COFFEE REBREWER: USING ASPECT-BASED SENTIMENT ANALYSIS TO FIND THE BEST COFFEE IN TOWN

Georgia Institute of Technology

ANHTUAN HO, BO WANG, JASON YOUNG, DONNA YUNG
GEORGIA INSTITUTE OF TECHNOLOGY

Introduction

- We created an interactive guide for the best tasting coffee in Philadelphia by analyzing Yelp reviews to identify coffee shops serving the highest rating of coffee, focusing on coffee quality only.

- Users can evaluate and compare different coffee shops based on coffee quality, enabling coffee enthusiasts in Philadelphia to make informed decisions when searching for the best cup of joe in town.

Methods

Aspect-based sentiment analysis (ABSA)

- Determine sentiment of text with regards to a specific aspect.

- Pretrained model from PyASBA library

- Aspect = ['coffee', 'espresso', 'latte', 'etc']

Sentiment scoring

- Model outputs 3 sentiment scores: 'negative', 'positive', 'neutral'

- If highest score is 'positive', then review classified as positive

Custom metrics for each coffee shop:

$$1)\ Average\ PSS = \frac{sum\ of\ PSS}{total\ number\ of\ reviews}$$

$$2)\ Positive\ sentiment\ ratio = \frac{\#\ positive\ reviews}{total\ number\ of\ reviews}$$

where PSS refers to Positive Sentiment Score

$$Composite\ score = (0.6 \times Metric\ 1) + (0.4 \times Metric\ 2)$$
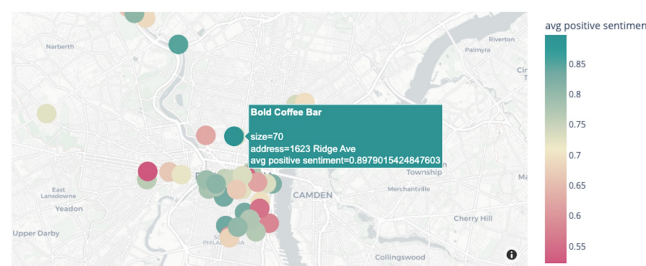
Interactive visualizations

- Geographic heatmap

- Line graph

- Top 10 table

Data

- **Source:** Downloaded from the Yelp official site

- **Characteristics:**
  - Total file size: 8.65 G
  - Number of businesses: 150,346
  - Number of reviews: 6,990,280
  - Original data format: JSON

Interactive Visualizations


The Landscape of Coffee Quality

Top 10 Coffee Shops by Sentiment Analysis

name	composite_score
Bold Coffee Bar	0.9244552112051416
Thunder Mug Cafe	0.8955540972727317
Pilgrim Roasters	0.8356592198212942
Grindcore House	0.8020097457562332
Konditori	0.7968630091390676
Ray's Café & Teahouse	0.7937139143335059
Knockbox Cafe	0.7934226261377336
La Colombe Coffee	0.7927396723202295
Vineyards Cafe	0.7690703160092234
Old City Coffee	0.7688304425257704

Coffee Sentiment Over Time
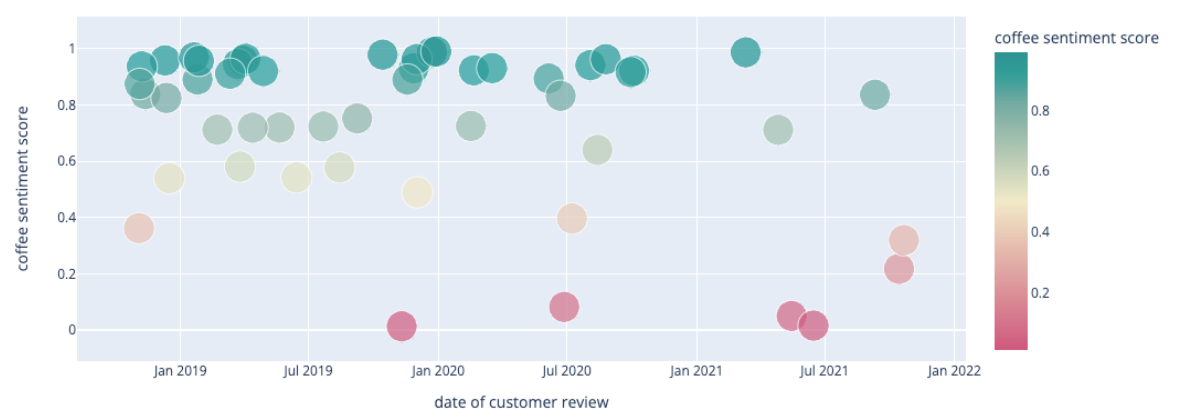
Experiments and Results

- **Experiment 1:** Using a random sample dataset of 20 coffee shops, we compared our results with the existing ratings on Yelp using Spearman's rank correlation. The correlation between our results with the average star ratings from recent reviews is 0.777.

- **Experiment 2:** Using linear regression, we tested our coefficients for *Metric 1* and *Metric 2* in our algorithm. The correlation between the fitted result and average star ratings from recent reviews is 0.824.

- **Experiment 3:** Using 10 randomly selected reviews per coffee shop, we manually rated each and compared our results. The correlation between human ratings and our results is 0.977.

- **Experiment 4:** Using the "useful" count as a weight for each review yielded a coefficient between the weighted ranks and human ranks of 0.734 which is a lower score and had no added benefit.