Kieu Trinh

OMSBA 5300

8/9/2025

**Effect of the College Scorecard Release on College Search Interest**

**Introduction**

The release of the College Scorecard in September 2015 provided prospective students with accessible information about U.S. colleges, including median graduate earnings. This project analyzes whether the release shifted online search interest toward colleges with higher graduate earnings compared to those with lower earnings. Using the provided dataset and a difference-in-differences approach, I compare changes in search activity for high-earning and low earning colleges before and after the Scorecard launch to address this question.

**Data Sources and Preparation**

* Google Trends data (trends\_up\_to\_....csv): provides a 0–100 search interest index for keywords representing U.S. colleges, recorded by month or week.
* The College Scorecard data (Most+Recent+Cohorts+(Scorecard+Elements).csv): contains institutional characteristics, including median graduate earnings and predominant degree type, allowing the sample to be restricted to bachelor’s-degree–granting colleges.
* The ID link file (id\_name\_link.csv): matches school names in the Google Trends data to the Scorecard’s unique identifiers (unitid, opeid).

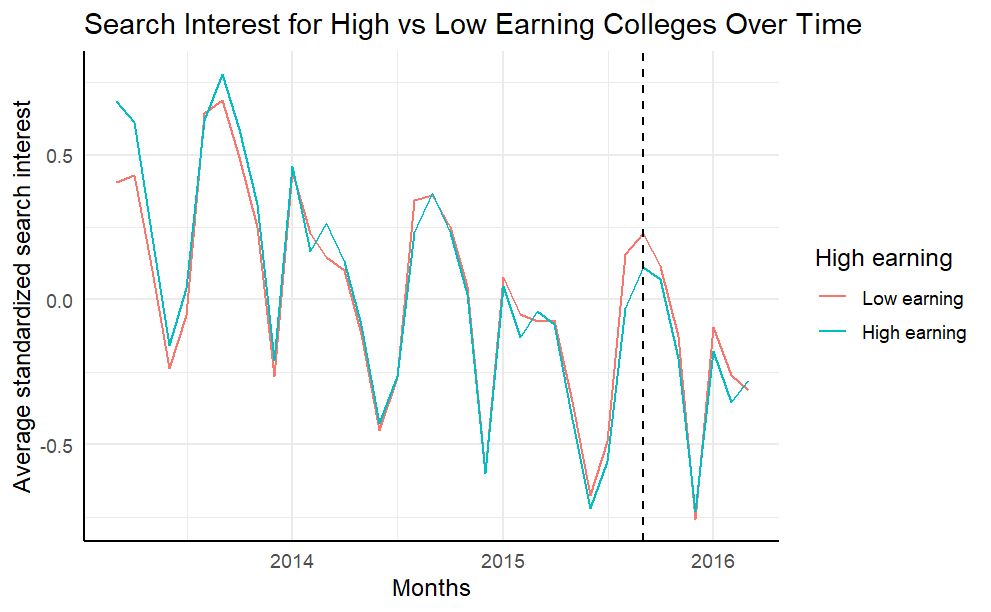
After importing the files, I removed missing dates, converted them to monthly periods, and standardized each keyword and school series for comparability. Median earnings were cleaned and converted to numeric values, and schools with duplicate names were dropped. The datasets were then merged to create a monthly college level panel containing standardized search interest, graduate earnings, and indicators for high-earning status and the post Scorecard period.

**Analytical Approach**

I use a difference-in-differences approach to estimate the effect of the College Scorecard release on online search interest. Colleges are classified as high earnings if their median graduate earnings, measured ten years after entry, are above the sample median and low earning otherwise. I choose the median split because it creates two groups of roughly equal size, which helps maintain statistical power, and avoids creating very small treatment or control groups that could result from using quartiles or fixed earnings thresholds. The post period begins in September 2015, when the Scorecard was released. The outcome variable is the average standardized Google Trends search interest (avg\_index) at the college month level. The model includes an interaction term between the high-earning indicator and the post period to capture the treatment effect, along with college fixed effects to control for time-invariant differences between institutions and month fixed effects to control for common time shocks.

**Analysis Results**

Before running the DID regression, I graphed the average standardized Google Trends search interest for high earning and low earning colleges over time, we can look at the graph below. I picked the line graph to clearly illustrate continuous changes and seasonal patterns in search interest, making it easier to compare trends between the two groups across months. The dashed vertical line marks September 2015, the release date of the College Scorecard. Looking at the graph, I can see both groups display nearly identical seasonal patterns before and after the release, with no visible widening of the gap between them. Therefore, at this point, I can say the release of scoreboard did not shift student interest to high-earnings colleges relative to low earnings ones. Additionally, the parallel movement of the lines supports the parallel trends assumption used in the DID regression and suggests no need for nonlinear terms in the time trend.



A screenshot of a computer screen

AI-generated content may be incorrect.

The introduction of the College Scorecard decreased search activity on Google Trends for colleges with high-earning graduates by 0.067 standard deviation units relative to what it did for colleges with low-earning graduates, with a standard error of 0.024. This result comes from the high\_earning × post\_scorecard coefficient in my regression. The outcome variable, avg\_index, is standardized within each college and keyword pair, so the coefficient is expressed in standard deviation units. A value of −0.067 means that after the Scorecard release, high earning colleges saw search interest drop by about 6.7 percent of their own historical variability compared to low earning colleges. While statistically significant at the 1 percent level, the effect size is small in practical terms. I include college fixed effects to control for all time-invariant characteristics of each institution such as size, location, and baseline popularity, and month fixed effects to control for shocks that affect all colleges equally in each month such as seasonal application cycles or national events. This isolates the effect of the Scorecard release on the relative change in search interest for high versus low earning colleges. While bias is possible if other events around September 2015 affected the two groups differently, the fixed effect structure and the parallel pre-treatment trends visible in the graph make large bias unlikely. Any remaining bias would likely be small, as there is no clear evidence of differential shocks in the post period.