Using Customer Knowledge Surveys to Explain Sales of Postgraduate Programs: A Machine Learning Approach

Eva Asensio, Alejandro Almeida*, Aida Galiano, Juan-Manuel Martín-Álvarez*

Universidad Internacional de La Rioja (UNIR), Logroño (Spain)

UNIR
LA UNIVERSIDAD
EN INTERNET

Abstract

Universities collect information from each customer that contacts them through their websites and social media profiles. Customer knowledge surveys are the main information-gathering tool used to obtain this information about potential students. In this paper, we propose using the information gained via surveys along with enrolment databases, to group customers into homogeneous clusters in order to identify target customers who are more likely to enroll. The use of such a cluster strategy will increase the probability of converting contacts into customers and will allow the marketing and admission departments to focus on those customers with a greater probability of enrolling, thereby increasing efficiency. The specific characteristics of each cluster and those postgraduate programs that are more likely to be selected are identified. In addition, better insight into customers regarding their enrolment choices thanks to our cluster strategy, will allow universities to personalize their services resulting in greater satisfaction and, consequently, in increased future enrolment.

I. Introduction

WHEN informing themselves about postgraduate programs in public and private universities, potential students commonly use the Internet [1]. Furthermore, the presence of universities on social networks, such as Instagram or Twitter, allows interaction with potential students.

Many universities have websites and social media profiles that allow them to collect information about potential students. In addition, some of these universities send customer knowledge surveys entitled, for example, "we want to know more about you" to potential students who are interested in the programs on offer. The data generated through these platforms is highly valued [2].

Generally, the use of the Internet facilitates communication between the potential student and postgraduate university services [1]. In many cases, universities offer unrequested information about their programs, for example, via social networks [3]. Therefore, for both universities and potential students, the consensus is that the use of the Internet generates value in the educational market [4].

The connections made through websites or social networks used in combination with customer knowledge surveys which some potential students voluntarily fill out, provide useful sales-optimizing information. In this context, the analysis of the information gathered contributes to building a better customer relationship management strategy [5].

Regarding the use of digital platforms, it must be taken into consideration that potential students are saturated with information from many sources. Immersed in an environment of over-information [6] they do not only receive information about university degrees, and pursuing a postgraduate degree is not even a primary need in many cases [7], [8]. Unsurprisingly in this context, there is a high probability that potential students will not notice or ignore the publicity they receive from universities. Thus, it is very important for the Universities to know their target customers well in order to access them effectively [9]. As mentioned above, postgraduate training may not be a primary need, so knowing customers well is essential to sales maximization [10]-[12].

These challenges demand that university marketing departments consider the fundamental characteristics of those potential students who have made the decision to start the admission process in postgraduate programs under offer [13]. For this reason, it is clearly useful to take into account the available data relating to enrolled postgraduate students and the information provided by those potential students who have answered the customer knowledge survey "we want to know more about you". Nevertheless, being in possession of this data alone does not generate advantages for the universities [14], and presently, few universities apply Business Analytics to generate competitive advantages [15] from the analysis of such data.

Despite data analysis being an important issue for all universities, given the high cost of academic programs, it is mostly private universities which use it for the design of effective strategies to attract

* Corresponding author.

E-mail address: alejandro.almeida@unir.net (A. Almeida), juanmanuel.martin@unir.net (J.-M. Martín-Álvarez).

new customers [8] and to design new programs that satisfy them more. In this sense, learning objectives and methodologies could be modified adapting them better to the characteristics of the customer.

This article analyses the data generated by the "we want to know more about you" surveys that the International University of La Rioja (UNIR) carried out with potential students some of whom finally enrolled. Attention is not limited to enrolled students, since the determining factors of those who finally decided not to enroll is also interesting. The analysis is carried out through the application of Machine Learning techniques, specifically cluster analysis with mixed data, to detect critical characteristics of both categories of students: those who finally enrolled and those who did not.

Machine Learning techniques applied to customer knowledge surveys allow us to identify three targets which are statistically different. This result is achieved by applying a statistical significance test to compare the proportions of enrolments within the distinct clusters. Target customers, or clusters, have differing enrolment probabilities. We identify the characteristics of these clusters and the probability of enrolment and we find that they belong to identifiable segments of the population. We consider that this find is crucial in the adaptation of marketing strategies and product characteristics to each target.

Our objective is to obtain more and better knowledge of the customer, the potential student, detecting the particular characteristics of those who are more likely to enroll in currently offered postgraduate programs. This knowledge will allow the identification of different targets and allow the marketing department to segment efforts and strategies by better adapting them to each target.

This document is structured as follows: section II describes the methodology undertaken; section III outlines the main results. In section IV there is a discussion of these results. Finally, section V draws conclusions.

II. Methodology

We collect data from customers who enquire about university programs and apply a specific methodology to that data. As part of this process, potential students fill out one of the data sources used in this work: a questionnaire. The other main data source comes from enrolments providing information about which students finally signed up.

Each dataset allows the segmentation of students and highlights differences in the conversion rate of each target cluster, hence moving us towards our objective.

A. Data Collection

To collect information about students, universities can use different methods such as the extraction of information from social networks or via knowledge surveys.

In part, the data used in this study is obtained through a survey carried out among students who are interested in a postgraduate program. The objective of the survey is to advise them on making the best possible choices so as to increase their satisfaction with the program. Importantly, this method allows the university to obtain direct and useful information from students without having to extract information from third parties.

This information was collected from May to October of 2020. A total of 16,272 surveys were filled in. The questionnaire poses 17 questions the responses to which provide data such as, response time in addition to other variables related to individual characteristics. It is important to note that not all the students that voluntarily completed the survey enrolled in one of the postgraduate programs.

The second data collection source is enrolment information: information about which of these students finally enrolled is collected. The availability of this information allows us to know which students are enrolled or not. In other words, we can compute the conversion rate, which is the probability of enrolling on a postgraduate program. Furthermore, we discover the characteristics of the students who enroll or not, which facilitates an evaluation of the appropriateness of degree courses to such

B. Data Preparation

To carry out our analysis we need to consider which data will be useful. First, nine variables that provide valuable information to segment the students were selected. Second, we carry out a data-cleaning process eliminating those students that present many null values or some outliers that may distort the analysis. This process and subsequent analysis are carried out in R, the language and environment for statistical computing, and its integrated development environment RStudio.

After the data cleaning process, the total number of surveys we work with is 11,859 of which 2,073 end up converting and enrolling in a postgraduate degree, that is a conversion rate of 17,48%.

The values taken by each of the selected variables are described below:

- **Age (A)** takes values from 19 to 79 years old, the average age being 35.12.
- **Average grade (AG)** values added in 3 groups. Between 5 and 6.4, between 6.5 and 8.4 and more than 8.5.
- **Working (W)** dichotomist variable reporting YES or NO.
- **Job (J)**: eight different values are used (middle manager, administrative, other company levels, consultant, specialist analytics, upper management and two categories of public workers, A-B and C-D).
- **Professional Experience (PE)** values added in 4 groups. From 1 to 3 years, from 3 to 5 years, more than 5 years or without professional experience.
- **Company size (CS)** values added in 3 groups. Large (more than 250 employees), medium (between 50 and 250 employees) or small (less than 50 employees).
- **Budget (B)** values added in 3 groups. Less than 4500, between 4500 and 5500 or more than 5500.
- **Survey Time (ST)** takes values between 22.75 seconds and 980 seconds with a time average of 146.95.
- **Country (C)**: we find students from 54 different countries, 6 being the most represented (Ecuador, Colombia, Spain, Mexico, Guatemala and Peru).

In addition to these variables, we also incorporate the binary variable **Matriculate (M)** that defines whether the student has enrolled in a program.

The characteristics of those customers who filled in the knowledge survey asking for information about a postgraduate program offered by UNIR are the following:

They are between 26 and 30 years old. Mostly they are from Ecuador (27.83%) but also from Colombia (18.55%), Spain (16.02%), Mexico (15.18%) and Guatemala (7.60%). In the labour market, they cover different strata: middle manager (26.14%), administrative (22.77%), other company levels (16.02%), consultant (11.81%), specialist analytics (10.96%), upper management (9.27%) and public workers (2.52%). 48.06% of them work in big companies (with more than 250 employees), 27.83% in small ones (less than 50 workers) and 24.45% in medium companies. They have more than 5 years in-work experience

(62.40%) and only the 0.84% have no experience. Finally, 51.43% report a high grade in previous studies (more than 8.5 points) and 44.70% a score between 6.5 and 8.5 points. Fig. 1 represents each of the 9 variables described above.
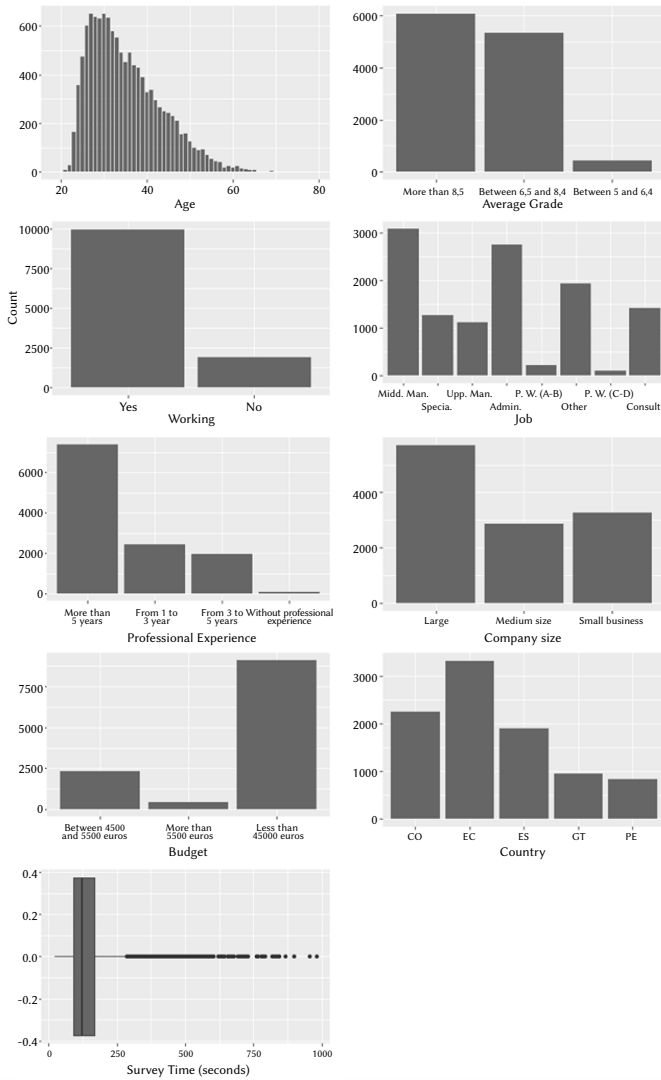

Fig. 1. Representation of the variables used.

C. Clustering Students

In marketing, segmentation is used to identify groups of customers that share homogeneous characteristics. For them, similar products and marketing strategies should be generated [16]-[18]. Among the objectives of segmentation are also the identification of target customers, the definition of the specific characteristics of customer groups, commonly known as the buyer persona [19]-[20].

Machine learning is used to segment students interested in pursuing a program at the university under study. These unsupervised learning techniques are used to gain insights into customers and can create clusters of students with a similar profile. Therefore, we used this approach to segment the potential students at this university into clusters.

Clustering techniques are tools that help to understand the different subgroups that exist within a data set. Specifically, these techniques aim to group the elements that are close enough to each other and far enough from other elements [21]. However, the choice of how to measure the distance between two elements is something

that is not agreed upon in the literature and there are many subjective alternatives. Distance is a numerical measure of how far two individuals are apart, in other words, it measures the proximity or similarity between individuals [22].

The most common way to measure distance is the Euclidean distance, although there are other alternatives such as the Manhattan distance used for particular types of problems. However, our data contains mixed data types (numerical and categorical) where these distances are not applicable, therefore traditional clustering algorithms such as K-means or hierarchical clustering are not valid.

Therefore, for our case study, we use the Gower distance, which is a measure of distance that can be calculated for two individuals whose attributes are mixed. The Gower distance is computed as the average of the dissimilarities between individuals. Each Gower distance lies between [0 1].

$$d(i,j) = \frac{1}{p}\sum_{i=1}^{p} d_{if}^{(f)}$$

The partial dissimilarity $d_{(i,j)}^{(f)}$ depends on the type of variable that we are measuring. In the case of numerical variables, partial dissimilarity is the ratio between the absolute differences between observations and the maximum observed range of all individuals. In the case of categorical variables, the partial dissimilarity is 1 if the observations are different and 0 if not.

The selected clustering algorithm should fit well with the Gower distance. To do this, we select the k-medoids algorithm. The k-medoid algorithm, Partitioning Around Medoids (PAM) is a classic partitioning method similar to the well-known k-means method but, instead of iterating over the centroids, it iterates over the medoids, that is, it tries to find the most representative object for each cluster [21]. The algorithm clusters the objects in a total of k clusters where k must be given a priori. The selection of the optimal number of clusters (k) must be made considering statistical information obtained in the data, although if there is any reasoned justification a priori, the number of clusters may vary for different reasons. To select the optimal number of clusters in which to divide our data we use the silhouette width. The silhouette width is one of the most commonly used options for measuring the similarity between each point in a cluster and compares this similarity with the closest point of the neighbouring cluster. This metric lies between [-1 1] where higher values mean greater similarities [23].

Fig. 2 shows the result of the measurement for values of k between 2 and 10 where it can be observed that segmenting the students into 2 or 3 groups maximizes the similarity within the clusters and the dissimilarity between clusters. We have divided the sample into 2 and 3 groups following the results found in the silhouette analysis, however, using k=3 produces useful results for marketing and academic strategies as shown in the following sections.
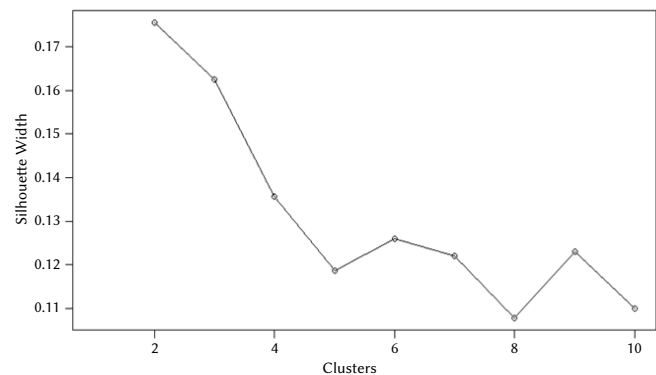

Fig. 2. Optimal number of clusters. Silhouette Width.

III. Results

This section presents the results of the k-medoids clustering technique using the Gower distance. As defined in the Methodology section, the k-medoids machine-learning algorithm for clustering data defines the most representative object of each cluster, that, in our case of study can be defined as the student profile in each cluster. Table I shows the characteristic that defines the profile of each cluster using k-medoids (K=2). Furthermore, along with these characteristics, we present the conversion rate (percentage of respondents who finally enrol in a postgraduate program).

TABLE I. Characteristics of the K-medoids (K=2)

Variable	Student profile	
	Cluster 1	Cluster 2
A	34.35	36.07
AG	>8.5	6.5< x <8.5
W	Yes	Yes
J	Middle-Manager	Clerical
PE	>5 years	>5 years
CS	> 250 empl.	> 250 empl.
B	<4500€	<4500€
ST	148.47	145.07
C	Ecuador	Spain
Conversion Rate[1]	17.52%	17.43%

[1] After performing a statistical test for equality of proportions, no differences were observed in terms of the conversion tare.
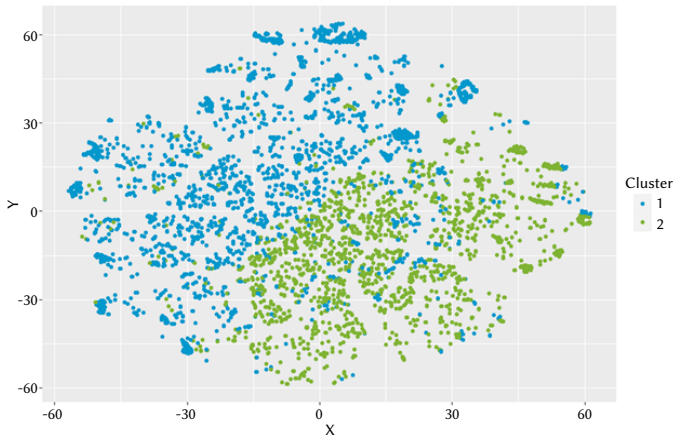

Fig. 3. Cluster plot from the "we want to know more about you" survey identifying customer personal characteristics (k=2).

To graph the division of students with k=2, Fig. 3 uses the t-Distributed Stochastic Neighbour Embedding technique [21], that helped us to visualize our multi-dimensional data into a two-dimensional plot.

Dividing the sample into two groups does not produce very useful results in terms of developing marketing or academic strategies. As can be seen in Table I, although there are differences between both clusters, the conversion rate is the same. This result, together with the small dissimilarity loss as observed in Fig. 3, leads us to perform the clustering with k = 3.

Table II shows the characteristic that defines the profile of each cluster using k-medoids (k=3) and the conversion rate for each cluster. Along with this, Fig. 4 shows in two dimensions the division into three groups of the students surveyed.

TABLE II. Characteristics of the K-medoids (K=3)

Variable	Student profile		
	Cluster 1	Cluster 2	Cluster 3
A	38.35	28.66	36.14
AG	>8.5	>8.5	6.5< x <8.5
W	Yes	Yes	Yes
J	Middle-Manager	Clerical	Middle-Manager
PE	>5 years	<3 years	>5 years
CS	> 250 empl.	<50 empl.	> 250 empl.
B	<4500€	<4500€	<4500€
ST	152.33	143.58	144.08
C	Ecuador	Ecuador	Spain
Conversion Rate[1]	16.24%	16.46%	19.25%

[1] After carrying out a statistical test of equality of proportions, differences are observed in terms of the conversion tare.
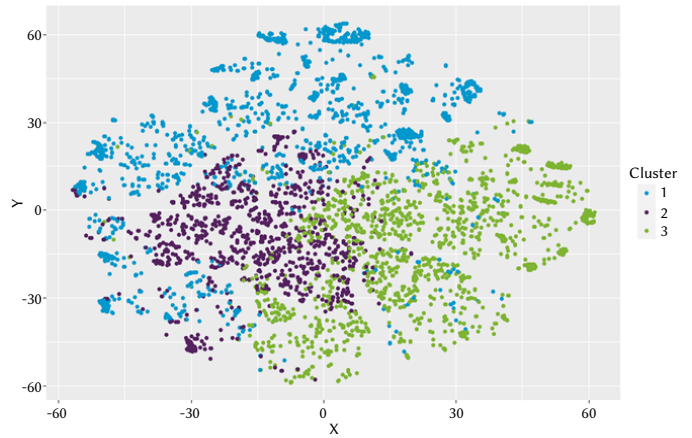

Fig. 4. Cluster plot from the "we want to know more about you" survey identifying customer personal characteristics (k=3).

For a better visualization of our results, we represent in Fig. 5, the variables in which we find differences between clusters that can be considered as the main discriminatory variables between the 3 clusters.

Cluster 1 present the lowest conversion rate. It includes people mostly from Ecuador (43.37%) and they are under 40 years old and working in middle management with more than 5 years' experience (88.41%). The conversion rate for this cluster is 16.24%. The previous mark register is the highest (up to 8 points) for the 95.40% of the customers included in this cluster. Customers included in this cluster can be considered as consolidated in the labour market and their professional career.

Cluster 2 contains the youngest customers with quite a high conversion rate (16.46%), not statistically significantly different from the previous cluster. Customers inside this cluster are from Ecuador (39.86%) and from Mexico (20.14%). They are working in administrative roles with low experience (63.47% have between 1 and 3 years' labour experience).

Cluster 3 contains people from Spain (31.14%) with an average age of 36.14. The conversion rate of this cluster is the highest (19.25%) and following a statistical test to compare its value with values of cluster 1 and 2, we find that it is statistically higher than the proportion registered by previous clusters. They work in big companies as middle managers with more than 5 years' experience (88.41%). The mark registered in previous studies is not high; 94.10% registers a mark between 6.5 and 8.5 points.
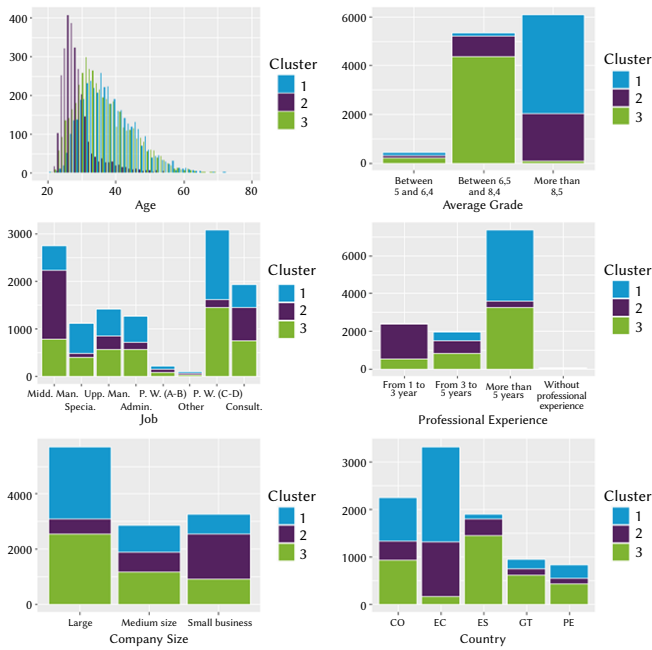
Fig. 5. Graphic representation of the composition of each cluster.

IV. DISCUSSION

In the previous sections of this paper, the application of cluster analysis allowed us to identify similar student profiles at a private university and also to identify those postgraduate programs which each cluster are more likely to enrol in.

Now, we discuss the implications of our analysis from a marketing, business and academic perspective.

From a marketing point of view, it is important to mention the high cost per contact for the marketing and the admission department. Clustering potential students will increase the probability of a successful conversion for each contact thereby allowing these departments to focus on those customers with more probability of enrolling. In short, a direct increase in efficiency.

As indicated in the results section, we see that postgraduate programs can be advertised by targeting three client clusters. This analysis indicates that there is a group of students with a greater propensity for enrolment. Therefore, this first result makes it possible to prioritise the resources available to rank information requests from potential students. Along these lines, it seems reasonable to give response priority to those students with a greater propensity to convert.

When the proportion of students from each cluster enrolled is detailed (Table III), it can be seen how there are clear preferences of some student clusters for some courses. Thus, while the students in cluster 3 seem to be interested in, for example, Program 1 (60.83% of the enrolled students belong to cluster 3), the students in clusters 1 and 2 seem to consider the programs 18 and 19 of interest. Therefore, the promotion of courses must be directed to those profiles that principally have historically enrolled in them.

Moreover, the fact that groups of customers with specific profiles have been identified as more likely to convert will allow departments, with this improved knowledge, to personalise messages and better choose the channels used to reach them. Each contact could be guided towards a specific profile since we also know those programs in which contacts in their cluster have a higher conversion rate.

Companies have strategic objectives that shape the future. In this sense, the results seem to indicate that the university should think about a future in which the student profile of cluster 3 will increase its presence, given the greater propensity to enrol observed.

The results obtained could be improved if more data, for instance, geographic information such as the city of residence, were requested in the "we want to know more about you" surveys. Knowing the city of residence would give us approximate information on the income level of that person, knowing the average salaries of that geographical area. This fact would also allow us to link the message to the pricing strategies, further increasing the success of each contact.

The results achieved in this paper seem to recommend the integration of this cluster analysis in the balanced scorecard used for strategic decision-making. Also, the integration of this algorithm in the Customer Relationship Management (CRM) of the university would help to execute what is stated in this section and improve the efficiency of the marketing department.

Regarding the academic perspective, the results will allow a personalization of programs while they will contribute to the accurate perception of the programs by potential clients. Better customer knowledge will allow the personalization of the service which may lead to greater satisfaction and, as a consequence, to an increase in future enrolment.

The University performs innovation processes to improve their programs, content and methodology. These processes should take into account the information referring to those clusters most likely to choose

In brief, our set of results found three different segments one of which had a significantly higher conversion rate. These results can be helpful in adopting product design strategies and marketing strategies.

Using the segmentation provided for the cluster analysis we compute (Table III) the proportion of enrolment in each program offered. We observe that for all offered programs the conversion rate of cluster 3 is significantly higher for programs 1 to 7 and 10 to 12. However, other programs have a higher propensity to receive students included in cluster 1 (program 8, 9, 13, 15 and 16) and cluster 2 (program 14, 17 y 19).

The fact that we are able to identify those postgraduate programs that convert better for each target is valuable information for many marketing, academic and business purposes.

TABLE III. PROPORTION OF ENROLLMENT IN EACH PROGRAM OFFERED

Program	Student profile		
	Cluster 1	Cluster 2	Cluster 3
Program 1	19,57%	19,57%	**60,87%**
Program 2	28,57%	20,41%	**51,02%**
Program 3	18,64%	31,64%	**49,72%**
Program 4	24,10%	27,18%	**48,72%**
Program 5	30,77%	23,08%	**46,15%**
Program 6	37,93%	17,24%	**44,83%**
Program 7	36,31%	19,27%	**44,41%**
Program 8	**50,00%**	6,25%	43,75%
Program 9	**43,01%**	17,20%	39,78%
Program 10	27,78%	33,33%	**38,89%**
Program 11	36,54%	25,00%	**38,46%**
Program 12	30,77%	30,77%	**38,46%**
Program 13	**38,33%**	25,83%	35,83%
Program 14	25,00%	**40,91%**	34,09%
Program 15	**50,00%**	25,00%	25,00%
Program 16	**54,35%**	21,74%	23,91%
Program 17	29,41%	**47,06%**	23,53%
Program 18	**55,56%**	26,67%	17,78%
Program 19	31,58%	**52,63%**	15,79%

each program-type in order to focus the improvements on the needs of each program's specific cluster. For example, if we want to create or to improve a program like program 1, academics should consider the specific characteristics of cluster 3 containing under 40-year-old Spanish students who do not at a high professional status. Meanwhile, if changes are made in Program 18, then the cluster 1 profile should be considered (professional students mostly from Ecuador).

V. Conclusion

The promotion of postgraduate programs represents an important part in the business management of private universities. Given the volatility of the environment and the fact that some graduate programs are ephemeral, attracting and enrolling new students is a priority in university management. In this context, it is essential to know the preferences of potential students to be able to communicate successfully with them.

The presence of universities on the Internet through web pages and social networks, facilitate communication between universities and their potential clients. Thus, the interactions of potential students with the university through the Internet helps to provide data that reflects preferences according to the profile of each potential student. Consequently, it is possible to observe the characteristics that drive the decision-making of students when enrolling in a particular program.

The detection of common characteristics shared by the students who enrol in the programs facilitates sales optimisation and increasing business efficiency due to the possibility of reallocating resources to profiles with a greater propensity to buy the program on offer. In other words, promotion strategies that incorporate the cluster analysis presented in this paper are more efficient due to the greater precision in the attention to the potential student. In addition, using this tool eliminates unnecessary information that can confuse future consumers. In an ideal scenario, we can generate specific marketing by including this algorithm in the university's CRM. By doing so, each program may be promoted based on the characteristics of those students who most frequently convert to said program. In this article, cluster analysis has been used to group potential students and design business strategies based on the characteristics of each group of those potential students. Specifically, we propose the development of marketing strategies for groups of students considering their propensity to enrol, both globally and at the program level.

Regarding the groups detected, we observe that, although there is a group with a greater propensity to enrol at the global level, the analysis at the program level reveals heterogeneous behaviour. Thus, it is observed how clusters 1, 2 and 3 have more presence in certain programs. This is because, since they are groups made up of homogeneous students, each group has a specific need. Thus, for example, while young profiles have more technological skills and are interested in programs with a high presence of technology, other profiles are consolidated at the work level and need to acquire soft skills, for example.

Although data-driven business management offers itself as a solution to improve the design of promotional strategies, we also observed some problems in our analysis case. These problems are related to data collection. For example, data collection ignores variables related to the potential student's disposable income. In the case of Spanish students, if the postal code were known, it would be possible to obtain the average income of the area in which the potential student resides. Thus, by comparing this data with the price of the program offered, an affordability index could be calculated. It should not be forgotten that in a global context, affordability is a quantity that can better explain conversion than just price. In other words, the degree is not as

affordable for a potential student residing in a region with a low per capita income as one residing in a wealthy region. The strength of the data used is that the registration process is monitored, which makes it easier to follow the student who is interested in a certain program to predict if they will enrol or not. In this sense, this paper is an important tool for data-based business decision-making that generates efficiency in the sale of programs and personalization of the service provided. Innovations or changes in the methodology, pedagogy or contents of a degree should take into account the profile of the cluster that is more likely to be enrolled in each particular program.

References

[1] R. M. C. Croda, D. E. G. Romero, F. R. C. Villar, "The promotion of graduate programs through clustering prospective students," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 6, pp. 23-32, 2019, doi: 10.9781/ijimai.2019.07.001.

[2] C. Dimartino, and S. B. Jessen, "Selling school: The marketing of public education," Teachers College Press, 2018.

[3] M. Coccoli, A. Guercio, P. Maresca, and L Stanganelli, "Smarter universities: A vision for the fast changing digital era," *Journal of Visual Languages & Computing*, vol. 25, no. 6, pp. 1003-1011, 2014.

[4] D-E. Gibaja-Romero, "Interacciones en economías de plataformas," in Perspectivas de la Industria 4.0, AlfaOmega, 2019.

[5] A. S. S. Tota, and M. C. U. Aguirre, "Marketing digital en universidades privadas en el estado Zulia," *Poliantea* vol. 13, no. 24, pp. 5-26, 2017.

[6] M. Yadav, "Social media as a marketing tool: Opportunities and challenges," *Indian Journal of Marketing*, vol. 47, no. 3, pp. 16-28, 2017.

[7] D. Vergidis, and C. Panagiotakopoulos, "Student Dropout at the Hellenic Open University: Evaluation of the Graduate Program, Studies in Education," *The International Review of Research in Open and Distributed Learning*, vol. 3, no. 2, 2002.

[8] D. Airey, and D. P. Stergiou, "Returning to education: A stressful experience?," in *Lifelong Learning for Tourism*, Routledge, pp. 51-69, 2017.

[9] A. Yasmin, S. Tasneem, and K. Fatema, "Effectiveness of digital marketing in the challenging age: An empirical study," *International Journal of Management Science and Business Administration*, vol. 1, no. 5, pp. 69- 80, 2015.

[10] J. Avorn, "Academic detailing: "marketing" the best evidence to clinicians," *Jama*, vol. 317, no. 4, pp. 361-362, 2017.

[11] F. Pucciarelli and A. Kaplan, "Competition and strategy in higher education: Managing complexity and uncertainty," *Business Horizons*, vol. 59, no. 3, pp. 311-320, 2016.

[12] M. E. M. Rosay, "La deserción en el posgrado: estudio comparativo entre maestristas de una universidad pública y privada," in Congresos CLABES, 2017.

[13] G. R. Johnson, C. Jubenville, and B. Goss, "Using institutional selection factors to develop recruiting profiles: Marketing small, private colleges and universities to prospective student athletes," *Journal of Marketing for Higher Education*, vol. 19, no. 1, pp. 1-25, 2009.

[14] R. Pizarro-Milian, "What's for sale at Canadian Universities? A mixed-methods analysis of promotional strategies," *Higher Education Quarterly*, vol. 71, no. 1, pp. 53-74, 2017.

[15] C. Lubienski, "Marketing schools: Consumer goods and competitive incentives for consumer information," *Education and Urban Society*, vol. 40, no. 1, pp. 118-141, 2007.

[16] K.K. Tsiptsis and A. Chorianopoulos, "Data mining techniques in CRM: inside customer segmentation" in John Wiley & Sons. 2011.

[17] C. Marcus, "A practical yet meaningful approach to customer segmentation", *Journal of consumer marketing*, 2008.

[18] R. S. Wu and P. H. Chou, "Customer segmentation of multiple category data in e-commerce using a soft-clustering approach," Electronic Commerce Research and Applications, vol. 10, no. 3, pp. 331-341, 2011.

[19] R. Burkholz, "Entwicklung einer Buyer Persona," in Marketing and Sales Automation (pp. 49-58). Springer Gabler, Wiesbaden, 2017.

[20] A. Revella, "Buyer personas: how to gain insight into your customer's expectations, align your marketing strategies, and win more business," In John Wiley & Sons, 2015.

[21] L. Rokach and O. Maimon, "Clustering methods" in Data mining and knowledge discovery handbook (pp. 321-352). Springer, Boston, MA, 2005.

[22] A. Bhat, "K-medoids clustering using partitioning around medoids for performing face recognition" *International Journal of Soft Computing, Mathematics and Control*, vol. 3, no. 3, pp. 1-12, 2014.

[23] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE" *Journal of machine learning research*, vol. 9, no. 11, 2008.

Eva Asensio Del Arco

Degree and PhD in Economic and Business Sciences from the Complutense University of Madrid (UCM). She has completed a Master in Humanities at the Francisco De Vitoria University (UFV). She taught at the UCM, at the UFV and at the Open University of Catalonia. Furthermore, she has worked for more than 20 years as a consultant in outstanding innovation projects in companies, foundations and public and private organizations. She has extensive teaching experience in public and private universities in the Business area.

Alejandro Almeida Márquez

Graduated in Economics and Business Administration, University of Extremadura and PhD in Economics, International University of Andalusia and the University of Huelva. He developed his thesis on spatial econometric models and has been a researcher at the Spanish Entrepreneurship Research Group of the University of Huelva. He is currently an undergraduate and graduate lecturer and member of the Data Analysis Applied to Economics and Business research group at UNIR.

Aida Galiano Martínez

PhD in Economics with mention of "European Doctor", University of Alicante, 2009. Affiliate research student, University College of London, UK (2007). Master's in quantitative economics from the International Quantitative Economic Doctorate (QeD) program, University of Alicante, 2005. Bachelor of Economics, University of Alicante. 2001. Experience in economic consulting in conducting economic sectorial and impact studies, Economic Strategies and Initiatives, Spin-off of the University of Zaragoza, 2008 – 2012. Associate Professor in Quantitative Analysis in Economics and Business at International University of La Rioja (UNIR). Accredited as Contratado doctor y professor de Universidad privada by ANECA (2018).

Juan Manuel Martín Álvarez

Phd in Economics with training in quantitative analysis for decision making. He has extensive experience as a teacher in public and private universities in the areas of Accounting, Finance, Statistics and Econometrics. He has worked in the development of business intelligence solutions for vending. Associate Professor in Quantitative Analysis in Economics and Business at International University of La Rioja (UNIR). Academic Coordinator of the Master in Business Intelligence at UNIR.