

The IDEA / College Factual DataThon, 25 April, 2015

The challenge is to explore the data and come up with an issue of your choosing related to the universities on the following list. Then devise and implement an analytics procedure to address the issue.

Example issues could be:

- 1) I am a high school senior who wants to study 19th century forensic methods. I like to bowl and want to be near gourmet barbecue restaurants. Which school is the best for me?
- 2) I am an employer looking for a data scientist with a background in accounting and mechanical engineering. The candidate must be multi-lingual and have at least an 8 handicap in golf. Which schools do I target to recruit at?
- 3) Come up with ratings, for 30 institutions (including RPI!), rating how good the culinary/food scene around each institution is.

You are encouraged to "run wild" with it, and follow your own ideas and creativity. Don't be limited by any one particular approach.

Suggested software:

- * Use anything that you feel comfortable solving the problem with.
- * In order for it to be computationally tractable you should use a programming language (so not excel).
- * Consider the following open source programming languages in order:
 - R
 - Python (pandas/numpy stack)
 - Julia

The DataThon Challenge involves the following universities:

- 1) Brown University
- 2) California Institute of Technology
- 3) California Polytechnic State University
- 4) Carnegie Mellon University
- 5) Columbia University
- 6) Cornell University
- 7) Georgia Institute of Technology
- 8) Harvard University
- 9) Harvey Mudd College
- 10) Massachusetts Institute of Technology
- 11) Princeton University
- 12) Purdue University
- 13) Rensselaer Polytechnic Institute
- 14) Rice University
- 15) Stanford University
- 16) University of California - Los Angeles
- 17) University of California - San Diego
- 18) University of California at Berkeley
- 19) University of Illinois - Urbana-Champaign
- 20) University of Maryland - College Park
- 21) University of Massachusetts – Amherst
- 22) University of Michigan - Ann Arbor
- 23) University of North Carolina - Chapel Hill
- 24) University of Pennsylvania
- 25) University of Southern California
- 26) University of Texas - Austin
- 27) University of Washington
- 28) University of Waterloo
- 29) University of Wisconsin - Madison
- 30) Virginia Tech

You will be provided with additional information related to the universities and their local areas at the start of the DataThon.

You are free to access and use any other public data you can find, starting immediately, to formulate an analytics goal and work plan.