University Retention Rate on an Institutional Level

By Daniel Immediato

NYC DSA – Python Project
February 27th, 2023

What is Retention Rate?

- Retention's most basic definition is: "the proportion of students from one year's fall term that re-enrolled during the following year's fall term."

- The average retention rate for the 2021-22 cohort was about 70%

https://collegescorecard.ed.gov/data/

Does it Matter?

Quality

- Retention dictates quality -higher retention means more prestige

Funding

- Colleges are encouraged to increase retention to receive federal funding

Completion

- Retention is directly correlated with completion rate

What Factors Influence Retention?

| Texas Case Study	College Scorecard
• The Texas University System in 2016 collected information from over 6,000 First Year students.	• The college scorecard is an institution-level data file updated annually by the Department of Education. It contains over 6,000 rows and 3,000 columns.
• This includes retention rate, GPA, High School GPA, the type of college attended, and whether they were instate or out of state.	• The data refers to various aspects of a university's "transparency". This includes completion rates, retention rates, average income of families, the demographics of the institutions, future success of the students, and so on.

The Given Factor: Overall GPA

- Educational success (probably) influences retention in some way

- The Texas case study allows us to see if GPA is an influencer, how much it acts as an influencer, and if we can apply that to other institutions.

- We can see here that overall GPA does impact retention rates.

Overall GPA and Retention
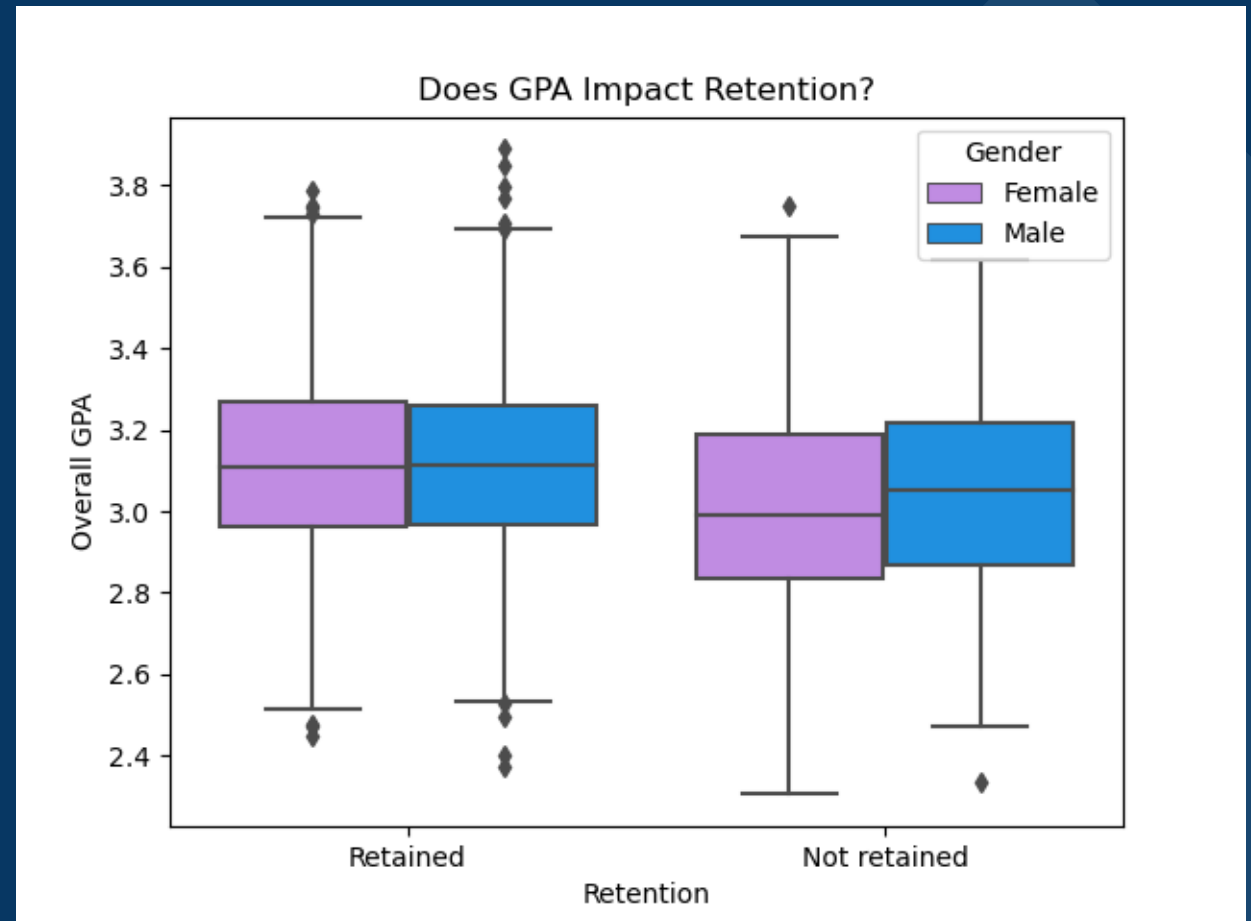
Retained Mean: 3.113918   Not Retained Mean: 3.023611

$P < 2.2e\text{-}16$

What About HS GPA?

- High School GPA shows the problems of outliers more prominently. There are so many outliers (more clearly noted in the violin plot to demonstrate density), that it caused the plot to be inflated around the center for the retained, being pulled by one extreme and the other.
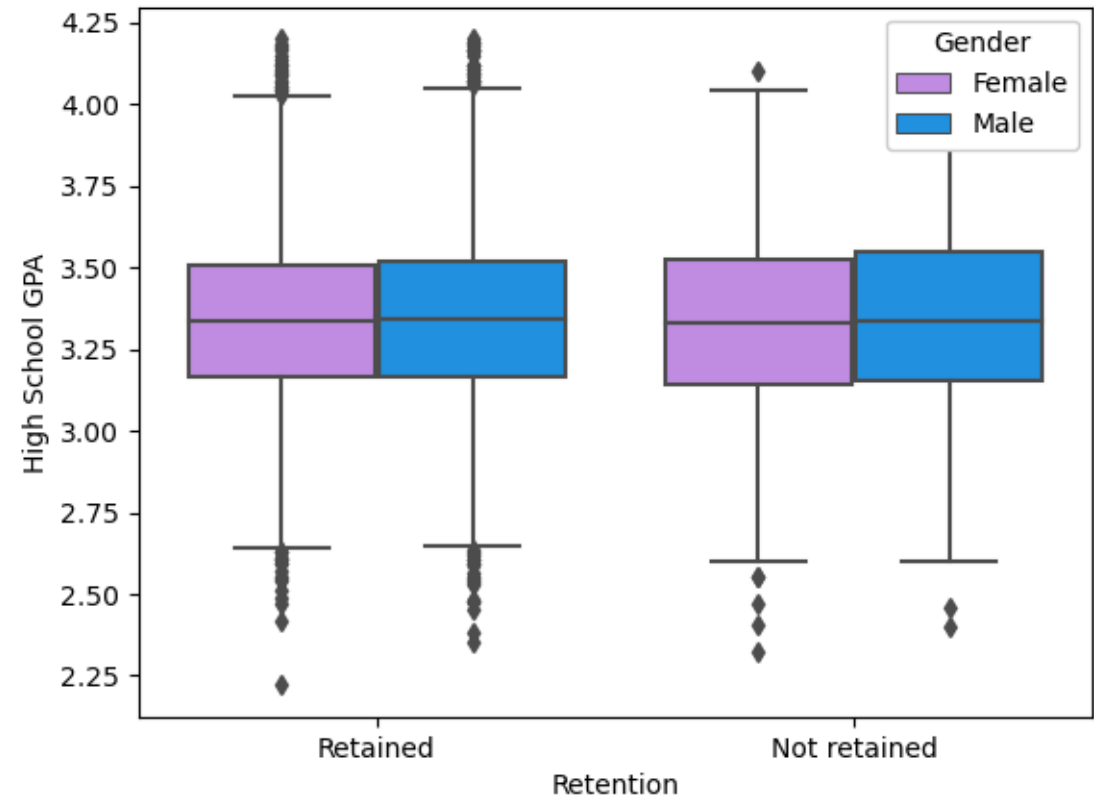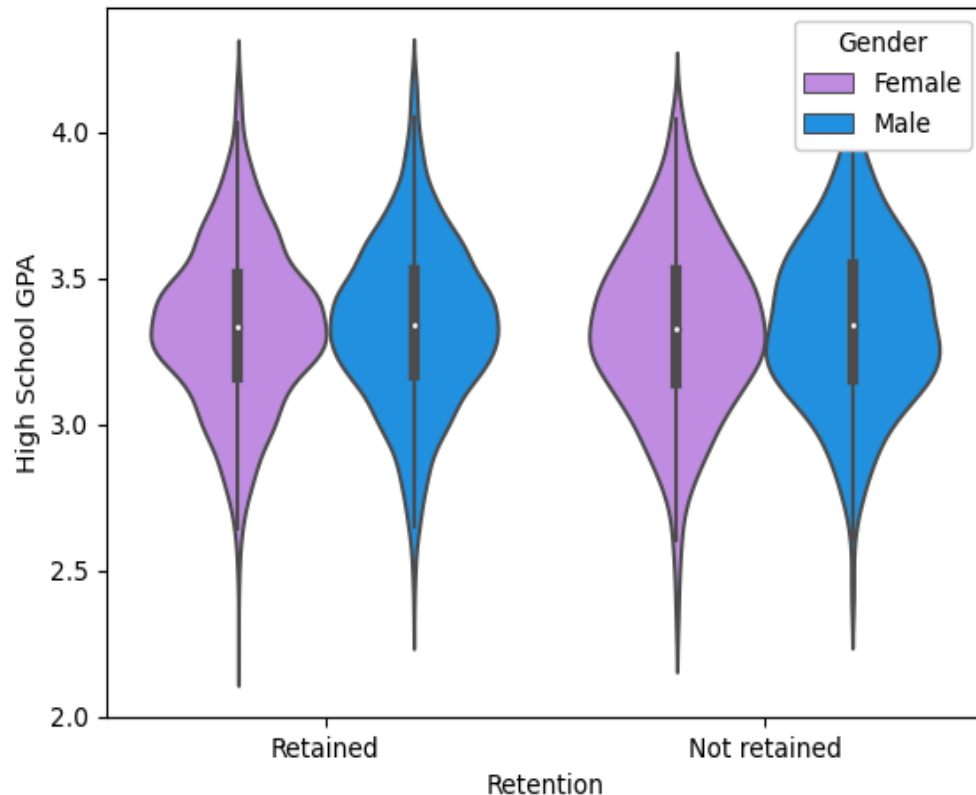
Retained Mean: 3.338728   Not Retained Mean: 3.340696
P=0.8433

Is There Hope?: Student Success

- Change in GPA is whether, over the year, the student's GPA changed in a positive or negative direction. Or didn't change at all.

- While most of the students had some sort of a change, there wasn't a significant difference when it comes to changes in GPA.

- This is likely due to "strength" of the changes not being taken into account
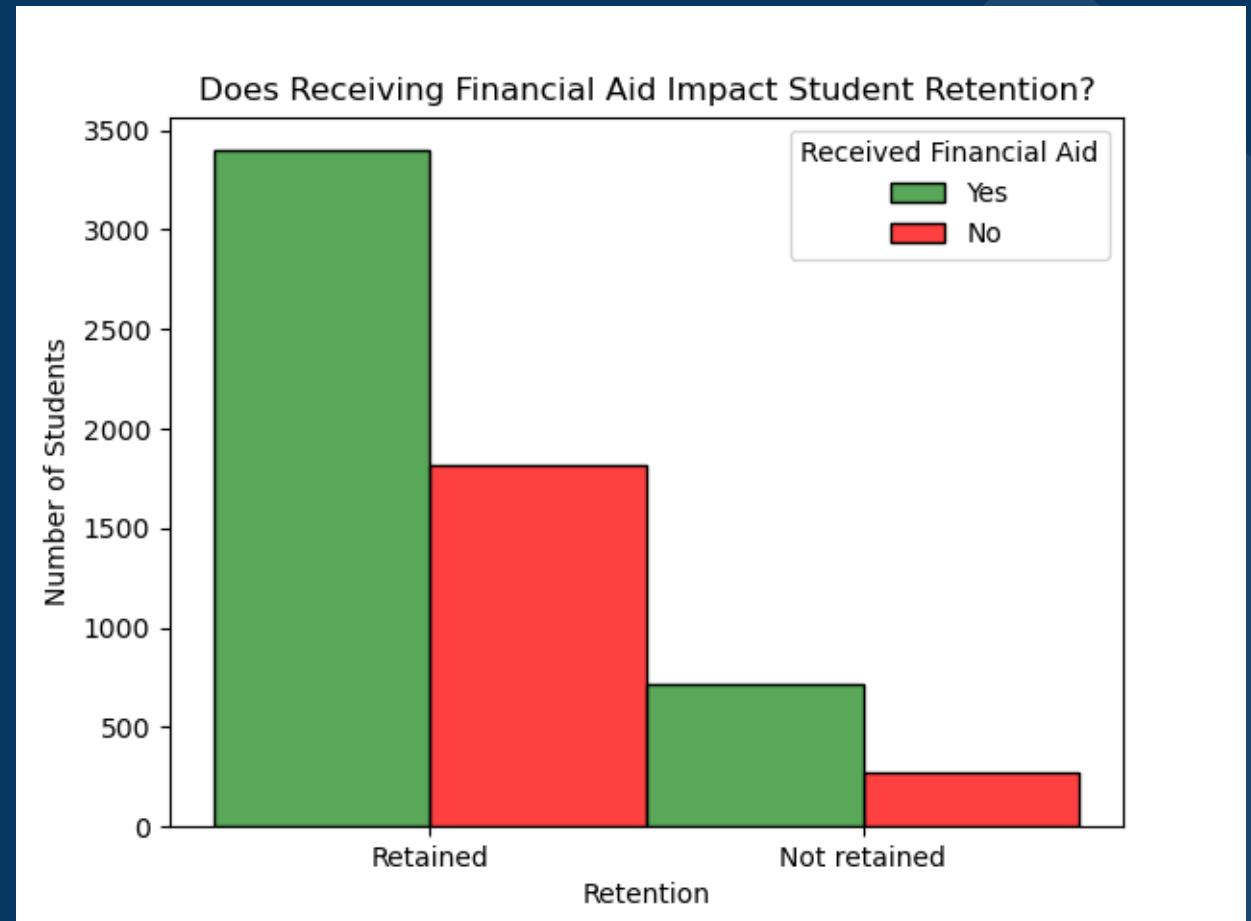
Direction of GPA Change Throughout First Year
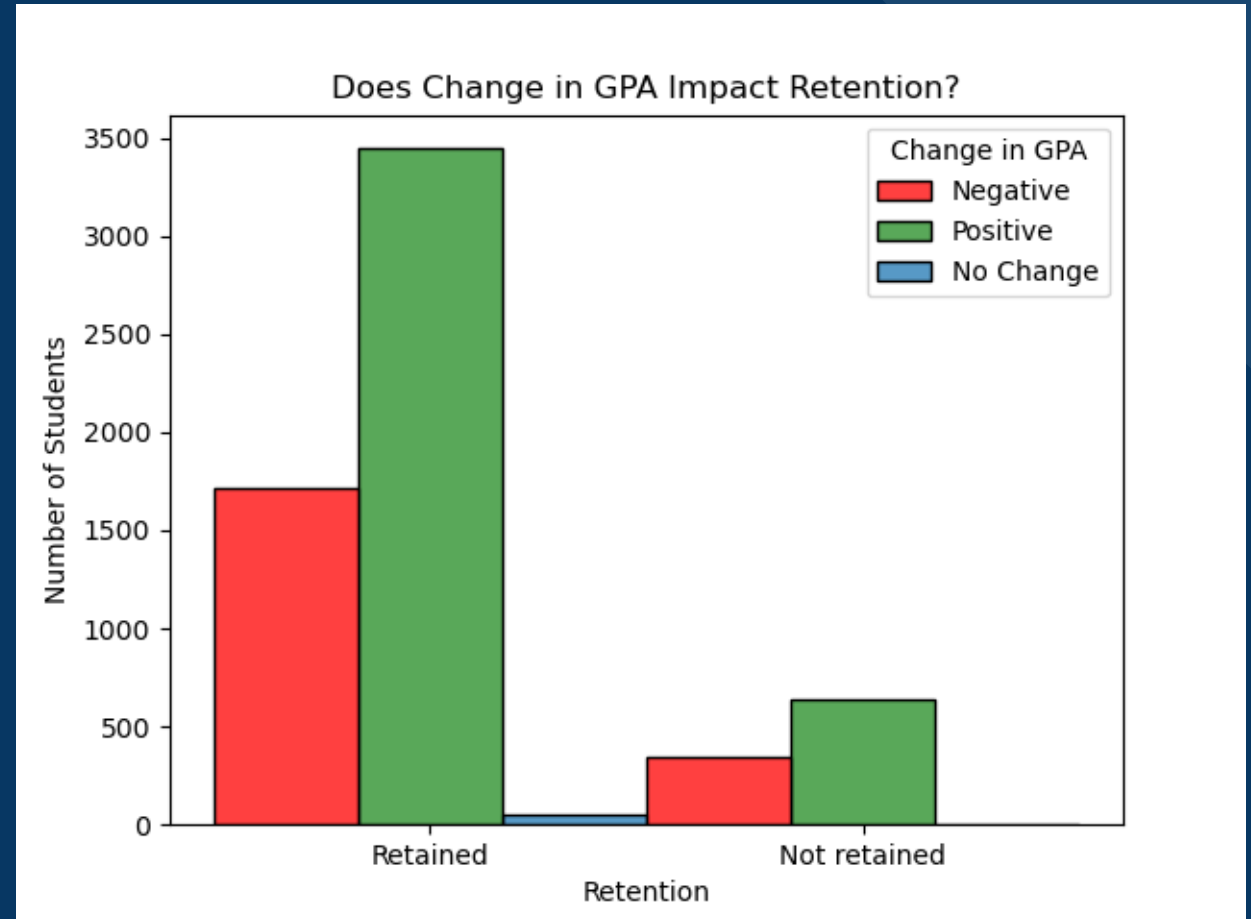
$X^2=2.969$
$P=0.2266$

Are Students Being Buried?: Debt

- Students often cite financial burdens as one of the main reasons as being unable to attend college. This was compounded during the coronavirus epidemic, which saw a national-level decrease in university retention rates.

- While the TUS case study does not provide actual university costs per study, nor cumulative debt, it does provide whether or not the student received any sort of financial aid.

- While the graph appears similar to GPA trends, the differences between means were enough to be statistically significant.

If the Student Received Financial Aid for First Year

Does Receiving Financial Aid Impact Student Retention?

Received Financial Aid
Yes
No

Number of Students

Retention
Retained   Not retained

X^2=17.193
P=3.376e-05

Differences of Retention

If GPA and debt influences retention, can that be applied on an institutional level?

TUS and College Scorecard measure retention differently.

- TUS was a categorical object

- College Scorecard was technically a "string with NaN values" and had to be converted to a float

- NaN values often had to be dropped, including if value was "Privacy Suppressed" (discussed in Future Works)

TUS Retention	College Scorecard
0 Retained	0 0.5590
1 Retained	1 0.8522
2 Retained	2 0.5000
3 Retained	3 0.8048
4 Retained	4 0.6315
5 Retained	5 0.8788
6 Retained	6 0.6094
7 Retained	7 NaN
8 Retained	8 0.6627
9 Retained	9 0.9243
10 Retained	10 0.7891
11 Retained	11 0.5733
12 Retained	12 0.4348
13 Retained	13 0.6252
14 Retained	14 0.5813
15 Retained	15 0.6238
16 Not retained	16 0.5618
17 Not retained	17 0.8064
18 Retained	18 0.5908
19 Retained	19 0.6411
Name: Retention, dtype: object	Name: RET_FT4_POOLED, dtype: float64

A Comparison: SAT Scores

- College Scorecard did not provide average GPA, so the closest thing it provided to academic achievement was SAT scores, which would therefore be an estimate of entrance GPA. And we already know that "change in GPA" might not matter.

- This is the most clear example of a linear positive regression we have. But, like with HS GPA, this should be taken with caution for two reasons: most colleges do not report SAT scores and a high SAT score usually means acceptance into a university that demands as such. That means the university is already successful at retaining students.

SAT Scores as Provided by the University (See Future Works)

Positive Regression
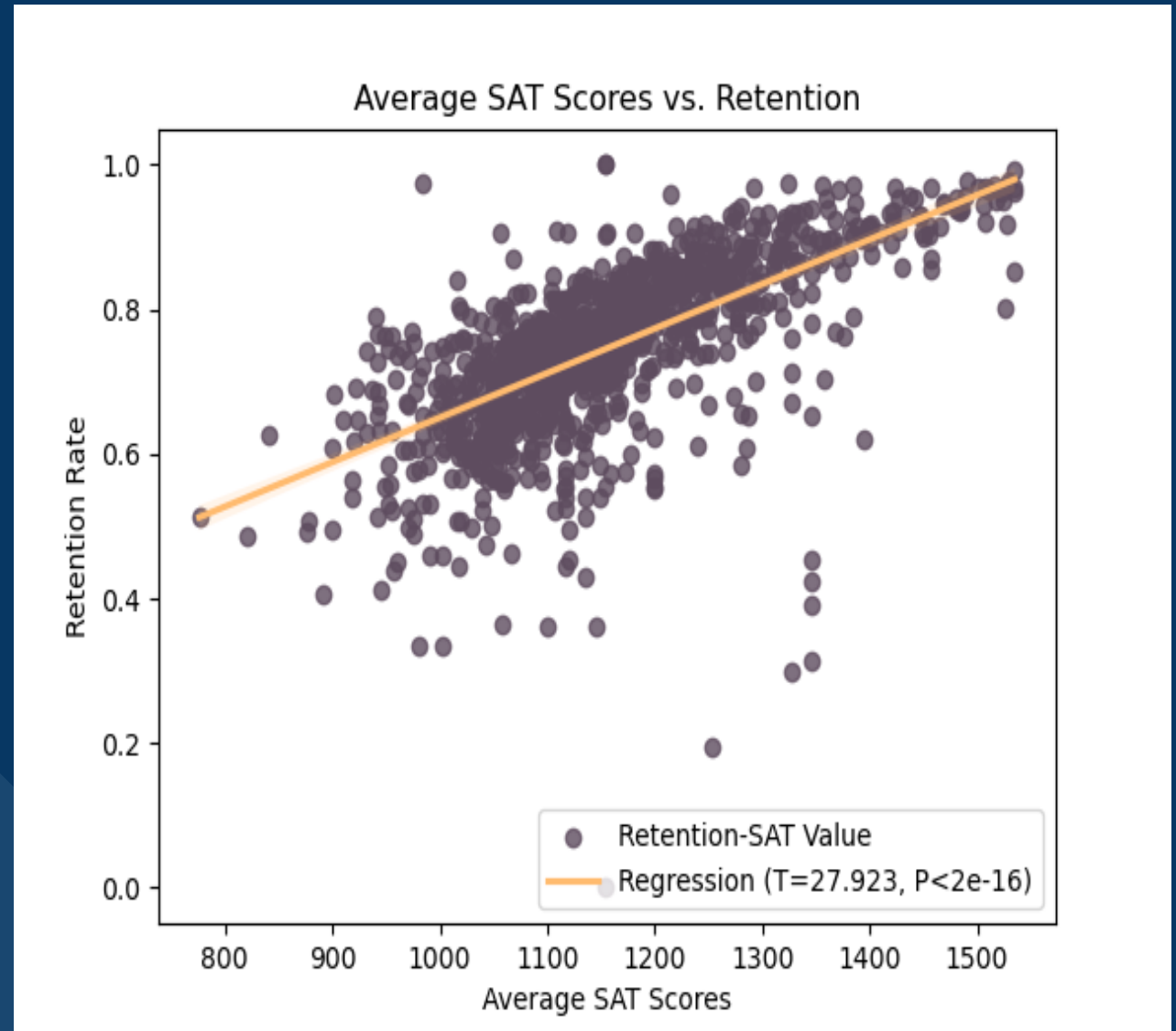
Easier for Some?: Family Income

- If one of the universal predicting factors in attending college is whether or not someone can afford it, then surely income would be positively correlated with retention.

- The cluster really just shows the average national income, but this does demonstrate that once moving away from the mean students are more likely to stay into their sophomore year.

- However, it should be noted that even those who have the mean family income still have a tendency to both have the greatest retention rates and also the worst.

Average Family Income Per Year

Retention vs. Family Income
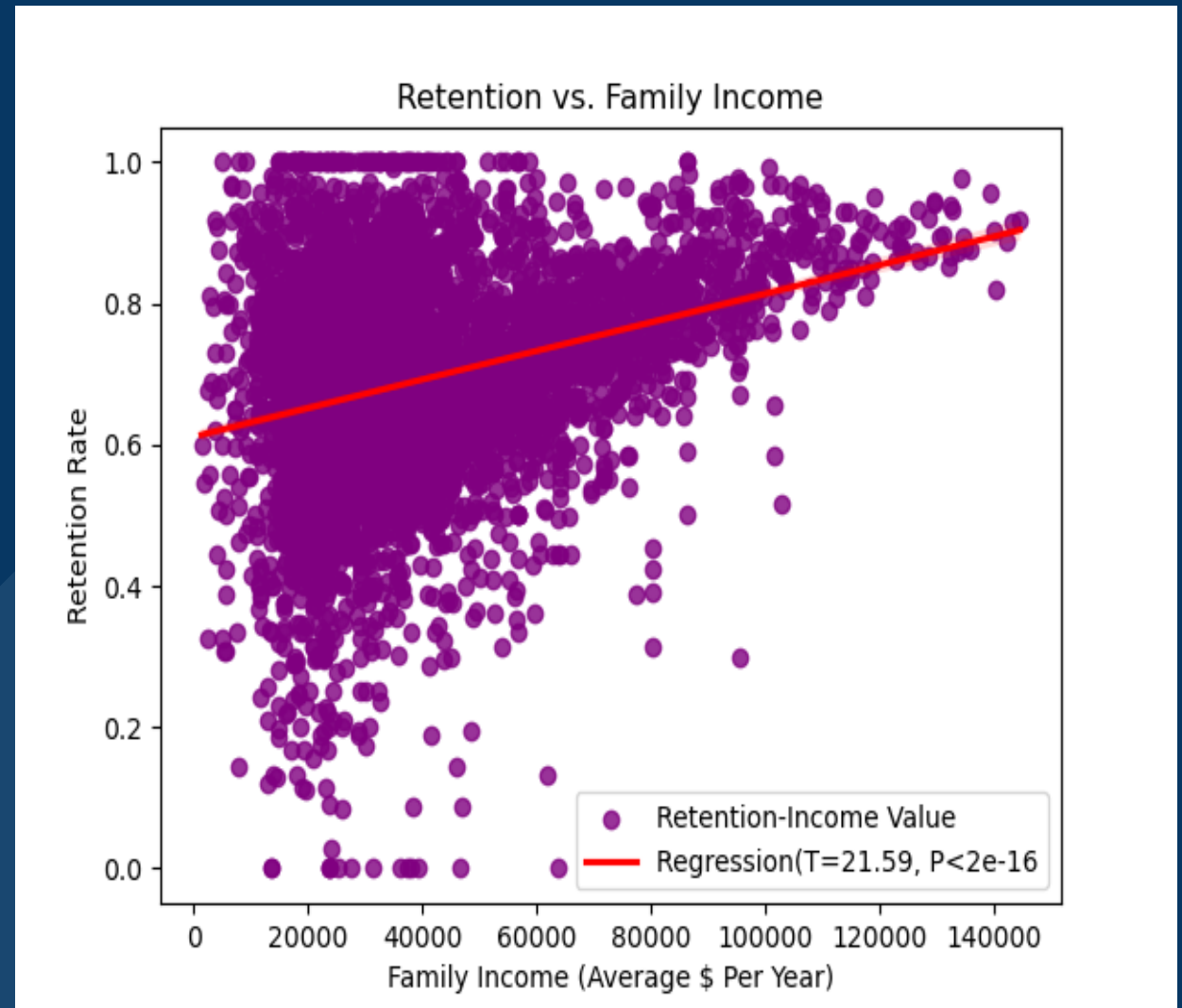
Not Always Expected: Tuition Cost

- One of the most significant factors on "why" one might attend a university would be tuition, but what about with retention? Unexpectedly, it would seem that the more costly a university is, the more likely students are to come back to it. This is probably due to: investment, where students feel obligated to complete a year they spent a large amount of money in, and faculty/reputation, universities that cost more usually have better facilities to deal with factors that affect student retention.

Retention vs. Instate Tuition

Outside Tuition vs. Retention
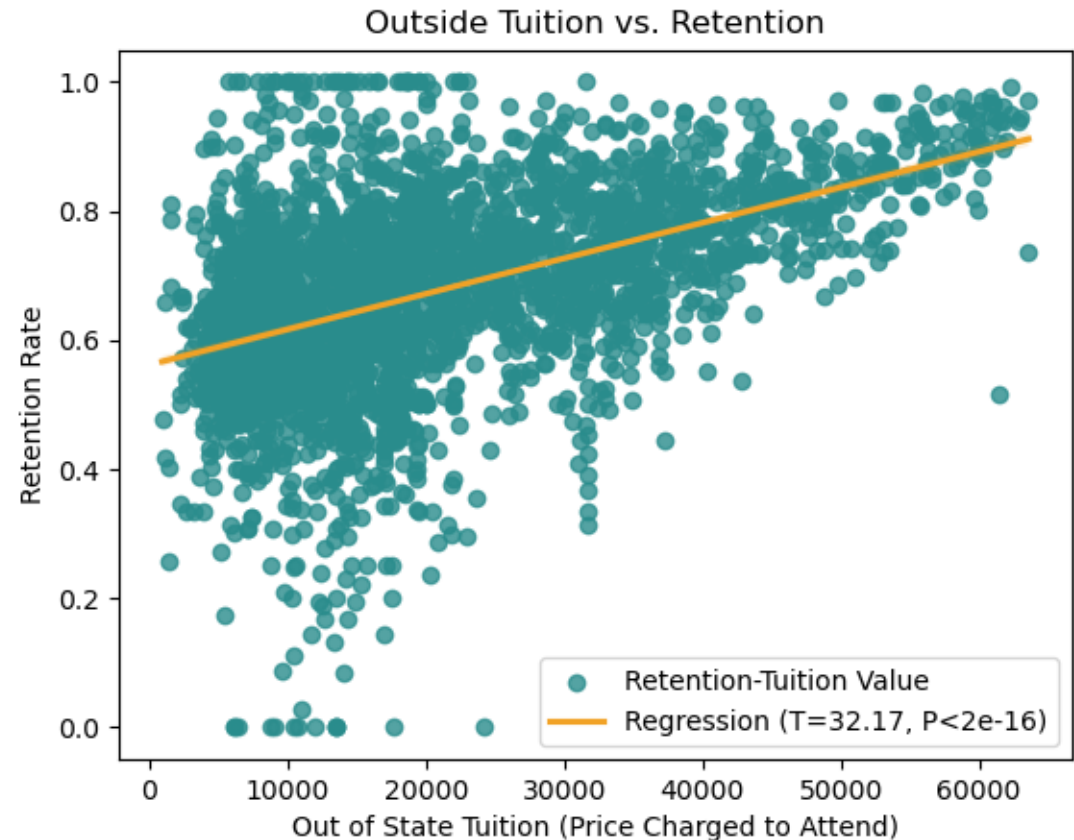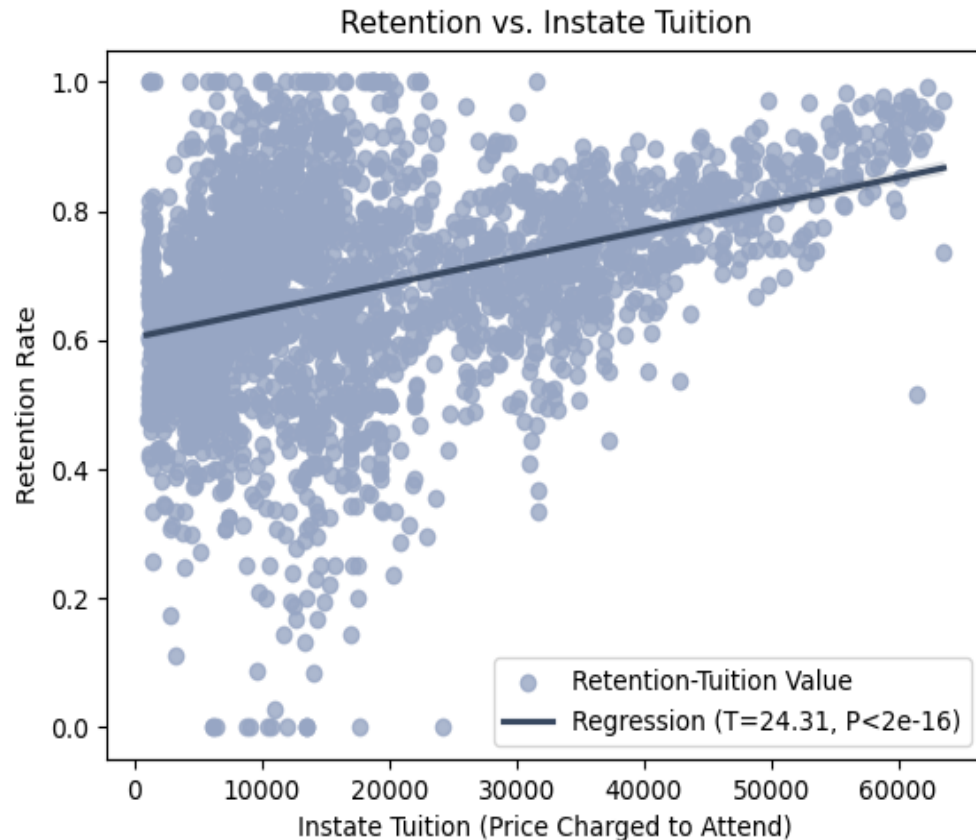
What Was That About Debt Again?

- College Scorecard has a somewhat similar variable as the TUS did for debt, except it's measured as a percentage/ratio. Here it's defined as "percent of undergraduates receiving federal loans".

- This graph is hard to read and all over the place, but is included so one can see that there is no statistical significance.

Percent of Undergraduates Receiving Federal Loans

Retention Rate vs. Student Loan Rate
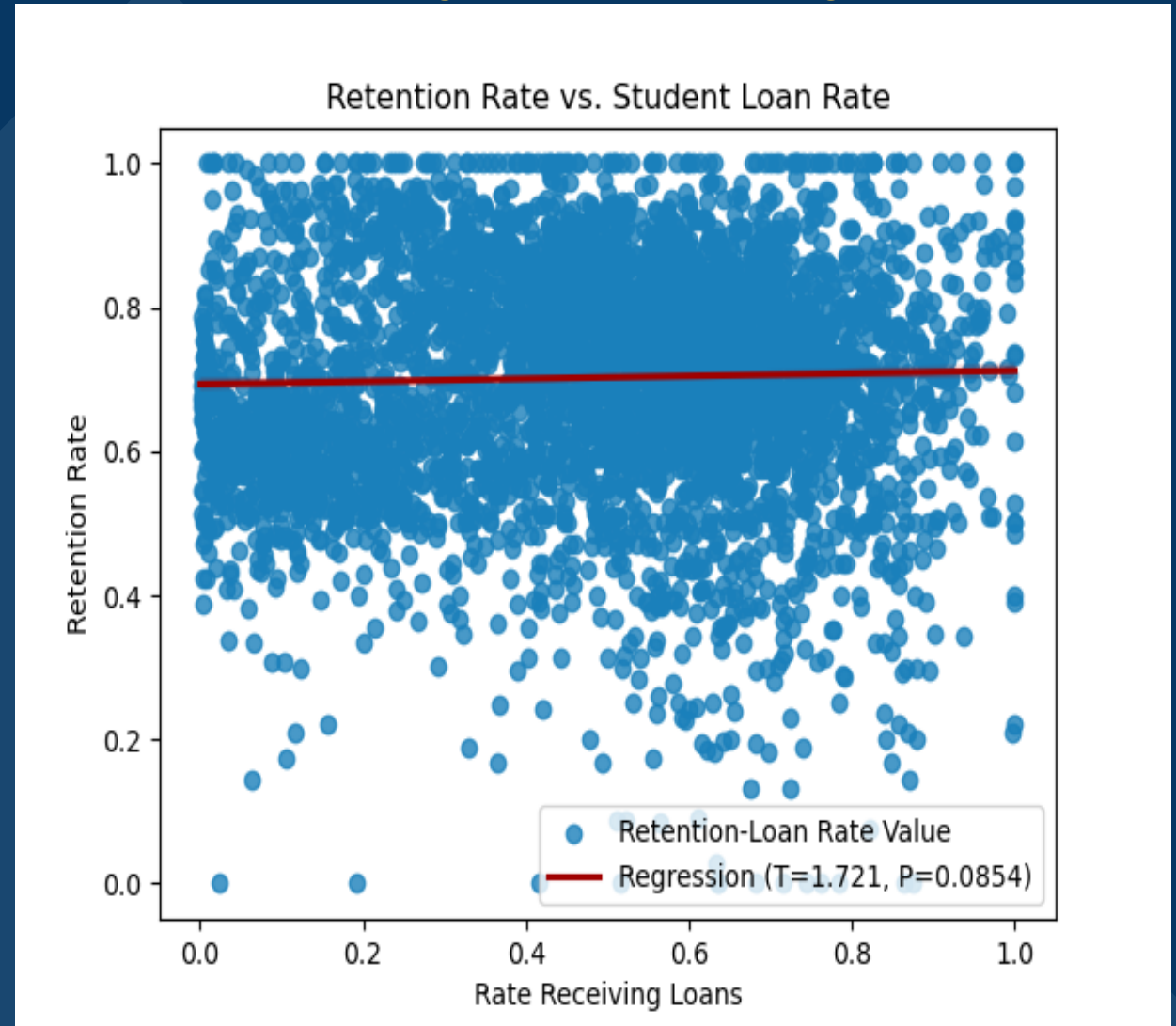
- A heat map can be used when there's large clusters, and we want to see the frequency of where the clusters are. The bar on the right shows the frequency/elevation level.

- We can see more clearly that, while most students are receiving loans, it's offset by a cluster not receiving loans and still having similar retention.

Percent of Undergraduates Receiving Federal Loans
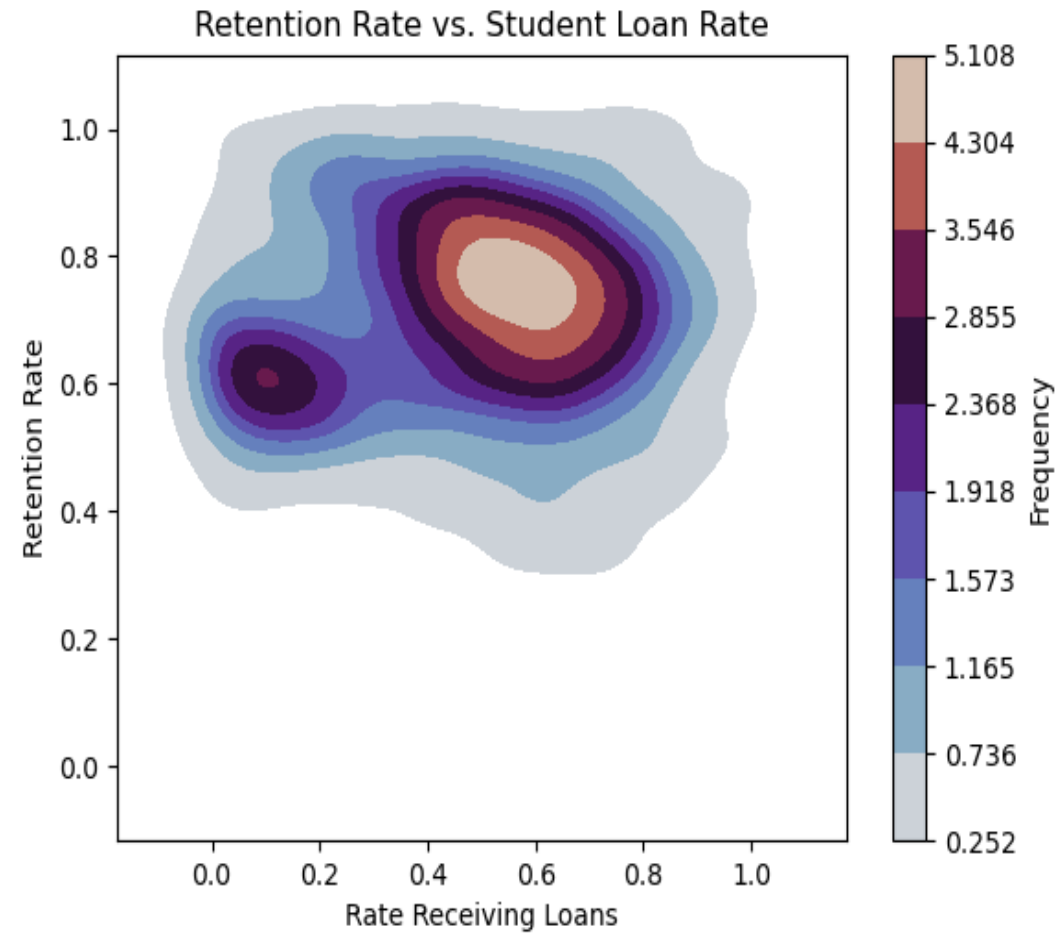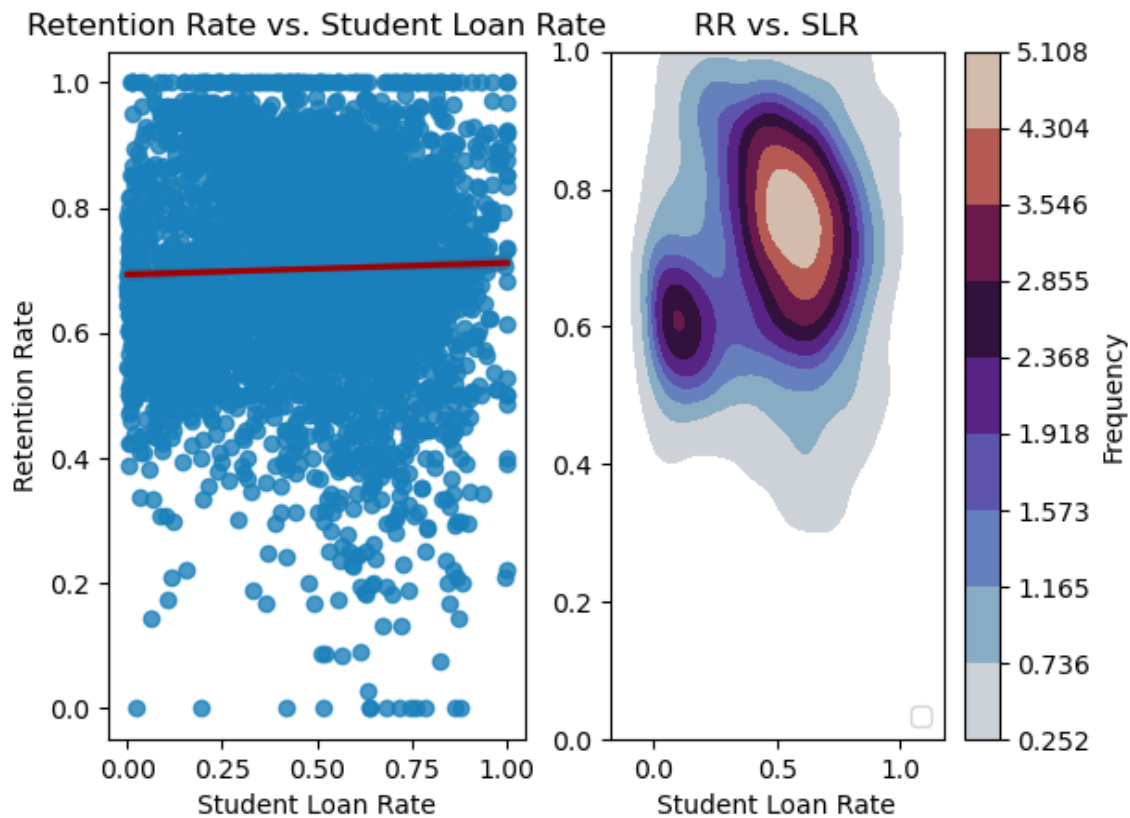
Can We Measure Loans Differently?: Default

- The default rate is a measure of the sense of debt burden of attending college and loan performance. It also measures a borrowers' behavior after they leave the university.

- Interestingly, for universities with a higher default rate, they also had a lower retention rate, even though these students would currently be experiencing a deference.

- Students experiencing severe financial burden may default, and if they don't have the right resources to deal with this then it show cases a college's inability to encourage success (See Future Works and the correlation between Retention Rate and Completion Rate).

Two-Year Cohort Default Rate

Retention Rate vs. Default Rate
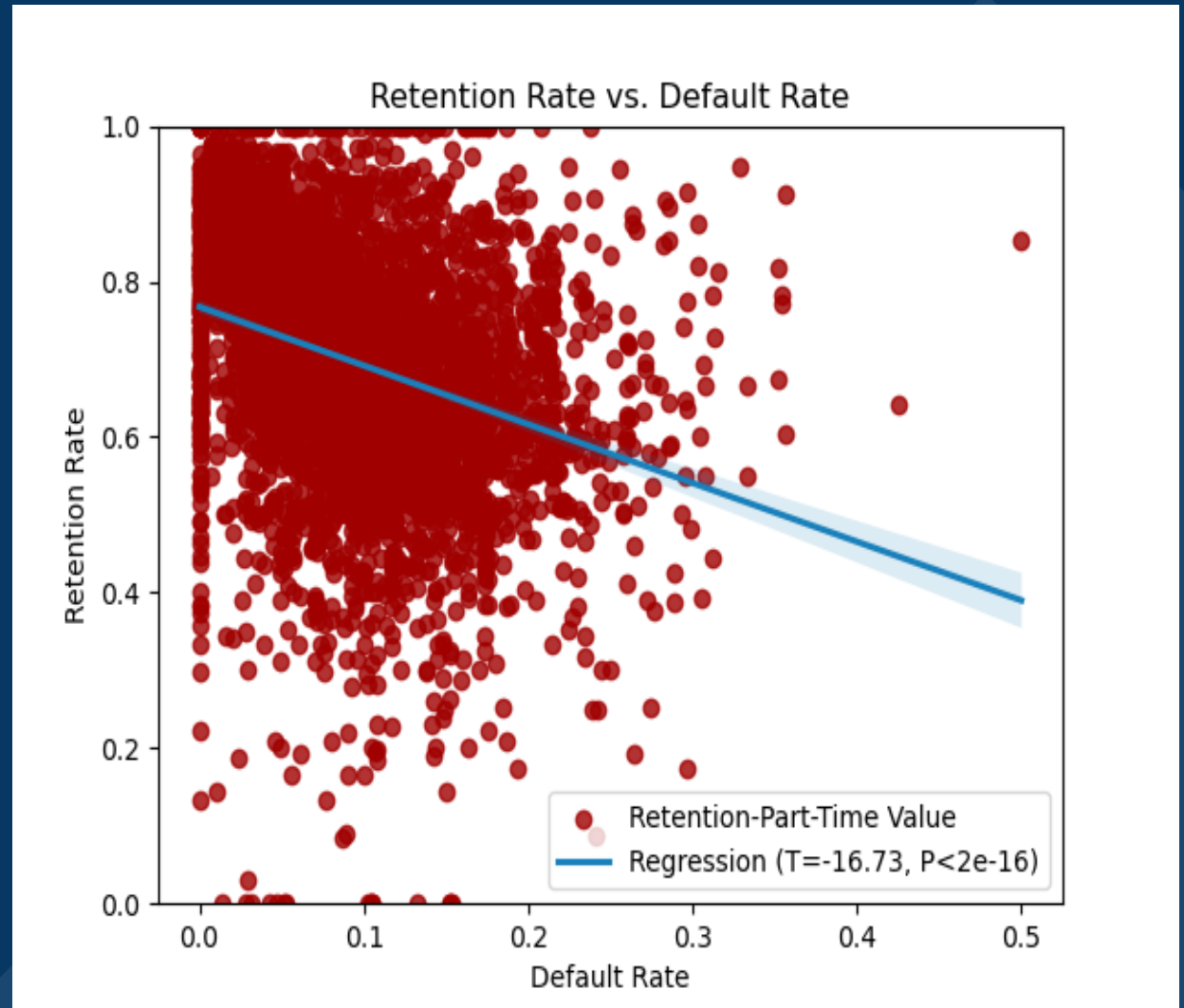
- A heat map more clearly shows the negative correlation, where the cluster begins in the "upper left" with a high value and then moves towards the bottom right. This shows that, if they have a low default rate, retention rates are high, but as soon as they do, retention rates start lowering.
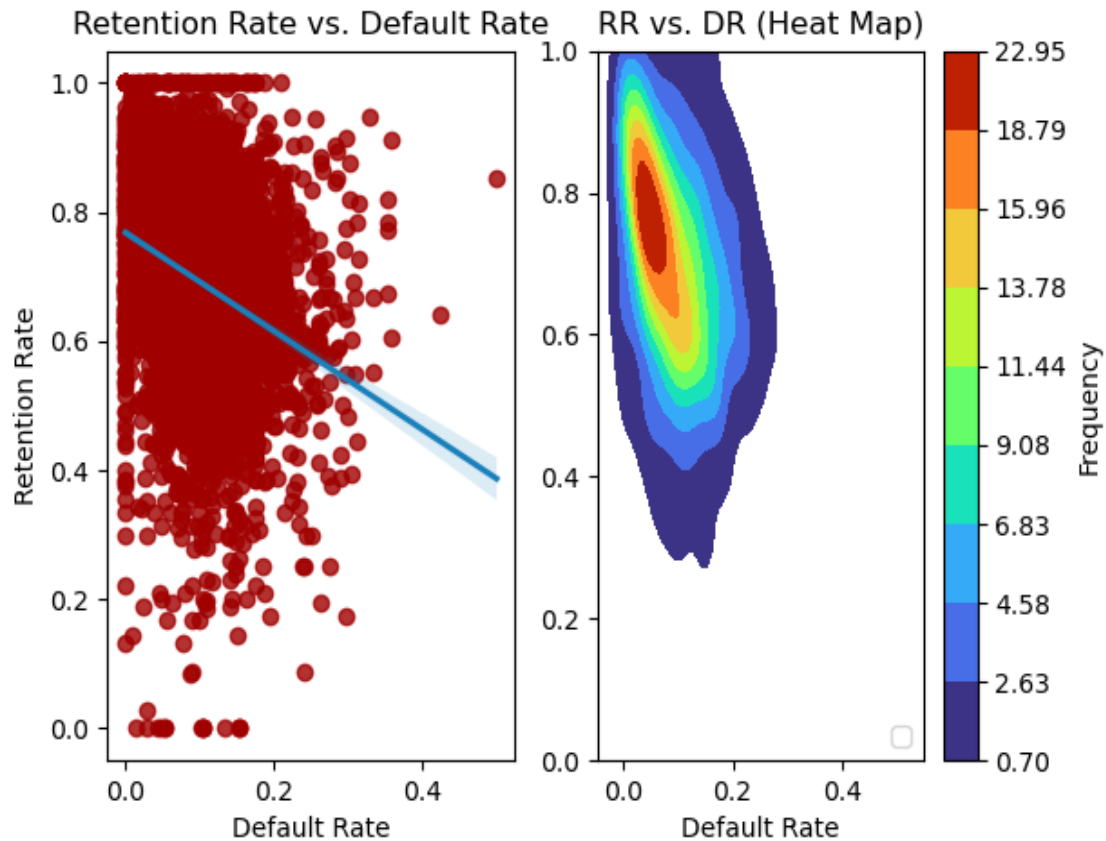
Two-Year Cohort Default Rate
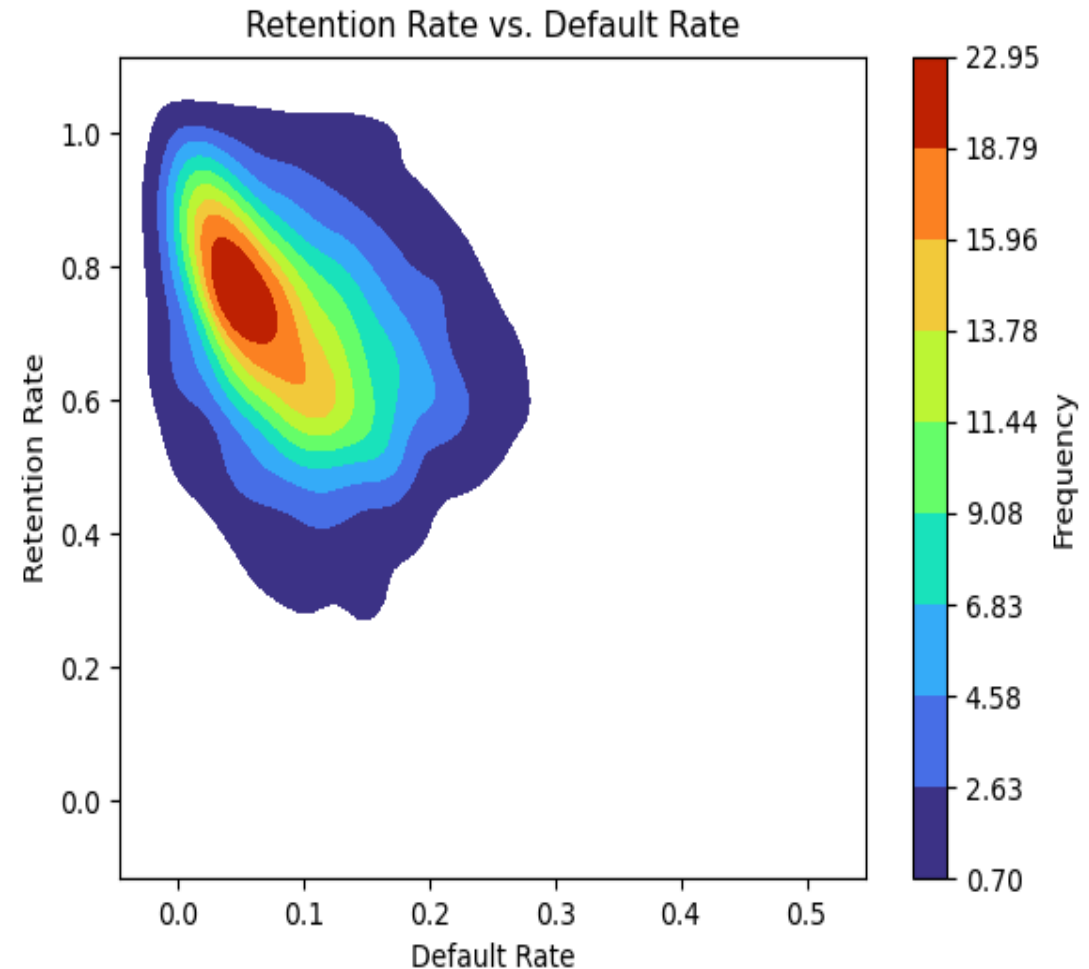
Does Part-Time Equal Commitment?

- Retention rate had a separate variable for part time students. However, there was a variable for the proportion of students that were seeking part-time at the university.

- It's no surprise that those individuals who invest more time in the university are more likely to return. These are the students are better able to take advantage of its facilities, attend its clubs, interact with its professors, and may also be the focus of university outreach programs.

Proportion Seeking Part-Time

Negative Correlation
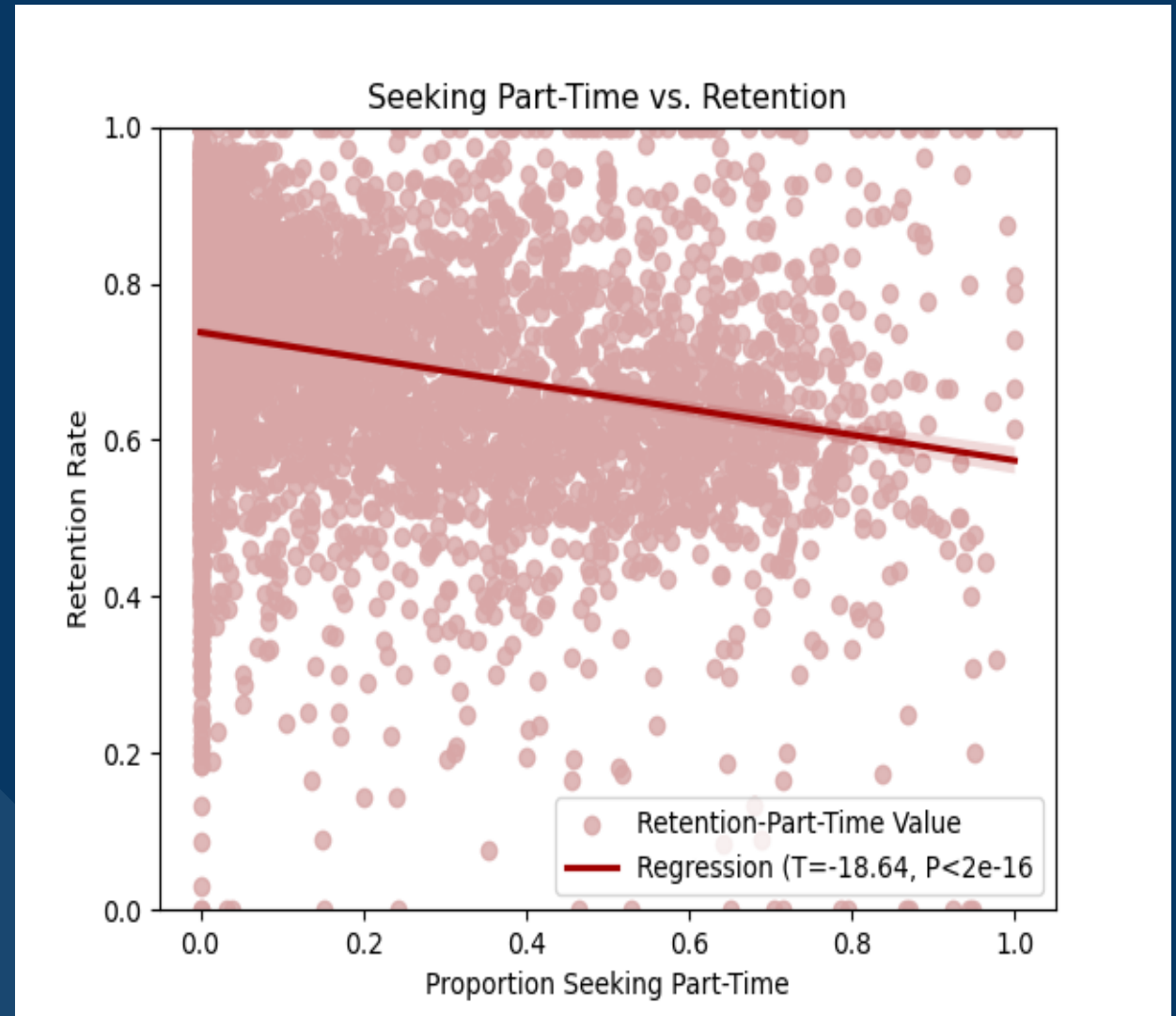
That Extra Push: The First Generation

- In the student's FAFSA, if at least one parent's status was "college" then it was coded as non-first generation.

- First generation students are often seen as "risk" because of their uncertainty in navigating the college landscape. They have less aid from their parents, and, if previous graphs are to go by, they might have less family income which means less retention.

- This graph also shows what is to be expected. First generation students tend to have lower retention rates than their counterparts.

% of Students Who Had No Parents In College

% as First Generation vs. Retention
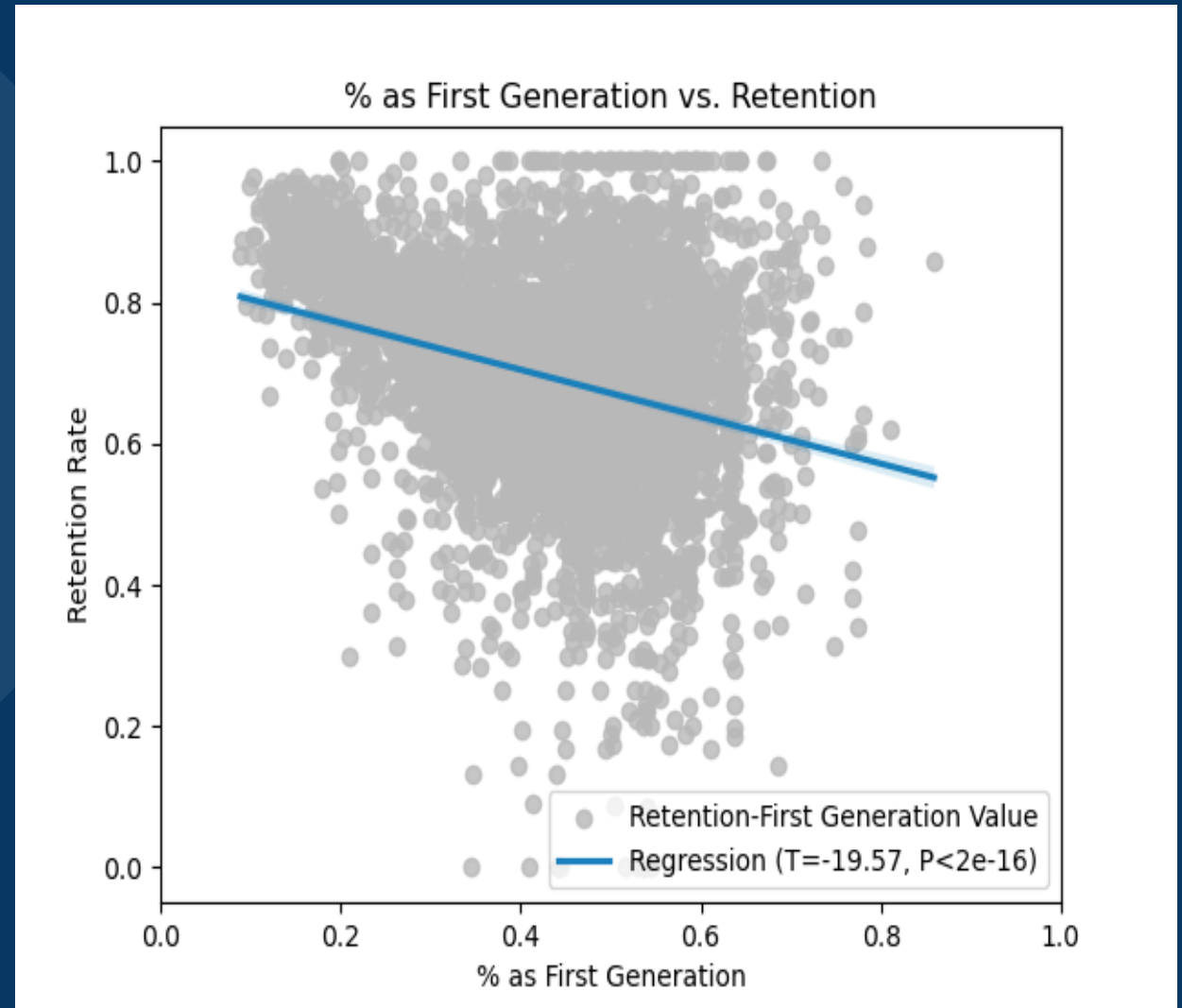
- When we break down the first generation into density, we can more easily see two things: most colleges have above 40% of their student body as first generation. However, when this rate decreases, retention increases almost unilaterally.

% of Students Who Had No Parents In College
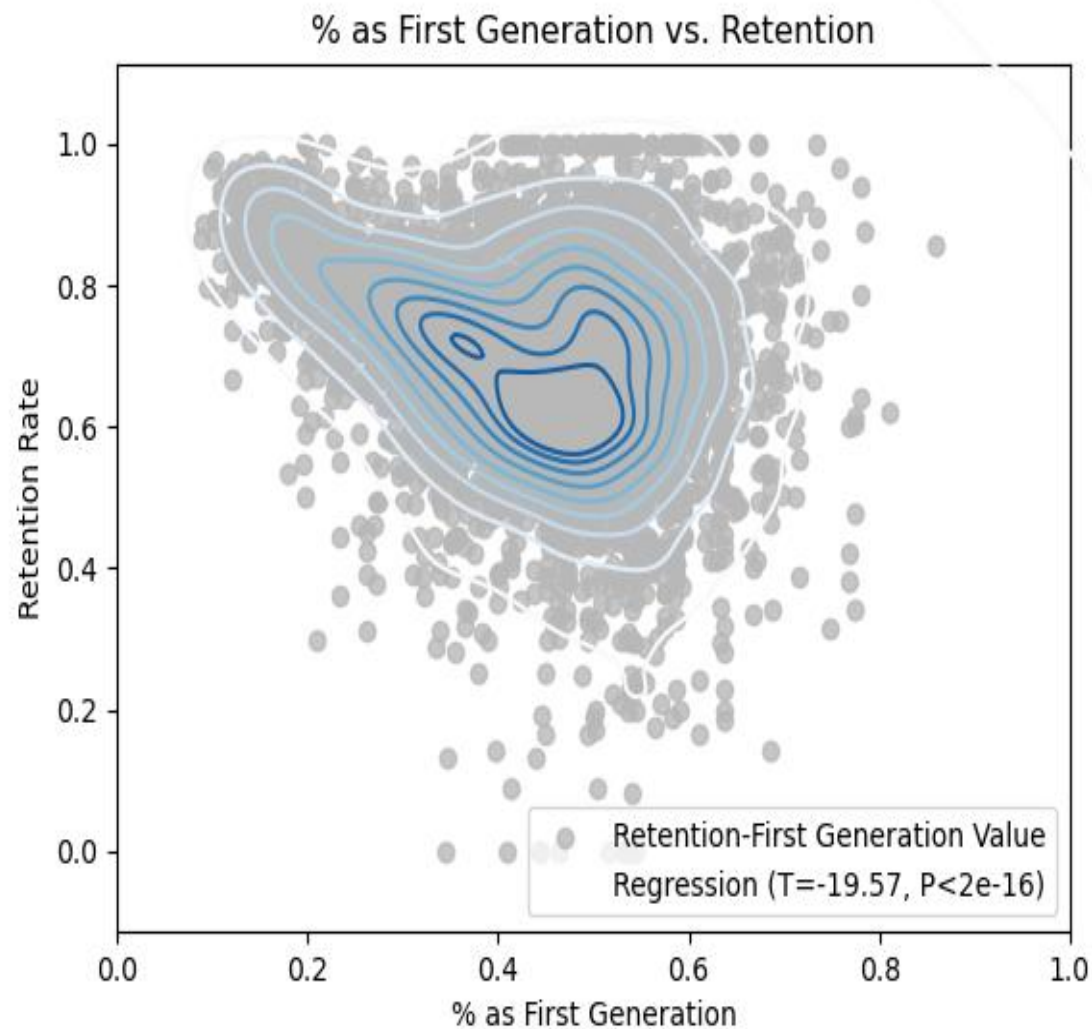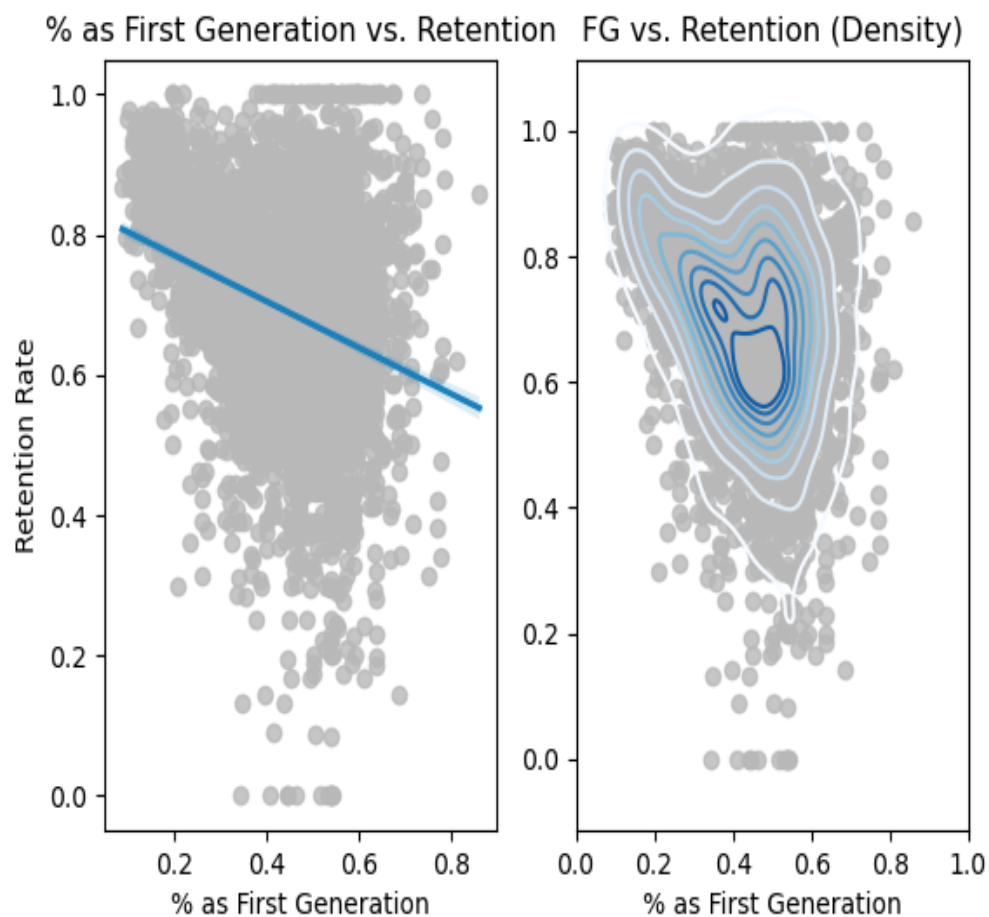
The Accumulation of Factors: Age

- Most people entering universities tend to be in their young adulthood. College Scorecard divides them into two groups, those 24 and below, and those above the age of 25.

- 25 and above was taken because it represents the accumulation of the independent variables taken: possible previous school debt, higher income, more financial burdens, and more distractions in general.

- This graph shows that, while not glaringly obvious, there is general negative correlation. However, it should be taken with a grain of salt because "above 25" could mean anything 25 and above for any university. IE, some universities might just happen to have a lot of 26 year-olds, which is too similar to 24 to mean anything.

Undergraduates Above 25

Above 25 vs. Retention
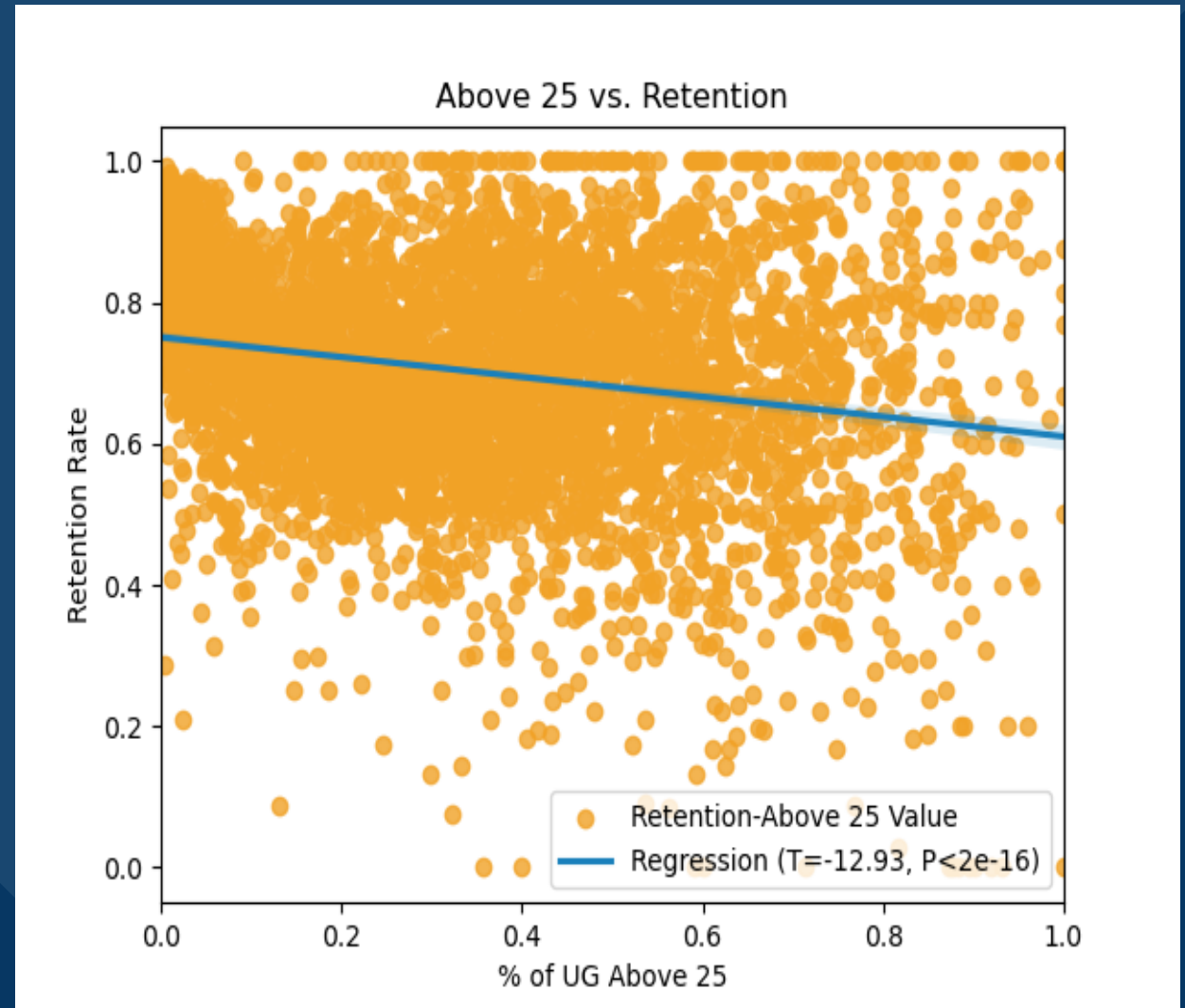
- The density plot more clearly shows A) that most undergraduates are under 25 and B) when the institution has most undergraduates under 25, the retention tends to be higher.

Undergraduates Above 25
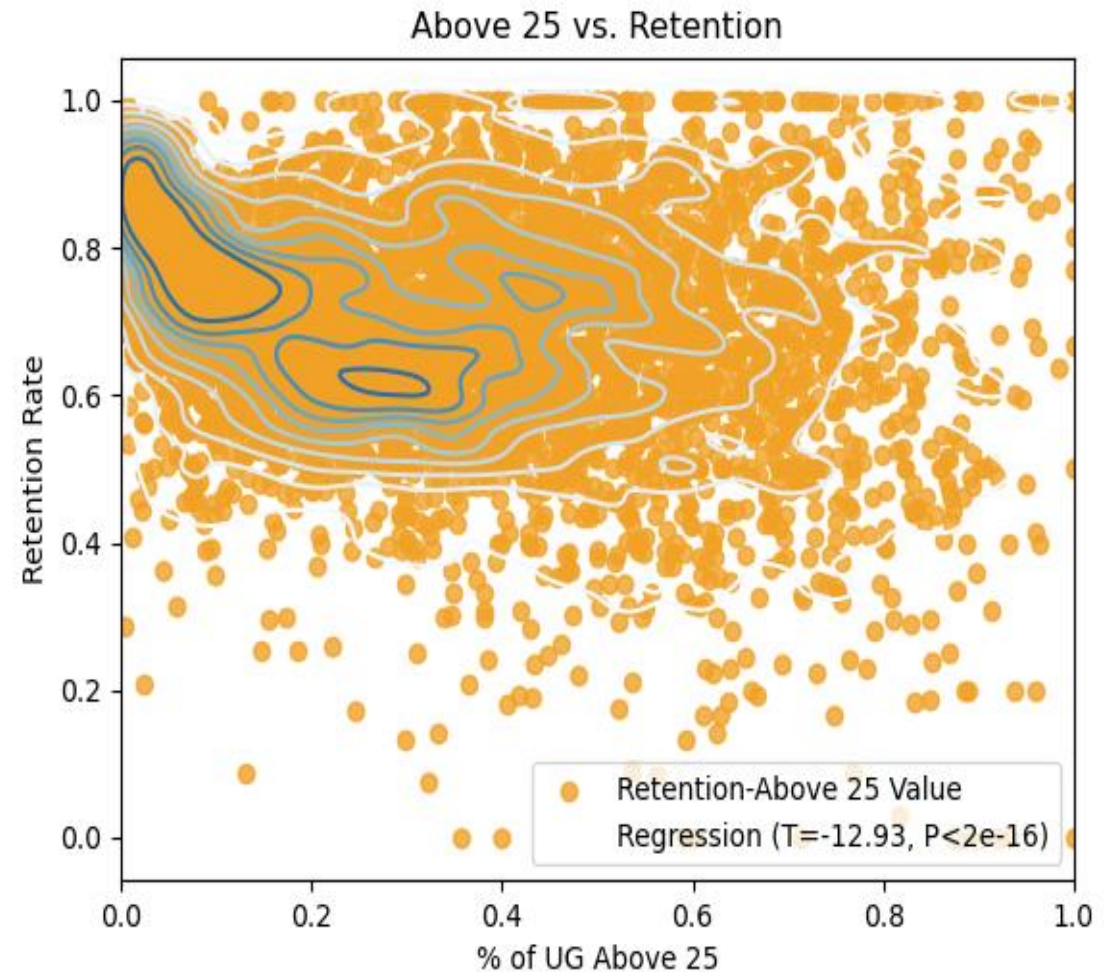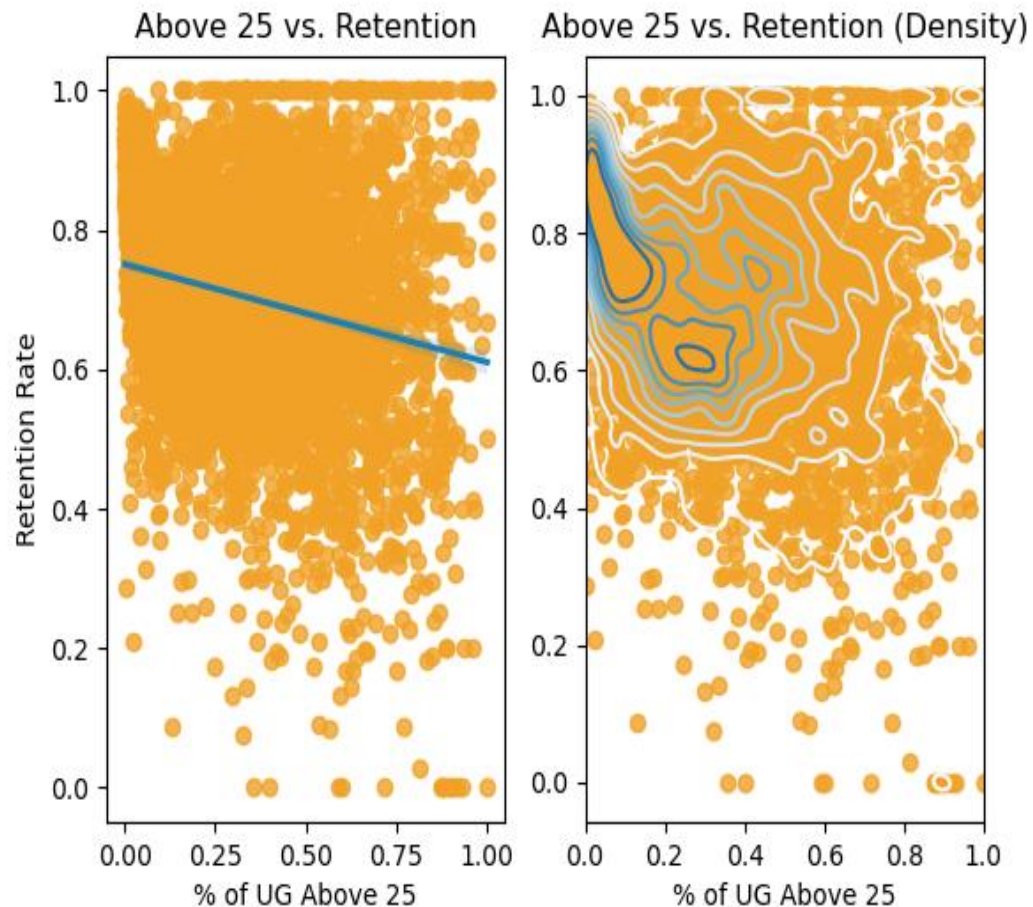
Conclusion: Does this Mean Anything?

What We Know

- GPA influences retention, but may be offset by large variances.

- Initial debt is not a negatively contributing factor, but a student's ability to handle it is (this includes family wealth).

- Those seeking part-time are less likely than those seeking full-time to maintain retention.

- A first generation student is less likely to be retained than a non-first generation student.

- Students who are more likely to have other burdens (those who are older) are more likely to not be retained.

Things to Consider

- The Texas case study was only one school system out of many, and therefore doesn't represent them all.

- There were many NaN's in the College Scorecard record. This includes "Privacy Suppressed" values, which is where the data was excluded.

- Retention was a ratio throughout, which doesn't always correlate well when the other variables are non-ratios that don't mean the same thing.

- "Retention" does not factor in transferring, and neither do many institutions. Often choosing to not even report it. Certain results, like SAT scores and "median debt", were also "estimated by the university", which can be unreliable.

Future Work: Completion

- Completion is defined as whether or not a student has achieved academic goals, within a certain defined percentage of expected time (usually 150%).

- It is directly correlated with retention rate. The higher retention of an institution, the higher the completion rate.

- In fact, it is so correlated that many of the same graphs look extremely similar.

- It'd be interesting to see if this applied to all factors.
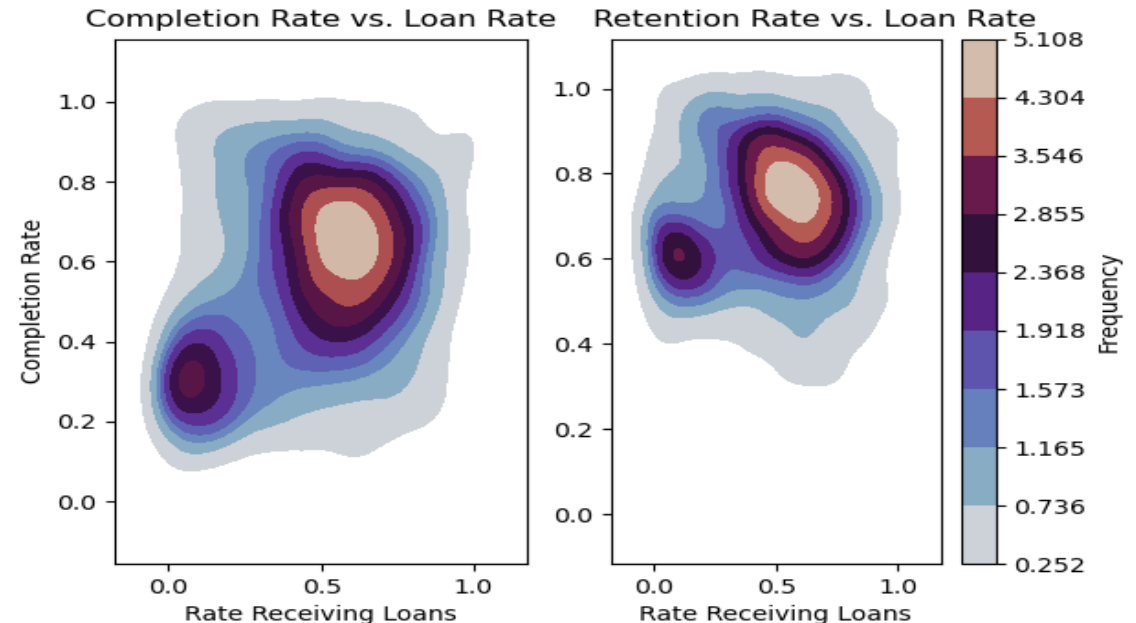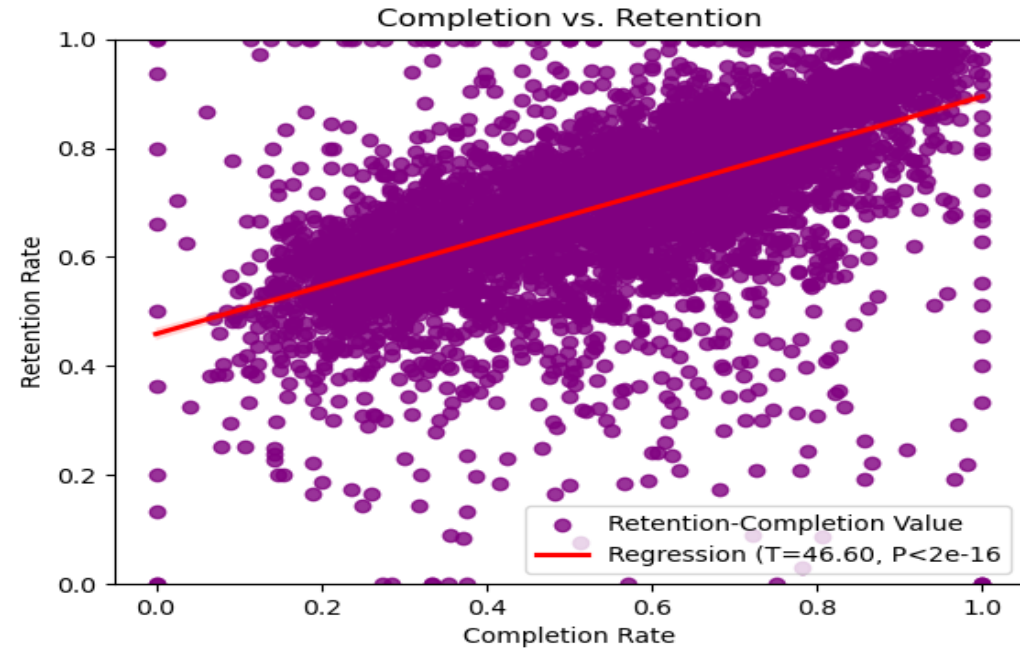
Other Future Work

Demographics

- How much do the various demographics of a university influence retention rate?

- This includes the correlation between staff and student body that share the same demographics.

Machine Learning

- Can we use machine learning to understand future student behavior in regards to retention?

- Can said machine learning identify "at-risk", "high-risk", and "traditional" students?

L4 vs. 4 Year

- Do schools that only have a less-than-4 year program have lower or higher retention rates than those with full four year programs?

Degree Type/Public or Private

- Does the type of degree one is pursuing impact retention rate? What about whether the university is a public or a private one?