The Equality of Upward Mobility Project

Assessing Higher Education Institutions through a Different Perspective: Upward Mobility

Hans Kullberg

May 2019

Abstract

The Opportunity of Upward Mobility has long been the backbone of the American Dream. It is the reason millions of immigrants flock to the United States as a beacon for determination and grit, to have a chance to succeed in the Land of Opportunity. Education has long been regarded as the cornerstone towards success with clearly defined metrics indicating that future success hinges on attaining higher levels of education. College Education, in particular, has long been regarded as the gateway to job attainment, career advancement and eventually, a higher income. Overtly or inadvertently, students elect to attain a college degree with the common goal of attaining higher levels of income mobility. As detailed in the groundbreaking inequality study and white paper, *Mobility Report Cards: The Role of Colleges in Intergenerational Mobility,* by Chetty, Friedman, Saez, Turner and Yagan and published by the National Bureau of Economic Research, access to university institutions is heavily skewed towards the already-wealthy. Furthermore, within the pool of institutions of higher learning, there exists a sizeable performance gap pertaining to upward mobility rates. That is, while it is well-known that colleges play a pivotal role in establishing upward mobility, some colleges are much more proficient at doing so than others. This paper utilizes machine learning and data science techniques to understand the *who* and *how* – which institutions are exceling in advancing upward mobility and what characteristics of these schools are propelling them to achieve these high rates. While the study, *Mobility Report Cards,* contains a rich database of parent and student income quintile information along with college characteristics ranging from selectivity, tier and graduation rates, it leaves the analysis to descriptive statistics rather than explanatory causes of students' future outcomes. Using a factor attribution approach, we find that schools with high levels of selectivity with a socially elite student body composition tend to have high upward mobility scores, typically carrying a high cost of tuition. The data demonstrates that upwardly mobile universities tend to have a proclivity to admit students from the highest income quintiles, spurring the question of whether universities actually contribute to the growing wealthy inequality divide. However, many alternative options exist within less selective, lower tier schools that sport an 'upward mobility score' on par with elite institutions. Trade, military and technical colleges offer standout options for achieving upward mobility. As the 'universal college debt-forgiveness' debate dominates political conversations with skyrocketing tuition prices, our study identifies institutions of higher learning that are 'worth their weight in salt' in addition to those that don't serve their students justice. As college-attendees become more cost-conscious and hesitant to take on exorbitant debt to attend university, students are placing more emphasis on 'return on investment' metrics to ensure their hard-earned money is actually worth the expense. Finally, while much emphasis and attention are paid to the established *U.S. News & World Report College*

Rankings, we present an alternative ranking system based on an institution's demonstrated ability to mobilize a student through higher social ranks from lower classes, manifesting in the American Dream.

I Introduction

The *Mobility Report Cards* project studied over 30 million college students between 1999-2013, observing key metrics such as parent's income quintiles and student's income quintiles 10 years after graduation when alumni are into their early thirties (approximately 31-34 years of age), serving as the basis for comparison measuring mobility on a holistic scale. Data is aggregated and compiled by each university, eliminating confidentiality concerns at an individual level. Data sources include federal income tax returns, the Department of Education and the Social Security.

Our study differs from *Mobility Report Cards* in 3 important ways:

1) **Emphasis on Mobility rather than Accessibility.** One of the major goals of the *Mobility Report Cards* study[1] was to identify schools admitting a large number of lower income students, granting access through those coveted doors. As such, the mobility rate measurement is calculated as the 'access rate', expressed as % of student body in Lowest Quintile, multiplied by the conditional probability of a student ending up in the Top Quintile given that he is from the Lowest Quintile. Under this methodology, which emphasizes both Accessibility and Mobility, a school with high percentage of low income students (high accessibility) but lower conditional probability of elevating them to the top quintile (low probability) could be equivalent to a school with a low percentage of low income students (low accessibility rate) but very high rates of elevating the few they have to the top (high mobility rate). In reality, this is really a measure of 'Accessimobility", encompassing both accessibility and upward mobility while holding both to the same regard of equal importance. Implicit in this assumption is a static acceptance rate whereby the percentage of low-income students accepted is equivalent to the overall acceptance rate. One possibility that cannot be discounted is that fewer lower income students apply to high mobility schools and thus, the yield rate of students hailing from Lower Quintile backgrounds is lower, penalizing the overall 'Upward Mobility Score' calculated in *Mobility Report Cards*. That is to say, it's feasible that lower-income students have a *higher*

[1] Mobility Report Cards: The Role of Colleges in Intergenerational Mobility Chetty, Friedman, Saez, Turner, and Yagan (2017) http://www.equality-of-opportunity.org/data/

acceptance rate to Upwardly Mobile institutions, but they may only apply in small numbers. The acceptance rate and yield of lower income students may vary across each university depending on the number that apply, making it difficult to determine which college is actually more accessible than others. That being said, the data strongly demonstrates that many colleges exhibit a strong degree of income segregation, supporting a continuation of the status quo, selecting mainly students from wealthy backgrounds. However, without full disclosure of acceptance and rejection rates by income quintile from universities, which may or may not be tracked, it's difficult to ascertain whether lower income students are disproportionately rejected when applying to university. Thus, while it is critical to promote social inclusion and socioeconomic diversity within a student body amongst social classes, it's difficult to assess whether colleges are disproportionately biased in their accessibility levels, and hence difficult to disaggregate the underlying driver of 'accessimobility' that is postulated in *Mobility Report Cards*. In defining their mobility rate, it's entirely possible to confound 'accessibility' with 'mobility', making it difficult to disentangle which one is driving the score when conducting attribution analysis. In this study, we're simply focused on the Mobility rate – that is, does the institution in question elevate its students to the next income level?

2) **Identification of the underlying explanatory causes attributable to higher rates of Upward Mobility**. Descriptive statistics describe the current state of affairs, but typically don't go far enough examining the underlying causes of higher mobility. By applying data science techniques, in particularly Binary Decision Trees and Random Forest Classifiers, we're able to deduce which attributes or features contribute more significantly to overall upward mobility score. This leaves out the guessing game while providing statistical underpinning of exactly what is contributing to upward mobility in the first place. Of course, individual characteristics of each student – perseverance, grit and determination just to name a few – can certainly contribute to overall upward mobility irrespective of institution. However, we're evaluating the overall picture, exploring driving factors leading to upward mobility from an institutional level.

3) **Holistic determination of 'mobility' focusing on 'jumps' along social classes.** Whereas *Mobility Report Cards* focuses solely on the number and conditional probability of students jumping from the bottom quintile (0-20%) of the income distribution to the top quintile (80-100%), we believe this is a narrow approach to defining income mobility. We take a more holistic stance on measuring income mobility, defined by

upward movement along income quintiles, and don't ignore the instances wherein a student jumps from the bottom to the fourth quintile (3 social class jumps) or 2nd lowest to the top quintile 3 social class jumps as well). In addition, we look at other 'social leaps' equivalent or greater than 2 jumps. To overlook these alternative scenarios would be to disregard a considerable amount of insight that can be gleaned at measuring upward mobility. To be sure, we weight a 4-social class jump greater than a 3 or 2-social class jump, but we don't discount the 2-social class jump entirely either, utilizing a weighted sum-product of class redistribution.

4) **Consideration of environmental factors that play a role in upward mobility.** "You are a product of your environment. So choose the environment that will best develop you toward your objective…Are the things around you helping you toward success – or are they holding you back?"[2] This quote by Clement Stone encapsulates the impact of a university's environment as a breeding ground for potential career success. While the lessons learned within the classroom play a central role to developing career-readiness, the environment on campus can be more critical in dictating future success. The connections one makes, the opportunities one takes, the friends one keeps and the relationships one builds are as essential if not moreso to helping build a future career as the grades earned within the classroom. While the *Mobility Report Cards* study collected a vast amount of information on college attributes ranging from endowment size and faculty salary to racial demographics and popular majors, very little attention was focused on incoming socioeconomic demographics of each school *in relation to* determining upward mobility. In this study, we acknowledge that the socioeconomic composition of the student body can have a significant role in opening doors and opportunities leading to income advancement.

While colleges play a pivotal role in shaping a student's future earnings potential, it should be noted that highly selective schools have the ability to select students with more natural talents relative to non-selective schools, confounding the ability to distill innate talent versus greater value-add from the institution itself. However, in analyzing driving factors contributing to an institution's Upward Mobility Score, this selectivity component can be captured and measured via a school's rejection rate[3] and its Barron's Selectivity Index. The notion implies that a higher degree of selectivity should emanate in an objectively more

---

[2] Quote from W. Clement Stone: "You are a product of your environment. So choose the environment that will beest develop you toward your objective. Analyze your life in terms of its environment. Are the things around you helping you toward success – or are they holding you back?"

[3] Rejection Rate is calculated as (1 – Acceptance Rate) for an institution.

talented group of incoming students, and hence higher potential for income mobility, regardless of institution attended.

Similar to the *Mobility Report Cards* study's stated goal of identifying colleges and institutions of higher learning that do an exemplary job of increasing income mobility, we are interested in identifying institutions that provide an affordable, scalable model producing students that move up the income ladder. Standout institutions include trade schools, military colleges and professional advancement schools delivering high value-added skills. While tuition price does have a positive correlation with overall upward mobility, there are several schools at the lower end of the tuition spectrum on par with elite, ivy league institutions with regards to upward mobility.

II Determination of Upward Mobility Score

The dataset[4] includes detailed information regarding each school's parent income profiles, student income profiles (10 years post-graduation) and conditional probabilities of attaining a higher income quintile given starting income quintile. Parent income quintile is expressed as a fraction of the student body makeup coming from each quintile. For example, 'Par_q2' represents the fraction of parents in the 2nd lowest income quintile). Top-tier, elite universities such as Harvard and Princeton fail to welcome substantial numbers of lower income students with the fraction of bottom-quintile students at 3.0% and 2.0%, respectively, against a national average of 12.5%. Looking at the 2nd Lowest rung (20-40% percentile) doesn't shift the numbers drastically only 5.3% and 4.1% of Harvard and Princeton's student body come from this class against a national average of 16.6% of a student body's population hailing from the 20-40% income demographic. As a reference, the national composite average representation from each of the income quintiles from bottom to top is 12.5%, 16.6%, 20.7%, 24.2% and 26.0%, confirming the general tendency for higher education to be skewed towards higher socioeconomic classes.

In addition to parental income, the dataset contains student income quintile statistics 10 years removed from graduation, coinciding roughly between the ages of 31-34. Studies have shown that the trajectory one's career takes by this checkpoint typically carries later in life, serving as a decent benchmark for projected lifetime income quintile rank. Student income quintile is expressed as a fraction of the student body composition within each quintile. For example, 'k_q4' represents the fraction of students in the 4th income quintile (60-80%). The

[4] Mobility Report Cards: The Role of Colleges in Intergenerational Mobility Chetty, Friedman, Saez, Turner, and Yagan (2017) http://www.equality-of-opportunity.org/data/

California Maritime Academy, which ranks as one of the most upwardly mobile institutions, sees a full 72.0% of their graduates ascending to the top income quintile, besting Princeton's 69.2% of graduates rising to the top bracket. A healthy 82.6% of graduates from the Saint Louis College of Pharmacy, a non-selective school, end up in the top income quintile while only 29.7% of students hail from the top quintile coming into the school, a drastic boost in upward mobility and class redistribution.

Finally, the most relevant descriptive statistic pertaining to measuring upward mobility is the *Conditional Probability* of rising to a particular income quintile given that one is coming from another income quintile. This is the most reliable measurement of upwards mobility, describing the contingent probability of ascending (or descending) an income class. For example, 'kq5_cond_pq1' represents the percentage of students climbing up to the 5th Quintile (highest) given that she is coming from the 1st Quintile. This particular conditional probability is the one *Mobility Report Cards* focuses on, but the dataset includes all permutations across various income quintiles as well (25 total permutations). Revisiting the Saint Louis College of Pharmacy, an outstanding 91.9% of students starting out in the lowest income quintile (0-20%) ascend to the top income quintile just 10 years out of school. By comparison, just 57.7% of bottom quintile incoming freshmen at Harvard end up in the top quintile, a considerable leap nonetheless. Nationwide, roughly 19.6% of incoming freshman from the lowest income quintile elevate their status to the top quintile after graduation.

While the 'kq5_cond_pq1' constitutes the largest leap, skipping up 4 income quintiles, a jump from the 1st to the 4th ('kq4_cond_pq1') is a remarkable accomplishment as well. We take a holistic approach in determining our upward mobility, accounting for any multi-class jump in income quintile. In deriving our measurement for an accurate representation of a school's **upward mobility score**, we use the following calculation:

(kq3_cond_pq1 + kq4_cond_pq2 + kq5_cond_pq3) X 1.0

+

(kq5_cond_pq2 + kq4_cond_pq1) X 1.25

+

(kq5_cond_pq1) X 1.5

Thus, we place more emphasis on the larger conditional income jumps but, at the same time, do not discount the scenarios where 2 or 3 income quintiles are skipped. This gives a holistic representation, primarily focused on the lower quintile but without ignoring other significant leaps. As the same methodology is applied uniformly across all schools with multipliers against each conditional probability, an objective comparative assessment can be made. Furthermore, in using conditional probabilities rather than counts, the acceptance rate of lower income students admitted into each institution is *not* taken into consideration. That is, we are indifferent to the accessibility or acceptance rates for lower income students. The summary statistics for the Upward Mobility Score across all school in the study are:

Mean: 0.8035
Median: 0.7359
Standard Deviation: 0.3199
Minimum: 0.1714
Maximum: 2.1737
25%: 0.5649
75%: 0.9981

Finally, we evaluated other potential metrics and calculations to determine upward mobility including a straightforward difference between students' versus parents' income quintiles as well as median salary levels 10 years out of graduation, but these did not accurately reflect the conditional probability of upward mobility coming from a lower class, which is our primary objective in this study. For curiosity, we also observed the mirror image of the Upward Mobility Score to derive a Downward Mobility Score for which many small, non-selective private colleges topped the list. In this calculation, the conditional probability of starting in the top quintile and finishing in the bottom (pq5_cond_kq1) would be most economically damaging and would thus carry the highest weight. Based on its non-performing service to its students, these schools may see future enrollment and tuition prices decline if potential students base their decisions on upward mobility.

We utilize this objective measurement of Upward Mobility Score across all of our data science models as the dependent variable, seeking to understand the relative importance of all independent variables that account for such a score. That is, we're answering the question – what factors play an important role contributing to upward mobility?

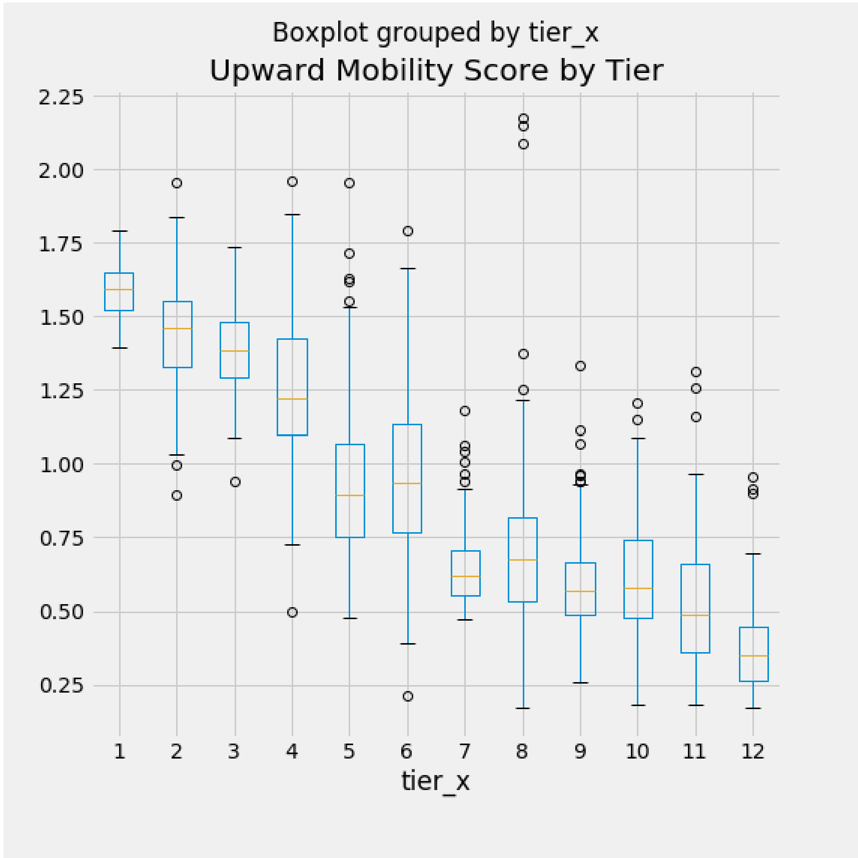
III Data Cleaning, Preparation, Extraction & Feature Engineering

In order to set up a robust, working dataframe with complete data – actual or imputed – many steps were taken to clean the data and extract key features. Feature engineering was applied to the dataset to create and build new features from existing variables using domain knowledge. This step of feature engineering, understanding relevant and related information contained within the dataset, oftentimes provides a significant predictive edge when building machine learning models. In our study, we're focused on determining the underlying attributes instrumental in attaining a high level of upward mobility as an educational institution.

Among the Features (x-variables) included in the dataset:

- School ID
- Name
- School Type (public, private non-profit, for-profit)
- School Tier (1 through 12, 1= Ivy Plus, 12=Less than 2-year school)
- 2-year or 4-year college
- Region (1-4)
- Barron's Selectivity Score (1-999, 1=Elite, 999=Non-selective)
- HBCU ? (Historically Black University)
- Student Size of School
- Tuition Sticker Price
- Public or Private
- Graduation Rate within 150% of average graduation time period
- Average Faculty Salary
- SAT Average (2013)
- Net Tuition for Bottom 20% Quintile
- Rejection Rate (1 – Acceptance Rate)
- Endowment Assets per Student
- Instructional Expenses per Student
- Asian/Pacific % Student Body
- Black % Student Body
- Hispanic % Student Body
- % Art & Humanities Majors (as a percentage of Student Body)
- % Business Majors
- % Health Majors
- % Multidisciplinary Majors
- % Public & Social Services Majors
- % STEM Majors
- % Social Sciences Majors
- % Trade & Personal Services Majors

In order to collate all data in one location for testing purposes, three separate raw data tables were merged together, joined by a common school ID. For categorical features with multiple classes (School Type, School Tier, Region, Barron's Selectivity Score), dummy variables were created with boolean integers (0 or 1) to assess this data in a numerical fashion, without giving more importance to one category over another. In a similar manner, much of the data had been formatted as *strings* and *objects* in a text-based format, which were converted to *floats* and *integers.*



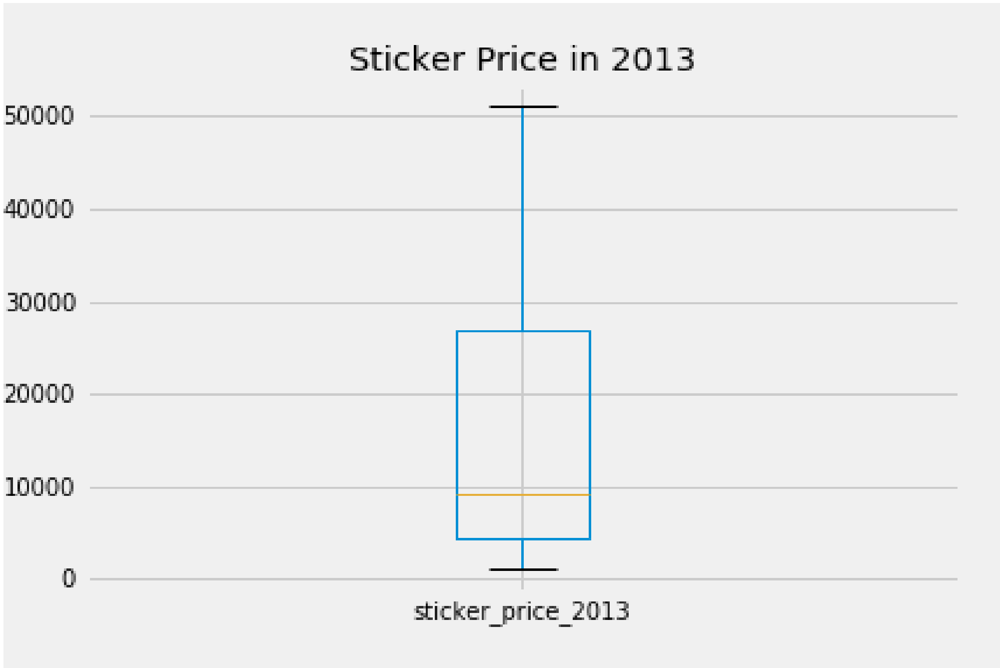
Upward Mobility Score grouped by Tier (Tier 1 = Ivy League)

Imputing Missing Data

Wrangling with missing data and imputing values is an essential element in the data science process in order to make predictions for missing or non-existent values. A couple of continuous variables have a significant number of missing values including Tuition Price ('sticker_price_2013'), SAT Average Score ('sat_avg_2013'), Graduation Rate ('grad_rate_150_p_2013'), and Rejection Rate ('scorecard_rej_rate_2013') with 100, 1245, 172 and 967 missing data points, respectively, out of the 2199 total schools. In order to understand the importance of each one of these features in terms of relevance towards predicting upward mobility, we used various regression models to determine importance and impute missing values where appropriate. The Tuition Price[5] (not including room & board) of the schools in our study exhibits a skewed, non-normal distribution with an

[5] Appendix, Exhibit E. Tuition Sticker Price vs. Upward Mobility Score

average of $15,876 per year but standard deviation of $13,745 and max of $51,008 in 2013 dollars.

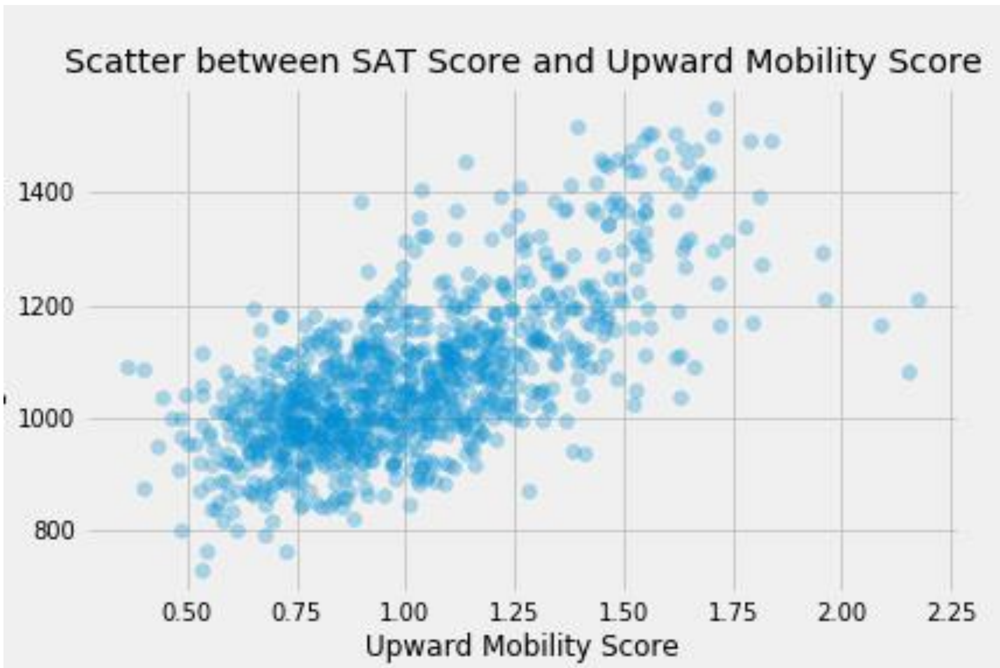


We use a K-Nearest Neighbors (KNN) Regression Model[6] with 6 predictive features to calculate a proxy Tuition Price by identifying the most similar institutions to the 100 schools with missing values. The K-Nearest Neighbors Regression Model is a simple algorithm that stores all available cases and predicts the numerical target (Tuition Price) based on a similarity measure using 'distance functions'. The 6 predictive features are Region, Barron's Selectivity Score, School Type, School Tier, Endowment Assets per student and Instructional Expenses per student. Permutations of hyperparameters (leaf size and number of neighbors) were evaluated with 30 leaves and 15 neighbors determined to result in an optimal root-mean squared error (RMSE). Results show the median and mean ($7,285 and $11,403) of the imputed sticker prices were indeed much lower than the aggregate median and mean of the entire dataset ($9,157 and $15,876), reflecting the right-tailed distribution. This validated the need to impute these values using a regression framework rather than simply relying on the mean or median. The predicted sticker prices for each school replaced each missing value (NaN) where applicable.

The Average SAT Score feature has a 67.5% correlation with the Upward Mobility Score and thus seems to be a pertinent feature to impute missing values, of which there over 50% of

[6] SciKit Learn K-Nearest Neighbors Regression Model documentation. https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsRegressor.html

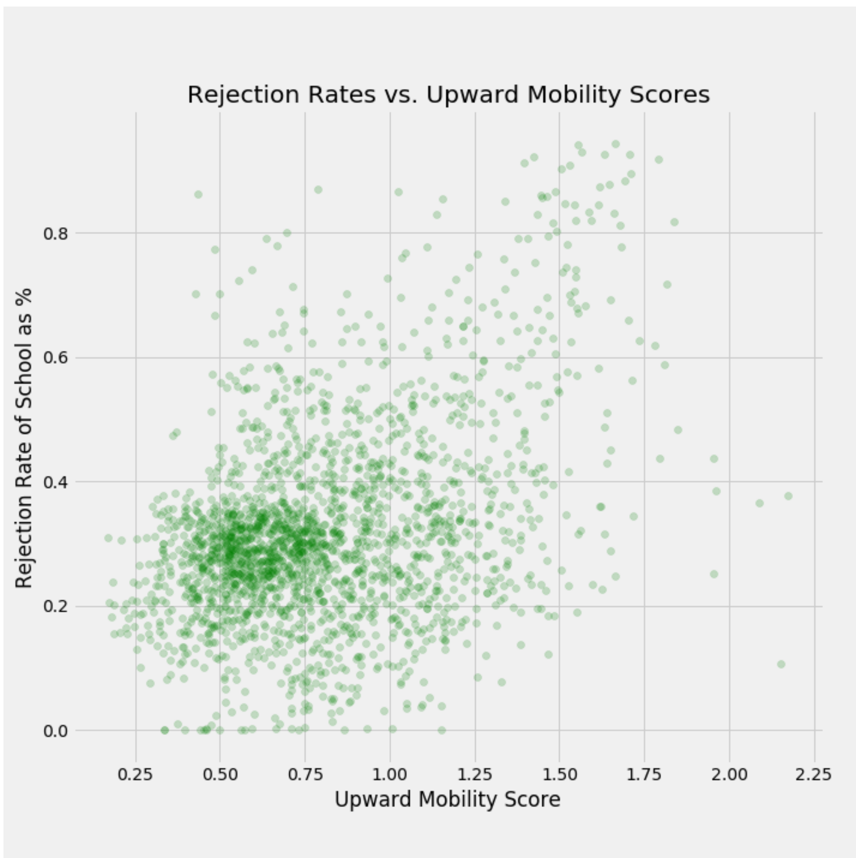
the dataset missing. Interestingly, a full 1245 out of 2199 schools have missing data for this feature, perhaps due to failure to report, lack of collection means or an institution's decision to hold that information confidential. However, when analyzing the relative feature importance of this variable using a binary decision tree regressor, the SAT Score only ranked as the 10th most important feature with a relatively low 1.4% degree of importance. Thus, when taking into consideration that more than half of the SAT Score data was missing along with its relative unimportance, it was decided to drop this particular variable for our overall upward mobility model.



Rejection Rate[7] is the inverse of Acceptance Rate and speaks to the relative selectivity of each school. A higher rejection rate signifies that a school can be more discerning in which students they admit and may post high barriers to entry (essay quality, test scores, etc.) for the lowly qualified. With 967 missing data points and a right-tail skewed distribution due to a small number of colleges with extremely high rejection rates, we're not able to simply plug in a mean or median value, especially given the significance of the feature. We employ a Random Forest Regression Model[8] to predict the rejection rate value for missing data. A Random Forest Model[9] combines multiple decision trees in determining the final output rather than relying on one individual decision tree. A distinct element of Random Forest models is a technique called Bootstrap Aggregation, commonly known as bagging, which

[7] Appendix, Exhibit B. Rejection Rate vs. Upward Mobility Score
[8] SciKit Learn Random Forest Regressor documentation. . https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html
[9] Random Forest in Python. *Towards Data Science.* https://towardsdatascience.com/random-forest-in-python-24d0893d51c0

involves training each decision tree on a *different sample of features* where the sampling is done with replacement such that the next tree may use an entirely different set of features. Using explanatory variables of School Type, Tier, Region[10], Barron's Selectivity, Instructional Expenses, Graduation Rate[11], Average Faculty Salary, Tuition and Tech-School Status, we fit our model on the non-null dataset to build our predictive model. Hyperparameter optimization was conducted with the number of estimators (unique decision trees) at 800 and Max number of features at 8. Predicted values for the missing data using our Random Forest Model were then inserted back into the main dataframe. Compared to the original, non-null dataset, the Median and Mean Rejection Rate calculated from our model was 0.28 and 0.27 versus 0.34 and 0.36 in the non-null dataset, reflecting the large dispersion due to the right-tailed nature of the data.



For missing data on non-important features or with very few missing datapoints within a feature, the distribution of the data was evaluated to assess whether a mean (normal distribution) or median (right or left-tail skewed distribution) should be applied and filled in. These relatively non-important or sparsely missing features include Graduation Rate within 150% of normal period (172 missing values), Average Faculty Salary (71 missing values),

[10] Appendix, Exhibit C. Upward Mobility Score by Region

[11] Appendix, Exhibit D. Graduation Rate vs. Upward Mobility Scores

Instructional Expenses per student (6 missing values), and relative fraction of Majors (11 missing values).

Measuring Social Eliteness

"Sometimes it's not *what* you know, but *who* you know that makes all the difference". This quote succinctly captures the reason for investigating the composition of a student body. Whether it's building a relationship or connection for a future job or potential business partner, some of the greatest value in attending college is acquired *outside* the classroom. Clubs, activities, events, study groups, parties and eating halls all provide ample opportunity to build and forge relationships that may change one's trajectory. Oftentimes, students from elite backgrounds have access to privileged opportunities that lower classes do not. In the business world, it is often connections to exclusive networks that can open doors and opportunities that very few outside people see. While this may be a commonly-agreed upon fact, it's difficult to discern which how these connections are made and which schools provide prime breeding ground to foster these relationships.

While the dataset already contains the *demographic* make-up of the student body between White, Black, Hispanic and Asian, it doesn't demarcate the *socioeconomic* make-up of the student body. While the *Mobility Report Cards* study does affirm that many institutions, especially elite universities, maintain the status quo by disproportionately selecting and educating the wealthy, it doesn't go far enough to quantify this using the data already available. Using the percentages depicting the fraction of students with parents from each income quintile, we derived a score to measure this 'Social Eliteness':

2 X ['par_top10pc'] + 1.5 X ['par_10-20pc'] + 1 X ['par_q4']

Here, we emphasize the top 10% percentile (90-100%) by placing 2x times the weighting as the 60-80% quintile and 1.5x times the weighting at the 80-90% percentile. We are using these multipliers against the percentage of students in each decile/quintile to magnify the significance of having socially elite students comprising the student body. This score presents a granular measure of the socioeconomic composition of the student body, enabling us to include it as a feature in our predictive model. Across the 2199 schools, this 'Social Eliteness Score' feature has an average of 0.70, standard deviation of 0.31, with a 25% and 75% cut-off of 0.45 and 0.90 respectively. The Social Eliteness Score has a correlation of 75.0% with the Upward Mobility Score, confirming the analysis that upwardly mobile institutions tend to cater to the already-wealthy classes, student potential notwithstanding.

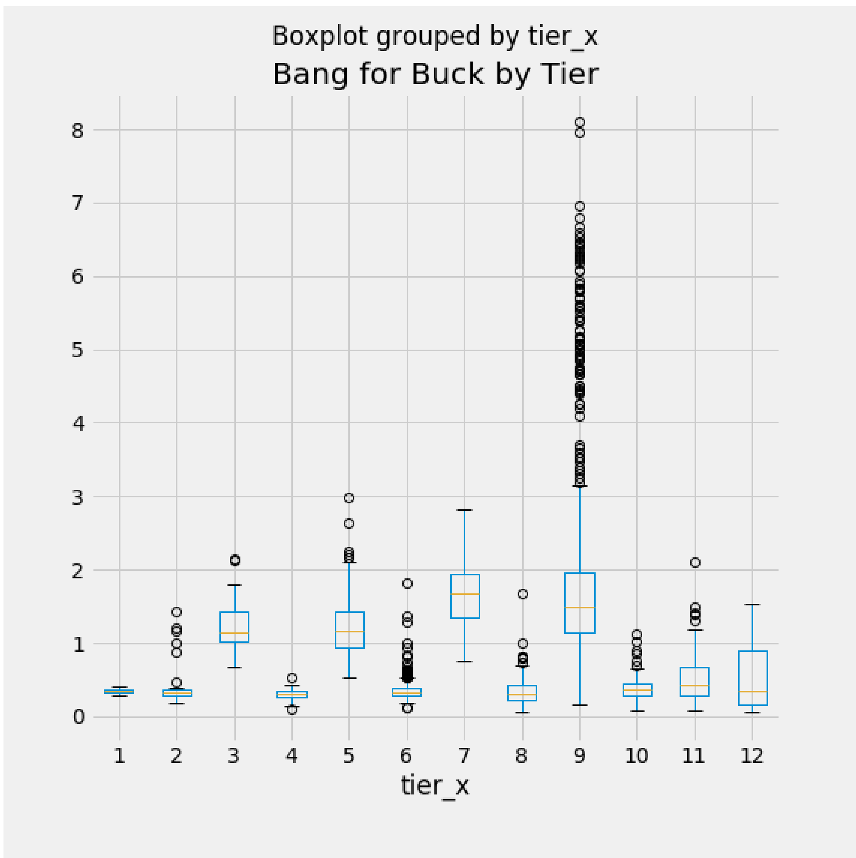
The Social Eliteness Score feature exhibits these summary statistics:

Mean: 0.6998
Standard Deviation: 0.3104
Minimum: 0.0693
Maximum: 1.664
25%: 0.4541
75%: 0.9053

Bang for the Buck

In an environment with persistently rising inflation, mounting student loans and questionable job prospects, students are asking themselves "am I getting my money's worth of education?" To address this issue, we take a look at the incremental upward mobility 'return' one receives for each incremental dollar spent on tuition, as the two series have a positive correlation, to derive a 'bang for the buck' metric. In order to scale the series appropriately given that Upward Mobility Score has an average around 0.8 while the average tuition per institution is $15,672, we multiply the Upward Mobility Score by 10,000.

This gives us a more granular perspective of the value of each tuition dollar spent. Depending on a student's prospective budget, financial aid and willingness to take on student debt, she can make an objective assessment in answering the aforementioned question using this metric. While Tier 1 (Ivy/Elite) schools have the highest distribution of Upward Mobility Schools, we see that their relative 'bang for the buck' is tightly bounded with no exceptional outliers. Conversely, one can find outliers along all tiers of schools, with Tier 9 being the most outstanding category. The Tier 9 category represents *2-year public and private not-for-profit universities* and thus, the denominator (Tuition Price) may drive this metric in many cases if tuition costs are very low. We explore some exceptional schools from these 2-year universities in our Appendix (Exhibit A).



IV Model Selection, Features & Hyperparameters

With Feature Engineering, Data Cleaning and Extraction steps completed, we can build both classification and regression-based models. Classification models distinguish the probability of a tuple, in this case higher education school/institution, belonging to a particular category or class. On the other hand, regression models predict point estimates of a continuous variable, in this case, Upward Mobility Score with each model seeking to attain a high degree of accuracy. While our Upward Mobility Scores (y-variable) are continuous, we can separate them into separate categories to be used in classification-based models. Using our distribution of Upward Mobility Scores, we objectively assign institutions to one of 5 classes using these cut-off points on Upward Mobility Score to derive 'Upward Mobility Buckets':

Class	Lower Bound	Upper Bound	# datapoints
Worst	0	0.5	330
Below Average	0.5	0.7	648
Medium	0.7	1	673
Good	1	1.3	357
Best	1.3	2.5	191

These demarcation cut-off points generate a roughly normal distribution with a right-tail at the top-end of the spectrum. We have the fewest number of datapoints in the 'Best' category, capturing the right-tail.

The overarching goal for our models is to be accurate with low variance while still being able to generalize across a larger population of data. That is, we do not seek to overfit our model, but would rather have the model work well to predict other institutions' Upward Mobility Scores and Buckets that are not included in our sample. To achieve these dual objectives, we apply a train-test split as well as cross validation across all models, using a training set and a testing set and iterating across multiple splits. Cross Validation[12] is a technique used to test the effectiveness of machine learning models across multiple samples of training and testing data, allowing us to understand if we're overfitting or underfitting to one particular dataset. Using a model trained on the training dataset, cross validation applies this model to the testing dataset. If results are fairly similar, the built model can be thought of as reliable to use on out-of-sample data.

We apply all of our features presented in numerical format as integers or floats with categorical variables converted to Booleans or classes. This set feature columns includes our 4 engineered features: **Social Eliteness Score, Tech School (Boolean), Tuition Sticker Price, Rejection Rates**, rounding out our total of 27 unique features. In all classification-based models, our y-variable is Upward Mobility Bucket while in all regression-based models, our y-variable is the continuous variable, Upward Mobility Score.

[12] SciKit Learn, *Cross-Validation: Evaluating estimator performance* https://scikit-learn.org/stable/modules/cross_validation.html

Regression Models

A regression model is the statistical technique for estimating the relationship of a dependent variable (Upward Mobility Score) across several variables that may help explain the variance of that dependent variable. Many of the input, explanatory variables may not be related to the dependent variable, but in conjunction with other variables, it may provide a boost to predictive accuracy. One of the metrics used to describe the accuracy of a regression model is the root-mean-square error, which measures the differences between values predicted by a model and the actual values observed. Differences in actual versus predicted values are squared to account for any negative values and then summed and divided by the number of values to derive an average. We then take the square root of this number to undo the previous squaring, which naturally emphasizes bigger deviations from actual value. The Root-Mean-Square Error (RMSE) can be used as a way to determine the quality of prediction as model generates and it is what we employ to gauge relative accuracy across each of our regression models.

A Linear Regression model is the simplest approach to regression models, using a linear approach to model the relationship between a dependent variable and multiple explanatory variables using a single line to estimate the predicted values. We use the linear regression model as a baseline approach to our dataset. Lasso and Ridge regression models are variations upon the linear regression model that help reduce overfitting and the complexity of a linear regression model using a technique called Regularization. In Ridge Regression, the cost function is altered by adding a penalty equivalent to the square of the magnitude of the coefficients. As ridge regression puts constraint on the coefficients via the penalty, effectively shrinking the coefficients and helping reduce multicollinearity. Lasso regression takes a similar approach with the only difference being that magnitudes of the coefficients are taken into account instead of taking the square of the coefficients. This type of regularization can completely eliminate some features to have a zero for a coefficient as the penalty reduces its importance. Lasso regression is frequently used to reduce over-fitting by identifying and selecting only important features when a dataset has many features to choose among. By tweaking the Alpha parameter in both Ridge and Lasso Regression, one can identify the most important features explaining variation in the dependent variable, in this case, Upward Mobility Scores.

A Binary Decision Tree Regression Model[13] predicts a target variable, here Upward Mobility Score, using a classification technique whereby it splits the dataset into separate, distinct categories using binary splits, most commonly dividing the data by the most significant, distinguishing feature first. While it's typically used for classification purposes, Decision

[13] SciKit Learn, *Decision Tree Regressor.* https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.html

Trees can be used to predict continuous variables using a Regression technique. The Depth of the tree pertains to the number of 'branches' a decision tree encompasses as the splitting increases. A Random Forest Regression Model[14] is an amplified, boosted version of the binary decision tree regressor using a bagging approach to feature selection in which certain subsets of features, specified by 'max features' are used to split each tree across several, up to thousands, of decision trees. The number of estimators represents the optimal number of decision trees to use within the Ensemble technique of Random Forests. Finally, a certain subset of the dataset is withheld, that can be used for out-of-sample testing which can be critical for particularly small datasets. An Out-of-Bag score[15] error is a method of measuring the prediction error of random forests and other machine learning models utilizing bootstrap aggregating to sub-sample data samples used for training.

Taking a look at the Train-Test Split and Cross Validation Root-Mean-Squared Error rates for each of the models which predict the Upward Mobility Score using the feature inputs, we find the following summary statistics:

Regression Models	Train-Test-Split RMSE	Cross Validation RMSE	Additional Notes
Linear	0.1550	0.1491	
Ridge	0.1793	0.1669	
Lasso	0.3300	0.3198	alpha = 0.01
Binary Decision Tree	0.2201	0.2140	Max Depth = 7, Max Features = 18
Random Forest	0.1414	0.0844	n_estimators = 900, Max Features = 10, oob score = 79.88%

While the Train-Test Split takes into account only 1 'out-of-sample' testing set, the Cross Validation metric accounts for 10 separate train-test splits, reporting the average RMSE score, and thus represents the most dependable metric to be used for generalization purposes. Not surprisingly, the Random Forest approach exhibits the lowest error rate to predict Upward Mobility Scores. However and quite surprisingly, the Linear Regression model comes in at second, outperforming both the Ridge and Lasso models, indicating many features are important to include to achieve optimal performance. In order to estimate Upward Mobility Scores with the highest degree of accuracy for schools and

[14] SciKit Learn, *Random Forest Regressor.* https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html
[15] SciKit Learn, *OOB Errors for Random Forests.* https://scikit-learn.org/stable/auto_examples/ensemble/plot_ensemble_oob.html

institutions not included in the sample, we would select the Random Forest Regressor which can be seen as our 'RF_Regressor_Predicted_UM_Score' in the Equality of Upward Mobility Project excel document.

Classification Models

While Regression Models predict and forecast a numeric, continuous variable such as the Upward Mobility Score, Classification models assign a class to each datapoint within a degree of probability. We evaluated the dataset through a number of different classification models, all generating varying degrees of accuracy, defined as predicting the correct class by applying the trained model on out-of-sample testing data. In our 10 cross-validation tests, we split the data into training (3/4ths) and testing (1/4ths) sets to observe out-of-sample accuracy. Another evaluation metric is the Precision Score, defined by the ratio of True Positives[16] divided by the sum of True Positives and False Positives. That is, measuring the number of 'correct' estimates for each prediction divided by number of total predictions for each particular class. Thus, each of the 5 classes generates a Precision Score such that we can observe the relative precision for each prediction with the overall Precision Score for the dataset represented by the sum-weighted average. A third evaluation metric for classification models is the Recall Score, which is a measure of the result relevancy, or how many truly relevant results are returned per class. Recall answers the question of 'how many actual class observations did the model capture correctly?'. The Recall Score is calculated as the number of True Positives divided by the summation of True Positives and False Negatives. In other words, it is the number of True Positives divided by the total number of actual observations in that particular class – those that the model correctly predicted and incorrectly missed. An ideal model would exhibit both a high level of Precision as well as Recall. The harmonic mean between Precision and Recall Scores is termed F1 which is calculated as 2 times the product of (Precision x Recall) divided by (Precision + Recall). The formulas for each of these evaluation metrics are as such:

Precision = True Positives / (True Positive + False Positives)

Recall = True Positives / (True Positive + False Negatives)

F1 = 2 x (Precision x Recall) / (Precision + Recall)

We use each of these evaluation metrics in assess the 'best' model to use, generating the highest degree of accuracy with lowest degree of variance on out-of-sample data. The 'Null Accuracy' score takes the majority class ('Medium') and assumes a naïve class prediction as

[16] SciKit Learn, *Precision-Recall Score* https://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html

with that class for all schools/institutions. The Null Accuracy is used as a base-case scenario assuming that we only know the most represented class in our dataset.

An additional evaluation metric to measure model accuracy is the Confusion Matrix[17], which succinctly depicts Precision and Recall Scores in a comprehensive, visual fashion. The Confusion Matrix is a table that describes the performance of a classification model on a set of test data in which the true values are known. The Predicted Classes are labeled as columns while the Actual Classes are labeled as Rows. Thus, through the Confusion Matrix, one is able to easily see how many times a model predicted each class and how many of those predictions proved accurate. Likewise, one can visualize the number of times the classification model incorrectly predicted the right class. In a multiclass model such as the one we're using here, a robust model should display a large number of datapoints on a diagonal dimension with the majority of Class 1 Predictions falling in the Actual Class 1 Bucket and so on. The Precision Score is calculated by dividing the True Positives in each column by the total number of observations in that column. Conversely, the Recall Score is calculated by dividing the True Positives in each Row by the total number of observations in each row.

Classification Models	CV Accuracy Mean	Precision Score TP/(TP+FP) Test Data	Precision Score for Each Class - Test Data	Recall Score for Each Class TP/(TP+FN) - Test	Optimal Hyperparameters
Null Accuracy ***(673 out of 2199)***	30.6%		[0.0%, 0.0%, 100.0%, 0.0%, 0.0%]	[0.0%, 0.0%, 100.0%, 0.0%, 0.0%]	
Logistic Regression	62.7%	61.8%	[69.6%, 61.3%, 58.0%, 53.7%, 85.3%]	[52.2%, 74.8%, 62.9%, 42.8%, 65.9%]	
KNN Nearest Neighbors	57.4%	61.5%	[76.3%, 60.9%, 55.5%, 57.3%, 85.7%]	[48.9%, 73.6%, 63.4%, 51.1%, 54.5%]	K=25
Random Forest Classification	64.8%	65.3%	[64.2%, 66.7%, 59.4%, 65.1%, 91.9%]	[68.4%, 63.1%, 67.7%, 64.4%, 61.8%]	n_estimators=900, Max_Features = 10, oob_score = 79.88%

[17] Data School, *Simple Guide to Confusion Matrix Terminology.* https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/

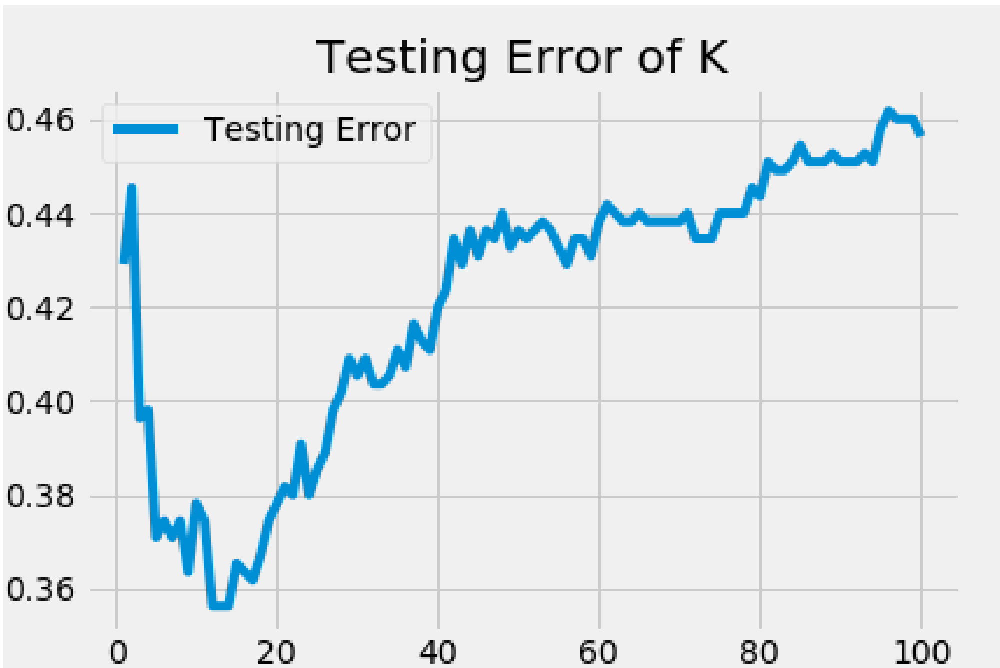
Logistic Regression[18] is a predictive analysis approach to describe data and explain the relationship between one dependent binary variable and one or more ordinal independent variables. While it is generally used to distinguish binary classes within a degree of probability along the logit function (sigmoid), logistic regression can be used to identify categorical outputs in a process called multinomial logistic regression. Logistic regression uses independent variables as inputs to assign a probability of the dependent variable belonging to a class, converting log-odds into a probability called the logistic function. We see that our Logistic Regression Model is highly precise (85.3%) in predicting Upward Mobility Bucket class 5 ('Best') scenarios but not as precise in predicting Upward Mobility Bucket Class 3 & 4 ('Medium' and 'Good') with only 58.0% and 53.7% probability respectively. Meanwhile, class 2 ('Below Average') generates the highest recall score at 74.8%, meaning that it correctly captures 3 out of every 4 Below Average institutions in the dataset. The Precision Score for on Testing Data using Logistic Regression is 61.8%while the Cross-Validation Accuracy Mean is 61.7%. The Confusion Matrix for the **Logistic Regression Classification Model** looks as such:

```
[[ 48,  37,   7,   0,   0],
 [ 11, 122,  25,   5,   0],
 [  8,  40, 105,  14,   0],
 [  2,   0,  41,  36,   5],
 [  0,   0,   3,  12,  29]]
```

While we had previously used K-Nearest Neighbors Regression model to calculate a predicted Tuition Price in our Data Cleaning process, here we're using K-Nearest Neighbors Classification Model to predict the correct Upward Mobility Bucket classes. Using the 27 explanatory, independent features, the KNN Classification Model utilizes a non-parametric technique, meaning it doesn't make any assumptions on the underlying data distribution. The KNN Classification approach is a common model to use when there is little or no prior knowledge about the distribution data. In our study, we're evaluating each out-of-sample (testing) datapoint and determining which bucket that should reside within based on similar features to other datapoints that are included in the training model. The KNN Model then assigns a deterministic probability on which class/bucket the datapoint falls into. One of the most important Hyperparameters that can be chosen is the number of nearest neighbors (K)

[18] SciKit Learn, *Logistic Regression.* https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

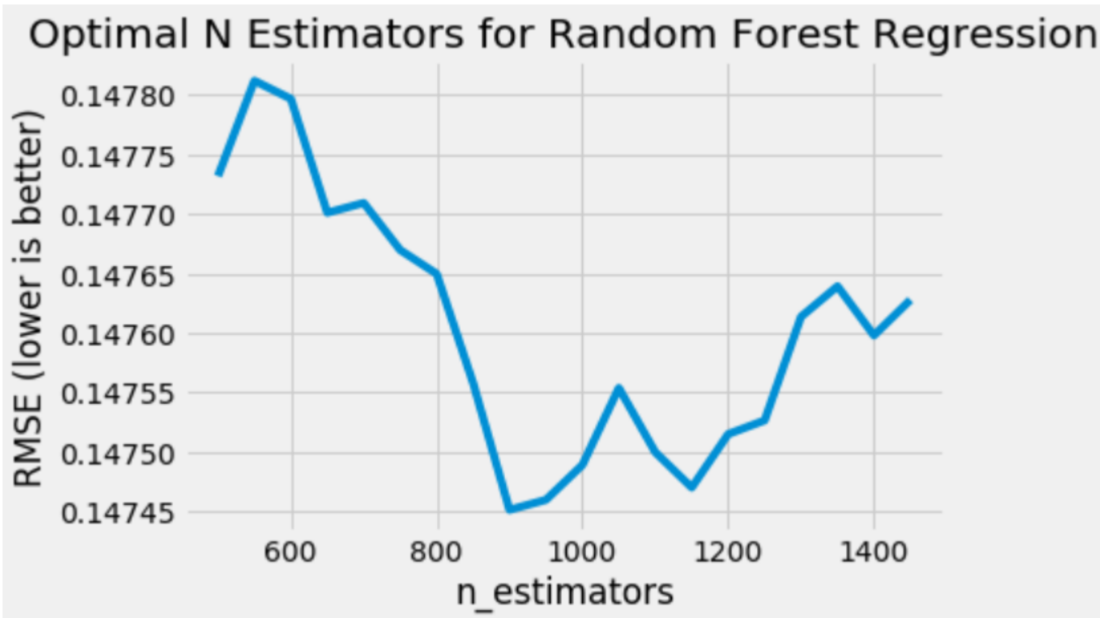
to select from. We ran through various tests optimizing the Testing Error of K and found K=15 to be optimal in terms of reducing the error. Like the Logistic Regression Model, the KNN Classification Model is precise at predicting the extremes ('Worst' and 'Best') in the dataset, with 76.3% and 85.7% precision while having more difficulty predicting the middle 3 Upward Mobility Buckets. The Recall Scores perform slightly worse than Logistic Regression with all 5 classes exhibiting marginally lower recall scores. Overall, the total Precision for the KNN Classifier Model was 61.5% on out-of-sample data while the Cross-Validation Accuracy Mean is 57.4%, marginally lower than Logistic Regression.

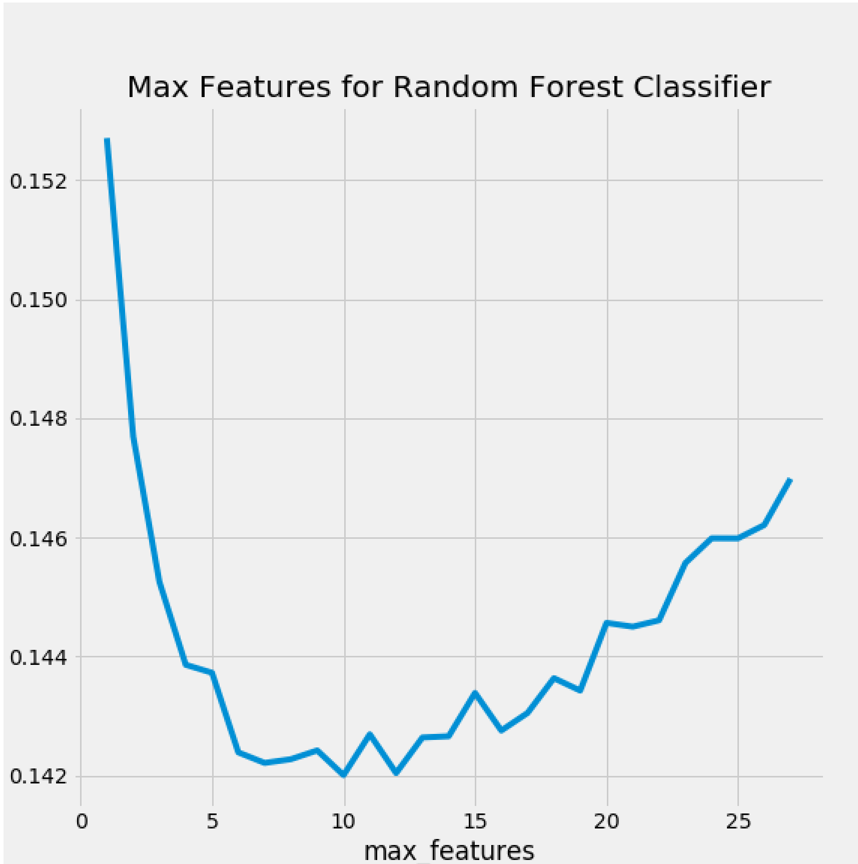


The Confusion Matrix for the **KNN Classification Model** looks as such:

```
[[ 45,  39,   7,   0,   1],
 [  7, 120,  36,   0,   0],
 [  6,  38, 106,  17,   0],
 [  1,   0,  37,  43,   3],
 [  0,   0,   5,  15,  24]],
```

The Random Forest Classification Model is a boosted ensemble model incorporating several binary decision trees to objectively derive a forecast based on a holistic aggregation of disparate tree models using out-of-bag, bootstrap aggregation. To tune and select our hyperparameters used in the model, we determined the optimal number of trees (n_estimators) to be 900 with a Maximum Features of 10. Using the Random Forest Classification Model, we see an uptick in precision across the board with all buckets save for class 1 ('Very Poor') indicating an increase against both Logistic and KNN models. The Precision Scores for each bucket are 64.2%, 66.7%, 59.4%, 65.1% and 91.9%, respectively, garnering an overall Precision Score of 65.3%. Likewise, the Recall Scores are mostly improved using Random Forests, save for the 2nd Class ('Below Average') with Recall Scores of 68.4%, 63.1%, 67.67%, 64.4% and 61.8%, respectively.





The Confusion Matrix for the **Random Forest Classification Model** looks as such:

```
[[ 52,  19,   5,   0,   0],
 [ 22, 106,  40,   0,   0],
 [  5,  32, 111,  16,   0],
 [  1,   2,  25,  56,   3],
 [  1,   0,   6,  14,  34]]
```

With an overall Cross Validation Accuracy Mean of 64.8%, the Random Forest Classification Model outperforms other classification models, we place most confidence in the Random Forest Classifier to predict out-of-sample data with a high degree of accuracy and low variance, thus use this for our production model. The probability-based scores can be found in the Equality of Upward Mobility excel document labeled as 'RF_Pred_Class[X]' for full exploration and breakdown of the Upward Mobility Bucket classification details. Each class is assigned a probability that the datapoint will fall into that particular bucket. Thus, for instances with False Positive predictions, there is a likelihood that the class with the second-highest probability could be the correct bucket. This can be observed in the Confusion Matrix as many False Positives in each column had an actual value one class higher or lower with very few predictions made as an extreme outlier.

V Feature Importance

Now that we have determined our optimal models for regression and classification purposes (Random Forests), we can seek to understand which explanatory variables drive the Upward Mobility Scores in the dataset. In order to draw objective conclusions from our models and understand the drivers to explain the results, it's important to have an interpretable model. Many machine learning models operate as black boxes that are difficult to objectively explain the results to a broader audience. In business cases, unexplainable models may be inexplicably thrown out if they're not easily understood or the underlying driving factors are too difficult to interpret.

Feature Importance[19] identifies the most important variables that help explain the underlying drivers of the model's logic. By knowing the most important features, we can work on improving the model by focusing on the most important variables or remove variables that don't contribute much to the overall prediction, based on its impurity score. Finally, feature importance allows one to make conclusive interpretations with respect to the ultimate outcome, given more granular insight into what the model deems most influential. In Random Forest models, every node is a condition of how to split values in a single feature such that similar values of dependent variables end up in the same set after the split. The Gini impurity / information gain, common for classification models, enables one to compute how much each feature contributes to decreasing the weighted impurity of the overall model. In essence, the more important the feature is, it will carry a higher importance weighting to the overall model. To ascertain the most important features, we used the Random Forest Regressor model over the Classifier model as the output is a continuous variable rather than distinct category, but the results in both cases hold many overlapping similarities.

We find **Barron's Selectivity Score** as the most important feature, perhaps a reflexive, self-reinforcing feature as the more selective the school actual is (based on Barron's calculations), the stronger the successful applicants should be from a talented application pool. For

[19] Towards Data Science, *Explaining Feature Importance by example of a Random Forest.* https://towardsdatascience.com/explaining-feature-importance-by-example-of-a-random-forest-d9166011959e

example, Yale (UMS = 1.567) has a Rejection Rate of 92.9%, or conversely an Acceptance Rate of just 7.1%, giving it the ability to select a highly talented individual who may have already demonstrated individual accomplishments indicative of future upward mobility down the road, based on her own merits. That is, while the school may do an exemplary job of selecting this individual, that student's future success may be moreso due to a byproduct of her own innate talents and ambition, qualitative intangibles that are difficult to measure. In that situation, very little of that student's success may be attributable to the programs Yale has taught her in the classroom or on campus, but moreso to her own self-drive, skill, grit and determination. Yale may have simply performed an exemplary job of selecting this high-achieving, self-motivated individual in the first place. Thus, the Barron's Selectivity Score may actually say more about the talent of the students selected to attend highly-selective schools moreso than the actual school itself. Notably, the Rejection Rate itself doesn't necessarily rank highly on the Feature Importance list, perhaps indicative Barron's Selectivity Score measuring a different aspect of selectivity apart from simply rejection rate.

Interestingly, the **Social Eliteness Score** ranked as the second most important feature, indicating Upward Mobility Scores are largely determined by the socioeconomic composition of the student body. As our model doesn't take into account the percentage of lower income students admitted but rather focuses on the conditional probability of a lower income student ascending to a higher income class post-graduation, the model identifies 'social eliteness' – being surrounded by a student population that is already elite – as one of the most predictive factors of upward mobility success. Thus, the environmental factors that a university institution fosters carries a significant weight on shaping the future outcome of each student. The mantra of "it's more about *who* you know rather than *what* you know" rings loudly as one's environment can certainly shape his direction. Socially elite schools, defined as having a significant percentage of the student body hailing from upper income quintiles, may benefit from the 'invisible advantages' that come with admitting a student from such a background. Perhaps the CEO of McKinsey, a premier consulting firm, has a son that attends Harvard and thus, decides to recruit heavily at the school. Perhaps a Stanford student's parent is a Microsoft Executive and helps his good friend land an internship at the company. Or perhaps the daughter of a Goldman Sachs managing director wants her best friend to be an analyst with her directly after graduation. These are the hidden benefits of a student population teeming with socially elite students. In terms of achieving upward mobility, schools with a wealthy student population create an environment in which beneficial connections and relationships can be fostered. In plain sight, the data says the number one factor for determining upward mobility success, save for Barron's Selectivity Score, is to be surrounded by already-elite people. Does this mean universities should focus on admitting some lower income students and surround them with 'socially elite' students? Does it mean there should be a 'social eliteness cap' placed on each school such that it would force high-income students to be spread across a larger number of schools, essentially 'equaling the playing field'? Are universities themselves, through their admissions practices, playing an integral role contributing to wealth inequality? Should institutions be required to admit a certain threshold of lower income students, particularly at public universities where state-funding could have more sway? The policy questions abound but the answer to the question - what is driving this wealth inequality gap between universities? - is abundantly clear: Upwardly Mobile universities are overwhelmingly populated by the socially elite who, in turn, help maintain the status quo ensuring it will continue to remain upwardly mobile.

The 25 top-rated Socially Elite universities as ranked by Social Eliteness Score are listed under Exhibit G with some surprising names at the top of the list. By far and away, the most unifying pattern in the socially elite universities are that they're private, for-profit 4-year institutions with very high ($40,000+) tuition prices. However, what may be surprising to some is that the list doesn't precisely follow what have been stereotyped as the 'most elite' universities in the country, namely Ivy League universities. With a gaudy 81.35% of all students hailing from the Top Income Quintile while only intaking 1.12% and 1.86% of its student body from the lowest 2 income quintiles, respectively, Washington & Lee University in Lexington, Virginia ranks at the #1 Socially Elite university in the country, perhaps surprising to many. Looking into the data further, 69.0% of Washington & Lee students come from the Top 10%, 54.3% hail from the Top 5% and 17.5% of all students come from the Top 1% of the US Population. Imagine the connections that can be cultivated when nearly 1 out of every 5 fellow students within your classroom came from the 1%. While one cannot be completely certain that Washington & Lee's admissions office intentionally favor wealthy applicants, the distribution of the data is almost too extreme to be attributable simply to chance. Princeton is the second highest-ranking school along the social eliteness factor with 76.8% of its students coming from the top income quintile while only 2.0% and 4.1% come from the lowest 2 income quintiles, respectively. Within those 76.8% of students in the top income quintile, 65.2% of the students are in the Top 10%, 52.6% are in the top 5% and, incredibly, a full 20.1% of students ascend onto campus from the Top 1%. With distributions distorted to these extremes, perhaps this means there are not enough qualified students applying from lower socioeconomic backgrounds. The more likely scenario, however, is that a student's 'financial ability to pay' may be perceived to be too low to attend these private, for-profit institutions, either overtly through financial aid forms like FAFSA or implicitly through a quick check on the student's zip code of residency. Many schools, especially private, for-profit institutions, require a 'financial willingness to pay' background check, requiring a notarized bank account statement before even stepping foot on campus. This can viewed as a roundabout manner of ensuring an institution only admits and accepts students of high-income financial backgrounds, all-but-guaranteeing a high Social Eliteness Score with a student body comprised of wealthy elite. Universities may be incentivized to attract and retain wealthy students to bolster their endowments as their endowment size is a critical factor in the *US News & World Report* ranking system. Rounding out the Top 10 universities are Davidson College, Middlebury College, Colby College, University of Richmond, Wake Forest University, Brown University, Colgate University and the University of Notre Dame. There are just 4 Ivy League universities in the Top 25 and a considerable amount of these schools are small, liberal arts colleges.

Following Barron's Selectivity and Social Eliteness Scores, the **Tier of School**, **Average Faculty Salary** and **Graduation Rate** are important features to determining upward mobility. The Tier of School feature may have some underlying multicollinearity to Social Eliteness Score, meaning this feature may be redundant as it correlates highly with the Social Eliteness Score. Faculty Salary perhaps implies higher quality staff may impart more wisdom to students with better teaching, leading to upward mobility. From an individualistic perspective, once a student has selected his university, the most imperative way to boost his upward mobility chances is by actually graduating on time. Understanding results is often as important, if not moreso, then having good results and through examination of the Important Features we understand which variables are most deterministic to the model's output. With these Feature Importance metrics, a student may be able to objectively assess the university she is planning on attending the next 2-4 years based on key criteria. If data is collected from out-of-sample colleges that are not included in this list, it is possible to compute said college's upward mobility score with the model provided. While much of the benefit a student receives in college can be attained within he classroom, a significant amount of one's college experience is defined by the relationships, opportunities and networks built outside the classroom, connections that may positively alter the trajectory of one's future career and path to upward mobility.

Feature	**Importance**
Barron's Selectivity Score	19.9%
Social Eliteness Score	17.8%
Tier of School	11.5%
Average Faculty Salary (2013)	9.1%
Graduation Rate in 150% of normal time	8.7%
% Asian/Pacific as Share of Student Body	3.4%
Instructional Expenses (2013)	3.4%
Tuition Price	3.1%
2-year or 4-year college	3.0%
% of STEM Majors	2.8%

% Hispanics as Share of Student Body	2.1%
% Art & Humanities Majors	1.9%
IPEDs Enrollment (School Size)	1.7%
% of Social Science Majors	1.7%
% of Healthcare Majors	1.6%
% Blacks as Share of Student Body	1.6%
Rejection Rate of School	1.3%
% of Multidisciplinary Majors	1.1%
Tuition Cost of Low Income Students (scorecard_netprice_2013)	1.1%
% of Business Majors	1.0%
% of Trade & Personal Majors	0.8%
% Public Social Majors	0.7%
Region	0.5%
Is a Tech School?	0.2%
Type of School	0.2%
Public or Private	0.1%
Historically Black College (HBCU)	0.0%

VI Key Takeaways & Analysis

While the dataset included a comprehensive set of features, it was not, by any means, an all-encompassing dataset. Some key additional missing features that could be considered for this study as it pertains to achieving upward mobility include:

- Teacher-Student Ratio
- Classroom Size
- Scholarship or Grant Opportunities Available
- Financial Aid Packages available to students (
- Descriptive information from Massive Open Online Courses (MooCs)
- Internship Opportunities / On-Campus Recruiting
- Percent of Students Matriculating to Graduate School
- Alumni Engagement
- Financial Debt upon Graduation (Liability side of the income statement)
- Other Global Universities & Institutions not included on list

Upward Mobility is a multifaceted concept that can be influenced by a wide variety of factors – a friend that helps with a job search, a course that paves the way for a new career, a teacher that gives a good recommendation, a scholarship that enables a student to start a job debt-free, an alumni connection, and the list goes on.

As with any dataset, there are exceptions to the rule and outliers that break the mold of convention. We use a SQL analysis technique to drill down into the data using GroupBy and SortBy calls to identify outliers and significant datapoints. Here are some interesting takeaways from the results:

Most Upwardly Mobile Institutions

The Saint Louis College of Pharmacy, which sees 91.9% of all students in the lowest quintile reaching the top quintile, ranks as the very top, most upwardly mobile institution. While this may be a surprise to many, 79% of all graduates end up in the Top 5% of Income and only 29.7% start in the top 20% quintile. Furthermore, the school accepts over 62% of all applicants, opening its doors to a wide array of applicants. The Massachusetts College of Pharmacy and Health Services (MCPHS) University takes the second spot with 91.3% of all lowest quintile incomers jumping to the top quintile within 10 years after graduation. Continuing the trend of Pharmacy schools, the Albany College of Pharmacy and Health Services ranks number 3 followed by Kettering University and Rose-Hulman Institute of Technology (Exhibit H). It's not until #12 that the Massachusetts Institute of Technology (MIT) that we find a Tier 1 school. The common theme to this list is a specialization of skills, most notably pharmacy and technology skillsets, that are valuable in building tangible knowledge to thrive in a tech-forward, progressive job market and economy.

Upward Mobility doesn't need to be Expensive

There are a handful of schools that offer high, upward mobility at a cheap price (Exhibit I). California Maritime Academy, Baruch College, California State Polytechnic University, New Mexico Institute of Mining & Technology and the University of Florida all offer a pathway to upward mobility at a tuition cost less than $7,000. The State University systems of California and New York are standout university systems across the nation, ranking highly amongst all public universities, which was also cited in the *Mobility Report Cards* study.

High Bang for the Buck = High Upward Mobility + Low Cost of Tuition

Following on the previous theme of providing upward mobility at a low cost, Maritime Academy and Colleges ranks near the top of the "Bang for the Buck" list (Exhibit K). We are using a minimum Upward Mobility Threshold of 1.25 as we're cognizant of very low cost tuition (<$2000) potentially impacting the Bang for the Buck ratio on the denominator side.

California Maritime Academy, Massachusetts Maritime Academy and SUNY Maritime College rank as the top universities on the Bang for the Buck list, a salute to Military academies.

Non-Selective Schools can also be Upwardly Mobile

While selective schools tend to yield the most talented students, non-selective schools with rejection rates below 20% (over 80% of applicants gain admittance) can also sport a high upward mobility score at a relatively low cost of tuition, driving a high 'Bang for the Buck' score (Exhibit J). Schools such as University of Mary Washington, the Citadel Military College, ITI Technical College and Iowa State University all accept a large number of applicants at a low Cost of Tuition while delivering a superb opportunity for Upward Mobility

Upward Mobility can occur in spite of Low Social Eliteness

While it does rate as an important factor in determining Upward Mobility, schools can have high upward mobility in spite of exhibiting low Social Eliteness Scores (Exhibit M). This is particularly the case when a college specializes in teaching a tangible skill such as pharmacy or technology. The Saint Louis College of Pharmacy, MCPHS University, New Jersey Institute of Technology, Illinois Institute of Technology and Milwaukee School of Engineering all rank highly in terms of Upward Mobility yet have Social Eliteness Scores lower than 0.9.

VII Conclusion

While achieving Upward Mobility may not be the ultimate objective of a college career, it undoubtedly plays a large role, consciously or subconsciously, in the decision-making process for prospective students. That all-important decision-making process has customarily relied upon school visits, brochures, pamphlets, word of mouth, reputation and the almighty *US News & World Report College Rankings*. For an additional fee to US News & World Report, prospective college students can view additional university data features such as Financial Aid, Campus Life, Test Scores, Clubs and even Alumni Salaries. There are many factors influencing a student's decision to attend a certain university and many students take a holistic approach. However, when it comes to determining which universities help their students achieve optimal upward mobility, there has been no definitive source for this answer. We have published this report with the explicit intention to offer prospective students a detailed analysis on Upward Mobility as an alternative to traditional rankings. The data reveals many institutions that perhaps 'fly under the radar' compared to mainstream elite schools but still perform an admirable job of helping lower income students transcend to the upper classes. We intend to maintain public access to this data from our GitHub Account, enabling any prospective student to query the results.

The College Board, which administers SAT Exams, made news recently in May 2019 by announcing that all test takers would be assigned an "adversity index" to help admissions officers "place students' SAT scores in the context of their socioeconomic advantages or disadvantages." This move may be a good first step forward in limiting the systemic bias for admitting a disproportionate share of wealthy students into the most upwardly mobile universities. Perhaps the key takeaways from the *Mobility Report Cards* study are being discussed within the ivory towers of the most elite institutions, shining a light into socioeconomically biased admissions practices. While not a fail-safe method as there are sure to many loopholes around this 'adversity index' score, it directly addresses a critical ailment of the admissions process and empowers admissions officers to make more informed judgments.

Do colleges perpetuate the growing trend of Inequality by self-selecting already-socially elite students on a disproportionate basis? There is irrefutable evidence suggesting that the university system as a whole, actually acts as an accelerant towards advancing economic inequality, contrary to popular opinion. What is abundantly clear is that the socioeconomic composition of the student body has a profound effect on determining its relative upward mobility score, moreso than almost any other factor. Thus, limiting access to these elite institutions to mainly the wealthy further perpetuates the wealth inequality gap, as the data suggests. In this vein, the popular view of college being the ticket to success should come with a caveat – it acts as a mobility booster for those lucky to attend upwardly mobile schools and acts as a retardant for others that take out unsurmountable loans to attend an institution that doesn't serve its students justice in terms of upward mobility. Thus, the university system may actually serve as ground zero for the 'hollowing out of the middle class' that has transpired in America, counterintuitively during a time that college attendance and graduation rates are near all-time highs. The discerning story says that upwardly mobile institutions selecting primarily the social elite propagates the widening divide unless corrective measures are made to amend current admissions policies. It is self-evident that the 'Equality of Access to Upward Mobility' through university education is anything but equal in its current state.

The question then becomes, why do schools tend to admit students of higher income demographic backgrounds? Endowments and fundraising may provide one answer as schools and universities are cognizant of the need to increase funds. Legacy admissions in which children of alumni receive preferential treatment in admissions may provide another clue. For-profit universities can be fraught with misguided intentions and misaligned incentives, biasing admissions officers to admit wealthier students. Placing a limit or cap on the number of students that may be admitted from the top quintile (say 50%), may create a system whereby the wealthy applicants are competing against primarily wealthy applicants for entry while leaving more space for lower income quintile applicants. This could create an income-based, multi-tiered admissions process, diversifying the socioeconomic make-up of the student body and, perhaps, slowing the wealth inequality gap.

By utilizing data science and machine learning models, particularly Random Forest Classification and Regression Models, we identify high-achieving schools on our Upward Mobility scale and uncover the driving attributes leading to these scores. While the *Mobility Report Cards* focuses on 'access + mobility', we isolate and define mobility as a weighted summation of the conditional probability of jumping from a lower to higher income quintile. We identify outliers and superstar institutions that offer promising alternatives to the 'pearly gates' of elite institutions. Even more, these institutions – especially those with low costs of tuition - may merit further inspection to ascertain what they are doing correctly and a possible emulation of their best practices. Trade, technical and military colleges offer upward advancement at rates on par, if not better, than Ivy League institutions. This is irrespective of tuition costs, which are more favorable at trade, technical and military colleges anyway. While the rising student loan crisis grows louder with public calls for debt forgiveness, perhaps students and government should be more discerning in selecting which schools do justice to providing upward mobility and are worth the price of admission. Students should make informed decisions on their college selection lest they graduate with poor job prospects along with untenable student debt. While simply graduating will improve the chances of upward mobility, the most important decision a student makes is deciding which institution to attend. We hope this study and analysis provides all prospective students a keen insight filled with objective data to maximize that decision.

APPENDIX

Appendix

Exhibit A. Stand out 2-year private and public non-profit universities (Tier 9 & 11):

Name	Tier	Tuition Sticker Price (2013)	Upward Mobility Score	Social Eliteness Score	Bang for the Buck
Perry Technical Institute	9	$9,327	1.3329	0.6748	1.43
Advanced Institute Of Hair Design	11	$6,241	1.3113	0.7561	2.10
ITI Technical College	11	$10,575	1.2562	0.5178	1.19
Crimson Technical College	11	$7,740	1.1629	0.2901	1.50
Pittsburgh Institute Of Aeronautics	9	$19,125	1.1145	0.5773	0.58
North Dakota State College Of Science	9	$4,438	1.0704	0.4397	2.41
Hesston College	9	$24,214	0.9685	0.6462	0.40
Universal Technical Institute of Rancho Cucamonga, CA	11	$7,368	0.9643	0.5267	1.31
Mitchell Technical Institute	9	$5,072	0.9617	0.4464	1.90
Ohlone College	9	$1,162	0.9410	0.9048	8.10
West Valley-Mission Community College District	9	$1,180	0.9389	0.8969	7.96
Southeast Technical Institute	9	$4,728	0.9301	0.5349	1.97
Northern Virginia Community College	9	$4,853	0.9241	0.8466	1.90
Linn State Technical College	9	$5,820	0.9056	0.4794	1.56

Exhibit B. Scatterplot between Rejection Rates vs. Upward Mobility Scores (41.0% Correlation)

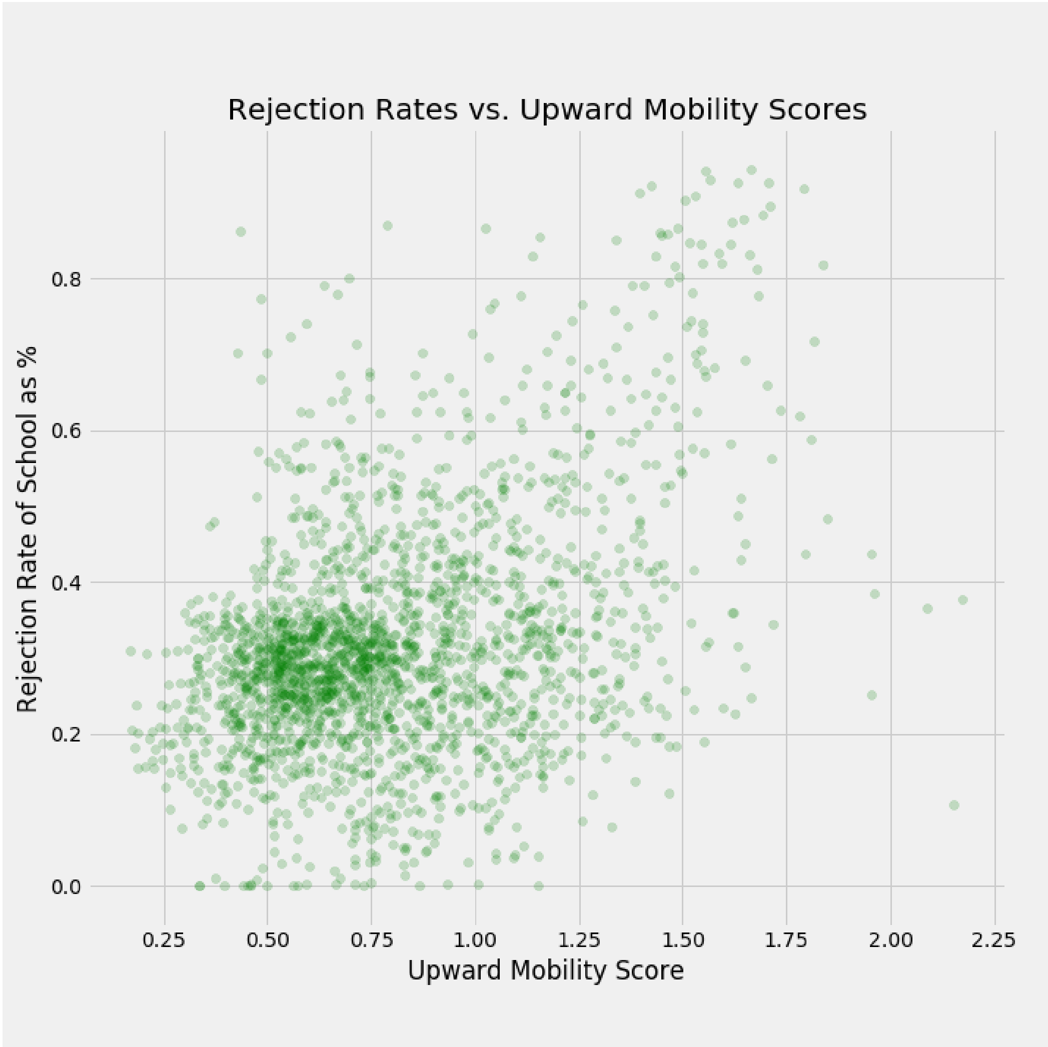


Exhibit C. Upward Mobility Score by Region

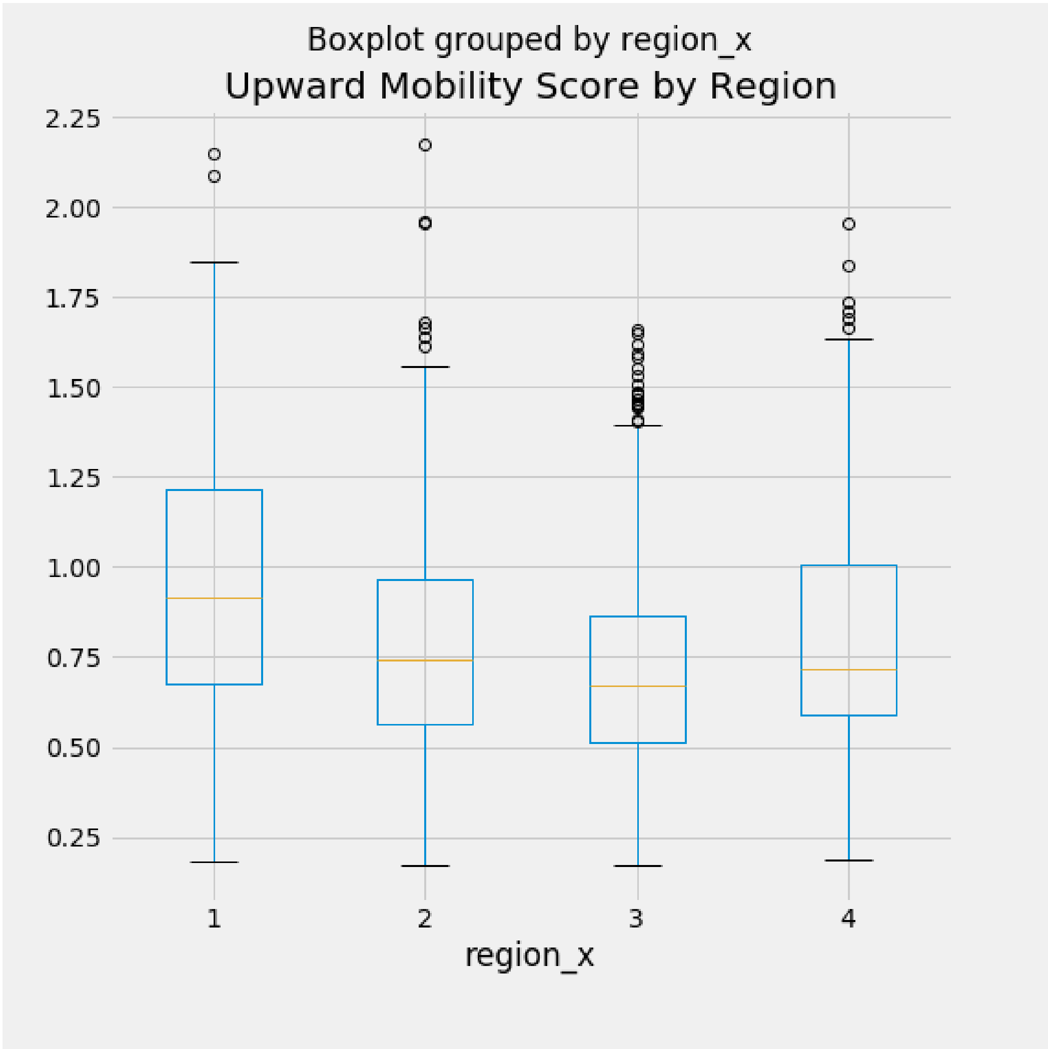


Exhibit D. Scatterplot between Graduation Rates vs. Upward Mobility Scores (68.8% Correlation)

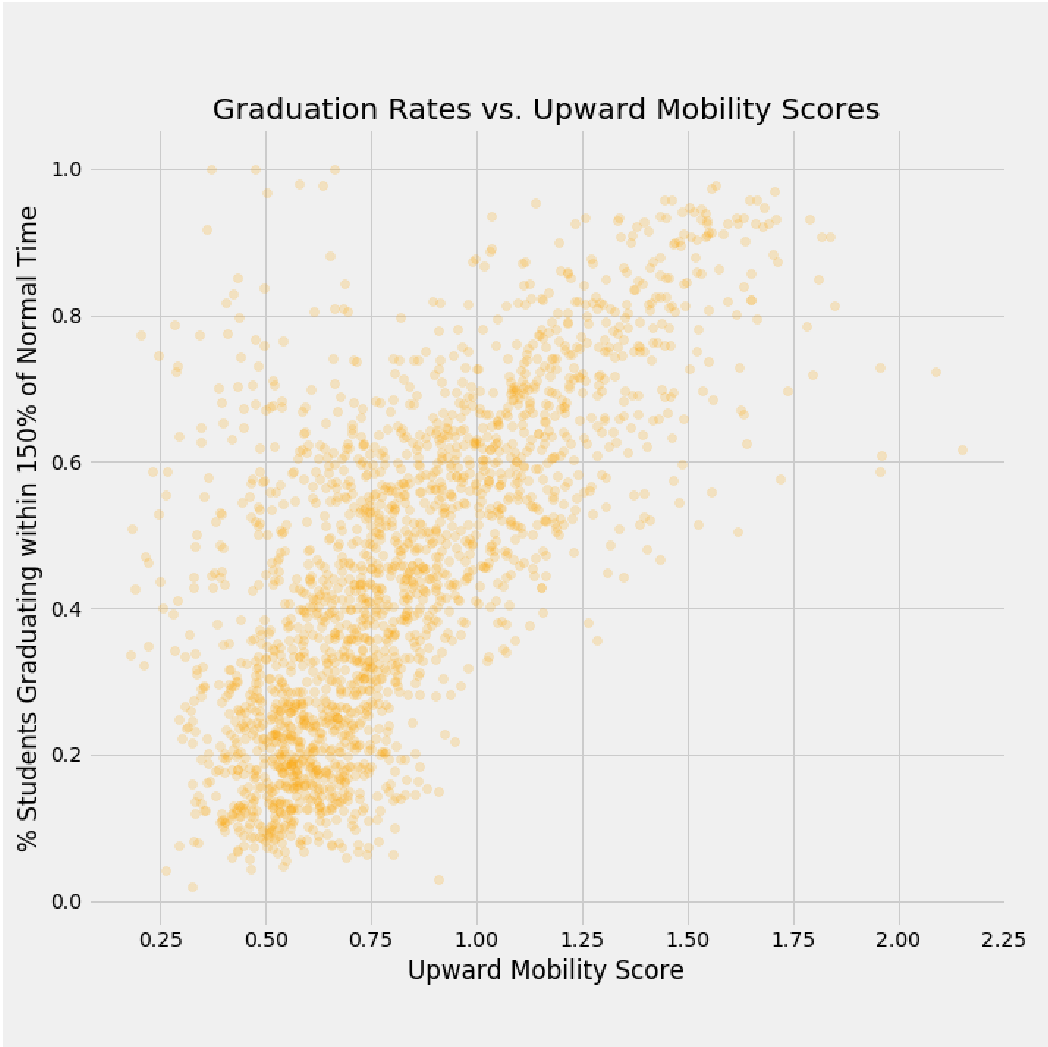


Exhibit E. Tuition Sticker Price vs. Upward Mobility Score

Exhibit F. Barron's Selectivity Index Score vs. Upward Mobility Score

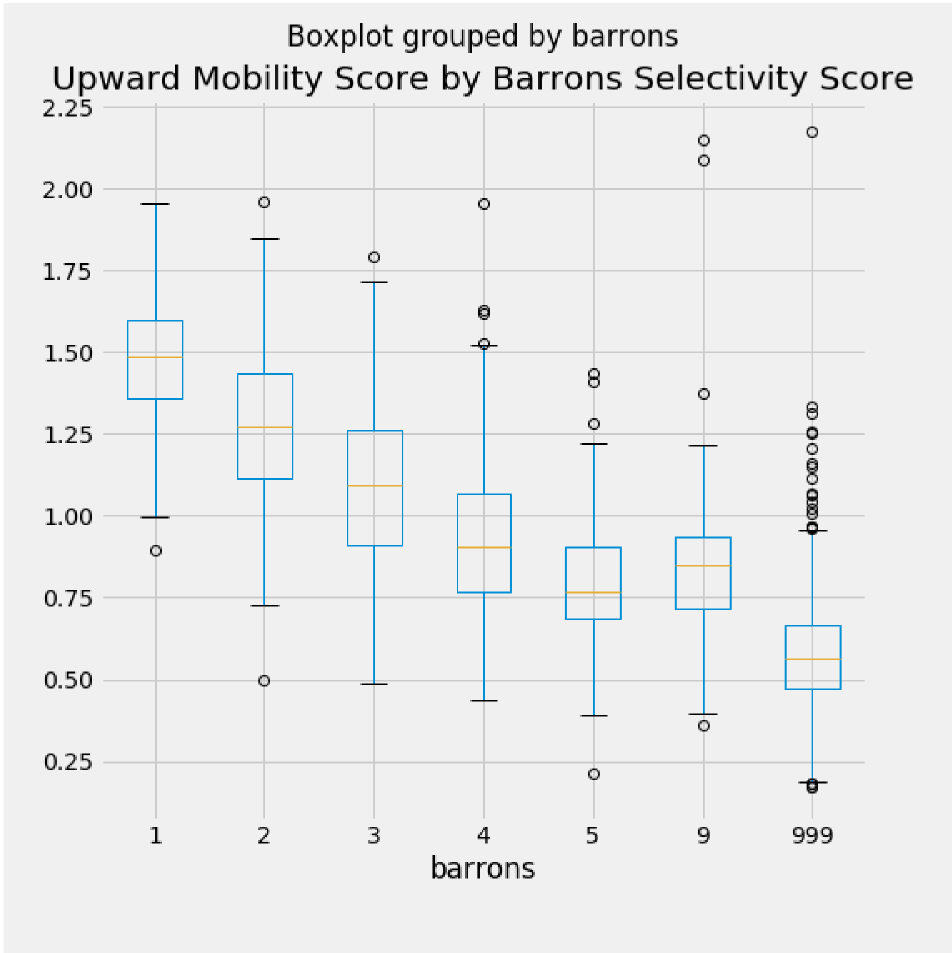


Exhibit G. Top-Rated Social Elite Universities ranked by Social Eliteness Score (2013)

	University/Institution	Social Eliteness Score	Upward Mobility Score	Cost of Tuition
1	Washington And Lee University	1.664	1.481	$ 45,617
2	Princeton University	1.586	1.706	$ 41,820
3	Davidson College	1.582	1.234	$ 45,377
4	Middlebury College	1.558	1.434	$ 45,957
5	Colby College	1.555	1.547	$ 47,350
6	University Of Richmond	1.552	1.306	$ 46,680
7	Wake Forest University	1.540	1.411	$ 46,200
8	Brown University	1.536	1.531	$ 47,434

9	Colgate University	1.531	1.367	$ 48,175
10	University Of Notre Dame	1.526	1.682	$ 46,237
11	Yale University	1.521	1.567	$ 45,800
12	Washington University In St. Louis	1.516	1.544	$ 46,467
13	Georgetown University	1.510	1.662	$ 46,744
14	Duke University	1.501	1.487	$ 47,243
15	Yeshiva University	1.497	1.223	$ 38,730
16	Bates College	1.489	1.336	$ 47,030
17	Dartmouth College	1.489	1.504	$ 48,108
18	Vanderbilt University	1.484	1.618	$ 43,838
19	Tufts University	1.482	1.680	$ 48,643
20	Harvard University	1.471	1.556	$ 43,938
21	Villanova University	1.470	1.638	$ 45,966
22	Kenyon College	1.467	1.034	$ 47,330
23	Williams College	1.466	1.139	$ 48,310
24	College Of The Holy Cross	1.464	1.556	$ 45,692
25	Trinity College of Hartford, CT	1.463	1.574	$ 49,056

Exhibit H. Most Upwardly Mobile Institutions (regardless of criteria)

	Name	Tier	Upward Mobility Score	Rejection Rate	Cost of Tuition	Bang for the Buck
1	Saint Louis College Of Pharmacy	8	2.1737	0.3777	$ 27,554	0.7889
2	MCPHS University	8	2.1513	0.1079	$ 29,500	0.7293
3	Albany College Of Pharmacy And Health Sciences	8	2.0897	0.3660	$ 30,131	0.6935
4	Kettering University	4	1.9599	0.3841	$ 36,980	0.5300
5	Rose - Hulman Institute Of Technology	2	1.9555	0.4378	$ 41,283	0.4737
6	California Maritime Academy	5	1.9550	0.2520	$ 6,536	2.9912
7	Worcester Polytechnic Institute	4	1.8471	0.4841	$ 44,222	0.4177
8	Harvey Mudd College	2	1.8379	0.8183	$ 48,594	0.3782
9	Babson College	4	1.8162	0.7179	$ 45,120	0.4025
10	Rensselaer Polytechnic Institute	2	1.8091	0.5880	$ 47,908	0.3776
11	University Of The Sciences In Philadelphia	6	1.7947	0.4365	$ 36,096	0.4972
12	Massachusetts Institute Of Technology	1	1.7910	0.9185	$ 45,016	0.3979
13	Stevens Institute Of Technology	4	1.7808	0.6184	$ 45,366	0.3925

14	Colorado School Of Mines	3	1.7360	0.6257	$ 16,918	1.0261
15	New Jersey Institute Of Technology	5	1.7185	0.3453	$ 15,648	1.0983
16	Bentley University	4	1.7139	0.5621	$ 42,511	0.4032
17	California Institute Of Technology	2	1.7093	0.8945	$ 43,362	0.3942
18	Princeton University	1	1.7059	0.9259	$ 41,820	0.4079
19	Lafayette College	2	1.7018	0.6586	$ 45,635	0.3729
20	Claremont Mckenna College	2	1.6921	0.8827	$ 47,395	0.3570
21	University Of Notre Dame	2	1.6822	0.7770	$ 46,237	0.3638
22	Tufts University	2	1.6795	0.8116	$ 48,643	0.3453
23	Stanford University	1	1.6641	0.9431	$ 44,757	0.3718
24	Saint John's University of Collegeville, MN	6	1.6637	0.2484	$ 38,704	0.4298
25	Georgetown University	2	1.6618	0.8303	$ 46,744	0.3555

Exhibit I. Most Upwardly Mobile Schools with Tuition Cost Under <$7,000

	Name	Cost of Tuition	Upward Mobility Score	Bang for the Buck
1	California Maritime Academy	$ 6,536	1.9550	2.9912
2	CUNY Bernard M. Baruch College	$ 6,561	1.4262	2.1738
3	California State Polytechnic University, Pomona	$ 6,904	1.3950	2.0206
4	New Mexico Institute Of Mining & Technology	$ 6,256	1.3493	2.1568
5	University Of Florida	$ 6,313	1.3467	2.1332
6	Advanced Institute Of Hair Design	$ 6,241	1.3113	2.1012
7	Southern Polytechnic State University	$ 5,840	1.2865	2.2030
8	San Diego State University	$ 6,866	1.2726	1.8535
9	California State University, East Bay	$ 6,564	1.2629	1.9239
10	California State University, Fullerton	$ 6,315	1.2331	1.9526
11	City College Of New York - CUNY	$ 6,389	1.2312	1.9271
12	California State University, Long Beach	$ 6,452	1.2285	1.9040
13	University Of Central Florida	$ 6,368	1.1994	1.8834
14	Florida International University	$ 6,497	1.1858	1.8251
15	Thomas A Edison State College	$ 5,871	1.1795	2.0091
16	CUNY John Jay College Of Criminal Justice	$ 6,359	1.1727	1.8442
17	CUNY Hunter College	$ 6,429	1.1720	1.8230

18	CUNY Queens College	$ 6,507	1.1663	1.7923
19	California State University, Bakersfield	$ 6,792	1.1576	1.7043
20	Florida State University	$ 6,507	1.1557	1.7760
21	California State University - Sacramento	$ 6,648	1.1524	1.7334
22	San Francisco State University	$ 6,468	1.1316	1.7496
23	CUNY Brooklyn College	$ 6,536	1.1231	1.7183
24	California State University, Northridge	$ 6,549	1.1011	1.6813
25	California State University, Los Angeles	$ 6,348	1.0904	1.7177

Exhibit J. Non-Selective Schools that are also Upwardly Mobile (Rejection Rate < 20%)

	Name	Rejection Rate	Bang for the Buck	Upward Mobility Score	Cost of Tuition
1	University Of Mary Washington	19.0%	1.5042	1.5511	$ 10,312
2	Citadel, The Military College Of South Carolina	19.6%	1.3015	1.4444	$ 11,098
3	ITI Technical College	17.4%	1.1879	1.2562	$ 10,575
4	Iowa State University Of Science & Technology	17.5%	1.5691	1.2131	$ 7,731
5	North Dakota State University - Main Campus - ...	15.9%	1.5197	1.1884	$ 7,820
6	University Of Iowa	19.8%	1.4491	1.1707	$ 8,079
7	Crimson Technical College	15.8%	1.5025	1.1629	$ 7,740
8	University Of Northern Iowa	16.7%	1.4953	1.1587	$ 7,749
9	Sonoma State University	18.1%	1.5653	1.1389	$ 7,276
10	Auburn University	17.3%	1.1049	1.1270	$ 10,200
11	University Of Colorado System	14.8%	1.1766	1.1222	$ 9,538
12	Shippensburg University Of Pennsylvania	16.6%	1.1427	1.1169	$ 9,774
13	University Of Oklahoma	19.6%	1.4505	1.1086	$ 7,643
14	Western Washington University	16.3%	1.2279	1.1008	$ 8,965
15	Kansas State University	4.1%	1.2121	1.0950	$ 9,034
16	University Of Nevada , Reno	16.0%	1.6098	1.0797	$ 6,707
17	Western Michigan University	17.2%	1.0046	1.0734	$ 10,685
18	Bloomsburg University Of Pennsylvania	11.4%	1.1998	1.0695	$ 8,914
19	University Of Wyoming	4.3%	2.6468	1.0502	$ 3,968
20	University Of Texas Of The Permian Basin	16.0%	1.9908	1.0452	$ 5,250
21	Montana Tech Of The University Of Montana	11.1%	1.8620	1.0404	$ 5,588

22	University Of Alabama In Huntsville	19.4%	1.1280	1.0330	$ 9,158
23	University Of Kansas	7.7%	1.0626	1.0315	$ 9,707
24	Texas A&M University - Corpus Christi	7.4%	1.3680	1.0303	$ 7,531
25	Valley City State University	17.1%	1.5408	1.0284	$ 6,674

Exhibit K. High Upward Mobility + Low Cost of Tuition = High Bang for the Buck Schools

	Name	Bang for the Buck	Upward Mobility Score	Cost of Tuition	Social Eliteness Score
1	California Maritime Academy	2.9912	1.9550	$ 6,536	1.0923
2	Massachusetts Maritime Academy	2.2445	1.6272	$ 7,250	1.0764
3	SUNY Maritime College	2.1738	1.6186	$ 7,446	1.0783
4	University Of Mary Washington	1.5042	1.5511	$ 10,312	1.3796
5	Binghamton University	1.7978	1.5497	$ 8,620	1.0286
6	California Polytechnic State University	1.7210	1.5347	$ 8,918	1.2514
7	University Of Florida	2.1332	1.3467	$ 6,313	1.0707
8	SUNY College At Geneseo	1.7020	1.3232	$ 7,774	1.1010
9	San Diego State University	1.8535	1.2726	$ 6,866	1.0014
10	Florida State University	1.7760	1.1557	$ 6,507	1.0141
11	Sonoma State University	1.5653	1.1389	$ 7,276	1.1437
12	California State University, Chico	1.5706	1.0997	$ 7,002	1.1214
13	University Of Nevada , Reno	1.6098	1.0797	$ 6,707	1.0493
14	Brigham Young University	1.8229	0.9114	$ 5,000	1.2554
15	Cascadia Community College	2.2674	0.8600	$ 3,793	1.0007
16	Raritan Valley Community College	1.7787	0.7926	$ 4,456	1.0169
17	Las Positas College	6.7913	0.7728	$ 1,138	1.0882
18	Saddleback College	6.1804	0.7058	$ 1,142	1.0014

Exhibit L. Best Overall Bang for the Buck Schools (Minimum Upward Mobility Score > 1.0)

	Name	Tier	Upward Mobility Score	Cost of Tuition	Bang for the Buck
1	California Maritime Academy	5	1.9550	$ 6,536	2.9912
2	University Of Wyoming	5	1.0502	$ 3,968	2.6468
3	North Dakota State College Of Science	9	1.0704	$ 4,438	2.4118
4	Massachusetts Maritime Academy	5	1.6272	$ 7,250	2.2445
5	Southern Polytechnic State University	5	1.2865	$ 5,840	2.2030
6	CUNY Bernard M. Baruch College	5	1.4262	$ 6,561	2.1738
7	SUNY Maritime College	5	1.6186	$ 7,446	2.1738
8	New Mexico Institute Of Mining & Technology	3	1.3493	$ 6,256	2.1568
9	University Of Florida	3	1.3467	$ 6,313	2.1332
10	Advanced Institute Of Hair Design	11	1.3113	$ 6,241	2.1012
11	California State Polytechnic University, Pomona	5	1.3950	$ 6,904	2.0206
12	Thomas A Edison State College	7	1.1795	$ 5,871	2.0091
13	University Of Texas Of The Permian Basin	7	1.0452	$ 5,250	1.9908
14	California State University, Fullerton	5	1.2331	$ 6,315	1.9526
15	City College Of New York - CUNY	5	1.2312	$ 6,389	1.9271
16	California State University, East Bay	5	1.2629	$ 6,564	1.9239
17	San Jose State University	5	1.4037	$ 7,323	1.9168
18	California State University, Long Beach	5	1.2285	$ 6,452	1.9040
19	University Of Central Florida	5	1.1994	$ 6,368	1.8834
20	Montana Tech Of The University Of Montana	5	1.0404	$ 5,588	1.8620
21	San Diego State University	5	1.2726	$ 6,866	1.8535
22	CUNY John Jay College Of Criminal Justice	5	1.1727	$ 6,359	1.8442
23	Florida International University	5	1.1858	$ 6,497	1.8251
24	CUNY Hunter College	5	1.1720	$ 6,429	1.8230
25	Binghamton University	3	1.5497	$ 8,620	1.7978

Exhibit M. High Upward Mobility (> 1.0) despite Low Social Eliteness Scores (< 0.9)

	Name	Social Eliteness Score	Upward Mobility Score	Cost of Tuition	Bang for Buck
1	Saint Louis College Of Pharmacy	0.8619	2.1737	$ 27,554	0.789
2	MCPHS University	0.7641	2.1513	$ 29,500	0.729
3	New Jersey Institute Of Technology	0.8230	1.7185	$ 15,648	1.098
4	Illinois Institute Of Technology	0.8726	1.6409	$ 40,052	0.410
5	Milwaukee School Of Engineering	0.8998	1.5553	$ 35,520	0.438
6	Pace University	0.6836	1.5263	$ 39,697	0.384
7	Maine Maritime Academy	0.8210	1.5199	$ 12,400	1.226
8	State University Of New York At Stony Brook	0.7230	1.4891	$ 8,430	1.766
9	Florida Institute Of Technology	0.8773	1.4803	$ 37,990	0.390
10	CUNY Bernard M. Baruch College	0.3940	1.4262	$ 6,561	2.174
11	Saint Francis College	0.6595	1.4092	$ 22,300	0.632
12	San Jose State University	0.8962	1.4037	$ 7,323	1.917
13	Saint John's University of Queens, NY	0.6768	1.3960	$ 38,892	0.359
14	California State Polytechnic University, Pomona	0.7810	1.3950	$ 6,904	2.021
15	Vaughn College Of Aeronautics And Technology	0.2080	1.3723	$ 21,642	0.634
16	New Mexico Institute Of Mining & Technology	0.8692	1.3493	$ 6,256	2.157
17	College Of Saint Scholastica	0.8415	1.3471	$ 32,842	0.410
18	Trine University	0.8012	1.3368	$ 25,902	0.516
19	Perry Technical Institute	0.6748	1.3329	$ 9,327	1.429
20	University Of Mary	0.7382	1.3158	$ 15,665	0.840
21	Advanced Institute Of Hair Design	0.7561	1.3113	$ 6,241	2.101
22	Wilkes University	0.7888	1.3067	$ 31,262	0.418
23	King's College of Wilkes-Barre, PA	0.8963	1.3031	$ 31,816	0.410
24	University Of California, Riverside	0.7560	1.3014	$ 13,407	0.971

Exhibit N. Social Eliteness & Rejection Rate affect on Upward Mobility Score (Multi-dimensional Scatterplot)

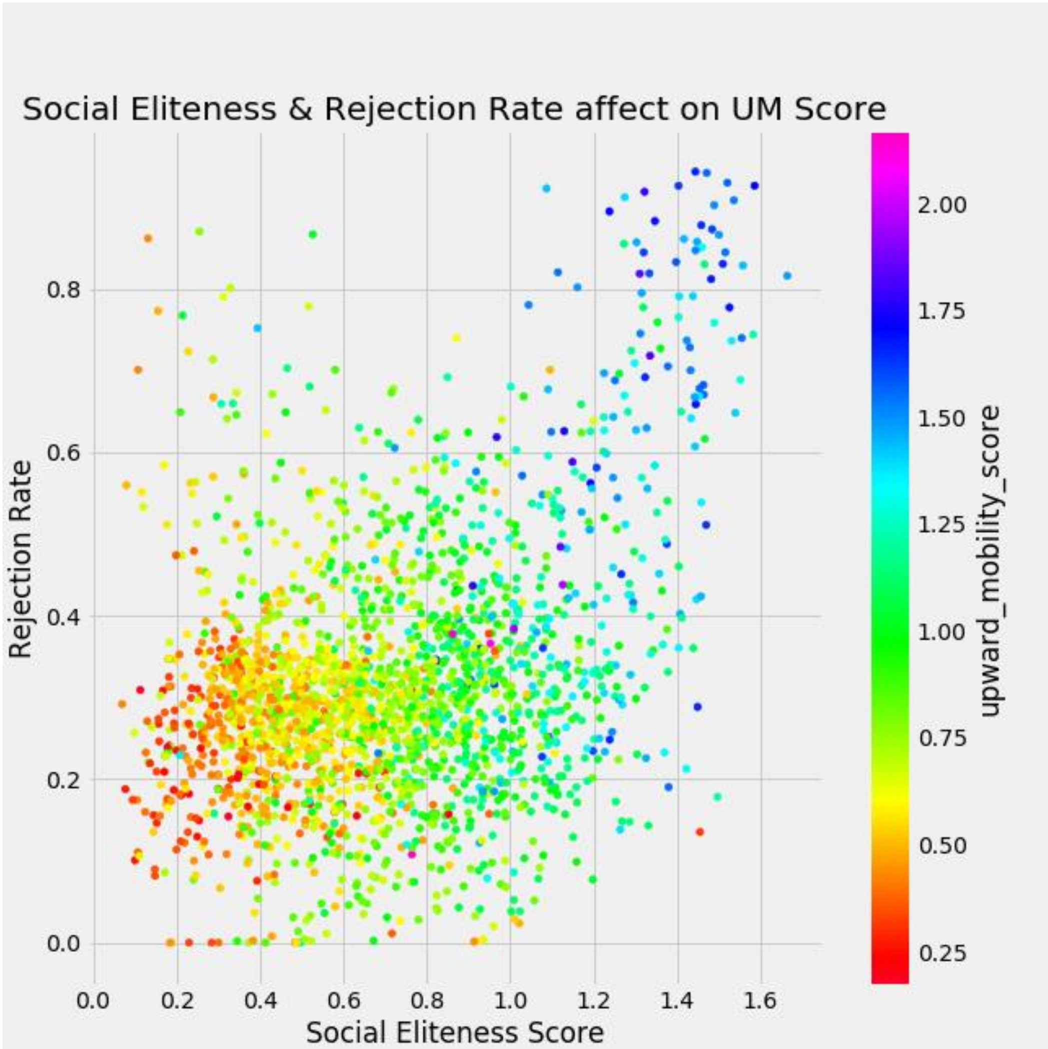


Exhibit O. Cost of Tuition & Graduation Rate on Upward Mobility Score (Multi-dimensional Scatterplot)

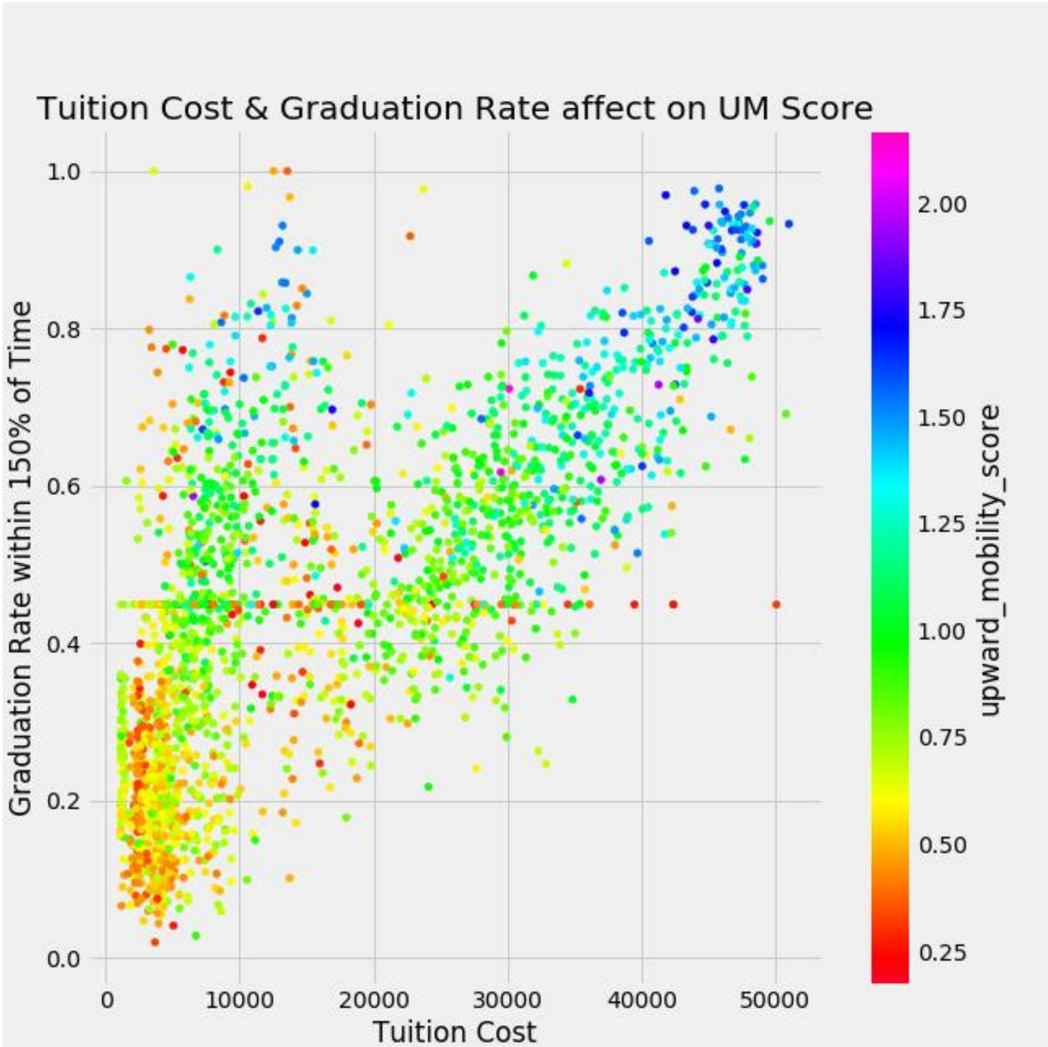


Exhibit P. Bang for the Buck & Social Eliteness effect on Upward Mobility Score (Multidimensional Scatterplot)

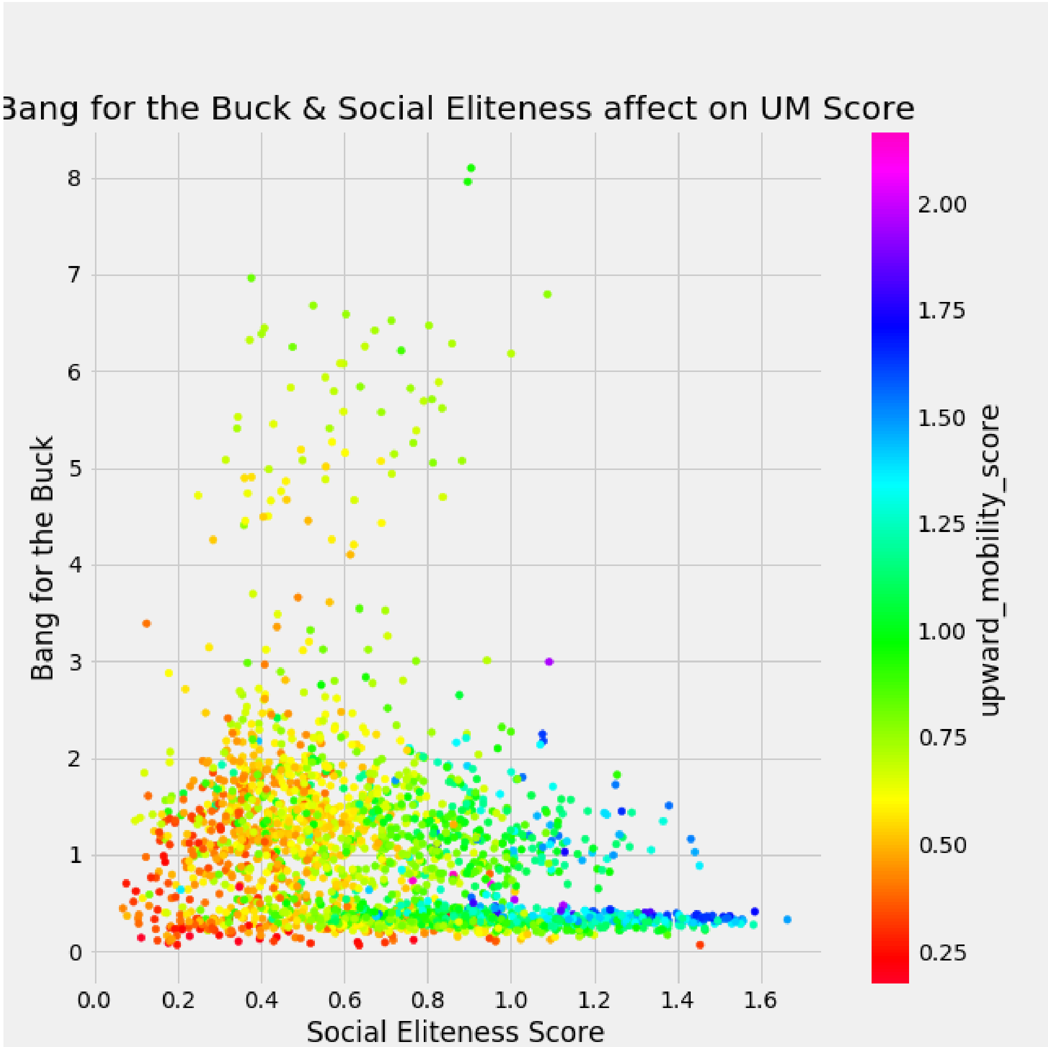
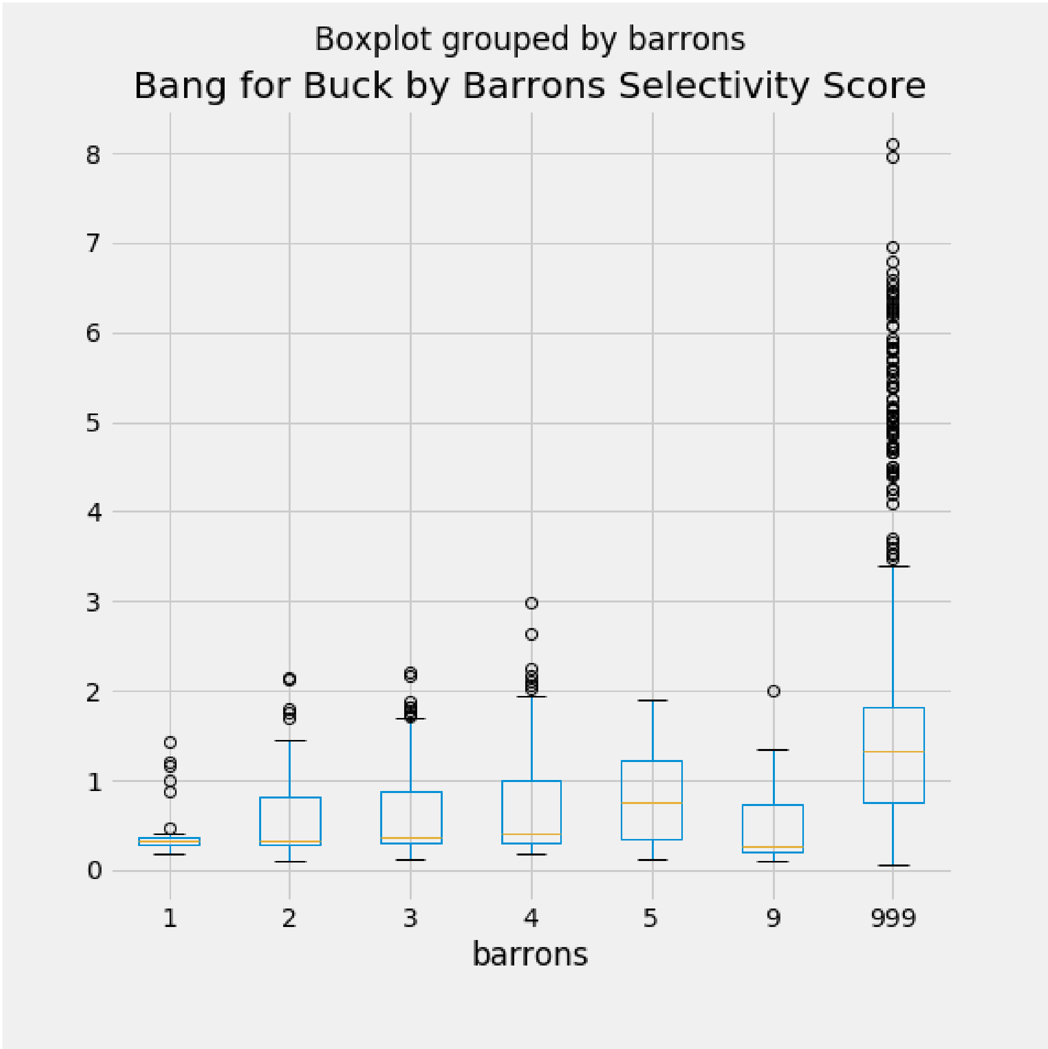


Exhibit Q. Bang for the Buck by Barron's Selectivity Score



Further Information & Resources

Git Hub Repository (Public Access):

https://github.com/SenorTodoPositivo/GA-FINAL-PROJECT---Equality-of-Upward-Mobility-Project.git

Equality of Opportunity Model (Jupyter Notebook):

Equality of Opportunity Project.ipynb

Equality of Opportunity Outliers (SQL):

Equality of Opportunity - Identifying Stars & Outliers.ipynb

Contact Information about Author:

Hans Kullberg
Open Dreams Organization
hans@open-dreams.org