**DSFBA project proposal**

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

*Title of your project proposal \**

What makes a top University?

*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.**

As students ourselves, we needed to research which institution to attend. Rankings are always a practical source when exploring options and many undergraduates rely on them. Nevertheless, different agencies use different scoring methods, making it difficult to identify which are the key characteristics that a top university should have. For this reason, we decided to conduct our own analysis and build our own rankings.

*Project objectives \**

**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 university and use these factors to predict each university’s ranking. To detect what features do the best universities have in common, an extensive data exploration will be done. We hope this analysis might also answer these research questions:

1. Do the top 50 colleges have :
   1. an international environment?
   2. the highest graduation rates?
   3. a better and more effective capacity to deliver teaching?
   4. higher reputation?
   5. more courses offered?
   6. other characteristics to be discovered…
2. Do the top 50 colleges add value to students who attend them? (i.e. higher graduate salary)
3. Does the campus size/ location or number of students matter?
4. Does the number of degrees offered influence the number of applying students?
5. Does the fact of being a public or private university influence students’ results?
6. Is there a difference in the gender or ethnicity composition of best and non-best universities?

Other questions may come up when analyzing the data.

Since American colleges always occupy the top 10 positions of all rankings, we decided to focus only on the American system, but we believe this analysis could be extended also to other countries. It could be interesting to apply the same model on universities of other countries and check if the rankings match or if there are differences in what it is considered as important.

We hope this analysis might help our siblings (or even other students) in the decision process of where to attend college. Knowing what are the key factors they should look for instead of simply relying on rankings, could be really beneficial for them.

*Data* \*

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

Data comes from a variety of sources:

1) Data regarding the American university system: US government Integrated Postsecondary Education Data System (IPEDS) <https://nces.ed.gov/ipeds/datacenter/DataFiles.aspx>

2) Data concerning rankings: these data have been obtained by scraping the following websites

1. Wall Street Journal/Times Higher Education College Rankings 2017: <https://www.timeshighereducation.com/rankings/united-states/2017#!/page/0/length/25/sort_by/rank/sort_order/asc/cols/stats>
2. QS

*Design overview* \*

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

Data cleaning and tidying, data transformation (dyplr package), joining tables, data exploration and visualization (ggplot package), creation of a predictive model, comparison of model’s results to different rankings’ sources.

N.B. This list is not exhaustive. Changes will be made depending on the results obtained from the exploratory analysis of the data.

*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.**

The following plan assumes that each of these tasks will be completed by the proposed date.

Nevertheless, the schedule and task division may be updated/modified according to the progresses made.

May 10: Data scraping , cleaning, tidying and transformation for rankings table (Lucy) + data cleaning, tidying and transformation for all the other data (Anna)

May 17: Data visualization and exploratory analysis (Each member will be assigned a specific number of variables to analyze)

May 19: complete the “project update” form (Anna)

May 24: Decide which variables to include for the prediction model (Anna+Lucy) and built the prediction model (Lucy). Compare the results to different rankings’sources (Anna)

May 30: Review of the report and preparation of the oral presentation (Anna+Lucy)

*Additional files*

Feel free to add additional files if you need to motivate your project proposal. Please be aware of the file size limit (10MB per file, max 5 files).