Austin Herrick, Camille Choe, Emily Ding -- Computer Science and Applications 12200

Cuisine Aggregator: What to Eat Where

Introduction:
When Jeffrey traveled to LA for the first time, he was overwhelmed by the simple question of "What to eat for dinner?" Like any good millennial, Jeffery turned to the trusty food review websites that had served him well in home city Chicago. However, he was struck by decision fatigue - there were too many restaurants with 4.5 star reviews. Did 4.5 stars mean that the food was good? Did 4.5 stars mean that the people at LA were lenient with reviews? (Not likely). What type of cuisine is a "must-try" in LA? In the end, Jeffery, dumbfounded and demoralized, settled for McDonald's.

Don't be like Jeffrey. Use Cuisine Aggregator.

The Goal:
Our goal is to create an aggregated way for users to search and understand restaurant reviews. Sites like Zagat, Yelp, or OpenTable contain large amounts of helpful information on particular restaurants, and are great tools for deciding whether a particular restaurant is worth checking out. But, without a place to start, or a way to compare one site to another, the helpfulness of these reviews is significantly undermined. Cuisine Aggregator will simplify this process for users, by allowing them to search for aggregated reviews of entire cuisines or price levels within a given city and compare it with baseline numbers (such as, the average review our chain restaurants). We want a user to find out whether Thai food or Mexican food is better in their city, or whether their city is a good place for high end Italian.

Questions to potentially answer:
1. How do the average reviews in a city compare to the averages of other cities?
2. How do chain restaurants compare with each other across cities?
3. What cuisines in this city stand out?
4. How do different cuisines in a certain price bracket compare within a city? What about across cities?

Source of Data:
Our sources of data comprise of the most popular restaurant review websites. This includes: Yelp, Zomato, OpenTable, TripAdvisor, Zagat, Gayot, and Dine. While it is unlikely that we will pull data from *all* the review sites, this would be possible if we were to expand this project further. This data may comprise of statistics such as: number of stars, location, cuisine type, price category, and keywords. Our final decision on which web-review site to use will depend on our ability to scrape data reliability extract the information we need.

Components:
- Web Scraping from review sites
- Extracting relevant information using Beautifulsoup
- Importing into SQL, building a database of necessary information
- Queries for aggregation computation
- Trie system for Autocompleting City Names (of valid metropolitan locations)
- Creating an interface for the user
- Data visualization

Timeline:
4th week: Submit Proposal, Introductory Presentation
5th week: Scrape Data and Extract Relevant Information
6th week: Check in with Professor
7th week: Export into SQL, build database. Queries for aggregation computation.
8th week: Check in with Professor
9th Week: Visualization and Interface. Autocomplete system for cities list.
March 14th: Final software Due
10th week: Final Presentation

Limitations / Potential Challenges:

- Initially scraping the necessary data may be extremely time consuming
    - We can solve this by finding faster internet connections, reviewing fewer cities, or aggregating fewer review sites.
- Restaurants may have repeated names
    - We can use address as a unique identifier within our database
- Different websites may use different sets of cuisine categories
    - We can solve this by choosing one website's cuisine labels and using that as the standard.