

Introduction

Yelp has been a popular website for people to review businesses, and get suggestions from other users on new businesses to discover. In this project we tried to build a recommendation system based on the information we got from the dataset.

Goals

- Cluster businesses based on similarity
- Build a network graph of related users
- Build a recommendation system to suggest businesses to a given user

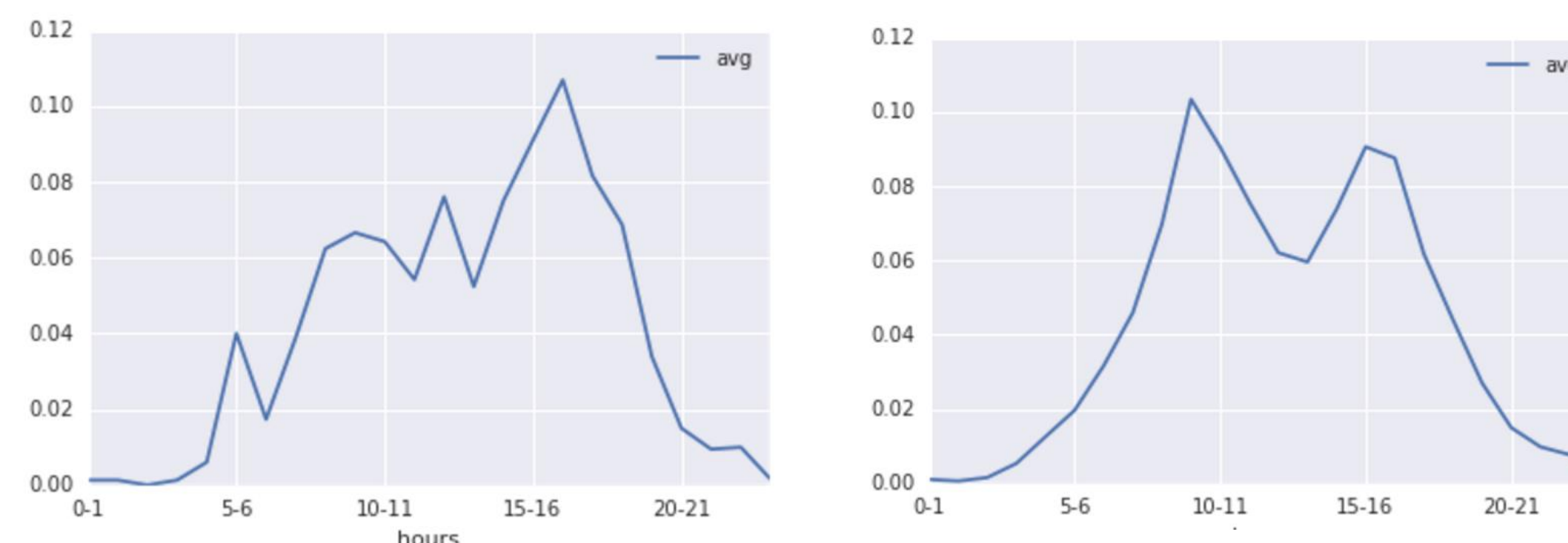
The dataset

A large dataset of multiple cities was provided by Yelp as part of a challenge, we only gathered the information we felt are necessary for our project, these were:

- Business related: business_id, name, address, attributes, check-in times
- User related: User_id, current friends user_ids.
- Review related: User_id, business_id, rating.

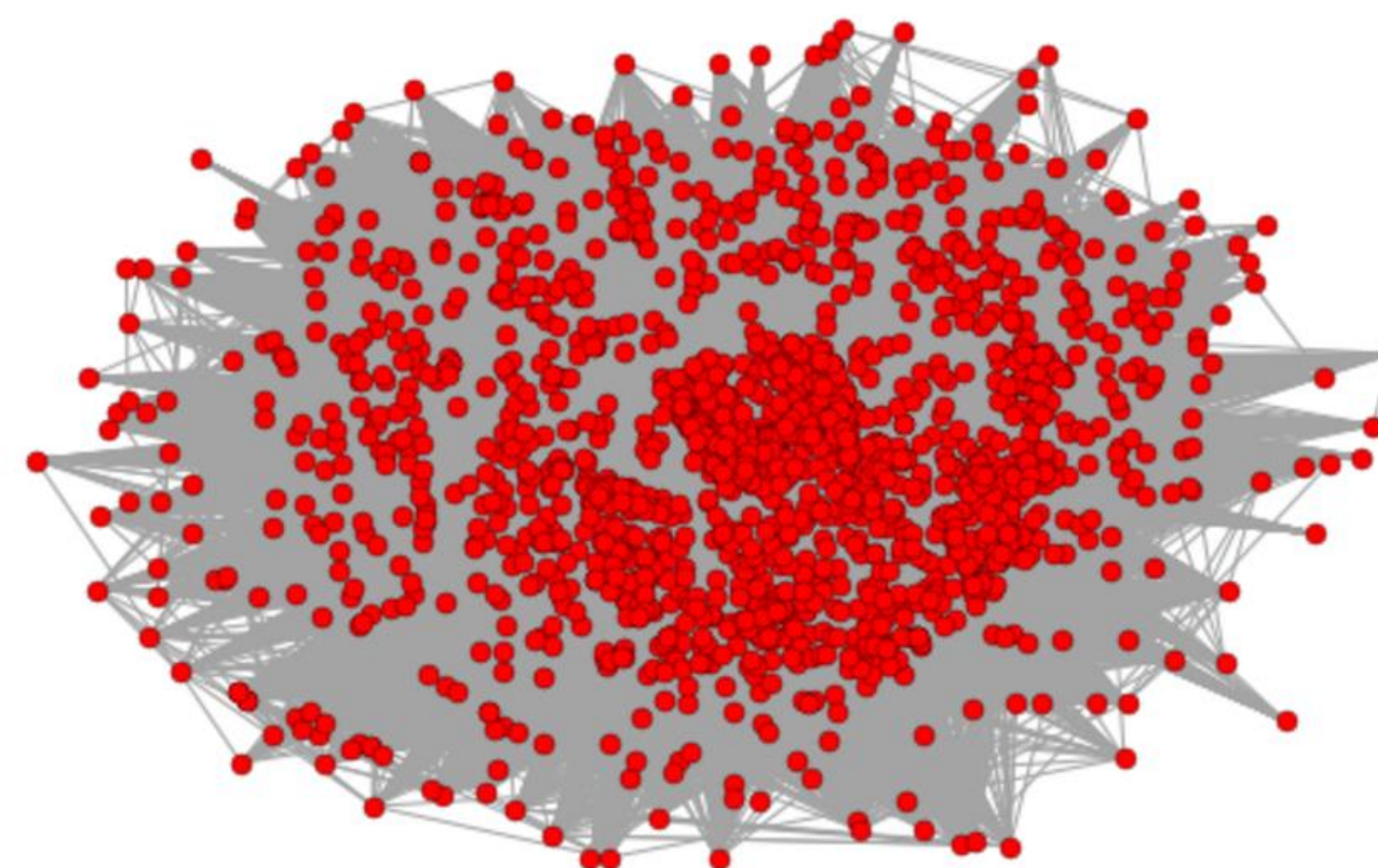
Techniques & results

I – Clustering businesses based on Check-in times:



2 different clusters of average check-in time series

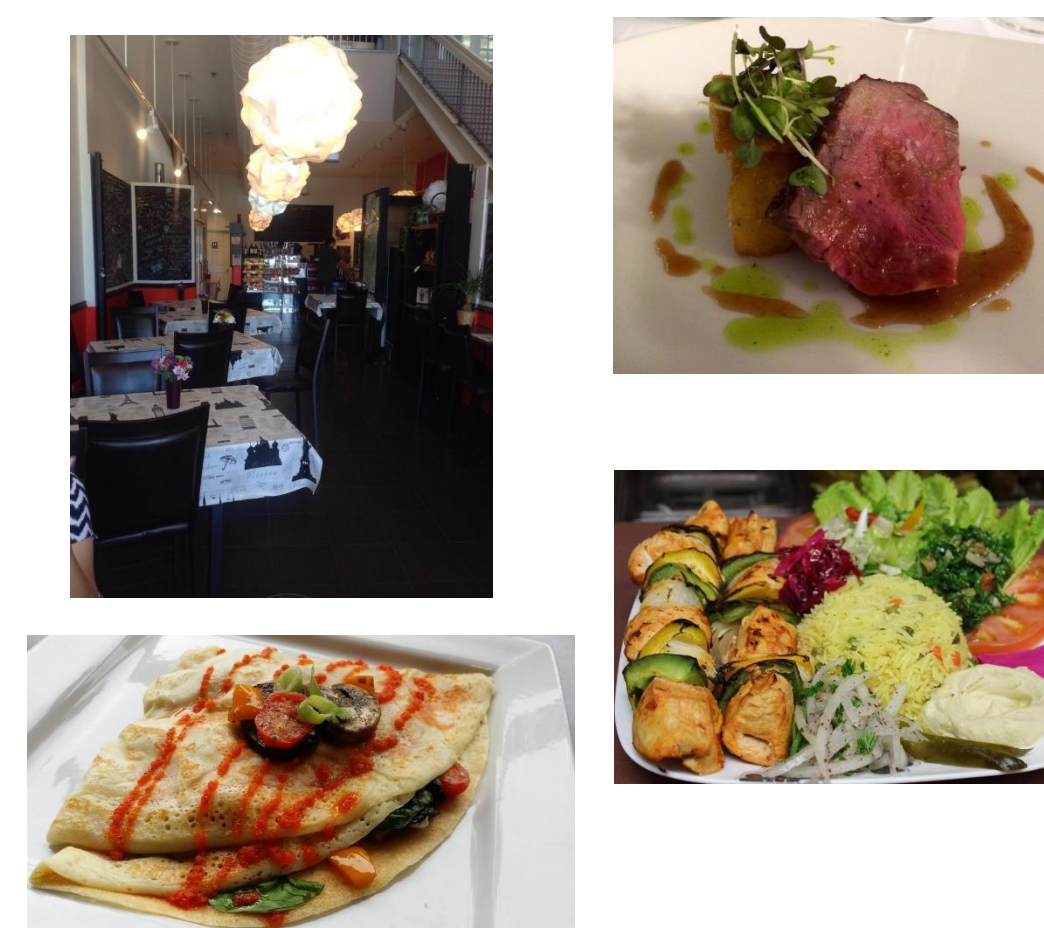
II – Creating a network of users:



A graph network of 1664 users in Waterloo with 600K edges between them

III – Recommending businesses:

We got user-based recommendation according to review history (in any given city).



Evaluation

To evaluate the recommendations we gave, we evaluated our rating prediction algorithm against the existing rating in our dataset. We used root-mean-square error (RMSE) to compute the error and it was 1.2 which is pretty good.

Conclusion

We implemented the recommending system according to the user's review and their connection to other users. We also evaluated our predictions according to existing data using root-mean-square error. Our system could be used to recommend restaurants in different cities and it could be used in travel website like TripAdvisor.

Future Work

- With more computing resources we could build a larger network of more users
- Build a recommendation system for users who can be potential friends