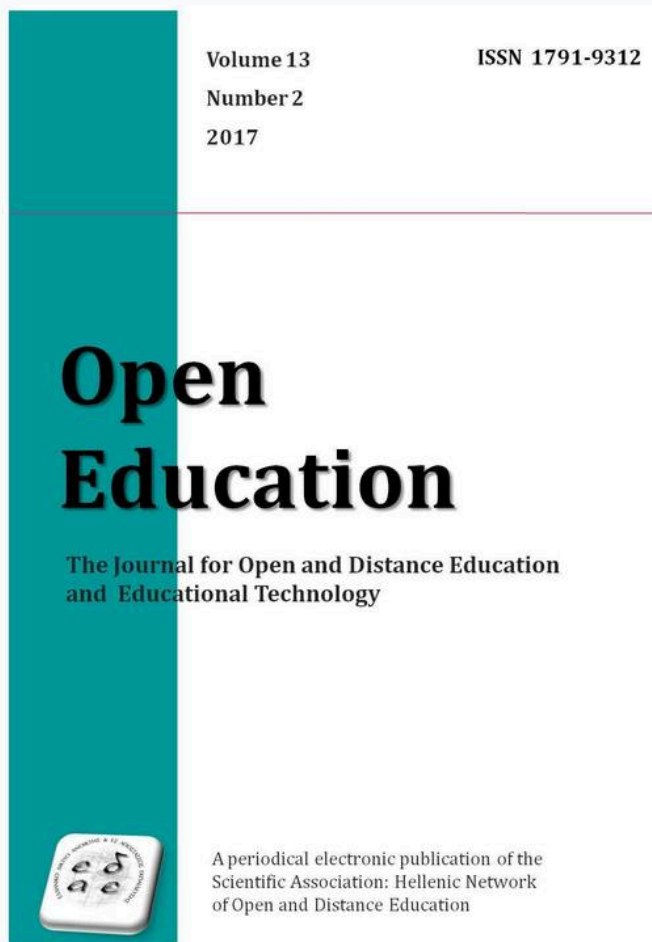


Ανοικτή Εκπαίδευση: το περιοδικό για την Ανοικτή και εξ Αποστάσεως Εκπαίδευση και την Εκπαιδευτική Τεχνολογία

Τόμ. 13, Αρ. 2 (2017)



Student Admission Data Analytics for Open and Distance Education in Greece

Vasileios Kagklis, Antonis Lionarakis, Gerasimos Marketos, Ghristos T. Panagiotakopoulos, Elias C. Stavropoulos, Vassilios S. Verykios

doi: [10.12681/jode.11482](https://doi.org/10.12681/jode.11482)

Βιβλιογραφική αναφορά:

Student Admission Data Analytics for Open and Distance Education in Greece

Ανάλυση Αιτήσεων Φοιτητών για την Ανοικτή και Εξ Αποστάσεως Εκπαίδευση στην Ελλάδα

Vasileios Kagklis

*Hellenic Open University,
Educational Content, Methodology and Technology Laboratory,
MSc, Eng.
kagklis@gmail.com*

Antonis Lionarakis

*Hellenic Open University,
School of Humanities,
Professor
alionar@eap.gr*

Gerasimos Marketos

*Hellenic Open University,
School of Science & Technology
PhD, MSc
gmarketos@gmail.com*

Chris T. Panagiotakopoulos

*University of Patras,
Department of Primary Education,
Professor
cpanag@upatras.gr*

Elias C. Stavropoulos

*Hellenic Open University,
Educational Content, Methodology and Technology Laboratory,
PhD
estavrop@eap.gr*

Vassilios S. Verykios

*Hellenic Open University
School of Science & Technology
Professor
verykios@eap.gr*

Abstract

Over the last few decades, distance learning has become very popular, as a result of the many pros it offers, along with its flexibility. The need for a better understanding of the data originating from such educational environments has led to the rise of the Educational Data Mining research field. However, most of the studies so far focus on the analysis of the data being collected during and/or after the distance learning courses. In this paper, we study the demographical data related to student applications for acceptance in distance learning programs offered by the Hellenic Open University, during the decade from 2003 to 2013. Our study aims at analyzing the data, and discovering patterns and knowledge that can be used to help the strategic placement of the university, and the improvement of the experience that offers to its students.

Moreover, we attempt to correlate the discovered findings with the social and financial status of the applicants' environment.

Keywords

Data analytics, admission demographics, registration applications, distance education

1. Introduction

The widespread use of computers and the Internet contributed in making distance learning popular (Bozkurt et al., 2015). Nowadays, more and more people are willing to enroll in distance learning programs, as they have several attractive features, such as the option of a self-paced study, the freedom to study whenever and wherever you like, and the fact that it does not matter where you live, as you can gain a degree from anywhere in the world. Therefore, there are almost no restrictions, and most of the time their quality is as good as the conventional courses, which require the physical presence of the students in a classroom. What is more, distance learning courses are often cheaper to register.

The Hellenic Open University (HOU) was officially established in 1997 and is the only university in Greece offering exclusively distance learning courses. It consists of four Schools (Humanities, Social Sciences, Science and Technology, and Applied Arts) and offers undergraduate and postgraduate distance learning courses to adult learners. Since its establishment, the HOU has evolved and has been attracting more and more students. People of all ages have shown great interest in studying through a distance learning program offered by the HOU. In Greece, the acceptance of a student in a public university is based on her performance in the Panhellenic exams. A lot of students may fail to obtain the required score or may have to move to a different city in order to study the subject of their choice. Thus, people are interested in the HOU not only due to financial reasons, but also because it provides a second chance towards studying the subject of their choice.

In this paper, we analyze the demographical data originating from applications of prospective students to distance learning programs offered by the HOU. The data used were collected in the period from 2003 to 2013. Our study focuses on discovering patterns and knowledge that can elevate the educational and learning process, either by providing information for targeted promotion of its distance learning programs, or by presenting an evaluation which can be used for improving services and so on.

For this purpose, we utilized features such as the registration year, the region of origin, the age, the gender, the educational level, the graduation grade from previous educational organizations and the graduation year of the applicants. Some of the subjects we study and discuss about are how the number of applications and the performance of the applicants changed during the decade 2003-2013, whether there is any correlation between performance, age, and gender, and whether the region of origin plays a role in the number of applications received.

2. Related Work

Distance education (White, 1982; Byrne, 1989) is an alternative way of learning from a distance, without the need of physical presence in a classroom. In the last decade, it has gained a lot of attention and popularity. In distance education, online discussions are used as a means of communication, as they provide many benefits to students and teachers. According to Tiene (2000), students have been found to be in favor of the self-paced, self-regulated feature of asynchronous discussions compared to face-to-

face discussions. Duffy, Gilbert, Kennedy, and Kwong (2002) report that students who have undertaken a degree by distance learning obtained significantly higher average marks than those who have undertaken it on campus.

The application of data mining in education is an emerging research field known as Educational Data Mining (EDM) (Baker, 2010). It studies the development of methods for exploring data that come from educational environments. The aim of EDM is to provide a better understanding of students' behaviour, to identify how students learn, and how the learning process can be augmented so as to improve the performance of the students (Berland, Baker, & Blikstein, 2014). Each educational problem has a specific goal and its own unique characteristics that require a different way of addressing. Therefore, the knowledge discovery process has to be adapted in the needs of the problem each time.

The problem of predicting low performance or even the possible drop-out of students (Pierrakes, Xenos, Panagiotakopoulos, & Vergidis, 2004; Romero & Ventura, 2010; Pal, 2012) has long been recognized and is one of the hottest topics in EDM. A lot of research has been devoted on how to apply EDM techniques effectively, in order to create models that can predict dropout rates and school failure (Romero & Ventura, 2010). Data mining techniques. Like correlation analysis, regression, classification and decision trees, have been used to evaluate the prediction results for students' academic success (Hämäläinen & Vinni, 2010).

Data from the online discussion fora can be used to apply text mining and natural language processing (Manning & Schütze, 1999), along with social network analysis and sentiment analysis techniques (Pang, Lee, & Vaithyanathan, 2002; Turney, 2002), in order to extract useful knowledge about the behavior of the students, their mood during the course, the collaboration or communication patterns, or even try to predict their final performance based on these findings. In (Wen, Yang, & Rosé, 2014), the authors apply sentiment analysis on students' posts, in order to identify students' opinion for specific features of a Massive Open Online Course (MOOC), and to evaluate if there is a connection between the sentiments and the students drop-out rate.

To the best of our knowledge, most of the existing studies focus on the analysis of the data being collected during and/or after the distance learning courses. In our analysis, we focus on the demographical data of student applications. Our goal is to analyze the data and to discover patterns and knowledge that can be useful to the HOU, mostly for shaping up the admission policy and the planning of student population. Finally, we attempt to correlate the discovered findings with the social and financial status of the applicants and the status of the Greek society.

3. Methodology

3.1 Description of the Dataset

Our dataset consisted of 794,800 student applications, which were submitted towards 83 different distance learning modules, the equivalent of a course, during a 10-year period (2003-2013). The applications were made by a total of 362,311 different applicants, who could apply once a year in multiple modules.

The dataset had initially 43 features, such as the first and the last name, the full address (country, city, zip code, street, and number), etc. However, a lot of features such as the first and last name of the applicants were sensitive and had no potential use or contribution to our analysis. Therefore, we considered only 8 of them, as no useful knowledge could be extracted from the rest of the features. The features we used are summarized in Table 1.

Table 1. List of features used in the analysis

Feature	Description
registration year	The year that the application was submitted
gender	The gender (male/female) of the applicant
age	The current age of the applicant (from birth year to registration year)
region of origin	The region from which the applicant originates
educational level	The educational level of the program (undergraduate/postgraduate) that the candidate applies for
field of study	The field of the program the candidate applies for
grade	The grade of the applicant in his/her previous educational level completed (before making the application)
graduation year	The year of graduation of the applicant in his/her last educational institute or organization

The registration year indicates the year in which the application was submitted. For the age of the applicants, we considered the number of years from the year of birth until the registration year. There were 52 unique, discrete values for the feature "region of origin" of the applicants. For the gender, the available options in the registration form were either male or female. The educational level refers to the educational level of the program that the applicant wants to register in. The educational level could be either undergraduate or postgraduate.

As already mentioned, the applications were submitted towards 83 different distance learning modules, which however cover 33 distinct learning subjects that fall under the four general study fields: Humanities, Social Sciences, Science and Technology, and Applied Arts. Table 2 summarizes the statistics for each one of these general fields of study.

Table 2. Number of modules and distinct learning subjects per field of study

Field	# Learning Subjects	# Offered Modules
Humanities	9	23
Social Sciences	7	16
Science and Technology	14	39
Applied Arts	3	5
Total	33	83

The grade indicates the applicant's performance in his/her previous graduation department or school. More specifically, for undergraduate programs, the graduation grade from the high school (grade of school certificate) was taken into account, and thus grades range from 10 to 20 (grades below 10 mean failure of graduation). Respectively, for postgraduate programs, the grade from the previous graduation institute or organization was taken into account, and thus grades range from 5 to 10 (grades below 5 mean failure of graduation). Lastly, the graduation year refers to the applicant's year of graduation from his/her previous department or school as well.

3.2 Technical Details

The data preprocessing was implemented in Python (Rossum, 1995). Mismatched feature-value pairs were aligned, while most of the rows with missing values were completely removed (approximately 6% of the initial records). The final dataset consisted of 747,714 clean records out of the 794,800 initial records. The data were

stored into an SQL database. The visualizations were created by using Tableau v. 9.3 (<http://www.tableau.com/>).

4. Results and Discussion

Figure 1 presents the total number of applications per gender during that period. The number of applications has its maximum peak in 2004, while its minimum peak occurs in 2012. The decreasing trend during 2010–2012 can be correlated with the financial crisis that stormed in Greece during that period. Moreover, observe that each year the university received more applications from female than male applicants. More specifically, there were totally 160,825 male applicants (44.39%) and 201,486 female applicants (55.61%). The percentage of female applicants was higher by approximately 11 units compared to the percentage of male applicants.

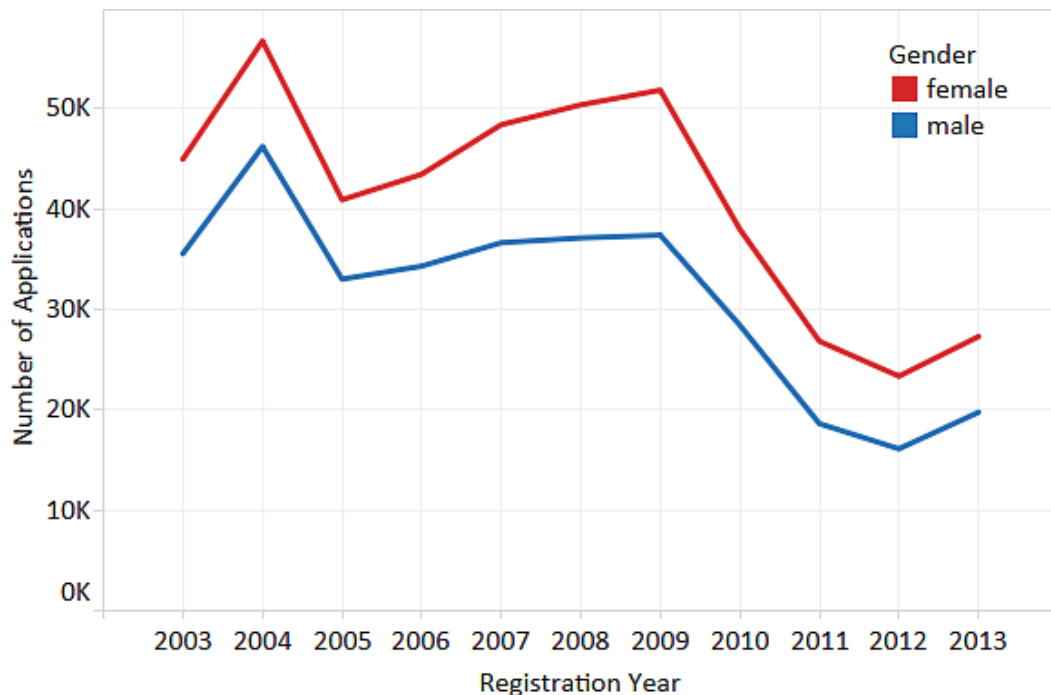


Figure 1. Number of applications per gender during 2003-2013.

The general decreasing trend in the number of applications can also be associated with the fact that after 2005 and until today, no new curricula directly connected with the market were introduced by the HOU. This is a major issue that many Greek universities have not solved yet. Moreover, in 2006 more universities, such as the Open University of Cyprus (2002), started offering competitive distance learning courses in Greek (History of the Open University of Cyprus, 2002).

Figure 2 shows the number of applications for each gender per field of study. There were more female applicants in Humanities, Social Sciences, and Applied Arts, whilst there were more male applicants in Science and Technology. These results can partially confirm the existence of the stereotypes about women and the field of Science and Technology in Greece (Smeding, 2012; Smyth & Nosek, 2015).

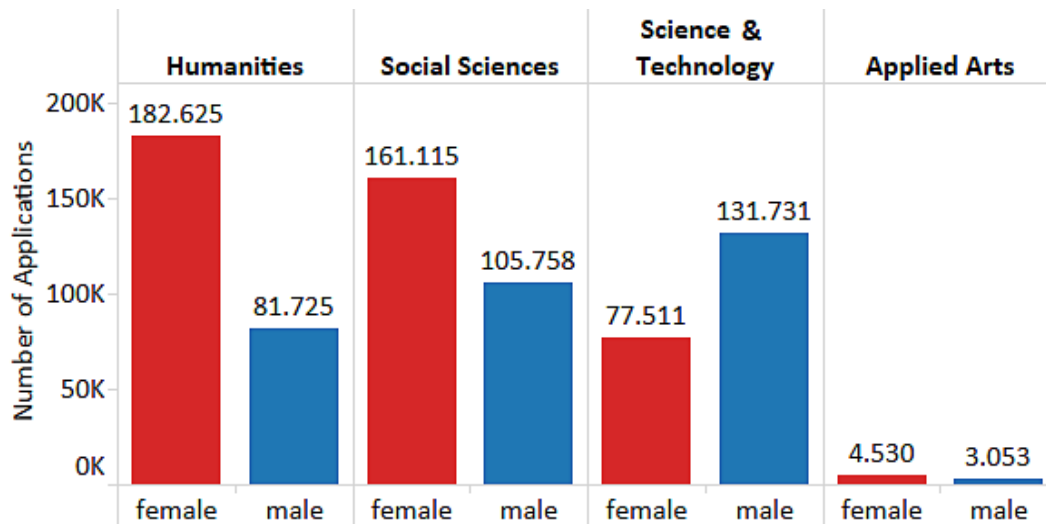


Figure 2. Number of applications for each gender per field of study.

Figure 3 and Figure 4 demonstrate the average grade for each gender, for undergraduate and postgraduate applications, respectively, during 2003–2013. For the undergraduate educational level, the descending trend starts in 2005 for both female and male applicants. Females have a higher average grade than males. For the postgraduate educational level, the descending trend starts in 2005, as well, for both female and male applicants. Again, the females have a higher average grade than the males. Notice, however, that after the minimum peak during 2006, the average grade has an ascending trend again until 2013, for both genders.

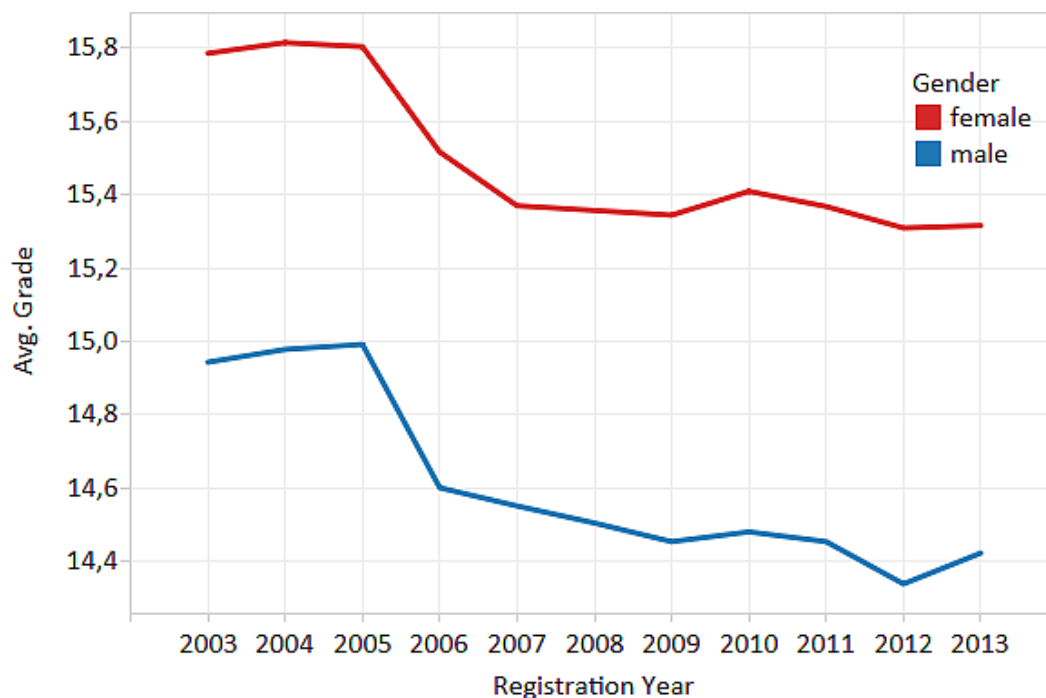


Figure 3. Average grade of male and female applicants to undergraduate programs, during 2003-2013.

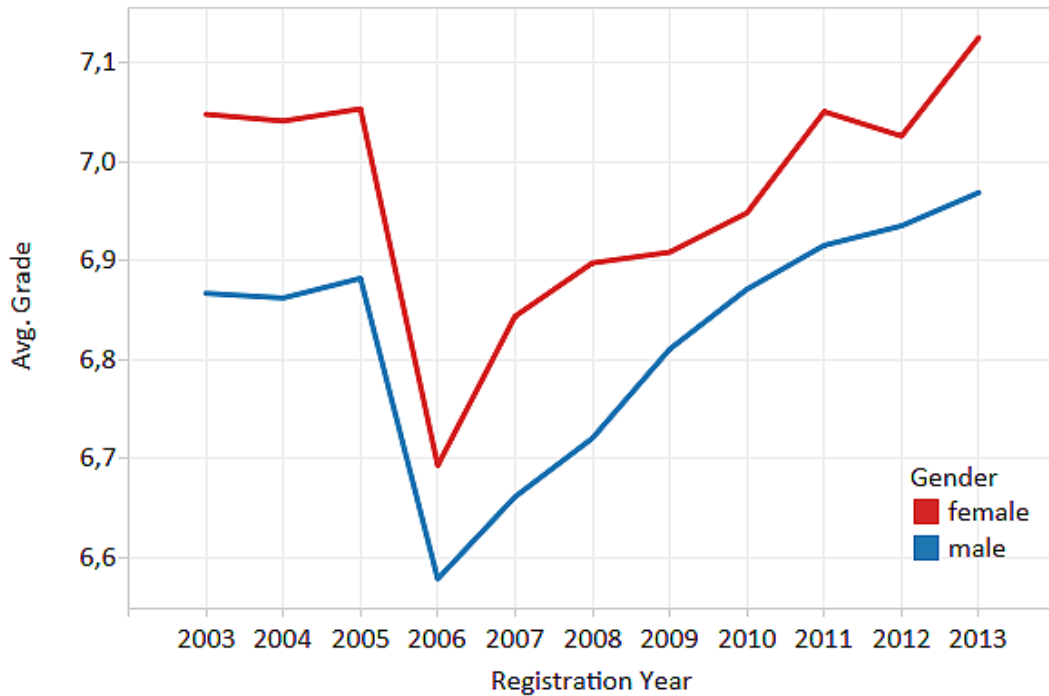


Figure 4. Average grade of male and female applicants to postgraduate programs, during 2003-2013.

The most important factor that impacts the formation of such patterns in the average grade of the applicants is the fact that the HOU is the only institute that does not take into account the grade of the applicants' previous degree or school certificate. The selection of the applicants that will eventually study in the HOU is done through a random, drawing-of-lots process. Therefore, it is a one way out for people with a low grade in their previous degree or their school certificate.

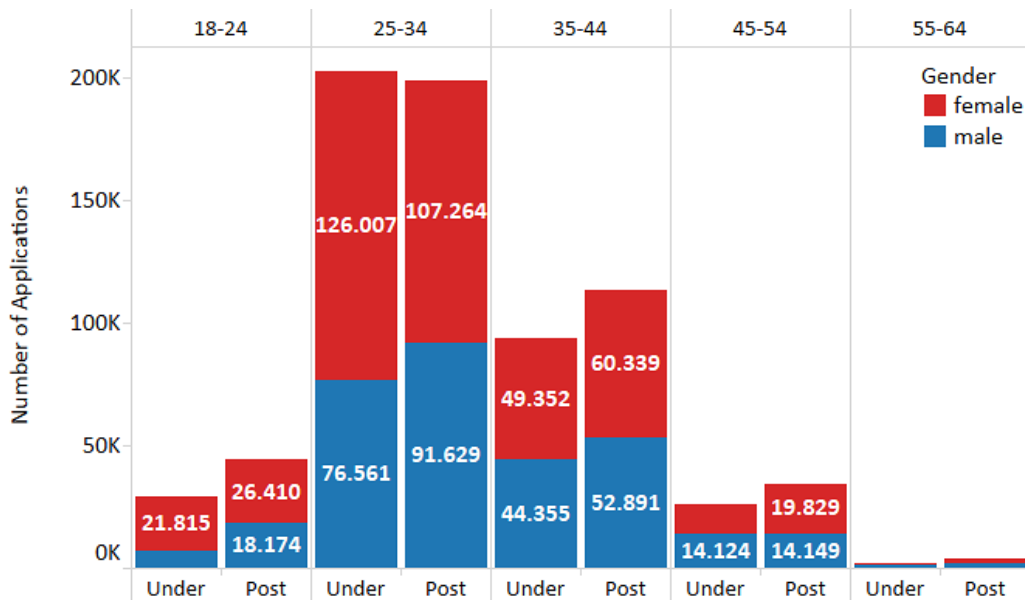


Figure 5. Number of applications per age group and educational level.

Figure 5 displays the number of applications made per age group, in combination with the educational level. We separated the applicants into five groups; 18–24, 25–34, 35–44, 45–54, and 55–64 years old. There was a sixth group, 65+, but the applications were very few (less than 100) compared to the rest of the groups, so we did not

include it in our analysis. We chose to divide ages in these divisions because this setup has been used in demographics by numerous newspapers and advertising executives. Females had shown more interest than the males almost in all age groups. The 25–34 age group held the majority of the applications. This can be explained by the fact that many students enter a public university to study a subject that quite often is not their first option. Then, after graduating or maybe even finding a job, they decide to either study or make postgraduate studies related to the subject that was their first option.

Figure 6 shows the average age of each gender per graduation year, but only for applicants who have graduated from a university or institute, and applied for a postgraduate program. Apart from the expected descending trend of average age as the graduation year increases, we notice that females had almost always been graduating at an earlier age than the males. One could suggest that females put more effort in their studies, they were more eager to learn and were more interested in enrolling in a distance learning program of the HOU than males, regardless of the age or the educational. However, we can overlook the fact that males have the obligation of military service for approximately one year. Female applicants also had a better average grade than the male applicants, in both undergraduate and postgraduate levels.

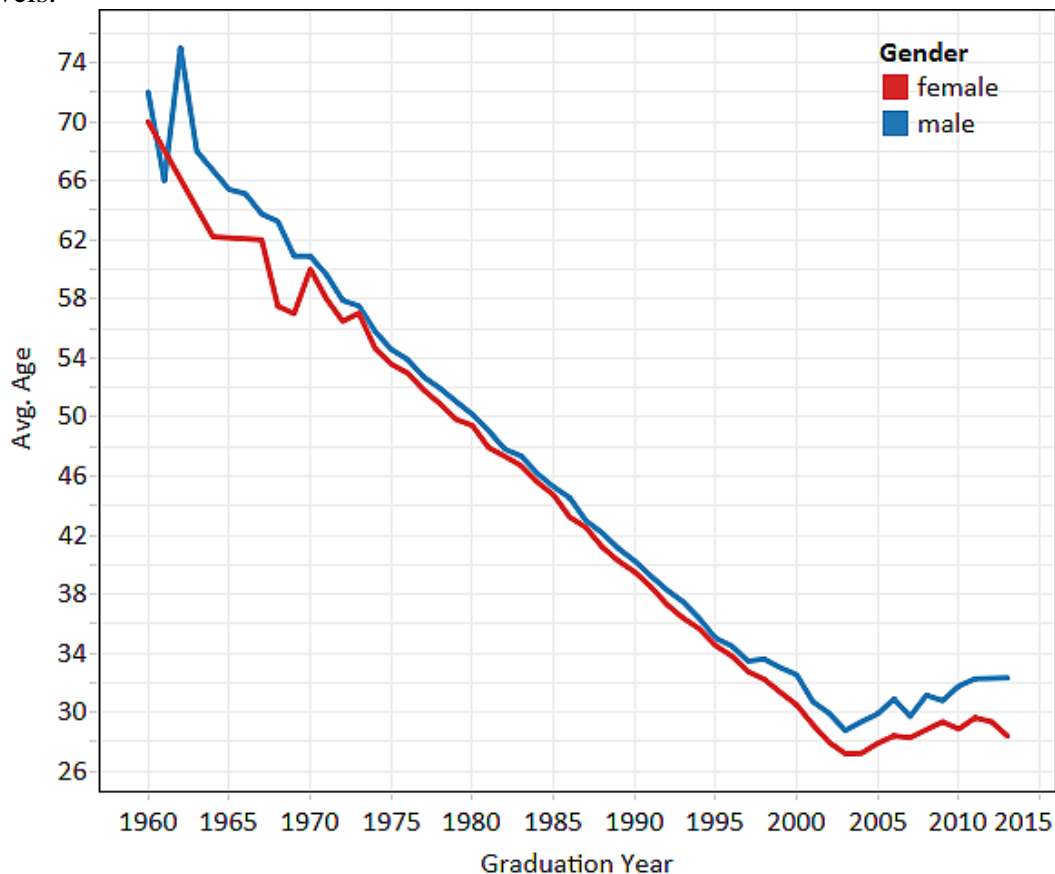


Figure 6. Average age of each gender per graduation year, for people that applied to postgraduate programs.

Another point to observe in Figure 6 is the gradual increase of the average age of graduation from 2010 and on. We suspect that the financial crisis played its role again. Many students were forced to find a job to cover their expenses and to continue their studies, especially those who had moved to a different city in order to study.

Working in a full-time job while studying may increase the time to graduation (Ehrenberg & Sherman, 1987; Orszag, Orszag, & Whitmore, 2001).

Figure 7 demonstrates the distribution of the applicants per region. We scaled the results per 100 residents. The data for the number of residents per region were retrieved from the Eurostat Census Hub¹ (European Statistical System, 2011). The three regions with the largest percentages are the regions of Achaia (ACH, 5.47%), Attiki (ATT, 4.49%), and Thessaloniki (THS, 4.45%), which are marked with their region code. In these three regions, the three largest cities of Greece are located, with their size and number of residents having the following descending order; ATT, THS, ACH. The area with the largest percentage (almost 5.5%) is the region of Achaia (ACH), in which the base of the HOU is located.

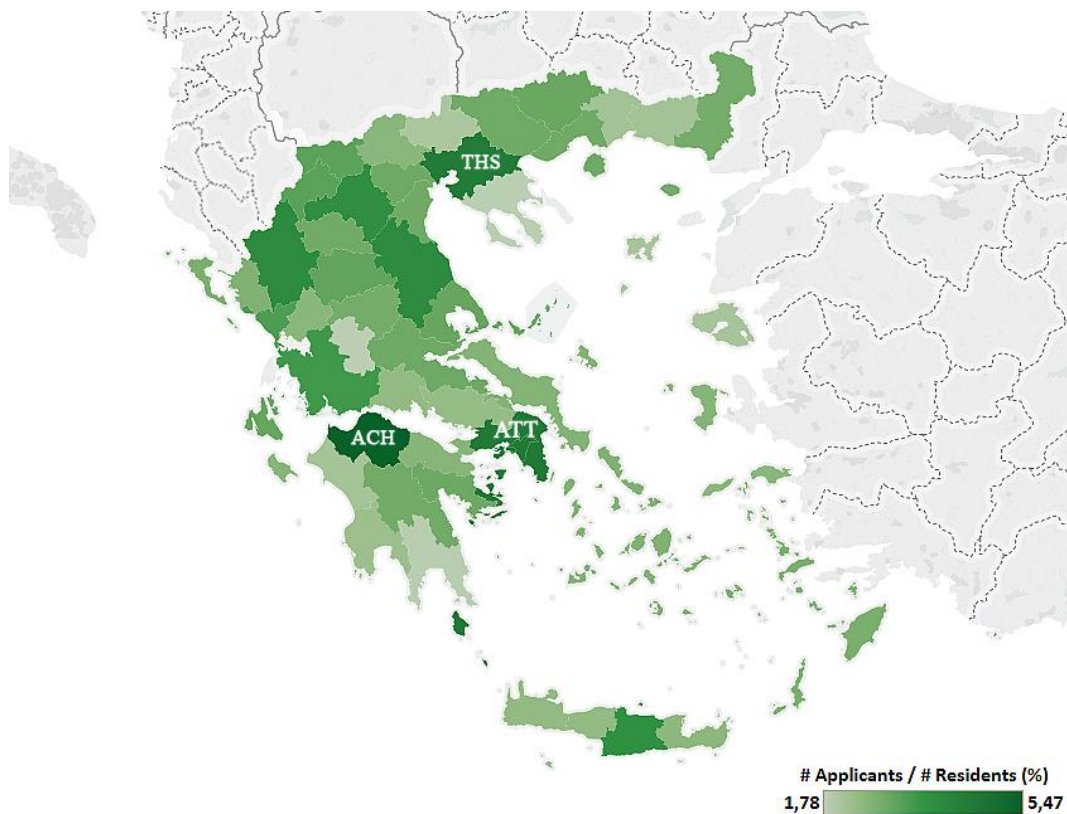


Figure 7. Number of applicants per 100 people for each region.

Remote areas in peripheral regions usually have no access to higher education. Therefore, we expected larger percentages of residents applying to the HOU from such regions. Additionally, despite the fact that regions with large cities do have public universities or other private institutes, in which people can study, they have large percentages as well.

Finally, Figure 8 shows the total number of applications, in log scale, per region, for years 2004, 2009, and 2013. Based on Figure 1, in 2004, 2009, and 2013, the global and local maximum peaks of the number of applications occur. Remote regions such as rural areas, islands, or regions with no universities become red as we move from 2004 to 2013, meaning that less applications were made from these areas. On the other hand, regions with large cities retained their number of applications. Thus, regions with large cities offer the majority of the applications.

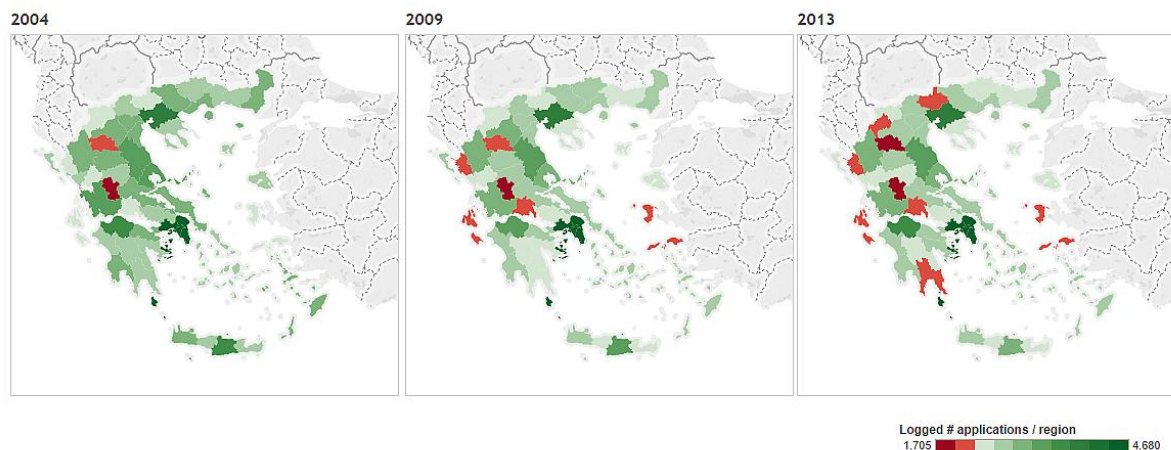


Figure 8. Number of applications in log scale per region, for years 2004, 2009, and 2013.

5. Conclusions & Future Work

In this paper, we analyzed the demographical data originating from applications of prospective students to distance learning programs offered by the HOU. The data we used are from the period 2003-2013. We utilized features such as the registration year, region of origin, the age, the gender, the educational level, the graduation grade from previous educational organizations and the graduation year of the applicants.

External factors (financial and societal) affected the number of applications received. The results indicated that more females applied to the HOU than males, which means that more females wanted to study the subject of their choice in a second-chance organization like the HOU. Moreover, females finished their previous studies approximately one year earlier on average than the male applicants – the same time as males' military service. Lastly, the vast majority of the applications came from the regions with large cities.

Thus, the HOU should aim at creating new, competitive curricula that will meet the needs of the modern market. Additionally, proper advertising should take place and focus on keeping the number of applicants from regions with large cities, while increasing the percentage of residents from peripheral regions applying to the HOU.

As a future work, we are going to use the graduation data (from the HOU) of the accepted applicants, and examine whether there is a correlation between the students' previous performance and their progress while studying in the HOU.

References

- Baker, R. S. (2010). Data mining for education. *International Encyclopedia of Education (3rd edition)*, 7, 112–118.
- Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning*, 19 (1–2), 205–220.
- Bozkurt, A., Akgun-Ozbek, E., Onrat-Yilmazer, S., Erdogdu, E., Ucar, H., Guler, E., Sezgin, S., Karadeniz, A., Sen, N., Goksel-Canbek, N., Dincer, G. D., Ari, S., & Aydin, C. H. (2015). Trends in distance education research: A content analysis of journals 2009-2013. *International Review of Research in Open and Distributed Learning*, 16 (1), 330–363.
- Byrne, T. C. (1989). The evolution of distance education. Calgary, Alberta: University of Calgary Press, 3 (2), 135.
- Duffy, T., Gilbert, I., Kennedy, D., & Kwong, P. W. (2002). Comparing distance education and conventional education: Observations from a comparative study of post-registration nurses. *Association for Learning Technology Journal*, 10 (1), 70–82.

- Ehrenberg, R. G., & Sherman, D. R. (1987). Employment while in college, academic achievement, and post-college outcomes: A summary of results. *The Journal of Human Resources*, 22 (1), 1–23.
- European Statistical System. Census Hub. (2011). <https://ec.europa.eu/CensusHub2/query.do?step=selectHyperCube&qhc=false>
- Hämäläinen, W., & Vinni, M. (2010). Classifiers for educational data mining. *Handbook of Educational Data Mining Chapman & Hall/CRC Data Mining and Knowledge Discovery Series*, 57-74.
- Manning, C., & Schütze, H. (1999). Foundations of statistical natural language processing. MIT Press.
- Open University of Cyprus. (2002). Retrieved from <http://www.ouc.ac.cy/>
- Open University of Cyprus. (2002). History of the Open University of Cyprus. Retrieved from <http://www.ouc.ac.cy/web/guest/university/genika/history>
- Orszag, J. M., Orszag, P. R., & Whitmore, D. M. (2001). Learning and earning: Working in college. Upromise.
- Pal S. (2012). Educational data mining and learning analytics: Applications to constructionist research. *International Journal of Information Engineering and Electronic Business (IJIEEB)*, 4 (2), 1–7.
- Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? sentiment classification using machine learning techniques. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- Pierrakeas, C., Xenos, M., Panagiotakopoulos, C., & Vergidis, D. (2004). A comparative study of dropout rates and causes for two different distance education courses. *International Review of Research in Open and Distance Learning (IRRODL)*, 5 (2), 117–131.
- Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, 40 (6), 601–618.
- Rossum, G. (1995). Python reference manual.
- Smeding, A. (2012). Women in science, technology, engineering, and mathematics (stem): An investigation of their implicit gender stereotypes and stereotypes' connectedness to math performance. *Sex Roles*, 67 (11-12), 617–629. doi: <http://dx.doi.org/10.1007/s11199-012-0209-4>
- Smyth, F. L., & Nosek, B. A. (2015). On the gender-science stereotypes held by scientists: explicit accord with gender-ratios, implicit accord with scientific identity. *Frontiers in Psychology*, 6.
- Tiene D. (2000). Online discussion: A survey of advantages and disadvantages compared to face-to-face discussions. *Journal of Educational Multimedia and Hypermedia*, 9 (4), 371-384.
- Turney, P. (2002). Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. *Proceedings of the Association for Computational Linguistics*, 417-424.
- Wen, M., Yang, D., & Rosé, C. P. (2014). Sentiment analysis in MOOC discussion forums: What does it tell us? *Educational Data Mining*.
- White, M. (1982). Distance education in Australian higher education a history. *Distance Education*, 3 (2), 255–278.

Notes

¹ In most of the European countries, a census takes place every 10 years. The data are from the last census, which took place in 2011 and covers the period 2001-2011.