

# Dish recommend based on yelp

Ran Wang, A20325697

Xiaocheng Hou, A20309864

# Problems

- What topics customer have in reviews?
- How similar between two kind of cuisines?
- How can we make a better recommendation for users?
- What dishes the users like in different cuisines?

# Approach

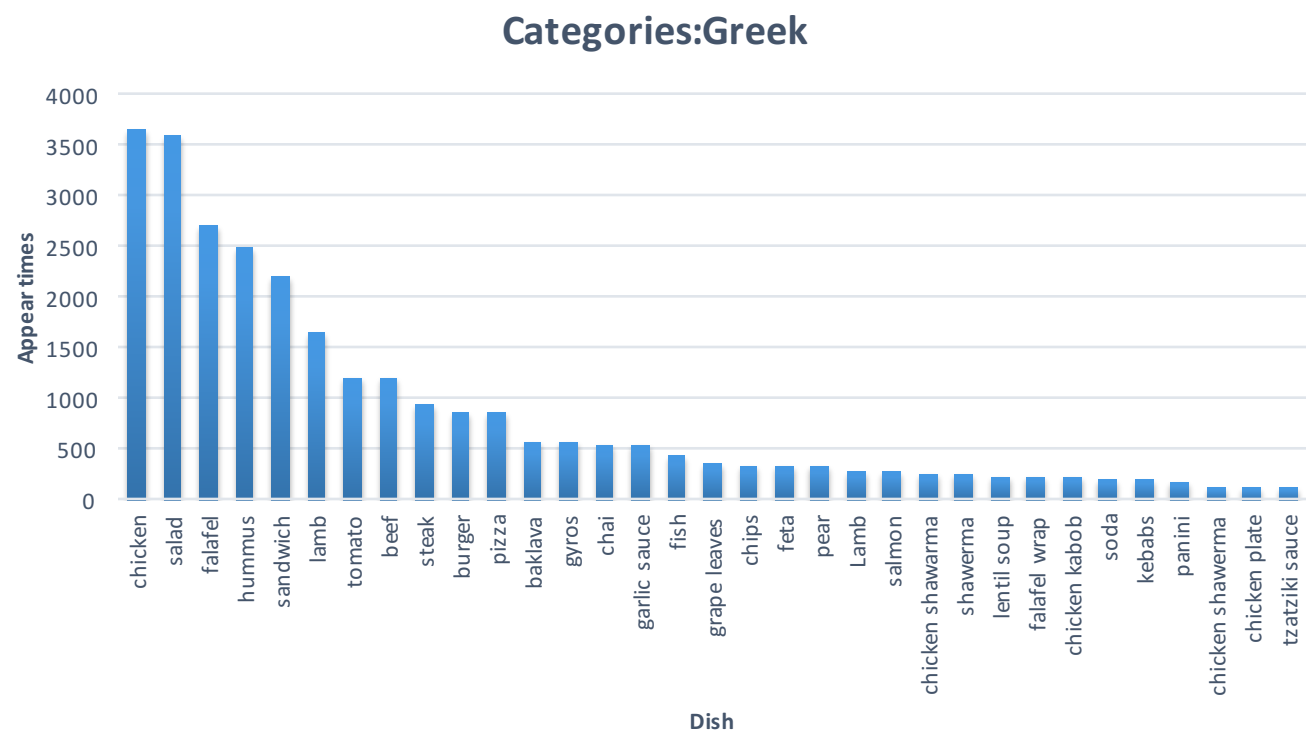
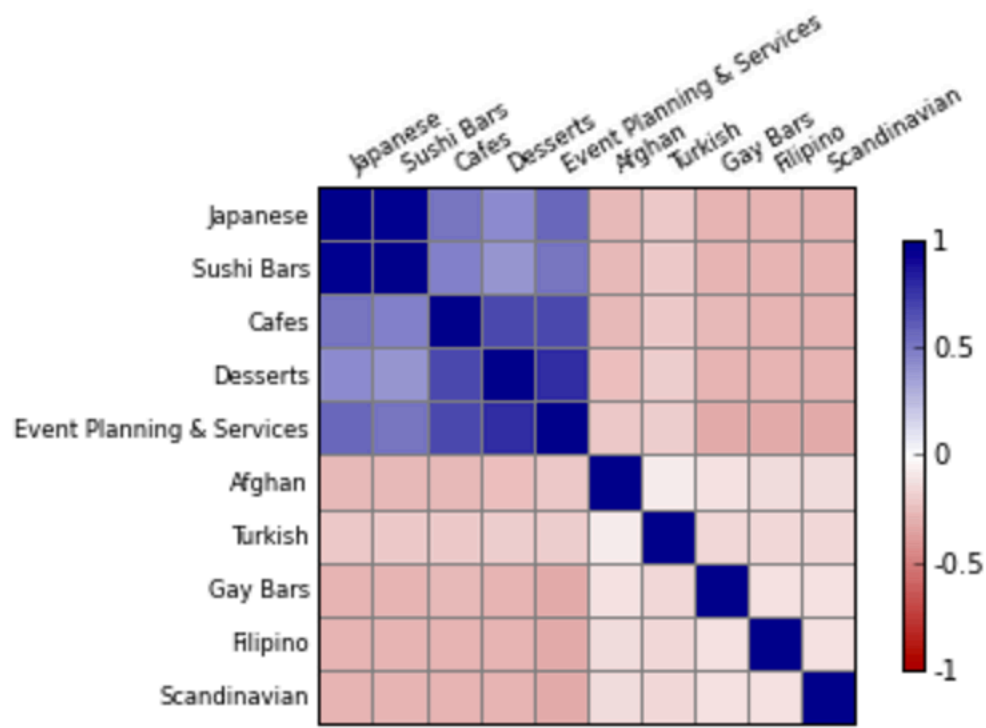
- Request Yelp academic challenge data from Yelp.
- Pre-processing data: split raw data to different objects
- Using tf-idf, document term frequency and other method to process the data.
- Using the LDA model to get the topics from reviews.
- Using K-means and other algorithms to cluster the data.
- Using similarity matrix to see the different kind of cuisines.
- Mining the favorite dishes.

# Data

- The raw data is yelp\_academic\_dataset.json which is 346.6 Mb.
- The raw data include three objects: user, business and review.
- There are 4597 business objects, 213670 review objects.
- We split the raw data to three files based on the objects.
- For convenience , we randomly made a sample reviews with 100000 entries for testing.
- After mining the dishes from review, we remove some entries from list.

# Results

Similarity between different cuisines and most popular dishes in Greek:



# Conclusion:

- We used K-means clustering to cluster the cuisine reviews and calculated the similarity between them.
- Based on the mentioned times, we found the most popular dishes in specific cuisines.
- But the accuracy was not very high when we mined the dishes from reviews.
- From our project:
  - It is necessary to find a higher quality pattern mining method to discover dishes.
  - For future work, we can do more sentiment analysis on reviews for dishes.