Analysis of College Affordability on Graduation

Ash Chigavazira, Lasya Maddula, Daniil Gavrilov, Precious

Gilbert, Zack Baldorado

The University of North Carolina at Charlotte

DTSC 1302

November 22, 2024

Uncovering the Variables that Impact Graduation Rates

Introduction

The national dialogue about enhancing higher education is dominated by college access, but completion is becoming a more important indicator of success. However, graduating college is challenging due to a number of issues, such as academic rigor, demanding schedules, and—above all—the financial burden. In fact, Research shows that in the past low college enrollment was due to the lack of financial assistance (Terriquez & Gurantz, 2014). College degrees are quite crucial for upward mobility, but it has also never been more costly. This report centers on the most impactful factors that affect graduation rates. More specifically, we are identifying the correlation between college affordability in both private and public out-of-state institutions and comparing graduation rates. Our research question asks, "What impact does college affordability in out-of-state institutions have on graduation rates?" The study will initially concentrate on the background and ramifications of the research in order to accomplish this. A synthesis of the literature on the subject of the study is part of this. The possible stakeholders, potential conflicts, and our ethical frameworks and considerations will then be the main topics. This will clarify the measurement of the study question through conceptualization and operationalization. Lastly, the study will give a summary of the methods used to manipulate the data. A synopsis of sources that support our theory with similar plots as well as our own plots will be showcased in this. Therefore, we hypothesize that the lower the affordability, the lower the graduation rate suggesting a strong negative correlation between the two variables. This is due to the idea that students might be forced to take on a significant amount of debt or have to carry the burden of working a job during the school year since they are experiencing financial

strain. Working a job during college can take away significant amounts of time used to study, build relationships with professors, and do well in school, leading to fatigue and dropout. Overall, our research will either support or dispute our theory that college affordability has a negative impact on graduation rates.

Context and Implication

In 2006, the U.S. Secretary of Education's Commission on the Future of Higher Education highlighted a critical issue in higher education: the persistent gap in college attendance and graduation rates between low-income students and their more affluent peers. This report underscored the need for higher education institutions to improve outcomes, particularly by addressing disparities in degree completion (Cook, 2010). While graduation rates have become a significant topic in recent years, obtaining accurate national-level graduation data remains a complex challenge. Most common models calculating the graduation rates exclude roughly forty-eight percent of enrolled students. While these databases may not fully account for evolving enrollment patterns or diverse student demographics, they still offer valuable insights into graduation rates (Cook, 2010).

High out-of-state tuition can push them to work full-time jobs just to cover their costs, leaving them with less time to study and affecting their academic performance. This financial strain can lead to burnout, and in many cases, it increases dropout rates. Since they're already juggling multiple responsibilities, some students are forced to delay their graduation or drop out altogether, making it harder for them to complete their degree on time. Research shows that students from lower-income backgrounds are especially vulnerable to these challenges, as they may not have the

financial support to attend expensive out-of-state institutions, limiting their academic potential and increasing the likelihood of not finishing their degree (Dale & Krueger, 2001).

The internationalization of higher education is an increasing trend that brings both advantages and challenges. It improves academic exchanges, facilitates the sharing of knowledge, and helps countries by providing education where local institutions may not meet demand. Western universities play a central role by creating global networks and using digital tools to attract international students, generate income, and enrich the learning experience. For students, studying internationally helps develop a global mindset, cultural understanding, and language skills. However, this trend also carries risks, including the commercialization of education, a "brain drain" where talented students leave developing countries, and a loss of cultural identity (Jibeen, 2015). As universities focus more on making money, especially by charging international students high fees, the benefits tend to go to developed countries, while developing nations struggle with issues like growing social inequality and limited improvements in education. Moreover, the standardization of curricula and the loss of cultural diversity can harm the values of the host countries. While internationalization offers significant opportunities, it also raises concerns about fairness, cultural preservation, and the overall quality of education.

With increasing tuition costs and economic developments, students tend to rely more on loans to afford higher education. Research presented in the journal *Social Forces* states that when loans exceed the $10,000 mark, the likelihood of graduation begins to decrease (Rachel, 2012). This finding directly challenges our research where our team noticed a clear trend that higher education costs have a relation with higher graduation rates. However, the data presented to our team does not specify how students find a way to afford to attend their institutions. The current government

support for higher graduation largely consists of loans rather than grants, putting many students in a challenging situation. By 2008 approximately two-thirds of all students relied on loans with a median college debt of $15,000 (Rachel, 2012). These loans were given out without accounting for creditworthiness. Initially, student loans were handled by the government, but private lenders entered the market and expanded the student debt. This led to students who were more likely to graduate but with a burden of debt. Public institutions offer lower tuition costs for accessible education focusing on enrolling larger classes with a lesser income. Student debt is one of the factors influencing public school completion. Private schools, however, charge higher tuition costs but with better educational opportunities, focusing on smaller classes and retention of students. Consequently, private schools report higher completion rates (Rachel, 2012).

The primary objective of our model was to predict graduation rates by examining key factors such as out-of-state tuition, the proportion of students in the top ten percent of their high school class, and alumni engagement. These factors were selected based on our analysis, which identified them as the most significant contributors to graduation rates. The percentage of students in the top ten percent of their high school class reflects a university's ability to attract high-achieving individuals, which often enhances its academic reputation. Institutions that draw top performers tend to create environments of intellectual rigor, as these students seek to surround themselves with equally driven peers and access resources that support their academic and personal growth. Alumni engagement also plays a pivotal role in influencing graduation rates. Through mentorship, financial contributions, and professional networking opportunities, alumni provide vital support that aids student success. Their involvement can inspire students, reduce financial barriers, and foster a sense of community that encourages persistence through graduation. Out-of-state tuition serves as a

measure of the investment required to access the academic resources that ensure student success. It is arguably the most influential factor, as it can either attract or deter prospective students depending on their financial circumstances. We deliberately chose not to include whether a school is public or private in our model to avoid bias in favor of private institutions. Private schools often have more resources and correspondingly higher tuition rates, which could skew the analysis. Various demographic information regarding background information such as full time enrollment, percentage of commuters, and gender all play a role that affects the graduation rate in public vs private schools. When comparing multiple models, nontraditional full time enrollment affects rates in both public vs private schools and the data can be misleading about the context of why private schools seem to have higher graduation rates for traditionally enrolled students although public institutions do better to allocate resources for a more diverse population (Scott et al., 2006). Excluding this variable allows for a more equitable comparison of graduation rates relative to the financial investment, offering a clearer picture of how efficiently resources are utilized to promote student success.

As a team our primary goal was to predict graduation rates based on by examining key factors such as Out of State Tuition Costs, proportion of students in the top ten percent of their high school class, alumni engagement, and College Acceptance Rates. While trying to examine these things it can call for some ethical implications that are determined by biases, demographics, and more. A big implication that could affect data collection was the fact that many students don't want to disclose private and sensitive information. Sensitive information could include the amount of their tuition, student debt, or even the amount of financial aid received by students. The likes of student debt and the amount of financial aid can leave a financial burden on graduated students for the rest

of their lives. This raises the question if the quality of education is even worth the amount of money being spent on tuition. The cost of post high school education makes education less accessible to lower income families. This can affect the number of students enrolling in colleges knowing that they can't afford it. Lower income students who can't keep up with the expenses of college tend to drop out of school affecting schools graduation rates. For the public and state universities, government funding creates a lot of tension between the impact of student services. Governments tend to fund certain schools over others based on demographic, size, and geolocation rather than funding schools that require funding.

There are several ways these implications can be treated fairly. The government funding can equally distribute funding throughout the state wide universities, making it easier for schools to students in need. These lower income families would not have to worry about the quality of education being worth the financial burden they are facing. Colleges can also offer student support services to students who may feel like dropping out of school, ensuring they do receive the education that they have been paying for.

Measurements

The original goal of this research project is to determine and calculate which variables had the highest correlation with graduation rates. It is important to outline the terms used throughout this research project. Although the cost of higher education is rising, affordability is thought to be a major predictor of college attendance and academic achievement. Textbooks and course materials alone have grown about three times faster than inflation over the last thirty years (U.S.

Bureau of Labor Statistics, 2016). This is important to consider when we discuss the realm of affordability, we must also consider that affordability expands to textbooks and other miscellaneous costs, not just tuition. And although the cost of attending college has decreased overall, particularly for those from lower- and lower-middle-income families (Perna & Li, 2006). Low- and lower-middle-income students frequently work while they study to offset these expenses, increasing their risk of suffering from the detrimental effects of extracurricular employment, which has been linked to an average of 15 hours per week. The amount of time spent in an occupation to pay for college is putting the student in debt which can have adverse effects on their time to study, make connections, or even do work related to their major.
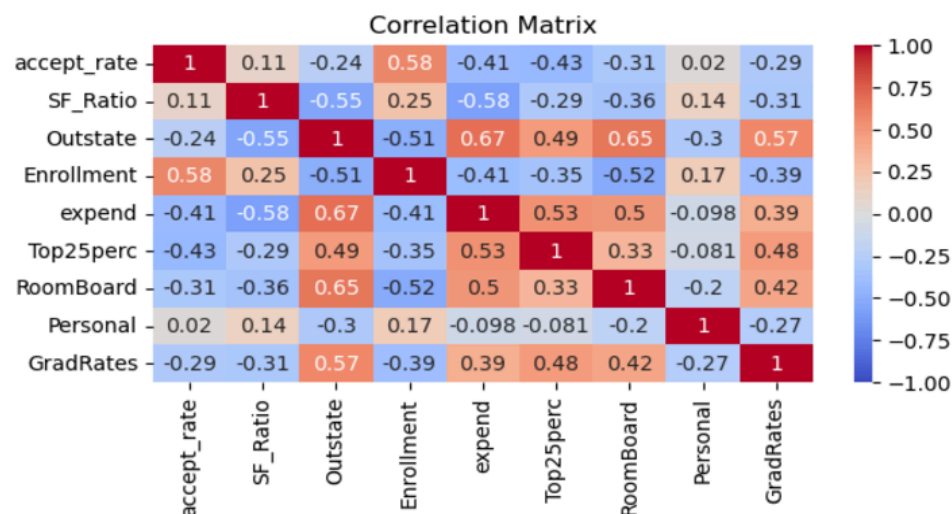
In an effort to study the correlations between various factors that affect post secondary success, we collected data from 777 colleges and universities throughout the United States. In our first analysis we ran a test using several factors we thought played a role in graduating college. In our first model, we tested acceptance rates, student faculty ratio, out of state tuition

costs, number of new students enrolled, instructional expenditure per student, new students from the top 25% of H.S. class, and estimated personal spending versus graduation rates to form our regression model, lm(grad rates ~ acceptance rates + student/faculty ratio + out of state tuition + enrollment rates + expenditure + top 25% students + room & board + personal spending). First we created a correlation matrix to get an idea for the relationships between our variables. Our correlation heatmap showed strong relationships between graduation rates versus out of state tuition costs, and the percentage of students coming from the top 25 percent of students attending the respective college with rates of 0.57, 0.48, and .42 respectively. Interestingly enough there are also strong correlations between the independent variables themselves. Expenditure displayed high correlations with top performing student enrollments, out of state tuition costs, and room and board. From here we created our hypothesis to test if the coefficients are equal to zero. After conducting an Ordinary Least Squares test our results in the performance of the model revealed that our model explained a good fit for the data but in contrast, it also showed that several variables were insignificant, because they displayed high p values. Variables such as full time undergrad students, enrollment rate, expenditure rate, and estimated personal spending with p values of 0.128, 0.853, 0.896, and 0.090 respectively. Although the adjusted R^2 was high with a value of 0.96 that signaled that there was a problem with our model as there were too many predictors. To inspect for multicollinearity present in our model. We next focused on analyzing this further by removing these variables through a VIF test (Variance inflation Factor), and opted for a simpler model after removing variables that were inflating our Adj. R-squared. We narrowed our model down to two independent variables to create a new regression model, lm(Grad Rates ~ Accept rates + Out of State tuition costs). After running tests we found that our new model displayed little multicollinearity between the independent variables but also did well

to explain the fit of the data, resulting in an adj. R-squared of 0.96 with low rates of 5.2 for VIF.

Data

    The dataset used for this research is 'Dataset C: College Affordability". The dataset contains many institutions and various information about them. Loading the dataset into a data framework was necessary to make manipulations but it is important to note that we did not have any missing values. Doing that allowed us to pull the necessary information to include in modeling and creating visualizations. Visualization allowed us to notice a clear pattern with a graduation rate and an "Outstate" variable which simply means out-of-state tuition. That gave us a clue that graduation rate and attending costs have a strong positive correlation. We decided to exploit it and build a statistical model to identify the exact relation and what other factors contribute to the graduation rate presented in the dataset.
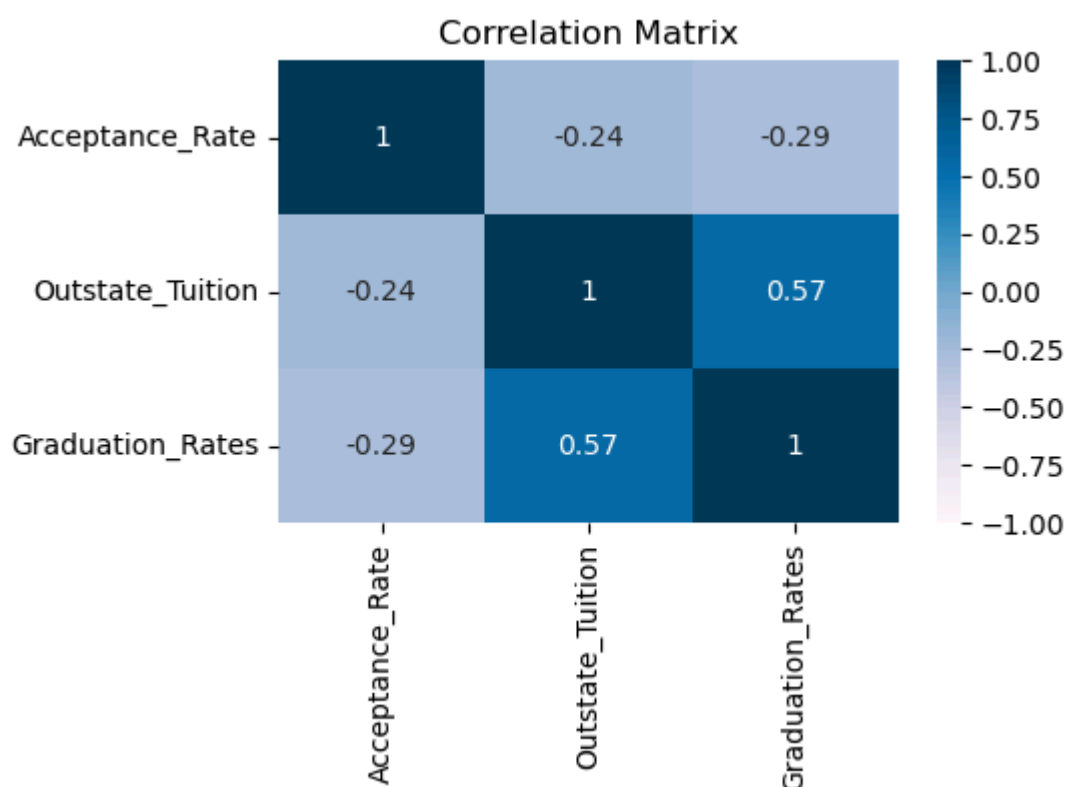
Models and Interpretations

Figure 2 - Correlation Heat Map using the factors within the "College Affordability" dataset.

Correlation Heatmap results
Variables Correlation
Acceptance Rate 1
Affordability (Outstate tuition) 1
Applications
Enrollment Rates
Top 10 Percent Ranked

Top 25 Percent Ranked

To find the best variables to predict total yearly compensation, a correlation heat map is used. The variables that explain graduation rates are presented at the table above, which have been extracted from the heatmap. Out of these variables, the best suited for a model are College Affordability and Acceptance Rate. Hence, both of the two independent variables are used to predict the graduation rates, the dependent variable. Coincidently, we did not have missing data so that did not impact the independent variables, which assisted in the creation of a robust and reliable model. The correlation between a variable and itself is always represented by the diagonal cells, and it is always 1. The correlation's strength is shown by the color intensity. Generally speaking, stronger correlations (whether positive or negative) are shown by darker hues, whereas weaker correlations are indicated by lighter colors. Within each cell, the actual correlation coefficient—which ranges from -1 to 1—is frequently shown. A significant association is indicated by a number near 1 or -1 and a linear association that is

weak or nonexistent is indicated by a value approaching 0. In this particular heatmap, acceptance rate in relation to out of state tuition has a weak negative correlation (-0.24) that can be explained by colleges charging significantly higher tuition for non-resident students. It is less likely to accept a large pool of out-of-state applicants, which lowers the acceptance rate for those students because of the cost barrier. Contrary to popular belief, private higher education does not lead to higher graduation rates. One may think that attending a smaller private university may afford a much higher rate of success in comparison to larger public universities but despite private universities ability to attract academically prepared students, they tend to have lower graduation rates when compared to public institutions when you factor in time completion (Cook & Pullaro, 2010, p. 28). In addition students on average take 5 years to graduate instead of 4 years to degree completion, and this provides a more accurate estimate of graduation rates. In fact when we take this in consideration public universities are shown to have higher graduation rates than private universities and students typically have a higher chance of success despite the systemic and financial hurdles students may face.

Additionally, the acceptance rate and graduation rate have a weak negative correlation (-0.29) as well because private colleges which typically tend to be more selective, tend to admit a larger range of students with different academic readiness, which might result in a lower graduation rate. Lastly, out-of -state tuition has a positive relation to graduation rates (0.57) because these universities typically have greater resources and higher academic standards due to the increased revenue from higher out-of-state tuition fees—there is frequently a positive correlation between out-of-state tuition and graduation rates.

The dataset is trained on multiple different regression models (Linear regression, Random Forest, KNN, XGBoost, LightGBM, and SVM). To find the best-performing model, statistical analysis is performed to test the utility of each model.
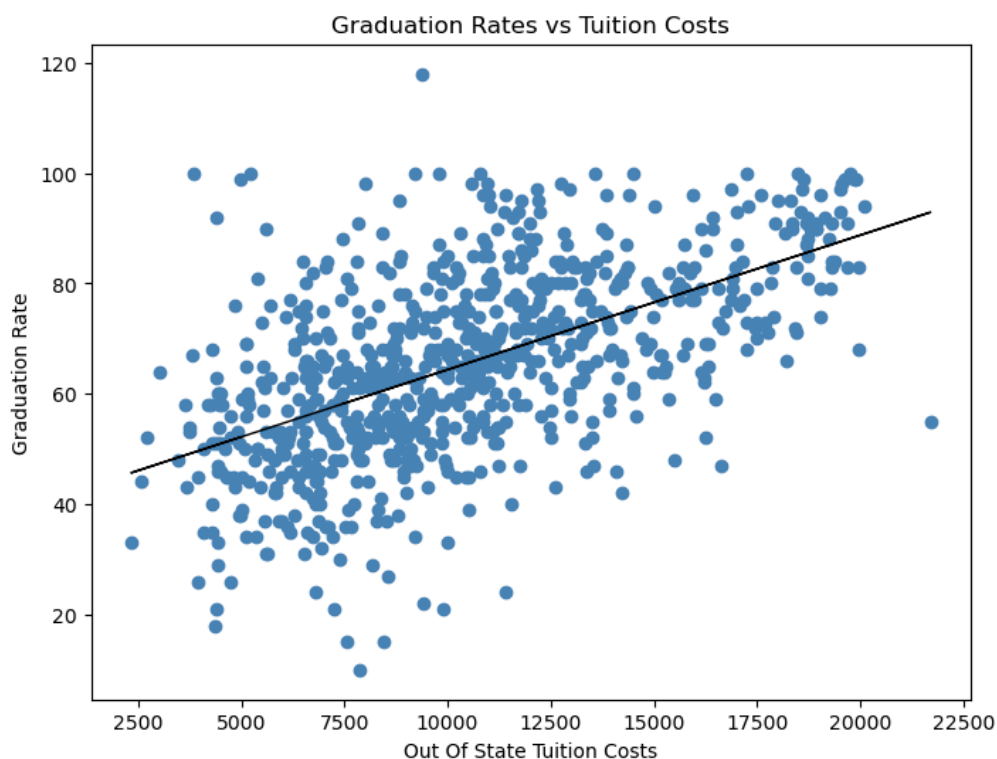

Figure 3 - Comparison of Out of State Tuition and Graduation Rate Scatter Plot

In Figure 3 (pictured above) the independent variables are the Out of State Tuition Costs. The dependent variable is the Graduation Rate, ranging from 9-99%. The mean of Out of State Tuition Costs came out to be $10,440 dollars, while the mean of Graduation Rates was 65.432%. While the Out of State Tuition increases our dependent variable of Graduation Rates tends to increase. This means that there is a pretty strong relationship between the independent variables (Out of State Tuition Costs) and the dependent variable (Graduation Rates). However, the data consists of some outliers who do not fall close to our line of best fit. Most of these schools are private schools whose cost stays the same whether students are in state or out of state. This

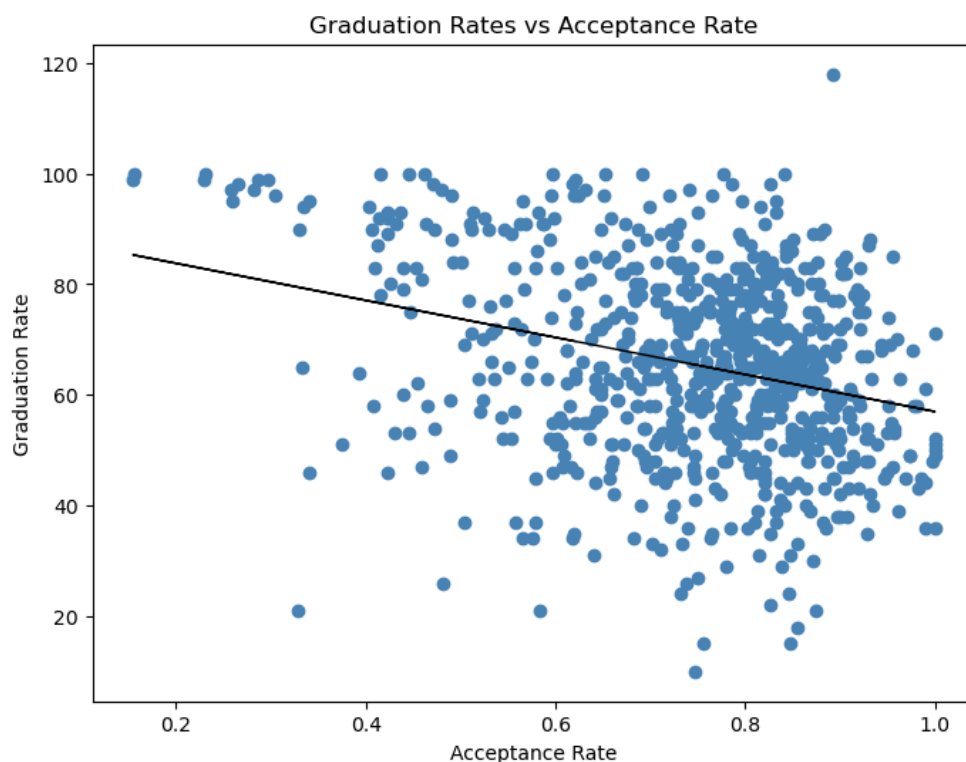suggests that out-of-state tuition may not influence graduation rates.

Figure 4 -Comparison of Acceptance Rate and Graduation Rate Scatter Plot

In Figure 4 (pictured above) the independent variables are the Acceptance Rates ranging from 19-99%. The dependent variable is the Graduation Rate, ranging from 9-99%. The mean of the Acceptance Rates came out to be 79%, while the mean of Graduation Rates was 65.432%. While the Independent Variable of Acceptance Rate increases our dependent variable of Graduation Rates tend to decrease. This means that there is a negative linear correlation between the independent variables (Acceptance Rate) and the dependent variable (Graduation Rates). However, the data consists of some outliers who are the same private schools from Figure 3.
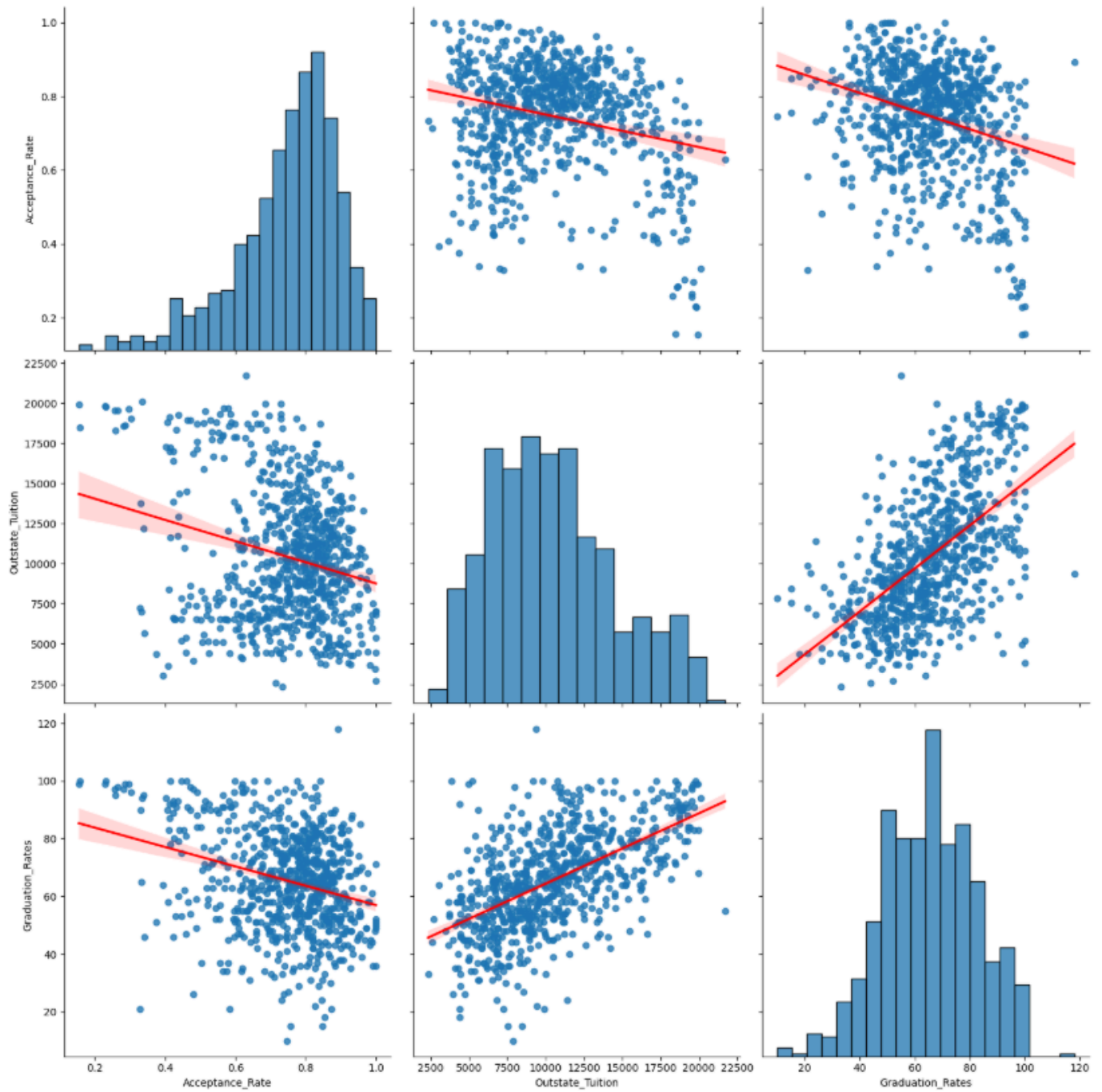
Figure 5 -Pairplot of variables Graduation rate, Tuition, and Acceptance Rate

Figure 5 is a visual of a pairplot comparing the relationships and distributions between

the variables. The diagonal graphs display distributions of the variables themselves. Acceptance rates distribution plot reveals that this variable is skewed to the left with approx mean acceptance rate of 80%. This means that 80% of the students who apply, get accepted to the institutions they apply to. Conversely, the rightly skewed Out of State tuition plot in the middle shows a mean cost of $10,000 per year with some institutions costing as much as $17,000 per year. Lastly, graduation rates follow a bell curve with a mean rate of approximately 70%. In addition to the relationships between the variables, graduation rates show a high positive relationship with tuition costs but a steady negative relationship with acceptance rates. These graphs could express that the higher the tuition costs, the graduation rates tend to be higher.

OLS Regression Results

Dep. Variable:	Grad_Rates	R-squared (uncentered):	0.942
Model:	OLS	Adj. R-squared (uncentered):	0.942
Method:	Least Squares	F-statistic:	6325.
Date:	Fri, 15 Nov 2024	Prob (F-statistic):	0.00
Time:	23:08:35	Log-Likelihood:	-3269.4
No. Observations:	777	AIC:	6543.
Df Residuals:	775	BIC:	6552.
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
accept_rate	35.5626	1.751	20.307	0.000	32.125	39.000
Outstate	0.0036	0.000	30.213	0.000	0.003	0.004

Omnibus:	24.867	Durbin-Watson:	1.964
Prob(Omnibus):	0.000	Jarque-Bera (JB):	35.733
Skew:	0.297	Prob(JB):	1.74e-08
Kurtosis:	3.867	Cond. No.	3.35e+04

Figure 6 -OLS Regression Model for Graduation Rates

In figure 6, we used an Ordinary Least Squares Regression Model(OLS). The dependent variable

is the percentage of Graduation Rates. The two independent variables selected were Acceptance Rates and Out of State Tuition Cost. Graduation Rates performed with an .942 R-Squared value. This tells us that there is about 94.2% variance in our dependent variable of Graduation rate. This means that there is a relatively strong correlation between our Grad Rates and our Out of state Tuition cost and acceptance rates.Our acceptance Rate had a coefficient of 35.5626 and a standard error of 1.751, meaning that for every time the acceptance rate increases by 1 the graduation rate increases by 35.5626. The small standard error helps us determine that our coefficient is fairly precise. Our Out of State tuition coefficient was .0036, meaning when the out of state tuition rate increases by 1 dollar the graduation rate increases by .0036 of an percent, that small change assumes that there is a positive relationship between the tuition rate and graduation rate.performed. The test computed a P value<.0001 which strongly indicates the rejection of the null hypothesis at a reasonable level of significance. This very low P value suggests that the model predictors collectively provide a significantly better fit to the data than a model without any predictors. A total of 777 schools were tested in this model.

Conclusion

Overall, this research investigated which variables had the highest impact on the rate of graduation across many out of state institutions. Multiple studies were evaluated and considered when creating the models. Some of the most relevant studies included the possible correlations between the independent variables which assisted in understanding the results of our model. The research used multiple regression models to display the correlation between different variables and is used as a predictor for percentage of graduation rate based on the given independent variables.

Before creating the models, a correlation heat map is used to determine which variables have the highest impact on graduation rate. The heat map analysis revealed that the largest factors correlating to success were the affordability of the college, otherwise known as the tuition. In the research project, we looked at 777 colleges and compared all of their affordability rates in relation to their graduation rates . As cited in Daniel Lang's *Does the Level of Tuition Fees Affect Student Retention and Graduation*? It is generally accepted that the price-response relationship is elastic and that the availability of financial aid and the cost of attendance both directly or indirectly influence students' persistence. In other words, persistence will increase if costs are reduced or financial aid is expanded, and vice versa. Our definition of graduation is measured by if a student completes their schooling. Finally, we decided that affordability, which is a variable representing how much tuition/overall costs a student has to pay, concluded as a logical correlator for graduation rates.

It is important to note that while the independent variable, affordability, was proven to be statistically significant, our hypothesis was in fact disproven. We theorized that as the cost of college increased, the graduation rate would decrease due to the time spent working to pay off the debt accumulated from attending that college. Although our hypothesis was refuted, our regression analysis revealed that when we utilize acceptance rates to display graduation rates for 777 private and public out-of-state universities, the results indicate that graduation rates rise in tandem with affordability. College tuition has a strong positive correlation with graduation rates which although is counterintuitive, can be explained by the idea that increased expense can encourage students to finish their coursework as soon as possible in order to get the most out of their tuition money. Additionally, higher tuition may be a tactic used by some universities to draw in more driven students who are dedicated to finishing their degrees.

Ultimately, graduation rates can depend on a multitude of factors and affordability plays an important role in deciding whether a student graduates or not, trends typically show that as tuition increases so does graduation rates.

References

1. Weisen, S., Maruyama, G., Hufnagle, A. S., Peczuh, M., & Do, T. (n.d.). Analyses of social issues and public policy - SPSSI journals - wiley. How are first-generation students doing throughout their college years? An examination of academic success, retention, and completion rates. https://spssi.onlinelibrary.wiley.com/doi/10.1111/asap.12137

2. Spica, E., & Biddix, J. P. (2021, July 23). Prices they pay: Academic achievement and progress to graduation barriers experienced by community college students due to the cost of course materials - innovative higher education. SpringerLink. https://link.springer.com/article/10.1007/s10755-021-09557-7

3. Cook, B., & Pullaro, N. (2010, September 1). *College graduation rates: Behind the numbers*. VTechWorks Repository. https://vtechworks.lib.vt.edu/items/d0842ecc-130e-4b97-9b2e-6c975c9df316

4. Dwyer, R. E., McCloud, L., & Hodson, R. (2012). Debt and Graduation from American Universities. *Social Forces*, *90*(4), 1133–1155. https://doi.org/10.1093/sf/sos072

5. Lang, D. (n.d.). does the level of tuition fees affect student retention and graduation?. DOES THE LEVEL OF TUITION FEES AFFECT STUDENT RETENTION AND GRADUATION? https://www.researchgate.net/publication/263735066_DOES_THE_LEVEL_OF_TUITION_FEES_AFFECT_STUDENT_RETENTION_AND_GRADUATION

6. Jibeen, T., & Khan, M. A. (n.d.). *Internationalization of Higher Education: Potential benefits and costs.* https://eric.ed.gov/?id=EJ1091722

7. Cook, B., & Pullaro, N. (2010, September 1). College Graduation Rates: Behind the numbers. https://vtechworks.lib.vt.edu/items/d0842ecc-130e-4b97-9b2e-6c975c9df316

8. Scott, M., Bailey, T., & Kienzl, G. (2006). Relative success? Determinants of college graduation rates in public and private colleges in the U.S. *Research in Higher Education*, *47*(3), 249–279. https://doi.org/10.1007/s11162-005-9388-y

9. Douglas A. Webber, decades, A. the last two, Pfeifer, C., Astin, A., Blose, G. L., Casella, G., Coleman, J. S., Constantine, J., Groot, H. D., Donlan, R., & Ehrenberg, R. G. (2010, April 20). Do expenditures other than instructional expenditures affect graduation and persistence rates in American higher education?. Economics of Education Review. https://www.sciencedirect.com/science/article/pii/S0272775710000488

10. Hemelet, S. (n.d.). *Sage journals: Discover world-class research*. https://journals.sagepub.com/doi/full/10.3102/0162373711415261?casa_token=7JNHU1jzFT0AAAAA%3As3lxz_G32PJqTD-FyR5Kz2TbhnUQN8llnsCkG6UsF30E9Pyobuc6mkmDzy4SsavyPPIfvNIvFKPj. https://journals.sagepub.com/

1. Group Git repository:

   https://github.com/Lasya5597/Group16dtsc/blob/a07e425fdfad92d43732cbe04547de742be142d8/College%20Affordability%20Project%20%20(5).ipynb