Hootsuite Data Scientist Take-Home Dataset “Random Acts of Pizza”

For you to show us how awesome you are with all things data, we’ve found an interesting dataset you can work your magic on: “Random Acts of Pizza”

> “This dataset contains a collection of 5671 textual requests for pizza from the Reddit community "Random Acts of Pizza" together with their outcome (successful/unsuccessful) and meta-data. All requests ask for the same thing: a free pizza. The outcome of each request --- whether its author received a pizza or not --- is known. Meta-data includes information such as: time of the request, activity of the requester, community-age of the requester, etc.” [1]

The URL at the bottom of this document will take you to a page where you can download and read up on the fields in the dataset.

The task asked of you is rather open-ended. **Explore the data and tell us about interesting patterns that distinguish successful and unsuccessful pizza requestors**. Given the time frame provided and the chances are that you have a day job, we do not expect you to build a 99% accurate pizza predictor that will in real time tell me if my RAoP post will score me a free pizza. To get you started you may want to consider (These are not requirements!):

- Are there certain characteristics about the requestor that correlate with a successful pizza request?
- Are there well defined groups of RAoP requestors?
- What, if anything, can you determine from the request texts?
- If you were to build a RAoP predictor how would you go about designing it, and what would be good features to use?
- Anything else about the data that is noteworthy, and why.

You are free to use any tools, platforms, analytic stacks, languages, magic, or voodoo you see fit to perform this analysis. The only concrete asks we have of you are that you submit the following 2 items before you begin your presentation:

1. A copy of the representation itself
2. A copy of the code you used to derive your results
3. (Optional): Any other artifacts you feel will complement the above 2 mandatory items

Logistics:

- If you have your own laptop that can connect to a screen via a Lightning port, then great, you can plug it in directly to our big screen! Otherwise, please have a copy of your presentation for us so we can pull it up on the screen with our laptops (You’ll need to submit a copy of your presentation anyways). Your best bet is to have the presentation on Google Slides or Keynote; not all of our machines can display Powerpoint properly
- Send the presentation and your source code to us through the interwebs or have it ready on a USB key
- You will have 20 minutes to present your findings followed by a 10 minute Q&A

[1] This dataset is publicly available at https://snap.stanford.edu/data/web-RedditPizzaRequests.html