

Description

Here at Granify, one of the main roles of a data scientist is being able to bridge the gap between raw data and knowledge. Even when data is cleaned and aggregated, there is still the open question of “What does this mean for the company and how do we act on it”. As a major step in us getting to know you better, we would like to see an example of your approach to turning data into information.

Problem and Data Description

The data-set attached contains generated user data for a hypothetical e-commerce site. Each line in the data-set represents a single session on the site (by an unknown user). Tabs are used to delimit values (the file format is .tsv). Each user was selected to be shown a particular marketing strategy, based on some prediction. All strategies were suppose to achieve the same goal (i.e. get a positive response from the user). Here is what each field in the data-set means:

- AD_ID - A positive integer designating which conversion strategy the user was presented
- RESPONSE - 1 if the user responded to marketing, 0 otherwise
- TIME - Day and hour in a 14 day range designating the start of the session. All times are in the user's local time.
- CONTROL - 1 if the user was part of a control group (and thus selected for a marketing strategy but not actually shown one), 0 otherwise

The remaining fields are discrete features that describe the session in some way. Features are named FEATURE_1, FEATURE_2, etc. and each feature can have values 1, 2, ..., N where N will vary for each feature.

The problem is given such a data-set, what can we learn from it? Questions you might want to consider include but are not limited to:

- What marketing strategies are effective and in what situations?
- What features are important for evaluating marketing strategy effectiveness?
- How are features correlated?
- What actionable conclusions can be drawn from this data-set?

These questions are simply a guideline to follow. Here at Granify, we appreciate creativity and would love to see how creative you can be when analyzing a data-set.

Objective

Prepare a brief (1-2 page) report that shows your findings. We want to get a sense of how you present data and what you consider important in a knowledge discovery task. What we would love to get out of your report:

- Learn how you analyze data
- What statistics are interesting for this problem
- What ways you can organize/display data that lead to interesting findings
- What actionable conclusions you can draw from this data-set

We don't have any hard restrictions about how your report should be formatted or what tools you use to analyze the data. Just be sure to present your findings in an easy-to-understand format (with clear sections presenting observations and conclusions). You can mention what techniques you used, but you don't have to give details (i.e. don't explain what k-means clustering is to us, just say that you used it).

Deliverables

The main deliverable expected is a brief report detailing your findings. It should be at least one page long and shouldn't be more than 3 pages long (brevity is appreciated, but take as much space as you have meaningful information to deliver) and should be sent as a PDF. If you have some short programming scripts that result from this task, feel free to share them with us (although code is not expected and avoid sending us large projects).

We expect this task to take somewhere between 4 and 8 hours. We do not expect you to spend days analyzing this data for us. There is no deadline, but we appreciate punctual responses. If there are additional analyses you would perform but have ran out of space or time, feel free to explain an additional processes you would use and what you would expect to learn from them.