SESUG 2024 Paper 55

Empowering Educational Institutions: Utilizing SAS® Tools for the Development of Student Retention Dashboard

Raj Bhosale ME, MBA, PMP
NC State University

ABSTRACT

In the evolving landscape of data management in a Higher-Ed setting, the need for effective tools to enhance student retention strategies is paramount. This paper presents a comprehensive methodology for harnessing the capabilities of SAS® Data Integration Studio and SAS® Visual Analytics to develop actionable dashboards.

SAS® Data Integration Studio serves as the backbone of our data management framework. It allows for the extraction, transformation, and loading (ETL) of data from multiple sources, including academic records, student engagement metrics, and demographic information. This process ensures that the data used in retention analysis is clean, consistent, and comprehensive. By automating the ETL processes, institutions can streamline their data workflows, thereby enhancing the reliability of the insights generated.

On the visualization front, SAS® Visual Analytics plays a pivotal role in transforming the processed data into intuitive and interactive dashboards. These dashboards are designed to cater to various stakeholders within educational institutions, such as graduate program directors, department heads, and faculty members. They provide real-time visual representations of key indicators like engagement levels, academic performance, and attrition rates. The interactive nature of these dashboards allows users to drill down into specific data points, facilitating a deeper understanding of underlying trends and enabling proactive intervention strategies.

INTRODUCTION

Student retention is a critical aspect of higher education, directly impacting institutional success and the academic and professional futures of students. The ability to retain students through to the completion of their degrees is a key indicator of an institution's effectiveness, affecting its reputation, financial stability, and overall performance. High retention rates are often correlated with a supportive learning environment, student satisfaction, and robust academic programs, while low retention rates can signal underlying issues that need to be addressed.

In the increasingly data-driven landscape of higher education, the use of meaningful visualizations has emerged as a powerful tool for leadership to identify trends, uncover insights, and make informed decisions. Visualizations transform complex data sets into accessible, intuitive formats, allowing administrators to quickly grasp patterns and correlations that might not be evident in raw data. By leveraging visual analytics, higher education leaders can better understand factors influencing student retention, such as demographic trends, academic performance, engagement levels, and support services utilization.

This paper explores the importance of student retention in higher education and demonstrates how SAS® Data Integration Studio and SAS® Visual Analytics can help create meaningful dashboards that can contribute towards short- and long-term strategic goals of the institution.

The paper can be broken down into three major steps:

1. Data Collection

2. Data Manipulation
3. Data Upload

DATA COLLECTION

This step mainly focuses on looking at the business requirement and determining the different sources of data. Here is the overview of the business requirement that we had on the scope of work:

1. Create a Dashboard to display cohort-based retention data for the past 10 years.
2. Ability to drill down by various elements like college, program and degree type.
3. Ability to drill down by demographic elements like gender and ethnicity.

The subsequent steps involved are identifying various data sources and tables, and determining how to integrate this data to form a unified dataset that serves as the primary data source for the dashboard. The Student Information System is the main data source for this step, with the following tables being of particular interest:

1. End of term Enrollment table: This table holds term-by-term information for students captured at the end of each term. Primary elements used from this table are program, degree, location (distance/on-campus) etc.
2. Degrees table: This table holds information about what Degrees were awarded to students. This table has a field called completion term that tells when a student was awarded a degree.
3. Bio table: This table holds demographic information about students. Some of the fields in this table are gender, ethnicity, residency etc.
4. Session table: This table holds session data for each term such as start and end dates.

DATA MANIPULATION

The end goal of this step was to combine all tables and create a dataset that will consist of one row per student per degree. This is how the structure of the table looks like:

#	Name	Description	Type	Length	Informat	Format	
1	EMPLID	EMPLID	Character	11	(None)	$11.	Y
2	STDNT_CAR_NBR	STDNT_CAR_NBR	Numeric	8	(None)	BEST12.	Y
3	STARTING_ACAD_YEAR	STARTING_ACAD_Y...	Character	8	(None)	$8.	Y
4	STARTING_TERM	STARTING_TERM	Character	4	(None)	$4.	Y
5	STARTING_TERM_DESCR	STARTING_TERM_D...	Character	30	(None)	$30.	Y
6	STARTING_PLAN	STARTING_PLAN	Character	10	(None)	$10.	Y
7	STARTING_PLAN_DESCR	STARTING_PLAN_D...	Character	30	(None)	$30.	Y
8	STARTING_ACAD_PROG	STARTING_ACAD_P...	Character	5	(None)	$5.	Y
9	STARTING_PROG_DESCR	STARTING_PROG_D...	Character	30	(None)	$30.	Y
10	STARTING_COLLEGE	STARTING_COLLEGE	Character	50	(None)	$50.	Y
11	STARTING_DEGREE	STARTING_DEGREE	Character	8	(None)	$8.	Y
12	ENDING_TERM	ENDING_TERM	Character	8	(None)	$8.	Y
13	ENDING_TERM_DESCR	ENDING_TERM_DES...	Character	30	(None)	$30.	Y
14	ENDING_PLAN	ENDING_PLAN	Character	10	(None)	$10.	Y
15	ENDING_PLAN_DESCR	ENDING_PLAN_DES...	Character	30	(None)	$30.	Y
16	ENDING_ACAD_PROG	ENDING_ACAD_PROG	Character	5	(None)	$5.	Y
17	ENDING_PROG_DESCR	ENDING_PROG_DES...	Character	30	(None)	$30.	Y
18	ENDING_COLLEGE	ENDING_COLLEGE	Character	50	(None)	$50.	Y
19	DEGREE	DEGREE	Character	8	(None)	$8.	Y
20	DEGREE_PLAN	DEGREE_PLAN	Character	10	(None)	$10.	Y
21	DEGREE_TYPE	DEGREE_TYPE	Character	20	(None)	$20.	Y
22	COMPLETION_TERM	COMPLETION_TERM	Character	4	(None)	$4.	Y
23	MASTERS_DEGREE	MASTERS_DEGREE	Character	8	(None)	$8.	Y
24	MASTERS_COMPLETION_TERM	MASTERS_COMPLET...	Character	4	(None)	$4.	Y
25	ENDING_WITHDRAW_CODE	ENDING_WITHDRA...	Character	30	(None)	$30.	Y
26	ENDING_WITHDRAW_DESCR	ENDING_WITHDRA...	Character	30	(None)	$30.	Y
27	ENROUTE_MASTERS	ENROUTE_MASTERS	Character	8	(None)	$8.	Y
28	NC_GENDER	NC_GENDER	Character	10	(None)	$10.	Y
29	TUITION_RES	TUITION_RES	Character	5	(None)	$5.	Y
30	NC_TUI_RES_CODE	NC_TUI_RES_CODE	Character	3	(None)	$3.	Y
31	NC_CIT_COUNTRY	NC_CIT_COUNTRY	Character	3	(None)	$3.	Y
32	ETHNICITY	ETHNICITY	Character	30	(None)	$30.	Y
33	SESS_BEGIN_DT	SESS_BEGIN_DT	Numeric	8	(None)	BEST12.	Y
34	SESS_END_DT	SESS_END_DT	Numeric	8	(None)	BEST12.	Y
35	LATEST_TERM_SYSTEM	LATEST_TERM_SYS...	Character	8	(None)	$8.	Y
36	TIME	TIME	Numeric	8	(None)	BEST12.	Y
37	INTERIM_STATUS	INTERIM_STATUS	Character	50	(None)	$50.	Y
38	STATUS	STATUS	Character	50	(None)	$50.	Y
39	PROG_ACTION	PROG_ACTION	Character	4	(None)	$4.	Y
40	REQ_TERM	REQ_TERM	Character	4	(None)	$4.	Y

Some of the important fields from the output table are highlighted in yellow. These fields indicate the following:

1. STDNT_CAR_NBR: This is the student career number. For e.g. if the student begins Masters degree at NC State, they will have a career number of 1. Once they finish Masters and move to a PhD or a second masters, their career number on the enrollment stack changes from 1 to 2. Here is an example:

EMPLID	**STDNT_CAR_NBR**	**STARTING_ACAD_YEAR**	**STARTING_TERM_DESCR**	**STARTING_PROG_DESCR**
XXXXXX	1	2014	2014 Fall Term	Civil Engineering -MS
XXXXXX	2	2017	2018 Spring Term	Civil Engineering -PHD

2. STARTING_TERM/PROGRAM/COLLEGE: This is the starting semester, the relevant program and the college for that career number.

3. ENDING_TERM/PROGRAM/COLLEGE: This is the ending semester, the relevant program and the college for that career number.

4. ENROUTE_MASTERS: This field is specifically for doctoral students to see if they received an enroute master's while they pursued their doctoral degree.

5. Status: This field is calculated based on an expression logic:

```
     CASE
     WHEN ((DEGREE  IS NOT NULL) AND (STARTING_PLAN = DEGREE_PLAN)) THEN
'COMPLETED SAME PLAN'
     WHEN ((DEGREE IS NOT NULL) AND (STARTING_PLAN <>  DEGREE_PLAN))
THEN 'COMPLETED DIFFERENT PLAN'
     WHEN ((ENDING_TERM = LATEST_TERM_SYSTEM) AND (DEGREE IS NULL)) THEN
'ACTIVE'
     WHEN ((COMPLETION_TERM IS NULL) AND (INPUT(ENDING_TERM, 4.)) <=
(INPUT(LATEST_TERM_SYSTEM, 4.) - 10) AND (STARTING_DEGREE  NOT IN
('PHD','EDD','DDES','DVM'))) THEN DISCONTINUED'
     WHEN (((INPUT(ENDING_TERM, 4.)) > (INPUT(LATEST_TERM_SYSTEM , 4.) -
10)) AND (LATEST_TERM_SYSTEM <> ENDING_TERM)) THEN DISCONTINUED'
     WHEN ((COMPLETION_TERM IS NULL) AND (INPUT(ENDING_TERM, 4.)) <=
(INPUT(LATEST_TERM_SYSTEM, 4.) - 10) AND (STARTING_DEGREE  IN
('PHD','EDD','DDES','DVM') AND (MASTERS_DEGREE IS NULL))) THEN
DISCONTINUED'
     WHEN ((COMPLETION_TERM IS NULL) AND (INPUT(ENDING_TERM, 4.)) <=
(INPUT(LATEST_TERM_SYSTEM, 4.) - 10) AND (STARTING_DEGREE  IN
('PHD','EDD','DDES','DVM') AND (MASTERS_DEGREE IS NOT NULL) AND
(MASTERS_COMPLETION_TERM > STARTING_TERM))) THEN DISCONTINUED WITH
ENROUTE MASTERS'
WHEN ((COMPLETION_TERM IS NULL) AND (INPUT(ENDING_TERM, 4.)) <=
(INPUT(LATEST_TERM_SYSTEM, 4.) - 10) AND (STARTING_DEGREE  IN
('PHD','EDD','DDES','DVM') AND (MASTERS_DEGREE IS NOT NULL) AND
(MASTERS_COMPLETION_TERM <= STARTING_TERM))) THEN DISCONTINUED'
ELSE 'N/A'
     END
```

As seen from the above expression, the status of the student will be from one of the following:

1. Completed same plan
2. Completed different plan
3. Active
4. Discontinued

5. Discontinued with enroute masters

Here is a simplified visual of how SQL joins are achieved inside of the SAS DI job:

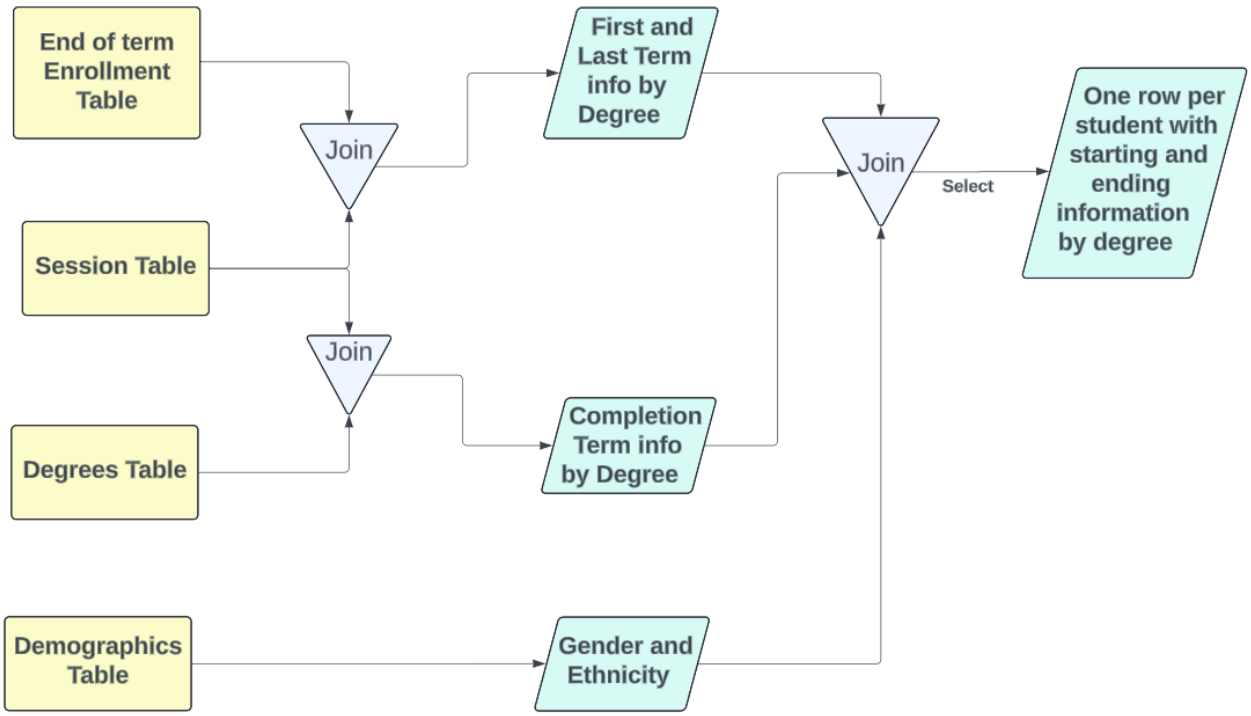
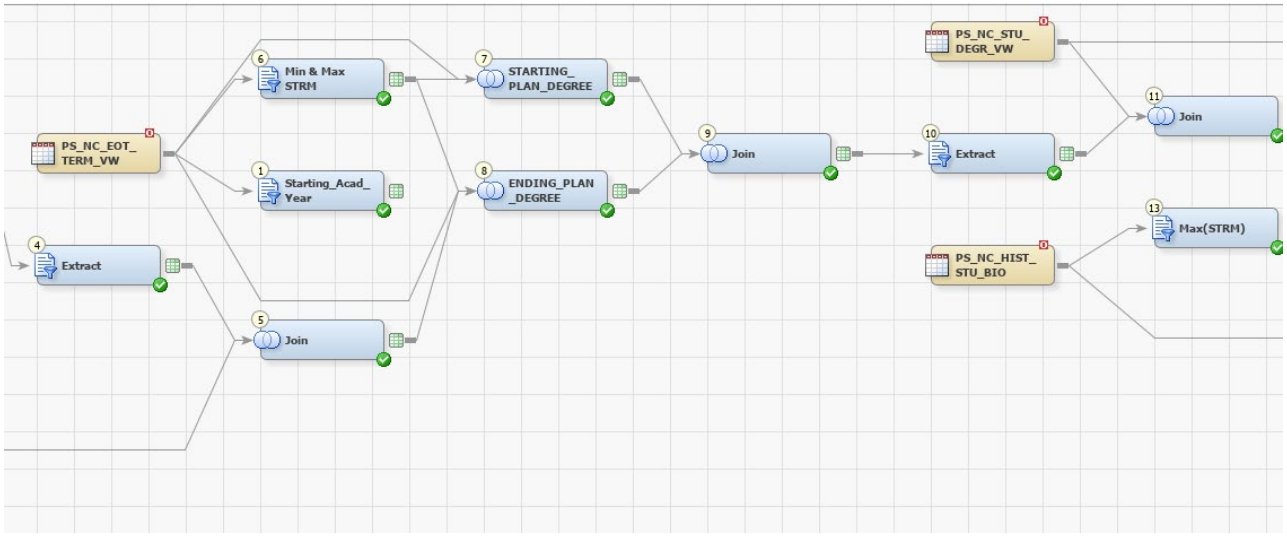
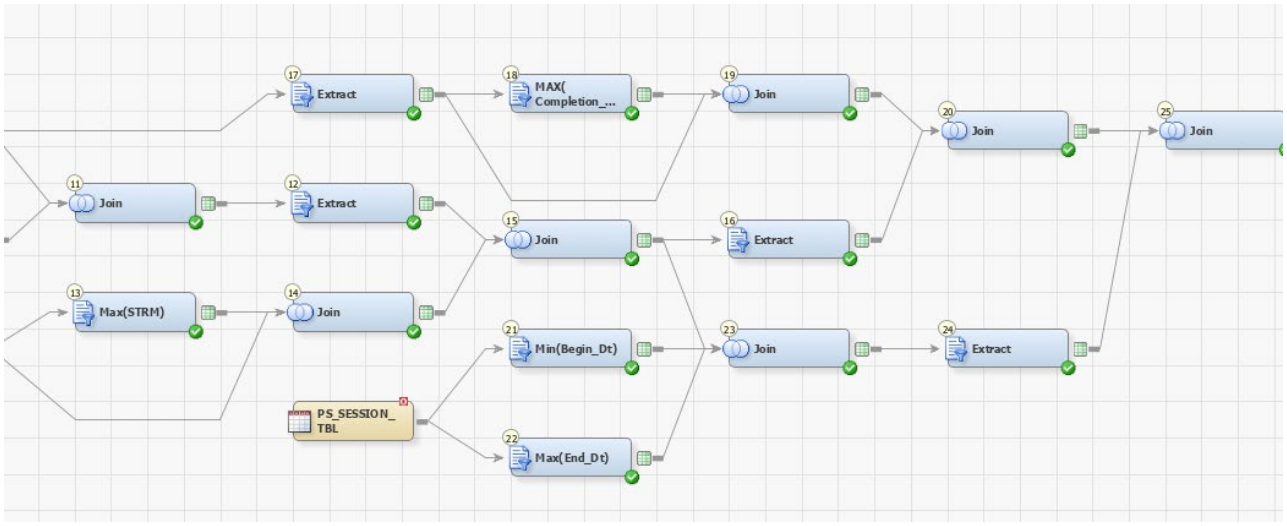


Figure 1: Display of how Joins are used to obtain Retention Data (One line per student degree)

Here is how the job looks in actual inside of the DI Studio:

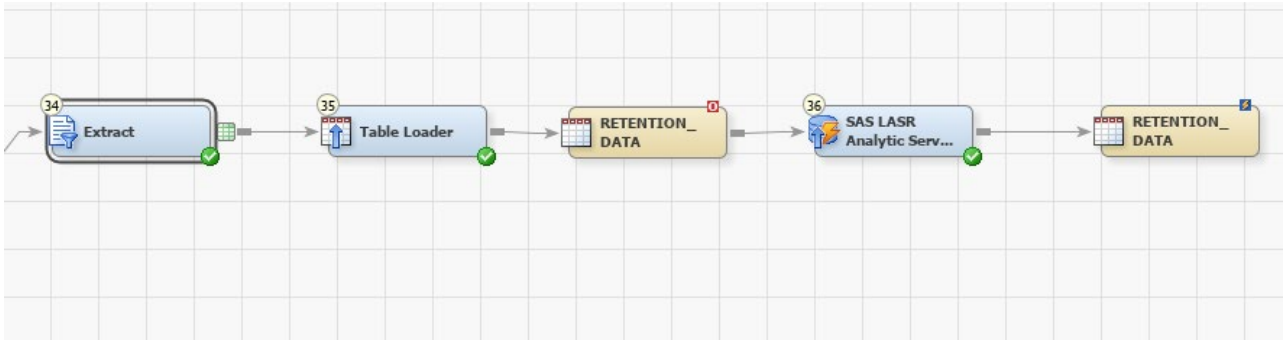




Display 1: Display of the SAS® DI Job

DATA UPLOAD

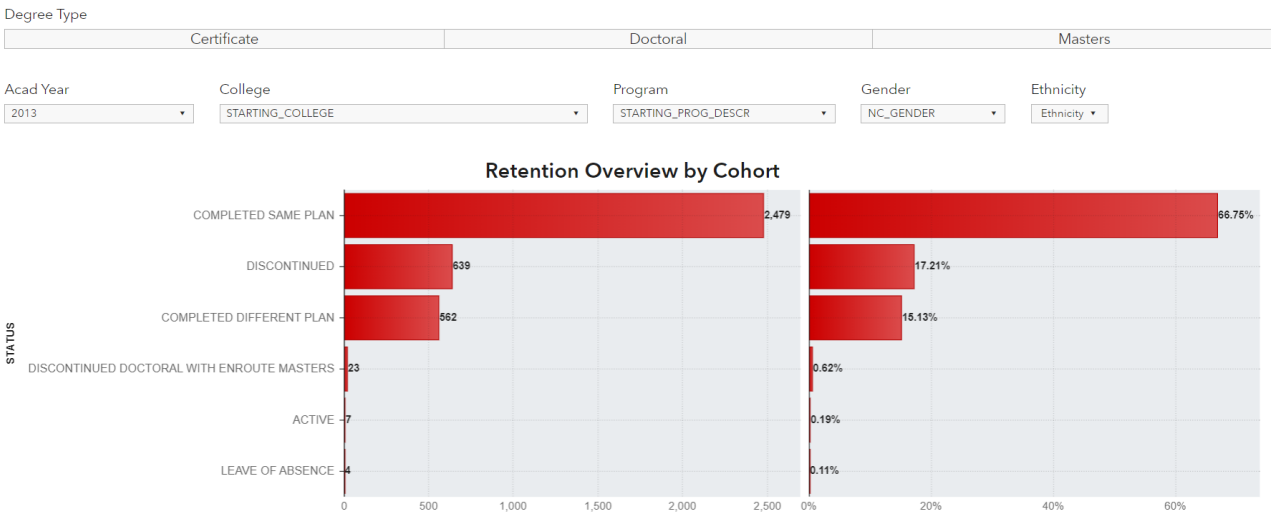
The final step is to upload this data to the SAS® LASR Server. Once the desired data set is obtained it is first converted into an Oracle table using the "Table Loader" function and then pushed to the SAS LASR server:



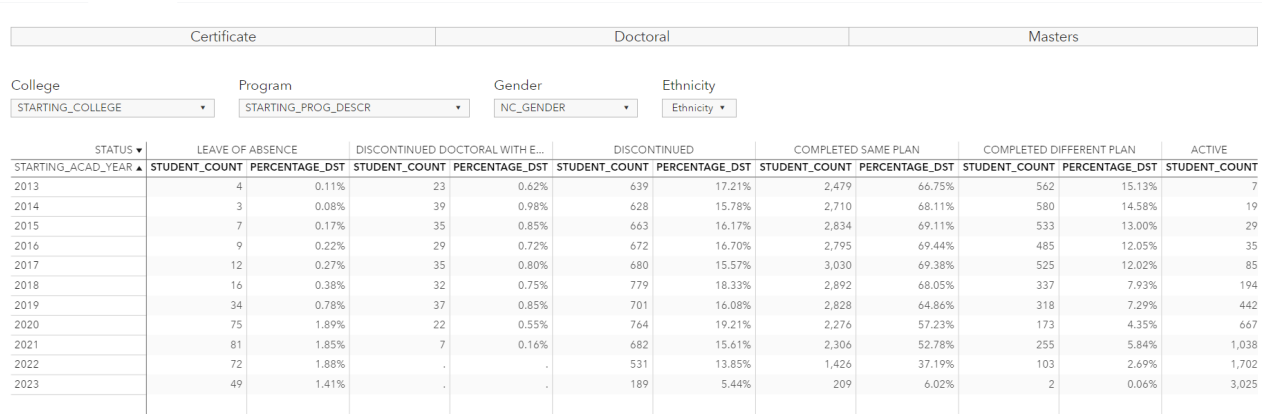
Display 2. Screenshot of how the data was uploaded to SAS® LASR

This task is scheduled to run once per term. It executes one day after the Census Day of Enrollment, which is the final day for students to drop or add classes. At this point, students' enrollment schedules are finalized, allowing us to view the new term's numbers on the dashboard.

Finally, here are screenshots of the final retention dashboard:



Display 3. Screenshot of Retention Dashboard



STATUS	LEAVE OF ABSENCE		DISCONTINUED DOCTORAL WITH E...		DISCONTINUED		COMPLETED SAME PLAN		COMPLETED DIFFERENT PLAN		ACTIVE
STARTING_ACAD_YEAR	STUDENT_COUNT	PERCENTAGE_DST	STUDENT_COUNT	PERCENTAGE_DST	STUDENT_COUNT	PERCENTAGE_DST	STUDENT_COUNT	PERCENTAGE_DST	STUDENT_COUNT	PERCENTAGE_DST	STUDENT_COUNT
2013	4	0.11%	23	0.62%	639	17.21%	2,479	66.75%	562	15.13%	7
2014	3	0.08%	39	0.98%	628	15.78%	2,710	68.11%	580	14.58%	19
2015	7	0.17%	35	0.85%	663	16.17%	2,834	69.11%	533	13.00%	29
2016	9	0.22%	29	0.72%	672	16.70%	2,795	69.44%	485	12.05%	35
2017	12	0.27%	35	0.80%	680	15.57%	3,030	69.38%	525	12.02%	85
2018	16	0.38%	32	0.75%	779	18.33%	2,892	68.05%	337	7.93%	194
2019	34	0.78%	37	0.85%	701	16.08%	2,828	64.86%	318	7.29%	442
2020	75	1.89%	22	0.55%	764	19.21%	2,276	57.23%	173	4.35%	667
2021	81	1.85%	7	0.16%	682	15.61%	2,306	52.78%	255	5.84%	1,038
2022	72	1.88%	.	.	531	13.85%	1,426	37.19%	103	2.69%	1,702
2023	49	1.41%	.	.	189	5.44%	209	6.02%	2	0.06%	3,025

Display 4. Screenshot of Retention Dashboard Trends

CONCLUSION

As demonstrated above, SAS® Data Integration Studio serves as a very effective tool towards Data Manipulation. With tables from different systems available under a single roof, the data can be successfully massaged and manipulated to the desired output using several transformations like Extract, Join, Table Uploader, Append and many others. SAS® Data Integration Studio provides the user with a stable dataset that can then be used for Dashboards and other visual reports using SAS® Visual Analytics.

CONTACT INFORMATION

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

Raj Bhosale (Director of Information Management)
NC State University
pnbhosal@ncsu.edu