

# Idea

**Problem to Solve:** Yelp restaurant review is an overall review, does not tell if a specific dish is good, what if you have a refined palate for margarita pizza? Turkish Coffee? Gluten free pancakes?

-> We want to find the places serve the best type of dish you want!

# Gather Data

- Write scraping script (real-time scraping)
- Use tools like Octoparse (click and drag)
- Download data online: business info, reviews... (offline data)

Will need data cleaning, strip punctuations and text parsing...

# Process Review Data

To search in review data and looking for the keyword user input, keep track of a counter to count frequency people liked it.

- **Main approach:** If review is 5 stars + mentioned keyword “croissant”  
(naively assume “croissant” mentioned here is the reason it got 5 stars, but they could also say: “everything is great, but I haven’t tried their croissant...”)
- **Additional approach:** Use Sentiment analysis to tell if a review includes the food you are searching for is classified as positive, negative or neutral.  
*Example: “Wow! Their croissant is amazing” -> positive;*  
*“Don’t try their almond croissant!” -> negative*

# Design UI

- **User input:**

*“croissant”*

- **Output:** The info of top 3 places with best croissant, sorted by frequency of people positively mentioned it

*#1: “Le Croissant Café, 1151 Broadway Ave, Burlingame, CA 94010, **72** positive comments on “croissant”*

*#2: “Farm: Table, 333 Geary Blvd, San Francisco, CA 94108, **57** positive comments on “croissant”*

*#3: “Acme Bread Company, 1 Ferry Bldg, San Francisco, 94111, **40** positive comments on “croissant”*

- **Near term:** Command Line

- **Future work:** web application or mobile app