This is the form your group must fill out for the project proposal (due May 5 2019 by 11:59pm CET).

El nombre y la foto asociados a tu cuenta de Google se registrarán cuando subas archivos y envíes este formulario. ¿No es tuya la dirección analucy.bejarano.m@gmail.com? Cambiar de cuenta

\*Obligatorio

Title of your project proposal \*

What makes a top University in USA?

Group member 1's name \*

Ana Lucy Bejarano Montalvo

Group member 2's name \* \*

Anna Alfieri

Background and motivation \*

**Discuss your motivations and reasons for choosing this project, especially any background or research interests that may have influenced your decision.**

Our aim is to understand the characteristics that make a top university. As foreign students ourselves, we needed to research which would be the best institution to attend. Rankings are always a practical source for exploring options however the procedures are not transparent and limit the decision to a simple selection that does not allow to explore the data. In that sense we want to present the information available about the universities so that people can make decision considering the data presented for the national Center of Education Statics; data that is also used for the companies that do the rankings in the United States. Visualizing the characteristics of the universities allow us and others to decide base on the source of data and give us the opportunity to create our own ranking to compare it with popular rankings.

**What are the scientific and inferential goals for this project? What would you like to learn and accomplish? List the benefits. What are some optional features (features or calculations which you consider would be nice to have, but not critical)?**

The goal of this project is to identify which are the most important characteristics that lead to a top USA Universities in 2017 to predict their rankings.

To identify the central features, we want to answer the following questions:

The top 50 colleges at USA in 2017

**Institution characteristics**

1. Have an international environment?
2. Generally, have large campus?
3. Are allocated in a particular region?
4. Are mostly private universities?

**Educational offerings**

1. Offer high degrees such as Post-master's certificate and Doctor's degree?

**Cost and investment**

1. Are the most expensive considering tuition and fees?
2. Are the most expensive considering the average cost of books and supplies required for an academic year?
3. Provide high salaries to their instructional staff?
4. Are the ones that expend the most in their libraries?

**Application and admission**

1. Receive more applications compared with the other universities?
2. Have the lowest rate of admission related to the number of applicants?
3. Have mostly a non-open admission policy?
4. Usually Requires GPA, rank and records to admit students or others test, for example foreign language?

**Grades and graduation rate**

1. Have the highest graduation rate among the universities?
2. The highest scores on…..SAT is a problem

**Benefits**

* Students and/or families can decide themselves which is their top university considering the information available
* The data presented will reduce the unintended bias generated by the ranking companies
* People can verify the information directly from the available files

**From where and how are you collecting your data?**

The data has been collected mainly downloading files available in the web page of the National Center for Education Statics. The rankings will be extracted from the Times higher Education and The QS World University Ranking web page. Links presented below.

* National Center for Education Statics

<https://nces.ed.gov>

* Times higher Education

https://www.timeshighereducation.com

* The QS World University Ranking

<https://www.qs.com/>

**List the statistical and computational methods you plan to use.**

The methods that we plan to use are: Classification tree, random forest and unit weighted model. We plan to compare the performance of these three models considering the accuracy but also their performance outside a training set.

Schedule/timeline \*

**Make sure that you plan your work so that you can avoid a big rush right before the final project deadline, and delegate different modules and responsibilities among your team members. Write this in terms of weekly deadlines.**

**Deadlines**

05.14 Wrangling

05.19 Visualization and Analysis

05.26 Modeling (II)

05.28 Final Analysis

05.30 Details

Responsibilities

**Data loading**

Download data files with the variables selected -> Team member 1

Extract ranking from webpages -> Team member 2

**Wrangling data and analysis**

Institution characteristics, Educational offerings, Grades and graduation rate -> Team member 1

Cost and investment, Application and admission -> Team member 2

**Modeling**

Classification three and random forest -> Team member 1

Weighted model -> Team member 2

**Final Analysis**

Team member 1 and 2