

# The Hunt for Fake Reviews

## Instructions

In this assignment, you are asked to apply your scraper-building skills to a real-world task. You are assigned one dataset containing restaurant links of a certain US city. You have to scrape "not recommended" Yelp review data for these restaurants.

"Not recommended" reviews are reviews that are filtered out by Yelp's internal algorithm because they are deemed fake. The overarching goal of this assignment is to scrape and analyze fake reviews.

The assignment is divided into two steps:

### *Step 1:*

Scrape the "not recommended" review data from Yelp.

### *Step 2:*

The main goal of *Step 2* is to describe the dataset and present some of its interesting characteristics to the reader.

Present descriptive analyses of the data you collected. Start with very basic analyses such as a table for descriptive statistics, containing the mean, standard deviation, minimum, and maximum for each variable. Afterward, present more detailed analyses of your data in the form of figures and charts, allowing a more nuanced insight into the industry and its online ratings. You can also run appropriate regression or machine learning techniques you picked up during your master programme. Get creative and think about which insights could be valuable for which audience for which purpose!