

Restaurant Recommendation System

 Janhavi Kashyap

March 3, 2021

INTRODUCTION

Recommendation engines are systems that make recommendations tailored to fit a user's preference and thereby improve user experience. In this project, we aim to build a recommendation system that recommends restaurants with the Yelp academic dataset.

APPROACH

Collaborative filtering: Adopting a matrix-factorization based approach that decomposes users and restaurants into a set of latent factors, taking a global view to ratings

This project is divided into three parts:

01

DATA CLEANING AND PREPROCESSING

- Loading the data
- Dropping insignificant columns
- Converting the dates and time into datetime objects
- Keeping only the columns where a business was open
- Creating a restaurants dataframe from the business dataset pertaining to only restaurants.

02

EXPLORATORY DATA ANALYSIS

Performing EDA on:

- Business dataset: Explored the restaurant data-subset by location as well as restaurant characteristics.
- User, check-in, review and tips datasets for restaurants
- Additional analysis: Tried to find the influence of user-based review systems like Yelp; analyzed restaurant characteristics from a Chain vs non-chain restaurant pov

03

MODELLING

- Built a restaurant recommendation system with collaborative filtering
- Matrix-factorization was employed
- SGD and Adam optimizers were used to fit the model

METHODOLOGY

THE DATA

- Converted the available code: `json_to_csv` convertor on Yelp's Github from python2 to python3 to load our dataset.
- The restaurant data was merged with the user data to give a total of more than 4 million data points. It had to be reduced to around a million consisting of data from only the restaurants that have more than fifty reviews for SGD.
- The most recent of the reviews were used as the test set, the second most recent as the validation and the rest of the data was used for training.

IMPLEMENTATION

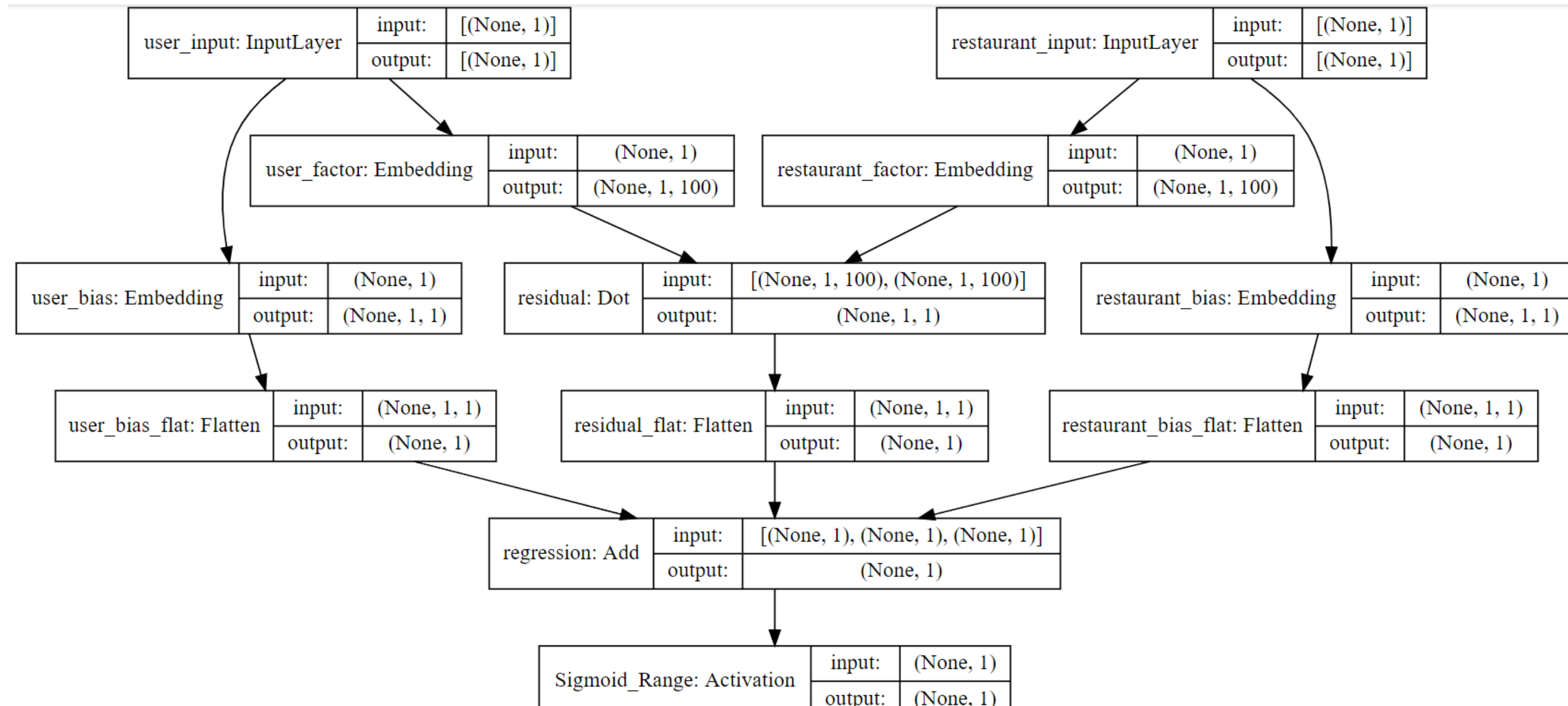
Four functions were used for improving the recommendations:

- **predict_ratings:** takes a userid as an input and predicts ratings for all the restaurant which are not rated by that user.
- **recommend_restaurants:** takes userid as input and recommends restaurants based on predicted ratings.
- **recommend_based_on_friends:** calculates the average rating of all the friends of a user on the restaurants they have rated and recommends restaurants to a user based on that average ratings.
- **recommend_based_on_location:** recommends the user the restaurants using the above functions but also takes a user's current location into consideration.(Only recommends those restaurants which are located in users current location)

RESULTS

Adam MSE: **1.7752**

SGD MSE: **2.9112**



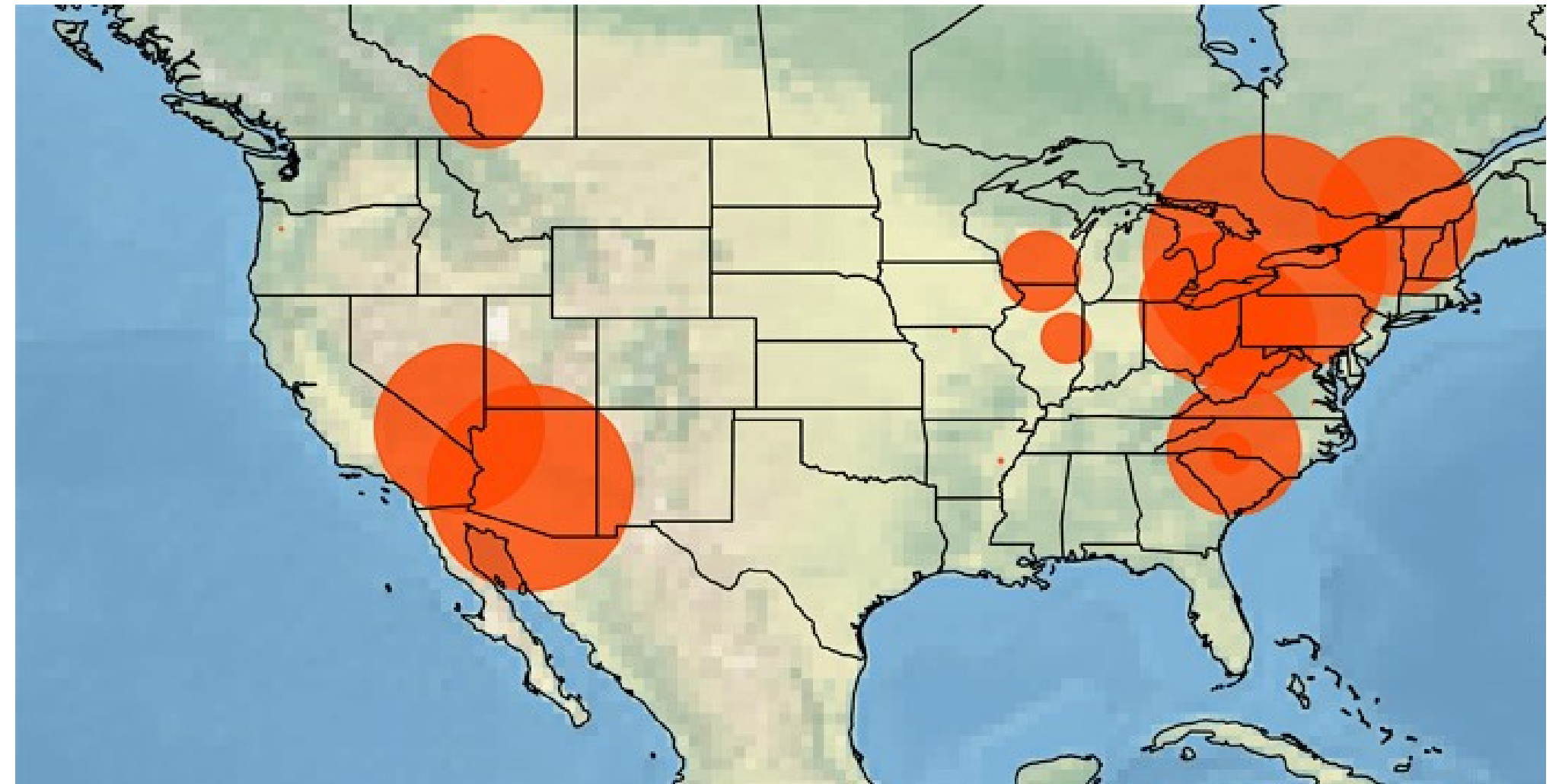
Model architecture

OBSERVATIONS

- Adam gave better scores on the test data compared to SGD
- The predicted ratings we got were all below 4 probably due to not having taken into consideration more features such as attributes from the business dataset that would have helped improve it

RESULTS

- The largest number of reviewed are from **Toronto, Ohio**
- **Las Vegas, Nevada** has the most top-rated restaurants
- **Bacchanal Buffet, Las Vegas** is the most reviewed restaurant
- **Earl of Sandwich, Las Vegas** is the best restaurant on the basis of review counts and ratings.



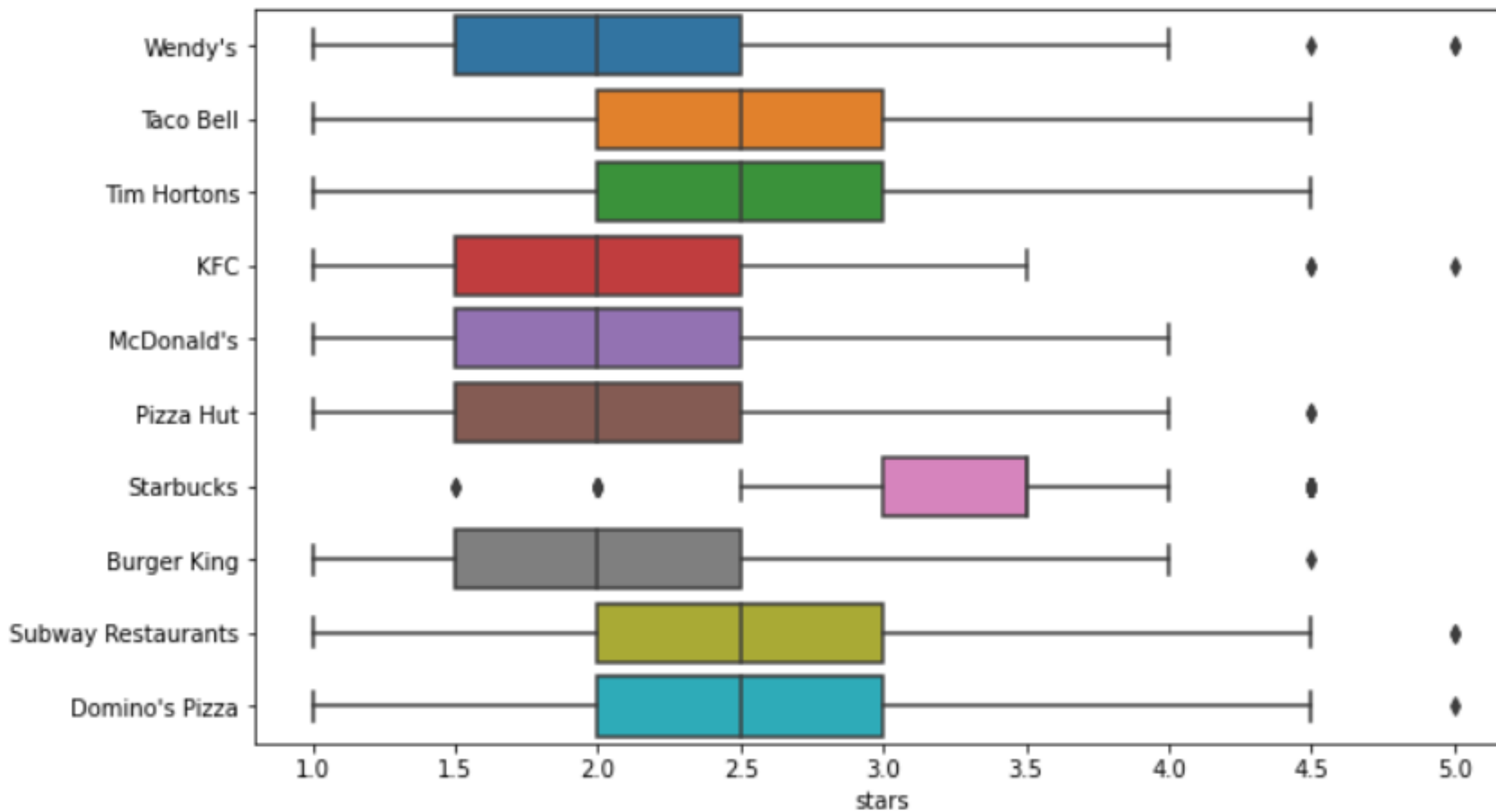
B A C C H A N A L
B U F F E T
CAESARS PALACE



RESULTS

- The restaurant with the most number of storefronts is **McDonald's**
- The most reviewed restaurant among the biggest chain restaurants is **McDonald's**

name

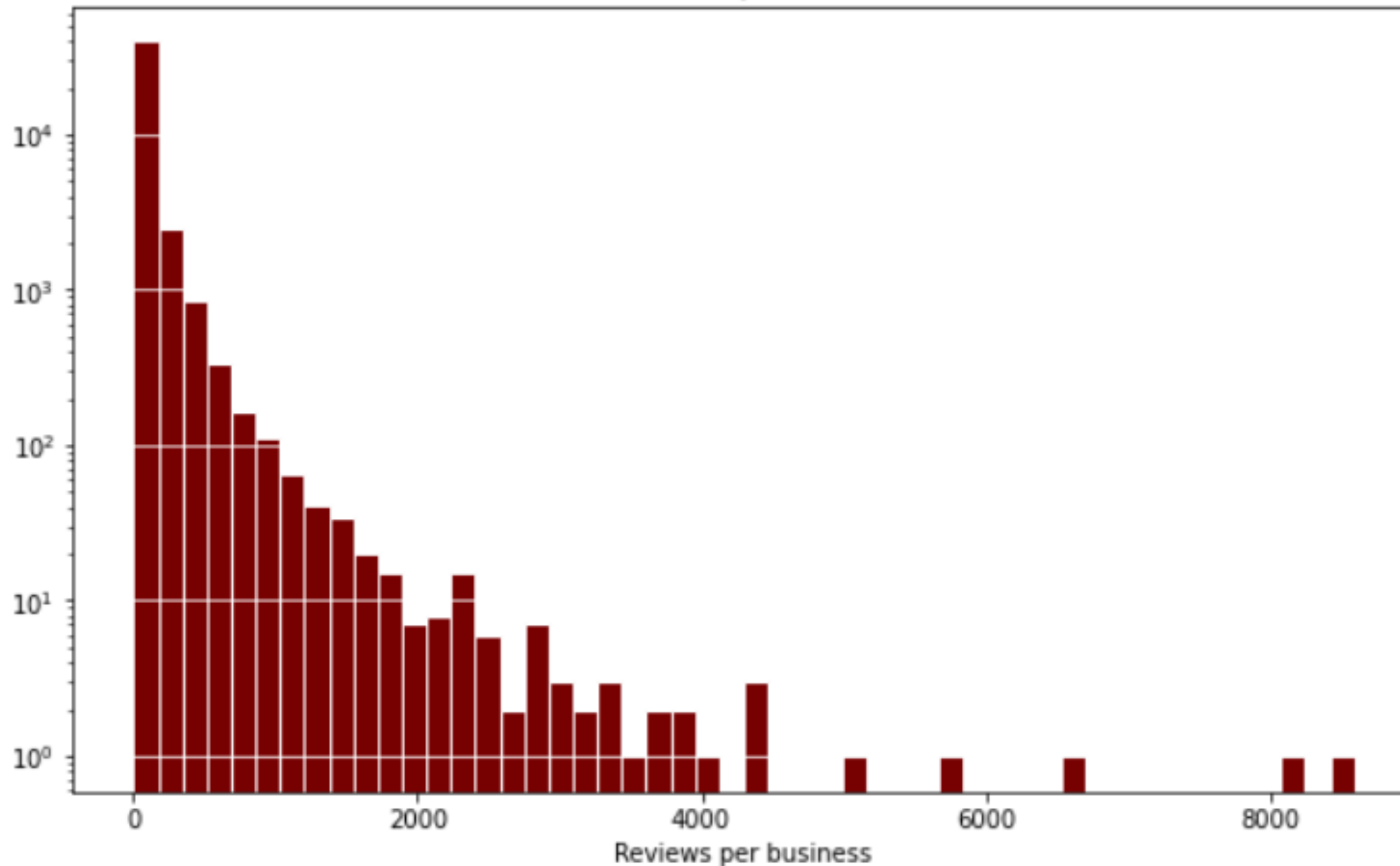


Despite being the most rated, McDonald's has an average rating of **1.9 stars** only which appears to be the case for other big restaurant chains as well.



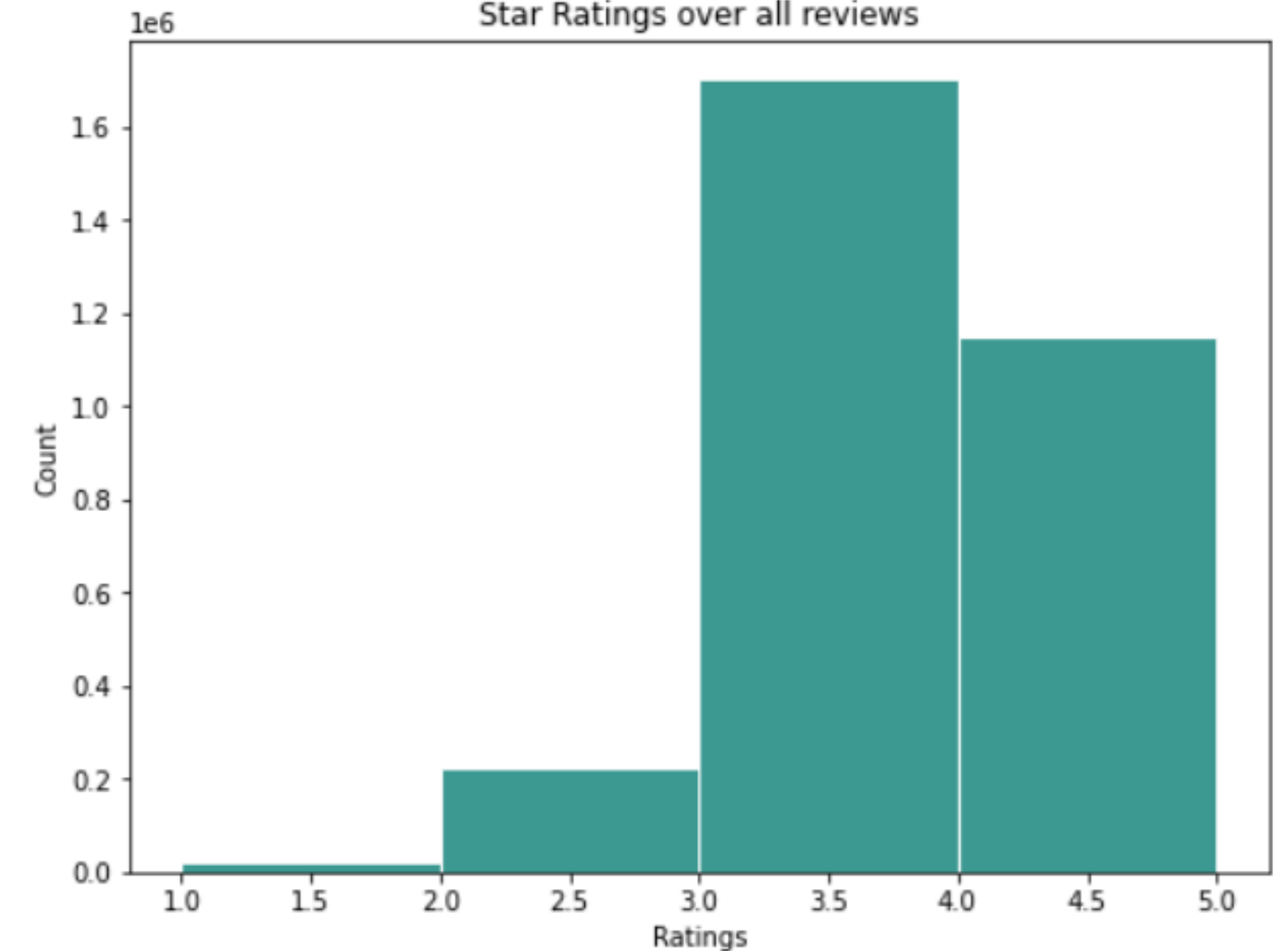
RESULTS

Review Count per Restaurant



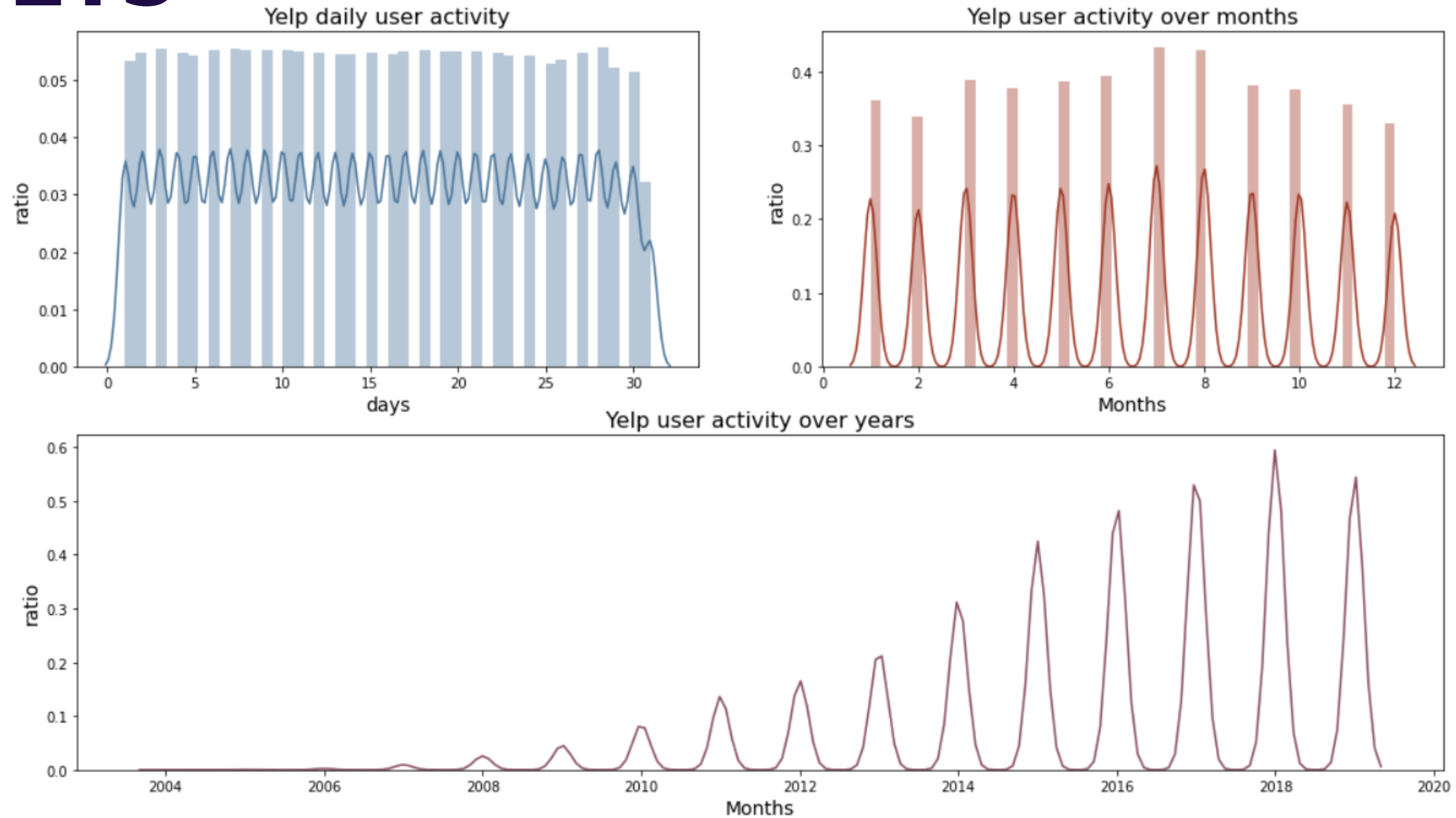
It can be observed that very few of all the restaurants have been rated by the users and despite taking into account only the restaurants with more than 50 reviews, we still have very little data from a user-item perspective.

Star Ratings over all reviews



The figure shows that the maximum user ratings lie between 3 and 4 stars.

RESULTS



A consistent user-base indicates that **Yelp is indeed a popular place to leave online reviews**

CONCLUSION

01

I
—

The ratings for SGD can be improved by using geo-location, attributes such as vegan, alcohol etc which further narrow down the preference and hence improve the predicted rating of the user

02

II
—

A comparison with user-user, item-item and user-item can help further analyze why our MSE from SGD and Adam is low.

03

III
—

Most reviews are for restaurants in Las Vegas, perhaps because of it being a popular city to have fun. Our predictions would be biased towards Sin City because it's home to the most well-rated restaurants