CS 4266 Final Project Report

Student Debt in the U.S. Higher Education System

Alan Wilms, Austin Wilms, Able Shi, Kevin Yang

Introduction:

What is the problem we are trying to address?

Applying to colleges can be a difficult and stressful process. Applicants must weigh important aspects for each school including location, cost of attendance, quality of degree programs, facilities, and project income after graduation. With some many schools to choose from, it can be overwhelming for an applicant to consider so many factors when applying to schools. This is a perfect problem for a big data application. Using big data principles, we can collect massive amounts of information on colleges and universities across the United States and conduct queries on that data to provide analyses and produce data models to help prospective students better decide on which schools to apply and enroll. For the purpose of this project, we decided to focus on factors such as student debt and projected income. The goal of this project was to analyze information gathered on thousands of colleges for many years in order to provide recommendations for college applicants based on student debt and projected income after graduation.

Data Collection:

How was the data collected?

Information on U.S. colleges and universities were provided by the U.S. Department of Education College Scorecard. This data was obtained by downloading a directory containing detailed information from the 1996-1997 to 2015-2016 school years.

How is the data composed (i.e. columns, json file, etc)?

Data on the schools were separated into school years and stored in comma-separated values (.csv) files.

What does the data describe?

The college information is organized into ten major categories: root, school, academics, admissions, student, cost, aid, repayment, completion, and earnings. The root category contains school ID and location. The school category includes basic information about a school's description such as the institution name, URL, and whether or not it is a public or private school for example. The academics category describes the metrics for measuring a school's academics such as the degree programs offered. The admissions category covers information about a school's acceptance rate and average standardized test scores. The student category describes the demographics of the student body for a school. The cost category includes information on the costs to students attending a school such as tuition and additional fees. The aid category describes the financial aid given to students. The repayment category details information about the student debt burden of attending a school and the student loan repayment rate for a school. The completion category describes the retention and graduation rates for

a school. Lastly, the earnings category includes information on the earnings and employment prospects of former students at a given school.

For more detailed description about the different information categories, the U.S. Department of Education provides a data dictionary which contains additional information about the type of data collected for each school. Additionally, the data dictionary includes a key in order to interpret the code names for every, detailed information category (i.e. column names in the .csv files).

The entire raw data for our project can be found at Kaggle. It is available for download in .csv format. A respective "data dictionary" (.csv format) with comprehensive descriptions for the data sources, columns, and format can be downloaded at the college scorecard website.

Big Data Problems:
Because the dataset had over 1,700 columns and over 7,000 rows per .csv that each represented a year, it strained the technical feasibility of various SQL systems. In particular, Amazon RDS supports MySQL, PostgreSQL, and Microsoft SQLServer among others and, as such, we expended significant time and energy testing these options.

According to the official specifications MySQL has a "hard limit of 4096 columns per table, but the effective maximum may be less for a given table" [1]. We encountered this effective limit, as MySQL Workbench's built-in Data Import Wizard failed to successfully import from one of the .csv files to a SQL database. To work around this, we made the design compromise to split each table into two parts column-wise (ie, splitting table into two parts based on columns), breaking apart the first 1000 columns. However, this now resulted in an error due to too many rows (resulting in a number of rows that is infeasible and also much greater than imported number of rows). After some investigation, our conclusion was that this was perhaps caused by the underlying B-tree page representation of MySQL's storage system [2]. If the columns are too long (too many bytes), behind the scenes, MySQL creates overflow pages and invisibly links the two pages. But perhaps this is triggering an unexpected increase in the number of rows. Row format compression and changing the underlying storage engine [3] were also unsuccessfully attempted as workarounds.

After failed attempts using MySQL, we decided to try using SQL Server next. The SQL server standard has a 1024 column limit for a nonwide table and a 30000 column limit for a wide table. The number of rows for a table is dependent on the amount of storage that is available. While SQL Server was much more efficient for data import execution, we ran into similar problems of unwanted data/column mutation behavior like before. When we attempted to create a wide table in SQL Server, it created an embedded XML for columns identified as "spares". While this behavior is geared specifically toward tables with many columns that have simply just "NULL" values, this significantly alters the visualization and format of our desired tables. In addition, these "NULL" columns weren't

constant and varied from table to table in the entirety of our dataset. This consequently makes querying much more difficult and in the end was a tradeoff we were reluctant to take on. We decided to attempt using scripts again, but unfortunately, the method didn't yield much fruition. We were able to split each table in our dataset into two separate "a" & "b" tables but we ran into an unexpected issue where the number of columns exceeded the SQL server limit of 8000. This came to us as a surprise because it was reported that the number of columns we had was ~11000 when we expected we had only 7850. We theorized that the issue stemmed from not necessarily the number of rows but rather the size of each row. As such, we were left with limited options as the schema was automatically defined by the import wizard. There was an additional option to minimize the data type of each column but we ran into a new obstacle about the truncation of certain columns. Despite our efforts after this point, we were not able to arrive at a successful alternative solution.

Our final solution to these problems was manually analyzing the data dictionary and only keeping the columns relevant to our querying. From there, the default SQL Server type identification unfortunately needed to be discarded because it erred toward assigning a TEXT label to each attribute. This meant that ordering by years, percentages, or any other mathematical value proved to be largely impossible, especially for floating point values. Furthermore, the occasional "PrivacySuppressed" attribute value needed to be replaced with NULL because it would other break the column type uniformity.

As a result, this required a large sequence of manual create table script executions. By consulting the data dictionary, columns of interest were identified. Then a Python script was utilized to parse the .csv files that contained the College Scorecard data and "PrivacySuppressed" values were replaced with NULL values in this way. These modified and condensed data tables were then uploaded to SQL Server successfully.

Methodology:

The relevant attributes for the questions that we wanted answered were uploaded to a Microsoft SQL Server instance hosted on Amazon RDS. This server was made available to all team members at the following address and with the following credentials:

Public Endpoint:	collegescorecard.cuvwexwhqjuk.us-east-2.rds.amazonaws.com
Username:	adam
Password:	*adamadam*

Analysis on the data uploaded to SQL Server was conducted using Microsoft SQL Server Management Studio on Windows and SQL Pro for MSSQL on Mac.

Prestige was calculated as a summed rank of the admission rate rank (where the most selective university would receive a rank of 1) and the better of the SAT and ACT test score rank (where the highest test score would receive a rank of 1). For example, Harvard University has an admit rate rank of 2, and ACT score rank of 2, and SAT score rank of 7, giving it a prestige score of 4, which happens to be the best score and deserving a prestige rank of 1. Prestige and Prestige Rank, which was particularly helpful for joins with other tables, was created as a view for easy access.

Findings:

What are the schools having students with less debt (<$10000)?

Institution Name	State	Median Graduate Debt	Least Debt Ranking	Expected Earnings After 10 Years
New Age Training	NY	1270	1	NULL
South Florida Institute of Technology	FL	1441	2	15500
Automeca Technical College-Aguadilla	PR	1750	3	18300
Automeca Technical College-Bayamon	PR	1750	3	18300
Automeca Technical College-Caguas	PR	1750	3	18300
Automeca Technical College-Ponce	PR	1750	3	18300
Coachella Valley Beauty College	CA	1914	7	NULL
Alliant International University-Fresno	CA	2000	8	58000
Alliant International University-Los Angeles	CA	2000	8	58000
Alliant International University-San Francisco	CA	2000	8	58000
Alliant International University-Irvine	CA	2000	8	58000
Alliant International University-Sacramento	CA	2000	8	58000
West Virginia School of Osteopathic Medicine	WV	2000	8	220700
Alliant International University-San Diego	CA	2000	8	58000
Ponce Paramedical College Inc	PR	2260	15	15500
Centro de Estudios Multidisciplinarios-San Juan	PR	2334	16	18500
Centro de Estudios Multidisciplinarios-Humacao	PR	2334	16	18500
Laredo Community College	TX	2334	16	29200
Centro de Estudios Multidisciplinarios-Bayamon	PR	2334	16	18500
San Bernardino Valley College	CA	2400	20	27000

Table 1: Prestige Rank, Median Graduate Debt, and Expected Earnings After 10 Years for 2013-2014
Generated by [E]

This data is for the year 2013 - 2014 because there is no future expected income information yet for the year 2015 - 2016. There is a general trend of community colleges and specialized institutions, such as beauty, design, and medical schools, that result in less graduate student debt. Furthermore, these institutions typically exist in lower cost-of-living areas such as Puerto Rico and southern states like Georgia, Arkansas, and Florida. Looking at the median earnings of students working and not enrolled 10 years after entry for the year 2013-2014, we found that, with the exception of the medical and law schools, the graduates typically earn less than $20,000. For example, according to Table 1, the South Florida Institute of Technology has median earnings of students working and not enrolled 10 years after entry of $15,500 in 2013-2014. This is less than half of the annual real median personal income of $31,099 in the United States in 2016 [4]. Furthermore, these schools have relatively low Prestige rank, which speaks to a potentially much lower quality of education provided by schools whose redeeming quality is lowering student debt.

What are the schools having students with higher debt (>$10000)?

Institution Name	State	Median Graduate Debt	Most Debt Ranking	Expected Earnings After 10 Years
Stenotype Institute of Jacksonville Inc-Jacksonville	FL	53831.5	1	28700
Southwest University of Visual Arts-Tucson	AZ	50624	2	28500
Southwest University of Visual Arts-Albuquerque	NM	50624	2	28500
Mt Sierra College	CA	45250	4	39500
Northwest College of Art & Design	WA	45000	5	28300
Newschool of Architecture and Design	CA	44332.5	6	NULL
Boston Architectural College	MA	44225	7	46400
Benedict College	SC	43375	8	22100
Florida Memorial University	FL	43250	9	25900
Platt College-Aurora	CO	42369.5	10	26500
Beulah Heights University	GA	42300.5	11	27700
Everglades University	FL	42200	12	41300
Everglades University-Orlando	FL	42200	12	41300
Everglades University-Sarasota	FL	42200	12	41300
Digital Media Arts College	FL	41883.5	15	NULL
International Academy of Design and Technology-Troy	MI	41472	16	26100
International Academy of Design and Technology-Nashville	TN	41472	16	26100
Sanford-Brown College-Chicago	IL	41472	16	26100
Collins College	AZ	41472	16	26100
Victory University	TN	40250	20	37700

Table 2: Prestige Rank, Median Graduate Debt, and Expected Earnings After 10 Years for 2013-2014 Generated by [E] after changing ASC to DESC

Many for-profit schools, such as DeVry University, and again specialized schools are among the schools with the highest median graduate debt. However, despite a median graduate debt that can reach upward of 20 times that of the schools with the lowest median graduate debt, the expected income for graduates still hardly reaches the $31,099 real median personal income level in the United States. As such, these school are unequivocally a bad financial investment for every type of student.

How does the median graduate debt compare to the prestige of the school?

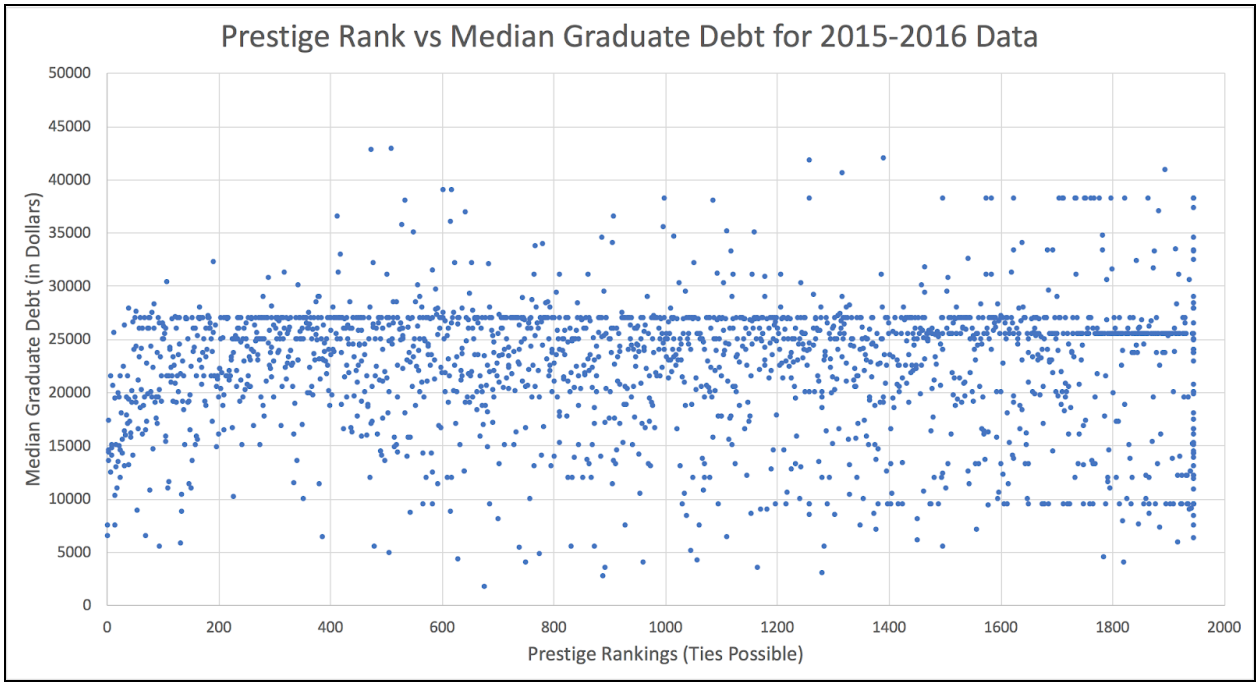

Figure 1: Scatterplot of Median Graduate Debt based on School Prestige for the year 2015-2016 Generated by [A]

Figure 1 displays the median graduate debt when compared to a school's prestige for the year 2015-2016. At first glance, there appears to be no clear correlation between a school's rank in prestige and its graduate debt; at the very least, it appears that the top 100 schools are below the $26,000 median graduate debt. However, organizing the average median graduate debt across different prestige ranking strata, reveals more details. According to Table 3, the universities ranked in the top 25 results in $7,562.53 less than the overall average median graduate debt and those ranked in the next 25 result in ~$3,461.07 less.

Prestige Rank	**Average Median Graduate Debt**
1-25	$14981.42
26-50	$19082.88
51-75	$20799.17
76-100	$21279.20
101-125	$22092.88
126-150	$20213.68
151-176	$22507.38
176-200	$23315.88
200-226	$22431.94
226-250	$23175.80
251-276	$23737.40
276-300	$24491.94
...	...

Average: $22543.95

Table 3: Average Median Debt Separated by Prestige

This means the up to a certain point, which appears to be the top 50, the prestige rank of the university is directly correlated with a lower average median graduate debt, so students who are seeking to minimize student loans should first seek to enroll in these top tier schools.

What are the schools in which students, based on their future income, will endure debt for a longer time (10+ years)?

Institution Name	Percentage of Graduates Employed	Median Graduate Debt	Median Student Earnings (After 10 Years)	Number of Years to Payoff
Gallipolis Career College	0.585	29937.5	13400	22.341
Benedict College	0.849	43375	22100	19.627
Stenotype Institute of Jacksonville Inc-Jacksonville	0.784	53831.5	28700	18.757
Livingstone College	0.852	39500	21400	18.458
Southwest university of Visual Arts-Tucson	0.783	50624	28500	17.762
Southwest University of Visual Arts-Albuquerque	0.783	50624	28500	17.762
Allen University	0.835	37525	21800	17.213
Florida Memorial University	0.865	43250	25900	16.699
Mountain State College	0.597	25337.5	15200	16.699
The School of Hairstyling	0.708	15000	9100	16.484
Arkansas Baptist College	0.788	36500	22300	16.368
Collins College	0.805	41472	26100	15.889
Sanford-Brown College-Chicago	0.805	41472	26100	15.889
International Academy of Design and Technology-Troy	0.805	41472	26100	15.889

International Academy of Design and Technology-Nashville	0.805	41472	26100	15.889
Central State University	0.853	38176	24400	15.645
Martin University	0.692	35868.5	23000	15.595
Beulah Heights University	0.749	42300.5	27700	15.271
Wright Career College	0.729	25000	17000	14.706
Wright Career College -OKC	0.729	25000	17000	14.706

Table 4: Average Length of Time for Debt Payoff (Years) For 2013_2014

The data in Table 4 lists the 20 college institutions that, on average, takes their respective graduates the longest to pay off their student debt compared to other college institutions for 2013- 2014. This table was generated by the query in Appendix [B]. We decided to use just the latest available annual data to tackle this question because there wasn't sufficient data in the general dataset; median graduate debt and median student earnings were only available in the years 2007-2008, 2009-2010, 2011-2012, 2012-2013 and 2013-2014. Due to this less than desired scarcity of necessary data, we thought it would be most accurate to just take the dollar values for debt and student earnings from the most recent available year since it would best reflect the current college economy. To calculate how long students from a particular college will take to pay off and break even with their student debt, we took data on a college's median graduate debt and its median student earnings after 10 years of graduation under the assumption they are not attending another educational institution. To us, both pieces of information are crucial in determining how long a student is actually still in debt after graduating from a certain college. This is because although some organizations have higher median graduates than other institutions, students from these organizations might have on average higher earnings than other schools. Another assumption we made was that graduates annually will deposit at least 10% of their annual earnings toward their debt repayment plan.

From Table 4, we can see that the institutions where students are in debt for the longest periods of time are generally schools that have high debt relative to their future student earnings. These schools typically appear to be religious-affiliated or art and design schools.

How does the debt of students attending schools with a tradition of minorities (non-white) compare to Ivy League schools?

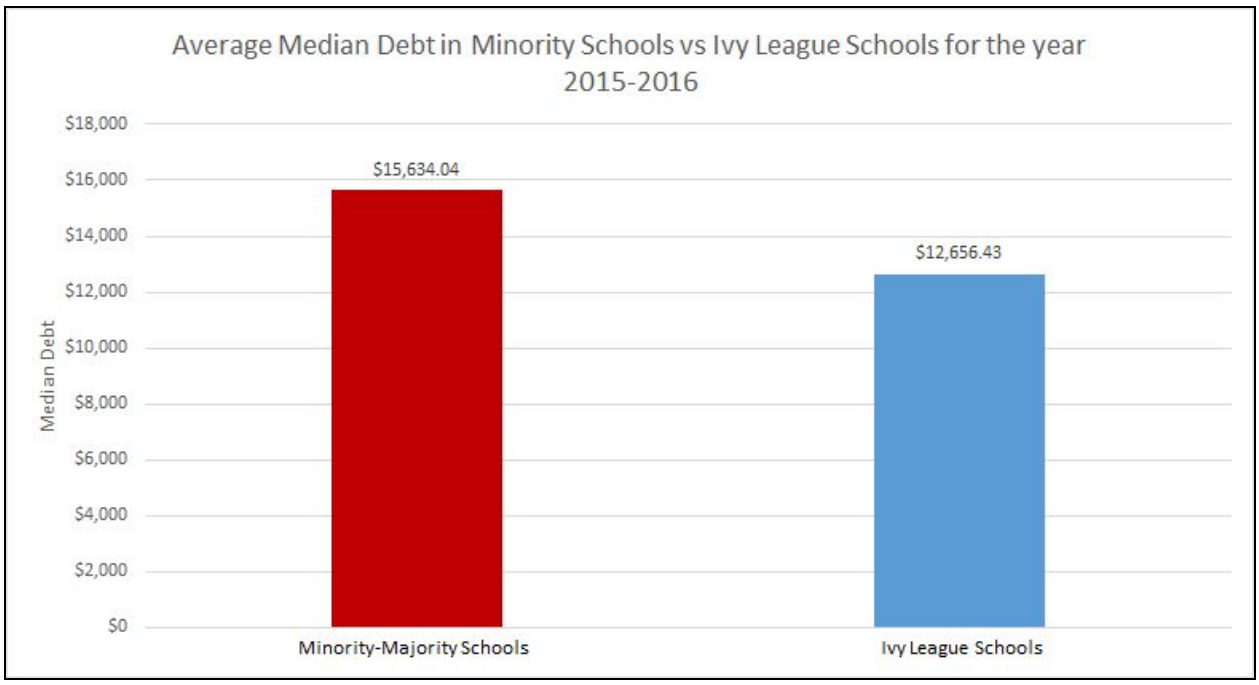

Figure 2: Comparison of Average Median Debt between minority-majority schools and Ivy League schools

For the purpose of this project, schools with a tradition of minorities or "minority-majority" schools were defined as schools where the caucasian students made up less than 50% of the total student body (Appendix [F]). Similarly, Ivy League schools consist of eight schools: Brown University, Columbia University, Cornell University, Dartmouth College, Harvard University, University of Pennsylvania, Princeton University, and Yale University (Appendix [G]). Figure 2 depicts the comparison of the average median debt of minority-majority schools and Ivy League schools for the most recent year of which data was available, 2015-2016. Based on our analysis, the average median debt after graduation is actually higher for minority-majority schools at $15,634.04 to that of Ivy League school graduates at $12,656.43. This would suggest that Ivy League schools have more robust financial aid packages for their students, allowing their students to attend without accruing massive amounts of student debt.

Is there a correlation between STEM degrees and income?

Figure 3 depicts the scatterplot comparing the median student earnings to the total percentage of STEM degrees offered at an institution. The graph also includes trendline to show a slight positive trend in median earnings as the percentage of STEM degrees increases, but as noted, the regression value is 0.2317 which would suggest a moderate correlation. The data was obtained from the school year 2013-2014 which was the latest year to report student earnings. Total percentage of STEM degrees was calculated by taking a sum of the percentages of all the different STEM majors offered by a

given institution (Appendix [H]). Therefore, we can say, that yes there is a small but present correlation between STEM degrees and income.

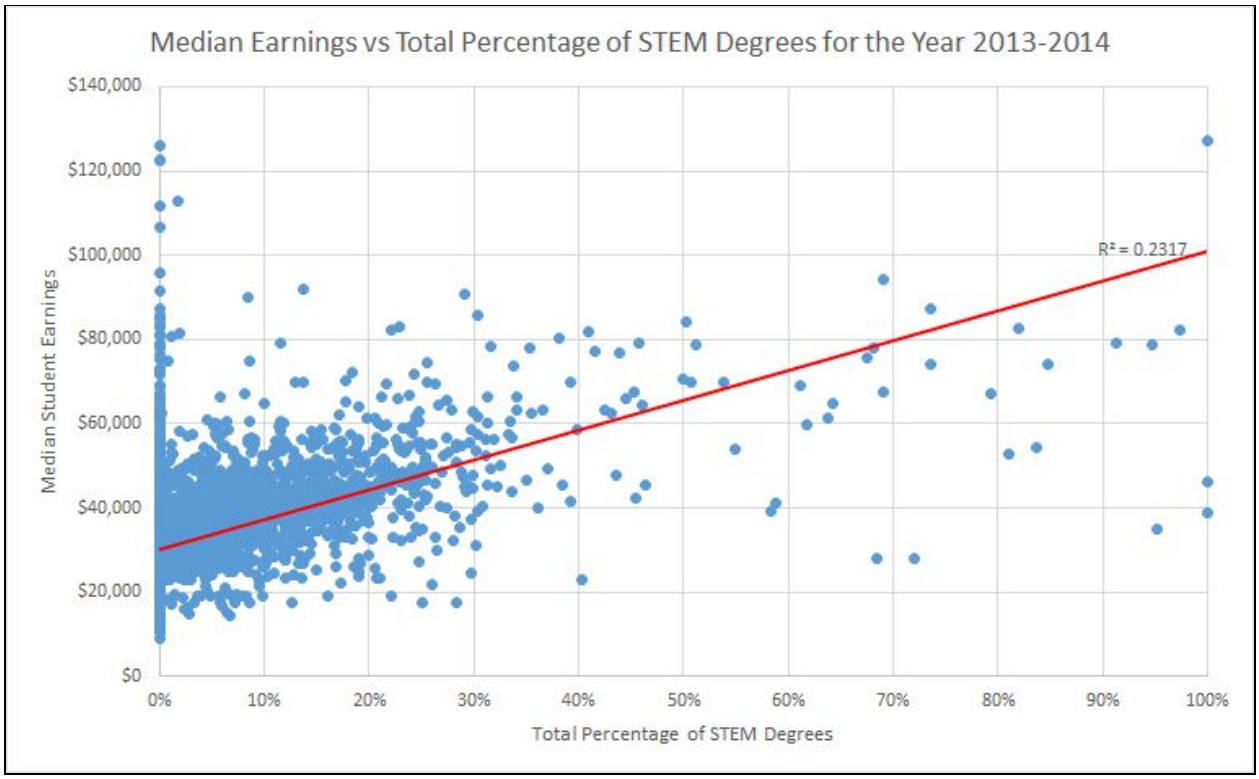

Figure 3: Comparison of Median Student Earnings to Total Percentage of STEM Degrees for the year 2013-2014

Debt over 20 years

Taking the median debt data from year academic year (with 1996-1997 excluded due to lack of data), a line graph (Figure 4) can be created to get a big-picture view on the state of student debts. This was created using the Appendix [C] query to first create a view of the median debts of each school for each year, and then calculating the average of each school from each year. As shown in Figure 4, debt has increased steadily over the last 20 years, which may be attributed to general inflation and increases in cost of living in the U.S. This highlights the importance enrolling in top tier schools as well as applying to high paying jobs after college to remedy this debt.

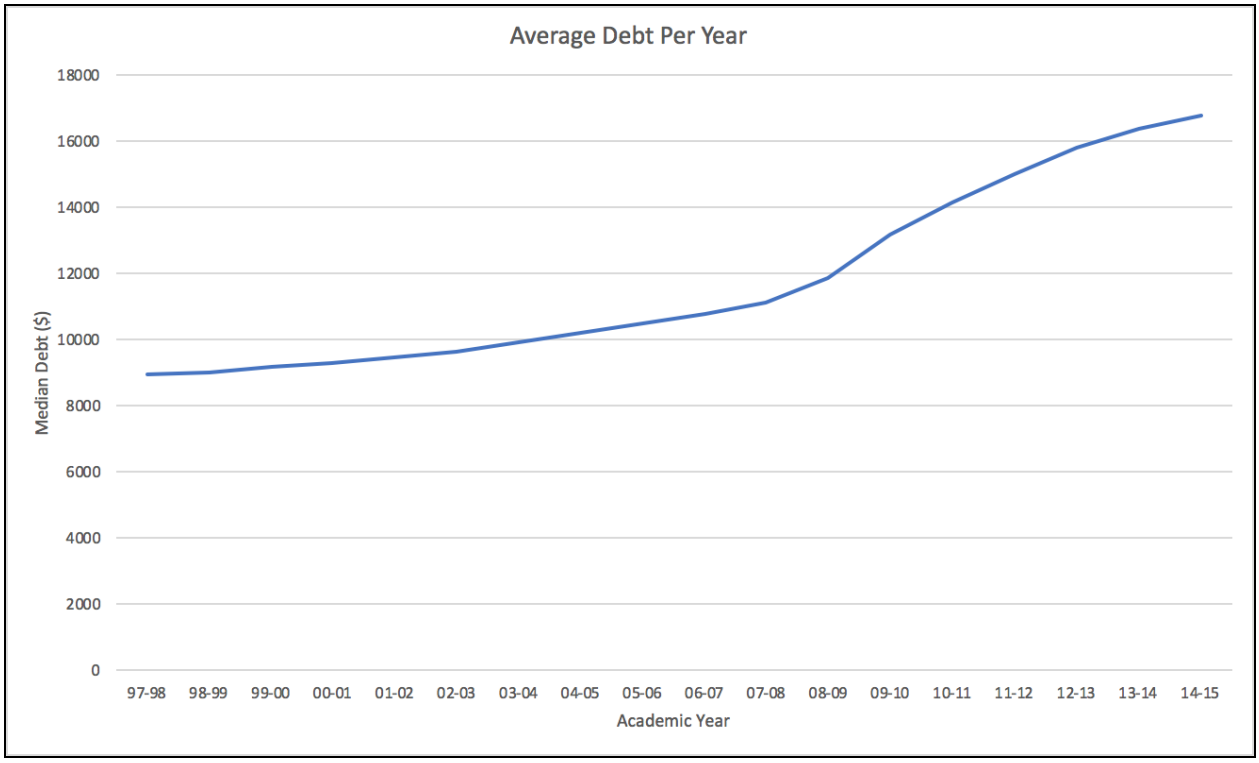

Figure 4: Median Debt Averaged Per Year

Conclusion:

What do our findings tell us? What are our recommendations?

As median student debt has steadily grown over the last 20 years, it has become increasingly important to be selective for aspects such as debt and earnings when applying to post-secondary education. Our findings suggest that a student enrolled in a prestigious school (top 50) would actually incur less student loan debt, where prestige includes selectivity rank and the test score rank of the school. The schools with the highest median graduate debt only offer less than a $15,000 increase in the expected earnings of graduates 10 years after enrollment, so student debt incurred *does not* generally correspond to a better education, as measured by this future income. Furthermore, the prestige of the school offers a better, more direct relationship with future income, so even the schools with the lowest median graduate debt may not be the best financial decision for student, as it will still take much longer for the recent graduate to pay off the debt.

When comparing student debt in schools with a tradition of minorities to those in Ivy League schools, we found that on average, in the most recent data available, students in Ivy League schools graduate with less student debt. This correlates with our findings on the relationship between school prestige and student debt, suggesting that students that attend higher ranked institutions will take on less average debt. Another observation revealed a moderate correlation between the percentage of STEM majors offered at an institution and the median student earnings.

In conclusion, religious-affiliated, art and design, and for-profit institutions should be avoided for their high debt and low future income for students, and the most prestigious schools and schools with a high percentage of STEM degrees offered should be particularly pursued for their financial soundness.

References:

[1]: https://dev.mysql.com/doc/refman/5.6/en/column-count-limit.html
[2]: https://dev.mysql.com/doc/refman/5.7/en/innodb-row-format-overview.html
[3]: https://dev.mysql.com/doc/refman/5.7/en/innodb-row-format-dynamic.html
[4]: https://fred.stlouisfed.org/series/MEPAINUSA672N

Appendix:

[A]:

```
CREATE VIEW [Prestige] AS
SELECT INSTNM AS 'Institution Name', GRAD_DEBT_MDN AS 'Median Graduate Debt',
RANK() OVER (ORDER BY GRAD_DEBT_MDN ASC) AS 'Median Graduate Debt Rank',
RANK() OVER (ORDER BY ADM_RATE ASC) AS 'Admit Rate Rank', RANK() OVER
(ORDER BY ACTCMMID DESC) AS 'ACT Score Rank', RANK() OVER (ORDER BY
SAT_AVG_ALL DESC) AS 'SAT Score Rank',
'Prestige' = CASE
        WHEN RANK() OVER (ORDER BY ACTCMMID DESC) >= RANK() OVER (ORDER
BY SAT_AVG_ALL DESC) THEN RANK() OVER (ORDER BY SAT_AVG_ALL DESC) +
RANK() OVER (ORDER BY ADM_RATE ASC)     ELSE RANK() OVER (ORDER BY
ACTCMMID DESC) + RANK() OVER (ORDER BY ADM_RATE ASC) END
FROM [dbo].[2015_16]
WHERE GRAD_DEBT_MDN IS NOT NULL AND ADM_RATE IS NOT NULL AND
ADM_RATE <> 0
ORDER BY 'Prestige' ASC;
```

*This query exempts the years 1996_97 to 2000_01 (inclusive) due to a lack of admission rate and/or test score data.

```
SELECT [Institution Name], RANK() OVER (ORDER BY [Prestige] ASC) AS 'Prestige Rank'
FROM dbo.Prestige;
```

[B]:

```
SELECT INSTNM AS 'Institution Name',
(COUNT_WNE_P10/(COUNT_NWNE_P10+COUNT_WNE_P10)) AS 'Percentage of Grads
Employed', GRAD_DEBT_MDN AS 'Median Graduate Debt', MD_EARN_WNE_P10 AS
'Median Student Earnings', GRAD_DEBT_MDN/(0.1 * MD_EARN_WNE_P10) AS 'Number of
Years to Payoff'
FROM [dbo].[2013_14]
```

```
WHERE GRAD_DEBT_MDN IS NOT NULL AND MD_EARN_WNE_P10 IS NOT NULL
AND COUNT_NWNE_P10 IS NOT NULL
ORDER BY 'Number of Years to Payoff' DESC;
```

[C]:

```
 CREATE VIEW [Average Debt] AS
SELECT t1.INSTNM AS 'Name', t1.GRAD_DEBT_MDN AS '2015-2016 Median Grad Debt',
t2.GRAD_DEBT_MDN AS '2014-2015 Median Grad Debt', t3.GRAD_DEBT_MDN AS
'2013-2014 Media Grad Debt', t4.GRAD_DEBT_MDN AS '2012-2013 Median Grad Debt',
t5.GRAD_DEBT_MDN AS '2011-2012 Median Grad Debt', t6.GRAD_DEBT_MDN AS
'2010-2011 Median Grad Debt', t7.GRAD_DEBT_MDN AS '2009-2010 Median Grad Debt',
t8.GRAD_DEBT_MDN AS '2008-2009 Median Grad Debt', t9.GRAD_DEBT_MDN AS
'2007-2008 Median Grad Debt', t10.GRAD_DEBT_MDN AS '2006-2007 Median Grad Debt',
t11.GRAD_DEBT_MDN AS '2005-2006 Median Grad Debt', t12.GRAD_DEBT_MDN AS
'2004-2005 Median Grad Debt', t13.GRAD_DEBT_MDN AS '2003-2004 Median Grad Debt',
t14.GRAD_DEBT_MDN AS '2002-2003 Median Grad Debt', t15.GRAD_DEBT_MDN AS
'2001-2002 Median Grad Debt', t16.GRAD_DEBT_MDN AS '2000-2001 Median Grad Debt',
t17.GRAD_DEBT_MDN AS '1999-2000 Median Grad Debt', t18.GRAD_DEBT_MDN AS
'1998-1999 Median Grad Debt', t19.GRAD_DEBT_MDN AS '1997-1998 Median Grad Debt',
t20.GRAD_DEBT_MDN AS '1996-1997 Median Grad Debt'
FROM [dbo].[2015_16] t1, [dbo].[2014_15] t2, [dbo].[2013_14] t3, [dbo].[2012_13] t4,
[dbo].[2011_12] t5, [dbo].[2010_11] t6, [dbo].[2009_10] t7, [dbo].[2008_09] t8, [dbo].[2007_08]
t9, [dbo].[2006_07] t10, [dbo].[2005_06] t11, [dbo].[2004_05] t12, [dbo].[2003_04] t13,
[dbo].[2002_03] t14, [dbo].[2001_02] t15, [dbo].[2000_01] t16, [dbo].[1999_00] t17,
[dbo].[1998_99] t18, [dbo].[1997_98] t19, [dbo].[1996_97] t20
WHERE t1.UNITID = t2.UNITID AND t2.UNITID = t3.UNITID AND t3.UNITID =
t4.UNITID AND t4.UNITID = t5.UNITID AND t5.UNITID = t6.UNITID AND t6.UNITID =
t7.UNITID AND t7.UNITID = t8.UNITID AND t8.UNITID = t9.UNITID AND t10.UNITID
= t11.UNITID AND t11.UNITID = t12.UNITID AND t12.UNITID = t13.UNITID AND
t13.UNITID = t14.UNITID AND t14.UNITID = t15.UNITID AND t15.UNITID =
t16.UNITID AND t16.UNITID = t17.UNITID AND t17.UNITID = t18.UNITID AND
t18.UNITID = t19.UNITID AND t19.UNITID = t20.UNITID;
```

```
SELECT AVG([2015-2016 Median Grad Debt]) AS '15-16 Median', AVG([2014-2015 Median Grad
Debt]) AS '14-15 Median', AVG([2013-2014 Media Grad Debt]) AS '13-14 Median',
AVG([2012-2013 Median Grad Debt]) AS '12-13 Median', AVG([2011-2012 Median Grad Debt])
AS '11-12 Median', AVG([2010-2011 Median Grad Debt]) AS '10-11 Median', AVG([2009-2010
Median Grad Debt]) AS '09-10 Median', AVG([2008-2009 Median Grad Debt]) AS '08-09 Median',
AVG([2007-2008 Median Grad Debt]) AS '07-08 Median', AVG([2006-2007 Median Grad Debt])
```

```
AS '06-07 Median', AVG([2005-2006 Median Grad Debt]) AS '05-06 Median', AVG([2004-2005
Median Grad Debt]) AS '04-05 Median', AVG([2003-2004 Median Grad Debt]) AS '03-04 Median',
AVG([2002-2003 Median Grad Debt]) AS '02-03 Median', AVG([2001-2002 Median Grad Debt])
AS '01-02 Median', AVG([2000-2001 Median Grad Debt]) AS '00-01 Median', AVG([1999-2000
Median Grad Debt]) AS '99-00 Median', AVG([1998-1999 Median Grad Debt]) AS '98-99 Median',
AVG([1997-1998 Median Grad Debt]) AS '97-98 Median', AVG([1996-1997 Median Grad Debt])
AS '96-97 Median'
FROM dbo.[Average Debt]
```

[D]:

```
SELECT '1-25' AS 'Rank', AVG(t2.GRAD_DEBT_MDN) AS 'Average Median Graduate Debt'
FROM dbo.[Prestige Rank] t1, dbo.[2015_16] t2
WHERE t1.UNITID = t2.UNITID AND (t1.[Prestige Rank] >= 1 AND t1.[Prestige Rank] <= 25)
UNION ALL
SELECT '26-50' AS 'Rank', AVG(t2.GRAD_DEBT_MDN) AS 'Average Median Graduate Debt'
FROM dbo.[Prestige Rank] t1, dbo.[2015_16] t2
WHERE t1.UNITID = t2.UNITID AND (t1.[Prestige Rank] >= 26 AND t1.[Prestige Rank] <= 50)
UNION ALL
SELECT '51-75' AS 'Rank', AVG(t2.GRAD_DEBT_MDN) AS 'Average Median Graduate Debt'
FROM dbo.[Prestige Rank] t1, dbo.[2015_16] t2
WHERE t1.UNITID = t2.UNITID AND (t1.[Prestige Rank] >= 51 AND t1.[Prestige Rank] <= 75)
UNION ALL
SELECT '76-100' AS 'Rank', AVG(t2.GRAD_DEBT_MDN) AS 'Average Median Graduate Debt'
FROM dbo.[Prestige Rank] t1, dbo.[2015_16] t2
WHERE t1.UNITID = t2.UNITID AND (t1.[Prestige Rank] >= 76 AND t1.[Prestige Rank] <=
100)
UNION ALL
SELECT '101-125' AS 'Rank', AVG(t2.GRAD_DEBT_MDN) AS 'Average Median Graduate Debt'
FROM dbo.[Prestige Rank] t1, dbo.[2015_16] t2
WHERE t1.UNITID = t2.UNITID AND (t1.[Prestige Rank] >= 101 AND t1.[Prestige Rank] <=
125)
UNION ALL
SELECT '126-150' AS 'Rank', AVG(t2.GRAD_DEBT_MDN) AS 'Average Median Graduate Debt'
FROM dbo.[Prestige Rank] t1, dbo.[2015_16] t2
WHERE t1.UNITID = t2.UNITID AND (t1.[Prestige Rank] >= 126 AND t1.[Prestige Rank] <=
150)
UNION ALL
SELECT '151-176' AS 'Rank', AVG(t2.GRAD_DEBT_MDN) AS 'Average Median Graduate Debt'
```

```
FROM dbo.[Prestige Rank] t1, dbo.[2015_16] t2
WHERE t1.UNITID = t2.UNITID AND (t1.[Prestige Rank] >= 151 AND t1.[Prestige Rank] <=
176)
UNION ALL
SELECT '176-200' AS 'Rank', AVG(t2.GRAD_DEBT_MDN) AS 'Average Median Graduate Debt'
FROM dbo.[Prestige Rank] t1, dbo.[2015_16] t2
WHERE t1.UNITID = t2.UNITID AND (t1.[Prestige Rank] >= 176 AND t1.[Prestige Rank] <=
200)
UNION ALL
SELECT '200-226' AS 'Rank', AVG(t2.GRAD_DEBT_MDN) AS 'Average Median Graduate Debt'
FROM dbo.[Prestige Rank] t1, dbo.[2015_16] t2
WHERE t1.UNITID = t2.UNITID AND (t1.[Prestige Rank] >= 200 AND t1.[Prestige Rank] <=
226)
UNION ALL
SELECT '226-250' AS 'Rank', AVG(t2.GRAD_DEBT_MDN) AS 'Average Median Graduate Debt'
FROM dbo.[Prestige Rank] t1, dbo.[2015_16] t2
WHERE t1.UNITID = t2.UNITID AND (t1.[Prestige Rank] >= 226 AND t1.[Prestige Rank] <=
250)
UNION ALL
SELECT '251-276' AS 'Rank', AVG(t2.GRAD_DEBT_MDN) AS 'Average Median Graduate Debt'
FROM dbo.[Prestige Rank] t1, dbo.[2015_16] t2
WHERE t1.UNITID = t2.UNITID AND (t1.[Prestige Rank] >= 251 AND t1.[Prestige Rank] <=
275)
UNION ALL
SELECT '276-300' AS 'Rank', AVG(t2.GRAD_DEBT_MDN) AS 'Average Median Graduate Debt'
FROM dbo.[Prestige Rank] t1, dbo.[2015_16] t2
WHERE t1.UNITID = t2.UNITID AND (t1.[Prestige Rank] >= 276 AND t1.[Prestige Rank] <=
300)
```

[E]:

```
SELECT INSTNM AS 'Institution Name', STABBR AS 'State', t1.[Prestige Rank],
GRAD_DEBT_MDN AS 'Median Graduate Debt', RANK() OVER (ORDER BY
GRAD_DEBT_MDN ASC) AS 'Median Graduate Debt Rank', MD_EARN_WNE_P10 AS 'Future
Income'
FROM [dbo].[2013_14] t2, dbo.[Prestige Rank] t1
WHERE GRAD_DEBT_MDN IS NOT NULL AND t1.UNITID = t2.UNITID;
```

[F]:

```
SELECT INSTNM AS 'Institution Name', UGDS_WHITE AS 'Percentage  of White Students',
GRAD_DEBT_MDN AS 'Median Graduate Debt'
FROM "2015_16"
WHERE UGDS_WHITE IS NOT NULL AND UGDS_WHITE < 0.5 AND
GRAD_DEBT_MDN IS NOT NULL
ORDER BY GRAD_DEBT_MDN DESC;
```

[G]:

```
SELECT INSTNM AS 'Institution Name', GRAD_DEBT_MDN AS 'Median Graduate Debt'
FROM "2015_16"
WHERE CONVERT(VARCHAR, INSTNM) IN ('Harvard University', 'Yale University',
'Princeton University', 'University of Pennsylvania', 'Dartmouth College', 'Columbia University',
'Brown University', 'Cornell University')
ORDER BY GRAD_DEBT_MDN DESC;
```

[H]:

```
SELECT INSTNM AS 'Institution Name', PCIP14 + PCIP26 + PCIP27 + PCIP40 + PCIP41 AS
'Total Percentage of STEM Degrees', GRAD_DEBT_MDN AS 'Median Graduate Debt',
MD_EARN_WNE_P10 AS 'Median Student Earnings'
FROM "2013_14"
WHERE PCIP14 IS NOT NULL AND PCIP26 IS NOT NULL AND PCIP27 IS NOT NULL
AND PCIP40 IS NOT NULL AND PCIP41 IS NOT NULL AND MD_EARN_WNE_P10 IS
NOT NULL
ORDER BY (PCIP14 + PCIP26 + PCIP27 + PCIP40 + PCIP41) DESC;
```