

# Final Project Proposal

Group Number: 070

Project Title: How to survive University

Team Name: Zuckerborgs

Team Member: Sarah Bender-Resnik salsabrsk

Da Gong Dagong3706

Christian Lim christianxlim

Alex Phan APhan1337

Yushan Wang YushanWang9801

Jiemin Tang jiemintang

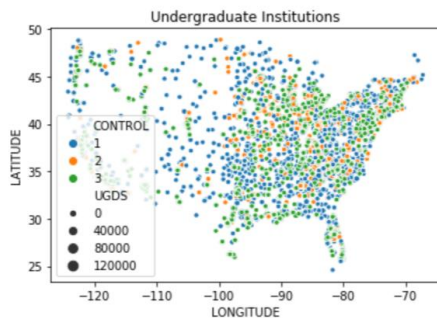
## Background:

For three million first-time freshman students in the US every year, choosing their dream college from offers is always a challenging problem. According to the College Scorecard Data provided by the U.S. Department of Education, there are 6418 colleges in America that can be divided into three main types: Private for-profit, Private nonprofit and Public.

```
In [23]: df=college[['CONTROL','LONGITUDE','LATITUDE','UGDS']]
#a=((df['LONGITUDE']>-120) and (df['LONGITUDE']<40))
df=df[(df['LONGITUDE']>-130) & (df['LONGITUDE']<-60)]
df=df[(df['LATITUDE']>23) & (df['LATITUDE']<55)]

#sns.despine(f, Left=True, bottom=True)
control = ["1", "2", "3"]
a=sns.color_palette()[0:3]
sns.scatterplot(x="LONGITUDE", y="LATITUDE",
                hue="CONTROL", size="UGDS",palette=a,
                data=df).set_title('Undergraduate Institutions')
#proj.scatterplot_us(college)
```

Out[23]: Text(0.5, 1.0, 'Undergraduate Institutions')



From this data, we can see that colleges are everywhere in America, which makes choosing one especially hard. Although every student all prefer to go to a college with a good reputation like Harvard or UCB, the fact is that the rate of acceptance is only 5.4% and 16.9% in 2019 for these two colleges. By this, students have to prove they are qualified for some good colleges using their SAT, ACT score, GPA, and their family fortune.

In order to help these applicants make their choice of college wisely, a successful guide generated from all college data would help them a lot in the application and save their time in finding every college.

Above all, we decided to generate a successful application guide from analyzing the College Scorecard. We will analyze the relationship between students' SAT score, College's completion

rate, College's type, College's student's size, and College's average salary for its graduate students and give students some advice for their application.

#### Data Science Question(s):

1. What type of college should one student choose to study since he/she will highly likely to graduate easily and on-time according to his/her SAT score?
2. What type of college should one student choose to apply since he/she will highly likely to receive an offer according to his/her SAT score?
3. What type of Predominant degrees-awarded school should one student choose to study since he/she will highly likely to graduate easily and on-time according to his/her SAT score?
4. What average annual net-cost of attendance in college would applicants expect to pay according to his/her SAT score?
5. What average first-year income in college would college graduates expect to get according to his/her SAT score?
6. What is the relationship between college size and its average completion rate?
7. What is the relationship between college type and its average completion rate?
8. What possibility one can graduate on-time from college could the applicant expect using his/her SAT score?
9. What is the relationship between college predominant degree-type and its average completion rate?

#### Hypothesis:

1. What is the relationship between college size and its average completion rate?  
H0: We can see that Large Size School has the lowest completion rate overall.  
H1: We can see that Large Size School has the highest completion rate overall.
2. What is the relationship between college type and its completion rate?  
H0: We can see that Private nonprofit School has the lowest completion rate overall.  
H1: We can see that Private nonprofit School has the highest completion rate overall.
3. What possibility one can graduate on-time from college could the applicant expect using his/her SAT score?  
H0: We can see that Predominantly certificate degree-granting School has the lowest completion rate overall.  
H1: We can see that Predominantly certificate degree-granting School has the highest completion rate overall.

#### Ethical Considerations:

The problem which this project intends to solve does not involve too many ethical concerns. Since it is the dataset published by the US government into publicity and it has a lot of null

values which represent the privacy some school wants to keep, there should be no ethical problem of this project. This dataset and the result of this project is in order to help general college applicants to choose their fittable college only.

Datasets:

This is the website provided by the U.S. Department of Education.:

<https://collegescorecard.ed.gov/>

This website allows users to check detail information of one specific college using the search engine it provides.

This is the download page of College Scorecard Data provided by the U.S. Department of Education.: <https://collegescorecard.ed.gov/data/>

This website allows users to download all college information in the US to execute further data analysis or information gaining. This dataset will be transferred into the pandas table and provide everything we need for our project.