Factors in Student Retention and Graduation Rate Data

By

Jahnvi Patel

Eberhardt School of Business, University of the Pacific

MSBA 285 : Capstone Project I

Dr. Lewis R. Gale

May, 2023

1. Abstract

This study aims to identify factors that impact student success in higher education by investigating the relationship between retention to the 2nd year and 4th year graduation rates. The study analyzes data from a cohort of students enrolled in a four-year degree program at a single institution. The variables examined include gender, region, high school GPA, major, ethnicity, first-term GPA, Pell Grant recipient status, athletic participation, and graduation cohort.

We employed Excel for generating pivot tables, cross-tabulation, and charts to scrutinize the data. Furthermore, we utilized RapidMiner to investigate and evaluate the data. Results show that retention to the 2nd year and the 4th year graduation rate is significantly associated with ethnicity, region, major, Pell grant & athletics. Students who participate in athletics and are Pell granted are more likely to be retained for 2nd year graduate in the 4th year.

These findings suggest that focus on diversity, academic preparedness, financial aid, and athletic participation may be important factors for both retention to the 2nd year and 4th year graduation rates. Institutions may consider implementing targeted interventions to support students who face academic and financial challenges, as well as programs that promote athletic participation and social integration on campus. Future research may investigate additional factors that impact student success, such as student engagement, institutional support, and peer networks.

2. The Problem Defined - The Introduction

The retention and graduation rates of students in higher education have become increasingly important measures of institutional success and student outcomes. The article "Multifaceted Concern" by Melanie Rivera-Cortez examines the problem of low retention rates at Wichita State University, where many students drop out before completing their degree. The university has taken steps to tackle this issue, such as providing academic support and mentorship programs. However, the article suggests that retention is a complex issue that requires addressing various factors like financial concerns, student engagement, and mental health. The article includes interviews with students and university officials discussing their perspectives and efforts to improve retention rates.

Institutions strive to retain students from year to year, and to support them in completing their degree program in a timely manner. However, not all students are successful in these goals, and there are a variety of factors that impact their ability to persist and graduate.

Using grad_6yr can provide a more comprehensive understanding of graduation rates and the factors that affect them, as it allows for a longer period for students to complete their degree. However, grad_4yr is often used as a benchmark for graduation rates, and many institutions aim to increase their graduation rates within this timeframe.

The purpose of this study is to investigate the factors that impact student retention and graduation rates, with a focus on retention to the 2nd year and 4th year graduation rates. Understanding the factors that contribute to student success is critical for institutions to develop effective interventions and support programs that can improve outcomes for all students.

The study examines a cohort of students enrolled in a four-year degree program at a single institution, and analyzes data on a variety of variables, including gender, region, high school GPA, major, ethnicity, first-term GPA, Pell Grant recipient status, athletic participation, and graduation cohort. The study aims to answer the following research questions:

- What factors are associated with retention to the 2nd year?
- What factors are associated with the 4th year graduation rate?

- Are there any differences in the factors that impact retention to the 2nd year and 4th year graduation rates?

By answering these research questions, this study aims to provide insights into the factors that impact student success and offer recommendations for institutions to improve retention and graduation rates.

3. Methods Section

The following methods were used to analyze the data in our research paper titled "Factors in Student Retention and Graduation Rate Data". The dataset consisted of 1587 students from 2005-2021.

We used Excel to create pivot tables, cross-tabulation, and charts to analyze the data. Additionally, we used RapidMiner to explore and analyze the data.

To analyze retention to the 2nd year, we examined the variable "retrn_2yr" and compared them across gender, region, major, athlete, Pell and ethnicity. We calculated probabilities and percentages to identify any patterns or relationships in the data.

Similarly, to analyze the 4th year graduation rate, we examined the variable "grad_4yr" and compared them across gender, athlete, Pell, region, major, and ethnicity. We also calculated probabilities and percentages to identify any patterns or relationships in the data.

We encountered missing values in the data, and we decided to replace them with their respective major's mean value (rounded to the nearest integer) to avoid bias. However, for the "accounting" major, we used the regional mean value as no mean value was available for that major.

4. Results Section

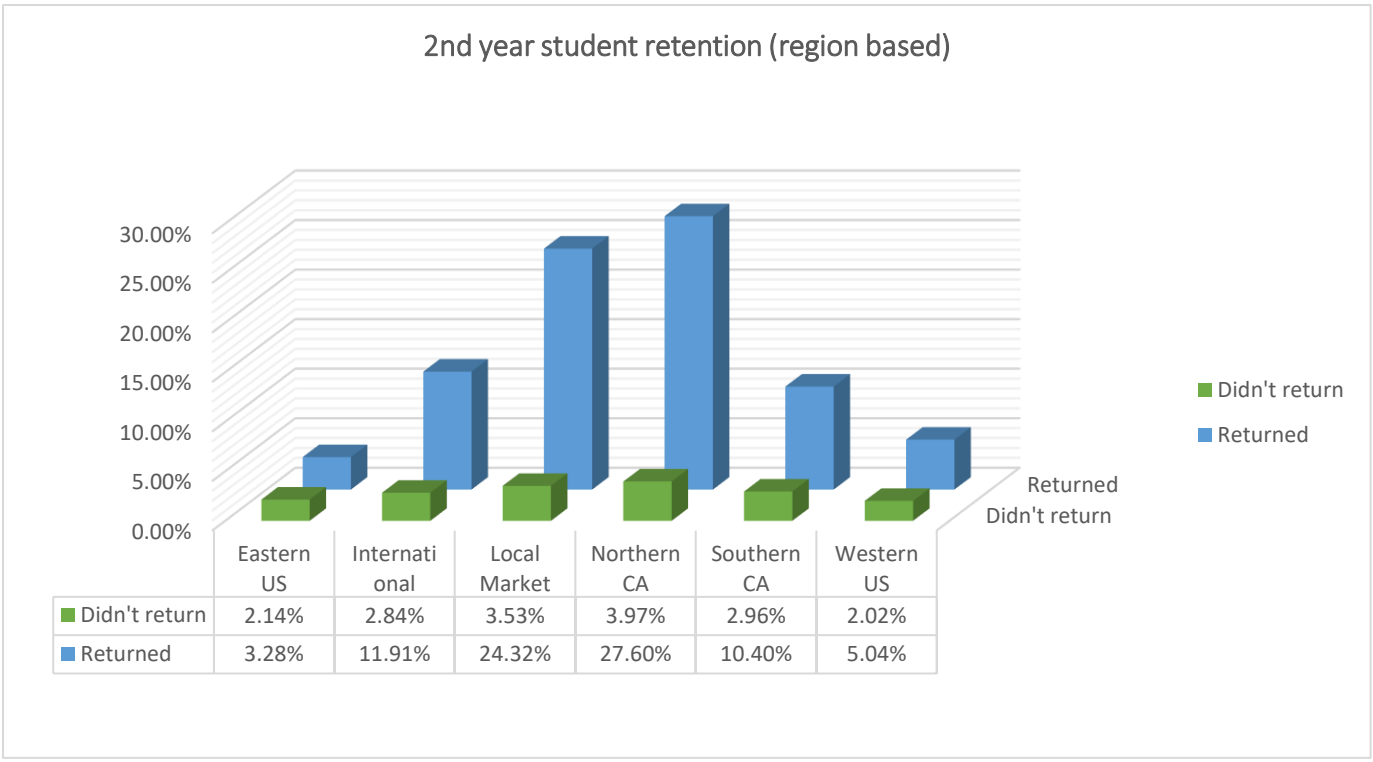
To predict 2nd year student retention, we can select the following attributes:

- term_code_key: This attribute is important to uniquely identify each student and keep track of their academic progress.
- gender: This attribute can be used to analyze the gender distribution of returning students and determine if there is any gender-based trend in retention.
- region: This attribute can help determine if students from certain regions are more likely to return for their second year of study.
- HSGPA: High School Grade Point Average (HSGPA) can be used to predict the academic performance of the student in their first year of college and their likelihood of returning for the second year.
- major: This attribute can be used to analyze the retention rates of students in different majors and identify majors with higher retention rates.

attribute	weight ↓
first_term_GPA	0.062
first_standing	0.058
region	0.026
HSGPA	0.015
ethnicity	0.009
major	0.007
Term date	0.006
athlete	0.004
gender	0.004
pell	0.002

- ethnicity: This attribute can help analyze if there are any ethnic disparities in retention rates and identify areas that need improvement.
- first_term_GPA: This attribute can be used to predict the academic performance of the student in their first year of college and their likelihood of returning for the second year.
- first_standing: This attribute can be used to analyze the academic standing of the student in their first year of college and determine if there is any correlation between academic standing and retention.
- pell: This attribute can help analyze if students receiving Pell grants have lower retention rates and identify areas that need improvement.
- athlete: This attribute can help analyze if student-athletes have higher or lower retention rates and identify areas that need improvement.
- grad_cohort: This attribute can be used to analyze if retention rates differ between different graduation cohorts.
- retrn_2yr: This attribute specifically indicates if the student returned for their second year of study or not and is the key attribute for predicting second-year student retention.

Analysis on 2nd year student retention



	Eastern US	International	Local Market	Northern CA	Southern CA	Western US
Didn't return	2.14%	2.84%	3.53%	3.97%	2.96%	2.02%
Returned	3.28%	11.91%	24.32%	27.60%	10.40%	5.04%

This table presents the percentage of college students who either returned to or did not return to a particular university, categorized by their region of origin. The regions are divided into six categories: Eastern US, International, Local Market, Northern CA, Southern CA, and Western US.

The data suggests that the highest percentage of returning students were from Northern CA (27.60%), the Local Market (24.32%), International (11.91%) and Southern CA (10.40%). Conversely, the lowest response rates were from the Eastern US (3.28%), and Western US (5.04%).

Overall, the table shows that 82.55% of students returned to university, while 17.45% did not return. These findings may be useful for educators and policymakers who are interested in understanding the factors that influence college retention rates and improving outcomes for students.

Row Labels	Eastern US	International	Local Market	Northern CA	Southern CA	Western US	Grand Total
2005	1.08%	0.00%	1.81%	1.08%	0.00%	1.08%	5.05%
2006	1.08%	0.36%	1.81%	1.08%	1.08%	1.08%	6.50%
2007	1.08%	0.00%	0.36%	2.89%	1.81%	1.08%	7.22%
2008	1.44%	1.44%	1.44%	2.17%	1.81%	2.17%	10.47%
2009	0.36%	1.81%	0.00%	0.36%	0.00%	2.17%	4.69%
2010	0.72%	1.81%	1.44%	2.89%	1.81%	1.44%	10.11%
2011	1.81%	1.08%	3.25%	1.44%	0.72%	1.08%	9.39%
2012	0.00%	1.44%	1.08%	0.72%	0.72%	0.00%	3.97%
2013	0.72%	0.36%	0.00%	1.81%	1.81%	0.00%	4.69%
2014	0.72%	2.89%	1.81%	1.44%	0.36%	0.36%	7.58%
2015	1.08%	1.08%	2.17%	3.25%	2.89%	0.00%	10.47%
2016	0.72%	0.72%	1.08%	0.72%	0.72%	0.00%	3.97%
2017	0.72%	1.44%	1.44%	1.08%	0.72%	0.36%	5.78%
2018	0.72%	0.36%	1.44%	1.44%	1.44%	0.72%	6.14%
2019	0.00%	1.44%	1.08%	0.36%	1.08%	0.00%	3.97%
Grand Total	**12.27%**	**16.25%**	**20.22%**	**22.74%**	**16.97%**	**11.55%**	**100.00%**

% of students who did not return

The table reveals that the highest percentage of students who did not return were from the Northern CA category, with 22.74% not returning. The second-highest percentage was from the Local Market category, with 20.22% not returning over the 15-year period. The lowest percentage of students who did not return were from the Western US category, with 11.55% not returning over the same period.

Row Labels	Eastern US	International	Local Market	Northern CA	Southern CA	Western US	Grand Total
2005	0.38%	0.31%	1.83%	2.44%	0.61%	0.38%	5.95%
2006	0.61%	0.69%	1.37%	3.51%	1.30%	0.38%	7.86%
2007	0.46%	0.31%	1.22%	1.83%	0.99%	0.92%	5.73%
2008	0.38%	0.76%	1.37%	2.67%	0.76%	0.31%	6.26%
2009	0.38%	0.84%	1.45%	2.29%	0.69%	0.46%	6.11%
2010	0.31%	0.61%	2.37%	2.82%	1.60%	0.46%	8.17%
2011	0.31%	1.60%	3.74%	2.60%	1.15%	0.38%	9.77%
2012	0.31%	1.68%	1.53%	1.83%	0.69%	0.23%	6.26%
2013	0.08%	1.07%	1.98%	2.52%	0.84%	0.53%	7.02%
2014	0.08%	2.82%	2.29%	2.06%	0.61%	0.38%	8.24%
2015	0.15%	1.45%	2.29%	2.21%	0.76%	0.38%	7.25%
2016	0.08%	0.69%	1.60%	1.45%	0.69%	0.08%	4.58%
2017	0.23%	0.23%	2.67%	1.98%	0.69%	0.31%	6.11%
2018	0.23%	0.92%	2.37%	1.60%	0.69%	0.38%	6.18%
2019	0.00%	0.46%	1.37%	1.60%	0.53%	0.53%	4.50%
Grand Total	**3.97%**	**14.43%**	**29.47%**	**33.44%**	**12.60%**	**6.11%**	**100.00%**

% of students who returned

The data reveals that the highest percentage of returning students came from the Northern CA market, with an average of 33.44%. The second-highest percentage of returning students came from Local Market, with an average of 29.47% over the period of 15 years.

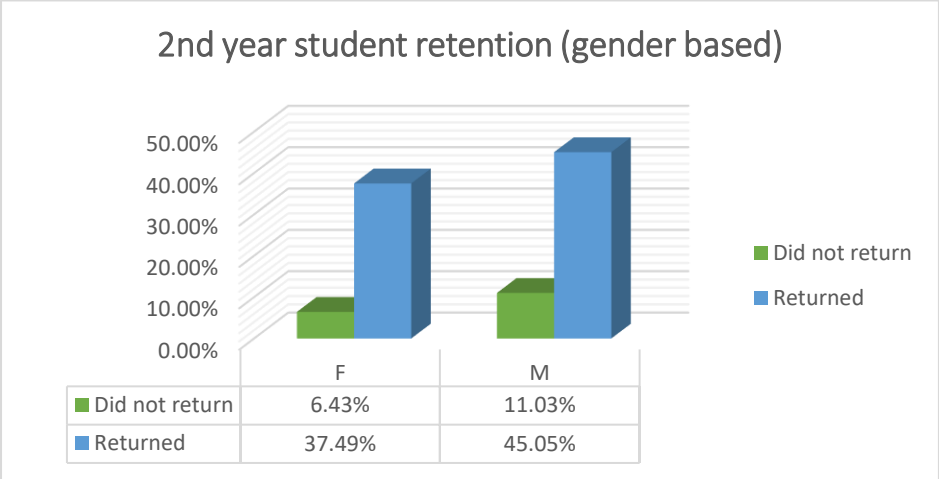
On the other hand, the lowest percentage of returning students came from the Eastern US market, with an average of 3.97%.

Out of 1587 students in their 2nd year, 82.55% (1310 students) returned and 17.45% (277 students) did not return.

2nd year	%	#
Did not return	17.45%	277
Returned	82.55%	1310
Grand Total	**100.00%**	**1587**

To get clearer picture we decided to break it down as followed:

- **Gender based.**

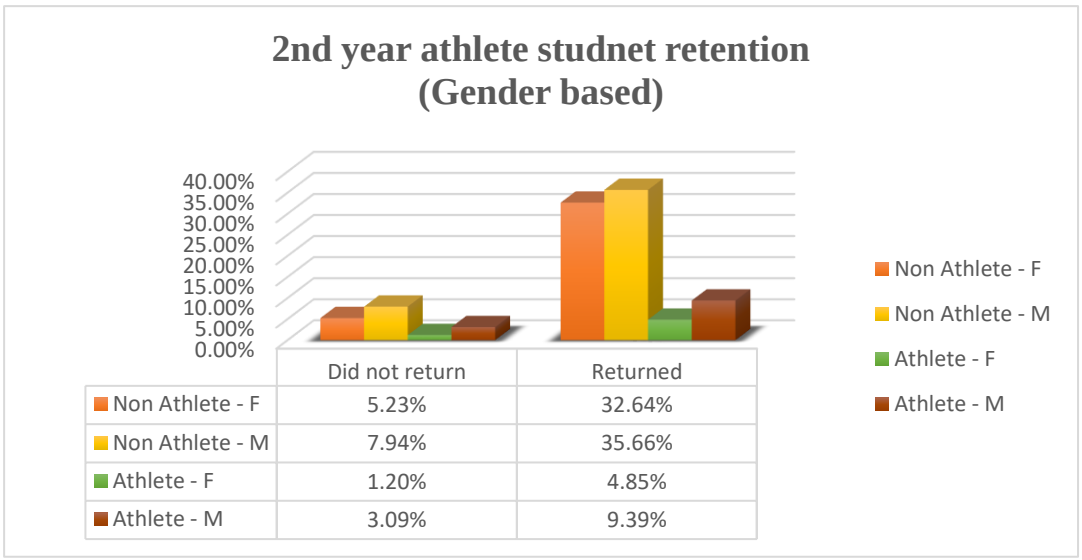


	F	M
Did not return	6.43%	11.03%
Returned	37.49%	45.05%

The probability of female students returning to the 2nd year is 85.4% and the probability of male students returning to the 2nd year is 80.4%.

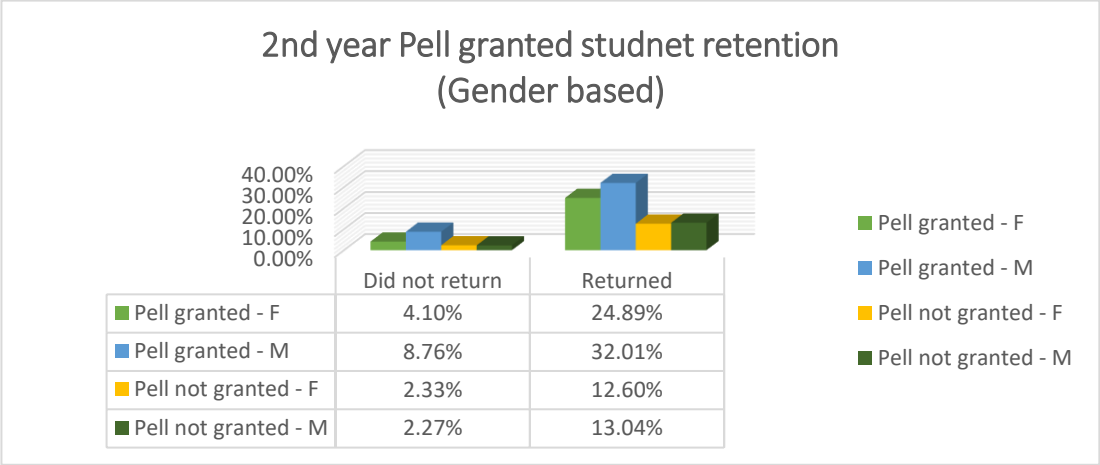
- **Gender & Athlete status based.**

The probability of female athletes returning to the 2nd year is 44.47% and the probability of male athletes returning to the 2nd year is 42.96%.



	Did not return	Returned
Non Athlete - F	5.23%	32.64%
Non Athlete - M	7.94%	35.66%
Athlete - F	1.20%	4.85%
Athlete - M	3.09%	9.39%

- **Gender & Pell grant status based.**



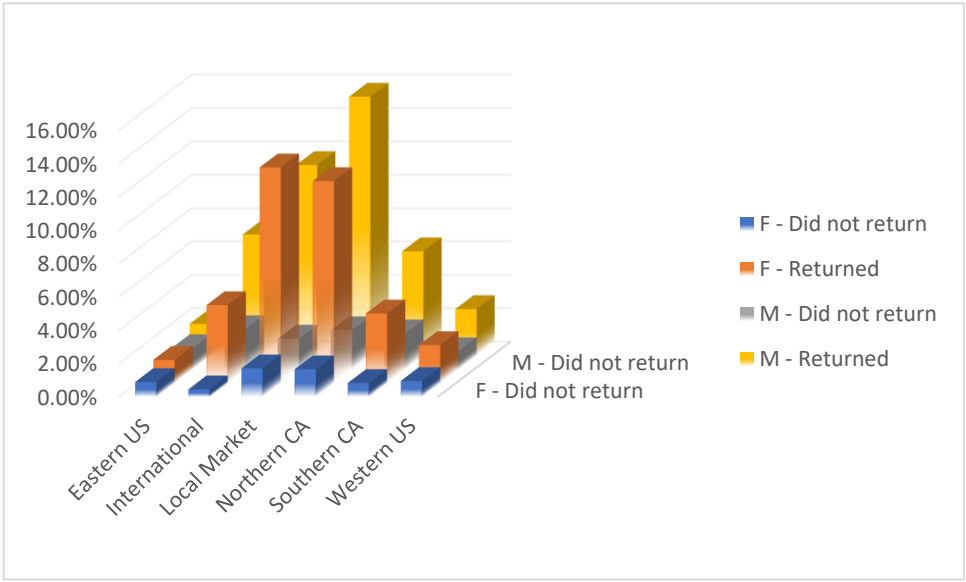
The probability of a female being Pell granted and returning to 2nd year is 85.81%, and the probability of a male being Pell granted and returning to 2nd year is 78.50%.

Furthermore, we've derived the detailed probability of students returning to 2nd year considering above 3 findings:

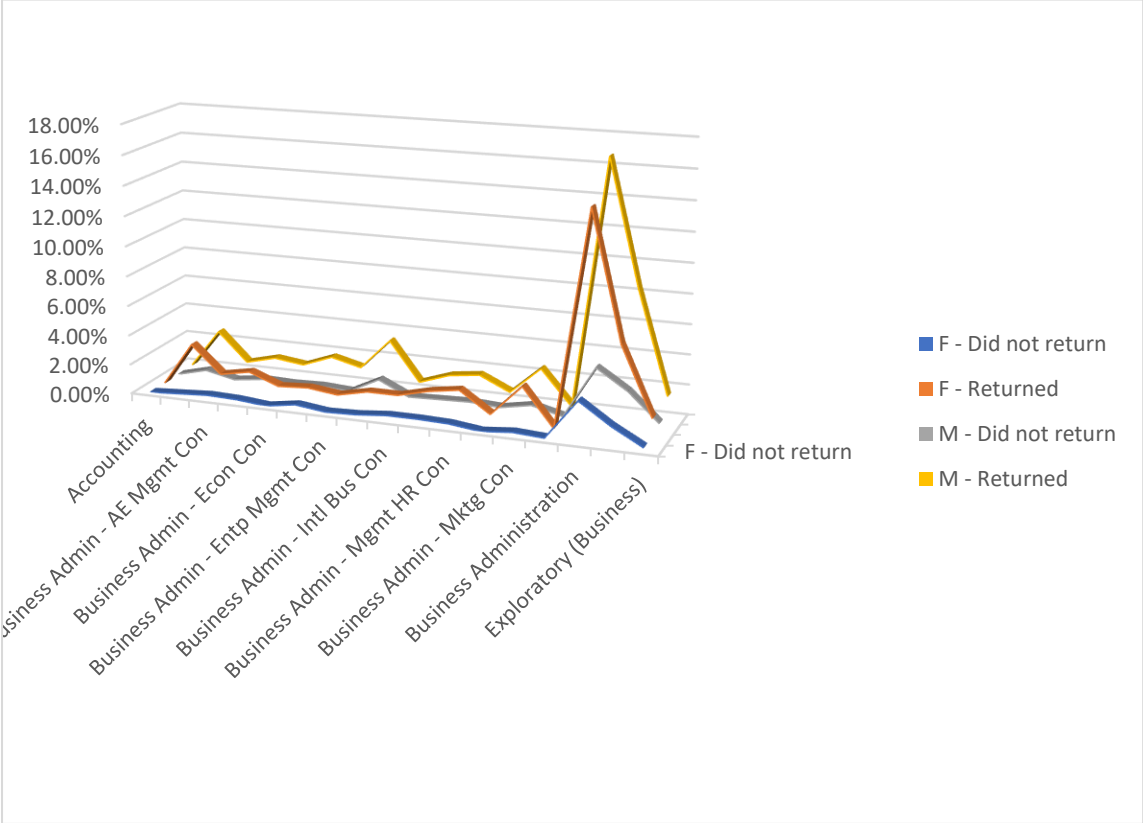
Labels	Pell granted				Pell not granted			
	F		M		F		M	
	Non-Athlete	Athlete	Non-Athlete	Athlete	Non-Athlete	Athlete	Non-Athlete	Athlete
Did not return	13.56%	16.67%	20.08%	25.61%	14.22%	41.67%	13.88%	20.59%
Returned	86.44%	83.33%	79.92%	74.39%	85.78%	58.33%	86.12%	79.41%

- **Gender & region based.**

The table provides information on the count of students who returned and did not return after two years in various regions, classified by gender. The regions are Eastern US, International, Local Market, Northern CA, Southern CA, and Western US. The data shows that the percentage of students who returned after two years is generally higher for males than females in all regions, except Local Market.



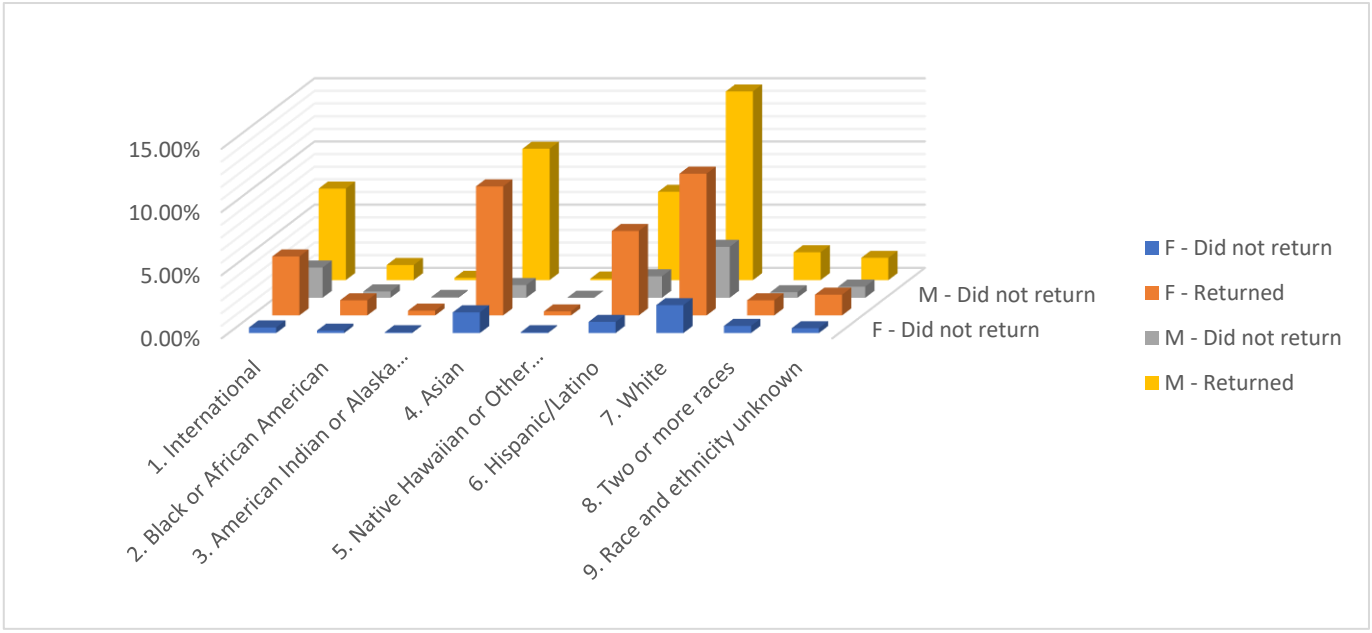
- **Gender & major based.**



The data is further broken down by majors. The majors with the highest percentage of students who returned after two years were Business Administration and Business Administration General. The majors with the lowest percentage of students who returned were Accounting and Business Admin – Econ Con.

- **Gender & ethnicity based.**

The data is broken down with each row representing a different racial/ethnic group. The table shows that the percentage of students who returned after two years varies significantly across different racial/ethnic groups, with the highest return rates among White and Asian students and the lowest among Native Hawaiian or Other Pacific Islander and American Indian or Alaska Native students.



Missing values:

When you have missing values in your dataset, it's important to carefully consider how to handle them. Ignoring them altogether may bias your analysis and potentially lead to inaccurate results. You may consider replacing the missing values with reasonable estimates or imputations, such as mean imputation or multiple imputation, to minimize the impact of missing data on your analysis.

Major	Eastern US	International	Local Market	Northern CA	Southern CA	Western US	Grand Total
Accounting			1				1
Business Admin - Acct Con	2	13	40	30	10	3	98
Business Admin - AE Mgmt Con		4	9	12	7	1	33
Business Admin - Bus Law Con	1		22	21	4	2	50
Business Admin - Econ Con	2	4	7	9	6	2	30
Business Admin - Entp Con	1	6	17	18	7	1	50
Business Admin - Entp Mgmt Con	4	3	15	3	6	1	32
Business Admin - Finc Con	5	24	21	23	10	6	89
Business Admin - Intl Bus Con	1	11	9	7	5	2	35
Business Admin - Int'l Bus Con	6	10	15	16	3	5	55
Business Admin - Mgmt HR Con	1	20	14	20	7	2	64
Business Admin - Mgmt IS Con	2	1	9	8	1		21
Business Admin - Mktg Con	4	15	22	33	10	8	92
Business Admin - Real Est Con	1			2	2		5
Business Administration	31	88	162	187	86	39	593
Business Administration Gen'l	23	34	56	96	39	36	284
Exploratory (Business)	2	1	23	16	9	4	55
Grand Total	**86**	**234**	**442**	**501**	**212**	**112**	**1587**

Missing values of "region" was replaced with the mean region with respect to the major.

Major	Average of grad_4yr
Accounting	#DIV/0!
Business Admin - Acct Con	0
Business Admin - AE Mgmt Con	0
Business Admin - Bus Law Con	0
Business Admin - Econ Con	1
Business Admin - Entp Con	1
Business Admin - Entp Mgmt Con	0
Business Admin - Finc Con	0
Business Admin - Intl Bus Con	0
Business Admin - Int'l Bus Con	0
Business Admin - Mgmt HR Con	1
Business Admin - Mgmt IS Con	1
Business Admin - Mktg Con	1
Business Admin - Real Est Con	0
Business Administration	1
Business Administration Gen'l	1
Exploratory (Business)	0
Grand Total	**0.53**

grad_4yr: No Record

Count of major	Column Label						
Major	Eastern US	International	Local Market	Northern CA	Southern CA	Western US	Grand Total
Accounting			1				1
Business Admin - Acct Con			14	6	1	1	22
Business Admin - AE Mgmt Con			2	1			3
Business Admin - Bus Law Con			7	6	2	2	17
Business Admin - Econ Con			3		2		5
Business Admin - Entp Con		1	1				2
Business Admin - Entp Mgmt Con	3	2	11	2	5	1	24
Business Admin - Finc Con	1	5	6	5	4	4	25
Business Admin - Intl Bus Con	1	6		4	3	2	16
Business Admin - Mgmt HR Con		6	2	6	2	1	17
Business Admin - Mgmt IS Con			1	5			6
Business Admin - Mktg Con		3	4	10	3	2	22
Exploratory (Business)			4	2	1	1	8
Grand Total	**5**	**23**	**56**	**47**	**23**	**14**	**168**

Missing value data for 4th year graduation.

In order to deal with the missing values in our dataset, we decided to replace them with the mean value of the respective major that the student belongs to. We rounded this mean value to the nearest integer. Although we could have used the mean values of the regions, we opted not to do so because it could potentially introduce bias into our analysis.

Region	Average of grad_4yr
Eastern US	0
International	1
Local Market	1
Northern CA	1
Southern CA	0
Western US	0
Grand Total	**0.53**

In the case of the "accounting" major, no mean value was available, so we replaced the missing values with the regional mean value (Local Market).

To predict 4th-year graduation rates, we can select the following attributes:

- term_code_key: This attribute is important to uniquely identify each student and keep track of their academic progress.
- region: This attribute can help determine if students from certain regions are more likely to graduate in their fourth year of study.
- HSGPA: High School Grade Point Average (HSGPA) can be used to predict the academic performance of the student in their first year of college and their likelihood of graduating in their fourth year.
- major: This attribute can be used to analyze the graduation rates of students in different majors and identify majors with higher graduation rates.

attribute	weight ↓
retrn_3yr	0.347
retrn_4yr	0.259
grad_4yr	0.147
first_term_GPA	0.062
first_standing	0.058
region	0.026
HSGPA	0.015
ethnicity	0.009
major	0.007
Term date	0.006
athlete	0.004
gender	0.004
pell	0.002
grad_cohort	0

- ethnicity: This attribute can help analyze if there are any ethnic disparities in graduation rates and identify areas that need improvement.
- first_term_GPA: This attribute can be used to predict the academic performance of the student in their first year of college and their likelihood of graduating in their fourth year.
- first_standing: This attribute can be used to analyze the academic standing of the student in their first year of college and determine if there is any correlation between academic standing and graduation rates.
- pell: This attribute can help analyze if students receiving Pell grants have lower graduation rates and identify areas that need improvement.
- athlete: This attribute can help analyze if student-athletes have higher or lower graduation rates and identify areas that need improvement.
- grad_cohort: This attribute can be used to analyze if graduation rates differ between different graduation cohorts.
- retrn_4yr: This attribute is important to determine if the student has graduated in their fourth year or not.

The return rates of the earlier years can help identify students who may be at risk of not graduating in their 4th year and targeting interventions towards those students may improve overall graduation rates. However, if you must choose a limited number of attributes, then the attributes listed above are likely to be the most important ones for predicting 4th-year graduation rates.

Analysis on 4th year graduated students

The data table specifically shows the number and percentage of students who did not graduate versus those who graduated, among a total of 1587 students who were in their fourth year of study.

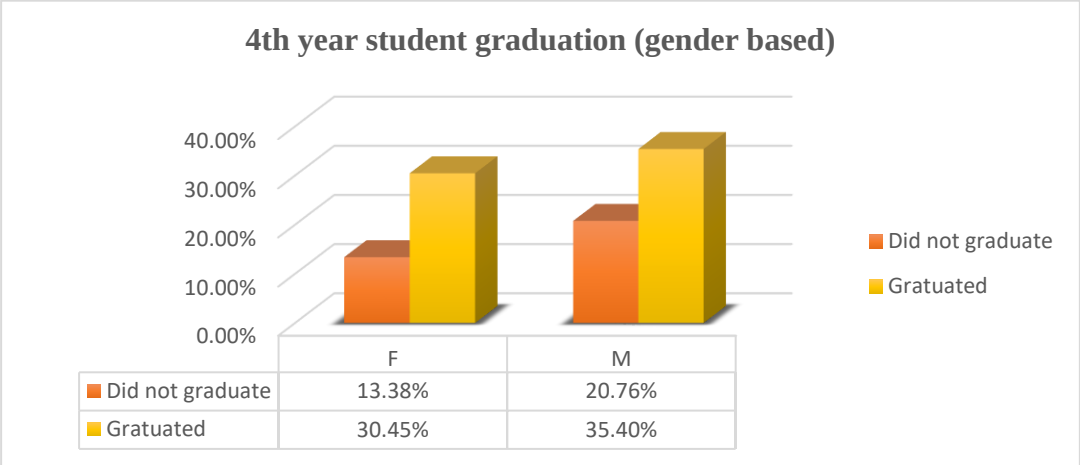
4th year	#	%
Did not graduate	781	49.21%
Graduated	806	50.79%
Grand Total	**1587**	**100.00%**

According to the data, 781 students (or 49.21% of the total) did not graduate, while 806 students (or 50.79% of the total) did graduate.

To get clearer picture we decided to break it down as followed:

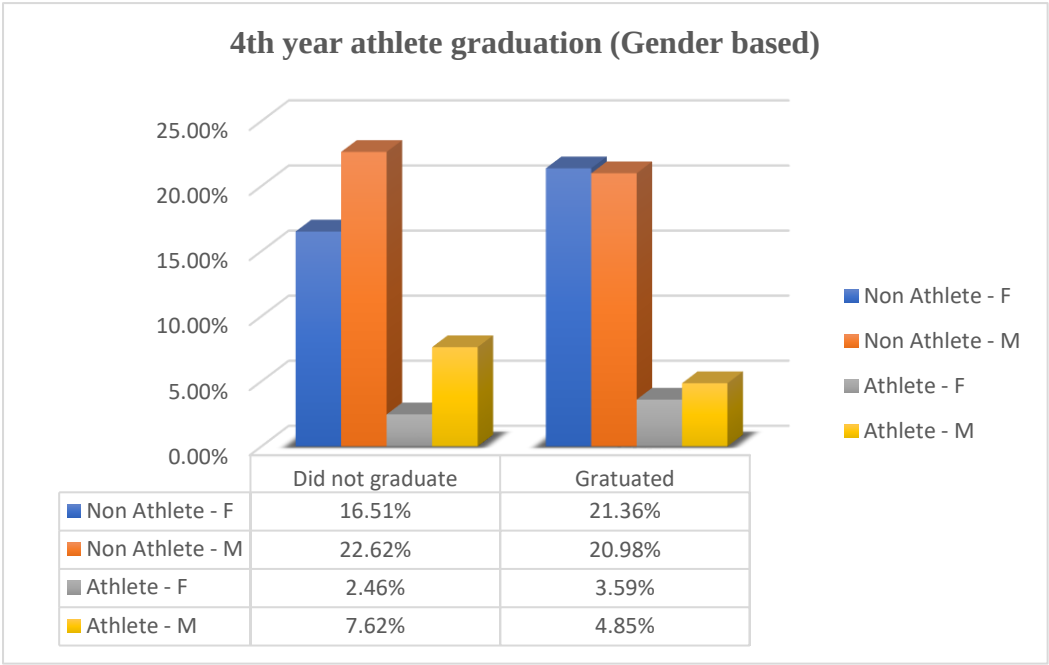
- **Gender based.**

The probability of a female student graduating 4th year is approximately 69.48% and the probability of a male student graduating 4th year is approximately 63.01%.



	F	M
Did not graduate	13.38%	20.76%
Gratuated	30.45%	35.40%

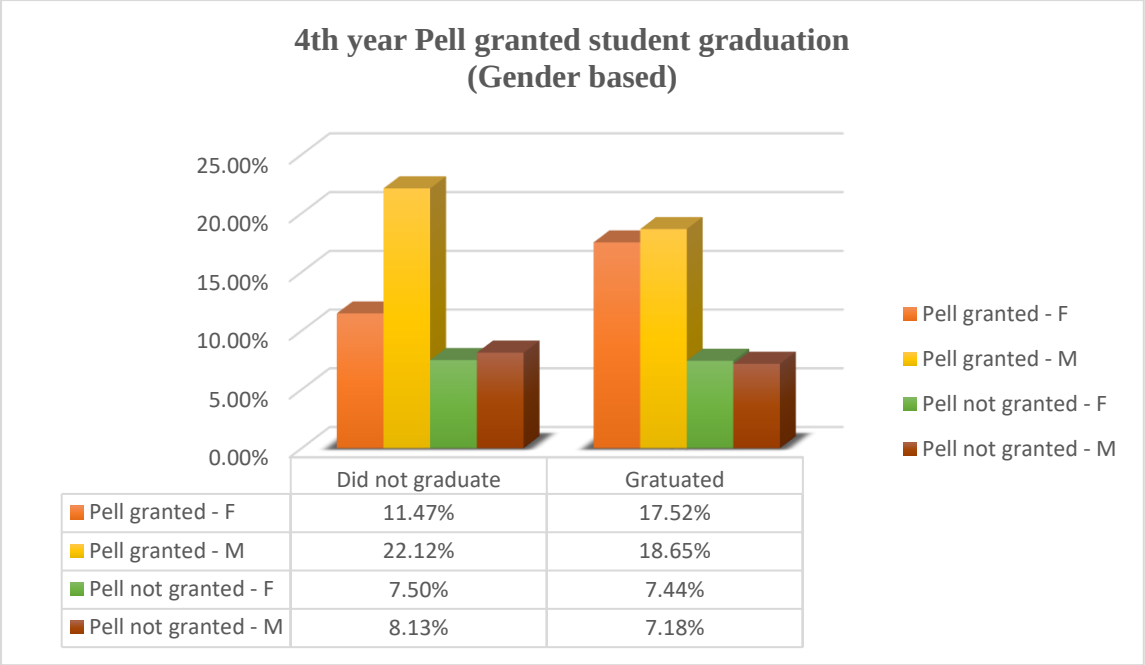
- **Gender & Athlete status based.**



	Did not graduate	Gratuated
Non Athlete - F	16.51%	21.36%
Non Athlete - M	22.62%	20.98%
Athlete - F	2.46%	3.59%
Athlete - M	7.62%	4.85%

The probability of a female athlete student graduating 4th year is approximately 59.26% and the probability of a male athlete student graduating 4th year is approximately 38.86%. And the probability of a female non-athlete student graduating 4th year is approximately 56.39% and the probability of a male non-athlete student graduating 4th year is approximately 48.03%.

- **Gender & Pell grant status based.**



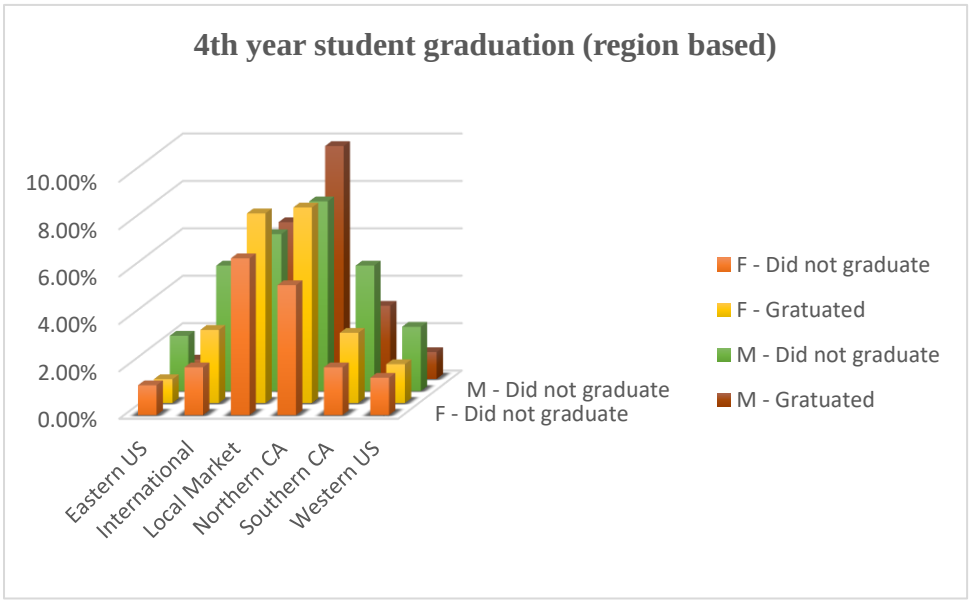
	Did not graduate	Gratuated
Pell granted - F	11.47%	17.52%
Pell granted - M	22.12%	18.65%
Pell not granted - F	7.50%	7.44%
Pell not granted - M	8.13%	7.18%

The probability of a female Pell granted student graduating 4th year is approximately 60.46% and the probability of a male Pell granted student graduating 4th year is approximately 45.75%. And the probability of female non-Pell granted students graduating 4th year is 12.59%, and the probability of male non-Pell granted students graduating 4th year is 17.51%.

Furthermore, we've derived the detailed probability of students graduating after 4th year considering above 3 findings:

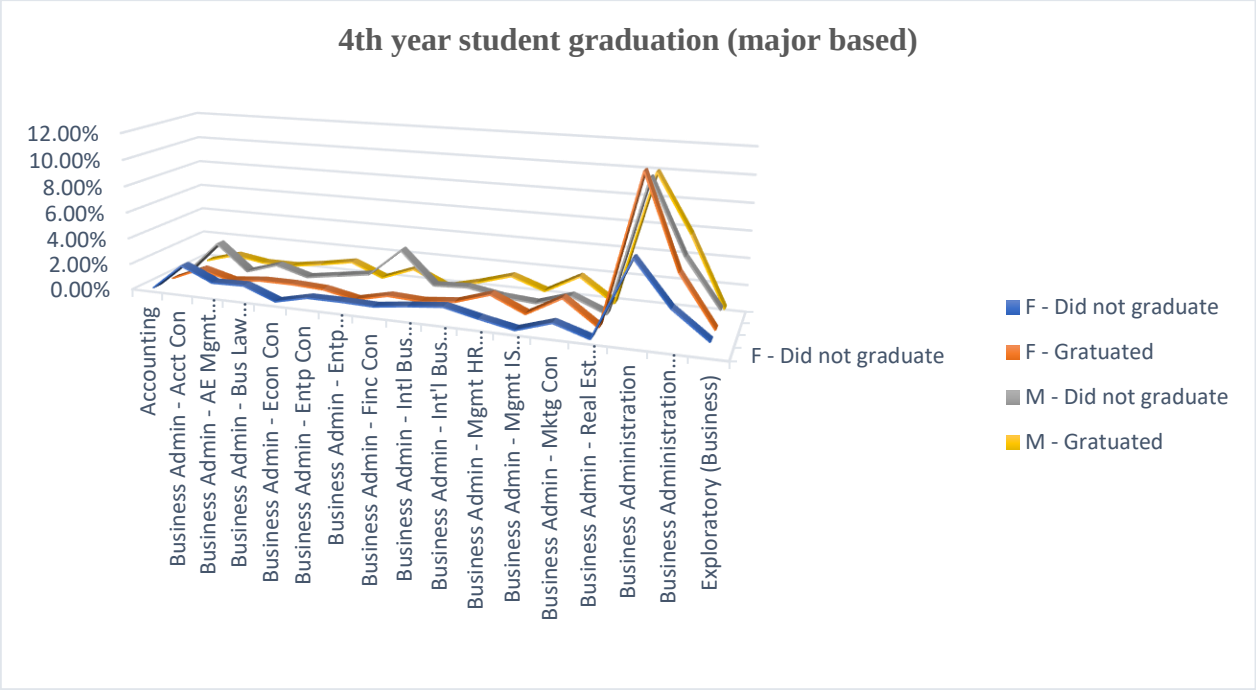
Labels	Pell granted				Pell not granted			
	F		M		F		M	
	Non Athlete	Athlete	Non Athlete	Athlete	Non Athlete	Athlete	Non Athlete	Athlete
Did not graduate	39.89%	38.10%	51.97%	60.98%	49.78%	58.33%	51.67%	61.76%
Gratuated	60.11%	61.90%	48.03%	39.02%	50.22%	41.67%	48.33%	38.24%

- **Gender & region based.**



The table shows the distribution of students who did and did not graduate from a 4-year program, disaggregated by gender (F and M) and region (Eastern US, International, Local Market, Northern CA, Southern CA, and Western US). The data shows that the highest percentage of graduates comes from the Northern CA region for both genders, with 8.25% of females and 9.83% of males graduating followed by the Local Market. while the Eastern US region has the lowest graduation rates.

- **Gender & major based.**



This table provides information on the percentage of students who graduated in four years, broken down by gender and by major within the Business Administration program. Here are some possible interpretations:

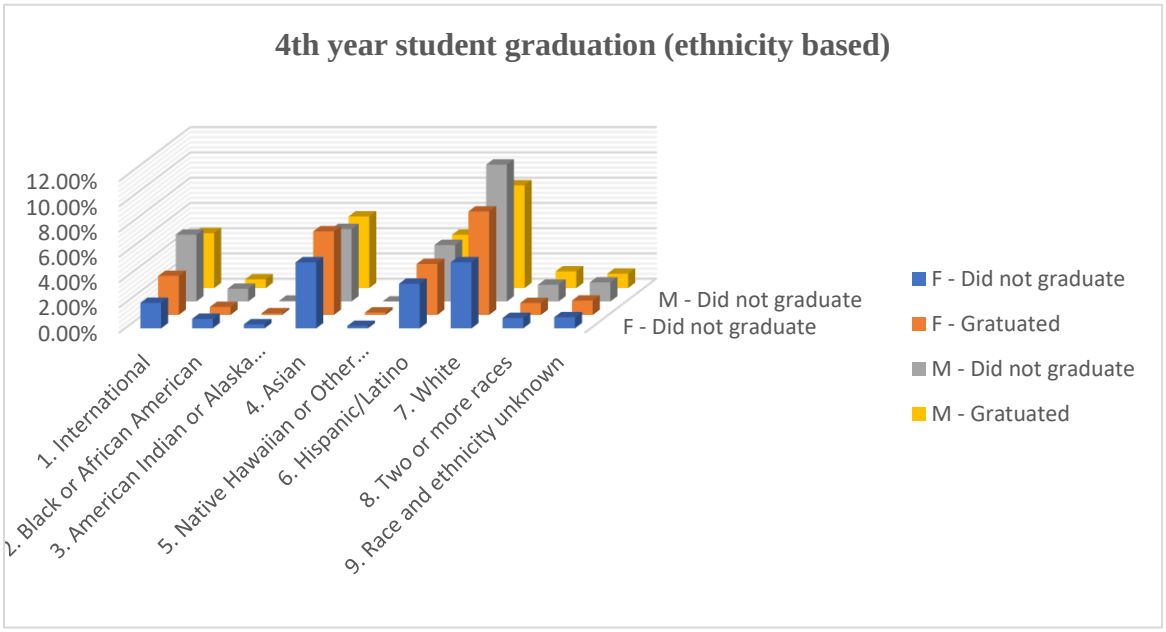
The overall graduation rate for the Business Administration program is 25.83% for male students and 24.95% for female students.

Among the specific majors, the highest graduation rate for male students is in the Business Administration, at 10.08%. For female students, the highest graduation rate is 11.15%.

There is a major where no female students graduated in four years (e.g. Accounting), whereas for male students, every major had at least one graduate.

- **Gender & ethnicity based.**

Overall, the table indicates that a higher percentage of males graduated in 4 years compared to females. Among racial/ethnic groups, White students had the highest percentage of graduates, while American Indian or Alaska Native & Native Hawaiian or Other Pacific Islander students had the lowest percentage.



5. Discussion Section

There are various factors that can impact student retention within academic institutions. It is important to identify the factors that are most relevant to a specific institution, as these may differ depending on the institution's location, country, and the programs offered.

Ethnicity and region have a significant impact on student retention and graduation rates. Research has shown that students from certain ethnic and regional backgrounds are more likely to drop out or not graduate on time compared to their counterparts. Therefore, it is crucial for institutions to take into consideration the diverse backgrounds of their students and provide appropriate support. For instance, providing culturally competent academic advising, mentoring programs, and financial aid opportunities can positively impact the retention and graduation rates of students from diverse backgrounds. Additionally, creating a welcoming campus environment that celebrates diversity and provides a sense of belonging can help students feel more connected to the institution, and thus, more likely to stay and graduate.

We discovered that female students are less likely to return for their second year and hence graduate less than male students after 4th year, thus we should recommend special tactics to boost female retention rates.

- Improve communication: Make sure that communication with female students is clear, frequent, and in a variety of formats. This could include newsletters, emails, social media, and face-to-face meetings.
- Create a supportive environment: Ensure that female students feel supported and valued. This could include creating mentoring programs, encouraging peer-to-peer support networks, and offering resources that address female-specific concerns.
- Provide flexible learning options: Offer flexible learning options, such as online courses or part-time programs, that accommodate the needs of female students who may have additional caregiving responsibilities or other time constraints.
- Address financial concerns: Provide financial aid options or scholarships to alleviate any financial concerns that female students may have.
- Address barriers to academic success: Address any barriers that female students may face in achieving academic success, such as addressing gender bias, providing tutoring or academic support, and offering workshops on time management and study skills.
- Conduct outreach: Engage in targeted outreach efforts to encourage female students to participate in surveys and other forms of data collection, to better understand their needs and to inform programmatic interventions.

Financial aid and grants are crucial factors for student retention in higher education institutions. The cost of tuition plays a vital role in students' decision to continue their education or drop out entirely. Providing financial support in the form of scholarships, discounts, and grants can motivate students to continue their studies. To ensure student retention, institutions must provide comprehensive support to their students throughout their academic journey. This support should start from the time of enrollment and include monitoring student activities and involvement, which will enable the institution to provide the appropriate support to retain students based on their interests.

6. Reference

Rivera-Cortez, Melanie. "'Multifacted Concern': What Student Retention Looks like at Wichita State." The Sunflower, thesunflower.com/71584/news/retention-is-a-multifacted-concern-at-wichita-state/.