

## UCF Visual Analytics - Dataset Development for the Undergraduate Attrition Study Dashboard

Scott Milbuta and Michelle Parente, University of Central Florida.

### ABSTRACT

The office of Institutional Knowledge Management (IKM) at the University of Central Florida (UCF) has been gradually developing many insightful and impactful applications in SAS® Visual Analytics. This allowed existing information once held in discrete and disconnected set of reports to be combined and quickly seen in detail across several new dashboards. The Attrition Study Dashboard in particular required a different approach in its development as the questions being asked required more attention in order to surface properly in Visual Analytics. To do this, development of the datasets that drives this information in our Visual Analytics environment had to be redeveloped from our existing BI Portal delivered datasets. New challenges and insights were learned in executing this vision and process leading to solutions in order to build a master dataset for our Attrition Study Dashboard. This process is presented here and showcases the end result.

### INTRODUCTION

The University of Central Florida needed a way to investigate attrition at our institution at a very high level but also contain the ability to slice into almost any cross sectional segment of the data. Our traditional reports regarding this issue answered only some questions regarding attrition. Only by combining several reports a better picture of our institutions attrition could be seen, yet the process was cumbersome and far from ideal.

Approximately three years ago UCF and IKM decided we needed a better way to surface this data in a single location that allowed for all of the disjointed points of information to be coalesced into one cohesive report and surfaced for our users. This initiative required new ways to put our data together in order to be able to answer those questions that were elusive in the traditional reports.

Development for this in SAS® Visual Analytics required many of the tables traditionally used for our functional reports to be combined. This process involved many iterations, modifications and permutations of the dataset structure and layout in order to fit the data properly for consumption in our Visual Analytics platform and its use for the reports eventual development. Some of the key considerations that were needed during this development were pre calculated data points, considerations regarding the times frame of data, and the type quantitative and qualitative elements. This was necessary in order to provide the dashboard developers the measures they needed.

Once all the elements were identified they were assembled and the eventual result was a fully functional albeit unique set of data used in our Attrition Study Dashboard.

### INSTITUTIONAL KNOWLEDGE MANAGEMENT

The reporting needs at an institution the size of the University of Central Florida requires a team dedicated to providing data that is delivered quickly, accurately. The office of Institutional Knowledge Management is the official source for all of UCF academic institutional data and is tasked with providing our community with easy to access, insightful and actionable data to the public and the university community as a whole.

To help accomplish this goal, IKM has a full SAS® Enterprise Business Environment in place supporting our data warehouse. This includes a complete set of tools and applications including a SAS® Visual Analytics suite. This allows IKM to disseminate data to our users in almost any form they wish to see and access.

Our Current data warehouse contains over 18 years of academic information for students based on state census files. It also contains data from our current Oracle12 environments and other operational and silos

of data stores located at the department level around campus and brings these together for cohesive reporting purposes.

## PEGASUS MINE PORTAL / IKM ANALYTICS PORTAL

The user community at UCF has traditionally accessed the university data warehouse through the SAS® Information Delivery Portal. This is also referred to as the Pegasus Mine Portal (PMP). This environment offers the ability to run pre-built dynamic reports that provides a myriad of academic data for use in operational decision-making, strategic planning, forecasting, and performance measurement. Image 1 is an example of our current Pegasus Mine Portal and the type of offerings available to our UCF consumers. This is just one tab of over a dozen available reporting areas. As you can see in this example currently when attempting to bring together disparate facts regarding grades, enrollment and demographic information users would have to navigate a multitude of locations. This area serves our community with excellent capability but just did not meet the necessary needs to be able to see across so many elements at once.

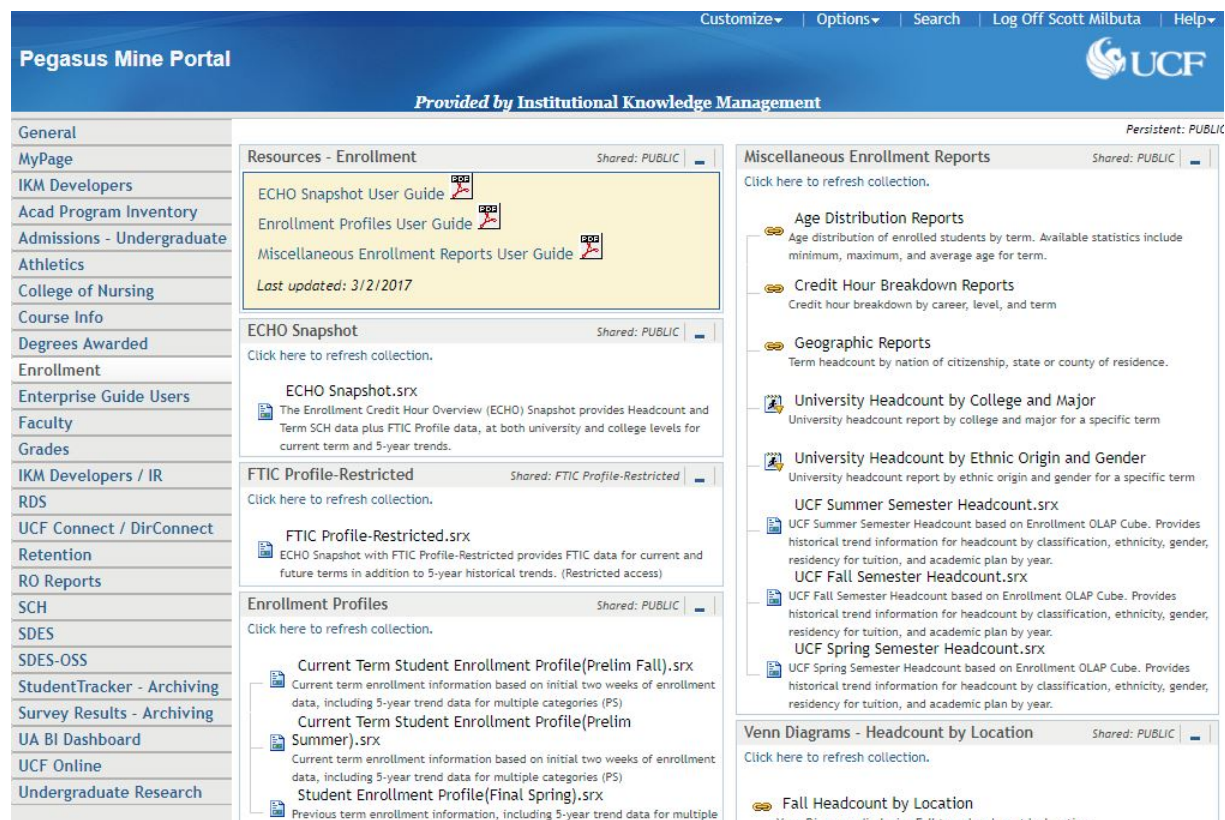
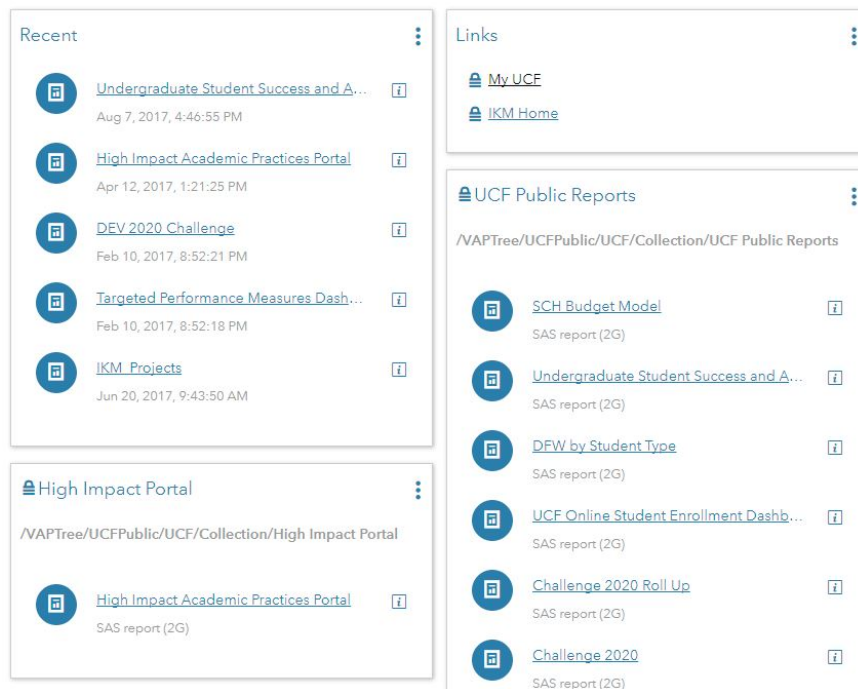


Image 1. UCF Pegasus Mine Portal Visual Example

Within the last three years we have expanded the way our users' access data by rolling out a new set of reports for administrative and executive communities on campus located in our Visual Analytics environment. This new environment has opened up whole new areas of report and capabilities on how to show traditional reports once located in the PMP in new dimensions. Image 2 is an example of what our users see when entering our analytics environment and a fraction of the reports and dashboards currently available for consumption.



**Image 2. UCF Visual Analytics Portal Example**

## DASHBOARD REQUIREMENTS

The requirements for the attrition study dashboard were very demanding in its concept. The following broad elements needed to be combined in order to deliver the necessary information to the university. The following is a list of some of the more important needs:

- 10 academic years of standard demographic and academic data
- Broad category views, Graduated, Enrolled, Attritted
- DFW course correlation between student types
- Entry college migration and results with drill down
- Major entry migration and results with drilldown
- Factors for progression and first term and last term impacts
- Attritted student academic status
- Impact summary on attrition
- Trending

All of these elements in the dashboard needed to be transitioned to seamlessly by the users, and convey the data clearly when presented. The data behind them also needed to be compiled in a way to allow for all of the cross sectional views required in order to accomplish our goals.

In the end two primary datasets were developed for this project. The first being a master student dataset containing aggregated course elements, and a course dataset with the atomic student course information.

When we received the specification of the requirements for the data needs we realized that our traditional tables we had in our data warehouse would only suffice as so far in collecting the core foundation data. To bring those elements together we had to begin to look at how we wanted to interpret that data for use within the design of the dashboard.

As any Visual Analytics developer can attest to, attempting to perform N-way views on data can sometimes be a challenge especially if the source data is not maintained in that way originally. One set of data that required extensive pre-calculation and manipulation was the student data that was brought in from many sources. Image 3 shows in the highlighted green boxes, some of the new columns that had to be created in order to properly surface comparative graphing objects and panels in the report. The overall dataset for this project has over 80 columns of data with over 40% of it in custom recoded or compiled data. This was necessary in order to bring the data down to one row per student and reduce the overall data set size for performance.

| STATUS    | BRANCH ID | ENROLLED | ATTRITD | PULL_YN | PULL_FIS_YEAR | MO_YEAR | LOW_INQ_MO | LOWING | LOW_INQ_MO | LOWING_SP   | FSGN        | FSGN_FIG  | RESIDE_COUNTRY | COMPLT_COUNTRY | SUPPLY | SCORSEN_T | NUMBERS_NT | DEPLDONT | INCLNT | PLDNT | FLOAT | SAPCON_EX_DATE |       |
|-----------|-----------|----------|---------|---------|---------------|---------|------------|--------|------------|-------------|-------------|-----------|----------------|----------------|--------|-----------|------------|----------|--------|-------|-------|----------------|-------|
| ENROLLED  | 0         | 1        | 0       |         |               |         |            |        |            |             |             |           | 8              | 4              | 4      | 5         | 4          | 1        |        | 1     |       |                |       |
| GRADUATED | 1         | 0        | 0       | Y       | 2012          |         |            |        |            |             | First_Gen   | 1         | 10             | 10             | 0      | 20        | 6          |          |        |       |       |                |       |
| ENROLLED  | 0         | 1        | 0       | Y       |               |         |            |        |            |             |             |           | 6              | 4              | 1      | 4         | 3          |          |        |       |       |                |       |
| ENROLLED  | 0         | 1        | 0       | Y       | 2009          |         |            |        |            |             | First_Gen   | 1         | 26             | 16             | 8      | 19        | 3          | 3        | 1      | 1     | 1     | 580.0          |       |
| GRADUATED | 1         | 0        | 0       |         |               |         |            |        |            |             |             |           | 17             | 6              | 11     | 11        | 9          |          |        |       |       |                |       |
| ATTRIT    | 0         | 0        | 0       | Y       | 2013          | 2013    | YES        |        | 1          | LOWING ENT. |             |           | 2              | 2              | 0      | 7         | 3          |          |        |       | 2     |                |       |
| GRADUATED | 1         | 0        | 0       | Y       | 2012          | 2012    | YES        |        | 1          | LOWING POS. |             |           | 9              | 7              | 2      | 16        | 4          |          |        |       |       |                |       |
| ATTRIT    | 0         | 0        | 0       | Y       | 2013          | 2013    | YES        |        | 1          | LOWING ENT. |             |           | 3              | 2              | 1      | 7         | 5          |          |        | 2     | 3     |                |       |
| ATTRIT    | 0         | 0        | 0       | Y       |               |         |            |        |            |             | First_Gen   | 1         | 2              | 2              | 0      | 5         | 2          | 5        |        |       | 5     | 560.0          |       |
| GRADUATED | 1         | 0        | 0       |         |               |         |            |        |            |             |             |           | 7              | 7              | 0      | 20        | 6          |          |        |       |       |                |       |
| GRADUATED | 1         | 0        | 0       | Y       | 2011          | 2011    | YES        |        | 1          | LOWING POS. |             |           | 13             | 12             | 0      | 23        | 5          | 5        |        | 4     |       |                |       |
| ENROLLED  | 0         | 1        | 0       | Y       | 2016          | 2016    | YES        |        | 1          | LOWING ENT. |             |           | 4              | 3              | 1      | 4         | 3          |          |        |       |       |                |       |
| ATTRIT    | 0         | 0        | 0       | Y       |               | 2013    | 2013       | YES    |            | 1           | LOWING POS. |           |                | 6              | 7      | 1         | 12         | 7        |        |       |       | 820.0          |       |
| ATTRIT    | 0         | 0        | 0       | Y       | 2013          | 2013    | YES        |        | 1          | LOWING ENT. |             |           | 2              | 2              | 0      | 6         | 3          |          |        |       |       |                |       |
| ATTRIT    | 0         | 0        | 0       | Y       | 2012          |         |            |        |            |             | First_Gen   | 1         | 1              | 1              | 0      | 2         | 2          |          |        |       |       |                |       |
| GRADUATED | 1         | 0        | 0       | Y       | 2008          | 2008    | YES        |        | 1          | LOWING ENT. |             |           | 9              | 9              | 0      | 36        | 3          | 5        | 1      | 2     | 2     |                |       |
| GRADUATED | 1         | 0        | 0       | Y       | 2010          |         |            |        |            |             | First_Gen   | 1         | 8              | 6              | 2      | 19        | 7          |          |        |       |       |                |       |
| ENROLLED  | 0         | 1        | 0       | Y       |               |         |            |        |            |             |             |           | 14             | 12             | 10     | 10        | 12         |          |        |       |       |                |       |
| GRADUATED | 1         | 0        | 0       | Y       | 2013          |         |            |        |            |             |             |           | 14             | 12             | 11     | 1         | 23         | 8        |        |       |       |                |       |
| GRADUATED | 1         | 0        | 0       | Y       | 2008          | 2008    | YES        |        | 1          | LOWING ENT. |             |           | 7              | 6              | 1      | 16        | 3          |          |        |       |       | 1050.0         |       |
| GRADUATED | 1         | 0        | 0       | Y       |               |         |            |        |            |             |             |           | 22             | 19             | 3      | 23        | 5          | 4        |        | 2     |       |                |       |
| GRADUATED | 1         | 0        | 0       | Y       | 2014          | 2011    | YES        |        | 1          | LOWING POS. |             |           | 14             | 9              | 8      | 9         | 7          |          | 3      | 2     |       |                |       |
| ENROLLED  | 0         | 1        | 0       |         |               |         |            |        |            |             | First_Gen   | 1         | 26             | 13             | 12     | 13        | 4          |          |        |       |       | 1070.0         |       |
| GRADUATED | 1         | 0        | 0       | Y       | 2013          |         |            |        |            |             |             |           | 7              | 6              | 1      | 20        | 11         | 1        |        | 1     |       |                |       |
| GRADUATED | 1         | 0        | 0       | Y       | 2008          | 2008    | YES        |        | 1          | LOWING ENT. |             |           | 9              | 9              | 0      | 29        | 8          | 6        |        | 3     | 3     | 1140.0         |       |
| GRADUATED | 1         | 0        | 0       | Y       | 2008          | 2008    | YES        |        | 1          | LOWING ENT. |             |           | 7              | 7              | 0      | 26        | 8          |          |        |       |       | 760.0          |       |
| GRADUATED | 1         | 0        | 0       |         |               |         |            |        |            |             | First_Gen   | 1         | 11             | 10             | 1      | 22        | 14         | 2        |        | 1     | 1     | 960.0          |       |
| ATTRIT    | 0         | 0        | 0       | Y       |               |         |            |        |            |             |             |           | 1              | 1              | 0      | 1         |            |          |        | 1     |       |                |       |
| GRADUATED | 1         | 0        | 0       | Y       | 2009          |         |            |        |            |             |             |           | 6              | 6              | 0      | 2         |            |          |        | 1     |       |                |       |
| ATTRIT    | 0         | 0        | 0       | Y       |               |         |            |        |            |             |             |           | 2              | 2              | 0      | 2         |            |          |        |       |       |                |       |
| ATTRIT    | 0         | 0        | 0       | Y       | 2009          | 2009    | YES        |        | 1          | LOWING ENT. |             |           | 5              | 4              | 1      | 11        |            | 7        |        | 7     |       | 1100.0         |       |
| GRADUATED | 1         | 0        | 0       |         |               |         |            |        |            |             |             |           | 8              | 8              | 0      | 26        | 4          |          |        |       |       | 860.0          |       |
| ATTRIT    | 0         | 0        | 0       | Y       |               | 2016    | 2016       | YES    |            | 1           | LOWING ENT. |           | 1              | 1              | 0      | 1         |            |          |        |       |       |                |       |
| ENROLLED  | 0         | 1        | 0       | Y       |               | 2015    | 2015       | YES    |            | 1           | LOWING ENT. |           | 7              | 7              | 0      | 16        | 16         | 8        | 1      | 2     | 1     |                |       |
| ENROLLED  | 0         | 1        | 0       |         |               |         |            |        |            |             |             |           | 1              | 1              | 0      | 2         | 2          |          |        |       |       | 1000.0         |       |
| GRADUATED | 1         | 0        | 0       | Y       | 2012          | 2014    | YES        |        | 1          | LOWING POS. |             | First_Gen | 1              | 9              | 8      | 1         | 27         | 12       | 8      |       | 4     | 3              | 800.0 |
| GRADUATED | 1         | 0        | 0       | Y       | 2011          | 2011    | YES        |        | 1          | LOWING ENT. |             |           | 10             | 10             | 0      | 24        | 10         | 1        |        |       |       | 1210.0         |       |
| GRADUATED | 1         | 0        | 1       |         |               |         |            |        |            |             | First_Gen   | 1         | 6              | 16             | 12     |           | 26         | 13       | 1      |       |       |                |       |

As indicated earlier we needed to create a secondary dataset more limited in scope that exclusively focused on courses. This data set was necessary in order to create some course level specific data by student in order to drill into academic performance information not easily reflected in aggregated data in the master table. The secondary table provided information on the following topics categories based on the student:

- STEM (Science, Technology, Engineering, Math) courses
- Excel Program Courses (Courses contributing to student success in first two years)
- Repeat Courses attempt and outcomes
- DFW percentages (Drop, Fail, Withdraw grades)

In addition to the above elements the course table also provided some broader measures for each course such as:

- Total Enrollment
- Class Size grouping
- Meeting Pattern

## Development Time and Process

The development of both datasets took approximately 1 months' time. This involved close collaboration with the analysts, developers and report builders in order to meet the needs of the project. Countless iterations of the data were created and tested in order to be able to generate the appropriate comparison elements needed. The process also was developed so as to seamlessly bring in new cohort information every term and add the data to the final datasets with minimal maintenance.

In addition, documentation was developed in order to be able to back track data elements to the source should it be necessary for explanatory reasons why certain categorical decisions were made. The overall development time from concept to completion for this project was just over a year.

Working the needs for the manipulation of certain data elements of information into both datasets, the Visual Analytic developers were able to bring all those points together in order to surface one of the most comprehensive reports at UCF regarding student attrition.

## DATASET IN ACTION

The following section shows the final datasets in action within the Visual Analytics environment. Using our standard BI environment we would not have been able to show such correlation between factors. The dataset allowed for very intuitive displays of complicated data and to clearly surface them in the Attrition Dashboard. The first sample in image 4 shows the summary page of the overall population of the cohorts.

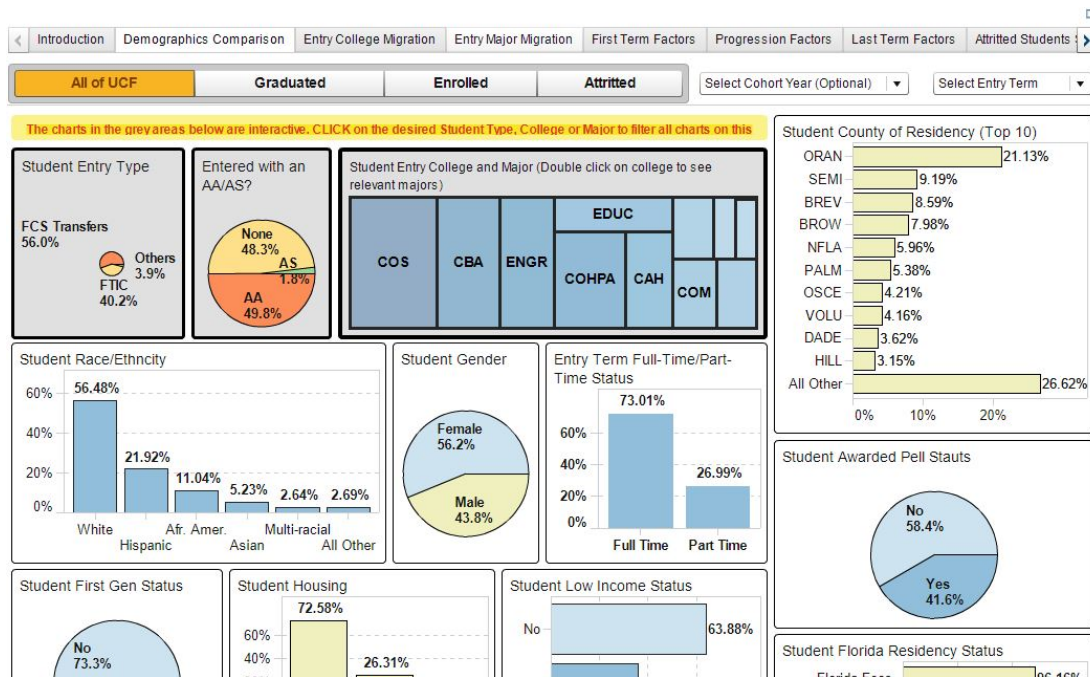


Image 4. Attrition Study Summary (image condensed for display)

Image 5 displays a sample of one of the three factor reports used in the dashboard. This one was selected as it contains an example of an insightful piece of data not previously realized when looking at all of the factors separately. The data when compiled and surfaced showed a spike in attrition for students with 3.0 or greater GPA during their first term. Additional elements not shown revealed a similar pattern of high GPA students attriting during entry into their third year.

This is a clear example of how the assembly of the data and then the development of the measures to use that data aided UCF in identifying a factor contributing to the attrition of students who for all other reasons should not be attriting.

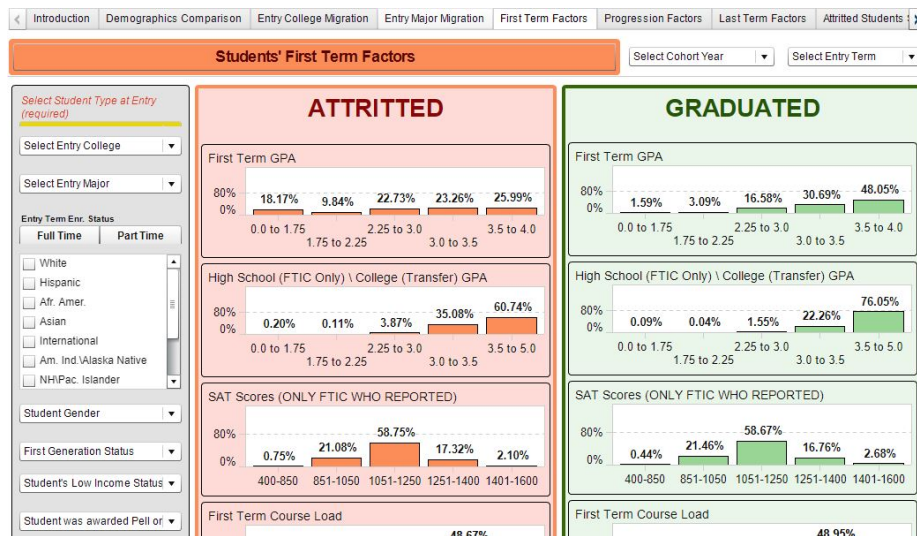


Image 5. Attrition study dataset sample

Image 6 shows a sample of the migration elements of the study. Users were able to trace a cohort's performance based on college of entry through most recent and then broad buckets of last known status. The function of this report is similar to a Sankey but limited in grouping points. At some point in the future when our environment migrates to the next version of Visual Analytics Sankey reports will most likely replace this element.

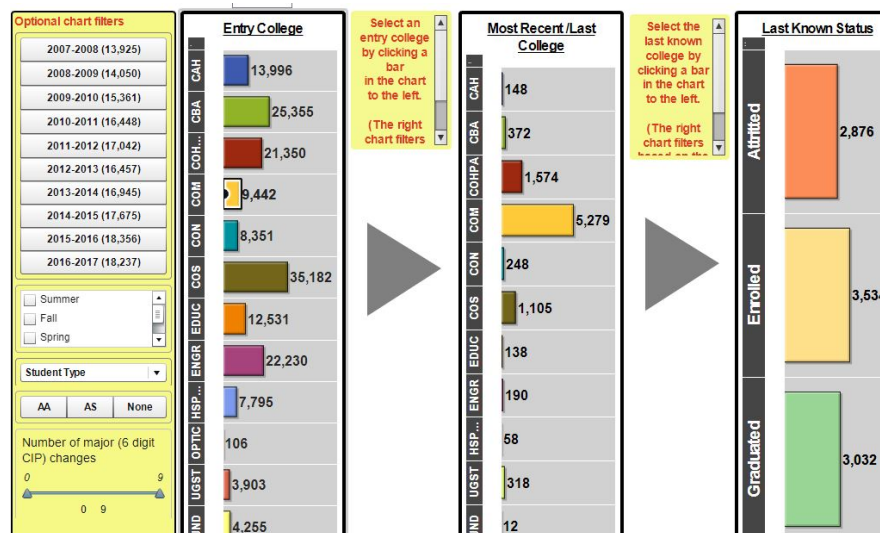


Image 6. Attrition study dataset sample

Image 7 is the main trends exploration for all cohorts. This display is a comprehensive data exploration tool for interested users to dig into the data. It displays all of the Enrolled, Attritted and Graduated information for each cohort year. It is filterable by over 40 separate selections all made possible by the careful development of the core datasets. This allows the user to see UCF as a whole within scope of the Attrition Study.

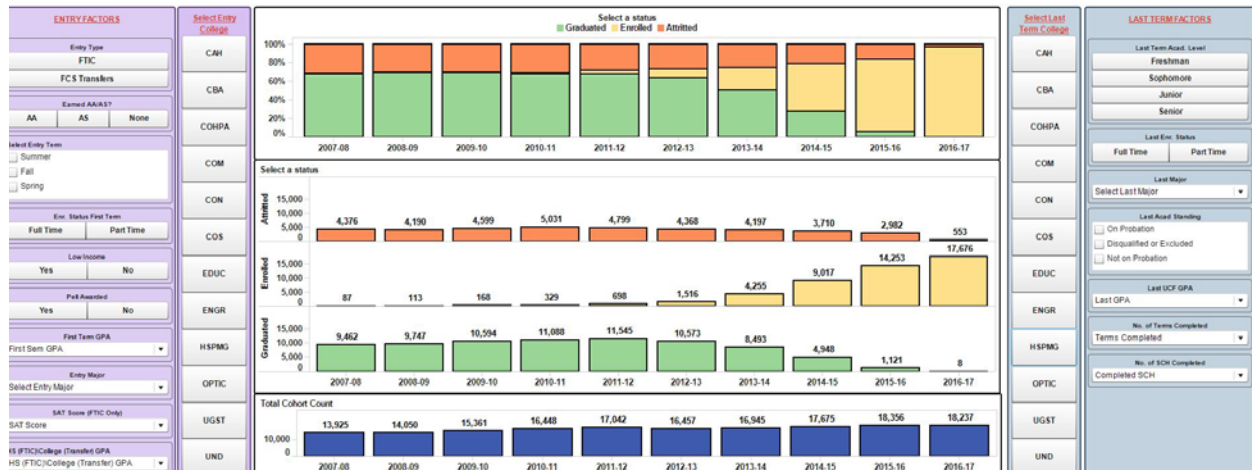


Image 7. Attrition study dataset sample

## CONCLUSION

The intent of this paper was to demonstrate that careful analysis and development of the datasets behind the Dashboard is just as important as the Dashboard themselves. Many reports today use datasets and information from sources that meet many functional needs but lack the necessary data points for proper analytical and insightful use.

Tools today such as Visual Analytics make that job much easier by being able to extrapolate hidden bits of data from our everyday datasets. Taking the time to ensure your data is just as useful and manicured as the Dashboard allows for new ways to see hidden information.

## ACKNOWLEDGMENTS

I would like to acknowledge Meghal Parikh and Andre Watts. They are the primary analysts and developers of the reporting portion of the Visual Analytics Dashboard. The images provided here are based on their excellent work.

## REFERENCES

Institutional Knowledge Management (<https://ikm.ucf.edu/about/about-ikm/>).

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Scott Milbuta  
Sr. Decision Support Architect/Developer  
[Scott.Milbuta@ucf.edu](mailto:Scott.Milbuta@ucf.edu)  
University of Central Florida  
Institutional Knowledge Management  
12424 Research Pkwy, Suite 215  
Orlando, FL 32826  
Web: <https://ikm.ucf.edu/>

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.