International Journal on Interactive Design and Manufacturing (IJIDeM)

A Data Analytics Approach to Contrast the Performance of Teaching (only) vs. Research Professors

--Manuscript Draft--

Manuscript Number:	IJDM-D-19-00277R1
Full Title:	A Data Analytics Approach to Contrast the Performance of Teaching (only) vs. Research Professors
Article Type:	Technical Paper
Corresponding Author:	Francisco Javier Cantu-Ortiz, Ph.D. Tecnologico de Monterrey Monterrey, Nuevo Leon MEXICO
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Tecnologico de Monterrey
Corresponding Author's Secondary Institution:	
First Author:	Mario D. Chavez, MSc.
First Author Secondary Information:	
Order of Authors:	Mario D. Chavez, MSc.
	Héctor G. Ceballos, Ph.D.
	Francisco Javier Cantu-Ortiz, Ph.D.
Order of Authors Secondary Information:	
Funding Information:	Instituto Tecnológico y de Estudios Superiores de Monterrey (GIEE-ISG-2014) — Not applicable Consejo Nacional de Ciencia y Tecnología (SNI_9804) — Dr. Francisco Javier Cantu-Ortiz
Abstract:	This research article presents a study to compare the teaching performance of teaching-only versus teaching-and-research professors at higher education institutions. It is a common belief that, generally, teaching professors outperform research professors in teaching-and-research universities according to student perceptions reflected in student surveys. This case study presents experimental evidence that shows this is not always the case and that, under certain circumstances, it can be the contrary. The case study is from Tecnologico de Monterrey (Tec), a teaching-and-research, private university in Mexico that has developed a research profile during the last two decades using a mix of teaching-only and teaching-and-research faculty members; during this time period, the university has had a growing ascendancy in world university rankings. Data from an institutional student survey called the ECOA was used. The data set contains more than 118,000 graduate and undergraduate courses for 5 semesters (January 2017 to May 2019). The results presented were derived from statistical and data mining methods, including Analysis of Variance and Logistic Regression, that were applied to this data set of more than nine thousand professors who taught those courses. The results show that teaching-and-research professors perform better or at least the same as teaching-only professors. The differences found in teaching with respect to attributes like professors' gender, age, and research level are also presented.
Response to Reviewers:	We have carefully revised and attended the reviewers' comments. The only and most important one was changing the type of paper from "Original" to "Technical", which has been already done.

A Data Analytics Approach to Contrast the Performance of Teaching (only) vs. Research Professors

Héctor G. Ceballos, Mario D. Chavez, Francisco J. Cantu-Ortiz*

Tecnologico de Monterrey, School of Engineering and Science, Ave. E Garza Sada 2501, Monterrey, N.L., 64989, Mexico,

ceballos@tec.mx, A00826797@itesm.mx, fcantu@tec.mx

* Corresponding author: Francisco J. Cantu-Ortiz, fcantu@tec.mx, +52 81 10508294

Abstract

This research article presents a study to compare the teaching performance of teaching-only versus teaching-and-research professors at higher education institutions. It is a common belief that, generally, teaching professors outperform research professors in teaching-and-research universities according to student perceptions reflected in student surveys. This case study presents experimental evidence that shows this is not always the case and that, under certain circumstances, it can be the contrary. The case study is from Tecnologico de Monterrey (Tec), a teaching-and-research, private university in Mexico that has developed a research profile during the last two decades using a mix of teaching-only and teaching-and-research faculty members; during this time period, the university has had a growing ascendancy in world university rankings. Data from an institutional student survey called the ECOA was used. The data set contains more than 118,000 graduate and undergraduate courses for 5 semesters (January 2017 to May 2019). The results presented were derived from statistical and data mining methods, including Analysis of Variance and Logistic Regression, that were applied to this data set of more than nine thousand professors who taught those courses. The results show that teaching-and-research professors perform better or at least the same as teaching-only professors. The differences found in teaching with respect to attributes like professors' gender, age, and research level are also presented.

Keywords: Student Evaluation of Teaching (SET), Teaching Professor, Research Professor, Data Science, ANOVA and Logistic Regression, Innovation in Higher Education

7. Acknowledgments

We would like to thank the Vice-rectory for Research and Technology Transfer and the Academic Vice-rectory of Tecnologico de Monterrey for providing us with the ECOA student survey data to perform this analysis, and the Intelligent Systems Group for its support in conducting this research. Also, the authors would like to acknowledge the financial and technical support of Writing Lab, TecLabs, Tecnologico de Monterrey, Mexico, in the production of this work.

A Data Analytics Approach to Contrast the Performance of Teaching (only) vs. Research Professors

Abstract

This research article presents a study to compare the teaching performance of teaching-only versus teaching-and-research professors at higher education institutions. It is a common belief that, generally, teaching professors outperform research professors in teaching-and-research universities according to student perceptions reflected in student surveys. This case study presents experimental evidence that shows this is not always the case and that, under certain circumstances, it can be the contrary. The case study is from Tecnologico de Monterrey (Tec), a teaching-and-research, private university in Mexico that has developed a research profile during the last two decades using a mix of teaching-only and teaching-and-research faculty members; during this time period, the university has had a growing ascendancy in world university rankings. Data from an institutional student survey called the ECOA was used. The data set contains more than 118,000 graduate and undergraduate courses for 5 semesters (January 2017 to May 2019). The results presented were derived from statistical and data mining methods, including Analysis of Variance and Logistic Regression, that were applied to this data set of more than nine thousand professors who taught those courses. The results show that teaching-and-research professors perform better or at least the same as teaching-only professors. The differences found in teaching with respect to attributes like professors' gender, age, and research level are also presented.

Keywords: Student Evaluation of Teaching (SET), Teaching Professor, Research Professor, Data Science, ANOVA and Logistic Regression, Innovation in Higher Education

1. Introduction

Assessment of teaching and methods of evaluation are active fields of research and present opportunities for innovation, spanning from elementary grades to higher education levels [1]. The phenomenon of digital transformation has brought new technologies that are revolutionizing modern organizations, including higher education institutions, where teaching and learning are experiencing novel pedagogical methods that require new ways to assess the quality of education from the viewpoints of administrators, teachers, and students [2]. Digital technologies and Artificial Intelligence have made it possible to combine traditional physical presence with online, virtual tools to deliver teaching as a worldwide phenomenon. Emerging technologies such as the

Internet of Things (IoT), Cloud Computing, Big Data, Machine Learning, and Data Analytics, Speech Recognition, Natural Language Understanding, Intelligent Tutoring Systems, and Virtual Reality, to name just a few, are transforming educational systems generally [3].

For instance, the emergence of MOOCs (Massive Online Open Courses) in 2006 as a disruptive innovator in distance learning is an example of technologies transforming higher education. MOOCs can offer unlimited student participation and open access to course content through the worldwide web. The MOOC approach challenges existing educational business models by selling a variety of services for teaching, training, certification, and competencies and skills demanded by employers [4]. Nonetheless, MOOCs face the same difficulties of many institutions, which is frequently referred to as "we are teaching 21st-century students with 20th-century professors and 19th-century pedagogy." This issue arises as many MOOC providers deliver traditional ways of teaching via lectures using digital technologies, mobile devices, and Artificial Intelligence methods [5, 6]. The main MOOC providers include Coursera, Udacity, Udemy, and Edx. For instance, by 2013, Coursera enrolled more than 5 million students, while Edx reported more than one million enrollments [7]. Even though MOOCs show a high potential to revolutionize massive education, they face a major problem having to do with attrition rates and course dropouts. Despite the large numbers of enrollments, only a very small percentage of students enrolled complete their studies. According to the study conducted by Katy Jordan, the average completion rate for MOOCs is approximately 15 percent [8]. Coursera reports completion rates of 7 to 9 percent [9], while Coffrin et al. report percentages that are even lower and lie between 3 and 5 percent [10]. Thus, we ask about the reasons for learner attrition and dropouts from MOOC programs and equate these numbers to student satisfaction and assessment, not just of teachers and instructors, but of the entire teaching and learning system of the MOOCs. Student profiles and motivation for learning are typically quite different from that of traditional university students, and this should be taken into careful consideration when evaluating MOOCs [11]. Although the profiles and the motivational dimensions of MOOC students are more complex and heterogeneous than those of traditional students, the lessons learned in teacher assessments were taken into account as well as the technologies provided by MOOC service providers in the evaluation of traditional teachers [12].

In addition to technology and platform issues, one component of the overall teaching evaluation is the assessment of the teacher's role, which is fundamental in the overall schooling system. Going beyond the processes of conventional teacher evaluation, Salazar and Lerner provide a framework for teacher evaluation that better prepares educators to serve culturally-and-linguistically diverse (CLD) learners. They address issues of theory, research, and practice and showcase a model concerned with issues of equity and excellence in evaluation. They present a five-tenet model to place the needs of CLD learners at the center, offering specific ways to assess and promote cultural responsiveness and confront unfairness and inequity; their model provides critical insight into the evaluation of the teacher role. It re-conceptualizes teacher evaluations in order to support CLD learners and their communities better while developing cultural competency and critical consciousness among all learners [13].

Another aspect of teacher evaluation is addressed by Kim Marshall, who shows how to break away from the typical, often ineffective evaluation approaches in which administrators use infrequent classroom visits or rely on standardized test scores to assess a teacher's performance. Marshall proposes a broader framework for supervision and evaluation that enlists teachers in improving the performance of all students. It includes thoughts on iPad and iPhone apps for classroom observations and a new chart on how principals can manage teacher evaluations. Also, it contains new thoughts on merit pay, a different approach to the test-score argument from Arne Duncan, and extensive tools and advice for time management, supervision, and evaluation practices that foster the professional development of teachers [14].

Aside from cultural and technological concerns, there exist several methods to conduct teacher evaluations that include the development of pedagogical materials, multimedia resources, scoring techniques, book publications, and learning platforms, among others. Nonetheless, one of the schemes more frequently used to evaluate teachers is to take into account the opinion of students about the quality of the service they receive, and this is commonly done by asking students to complete surveys of student satisfaction [15].

The research question posed in this study is whether or not teaching-only professors perform better than teaching-and-research professors according to student opinion in teaching-and-research institutions. The study contributes to the discussion of this issue by making a detailed analysis of data collected from a teaching evaluation survey called the ECOA in Spanish (student opinion survey) that is answered by students at a private teaching-and-research university in Mexico. The teacher assessment that analyzes the degree of student satisfaction through means of a survey has a long tradition at *Tecnologico de Monterrey;* it has been used as a measurement tool since the middle '70s to the present, evolving over the years. We believe that the results of this analysis could be relevant to academic, teaching-and-research institutions and could support the development of efficient research and teaching strategies in those institutions.

To address these issues in this article, we organized the sections as follows: In section 2, we do an overview of teacher evaluation in higher education and teaching-and-research institutions. In section 3, we describe the methodology and data employed in conducting this study. Section 4 describes the results of the experiments we performed, section 5 discusses the outcomes obtained, and section 6 summarized the conclusions of our study.

2. Teacher Evaluation in Teaching-and-research Universities

Student evaluation of teaching (SET) at higher education institutions spans undergraduate and graduate academic levels and has been an active field of research for the last few decades. The evaluation of teachers by students and the consequences that those evaluations have on teacher wages, faculty careers, and promotions, have been widely criticized and keep being a vigorous area of study. However, the student evaluation of teachers through surveys and questionnaires arguably provides an objective and measurable means of hearing the voice of the learners, who are the recipients of the teaching.

Tsinidu et al. present a study in which they identify the quality determinants for educational services provided by higher education institutions in Greece and measure their relative importance from the students' points of view. They describe a multi-criteria, decision-making methodology that was used for assessing the relative importance of quality determinants affecting student satisfaction. They use the *Analytical Hierarchical Process* (*AHP*) to measure the relative weight of each quality factor that contributes to the quality of educational services as it is perceived by students. Their study can be used to quantify internal quality assessments in higher education institutions [16].

Steve Stack analyzes the relationship between research productivity and student evaluations of teaching and reports that it has several shortcomings. He argues that research typically fails to check and adjust for nonlinear distributions in research productivity and that approximately 15% of researchers account for most articles and citations. Then, he highlights that the focus of analysis is typically the teacher and not the class, and those top researchers might disproportionately teach small classes at the graduate level, and student evaluations usually score the teacher higher in such classes. His study aims to correct those issues using data from 167 classes in the social sciences and from 65 faculty. He finds that the quality of research productivity measured by numbers of citations per year is not related to the student evaluation of teaching. And he finds that when the distribution of citations is corrected for skewness, a significant positive relationship between research productivity and student evaluation of teachers emerges. He concludes that this is the first systematic investigation to demonstrate a significant relationship between the quality of research (measured by numbers of citations) and student evaluation of teachers [17].

On the other hand, Spooren et al. present an overview of the state-of-the-art of student evaluation of teaching in higher education. Their study is based upon research reports published in peer-reviewed journals since 2000. They explore the traditional topics such as the dimensionality debate, the 'bias' question, and the questionnaire design; also, some recent research trends in student teaching evaluation, such as online, and some other biases, including the professors' characters and personalities, thus giving researchers some suggestions to formulate future research. Spooren et al. argue that teacher evaluation through student surveys remains a topical but delicate issue in higher education and in education research. They add that many stakeholders are not convinced of the usefulness and validity of student evaluation of teachers for both formative and summative purposes. They conclude that research on student evaluation of teaching has thus far failed to provide clear answers to several critical questions concerning the validity of teacher assessments [18].

The studies above present state-of-the-art in teacher evaluation in higher education institutions. This overview allows us to focus on universities of interest. These institutions are the teaching-and-research universities where teaching is typically done by professors who follow either a teaching-only track or a teaching-and-research path. Those universities are frequently included in world university rankings that display tables of top-1000 universities in the world like Shanghai, QS, Times Higher Education, U-Multirank, or US News and World Report. World university rankings have been in the landscape of higher education for the last two decades and have become a relevant factor in measuring the quality and performance of universities as well as public perception and reputation worldwide [19]. Some rankings focus on research-intensive universities, which is the case of the Shanghai ranking (Academic Ranking of World Universities), which takes into account Nobel prize awardees as well as publications in the journals *Nature* and *Science*. The *Times Higher Education World University Ranking* (THE WUR) also favors universities with a high research profile, and *US News and World Report's* Best Global University Ranking also is biased towards research-intensive universities. On the other hand, the QS World University Ranking (QS WUR) calculates university scores based on a more balanced combination of teaching and research indicators that include academic reputation, employer reputation, citation per faculty, students per faculty, and international students and professors; this is more suitable for teaching-and-research universities without excluding research-intensive institutions. Therefore, our research study focuses on higher education institutions ranked by QS WUR.

Student education in research-intensive universities, which are typically ranked in the top-100 of the QS WUR and other world university rankings, is mostly done by professors who educate their students following a research-based teaching approach. However, universities that are teaching-and-research and not only research-intensive found in the QS rankings between 101-1000 combine both teaching-only and teaching-and-research professors in their delivery of teaching [20].

Research output and its impact are among the most important metrics in world university rankings, and universities spend a significant amount of resources on funding research with internal and external resources in order to increase research output and the impact of their faculty members and research students. Thus, universities aim at balancing the proportion of teaching-only and research professors to be listed in the ranking tables and, at the same time, fulfill the mission of educating and preparing students for professional life. As a consequence, on top of their research activities, most research professors frequently spend at least half their time lecturing as teachers at graduate and undergraduate academic levels.

Nevertheless, the academic education and activities of a research professor may be different from the ones done by a professor who is fully dedicated to teaching students. For many years, there has been a debate about the role and importance of research activity with respect to teaching performance. In many cases, there is a general belief that at least in teaching-oriented universities, teaching-only professors outperform teaching-and-research professors based on student opinion, but in this study, evidence that this is not necessarily the case is provided. Additionally, research has not been a characteristic that students evaluate when they assess professor performance through opinion surveys at the end of an academic period. However, this theme is a subject of enduring debate.

3. Methodology and Data

The method employed to conduct this study is a data analytics methodology frequently used in data science projects. The methodology is called the Cross-Industry Standard Process for Data Mining, known as CRISP-DM, a common methodology used in data analytics and data mining studies [21]. CRISP-DM was conceived in 1996 and became a European Union project under the ESPRIT funding initiative in 1997 under the leadership of several companies that included Integral Solutions Ltd, Teradata, Daimler AG, NCR Corporation, and OHRA. The first version of the methodology was presented at the 4th CRISP-DM SIG Workshop in Brussels in March 1999 and was published as a step-by-step data mining guide later that year [22]. While many non-IBM data mining practitioners use CRISP-DM, IBM is the primary corporation that currently uses the CRISP-DM process model, and it has incorporated it into its SPSS (Statistical Package for the Social Sciences) modeler product [23]. CRISP-DM comprises six stages performed in cycles, namely, Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment [24].

In the following subsections, how this procedure was applied to the dataset in order to answer the research question of this study is explained.

3.1 Business Understanding.

Comprehensive and profound business understanding is crucial to have a successful project and deliver useful results. Business understanding is about comprehending and deciphering the problem domain in which the study is conducted. In this case, the evaluation of teaching performance in teaching-and-research universities typically ranked in the 101-1000 rank of world university rankings is the problem and the domain in which we do our analysis. More specifically, we present Tecnologico de Monterrey, a university ranked in position 158 of the QS World University Ranking 2020, as our case study in which we conducted the teacher performance analysis [25]. Nonetheless, the issues raised in the previous section regarding teaching evaluations arise to the surface when we get into the context and cultural aspects of particular institutions. Two of the authors have more than sixty years of combined experience in teaching both undergraduate and graduate courses and in managing courses and assessing their teachers. One of the authors has more than five-years' experience evaluating teachers as a student. Thus, we believe that there is a fairly good understanding of the problem domain and the relevance of the problem we are trying to solve [26].

3.2 Data Understanding.

Data understanding derives from a good comprehension of the business problem domain. For the case study, the data is taken from the ECOA student satisfaction survey administered by the Registry office at Tecnologico de Monterrey. The ECOA is responded to by students at the end of each semester in the 26 university campuses that comprise Tecnologico de Monterrey. The purpose is to get feedback about the professors' teaching performances in various academic periods of semesters and quarters. Student satisfaction is measured using three questions on the ECOA survey with a 1 - 10 scale (where 10 is the best and 1 is the worst score). The survey questions cover three areas: (1) Intellectual challenge, (2) Learning guide, and (3) Recommended Teacher. A record of the survey contains the average score and the number of students who answered each of those questions about their satisfaction for a given teacher in a given class. This score is the most relevant measure we have for teaching quality from the students' perspective.

Two data sets in our analysis were used. The first dataset is comprised of responses from the semester of August - December 2017 and contains 15,781 records. This dataset consists of 60 features that describe the professors' teaching and research activities and the attributes of the class. Professor attributes include Nationality, Gender, Age, Campus, School, Department, Maximum Professional Degree, Teaching Certificate, Semesters of Experience, membership on the National Researcher System (SNI), Total Teaching Hours, and the number of papers published in Scopus in the last 5 years (Conference Proceedings, Journal Articles, and Total publications). Class features describe the attributes of the class, such as the number of credits, number of students, and the level of students (undergraduate or graduate).

The second dataset contains responses collected during 5 semesters, from May 2017 to May 2019, with 118,818 records of about 9,469 professors, of whom 626 are considered as researchers, given that they were classified as such in the National Research System (SNI) of Mexico during the period 2016-2019. In this second dataset, the maximum proficiency level of researchers is included on a scale 1-4, where 4 is the upper proficiency level.

3.3 Data Preparation.

The two datasets with responses recorded on the ECOA survey were prepared as follows:

First, we omitted variables that had the same value in all the records and features that had a significant amount of null values. For example, in the first dataset, the variable, "age," had 5,284 null values, which is one-third of the data. There is no way to fill those values in a good manner, so we decided to drop them. Other features have a high amount of null values, but some of them actually represent a 0, such as the number of published documents by a professor. Next, the survey responses were aggregated by the professor, and four main variables were created for the first dataset:

- *sni_yn:* We defined a binary variable that is equal to 1 when the professor is a researcher and 0 when he/she is not. A professor was considered a researcher if he/she was a member of the National Researcher System (SNI) during the period 2016-2019.
- *score:* We calculated the average of the three student satisfaction scores of a course and decided to use it as the target variable.
- *weighted score:* To make an analysis of the individuals and not of courses, we summarized survey results by professor, so we took the arithmetic mean to summarize the scores that a professor gets in all of his/her courses, weighted by the number of students per group.
- *score category:* The binary variable takes the value 1 if it is above the median and 0 otherwise. This is the classification feature that we use in the logistic regression model. We say that a professor above the median is good and bad if below it.

In the second dataset, responses were also aggregated at the professor level. We defined another four variables, one for distinguishing the researcher proficiency and three others to break down the weighted score into each one of the questions that comprise student satisfaction.

- *SNI:* We classified researchers according to the maximum proficiency level they reached in the period 2016 - 2019. This variable is in the scale 1 - 4, where 1 represents the lower proficiency level, and 4 represents the upper one.
- 05. RET: It is the professor weighted average score for the Intellectual Challenge question, on the scale 1 - 10, where 10 is the maximum score.
- 06. APR: It is the professor weighted average score for the Learning Guide question, on the scale 1 - 10, where 10 is the maximum score.
- 08. REC: It is the professor weighted average score for the Recommended Teacher question, on the scale 1 - 10, where 10 is the maximum score.

3.4 Modeling

The next step of the CRISP-DM methodology was the modeling task with the purpose of finding interesting patterns and hidden knowledge that may shed light on the answer to the research question of the study. In order to do this task, we used statistical methods and machine learning algorithms to identify 1) significant differences between researchers and professor evaluations, and 2) the more meaningful variables that predict the assessment of a professor. As a working (null) hypothesis, we assumed that teaching-only professors are better evaluated by students compared to teaching-and-research professors (Ho). The alternate hypothesis is that teaching-only professors are evaluated either worse or the same by students than teaching-and-research professors (Ha). We prepared data and applied data mining methods and tools in order to test our hypotheses.

A careful design of experiments was conducted that were carried out using various modeling methods. We applied statistics and machine-learning techniques to the dataset to obtain results about the performance of teaching-only and teaching-and-research professors considering various attributes of both types of professors. The main results are reported in the next section in which we confirm the alternate hypothesis (Ho). Although we used several techniques for identifying the attributes of groups on which researchers are best evaluated, we only reported a kind of Linear Multiple Regression since it yielded the most accurate technique. We also used statistical techniques like mean average calculations and analysis of variance to determine the significance of various mean comparisons. We also used the hold-out method and a simple form of cross-validation in order to train a model to test it with the test data that was set apart.

Regarding tools, there are different data mining packages, both open-source and proprietary, that have proved helpful in conducting modeling with datasets; these include Rapid Miner, Weka, Jupyter, SPSS, among others. In order to satisfy the constraints of the problem, we decided to mine data using Python's programming language on Jupyter Notebooks. We used the Pandas package to manage data as data frames and the Scikit-learn package to test classification models and to pre-process the data.

3.5 Evaluation

A common approach employed in data mining studies is the splitting of the data set into a training dataset and a testing dataset. The model is trained using the training set and one of the various modeling techniques. Once the model has been trained, we test the accuracy of the resulting model, making predictions over the testing dataset, and since we know the real answers, we are able to compare model predictions with the real answers in order to evaluate the accuracy of the trained model. We followed this approach to evaluate the models obtained by applying logistic regression and analysis of variance to validate the statistical significance in comparing various attributes of teaching-only versus teaching-and-research professors. In order to avoid biases and to make fair comparisons, we deleted the theses courses from the data set. These courses are taught by teaching-and-research professors, and typically, they receive high scores. We did the same with graduate courses so that the comparison is made using undergraduate courses taught by teaching-only and teaching-and-research professors. The results obtained are further explained in the Results section.

3.6 Deployment.

Deployment means putting the results into action. In our case, the results of the modeling task are classified in such a way that we can answer the research question regarding the comparison of teaching-only versus teaching-and-research professors and the various attributes that were used during the modeling phase. After their analysis and the corresponding validation with statistical methods, these results are given to university administrators in order to make the best use of both teaching-only and teaching-and-research professors in providing students the best possible service, by assigning teachers to classes in which they show a better performance.

4. Results

An initial exploratory analysis using the first dataset was performed. In this analysis, student satisfaction with the teaching they received was analyzed, calculating the average of the score obtained by the professor in the three questions of the survey. Next, we used the second dataset for analyzing the three questions separately. The second dataset that contains responses of five semesters is much larger than the former and allowed us to perform a longitudinal analysis over five periods of time.

4.1 Statistical Analysis

An exploratory analysis of the first dataset using statistical techniques to compare teaching-only with teaching-and-research professors was made. As shown in Figure 1, the mean ECOA score of teaching-and-research professors (Called SNI) is greater than the one for teaching-only professors (No SNI). If we split the data and analyze the responses of undergraduate and graduate students, we observe that graduate students evaluated professors better compared to undergraduate students (Figure 2). Additionally, teaching-and-research professors received a higher score from graduate students (9.21 > 8.98), whereas undergraduate students do an equal evaluation of both types of professors (Figure 3). The mean scores shown in Figures 1, 2, and 3 are significantly different according to the ANOVA results shown in Figure 4, which illustrates that the level of significance of these tests shows a high value.

4.2 Data Mining Analysis

In order to apply data mining techniques to find patterns in the datasets, we determined if some specific characteristics prevail when professors are evaluated as good or bad (above/below the mean) by student opinion. To accomplish this, we used Logistic Regression, also called logit regression, a technique used to find the probability that a binary goal variable takes the value yes or no, based on the attributes that define the goal variable. It uses a logistic function whose coefficients are calculated from the data to best fit the classification of goal variable as a yes or no. Next, we ranked the variables that provide better accuracy individually running a Recursive Feature Elimination algorithm that recursively removes features using the remaining attributes to work out the combination of attributes that contributes to the prediction of the goal variable. Making combinations of the attribute variables and using 80% of the data for training and 20% for testing, we found multiple models with accuracy ranging between 0.5 and 0.8. Figure 5 shows the ROC curve (Receiver Operating Characteristic) of these models, which covers 80% of the area and shows that the model built is a reliable approximation of the unknown true model.

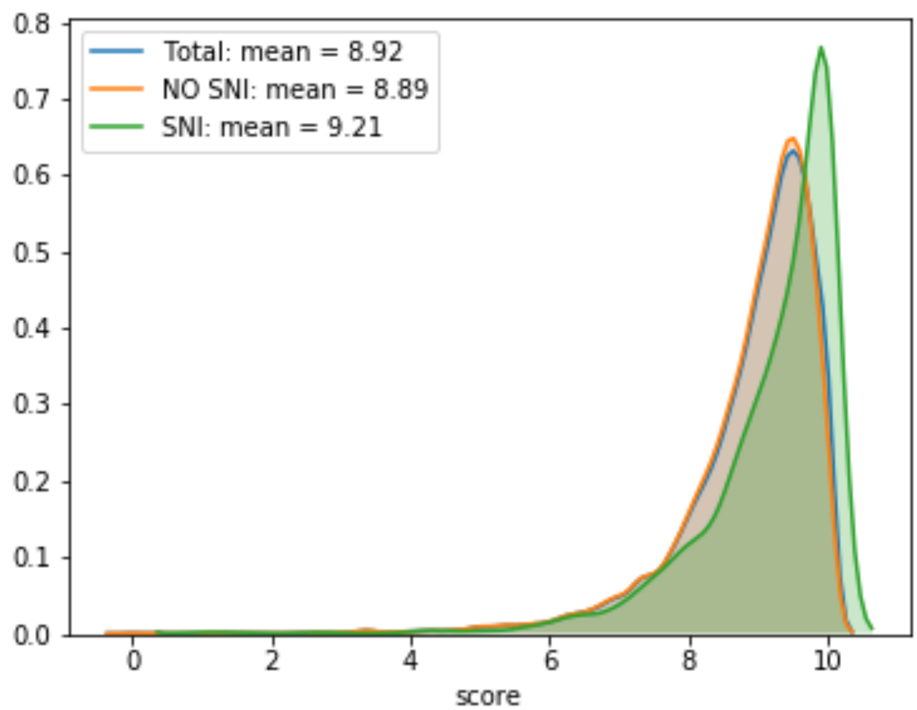


Figure 1: Score distribution of teacher-only professors (NO SNI) and teaching-and-research professors (SNI)

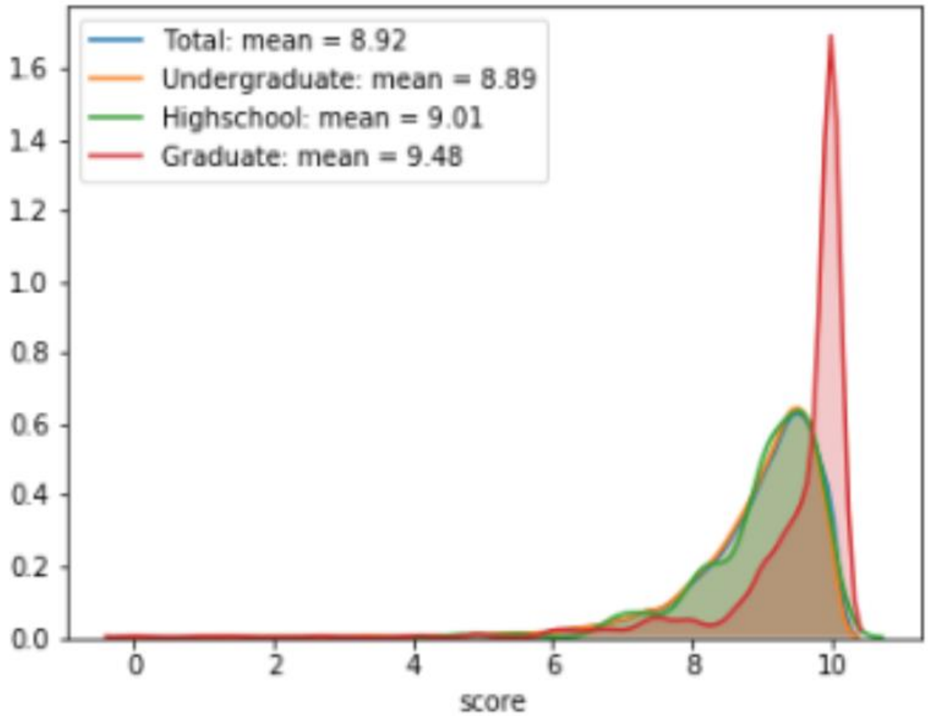


Figure 2: Score distribution of professors at the three academic levels (Highschool, Undergraduate, and Graduate), and overall (Total).

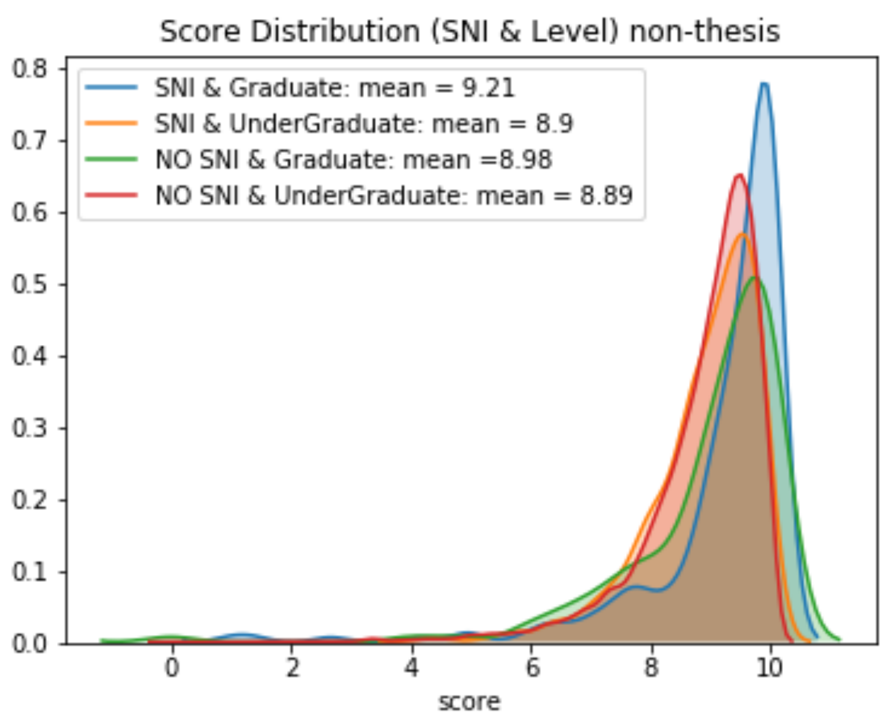


Figure 3: Score distribution of Teacher-only Professors (NO SNI) and Teaching-and-Research Professors (SNI) in the graduate and undergraduate groups.

Course Level	SNI - NO SNI	Academic Level	SNI / NO SNI & Academic Levels
F-statistic	6.678	20.233	8.498
P-value	0.0097676	0.0000069	0.0000123
R squared	0.0004397	0.0013470	0.0016968

Figure 4: ANOVA results for the comparison of average scores of Teaching-and-Research professors (SNI) versus Teacher-only professors (NO SNI), in groups of different Academic Levels, and comparing both dimensions.

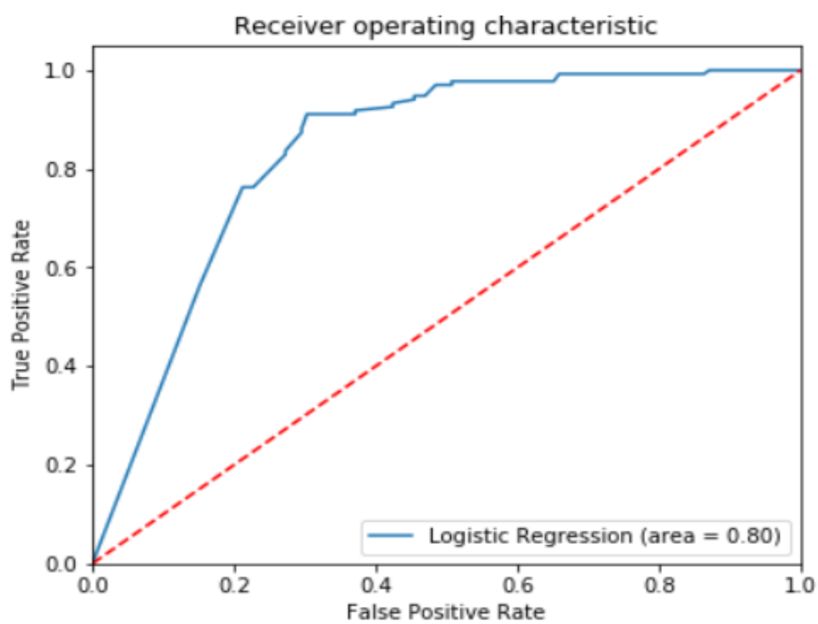


Figure 5: Receiver Operating Characteristic (ROC) Curve for the Logistic Regression models built using Feature Elimination on the variables of the first dataset.

The ranking of variables on groups of graduate students show that they score professors higher who: 1) have a higher responsibility percentage in the group, and 2) are researchers. This second variable proves the statement that a professor's research productivity is a valuable indicator of a teacher's educational skills and knowledge of the subject matter and is reflected in student evaluation of teaching [27]. The full list of ranked features is shown in Table 1.

Table 1. Professor and group features ranked by the Recursive Feature Elimination algorithm on groups of graduate students.

Rank	Feature	Professor / Group

1	Percentage of responsibility of the group	Professor
2	Is a teaching-and-research professor	Professor
3	Class transmitted to multiple campuses	Group
4	Number of senior students	Group
5	Number of hours in the classroom	Professor
6	Number of credits	Group
7	Foreign nationality	Professor
8	Number of undergraduate students attended	Professor
9	Number of scientific publications	Professor
10	% of participation in the survey	Group
11	Number of graduate students attended	Professor
12	Number of laboratory hours	Group
13	Main professor	Group
14	Total number of students attended	Professor
15	Number of teaching hours	Group

On the other hand, in groups with undergraduate students, the average score is positively correlated for professors who: (1) also teach at high school, (2) have a foreign nationality, (3) teach more groups, (4) have a greater percentage of responsibility in the group, and (5) is a teaching-only professor. The full list of ranked professor-and-group features is shown in Table 2. Table 3 shows the ranking for professor features only.

Table 2. Professor and group features ranked by the Recursive Feature Elimination algorithm on groups of undergraduate students.

Rank	**Feature**	**Professor / Group**
1	% of participation in the survey	Group
2	Number of high school students attended	Professor
3	Foreign nationality	Professor
4	Number of Teaching hours	Professor

5	Percentage of responsibility of the group	Professor
6	Main professor	Professor
7	Number of credits	Group
8	Is a terminal group	Group
9	Is a teaching-only professor	Professor
10	Total number of students attended	Professor
11	Number of undergraduate students attended	Professor
12	Certified in the teaching abilities program	Professor
13	Number of hours in the classroom	Professor
14	Number of senior students	Group
15	Class transmitted to multiple campuses	Group
16	Number of graduate students attended	Professor
17	Number of laboratory hours	Group
18	Number of scientific publications	Professor

Table 3. Professor features ranked by the Recursive Feature Elimination algorithm on groups of undergraduate students.

Rank	Feature
1	Number of high school students attended
2	Has a Ph.D.
3	Certified in the teaching abilities program
4	Number of undergraduate students attended
5	Number of graduate students attended
6	Total number of students attended
7	Has a Masters
8	Number of scientific publications
9	Is a teaching-only professor

10	Foreign nationality

In both cases, the percentage of senior students in the group is ranked among the top 6 features, which indicates that the maturity of students contributes most to scoring professors higher. A similar study also indicated that the older the student evaluating the professor, the higher the score will be given [28]. Additionally, we observe that the percentage of responsibility in both populations indicates that team teaching seems to affect the professors' evaluations in a negative way.

Finally, applying a Logistic Regression to a sample of graduate students only, the model with the highest accuracy (0.8) had a single variable, namely, the percentage of senior students. With this single variable, it was possible to forecast if a professor would be qualified above or below the average in 80% of the cases. The confusion matrix of this model is illustrated in Figure 6, where it can be observed that it is more likely to predict bad professors as good ones (38) than to predict good professors as bad ones (15).

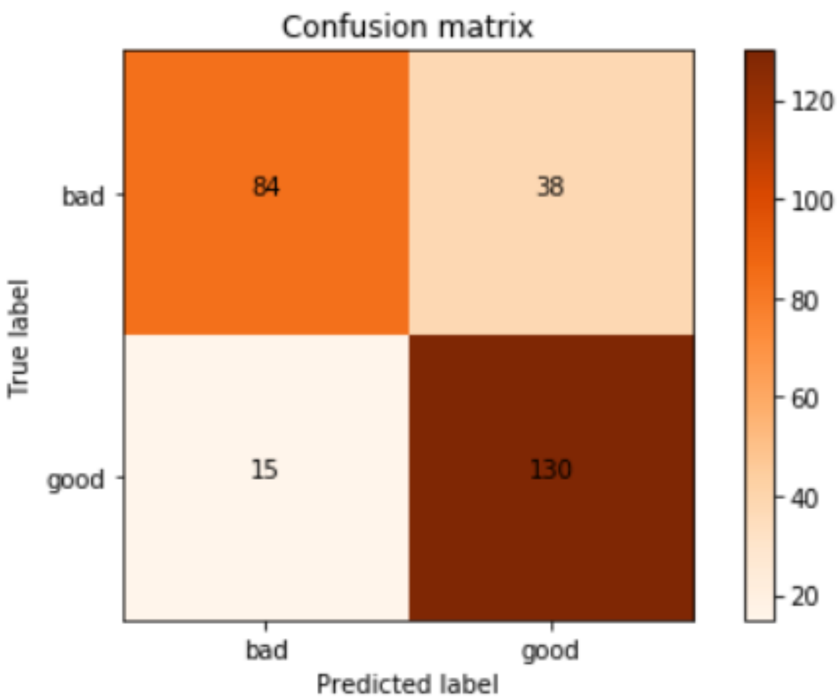


Figure 6: Confusion matrix for the Logistic Regression model based on the number of senior students in the group.

4.3 Analysis of the satisfaction dimensions

In order to make a deeper analysis of student satisfaction in the three dimensions measured by the ECOA survey, we used the second dataset. As mentioned before, this dataset contains responses from 5 semesters. Analyzing the average score for the three questions on the five semesters we observe that both Intellectual Challenge (05. RET) and Learning Guide (0.6. APR) have similar scores, 9.06 and 9.01, respectively, whereas Professor Recommendation (08. REC) is lower, with 8.73 average.

Nevertheless, when we compare the responses for full-time professors and researchers, we observe differences. As shown in Figure 7, the average score obtained by researchers is higher than those obtained by full-time professors in the three questions. According to a 1-way ANOVA test, these differences are significant.

Figure 7: Differences between the average score of full-time professors (FT Professor) and researchers (Researcher) in the three dimensions: Learning Guide (06. APR), Intellectual Challenge (05. RET), and Professor Recommendation (08. REC).

Next, the evolution of the three scores in the five semesters was analyzed. Figure 8 shows the different average scores obtained. It can be observed that the three of them increase over time. The factor 08. REC passes from 8.70 to 8.81 (Δ=0.11); 05. RET goes from 9.02 to 9.12 (Δ = 0.10); and 06. APR passes from 8.98 to 9.06 (Δ = 0.8). A 1-way ANOVA test demonstrated that the means of the three semesters are significantly different between every pair of semesters except for the first two.

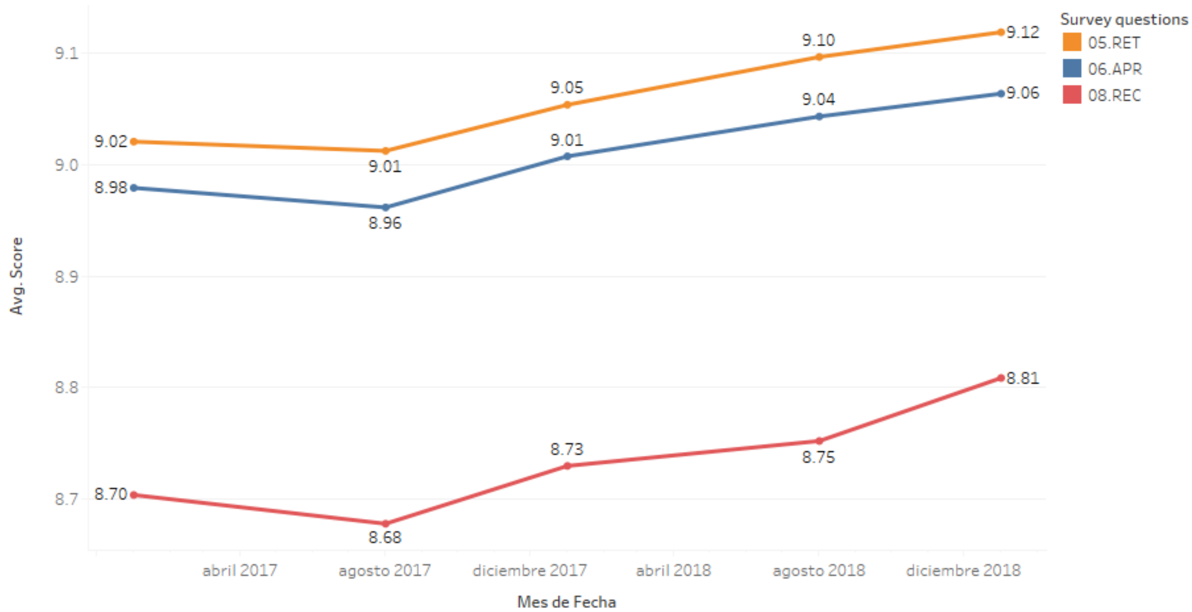


Figure 8: Temporal evolution of the three dimensions: Learning Guide (06. APR), Intellectual Challenge (05. RET), and Professor Recommendation (08. REC).

As shown in the previous subsection, graduate students scored their professors higher than undergraduate students. In the three factors, this difference is significant. As Learning Guide (06. APR), undergraduate students scored professors with 9.00 on average, in contrast to graduate students who scored professors with 9.33 ($\Delta = 0.33$). In terms of the Intellectual Challenge (05. RET), undergraduate students scored 9.05, whereas graduate students scored 9.42 on average ($\Delta = 0.37$). The difference is very notable in the Professor Recommendation score (08. REC), where undergraduate students give professors an 8.71 score, and graduate students score them with 9.28 ($\Delta = 0.57$).

Nevertheless, when both dimensions (the students' academic levels and the research activities) were broken down, the differences between researchers and full-time professors vanish at the undergraduate course level. Figure 9 shows the differences in satisfaction scores between researchers and full-time professors for both undergraduate and graduate students. Evidently, graduate students score researchers higher in the three scores. Nevertheless, undergraduate students scored researchers only slightly higher in the Intellectual Challenge factor, but in the Learning Guide and Professor Recommendation, both are scored the same. Once again, the biggest difference is in the Professor Recommendation (08. REC), where the researchers are scored 0.68 points higher than a full-time professor.

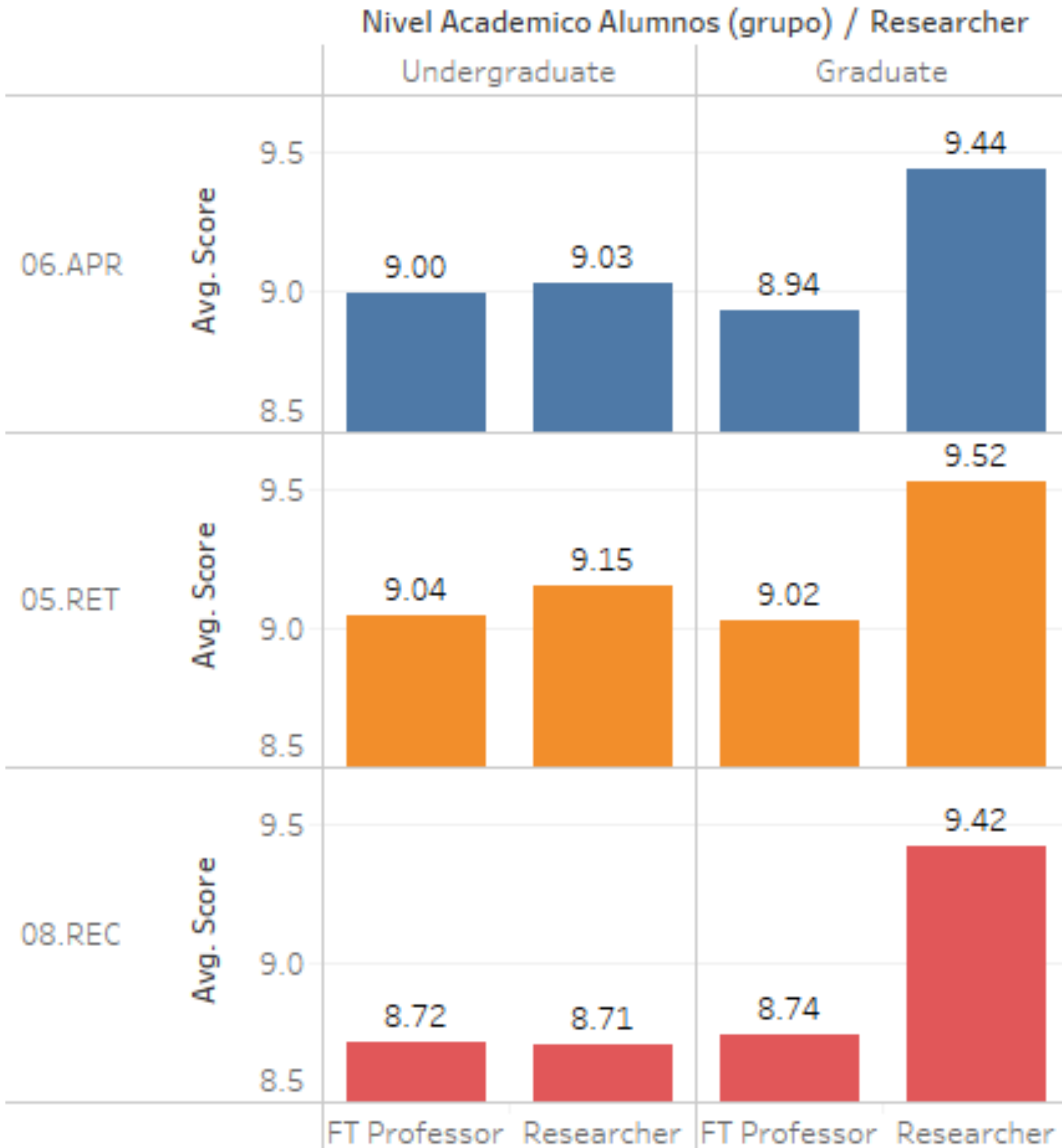


Figure 9: Differences between teaching-only professors (FT Professor) and teaching-and-research professors (Researcher) with undergraduate and graduate students, in the three dimensions: Learning Guide (06. APR), Intellectual Challenge (05. RET), and Professor Recommendation (08. REC).

Figure 10 shows the average scores obtained by women and men when they are evaluated as a professor in the three dimensions. Whereas in the Intellectual Challenge, there is practically no difference, in the Learning Guide and Professor Recommendation, women slightly overcome men by 0.07 and 0.04 points, respectively.

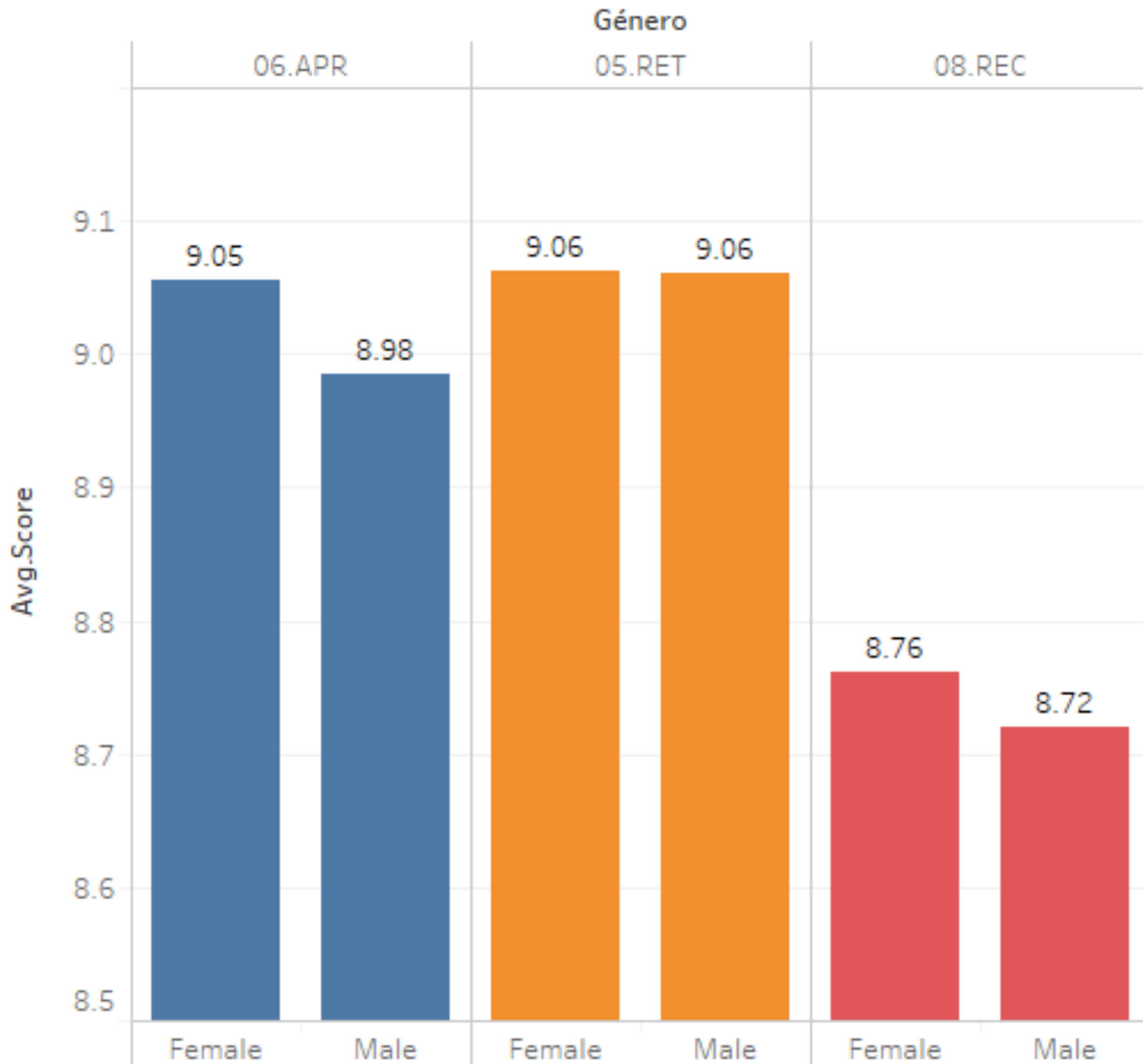


Figure 10: Differences in gender in the three dimensions: Learning Guide (06.APR), Intellectual Challenge (05.RET), and Professor Recommendation (08.REC).

Nevertheless, by splitting the sample between researchers and full-time professors, the differences become more notable. Figure 11 shows the differences between women and men once they are divided into researchers and full-time professors; the bar sizes indicate the number of professors evaluated. On the one hand, male researchers overcome female researchers in the three dimensions; the Intellectual Challenge and Professor Recommendation display the highest differences, 0.10 and 0.09 points, respectively. On the other hand, full-time female professors overcome male professors in only two dimensions: Learning Guide ($\Delta = 0.09$) and Professor Recommendation ($\Delta = 0.07$); at Intellectual Challenge, both men and women are evaluated the same [29].

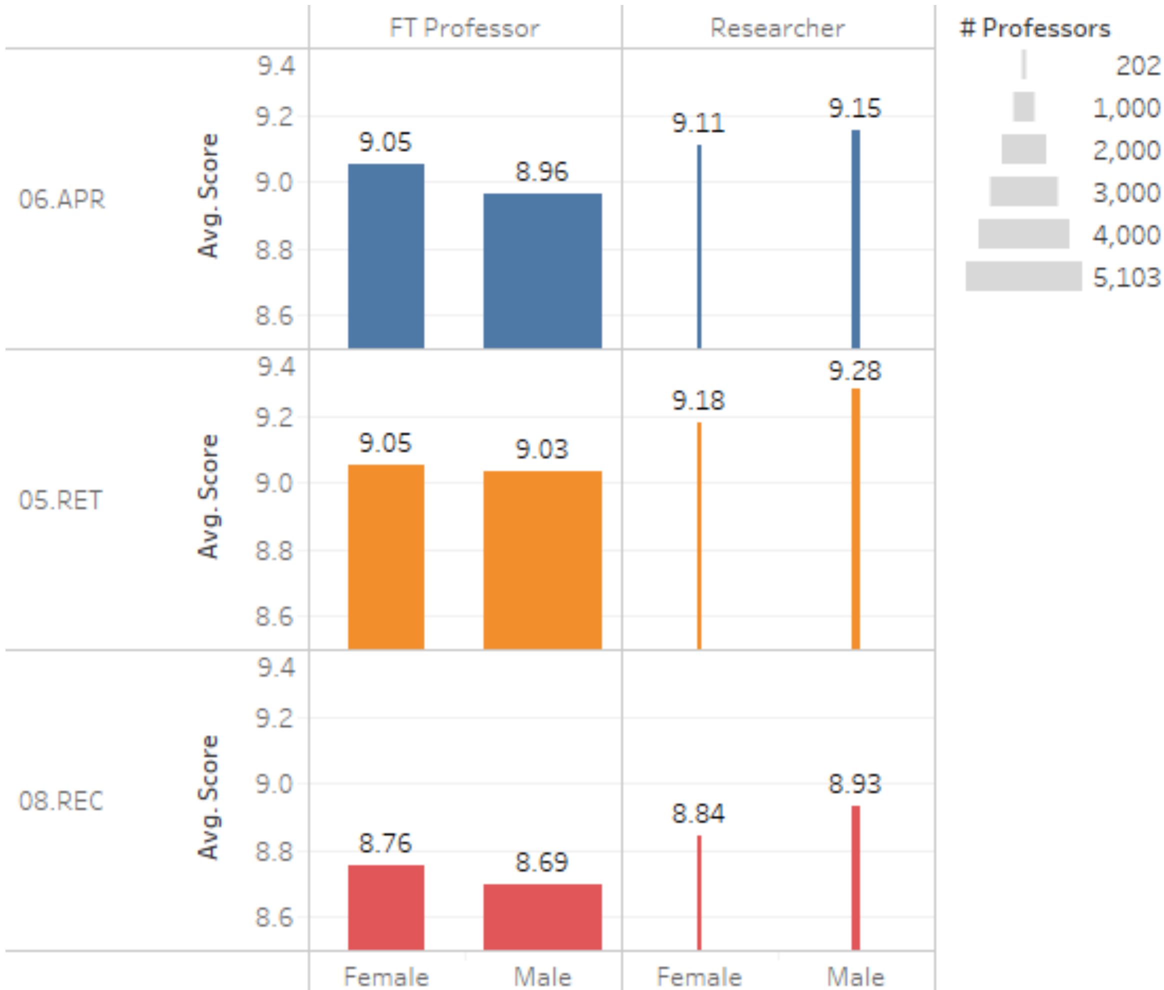


Figure 11: Differences by gender between teaching-only professors (FT Professor) and teaching-and-research professors (Researcher), in the three dimensions: Learning Guide (06. APR), Intellectual Challenge (05. RET), and Professor Recommendation (08. REC). The bar size indicates the number of professors in each group.

Next, how a professor's age influences student satisfaction was analyzed. Figure 12 shows the distribution of professors by age, divided into segments of 10 years. It can be observed that most of the professors are between 30 and 49 years old.

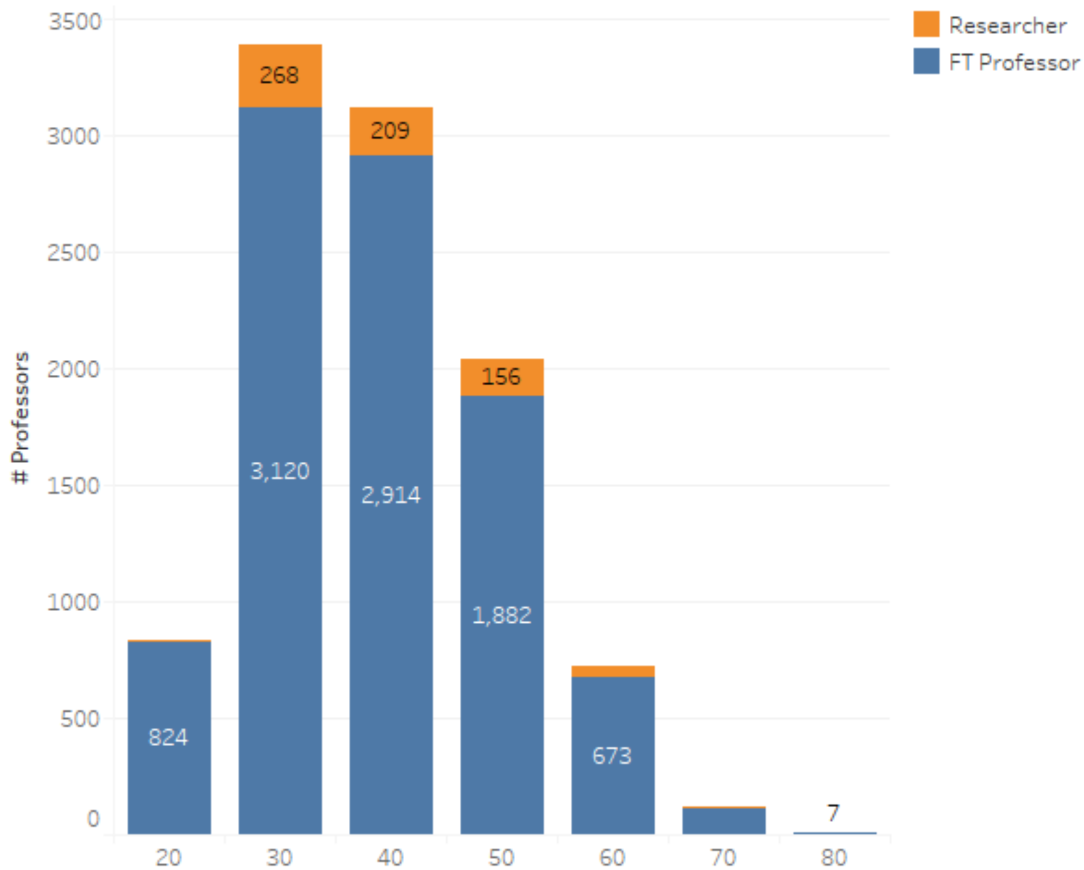


Figure 12: Distribution of professors by age (in 10-year bins).

The sample of full-time professors and researchers was split in order to observe how aging affects teaching quality. For full-time professors, aging affects the students' perception of teaching in the three dimensions, as shown in Figure 13. For Learning Guide, the difference between the youngest and oldest full-time professors is only -0.07 points, but for Intellectual Challenge, it is -0.13 points, and for Professor Recommendation, the difference is -0.34 points.

Figure 13: Average satisfaction scores of teaching-only professors by age, in the three dimensions: Learning Guide (06.APR), Intellectual Challenge (05.RET), and Professor Recommendation (08.REC).

In contrast, for researchers, aging improves teaching quality according to the students' responses. As shown in Figure 14, in Intellectual Challenge, the difference between the youngest and the oldest researchers is +0.22 points, for Learning Guide, it is +0.21, and for Professor Recommendation, it is +0.27.

Figure 14: Average satisfaction scores of teaching-and-research professors by researcher age, in the three dimensions: Learning Guide (06. APR), Intellectual Challenge (05. RET), and Professor Recommendation (08. REC).

Finally, the differences in student satisfaction due to the proficiency level of researchers were examined. For that purpose, we used the three levels of proficiency conferred by the National Research System of Mexico (SNI) upon our researchers. Figure 15 makes it evident that teaching quality grows with the proficiency level of the researcher in the three dimensions; the quality of the two upper levels being practically the same. For Intellectual Challenge, the difference between the lowest and the highest proficiency level is +0.35, whereas for Learning Guide, it is +0.32, and for Professor Recommendation, it is +0.46.

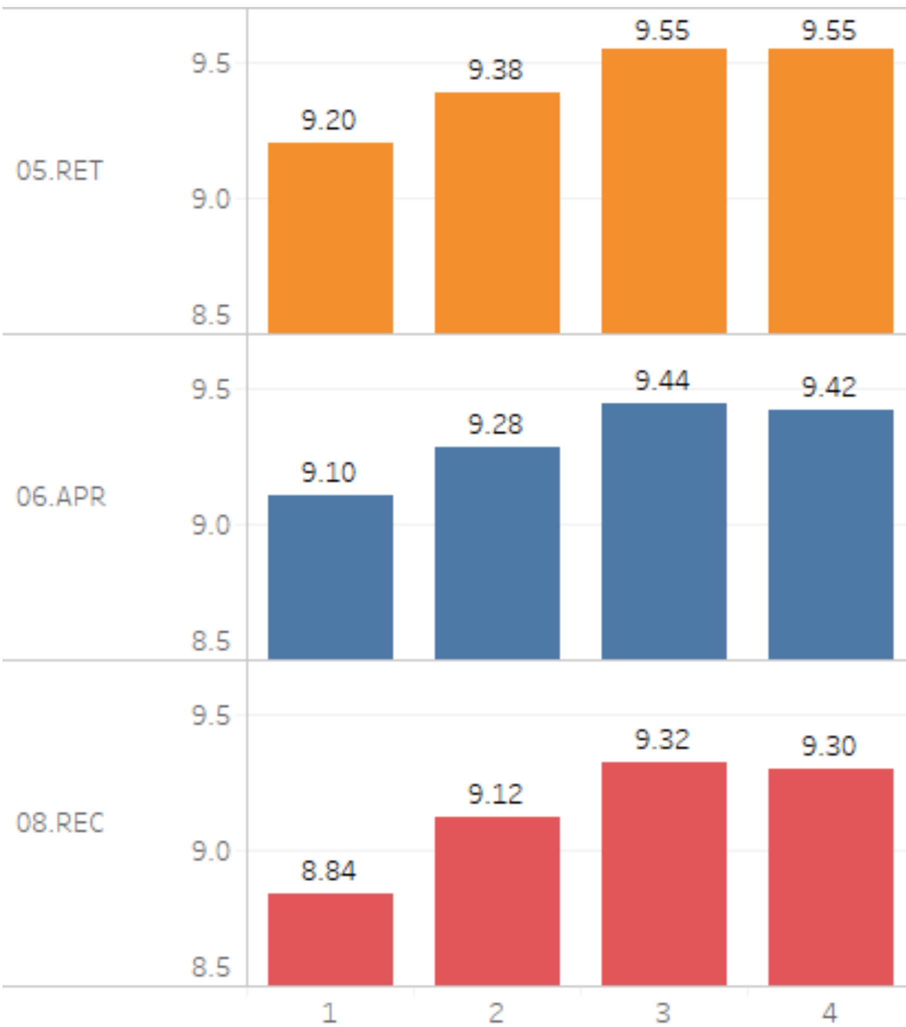


Figure 15. Average satisfaction scores of teaching-and-research professors by proficiency level in the three dimensions: Learning Guide (06. APR), Intellectual Challenge (05. RET), and Professor Recommendation (08. REC).

5. Discussion

5.1 Significance of the results.

A summary of the mean score for both teaching-only and teaching-and-research professors at different levels and obtained from the first dataset (August-December 2017) is shown in Figure 16. The advantage of teaching-and-research professors is observed at all levels, except in undergraduate groups. The ANOVA analysis helps us by stating that these means are indeed significantly different from each other, so the initial hypothesis is rejected. Additionally, note that graduate students evaluate the teaching-and-research professors even higher when thesis groups are considered.

Course Level	**Teaching-only professors**	**teaching-and-research professors**
Undergraduate	8.89	8.90
Graduate	9.89	9.21
Undergraduate and Graduate	8.90	9.21
Graduate and Thesis	9.18	9.57

Figure 16: Professors' evaluation means results classified by different levels. Note that a professor might be teaching simultaneously in undergraduate and graduate groups, but his/her evaluation is accounted for in the corresponding level.

This result is confirmed by the analysis of the second dataset, even when we split the weighted score into the three satisfaction dimensions of the ECOA survey. In the first place, we observed that Professor Recommendation is scored lower than the Learning Guide and Intellectual Challenge overall. And when we compare the scores obtained by teaching-only and teaching-and-research professors, we observed again that the latter obtained a higher score in all the three dimensions.

The ranking of professor and group characteristics using the Recursive Feature Elimination (RFE) algorithm permitted us to identify the relevance that students would assign to them (see Tables 1-3). Table 4 shows the comparison of the rankings obtained by RFE by analyzing the responses of the undergraduate and graduate students. The column "Agreement" has the absolute difference of both rankings: the lower this number, the more similar is the relevance that both levels of students assigned to a given feature. At the top of Table 4 are the features with more agreement, meanwhile features ranked by only one type of students are at the bottom.

Table 4. Agreement/Disagreement between graduate and undergraduate students' ranking of Professor and Group features based on the Recursive Feature Elimination algorithm.

Feature type	**Feature**	**Undergraduate Rank**	**Graduate Rank**	**Agreement**

Group	Number of credits	7	6	1
Professor	Number of undergraduate students attended	11	8	3
Professor	Foreign nationality	3	7	4
Professor	Percentage of responsibility of the group	5	1	4
Professor	Total number of students attended	10	14	4
Professor	Number of graduate students attended	16	11	5
Group	Number of laboratory hours	17	12	5
Professor	Main professor	6	13	7
Professor	Number of hours in the classroom	13	5	8
Group	% of participation in the survey	1	10	9
Professor	Number of scientific publications	18	9	9
Group	Number of senior students	14	4	10
Group	Number of teaching hours	4	15	11
Group	Class transmitted to multiple campuses	15	3	12

Professor	Number of high school students attended	2	-	
Group	Is a terminal group	8	-	
Professor	Is a teaching-only professor	9	-	
Professor	Certified in the teaching abilities program	12	-	
Professor	Is a teaching-and-research professor	-	2	

The features that undergraduate students rank high and graduate students rank low are: foreign nationality (3 vs. 7); the main professor is in team-teaching classes (6 vs. 13); percentage of participation on the survey (1 vs. 10), and the number of professors' teaching hours (4 vs. 15). Additionally, four features were ranked only in undergraduate groups, namely, the number of high school students attended by the professor (ranked 2nd), terminal groups (ranked 8), teaching-only professors (ranked 9), and certification in the teaching abilities program (ranked 12).

The features that graduate students rank high and undergraduate students rank low are the percentage of responsibility of the group in team-teaching groups (1 vs. 5), the number of hours spent in the classroom (5 vs. 13), the number of scientific publications made by the professor (9 vs. 18), the number of senior students in the group (4 vs. 14), and whether the class is transmitted to multiple campuses (3 vs. 15). Additionally, the unique feature that was ranked only at graduate groups was being a teaching-and-research professor (ranked 2nd).

As it can be seen and could be expected, features related to the professor's research activities are more appreciated by graduate students: 1) being a teaching-and-research professor (ranked 2), 2) classes with more hours of theory (ranked 5), and 3) the number of scientific publications made by the professor (ranked 9). The prominence of the ranking obtained by classes transmitted to multiple campuses at graduate groups can be explained by the relatively higher number of these groups in graduate programs.

On the other hand, undergraduate students appraised professor features such as 1) being a teaching-only professor (ranked 9), 2) the number of high school students attended by the professor, and 3) being certified in the teaching abilities program (ranked 12). Another feature ranked high by undergraduate students was having a foreign nationality (ranked 3), which seems to be appealing to Mexican students. And finally, features related to the maturity of students were also ranked high in undergraduate groups: being a group in the last semesters of the program (ranked 8th).

5.2 Related work.

Our results confirm the correlation found by Stack [18], Ting [30], and Spooren [27] about student perception of teaching quality and research activities. Stack and Ting correlate teacher's productivity in terms of papers and citations with higher scores on student evaluations. Our results are consistent with their findings as long as those professors we classified as teaching-only have very few scientific publications in comparison to those classified as teaching-and-research professors. The latter has at least one research product by year, and depending on their discipline, this number varies. Nevertheless, the differences between both kinds of professors, given the average scores of the three teaching dimensions, are more evident. We can attribute this pronounced difference to the definition of researcher we used, which is supported by CONACYT (the Mexican Council for Science and Technology), an external agency.

Furthermore, our results contribute to the global discussion by analyzing the effect of gender on teaching quality. In our analysis, female professors are more highly evaluated than male professors as Learning Guide (06. APR) and for Professor Recommendation (08. REC), confirming the observations of Basow and Montgomery [33] and Smith and colleagues [34]. And whereas this difference is also observed among teaching-only professors, it is reversed among teaching-and-research professors, where male researchers are better evaluated by students in the three teaching dimensions.

On the other hand, we observed that aging seems to improve the teaching quality of teaching-and-research professors, whereas it negatively affects the teaching quality of teach-only professors. This pattern found for teach-only professors is consistent with the results of Spooren [27], McPherson, and Jewell [31] and McPherson and colleagues [32], where they notice that younger teachers receive higher evaluations by students. Nevertheless, when we analyze teaching-and-research professors, we found the opposite trend. This could be attributed to the experience accumulated through years of research, as pointed out by McPherson [32]. In both cases, the dimension "Professor Recommendation" showed a higher difference between the youngest and the oldest professors.

In terms of proficiency level, we also found differences among teaching-and-research professors. In this case, an increase in the proficiency level is correlated with the quality of teaching in the three dimensions. An important finding is that researchers classified in the two upper levels of proficiency have similar scores, which are significantly higher than the lower two.

Finally, we would like to point out that we found evidence of an improvement in the teaching quality of the professors overall (see Figure 8). The causes of this improvement must be further studied. By now, we can only hypothesize that this variation is attributed to institutional efforts in educational innovation and training for all the professors.

6. Conclusions and Future Work

We present this study in which we try to answer the research question regarding the teaching performance of teaching-only and teaching-and-research professors: Does the former perform better than the latter according to student opinion in teaching-and-research institutions? The context in which this question is addressed is given by teaching-and-research universities ranked in the band 101 - 200 of the QS World University rankings, although the same question can be asked for universities with a teaching-and-research orientation independently of the ranking band.

The research question was approached by applying the methodology Cross-Industry Standard Process for Data Mining, known as CRISP-DM, which is a common method used in data analytics studies. The six steps were applied, which define the methodology that goes from business understanding to system deployment, including data understanding and preparation, as well as modeling and model evaluation. The modeling phase of the CRISP-DM methodology applied data mining and statistical methods that included logistic regression and analysis of variance. The calculation of coefficients of the logistic function using the training data yielded a model that was applied to the holdout data that was set apart to test the model. The accuracy of the resulting model was evaluated using procedures like ROC curves and the confusion matrix, which showed a statistically significant prediction capacity.

Once the model was found, the experiments were carried out using different variables to perform runs with data drawn from the undergraduate level, graduate level, the age of teachers, and teacher gender. In all cases, the performance of research professors was better or at least the same compared to teaching-only professors. With these results, we found a negative answer to the research question (The Null Hypothesis) and provided evidence in favor of the research professor yielding higher in teaching activity (The Alternate Hypothesis).

The results showed that in general, teaching-and-research professors perform better or at least the same as teaching-only professors in the graduate or undergraduate academic levels using data of student survey results from five semesters in a teaching-and-research university ranked in position number 158 of the QS World University Rankings 2020. We hope that these results contribute to revising the belief that research professors in teaching-and-research institutions are not good teachers in general.

Independently of teaching-only versus research-and-teaching professor behavior, we believe that the results obtained may be useful information in scheduling teacher classes in academic periods in higher education institutions. By identifying the group attributes where researchers are best evaluated, we could recommend a better group assignment for them, i.e., graduate courses and undergraduate courses of terminal semesters. In this way, we would improve the learning process and satisfaction of students.

As for future work, a non-deterministic technique known as matching [29] can be applied to account for the differences in student satisfaction for these types of professors. The matching procedure allows us to compare the sub-samples of teaching-only and teaching-and-research professors with the same characteristics [27], in order to determine if they have an equivalent evaluation teaching the same population of students or the same class. A further line of research is to consider the function that calculates the professor's score as a linear optimization problem subject to a set of constraints given by academic level (undergraduate or graduate), gender (male or female), teacher seniority (age and degrees), semester in which a course is taught, students' major, geography (area of the country in which the campus is located), and some other attributes. The result of the linear optimization problem could be used in scheduling teacher workload to maximize teacher performance and improve service to students, as well [35, 36].

Finally, the educational model is continually evolving [37]. The new educational model at Tecnologico de Monterrey started in August 2019, introducing a set of innovations in pedagogical methods whose elements have been smoothly introduced since 2012 and 2019, are now fully implemented and deployed [38]. The role of a teacher, either teaching-only or teaching-and-research, has been fully revised, and new features of teacher activity need to be evaluated in order to determine the level of teaching quality and satisfaction [39]. Thus, a new model for evaluation of teaching performance needs to be designed and implemented, and once it is running, we will need new data generated from student evaluations of teaching in order to conduct new experiments in the evaluations of teacher-only versus teaching-and-research professors [40].

8. References

[1] Charlotte Danielson (2013). The Framework For Teaching Evaluation Instrument: The Newest Rubric Enhancing The Links To The Common Core State Standards, With Clarity Of Language For Ease Of Use And Scoring. Copyright material by Charlotte Danielson.

[2] Arthur M. Lange (2017). Information Technology and Organizational Learning: Managing Behavioral Change in the Digital Age /Third edition). CRC Press, Taylor and Francisco, Boca Raton, Florida.

[3] Association for the Advancement of Artificial Intelligence (2016). Artificial Intelligence and Life in 2030. One Hundred Year Study on Artificial Intelligence. Report of the Study Panel 2015. https://ai100.stanford.edu/2016-report Retrieved September 18, 2019.

[4] Kaplan, Andreas M.; Haenlein, Michael (2016). "Higher education and the digital revolution: About MOOCs, SPOCs, social media, and the Cookie Monster." Business Horizons. 59 (4): 441–50.

[5] Yousef, A. M. F., Chatti, M. A., Schroeder, U., Wosnitza, M., Jakobs, H. (2014). MOOCs - A Review of the State-of-the-Art. Int Conf on Computer Supported Education 2014. Barcelona, Spain, pp. 9–20.

[6] Shirky, Clay (8 July 2013). "MOOCs and Economic Reality." The Chronicle of Higher Education.

[7] Fowler, Geoffrey A. (2013). An early report card on MOOCs. Wall Street Journal.

[8] Katy Jordan (2017). MOOC completion rates. www.katyjordan.com. Retrieved Sep 16, 2019.

[9] D Koller (2012). “MOOCs on the Move: How Coursera Is Disrupting the Traditional Classroom" (text & video). Knowledge @ Wharton. U of Pennsylvania. Retrieved Sept 16, 2019.

[10] Coffrin, Carleton; Corrin, Linda; de Barba, Paula; Kennedy, Gregor (2014). Visualizing Patterns of Student Engagement and Performance in MOOCs. Proc of the Fourth International Conf on Learning Analytics and Knowledge. New York, NY, USA: ACM. pp. 83–92.

[11] Breslow, Lori; Pritchard, David E.; DeBoer, Jennifer; Stump, Glenda S.; Ho, Andrew D.; Seaton, Daniel T. (2013). Studying Learning in the Worldwide Classroom Research into edX's First MOOC. Research & Practice in Assessment, vol. 8, pp 13-25.

[12] Hew, Khe Foon (May 2016). Promoting engagement in online courses: What strategies can we learn from three highly rated MOOCs. British J of Educational Technology. 47 (2): 320–341.

[13] Salazar, M.C, and Lerner, J. (2019). Teacher Evaluation as Cultural Practice: A Framework for Equity and Excellence. Language, Culture, and Teaching Series, Routledge, New York.

[14] Kim Marshal (2013). Rethinking Teacher Supervision and Evaluation: How to Work Smart, Build Collaboration, and Close the Achievement Gap, 2nd ed. Copyright material Jassey Bass, Willey.

[15] Grubera, T., Reppelb, A. & Voss, R. “Understanding the characteristics of effective professors: the student’s perspective.” J of Marketing for Higher Education (2010) 175–190.

[16] Tsinidou, M., Gerogiannis, V. & Fitsilis, P. "Evaluation Of The Factors That Determine Quality In Higher Education: An Empirical Study." Quality Assurance in Education (2010). 227-244.
[17] Spooren, P., Brockx, B. & Mortelmans, D. (2013). On the Validity of Student Evaluation of Teaching: The State of the Art. Review of Educational Research. Vol. 83, pp 598-642.
[18] Stack, S. (2003). Research Productivity And Student Evaluation Of Teaching In Social Science Classes. Research in Higher Education, *44*, 539–556.
[19] Cantú-Ortíz F. (Ed.). Research Analytics. Taylor and Francis Group - CRC Press (2018).
[20] Cantu-Ortiz, F., Bustani, A. Moreira, H., Molina, A. (2009). A Research-Based Development Model: The Research Chairs Approach. The J of Knowledge Management, 13(1), pp 154 – 170.
[21] Shearer C. (2000). The CRISP-DM model: the new blueprint for data mining, J Data Warehousing; Vol. 5, pp. 13—22.
[22] CRISP-DM (2019). Cross-industry Standard Process for Data Mining. https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining. Retrieved September 18, 2019.
[23] Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, and Rüdiger Wirth (2000); CRISP-DM 1.0 Step-by-step data mining guides
[24] Harper, Gavin, and Stephen D. Pickett (2006). Methods for mining HTS data. Drug Discovery Today. 11 (15–16), pp 694–699.
[25] QS World University Rankings 2020 www.topuniversities.com/university-rankings/world-university-rankings/2020
[26] Provost F., and Fawcett, T. (2015). Data Science for Business: What you need to know about Data Mining and Data Analytics Thinking. O' Reilly Media Inc., Sebastopol, CA, USA.
[27] Spooren, P. (2010). On the credibility of the judge. A cross-classified multilevel analysis on student evaluations of teaching. Studies in Educational Evaluation, *36*, 121–131. doi:10.1016/j.stueduc.2011.02.001
[28] Benton, S. & Cashin, W. (2011). IDEA Paper No. 50: Student ratings of teaching: A summary of research and literature.
[29] Djurdjevic, D. & Radyakin, S. "Decomposition of the Gender Wage Gap Using Matching: An Application for Switzerland." Swiss J Economics Statistics (2007) 143: 365. https://doi.org/10.1007/BF03399243
[30] Ting, K. (2000). A multilevel perspective on student ratings of instruction: Lessons from the Chinese experience. Research in Higher Education, 41, 637–661. http://doi:10.1023/A:1007075516271
[31] McPherson, M. A., & Todd Jewell, R. (2007). Leveling the playing field: Should student evaluation scores be adjusted? Social Science Quarterly, 88, 868–881. http://doi:10.1111/j.1540-6237.2007.00487.x
[32] McPherson, M. A., Todd Jewell, R., & Kim, M. (2009). What determines student evaluation scores? A random-effects analysis of undergraduate economics classes. Eastern Economic Journal, 35, 37–51. http://doi:10.1057/palgrave.eej.9050042
[33] Basow, S. A., & Montgomery, S. (2005). Student ratings and professor self-ratings of college teaching: Effects of gender and divisional affiliation. J of Personnel Evaluation in Education, 18, 91–106. http://doi:10.1007/s11092-006-9001-8
[34] Smith, S. W., Yoo, J. H., Farr, A. C., Salmon, C. T., & Miller, V. D. (2007). The influence

of student sex and instructor sex on student ratings of instructors: Results from a college of communication. Women's Studies in Communication, 30, 64–77. http://doi:10.1080/07491409.2007.10162505

[35] Chaerani, D. and Dewanto, S. (2014). Active Learning based on Research for Linear Optimization Course at the Department of Mathematics, Universitas Padjadjaran. Int Conf on Advances in Education Technology.

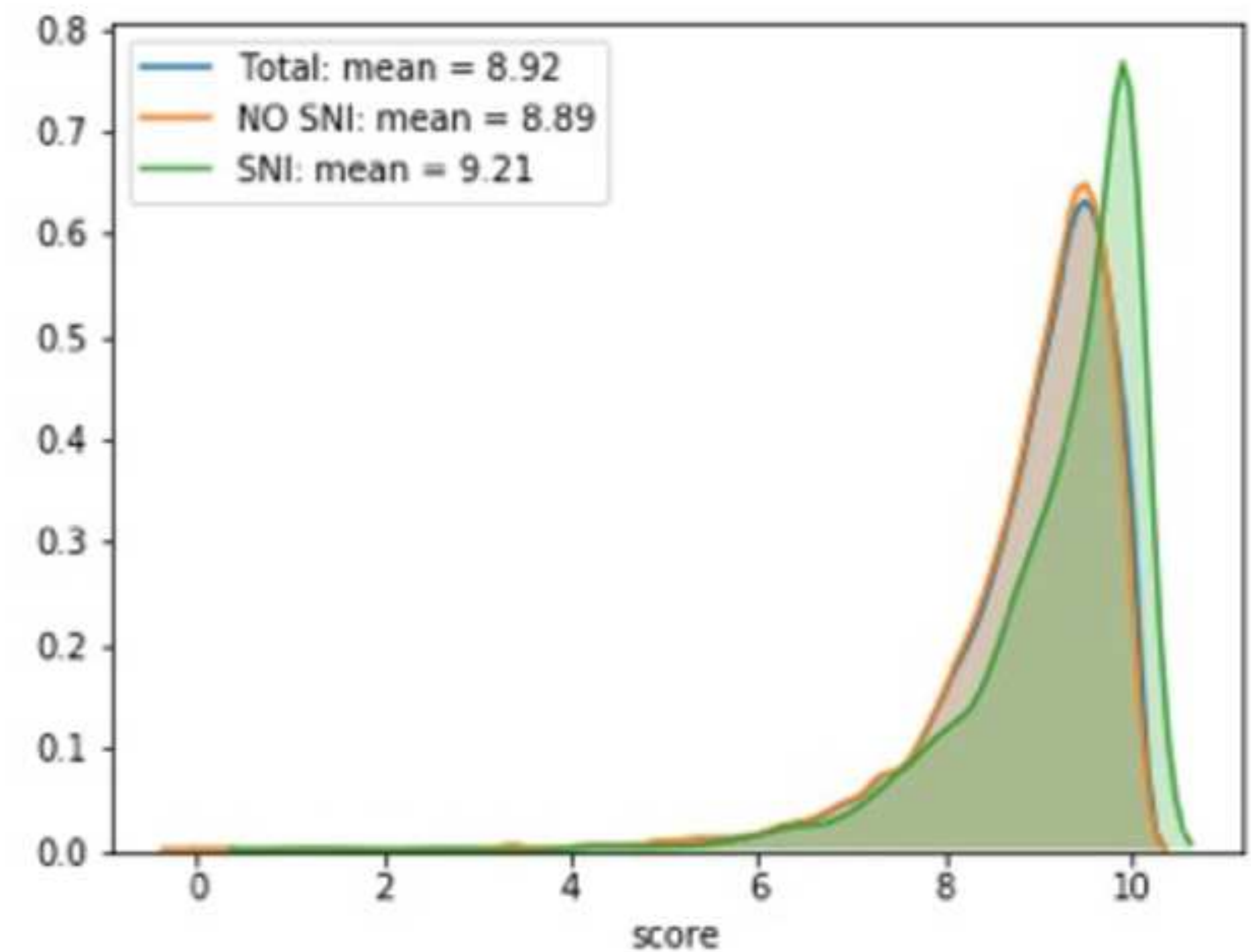
[36] Bucco, G. Bornia-Poulsen, C., Bandeira, C. (2017). Development of a Linear Programming Model for the University Course Timetabling Problem. Gest. Prod, Sao Carlos, v. 24, n. 1, pp 40-49, http://dx.doi.org/10.1590/0104-530X2133-15

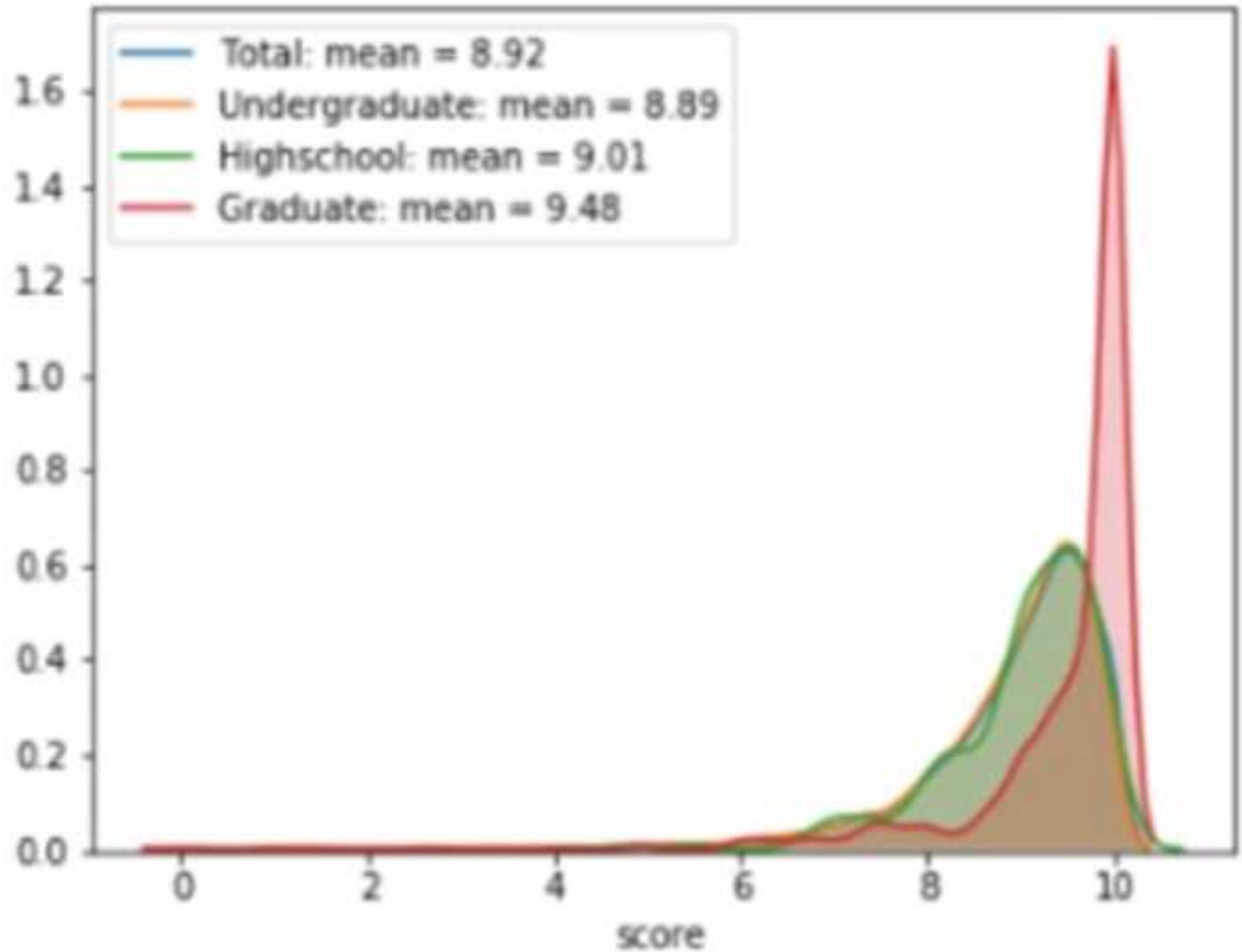
[37] Tec21 (2019). Educational Model of Tecnologico de Monterrey, Tec 21.

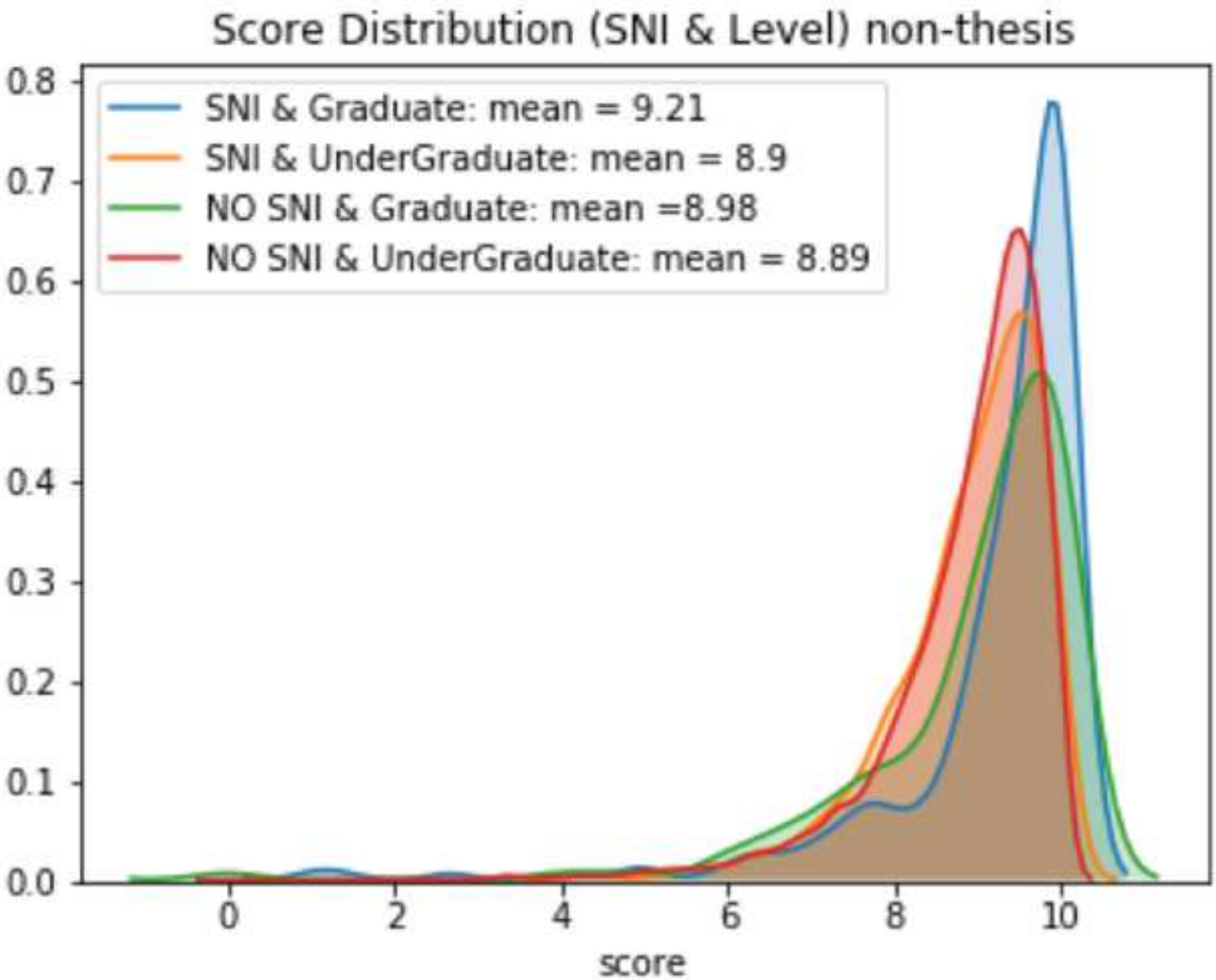
[38] Akela, D. (2010). Learning Together: Kolb's Experimental Theory and its Application. J of Management and Organization, v. 16, n. 1, pp 100-112.

[39] Garcia, N. Guajardo, B. Valenzuela J. (2017). Blueprint de un Sistema de Innovación Educativa en las Instituciones de Educación Superior: El Caso del Tecnológico de Monterrey y su Modelo al 2021. 4º Congreso Internacional de Innovacion Educativa, Monterrey, México.

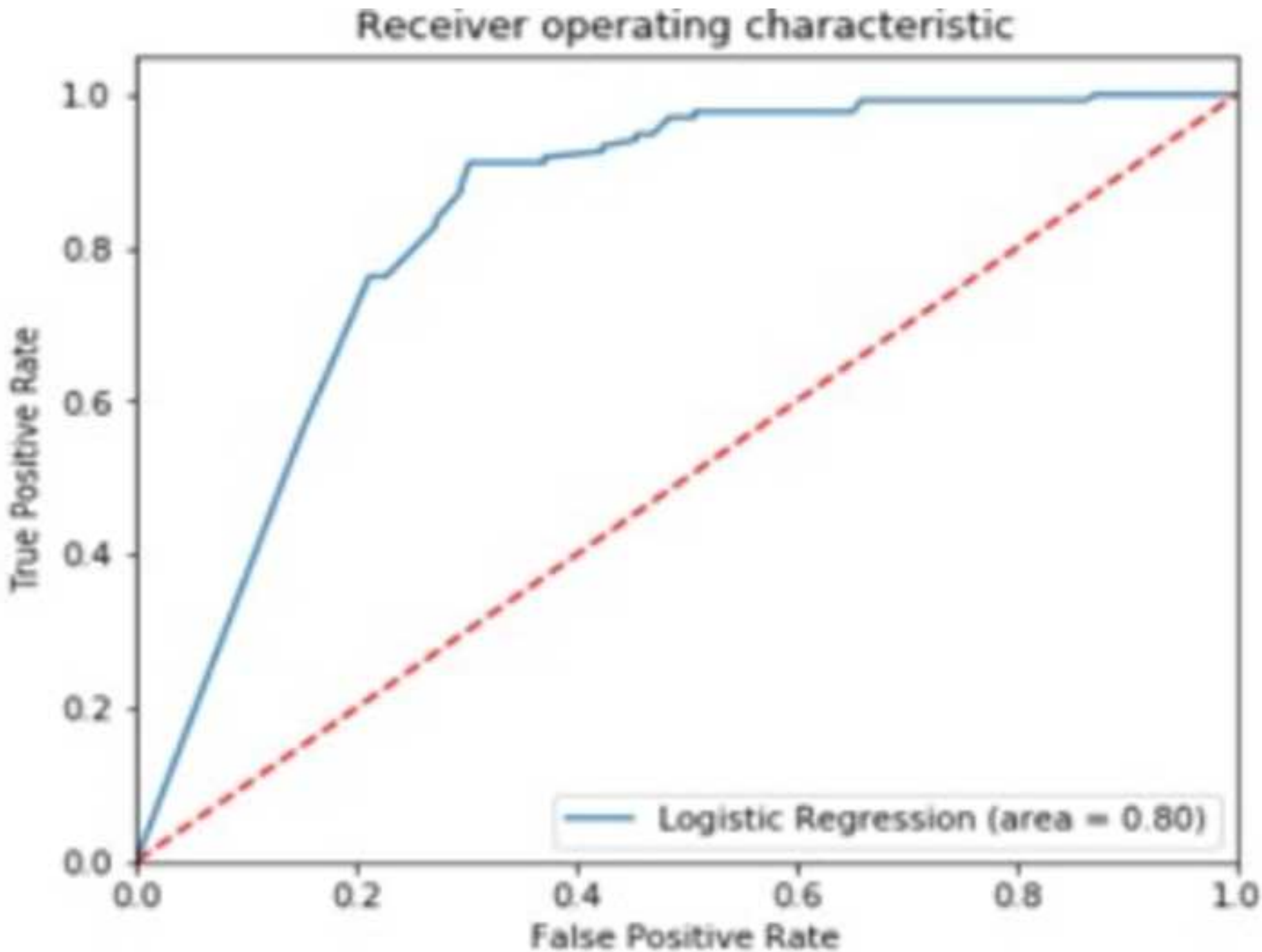
[40] Bridgstock, R. and Tippet, N. (2019). Higher Education and the Future of Graduate Employability. Edward Elgar Publishing Limited: Cheltenham, UK.

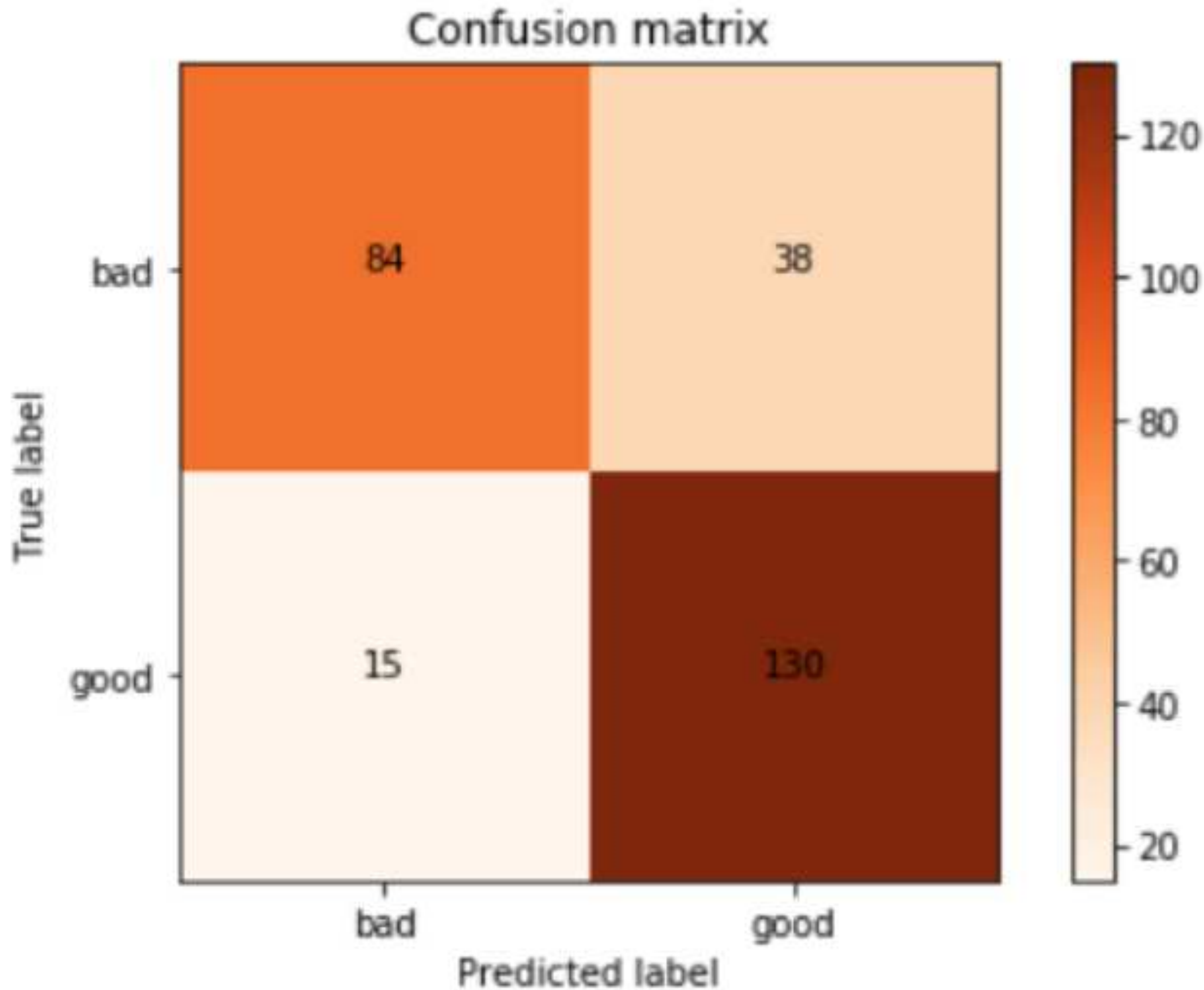

Total: mean = 8.92
NO SNI: mean = 8.89
SNI: mean = 9.21
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
2
4
6
8
10
score


Total: mean = 8.92
Undergraduate: mean = 8.89
Highschool: mean = 9.01
Graduate: mean = 9.48
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
score

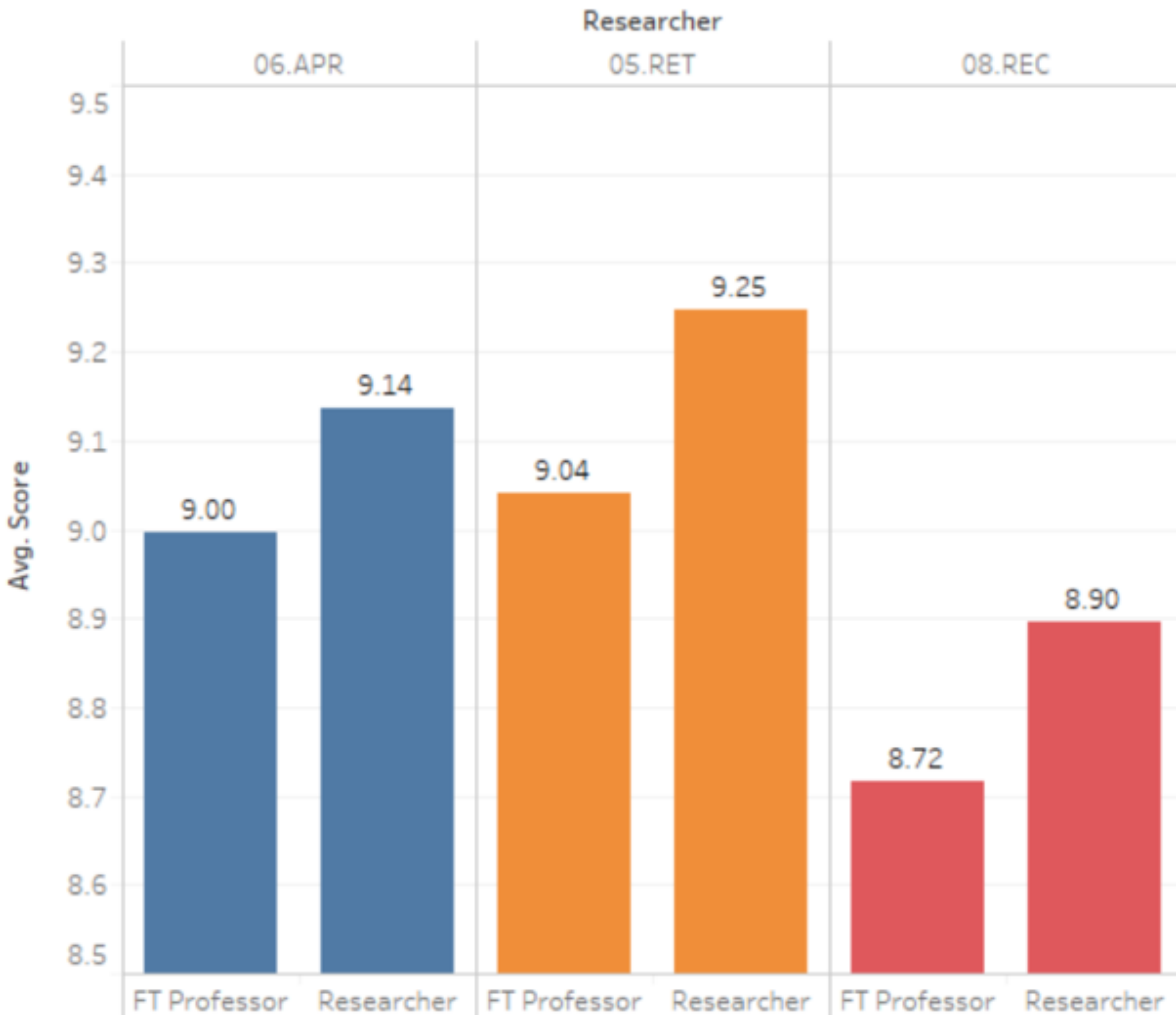

Score Distribution (SNI & Level) non-thesis
SNI & Graduate: mean = 9.21
SNI & UnderGraduate: mean = 8.9
NO SNI & Graduate: mean =8.98
NO SNI & UnderGraduate: mean = 8.89
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
2
4
6
8
10
score

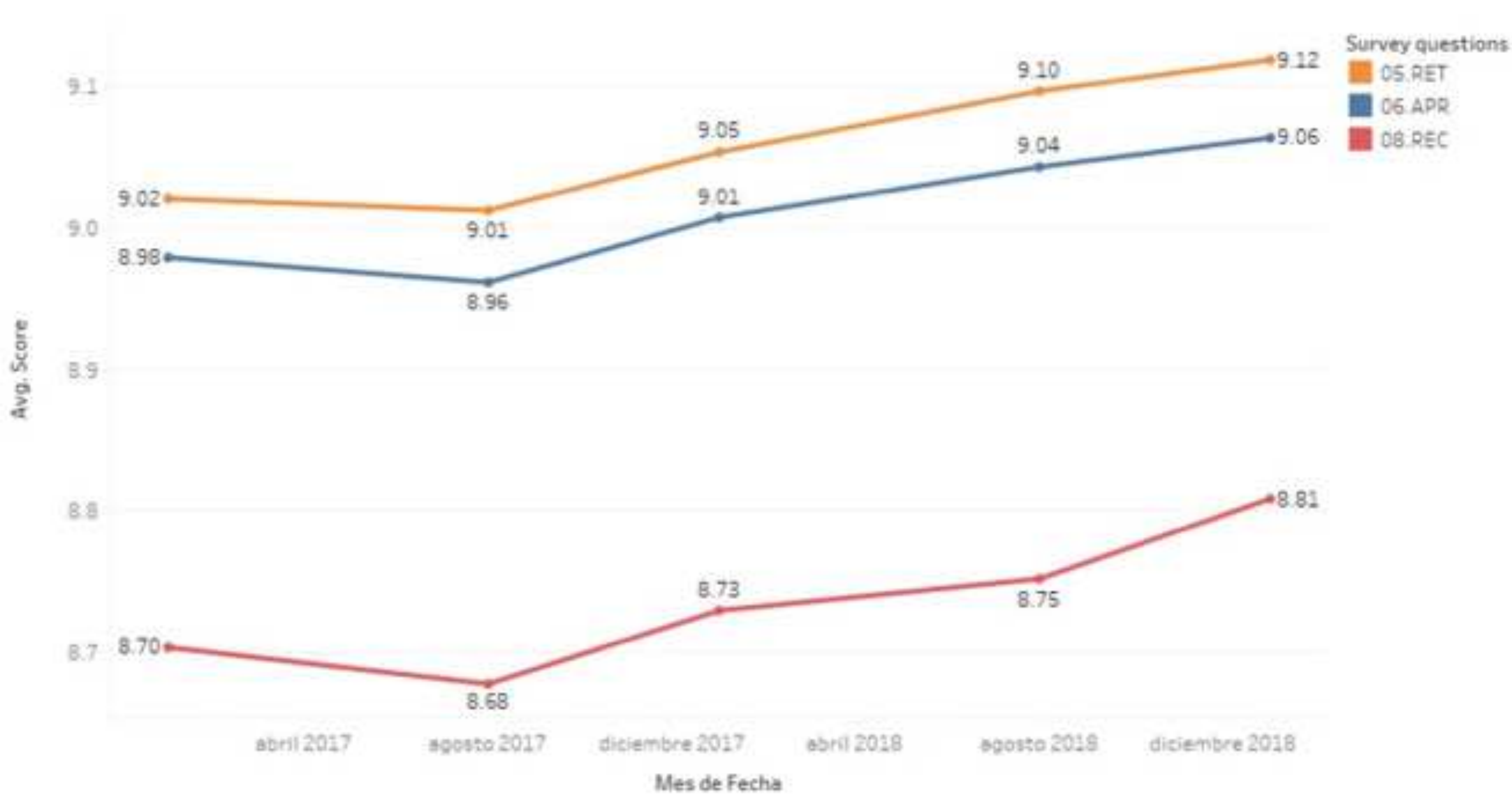
Course Level	**SNI - NO SNI**	**Academic Level**	**SNI / NO SNI & Academic Levels**
F-statistic	6.678	20.233	8.498
P-value	0.0097676	0.0000069	0.0000123
R squared	0.0004397	0.0013470	0.0016968


Receiver operating characteristic
1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate
Logistic Regression (area = 0.80)


Confusion matrix
True label
bad
good
84
38
15
130
bad
good
Predicted label
120
100
80
60
40
20

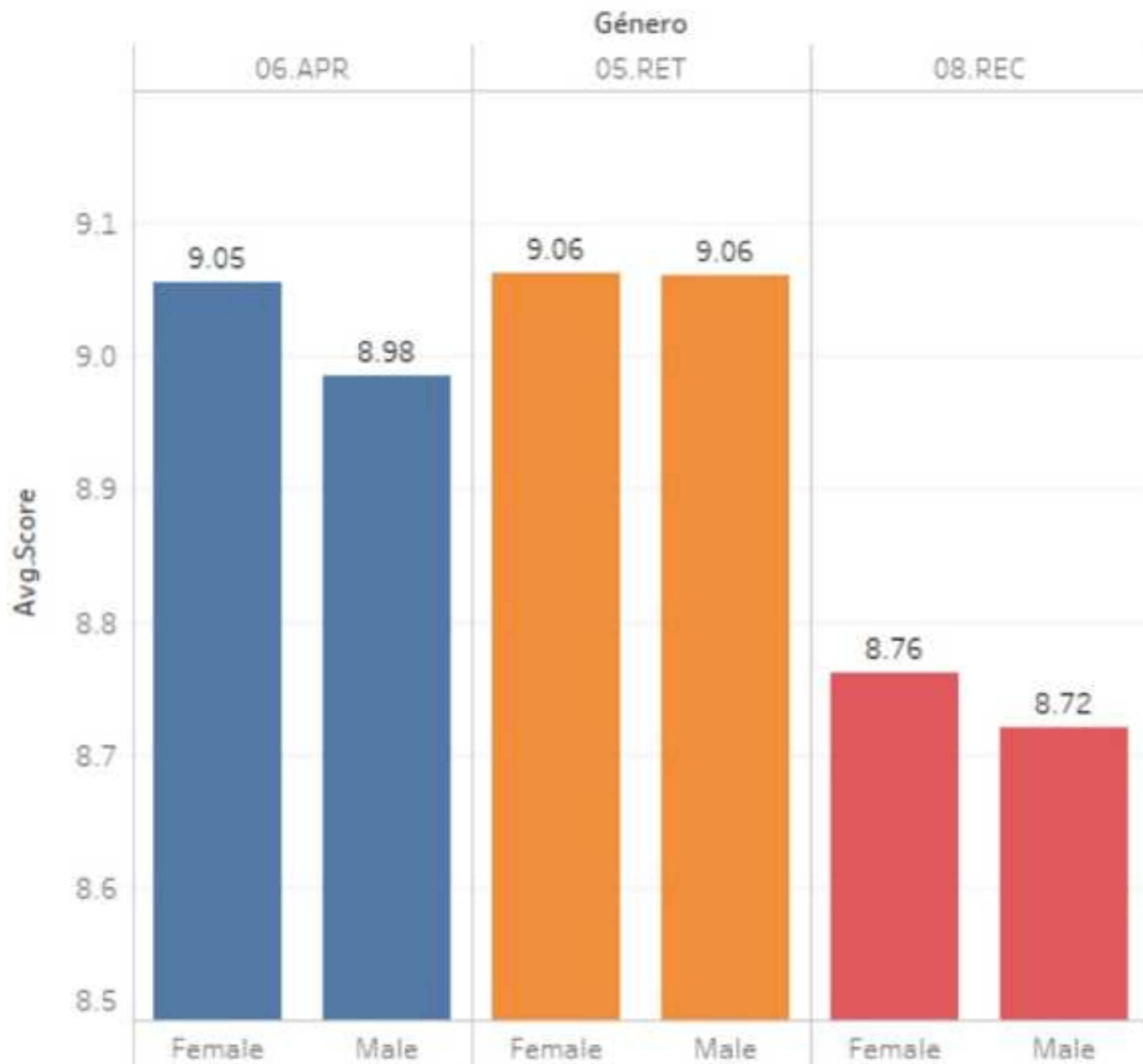
Figure

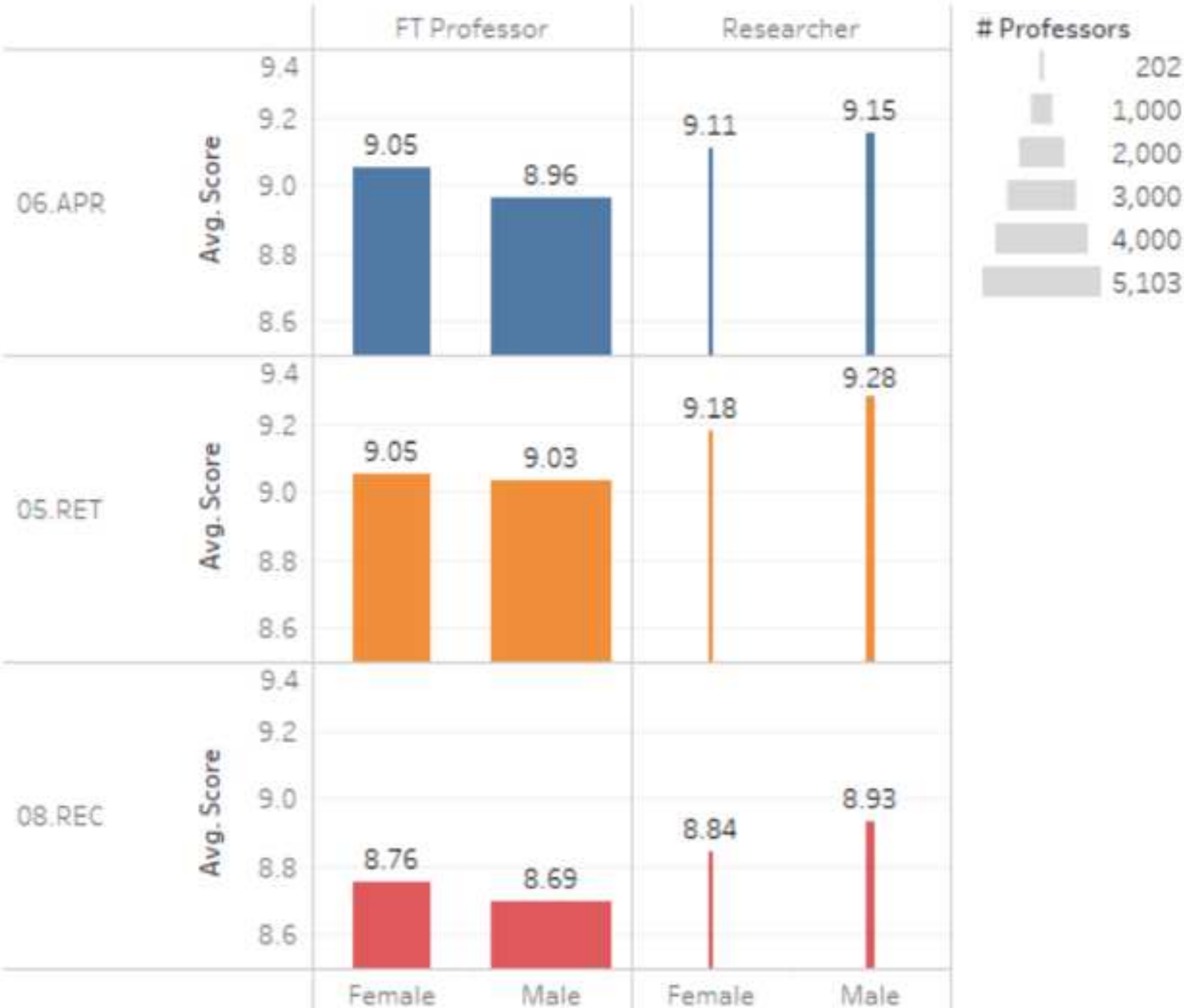

Researcher
06.APR
05.RET
08.REC
Avg. Score
9.5
9.4
9.3
9.2
9.1
9.0
8.9
8.8
8.7
8.6
8.5
9.00
9.14
9.04
9.25
8.72
8.90
FT Professor
Researcher
FT Professor
Researcher
FT Professor
Researcher


Survey questions
05.RET
06.APR
08.REC
Avg. Score
9.1
9.0
8.9
8.8
8.7
9.02
9.01
9.05
9.10
9.12
8.98
8.96
9.01
9.04
9.06
8.70
8.68
8.73
8.75
8.81
abril 2017
agosto 2017
diciembre 2017
abril 2018
agosto 2018
diciembre 2018
Mes de Fecha

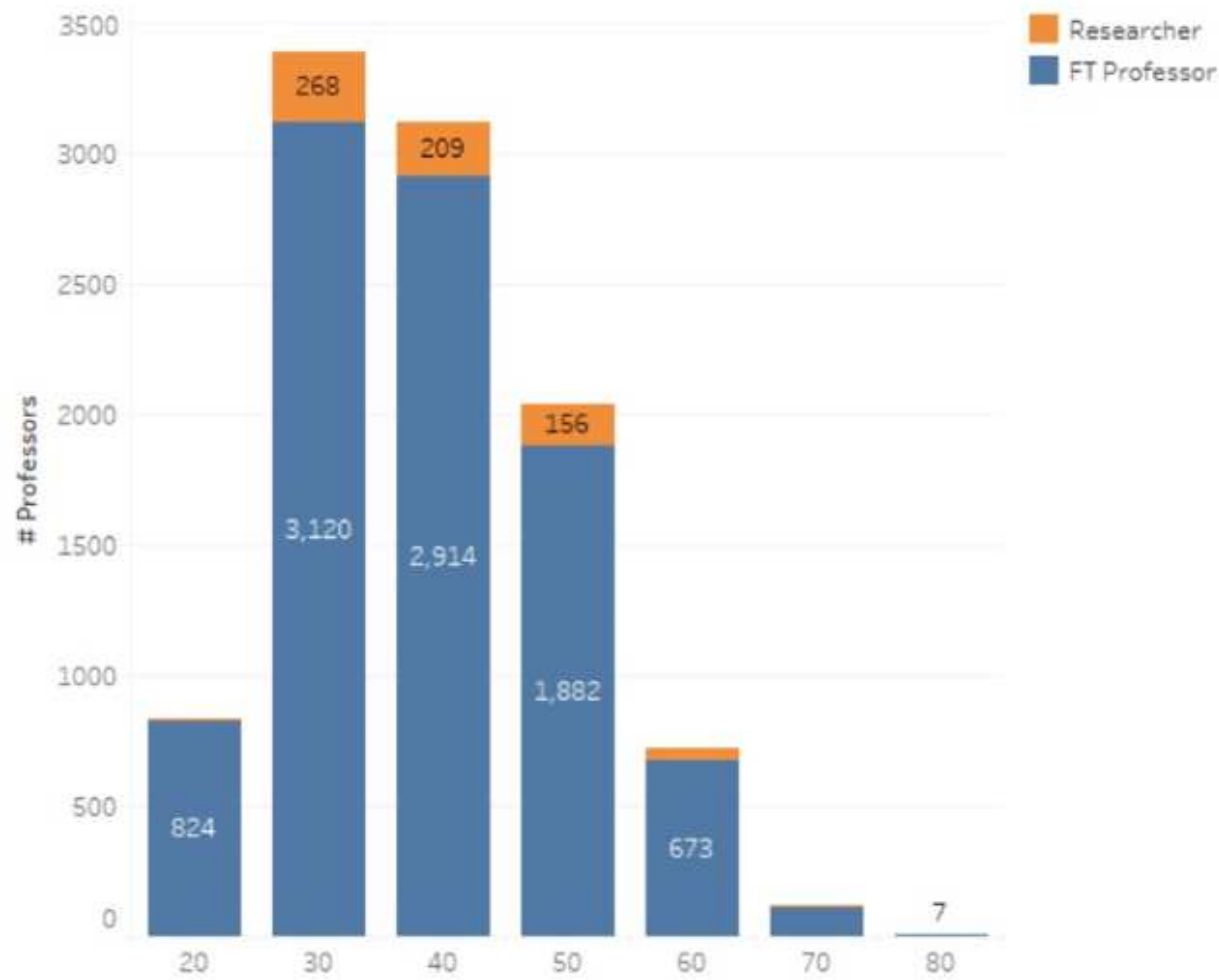
Figure


Nivel Academico Alumnos (grupo) / Researcher
Undergraduate
Graduate
06.APR
05.RET
08.REC
Avg. Score
9.5
9.0
8.5
9.00
9.03
8.94
9.44
9.04
9.15
9.02
9.52
8.72
8.71
8.74
9.42
FT Professor
Researcher


Género
06.APR
05.RET
08.REC
Avg.Score
9.1
9.0
8.9
8.8
8.7
8.6
8.5
9.05
8.98
9.06
9.06
8.76
8.72
Female
Male
Female
Male
Female
Male

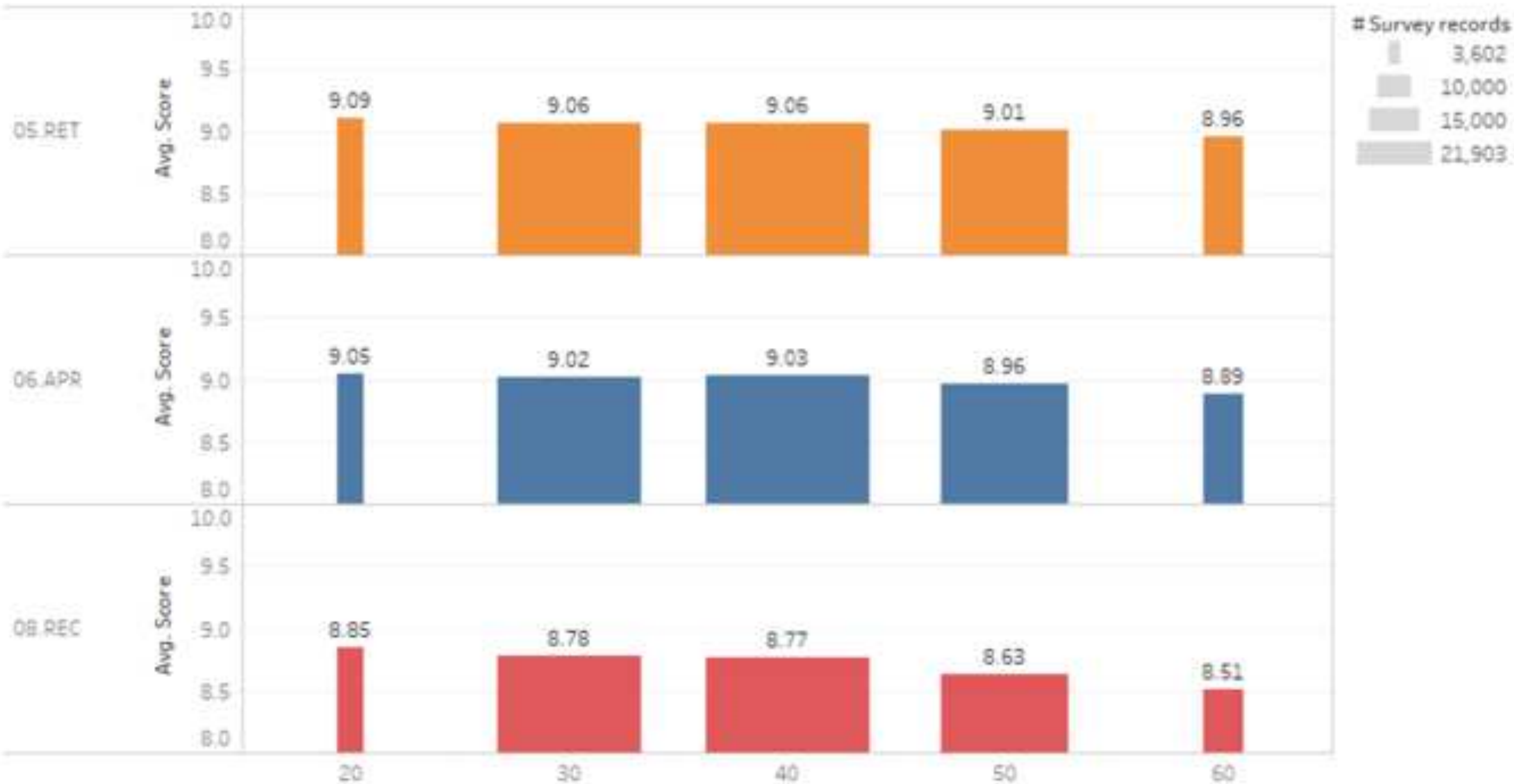


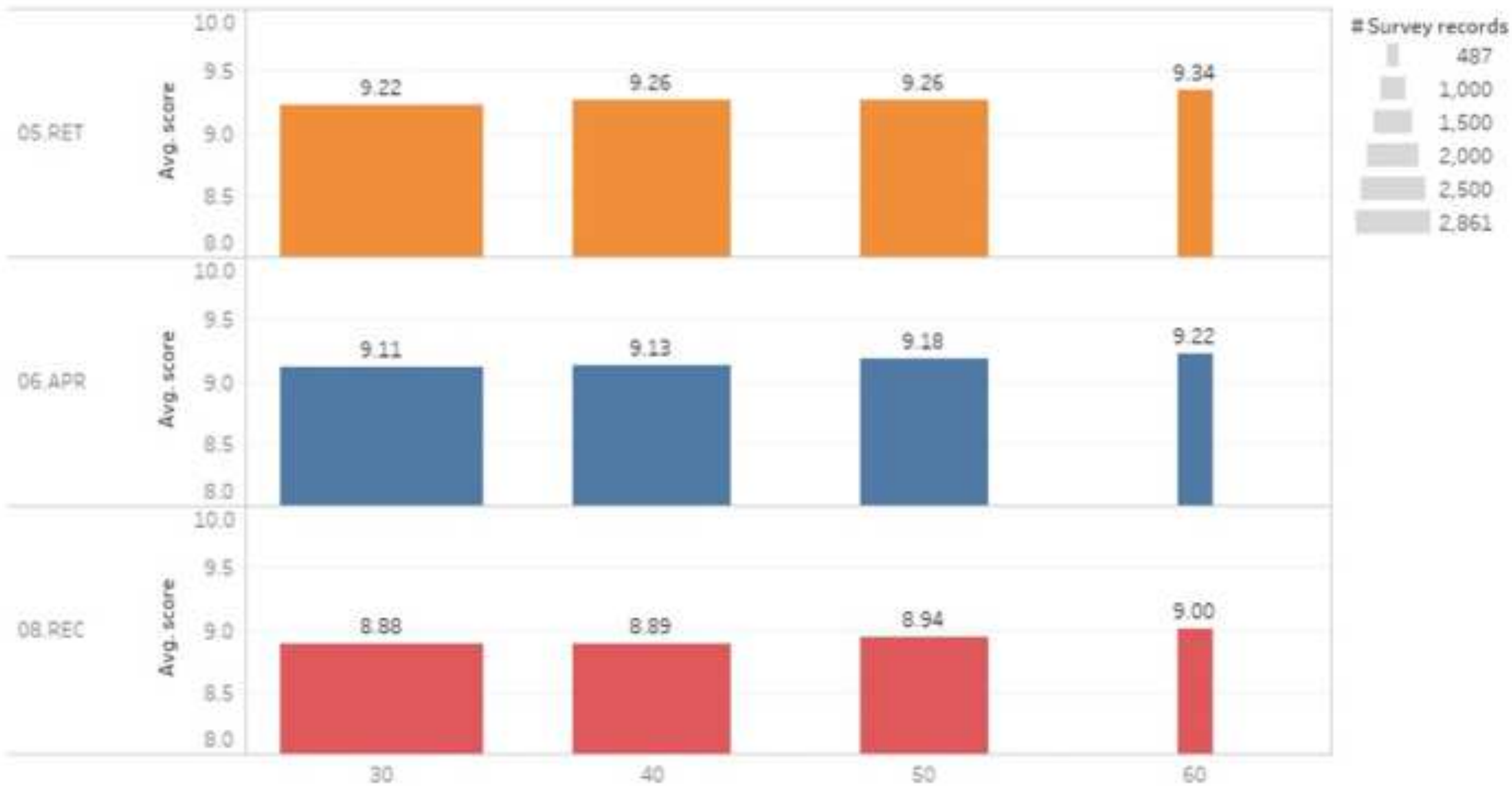
FT Professor
Researcher
Professors
202
1,000
2,000
3,000
4,000
5,103
06.APR
05.RET
08.REC
Avg. Score
9.4
9.2
9.0
8.8
8.6
9.05
8.96
9.11
9.15
9.05
9.03
9.18
9.28
8.76
8.69
8.84
8.93
Female
Male
Female
Male



Researcher
FT Professor
Professors
3500
3000
2500
2000
1500
1000
500
0
268
3,120
209
2,914
156
1,882
824
673
7
20
30
40
50
60
70
80

Figure


05.RET
06.APR
08.REC
Avg. Score
10.0
9.5
9.0
8.5
8.0
9.09
9.06
9.06
9.01
8.96
9.05
9.02
9.03
8.96
8.89
8.85
8.78
8.77
8.63
8.51
20
30
40
50
60
Survey records
3,602
10,000
15,000
21,903


05.RET
06.APR
08.REC
Avg. score
10.0
9.5
9.0
8.5
8.0
9.22
9.26
9.26
9.34
9.11
9.13
9.18
9.22
8.88
8.89
8.94
9.00
30
40
50
60
Survey records
487
1,000
1,500
2,000
2,500
2,861


05.RET
06.APR
08.REC
9.5
9.0
8.5
9.20
9.38
9.55
9.55
9.10
9.28
9.44
9.42
8.84
9.12
9.32
9.30
1
2
3
4

Course Level	Teaching-only professors	teaching-and-research professors
Undergraduate	8.89	8.90
Graduate	9.89	9.21
Undergraduate and Graduate	8.90	9.21
Graduate and Thesis	9.18	9.57

Rank	Feature	Professor / Group
1	Percentage of responsibility of the group	Professor
2	Is a teaching-and-research professor	Professor
3	Class transmitted to multiple campuses	Group
4	Number of senior students	Group
5	Number of hours in the classroom	Professor
6	Number of credits	Group
7	Foreign nationality	Professor
8	Number of undergraduate students attended	Professor
9	Number of scientific publications	Professor
10	% of participation in the survey	Group
11	Number of graduate students attended	Professor
12	Number of laboratory hours	Group
13	Main professor	Group
14	Total number of students attended	Professor
15	Number of teaching hours	Group

Rank	Feature	Professor / Group
1	% of participation in the survey	Group
2	Number of high school students attended	Professor
3	Foreign nationality	Professor
4	Number of Teaching hours	Professor
5	Percentage of responsibility of the group	Professor
6	Main professor	Professor
7	Number of credits	Group
8	Is a terminal group	Group
9	Is a teaching-only professor	Professor
10	Total number of students attended	Professor
11	Number of undergraduate students attended	Professor
12	Certified in the teaching abilities program	Professor
13	Number of hours in the classroom	Professor
14	Number of senior students	Group
15	Class transmitted to multiple campuses	Group
16	Number of graduate students attended	Professor
17	Number of laboratory hours	Group
18	Number of scientific publications	Professor

Rank	Feature
1	Number of high school students attended
2	Has a Ph.D.
3	Certified in the teaching abilities program
4	Number of undergraduate students attended
5	Number of graduate students attended
6	Total number of students attended
7	Has a Masters
8	Number of scientific publications
9	Is a teaching-only professor
10	Foreign nationality

Feature type	Feature	Undergraduate Rank	Graduate Rank	Agreement
Group	Number of credits	7	6	1
Professor	Number of undergraduate students attended	11	8	3
Professor	Foreign nationality	3	7	4
Professor	Percentage of responsibility of the group	5	1	4
Professor	Total number of students attended	10	14	4
Professor	Number of graduate students attended	16	11	5
Group	Number of laboratory hours	17	12	5
Professor	Main professor	6	13	7
Professor	Number of hours in the classroom	13	5	8
Group	% of participation in the survey	1	10	9
Professor	Number of scientific publications	18	9	9
Group	Number of senior students	14	4	10
Group	Number of teaching hours	4	15	11
Group	Class transmitted to multiple campuses	15	3	12
Professor	Number of high school students attended	2	-	
Group	Is a terminal group	8	-	
Professor	Is a teaching-only professor	9	-	
Professor	Certified in the teaching abilities program	12	-	
Professor	Is a teaching-and-research professor	-	2	