

NYC Restaurant Sentiment Mapping from Yelp Reviews



TeamBD (#110): Yu-Ying Liao, Matthew Matsuo, Shinnosuke Jay Nonaka, Akira Taniguchi

Introduction

Overwhelmed by the endless number of restaurants in **NYC**? We designed a unique dashboard to let users compare **restaurant review sentiment**. This will better inform the public and improve the consumer dining experience while simultaneously allowing businesses to scope out their competition. Furthermore, restaurants are clustered by **geography** and **sentiment** to provide more insight on localized review patterns.

Approaches

Our methodology uniquely combines both natural language processing (NLP) and clustering methods (K-means) to deliver a comprehensive dashboard of restaurant reviews in NYC. The NLP algorithm scores each cleaned user review as either positive or negative. Each review's score and label is then aggregated to the restaurant level and used as an input to the K-means clustering algorithm, which groups restaurants by geography and other dimensions.

Yelp Dataset

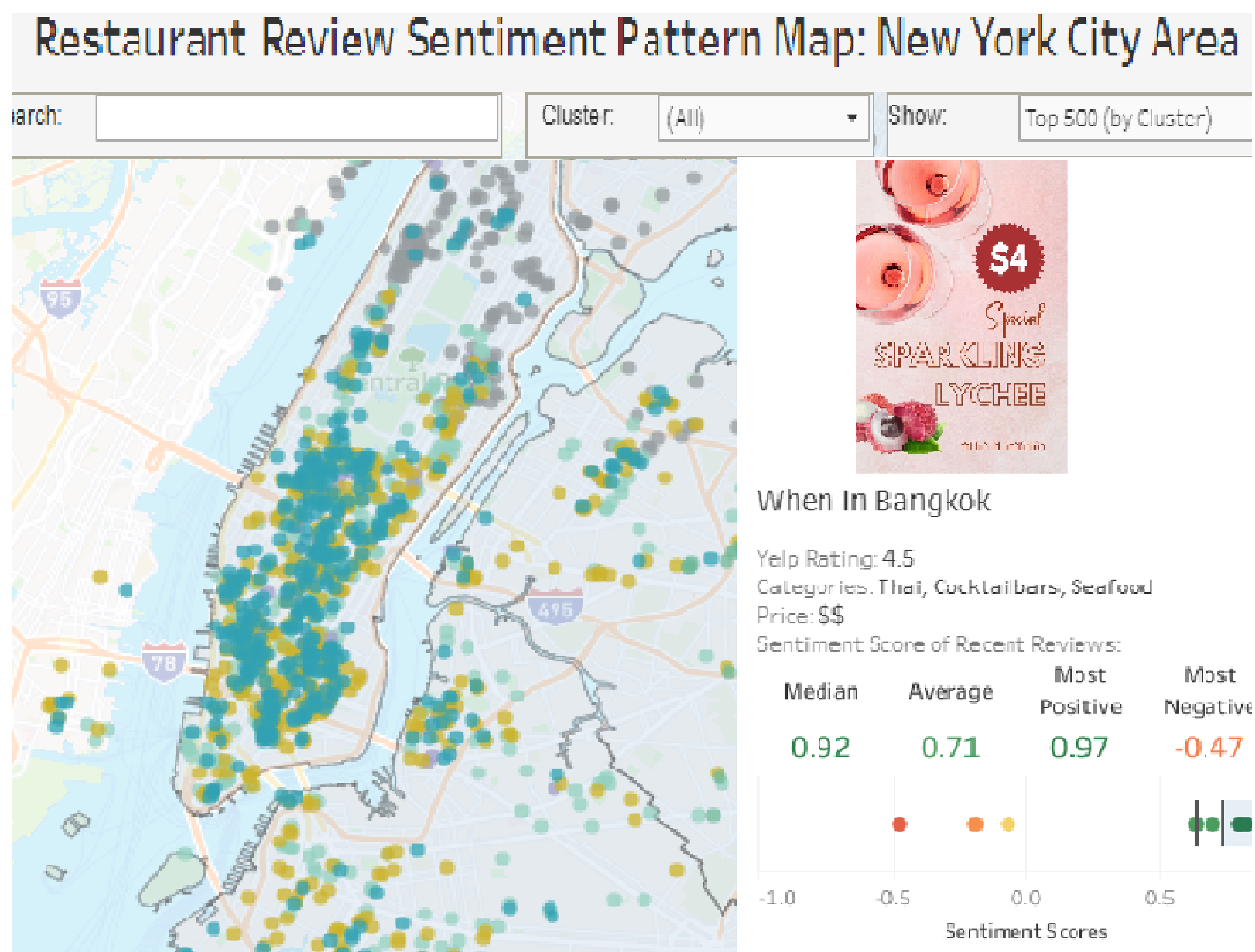
Users post their thoughts and experiences on Yelp. Some data can be accessed via their API, but web scraping is needed to collect all reviews.

Location	NYC
Restaurants	18,198
Reviews	315,711

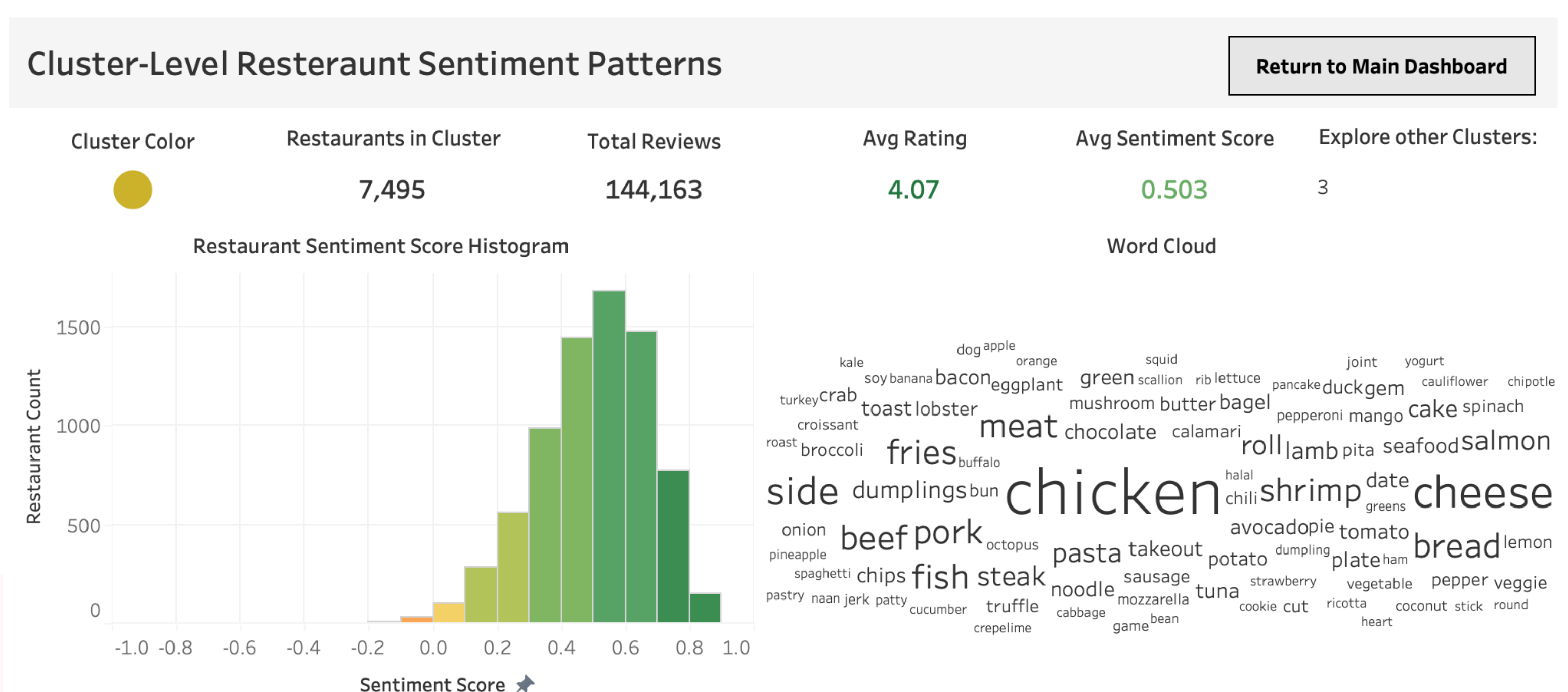
Experiments

We compared seven pre-trained NLP models and landed on a model with a low relative training time, positive and negative sentiment labels, and a non-heavily skewed distribution of output ranging from -1 to 1. This model also labeled **82% of one-star reviews as negative** and **98% of five-star reviews as positive**.

Our clustering algorithm tests involved several iterations of K-means, DBSCAN, and hierarchical clustering using variables such as latitude, longitude, review count, rating and average sentimental score. We settled on **K-means with seven clusters** and a **silhouette score of 0.3** as our final model.



The main dashboard (left) and tooltip (right)



Cluster-level dashboard (above)

Results

- The final deliverable was an interactive Tableau dashboard.
- Main dashboard provides an intuitive map of the NYC area, color-coded by sentiment clusters, allowing the users to explore sentiments geospatially.
- The tooltip provides detailed insights into individual restaurants and their sentimental characteristics.
- The cluster-level dashboard conveys important information on each sentiment cluster such as summary statistics, sentiment distribution graphs, and word clouds.