

POPULAR HIGHER EDUCATION TYPES IN THE EAST AND THE WEST OF THE U.S.

Phuc (Peter) Luong

Section 1: Introduction

International students and out-of-state domestic students often would like to have a thorough understanding of the U.S. university system at a local level before enrolling. Besides ranks and popularity, specific majors and academic types play important roles in the students' decisions to attend college.

In addition to students, policy makers and educators are also interested in learning more about the academic distribution of a region before establishing new institutions. This knowledge helps not only compete with other universities and colleges, but also to have an overall understanding of the general demand of the local areas and nearby places.

However, one often wonders if there is a difference between different U.S. regions regarding specific education subjects. In other words, the paper examines whether there is a significant difference in academic type distribution of colleges and universities between the eastern and western regions of the U.S. "Are we more likely to see Technology-focused majors in the West coast relative to the East coast?" is an example of one of the interested questions.

For this project, I am going to utilize Foursquare API to find location data of a random selection of higher education departments throughout the East and West of the U.S. By using Foursquare API, I am also able to find location of university departments and their respective categories. The choices of region are completely arbitrary. Furthermore, five cities are drawn randomly for each region and Foursquare API will help obtain information about the academic type distribution of nearby cities and towns that house the corresponding universities. Later on, I will test formally, using Chi-square test for independence and multinomial logistic regression, to understand whether demographic characteristics of a region can predict the categorization of the higher education subjects of that region.

Section 2: Data Collection and Features

2.1. Data Sources and Collection Methods

As mentioned before, the main data source for this analysis comes from Foursquare API (for developers), a global database containing rich information of venues from many regions across the globes. The venues are also sorted to more than 900 categories, which are categorized mainly by the Foursquare consumer community. This aspect is very important in my analysis since all the assumptions about the population distribution (classes/categories of all universities

in the U.S) are based on the accuracy of venues' categorical classification. For instance, I will use the labelling for department of a university based solely on the data given by API Foursquare search.

For the locations and demographic data of unincorporated cities and towns, I scraped data and manually filled some data points in an Excel table containing relevant information. In addition to Foursquare API, I used sources from the following websites: Wikipedia, Google Maps and <https://simplemaps.com/>. The data tables can also be accessed through my GitHub account.

2.2. Data Description and Summary

The first data set, a raw dataset from Foursquare search results, concatenates different results obtained from Foursquare API data search. Relevant variable includes: name of the institution's department, coordinates (latitude and longitude), city, state, region and department type, which is the main dependent variable for this entire analysis. For the rest of this report, I will refer to this dataset as the main data. Later on, I will also merge the main data with data from Simplemaps.com, a city/town-level data that includes basic demographic information about the regions such as population, density and coordinates (cities and town level). This will be covered more in the later regression analysis.