduates* cases were considered here. We split enrollments into two groups: *Long* duration (students who took more than five years to graduate) and *Short* duration (students who took five years or less to graduate).
- Enrollment variables concerning graduation "GPA". Here, we considered only the *Graduates* cases too. The enrollments were further separated into two groups: *High* GPA (students who scored more than 0.8) and *Low* GPA (students who scored 0.8 or less).

Table 2 shows the results, while Figure 9 shows the IG of each variable concerning the defined outcome.

When testing the correlation of variables concerning "Enrollment situation" and "GPA", the *Admission Score* shows the greatest IG among the statistically significant variables ($p<0.01$). Enrollments with higher *Admission scores* correlate to graduated and *High* GPAs students, while lower admission scores suggest *Dropout* and *Low* GPAs. This reinforces Voelkle and Sander [2008], indicating that admission score affects dropout through an influence on university grades. Students with lower admission scores perform poorly in exams and are likely to dropout. "Admission score" has no statistically significant correlation to the "Duration" of the enrollment.

The *Program* chosen by the student also correlates to "Enrollment situation", "Duration", and "GPA" (see Table 2). *Agronomy*, *Accounting*, and *Management* programs have been offered since 1986 and are mostly offered at night. *Chemistry* and the four *Engineering* programs started in the second semester of the 2000s, and are offered in the *mornings and afternoons* or *afternoons and nights*. These two groups of programs are characterized by variables "Admission year" and "Shift". As shown in Figure 9, these variables, along with "Program", have, in general, a *high* IG considering the three outcomes.

Table 2. Correlation measurement for enrollments (12,185 total).

Variable	Enrollment situation			Graduation duration			GPA		
	Dropout (5,415)	Graduated (4,266)	P	Short (2,750)	Long (1,516)	P	Low (3,070)	High (1,196)	P
Program			<.001			<.001			<.001
Agronomy	515 (9.51%)	937 (21.96%)		658 (23.93%)	279 (18.40%)		738 (24.04%)	199 (16.64%)	
Accounting	485 (8.96%)	880 (20.63%)		726 (26.40%)	154 (10.16%)		515 (16.78%)	365 (30.52%)	
Chemistry	684 (12.63%)	303 (7.10%)		192 (6.98%)	111 (7.32%)		225 (7.33%)	78 (6.52%)	
Civil Engr.	646 (11.93%)	552 (12.94%)		237 (8.62%)	315 (20.78%)		430 (14.01%)	122 (10.20%)	
Computer Engr.	768 (14.18%)	138 (3.23%)		24 (0.87%)	114 (7.52%)		108 (3.52%)	30 (2.51%)	
Electrical Engr.	907 (16.75%)	338 (7.92%)		126 (4.58%)	212 (13.98%)		301 (9.80%)	37 (3.09%)	
Management	617 (11.39%)	767 (17.98%)		667 (24.25%)	100 (6.60%)		444 (14.46%)	323 (27.01%)	
Mechanical Engr.	793 (14.64%)	351 (8.23%)		120 (4.36%)	231 (15.24%)		309 (10.07%)	42 (3.51%)	
Shift			<.001			<.001			<.001
Morning and aft.	3,482 (64.30%)	2,203 (51.64%)		1,102 (40.07%)	1,101 (72.63%)		309 (10.07%)	42 (3.51%)	
Aft. and night	831 (15.35%)	416 (9.75%)		255 (9.27%)	161 (10.62%)		335 (10.91%)	81 (6.77%)	
Only night	1,102 (20.35%)	1,647 (38.61%)		1,393 (50.65%)	254 (16.75%)		959 (31.24%)	688 (57.53%)	
Admission type			<.001			<.001			<.001
SISU	3,499 (64.62%)	1,769 (41.47%)		787 (28.62%)	982 (64.78%)		1,338 (43.58%)	431 (36.04%)	
Entry test	1,307 (24.14%)	1,897 (44.47%)		1,531 (55.67%)	366 (24.14%)		1,239 (40.36%)	658 (55.02%)	
Sisu waiting list	241 (4.45%)	14 (0.33%)		9 (0.33%)	5 (0.33%)		13 (0.42%)	1 (0.08%)	
Program change	277 (5.12%)	423 (9.92%)		286 (10.40%)	137 (9.04%)		354 (11.53%)	69 (5.77%)	
Transfer	70 (1.29%)	147 (3.45%)		124 (4.51%)	23 (1.52%)		117 (3.81%)	30 (2.51%)	
Other	21 (0.39%)	16 (0.38%)		13 (0.47%)	3 (0.20%)		9 (0.29%)	7 (0.59%)	
Admission score	540.22 (197.22)	584.28 (210.48)	<.001	579.25 (228.59)	593.39 (172.56)	.02	578.54 (211.13)	599.01 (208.19)	.004
Admission age	20.89 (5.14)	20.28 (3.60)	<.001	20.59 (3.81)	19.72 (3.09)	<.001	20.21 (3.45)	20.45 (3.95)	.06
Admission season			<.001			<.001			<.001
Fall	3,497 (64.58%)	3,536 (82.89%)		2,512 (91.35%)	1,024 (67.55%)		2,449 (79.77%)	1,087 (90.89%)	
Spring	1,918 (35.42%)	730 (17.11%)		238 (8.65%)	492 (32.45%)		621 (20.23%)	109 (9.11%)	
Admission quota			<.001			<.001			<.001
No quota policy	1,012 (18.69%)	2,203 (51.64%)		1,752 (63.71%)	451 (29.75%)		1,526 (49.71%)	677 (56.61%)	
Quota group	2,017 (37.25%)	1,007 (23.61%)		525 (19.09%)	482 (31.79%)		726 (23.65%)	281 (23.49%)	
No quota group	2,386 (44.06%)	1,056 (24.75%)		473 (17.20%)	583 (38.46%)		818 (26.64%)	238 (19.90%)	

Nightly programs (*Agronomy*, *Accounting*, and *Management*) generally have a higher proportion of graduates, short duration, and high GPA enrollments. In comparison, daily programs (*Chemistry* and *Engineering*) have a higher proportion of dropouts, long duration, and low GPAs. The difference in outcomes of these two groups of programs can be related to the programs' profiles. *Chemistry* and *Engineering* programs have complex math and physics curricula, making it more challenging for students who do not have a solid background in these fields to advance in the program [Paura and Arhipova, 2014].

It can be further observed that "Admission type", "Age", "Season", and "Quota" are also correlated to "Enrollment situation", "Duration", and "GPA". Students admitted via *entry test* have a high proportion of graduates, short duration, and high GPA. In contrast, students from *SISU* have a higher proportion of dropouts, long duration, and low GPAs during their studies. Entry tests were used only when *Agronomy*, *Accounting*, and *Management* programs were offered, which, in summary, is correlated to graduates/short duration/high GPA students. *Transference* and *Program change* students correlate to graduates, indicating that when a student moves from one program to another, this option does not usually lead to dropout. Almost all students admitted via the *waiting list* of SISU have dropped out.

The waiting list is a mechanism that allows students who have not gotten a high admission score to enter the university when there is a vacancy, independent of scores. And, as we stated before, "Admission score" is highly correlated to "Enrollment situation". Besides, it can be observed that older students have a higher proportion of dropouts but a lower proportion of long-duration enrollments. Students admitted in *Spring* typically are the ones who did not have admission scores high enough to be admitted in *Fall*. For that reason, *Fall* students have a higher proportion of graduates, while *Spring* students have a higher proportion of dropouts. *Quota group* students, i.e., those who fit within a set of governmental rules such as financial conditions, ethnicity, etc., also evidence a higher correlation with dropout.

5.3.1 Practitioners: Enrollment Correlation

The correlation analysis of enrollment variables reveals actionable insights that can support program-level and institutional decision-making. For example, understanding that students admitted, e.g., via SISU or waiting lists, tend to lead to higher dropout rates or lower GPAs allows coordinators to design targeted mentoring or onboarding strategies to support these groups. Also, knowing that students of Spring enrollment have higher dropout rates because of their lower GPAs, suggest offering more preparatory courses. It is also possible to plan the number of seats for admissions via the entry tests and via SISU, since this is optional, and students who enroll via the entry test tend to graduate more.

Table 3. Correlation measurement for students (11,290 total)

Variable	Enrollment situation			Graduation duration			GPA		
	Dropout (4,545)	Graduated (4,249)	P	Short (2,719)	Long (1,530)	P	Low (3,060)	High (1,189)	P
Gender			<.001			<.001			<.001
Female	1,409 (31.00%)	1,805 (42.48%)		1,267 (46.60%)	539 (35.23%)		1,169 (38.20%)	637 (53.57%)	
Male	3,136 (69.00%)	2,444 (57.52%)		1,452 (53.40%)	991 (64.77%)		1,891 (61.80%)	552 (46.43%)	
Brazilian	4,523 (99.52%)	4,239 (99.76%)	.08	2,711 (99.71%)	1,528 (99.87%)	.47	3,054 (99.80%)	1,185 (99.66%)	.62
Same state	2,354 (51.79%)	2,798 (65.85%)	<.001	1,861 (68.44%)	956 (62.48%)	<.001	2,037 (66.57%)	780 (65.60%)	.57
Same city	721 (15.86%)	955 (22.48%)	<.001	645 (23.72%)	312 (20.39%)	.001	688 (22.48%)	269 (22.62%)	.95
High school type			<.001			<.001			<.001
Not informed	461 (10.14%)	183 (4.31%)		123 (4.52%)	32 (2.09%)		129 (4.22%)	26 (2.19%)	
Private	1,538 (33.84%)	1,224 (28.81%)		662 (24.35%)	559 (36.54%)		942 (30.78%)	279 (23.47%)	
Public	2,546 (56.02%)	2,842 (66.89%)		1,934 (71.13%)	939 (61.37%)		1,989 (65.00%)	884 (74.35%)	
Ethnic group			<.001			<.001			<.001
Yellow	47 (1.03%)	30 (0.71%)		18 (0.66%)	11 (0.72%)		23 (0.75%)	6 (0.50%)	
White	2,293 (50.45%)	2,218 (52.20%)		1,201 (44.17%)	1,020 (66.67%)		1,636 (53.46%)	585 (49.20%)	
Indigenous	8 (0.18%)	1 (0.02%)		0 (0.00%)	1 (0.07%)		1 (0.03%)	0 (0.00%)	
Not declared	1,660 (36.52%)	1,645 (38.71%)		1,337 (49.17%)	306 (20.00%)		1,120 (36.60%)	523 (43.99%)	
Brown	494 (10.87%)	311 (7.32%)		140 (5.15%)	171 (11.18%)		246 (8.04%)	65 (5.47%)	
Black	43 (0.95%)	44 (1.04%)		23 (0.85%)	21 (1.37%)		34 (1.11%)	10 (0.84%)	

Furthermore, correlations between admission scores and academic success may encourage the adoption of stricter cut-offs, additional entrance evaluations, or preparatory leveling modules. Programs with historically low retention or performance could also benefit from adjusting their selection processes to prioritize profiles associated with more successful trajectories.

By observing that some programs or shifts (e.g., night programs) have systematically better outcomes in terms of graduation rates and GPA, decision-makers might reconsider resource allocation, scheduling strategies, or curricular structure to align with more favorable student performance patterns. For example, one may reduce the number of courses in daytime programs or stimulate common subjects to be taught in nighttime programs. Adjusting the percentage of students admitted by each path, e.g., shortening access via waiting lists, may be a strategic option to reduce dropout and improve academic outcomes in specific contexts.

Finally, in the macro context, if agencies responsible for public policies for higher education access data of universities showing that students admitted via SISU waiting lists experience high dropout rates, they could redesign these options in the SISU platform's system or maybe replan the calendar. What happens in these cases of students admitted via waiting lists is that they start classes when the semester is already ongoing. This alone leads students to try to catch up with their peers and try to understand already covered content on their own, which creates insecurity. This, added with the fact that students on waiting lists have lower GPAs, increases the overall dropout rates.

5.4 Correlation for student sociodemographic variables

Considering the students from all undergraduate programs, the following correlations were measured for student variables ("Gender", "Country", "State", "City", "High school type", and "Ethnic group"):

- Student x situation (*Dropout* and *Graduates*, excluding *Attendees*).
- Student x duration of the program. We considered only the first completed graduation for students with more than one.
- Student x grades. We also considered only the first completed graduation.

Table 3 summarizes the results, while Figure 9 shows the IG of each variable concerning the defined outcome.

The variable "Gender" has the greatest grade IG. *Male* students have a higher dropout rate, longer duration, and lower GPAs. This aligns with previous studies Freitas *et al*. [2023]; Salgado *et al*. [2025] and leads to the belief that, in absolute numbers, male enrollment is superior compared to female, but the graduate index is slightly higher in favor of women. Students from the *Same state* and *City* tend to be more successful in graduating with short-duration enrollment. The variable "Same state" has a higher IG considering duration, which reflects the difficulty of students who migrate from other cities or states, maybe due to living away from their families or due to financial support reasons.

The "High school type" has the second higher IG considering all outcomes. *Private* school students have a higher proportion of dropouts, long duration, and low GPAs. Concerning the "Ethnic groups", *not declared*, *white*, and *black* students have a higher proportion of graduates. Furthermore, *not declared* students are associated with short duration and high GPA. We highlight that *not declared* was the only status option when just *Agronomy*, *Accounting*, and *Management* programs were offered, so the graduates, short duration, and high GPA correlation were somehow expected.

5.4.1 Practitioners: Correlation analysis for student sociodemographic variables

By analyzing student-related variables, program managers can identify key demographic and academic background characteristics associated with success or dropout. For instance,

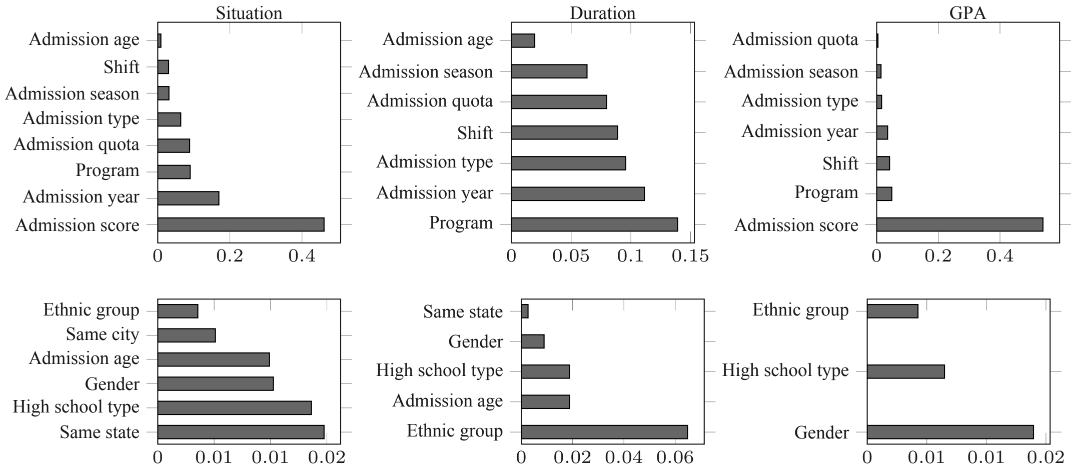

Figure 9. Information Gain of each variable with respect to the defined outcome.

understanding that male students (or those from outside the state or city) are more likely to drop out can be used to develop support for academic mentoring, housing, or integration workshops aimed at reducing these vulnerabilities.

Likewise, the strong correlation between high school type and performance suggests the need for differentiated pedagogical strategies in the early semesters. Students from public schools, although more likely to persist, may still require structured academic reinforcement. In contrast, those from private schools with higher dropout rates may need early engagement strategies to maintain motivation and performance. Students entering through competitive mechanisms with higher admission scores tend to follow more linear paths with fewer retakes, while those admitted via alternative routes or with lower entrance grades face additional challenges in adapting to course rigor, pacing, and prerequisite chains. These disparities help clarify why programs with more heterogeneous entry profiles show increased fragmentation in early semesters and higher attrition. Such observations underscore the interplay between prior academic background, curricular design, and institutional entry pathways, offering an understanding of how performance differences materialize and where reforms or targeted support initiatives may be most effective.

The discovered relationships between ethnic groups and outcomes also raise awareness of the need for inclusive monitoring and support, especially in the context of quota-based admission policies. Institutional efforts to better understand the intersection between student identity and performance can contribute to more equitable academic environments.

These findings also help refine outreach, onboarding, and retention policies. For example, course coordinators could request tailored reports from institutional data systems to guide the assignment of tutors, inform curricular revisions, and adjust policies for student admission and follow-up according to their profiles and risks.

5.5 Process map analysis

Figure 10 shows the process models for the programs Agronomy (Ag), Accounting (Ac), Chemistry (Cm), Civil Engr. (Cv), Computer Engr. (Cp), Electrical Engr. (El), Management (Mn), and Mechanical Engr. (Mc).

Each node is colored according to the average grade obtained by the students. Redder means closer to the descending grades, while greener means closer to the maximum grade. Ag, Ac, Mn, and Cv present more green and yellow colors, while Mc, Cm, El, and Cp have a considerable amount of orange and red. This reflects the results in Table 2. Beyond the visual interpretation of trajectories, these patterns reveal structural and contextual differences across programs that help explain their distinct dropout rates. Some programs (e.g., Ag, Ac, Mn, and Cv) exhibit predominantly green and yellow nodes, which may reflect a combination of more stable curricular structures, lower early-course failure rates, and student cohorts whose prior preparation better aligns with program demands. In contrast, the larger concentration of orange and red nodes suggests a higher academic burden in foundational courses, stronger mathematical or laboratory requirements, or curricular sequences that amplify the impact of early failures on long-term progression.

Remark that Ag, Ac, Mn, and Cv have a higher proportion of graduates in comparison with dropouts, and Cp has the worst index of graduates among all others. Most programs present a linear path, but the older courses (Mn, Ac, and Ag) have parallel paths, indicating a change in curricula. Furthermore, some models present a spaghetti-like behavior at the beginning, suggesting a trend for redoing courses.

5.5.1 Practitioners: Process map analysis

Process maps allow course coordinators to visualize student trajectories and detect critical patterns, such as high failure or retake rates in specific courses. These insights can inform targeted actions like revising prerequisites of programs, offering academic support, or adjusting course sequencing to reduce delays and dropouts.

Color-coded maps also highlight low-performance areas, guiding interventions in course content, teaching methods, or instructor allocation. This can draw professors' attention to the performance of their courses with respect to others, so that these professionals can progressively reevaluate their practice and methods in classes. Moreover, complex or fragmented paths in older curricula may highlight the need for curricular simplification or clearer equivalency rules.

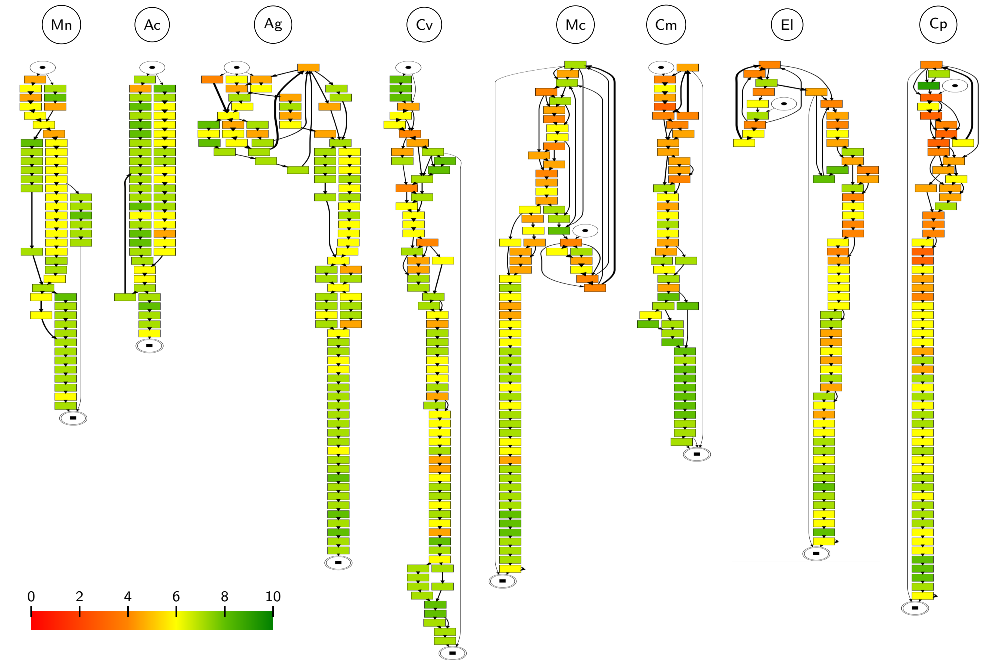

Figure 10. Process models for each program - colors represent grades.

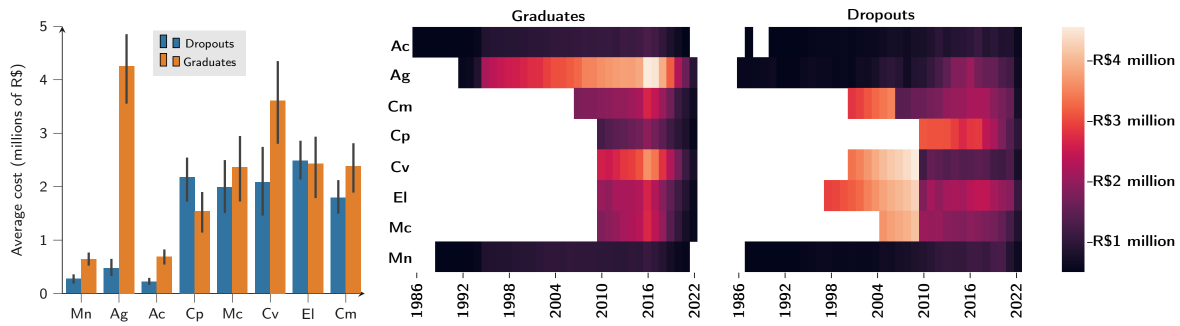

Figure 11. Graduates vs. Dropouts - Average cost by program and heatmaps by year.

By translating raw academic data into visual models, this analysis equips managers with concrete, data-driven evidence to refine curricular planning and improve student flow through the program.

5.6 Plot analysis

To illustrate the plot analysis, we calculated the practiced hourly cost for each program based on the annual cost and the total number of hours of courses taken by students in each program, as shown in Table 6. Then, we split the cost for graduates and dropouts. Figure 11 (left) shows the average cost for each program, and (center and right) shows a cost heatmap for graduates and dropouts through the years.

Ag is the program with the highest average cost. Its heatmap shows an increasing cost peaking around 2016 for both graduates and dropouts. However, since most Ag students graduate, the average cost for dropouts is 9 times smaller than for graduates. Mn and Ac, in contrast, are the programs with the smallest cost. The other programs (Cp, Mc, Cv, El, and Cm) have intermediate cost, approximately three times more expensive than Mn and Ac. This difference can be explained in terms of the programs' curricula, as Mn and Ac do not require labs and have a shorter total time for completion. Besides, while Mn and Ac have at least twice the cost for graduates compared to dropouts, Cp and El have the cost for dropouts surpassing that of graduates. The heatmaps also indicate that Mc, El, Cv, Cp, and Cm have a higher cost with dropout students in the first years of the programs, as the number of graduates in these years is low.

5.7 Specific program analysis

A detailed analysis was conducted for the Cp program because it has the highest dropout rate among all the evaluated programs. Although the case study focuses on this specific program, the same analytical procedure can be adapted to other levels of granularity, from institution-wide analyses, including multi-campus structures, to program-level evaluations or even course-level investigations within a curriculum.

For enrollment variables, we measure the correlation concerning the same four outcomes as before (enrollment situation, duration, and GPA). Table 4 summarizes the results.

The following IG were found for the statistically significant

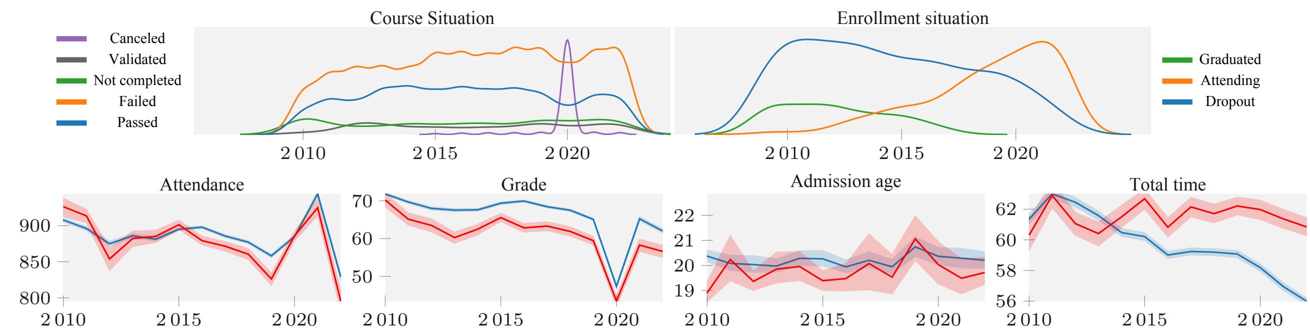

Figure 12. Variables progress over time. Red lines correspond to the selected program.

Table 4. Correlations for Cp enrollments (1,282 total) and students (1,183 total). Some intraclass proportions are omitted for clarity.

Variable	Enrollment situation			Duration	GPA
	Dropout (768)	Graduated (138)	P	P	P
Admission type			<.001	.046	.94
SISU	623 (81.12%)	110 (79.71%)			
Entry test	66 (8.59%)	23 (16.67%)			
Sisu waiting list	72 (9.38%)	1 (0.72%)			
Program change	3 (0.39%)	0 (0.00%)			
Transfer	3 (0.39%)	3 (2.17%)			
Other	1 (0.13%)	1 (0.72%)			
Admission score	571.55 (143.92)	584.82 (142.73)	.38	.43	.07
Admission age	19.95 (3.75)	18.93 (1.74)	<.001	.88	.22
Admission season			.008	.38	.07
Fall	370 (48.18%)	84 (60.87%)			
Spring	398 (51.82%)	54 (39.13%)			
Admission quota			.003	.49	.68
No quota policy	4 (0.52%)	4 (2.90%)			
Quota group	319 (41.54%)	69 (50.00%)			
No quota group	445 (57.94%)	65 (47.10%)			
	Dropout (669)	Graduated (138)	P		
Gender			.69	.28	.39
Female	81 (12.11%)	18 (13.04%)			
Male	588 (87.89%)	120 (86.96%)			
Brazilian	665 (99.40%)	137 (99.28%)	>.99	>.99	>.99
Same state	321 (47.98%)	82 (59.42%)	.02	.57	.78
Same city	95 (14.20%)	22 (15.94%)	.69	>.99	.87
High school type			.29	.30	.98
Private	248 (37.07%)	44 (31.88%)			
Public	421 (62.93%)	94 (68.12%)			
Ethnic group			<.001	.43	.75
Yellow	14 (2.09%)	3 (2.17%)			
White	376 (56.20%)	108 (78.26%)			
Indigenous	2 (0.30%)	0 (0.00%)			
Not declared	192 (28.70%)	8 (5.80%)			
Brown	76 (11.36%)	16 (11.59%)			
Black	9 (1.35%)	3 (2.17%)			

variables to enrollment situation: "Admission age" (.027), "Type" (.024), "Quota" (.007), "Season" (.001), and "Ethnic group" (.038). We could not find any statistically significant variables related to duration and GPA, possibly due to the sample size.

While the timeline plot of all programs (Figure 8) shows that the number of passed courses is greater than failed, it is possible to observe in Figure 12 that Cp students have more failed courses. In the same way, it is possible to observe that Cp students have more than twice the dropout rate of graduates and lower grades compared to all other programs.

We performed process map analysis to obtain the process models for graduates (Figure 13) and dropout (Figure 14) students' paths, using two levels of abstraction: by course (models 1, 2, and 3) and by semester (models 4, 5, and 6). In Figure 13, models 3 and 6 are colored based on frequency: bluer means higher frequency; models 1 and 4 are colored based on attendance, and models 2 and 5, based on grades. In models 1, 2, 4, and 5, redder means falling grade/attendance, while greener means closer to maximum grade/attendance. Models 1 to 6 are colored similarly in Figure 14.

When analyzing the models in Figures 13 (graduates) and 14 (dropouts), we observe that graduates follow a more straightforward path from semester 1 to 9, while dropouts tend to go from 1 to 3, or only attend semester 1. This suggests students' difficulty in staying at the university, especially at the start point, when courses that are more dependent on students' high school backgrounds are offered. Furthermore, in the first semesters of the program, usually, more general courses are offered since they serve as a background for specific courses coming later. Thus, perhaps the lack of identification with the chosen course in these early semesters, as well as the frustration and insecurity of pursuing the wrong path, leads the students to drop out. For the sake of clarity, we only show aliases in replacement for the names of courses in this figures [4].

For graduates (Figure 13), the most retaken semester is the second (a), but this does not reflect lower grades/attendance. The lowest attendance can be observed in the third and last semesters, and the lowest grades are around semesters 3-5 and semester 1. When analyzing courses, we found that the *U* course is the most retaken, and it has lower grade/attendance, while *V* has also been retaken but with not so low grade/attendance (b). This reveals two courses where students behave differently: in U, students fail the course and retake it without going to classes (high frequency - low grade/attendance), while in *V*, students go to classes and almost pass it (high frequency - low grade - high attendance). *W* also has a pattern of high frequency-low grade/attendance (c). Courses *U, V*, and *W* should have been further processed to understand why this behavior occurs.

For dropouts (Figure 14), the most retaken semester is the first, and students tend to drop out in semester 1. Dropouts have poor attendance and grades in almost all courses and

[4] Complex variables (U), Electric circuits analysis (V), Program conclusion work 2 (W), Integral and differential calculus 1 (X), Physics 1 (Y), Analytic geometry and linear algebra (Z).

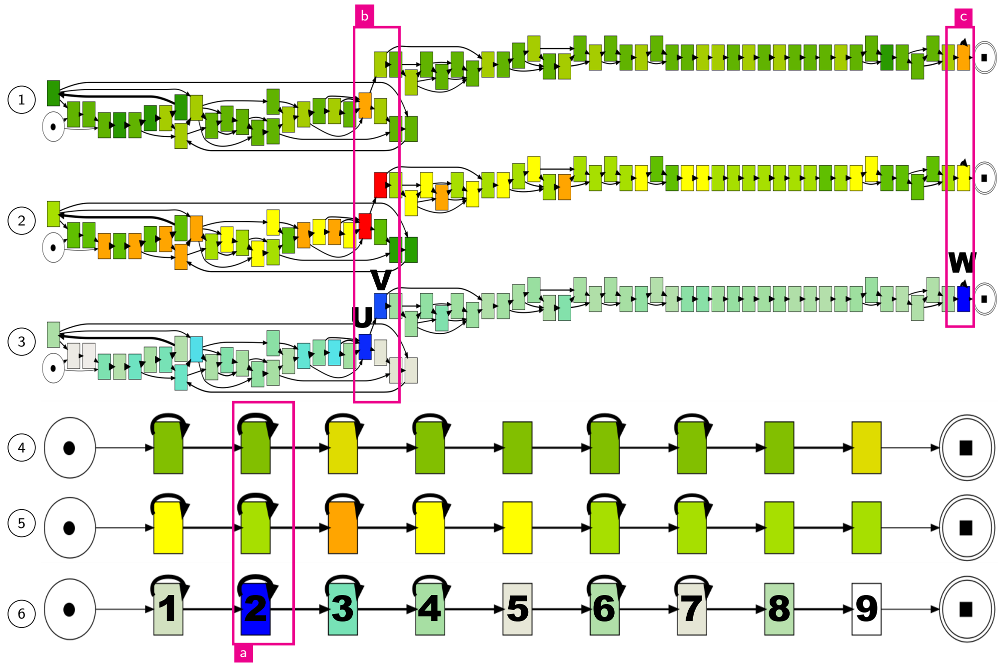

Figure 13. Process models for graduates

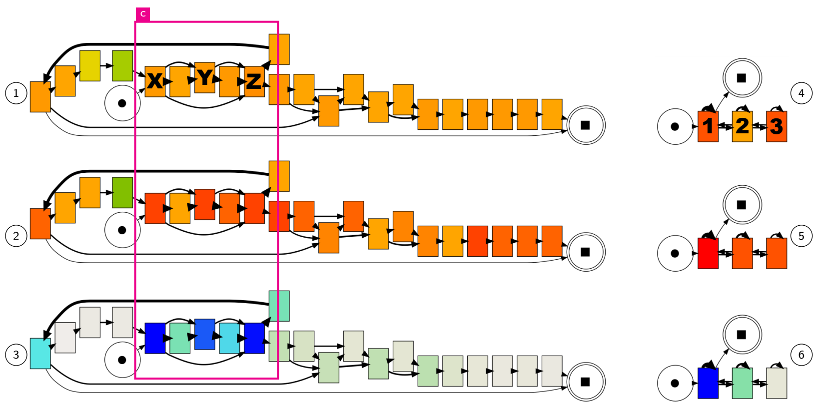

Figure 14. Process models for dropouts

semesters. The most retaken courses are *X*, *Y*, and *Z*, which, in general, have the lowest grades (c). This reinforces the importance of a high school background, especially in Math and Physics, subjects on which these courses depend.

When analyzing student grades by course, we identified significant drifts in some of them. The drifts detected on course V are shown in Figure 15 (top). They indicate three behaviors for the grade's time series: the grades go down from 2010 to 2014, then the first drift occurs; the grades go up until late 2017, then the second drift occurs; finally, the grades have a more oscillating behavior from 2018 onward.

We explored further variables from course V to analyze if some of them have similar behaviors. It turns out that course V had several professors during the analyzed timespan. The KDE of variable professors is shown in Figure 15 (bottom). The declining behavior (before the first drift) matches when Professor 5 lectured on course V, and the behavior changes for Professors 1 and 3. When Professor 2 replaces Professor 4, the second drift occurs. Further processing may help to identify specific features that differentiate their classes.

5.7.1 Practitioners: Plot analysis

This visual resource allows for building and analyzing student pathways (and all correlated variables) through graphical representations, such as process maps, which enhance the understanding of academic flows.

These insights allow managers and coordinators to quickly identify critical bottlenecks, deviations, and significant patterns, supported by quantitative metrics like dropout rates and average course duration.

Based on these insights, they can reconfigure the trajectories based on targeted improvements that have a higher chance of reducing unsuccessful trajectories. This tends to empower educational leaders to make non-empirical data-driven decisions that effectively improve program quality and student success.

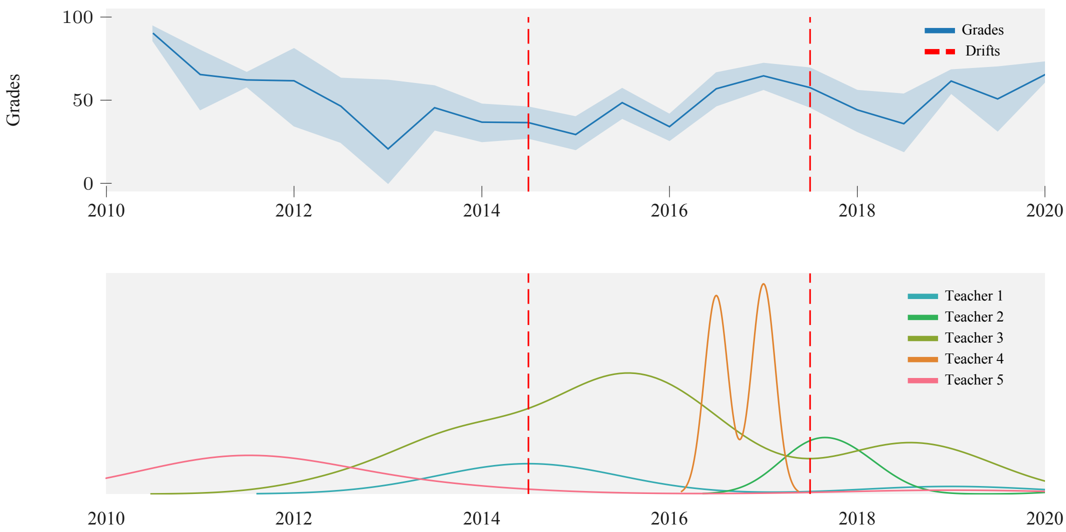

Figure 15. Average grades and detected drifts for course V (top) and the professor KDE in the same timespan (bottom).

6 Discussions and limitations

A deeper interpretation of the observed patterns can be articulated through established educational theories on student retention. According to Tinto's model Tinto [1975], students' likelihood of persisting in higher education is strongly influenced by their academic and social integration, with early academic performance acting as a critical determinant of long-term success. The trajectories uncovered in our process maps, particularly the fragmentation and repeated failures concentrated in the initial semesters of certain programs, reflect precisely the forms of weak academic integration that Tinto identifies as precursors to attrition.

Similarly, Bean's model Bean [1980] emphasizes the role of background characteristics and institutional conditions in shaping students' intentions to remain enrolled. The strong associations we found with admission scores, high school type, gender, and place of origin align with Bean's view that students' pre-university preparation and demographic contexts influence both their academic adjustment and their perceived ability to succeed. Together, these theoretical perspectives help contextualize our findings, suggesting that the differences in dropout rates and performance across programs and admission routes are not merely structural features of the curricula but manifestations of broader mechanisms of integration, preparedness, and institutional fit described in the retention literature[5].

While the proposed approach shows promising results, some limitations must be noted. Although this study examines the associations between multiple explanatory variables and student dropout, it does not model the possible causal relationships among the explanatory variables themselves. These variables may interact in ways that are not explicitly captured by our approach, e.g., when a factor simultaneously influences both dropout and another predictor. As a result, some correlations reported here may partially reflect omitted variable bias or confounding effects inherent to observational educational data. Therefore, the findings should be interpreted as descriptive relationships rather than causal claims, and conclusions about "drivers" of dropout must be made with caution. Future work could incorporate causal modeling or structural representations of dependencies among predictors to further refine these insights.

It is also worth noting that the effectiveness of this study depends on the quality and consistency of the academic records. Another consideration is that the case study considered a single institution. Although the proposed approach can be extended to other academic settings, this was considered beyond the scope of this paper. Another limitation to be highlighted is that the analysis requires careful preprocessing and domain expertise to ensure meaningful interpretation. Due to contextual differences in curricula and policies, results may not generalize without some extent of adaptation. Broader validation across diverse institutions would strengthen the framework's applicability.

7 Conclusion

This study presents a methodology for curriculum mining, validating it through a case study of a Brazilian public university. The methodology can be used to help people understand: what are the variables with a higher correlation to the successful and less-successful students (in one program or overall); how data progress overtime and what drift points occurred; what different paths graduated, and dropout students tend to take; how grades, attendance, and frequency progresses along these paths; and how costs are quantified for dropout and graduates. The results were discussed with educators and specialists from the university. Specifically, in this case study, we found that admission scores, program, high school type, gender, and location are the variables with a higher correlation to successful and less successful students.

From the analysis conducted in this paper, we aim to facilitate comparisons between different campuses or universities, enabling them to exchange efficient practices towards common goals. For example, it should be possible to identify and correct problems proactively, direct funding to specific areas,

[5] To learn more about the Tinto and Bean models, as well as others, consult Costa and Gouveia [2018].

and conduct targeted advertising campaigns to attract more new students of a certain gender, age, or educational background. This information can also be used to plan the need for staff positions on campuses, to calculate the percentage of vacancies for new students entering tests, or to focus on those with the best graduate rates, such as SISU.

In a more limited context, program coordinators can make more assertive decisions to keep students in the program and analyze overall course performance. For instance, they can offer reinforcement courses for students entering in the spring if they have lower GPAs or organize support groups for socially vulnerable students, such as those who live far from their families and have a higher dropout rate.

In a broader context, the data can be used to inform public policy decisions that encourage university admission of students from public schools and propose affirmative actions, such as promoting gender equity in areas where male presence is the majority. Regarding SISU, the data could be useful for redesigning the options and the number of them in the system, as well as the admission schedules. This is because if a student enters university while classes are already underway, it requires more effort and support to stay on track and avoid dropping out. Moreover, the methodology opens opportunities for further developments, including validating the approach across different institutions, incorporating analytical components capable of anticipating risks earlier in the academic trajectory, and connecting the framework to continuous monitoring environments, thereby aligning it with emerging directions in educational data science.

Declarations

Acknowledgements

We thank CNPq, CAPES, FINEP, Araucária Foundation, UTFPR, IFSC, IFPR, and PUC/PR.

Funding

This research has no specific funding to report.

Authors' Contributions

LF is the main contributor and writer of this manuscript. SCF, JJF, and DMVS helped with the experiments, data analysis, and writing. MT and LT were responsible for the validation and writing. EES was responsible for the supervision, Funding acquisition, and review. All authors read and approved the final manuscript.

Competing interests

The authors have no competing interests to declare.

Availability of data and materials

The datasets generated and analyzed during this study can be made available upon request.

References

Araque, F., Roldán, C., and Salguero, A. (2009). Factors influencing university drop out rates. *Computers & Education*, 53(3):563–574. DOI: 10.1016/j.compedu.2009.03.013.

Awang Long, Z., Faizuddin Mohd Noor, M., *et al*. (2023). Factors influencing dropout students in higher education. *Education Research International*, 2023(1):7704142. DOI: 10.1155/2023/7704142.

Bäulke, L., Grunschel, C., and Dresel, M. (2022). Student dropout at university: A phase-orientated view on quitting studies and changing majors. *European Journal of Psychology of Education*, 37(3):853–876. DOI: 10.1007/s10212-021-00557-x.

Bean, J. P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in higher education*, 12(2):155–187. DOI: 10.1007/bf00976194.

Behr, A., Giese, M., Teguim K, H. D., and Theune, K. (2020). Early prediction of university dropouts–a random forest approach. *Jahrbücher für Nationalökonomie und Statistik*, 240(6):743–789. DOI: 10.1515/jbnst-2019-0006.

Bifet, A. and Gavaldà, R. (2007). Learning from time-changing data with adaptive windowing. In *Proceedings of the 7th SIAM International Conference on Data Mining*, pages 443–448, Minneapolis, Minnesota, USA. Society for Industrial and Applied Mathematics. DOI: 10.1137/1.9781611972771.42.

Boneau, C. A. (1960). The effects of violations of assumptions underlying the t test. *Psychological Bulletin*, 57(1):49. DOI: 10.1037/h0041412.

Brasil (2012). Lei nº 12.711, de 29 de agosto de 2012. Available at:`https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/l12711.htm`.

Briskiewicz, L. B. (2016). Identificação dos gastos dos cursos de graduação da universidade tecnológica federal do paraná câmpus pato branco e mensuração do custo ideal por aluno. Available at:`https://repositorio.utfpr.edu.br/jspui/bitstream/1/14181/3/PB_COCTB_2016_2_16.pdf`.

Cerdeira, J. M., Nunes, L. C., Reis, A. B., and Seabra, C. (2018). Predictors of student success in higher education: Secondary school internal scores versus national exams. *Higher Education Quarterly*, 72(4):304–313. DOI: 10.1111/hequ.12158.

Chapela-Campa, D., Dumas, M., Mucientes, M., and Lama, M. (2022). Efficient edge filtering of directly-follows graphs for process mining. *Inf. Sciences*, 610:830–846. DOI: 10.1016/j.ins.2022.07.170.

Chen, Y., Johri, A., and Rangwala, H. (2018). Running out of STEM: a comparative study across stem majors of college students at-risk of dropping out early. In *Proceedings of the 8th international conference on learning analytics and knowledge*, pages 270–279. DOI: 10.1145/3170358.31704.

Corder, G. W. and Foreman, D. I. (2014). *Nonparametric statistics: A step-by-step approach*. John Wiley & Sons. Book.

Costa, O. S. d. and Gouveia, L. B. (2018). Modelos de retenção de estudantes: abordagens e perspectivas.

REAd. Revista Eletrônica de Administração (Porto Alegre), 24(3):155–182. DOI: 10.1590/1413-2311.226.85489.

Freitas, M., Lara, G., Southier, L., Favarim, F., Dosciatti, E., Teixeira, L., and Teixeira, M. (2023). Assimetria de gênero na computação: um estudo de caso em uma universidade pública brasileira. In *Anais do XXXIV Simpósio Brasileiro de Informática na Educação*, pages 1007–1017, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/sbie.2023.234686.

Härdle, W., Müller, M., Sperlich, S., Werwatz, A., *et al.* (2004). *Nonparametric and semiparametric models*, volume 1. Springer. DOI: 10.1201/b10510-5.

IBGE (2022). Ipca - Índice nacional de preços ao consumidor amplo. Available at:`https://www.ibge.gov.br/estatisticas/economicas/precos-e-custos/9256-indice-nacional-de-precos-ao-consumidor-amplo.html`.

Inep (2020). Censo da educação superior 2019. Available at:`https://www.encurta.ae/7bvcf`.

Inep (2022). Indicadores de fluxo da educação superior. Available at:`https://www.encurta.ae/gd6by`.

Levy, Y. (2007). Comparing dropouts and persistence in e-learning courses. *Computers & education*, 48(2):185–204. DOI: 10.1016/j.compedu.2004.12.004.

Maimon, O. Z. and Rokach, L. (2014). *Data mining with decision trees: theory and applications*, volume 81. World scientific. DOI: 10.1142/9789812771728.

Marques, F. T. (2020). The return to higher education of dropout students in Brazil. *Cadernos de Pesquisa*, 50:1061–1077. DOI: 10.1590/198053147158.

Morais, D. C. S. d., Pontual Falcão, T., and Tedesco, P. (2024). Promoting children's participation in a participatory design process in a rural school: A new role needed? *Journal of the Brazilian Computer Society*, 30(1):116–132. DOI: 10.5753/jbcs.2024.3114.

Nakatumba, J. and van der Aalst, W. M. (2009). Analyzing resource behavior using process mining. In *Int. Conf. on BPM*, pages 69–80. Springer. DOI: 10.1007/978-3-642-12186-9_8.

OECD (2022). *Education at a Glance 2022*. DOI: 10.1787/3197152b-en.

Ortiz-Lozano, J. M., Rua-Vieites, A., Bilbao-Calabuig, P., and Casadesús-Fa, M. (2018). University student retention: Best time and data to identify undergraduate students at risk of dropout. *Innovations in education and teaching international*. DOI: 10.1080/14703297.2018.1502090.

Paura, L. and Arhipova, I. (2014). Cause analysis of students' dropout rate in higher education study program. *Procedia - Social and Behavioral Sciences*, 109:1282–1286. DOI: 10.1016/j.sbspro.2013.12.625.

Salgado, L. C. C., Moro, M. M., Araujo, A., de Figueiredo, R. V., Cappelli, C., Nakamura, F., and de Santana, T. S. (2025). Wit comes of age: The successful story of the women in information technology workshop. *Journal of the Brazilian Computer Society*, 31(1):36–49. DOI: 10.5753/jbcs.2025.4506.

Saqr, M., López-Pernas, S., Helske, S., and Hrastinski, S. (2023). The longitudinal association between engagement and achievement varies by time, students' profiles, and achievement state: A full program study. *Computers & Education*, 199:104787. DOI: 10.1016/j.compedu.2023.104787.

Sato, D. M. V., Barddal, J. P., and Scalabrin, E. E. (2021). Interactive Process Drift Detection Framework. In *ICAISC 2021: Artificial Intelligence and Soft Computing*, volume 12855 LNAI, pages 192–204, Zakopane, Poland. Springer, Cham. DOI: 10.1007/978-3-030-87897-9_18.

Sato, D. M. V., De Freitas, S. C., Barddal, J. P., and Scalabrin, E. E. (2022). A Survey on Concept Drift in Process Mining. *ACM Computing Surveys*, 54(9):1–38. DOI: 10.1145/3472752.

SISU (2022). Sistema de seleção unificada. Available at:`https://acessounico.mec.gov.br/sisu`.

Tayebi, A., Gomez, J., and Delgado, C. (2021). Analysis on the lack of motivation and dropout in engineering students in Spain. *IEEE Access*, 9:66253–66265. DOI: 10.1109/ACCESS.2021.3076751.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of educational research*, 45(1):89–125. DOI: 10.3102/00346543045001089.

Van Der Aalst, W. (2016). Data science in action. In *Process mining*. Springer. DOI: 10.1007/978-3-662-49851-4_1.

Voelkle, M. C. and Sander, N. (2008). University dropout: A structural equation approach to discrete-time survival analysis. *Journal of Individual Differences*, 29(3):134. DOI: 10.1027/1614-0001.29.3.134.

Xenos, M., Pierrakeas, C., and Pintelas, P. (2002). A survey on student dropout rates and dropout causes concerning the students in the course of informatics of the Hellenic Open University. *Computers & Education*, 39(4):361–377. DOI: 10.1016/S0360-1315(02)00072-6.

Zhang, J., Gao, M., and Zhang, J. (2021). The learning behaviours of dropouts in moocs: A collective attention network perspective. *Computers & education*, 167:104189. DOI: 10.1016/j.compedu.2021.104189.

Zhao, B. (2017). Web scraping. *Encyclopedia of big data*, pages 1–3. DOI: 10.1007/978-3-319-32001-4_483-1.

A Complementary datasets

Table 5. Yearly cost by program (Brazilian Real - R$)

Year	Mn	Ag	Ac	Cp	Mc	Cv	El	Cm
1986	-	-	17,201.53	-	-	-	-	-
1987	-	78,007.08	17,364.68	-	-	-	-	-
1988	17,171.31	80,898.51	18,008.33	-	-	-	-	-
1989	19,078.75	89,884.96	20,008.75	-	-	-	-	-
1990	24,572.70	115,768.38	25,770.50	-	-	-	-	-
1991	32,188.17	151,646.90	33,757.20	-	-	-	-	-
1992	35,386.38	166,714.47	37,111.30	-	-	-	-	-
1993	46,270.07	217,990.40	48,525.53	-	-	-	-	-
1994	144,961.14	682,949.76	152,027.32	-	-	-	-	-
1995	687,387.57	3,238,462.34	720,894.55	-	-	-	-	-
1996	756,617.51	3,564,622.70	793,499.13	-	-	-	-	-
1997	790,584.47	3,724,649.90	829,121.82	-	-	-	-	-
1998	810,450.89	3,818,245.78	849,956.64	-	-	-	2,672,289.74	-
1999	816,939.84	3,848,816.94	856,761.90	-	-	-	2,693,685.69	-
2000	853,986.83	4,023,354.96	895,614.75	-	-	-	2,815,840.26	-
2001	880,655.77	4,148,999.30	923,583.68	-	-	3,117,758.45	2,903,775.43	2,543,157.39
2002	917,465.74	4,322,420.69	962,187.97	-	-	3,248,075.66	3,025,148.49	2,649,457.22
2003	984,704.70	4,639,201.01	1,032,704.52	-	-	3,486,119.69	3,246,854.70	2,843,629.87
2004	1,041,349.43	4,906,069.13	1,092,110.42	-	-	3,686,657.28	3,433,628.68	3,007,208.50
2005	1,092,717.36	5,148,076.82	1,145,982.29	-	3,437,109.81	3,868,513.55	3,603,003.49	3,155,548.77
2006	1,134,620.32	5,345,492.63	1,189,927.83	-	3,568,914.33	4,016,861.32	3,741,169.62	3,276,556.13
2007	1,159,150.12	5,461,058.91	1,215,653.35	-	3,646,072.08	4,103,703.41	3,822,051.42	3,347,393.27
2008	1,195,872.73	5,634,068.73	1,254,166.02	-	3,761,581.96	4,233,711.34	3,943,136.44	3,453,440.82
2009	1,248,183.30	5,880,517.47	1,309,026.48	4,066,834.97	3,926,123.29	4,418,904.83	4,115,619.42	3,604,503.25
2010	1,289,384.72	6,074,628.14	1,352,236.29	4,201,077.59	4,055,721.14	4,564,768.96	4,251,472.37	3,723,484.71
2011	1,350,513.10	6,362,619.89	1,416,344.39	4,400,246.28	4,247,998.63	4,781,179.87	4,453,030.23	3,900,011.22
2012	1,424,805.04	6,712,628.61	1,494,257.72	4,642,304.51	4,481,681.69	5,044,193.32	4,697,992.11	4,114,551.45
2013	1,498,886.78	7,061,647.04	1,571,950.61	4,883,677.89	4,714,703.61	5,306,462.63	4,942,260.93	4,328,484.68
2014	1,582,134.79	7,453,850.16	1,659,256.57	5,154,916.84	4,976,557.74	5,601,183.00	5,216,753.58	4,568,888.25
2015	1,683,549.63	7,931,641.96	1,765,614.92	5,485,347.01	5,295,555.09	5,960,218.83	5,551,147.48	4,861,753.99
2016	2,076,995.18	9,785,266.69	2,178,239.13	6,767,272.61	6,533,126.31	7,353,121.97	6,848,450.65	5,997,945.90
2017	2,126,659.89	10,019,250.12	2,230,324.77	6,929,090.34	6,689,345.19	7,528,948.43	7,012,209.50	6,141,367.64
2018	2,189,793.00	10,316,686.70	2,296,535.33	7,134,790.86	6,887,928.51	7,752,456.63	7,220,377.53	6,323,683.41
2019	2,262,353.99	10,658,540.47	2,372,633.33	7,371,209.31	7,116,166.93	8,009,342.06	7,459,631.98	6,533,225.01
2020	2,338,450.44	11,017,050.68	2,452,439.12	7,619,147.00	7,355,526.02	8,278,743.95	7,710,543.85	6,752,976.29
2021	2,507,815.53	11,814,973.86	2,630,059.98	8,170,972.91	7,888,258.86	8,878,341.97	8,268,989.29	7,242,068.74
2022	2,592,666.43	12,214,728.62	2,719,046.98	8,447,434.40	8,155,154.84	9,178,737.00	8,548,767.12	7,487,101.14

Table 6. Practiced hourly cost for each program, based on the annual cost and the total number of hours of courses taken by students in each program (Brazilian Real - R$)

Ano	Mn	Ag	Ac	Cp	Mc	Cv	El	Cm
1986	-	-	26.67	-	-	-	-	-
1987	-	573.58	6.60	-	-	-	-	-
1988	286.19	1189.68	7.93	-	-	-	-	-
1989	30.67	330.46	3.92	-	-	-	-	-
1990	22.18	851.24	0.92	-	-	-	-	-
1991	1.42	1115.05	0.88	-	-	-	-	-
1992	0.64	4.99	0.56	-	-	-	-	-
1993	0.54	4.05	0.50	-	-	-	-	-
1994	1.34	10.12	1.39	-	-	-	-	-
1995	5.35	32.45	5.76	-	-	-	-	-
1996	6.34	28.12	5.77	-	-	-	-	-
1997	7.46	26.08	6.33	-	-	-	-	-
1998	7.68	24.02	6.97	-	-	-	10479.57	-
1999	7.80	22.40	7.45	-	-	-	13813.77	-
2000	7.80	23.14	7.96	-	-	-	14440.21	-
2001	8.65	24.77	8.88	-	-	51962.64	38717.01	10596.49
2002	9.31	26.50	9.68	-	-	18044.86	100838.28	8410.98
2003	10.72	30.90	10.06	-	-	33201.14	72152.33	11151.49
2004	10.12	39.24	10.16	-	-	122888.58	8804.18	11792.97
2005	11.34	31.11	11.15	-	28642.58	16118.81	48040.05	42073.98
2006	10.96	33.47	11.20	-	12522.51	22315.90	11336.88	54609.27
2007	11.27	33.81	10.83	-	141.98	157.14	151.76	109.18
2008	11.58	33.55	11.14	-	68.33	67.25	65.12	51.84
2009	9.42	34.30	10.41	-	37.41	36.02	36.44	31.46
2010	12.54	35.41	11.81	29.96	14.81	13.16	12.36	23.71
2011	12.95	38.62	12.06	29.01	20.16	18.29	20.97	23.63
2012	14.68	40.12	12.63	27.74	20.16	17.61	20.45	26.65
2013	14.61	38.18	13.39	27.14	19.63	17.88	19.95	28.90
2014	16.65	35.26	14.26	28.11	19.67	17.89	18.99	34.78
2015	18.72	34.33	16.05	27.32	19.90	18.06	20.28	33.48
2016	21.69	39.17	18.88	33.61	22.95	20.73	24.25	48.41
2017	20.44	36.22	19.59	34.39	22.16	21.22	23.82	51.04
2018	21.69	34.94	20.53	34.98	21.86	22.71	26.43	50.13
2019	22.87	34.07	22.49	32.36	23.27	23.29	27.87	47.98
2020	26.28	36.98	22.03	35.11	27.20	26.36	32.48	56.61
2021	31.32	42.84	24.36	39.20	34.50	31.28	42.84	71.52
2022	32.07	42.96	26.36	37.73	34.20	34.52	49.23	80.65