Springboard

Yelp Review Case Study

The Business Problem

You're a data scientist at a marketing and design agency (congrats on the new job!) The company is expanding into a new sector: restaurants. **They really want to offer a "domain expertise" value-add in order to attract more restaurants as clients.** They're planning their next advertising campaign to pitch to clients and will be creating newspaper ads, magazine ads, and leaflets.

They know from customer surveys that a lot of diners select restaurants after checking Yelp, which means restaurants are very keen to maintain — if not improve — their Yelp ratings. On top of this, **the marketing department suspects that there's a link between specific words in a review and the rating linked to the review.**

Your company would like you to create a list of the top "favorable" words that are found in Yelp reviews so that they can use them in their ads. The hope is that people who see these favorable words in ads for a restaurant will draw people to that restaurant and also leave a more positive review after their meal.

You're a key influencer for this project. You meet up with the Head of Campaigns and decide that **your data science project should:**
1. Identify the category of restaurants to approach.
2. Produce a small set of "favorable" words or phrases to use in an ad campaign.

The Data

The Yelp Challenge Dataset is a freely available dataset used for educational purchases (what, did you think we'd be asking you to hack Yelp's database?) If this were a real business situation and you were working for a real marketing company, you would've had to reach an agreement with Yelp for such commercial use of the data.

In any case, for this case study, you have a bunch of data to work with. The key, however, will be to identify that data that will address your company's specific needs. Given what

you and the Head of Campaigns decided should be the scope of your project, you'll need to:

1. Identify the best (in this case, the most popular) type of restaurant category
2. Find a connect between review text (the specific "favorable" words) and review ratings

Your first question should be: "Which data file or files would I be interested in and why?"

As both a hint and guidance, consider this key learning objective: this part of the case study is meant to help you learn how to relate a data dictionary to the question of interest in order to identify key parts of the dataset.

You'll download your dataset in the next resource in this subunit.