Culinary Tourism

Food / Restaurant Recommender

Group 4

Aniketh Satyanarayana, Sejoon Park, Swetha Vijaya Raju, Yong Zhao

Agenda

Problem Statement

Model Formulation and Deployment

Introduction

Project Scope

Impact

Food & Culture

Food and culture are intertwined and central in a person’s identity. Although there are various cultures around the world, similar types of food exist between different nations. The “melting pot” of America allows different types of cuisines to be available in metropolitan cities. We plan to investigate how a person’s food preferences can help identify other cuisine foods with similar features or even restaurants with similar ambiance that the customer enjoyed.

Problem Statement

- Build a food and restaurant recommendation system based on similar food features and ambience related features that the customer enjoyed
- Analyze the sentiment of different restaurants based on the reviews.
- Help businesses providing more helpful recommendations that the customer will enjoy

Project Scope

1. A Search Engine model where one gives features and entities to look for restaurants / food
2. A recommendation system that clusters similar reviews to identify similar food and ambience features.
3. Sentiment analysis of the restaurant reviews and to predict the rating of a restaurant based on its reviews
4. A food glossary of different cuisines to look explore different culture
5. A QnA system that answers a given question based on the reviews

Datasets

- Yelp Business
  - 150k rows
  - Split business citywise
  - Got the top 10 cities in the dataset
- Yelp Reviews
  - 7 million rows
  - If business_id in reviews dataset, then write the rows line by line to 10 different files.

		restaurant_count
city	state	
Philadelphia	PA	14567
Tucson	AZ	9249
Tampa	FL	9048
Indianapolis	IN	7540
Nashville	TN	6968
New Orleans	LA	6208
Reno	NV	5932
Edmonton	AB	5054
Saint Louis	MO	4827
Santa Barbara	CA	3829

Santa Barbara

	business_id	name	categories	review	latitude	longitude	total_rating	rating
269194	zxW8zECvT_SqejieMMjb5A	Chilango's Mexican Restaurant	Restaurants, Mexican	Sabor! \n\nChilangos is part of the Milk&Honey...	34.416792	-119.695894	3.5	4.0
269195	zxW8zECvT_SqejieMMjb5A	Chilango's Mexican Restaurant	Restaurants, Mexican	The salsa roja has improved! WOOHOO!! And now ...	34.416792	-119.695894	3.5	4.0
269196	zxW8zECvT_SqejieMMjb5A	Chilango's Mexican Restaurant	Restaurants, Mexican	This review is solely based on one time, eatin...	34.416792	-119.695894	3.5	5.0
269197	zxW8zECvT_SqejieMMjb5A	Chilango's Mexican Restaurant	Restaurants, Mexican	Very fresh tasty food. Has been in business f...	34.416792	-119.695894	3.5	4.0
269198	zxW8zECvT_SqejieMMjb5A	Chilango's Mexican Restaurant	Restaurants, Mexican	mmm decent mexican & cheap on state street!!! ...	34.416792	-119.695894	3.5	5.0

269630 rows × 9 columns

Search Engine

A Search Engine model where one gives features and entities to look for restaurants / food .

POS Tagging

Steps:

- Tokenize reviews
- Clean review tokens
- Convert clean tokens to string
- Use Spacy for,
    - tagging adjectives and nouns
    - extracting noun_chunks
- Map adjectives and nouns to respective noun_chunks

POS Tagging

"

Ssooooo slow. Waited 20 min for the server to even come over to us. Forgot our drinks. Then was offered my check before i even got any of my food.. ......and we never even got everything we ordered. When i did get my food there was a hard layer on top from sitting in the window..... Also they were out of fettuccine noodles ??

"

"

ssooooo slow waited min for the server to even come over to us forgot our drinks then was offered my check before i even got any of my food and we never even got everything we ordered when i did get my food there was a hard layer on top from sitting in the window also they were out of fettuccine noodles at an italian restaurant

"

POS Tagging

"

ssooooo slow waited min for the server to even come over to us forgot our drinks then was offered my check before i even got any of my food and we never even got everything we ordered when i did get my food there was a hard layer on top from sitting in the window also they were out of fettuccine noodles at an italian restaurant

"

Adjectives

{'fettuccine', 'hard', 'italian'}

Adjective mapping

```
{
 'hard': ['a hard layer'],
 'italian': ['an italian restaurant'],
 'fettuccine': ['fettuccine noodles']
}
```

Nouns

{'check', 'drinks', 'food', 'layer', 'min', 'noodles', 'restaurant', 'server', 'top', 'window'}

Noun mapping

```
{'noodles': ['fettuccine noodles'],
'top': ['top'],'drinks': ['our drinks'],
'server': ['the server'],
'restaurant': ['an italian restaurant'],
'layer': ['a hard layer'],
'window': ['the window']}
```

Bag of Words

Adjectives

{

‘clean’, ‘cute’, ‘spicy’,
‘hot’, ‘cold’, ‘salty’,
'sanitary', 'lurk', 'habit',
'dilled', 'unflattering',
'fascinating', 'emerald',
'rockford', 'shabby', 'quiet',
'pager', 'chewed',
'piquante',

'hellbound', ...

}

Nouns

{

‘pakoras’, ‘food’, ‘tacos’,
‘burger’, 'noodles',
'competition', 'basura',
'sparks', 'augh', 'cycles',
'mashawi', 'chillers',
'fraternities', 'fowls',

'stunk', 'sono’, ...

}

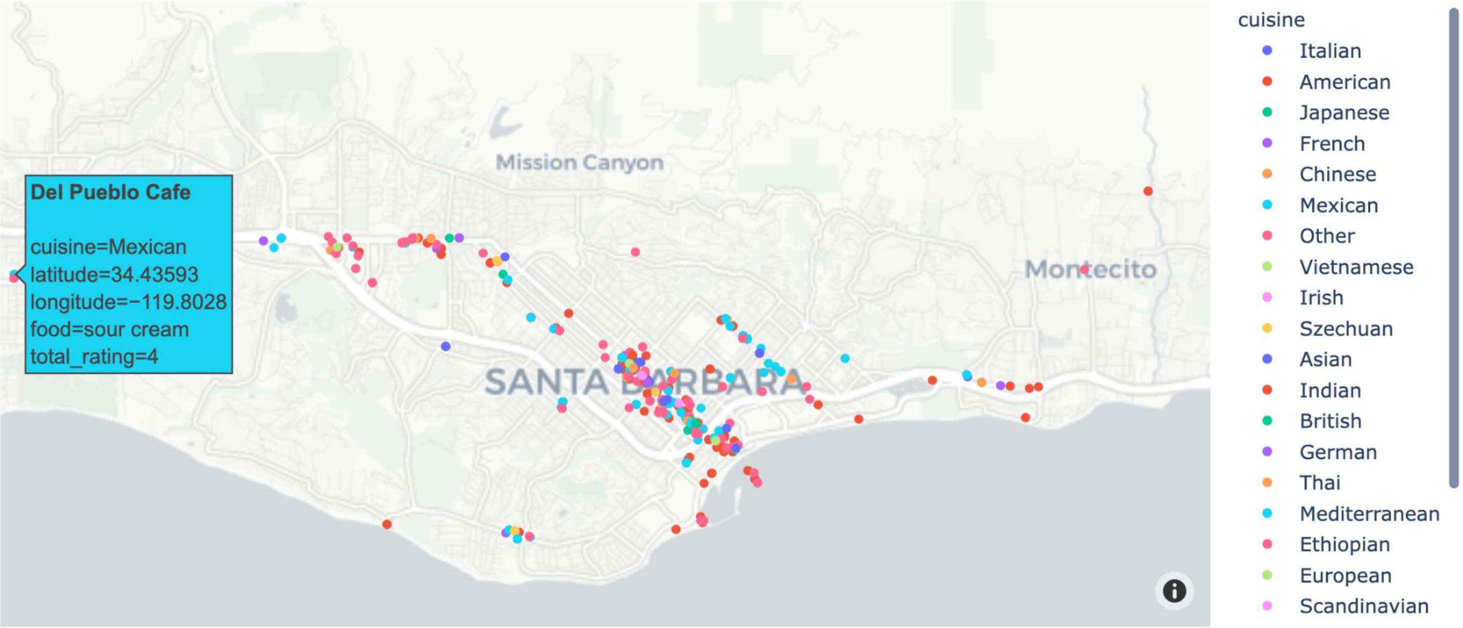
Search Engine

Steps:

- Enter any feature / food item
- Checks if the given input is in adjective or noun bag of words
- Geoplots the resulting restaurants with the given the feature or food item in the reviews
- The restaurants are colored based on the cuisine type
- Enter feature: “sour”, “spicy”, “cute”, “burger”

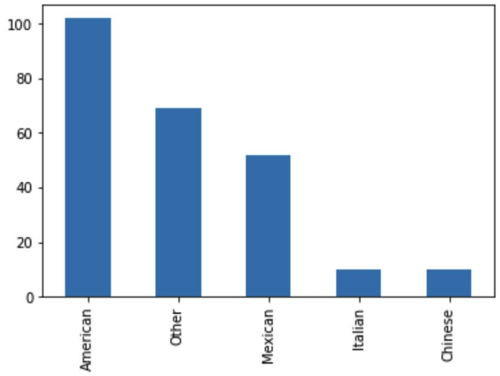
Search Engine

Input : “sour”

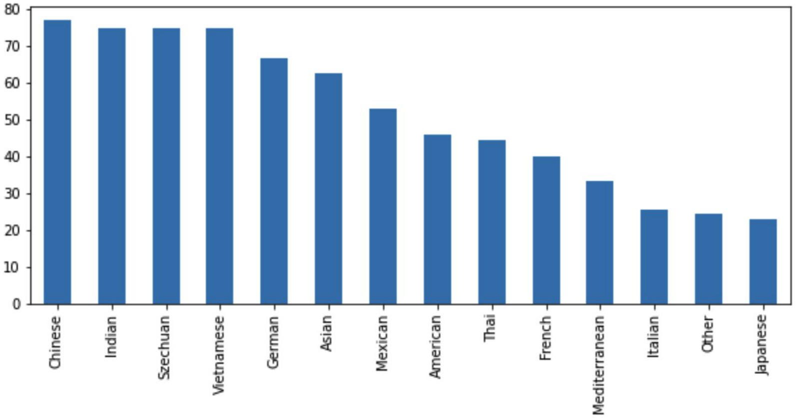


“Sour” cuisines

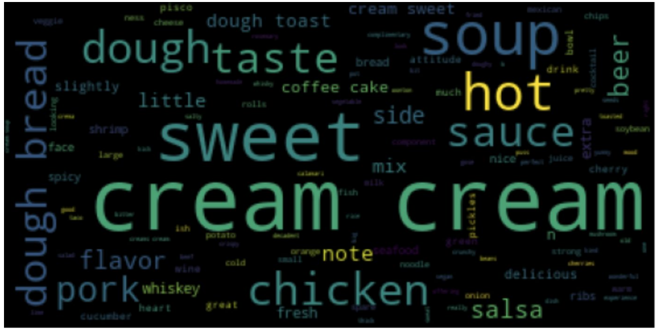
Total count



Normalized

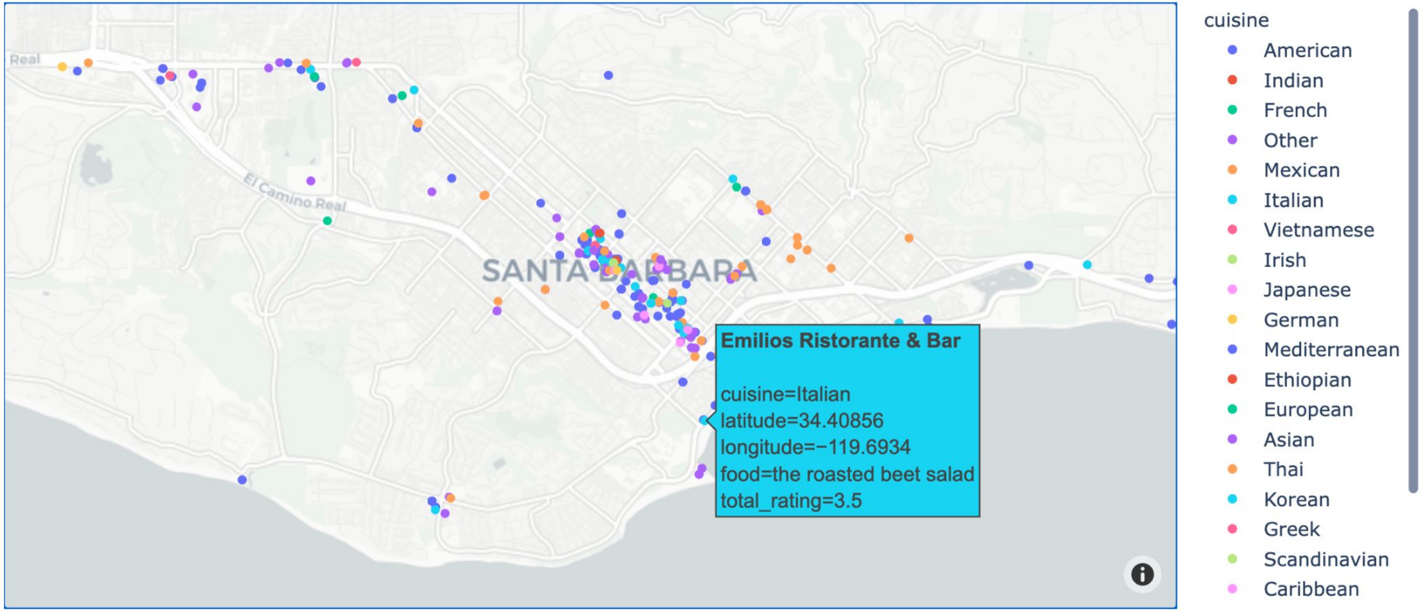


“Sour” cuisines



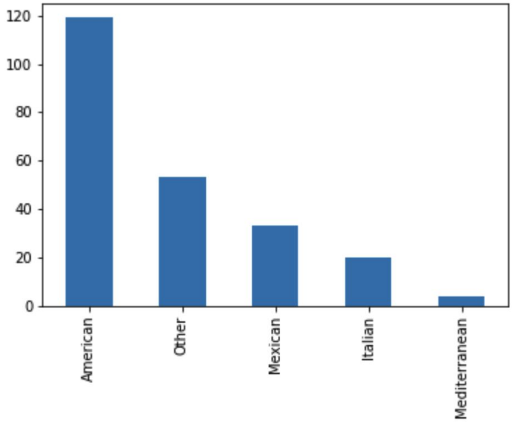
Search Engine

Input : “roasted”

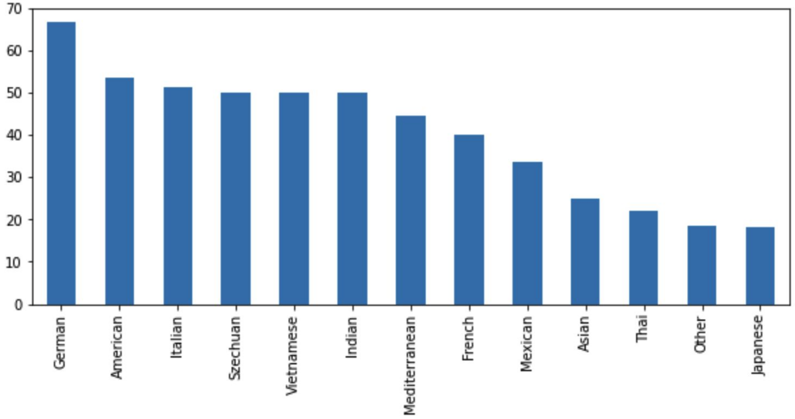


“Roasted” cuisines

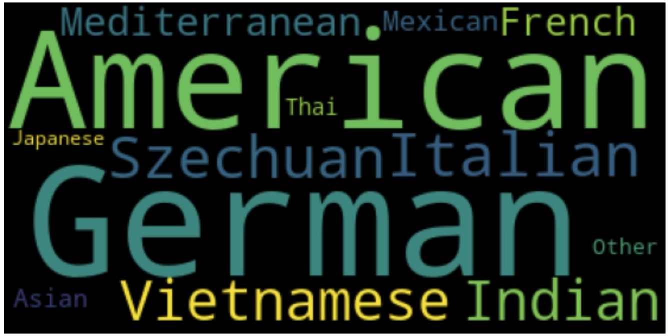
Total count

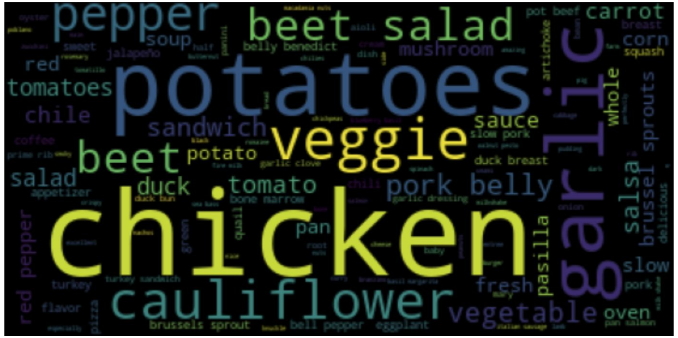


Normalized



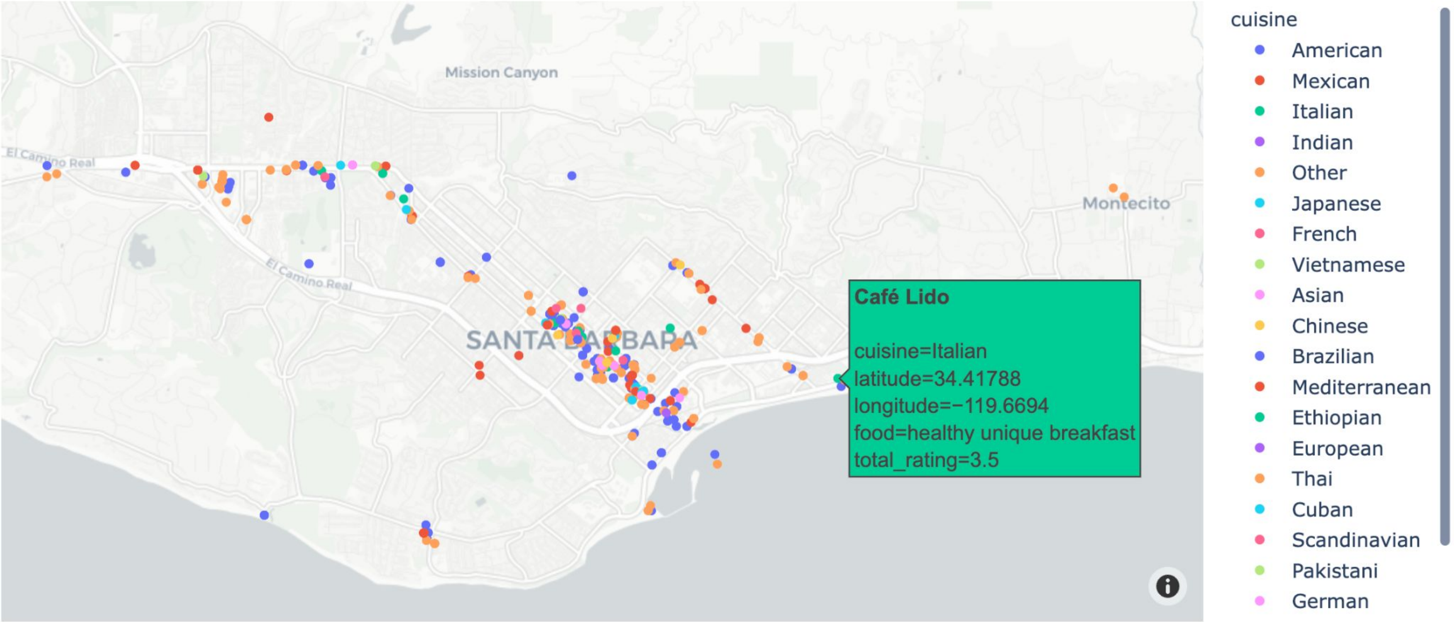
“Roasted” cuisines





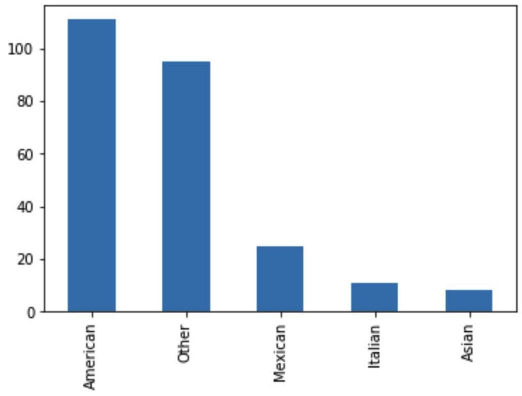
Search Engine

Input : “healthy”

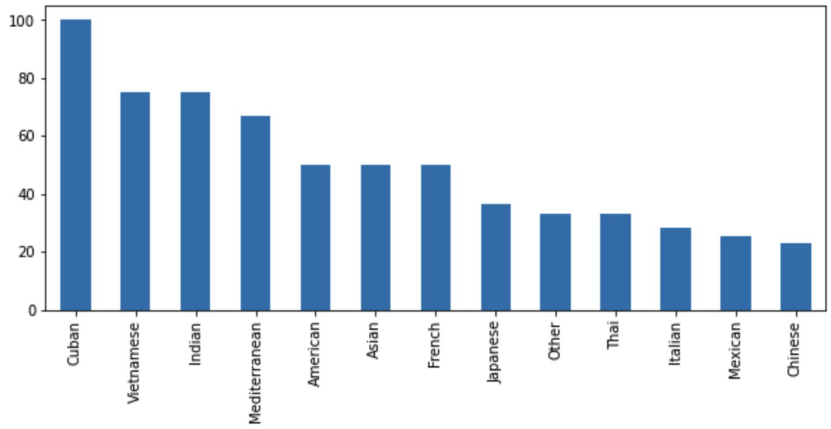


“Healthy” cuisines

Total count



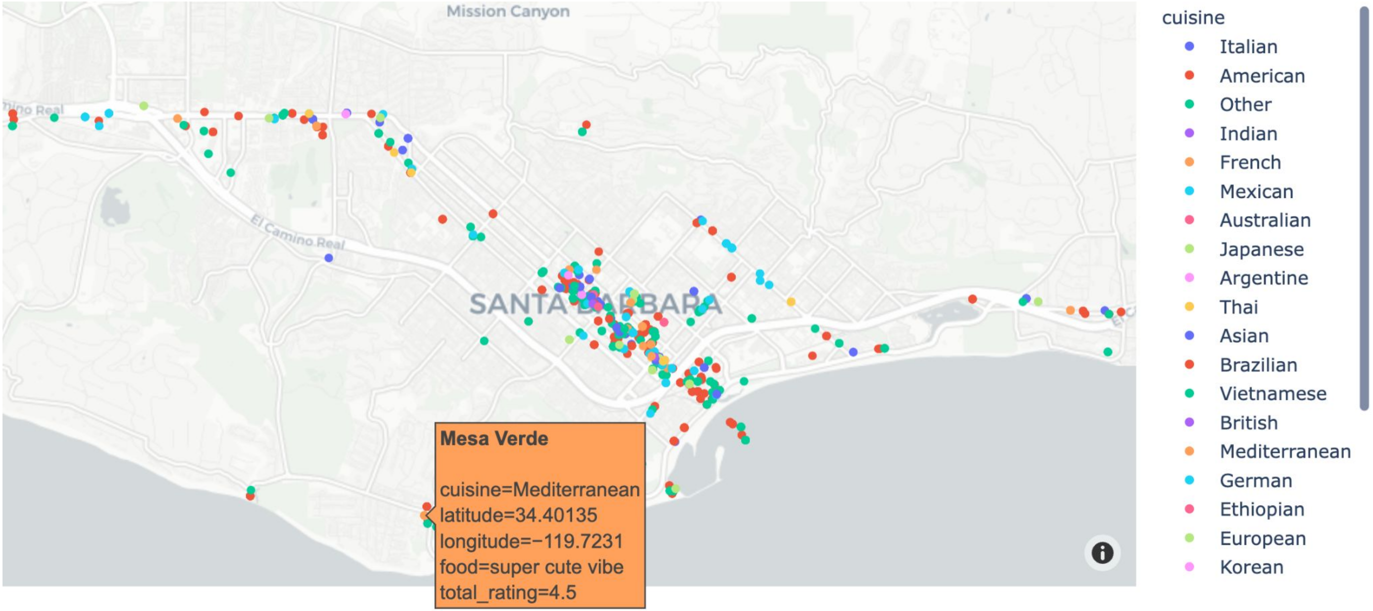
Normalized



“Healthy” cuisines

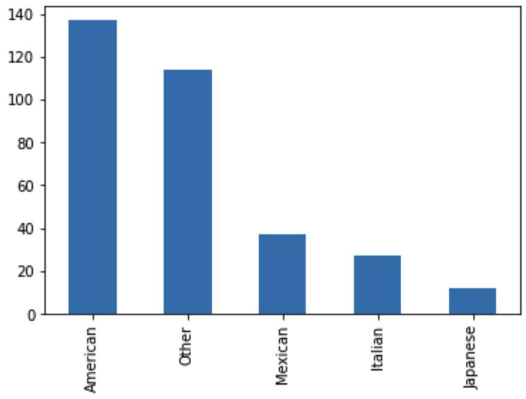
Search Engine

Input : “cute”

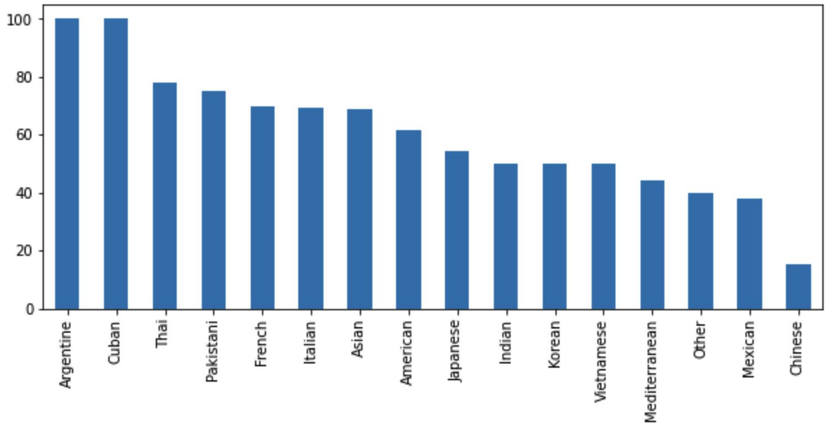


“Cute” Ambience

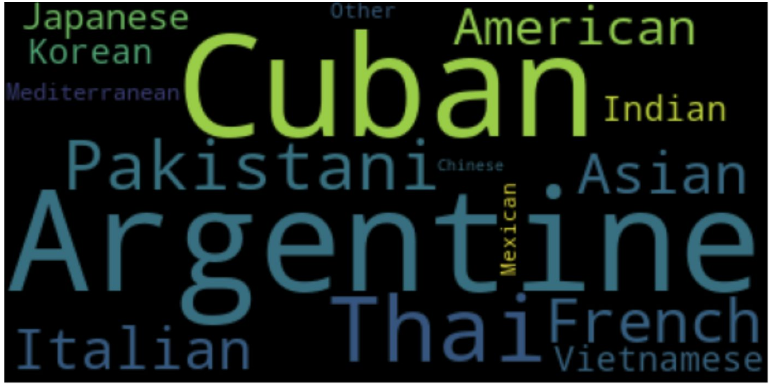
Total count

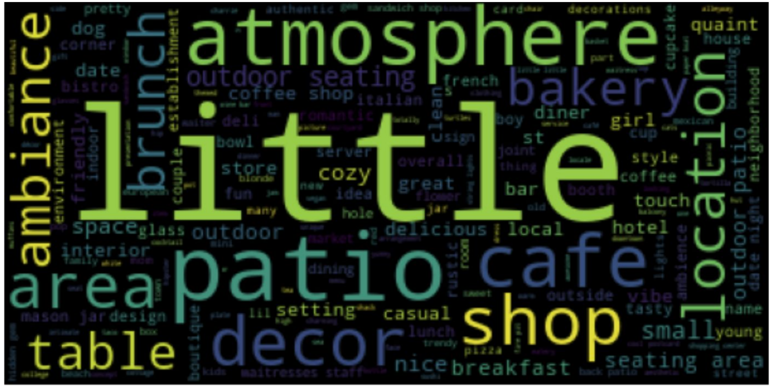


Normalized



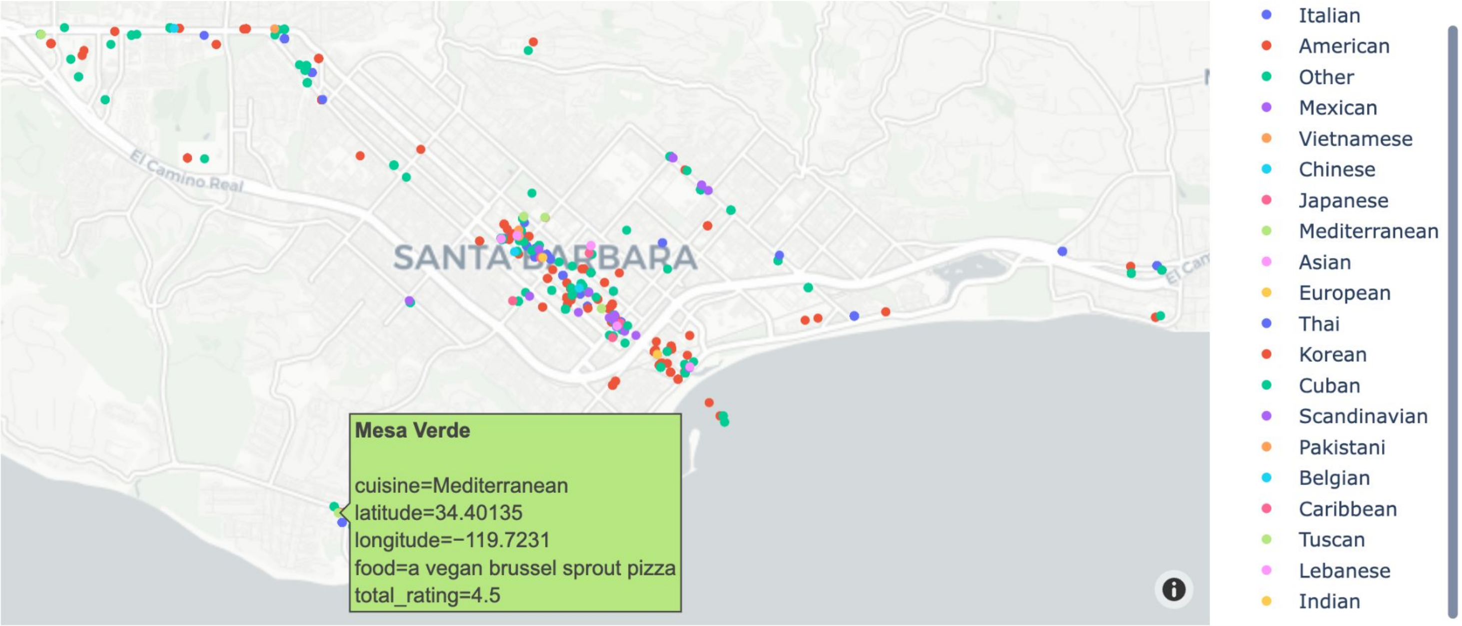
“Cute” Ambience





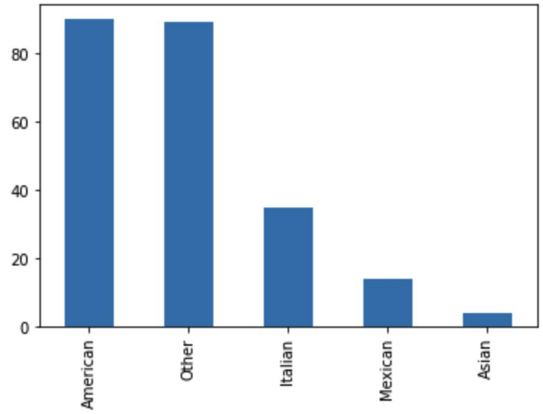
Search Engine

Input : “pizza”

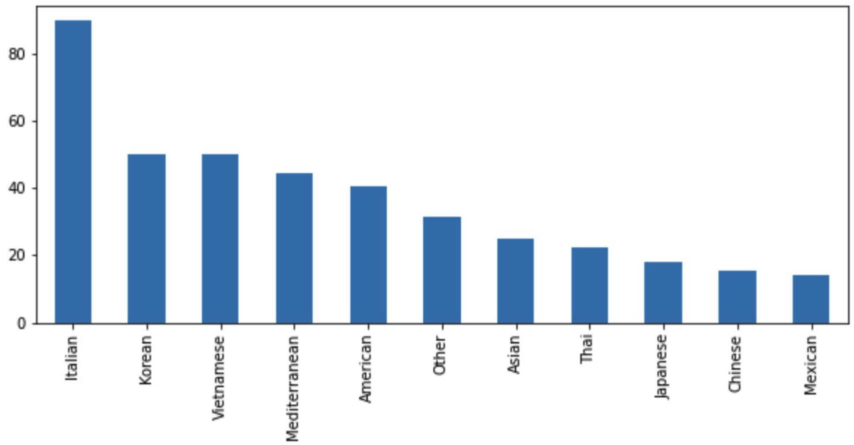


“Pizza” places

Total count



Normalized



“Pizza” places

Clustering

Cluster similar reviews to identify similar food / ambience features of different cuisines.

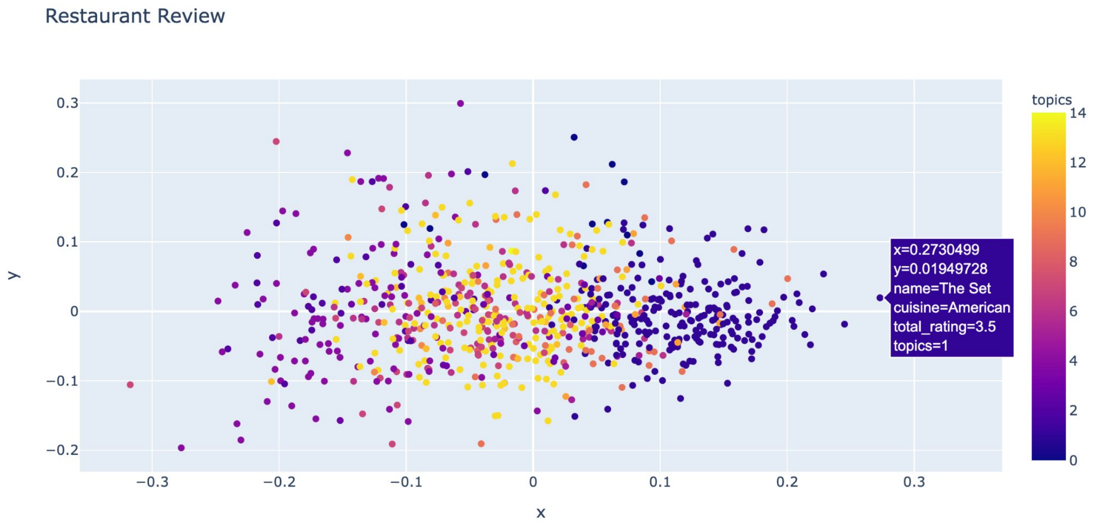
Clustering t-SNE

Steps:

- Tokenize reviews
- Clean review tokens
- Convert clean tokens to string
- Identify the restaurant reviews
- Extract cuisine type from the Categories column.
- Combine all the reviews of a restaurant.

Clustering t-SNE

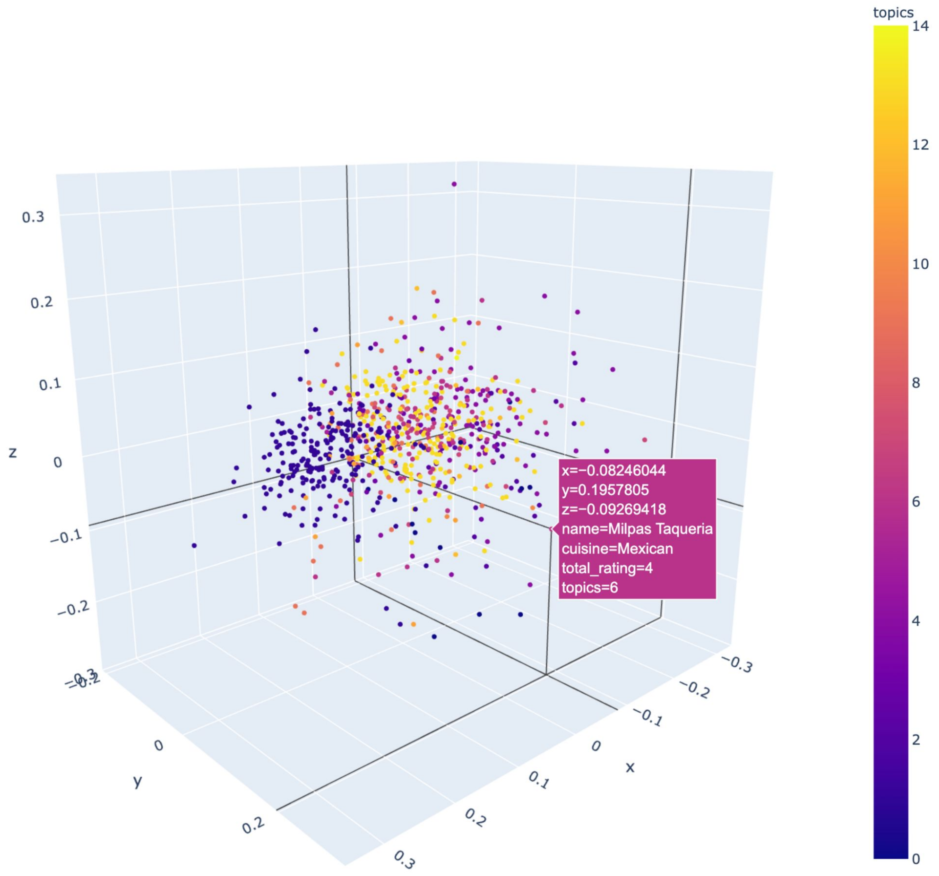
- Used TextHero for,
  - Generated term frequency - inverse document frequency(tf-idf) from clean review
  - K-Means Clustering on clean review to generate the topics.
  - Principal Component Analysis(PCA) on tf-idf
  - Generated Scatterplot



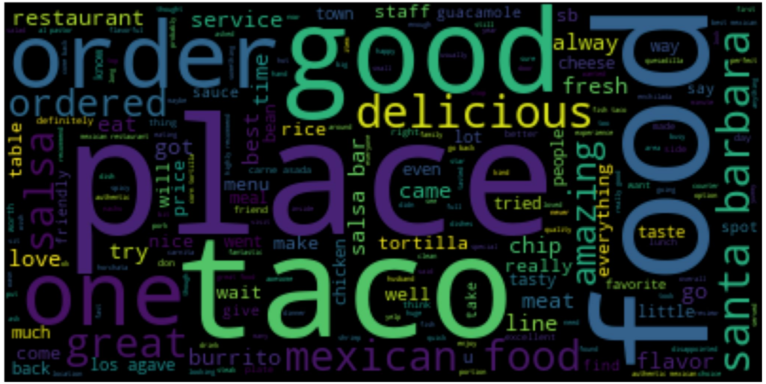
Cluster count : 15

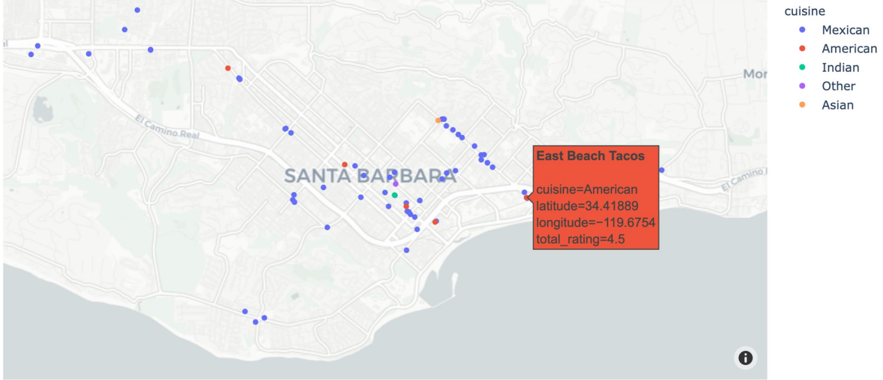
Clustering t-SNE

- Group the topics and combine their reviews
- Generate Word Clouds to see word frequency in each cluster.
- Geoplot restaurants by cuisine type for each cluster.

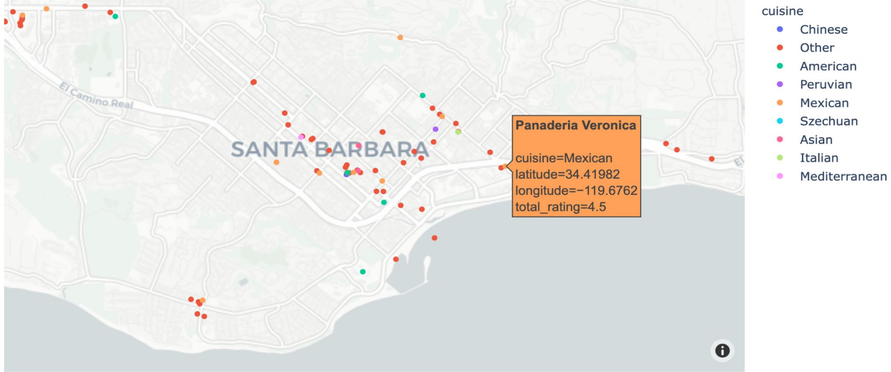


Clustering t-SNE

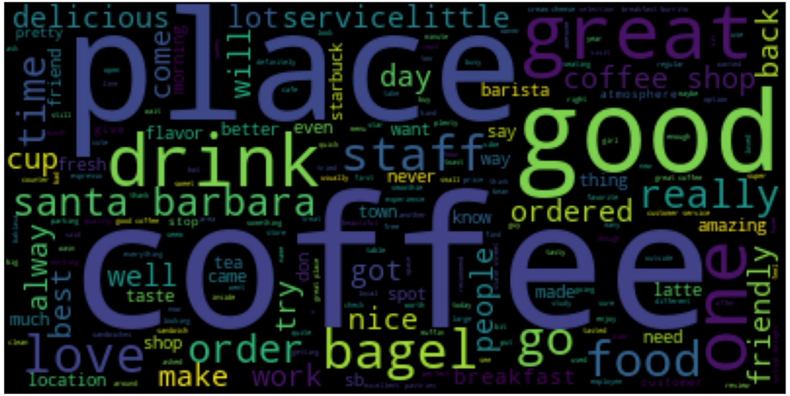


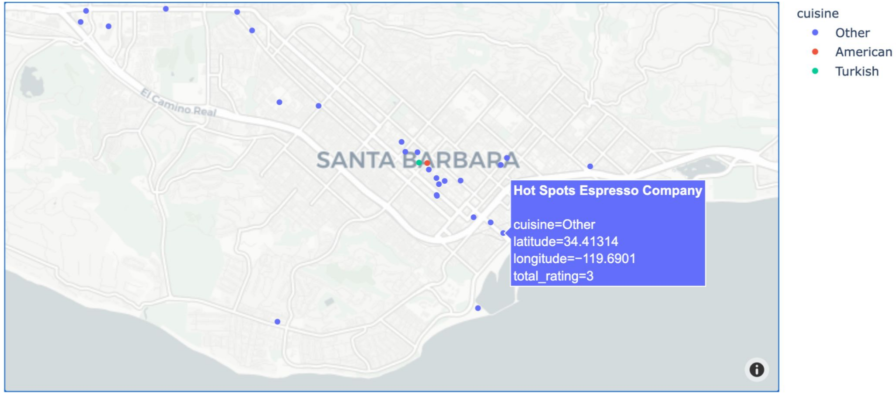


Clustering t-SNE



Clustering t-SNE



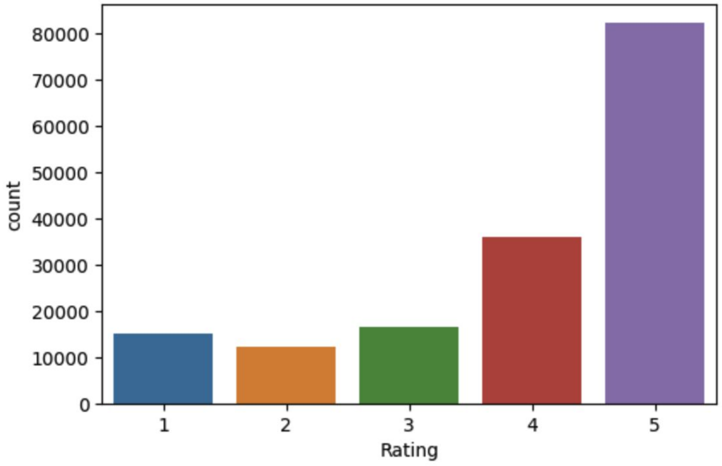


Sentiment Analysis

Sentiment analysis of the restaurant reviews and to predict the rating of a restaurant based on its reviews.

Sentiment Analysis

- Predict the rating of a restaurant based on the sentiment of the reviews
- Use Harvard Inquirer Dictionary
- Tokenize the reviews
- Clean and stem both the review tokens as well as the dictionary.
- Count the number of positive and negative words in each review using the dictionary
- Normalize positive and negative word count
- Fit a multi-category ML model - AutoGluon



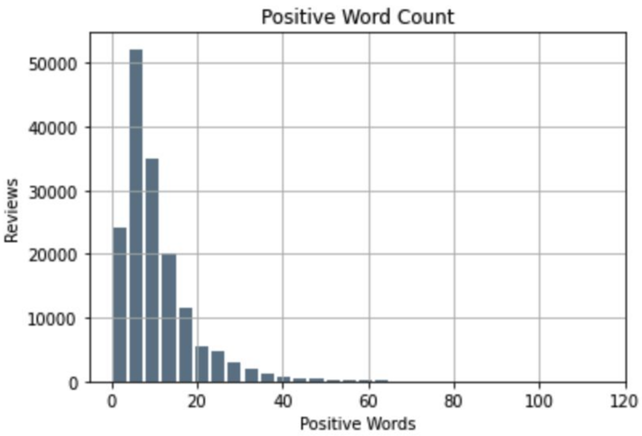
Sample size : 162,021

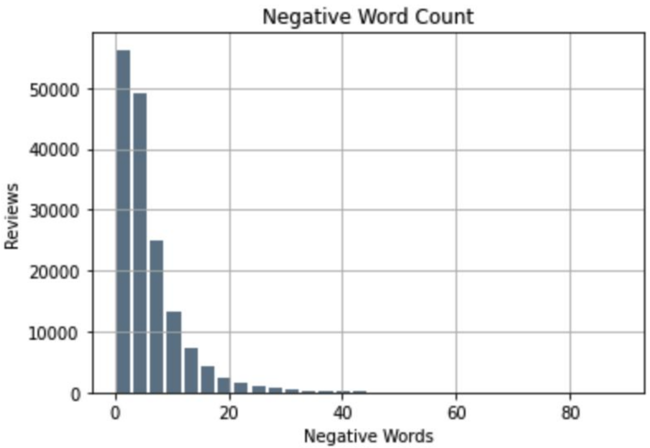
Sentiment Analysis

	Review	Rating	Review_Words	Stemmed_Review	Pos_Count	Neg_Count	Total_Tokens	Pos_Count_Normalized	Neg_Count_Normalized
1	$5 for a cup of coffee that was not refilled. ...	2	[$5, for, a, cup, of, coffee, that, was, not, ...	[for, a, cup, of, coffe, that, wa, not, refil,...	1	3	46	0.021739	0.065217
2	********??\nWe have been a long time patron of...	3	[********??, We, have, been, a, long, time, pa...	[we, have, been, a, long, time, patron, of, th...	8	8	94	0.085106	0.085106
3	2 stars for the waiting staff, 3 stars for the...	3	[2, stars, for, the, waiting, staff,, 3, stars...	[star, for, the, wait, staff, star, for, the, ...	6	6	66	0.090909	0.090909
4	2.5-3 Overall\n\nPrice\nIt's Santa Barbara, so...	2	[2.5-3, Overall, Price, It's, Santa, Barbara,,...	[overal, price, it, santa, barbara, so, be, pr...	33	17	304	0.108553	0.055921
5	A friend and I went here for lunch for the fir...	5	[A, friend, and, I, went, here, for, lunch, fo...	[a, friend, and, i, went, here, for, lunch, fo...	15	8	144	0.104167	0.055556
...	...	...	...	...	...	...	...	...	...
162017	Sabor! \n\nChilangos is part of the Milk&Honey...	4	[Sabor!, Chilangos, is, part, of, the, Milk&Ho...	[sabor, chilango, is, part, of, the, milk, fam...	3	2	83	0.036145	0.024096
162018	The salsa roja has improved! WOOHOO!! And now ...	4	[The, salsa, roja, has, improved!, WOOHOO!!, A...	[the, salsa, roja, ha, improv, woohoo, and, no...	2	0	21	0.095238	0.000000
162019	This review is solely based on one time, eatin...	5	[This, review, is, solely, based, on, one, tim...	[thi, review, is, sole, base, on, one, time, e...	16	14	187	0.085561	0.074866
162020	Very fresh tasty food. Has been in business f...	4	[Very, fresh, tasty, food., Has, been, in, bus...	[veri, fresh, tasti, food, ha, been, in, busi,...	1	0	21	0.047619	0.000000
162021	mmm decent mexican & cheap on state street!!! ...	5	[mmm, decent, mexican, &, cheap, on, state, st...	[mmm, decent, mexican, cheap, on, state, stree...	18	3	107	0.168224	0.028037

162021 rows × 9 columns

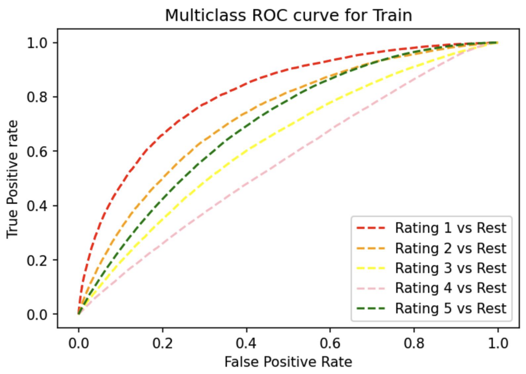
Sentiment Analysis

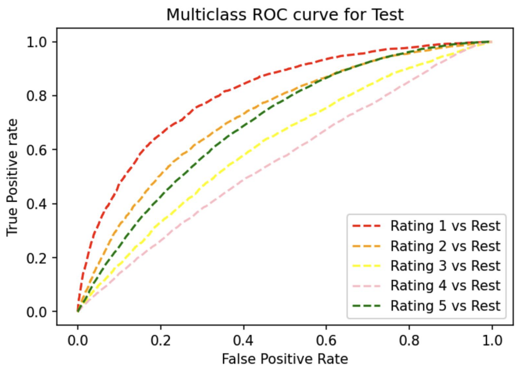




AutoGluon

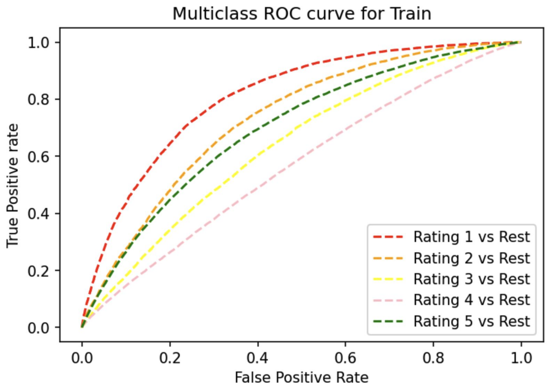
- Features : Pos_Count, Neg_Count
- Label : Rating
- Train test split : 0.2
- Train Accuracy : 54%
- Test Accuracy : 53%

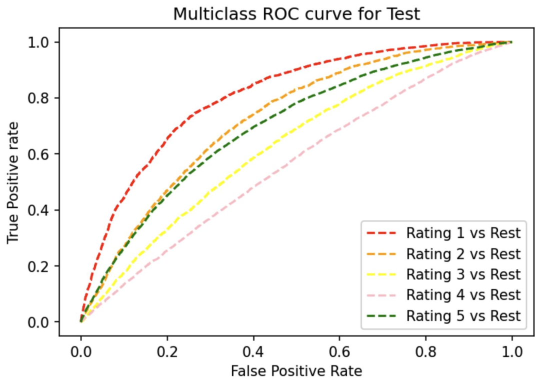




AutoGluon

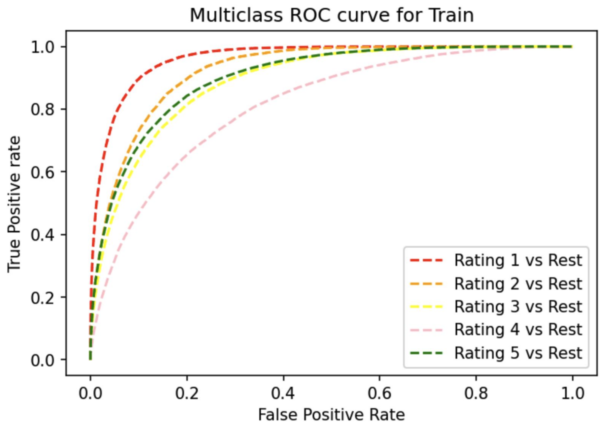
- Features : Scale_Pos_Count, Scaled_Neg_Count
- Label : Rating
- Train test split : 0.2
- Train Accuracy : 53%
- Test Accuracy : 52%

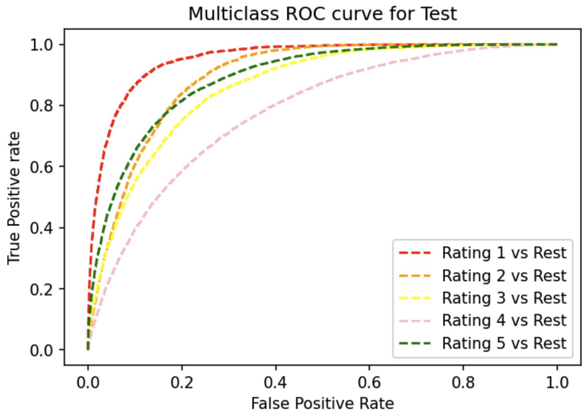




AutoGluon

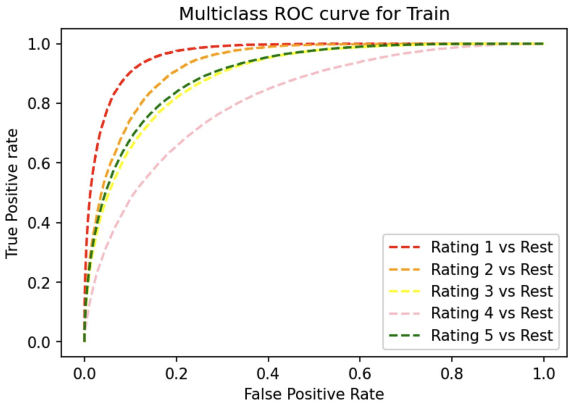
- Features : Review
- Label : Rating
- Train test split : 0.2
- Train Accuracy : 69%
- Test Accuracy : 66%

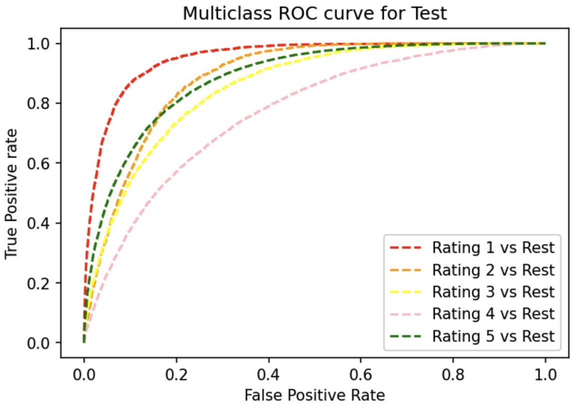




AutoGluon

- Features : Review, Pos_Count, Neg_Count
- Label : Rating
- Train test split : 0.2
- Train Accuracy : 70%
- Test Accuracy : 65%





Food Dictionary

A food glossary of different cuisines to look explore different culture.

Food Dictionary

- Created food dictionaries for different cuisines
- NLP techniques used
  - Web scraping
  - Beautiful Soup
- Deployed a food glossary using streamlit

Food Dictionary

text	date	Review_Words	Stemmed_Review	French_Food_Words	Food_Words	Pos_Words
Family diner. Had the buffet. Eclectic assortm...	2014-02-05 20:30:30	[Family, diner., Had, the, buffet., Eclectic, ...	[famili, diner, had, the, buffet, eclect, asso...	[]	[chicken]	[fresh, good, friendli, attent, good, casual, ...
Cute interior and owner (?) gave us tour of up...	2017-01-14 20:54:15	[Cute, interior, and, owner, (?), gave, us, to...	[cute, interior, and, owner, gave, us, tour, o...	[]	[salad, salad]	[cute, will, great, beauti, like, good, fill, ...
The bun makes the Sonoran Dog. It's like a snu...	2011-10-27 17:12:05	[The, bun, makes, the, Sonoran, Dog., It's, li...	[the, bun, make, the, sonoran, dog, it, like, ...	[]	[bacon]	[make, like, like, like, favorit, soft, indulg...
Upland is a brewery based out of Bloomington, ...	2014-11-12 14:12:20	[Upland, is, a, brewery, based, out, of, Bloom...	[upland, is, a, breweri, base, out, of, bloomi...	[]	[burger, pizza, pizza, pizza, pizza, burger]	[popular, open, good, excit, pretti, light, go...

QnA System

A QnA system that answers a given question based on the reviews

[Demo]

QnA system

Enter a row number

499.00

This is a great place to hang on the patio with friends and tell a few tales over some ale. Start with the Carolina Sweets, move the cheese steak or burger, and end with the cookie!

Type in your question here

How is the experience?

Answer: a great place to hang on the patio

Impact

Provide recommendations that will have a high success rate

Help people explore new cuisines that they will enjoy

Increase foot traffic for restaurants, help target customers with higher success rate

Future Work

- Train Spacy to identify food entities by using food dictionary of different cuisines
- Identify synonyms of features to produce better search results
- Create a food dictionary with food descriptions.
- Cluster based on food description to identify the similar foods and culture