**College Analysis**

This project is a part of the [Data Science Working Group](http://datascience.codeforsanfrancisco.org/) at [Code for San Francisco](http://www.codeforsanfrancisco.org/). Other DSWG projects can be found at the [main GitHub repo](https://github.com/sfbrigade/data-science-wg).

**-- Project Status: [ Completed]**

**Project Intro/Objective**

The goal of this project involves working with an unsupervised dataset of University enrollment stats and records from the early 90’s to 2018 to determine if a predictive model can be created to ascertain if a tuition decrease of 7% would bring a rise in overall enrollment.

**Methods Used**

* Data Visualization
* Predictive Modeling
* Data Analysis

**Technologies**

* R
  + dplyr / tidyr
  + ggplot
  + Cluster
  + Factoextra
  + Tibble
  + Tidyr
  + Tidyverse
* Rstudio

**Project Description**

**Analysis**

**Now you don’t have to follow these guidelines when conducting your analysis, but this was the general though process I went through creating my analysis.**

* Exclude all universities that do not have Title IV students. Generate unsupervised clusters of universities. Describe the features of each cluster – especially the features of the cluster to which your alma mater belongs (perform the analysis for The Ohio State University main campus if your alma mater is not represented in the dat a).
* Use your alma mater’s cluster to create a regression model and evaluate the relationship between average cost and enrollment. Use Number of Title IV students (combine NUM4\_PUB and NUM4\_PRIV) as the dependent variable. Use as many independent variables as you deem necessary, however, you must include at least one of the following as an independent variable:
  + Net Tuition revenue per full-time equivalent student (TUITFTE)
  + Average Net Price for Title IV institutions (merge both public and private fields, NPT4\_PUB and NPT4\_PRIV respectively).
* Assume other tests have been performed that confirm a causal relationship between cost and enrollment. Use the regression model to evaluate the effect on enrollment if your alma mater did a one-time price decrease of 7%.
  + How will this affect the total net tuition revenue of the school?
  + If this decreases the total net tuition revenue, what other variables could be changed to offset this loss?

**Needs of this project**

* data exploration/descriptive statistics
* data processing/cleaning
* statistical modeling

**Getting Started**

1. Clone this repo (for help see this [tutorial](https://help.github.com/articles/cloning-a-repository/))
2. The Data for this project contains sensitive information. Please reach out to the leads for access
3. Download data from this link <https://collegescorecard.ed.gov/data/>
4. Download CollegeScorecardDataDictionary to understand how data was reported