



Proposal: Yelp Dataset Analysis

STAT 628 Group 9: Yezhou Li, Chushí Shi, Lu Chen, Fangfei Lin



Outline

- Motivation and Goal
- Overview of Business
- Food Map: Heat Map
- Word Cloud: One-Star Restaurant vs. Five-Star Restaurant
- Yelp User Behaviors and Social Network
- Future Work

Motivation

Our dataset has 3172 Asian restaurants in Great Toronto Area (Downtown, North York, Vaughan, Scarborough, Mississauga, Richmond Hill). We only focus on the following five categories:

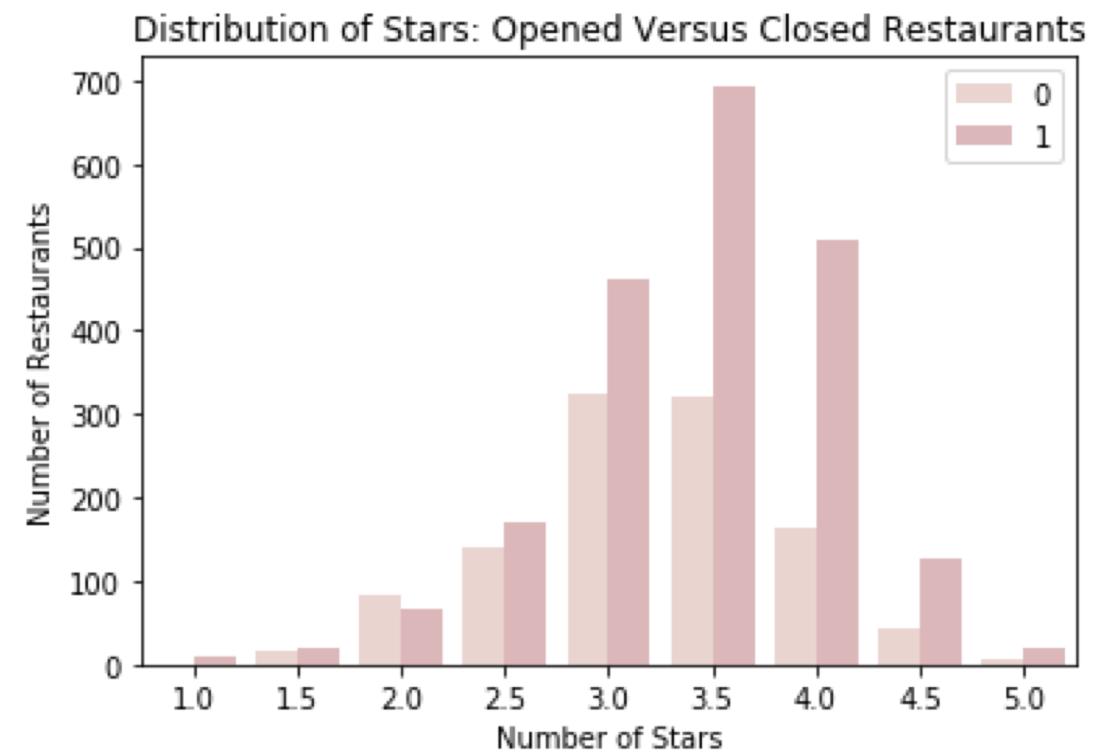
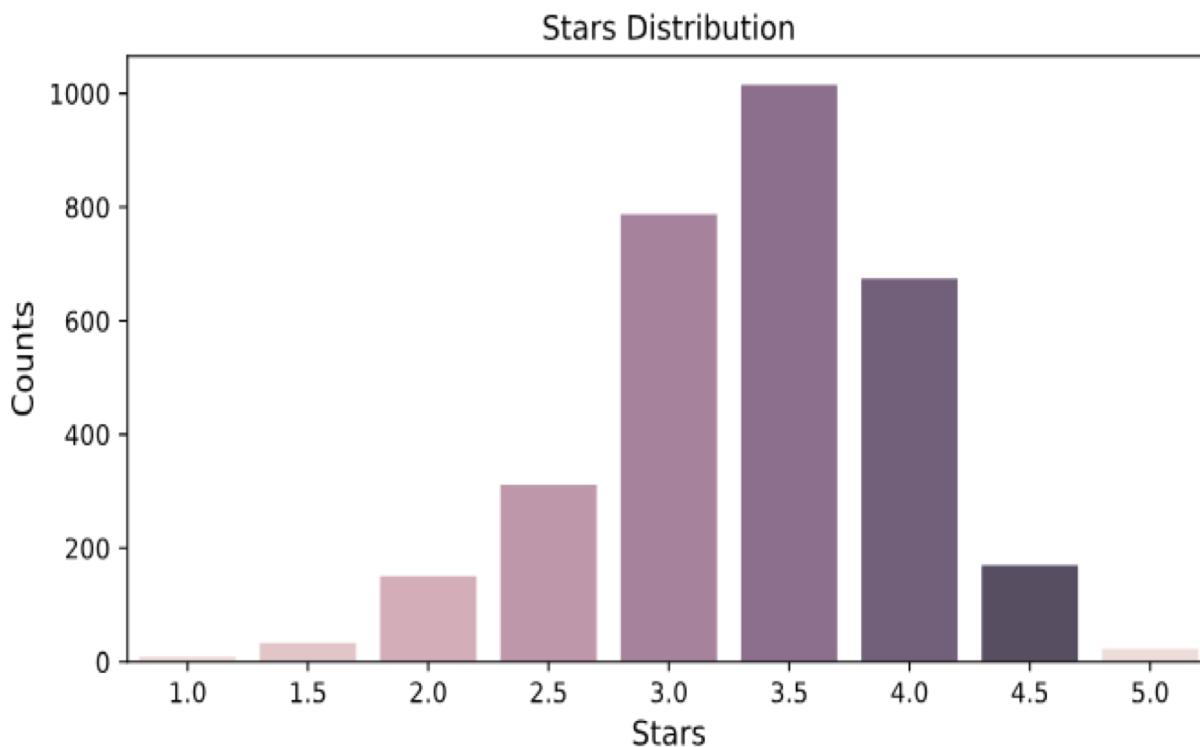
- Chinese
- Japanese
- Korean
- Thai
- Vietnamese

Our Goal

Give **ADVICE** to future business owners on:

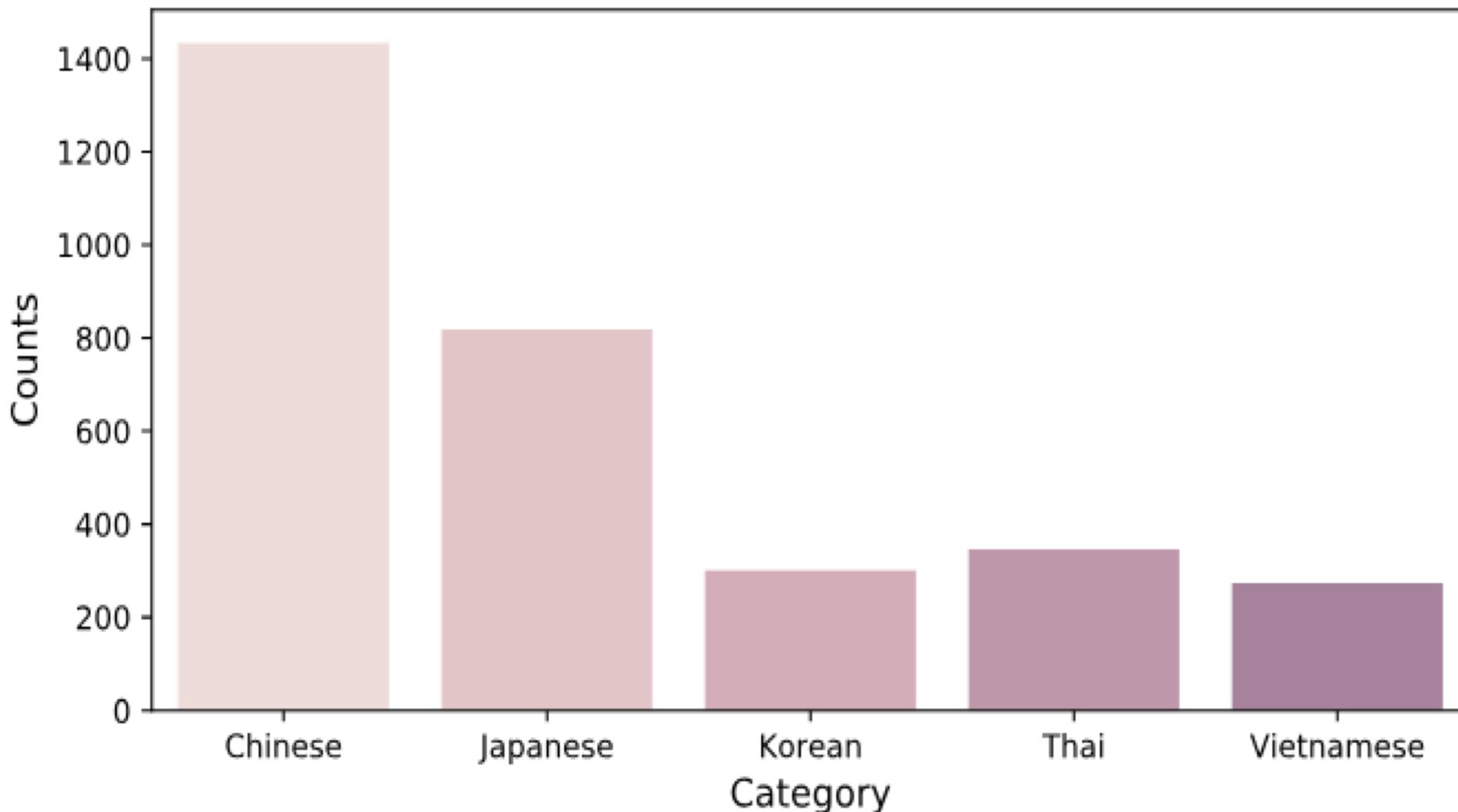
- Location
- Food
- Operations
- Social media advertisement

Overview of Business

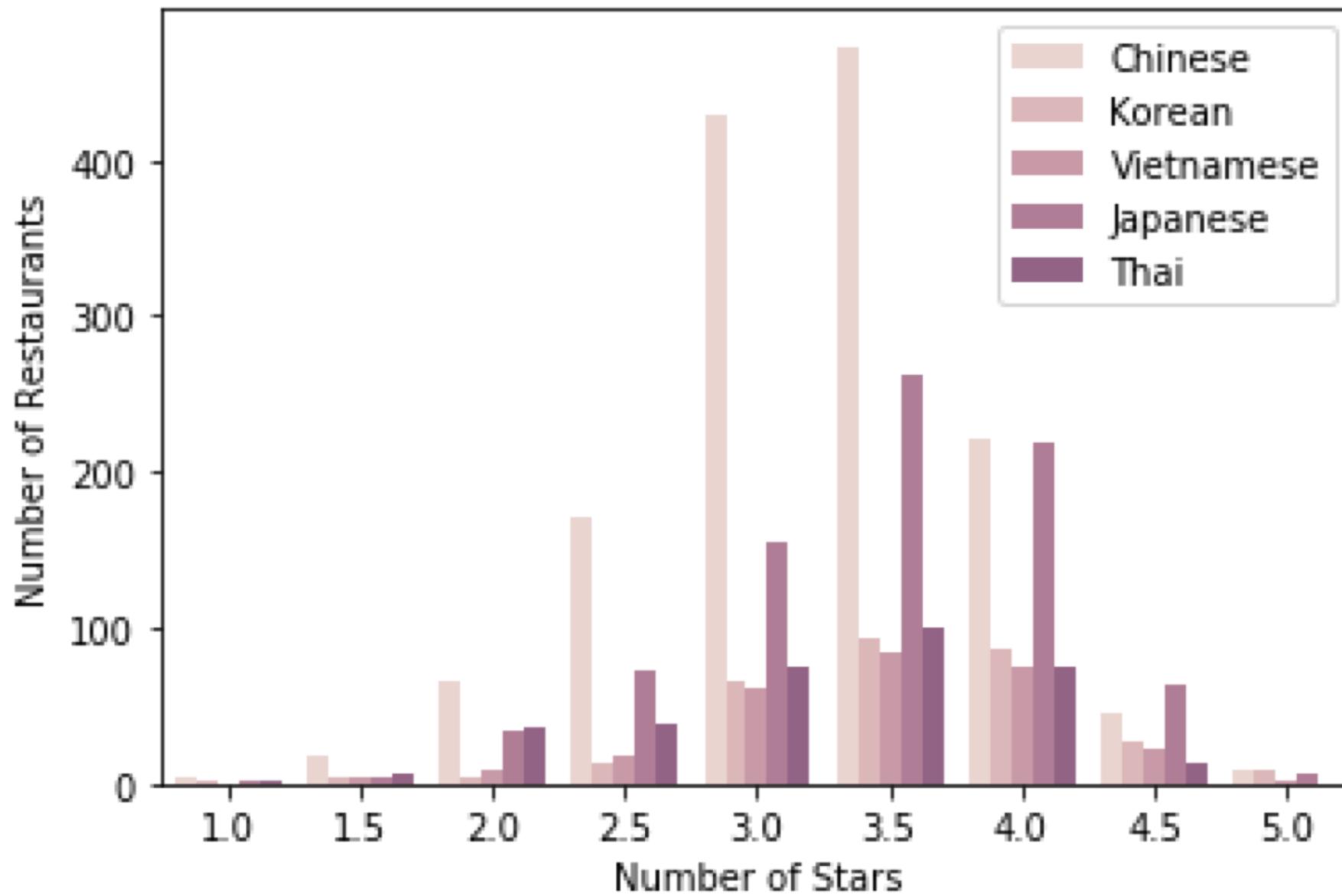


Overview of Business

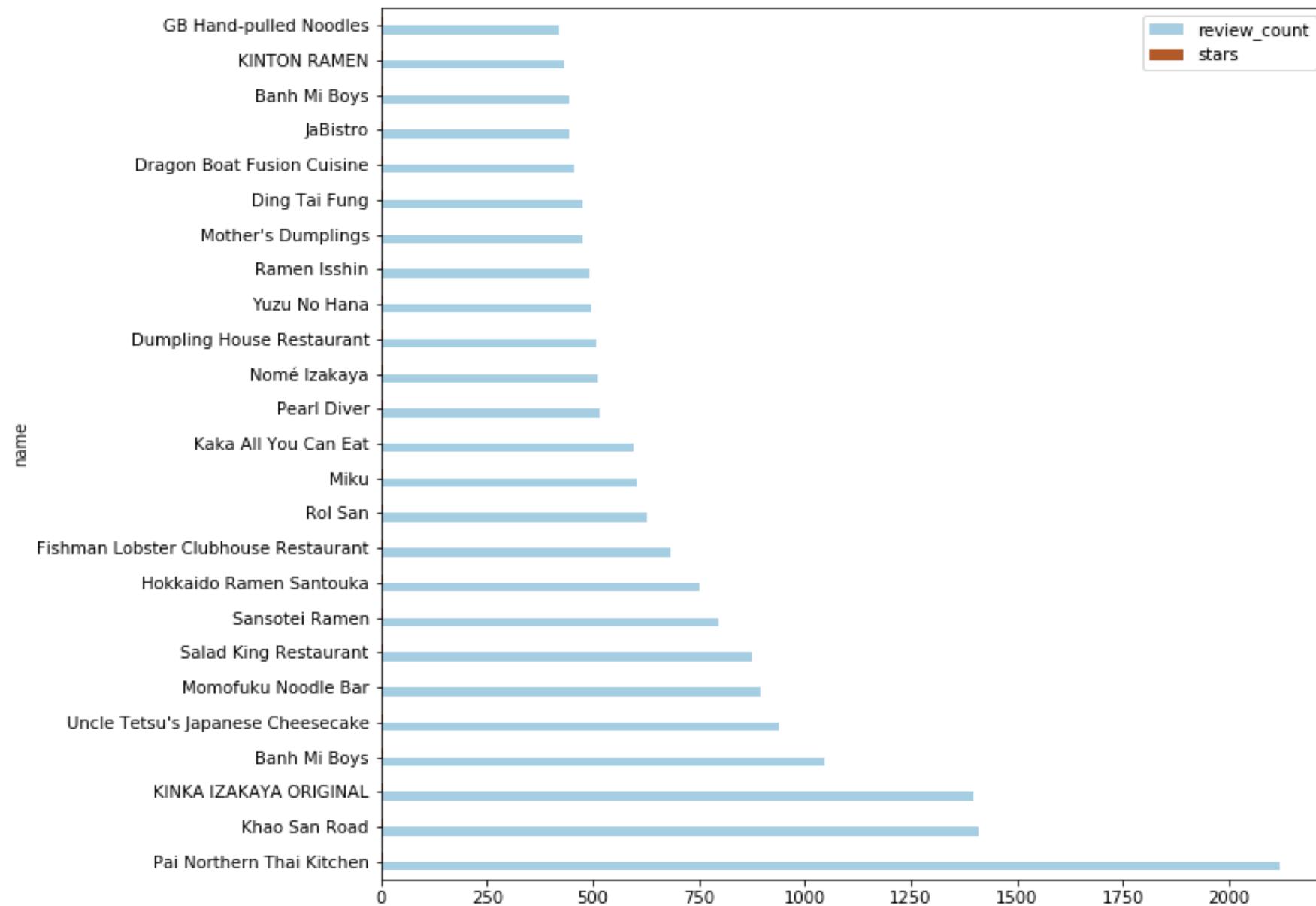
Category Distribution



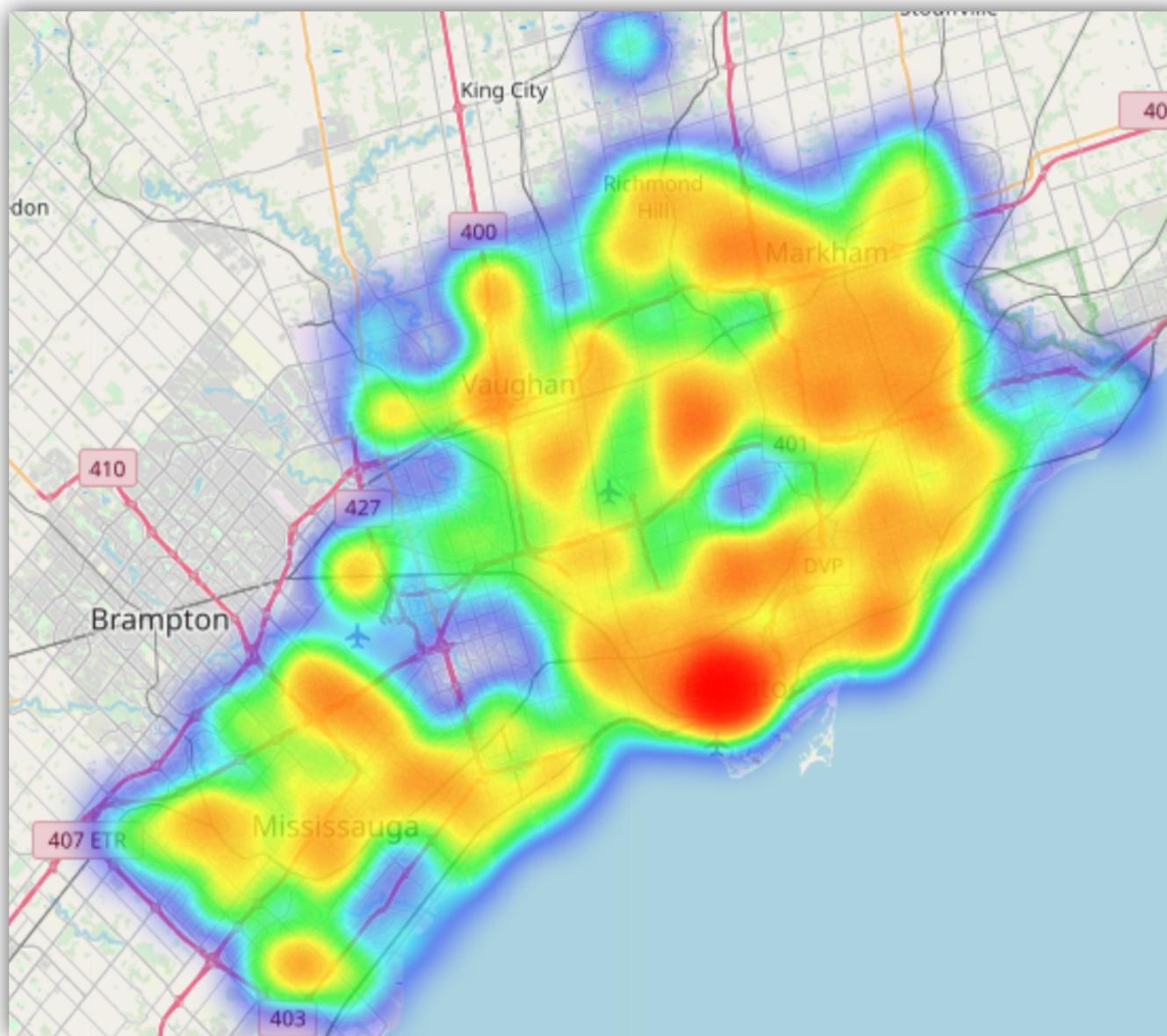
Overview of Business



Overview of Business: Top 30 Restaurants



Food Map



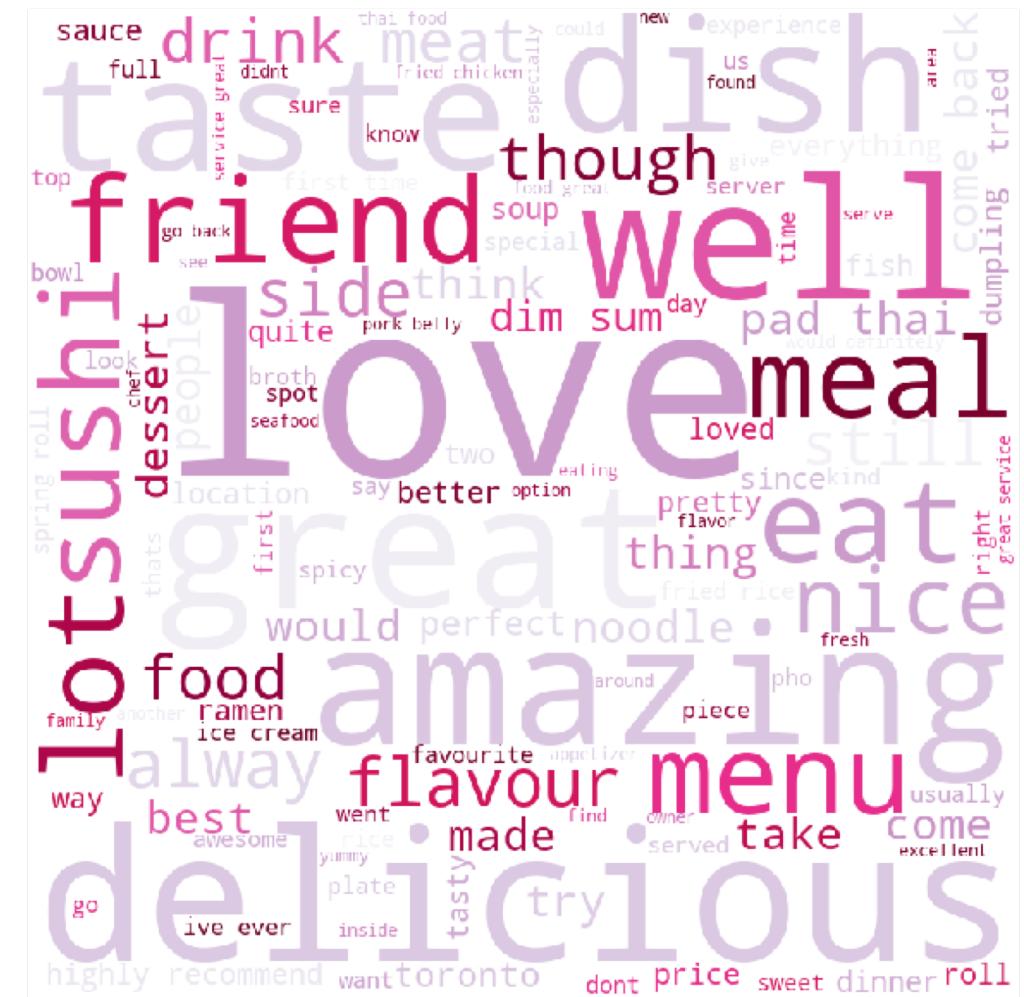
Food Map

- We visualize five categories of Asian restaurants, and we take Chinese restaurants as our example.

Word Cloud of Reviews



One-Star Comment

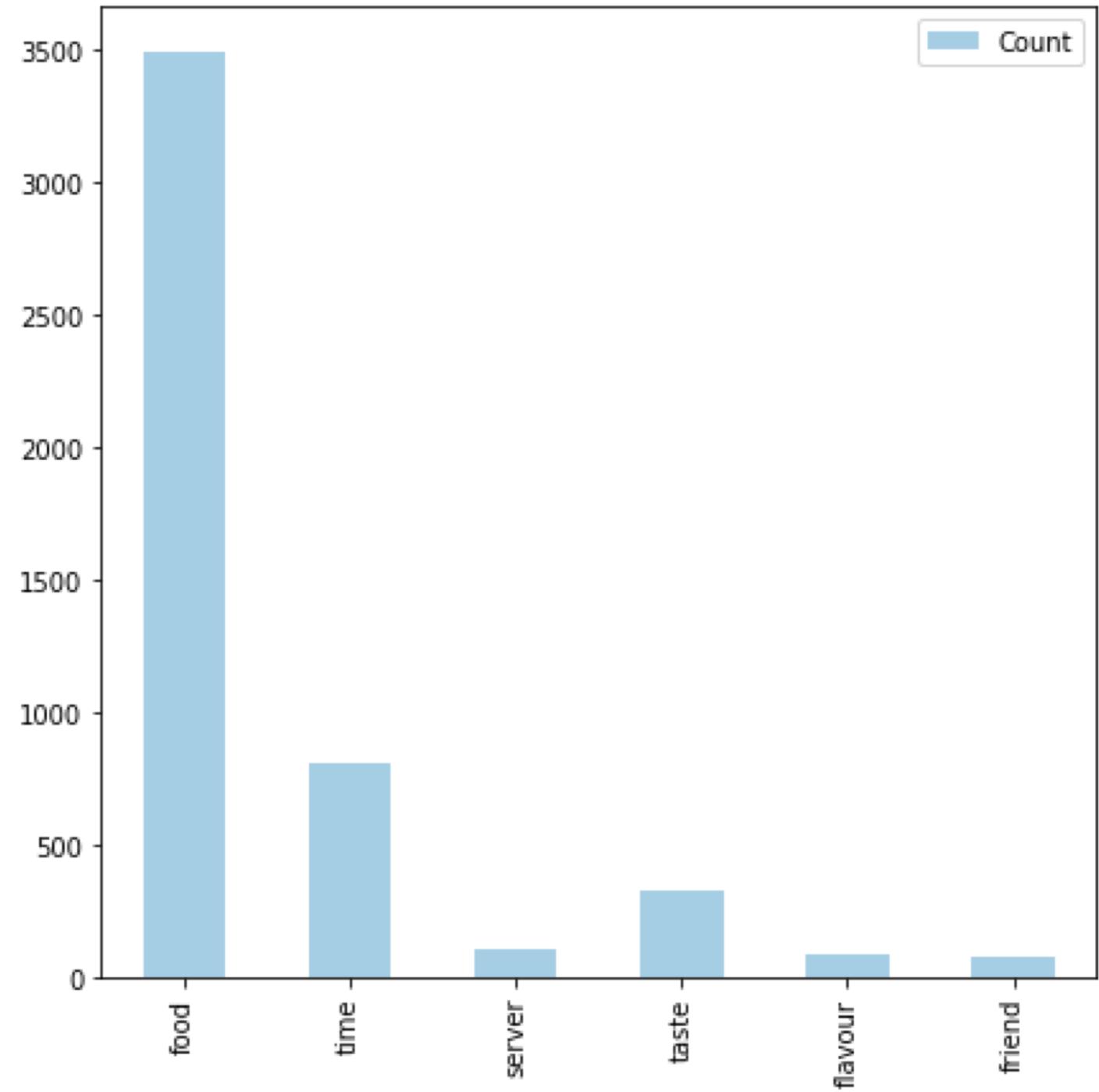


Five-Star Comment

Tips

People most care about:

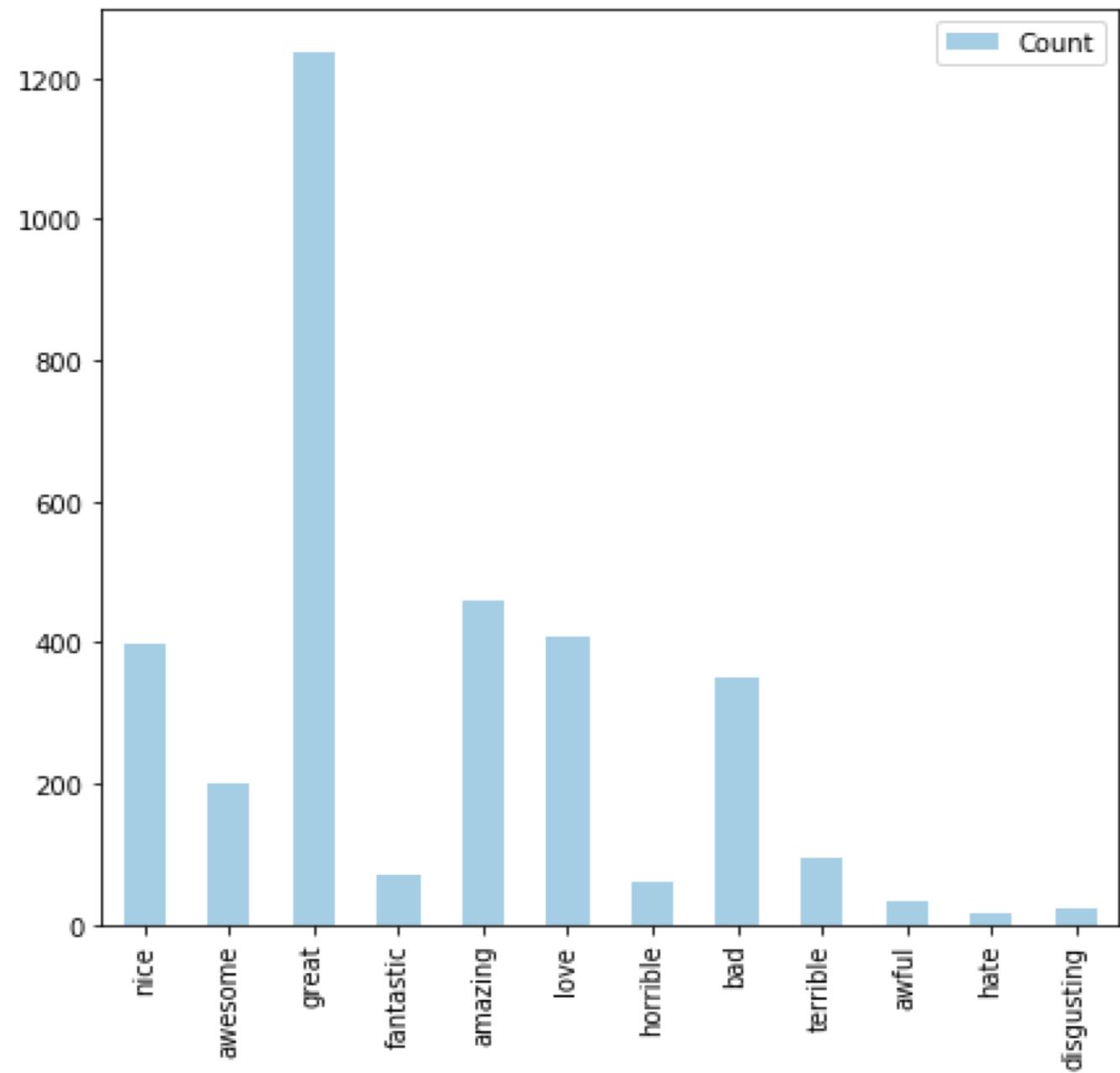
- Food
- Time of Waiting
- Service



Tips

Key Words:

- Positive: Nice, Awesome, Great, Fantastic, Amazing, Love
- Negative: Horrible, Bad, Terrible, Awful, Hate, Disgusting
- More positive comments



Word Cloud of Reviews



Chinese



Korean

Word Cloud of Reviews



Vietnamese



Thai

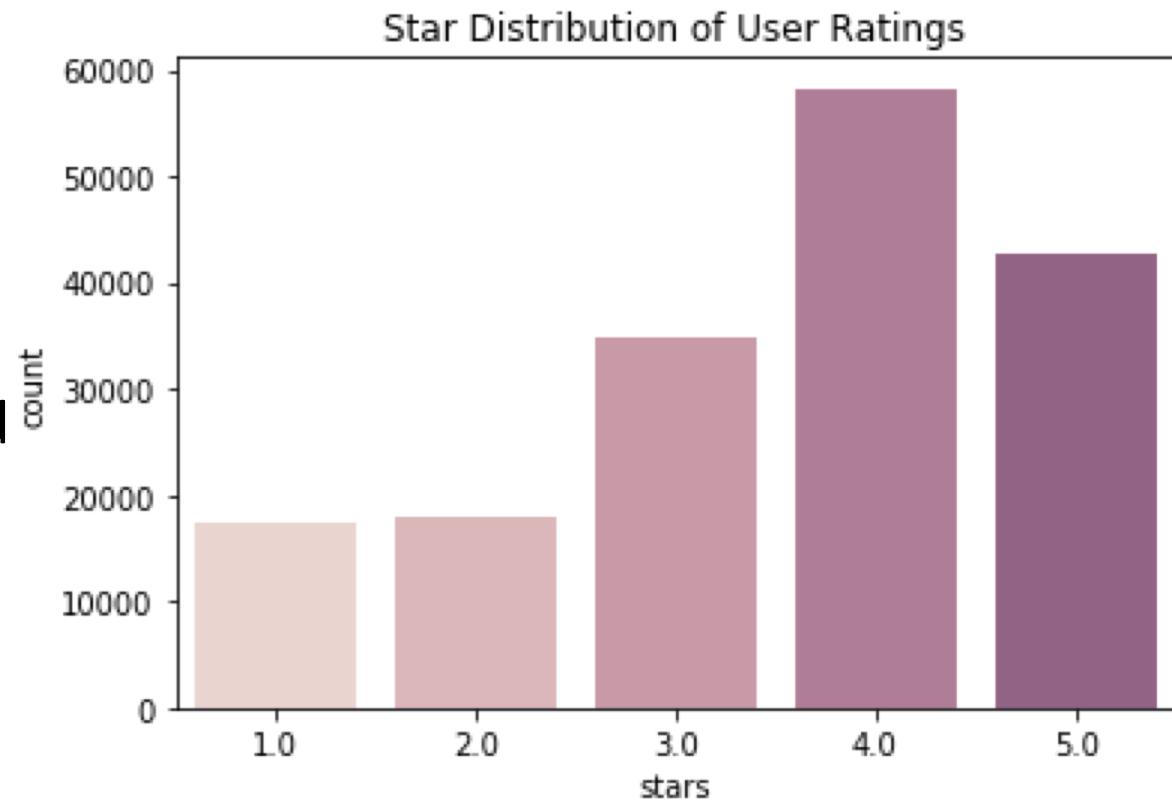
Word Cloud of Reviews



Japanese

Yelp User Behaviors & Network

- 51702 Users
- Only 34 of them have more than 500 followers.
- Average star ratings is 3.5, and the standard deviation is 1.1.
- Social Network



Our Future Work

Further Word Processing:

- Latent Dirichlet Allocation
- Word2Vec
- GloVe

Sentiment Analysis:

- SVM
- Xgboost
- LSTM

Social Network Advertisement

- Reconstruct Social Network
- Graph Clustering/Density

Thank You all!