BI–based Methodology for Analyzing Higher Education: A Case Study of Dropout Phenomenon in Information Systems Courses

André Menolli*
menolli@uenp.edu.br
Northeast University of Paraná
State University of Londrina
Bandeirantes, Brazil

Flávio Horita
Federal University of ABC
Santo André, Brazil
flavio.horita@ufabc.edu.br

José Jorge L. Dias Jr.
Federal University of Paraíba
João Pessoa, Brazil
jorge@dcx.ufpb.br

Ricardo Coelho
Northeast University of Paraná
Bandeirantes, Brazil
rgcoelho@uenp.edu.br

ABSTRACT

Through a Design Science Research method, this paper presents a BI-based methodology to analyze the Information System (IS) courses, in particular the dropout phenomenon and their enrolled students, using open public data from the Anísio Teixeira National Institute for Educational Studies and Research. Data related to Higher Education in Brazil were analyzed and understood, to create a dimensional model. Based on this, a Data Warehouse was created, and the original data was then extracted, transformed and loaded, and a Business Intelligence (BI) environment was deployed and set. This BI environment was utilized for carrying out several descriptive statistics analyses about the dropout phenomenon. Results showed differences in dropout rates in the dimensions of race, type of city where the course is located, type of institution (public or private), teaching modality (on-site or e-learning), year of entry into course, and course period. Thus, this paper increase the knowledge about IS courses and encourages discussion about dropout phenomenon.

CCS CONCEPTS

• **Information systems** → **Enterprise applications**; • **Applied computing** → *Enterprise information systems*; *Education*; • **Social and professional topics** → User characteristics.

KEYWORDS

Information System Courses, Higher Education, Business Intelligence, Dropout, Data Model, OLAP, ETL process.

ACM Reference Format:
André Menolli, Flávio Horita, José Jorge L. Dias Jr., and Ricardo Coelho. 2020. BI–based Methodology for Analyzing Higher Education: A Case Study of Dropout Phenomenon in Information Systems Courses. In *XVI Brazilian Symposium on Information Systems (SBSI'20), November 3–6, 2020, São Bernardo do Campo, Brazil.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3411564.3411636

SBSI'20, November 3–6, 2020, São Bernardo do Campo, Brazil
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-8873-3/20/11…$15.00
https://doi.org/10.1145/3411564.3411636

1 INTRODUCTION

Dropout in the Higher Education system is a complex and common phenomenon in private and public Higher Education Institutions (HEI). That is why it has been addressed by several works over the past few years, e.g., [12, 20, 22, 23]. The causes of this phenomenon have many routes and consequences, just to mention some, the high expectations on the courses [4] or even a conflict understanding of the area [2].

Interestingly, there has been an increasing demand for IT professionals worldwide and in Brazil that leads to a high demand in the companies. For example, several organizations create the area of "data science" in which data engineers and analysts are responsible for implementing machine learning algorithms to give sense to the existing data, to provide valuable insights, and to support strategic decisions. IT turned into a fundamental area within organizations. But, at the same time, HEI courses are frequently facing problems with unsatisfied students, low grades, and a fast-outdated curriculum, which led to high rates of dropout in computer courses [8, 20]. Understanding the characteristics of such a phenomenon then becomes crucial to provide effective measures to reduce the issue.

Nevertheless, the challenges now are threefold. To start with, the available systems for data sharing are very confusing with not a clear data model and visual interface that would support a proper analysis, and thus decision-making. Secondly, the dropout phenomenon in IS courses has been a critical issue all over the country as it remains increasing in the last couple of years, even though some studies have been investigated the matter. Last but not least, there is a continuous need to take the analysis of such phenomenon to a broader level (e.g., National Level - Brazil), and from that understand the overall impact of public polities and existing efforts.

In this manner, this paper presents a BI-based methodology for examining the Brazilian Information Systems (IS) courses. In particular, we analyze the phenomenon of dropout from different perspectives (e.g., race, sex, and regions). For doing so, we utilize public data from the Anísio Teixeira National Institute for Educational Studies and Research (INEP), as well as rely on Business

Intelligence tools to gather, filter, structure, and analyze the data, such as Mondrian. Hence, the key contributions are the following:

- **Research methodology.** The overall BI - based research methodology can be recognized as an important contribution of this work. Not only as it analyzes the data itself and raises some insights but also because it provides a way of proper structuring the data.
- **Dropout phenomenon in Brazil.** This analysis utilized data from all Brazilian IS courses instead of focusing on a specific university, which made it possible to establish an overall panorama of dropout, as well as to raise interesting and important issues of IS courses as a whole.

This paper is structured as follows. Section 2 reviews the theoretical background and outlines some related works. Section 3 introduces the research method employed in this work, while Section 4 describes the BI-Based Methodology, considering data sources, Extract-Transformation-Load (ETL) process, as well as the employed analytical tools. Section 5 introduces the results obtained in this work. Ultimately, Section 6 also draws some final considerations and provides recommendations for future works.

2 BACKGROUND

This section first outlines the theoretical background of the work, mainly, the overall dropout panorama in undergraduate courses in Brazil. Later, it also details the existing related works.

2.1 Dropout in Undergraduate Courses

Dropout is defined as the abandonment of a certain program, regardless of the motivation of the leaving [15].

Dropout is a complex phenomenon with multifactorial causes such as personal and individual issues, academic and pedagogical aspects, and university management [3]. Besides, other problems encompass social, financial, and other institutional matters. Hence, dropout has being shown a multidimensional phenomenon, and regional aspects must be considered [12, 16, 25].

There are three different types of dropout: 1) course dropout, when student leaves the course in several situations. He/she does not enroll in his/her course or do transfer for another one; 2) Institution dropout, when student leaves the education institution; and 3) Higher education system dropout, when the student leaves, temporarily or definitely, the higher education.

2.2 National Census of Higher Education

The main data source about higher education in Brazil is available at INEP, named the Census of Higher Education.

This census gathers information on higher education institutions, their distance (e-learning) or on-site undergraduate courses, as well as numbers about their enrollment, vacancies offered, and graduated students. Data are collected from completion questionnaires by Higher Education Institutions and by importing data from the electronic system from the Ministry of Education (also known as e-MEC).

During the questionnaire completion period, institutional researchers may make any necessary changes or additions to the data of their respective institutions at any time. After this period, INEP checks the consistency of the information collected. The Census system is then reopened for checking and data validation by HEIs.

2.3 Related Works

The analysis of the dropout phenomenon in Information Systems Courses has been studied several times before in literature. Saraiva et al. [22] carried out a study to investigate the dropout from a social and human perspective. Analyzing data collected through a survey with 54 students that quite their course, they indicated that the two reasons were the overall structure of courses (i.e., degradation or lack of proper infrastructure, or the period of course that makes problematic to combine it with trainee programs) and the lack of interest (i.e., students on their second undergraduate course or those how simply decided to change to other courses). This is in line with the findings of Sihessarenko et al. [24] that also further reasons like the professors' didactics, unachieved expectations about the course, inadequate classroom teaching, and lack of proper guidance by the course coordinator.

Other line of work put a light on the question of how to determine the dropout in IS courses. Damasceno e Carneiro [4] established a methodology that comprises the following elements: a) the process for data structure; b) a dropout index that consider the number of students exist against the number of new arrivals; c) and, ultimately, the analysis from distinct lens of analysis (e.g., the student profile, period of course, etc). Through a case study at a Federal University in Brazil, they were able to estimate the dropout rates, which were very high (e.g., 69% in 2009 and 53% in 2013). With a focus on a Computer Science course instead, Rodrigues et al. [20] examined the dropout phenomenon at Federal University of Rio Grande do Sul. They determined the dropout rate using a formula that takes the number of students exists per year (or semester) against the number of new arrivals of the same year (or semester).

Although these previous works have addressed important and relevant questions, they conducted studies restricted to their universities (e.g., Federal University of Minas Gerais - UFMG). Furthermore, they focused their analysis on either establishing a formula to estimate dropout rates or examining the phenomenon itself from different perspectives. This work instead extends the analysis to a National Level of Brazil, as well as describes a BI-based approach to conduct such analysis, including data sources, Extract-Transformation-Load (ETL) process, and data model.

3 RESEARCH METHOD

This work adopted the Design Science Research (DSR) method as a means to develop the BI-Based Methodology. In particular, DSR method was choose since it focus on "designing and evaluating IT artifacts that can be used to address practice and research problems" [13]. We understand an artifact as "a thing that can be transformed into a material (e.g., model) or process (e.g., method)" [14] and, therefore, our methodology comprises the artifact utilized to investigate the dropout phenomena in IS courses (i.e., problem of practice).

For proper elaborating and evaluating the artifact, we followed the six-steps DSR method proposed by Peffers et al. [19]. First, we identifed the **research problem** that was already explored in Section 1, i.e., a comprehensive and suitable methodology to support in the analyzes of open education data. This is relevant to both practice and literature as it would help universities in establishing more attractive lectures or even more inclusve policies (e.g., for social vulnerable communities). Next, we developed the **IT artifact** (i.e., BI-Based Methodology) to address the identified problem (see Section 4. For doing so, we established a software architecture that utilizes existing analytics (e.g., Mondrian), ETL (e.g., Kettle), storage (e.g., Postgres), and web-based visualization (e.g., OLAP) tools. Having developed the IT artifact, we were ready to move to **evaluation** (see Section 5). Here, we carried out a case study method in which the context was the Brazilian IS courses and the unit of analysis was the dropout rates of these courses [26]. Through the results allowed by the our proposed methodology, we definitely would be able to understand the critical factors about the problem and thus proposed more accurate actions to managers and directors of universities. Ultimately, this paper itself consists the **communication** of obtained results, i.e., the last step of Peffers et al. [19]'s DSR method.

4 BI-BASED METHODOLOGY

In order to build an overview of Information Systems courses in Brazil, we constructed a DW (Data Warehouse), and from this we deployed a BI (Business Intelligence) solution, using Pentaho Business Analytics[1].

The proposed solution has the purpose of enabling a different analysis of the data. To reach this goal, several steps were needed to make data available to the final user. Figure 1 shows the architecture used to implant the BI solution with its steps. The next sections describe each layer of the architecture used.

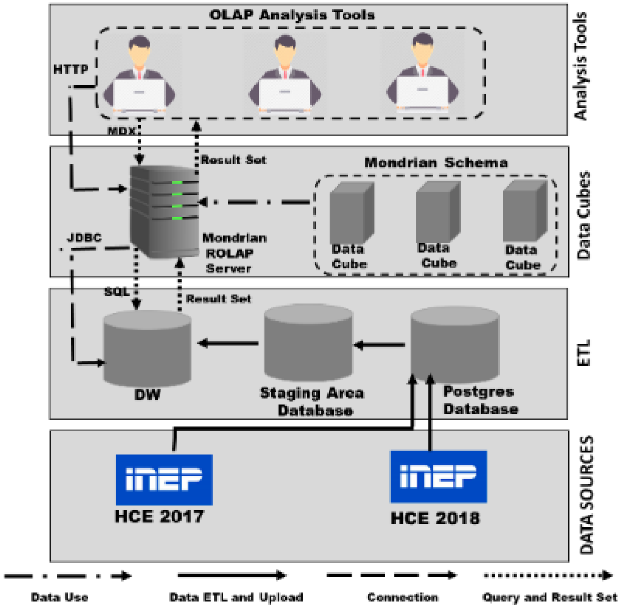


Figure 1: Architecture to Deploy BI Solution

4.1 Data Sources

The first step was to download and understand Brazil's open data from higher education courses available on INEP site. In INEP site, there are available to download different data about higher education and high school courses; for instance, the data about ENADE National Student Performance Exam), ENEM (National High School Exam), HEC, among others.

In this stage of the research, we have been working just with HEI data. The HCE are data compiled provided by all higher education institutions and presented information about all higher education courses in Brazil. The data are submitted every year by the institutions, and then compiled and become available by INEP. There is HDE data from 1995 until 2018. In this stage, we just have been working with the most recent data from the years 2017 and 2018.

4.2 Extract-Transformation-Load (ETL) process

The HCE data are available on CSV (Comma-Separated Values) format. Hence, the first task to start building the DW was to understand the data and import them to a database in a DBMS (DataBase Management System). Each HCE file brings information about each student in Brazil in the Censu's year. Thus, each student table created contain more than 11 million records. To create a faster DW and at the same time and address the research requirement, we delete all data that are not from the computer science area, since in Brazil Information System courses are classified within such area. Furthermore, we deleted information about courses non-higher education courses, such as sequential courses.

After we have had all data in Postgres tables, we migrate them to tables in the staging area. This area is an intermediate step between the extraction of the data source and the load in DW. It is responsible for filtering the data, as well as for conducting transformations and integrations. According to Dumoulin [9], a fundamental concept that greatly simplifies DW projects and facilitates maintenance is the use of data staging areas. Starting from the logical project of the staging area, it is possible to have a good idea of the attributes and source tables necessary to populate the DW. The data staging should just contain the necessary information to populate the warehouse.

The staging area, defined in this work, uses a normalized data model to allow the best data consistency and to facilitate the integration process among different databases. Furthermore, in this step, it was created one table for each table in Postgres Database.

In this step, several data processes and transformations rules were applied. The ETL tool for Data Integration (Kettle) platform was used. The Kettle can be downloaded free of charge through the official website of Hitachi [2], and it is available under Apache 2.0 license. Kettle provides the ETL capabilities that facilitate the process of capturing, cleaning, and storing data. As an example of transformation applied to data, Figure 2 depicts the transformations on the student's data before storing them. Among the rules applied we can mention: (1) transform numeric data to string (for instance 1 -> male; 2 -> female), replace some null fields, concat fields, define number range, among others.

Once the staging area was created and the most data processing and transformation were performed, we created a DW using a multidimensional model.

[1] https://www.hitachivantara.com/en-us/products/data-management-analytics/pentaho-platform/pentaho-business-analytics.html

[2] https://community.hitachivantara.com/

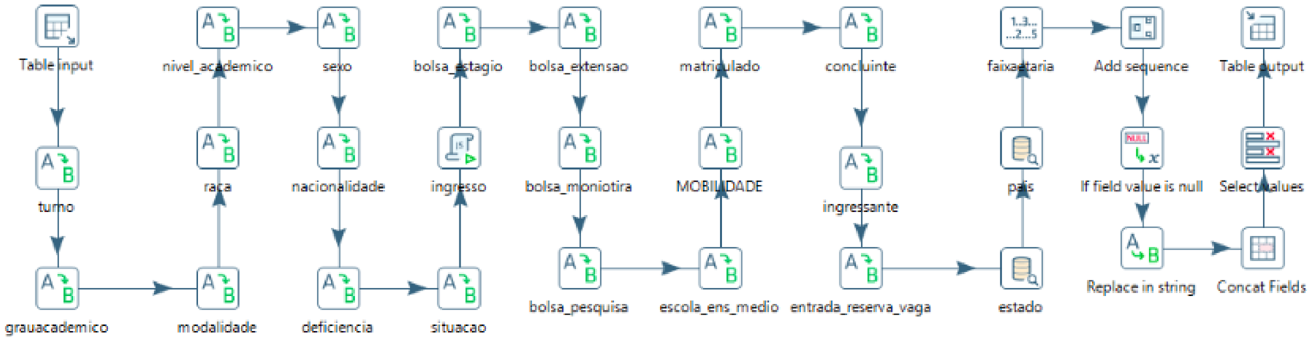


Figure 2: ETL Process

The dimensional model was used because, according to Song et al. [21], there are two main advantages of using a dimensional model in data warehouse environments. First, a dimensional model provides a multidimensional analysis space in relational database environments. Second, a typical and not normalized dimensional model has a simple schema structure, which simplifies the processing of query and improves performance.

In relation to the definition of the DW data model, the first step is the formulation of the data marts that are in a single data source initially. In the second step, the dimensions are identified for these data marts; and the intersection between them and their dimensions are marked. Next, the granularity of the fact tables of these data marts must have to be declared. Also, it is verified which the dimensions are related to these tables. Finally, the facts that compose the fact tables are defined.

Following this implementation model, there is the possibility of superposition of facts in related to replication of dimensions of the data marts. To solve this problem was used a common data area, which enables to share dimensions and facts used by several data marts.

The final physical dimensional model is composed of four fact tables (gray color) and five dimensions (white color), as shown in Figure 3. In the design, we choose to use dimensions unnormalized, with many attributes, but that can be used more than once with different names in the cubes in the Data Cubes layer. This kind of dimension is called the role-playing dimension. For instance, we used the table dim_data to create a dimension called Census Year (Ano do Censo) and other called Year of Admission (Ano de Ingresso).

4.3 Data Cubes

Once the physical DW was done, it was possible to create the Cubes. Our solution has used the Mondrian ROLAP server, which is an open source OLAP (online analytical processing) server, written in Java. The Mondrian receives and parses the MDX language into Structured Query Language (SQL) to retrieve answers to dimensional queries.

The Mondrian Server requires the Mondrian Schema that describes a logical model. It consists of cubes, hierarchies, and members, and a mapping of this model onto a physical model. The first task before creating the Mondrian schema was to define the cubes' granularity and to understand which dimensions were common to the cubes. We used the Bus matrix to define it, and the matrix is shown in Table 1.

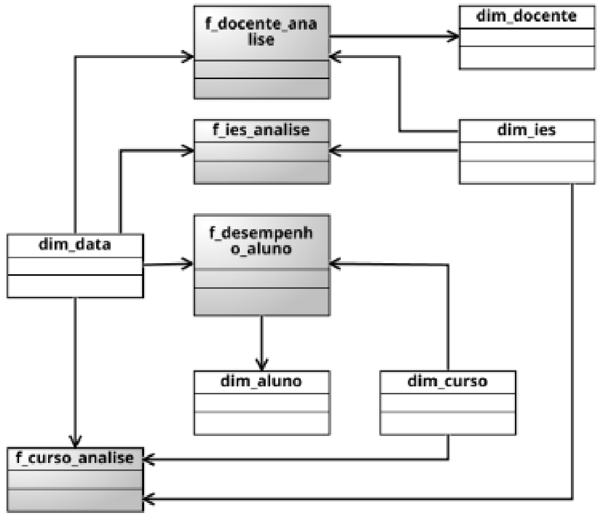
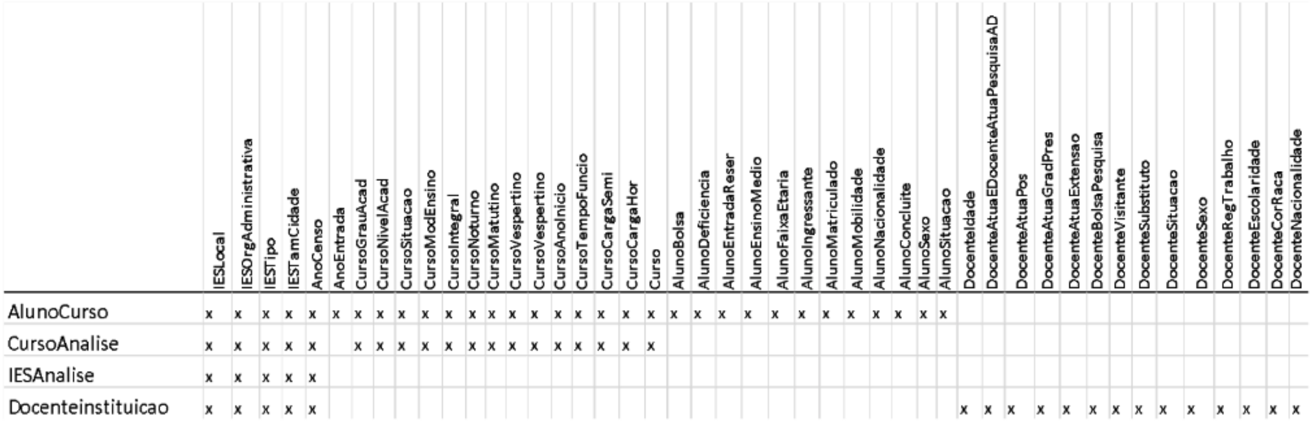


Figure 3: Dimensional Data Model

After that, the cubes, measures, calculated measures, dimensions, hierarchies, members, and levels were defined in order to create the Mondrian Schema, as shown in Figure 4 (it was used the Schema Workbench tool). The Mondrian Schema is composed of four cubes with the following granularity, number of dimensions, and number of measures:

- AlunoCurso - student per course by year, 32 dimensions and 28 measures
- CursoAnalise - course by year, 21 dimensions and 40 measures
- IESAnalise - university by year, 5 dimensions and 57 measures
- DocenteInstituicao - professor by year, 20 dimensions and 49 measures.



	IESLocal	IESOrgAdministrativa	IESTipo	IESTamCidade	AnoCenso	AnoEntrada	CursoGrauAcad	CursoNivelAcad	CursoSituacao	CursoModEnsino	CursoIntegral	CursoNoturno	CursoMatutino	CursoVespertino	CursoVespertino	CursoAnoInicio	CursoTempoFuncio	CursoCargaSemi	CursoCargaHor	Curso
AlunoCurso	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
CursoAnalise	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x
IESAnalise	x	x	x	x	x															
Docenteinstituicao	x	x	x	x	x															

	AlunoBolsa	AlunoDeficiencia	AlunoEntradaReser	AlunoEnsinoMedio	AlunoFaixaEtaria	AlunoIngressante	AlunoMatriculado	AlunoMobilidade	AlunoNacionalidade	AlunoConcluite	AlunoSexo	AlunoSituacao	DocenteIdade	DocenteAtuaEDocenteAtuaPesquisaAD	DocenteAtuaPos	DocenteAtuaGradPres	DocenteAtuaExtensao	DocenteBolsaPesquisa	DocenteVisitante	DocenteSubstituto	DocenteSituacao	DocenteSexo	DocenteRegTrabalho	DocenteEscolaridade	DocenteCorRaca	DocenteNacionalidade
AlunoCurso	x	x	x	x	x	x	x	x	x	x	x	x														
CursoAnalise																										
IESAnalise																										
Docenteinstituicao													x	x	x	x	x	x	x	x	x	x	x	x	x	x

Table 1: Bus Matrix

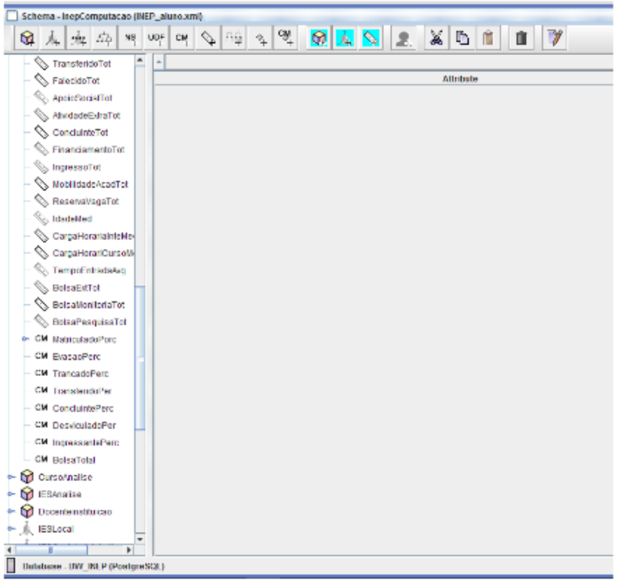
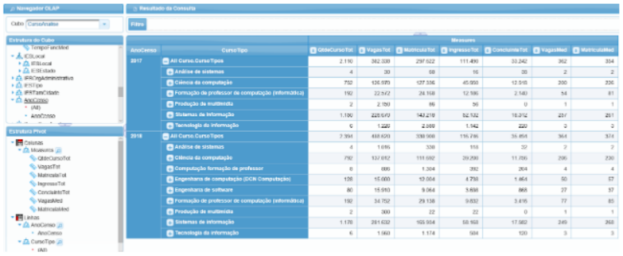


Figure 4: Mondrian Schema



		Measures						
AnoCenso	CursoTipo	QtdeCursoTot	VagasTot	MatriculaTot	IngressoTot	ConcluinteTot	VagasMed	MatriculaMed
2017	All Curso.CursoTipos	2.110	382.338	297.622	111.490	33.242	362	354
	Análise de sistemas	4	30	60	16	30	2	2
	Ciência da computação	752	126.970	127.336	45.050	12.518	200	226
	Formação de professor de computação (informática)	192	22.572	24.168	12.186	2.140	54	81
	Produção de multimídia	2	2.150	86	56	0	1	1
	Sistemas de informação	1.160	228.670	143.218	52.132	18.312	257	261
	Tecnologia da informação	6	1.220	2.588	1.142	220	3	3
2018	All Curso.CursoTipos	2.394	488.620	330.908	116.746	35.494	364	374
	Análise de sistemas	4	1.016	338	118	32	2	2
	Ciência da computação	792	137.012	111.692	39.280	11.756	206	230
	Computação formação de professor	8	888	1.304	392	204	4	4
	Engenharia de computação (DCN Computação)	128	15.000	12.004	4.738	1.464	50	57
	Engenharia de software	80	15.910	9.064	3.698	868	27	37
	Formação de professor de computação (informática)	192	34.752	29.138	9.832	3.416	77	85
	Produção de multimídia	2	300	22	22	0	1	1
	Sistemas de informação	1.178	281.632	165.934	58.168	17.582	249	268
	Tecnologia da informação	6	1.660	1.174	604	120	3	3

Figure 5: Query on Computer Courses Data

4.4 Analysis Tools

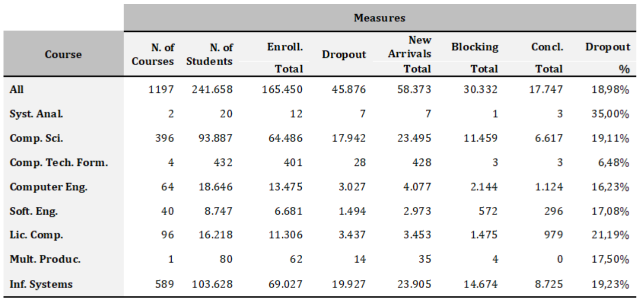
In the Analysis Tools layer, we set the Pentaho Community tools to enable final users to make OLAP queries. The OLAP tools allow analysis and management, providing a great profit of performance, as fast access to a great variety of data views organized through the multidimensional database. For example, Figure 5 displays an analysis performed in the OLAP BI tool deployed.

5 A CASE STUDY OF DROPOUT IN BRAZILIAN INFORMATION SYSTEM COURSES

The BI tool built for this work enables various analyzes. We have analyzed especially the dropout of information systems courses as a way of validating all the elaborated methodology as well as contributing to the characterization of these courses in Brazil. Therefore, all graphs and tables presented below were developed from the data generated by the BI tool.

Table 2 shows the numbers of courses, students, enrollments, entrants, graduated, and dropout group by course type in the Computing area in the year 2018. As can be seen, Information Systems is the most popular course in the area with 589 courses and 103,628 students. Besides, It has the most significant number of enrollments (69,027) and graduates (8,725). Almost half of the courses are IS (49,2%). According to the Computer Higher Education Report [5], by 2014, 604 IS courses were registered in Brazil. However, the highest dropout rate is in the Teaching Degree in Computer Science course (21.19%), while IS course dropout is 19.23%.

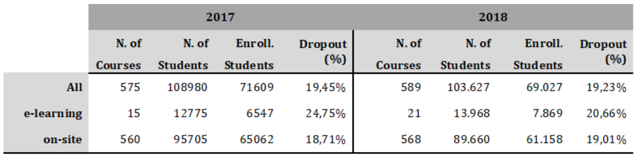
Table 3 presents the specific numbers of the IS courses. As can be seen, the number of both e-learning and on-site courses increased from 2017 to 2018. However, it is noteworthy that the number of students and registered students decreased in on-site mode while increasing in distance mode. These numbers are consonant with



	Measures							
Course	N. of Courses	N. of Students	Enroll.	Dropout	New Arrivals	Blocking	Concl.	Dropout
			Total		Total	Total	Total	%
All	1197	241.658	165.450	45.876	58.373	30.332	17.747	18,98%
Syst. Anal.	2	20	12	7	7	1	3	35,00%
Comp. Sci.	396	93.887	64.486	17.942	23.495	11.459	6.617	19,11%
Comp. Tech. Form.	4	432	401	28	428	3	3	6,48%
Computer Eng.	64	18.646	13.475	3.027	4.077	2.144	1.124	16,23%
Soft. Eng.	40	8.747	6.681	1.494	2.973	572	296	17,08%
Lic. Comp.	96	16.218	11.306	3.437	3.453	1.475	979	21,19%
Mult. Produc.	1	80	62	14	35	4	0	17,50%
Inf. Systems	589	103.628	69.027	19.927	23.905	14.674	8.725	19,23%

Table 2: The numbers of courses, students, enrollments, entrants, graduated, and dropout group by course type in the Computing area in the year 2018.

INEP data showing the growth of distance education in higher education in Brazil [6]. On the other hand, dropout was higher in the distance modality in the two years.



	2017				2018			
	N. of Courses	N. of Students	Enroll. Students	Dropout (%)	N. of Courses	N. of Students	Enroll. Students	Dropout (%)
All	575	108980	71609	19,45%	589	103.627	69.027	19,23%
e-learning	15	12775	6547	24,75%	21	13.968	7.869	20,66%
on-site	560	95705	65062	18,71%	568	89.660	61.158	19,01%

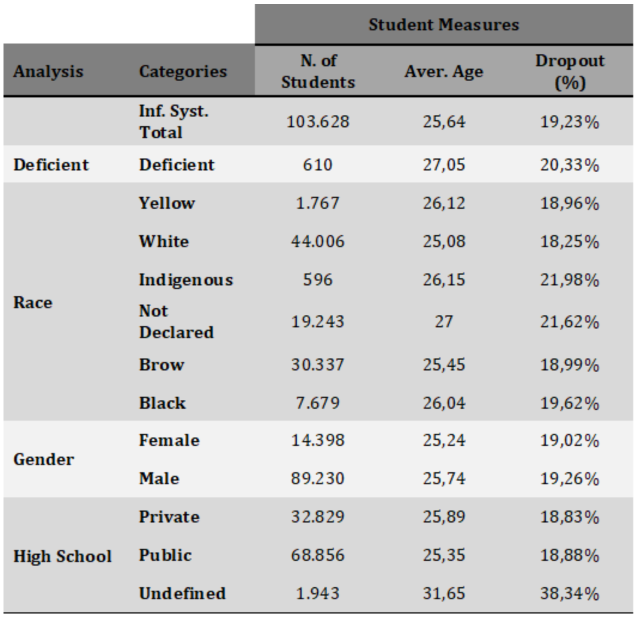
Table 3: The specific numbers of the IS courses

The population of disabled students in IS course corresponds to 0.59%. Regarding race, the IS courses have the majority of white students (42.47%), followed by Pardo (29.27%) and blacks (7.41%). Indigenous people, for example, correspond to 0.58% of students, and Yellows correspond to 1.71%. It is noteworthy two aspects: the first is about the dropout rate of indigenous and black people is higher than that of white students, maybe because they are social vulnerability groups. The second is the difference between the average ages of the black, yellow and indigenous student groups and the white students group. The former being one year older than the latter. Table 4 shows these numbers.

On the gender dimension, IS courses follow the pattern of technology and engineering courses, with a low rate of women (13.9%) [17]. There is no difference about the average age (25.74 and 25.24) and dropout rates (19.02% and 19.26%) in gender.

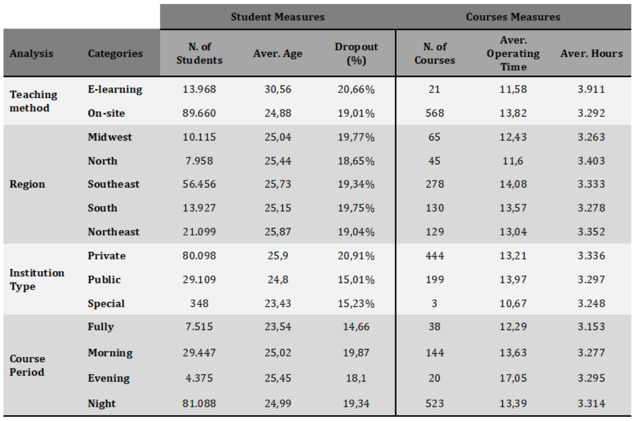
Two-thirds of students came from public high school. These numbers reflect the policy to expand access to lower-income student groups from the public high school system of the last decade [1]. Unlike common sense, public universities mostly receive low-income students. Most students are concentrated in a monthly household income range of one to two minimum wages [7]. Like any public policy, it is relevant to evaluate the concrete results. In this context, the government, society, and academia are responsible for carefully monitoring the variables involved in this process. Moreover, there is no significant difference in dropout rate (18.83% and 18.88%) either considering this dimension.

Most students are located in the Southeast (51,5%), followed by the Northeast (19,3%), South (12,7%), Midwest (9,2%), and North (7,3%). There are no significant differences between the average ages. However, the North has a slightly lower dropout rate than the others.



		Student Measures		
Analysis	Categories	N. of Students	Aver. Age	Dropout (%)
	Inf. Syst. Total	103.628	25,64	19,23%
Deficient	Deficient	610	27,05	20,33%
Race	Yellow	1.767	26,12	18,96%
	White	44.006	25,08	18,25%
	Indigenous	596	26,15	21,98%
	Not Declared	19.243	27	21,62%
	Brow	30.337	25,45	18,99%
	Black	7.679	26,04	19,62%
Gender	Female	14.398	25,24	19,02%
	Male	89.230	25,74	19,26%
High School	Private	32.829	25,89	18,83%
	Public	68.856	25,35	18,88%
	Undefined	1.943	31,65	38,34%

Table 4: Analysis of students per race, gender and high school.



		Student Measures			Courses Measures		
Analysis	Categories	N. of Students	Aver. Age	Dropout (%)	N. of Courses	Aver. Operating Time	Aver. Hours
Teaching method	E-learning	13.968	30,56	20,66%	21	11,58	3.911
	On-site	89.660	24,88	19,01%	568	13,82	3.292
Region	Midwest	10.115	25,04	19,77%	65	12,43	3.263
	North	7.958	25,44	18,65%	45	11,6	3.403
	Southeast	56.456	25,73	19,34%	278	14,08	3.333
	South	13.927	25,15	19,75%	130	13,57	3.278
	Northeast	21.099	25,87	19,04%	129	13,04	3.352
Institution Type	Private	80.098	25,9	20,91%	444	13,21	3.336
	Public	29.109	24,8	15,01%	199	13,97	3.297
	Special	348	23,43	15,23%	3	10,67	3.248
Course Period	Fully	7.515	23,54	14,66	38	12,29	3.153
	Morning	29.447	25,02	19,87	144	13,63	3.277
	Evening	4.375	25,45	18,1	20	17,05	3.295
	Night	81.088	24,99	19,34	523	13,39	3.314

Table 5: Analysis per teaching method, region in Brazil, type of university, and period of lectures.

73% of students are in private higher education institutions. The private sector representative in higher education has continued to expand since 2001 [1, 18]. The difference of the dropout rate between the private (20,91%) and the public (15,01%) institutions stands out. This result is consistent with the evidence found in other analyzes [16, 23]. Despite this difference, dropout factors differ in both contexts [10–12, 25].

Considering the course period dimension, most students are on the night shift (66.23%), followed by the morning (24.05%), fully

(6.14%), and evening (3.57%). It is noteworthy that in the course fully the average age (23.5) and dropout rate (14.66%) is lower than the others. Possibly because they are students, who dedicate themselves exclusively to the course, not competing, for example, with employment.

Figure 6 depicts the rate of dropout behavior over the years that students entered in the course. For purposes of comparison, we have analyzed both the information systems course and the computer science course. The graph shows the dropout rates for 2017 and 2018, as well as the average of these two years. As can be seen, dropout increases substantially in the first year (20.94%), decreasing slightly, and then remains constant up to fifth year. From the fifth year, the dropout again increases. This behavior is almost the same for both courses. However, IS course rates worsen more than CS from the fifth year.

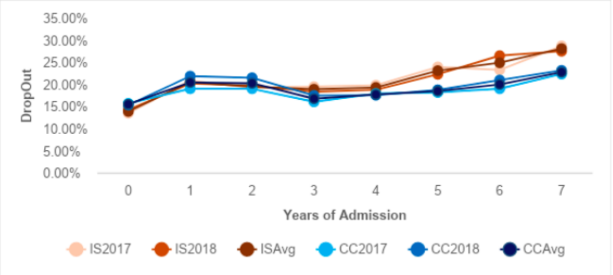


Figure 6: The rate of dropout behavior over the years.

Table 6 presents the dropout rates comparing the city size and type (if the city is state capital or not) in the years of 2017 and 2018. Most courses are located in cities with over one million inhabitants (37.22% in 2017 and 41.26% in 2018). We could see an increase in the number of courses in large cities and a decrease in the amount in smaller ones. Even with growth in the number of courses, the number of students and enrollment declined in almost all groups. In addition, most are located outside the capital of your state (63.65% in 2017 and 62.43% in 2018).

Dropout rates are higher in the groups with the highest concentration of courses. Also, courses outside the capital have a slightly lower dropout rate (18.14% in 2017 and 18.34% in 2018) than those located in the capital (19.34% in 2017 and 19.81% in 2018). Dropout rates may be lower in smaller cities where there is less competition between courses. However, further analysis is recommended to evaluate this hypothesis.

6 CONCLUSIONS

This paper has presented a methodology for analyzing Higher Education courses utilizing concepts and tools from BI. We particularly employed it to examine the dropout phenomenon of IS courses from different perspectives not only to explore the potential of the overall methodology, but also to raise some insights to work on the issue. Results findings showed that the BI-based methodology allowed us to make flexible and interesting analysis, as well as to investigate the positive and negative issues of the courses; for instance, the low number of indigenous people (i.e., 0.58%) and women (i.e.,13.9%), and the differences of dropout with regard to race, type of city, learning modality (on-site and e-learning), type of institution (private or public), year of entry into course, and course period.

Thus, we highlight the following contributions: 1) presentation of a diagnosis about the profile of IS courses in Brazil, which allows us to understand the general overview of these courses; 2) the process reported in the methodology may assist other researchers in the systematization and organization of Higher Education census data in order to perform new analyses.

We point out two limitations in this paper. The first is related to the scope of analysis that considered only the last two years (2017 and 2018) of the census. The second one is associated with the construction of the census database itself, since HEI inform their data to INEP which may cause a possible lack of standard among the concepts used. For instance, each institution may have an interpretation of what to report on dropout.

Future work lines should focus on adapting the data collection process to include further data, e.g., from other courses and previous years. Other data sources should be also considered as they may supplement existing data and expand the horizon of analysis to cover other issues, e.g., courses modules, student grades, or even socio-economic variables. We believe that might have a tremendous impact on data analysis and would also provide interesting insights to academic managers. Ultimately, future studies could analyze how public policies would affect the dropout of undergraduate courses, and thus propose some refinements, as well as directions for improvements.

ACKNOWLEDGMENTS

Thanks for Fundação Arauc'aria (Foundation in Support of the Scientific and Technological Development of the State of Paran'a, Brazil) for support.

REFERENCES

[1] Aparecida da Silva Xavier Barros. 2015. Expansão da educação superior no Brasil: limites e possibilidades. *Educação & Sociedade* 36, 131 (2015), 361–390.

[2] Rita Cristina Galarraga Berardi and Silvia Amelia Bim. 2017. A crise de identidade dos cursos de Sistemas de Informação é percebida "além-muros" das universidades no sul do Brasil? *iSys-Revista Brasileira de Sistemas de Informação* 10, 4 (2017), 24–44.

[3] Francisco José da Costa, Marcelo de Souza Bispo, and Rita de Cássia de Faria Pereira. 2018. Dropout and retention of undergraduate students in management: a study at a Brazilian Federal University. *RAUSP Management Journal* 53, 1 (2018), 74–85.

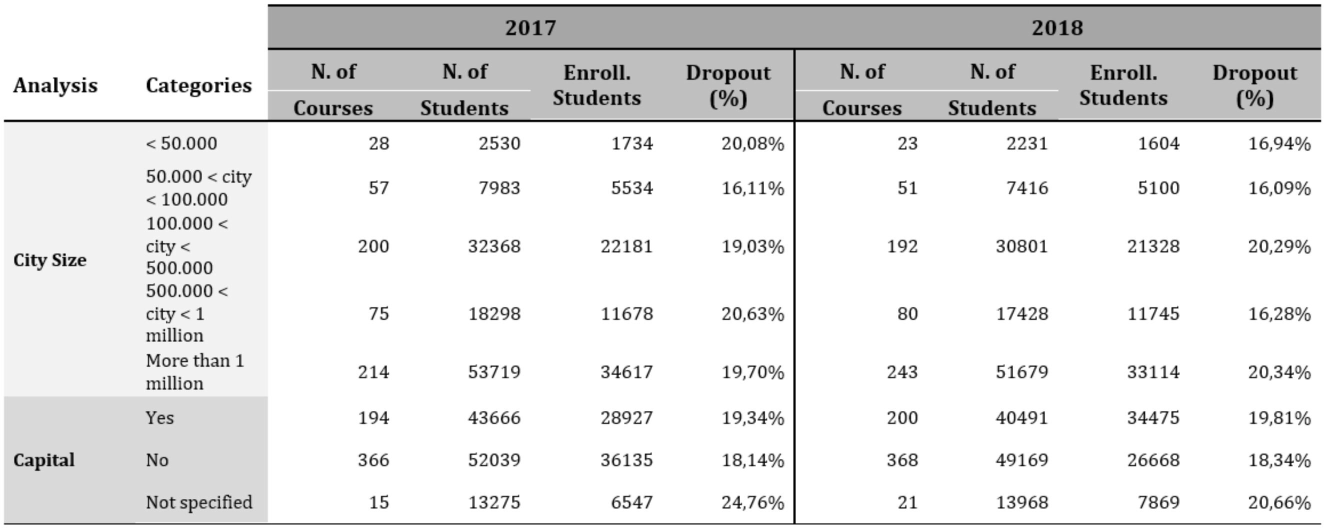
[4] Ieza Damasceno and Murillo Carneiro. 2018. Panorama da Evasão no Curso de Sistemas de Informação da Universidade Federal de Uberlândia: Um Estudo Preliminar. In *Proceedings of the Brazilian Symposium on Computers in Education (SBIE)*. Fortaleza, Brazil, 1766–1770. https://doi.org/10.5753/cbie.sbie.2018.1766

[5] SBC (Sociedade Brasileira de Computação). 2014. Relatório sobre a Educação Superior em Computação - Estatísticas 2014. http://www.sbc.org.br/documentos-da-sbc/summary/133-estatisticas/1007-estatisticas-da-educacao-superior-2014.

[6] INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira). 2019. Censo da Educação Superior 2018 – Notas Estatisticas. http://download.inep.gov.br/educacao_superior/censo_superior/documentos/2019/censo_da_educacao_superior_2018-notas_estatisticas.pdf

[7] FONAPRACE (Fórum Nacional de Pró-reitores de Assuntos Comunitários e Estudantis). 2019. V Pesquisa Nacional de Perfil Socioeconômico e Cultural dos(as) Graduandos(as) das IFES - 2018. http://www.andifes.org.br/wp-content/uploads/2019/05/V-Pesquisa-Nacional-de-Perfil-Socioeconomico-e-Cultural-dos-as-Graduandos-as-das-IFES-2018.pdf

[8] Luciano Antonio Digiampietri, Marcelo de Souza Lauretto, and Fábio Nakano. 2016. Estratégia de Análise Quantitativa para Revisão de Pré-requisitos em



Analysis	Categories	2017				2018			
		N. of Courses	N. of Students	Enroll. Students	Dropout (%)	N. of Courses	N. of Students	Enroll. Students	Dropout (%)
City Size	< 50.000	28	2530	1734	20,08%	23	2231	1604	16,94%
	50.000 < city < 100.000	57	7983	5534	16,11%	51	7416	5100	16,09%
	100.000 < city < 500.000	200	32368	22181	19,03%	192	30801	21328	20,29%
	500.000 < city < 1 million	75	18298	11678	20,63%	80	17428	11745	16,28%
	More than 1 million	214	53719	34617	19,70%	243	51679	33114	20,34%
Capital	Yes	194	43666	28927	19,34%	200	40491	34475	19,81%
	No	366	52039	36135	18,14%	368	49169	26668	18,34%
	Not specified	15	13275	6547	24,76%	21	13968	7869	20,66%

Table 6: Analysis of number of students, courses, enrolled students and dropout per type of city

uma Matriz Curricular do Curso de Bacharelado em Sistemas de Informação. *iSys-Revista Brasileira de Sistemas de Informação* 9, 2 (2016).

[9] Rob DuMoulin. 2005. Architecting Data Warehouses for Flexibility, Maintainability, and Performance.

[10] Vera Lucia Felicetti and Paulo Fossatti. 2014. Alunos ProUni e não ProUni nos cursos de licenciatura: evasão em foco. *Educar em Revista* 51 (2014), 265–282.

[11] Jocimar Fernandes, Ailton da Silva Ferreira, Denise Cristina de Oliveira Nascimento, Eduardo Shimoda, and Giovany Frossard Teixeira. 2010. Identificação de fatores que influenciam na evasão em um curso superior de ensino à distância. *PerspectivasOnLine 2007-2011* 4, 16 (2010).

[12] Rosangela Fritsch, Cleonice Silveira da Rocha, and Ricardo Ferreira Vitelli. 2015. A evasão nos cursos de graduação em uma instituição de ensino superior privada. *Revista Educação em Questão* 52, 38 (2015), 81–108.

[13] Shirley Gregor and Alan R. Hevner. 2013. Positioning and presenting Design Science Research for maximum impact. *MIS Quarterly* 37, 2 (2013), 337–356.

[14] Alan Hevner, Salvatore March, Jinsoo Park, and Sudha Ram. 2004. Design science in information systems research. *MIS quarterly* 28, 1 (2004), 75–105.

[15] Gérard Lassibille and Lucía Navarro Gómez. 2008. Why do higher education students drop out? Evidence from Spain. *Education Economics* 16, 1 (2008), 89–105.

[16] MBCM Lobo. 2012. Panorama da evasão no ensino superior brasileiro: aspectos gerais das causas e soluções. *Associação Brasileira de Mantenedoras de Ensino Superior. Cadernos* 25 (2012).

[17] Carolina Santana Louzada, Wesckley Faria Gomes, MASN Nunes, Edilayne Meneses Salgueiro, Beatriz Trinchão Andrade, and PS Lima. 2014. Um mapeamento das publicações sobre o ingresso das mulheres na computação. In *CLEI 2014: Conferência Latino-americana em Informática-VI Congresso da Mulher Latino-americana na Computação. Montevidéu.*

[18] Deise Mancebo, ANDRÉA ARAUJO DO VALE, and TÂNIA BARBOSA MARTINS. 2015. Políticas de expansão da educação superior no Brasil 1995-2010. *Revista Brasileira de Educação* 20, 60 (2015), 31–50.

[19] Ken Peffers, Tuure Tuunanen, Marcus A Rothenberger, and Samir Chatterjee. 2007. A design science research methodology for information systems research. *Journal of management information systems* 24, 3 (2007), 45–77.

[20] Francisco Scheffel Rodrigues, Christian Puhlmann Brackmann, and Dante Augusto Couto Barone. 2015. Estudo sobre a evasão no curso de ciência da computação da ufrgs. *Revista Brasileira de Informática na Educação* 23, 1 (2015), 1–6.

[21] William Rowen, Il-Yeol Song, Carl Medsker, and Edward Ewen. 2001. An analysis of many-to-many relationships between fact and dimension tables in dimensional modeling. In *Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW 2001)*. Interlaken, Switzerland, 1–13.

[22] Juliana Saraiva, Vanessa Dantas, and Amanda Rodrigues. 2019. Compreendendo a Evasão em uma Década no Curso Sistemas de Informação à luz de fatores humanos e sociais. In *Proceedings of the IV Workshop sobre Aspectos Sociais, Humanos e Econômicos de Software (WASHES)*. Belém, Brazil, 21–30.

[23] Roberto Leal Lobo Silva Filho, Paulo Roberto Motejunas, Oscar Hipólito, and Maria Beatriz Carvalho Melo Lobo. 2007. A evasão no ensino superior brasileiro. *Cadernos de pesquisa* 37, 132 (2007), 641–659.

[24] Michelli Slhessarenko, Claudio Reis Gonçalo, Joana Carlos Beira, and Priscila Cembranel. 2014. A evasão na educação superior para o curso de bacharelado em sistema de informação. *Revista Gestão Universitária na América Latina-GUAL* 7, 1 (2014), 128–147.

[25] Maria Izabel de Quadros Vivas. 2011. Evasão na Educação Superior: uma aproximação com o fenômeno na universidade pública. (2011).

[26] R. K. C Yin. 2002. *Case study research: design and method.* S. P. Inc.