George Mason University
AIT 590: Introduction to Natural Language Processing

The Geography of Taste

By Sriya Bhavaraju, Varalakshmi Vakkalagadda, Krishna Chaitanya Ashish
Team 04, Section-01

Under the guidance of Dr. Duoduo Liao
Date of Submission: 12.07.20

1. ABSTRACT

Arguably, the culinary diversity of a culture is an important factor used when defining it. It is also safe to say that those choices, which can be defined as the taste, can be linked to the socio-cultural context and not always defined by an individual's choice. In this paper, we used the taste as an important factor to understand the socio-economic background of a region. Food and drink choices made by people within a certain region can help us determine the diversity of the it and distinguish themselves from the neighboring regions. This data can be used to help boost business sales as well as help new establishments understand the layout and demands of the region. Similar studies have been performed previously to analyze the diversity of regions. In those studies, they primarily considered the different types of restaurants established in a particular region to determine the diversity of the city. In our method, we chose a Yelp dataset which contains reviews written by customers and their ratings. We chose to analyze from five cities including Pittsburg, Charlotte, Vegas, Cleveland and Phoenix. We used Natural Language Processing techniques to pre-process the reviews collected and identify the most optimal features. The results we derived provided us with the most common food and drink choices of each city. These features help us determine the most common choice of restaurant for the specific city. They will also help us better understand the layout of the region and taste of the people in it. Some of our ideas for future work on this dataset includes topic modeling using LDA, and name entity recognition from Spacy instead of manually recording the names of food.

2. INTRODUCTION

Several industries including marketing-commerce are adopting various machine learning techniques to improve their business model. Machine learning and artificial Intelligence has become a major part of our life, the recommendations we get on social media based on our search history are all derived from the usage of AI/ML. Consider the music app Spotify, it uses our search and playlist to suggest new songs that we might enjoy listening to. We can use the same concept and apply to food as well. We can train a model to understand the food choices made by the customers to provide better insight about their preferences.

Diversity is an important characteristic that defines a city and can help their economy if they can understand it. The United States is a diverse country with many religions and races, and it is only growing faster as new communities are developing and trying to establish themselves here. Emphasis on diversity and seeing the city as a melting pot, which is championed by postmodern thinking, has not addressed the gaps between ethnic and economic groups [12]. Food is an essential part of our daily life and people have increasing number of options to choose from. In this study we propose using taste as an indicator of social status and combines different communities and races together. Taste is an individual choice and can be vastly diverse, performing analysis on this will give us a detailed outline of diversity within a region. This analysis can help us understand the cultural diversity and help keep track of the culinary choices of the people.

This study can also help upcoming businesses and establishments understand the layout and preferences of the city. These factors play a very important role in the development and flourishment of a business. The new data sources have enabled the researchers to go beyond basic demographics such as race or income and delve into a multitude of sociospatial phenomena in modern cities []. Our research aims to answer the following questions:

1) To what extent is taste a good indicator of diversity of a place.

2) Can the conclusions drawn on a city based on taste actually help us understand the eating routines?

3. RELATED WORKS

The concepts of machine learning and artificial intelligence are not new, they are being used in every aspect of life to understand human nature better and make it easier. Lot of studies have been conducted where users' choices can be predicted using trained model. [1] one such study used GPS traces from a vehicle to build a network and used community detection algorithm to identify the overlap of communities in an area. Most of the research in this area uses data from social media like tags, images and check-ins to understand the preferences and geography of the area better. Some studies have used different labels for different clusters of users, in this model a rank is generated from a correlation value, if the co-relation value is higher than the pr-defined threshold the food is recommended [2].

4. OBJECTIVE

The main objective of the study is to understand the diversity of taste of five American cities including Pittsburg, Charlotte, Vegas, Cleveland and Phoenix by examining the reviews of the restaurants extracted from Yelp. In United State, restaurants expenditure exceeds spending in higher education, computers, books, magazines, movies and recorded music [3]. Yelp is a social media platform which allows users to get details of the local business and write reviews about their services. Yelp users have generated nearly 127 million reviews for different businesses across the world [4]. Our aim through this study was to draw the following conclusions; firstly, user's choice of taste reflects their type of restaurant they visit. Secondly, user's rating as well will help us understand their taste.

5. DATASETS

For your study we chose a Yelp dataset which contains reviews for five hundred restaurants from the five cities. We chose these cities as they had the highest review count. The dataset contains 30199 reviews, reviewed by 966 unique users. To reduce ambiguity, we only considered users who have reviewed more than twenty restaurants. The dataset contains two main parts; user information, which contains information about the users and business information which contains information about the restaurant and its rating.

5.1 SCHEMA

Attribute	Datatype
USER ID	A unique ID for every user
BUISSNESS ID	A unique ID for every business
RATING	Rating given to the restaurant by the user
REVIEW	Detailed review given by the user to the restaurant.
CITY	The location of the restaurant.
LATITUDE	Latitude of the restaurant.
LONGITUDE	Longitude of the restaurant.
REVIEW COUNT	The number of reviews the user has given.
STAR RATING	The stars given to a restaurant by the user

postal_coc	latitude	longitude	Restaurant	review_co	user_id	review_id	User-Resta	text	date	user-reviewCount
85016	33.5082	-112.037	4	3515	h0fffFM3G	F6MW-SXe	5	Great breakfast joint!	1/4/2015 4:11	30
85016	33.5082	-112.037	4	3515	3cC726zw	71Je4Eb7k	4	I should have had the pancakes.	4/19/2014 21:39	25
85016	33.5082	-112.037	4	3515	dEk7mXM	Bx8IxSpyze	3	95 minute wait for breakfast? Really?	11/30/2014 23:45	48
85016	33.5082	-112.037	4	3515	771OWzbz	egxw1AeF	2	I so wanted to fall in love with the cuten	12/7/2013 20:25	26
85016	33.5082	-112.037	4	3515	_jYEC7fvq	6ZgMgiayp	4	OMG, this place is so so so so so yummy.	4/19/2014 18:05	23
85016	33.5082	-112.037	4	3515	7sNE58P4	axhqVPzfk	5	I always enjoy the pancakes with Sunnys	4/13/2016 3:50	21
85016	33.5082	-112.037	4	3515	EMM4HqI	ITgv-LfFku	2	This review is based solely on wait	2/24/2014 7:11	42
85016	33.5082	-112.037	4	3515	SCo1UBoe	_0n-camY	4	I meant to write this review earlier but I	9/25/2015 22:24	49
85016	33.5082	-112.037	4	3515	KMQ-5KpT	xWnVBSTC	4	Went there recently for a late morning b	7/16/2016 19:56	38
85016	33.5082	-112.037	4	3515	qN-Hw7oc	v96BuE7T	3	Really great place for breakfast or brunc	10/14/2016 21:29	21
85016	33.5082	-112.037	4	3515	v6SrDjzqb	id382WQK	5	Now HERE is were the benny's are at.	3/16/2014 2:05	39
85016	33.5082	-112.037	4	3515	EMM4HqI	SBUAbksL	4	This is an update to my earlier, less-	4/23/2014 3:43	42

Fig 1: Sample of the dataset before pre-processing.

5.2 PRE-PROCESSING OF DATA

To apply NLP concepts, we processed the data to get a clean data set so that the model can be trained efficiently. The pre-processing steps are as follows; Firstly, all the numerical data is removed, then punctuation's and stop words (most repeated words like to, the etc.) are removed. The next step is to lemmatize the text by adding POS tags to the word. Next, we select featured from the data. The most relevant features are extracted from top 1000 frequent words which are plotted using frequency distribution. The list of selected features can be found in appendix.

B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	business	aioli	albacore	american	angus	apple	argula	asiago	asian	authentic	avocado	bacon	bake	banana	basil	bbq
gg', 'bacon', 'american',	-050d_XIo	0	0	0.015693	0	0.016032	0	0	0	0	0	0.316362	0.033251	0	0	
pas', 'bread', 'flatbread'	-7H-oXvCx	0.086986	0	0.027238	0	0.027826	0.038104	0	0	0	0	0.160155	0	0	0.017046	0.0716
:hili', 'mayo', 'cheese', 's	01fuY2NN	0	0	0.008753	0	0	0	0	0.020777	0	0	0.044115	0.009273	0.059728	0.032867	
lic', 'lemon', 'green', 'av	042IHd5Kj	0.100976	0	0	0	0	0	0	0	0	0.124086	0.066397	0	0	0	0.0831
ll', 'panang', 'curry', 'sau	0PCBt3JKD	0	0	0	0	0	0	0	0.03269	0.010201	0	0	0	0.010442	0.068951	
/ocado', 'cheddar', 'fren	0VUbtKyg-	0	0	0	0	0	0	0	0	0	0.084196	0.202736	0	0	0	0.1269
', 'waffle', 'sandwich', 'r	0d0i0FaJq	0	0	0	0	0	0	0	0	0	0.013702	0.13197	0	0.059559	0	0.0688
neddar', 'bread', 'french'	1MvR4NJC	0	0	0	0	0.070333	0	0	0	0	0	0.046264	0.05835	0	0	
'shot', 'cheese', 'tasty', '	1RECFIhw	0.029388	0	0.023005	0	0	0	0	0.027303	0	0	0.115944	0	0	0	0.0967
iscuit', 'apple', 'chicken'	1gQrMs-I8	0.016771	0	0	0	0.093884	0	0	0.015581	0.014587	0.027479	0.066167	0	0	0.016432	0.0690
ndwich', 'beef', 'turkey',	1veVZUaw	0	0	0	0	0	0	0	0	0.029203	0	0.022077	0	0	0	
'pad', 'sauce', 'thai', 'spi	228mfqUu	0	0	0.028963	0	0	0	0	0.068749	0.064363	0	0	0	0.065878	0.181261	
sauce', 'fry', 'yummy', 'r	22MYhTXv	0	0	0	0	0	0	0	0	0.027874	0	0	0	0	0	
omato', 'bean', 'cilantro'	2g4ZTD3eI	0	0	0	0	0	0	0	0.018325	0.085778	0.24238	0.025939	0	0	0	
dessert', 'egg', 'incredibl	2iTsRqUsP	0.075558	0	0.034503	0	0.020141	0	0	0	0	0	0.252556	0.093994	0.028028	0.030847	0.0311
y', 'margarita', 'taco', 'b	2nn6Sx5EF	0	0	0.071697	0	0	0	0	0	0	0	0	0.075958	0	0	
tato', 'sandwich', 'patty	2um9bzTS	0	0	0.041553	0	0	0	0	0	0	0.152203	0	0	0	0.052009	

Fig 2: Sample of the dataset after pre-processing.

6. METHODOLOGY

6.1 Workflow

Our assumption while working with the Yelp dataset is that, if a person writes a review about a place or food or the drink, the person has purchased the drink or considered purchasing it and therefore can be an indicator of one's taste. The following diagram summarizes the workflow of the study:

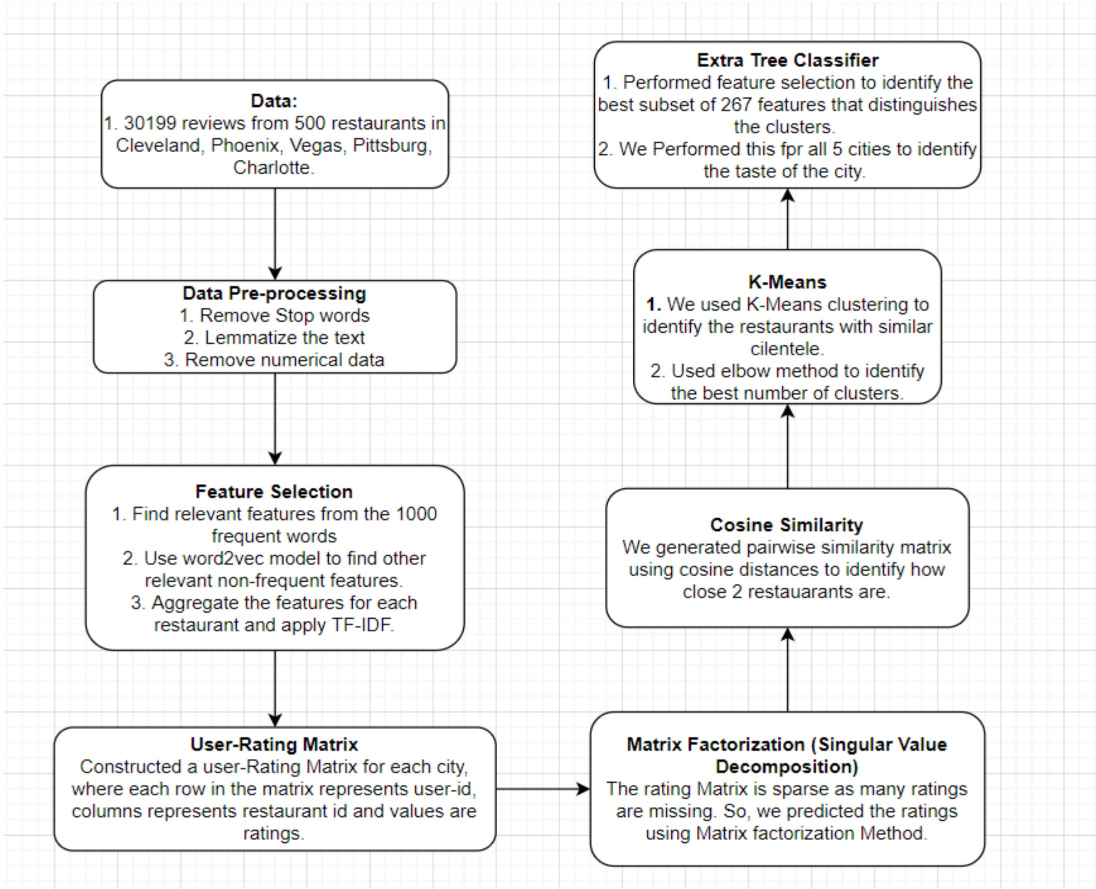


Fig: Methodology

6.2 Feature Selection

Feature selection is a process where choose features which contribute most to our prediction variable or the output we are interested in [6]. Feature selection improves the accuracy of the model and reduces overfitting of data and significantly reduces the training time. According to Bourdieu's theory of distinction, food, drink, and interior decoration are among the best indicators of taste reflecting one's everyday choice [5]. Therefore, we are interested in adjectives which describe these three categories of food. While selecting the features we are interested in the most repeated words regardless the way they are embedded in the sentence, we also avoided selecting words that are too generic for better results.

After pre-processing the data, we had 1000 frequent words from which we had to select our features. The most relevant features are extracted from these 1000 words and a frequency distribution is plotted. Although we expect to get specific words from the comments, example different types of cheese (ricotta, gouda, peperjack) they are not part of the frequent words. To solve this problem, we use the Word2Vec model. This algorithm uses a neural network model in order to learn word associations from a large corpus of text. This algorithm once trained can be used to detect synonymous word or suggest additional words for a sentence.

```
provolone', 0.5979742407798767),
smoke_gouda', 0.5545506477355957),
manchego', 0.5438302159309387),
white_cheddar', 0.5392850637435913),
cheesy_goodness', 0.5315919518470764
jack_cheese', 0.519109845161438),
asiago', 0.514570951461792),
havarti', 0.513709306716919),
gruyere_cheese', 0.5108760595321655)
gouda', 0.5107713341712952)]
```

Fig. 3: Ten most similar words to cheese generated from Word2Vec model

The reviews are then binarized based on the 267 features selected and aggregated them for each restaurant. Once all the features are extracted, we weigh them using TF-IDF method.

FOOD	Cheese, chicken, burger, sauce, salad, pizza, dessert, pork, egg, meat, steak, beef, shrimp, bacon, fish, bbq, pasta, thai, mexican, seafood, tea, Italian, vegetarian, Asian, goat, vegan, American, chinese
DRINK	Beer, wine, cocktail, shake, juice, coffee
FOOD ADJECTIVES	Fry, tasty, hot, spicy, crispy, grill, toast, roast, yummy, salt, spice, creamy, juicy, fried, crisp, bake, crunchy, sour, greasy, chewy
AMBIANCE	Friendly, fantastic, pricey, classic, incredible, authentic, overly, splendid

Fig 4: Features selected from the top 1000 words.

6.3 User Rating Matrix

Our main aim is to identify the choice/taste of people in each city. Considering all 1000 features for clustering might cause ambiguity as they may be some repeated words in the review which are not important but may be considered. This reduces the dimensionality of the cluster. To address this problem and to determine the choice we performed the following steps to find the clusters of restaurants that have been visited by similar users. As the choice of the restaurant represents their taste for food, drink and ambience we generated a user-rating matrix for each city separately.

	user_id	01fuY2NNscttoTxOYbuZXw	1MvR4NJQbHy0i7ME1loYpw	22MYhTXwSXaS4rW2VOrR-w	2um9bzTSvD3f2tkdBoA38g	4
0	0EgbVtQfdNQq3Tapxwl_sw	4.5	3	4.0	0.0	
1	0KGCdp9W8jz9KJRsABb66g	5.0	0	4.0	3.0	
2	146udnwO4772RFXD1i5ruQ	0.0	0	0.0	0.0	
3	1gEx8Si6rkStWBp_IXnGCw	4.0	5	0.0	0.0	
4	1sGYXSkJHPhJ6wQtc-RbZw	0.0	0	0.0	0.0	
...	...	...	...	...	...	
226	z6v7O3QGZmD17zA6o1Cipg	5.0	5	0.0	0.0	
227	zSwb7qNpSgU3ekHMpiHsOA	0.0	0	0.0	0.0	
228	zZYHZwmBl9Af4pl-aLXBFA	4.0	4	0.0	3.0	
229	zdJ5RyU4nUWj8bceNxLrZg	5.0	5	0.0	0.0	
230	zox0bzk1LvmLb77fmcRjUQ	0.0	0	0.0	4.0	

Fig 5: User-rating Matrix

The rating matrix is sparce as not all the restaurants have ratings. To overcome this, we applied matrix factorization on the original matrix to predict the ratings using singular value decomposition method. SVD at a high level is an algorithm that decomposes matrix r into best lower rank.

$$R = U\Sigma V^T$$

Here, R is the user rating matrix, U is the users features, sigma is the diagonal matrix of weights and V^T is the restaurants feature matrix. In the above formula U represents the how much users like each feature and V^T represents how relevant each feature is to each restaurant.

From this matrix, we generate pairwise similarity matrix to identify how similar two restaurants, this is done using cosine similarity.

$$\cos(\mathrm{A},\mathrm{B}) = \frac{\mathrm{AB}}{|\mathrm{A}||\mathrm{B}|} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{A_i B_i}}{\sqrt{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{A_i^2}}\sqrt{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{B_i^2}}}$$

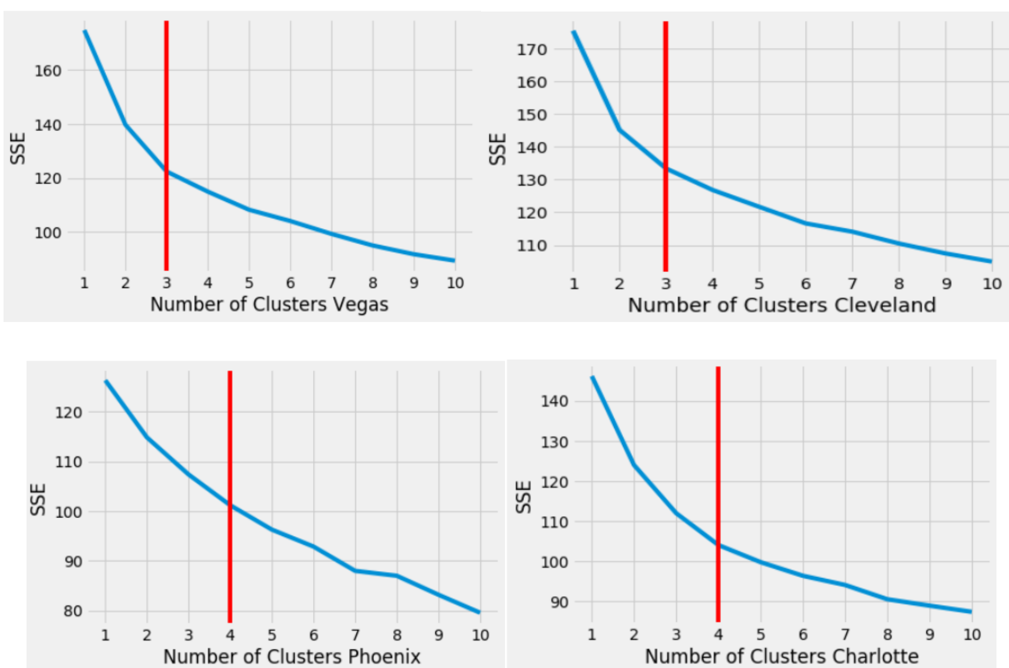
In the above formula A and B are vector attributes which in our study are restaurant A and restaurant B. Every element A and B is 1 if the user has reviewed the restaurant and 0 otherwise [9].

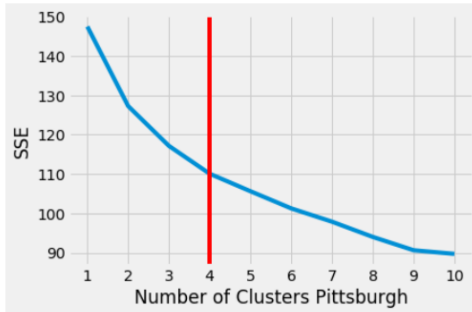
	01fuY2NNscttoTxOYbuZXw	1MvR4NJQbHy0i7ME1loYpw	22MYhTXwSXaS4rW2VOrR-w	2um9bzTSvD3f2tkdBoA38g	4vDffQ7eSQ7n39cyroOSKQ
0	4.486901	1.750658	4.265819	0.688099	1.415001
1	4.617026	-0.245718	1.810118	1.222820	0.341449
2	0.311160	0.576230	-0.034477	0.056776	0.973495
3	3.652034	2.788615	-0.399805	0.194463	1.706115
4	-0.120374	0.521971	-0.235889	-0.155580	0.198050
...	...	...	...	...	...
226	4.588964	3.787393	0.906785	-0.647193	3.541973
227	0.188144	0.378437	-0.206056	0.010623	0.004383
228	3.677652	1.908820	0.076987	2.068877	0.846049
229	5.563245	4.003370	-1.395483	0.450870	0.707414
230	-0.246256	0.171334	-1.388297	2.244128	-0.403487

Fig 6: User-Rating Matrix after applying SVD

6.4 K-Means Clustering

K-means clustering method is an unsupervised machine learning technique used to identify clusters of data object in a dataset [8]. In K-means clustering the center points of a cluster are chosen and called as centroids. The first step is to choose k centroids where k is the number of clusters we choose. In the next step each data point is assigned to its nearest centroid then we compute the mean of all the points for each cluster and fix a new centroid. In our study we cluster the similar restaurants together using K-means and identify optimal number of clusters using the elbow method. The below figures represent the optimal number of clusters for each city in our study.





6.5 Extra Tree Classifier

Extremely Randomized Tree Classifier is a technique which aggregates the results of multiple de-correlated decision trees collected to output its classification result [10]. The concept is very similar to Random Forest Classifier the only difference is the construction of the decision tree. Each decision tree is constructed from the original training sample and at each test node the tree is provided with a random samples of K features. Based on the Gini Index the decision tree will split at the best feature selected from the feature set. In our study we performed feature selection to identify the subset of 267 features that distinguish the clusters. We performed this on all five cities to identify the prominent taste in each city.

7. RESULTS

The table below shows the features selected for each city and their F1-scores. F-score is used to measure the precision. It is most commonly used for evaluating information in a retrieval system. An F1-score is considered perfect when it is 1 and a failure when it is zero from the table, we can see that all cities have an F1-score more than zero.

City	Feature	F-score
PITTSBURG	Food: bucatini, spaetzle, brownie, burrito, Taiwanese, turkey, biancoverde, hash brown, sheep, BBQ Drinks: mimosa Ambience: overly	0.40
VEGAS	Food: Pancake, halibut, oyster, tagliatelle, focaccia, shank Drinks: malbec, sauvignon Dessert: honey, truffle, cake(carrot)	0.73
CHARLOTTE	Food: waffle, chicken, vegan, pizza, beef, burger, white_cheddar, muffin, jalapeno, toast, avocado	0.45

	Drinks: gin, negroni	
PHOENIX	Food: albacore, skirt steak, chocolate, oyster, calamari, carbonara, marinara_sauce, fettuccine, smoke_guda, roast, risotto, bacon, endive Dessert: Oreo	0.39
Cleveland	Food: beef, pretzel, biscuit, biancoverde, sour, protobello_mushroom, spaetzle, mushroom, hot-dog, shank, tuna, steak Dessert: gelato Drinks: gin	0.59

Table 1: Feature importance generated for each city using Extra Tree Classifier

Our main aim with this study was to draw conclusions on users' choices based on the reviews provided by them. From table 1, we can draw the following conclusions.

- Diners in Vegas prefer medium-high end eats.
- Diners in Charlotte are more inclined towards fast-food.
- Diners in Pittsburg give importance to ambiance and prefer different kinds of pasta.
- Majority of the dinners in Phoenix prefer to eat sea food and Italian food.
- Diners in Cleveland lean towards picking beef.

7.1 Limitation

One of the setbacks we initially faced in our study was, we could not work with the most recent data. Our initial aim was to web-scrape yelp reviews and then perform our analysis on them. As user-id is a private information we did not have access to it and were unable to web-scrape the data. Instead, we used 2015-2017 yelp review data set. Yelp reviews maybe biased, as the entire population of the city does not write the review the reviews which are posted maybe biased and do not always provide the most accurate conclusions. In case of Phoenix, we can see in table 1 that the F-score is low compared to other cities, this may be due to data bias or similar food taste. We could not consider more than 100 restaurants in each city as our processor did not support the huge amount of data.

8. CONCLUSION

In this study we extracted features from a yelp data set to determine the taste and choice of restaurants made by the people in Vegas, Charlotte, Pittsburg, Phoenix and Cleveland. We chose these cities because they had the highest number of reviews. We obtained 30199 reviews for 500 restaurants and processed them. After applying various NLP techniques, we found the most common dishes in each city and the type of cuisine they prefer. Vegas being a destination which is most attracted by tourist, most of the dishes that showed up on the feature selection were high end eats. Similarly, people in Cleveland and Phoenix prefer Italian food. The dinners in Pittsburg gave a lot of importance to ambience. This helps understand the neighborhood better and develop it according to the user's choice. This study is a vast improvement over the previous projects on this topic, previously only the type of restaurants was taken into consideration and one of the major drawbacks with this it does not reflect the user's choice and possibly give false results about the diversity of the region. Our study aims at considering only the restaurants which are visited by the users and considers their review as a result is a more appropriate description of diversity of the place.

Future work can focus on topic modeling using LDA, this would involve counting words and grouping similar word patterns within the unstructured data by detecting patters such as word frequency and distance between words [11]. Next, we would like to use name entity recognition from Spacy instead of manually recording the names of food.

Appendix

List of features selected from the reviews.

Steak types	Chorizo, rib , brisket, hanger_steak, filet, spencer, ribeye, skirt_steak, tomahawk, filet_mignon
Meat types	Duck, meatball, lamb, turkey, calamari, rotisserie_chicken, chicken_breast, pork_loin, angus_beef, beef_patty, spare_rib, shank, sheep
Sweets	Cream, chocolate, cake, waffle, biscuit, truffle, pudding, honey, crepe, caramel, jam, sugar, vanilla, pastry, cheesecake, gelato, cinnamon, cookie, chocolate_mousse, tiramisu, chocolate_brownie, oreo, whipped_cream, muffins, croissant
Veggie types	Potato, onion, tomato, corn, bean, mushroom, garlic, lemon, pickle, vegetable, chili, spinach, lettuce, peanut, olive, basil, cucumber, jalapeno, cilantro, carrot, endive, mixed_green, portobello_mushroom
Fruit types	Avocado, banana, fruit, strawberry, apple, orange, pineapple, coconut, mango, papaya, kiwi, guava
Liquor types	Margarita, Ipas, hefeweizen, chianti, vino, malbec, pinot, sauvignon, negroni, gimlet, mojito, gin
Dressing types	Salsa, pepper, guacamole, cheddar, aioli, parmesan, mozzarella, mayo, mayonnaise, marinara_sauce, creme_fraiche
Breakfast types	Bread, pancake, pie, burrito, donut, flatbread, omelet, scramble_eggs, cereal
Latin types	Taco, sandwich, roll, rice, bun, dog, sausage, tacos, tortilla, savory, caesar, clam_chowder, etouffee
Italian types	Pasta, risotto, dumpling, biancoverde, tagliatelle, carbonara, pappardelle, spaghetti, penne, bucatini, lasagna, fettuccine,linguini, spaghetti_meatball, spaetzle, focaccia
Ethnic food	
Fast food	Chip, pretzel, patty, cheeseburger, gourmet_burger, delux_burger, ronin_burger, hamburger, sausage_pepperoni, French_fry, parmesan_fry, rotisserie, hashbrowns, croquette, tots

Sea food	Shrimp, crab, salmon, lobster, tuna, oyster, scallop, mussel, squid, tilapia, Hamachi, sear_tuna, halibut, albacore, tilapia
Asian	Sushi, noodle, ramen, tofu, kalbi, congee, dim_sum
Drinks	mimosa, lime, shot, milkshake
Cheese	Provolone, smoke_gouda, manchego, asiago, white_cheddar, gruyere_cheese, cheddar_cheese, Havarti, jack_cheese, manchego_cheese
Mexican	Tinga, refried_bean
Salad	Cobb_salad, argula_salad, Caesar_salad
Thai	tom_kha, panang_curry,curry, massaman, pad_see, kee_mao, pad_thai
Coffee	Milk_tea, iced_coffee, lattes,French_press, matcha, cappuccino
Ethnic food	Taiwanese

9. REFERENCES

[1] Rinzivillo, S.; Mainardi, S.; Pezzoni, F.; Coscia, M.; Pedreschi, D.; Giannotti, F. Discovering the Geographical

Borders of Human Mobility. *Künstl. Intell.* **2012**, *26*, 253–260. [CrossRef]

[2] Xinke Li, Wenyan Jia, Zhaofang Yang,Yuecheng Li, Ding Yuan, Hong Zhang, Mingui Sun Application of Intelligent Recommendation Techniques

[3] Schlosser, E. *Fast Food Nation: The Dark Side of the All-American Meal*; Houghton Mifflin Harcourt: Boston, MA, USA, 2012.

[4] Yelp. *Yelp Information*; Yelp: San Francisco, CA, USA, 2017.

[5] Bourdieu, P. *Distinction: A Social Critique of the Judgment of Taste*; Routledge: Abingdon, UK, 1984; Volume 1.

[6] Shaikh, R. (2018, October 28). Feature Selection Techniques in Machine Learning with Python. Retrieved December 08, 2020, from https://towardsdatascience.com/feature-selection-techniques-in-machine-learning-with-python-f24e7da3f36e

[7] Word2vec. (2020, December 07). Retrieved December 08, 2020, from https://en.wikipedia.org/wiki/Word2vec

[8] Real Python. (2020, November 07). K-Means Clustering in Python: A Practical Guide. Retrieved December 08, 2020, from https://realpython.com/k-means-clustering-python/

[9] The Geography of Taste: Using Yelp to Study Urban Culture. (n.d.). Retrieved October 21, 2020, from https://www.researchgate.net/publication/327642523_The_Geography_of_Taste_Using_Yelp_to_Study_Urban_Culture

[10] AlindGuptaCheck out this Author's contributed articles., AlindGupta, & Check out this Author's contributed articles. (2020, July 01). ML: Extra Tree Classifier for Feature Selection. Retrieved December 08, 2020, from https://www.geeksforgeeks.org/ml-extra-tree-classifier-for-feature-selection/

[11] Ganegedara, T. (2020, December 05). Intuitive Guide to Latent Dirichlet Allocation. Retrieved December 08, 2020, from https://towardsdatascience.com/light-on-math-machine-learning-intuitive-guide-to-latent-dirichlet-allocation-437c81220158

[12] Madanipour, A.; Cars, G.; Allen, J. *Social Exclusion in European Cities: Processes, Experiences, and Responses*; Psychology Press: Abingdon, UK, 2000; Volume 23.