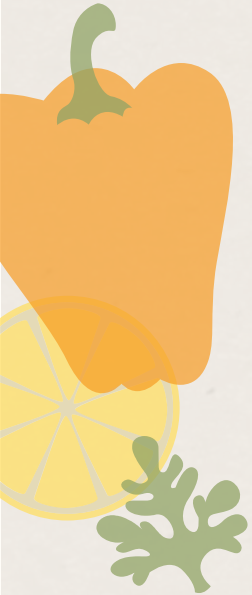


Restaurant Recommendation System

NLP Project - Group 24

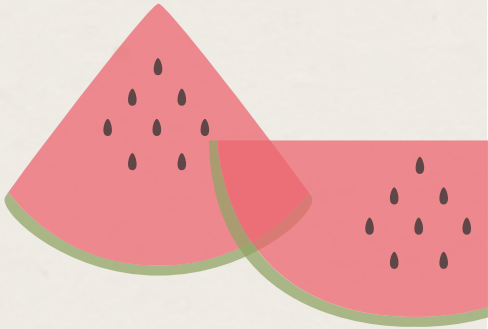


Table of Contents

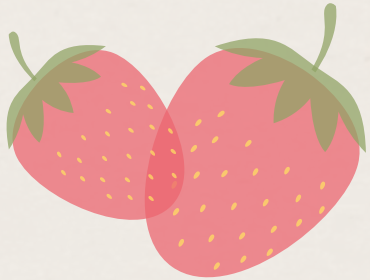
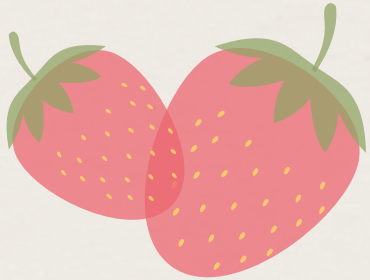


Table of Contents



01 Introduction

Introduction

- We aim to build a recommendation system that will enable us to make sophisticated restaurant recommendations for users by applying learning algorithms to develop a predictive model of customers' restaurant ratings.
- This will retrieve restaurants based given cuisine as the input.

02 Dataset

Dataset

- Primary Dataset : Yelp Open Dataset
- The Yelp dataset is a subset of Yelp's businesses, reviews, and user data that has been made publicly available for use for personal, educational, and academic purposes.
- This dataset contains 5,200,000+ reviews, 174,000+ businesses and 11 metropolitan areas.
- Yelp Dataset Source

Dataset

File Name	Description
business.csv	Details about all the business ids
business_hours.csv	Details about the weekly timings of the businesses
reviews.csv	User reviews for the businesses
users.csv	List of all the users and their details
tip.csv	Details about small compliments given by the users
checkin.csv	Mapping of business with date

Business Data

Field	Description
business_id	business_id
name	business name
city	City located in
state	State located in
latitude	latitude
longitude	longitude
stars	Star rating
is_open	Open or not
categories	List of categories of business

Review Data

Field	Description
review_id	review_id
user_id	user_id
business_id	Business_id
stars	Star Rating
text	Review text
date	Review given on this date
useful	No of useful votes received
funny	No of votes received for the review being funny

03 Objectives

Objectives

Predict Ratings

To predict rating of restaurants listed in the Yelp dataset based on the reviews given by the users.

Recommend Restaurants

Recommending restaurants to the users using the predicted stars and sentiment polarity values.

Graphical User Interface

A GUI using Tkinter which takes cuisine as input from the user to predict 10 best restaurants.

04

Pre-processing

Open & Closed businesses

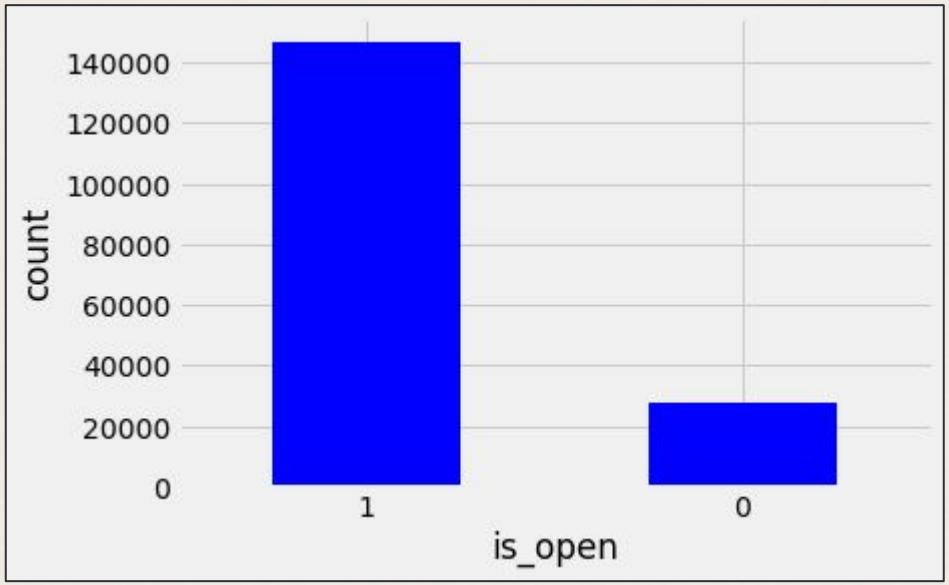


Fig 1. Plot to show no of open and closed businesses

- Total - 1,74,567 businesses
- Open - 1,46,702 businesses
- Closed - 27,865 businesses

Thus, consider only the businesses that are open.

Restaurants in States

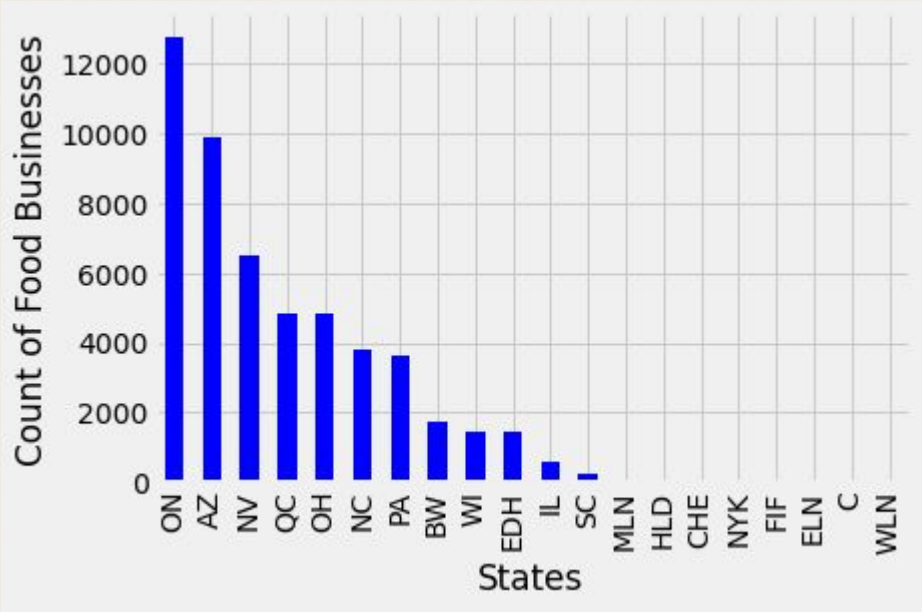


Fig 2. Plot to show the count of food businesses or restaurants in each state

- ON - Ontario State is having the highest number of food/restaurant related businesses.

  Thus we consider only the restaurants located in ON.

Cuisine related restaurants

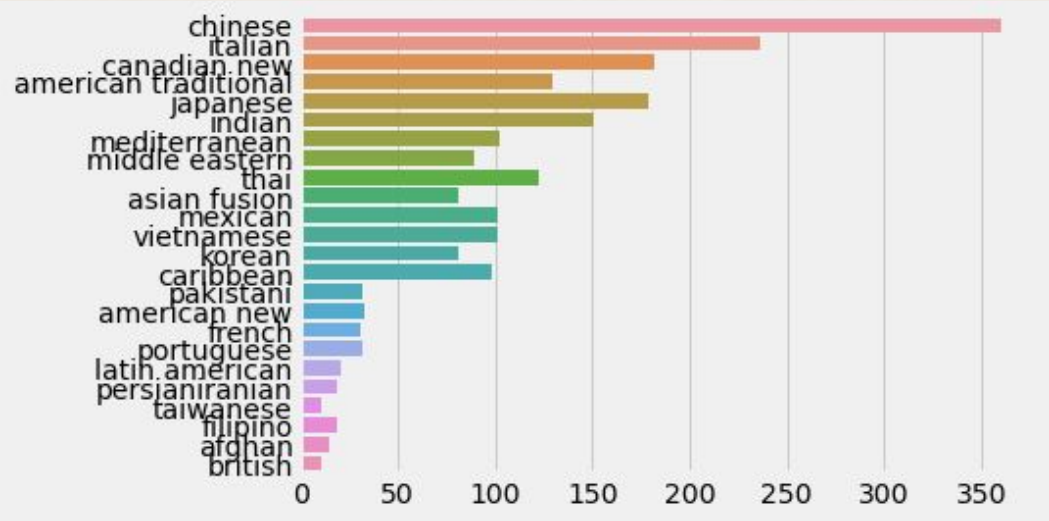


Fig 3. Plot to show the count of each cuisine based businesses

- Considering only the top cuisine related restaurants

Categories split

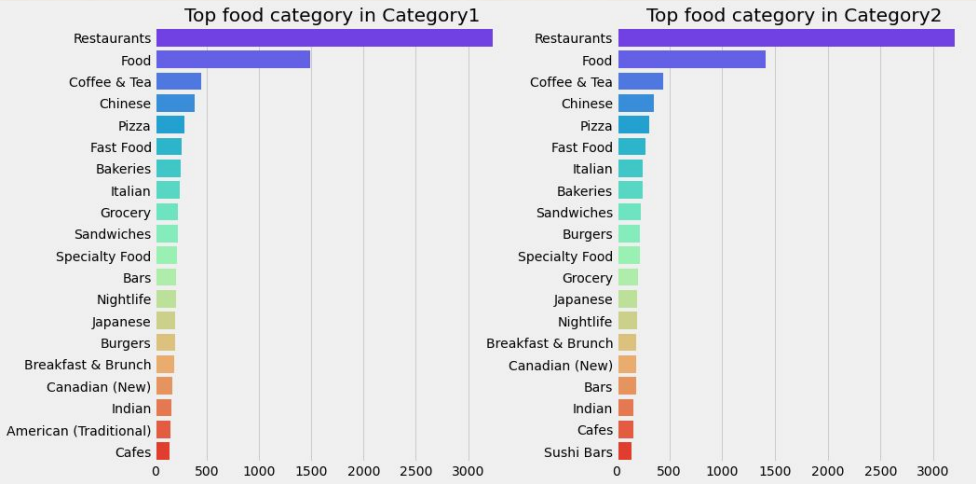
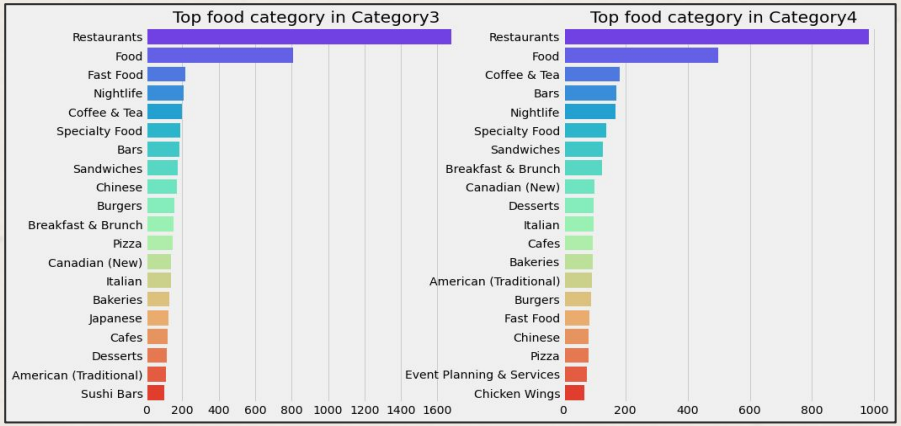
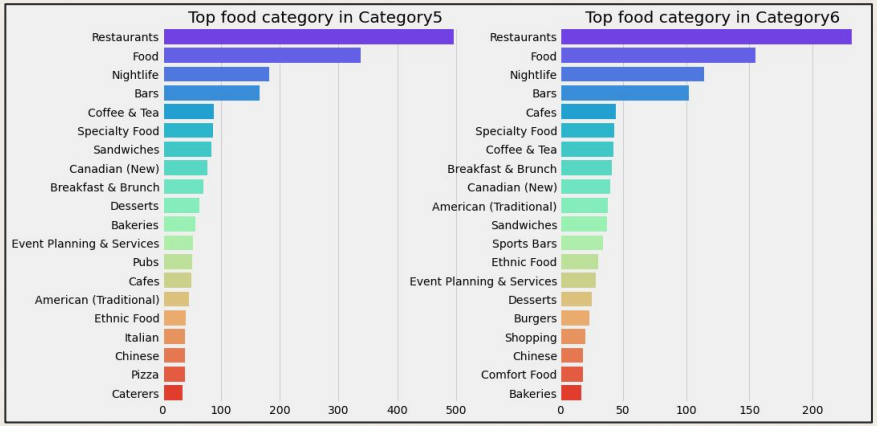


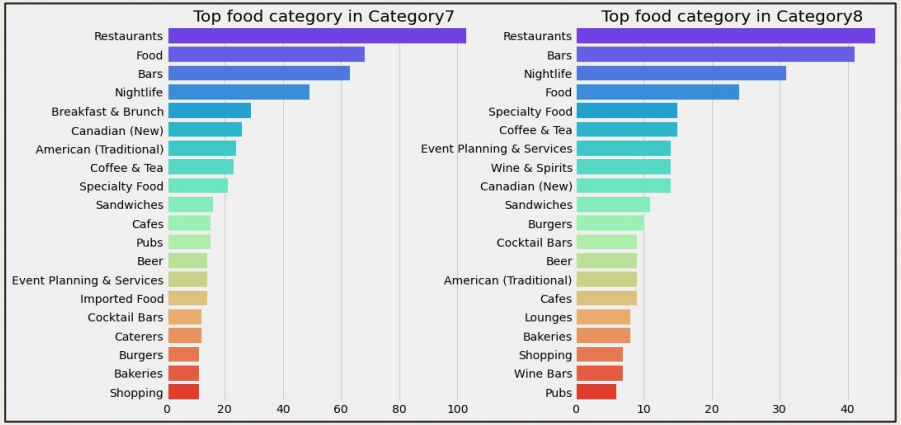
Fig 4. Plot to show the count of different categories

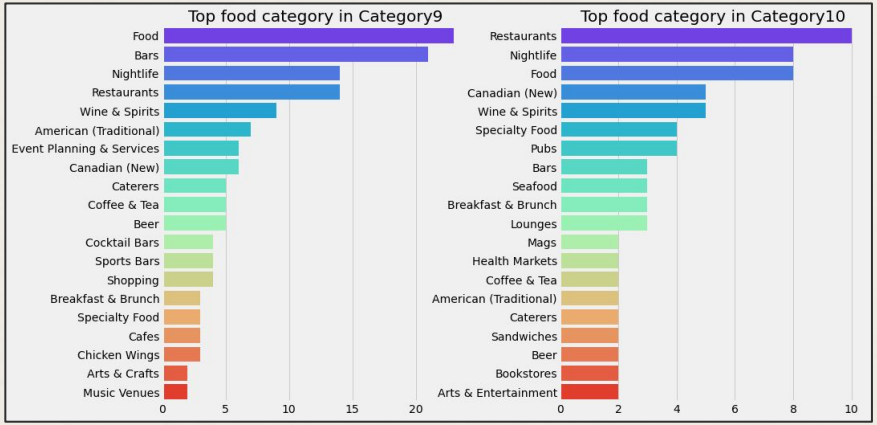
- The categories column in each business consists of a minimum of one category and a maximum of 10 categories.
- Thus, have split the categories into category1, category2, etc for analysis.

Categories split









Review Counts of cities

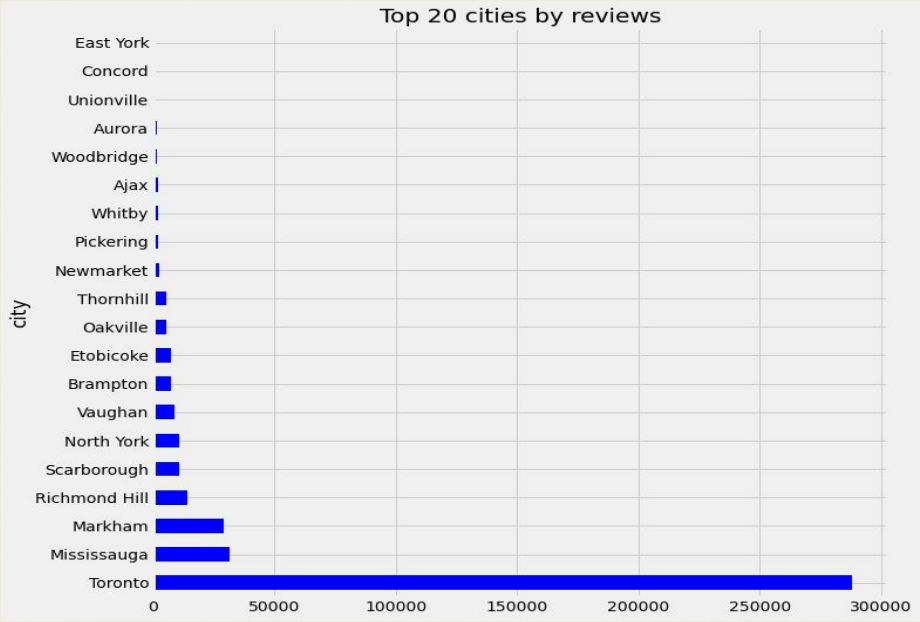


Fig 5. Plot to show the count of user reviews in top 20 cities of ON (Ontario)

- Top 20 cities by reviews
- Toronto city in Ontario has the highest number of food or restaurant related reviews.

Process the text

- Convert all the text in the categories, reviews and cities to lowercase.
- Also, pre-process the reviews
  - Perform case normalization
  - Remove punctuations and stopwords
  - Return cleaned text as list of words

```
y=preprocess('My hubby likes the apple pies there; crust is delicious.')
y
```

```
'hubby likes apple pies crust delicious'
```

Fig 6. Processed text output for a given string

WordCloud

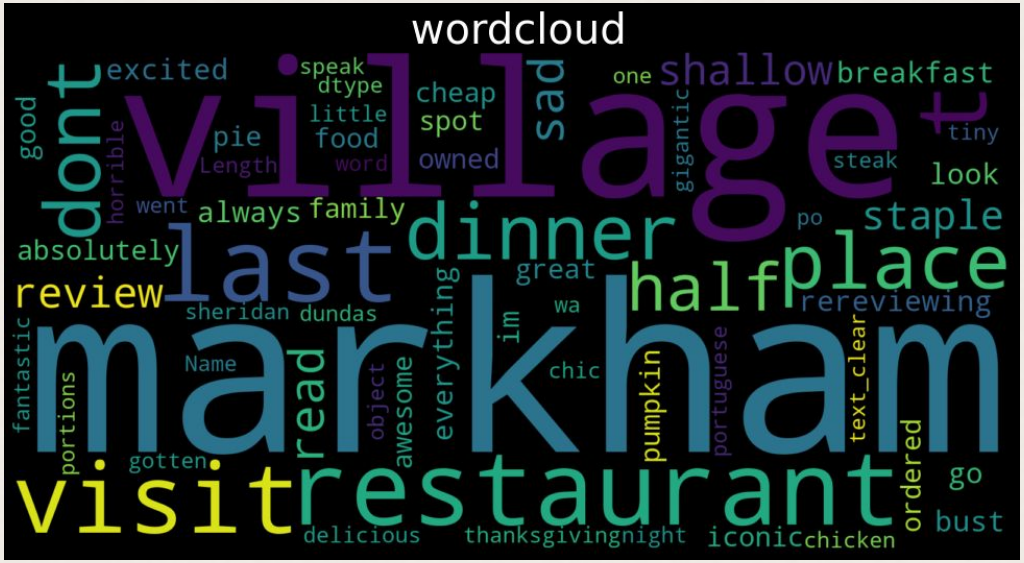


Fig 7. Word cloud for the most frequently used words in the user reviews.

05

Predict Ratings

Data Preparation

- To predict the ratings of the reviews, we need to take the text of each review and stars rating.
- So load them into two different lists maintaining their corresponding index.
- Vectorize the complete data using CountVectorizer
- Split it into train and test data.

Gaussian Naive Bayes

- P(“food was great”) = P(“food”)*P(“was”)*P(“great”)
- For each class c ,

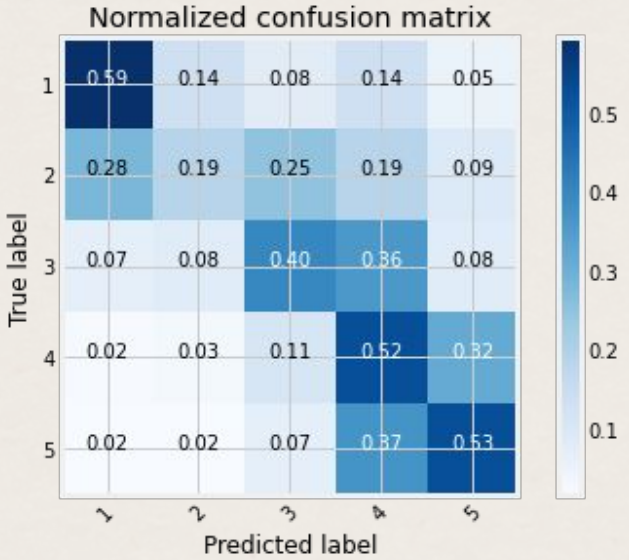
  P(“food was great” | c) = P(“food” | c)*P(“was” | c)*P(“great” | c)
- The class with maximum posterior probability is considered

$$P(x_i \mid y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$

Classifier and Predictions

- Gaussian Naive Bayes classifier was used to classify reviews into the five star ratings.
- The model gives an accuracy of 0.40

	precision	recall	f1-score	support
1	1.00	0.44	0.62	9
2	0.00	0.00	0.00	6
3	0.24	0.18	0.21	22
4	0.41	0.74	0.53	62
5	0.40	0.12	0.18	52
accuracy			0.40	151
macro avg	0.41	0.30	0.31	151
weighted avg	0.40	0.40	0.35	151



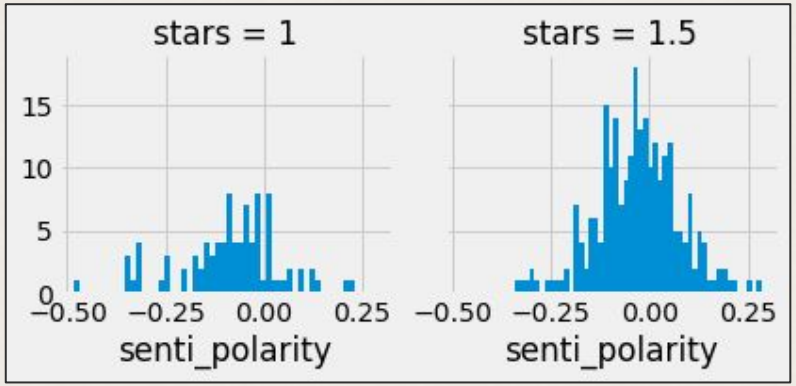
06

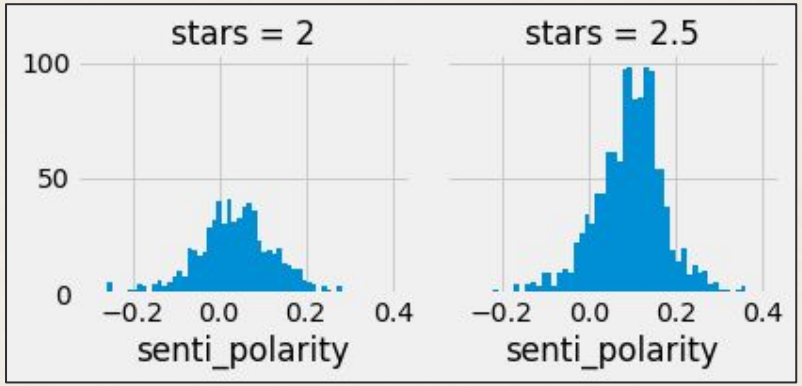
Recommend Restaurants

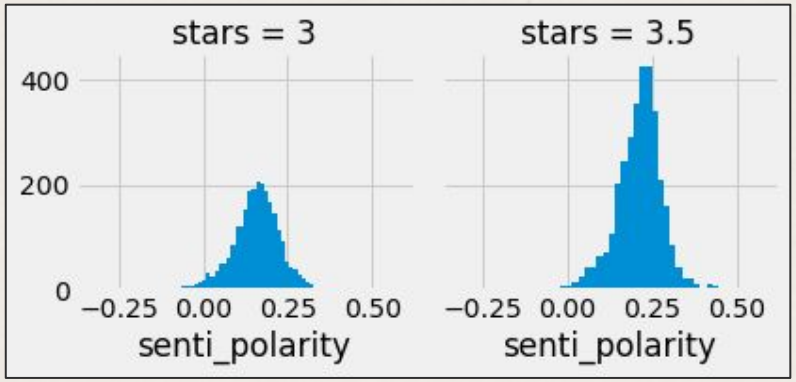
Compute Sentiment Polarity

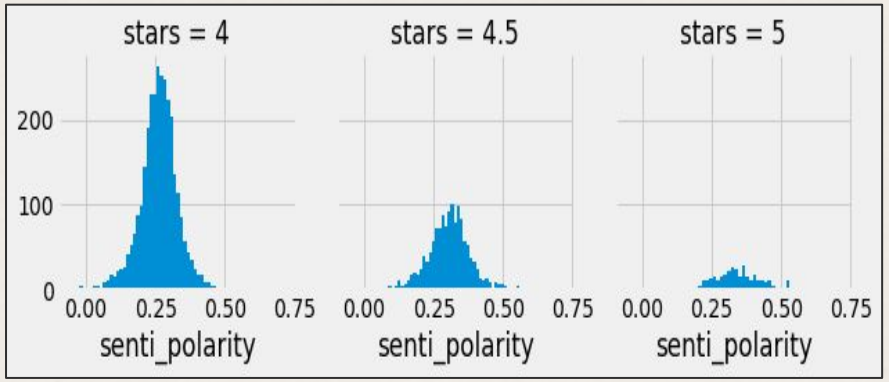
- For each of the processed reviews, sentiment polarity value was computed.
- These were grouped by business_id and the mean of it were appended as a new field in the dataframe.
- The polarity values range between -1.0 to 1.0.
- Then, joined the updated data frame to the restaurants data on business_id.

Sentiment Polarity vs Ratings









Sentiment Polarity vs Ratings

- Stars more than 3.5 are all distributed between the range 0 and 0.5 of sentiment polarity values.
- Consider only those business_ids that have values greater than 3.5 stars and polarity values greater than 0.
- This data can be used to recommend the restaurants to the users.

Indexer

- A term to document index was maintained.
- Here, cuisine to list of business_ids were stored.
- All the business_ids were sorted in descending order of their sentiment polarity values.
- Thus, given a cuisine (term), related restaurants were retrieved efficiently.

07 Results

Highest Rated Restaurant

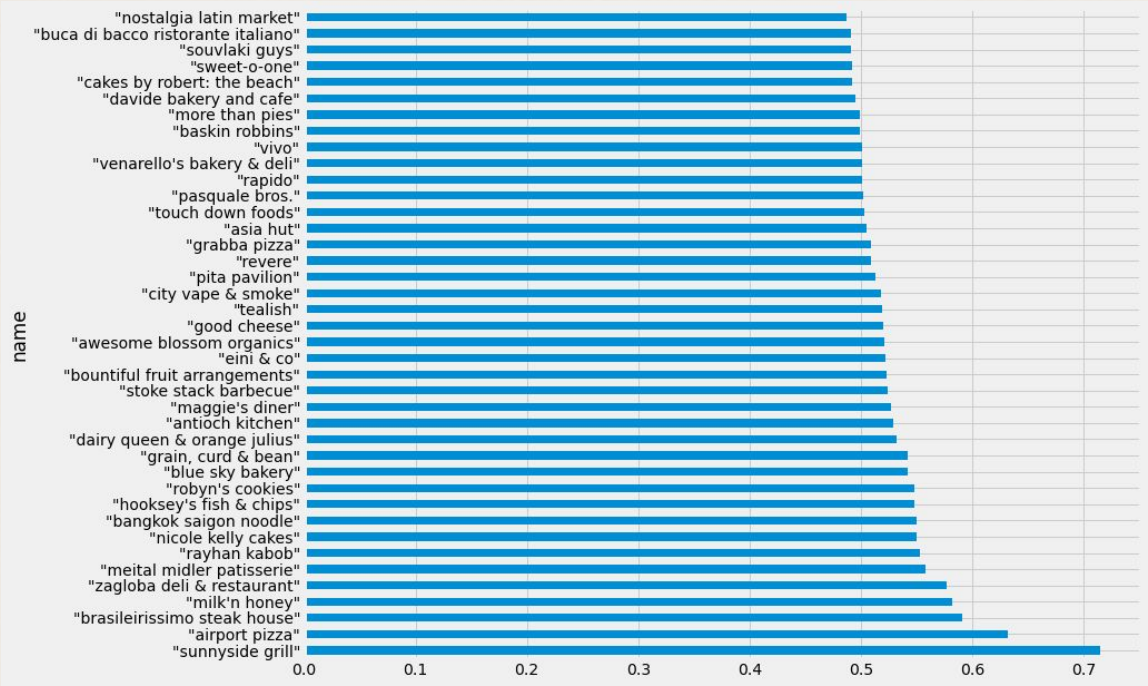


Fig 8. Plot for top 40 restaurants alongside polarity values

- Irrespective of the cuisine, the restaurant with maximum positive reviews (highest sentiment polarity values) is **SunnySide Grill**

Best Restaurants for Indian cuisine

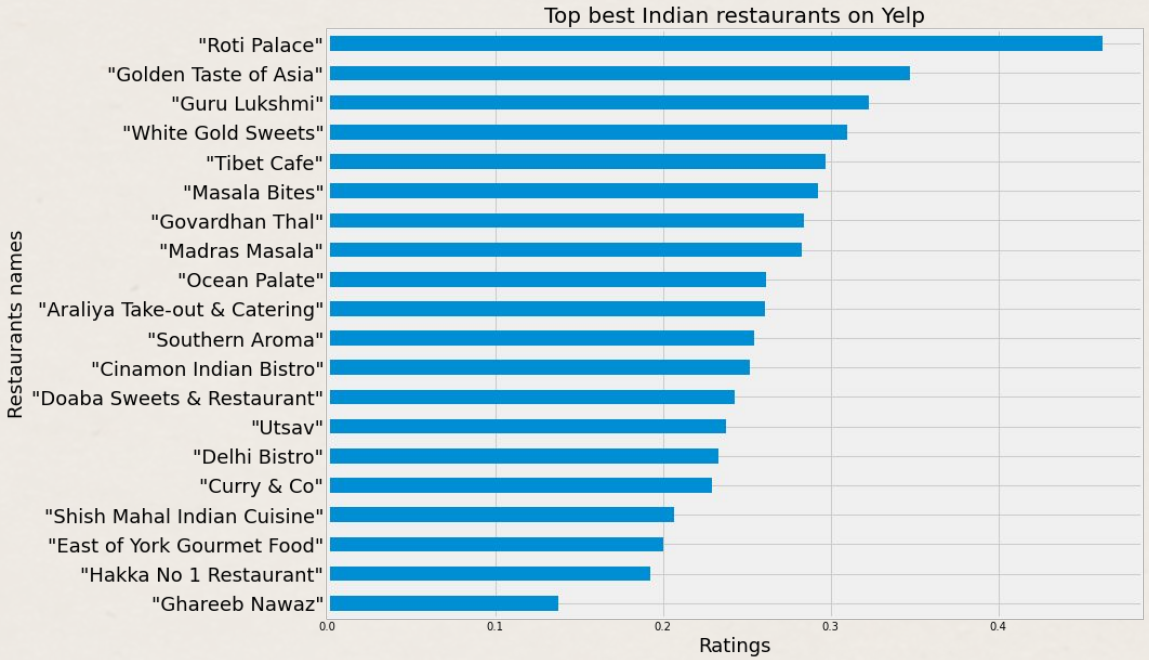


Fig 9. Plot for top 20 Indian restaurants alongside ratings

- For Indian cuisine, **Roti Palace**, is the best recommendation.

Best Restaurants for Italian cuisine

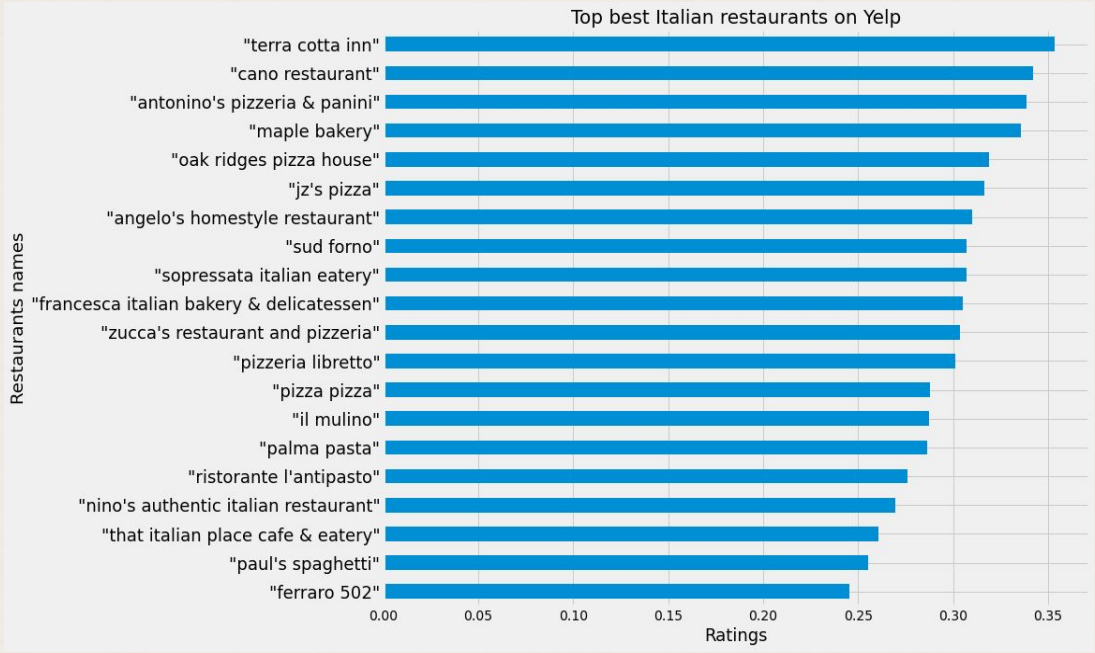


Fig 10. Plot for top 20 Italian restaurants alongside ratings

- For Italian cuisine, **Terra Cotta Inn**, is the best recommendation.

Top 10 Restaurants for Thai Cuisine

	name	address	city	state	senti_polarity
100	"bangkok saigon noodle"	"136 holland street e, unit 1"	bradford	on	0.550298
4635	"thai wok n roll"	"30 broadleaf ave"	whitby	on	0.480556
3881	"thai cuisine experts"	"6970 lisgar drive, unit 2a"	mississauga	on	0.416845
3902	"bolton thai cuisine"	"18 king street e, unit l5"	caledon	on	0.396744
1887	"feng wok 'n' roll"	"28 dundas street e, unit 4"	mississauga	on	0.383282
4549	"thai express"	"178-200 windflower gate"	vaughan	on	0.378685
2766	"silk restaurant"	"446 parliament street"	toronto	on	0.359119
1530	"pho tai loi"	"1550 s gateway road, suite 1b"	mississauga	on	0.359035
538	"river tai restaurant"	"92 harbord street"	toronto	on	0.358654
4234	"bach yen"	"738 gerrard street e"	toronto	on	0.343499

Fig 11. Final recommendation of restaurants for thai as input

When a user enters a cuisine (**Thai**, here) top 10 restaurants are displayed in descending order of their ratings.

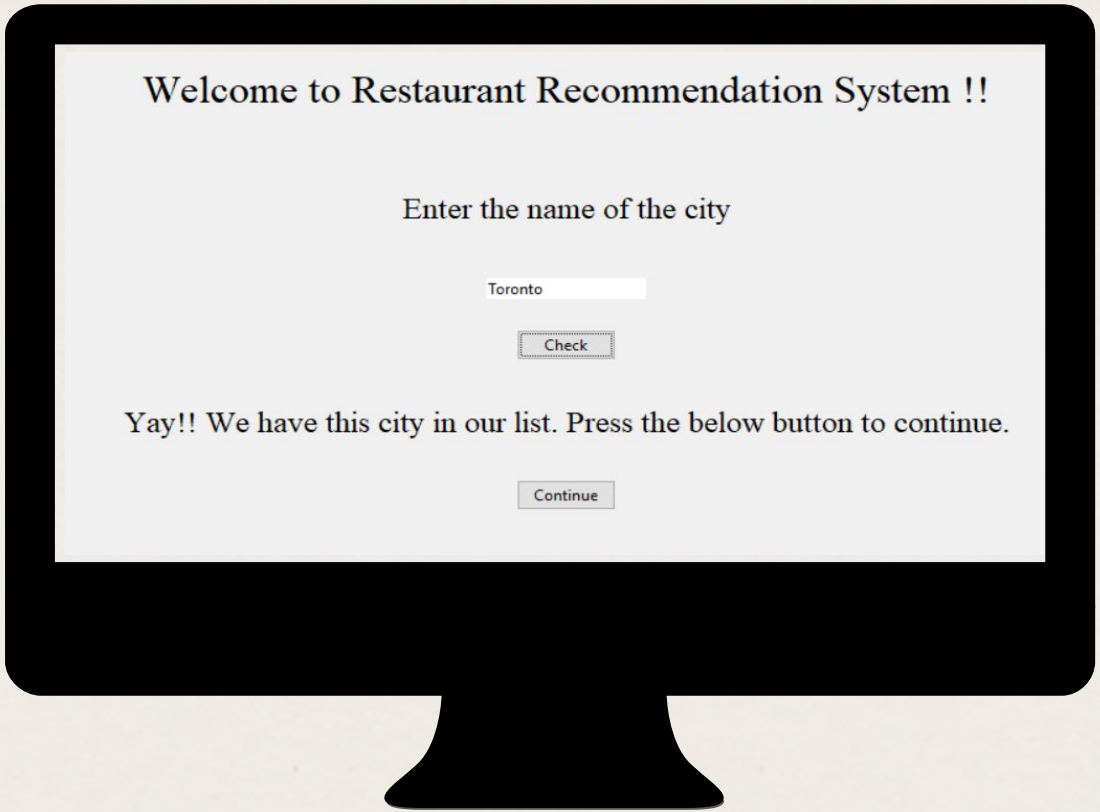
08

GUI Integration

GUI (Tkinter)

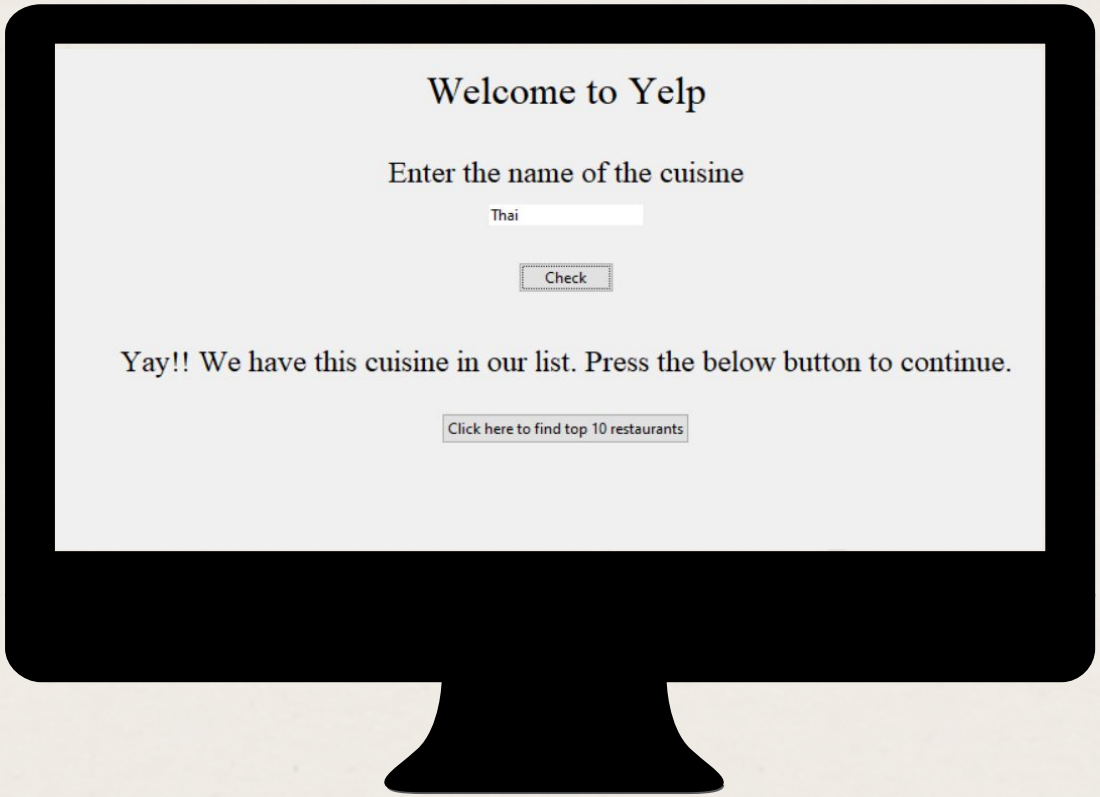
This is the start point of the UI. It first asks for the city as the input.

Since the city exists in Ontario, we can continue.



GUI (Tkinter)

Then we give Thai as the cuisine for input. Since, the cuisine exists in our list, we proceed to view the top 10 restaurants.



GUI (Tkinter)

The top 10 restaurants.

	name	address	city	state	senti_polarity
100	"bangkok saigon noodle"	"136 holland street e, unit 1"	bradford	on	0.550298
4635	"thai wok n roll"	"30 broadleaf ave"	whitby	on	0.480556
3881	"thai cuisine experts"	"6970 lisgar drive, unit 2a"	mississauga	on	0.416845
3902	"bolton thai cuisine"	"18 king street e, unit l5"	caledon	on	0.396744
1887	"feng wok 'n' roll"	"28 dundas street e, unit 4"	mississauga	on	0.383282
4549	"thai express"	"178-200 windflower gate"	vaughan	on	0.378685
2766	"silk restaurant"	"446 parliament street"	toronto	on	0.359119
1530	"pho tai loi"	"1550 s gateway road, suite 1b"	mississauga	on	0.359035
538	"river tai restaurant"	"92 harbord street"	toronto	on	0.358654
4234	"bach yen"	"738 gerrard street e"	toronto	on	0.343499

09 Contributions

Our Contributions

Suma Shreya

- Data cleaning & filter based on cuisines
- GUI using Tkinter

Hemanth Pasupuleti

- Indexer using tf-idf
- Naive Bayes Classifier

Rukmini Meda

- Recommending restaurants using polarity values

10

Team Members

Our Team

Suma Shreya
S20180010170

Hemanth Pasupuleti
S20180010131

Rukmini Meda
S20180010102

Thanks

CREDITS: This presentation template was created by **Slidesgo,** including icons by **Flaticon** and infographics & images by **Freepik**.

Please keep this slide for attribution