

What software is in almost every desktop pc? MS Excel: Academic analytics of historical retention patterns for decision-making using customized software for Excel

Dr. Mérida C. Mercado

Inter American University of Puerto Rico
Arecibo Campus
memercado@arecibo.inter.edu

Prof. Nicolás Ramos

Inter American University of Puerto Rico
Arecibo Campus
niramos@arecibo.inter.edu

Abstract - Analysis of historical patterns in student data, from admission to graduation, in the institutional database provides information for understanding retention and developed metrics for evaluating the effectiveness of interventions. Inter American University of Puerto Rico, Arecibo Campus, has developed a strategy for integrating academic analytics using Excel. A faculty driven initiative took advantage of existing resources since costly specialized software was not available to analyze data of cohorts from 1995 to the present. The customized software (ERDU 2.0) transforms data into accessible information (interactive worksheets - tables and graphs) for academic, student services and administrative personnel. This session will outline the critical steps for integrating analytics for decision-making using existing computing infrastructure and developing customized software for Excel. Institutions will identify how to adopt this model to their own technological infrastructure for a more data-driven decision making in retention. Individuals will learn how to: involve people from across campus in evidence based management when the customized software removes technological barriers of the user and evaluate their own process to turn data into strategic action.

Introduction

As faculty collaborating in campus discussions related to retention, we found that evidence to support an approach described in the literature or to challenge its applicability to our institutional circumstances was not easily available. In many instances, the discussions about characteristics student's attrition is based in anecdotal evidence. Also, other considerations like politics, personalities, experiences, beliefs or tradition, in many circumstance, prevail in the decision making process. We identified a need for evidence about student performance and focused on the student data base as the fundamental source of information to describe and research student's academic and socioeconomic background as well as the factors of student's performance associated with retention. Caison (2007) found that the independent variables drawn from institutional databases out-perform variables drawn from the institutional integration survey scale developed by Pascarella and Terenzini in 1980.

Institutional Background

Inter American University of Puerto Rico (IAUPR), established in 1912, is the largest private institution of higher learning in the island of Puerto Rico. It has nine campuses and two professional schools. The units of this multi campus system are distributed throughout the island and serve both large urban populations and students who reside in small communities or in rural areas. The university system has maintained a central information system since August 1995 for managing student data for all campuses and this centralized system provides the foundation for this project. The data stored for all campuses using the same information structure creates a database that can be mined and analyzed for the identification of historical patterns associated with retention by cohort and campus.

Institutional student data, from admission to graduation, from Fall 1995 to the present by semester, provides value information on historical patterns for understanding if student retention behavior exhibit consistent characteristic or if there are variations by cohort or according to other demographic, academic or performance characteristics recorded in the database. Software tools for analytics designed to convert data into information relevant to discussions are available but most of it are integrated into costly specialized software that requires trained and dedicated technical personnel. The cost of this technology is a barrier to the use of information by those with a need relevant data and a small budget. These tools can also be rejected by users if they challenge their quantitative and technological skills beyond their expertise or if they depend on the availability of technical support. As faculty members at Inter American University of Puerto Rico - Arecibo Campus (IAUPR-AC) we convinced campus authorities to provide support a project which main goal was identify patterns in admission data of students. Since 2006, we have researched and developed two customized software the first one for the quantitative descriptions the profiles of high schools and academic programs based on admission data and the second one for analyzing historical patterns related to academic performance and retention for student cohorts admitted to the Campus since 1995. With the sponsorship of the Vice Presidency for Academic and Student Affairs and Systemic Planning of the IAUPR and our campus we initiated a collaborative project with other five campus of the system, the research and development process was extended to analyze retention patterns for those campus derived from it particularly data base by cohort from 1995 to the present.

The study and research of student retention in universities has a long tradition in the United States (Tinto, 1973; Astin, 1975; Pascarella, & Terenzini, 1983; Cabrera, Castaneda, Nora & Hengstler, 1992). In Puerto Rico, until now, the studies and research on the subject retention have been directed mainly to perceptions and few are oriented to descriptive analysis on evidence base of the profile of students entering universities (Centro de Investigaciones Comerciales e Iniciativas Académicas, 2004; Ramos, Mercado & Rosa, 2008). On the other hand, plans to retention at the universities of Puerto Rico, thanks to various federal proposals have focused on facilitating the transition from high school student to college and, especially, to support first-generation students achieve success in college. Moreover, these proposals are oriented to service and not for research, much less, the development of computer applications to facilitate research for the description, analysis, modeling and prediction of retention. Our project was designed to make effective use of the data and computing resources available to faculty and staff at the university. The project uses MS Office, in particular the capabilities of Excel for data mining, statistical analysis and representation of results that can be transferred to word processing and presentation applications. The customized software for Excel (ERDU 2.0) transforms data into accessible information and results for academic management, student services and administrative personnel who have basic skills in productivity software. The software, programmed in Visual Basic, allows the user to generate Excel interactive worksheets with updated information relevant to their role. The information is presented in tables and graphs in Excel for use as presented or for further analysis with Excel functions or with specialized statistical software.

Learning Objectives

The tutorial has the following objectives:

- Describe how the need for information led to the development of a process using existing PCs and software that allows for research and provides decision makers with a tool for evidence based management of retention
- Explain the process of data retrieval from the institutional database provides for research using data mining and statistical techniques to describe retention patterns for cohorts from August 1995 to the present
- Describe how ERDU 2.0, allows for the reproduction of significant patterns and analysis so that decision-makers using a Windows PC with Excel can reproduce the analysis for subsequent semesters

- Illustrate how users can use the capabilities of the customized software for Excel (ERDU 2.0) to answer questions according to their needs and can integrate the information into reports prepared in Word or PowerPoint

Data retrieval, data mining, answering research questions and modeling retention

The project was developed as an empirical process of testing and refinements but as it has matured it is characterized by a process that starts with the retrieval of data warehoused in the institutional database for all cohorts admitted to a campus since August 1995. The data obtained follows each student and obtains data about the student for every semester that has passed from first enrollment to the present. Thus, we can identify enrollment patterns of individuals in cohorts including those who were not retained, those who have graduated and who have enrolled for graduate studies, and stop outs including those who may have returned many years later. The variables obtained from the database include those who are invariants for the student like academic data obtained at admission and variables that vary according to student performance each semester like GPA and credit earned.

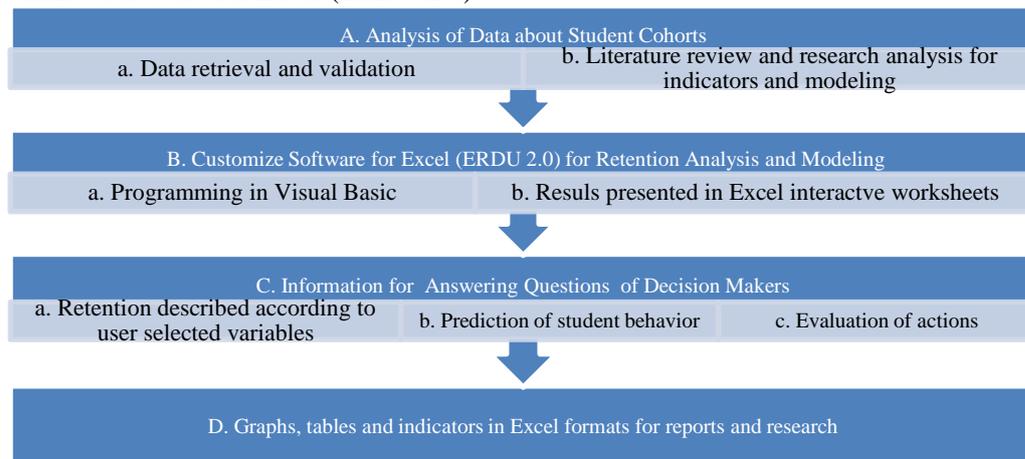
The queries are developed with the aid of personnel responsible with the management of the institutional database. They are dynamic since as the project progresses there are additional requests from users or the need to respond to answer new research questions due to topics in the literature or to findings from previous data analysis. At present each query has 69 variables for each student admitted in the cohort. Research continues to identify relations between variables and to interpret the data applying models from other fields such as the area of customer profitability developed for marketing. The Illinois Institute of Technology (Demski, 2011), unlike many other universities, decided to develop their own personal computer system for the retention (which bears some similarities to our project). Their early warning system was developed by personnel of the university, allowing greater freedom to revise and update based on the recommendations of its users.

Programming in Visual Basic has been developed so that useful and meaningful patterns identified and modeled in their search are presented in Excel spreadsheets. These are the activities of researchers and the sponsor and the crucial linking the project between research and development. Moreover, this process is conceptualized to users and their different roles in the institution and its need for information to make decisions based on evidence related to retention.

Examples of Information for Decision Making

The users authorized to manage student information can access ERDU 2.0 from their desktop pc. In the software there are no predefined results. The user can solicit the analysis for subsets of the population by selecting variables from the menu. Also can describe retention; seek patterns or model behavior for all cohorts or segments of the cohorts. Thus, for example, the personnel of nontraditional program like Services Program for Adult Students (AVANCE) can explore retention in their population or a chairman of academic department can explore retention and graduation rate in their academics programs. This segmentation options encourages particular offices or groups with some skills in using Excel to use the software for informed decision making. In the next diagram we illustrate this point.

Diagram 1: Sequence followed in the project so that the research about retention is transformed into customized software for Excel (ERDU 2.0).



The capabilities of the program ERDU 2.0 that will be explained to the audience in a discussion of the following questions:

- How many first times university students from high, medium and low income families of the cohort of Fall 2002 are still enrolled 8 semesters later?
- What is the probability that students who enroll in the third semester for the cohort of Fall 2002 will graduate 12 semesters (6 years) later?
- How does the survival pattern of students for all cohorts who have completed 12 semesters compare to the pattern that is exhibiting the cohort of students admitted in August 2010?

Retention by academic and demographic variables

Retention according to student characteristics is a core topic in retention and institutions are asked to indicate retention by demographics in many evaluative and accrediting processes. The database has the information of family income and the user can select to run the analysis on first time university students (regular) rather than on all the population that includes transfers, distance and adult students. ERDU produces a table indicating numbers of students in the cohort and tracks how they are distributed according to retention and graduation at any subsequent semester. The user can then explore the frequencies for that semester according to the variables and compare the results from previous semesters. Sections of the table can be selected and copied to be integrated into Word or PowerPoint for reporting since it is an Excel table.

Identifying the probability of graduation from historical data from observed behavior in the first two years of study

The first year of study is crucial since it is the time when many students leave yet understanding how those who enroll for the second year progress in their education is important. The program analyzes the probability of graduation of this student population based on observed patterns for all cohorts. The observed probability is an important reference point for predicting what will occur if the institution does not intervene to promote retention in these student populations.

Evaluating the effectiveness of institutional interventions for retention based on modifying historical patterns

The program generates a graph showing the survival of students for 12 semesters for cohorts from August 1995 to August 2005. Research indicated that in all cohorts exhibit a similar pattern even when the size of the cohort varied. This constancy led to developing a model for predicting what will be the survival behavior of the next cohort assuming a similar institutional environment. It also allows for determining the success or failure of retention interventions that if effective should alter the historical pattern.

Conclusion

The customized programming is the tool that allows us to go beyond reporting based on the work of institutional technical personnel since the users can request information related to their needs. The software allows users to reproduce the analysis as new data is integrated into the institutional database every semester. The software does not have predefined values, it uses the data provided by the query and show results in table and graph based in our literature review and research findings. The variables available in ERDU for segmenting populations when requesting information will be described to the audience along with other areas of analysis that have been integrated.

References

- Astin, A. W. (1975). *Preventing students from dropping out*. San Francisco, CA: Jossey-Bass.
- Cabrera, A., Castaneda, M., Nora, A., & Hengstler, D. (1992). The convergence between two theories of college persistence. *Journal of Higher Education*, 63, 143-164.
- Caison A. L. (2007). Analysis of institutionally specific retention research: A Comparison Between Survey and Institutional Database Methods. *Research in Higher Education*.48(4), 435-451. doi: 10.1007/s11162-006-9032-5
- Centro de Investigaciones Comerciales e Iniciativas Académicas (2004). *Perfil del Ingresado a las Instituciones de Educación Superior de Puerto Rico, 1996-2002*: Preparado para el Consejo de Educación Superior. Río Piedras, Puerto Rico: Universidad de Puerto Rico Facultad de Administración de Empresas.
- Demski, J. (2011, march). Shining a Light on Retention. *Campus Technology*. Retrieved April10, 2011 from <http://campustechnology.com>
- Pascarella, E. T., & Terenzini, P. T. (1983). Predicting voluntary freshman year persistence/withdrawal behavior in a residential university: A path analytic validation of Tinto's model. *Journal of Educational Psychology*, 75, 215-226.
- Ramos, N, Mercado, M. & Rosa, H. (2008). Análisis cronológico del trasfondo académico de os estudiantes de nuevo ingreso matriculados en el Recinto de Arecibo, desde el año académico 1995-1996 al 2006-2007, por programa académico y escuela tributaria. *Espacio Abierto*. Retrieved June1, 2011 from <http://www.arecibo.inter.edu/biblioteca/abierto.htm>.
- Tinto, V. (1973). College proximity and rates of college attendance. *American Educational Research Journal*, 10(4), 277-93.