

# The Promotion of Graduate Programs through Clustering Prospective Students

Rosa-María Cantón-Croda\*, Damián-Emilio Gibaja-Romero\*, Fernando-Rey Castillo-Villar

Deanship of Graduate Studies in Engineering and Businesses, UPAEP-University Puebla, Puebla (Mexico)

Received 25 January 2019 | Accepted 10 June 2019 | Published 1 July 2019



## ABSTRACT

The promotion of academic programs, particularly at graduate levels, emerges as a response to market changes. In general, graduate programs are not a first order necessity which makes necessary the right promotion of such programs guarantee the attraction of prospective students, which enroll in some of them, which is essential for the financial sustainability of universities. Notably, the last one is a crucial problem for private universities. In this paper, we analyze the prospective students that enroll in a private to design better promotion strategies by using on data gathered by online sources. Specifically, we use clustering techniques to define marketing strategies based on segments of students. We find that age and city are crucial to promoting graduate programs while marital status and sex does not impact the decision of students in the university that we analyze.

## KEYWORDS

Clustering Methods,  
Decision Making,  
Machine Learning.

DOI: 10.9781/ijimai.2019.07.001

## I. INTRODUCTION

**N**OWADAYS, the Internet's massification simplifies the way that consumers get information about academic programs at any education level. At the same time, online platforms, like social networks and own web-pages, allow that universities interact more directly with the society, and particularly with their prospective students. Through these interactions, universities get and generate data that summarizes students' preferences and needs. Hence, it is possible for universities the alignment of their online marketing strategies to what prospective students need or are interested [2]. For example, Pham et al. [1] observe that universities promote their academic programs through YouTube by considering students targeted segments, i.e., people that search for cuisine tutorials receive the publicity of gastronomic academic programs, while people that see space programs receive information about engineering programs. In a different market, Freeman et al. [10] provide empirical evidence about the positive impact of promoting nutrient-poor food through Facebook; and they observe that the social network offers unique digital marketing tools that increase the engagement of consumers on these products.

In general, we observe that digital platforms are structures that facilitate the sharing of information between consumers and providers of education services, and vice versa, even if no transaction is made; hence, the value of platforms relies on being knowledge intermediaries for all agents, regardless of the market's side where they stand [25]. Consequently, digital platforms are structures that universities can use to disseminate advertising about all their academic programs, even if prospective students do not ask directly for information related to education programs [26].

Although digital platforms facilitate the communication between universities and prospective students, it is essential to remark that the last ones are exposed to advertisements from different products, which exposes them to excess information [7]. Moreover, obtaining a graduate certificate/master's degree/doctoral degree does not represent a first order necessity since people face family obligations or employment commitments when they want to enroll in a graduate program [11]-[12]. Excess of information and other priorities are factors that overshadow graduate programs, i.e., the probability of ignoring graduate programs' advertising is high. Hence, such factors cause a negative impact on the effectiveness of digital promotion strategies [13].

The designing of digital marketing campaigns must identify those features of prospective students that drive their decision to start and admission process, but also enroll in an academic program [6]. In other words, the marketing design must consider data from prospective consumers that become active consumers of the educations service.

The majority of universities are aware of having web pages and social networks profiles to interact with their community (prospective students, students, professors, and the society in general). Also, universities value the importance of storing data generated by the interaction in such platforms [8]. Although the theory emphasizes that owning data is distinct from having relevant information for the planning of marketing strategies [16], not all universities know how to treat information from digital sources [22].

Notably, for private universities, designing effective strategies to promote graduate programs is relevant given the high costs of their academic programs and the fact that such education level represents an opportunity for professional improvement or it is considered a personal achievement [12]. As we mentioned before, it is not a first order necessity that private universities need to understand if they want to attract new students that successfully enroll in the university [4] [6] [17].

In the case of Mexico, private universities face an additional problem. The National Council of Research and Technology

\* Corresponding author.

E-mail addresses: rosamaria.canton@upaep.mx (R. M. Cantón-Croda), damianemilio.gibaja@upaep.mx (D. E. Gibaja-Romero).

(CONACYT, its acronym by its name in Spanish, Consejo Nacional de Ciencia y Tecnología) provides scholarships to study a program that holds a CONACYT's quality certification, no matter if the university that offers them is private or public. This is a public strategy that pursues the generation of high-skilled human capital, but it also generates incentives, among prospective students, to apply to those graduate programs with the CONACYT's certification. Naturally, the largest population of graduate students belongs to public universities [27]. Consequently, private universities should make an additional effort to promote non-certified graduate programs.

We analyze the promotion's dataset of a Mexican private university that offers graduate programs with and without the CONACYT's certification. Mainly, we restrict our attention to the data generated by the university's web page since it includes those prospective students that enroll in a graduate program of the university. We analyze the dataset through machine learning techniques, particularly clustering algorithms to identify the features that drive students' decision making to enroll at a graduate program. From the marketing perspective, the identification of clusters facilitates the development of promotion campaigns by focusing on groups of graduate programs instead of launching a general campaign, which may confuse, or specific campaigns that generate an excess of information. From a technical point of view, this methodology produces better results when data is not generated by a controlled survey; for these datasets, machine learning provides more explicit logical rules for marketers [28]. We observe that age and the city, where prospective students live, are features under we segment graduate programs. Surprisingly, marriage status and sex do not provide any information about the students that enroll in a graduate program of this university. By the previous results, the design of marketing campaigns must consider segments of age and city.

The paper is organized as follows. In section II, we present a brief literature review about how private universities design marketing campaigns; we focus on machine learning to generate data-driven marketing strategies. Part III discusses the data treatment and the clustering method that we implement to partition data. Section IV presents the implementation of the clustering analysis and the main results. In the last section, we offer the conclusions.

## II. LITERATURE REVIEW

Our paper is closely related to the literature that analyzes the selling of educational services. We present a brief review of how marketing and machine learning contribute to promoting universities since our objective is to propose marketing strategies to promote graduate programs.

### A. Marketing Private Educational Services

Marketing schools have called the attention of academics and practitioners since governments, in recent years, pursues high-skilled workers generation boosting competition among universities. Some of these policies facilitate the creation of universities, while others provide financial resources to consolidate research activities. Hence, marketing schools is an area of study that emerges from understanding how universities respond to market changes due to the government's intervention; mainly, how they attract new students under public policies that prioritize strategic economic sectors [9, 29]

In New Orleans, for example, school-choice policies promote hierarchies among universities to benefit advantaged students; i.e., the government provides scholarships to those students that enroll in high ranked universities [21]. Thus, public policy generates an "unfair" competition among universities for the attraction of well-qualified students. As a reactive response, universities invest in marketing strategies to avoid disadvantaged students. Lubienski [30] finds a similar result in Michigan, where the government does not discriminate

among universities. However, schools invest in marketing strategies to attract better-performing students.

Concerning the promotion of private universities, we recall that the successful attraction and retention of students are crucial tasks for financial sustainability since a significant part of their financial resources come from their students' population. Around the globe, whenever private universities do exist, these academic institutions compete against public universities to attract new students since the last ones offer more accessible tuition fees Turner [19]. Hence, graduate programs, offered by private universities, deal with extra pressure to attract new students since getting a graduate certificate is a complementary achievement [11], and financial support is scarce [12].

To guarantee their financial sustainability, private universities invest in marketing campaigns (digital or not) to attract more students. It is worth to mention that the success of this investment depends in increasing the number of prospective students, but also in generating a significant conversion rate (the percentage of prospective student that enroll to some academic program) [23].

The promotion of graduate programs requires that universities transform their data into valuable information to design effective marketing strategies. Moreover, the data analysis must consider that students do not only find the information shared by a university about their graduate programs; socio-psychological characteristics, and economic aspects as well, play a significant role in how they choose a graduate program and complete the enrollment process [20].

Typically, in the designing of marketing campaigns for graduate programs, marketers focus on showing that people with a graduate certificate increase their life's quality. So, marketing strategies usually show where former students work and how their salary increases when they get a graduate certificate [14] [15]. The previous strategy communicates the discharge profile of graduate programs, which indicates if the graduate programs help prospective students to fulfill specific objectives in their personal or professional lives [3] [4]. However, researchers agree that marketing strategies also should communicate the admission profile since prospective students need to know if the program is suitable for them [5]. For example, Turner [19] finds that universities in the United States use racial diversity as a selling point for their services; which means that promotion strategies show a welcoming environment where prospective students are welcome regardless of their socioeconomic features. At the same time, universities show the future gains of studying in a multicultural environment. In other words, admission and discharge profiles are necessary for the development of marketing strategies.

### B. Market Segmentation and Machine Learning

Market segmentation has been regarded as one of the most important strategic elements of marketing [31]. Canhoto et al. [32] refer to this concept as "the practice of grouping customers" as its primary purpose is to divide the market into different groups of consumers and then to target one or more of them with specific marketing tactics. Market segmentation brings multiple benefits to firms such as a full understanding of the market, accurate predictions of consumer behavior, and identification of new market opportunities [33]. Besides, market segmentation leads to a better allocation of financial resources to those consumer groups that the firm can satisfy [31].

However, it seems that market segmentation has lagged the current needs of marketing practitioners and new technological advances [34] [35]. Notably, the digital revolution has transformed the way consumers communicate their interests and needs so that it challenges the traditional conceptualizations of marketing strategies. Regarding market segmentation, there is a need to explore more about new segmentation variables, data analysis techniques, and segmentation models [35]. In this sense, machine learning techniques help to

improve market segmentation using data from different sources, which is a useful capacity to understand how the data “behave” [18].

For firms, the possibility of discovering well-delimited segments may generate what Kumar et al. [36] call “pockets of growth.” The authors assert that dividing the market into segments enables firms to identify lucrative hot spots which may be attended through tailored offers and messages. This customized process may become a sustainable competitive advantage for firms that are competing in dynamic market environments [36]. Unfortunately, there is a substantial gap between the large amounts of information available in the market and the lack of knowledge and skills of marketers to analyze and identify patterns from it [23].

### III. METHODOLOGY

In this paper, we analyze data of prospective students that ask for information about graduate programs in a private university. The dataset comes from the promotion’s webpage of the graduate department.

#### A. Acquiring Data

Universities can get prospective students’ data from different sources, like social networks and specialized survey companies. In this paper, we use the dataset gathered from the university’s web page because universities have total control over the information that they share in their web page, the data that they get by interacting with its community (professors, students, and prospective students). The fact that such dataset belongs to the university represents an advantage over data from third parties to generate effective marketing strategies. Note that third parties summarize what prospective students search for, but it is baffling that such datasets point out those students that enroll in the university.

Consequently, the identification of programs that need, or not, marketing campaigns is not direct. On the contrary, the promotion’s dataset summarizes the features that make appealing a graduate program for a prospective student, and we can observe if a prospective student enrolls or not in a graduate program. In other words, it is easier to identify if a marketing strategy succeeds or not in promoting a specific program.

#### B. Prepare the Data

We build our data by deleting all prospective students that do not enroll in any graduate program from the promotion’s dataset. The dataset includes features like age, marital status, sex, address, and the graduate program where prospective students enroll. We identify that the university offers 21 Ph.D. programs and 25 masters programs in business, engineering, and social sciences. By using the software R, we split the promotion’s dataset into two datasets: the first one, *datasetPHD* only includes data of Ph.D. programs, and the second one *datasetMASTER* summarizes data of masters programs. We remark that we only use the software R to analyze the dataset.

Our primary objective is to segment prospective students to improve the designing of marketing strategies. In the first analysis, we consider the whole *datasetPHD* and *datasetMASTER*. The values that each variable can take are described below

- **Age (A).** From 27 to 76 in the Ph.D.’s dataset, and from 21 to 63 in the masters’ dataset.
- **Marital status (MS).** In both datasets, we have the values of single, married, divorced, and NC (no answer).
- **Sex (S).** In both datasets, we find *male* and *female* values.
- **City (C).** In *datasetMASTER*, we find that students come from 92 cities, while Ph.D. students come from 114 cities. In the appendix, we include all the values that this variable can take at each dataset.
- **Graduate programs (GP).** This is the endogenous variable

of our exercise since our objectives pursue the determination of factors that drive the enrollment of students to an academic program. In the appendix, we present all the graduate programs that university offers.

#### C. Pre-Process Data

Given the values that each variable can take, we decide to convert qualitative features into factors to simplify the analysis. So, we transform qualitative data into quantitative data by using the function **InsFactor** of Software R. Thus, the variables **GP** runs from one to 21 in *datasetPHD*, and **GP** runs from one to 25 in *datasetMASTER*.

Also, we find that eight Ph.D. programs (*Desarrollo Económico y Sectorial Estratégico*, *Desarrollo Económico y Sectorial Estratégico online*, *Ingeniería Mecatrónica*, *Ingeniería Mecatrónica online*, *Logística y Dirección de la Cadena de Suministro*, *Logística y Dirección de la Cadena de Suministro online*, *Planeación Estratégica y Dirección de Tecnología*, *Planeación Estratégica y Dirección de Tecnología online*, and eight masters’ programs (*Dirección y Mercadotecnia online*, *Ingeniería Mecatrónica online*, *Logística y Dirección de la Cadena de Suministro*, *Logística y Dirección de la Cadena de Suministro online*, *Planeación Estratégica y Dirección de Tecnología*, *Planeación Estratégica y Dirección de Tecnología online*, *Psicología del Trabajo y de las Organizaciones*, *Sistemas Integrados de Manufactura y Estrategias de Calidad*) own the CONACYT’s certification. So, we split the previous datasets into two datasets, one that only considers graduate programs with the CONACYT’s certification (*datasetMASTER1* and *datasetPHD1*), and the other includes only graduate programs with no graduate certification (*datasetMASTER2* and *datasetPHD2*).

Finally, we standardized the data since variables take values in different ranges. Through this process, we simplify the comparison between variables.

#### D. Clustering and Segmentation

Segmentation is a marketing technique that serves to identify the right customers of a product. In our case of analysis, we search for determining the characteristics of students that enroll in a particular program and enhance marketing strategies. However, segmentation processes depend on the data that we have in hand. For example, *demographic segmentation* uses data like age, gender, and income, while *psychographic segmentation* generates a partition of customers based on values and lifestyle [24]. Note that the promotion’s dataset allows us to segment prospective students into demographic clusters.

We recall that our dataset comes from the promotion’s web-page [28]. Hence, unsupervised learning techniques are appealing to create clusters of prospective students and characterize students enrolled in a specific program. In other words, we segment the students of this university through clustering techniques.

Although it is possible to relax the segmentation by no creating a partition of students, i.e., the construction of non-exclusive clusters, digital marketing supports the identification of well-delimited segments to avoid overlapping advertisement, and hence, an excess of information [7].

Clustering analysis is a tool to describe the features of students that are “close enough.” However, it is important to mention the lack of a general agreement in the meaning of “close enough,” which implies that results are subjective. Even more, we can find different clusters, even if we use the same dataset since the methods’ application relies on the selection of two key aspects:

1. **The way that we measure the similarity between variables.** Remember, we transform all qualitative features into quantitative data by using the factor function **InsFactor**. It is worth to mention that **city** indicates a spatial feature of prospective students;



however, the factor transformation does not care on how far the prospective student is from the university, instead, the city variable summarizes a student's feature. In other words, we only care in which place the prospective students are, and not in the geographical features of the city where the prospective student lives. By the previous discussion, the classical *Euclidean distance* is an appealing measure of similarity between the **GP** variable and the other variables. So, we compute the dissimilarity level through the following formula

$$d(GP, y) = \sqrt{\sum(GP - y)^2}$$

where  $y \in \{A, C, Ms, S\}$ . As usual, the dissimilarity level runs from zero to infinity, i.e., as the distance between variables increases, they do not share similar features.

2. **The clustering algorithm.** We use cluster analysis to identify groups of prospective students that share similar features when they enroll in an academic program. Since the university pursues the financial sustainability of all its graduate programs, no institutional strategy emphasizes the promotion of a specific graduate program. So, we use agglomerative clustering to identify those academic programs, with a high level of similarity, concerning other variables in the promotion's dataset. We search for clusters through the *k-means algorithm* since we consider that data distribute in a Euclidean space. This algorithm partitions the students' database into *k clusters* in a way that minimizes the within-cluster sum squares.

Despite the subjectivity of the clustering results, it is worth to mention that the mentioned aspects provide clustering methods with the flexibility to design marketing strategies without any bias. This feature arises since the algorithm freely searches for groups of graduate programs with similar characteristics, i.e., no pre-established set, of academic programs, insides in the finding of clusters. In practical terms, the marketer learns the students' behavior from the promotion's dataset.

### E. Exploring the Data

First, it is necessary to determine if our datasets admit the creation of clusters, i.e., if elements in our datasets are "close enough." Hence, we explore the dissimilarity level of these elements.

By exploring the data, we observe that graduate programs concentrate, in proportion, almost the same number of students with some specific marital status or sex. In other words, the probability that someone married enrolls, in some graduate program, is the same for all academic programs; similarly, the probability that a woman enrolls in the *master in Dirección de Organizaciones* is equal to the probability that a woman enrolls in the *masters in Logística y Dirección de la Cadena de Suministro*. We find that previous observation hold for enrolled students in Ph.D. programs. In other words, data distribute almost uniformly when we analyze it by the variables **MS** and **S**.

Consequently, we find a single cluster when we measure the distance between programs by considering the marital status and sex. Based on this observation, the marketing proposal relies on promoting all academic programs at the same time, which compromise the success of the campaign by excess information [7] [13]. Therefore, we focus on clustering our datasets through the variables of **city** and **age**.

## IV. CLUSTERING IMPLEMENTATION

In this section, we first group graduate programs by considering the variable **city** to exemplify the clustering methodology. Finally, we present the results for the variable **age**.

### A. The possibility of Clusters' Generation

First, we use the dissimilarities to check if the datasets admit the possibility two generate different clusters, i.e., we first verify that such

variables do not present a uniformly distributed behavior. In Fig. 1 and 2, we observe the heat map of the dissimilarity matrix for Ph.D.'s and master's programs, respectively, when we try to group them by the **city** of precedence. In both cases, we observe that it is possible to group data since the dissimilarity level is low for blocks of graduate programs (the red color illustrates this feature). Even more, the diagonal of both figures indicate us that graduate programs differ concerning the location where their students come from. In other words, it is possible to group graduate programs by city, but the distribution is not uniform.

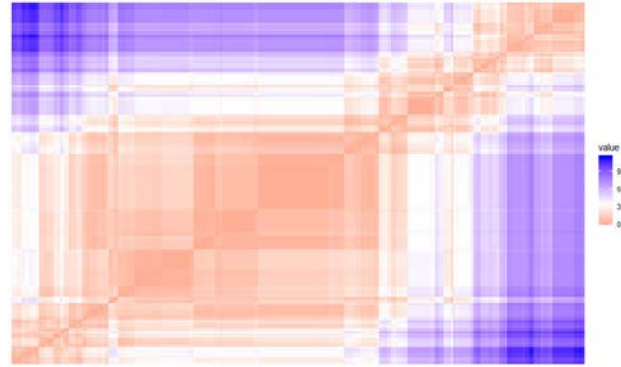


Fig. 1. Dissimilarity's heat map of Ph.D.'s programs concerning variable C.

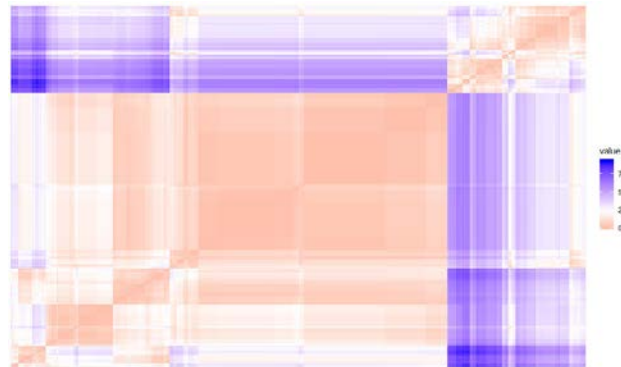


Fig. 2. Dissimilarity's heat map of master's programs for variable C.

### B. The Optimal Number of Clusters

By the previous section, we observe that graduate programs can be grouped when we consider the variable **City**. However, the dissimilarity matrix does not indicate the optimal number of clusters.

Given that clustering is an unsupervised machine learning technique, the number of clusters is a parameter that we need to indicate as an input for our algorithm. In the literature, we can find a broad discussion about the criteria to determine the optimal number of clusters [33] [37]. In general, the criteria that we use to establish such number depends on the algorithm and the distance that we use to group our data [18] [28].

We need to indicate the number of clusters that we want to compute since clustering is an unsupervised machine learning technique. To do it, remember that we use the *k-means algorithm* to group data concerning the Euclidean distance. Under this setting, the number of clusters that we set indicates the number of centroids that the algorithm needs to find. In other words, the algorithm searches for *k* centroids, which correspond to points that minimize the sum of squares within the cluster.

During the *k-means* algorithm, centroids do not only minimize the sum of squares within data elements in a group; the algorithm also pursues the separation with other clusters. Both criteria establish that the sum of intergroup squares must increase when we establish the number of centroids to find. If such sum decreases, the distance between clusters is not enough to separate the data.

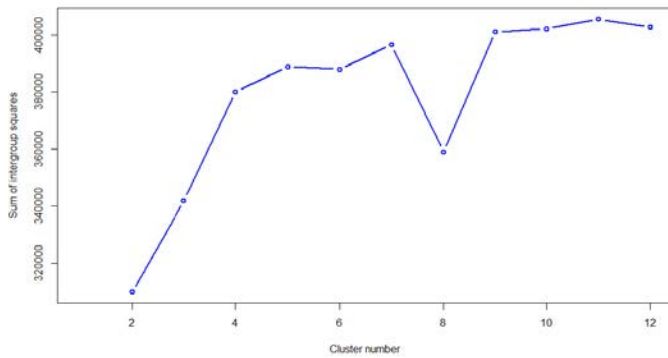


Fig. 3. Sum of intergroup squares concerning the number of clusters for the datasetPHD.

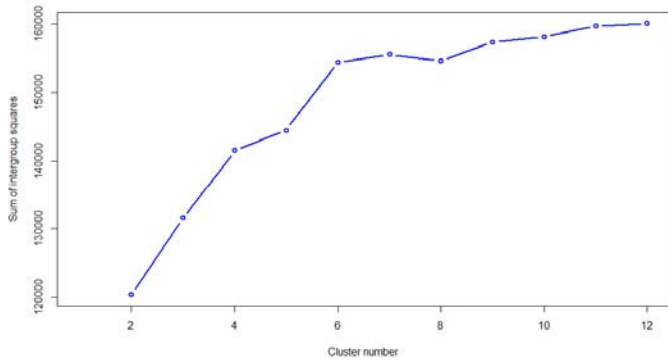


Fig. 4. Sum of intergroup squares with respect to the number of clusters for the datasetMASTER.

For our case, we use this criterion to set the number of  $k$  centroids, which we can find by computing the *sum of intergroup squares* through the instruction `betweens` in the software R. Below, we write the code that we use to compute the sum of intergroup squares iteratively:

```
sumbt1 = kmeans(ProgramCity, centers=1)$betweens
for(i in 2:10) sumbt1[i] = kmeans(ProgramCity, centers = i)$betweens
```

Note that the previous code also establishes the sum of intergroup squares as a function of the number of centroids. We illustrate how the sum of intergroup squares changes when the number of clusters increases in Fig. 3 and Fig. 4 for the *datasetPHD* and *datasetmasters*, respectively.

In the case of doctorate programs, we observe that the sum of squares decreases when the number of clusters increases to eight (see Fig. 3). In other words, under this criterion, the optimal number of clusters to partition the data of doctorate programs is seven. Concerning masters programs, we observe that the sum of squares decreases when the number of clusters is eight. Hence, we choose to compute seven clusters for the *datasetMASTER*.

Also, we verify if masters' data overlap if we consider seven clusters. We use the function `fviz` to verify if clusters are disjoint or not. This function uses the Principal Component Analysis to verify if the  $k$ -means algorithm outputs well-delimited clusters. Although the standardization that such instruction does is not easy to interpret, graphically the function illustrates the region that each cluster cover when we reduce to two dimensions, the dimension of a dataset [38]. Fig. 5 and Fig. 6 illustrate the graphical results of the function `fviz` when we search for seven clusters to group doctorate students and seven clusters to group masters students. When the  $k$ -means algorithm searches seven clusters in both datasets, in Fig. 5 and Fig. 6, we observe that no cluster overlap.

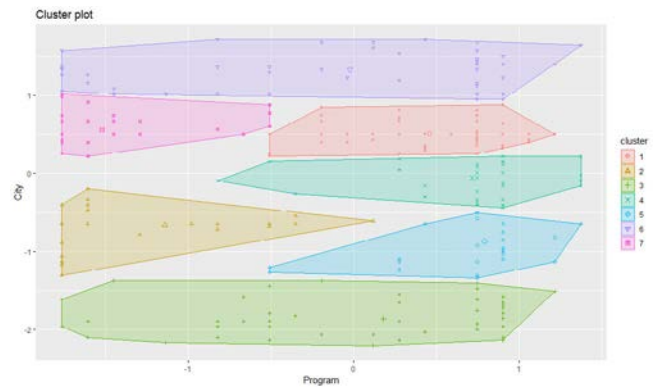


Fig. 5. Partition delimitation of datasetPHD.

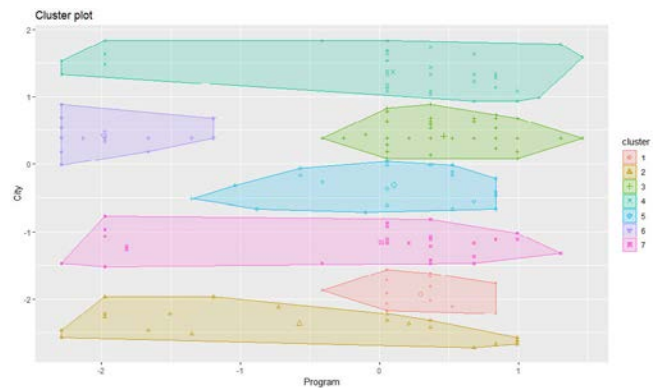


Fig. 6. Partition delimitation of datasetMASTER.

### C. Clusters

In the previous sections, we discuss the possibility to cluster data since the dissimilarity level is low. Moreover, in terms of the distance between clusters, we find that seven is the “optimal” number of clusters that we can find. Finally, the Principal Component Analysis ensures the generation of non-overlapped clusters when we apply the  $k$ -means algorithm.

Table 1 shows the centroids of the seven clusters that the  $k$ -means algorithm finds for Ph.D. programs, while Fig. 7 graphically shows these clusters. Remember, we transform the qualitative information into quantitative information with the `InsFactor` instruction. To simplify the analysis, we rounded the centroids value to the nearest integer; in brackets, we include the program and the city that corresponds to each factor. In this table, we observe that *Mecatrónica* and *Planeación* appear twice. In the case of *Mecatrónica*, factor 12 refers to the face-to-face modality, while factor 13 corresponds to the online modality. Hence, students around (13, 10) are interested in the online modality of the *Mecatrónica* program, while students grouped around (12, 103) searches the face-to-face modality. Concerning the *Planeación* program, this is an element of two centroids given its high demand. So, these clusters indicate the cities where their students live. Specifically, the demand for this program comes around *Apodaca* (6) and *Miguel Hidalgo* (63). Also, we can observe that *Mecatrónica* and *Planeación* appear twice. In the case of *Mecatrónica*, factor 12 refers to the face-to-face modality, while factor 13 corresponds to the online modality. Hence, students around (13, 10) are interested in the online modality of the *Mecatrónica* program, while students grouped around (12, 103) searches the face-to-face modality. Concerning the *Planeación* program, this is an element of two centroids given its high demand. So, these clusters indicate the cities where their students live. Specifically, the demand for this program comes around *Apodaca* (6) and *Miguel Hidalgo* (63).

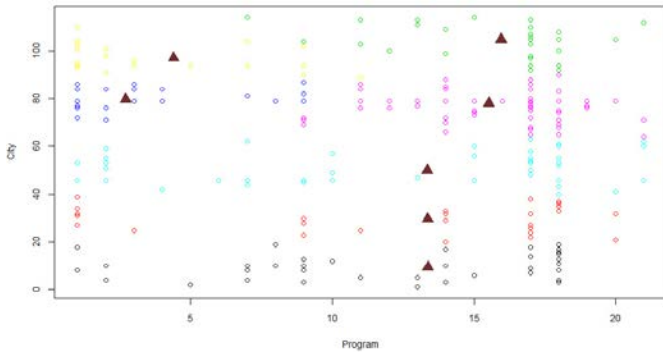


Fig. 7. Cluster visualization of datasetPHD.

TABLE I. THE CENTROIDS OF THE DATASETPHD CLUSTERS

Cluster	Program	City
1	15 (Logística)	79 (Puebla)
2	4 (Finanzas)	45 (Guadalupe)
3	17 (Planeación)	6 (Apodaca)
4	13 (Mecatrónica)	10 (Benito Juárez)
5	17 (Planeación)	63 (Miguel Hidalgo)
6	12 (Mecatrónica)	103 (Tlaxcala)
7	2 (Desarrollo Económico)	80 (Puerto la Cruz)

By data in Table II, we can say that *Logística* master is the one under students feel the most significant interest, and its students come from cities near to *Río Blanco*, *Torreón*, and *Iztacalco*. This calls our attention since *Logística* is not the program with the largest population of students. We can conclude that *Logística* represents a point of attraction for students, but they enroll in an academic program near to it. Notably, we observe that *Mercadotecnia* and *Planeación*, the programs with the largest population of students, also belong to the clusters where *Logística* is an element of the centroid, as Fig. 8 illustrates.

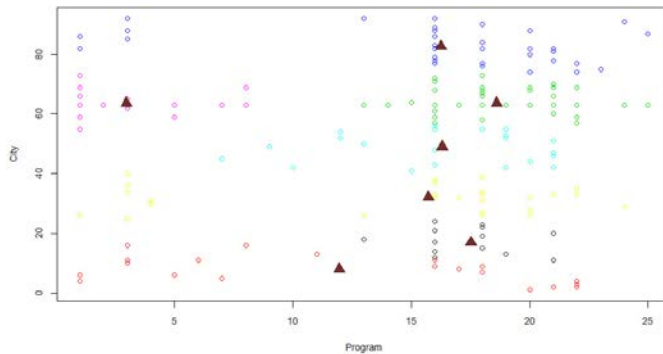


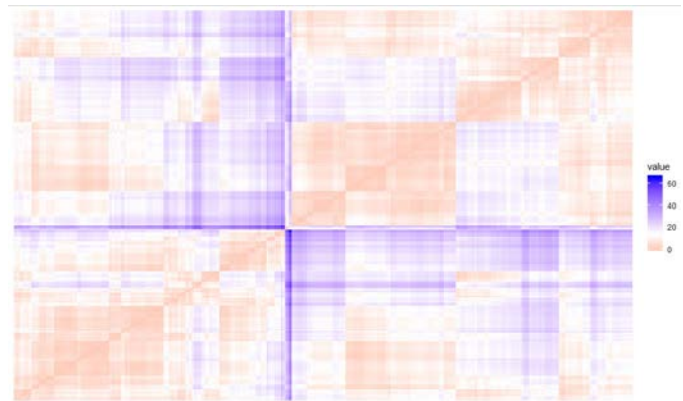
Fig. 8. Cluster visualization of datasetMASTER.

TABLE II. THE CENTROIDS OF THE DATASETMASTER CLUSTERS

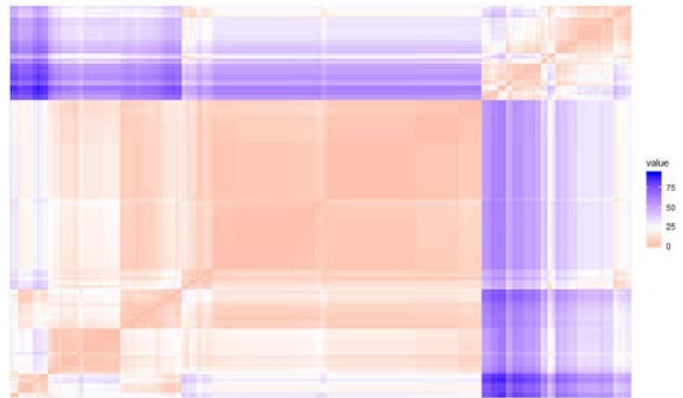
Cluster	Program	City
1	17 (Planeación)	16 (Boca del Río)
2	11 (Ingeniería de Software)	7 (Apan)
3	16 (Logística)	64 (Río Blanco)
4	16 (Logística)	82 (Torreón)
5	16 (Logística)	49 (Iztacalco)
6	2 (Administración de Empresas de Servicio)	63 (Puebla)
7	15 (Mecatrónica)	32 (Cuautla)

#### D. Clustering Data by Age

We apply the previous procedure to support clustering by using the variable *age*. So, we compute the dissimilarity matrix, whose heatmap is illustrated in Fig. 9. In datasets, masters and Ph.D. students, Fig. 9.a and 9.b respectively, it is possible to create clusters when the distance between points is measured concerning the age. It is worth to mention that dissimilarity is more significant than in the city case, which indicates that the variance along with graduate programs increases. In other words, age is a factor that drives the decision to enroll in a graduate program in opposition to marital status.



(a) Heat map of datasetPHD

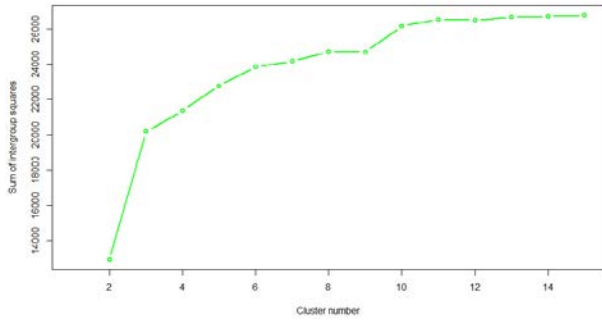


(b) Heat map of datasetMASTERS

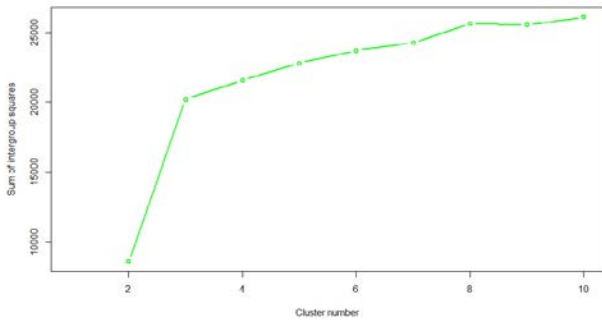
Fig. 9. Heat map of the dissimilarity matrix when we measure distances by age.



In both cases, the dissimilarity level does not establish the optimal number of clusters to compute through the *k*-means algorithm. Hence, we use the criterion of the sum of inter-group squares (see Fig. 10). By Fig. 10.a, we observe that seven Ph.D. clusters contribute to differentiating academic programs concerning age, and Fig. 10.b suggests that eight clusters are the optimal number to group masters' programs.



(a) Number of clusters for datasetPHD



(b) Number of clusters for datasetMASTERS

Fig. 10. The sum of intergroup squares as a function of the number of clusters.

For the *datasetPHD*, Table III indicates the centroids of the seven clusters that groups Ph.D. programs.

Fig. 11 graphically shows the clusters of doctorate programs. By Table III, we observe that *Planeación* is the program that drives the decision of Ph.D. students. This is not surprising since *Planeación* has the CONACYT's certificate. However, it is worth to mention five years separate the centroids of clusters 4, 5, 6, and 7; this is a fact of interest since generations of students have different abilities to interact digitally. Usually, we consider that generations are separated by 10-15 years [13], in opposition to our findings. The previous observation may reflect the fast-changing in digital platforms/networks and accessibility to the internet as well. Finally, we note that older students are interested in the *Tecnologías de Información* program.

TABLE III. CENTROIDS FOR CLUSTERING PH.D. BASED ON AGE

Cluster	Program	Age
1	3 (Organizaciones)	45.11
2	3 (Organizaciones)	34.32
3	21 (Tecnologías de Información)	58.33
4	19 (Manufactura)	46.27
5	18 (Planeación)	40.203
6	18 (Planeación)	31.05
7	16 (Planeación)	35.37

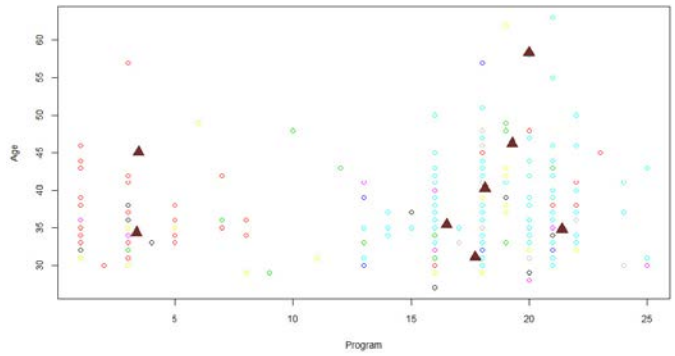


Fig. 11. Clusters of Ph.D. programs by age.

Table IV presents the centroids of the eight clusters that the *k*-means algorithm finds; we can observe them in Fig. 12.

TABLE IV. CENTROIDS FOR CLUSTERING MASTERS' PROGRAMS BASED ON AGE

Cluster	Program	Age
1	3 (Agronegocios)	36.3
2	15 (Mecatrónica)	33
3	20 (Psicología del Trabajo)	42.41
4	21 (Manufactura)	36.32
5	17 (Logística)	37.17
6	18 (Planeación)	58.2
7	16 (Logística)	47
8	19 (Planeación)	31

We observe that *Planeación* y *Logística* are programs that attract the attention of masters' students. Both programs appear twice as an element of the cluster's centroids due to the online and face-to-face modalities. The online modalities correspond to the clusters of this program with the highest age. Also, it is essential to mention that *Agronegocios* is the only program with no CONACYT's certification that appears in Table IV.

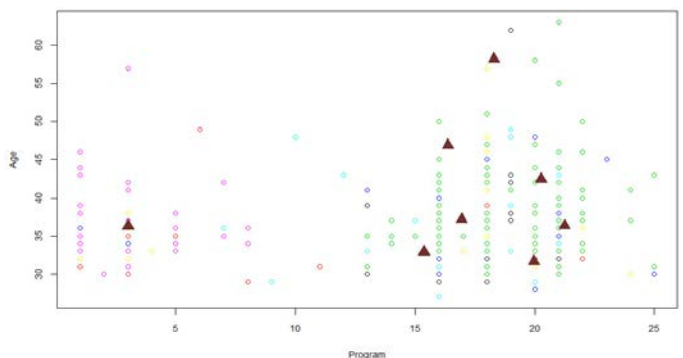


Fig. 12. Clusters of masters' programs by age.

## V. DISCUSSING MARKETING STRATEGIES

In the previous sections, we illustrate the application of clustering methods to find similarities among graduate programs of a private university. Given the high costs of personal marketing and the probability to generate excess information, grouping graduate programs contribute to focus on segments of people interested in similar programs.

### A. Doctorate Recommendations

For Ph.D. programs, we observe that it is possible to promote them by focusing on seven clusters of customers when we group them by age or city. This analysis indicates that a program, the centroid's element, may serve as an attention focus in the promotion of groups of programs.

When we group Ph.D. programs by city, we observe that *Planeación* is the centroid for programs in engineering, and it attracts students from two cities, *Miguel Hidalgo* and *Apodaca*. Hence, promotion in cities where *Planeación* is the centroid must prioritize the marketing on this program. Even more, such marketing strategies serve to attract students for the other Ph.D. programs in the cluster.

Programs with a business/economic orientation are grouped in two different clusters. In the first one, *Finanzas* should drive the marketing strategies, while *Desarrollo Económico* is the focus for the designing of marketing strategies for the other cluster.

In the analysis by age, we observe that programs with the CONACYT's certificate (*Planeación* and *Manufactura*) dominate the formation of the cluster. In other words, Ph.D. students search for CONACYT programs, which is intuitive given the economic incentive that such an institution provides. However, *Organizaciones* is a non-certificated program that attracts two segments of students in generations separated by ten years. In the case of *Planeación*, students' generations separated by five years; this suggests that the university should take care of how it interacts with its students through online platforms. So, it is necessary to understand the online applications that students in each cluster use to communicate with the university. Since the promotion's dataset does not provide details about students' online habits, we suggest asking for such information. For example, the promotion webpage may include a question such as *from which online platform do you get university's information?*

The promotion's dataset does not provide complete information about the labor situation of students enrolled in *Planeación* and *Organizaciones*. Given that *Planeación* is a research program and *Organizaciones* is a business, we conjecture that their students have academic and management duties, respectively.

### B. Masters Recommendations

In the master's dataset, the clustering analysis finds that CONACYT's programs lead the attraction of students since five clusters, of seven, have their centroids in one of these programs when we cluster by the city. Concerning age, we find that seven of eight clusters have their centroid in a program with the CONACYT's certification. Comparing these results with the ones obtained for the Ph.D. case, we observe that the CONACYT's certification drives the decision of students to enroll in a masters' program. Hence, the university must focus their attention on similar programs to *Organizaciones*, *Tecnologías de Información*, and *Agronegocios* to attract students that do not search for the CONACYT's scholarship.

Finally, in the masters' case, it is essential to remark that older students search for the online modality, while younger students are interested in the face-to-face scheme. These observations suggest that older students want flexibility in the way that they take courses, and consequently, marketing strategies must promote such a feature if the promotion's campaigns want to attract older students. On the contrary,

marketing strategies must show the benefits of face-to-face classes to increase the interest of younger students.

Given the importance of the CONACYT's certification, Appendix B presents the same analysis for these programs. In this case, the variable age is the one that contributes to differentiating among CONACYT's programs since the distribution concerning city is almost uniform.

## VI. CONCLUSION

The promotion of graduate education represents a significant problem for marketing designers since the attraction of new students, for these programs, highly depends on economic fluctuations. In other words, the demand for these academic programs is characterized by its volatile since getting a graduate certification is not a first order necessity. Hence, the attraction and enrollment of new students require that universities successfully communicate with their prospective students.

Online platforms, like web-pages and social networks, facilitate the communication between universities and their community. So, online interactions, through these platforms, contribute to gathering data that reflects the preferences and needs of the whole community. Consequently, it is possible to observe the features that drive the decision making of students that enroll in a particular program.

The identification of features that determine how prospective students become students of a program is a strategy that enhances the designing of marketing campaigns. In other words, promotion strategies that incorporate such features are more efficient since they identify the objective public more precisely and eliminate unnecessary information that confuses prospective consumers. In an ideal scenario, we can generate specific marketing to promote each program based on these features. However, personal marketing is too costly and difficult to achieve with the technology in hand. In this paper, we use clustering analysis to group academic programs and design marketing strategies based on the features that diminish the dissimilarities between programs. In other words, we propose the development of marketing strategies for groups of programs b considering how close they are with respect to variables like age and city.

Concerning the age variable, we observe that it is possible to group students in generations separated by five years, which may reflect how fast online interactions change. Typically, generations are separated by 15-20 year, which represents a considerable gap in an environment where online interactions change rapidly. We propose marketing strategies that include the online habits of such prospective students. We conjecture that five years of separation, between generations, is due to the velocity with Internet users change from one application to another. In this sense, we can segment promotion by considering the habits.

Also, we observe that city is a feature that drives the decision to enroll in a graduate program: non-common programs (like Economic Development or Strategic Planning) attract students that live in cities far away of the city where the university that we analyze is located. In opposition, students that live close to the university choose to enroll in more popular academic programs like Management and Mechatronics. Hence, if the university wants to attract students from cities that are not close to the university, the promotion must focus on those programs that distinguish the university concerning other universities. On the contrary, to maintain local demand, the promotion needs to focus on the advantages that the university provides by enrolling in classical/typical/common graduate programs.

Although data-driven marketing appears as a solution to improve the designing of promotion strategies, we also observe some problems



in our case of analysis. Such problems are related to data collection. For example, data collection ignores variables like job status and academic antecedents, which the literature points out as features that drive the academic career of students. Also, the registration process is not watched out, which compromised the quality of the dataset. In a first experiment, we try to do a dynamical analysis, but not all datasets satisfy data mining quality standards. Hence, online marketing must ensure how the data is collected.

APPENDIX

A. Graduate Programs

The university offers **masters programs** in *Administración de Empresas de Servicio, Administración de Empresas de Servicio online, Agronegocios, Agronegocios online, Arquitectura, Ciencia y Gestión Política online, Derecho Corporativo y Bancario, Desarrollo Económico y Sectorial Estratégico, Dirección de Organizaciones online, Dirección y Finanzas online, Dirección y Gerencia de Sistemas de Salud, Dirección y Mercadotecnia online, Dirección y Producción Publicitaria, Diseño y Comunicación Persuasiva, Ingeniería Mecatrónica online, Logística y Dirección de la Cadena de Suministro, Logística y Dirección de la Cadena de Suministro online, Planeación Estratégica y Dirección de Tecnología, Planeación Estratégica y Dirección de Tecnología online, Psicología del Trabajo y de las Organizaciones, Sistemas Integrados de Manufactura y Estrategias de Calidad, Tecnologías de Información y Análisis de Decisiones, Tecnologías de Información y Análisis de Decisiones online, Ventas Estratégicas y Negociación, Vivienda.*

The university offers **PhD programs** in *Desarrollo Económico y Sectorial Estratégico, Desarrollo Económico y Sectorial Estratégico online, Dirección de Organizaciones, Dirección de Organizaciones online, Dirección y Finanzas, Dirección y Finanzas online, Dirección y Mercadotecnia, Dirección y Mercadotecnia online, Ingeniería Biomédica, Ingeniería de Software, Ingeniería Mecatrónica, Ingeniería Mecatrónica online, Logística y Dirección de la Cadena de Suministro, Logística y Dirección de la Cadena de Suministro online, Planeación Estratégica y Dirección de Tecnología, Planeación Estratégica y Dirección de Tecnología online, Sistemas Integrados de Manufactura y Estrategias de Calidad, Tecnologías de Información y Análisis de Decisiones, Tecnologías de Información y Análisis de Decisiones online.*

B. The CONACYT's Certification

Table V presents the centroid of the clusters that we get when we cluster Ph.D. programs by age. Fig. 13 illustrates where the centroids are located when we analyze grouped Ph.D. programs by age.

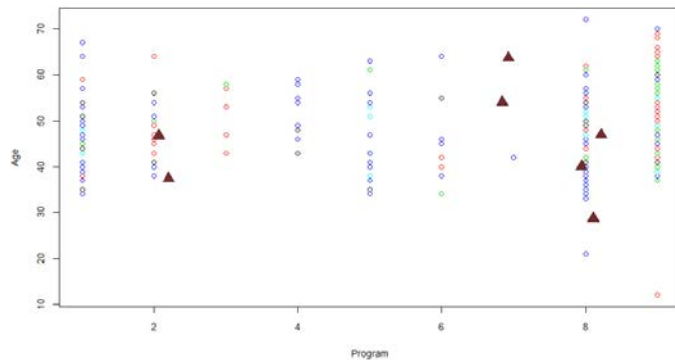


Fig. 13. Clusters of Ph.D. programs with the CONACYT's certification.

TABLE V. CENTROIDS OF CLUSTERS WHEN WE CONSIDER THE CONACYT'S CERTIFICATION

Cluster	Program	Age
1	6 (Logística)	54
2	8 (Planeación)	29
3	2 (Desarrollo Económico)	47
4	2 (Desarrollo Económico)	37
5	7 (Planeación)	40
6	6 (Logística)	64
7	8 (Planeación)	47

Table VI presents the centroid of the clusters that we get when we cluster masters programs by age. Graphically, this situation is presented in Fig. 14.

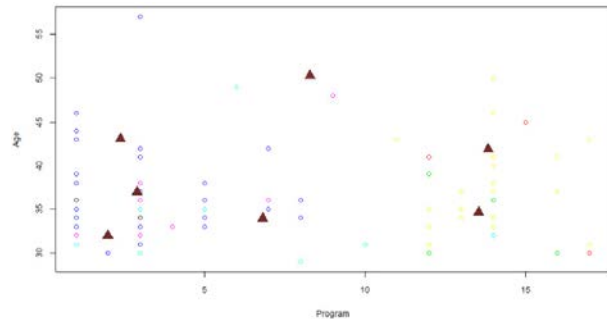


Fig. 14. Clusters (by age) of masters programs with the CONACYT's certification.

TABLE VI. CENTROIDS OF MASTERS PROGRAMS WITH CONACYT CERTIFICATION

Cluster	Program	Age
1	3 (Logística)	36.9
2	13 (Planeación)	41.9
3	8 (Desarrollo Económico)	50.2
4	2 (Desarrollo Económico)	43.07
5	2 (Planeación)	31.9
6	6 (Logística)	33.9
7	13 (Planeación)	34.6

REFERENCES

- [1] H. H. Pham, K. Farrell, H. M. Vu, and Q. H. Vuong, "Using YouTube videos to promote universities: a content analysis," *Working Papers* CEB 17-004, Université Libre de Bruxelles, 2017.
- [2] 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.
- [3] J. R. Posselt, "Inside graduate admissions," in Harvard University Press, 2016.
- [4] J. Avorn, "Academic detailing: "marketing" the best evidence to clinicians," *Jama*, vol. 317, no. 4, pp. 361-362, 2017.
- [5] 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.

- [6] 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.
- [7] M. Yadav, "Social media as a marketing tool: Opportunities and challenges," *Indian Journal of Marketing*, vol. 47, no. 3, pp. 16-28, 2017.
- [8] C. Dimartino, and S. B. Jessen, "Selling school: The marketing of public education," Teachers College Press, 2018.
- [9] 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.
- [10] B. Freeman, B. Kelly, L. Baur, K. Chapman, S. Chapman, T. Gill, and L. King, "Digital junk: food and beverage marketing on Facebook," *American journal of public health*, vol. 104, no. 12, pp. 56-64, 2014.
- [11] 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.
- [12] D. Airey, and D. P. Stergiou, "Returning to education: A stressful experience?," in *Lifelong Learning for Tourism*, Routledge, pp. 51-69, 2017.
- [13] 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.
- [14] K. H. Mok, "Massification of higher education, graduate employment and social mobility in the Greater China region," *British Journal of Sociology of Education*, vol. 37, no. 1, pp. 51-71, 2016.
- [15] Y. J. Lee, and M. Sabharwal, "Education-job match, salary, and job satisfaction across the public, non-profit, and for-profit sectors: Survey of recent college graduates," *Public Management Review*, vol. 18, no. 1, pp. 40-64, 2016.
- [16] 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.
- [17] M. E. M. Rosay, "La deserción en el posgrado: estudio comparativo entre maestrías de una universidad pública y privada," in *Congresos CLABES*, 2017.
- [18] C. X. Ling and C. Li, "Data mining for direct marketing: Problems and solutions," in *Kdd*, vol. 98, pp. 73-79, Aug. 1998.
- [19] E. O. Turner, "Marketing diversity: selling school districts in a racialized marketplace," *Journal of Education Policy*, vol. 33, no. 6, pp. 793-817, 2018.
- [20] A. M. Kalimullin, and S. G. Dobrovorskaya, "Higher Education Marketing Strategies Based on Factors Impacting the Enrollees' Choice of a University and an Academic Program," *International Journal of Environmental and Science Education*, vol. 11, no. 13, pp. 6025-6040, 2016.
- [21] H. Jabbar, "Selling schools: Marketing and recruitment strategies in New Orleans," *Peabody Journal of Education*, vol. 91, no. 1, pp. 4-23, 2016.
- [22] C. Lubienski, "Marketing schools: Consumer goods and competitive incentives for consumer information," *Education and Urban Society*, vol. 40, no. 1, pp. 118-141, 2007.
- [23] C. Kanich, C. Kreibich, K. Levchenko, B. Enright, G. M. Voelker, V. Paxson, and S. Savage, "Spamalytics: An empirical analysis of spam marketing conversion," in *Proceedings of the 15th ACM conference on Computer and communications security*, pp. 3-14, Oct. 2008.
- [24] N. Larsen, "Market segmentation: A framework for determining the right target customers," Aarhus school of business, Denmark, unpublished, 2010.
- [25] D-E. Gibaja-Romero, "Interacciones en economías de plataformas," in *Perspectivas de la Industria 4.0*, AlfaOmega, 2019.
- [26] 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.
- [27] E. Blanco-Livera and F. Martínez-Montes, "Las instituciones de educación superior que imparten programas en México," in *Diagnóstico del Posgrado en México: Nacional*, pp. 28-39, 2015.
- [28] R. Flores-Lopez, and J. M. Ramon-Jeronimo, "Marketing segmentation through machine learning models: An approach based on customer relationship management and customer profitability accounting," *Social Science Computer Review*, vol. 27, no. 1, pp. 96-117, 2009.
- [29] A. Zancajo, "Educational providers in the marketplace: Schools' marketing responses in Chile," *International Journal of Educational Research*, vol. 88, pp. 166-176, 2018.
- [30] C. Lubienski, "Marketing schools: Consumer goods and competitive incentives for consumer information," *Education and Urban Society*, vol. 40, no. 1, pp. 118-141, 2007.
- [31] L. Quinn, T. Hines, and D. Bennis, "Making sense of market segmentation: a fashion retailing case," *European Journal of Marketing*, vol. 41, no. 5/6, pp. 439-465, 2007.
- [32] A. I. Canhoto, M. Clark, and P. Fennemore, "Emerging segmentation practices in the age of the social customer," *Journal of Strategic Marketing*, vol. 21, no. 5, pp. 413-428, 2013.
- [33] M. N. Tuma, R. Decker, and S. W. Scholz, "A survey of the challenges and pitfalls of cluster analysis application in market segmentation," *International Journal of Market Research*, vol. 53, no. 3, pp. 391-414, 2011.
- [34] L. Quinn, and S. Dibb, "Evaluating market-segmentation research priorities: Targeting re-emancipation," *Journal of Marketing Management*, vol. 26, no. (13-14), pp. 1239-1255, 2010.
- [35] S. Dibb, and L. Simkin, "Implementation rules to bridge the theory/practice divide in market segmentation," *Journal of marketing management*, vol. 25, no. (3-4), pp. 375-396, 2009.
- [36] V. Kumar, Z. Rahman, A. Kazmi, and P. Goyal, "Evolution of sustainability as a marketing strategy: Beginning of new era," *Procedia-Social and Behavioral Sciences*, vol. 37, pp. 482-489, 2012.
- [37] A. Ansari, and A. Riasi, "Taxonomy of marketing strategies using bank customers' clustering," *International Journal of Business and Management*, vol. 11, no. 7, pp. 106, 2016.
- [38] C. Ding, and X. He, "K-means clustering via principal component analysis," in *Proceedings of the twenty-first international conference on Machine learning*, p. 29, Jul. 2004.



Rosa María Cantón Croda

Rosa María Cantón Croda, Ph.D. in Computer Science from the Tecnológico de Monterrey, Ciudad de México, Master in Information Technologies from the Tecnológico de Monterrey, Veracruz, was graduated in Computer Science Administrator from the Tecnológico de Monterrey, Monterrey. She was Manager of Systems in transport and construction companies. More than twenty years of experience in the academic administration of Tecnológico de Monterrey and currently Dean of postgraduate studies in Engineering and Business at UPAEP. She has written some articles for national and international congresses. She was certified in PMBook, Positive Psychology by Tecmilenio and in Project-Based Learning by Aalborg University in Denmark. Currently, her research interests are Big Data and Business Intelligence.



Damián Emilio Gibaja Romero

Damián Emilio Gibaja Romero, Ph.D. in Economics from El Colegio de México, Ciudad de México, Master in Economics from El Colegio de México, México. He has written some articles for national and international congresses. He did research visits at the Paris School of Economics and the University of Glasgow. Currently, his research interests are Game Theory and Mathematical Economics.



Fernando Reay Castillo Villar

Fernando Reay Castillo Villar, Ph.D. in Social Sciences, a Master's Degree in Political Analysis, and a Bachelor's Degree in Marketing from Tecnológico de Monterrey (ITESM). During his doctoral studies, he made a stay at the Metropolitan University of Manchester (MMU). Professor in the Graduate Programs of Management and Marketing at UPAEP-University. His areas of interest are brand development as well as the implementation of sociological concepts in Marketing. He has published research papers in international journals.