Effects of College Attendance Cost on Graduate Median Salary

Eric Livezey
Wentworth Institute of Technology

Abstract—**This document outlines a study into how the cost of college attendance affects the median salary of graduates. This study only covers universities in the United States.**

Keywords—College, Tuition, Salary

I. Introduction

There are many factors which determine which college students may wish to attend but perhaps some of most important factors for many students is the price. College in the U.S. is typically quite expensive with many private universities costing upwards of $50,000 a year in total fees. Public universities are typically significantly cheaper, but it can be hard for the casual observed to accurately determine whether or not a more expensive college might actually be more valuable. Perhaps the best way to find this out is to compare the cost of attendance to the median salary of its graduate students. If graduating students from the more expensive colleges actually make enough money to offset the increase in price, then it might make sense to say that the more expensive college is worth the price hike. First, we need to determine whether or not there is an actual correlation between these two statistics. If there is no correlation between cost and salary, then it would make any findings meaningless.

Once it has been determined that there is a meaningful correlation, we can try to find out how this correlation relates to specific colleges. We can do this by breaking down colleges into different categories. Since public and private schools tend to have significantly different attendance costs, it makes sense to separate them. It might also make sense to distinguish the predominant type of degree for a college since it affects how many years the students attend the university which will thus directly affect the cost. One might also wonder if this correlation remains true regardless of how high tuition is. If there is point where tuition keeps increasing but the salaries begin to plateau, then it might make sense to say that colleges with more expensive tuition than that point would typically be less worthwhile than cheaper ones. Lastly, while tuition may be one of the most important factors there are many other factors which might determine the salary of graduates, one being admission rate. It is a common stereotype that colleges with lower admission rates are typically "better" than ones with higher rates. One might assume that this means it would be more worthwhile to attend a more selective college, or at least that graduates at these colleges would tend to make more money. This project will try to prove whether this assumption is true or not.

II. Datasets

A. College Scorecard

The College Scorecard is a dataset provided by the U.S. department of education. This dataset provides a wealth of information about every official college in the country. This data is sourced from the colleges themselves and from various government departments.

B. Character of the datasets

The College Scorecard dataset can be downloaded in the form of several CSVs. Only the data set for the most recent available academic year (2022-2023) was used for this project. This dataset contains 6,429 universities and 3,306 different columns. The data was mostly used as-is, though there were a few additional columns added for ease-of-use. All relevant columns are outlined in table below.

TABLE I.   RELEVANT DATASET COLUMNS

Column Name	Column Information	
	Description	*Custom (yes or no)*
PREDDEG	Predominant undergraduate degree awarded	No
CONTROL	Control of institution (Public, Nonprofit Private, or For-profit Private)	No
ADM_RATE	Admission Rate	No
NPT4_PUB	Average net price for Title IV institutions (public institutions)	No
NPT4_PRIV	Average net price for Title IV institutions (private for-profit and nonprofit institutions)	No
MD_EARN_WNE_P*	Median earnings of students working and not enrolled 6-10 years after enntry where * is the number of years.	No
AVG_YRS	Average number of years a student attends the college before graduating. Calculated by taking the national average number of years to complete the predominant degree	Yes
NET_PRICE	Average of public and private net prices for the college.	Yes

a. Sample of a Table footnote. (*Table footnote*)

Fig. 1.   An list of used columns from the Colege Scorecard dataset.

III. Methodology

There were several methodologies used to analyze and process data in this project. This section contains an overview of each.

A. Linear Regression

Linear regression is a model which estimates the relationship between two metrics. Linear regression is very useful for finding linear relationships. It will generate a line which best fits the data, and a value called r which represents how strong the relationship is from 0-1. The biggest disadvantage is that it can only discern linear relationships so if data follows some of curve, then it won't be very helpful. All of the relationships visualized in this project were expected to be linear though, so it was used to generate most of the visualizations which show some sort of correlation. The matplotlib module does not natively support linear regression so the SciPy and Seaborn were used to calculate the r value and generate the visualizations respectively.

B. Calculating Total Cost

In order to compare cost of a college to the salary of its graduates we needed to have a metric which measures the total cost. The College Scorecard dataset only provides the average net cost for private and public universities for a single academic year. We define total cost as the total price spent to attend college to attain a degree. This means that if someone took 5 years to attain their degree, then their total cost would be the net annual price from the dataset × 5. We used the national average time to complete the predominate degree for each school for this calculation. A certificate tends to take about 1 year to complete, an associate's degree takes about 3 years while a bachelor's degree takes about 5 years on average [1]. We used these values to create the AVG_YRS column. So, to get the average total cost we used the following formula.

$$NET\_PRICE \times AVG\_YRS = TOTAL\_COST \quad (1)$$

This value was saved in a column for each institution labeled as TOTAL_COST. This can be somewhat problematic with some non-standard length schools though as the cost will often be significantly under or overestimated. Most colleges run fairly close to the average length though, so using that value will create accurate results for the majority of institutions.

C. Calcuating Return on Investment (ROI)

In this project we define ROI as the total amount of money accrued over a period of time minus the cost of education. This is a useful metric for estimating the "value" attending a particular college. If someone attends college for 5 years and works for 5 years, then their 10-year ROI would be the income from those 5 years of work minus the price of those 5 years of education. The College Scorecard dataset provides the median income for students 5-10 years after enrollment for most institutions. Unfortunately, not every year is always present, and it does not provide any data for before or after this time period. So, we estimated the missing years using some math. If we assume that income increases linearly for the first 10 years, then we can estimate the missing years by calculating a linear regression line for those first 10 years using the present data and then just assuming that each missing year falls exactly on that line. For years greater than 10, we just assumed that the salary remains the same. Salary actually tends to grow over the course of a person's life, so this approach is somewhat conservative, which is one of its biggest disadvantages.

IV. RESULTS

A. Correlating Cost and Earnings

The first question to answer is whether total cost and median earnings are even related. It certainly seems like they would be, but we need some sort of statistical proof. In order to do this, we first divide colleges by institution type (public, private nonprofit, or private for-profit). Fig. 2 and Fig. 3 show the average total cost and median earnings for these categories respectively.
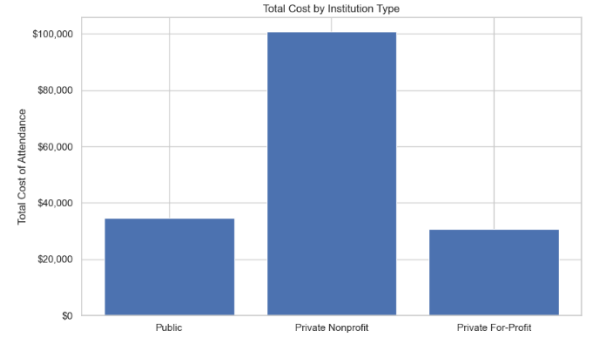

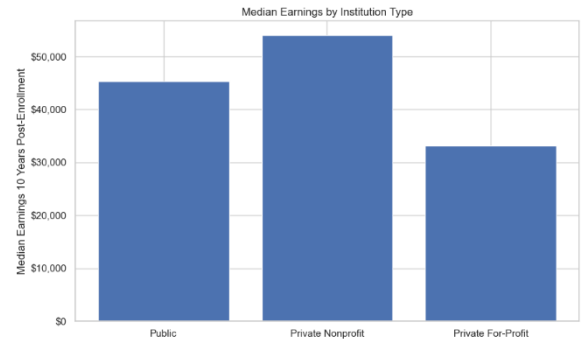Fig. 2.   Bar Graph of Total Cost by Instituation Type

Fig. 3.   Bar Graph of Median Salary by Instituation Type

These two visualizations alone will not completely tell of if they correlate or not but they give us some useful information. It's very important to note just how much more expensive private nonprofit institutions are compared to public and private for-profit. The earnings is also highest for private nonprofit but by nearly as much. Private nonprofit institutions are about 3 times as expensive as public institutions on average but only have less than 150% of the earnings 10 years post enrollment. This will be a major factor it the ROI calculation and can explain a lot of the differences in the data. We can also kind of see some sort of correlation between the cost and the earnings in that they are ranked in the same order for both by instituation type. The only reason thet the graphs looks so significantly different because of the massive increase in price for private institutions as noted earlier. But, in order to definitively see whether they correlate we need to calculate a regression line. Fig. 4 shows the result of applying a linear regression model to each instituation type.
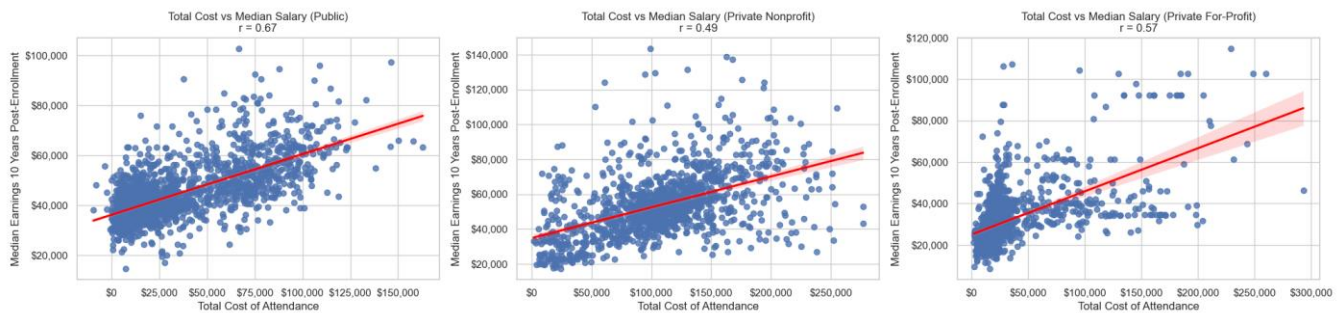
Fig. 4.   Bar Graph of Median Salary by Instituation Type

From this these graphs we can determine that there a correlation between the cost of attendance and salary. The correlation is strongest with public schools which has an r value of 0.67. This is likely due to the relatively low and more standardized attendance costs seeing as they are more closely related to each other than with private schools. It might seem somewhat surprisiing that private for-profit institutions have a stronger correlation than private nonprofit instititions but this is mostly explained by the degrees most offered by each instition type. For-profit institutions offer a large number of certificates. This results in a much more consistent total cost since programs are mostly the same length as well as more consistent salaries since certificates tend to make less money than more advances degrees. Nonprofit institutions on the other hand offer mostly bachelor's degrees which can vary significantly in terms of length. The average length of time to complete a bachelor's degree may be 5 years but many institutions take only 4 years to complete and in some rarer cases they can take up to 7 or even 8 years. This can skew the attendance cost a bit since we are always assuming that a bachelor's degree takes 5 years. We can also visually see that the data does not plateau at any point so we can assume that there is not some significant point in where returns diminish.

B. Return on Investment

For this project, ROI is defined as the total amount of income generated over a period of time after enrollment minus the total cost of education. The methodology for calculating ROI is detailed in Part C of Section III. The first visualization is simply a summary of the ROIs for each institution type. The 10 and 30 year ROIs are shown in Fig. 5 and Fig. 6 respectively.
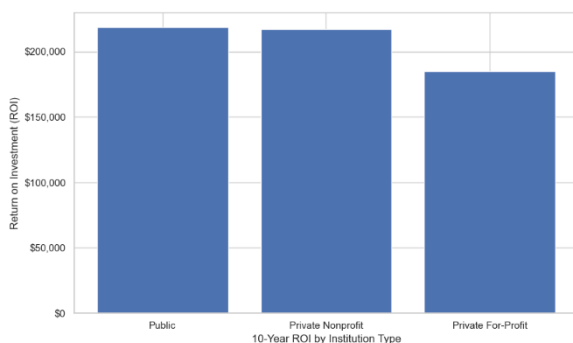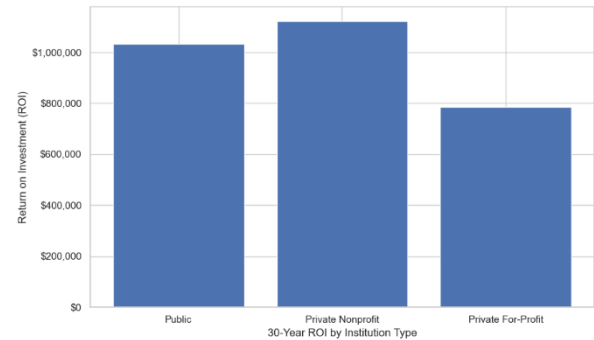

Fig. 5.   10-year ROI by institution type

Fig. 6.   30-year ROI by institution type

From these figures we can see how ROI differs between dfferent institution types. At 10 years post-enrollment, the ROI of public schools is actually the highest. Even though private nonprofit institutions tend to lead to higher salaries, the massive increase in total cost makes it so the price does not pay off until quite a bit later. Private nonprofit instiions actually only begin to overtake public institutions in ROI at around 20 years post-enrollment. The difference is extremely suttle at that point though so it is much easier to see on the 30-year ROI graph. At this point, private nonprofit colleges clearly have a greater return on investment than public colleges. Private for-profit colleges have a significantly lower return on investment for nearly all periods. This is because that while for-profit colleges may often offer lower total cost than other groups, the lower volume of higher level degrees makes the median income lower. We can see a breakdown by degree is Fig. 7.
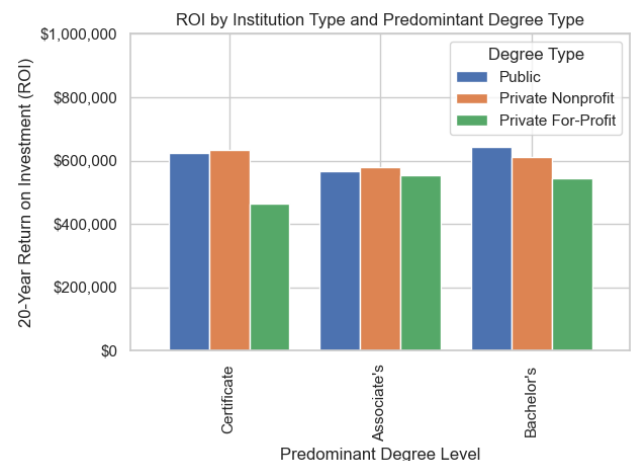

Fig. 7.   20-year ROI by institution type

We can see in this visualization how specific degree types affect the ROI. It may seem surprising that a certificate has nearly the same or even a higher ROI than a bachelor's and associate's in some cases but it makes sense when consider the time it takes to complete each degree. This difference is most pronounced in private nonprofit degrees. As shown in Fig. 2 the cost of private nonprofit institutions is nearly 3 times that of any other type. Because of this, the years it takes to complete a degree affects the ROI more significantly. This means that a certificate is massively less expensive than a bachelor's degree since it takes 5 times less time, especially at private nonprofit institutions. This explains why it can have a higher ROI despite having a significantly lower salary and why it takes more time to pay off higher degrees at nonprofit instititions.

C. Admission Rate vs Salary and ROI

The last thing we want to explore is whether acceptance rate has a significant effect on earnings or ROI. We can find this by simply plotting a regression line through the two statistics when compared to admission rate. Since we expect salary would increase with lower acceptance rates, the r value will be negative. This means that the closer it is to -1, the stronger the relationship. These visualizations can be seen in Fig. 8 and Fig 9.
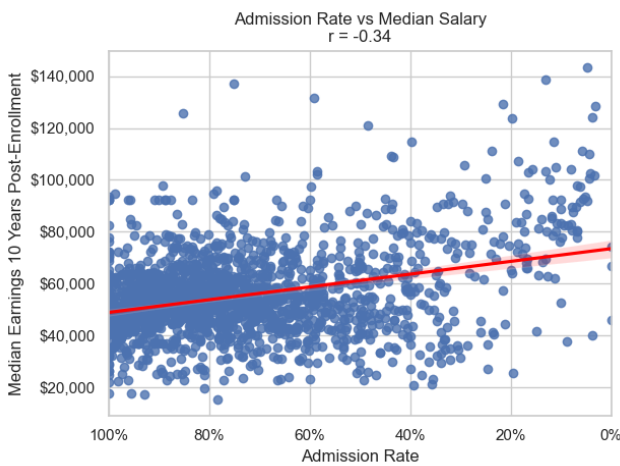

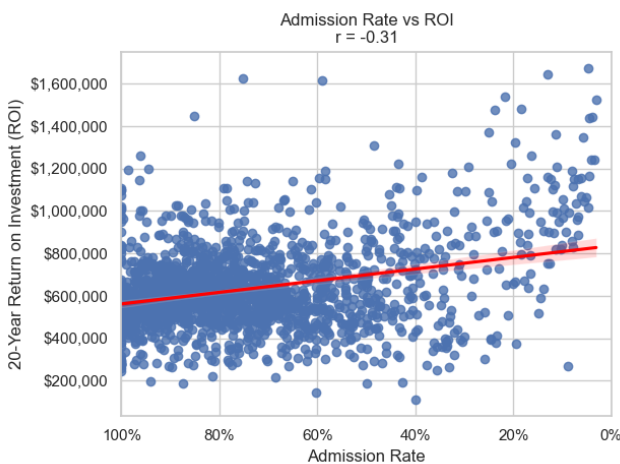Fig. 8. Median salary vs admission rate

Fig. 9. 20-year ROI vs admission rate

We can conclude from these visualizations that admission rate does not predict median salary or ROI very well. At least not linearly. It actually turns out that it might be related on a curve though. From just looking at the graphs it's pretty clear the salary increases significantly faster the closer the admission rate gets to 0. So it might be related exponentially but unfortunately linear regression will not illustrate that relationship.

V. DISCUSSION

While most of the findings are fairly conclusive, there are definitely some issues with the way some of the calculations were done. One of the biggest sources of error was the calculation of total cost. While assuming that every college takes the average amount of time to complete works fairly well for certificates and associate's degrees, it has issues when it comes to bachelor's degrees because of how much the length of time to complete it varies. It would be better if we had some actual data about how long it takes to complete a degree at each institution beyond just the average. This way we could more accurately determine relationships since colleges which take less time may charge more than ones which take longer.

Another issue with this is that ROI in the way we calculated it isn't the most ideal metric for value. While it does great for something that isn't extremely difficult to calculate, there are a lot of things it does not account for. What we really want to know is how much more money someone is making when attending a college and using their degree vs someone who attends a different college while accounting for all college related expenses. One of the reasons that this isn't accounted for is that the salary data is too general. Salary is only given for specific time frames after college enrollment for all graduates who are working and not enrolled. Ideally, we would only take the salary of students who got a job in their field. If we just take the median salary of every graduate, then it's not accounting for people who don't actually use their degree and thus shouldn't be considered when talking about value. Another reason that this isn't perfect is that we don't account for every expense. The biggest thing is debt. Ideally, we could integrate debt into our ROI calculations as an expense but it's too difficult to know how much debt students will have at specific points in time.

VI. CONCLUSION

In conclusion, this project illustrates some useful findings about college value. It was found that median salary is strongly correlated to attendance cost but much more so in public universities than in private ones. This means that it would be reasonable to say that broadly speaking public universities are safer than private universities since their tuition to salary rates are more consistent.

We can also conclude that private nonprofit universities will eventually have the highest return on investment but can take up to 20 years to reach that of public institutions. We can also say that private for-profit universities have a relatively low return on investment due to low post graduate salaries. Lastly, we can conclude that admission rate is only slightly correlated with median salary and ROI though it might be more correlated with a curve rather than a line.

All in all, these findings can help students better understand the value of certain universities. If a student finds themselves wondering if the increase in price in private non-profit universities is worth it, this data might be helpful.

REFERENCES

[1]   D. Shapiro., A. Dundar, P.K. Wakhungu, X. Yuan, A. Nathan, & Y. Hwang, "Time to degree: a national view of the time enrolled and elapsed for associate and bachelor's degree earners" (Signature Report No. 11), Herndon, VA: National Student Clearinghouse Research Center, 2016.