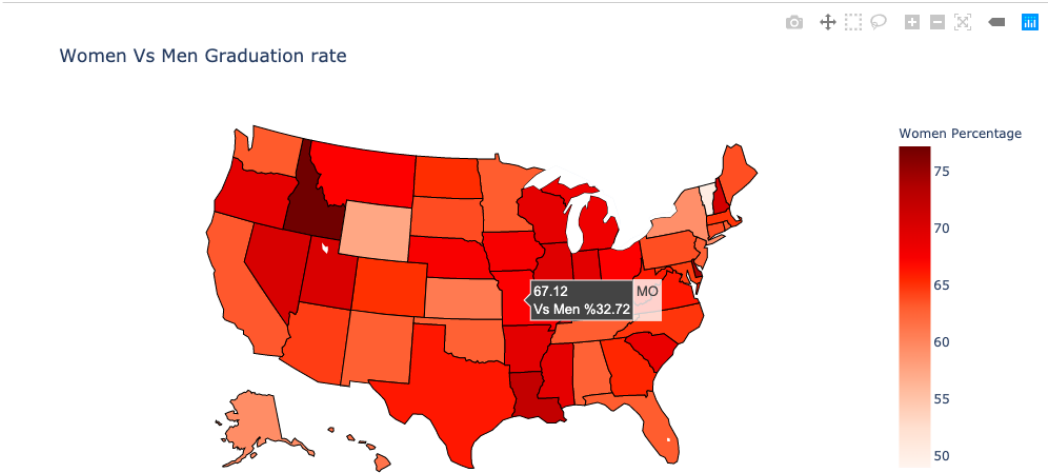
Student Loan Debts

We have extracted education loan data for the institution in the United States over last 10 years from 2007-17.

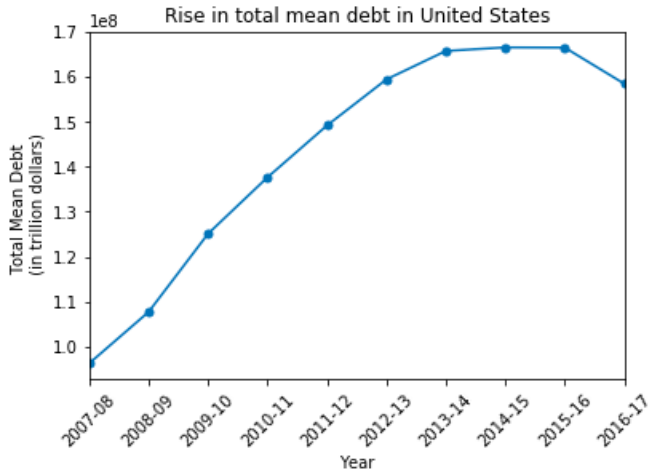
Visualization 1:

We have used the heatmap to visualize the percentage of male vs female students who have graduated from all the institutions in the United States. Comparatively, we have observed that there are a greater number of female students than male students.



Visualization 2

Here, we are performing a time series analysis of total mean debt for all the students who have graduated over the past 10 years.



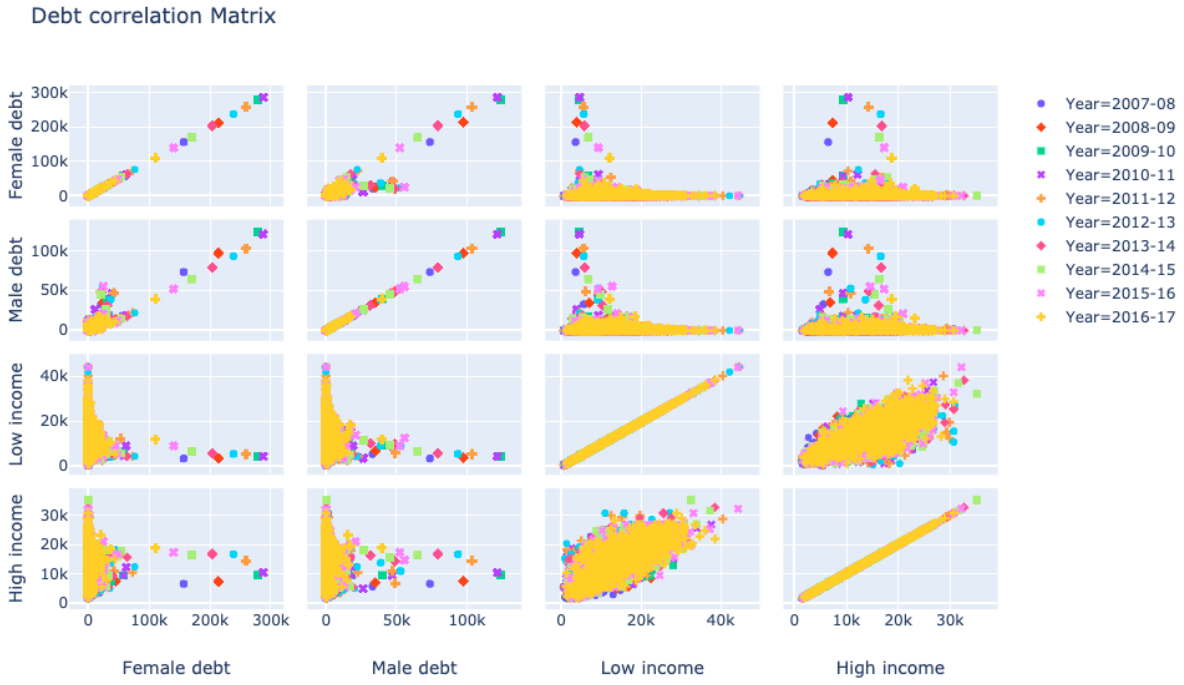
Inference:

We observe that the mean debt amount has risen steadily from less than a trillion dollars and risen to nearly 2 trillion dollars over a period of 10 years.

The amount reached a maximum of $1.7 trillion in the period 2014-2016. After 2016 the amount reduced a bit to $1.6 trillion and currently is stagnant at that figure.

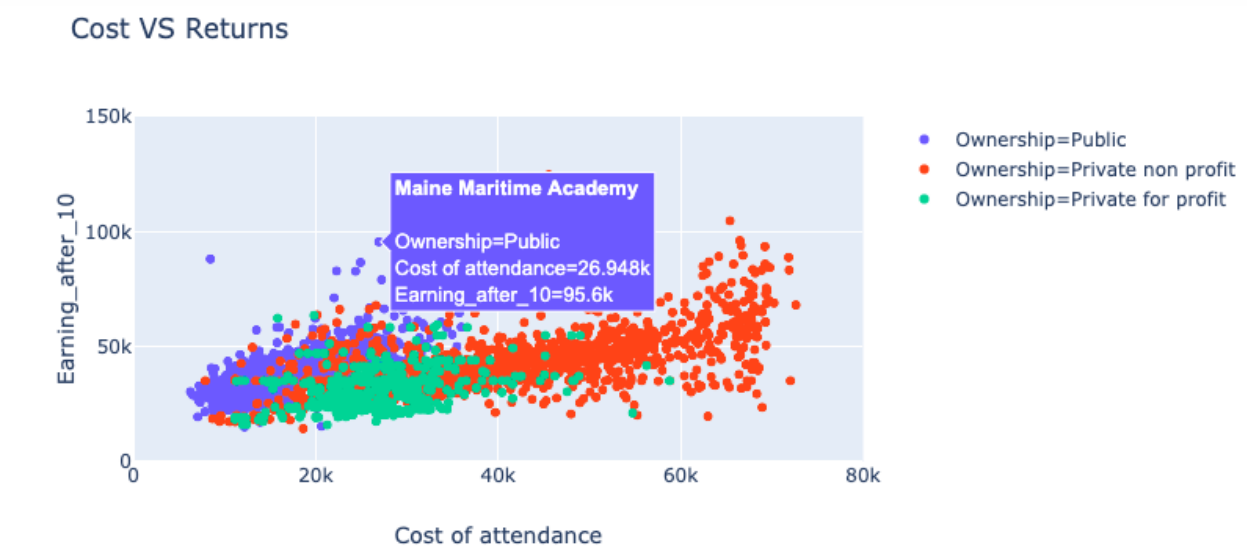
Visualization 3:

Let us see how the loan debt amount varies over the years for students from different groups. In the below correlation matrix we have found a relation between male, female, low income and high-income students loan debt. Male and female students belonging to high-income groups have comparatively low loan debt than students from low-income groups. We can see more data points for the year 2016-17 for all the correlation matrix, it suggests that a number of students from these groups have who are in loan debt have increased.



Visualization 4:

Let us check for all the students from the different groups mentioned above, where are they studying? What is the cost of attendance at those institutions & respective repayment rate? We have divided institutes into 3 types as public, private non-profit, private for profit.



Inference:

From the scatterplot, it can be seen that the cost of attendance for all the types of institutions varies majorly, something that can be confirmed with the statistical analysis performed above. The cost of attendance is highest for Private non-profit institutions and their earnings seem to be spread over a wider range, thereby possessing high variance with respect to earnings. Certain outliers can be seen to exist in this case.

The second-highest cost of attendance appears to be for Private for-profit institutions with their earnings over a period of 10 years lying close to each other in terms of variance.

The lowest cost of attendance can be observed for Public institutions, which is very low compared to the cost of attendance associated with Private nonprofit institutes. However, from the scatter plot it can be observed that the earnings from this investment are comparative to the earnings received by graduates from the Private nonprofit institutions. This goes to consolidate the high rate of repayment we observed in our statistical analysis above for Public institutions.

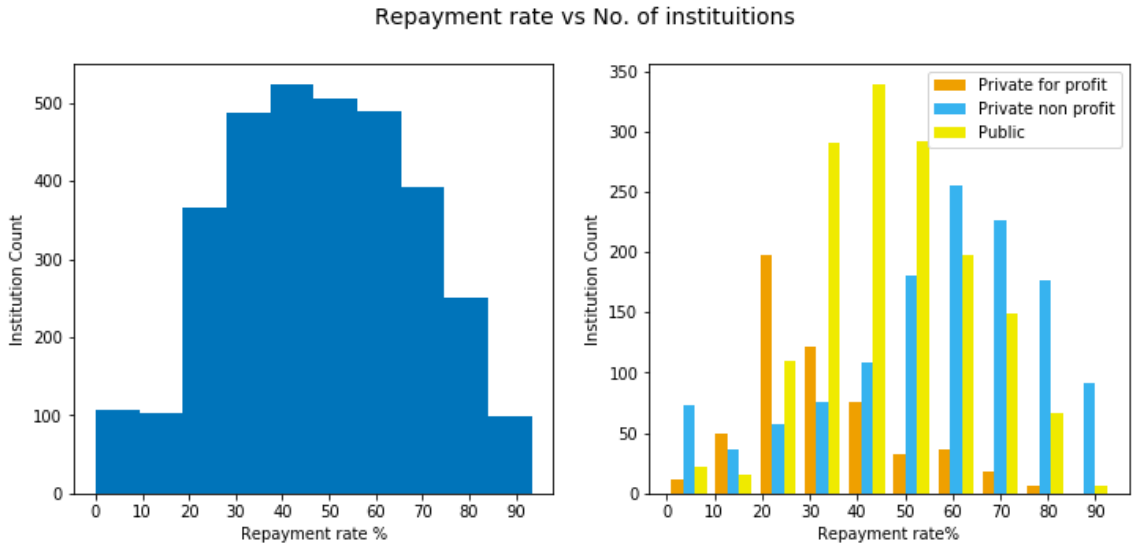
When we observe the outliers, we notice the universities 'St. Louis College of Pharmacy' and 'Albany College of Pharmacy' give the best rate of returns after a period of 10 years. We can generalize this by saying that pharmaceutical universities provide the best ROI.

Another interesting outlier is the public university 'The United States Merchant Marine Academy' which gives you the returns as high as Harvard and Stanford but at nearly 1/4th the price.

This answers our question whether the institution that a student graduates from affects the loan repayment rate because of the income earned after graduating.

Visualization 5:

Student Loan Debts



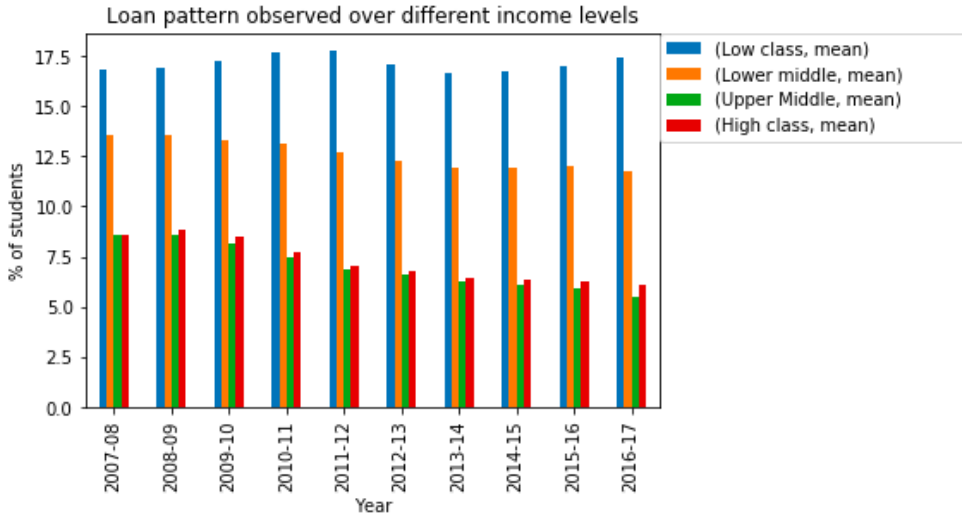
3-year repayment rate: It is the fraction of students who were able to make a payment towards their loan repayment amount in the duration of 3 years from the time they entered the repayment cycle.

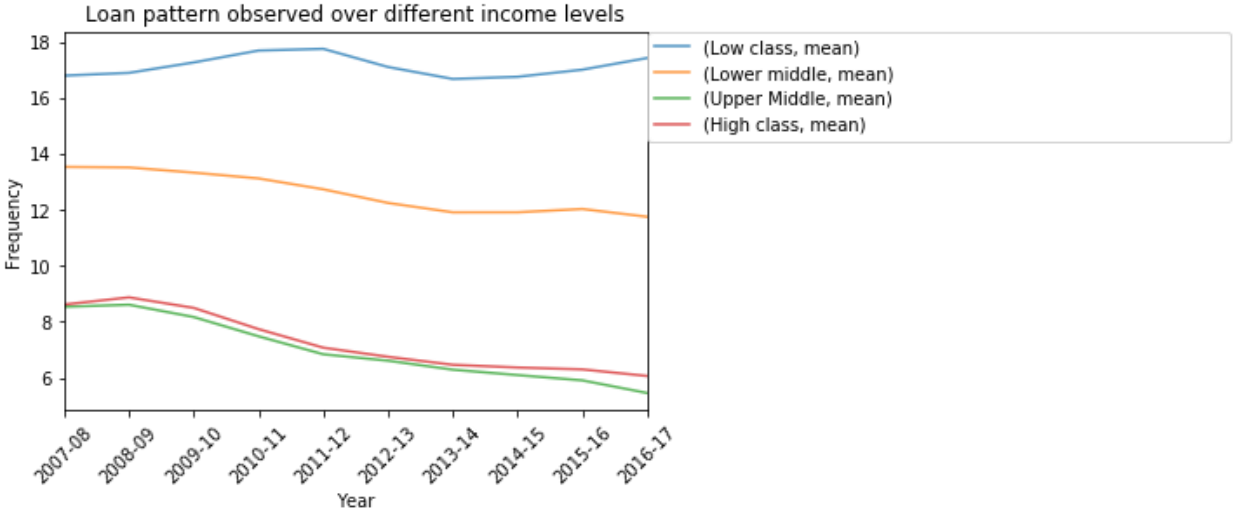
1. The majority of the private for-profit institutions have a repayment rate of only 20% which means that a large majority of students studying at for-profit institutions have difficulty in repaying their debts.
2. The repayment rate of public universities is a smooth normal curve with a mean of around 40%. This means that a major fraction of students enrolled at public universities is not able to pay off their debts.
3. This is a problematic situation wherein students from maximum number of institutions have not been able to contribute to their loan repayment due to possible reasons such as unemployment, insufficient funds, not enough return on investment etc.

Conclusion: The repayment rate at private non profit institutions is the best amongst the three with a mean of 60% which shows us that a major fraction of students is able to repay their debts.

Visualization 6:

Student Loan Debts





Inference:

The lower-income households take larger loans as compared to higher-income households. What's interesting to note over here is how the no of loans taken by higher-income households and upper-middle-income households have kept on reducing over the years while at the same time the loans taken by lower-income households have kept on increasing. This shows a high disparity between the median income of households.

As per our hypothesis testing, we know that private for-profit institutes and public institutes have low costs of attendance however their repayment rates are considerably lower than private non-profit institutes. Keeping income groups in mind, it can be assumed that students from low class and lower middle class will attend public or private for-profit institutes. This inherently will result in lower repayment rates (as per hypothesis) for students graduating from private for-profit institutes as compared to students graduating from private non-profits.

If this trend continues, it won't be surprising to see the disparity that exists between the 2 income groups to increase further. Hence, authorities should look into either reducing costs of attendance for private non-profit or improve repayment rates for other institutes.

Visualization 7:

OLS Regression Results

Dep. Variable:	Repay_rate	**R-squared:**	0.151
Model:	OLS	**Adj. R-squared:**	0.150
Method:	Least Squares	**F-statistic:**	294.4
Date:	Thu, 05 Dec 2019	**Prob (F-statistic):**	1.93e-118
Time:	12:36:16	**Log-Likelihood:**	836.14
No. Observations:	3321	**AIC:**	-1666.
Df Residuals:	3318	**BIC:**	-1648.
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.3291	0.008	41.079	0.000	0.313	0.345
Private non profit	0.2303	0.010	24.035	0.000	0.211	0.249
Public	0.1390	0.009	14.815	0.000	0.121	0.157

Omnibus:	149.133	**Durbin-Watson:**	1.537
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	178.756
Skew:	-0.490	**Prob(JB):**	1.53e-39
Kurtosis:	3.576	**Cond. No.**	5.20

From the linear model, we can see that the institution type is a significant independent variable that affects the value of the dependent variable considered here, which is the Repayment Rate. The base/ reference institute type here is Private for-profit. The inference from the regression model we can say that:

On average, the repayment rate for the Private Non Profit institute is 23.03% greater than the repayment rate observed for the Private For-Profit institute. On average, the repayment rate for Public institute is 13.9% greater than the repayment rate observed for Private For-Profit institute.

Hence, we can conclude that we reject our null hypothesis. The type of institution has effects on the repayment rate observed from these institutes.