

Do Yelp reviews tell the truth?

— A capstone project proposed by Rong Fang

Questions

- What are the useful reviews on Yelp?
- Who would be benefited from these reviews?

Objectives

- To develop a tool to refine the evaluation system of Yelp and to help users to find the most useful information for restaurant searching.

Data

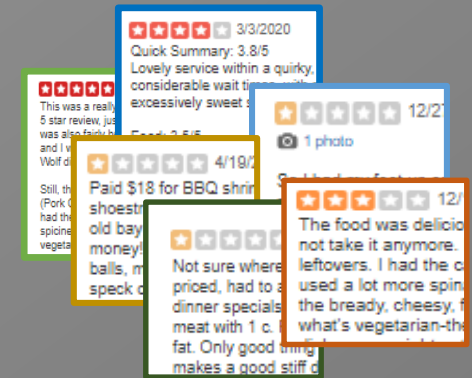
- Yelp open dataset

Methods

- Categorize the reviews with opinion mining: (e.g. service, attitude, overpriced, food)
- Sentiment analysis — label the subjectivity level of the reviews in terms of the relevant categories.
- Reevaluate the rating of the reviews based on the categories.

Expected Results

- A refined recommendation system that identifies the recommendation level of each category
- Order recommendations in a user's search based on her/his previous reviews



Random

Sentimental

Redundant

Biased

Opinion mining

Sentiment classification

Reevaluate

Informative

Precise

Relevant

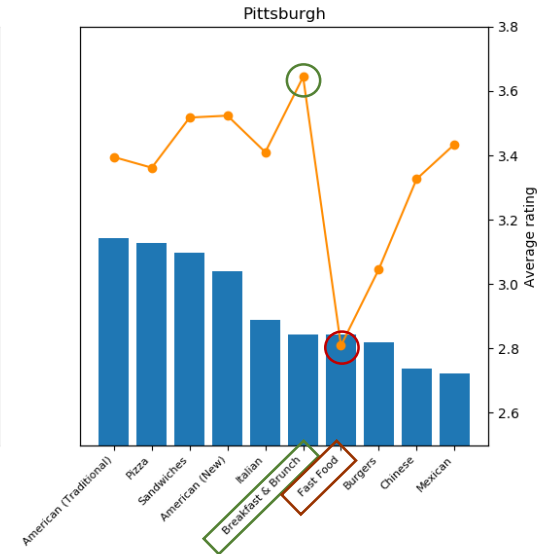
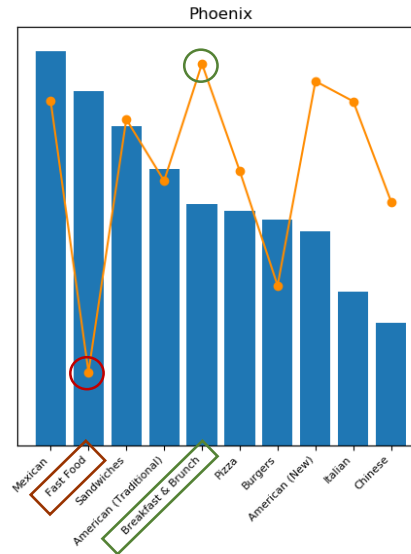
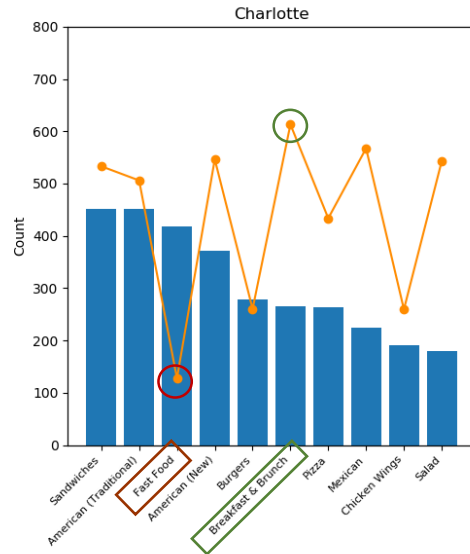
Neutral

Supporting libraries in python

- **Data frame process:**
Pandas
- **Cluster-computing framework:**
Pyspark
- **Visualization:**
Matplotlib, Seaborn
- **Opinion mining and Sentiment classification:**
NLTK, TextBlob, Stanford CoreNLP, and SpaCy

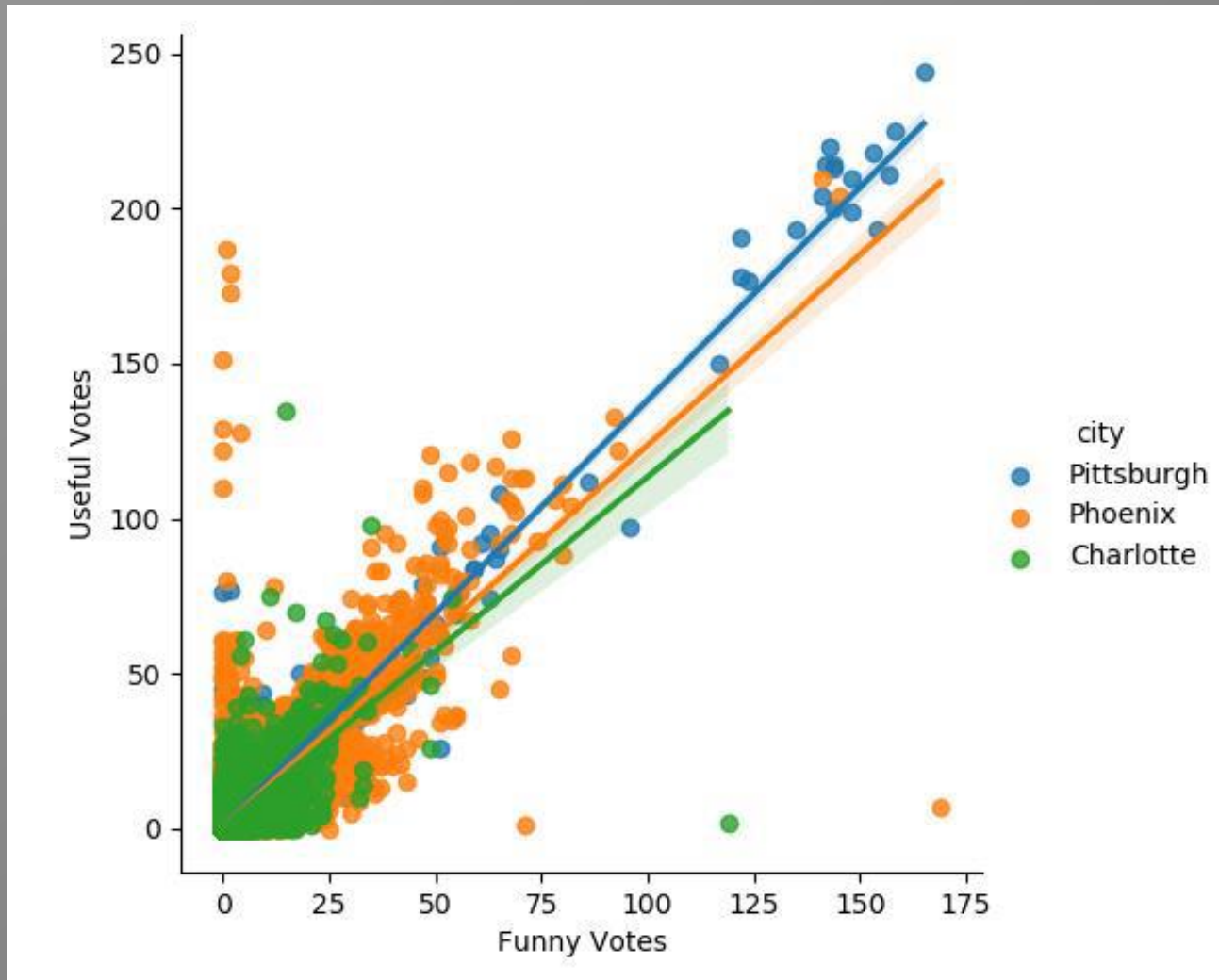
Preliminary Results

Numbers of restaurants based on their categories and average ratings



Preliminary Results

Numbers of votes for funny vs. votes for useful



Preliminary Results

Numbers of votes as useful against the rating given by reviews

