Postsecondary Institutions Expenses and Student Attendance Costs: Trends and Impact on Degree Completion: 2015-2021

Melinda Fang, Harrison Jia, Lingchao Mao, Patricia Zhou
December 4, 2022

1 Executive Summary

In this report we analyze the trends of institutional expenses, cost of attendance, and financial aid over 2015-2021 across all degree-offering post-secondary institutions in the United States. We use institutional expenses and costs items that have statistically significant impacts on college completion for different student subpopulations. This analysis will better inform universities, baccalaureate, and associate colleges the short-term impacts of their cost allocation on completion rates and provide insights for future allocation.

1.1 Background

Postsecondary education has never been more important as a key venue to educate our workforce over the future decades. As the gap between earnings of high school graduates and university graduates continue to grow (ref), increasing college completion is a common goal for students, institutions, and societal organizations. However, increasing completions is not simply a matter of increasing enrollment. Recent studies showed that 39 million Americans hold some postsecondary education or training without completion. About one out of three students drop out from college in the US, and two out of five were due to financial issues. Students who drop-out of college are almost 100 times more likely to default on their student loans, end up with debt and no degree. (link) (link2)

Institutional expenses can impact college completion from different perspectives. From the institution's side, the amount of money invested in different aspects of the institution's development directly affects the availability and quality of educational resources. From the student's angle, the cost of attendance and financial aid are key players for affordability. With this in mind, we propose the following questions:

> *What are recent trends in institutional expenses and attendance costs?*
> *How have the changes in institutional financials and attendance costs impacted post-secondary education completion, especially the underrepresented communities?*

Understanding the impact of the increase/decrease of these expense or cost items on completion rate/drop out rate is important for understanding their short-term impacts on completion rates and provide insights for future allocation.

1.2 Summary of key results

For the expenses/cost items that are both important for completion and increasing significantly, all top 3 of them are all awards/grants related. Financial aid largely showed effects on different student subpopulations, helping more underrepresented student groups. Public and private non-profitable schools should keep investing in offering students, especially undergraduates, grant fundings, as the higher number of grants recipients leads to a higher college completion rate.

For the expenses/cost items that are important for completion but currently showing a decreasing or a slower increasing trend (<4% annual increase), we suggest spending money dedicated to instructional use and institutional support. For private non-profitable schools, the fields like auxiliary enterprises and equipment (e.g. art, library, etc.) are expected to be invested for a long run. For public schools, we suggest the decision makers to support a wide range of purposes across the institution, with a vast majority of funds dedicated to administrative and academic support. Private for-profit schools should distribute money to academic support too.

Regardless of the changing trend, public service and student service are considered not important when it comes to college completion rate, therefore we think that schools are able to rationally cut fundings towards those areas. From the student's costs perspective, tuition was shown to have a significant effect on the completion rate of pirate non-profit institutions, and off-campus room and board expenses for all institutions in general.

2 Technical Exposition

In this section, we describe the statistical methods we used and walkthrough our findings in detail. We start by summarizing the temporal trend of institutional finances, student financial aid, and student fees over the recent five years (Part I). Then, we sought to answer the question "*How does cost impact college completion?*" using regression-based hypothesis testing (Part II), which an emphasis on cost items that impacted the underrepresented student population. Lastly, we use machine learning to model the complex relationships among the different types of data (institutional characteristics, institutional finances, financial aid, student fees) to predict completion (Part III). Feature contribution analysis of the machine learning models reveals the key predictors of completion.

2.1 Data Collection and Preprocessing

2.1.1 Data Sources

We used datasets published by the National Center for Education Statistics (NCES) which was a collection of surveyed information on US postsecondary education institutions between 2015-2021. Specifically, our analysis involved the following categories of datasets: Completions, Graduation Rates, Institutional Characteristics, Institution Finances, Student Financial Aid, and Outcome Measures. To facilitate our interpretations of the data above, we also looked at United States annual CPI data from FRED.

2.1.2 Institution Inclusion & Exclusion

Among all the post-secondary institutions present in NCES database, we excluded the institutions that satisfy any of the following criteria, for the reason that their financial pattern may differ significantly from the active universities, baccalaureate, or associate colleges:

- Inactive, new or merged with other institutions in 2021
- Not degree-granting institutions
- All programs offered completely via distance education
- Special focus institutions based on the Carnegie classification

After applying the filtering criteria above, we analyze a total of 2,511 institutions, of which 57% were public institutions, 56% offered degrees of primarily baccalaureate or above (Table 1). In terms of special services, 98% had programs for veterans and their families, 73% were members of an athletic association, and almost every institution has (had) counseling, employment, placement, and library services.

Table 1. Summary characteristics of the included institutions

	Characteristic	Number of institutions	% of total
	All institutions	2511	
Control	Public	1439	57.3%
	Private not-for-profit	928	37.0%
	Private for-profit	144	5.7%
Category	Primarily baccalaureate or above	1396	55.6%
	Associate's and certificates	862	34.3%
	Not primarily baccalaureate or above	245	9.8%
	Graduate with no undergraduate degrees	7	0.3%
Special Learning programs	Study abroad	1623	64.6%
	ROTC	946	37.7%
	Distance course	2330	92.8%
	Distance program	1905	75.9%
	Part-time	2405	95.8%
Student registered with disablities	>3%	8	0.3%
	less than 3%	1185	47.2%
	None or not applicable	1312	52.3%
Programs and Affiliations	Member of National Athletic Association, football, basketball, cross-country*	1825	72.7%
	Programs for veterans and their families**	2448	97.5%
Services	Remedial	1932	76.9%
	Counseling	2495	99.4%
	Employment	2352	93.7%
	Day care	733	29.2%
	Placement	2089	83.2%
	On campus housing	1600	63.7%
	Hospital	53	2.1%
	Library	2498	99.5%

2.1.3 Outcome Measures

We choose students' completion as the outcome of interest. We measure completion from the following perspectives:

(1) *Completion*: total number of degrees/awards offered at a given year;
(2) *Completion rate*: percentage of students of a given cohort graduated within 150% of normal time.

Information on students who did not complete the study can be further measured by:

(3) *Drop out rate*: percentage of students who did not graduate within 150% of normal completion time and are no longer enrolled;
(4) *Transfer out rate*: percentage of students who transferred out to another institution;
(5) Still enrolled but not completed on time.

Our analysis focuses on the outcome measures (1) to (2). This analysis can be extended to (3)-(5) in future work.

2.2 Trends over five recent years

An overview of annual change rates of institutional expenses, tuition costs, and total grant aids with respect to the numbers in 2015 are shown in Figure 1. The three trends lines were compared them with the change rate of United States Consumer Price Index (CPI). The published out-of-state tuition and fee increase rate is similar to the CPI, whereas the institutional expenses increased at a slower rate than CPI, and the total grant aid awarded to undergraduate students even slower and with large fluctuations. In the subsequent sections, we provide a more thorough view of spatio-temporal trends of institutional expenses and student attendance costs.

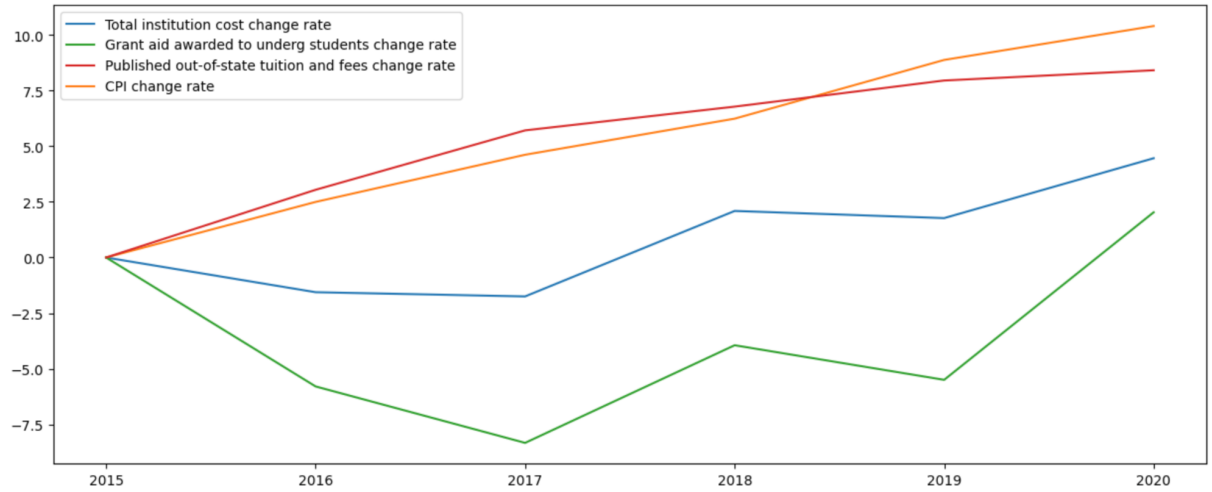

Figure 1. Trends of total institution cost, grant aids, and out-of-state tuition (relative to 2015) in comparison with CPI for public institutions

2.2.1 Institutional Expenses

In Table 2 to Table 4, after categorizing all institutions into public, private for-profit, and private not-for-profit, we summarize the five-year trend (2015 - 2020) of institution-reported financials for the seven big expense categories as well as grants. For all three categories of institutions, *Instruction,* followed by *instructional support* were the categories with the greatest expense amounts.

For public institutions, we observe that *instruction and institutional support* expenses have steadily increased across the years. *Student services* expenses had an increase slow down in 2019 and 2020, which might be due to reduced and/or restricted on- and off-campus student activities, cultural events, and athletics events due to COVID. *Auxiliary enterprise expenses*, which covers expenses related to residence halls, food services, student health services, faculty and staff parking/housing, is the only category that decreased across the years and noticeably, has a doubled decreasing rate in 2020. *Scholarship/fellowship* incurred expenses as well as institutional *grants* exhibit large fluctuations.

Table 2. Public institution's financial expenses (2015-2020) summarized as median amount across all institutions and change rate with respect to the previous year

Category	Expense Name	2015		2016		2017		2018		2019		2020	
		Amount		Amount	Change	Amount	Change	Amount	Change	Amount	Change	Amount	Change
Instruction	Current year total	18,004,721	-	17,514,724	-2.7%	18,056,402	3.1%	18,314,640	1.4%	18,386,901	0.4%	18,647,768	1.4%
	Salaries and wages	9,946,383	-	10,174,616	2.3%	10,307,035	1.3%	10,470,554	1.6%	10,593,305	1.2%	10,630,983	0.4%
Institutional Support	Current year total	6,691,303	-	6,594,253	-1.5%	6,920,047	4.9%	7,113,420	2.8%	7,324,796	3.0%	7,482,997	2.2%
	Salaries and wages	2,618,707	-	2,726,337	4.1%	2,865,155	5.1%	2,987,988	4.3%	3,054,255	2.2%	3,177,210	4.0%
Student Services	Current year total	4,561,653	-	4,569,822	0.2%	4,815,047	5.4%	5,013,106	4.1%	5,136,941	2.5%	5,141,970	0.1%
	Salaries and wages	2,197,460	-	2,228,686	1.4%	2,353,456	5.6%	2,465,754	4.8%	2,570,889	4.3%	2,586,756	0.6%
Scholarships & Fellowship Expenses	Current year total	3,211,238	-	2,887,964	-10.1%	2,820,478	-2.3%	2,928,742	3.8%	2,955,997	0.9%	3,648,863	23.4%
	Total gross scholarships and fellowships	9,752,022	-	9,215,261	-5.5%	9,049,198	-1.8%	9,426,125	4.2%	9,517,126	1.0%	10,029,996	5.4%
Academic Support	Current year total	3,694,728	-	3,759,797	1.8%	3,789,349	0.8%	3,908,351	3.1%	3,894,877	-0.3%	4,045,553	3.9%
	Salaries and wages	1,703,700	-	1,775,318	4.2%	1,779,558	0.2%	1,828,506	2.8%	1,899,967	3.9%	1,973,789	3.9%
Auxiliary Enterprises	Current year total	2,921,024	-	2,774,338	-5.0%	2,856,363	3.0%	2,761,637	-3.3%	2,685,691	-2.8%	2,486,521	-7.4%
	Salaries and wages	519,191	-	537,315	3.5%	542,988	1.1%	539,609	-0.6%	541,154	0.3%	522,576	-3.4%
Public Service	Current year total	209,984	-	214,023	1.9%	224,677	5.0%	235,517	4.8%	226,031	-4.0%	238,078	5.3%
	Salaries and wages	77,252	-	72,623	-6.0%	71,929	-1.0%	76,788	6.8%	77,000	0.3%	82,375	7.0%
Grants	Federal nonoperating grants	6,829,311	-	6,206,913	-9.1%	5,851,205	-5.7%	6,166,174	5.4%	6,149,517	-0.3%	6,857,412	11.5%
	Federal operating grants and	1,031,706	-	1,012,234	-1.9%	980,823	-3.1%	1,081,972	10.3%	1,054,324	-2.6%	1,270,223	20.5%
	Grants by state government	502,286	-	527,885	5.1%	571,291	8.2%	652,227	14.2%	712,387	9.2%	799,609	12.2%
	Institutaionl grants	388,335	-	437,233	12.6%	475,788	8.8%	537,379	12.9%	572,311	6.5%	669,682	17.0%
	Local/private operating grants	93,637	-	94,114	0.5%	96,862	2.9%	105,366	8.8%	117,168	11.2%	110,774	-5.5%
	Other federal grants	174,037	-	174,601	0.3%	172,004	-1.5%	183,091	6.4%	205,970	12.5%	563,745	173.7%
	Pell grants (federal)	6,370,288	-	5,776,781	-9.3%	5,412,361	-6.3%	5,710,489	5.5%	5,707,081	-0.1%	5,495,532	-3.7%
	State operating grants and	512,077	-	483,005	-5.7%	476,879	-1.3%	528,640	10.9%	533,834	1.0%	552,838	3.6%

The expense landscape of for-profit institutions is noticeably different from the one above. While expenses of public and non-profit private institutions generally increase over the years, the private for-profit institutions appear to manage to decrease most expenses.

Table 3. For-profit private institutions' financial expenses (2015-2020) summarized as median amount across all institutions and change rate with respect to the previous year ("year" being the ending financial year)

Category	Item Name	2015		2016		2017		2018		2019		2020	
		Amount		Amount	Change	Amount	Change	Amount	Change	Amount	Change	Amount	Change
Instruction	Current year total	751,616	-	691,168	-8.0%	664,206	-3.9%	585,272	-11.9%	592,581	1.2%	571,742	-3.5%
	Salaries and wages	423,500	-	370,491	-12.5%	367,752	-0.7%	326,873	-11.1%	325,000	-0.6%	340,958	4.9%
Institutional Support	Current year total	431,003	-	406,507	-5.7%	387,985	-4.6%	339,089	-12.6%	315,061	-7.1%	319,860	1.5%
	Salaries and wages	162,007	-	149,018	-8.0%	135,412	-9.1%	121,263	-10.4%	112,096	-7.6%	112,096	0.0%
Academic Support	Current year total	114,709	-	112,829	-1.6%	109,033	-3.4%	93,791	-14.0%	90,469	-3.5%	86,555	-4.3%
	Salaries and wages	54,245	-	52,406	-3.4%	48,011	-8.4%	42,000	-12.5%	39,237	-6.6%	39,473	0.6%
Grants	Total student grants	657,857	-	578,147	-12.1%	538,137	-6.9%	495,084	-8.0%	493,099	-0.4%	511,415	3.7%
	Pell grants	580,613	-	496,764	-14.4%	465,497	-6.3%	434,918	-6.6%	430,964	-0.9%	437,640	1.5%
	Other federal grants	5,366	-	3,750	-30.1%	2,809	-25.1%	0	-100.0%	0	-	0	-

Similar to the public institutions, non-profit private institutions show steadily increasing expenses over the years. The fastest increase was seen in *auxiliary enterprise* expenses, with 4.9% increase in 2019 and 8.5% in 2020. This suggests that non-profit institutions recently increased investment in self-supporting operations such as residence halls, food services, student health services, faculty and staff housing, etc. Similar to the public institutions, financials related to grants exhibit large fluctuations.

Table 4. Non-profit institutions' financials expenses (2015-2020) summarized as median amount across all institutions and change rate with respect to the previous year

Category	Item Name	2015		2016		2017		2018		2019		2020	
		Amount		Amount	Change	Amount	Change	Amount	Change	Amount	Change	Amount	Change
Academic Support	salaries and wages	632,501	-	686,888	8.6%	684,872	-0.3%	722,342	5.5%	744238	3.0%	758,000	1.8%
	total amount	1,472,000	-	1,545,289	5.0%	1,548,832	0.2%	1,608,511	3.9%	1678402	4.3%	1,680,880	0.1%
Auxiliary Enterprises	salaries and wages	148,915	-	158,361	6.3%	159,758	0.9%	156,092	-2.3%	163740	4.9%	177,668	8.5%
	total amount	2,206,743	-	2,227,162	0.9%	2,249,330	1.0%	2,230,870	-0.8%	2411047	8.1%	2,388,691	-0.9%
Institutional Support	salaries and wages	1,628,670	-	1,753,944	7.7%	1,793,219	2.2%	1,857,282	3.6%	1920373	3.4%	1,938,072	0.9%
	total amount	4,545,586	-	4,664,393	2.6%	4,741,189	1.6%	4,847,273	2.2%	5174516	6.8%	5,088,260	-1.7%
Instruction	salaries and wages	3,772,326	-	4,002,546	6.1%	4,063,547	1.5%	4,245,132	4.5%	4442516	4.6%	4,497,858	1.2%
	total amount	6,426,170	-	6,837,172	6.4%	6,974,828	2.0%	7,357,899	5.5%	7576673	3.0%	7,648,981	1.0%
Grants	total student grants	6,076,481	-	6,251,361	2.9%	6,906,706	10.5%	7,857,608	13.8%	8119368	3.3%	8,338,630	2.7%
	federal grants and contracts - total	222,115	-	217,435	-2.1%	196,064	-9.8%	196,623	0.3%	216764	10.2%	761,560	251.3%
	federal grants and contracts - unrestricted	94,189	-	106,123	12.7%	103,530	-2.4%	94,454	-8.8%	106569	12.8%	391,921	267.8%
	institutional grants (funded)	324,000	-	358,566	10.7%	337,906	-5.8%	380,284	12.5%	404810	6.4%	419,619	3.7%
	institutional grants (unfunded)	1,553,878	-	1,782,641	14.7%	2,235,708	25.4%	2,753,633	23.2%	2983154	8.3%	2,954,955	-0.9%
	other federal grants	97,585	-	102,938	5.5%	100,429	-2.4%	109,530	9.1%	119565	9.2%	196,368	64.2%
	pell grants	1,218,181	-	1,226,728	0.7%	1,239,020	1.0%	1,317,794	6.4%	1348037	2.3%	1,334,749	-1.0%
	private gifts, grants and contracts - temp. restricted	402,855	-	432,206	7.3%	449,328	4.0%	498,111	10.9%	418091	-16.1%	434,928	4.0%
	private gifts, grants, and contracts - perm. restricted	62,676	-	70,831	13.0%	74,667	5.4%	78,074	4.6%	16099	-79.4%	6,248	-61.2%
	private gifts, grants, and contracts - total	1,900,648	-	2,041,484	7.4%	2,100,292	2.9%	2,209,112	5.2%	2206016	-0.1%	2,286,307	3.6%
	private gifts, grants, and contracts - unrestricted	824,741	-	837,756	1.6%	831,484	-0.7%	912,741	9.8%	921552	1.0%	908,971	-1.4%
	state grants	98,825	-	96,217	-2.6%	129,400	34.5%	152,521	17.9%	150652	-1.2%	169,058	12.2%

2.2.2 Student Cost of Attendance

Next, we analyzed the cost of attendance from the students' perspective. *Tuition* had a stable 1.5-2.5% increase for each year for in-state and in-district students; and 0.2-1.1% for out-of-state students. *Student fees* started off by increasing at a similar rate as the institutional expenses of 3-7% each year, and started decreasing after COVID started, which also might be due to reduced operational expenses. The on-campus living expenses had a relatively slower increase than off-campus living expenses.

Table 5. Non-profit institutions' financials expenses (2015-2020) summarized as median amount across all institutions and change rate with respect to the previous year

Category	Item name	2015		2016		2017		2018		2019		2020		2021	
		Amount		Amount	Change	Amount	Change	Amount	Change	Amount	Change	Amount	Change	Amount	Change
In-district	Average tuition for full-time undergraduates	8,791	-	9,000	2.4%	9,360	4.0%	9,490	1.4%	9,576	0.9%	9,828	2.6%	10,065	2.4%
	Required fees for full-time undergraduates	576	-	600	4.2%	620	3.3%	660	6.5%	695	5.3%	716	2.9%	720	0.6%
	Per credit hour charge for part-time undergraduates	294	-	300	2.0%	310	3.3%	320	3.2%	324	1.1%	329	1.7%	329	0.0%
	Average tuition full-time graduates	11,376	-	11,610	2.1%	12,000	3.4%	12,154	1.3%	12,240	0.7%	12,600	2.9%	12,600	0.0%
	Required fees for full-time graduates	432	-	459	6.3%	490	6.8%	500	2.0%	546	9.2%	550	0.7%	570	3.6%
	Per credit hour charge part-time graduates	543	-	559	2.9%	578	3.4%	587	1.6%	600	2.2%	613	2.2%	622	1.5%
In-state	Average tuition for full-time undergraduates	8,926	-	9,128	2.3%	9,381	2.8%	9,552	1.8%	9,790	2.5%	9,960	1.7%	10,200	2.4%
	Required fees for full-time undergraduates	576	-	600	4.2%	622	3.7%	662	6.4%	700	5.7%	716	2.3%	720	0.6%
	Per credit hour charge for part-time undergraduates	298	-	305	2.3%	315	3.3%	325	3.2%	329	1.2%	333	1.2%	337	1.1%
	Average tuition full-time graduates	11,376	-	11,610	2.1%	12,000	3.4%	12,154	1.3%	12,240	0.7%	12,600	2.9%	12,600	0.0%
	Required fees for full-time graduates	432	-	459	6.3%	490	6.8%	500	2.0%	546	9.2%	550	0.7%	570	3.6%
	Per credit hour charge part-time graduates	543	-	559	2.9%	578	3.4%	587	1.6%	600	2.2%	613	2.2%	622	1.5%
Out-of-state	Required fees for full-time undergraduates	596	-	616	3.4%	650	5.5%	696	7.1%	730	4.9%	750	2.7%	741	-1.2%
	Average tuition for full-time undergraduates	13,931	-	14,086	1.1%	14,240	1.1%	14,400	1.1%	14,616	1.5%	14,646	0.2%	14,775	0.9%
	Per credit hour charge for part-time undergraduates	477	-	492	3.0%	497	1.0%	500	0.7%	500	0.0%	500	0.0%	501	0.2%
	Average tuition full-time graduates	15,240	-	15,660	2.8%	16,200	3.4%	16,203	0.0%	16,128	-0.5%	16,320	1.2%	16,368	0.3%
	Required fees for full-time graduates	432	-	460	6.5%	500	8.7%	500	0.0%	550	9.9%	550	0.1%	570	3.6%
	Per credit hour charge part-time graduates	691	-	701	1.4%	716	2.2%	723	1.0%	726	0.4%	734	1.1%	740	0.8%
Living Expenses and Supplies	Published out-of-state tuition and fees	14,966	-	15,091	0.8%	15,305	1.4%	15,558	1.7%	15,518	-0.3%	15,493	-0.2%	15,718	1.5%
	Books and supplies	1,200	-	1,200	0.0%	1,206	0.5%	1,220	1.2%	1,200	-1.6%	1,200	0.0%	1,200	0.0%
	On campus, other expenses	2,863	-	2,887	0.9%	2,938	1.7%	2,989	1.8%	3,000	0.4%	3,026	0.9%	3,046	0.6%
	Off campus (not with family), room and board expenses	8,652	-	8,859	2.4%	9,100	2.7%	9,240	1.5%	9,514	3.0%	9,626	1.2%	9,892	2.8%
	Off campus (not with family), other expenses	3,668	-	3,687	0.5%	3,696	0.2%	3,754	1.6%	3,800	1.2%	3,892	2.4%	3,942	1.3%
Financial Aid and Loans	# undergrad awarded federal, state, local, institutional / other aid	371	-	382	3.0%	383	0.3%	397	3.7%	411	3.5%	430	4.6%	436	1.4%
	% undergrad awarded federal, state, local, institutional / other aid	71%	-	69%	-2.8%	69%	0.0%	68%	-1.4%	66%	-2.9%	70%	6.1%	69%	-1.4%
	Total # of federal, state, local, institutional / other aid awarded to undergrad	1,735,114	-	1,859,169	7.1%	1,880,119	1.1%	2,060,117	9.6%	2,232,888	8.4%	2,449,900	9.7%	2,449,925	0.0%
	% undergrad awarded Pell grants	51%	-	48%	-5.9%	46%	-4.2%	45%	-2.2%	44%	-2.2%	43%	-2.3%	41%	-4.7%
	Total # of Pell grant aid awarded to undergrad	1,130,622	-	11,373,442	905.9%	1,122,213	-90.1%	1,167,639	4.0%	1,253,599	7.4%	1,248,985	-0.4%	125,117	-90.0%
	% undergrad awarded federal student loans	57%	-	55%	-3.5%	54%	-1.8%	51%	-5.6%	50%	-2.0%	49%	-2.0%	46%	-6.1%
	Total # of federal student loans awarded to undergrad	1,652,841	-	1,589,373	-3.8%	1,565,946	-1.5%	1,475,790	-5.8%	1,422,065	-3.6%	1,356,804	-4.6%	129,117	-90.5%
	% of full-time, first-time undergrad awarded other student loans	6	-	5	-16.7%	6	20.0%	6	0.0%	6	0.0%	6	0.0%	6	0.0%
	Total # of other student loans awarded to full-time, first-time undergrad	103,198	-	120,126	16.4%	123,501	2.8%	139,838	13.2%	165,195	18.1%	146,317	-11.4%	141,649	-3.2%

2.2.3 Visualization of Completion and Grants by state

By mapping the averaged statistics over the past 5 years on the US map, we are able to roughly identify if correlation between 4-year completion rates and grants provided by the state governments exists. From the following graphs it can be seen that states with higher government grants are more likely to have higher college completion rates. For instance, the states that are provided with most student grant fundings (above 10 million), including WA, CA, NY, NJ, SC, and FL, appear to have the highest graduation rates as well. The states with the lowest government support (below 2 million), such as AL, HI, and most of the midwestern states, tend to have lower college completion rates compared to the rest of the US.

Note that there are some exceptions to this positive correlation, for example, IA, which received average government grants below 2 million, had one of the highest college completion rates in the country. TN, on the contrary, had the largest amount fundings (over 14 million) by the state government but its college completion rate was not as satisfying (below 50%).

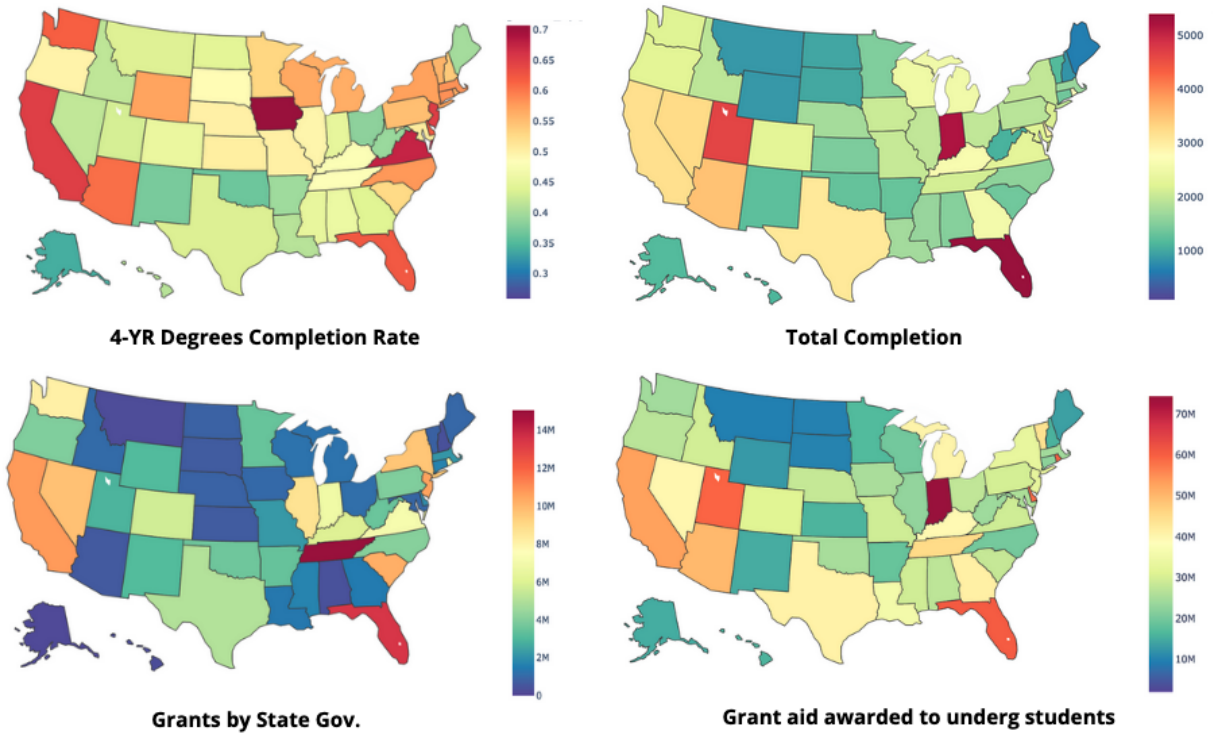

Figure 2. US Map colored by completion rate, number of completed students, institution received grants by state government, and grant aid awarrded to undergraduate students averaged from 2015 to 2021 and within each state

Geographically, the states with higher college completion rates are centralized on the west and east coasts, which is not surprising as students in more developed areas or at the higher end of the socioeconomic status spectrum are more likely to complete post-secondary academic degrees.

2.3 Impact of costs on outcome measures

In this section, we seek to answer the question "*How does institutional expenses/attendance costs impact college completion?*" Regression and ANOVA tests are used to examine if the impacts of each expense/cost item on the outcome measure are statistically significant. We conducted the analysis for outcome measures of the underrepresented student populations.

Since each institution has different baselines due to institution size, location, year established, etc., the absolute outcome measures as well as cost values may not be comparable across institutions. To test for a more "unbiased" effect of cost/expense items across all institutions, we computed the relative change of each variable with respect as:

$$relative\ change = \frac{(current\ year\ value - previous\ year\ value)}{previous\ year\ value} \cdot 100\%$$

Since most US academic institutions were impacted by COVID in the beginning of 2020, the pandemic may have had an impact on the financial patterns of years 2020-2021. In order to estimate an average effect of cost/expense on the outcome measures before and post-pandemic, we computed the average of variables for each institution from 2015-2020.

Then, this averaged outcome was used to fit linear regression model of the form:

$$y_{avg} = \beta_1 x_{avg} + \beta_0$$

The estimated coefficient indicates the magnitude and direction of the impact of cost/expense x on the outcome measure y, and the t-test of the coefficient indicates whether this effect is statistically significant.

Next, we describe our results split by institution type (non-profit private, for-profit private, public) and cost type (aid, finance, fees). For simplicity, we only display variables that were significant under significance level of $\alpha = 0.05$ correcting for multiple testing with sequential stepwise bonferroni. Using these selected combinations may give us both strong indicators of impact from certain costs on graduation rates but also trends within each type of institution.

2.3.1 Public institutions

2.3.1.1 Financial Aid

Significant results with correction for multiple testing with majority of the results (Table 10) were with benefits, grants and loans with undergraduate hispanic with a significant jump in grad rate. Just as significant is the rate for 2 year grants, perhaps to graduate earlier or transfer. We also see these programs to benefit the hispanic group much more than other groups. (Figure 4)

Table 10. Coefficients and p_values of avg fin aid (X) and avg Outcome (Y) over 2015-2021 for public Schools, * values are not actually 0 but smaller than 10^-5

Financial Aid Variable	Outcome	Coefficient	P-value
Percent of undergraduate students awarded Pell grants	completion rate-4yr-hispanic	1.2264	0.0000
Number receiving Post-9/11 GI Bill Benefits - all students	completion rate-4yr-hispanic	0.3853	0.0000
Number receiving Post-9/11 GI Bill Benefits - undergraduate students	completion rate-4yr-hispanic	0.3639	0.0000
Number of undergraduate students awarded Pell grants	completion rate-4yr-hispanic	0.7937	0.0000
Number of undergraduate students awarded federal student loans	completion rate-4yr-women	0.3055	0.0000
Total amount of federal, state, local, institutional or other sources of grant aid awarded to undergraduate students	completion rate-2yr-white	0.2378	0.0000
Percent of undergraduate students awarded federal, state, local, institutional or other sources of grant aid	completion rate-4yr-asian	1.4022	0.0000
Total amount of federal, state, local, institutional or other sources of grant aid awarded to undergraduate students	completion rate-4yr-hispanic	0.5239	0.0000
Total amount of Post-9/11 GI Bill Benefits awarded - graduate students	completion rate-4yr-hispanic	0.1521	0.0001
Percent of undergraduate students awarded Pell grants	completion rate-2yr	-0.2493	0.0001

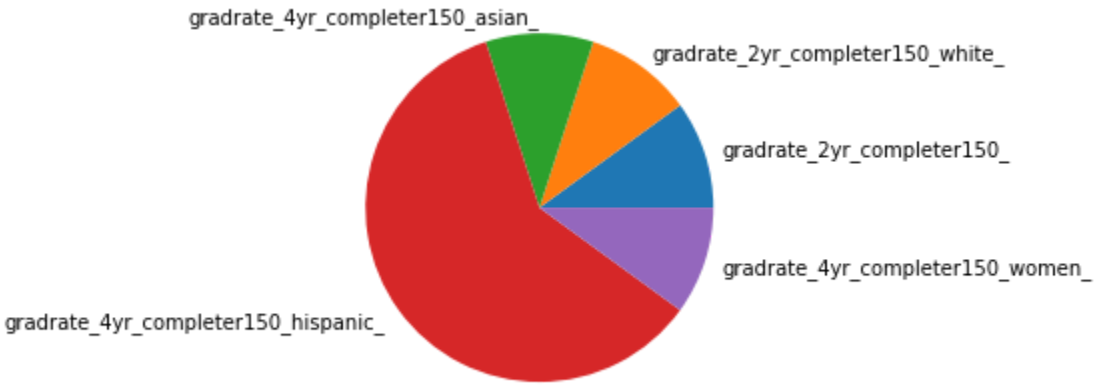

Figure 4. Under (a=0.05, with correction) Proportion of the Outcome variable with significance

2.3.1.2 Institutional Finance

Significant results with correction for multiple combinations (Table 11).. Similar to what we have seen for other categories, financial support, auxiliary, investment income, and similar columns seem to have significant impact on graduation rates for now both 2 and 4 year degrees. We also see a trend with public schools benefiting hispanic students' graduation rates much more so than other races.

Table 11. Coefficients and p_values of avg finance (X) and avg Outcome (Y) over 2015-2021 for public Schools, * values are not actually 0 but smaller than 10^-5

Finance Variable	Outcome	Coefficient*	P-value*
investment income	completion rate-4yr-hispanic	0.0000	0.0000
institutional support - all other	completion rate-2yr_hispanic	0.2140	0.0000
institutional support - all other	completion rate-4yr-hispanic	0.2072	0.0000
state operating grants and contracts	completion rate-4yr-black	0.1964	0.0000
auxiliary enterprises -- current year total	completion rate-2yr-black	0.0000	0.0000
academic support - interest	completion rate-4yr	0.2708	0.0000
pell grants (federal)	completion rate-4yr-hispanic	0.3126	0.0000
independent operations - operations and maintenance of plant	completion rate-4yr-white	0.0063	0.0000
total all revenues and other additions	completion rate-2yr_men	-0.1235	0.0000
institutional support - employee fringe benefits	completion rate-4yr-hispanic	0.3717	0.0000
academic support - salaries and wages	completion rate-4yr-hispanic	0.0989	0.0000
net scholarships and fellowship expenses	completion rate-4yr-hispanic	0.3025	0.0000
total noncurrent liabilities	completion rate-other	0.1659	0.0000

2.3.1.3 Fees

Significant results with correction for few combinations (Table 12). Mainly fees seem to only affect 4 year programs. This time we see off campus spending post covid have a big effect which might be again due to covid situations.

Table 12. Coefficients and p_values of avg fee (X) and avg Outcome (Y) over 2015-2021 for public Schools, * values are not actually but smaller than 10^-5

Finance Variable	Outcome	Coefficient	P-value*
Off campus (not with family), room and board	completion rate-4yr	-0.6499	0.0000
Tuition and fees of 5th largest program	completion rate-4yr	0.3590	0.0000
Tuition and fees of 6th largest program	completion rate-4yr	0.3551	0.0000

2.3.2 For-profit Private Institutions

2.3.2.1 Institutional Finance

The list of most significant variable–outcome pairs are shown in Table 6. Majority of significant outcomes were increased rate of graduation in 2 year programs with respect to federal funding or discounts.

Table 6. Coefficients and p-values of averaged institutional finance variables (X) and averaged outcome measures (Y) over 2015-2021 for Nonprofit Schools.* values are not actually 0 but smaller than 10^-5

Finance Variable	Outcome	Coefficient*	P-value*
federal appropriations	completion rate-2yr	0.0000	0.0000
discounts and allowances from state government grants applied to auxiliary enterprises	completion rate-2yr	0.4279	0.0000
federal appropriations	completion rate-2yr-black	0.0000	0.0000
equity, beginning of year	completion rate-other-women	0.2892	0.0000
academic support-total amount	completion rate-2yr_men	0.4788	0.0000

2.3.2.2 Financial Aid

No significant result with above written criteria. However, if we ignore the multiple testing correction, we get significant results ($\alpha = 0.05$) for many combinations with a significant proportion of them Black and for women. Figure 3 shows the different variations of the completion rate outcome tested and the proportion of tests a given outcome was significantly impacted by any of the cost/expense items among all the tests.

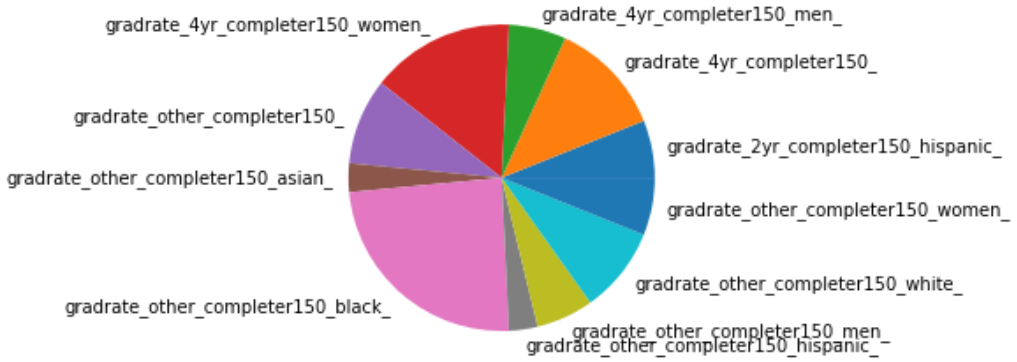

Figure 3. Proportions of the Outcome variable with significance

2.3.2.3 Fees

No significance to report. All p-values fail both with correction.

2.3.3 Non-profit Private Institutions

2.3.3.1 Financial Aid

Majority of the cost/expense-outcome pairs resulted in significant tests. As seen in Table 7 the most frequent pair were with white and asian undergraduate students given pell grants with significant jump in grad rate.

Table 7. Coefficients and p_values of avg fin aid (X) and avg Outcome (Y) over 2015-2021 for Nonprofit Schools, * values are not actually 0 but smaller than 10^-5

Financial Aid Variable	Outcome	Coefficient	P-value*
Percent of undergraduate students awarded Pell grants	completion rate-4yr-white	0.4778	0.0000
Total amount of federal, state, local, institutional or other sources of grant aid awarded to undergraduate students	completion rate-4yr-asian	0.7223	0.0000
Percent of undergraduate students awarded federal, state, local, institutional or other sources of grant aid	completion rate-4yr-white	0.5606	0.0000
Number of undergraduate students awarded federal, state, local, institutional or other sources of grant aid	completion rate-4yr-asian	0.8467	0.0000

2.3.3.2 Institutional Finance

Similarly, multiple combinations of expense/cost-outcome pairs had significant tests (Table 8). All outcome variables affected were 4 year graduation rates as it seems to largely be an institution wide result (regardless of race or gender despite the few results of them here as well). Finance variables most influential seem to revolve around gifts to the nonprofits as well as general increase in putting the money back into the school itself (auxiliaries, buildings, equipment etc). Increased tuition leads to higher dropout rates for men in these 4 year universities.

Table 8. Coefficients and p_values of avg finance (X) and avg Outcome (Y) over 2015-2021 for Nonprofit Schools, * values are not actually 0 but smaller than 10^-5

Finance Variable	Outcome	Coefficient*	P-value*
private gifts - temporarily restricted	completion rate-4yr-women	0.0000	0.0000
private gifts, grants and contracts - temporarily restricted	completion rate-4yr-white	0.1229	0.0000
private gifts - temporarily restricted	completion rate-4yr	0.0000	0.0000
discounts and allowances from pell grants applied to auxiliary enterprises	completion rate-4yr	0.1097	0.0000
auxiliary enterprises-total amount	completion rate-4yr	0.1082	0.0000
permanently restricted net assets included in total restricted net assets	completion rate-4yr	0.1882	0.0000
buildings - end of year	completion rate-4yr	0.1619	0.0000
equipment, including art and library collections - end of year	completion rate-4yr	0.2090	0.0000
private gifts - temporarily restricted	completion rate-4yr	0.0000	0.0000
equipment, including art and library collections - end of year	completion rate-4yr-men	0.2311	0.0000
equipment, including art and library collections - end of year	completion rate-4yr	0.1700	0.0000
discounts and allowances from pell grants applied to auxiliary enterprises	completion rate-4yr	0.0781	0.0000
auxiliary enterprises-total amount	completion rate-4yr	0.0968	0.0000
other revenue - permanently restricted	completion rate-4yr-white	0.0000	0.0000
total assets	completion rate-4yr	0.2649	0.0000
private gifts, grants and contracts - temporarily restricted	completion rate-4yr-men	0.0576	0.0000
discounts and allowances from pell grants applied to auxiliary enterprises	completion rate-4yr-white	0.1269	0.0000
tuition and fees - unrestricted	completion rate-4yr-men	-0.2589	0.0000
hospital revenue - unrestricted	completion rate-4yr-women	0.1942	0.0000

2.3.3.3 Fee

Significant results with correction for only 1 variable combination on campus, other expenses 2021-22 and gradrate_4yr_completer150 (on campus costs vs graduation rate). One hypothesis is that this significant change could be a result of covid and its complications.

Table 9. Coefficients and p_values of avg fee (X) and avg Outcome (Y) over 2015-2021 for Nonprofit Schools, * values are not actually 0 but smaller than 10^-5

Student Fee Variable	Outcome	Coefficient	P-value*
On campus, other expenses 2021-22	completion rate-4yr	-0.2200	0.0000

2.3.4 Key Observations

From our linear regression result on cost vs outcome measures, fee category is largely influenced by on and off campus costs for 4-year universities graduation rate in that higher relative costs lead to less graduation rates within post covid years 2020-2021. Finance cost

category largely influenced by federal appropriations and statewide discounts for for profit schools both increasing their overall balance sheet, by donations and auxiliaries of all kinds giving students better access to and quality of education, and by grants and other federal and statewide systems that reduce costs for students to stay and finish their degree. Unlike the previous two cost categories, financial aid largely showed effects on student types, helping more underrepresented student groups. An additional trend within the public institution category was the increased effect of these variables on the hispanic student population.

2.4 Prediction of Outcome Measures

The previous section analyzed the impact of individual institutional expenses and costs items on the outcome measures. Next, we will use machine learning to model the multivariate and interrelated relationships of different types of data including institutional expenses, attendance costs, and institutional characteristics, such as institution size, whether land grant, offering of special learning programs, services offered (e.g. day care, hospital, library), athletic and veteran programs. The purpose of developing a machine learning model is not only to predict completion but also to identify what variables are key predictors considering the interrelationship of all variables.

2.4.1 Prediction Model

We compared four machine learning models, two linear and two nonlinear. Their implementation and training details are provided below:

(1) **Principal Component Analysis** (PCA): we first used PCA to reduce the concatenated feature matrix into lower-dimensional principal components (PC), then trained a linear regression model using the ten PCs;
(2) **Partial Least Squares Regression** (PLS) can be seen as a supervised version of PCA where the outcome measure guide the decomposition;
(3) **XGBoost** is popular supervised boosted tree model that combines estimates of multiple weaker tree models. We tuned the hyperparameters of XGBoost via 5-fold cross-validation;
(4) **Encoder-decoder MLP**: fully-connected neural networks or Multi-layer perceptron (MLP) can easily overfit on this small dataset. Thus, we implemented a custom encoder-decoder architecture within the MLP. The encoder was composed of three linear-relu layers and can be seen as a nonlinear version of PCA, as it projects inputs into a lower-dimensional latent space. The latent space vector is then passed to a single linear layer to predict the outcome. The decoder has a symmetrical architecture as the encoder and is in charge of making sure that the latent vector contains enough information to reconstruct the original input. The entire network is trained to optimize reconstruction loss (difference between the inputs vs. decoder output) and prediction loss.

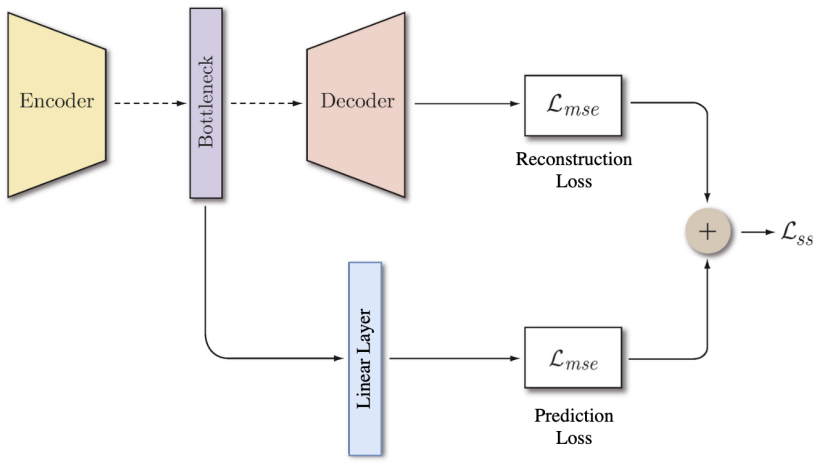

Figure 6. Flowchart of the custom encoder-decoder MLP

We split the entire 2015-2021 dataset into 80% training and 20% test set. The training set is used to learn and tune the models, and the test set is used to report model performance measured in Mean Squared Error (MSE) and correlation (Corr). The split is down such that all rows of the same institution are either in training or test to avoid data leakage. All models are trained using a total of 85 variables involving institutional expenses, attendance costs, and institution characteristics. Missing values were imputed using 0. We applied log-transform to all cost-related variables.

Table 13 compares the test performances of the four models for predicting completion, i.e. total number of awards/degrees. XGBoost achieved the best performance with 0.952 correlation. These results suggest that completion can be accurately predicted using these variables and a not too complex nonlinear model is the best fit for this data.

Table 13. Comparison of four machine learning models for predicting total completion

Model	Description and Hyperparameters	Test performance	
		MSE	Corr
XGBoost	learning_rate=0.015, max_depth=4, n_estimators=600	**195779**	**0.952**
PCA + Linear regression	n_components=10	858271	0.931
PLS Regression	n_components=10	787288	0.936
Encoder-decoder MLP	learning_rate = 0.01, batch_size=256, latent_dimension=5, weight_reconstruction=0.2	839227	0.945

Thus, we used XGBoost to train prediction models for multiple outcome measures, including total completion for STEM vs non-STEM majors, undergraduate vs graduate vs certificate, and the completion rate (i.e. percentage of students in the cohort who graduated within 150% of normal time) for 4 year and 2 year programs (Table 14). The model was in general more accurate for predicting the absolute number of completions than the completion rate.

Table 14. Prediction accuracy of outcome measures by degree type

Outcome Measure	Subpopulation	Sample size		Test performance	
		Training	Test	MSE	Corr
Completion	Total	8043	2009	195779	0.952
	STEM	7886	1964	126499	0.943
	Non-STEM	8043	2009	691470	0.909
	Certificate	3677	899	465123	0.955
	Undergraduate	8043	2009	2424365	**0.979**
	Graduate	2220	523	103244	0.950
Completion rate	4 year programs	3101	697	0.00642	**0.838**
	2 year programs	4342	1150	0.00801	0.597

2.4.2 SHapley Additive exPlanations (SHAP)

Deep learning solutions are known as black-box models and hard to interpret, thus to further investigate the relationships between the institutional attributes and completion rates and make the complex outputs interpretable, we used SHAP (SHapley Additive exPlanations), a popular eXplainable AI (XAI) tool, to quantify the importances of each input variable to the prediction model. Specifically, it deconstructs prediction into a sum of contributions from each feature as the sum of the SHAP values and a fixed base value. In our case, the base value is equal to the mean of the target variables.

$$f(x) = base\ value + sum(SHAP\ values)$$

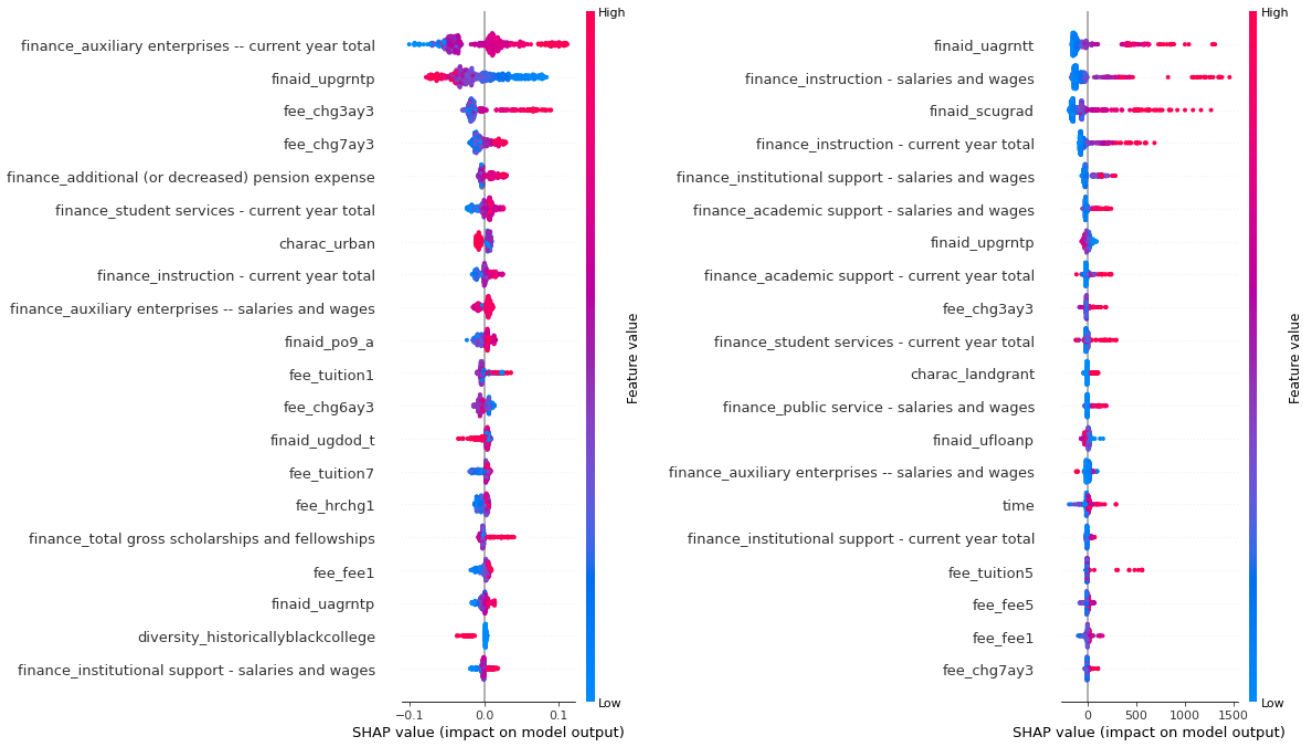

Figure 7. SHAP completion rate (left) and completion stem (right)

According to the beeswarm plots above, the top 5 most influential attributes are auxiliary enterprises, percentage of undergraduates awarded Pell grants, out-of-state tuition and fees, off-campus room and board, and pension expenses. For example, we see that the higher values of auxiliary enterprises have higher SHAP values, indicating that more auxiliary enterprises result in more predicted college completion rate. Likewise, higher values of out-of-state tuition, off-campus room and board, and pension expenses would lead to higher completion rates as well.

The reverse is seen for the percentage of undergraduates awarded the Pell grant, which is known as a federal grant awarded to undergraduate students who display exceptional financial need. In this case, lower percentages of Pell grant recipients lead to higher college completion rates.

The least important variables include operational salaries and wages, if the institution was historically a black college, percentage of undergraduates awarded grant aid, required fees for full-time undergraduates, and total gross scholarships and fellowships awarded.

From Figure 7, we can see that the most important features contributing to completion rates in the STEM field are slightly different – including total amount of Pell grant aid awarded, institutional salaries and wages, total number of financial aid in the cohort, current year's total instruction expenditure, and operational salaries and wages. They all appear to have positive relationships with predicted college completion rates in STEM.

2.5 Discussion

Table 15. Summary table of insights and recommendations

	Currently increasing a lot (> 4%) in 2020	**Currently decreasing or increasing (< 4%) in 2020**
Variable is important (p<0.05)	● Percentage of undergraduates awarded Pell grants (private non-profit, public) ● Total undergrads receipting award/grants (private non-profit, public) ● Percentage undergrads receiving awards/grants (private non-profit, public)	● Auxiliary enterprises (private non-profit) ● Equipment, including art and library (private non-profit) ● Institutional support (public) ● Academic support (public, private for profit)
Variable is not important (p>0.05)	● Public service (public)	● Student services (public)

Based on the modeling outputs, we compiled a confusion matrix to group the predictive features, differentiating by public schools, for-profit private schools, and non-profit private

schools. The columns indicate whether the feature is currently having an increasing trend in the most recent year (Part I), while the rows indicate whether the variable has statistically significant impact on completion (Part II and III). We set the increase rate threshold as 4%, as the inflation rate during the same period of time is approximately 4%, therefore if the increase of an attribute is more than 4%, we categorize it to the left column, otherwise to the right. In terms of rows, we adapted the regression and SHAP results from the previous sections and set the threshold for p-value of 0.05. Intuitively, if the p-value of an attribute is less than 0.05, we identify this feature as important, and vise versa.

For the features that are both important and increasing significantly (true positives), we see that the top 3 of them are all awards/grants related. Financial aid largely showed effects on different student subpopulations, helping more underrepresented student groups. To be specific, public and private non-profitable schools should keep investing in offering students, especially undergraduates, grant fundings, as the higher number of grants recipients leads to a higher college completion rate. Recalling the conclusion we drew from SHAP analysis, the students who receive Pell grants are likely to be in exceptional financial need, which is significant but also negatively correlated with our target variable. Thus offering more institutional awards and financial aid would help those who have financial problems to continue and finish their post-secondary education.

For the variables that are important but currently showing a decreasing or a slower increasing (< 4%) trend (true negatives), we assume that they might have a long-term effect. For instance, the investment in the current year may be reflected in the change of graduation rate in the future 3-5 years. Hence we suggest spending money dedicated to instructional use and institutional support, while the amounts of investment are subject to change depending on the actual endowment/profit of the year. For private non-profitable schools, the fields like auxiliary enterprises and equipment (e.g. art, library, etc.) are expected to be invested for a long run. For public schools, we suggest the decision makers to support a wide range of purposes across the institution, with a vast majority of funds dedicated to administrative and academic support. Private for-profit schools should distribute money to academic support too.

Regardless of the changing trend, public service and student service are considered not important when it comes to college completion rate, therefore we think that schools are able to rationally cut fundings towards those areas.

From the student's costs perspective, tuition was shown to have a significant effect on the completion rate of pirate non-profit institutions, and off-campus room and board expenses for all institutions in general.

2.6 Areas of further analysis

This study has several areas of improvement. First, some expenses have shorter-term impact while others have long-term impact. In all our analysis, we used the same time period for both the expense/cost item and the outcome measure, which only analyzes the short-term same-year effect. Future work can be improved by collecting data from more years and creating

a time lag variable to study the long-term effects. Second, one reason why prediction models performed better for absolute completion than the completion rate (%) is because of the unknown drop out or trans-fer out time of the students who did not complete the study. For example, a student enrolled for cohort 2020 who dropped out in 2018 will be measured within the completion rate of year 2020, which is being predicted by the costs/expenses of year 2019-2020 when the student was no longer enrolled in the institution. Third, the SHAP outputs exhibit clusters of data points which suggest similar groups of institutions that share similar patterns. Future work can develop prediction models for more granular types of institutions and can provide more personalized insights for each institution.