Impact of socio-economic factors on US College Admissions*

Saumya Bakshi

30 April 2022

Abstract

Colleges and Universities allow incoming students to explore and build on interests and passions. In this paper we look at effects of socio-economic factors on a colleges acceptance rate as it may influence the decisions of future applicants. We look at variables such as average SAT score, median family income, gender, age, average faculty salary.

1 Introduction

The value of college education is prime in a students academic life. College helps explore and build on ones interests and passions and college admissions can be rather tricky to navigate. Although a colleges admission rates are a focal point for high school students applying for colleges, it is imperative to explore what socio-economic factors may play into the admission of a candidate. It may also guide future applicants in their decision to apply for colleges or universities. This is explored using the US Department of Education College Scorecard Database, which contains information on over 1000 colleges and universities across the United States.

The tool is designed to weigh the students priorities when applying for colleges and providing a ranking on the basis of the same. Take for example a student that values the universities graduation rate, in this case, the tool will provide a higher ranking for the Prestigious IVY League Universities such as Harvard, Princeton, and Yale. Another student might find post-graduation earnings and average annual cost to be more salient, and for such students the tool results in a higher ranking for institutions such as State University of New York Downstate Medical Center, Kentucky's Berea College, and California's Samuel Merritt University. However, the underlying dataset that fuels the tool can also be used to determine what factors affect admission rates for a US College.

The main aim of this paper is to understand the factors that might affect admission rates in US Colleges. In order to do so, we look at variables such as average SAT score, median family income, gender, age. An applicant with a higher amount of disposable income as a measure of their median family income might be predisposed to a higher SAT score due to external coaching and may also have the ability to pay a higher tuition. Such applicants might be able to apply to more competitive schools with a lower admission rate. A college or university with working alumni base might be considered to create employable graduates. Such colleges might be considered more competitive and would have a lower admission rate.

For this report, I will be using the open access data provided by the US Department of Education College Scorecard Database. The dataset will be downloaded, processed, and analyzed in `R` (R Core Team 2020) primarily using the `dplyr` (Wickham et al. 2021) and `Tidyverse` (Wickham et al. 2019) packages. Figures and tables will be created with `ggplot2` (**citeGgplot2?**). The packages `knitr` (Xie 2014) are used to generate the R markdown report.

*Code available: https://github.com/Saumya510/Final-Project

2 Data

2.1 Data Source

This report uses the data from US Department of Education College Scorecard resources updated by the United States Government (Department of Education) (**citeOpenDataPortal?**). This data has been provided by the government for high school students to make an informed decision about their college choices on the basis of factors that are important to them. This tool pools data points that detail factors such as the student body's demographic and socio economic information, the students family demographic and and socio-economic information, the post graduate students' average salary, the schools neighborhood. The primary dataset used in this report, was extracted from the US Department of Education College Scorecard Portal by downloading this file and it contains all the reported information dated April 10, 2022.

2.2 Data Structure

The data contains 2989 categorical and continuous variables containing the above mentioned information for 6694 institutions. In order to analyse our variables of interest, we manually narrowed down the number of columns from 2989 to 42. The also helped create a dataset that can be analysed efficiently and allows for the sharing of resources. We removed any colleges that did not report any admission rate as that was our main factor of interest.

2.3 Data Analysis

For the purpose of this paper, we studied 5 variables namely SAT Average of Applicants, Age group of Applicants, Gender of Applicants, Tuition of the Institution and Median Household Income of the Applicants. These are all imperative factors in determining a high school applicant's preferred college. We have plotted scatter plots for each to see a visual representation of the changes in admission rate on the basis of these factors

We note that there is a stark negative correlation between the SAT average and the admission rate. As the college's student body's SAT Average increases, the university becomes increasingly competitive with a decreasing admission rate. Next we examine the applicants age and gender. The average age of entry for a college does not impact its admission rate. We do note that the average age of entry for a college is between 19 and 22. We also note that the percentage of female students in the college or the gender distribution of a college does not impact the admission rate. Having said that we also want to point out that we see a significant skewness in the data points. This implies that there are a significant number of colleges in the US that have a higher percentage of female.

We also note a negative correlation between the average faculty salary and the admission rate. As the college's average faculty salary increases, the university would be more focused at the quality of education provided at the institution. This would imply that the high quality of expectation would result in a decreasing admission rate. Finally, we examine the Median Household income of the attendees of the university and we note that the there is a significant correlation between the Median Household income and the admission rate of a college.

3 Figures

Figure 1: SAT Average Score VS Admission rate

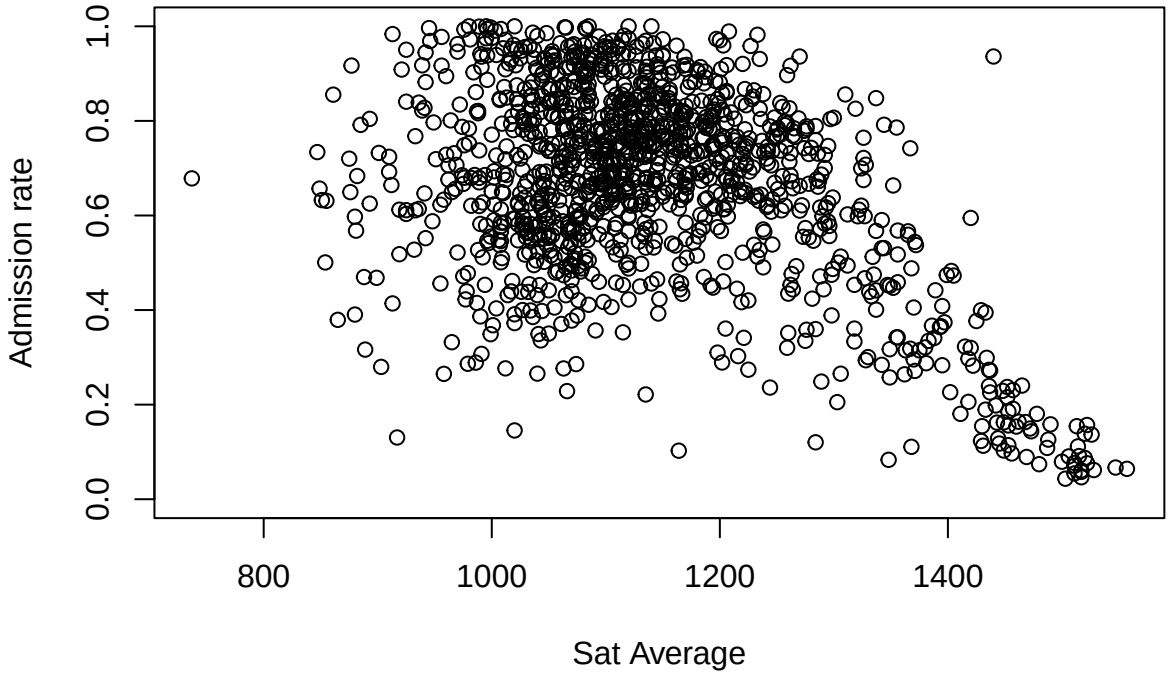


Figure 2: Average Age Entry VS Admission rate

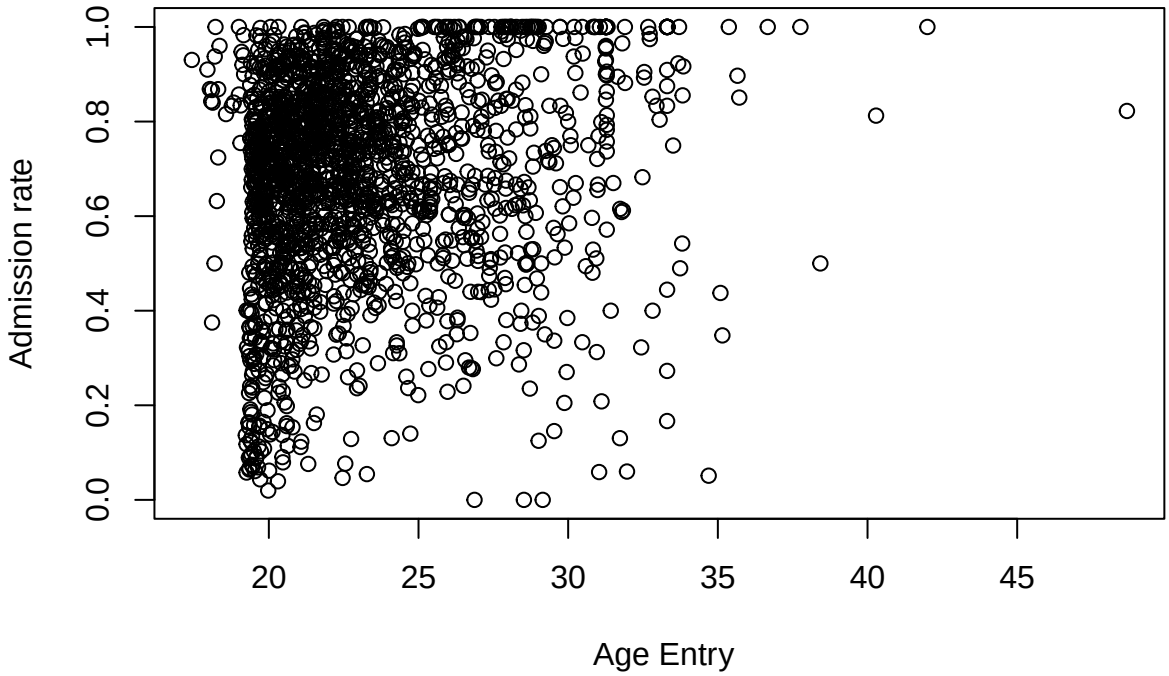


Figure 3: Percentage Female Population VS Admission rate

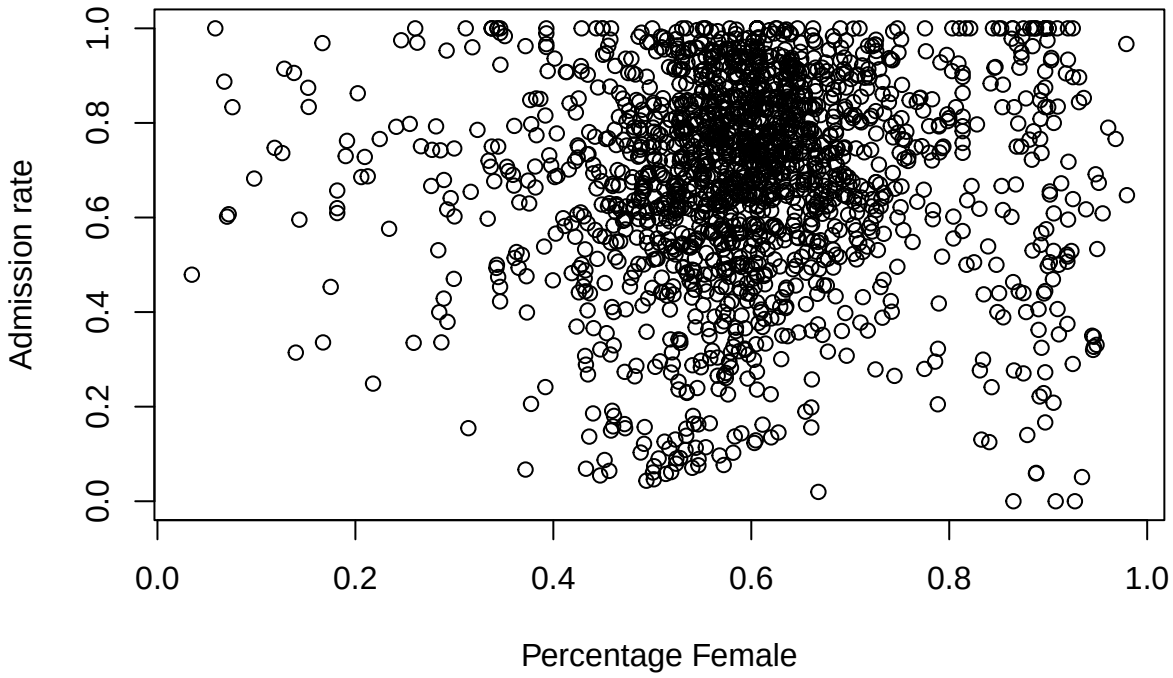


Figure 4: Average Faculty Salary VS Admission rate

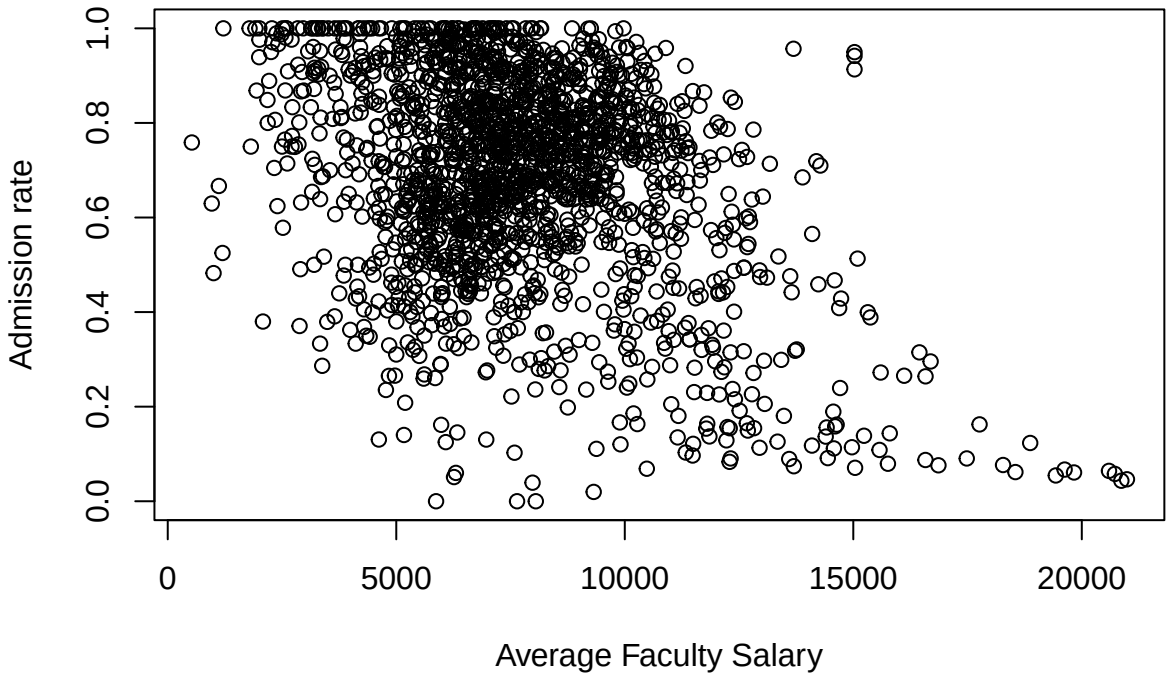
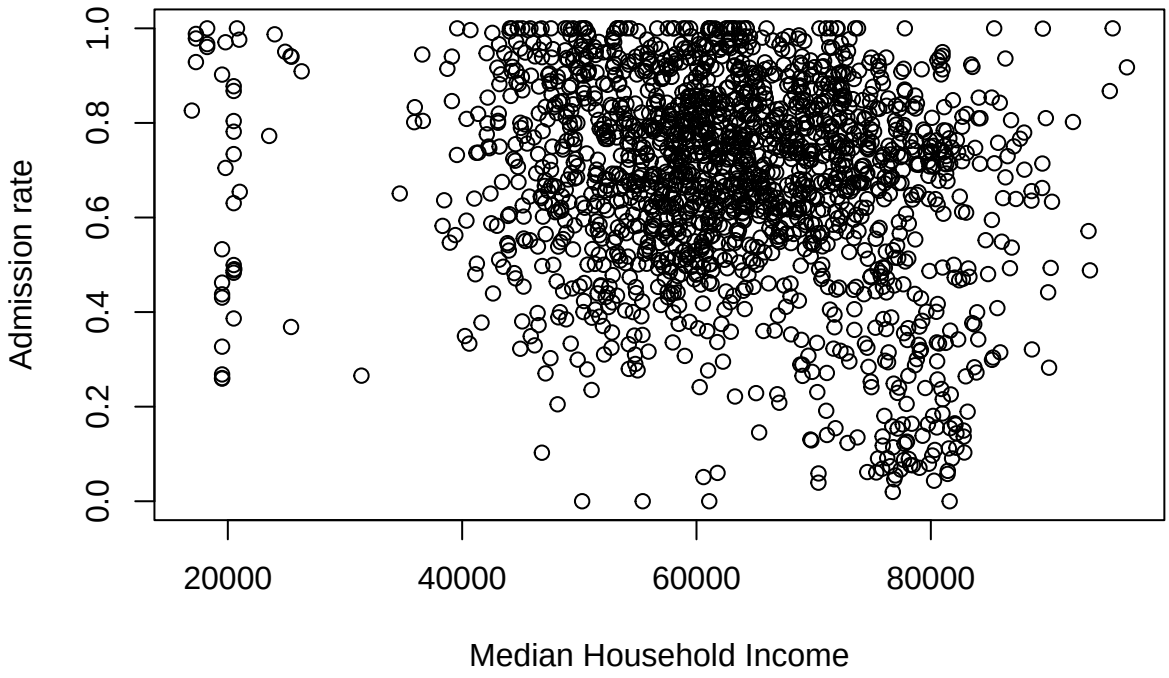


Figure 5: Median Household Income VS Admission rate



4 Results

In our analysis, we noted that there are significant impacts of Average SAT Scores and Average Faculty Salary on the admission rates. However factors such as average age of entry and gender do not have impact on the admission rate of the college.

5 Discussion

In this study, I discussed that SAT Scores and Average Faculty Salary have an impact on the admission rate in admission rates in US colleges.

5.1 Limitations:

The Data only looks at undergraduate admission rates and thus excludes any Transfer, Masters, or PhD students. This limits the exploration of the data to simply a subset of the student population.

In addition to that, many of the numerical variables describe the demographics of student's home neighborhood. As in it assumes if a student comes from a particular neighborhood, the student must also belong to that demographic. This is done to combine the data sets of the IPEDS and the US Treasury Data. However this includes the implicit bias that the students from a particular neighborhood as a hall mark of that demographic.

References

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.