Analysing University Application Rates using TCDSB Data

How a school's grade 9 graduation rates, literacy rates, Early Development Index and Library Activity Influence University Application rates

Sukhmani Khaira

21/12/2020

Code and supporting data is available at https://github.com/sukhmani-khaira/Univeristy-Application-Rates

Abstract
University Application Rates are influenced by multiple factors like school's resources, socio-economic status, student's self efficacy etc. By running a linear regression, this study analyses association between university application rates and the school's literacy rate, EDI score and graduation rate, and resources like library activity. The model finds a significant relation between the literacy rate and university application rate.

Keywords
Post-secondary Education, university application rates, linear regression, Toronto District School Board, EDI score, Literacy rate

Introduction

Post-Secondary Education in Canada has seen significant changes in the last 2 decades. Data from the 2016 Census on Canadian education showed positive highlights about the nature of post-secondary education in Canada, and has brought up some trends in Canadians pursuing university degrees and college diplomas. Canada ranks highly in the proportion of people who attend college or university and 'has the highest proportions of College and University Graduates in the OECD' (2016 Census: Education in Canada) largely due to it its huge college sector which is absent in most other developed nations.
40.7% of young women attained a bachelor's degree or higher (7.9 percentage points higher from 2011), while men were more likely to focus on careers in vocational skills, their proportion increasing by 59% between 2016 and 2011. (2016 Censes: Education in Canada). With an increased demand for skills and a changing labour market, it is important to study how different factors influence university applicants. Given how broad this can get, we start at the grassroots and look into schools and neighbourhoods of Toronto, and their university application rates. Since there are different types of post-secondary certifications available in Canada, it is important It is important to define that a university degree is one that is granted by universities or by other degree-granting institutions like colleges. (2016 Censes: Education in Canada). On the other hand, a 'college diploma' is a CEGEP or a non-university certificate.
It is a commonly accepted that a school's resources and performance greatly influence their students' achievement(Greenwald, 1996) and subsequently their post-secondary plans. Existing literature also supports that increased spending on resources can be associated with significant increase in student achievement (Greenwald, 1996). Hypothesising this in our mind, we look into university application rates, given a literacy rates, graduation rates and library activity. Analysing factors that influence post-secondary education benefits the larger educational field, and provides evidence towards adoption of support strategies/ teaching techniques to increase student engagement. While analysing observational data, our diagnostics show correlations and not causation. While specific techniques like Propensity Score matching and Regression Discontinuity provide for casual links, they cannot be applied to our data set as there is no 'treatment' group. While propensity score matching could have been used to factor in School Graduation Rates or EDI rates, but our observations are aggregated across neighbourhoods, so propensity score matching would not have been an effective method.
The data in use is TCDSB's Well-being-toronto-education dataset from 2008, and it is analysed by running a linear regression. The methodology section will layout the details of the dataset and the features of our regression model. This is followed by a results section and a discussion of our findings. Lastly, the weakness section lays out potential drawbacks and next steps lays out the direction that further research will take.

Methodology

-Data

The Toronto Catholic District School Board (TCDSB) administers schooling in 140 neighbourhoods of Toronto. Wellbeing Toronto-Education data is published by the Social Development, Finance and Administration of Toronto and is available on Toronto Public Library, TCDSB and Toronto District School Board (TDSB). The dataset contains 11 variables and 140 observations (1 for each neighbourhood) and looks at 9 variables that report the School Graduation rate (Grade 9), the Literacy Rate, University Application Rate, Library Activity etc. Each neighbourhood contains within itself smaller sub-areas (Census Tracts in this case). Data from Census Tracts is aggregated, and that combination defines the entire neighbourhood. Since most neighbourhoods in Toronto tend to be socio-economically homogeneous within themselves, this eliminates the possibility of 'averaging out' socio-economic differences in a neighbourhood when aggregating data.

As TCDSB uses specific language, the terminology for each variable is explained below:
1. Library Programs: Number of library and cosponsored programs
2. Library Program Attendance: number of people who attended library and Co sponsored programs
3. Library Open Hours: number of total hours library branches are open per year, including those on Sundays
4. Library Space: count of public space facilities (auditorium, meeting rooms, theatres etc)
5. Catholic University Applicants: Percentage of Catholic High School students applying to universities. This looks at the total number of 17 to 21 year olds attending the board as of March 31 of the schoolgirl who apply to Ontario universities, out of all 17 to 21 year old students (living in the census tracts) 6. Catholic School Literacy: Number of successful first time eligible students who passed Ontario Secondary School Literacy Test, out of all first time eligible students (living in census tracts) 7. Catholic School Graduation: number of students in grade 9 cohort graduating at the end of five years out of all students in the cohort (living in the census tract) 8. Early Development Instrument: For EDI children are tested on five areas of competence like physical health & well-being, social knowledge, emotional health etc. Based on categorization of scores into low/medium/high groups were Low = 0, Medium = 50 and High = 100
9. Library Activity: Total library activity like circulation of library materials, visits, info requests and book loans.

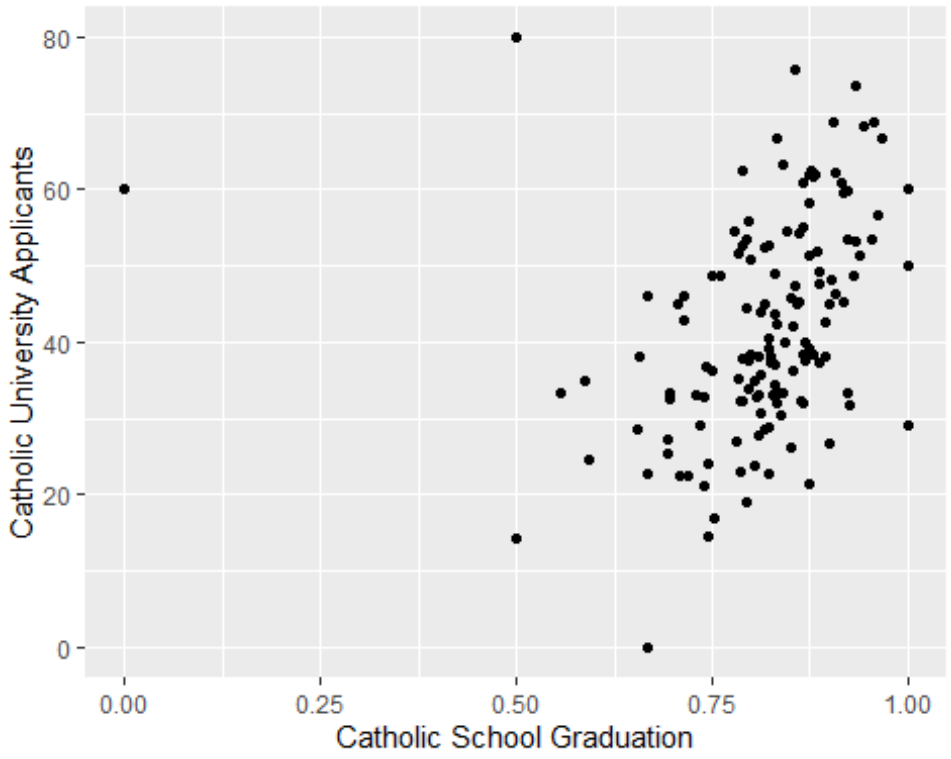

FIGURE 1

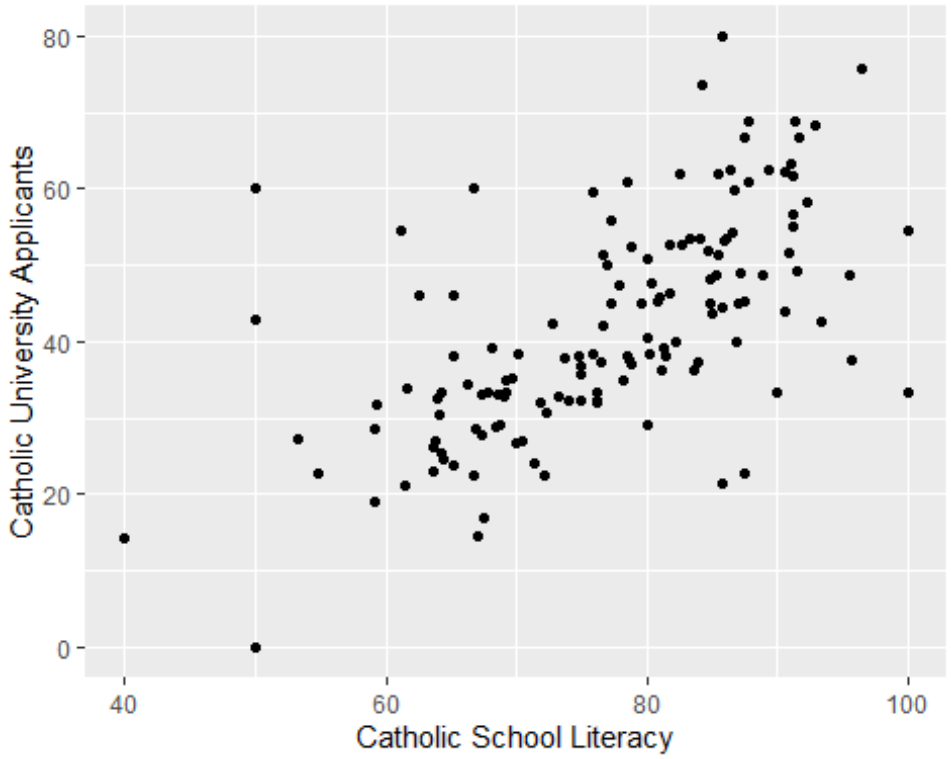

FIGURE 2

Plotting the raw data in Figure 1 and 2 for University Applicants against the School Literacy and School Graduation Rate shows linear and positive relations between the variables, but it is not possible to infer how strong these relations are graphically, especially since we have outliers that are skewing the graph

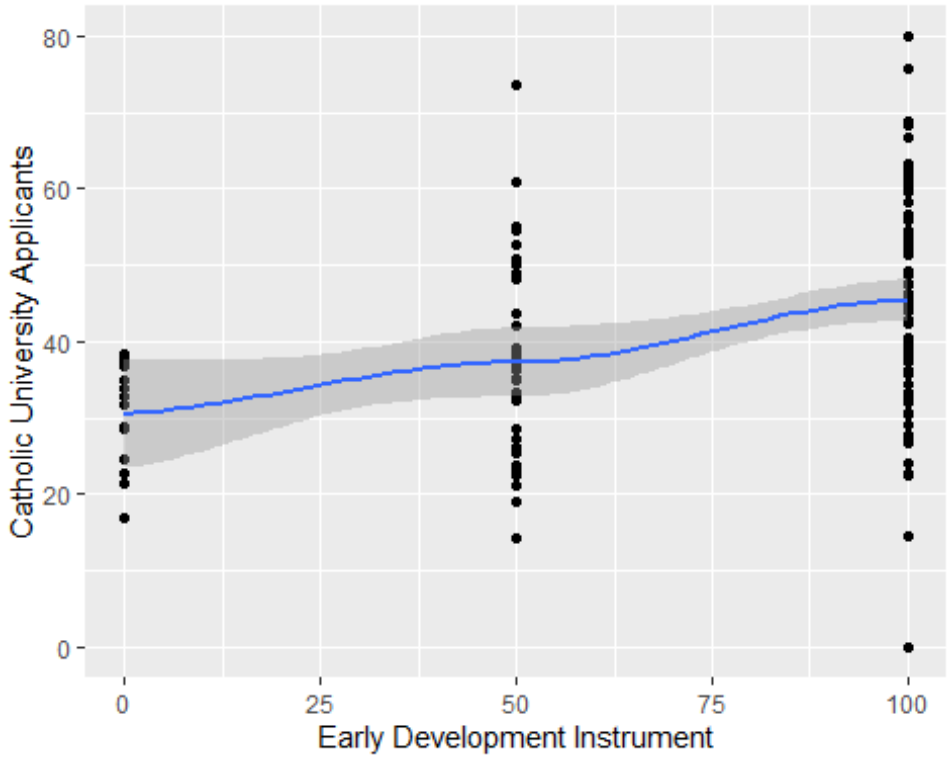

FIGURE3

Similarly, in Figure 3, University Application Rates v/s the Early Development Instrument (EDI) shows clear trends with schools with higher EDI showing higher university applicants. But it is still not possible to talk about corelations from these singular graphs.

Model

Existing research supports that a school's self-performance and resource availability are positively related to student outcomes, with effect sizes big enough to infer that an increase in spending (on school resources or programs) could be associated with significant increases in student achievement (Greenwald, 1996). To understand how a school's library activity, its graduation rate and literacy rate influence its university applications rate. We run a linear regression in R to study the correlation between University Application Rates and said independent variables. A linear regression is one of the most widely used methods of statistical analysis. It is a technique that allows us to predict a dependent variable from one or more independent variables (Yadav, Kumari, 2018) and thus allows us to measure the extent of association between variables.

Since we were analysing the university application rates from independent factors, the most appropriate way for diagnostics is running a linear regression.

```
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -28.743  -7.145  -1.035   6.056  31.817 
##
## Coefficients:
##                                             Estimate Std. Error t value
## (Intercept)                                 -18.4908     8.0176  -2.306
## `Catholic School Graduation`                 -5.0368     9.6874  -0.520
## `Catholic School Literacy`                    0.7425     0.1033   7.188
## `Library Activity`                            0.5347     0.5191   1.030
## as.factor(`Early Development Instrument`)50   4.2622     3.5765   1.192
## as.factor(`Early Development Instrument`)100  6.8755     3.3397   2.059
##                                            Pr(>|t|)
## (Intercept)                                  0.0226 *
## `Catholic School Graduation`                 0.6040
## `Catholic School Literacy`                  4.15e-11 ***
## `Library Activity`                           0.3048
## as.factor(`Early Development Instrument`)50   0.2355
## as.factor(`Early Development Instrument`)100  0.0415 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.03 on 134 degrees of freedom
## Multiple R-squared:  0.4187, Adjusted R-squared:  0.397 
## F-statistic:  19.3 on 5 and 134 DF,  p-value: 1.949e-14
```

-Regression Equation:

Catholic University Applicants = -18.4908 -5.0568(Catholic School Graduation) + 0.7425(Catholic School Literacy) +0.5347(Library Activity) + 4.2622(as.factorEDI_50) +6.8755(as.factorEDI_100)

Results

Intuitively it makes sense for a school with strong library activity, and higher literacy and graduation to have a higher rate of university applicants. After running a regression to predict how university application using explanatory variables like graduation rates, literacy rates and library activity, the coefficients support our intuition for all variables except School Graduation Rates.

We find that a higher Catholic School Literacy Rate and Library Activity is consistent with a higher University Applicant Rate. Additionally, schools with an EDI of 100 tend to have a higher application rate than schools with an EDI of 50, and consequently those with an EDI of 0. This could be alternately be seen as students from lower EDI schools having a lower tendency to apply to universities. While we cannot say that a higher literacy rate or EDI score leads to a higher University Application rate, we can see a correlation that is insightful and supportive of further research.

Table 1 below shows the Analysis of Variance (ANOVA) table and supports the significance of the model, especially the School Literacy Rate.

```
anova(model)
```

```
## Analysis of Variance Table
##
## Response: Catholic University Applicants
##                                          Df  Sum Sq Mean Sq F value    P
r(>F)
## `Catholic School Graduation`              1  2079.3  2079.3 17.1014 6.21
6e-05
## `Catholic School Literacy`                1  8877.9  8877.9 73.0175 2.52
2e-14
## `Library Activity`                        1   212.5   212.5  1.7473    0
.1885
## as.factor(`Early Development Instrument`)   2   564.3   282.1  2.3204    0
.1022
## Residuals                                134 16292.5   121.6
##
## `Catholic School Graduation`             ***
## `Catholic School Literacy`               ***
## `Library Activity`
## as.factor(`Early Development Instrument`)
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

TABLE1

Discussion

Surprisingly we see a significant negative coefficient on School Graduation but can be potentially be explained by how university applicants are defined in the data. According to the dataset, only those students are counted as university applicants who apply to Ontario Universities by March 31 of the school year. These do not count for students who apply late, apply to universities outside of Ontario or take a gap year.

The significance of Literacy Rate in our model is integral piece of our analysis: The school Literacy rate accounts for students who pass Ontario School Literacy Test Results (OSSLT) in their first time, thus it allows us to look into those who are left behind i.e. those who retake the OSSLT. The aim of the study is to allow for insights into university application behaviours, and additional support for students struggling with their OSSLT could potentially increase a school's university rates. This paper, and consequent research aims to highlight potential relations in how university application rates change, thus current study outcomes could help define relevant educational trends and support individual learning needs and abilities.

Weaknesses

-The first potential drawback to this dataset is that it uses data from 2008 Reference Period. While the Toronto Education system hasn't changed much in the last decade, the using a 12-year-old dataset could mean that out analysis isn't up to date.

-Secondly, the TDCSB university application rates only account for students who have applied to Ontario Universities before March 31 of the school year. This leaves out a lot of student who may take time off before heading to universities or choose to apply to only out of province universities.

-Lastly, the data for each neighbourhood is a mathematical aggregation of much smaller census tracts. Statistics Canada has very strict privacy policies which could mean that some smaller areas could have had their values repressed and the overall total might be slightly undercounted.

Next Steps

There is a lot of subsequent research that can stem from this analysis. This dataset was limited to Catholic schools in Toronto, but it can be extended to all TDSB schools, giving us a larger a bigger data set for each neighbourhood and a stronger outcome for the same analysis. A dataset that is more accurate would also allow us to perform propensity score matching between neighbourhoods using library activity and open hours, thus moving us closer to making causal links.

References

- Government of Canada, S. (2017, November 29). Education in Canada: Key results from the 2016 Census. Retrieved December 19, 2020 from https://www150.statcan.gc.ca/n1/en/daily-quotidien/171129/dq171129a-eng.pdf?st=8DR_7rod

- Greenwald, Rob, et al. "The Effect of School Resources on Student Achievement."Review of Educational Research, vol. 66, no. 3, 1996, pp. 361–396.JSTOR, www.jstor.org/stable/1170528. Accessed 9 Dec. 2020.

- wellbeing-toronto-educationpublished by Social Development, Finance & Administration https://open.toronto.ca/dataset/wellbeing-toronto-education/Contains information licensed under the Open Government Licence – Toronto

- Kumari K, Yadav S. Linear regression analysis study. J Pract Cardiovasc Sci 2018;4:33-6. https://www.researchgate.net/publication/324944461_Linear_regression_analysis_study

- Alexander, Rohan (2020, November 5). Telling Stories with Data https://www.tellingstorieswithdata.com/06-03-matching_and_differences.html