



Recommendation System with Yelp dataset

202035360 LeeDaHyeon
202135705 KooDoYeon
202135809 YoonHyeWon
202135842 ChoTaeWan

+

Index

01

Business Objective

02

Data Exploration

03

Data Preprocessing

04

Modeling

05

Evaluation and
Analysis

06

Learning Experience

02

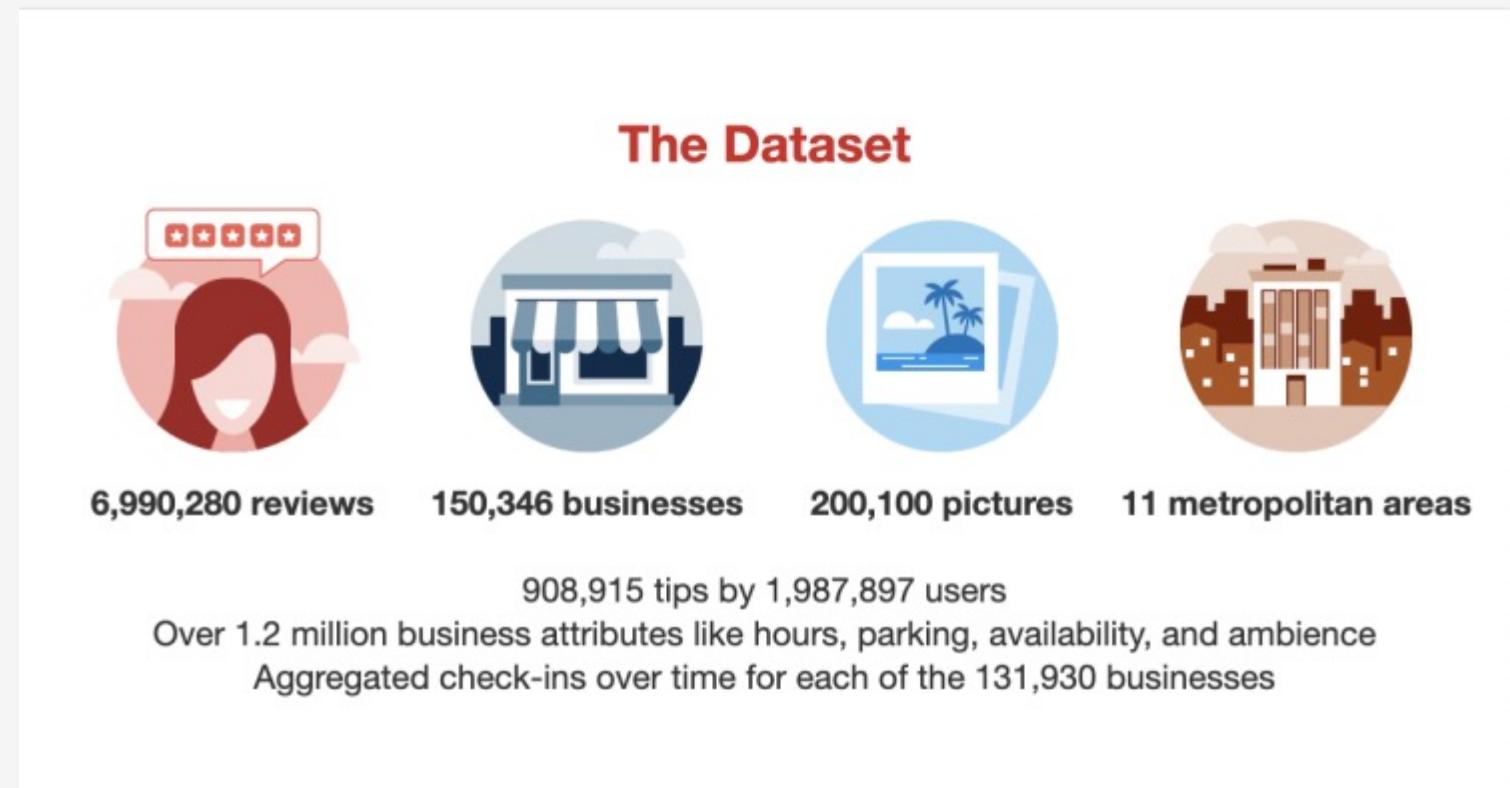
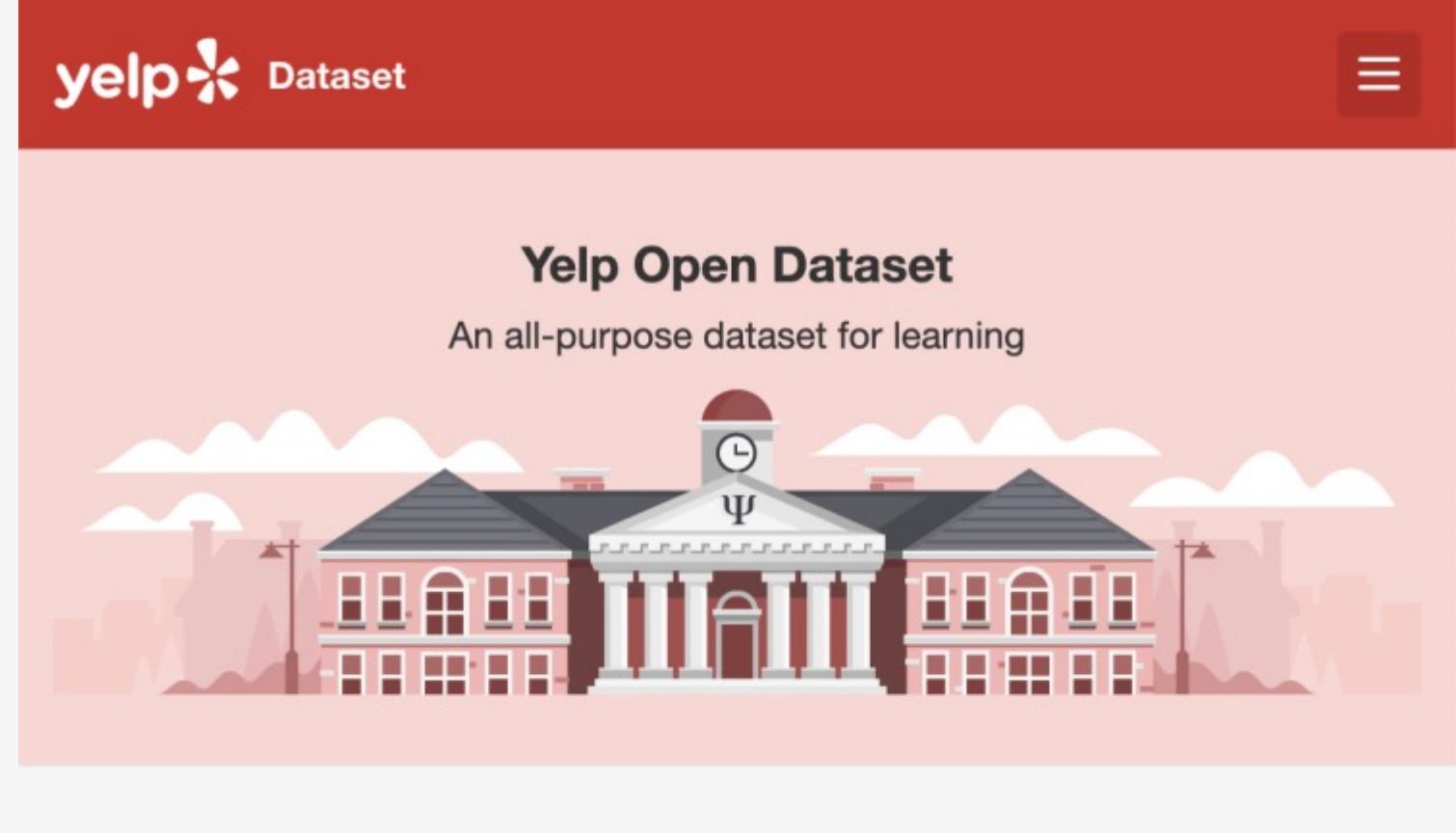
Business Objective

- **Yelp Dataset**

It provides reviews of restaurant.

User evaluation and recommendations using custom evaluation, user evaluation, user evaluation and recommendations of restaurants.

- Updates the dataset on a regular basis on its own.
- We can study and develop our recommendations system by looking at the updated dataset even after Machine Learning Term Project.



Data Exploration

Configuring datasets

- 1) Business.json - Restaurant data, such as location, category, and rating of a restaurant
- 2) Review.json - Review data for user-generated evaluations
- 3) User.json - User information data, such as user name, number of reviews, etc
- 4) Checkin.json - Business hours and number of people coming in at that time data
- 5) Tip.json - Short review data
- 6) Photos.json - Photo, company ID, caption of photo, and corresponding category data.

We need to convert to csv file for recommendation system implementation!

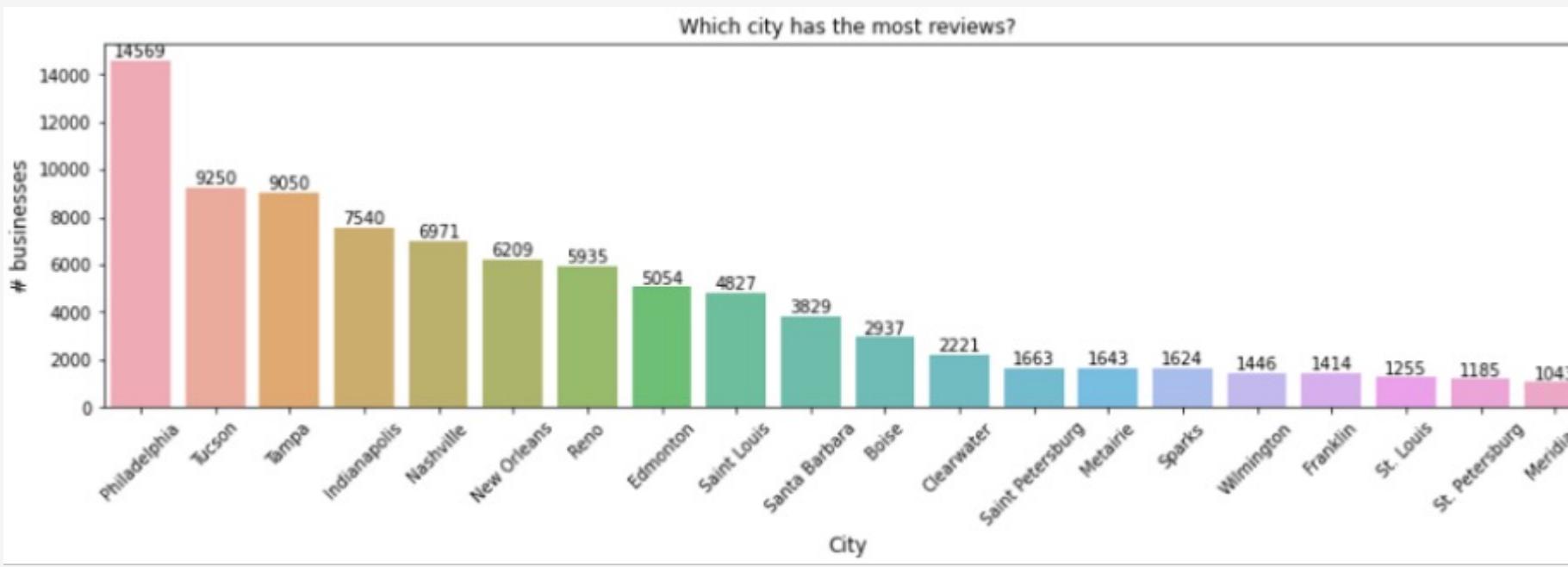
Data Exploration

Business.csv

```
[2]: import pandas as pd

df = pd.read_csv('data/business.csv')
df.head()
```

	business_id	name	address	city	state	latitude	longitude	stars	review_count	is_open	categories	hours
0	Pns2i4eNsfO8kk83dixA6A	Abby Rappoport, LAC, CMQ	1616 Chapala St, Ste 2	Santa Barbara	CA	34.426679	-119.711197	5.0	7	0	Doctors, Traditional Chinese Medicine, Naturop...	NaN
1	mpf3x-BjTdTEA3yCzrAYPw	The UPS Store	87 Grasso Plaza Shopping Center	Affton	MO	38.551126	-90.335695	3.0	15	1	Shipping Centers, Local Services, Notaries, Ma...	{'Monday': '0:0-0:0', 'Tuesday': '8:0-18:30', ...}
2	tUFrWirkIKI_TAnsVWINQQ	Target	5255 E Broadway Blvd	Tucson	AZ	32.223236	-110.880452	3.5	22	0	Department Stores, Shopping, Fashion, Home & G...	{'Monday': '8:0-22:0', 'Tuesday': '8:0-22:0', ...}
3	MTSW4McQd7CbVtyjqoe9mw	St Honore Pastries	935 Race St	Philadelphia	PA	39.955505	-75.155564	4.0	80	1	Restaurants, Food, Bubble Tea, Coffee & Tea, B...	{'Monday': '7:0-20:0', 'Tuesday': '7:0-20:0', ...}
4	mWMc6_wTdE0EUBKIGXDvFA	Perkiomen Valley Brewery	101 Walnut St	Green Lane	PA	40.338183	-75.471659	4.5	13	1	Brewpubs, Breweries, Food	{'Wednesday': '14:0-22:0', 'Thursday': '16:0-2...



Data Exploration

review.json

	review_id	user_id	business_id	stars	funny	cool	text	date
0	KU_O5udG6zpxOg-VcAEodg	mh_-eMZ6K5RLWhZylSBhwA	XQfwVwDr-v0ZS3_CbbE5Xw	3	0	0	If you decide to eat here, just be aware it is...	2018-07-07 22:09:11
1	BiTunyQ73aT9WBnpR9DZGw	OyoGAe7OKpv6SyGZT5g77Q	7ATYjTlgM3jUlt4UM3IypQ	5	0	1	I've taken a lot of spin classes over the year...	2012-01-03 15:28:18
2	saUsX_uimxRICVr67Z4Jig	8g_iMtfSiwikVnbP2etR0A	YjUWPpl6HXG530lwP-fb2A	3	0	0	Family diner. Had the buffet. Eclectic assortm...	2014-02-05 20:30:30
3	AqPFMleE6RsU23_auE\$xiA	_7bHUi9Uuf5__HHc_Q8guQ	kxX25Oes4o-D3ZQBkiMRfA	5	0	1	Wow! Yummy, different, delicious. Our favo...	2015-01-04 00:01:03
4	Sx8TMOWLNUJBWer-0pcmoA	bcjbaE6dDog4jkNY91ncLQ	e4Vwtrqf-wpJfwesgvdgxQ	4	0	1	Cute interior and owner (?) gave us tour of up...	2017-01-14 20:54:15

checkin.csv

```
[3]: df = pd.read_csv('data/checkin.csv')
df.head()

[3]:
      date    business_id
0 2020-03-13 21:10:56, 2020-06-02 22:18:06, 2020... ---kPU91CF4Lq2-WlRu9Lw
1 2010-09-13 21:43:09, 2011-05-04 23:08:15, 2011... --0iUa4sNDFiZFrAdlWhZQ
2 2013-06-14 23:29:17, 2014-08-13 23:20:22 --30_8lhuyMHbSOcNWd6DQ
3 2011-02-15 17:12:00, 2011-07-28 02:46:10, 2012... --7PUidqRWpRSpXebiyxTg
4 2014-04-21 20:42:11, 2014-04-28 21:04:46, 2014... --7jw19RH9JKXgFohspgQw
```

Data Exploration

tip.csv

	text	user_id	business_id	compliment_count	date
0	Avengers time with the ladies.	AGNUgVwnZUey3gcPCJ76iw	3uLgwr0qeCNMjKenHJwPGQ	0	2012-05-18 02:17:21
1	They have lots of good deserts and tasty cuban...	NBN4MgHP9D3cw--SnauTkA	QoezRbYQncpRqyrLH6Iqjg	0	2013-02-05 18:35:10
2	It's open even when you think it isn't	-copOvldyKh1qr-vzkDEvw	MYoRNLbSchwjQe3c_k37Gg	0	2013-08-18 00:56:08
3	Very decent fried chicken	FjMQVZjSqY8syIO-53KFKw	hV-bABTK-gh5wj31ps_Jw	0	2017-06-27 23:05:38
4	Appetizers.. platter special for lunch	ld0AperBXk1h6UbqmM80zw	_uN0OudeJ3Zl_tf6nxg5ww	0	2012-10-06 19:43:09

Data Exploration

user.csv

```
[5]: df = pd.read_csv('data/user.csv')
df.head()
```

	compliment_funny	user_id	average_stars	compliment_cool	compliment_plain	fans	yelping_since	compliment_note	friends	review_count	...
0	467	qVc8ODYU55ZjKXBgXdi7w	3.91	467	844	267	2007-01-25 16:47:26	232	NSCy54eWehBjyZdG2IE84w, pe42u7DcCH2Qml81NX-8qA...	585	...
1	3131	j14WgRoU_-2ZE1aw1dXrJg	3.74	3131	7054	3138	2009-01-25 04:35:42	1847	ueRPE0CX75ePGMqOFVj6IQ, 52oH4DrRvzzl8wh5UXyU0A...	4333	...
2	119	2WnXYQFK0hXEoTxPtV2zvg	3.32	119	96	52	2008-07-25 10:41:00	66	LuO3Bn4f3rlhyHlaNFTlnA, j9B4XdHUhDfTKVecyWQgyA...	665	...
3	26	SZDeASXq7o05mMNLshsdIA	4.27	26	16	28	2005-11-29 04:38:33	12	enx1vVPnfdNUdPho6PH_wg, 4w0cvMLtU6a9Lslggq74Vg...	224	...
4	0	hASlMy-EnncsH4JoR-hFGQ	3.54	0	1	1	2007-01-05 19:40:59	1	PBK4q9KEEBHhFvSXCUirlw, 3FWPpM7KU1gXeOM_ZbYMbA...	79	...

5 rows × 22 columns

compliment_cute	compliment_list	name	compliment_writer	compliment_photos	funny	useful	cool	compliment_more	compliment_profile
56	18	Walker	239	180	1259	7217	5994	65	55
157	251	Daniel	1521	1946	13066	43091	27281	264	184
17	3	Steph	35	18	1010	2086	1003	13	10
6	2	Gwen	10	9	330	512	299	4	1
0	0	Karen	0	0	15	29	7	1	0

Data Exploration

Ratings, reviews and locations largely influence a business.

So, we think these 3 the main factors one have to look into when we have this kind of dataset.

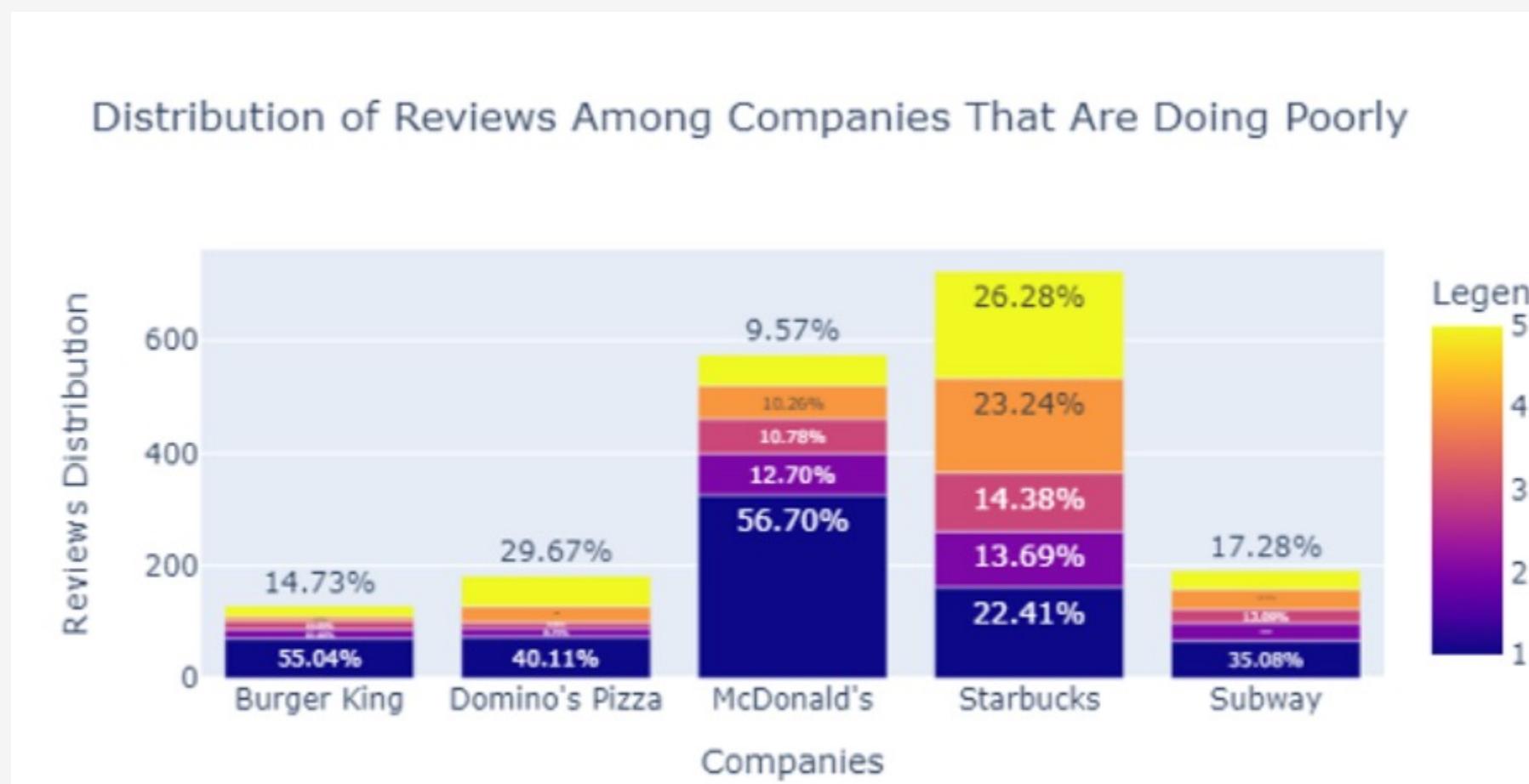
Hypothesis)

larger business with many locations, more difficult to maintain good rating.

Have more the number of reviews, stars get down.

Data Exploration

Grouping the dataset with avg rating and count of reviews and number of locations and sorting it city wise.



	review_count	stars	city
name			
McDonald's	17359	1.863442	263
Dunkin'	9864	2.302941	235
Starbucks	20692	3.126381	197
Wawa	4263	3.319218	167
Domino's Pizza	6291	2.281356	165
Taco Bell	8325	2.154795	164
Burger King	5255	2.028107	157
CVS Pharmacy	4481	2.456522	157
Subway	4123	2.586057	148
Wendy's	6164	2.015106	148
Pizza Hut	4588	2.033088	139
Walgreens	3974	2.624633	136
The UPS Store	4281	3.215302	131
Enterprise Rent-A-Car	4996	3.278017	126
Papa John's Pizza	2976	2.408163	114
Great Clips	1970	2.900000	100
Chick-fil-A	8004	3.385802	93
KFC	2923	1.859649	90
Applebee's Grill + Bar	5940	2.522321	89
Chipotle Mexican Grill	9383	2.426282	87

+

10

Data Exploration

Grouping the dataset with avg rating and count of reviews and number of locations and sorting it city wise.



	review_count	stars	city
name			
McDonald's	17359	1.863442	263
Dunkin'	9864	2.302941	235
Starbucks	20692	3.126381	197
Wawa	4263	3.319218	167
Domino's Pizza	6291	2.281356	165
Taco Bell	8325	2.154795	164
Burger King	5255	2.028107	157
CVS Pharmacy	4481	2.456522	157
Subway	4123	2.586057	148
Wendy's	6164	2.015106	148
Pizza Hut	4588	2.033088	139
Walgreens	3974	2.624633	136
The UPS Store	4281	3.215302	131
Enterprise Rent-A-Car	4996	3.278017	126
Papa John's Pizza	2976	2.408163	114
Great Clips	1970	2.900000	100
Chick-fil-A	8004	3.385802	93
KFC	2923	1.859649	90
Applebee's Grill + Bar	5940	2.522321	89
Chipotle Mexican Grill	9383	2.426282	87

+

11

Data Exploration

Find a company that is performing well with more than 1000 reviews

	review_count	stars	city
name			
European Wax Center	3317	4.107143	49
Nothing Bundt Cakes	1396	4.500000	29
Take 5 Oil Change	1032	4.179487	26
Trader Joe's	3556	4.348485	26
Discount Tire	2218	4.302632	17
Total Wine & More	1257	4.000000	15
Mission BBQ	2686	4.227273	10
Dutch Bros Coffee	1096	4.078125	8
Fleming's Prime Steakhouse & Wine Bar	2071	4.071429	7
The Capital Grille	2237	4.083333	6

Distribution of Reviews Among Companies That Are Doing Good

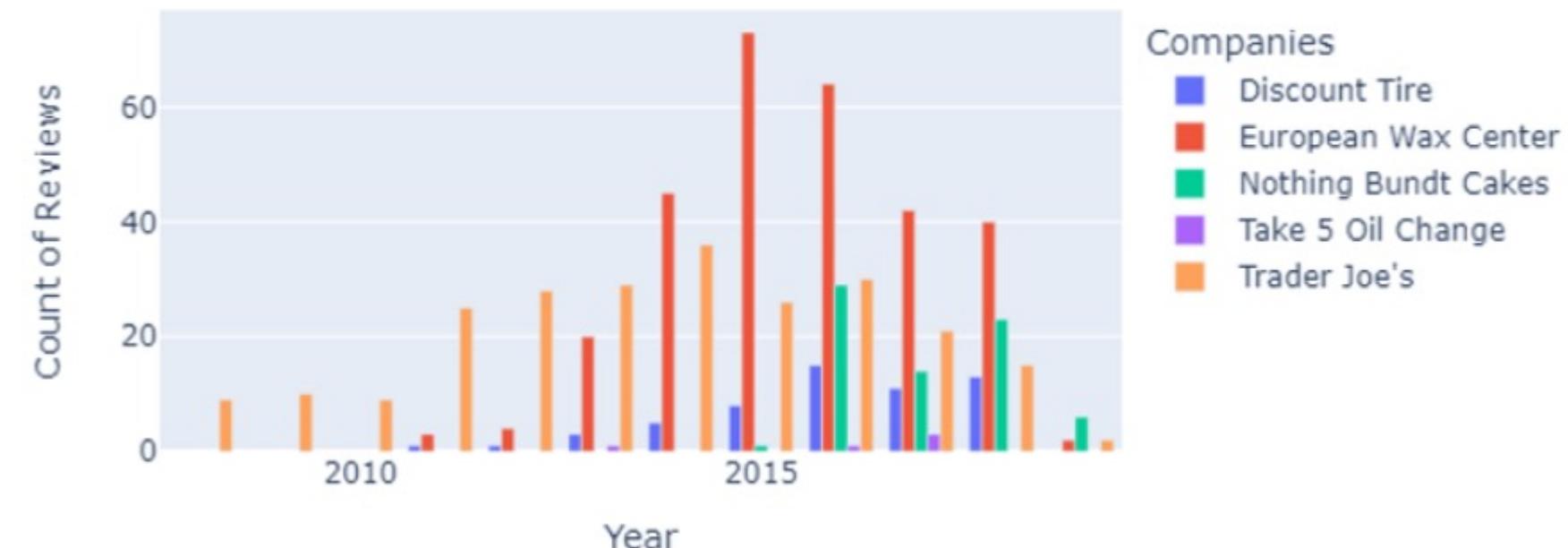


Data Exploration

Find a company that is performing well with more than 4 stars and more than 1000 reviews

	review_count	stars	city
name			
European Wax Center	3317	4.107143	49
Nothing Bundt Cakes	1396	4.500000	29
Take 5 Oil Change	1032	4.179487	26
Trader Joe's	3556	4.348485	26
Discount Tire	2218	4.302632	17
Total Wine & More	1257	4.000000	15
Mission BBQ	2686	4.227273	10
Dutch Bros Coffee	1096	4.078125	8
Fleming's Prime Steakhouse & Wine Bar	2071	4.071429	7
The Capital Grille	2237	4.083333	6

Distribution of Reviews overtime for selected good companies



We decided to proceed with collaborative filtering using only business data and review data!

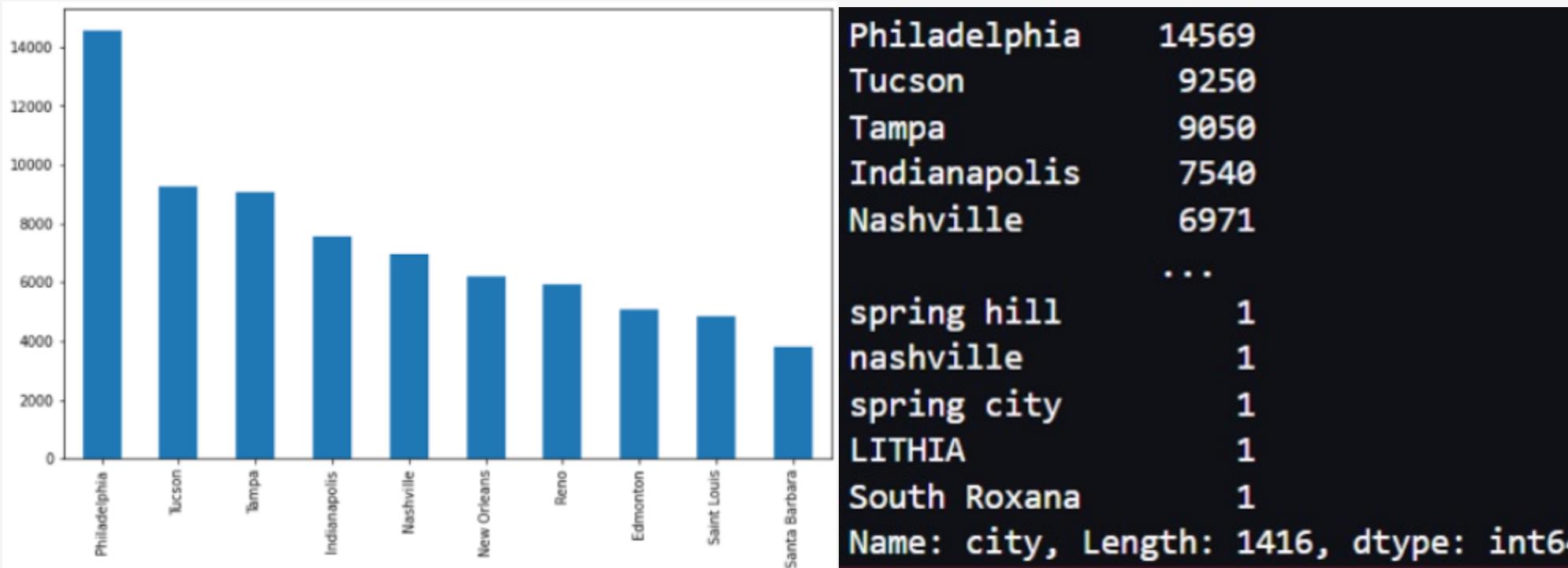
Data Preprocessing

1) Select the data for our model

Ratings, reviews and locations largely influence a business.

So, i think these 3 the main factors one have to look into when we have this kind of dataset.

1-1) Location Setting



We think it will serve the purpose of the recommendation system to conduct collaborative filtering within one area.

We chose "Philadelphia" because it has the highest number of reviews

Data Preprocessing

1-2) Businesses in Philadelphia and currently open business.

1-3) getting just restaurants from Philadelphia business.

1-4) Result.

	index	business_id	name	address	categories	stars
0	3	MTSW4McQd7CbVtyjqoe9mw	St Honore Pastries	935 Race St	Restaurants, Food, Bubble Tea, Coffee & Tea, B...	4.0
1	15	MUTTqe8uqyMdBl186RmNeA	Tuna Bar	205 Race St	Sushi Bars, Restaurants, Japanese	4.0
2	19	ROeacJQwBeh05Rqg7F6TCg	BAP	1224 South St	Korean, Restaurants	4.5
3	35	aPNXGTDkf-4bjhyMBQxqpQ	Craft Hall	901 N Delaware Ave	Entertainment, Arts & Entertainment, Brewpubs,...	3.5
4	82	ppFCk9aQkM338Rgwpl2F5A	Wawa	3604 Chestnut St	Restaurants, Automotive, Delis, Gas Stations, ...	3.0
...
3522	150178	auwFZzfhe2pvFw43OfsAfw	Stina Pizzeria	1705 Snyder Ave	Pizza, Restaurants, Mediterranean	4.5
3523	150210	K1SsvIPfFcHniNSPc3IG7g	Flip-N-Pizza	1308 W Girard Ave	Restaurants, American (Traditional), Chicken W...	4.0
3524	150237	OfRbGvHn-nAnQMqNxilpcA	Grilly Cheese	3306 Arch St	Specialty Food, Event Planning & Services, Res...	3.0
3525	150306	wVxXRFf10zTTAs11nr4xeA	PrimoHoagies	6024 Ridge Ave	Restaurants, Specialty Food, Food, Sandwiches,...	3.0
3526	150336	WnT9NizQgLlLjPT0kEcsQ	Adelita Taqueria & Restaurant	1108 S 9th St	Restaurants, Mexican	4.5

3527 rows × 6 columns

Data Preprocessing

2) Got dummies from categories columns: Remove nested properties

In column has nested attributes. In order to create a feature table, we need to separate those nested attributes into their own columns. Therefore, the following functions will be used to achieve this goal.

- Function that extract keys from the nested dictionary.
- Convert string to dictionary.
- Get dummies from categories.
- Concat all tables and drop restaurant column
- Map floating point stars to an integer

	Acai Bowls	Active Life	Adult	Adult Entertainment	Afghan	African	Airport Lounges	Airport Shuttles	American (New)	American (Traditional)	... Uzbek	Vegan	Vegetarian	Venues & Event Spaces	Vietnamese	Wine Bars	Wineries	Wraps	name	stars
0	0	0	0	0	0	0	0	0	0	0 ...	0	0	0	0	0	0	0	0	St Honore Pastries	4
1	0	0	0	0	0	0	0	0	0	0 ...	0	0	0	0	0	0	0	0	Tuna Bar	4
2	0	0	0	0	0	0	0	0	0	0 ...	0	0	0	0	0	0	0	0	BAP	5
3	0	0	0	0	0	0	0	0	0	1 ...	0	0	0	0	0	0	0	0	Craft Hall	4
4	0	0	0	0	0	0	0	0	0	0 ...	0	0	0	0	0	0	0	0	Wawa	3
...	
3522	0	0	0	0	0	0	0	0	0	0 ...	0	0	0	0	0	0	0	0	Stina Pizzeria	5
3523	0	0	0	0	0	0	0	0	0	1 ...	0	0	0	0	0	0	0	0	Flip-N-Pizza	4
3524	0	0	0	0	0	0	0	0	0	0 ...	0	0	0	0	0	0	0	0	Grilly Cheese	3
3525	0	0	0	0	0	0	0	0	0	0 ...	0	0	0	0	0	0	0	0	PrimoHoagies	3
3526	0	0	0	0	0	0	0	0	0	0 ...	0	0	0	0	0	0	0	0	Adelita Taqueria & Restaurant	5

3527 rows × 549 columns

Content Based Filtering

A method of recommending another item with content similar to the item when the user prefers a specific item.

We have a review and business(restaurant) data.

Through the review data, we will understand what characteristics the user likes and proceed with the process of recommending the restaurant with the characteristics.

Content Based Filtering

1) Gets the id value of the highly rated restaurant

Gets the data reviewed by a particular user, and gets the id value of the restaurant rated at a high rate.

'user_id_recommend': ID value of the target user. Proceed with the process of which restaurants to recommend to this user.

First, bring in the review data record left by the user, and then bring in business_id with high preference.

```
0      XQfwVwDr-v0ZS3_CbbE5Xw
45633   d_tRshM-w6S4QxE4VVi8tQ
Name: business_id, dtype: object
```

: The id value of the restaurant identified as the user's preference.

Content Based Filtering

2) Examine what characteristics the restaurants have.

- Restore information to the business dataset and select and output only important information.
- Since there is a lot of data, delete columns of the same meaning and identify overall preference characteristics.

Features of the user's preference:

- RestaurantsDelivery
- OutdoorSeating
- BusinessAcceptsCreditCards
- RestaurantsPriceRange2
- RestaurantsTakeOut
- WiFi
- Alcohol
- RestaurantsAttire
- GoodForKids
- RestaurantsTableService
- RestaurantsGoodForGroups
- WheelchairAccessible
- NoiseLevel
- lot
- casual
- lunch
- brunch
- breakfast
- American (New)
- Breakfast & Brunch
- Coffee & Tea
- Food
- Juice Bars & Smoothies
- Sandwiches

This is Characteristics of restaurants that are thought to be user-preferred →

Content Based Filtering

3) Use KNN for prediction model

- Verify that the data modified by preprocessing, data deleted due to data capacity issues, and selected columns are appropriate.
- The similarity between items (restaurant characteristics) is measured using KNN.
The target value was taken as a store's rating.

```
Score on training set: 0.8226975681177017
Score on test set: 0.7997428938723039
```

Content Based Filtering

4) Using Cosine similarity

- Look for a restaurant similar to 'reviewed_rest_feature' (characteristics of user preferred restaurant).
- A similarity is assigned to each restaurant data as a result of the calculation.
- The top five results of similarity are output.
(indicators are similar)

index	business_id	name	address	categories	attributes	stars	BusinessParking	Ambience	GoodForMeal	Dietary	Music
1867	XQfwVwDr-v0ZS3_CbbESxw	Turning Point of North Wales	1460 Bethlehem Pike	Restaurants, Breakfast & Brunch, Food, Juice B...	{'NoiseLevel': 'average', 'HasTV': 'False', ...}	3.0	{'garage': False, 'street': False, 'validated': ...}	{'touristy': False, 'hipster': False, 'romanti...}	{'dessert': False, 'latenight': False, 'lunch': ...}	0	0
3764	FYw5uf26ry8-MuPKD10z0Q	Josef's Vienna Bakery & Cafe	933 W Moana Ln	Grocery, Desserts, Breakfast & Brunch, Food, B...	{'WIFI': 'free', 'RestaurantsReservations': ...}	4.0	{'garage': False, 'street': False, 'validated': ...}	{'romantic': False, 'intimate': False, 'touris...}	{'dessert': False, 'latenight': False, 'lunch': ...}	0	0
32034	TZ9eAkKBKxj4HF_4rga9Gw	Turning Point - Bryn Mawr	925 W Lancaster Ave	American (Traditional), Juice Bars & Smoothies...	{'BusinessAcceptsCreditCards': 'True', 'Outdoo...}	3.5	{'garage': False, 'street': False, 'validated': ...}	{'touristy': False, 'hipster': False, 'romanti...}	{'dessert': False, 'latenight': False, 'lunch': ...}	0	0
71378	MZssFQddj218X-68N5XwaA	The Turning Point of Blue Bell	994 Dekalb Pike	Cafes, Restaurants, Breakfast & Brunch, Americ...	{'WiFi': 'free', 'RestaurantsPriceRange2': '...}	3.0	{'garage': False, 'street': None, 'validated': ...}	{'touristy': False, 'hipster': False, 'romanti...}	{'dessert': False, 'latenight': False, 'lunch': ...}	0	0
145958	U9hqH3gn7yGaSue3gQUUw	The Kettle & Grille	230 N Maple Ave	Restaurants, American (New), Breakfast & Brunch	{'RestaurantsTableService': 'True', 'BikeParki...}	4.0	{'garage': False, 'street': False, 'validated': ...}	{'romantic': False, 'intimate': False, 'classy': ...}	{'dessert': None, 'latenight': False, 'lunch': ...}	0	0

Content Based Filtering

5) Result

- Print out the characteristics of these recommended restaurants in detail.
- Save only features that correspond to at least 3 times out of 5 recommended restaurants.
- Out of more than 1300 features in total, about 25 can be selected as major features.

RestaurantsDelivery_True	5
OutdoorSeating_True	4
BusinessAcceptsCreditCards_True	5
BikeParking_True	3
RestaurantsPriceRange2_2	5
RestaurantsTakeOut_True	5
WiFi_free	3
Alcohol_none	4
RestaurantsAttire_casual	4
GoodForKids_True	5
RestaurantsTableService_True	5
RestaurantsGoodForGroups_True	5
WheelchairAccessible_True	3
NoiseLevel_u'average'	4
lot_True	5
casual_True	4
lunch_True	5
brunch_True	4
breakfast_True	5
Breakfast & Brunch	5
Coffee & Tea	4
Food	4
Juice Bars & Smoothies	3
Restaurants	3
Sandwiches	4

Content Based Filtering

Compared to previous features

Restaurant features that tend to be preferred by target users	the characteristics of recommended restaurants
<p>Features of the user's preference:</p> <ul style="list-style-type: none"> - RestaurantsDelivery - OutdoorSeating - BusinessAcceptsCreditCards - RestaurantsPriceRange2 - RestaurantsTakeOut - WiFi - Alcohol - RestaurantsAttire - GoodForKids - RestaurantsTableService - RestaurantsGoodForGroups - WheelchairAccessible - NoiseLevel - lot - casual - lunch - brunch - breakfast - American (New) - Breakfast & Brunch - Coffee & Tea - Food - Juice Bars & Smoothies - Sandwiches 	<pre> RestaurantsDelivery_True 5 OutdoorSeating_True 4 BusinessAcceptsCreditCards_True 5 BikeParking_True 3 RestaurantsPriceRange2_2 5 RestaurantsTakeOut_True 5 WiFi_free 3 Alcohol_none 4 RestaurantsAttire_casual 4 GoodForKids_True 5 RestaurantsTableService_True 5 RestaurantsGoodForGroups_True 5 WheelchairAccessible_True 3 NoiseLevel_u'average' 4 lot_True 5 casual_True 4 lunch_True 5 brunch_True 4 breakfast_True 5 Breakfast & Brunch 5 Coffee & Tea 4 Food 4 Juice Bars & Smoothies 3 Restaurants 3 Sandwiches 4 dtype: int64 </pre>

Collaborative Filtering

A method based on the idea that similar users may have similar restaurant preferences

This section makes recommendations to restaurant users using collaborative filtering techniques.

Implement machine learning techniques such as SVD and PCA to build a recommendation system.

Data Preprocessing

for Collaborative Filtering

- pull out needed columns from subset_review table
 - pull out names and addresses of the restaurants from rest table
- rest table: A table extracted from restaurants in Philadelphia
- combine df_review and restaurant table.
 - Complete combined_business_data.

	user_id	business_id	stars	date	name	address
0	_7bHUi9Uuf5__HHc_Q8guQ	kxXSOes4o-D3ZQBkiMRfA	5	2015-01-04 00:01:03	Zaika	2481 Grant Ave
1	kSMOJwJXuEUqzfmuFncK4A	kxXSOes4o-D3ZQBkiMRfA	2	2014-07-13 17:25:47	Zaika	2481 Grant Ave
2	mqBWACmaHflW4eh_Ofp16Q	kxXSOes4o-D3ZQBkiMRfA	5	2010-08-20 19:16:04	Zaika	2481 Grant Ave
3	Z-xgVb4nM42943m2wbBkFw	kxXSOes4o-D3ZQBkiMRfA	5	2017-01-02 14:25:26	Zaika	2481 Grant Ave
4	2SEoXb6r6hPKrl9V9VzBgA	kxXSOes4o-D3ZQBkiMRfA	5	2015-07-28 17:15:20	Zaika	2481 Grant Ave
...
16254	KCeHOGhibdbYSF2lLehZWA	duf2PORLOL6it1dTGhSnKA	1	2018-09-23 22:25:43	Junes Breakfast	6227 Lancaster Ave
16255	q9Yb2Bt3k99CcrcGWCWZew	SADPX8WX34EQDj_jqtacMw	1	2018-09-17 22:47:32	Bai Wei	1038 Race St
16256	9xfWUskLWXHFKC-xrm1K9g	MGNAYydr7SxPMDgFlnZ7sg	4	2014-04-27 16:48:14	Little Caesars Pizza	200 W Oregon Ave
16257	Ed0GuUa_1UjbSk8qaQsHIA	ELJcgId3RiCADVIT5QbUTA	3	2012-09-18 15:43:06	Hyon's Seafood	1320 S 21st St
16258	6a-782Y5FXqJ7KGmQtJsSQ	s55Hw6gHhQzZctinFhPu-g	3	2016-07-20 13:49:41	Urban Eatery	3400 Lancaster Ave

16259 rows × 6 columns

Data Preprocessing

for Collaborative Filtering

- Sorting based on star, then only head(5) of the business_id of the most famous restaurants (most rated restaurants) is output.

Most Popular Restaurants:

Restaurant 1:

Name: ['Village Whiskey']
Address: ['118 S 20th St']

Restaurant 2:

Name: ["Honey's Sit-N-Eat"]
Address: ['800 N 4th St']

Restaurant 3:

Name: ['HipCityVeg']
Address: ['127 S 18th St']

Restaurant 4:

Name: ['Penang']
Address: ['117 N 10th St']

Restaurant 5:

Name: ['El Camino Real']
Address: ['1040 N 2nd St']

Collaborative Filtering

Building a utility Matrix (User-Restaurant Matrix)

- This matrix contains each user, each restaurant, and the rating each user gave to each restaurant.
- Notice this matrix will be sparse because every user doesn't review every restaurant.
- After finish the configuring, the matrix is transposed to represent the user as a column and the restaurant as a row.

[20]:

	name	1 Stop Pizza	16th Street Seafood	2 in One Cafe	3J's Food Market	4 Seasons Mexican Restaurant	African Small Pot	All's Italiano Pizzeria & Ristorante	Allegro Pizza	Angelina's	Ants Pants Café	...	Wendy's	Wilson Famous Blue Ribbon Meats	Wilson's Market	Wings & More	Wishbone	Xochitl	Zaika	Zio Pizza Palace & Grill	Zorba's Taverna	iPho Vietnamese Restaurant	
	user_id																						
-4AjkzIHowEiBCMd4CZA		0	0	0	0	0	0	0	0.0	0	0.0	...	0	0	0	0	0.0	0	0	0	0	0	0
-6PFZka7og6Khaw6oyjvQ		0	0	0	0	0	0	0	0.0	0	0.0	...	0	0	0	0	0.0	0	0	0	0	0	0
S8M395r8NtOCvS2LRfDw		0	0	0	0	0	0	0	0.0	0	0.0	...	0	0	0	0	0.0	0	0	0	0	0	0
--UizzbnQlZg7bEv2oXEyg		0	0	0	0	0	0	0	0.0	0	0.0	...	0	0	0	0	0.0	0	0	0	0	0	0
ccVMj2PN6Z9qtdOdlung		0	0	0	0	0	0	0	3.0	0	0.0	...	0	0	0	0	0.0	0	0	0	0	0	0

5 rows x 279 columns

Collaborative Filtering

Decomposing the Matrix

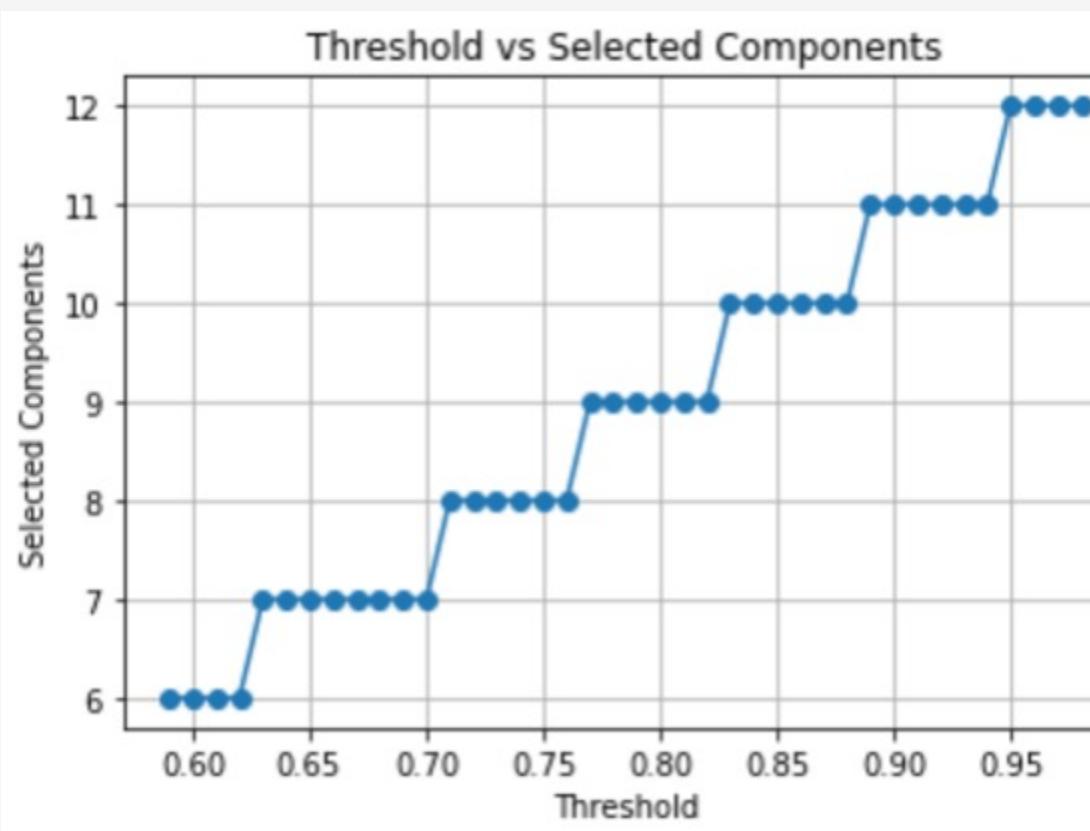
- Now we use Sklearn's truncatedSVD to compress the transpose matrix into the number of rows of 12 matrices.
- It compresses users arbitrarily with 12 components for their tastes.
- Select n_components from truncatedSVD
-> Dimension reduction in data and trade-offs in performance should be considered.

Collaborative Filtering

Decomposing the Matrix

- Create and apply PCA models to find the optimal values for `n_components`.

`n_components = 12`



Threshold: 0.59, 선택된 주성분의 개수: 6
 Threshold: 0.6, 선택된 주성분의 개수: 6
 Threshold: 0.61, 선택된 주성분의 개수: 6
 Threshold: 0.62, 선택된 주성분의 개수: 6
 Threshold: 0.63, 선택된 주성분의 개수: 7
 Threshold: 0.64, 선택된 주성분의 개수: 7
 Threshold: 0.65, 선택된 주성분의 개수: 7
 Threshold: 0.66, 선택된 주성분의 개수: 7
 Threshold: 0.67, 선택된 주성분의 개수: 7
 Threshold: 0.68, 선택된 주성분의 개수: 7
 Threshold: 0.6900000000000001, 선택된 주성분의 개수: 7
 Threshold: 0.7000000000000001, 선택된 주성분의 개수: 7
 Threshold: 0.7100000000000001, 선택된 주성분의 개수: 8
 Threshold: 0.7200000000000001, 선택된 주성분의 개수: 8
 Threshold: 0.7300000000000001, 선택된 주성분의 개수: 8
 Threshold: 0.7400000000000001, 선택된 주성분의 개수: 8
 Threshold: 0.7500000000000001, 선택된 주성분의 개수: 8
 Threshold: 0.7600000000000001, 선택된 주성분의 개수: 8
 Threshold: 0.7700000000000001, 선택된 주성분의 개수: 9
 Threshold: 0.7800000000000001, 선택된 주성분의 개수: 9
 Threshold: 0.7900000000000001, 선택된 주성분의 개수: 9
 Threshold: 0.8000000000000002, 선택된 주성분의 개수: 9
 Threshold: 0.8100000000000002, 선택된 주성분의 개수: 9
 Threshold: 0.8200000000000002, 선택된 주성분의 개수: 9
 Threshold: 0.8300000000000002, 선택된 주성분의 개수: 10
 Threshold: 0.8400000000000002, 선택된 주성분의 개수: 10
 Threshold: 0.8500000000000002, 선택된 주성분의 개수: 10
 Threshold: 0.8600000000000002, 선택된 주성분의 개수: 10
 Threshold: 0.8700000000000002, 선택된 주성분의 개수: 10
 Threshold: 0.8800000000000002, 선택된 주성분의 개수: 10
 Threshold: 0.8900000000000002, 선택된 주성분의 개수: 11
 Threshold: 0.9000000000000002, 선택된 주성분의 개수: 11
 Threshold: 0.9100000000000003, 선택된 주성분의 개수: 11
 Threshold: 0.9200000000000003, 선택된 주성분의 개수: 11
 Threshold: 0.9300000000000003, 선택된 주성분의 개수: 11
 Threshold: 0.9400000000000003, 선택된 주성분의 개수: 11
 Threshold: 0.9500000000000003, 선택된 주성분의 개수: 12
 Threshold: 0.9600000000000003, 선택된 주성분의 개수: 12

Collaborative Filtering

Generate a correlation matrix

- We calculated PearsonR coefficient for every restaurant pair in the result_matrix.
- The correlation-based on similarities between users' tastes.

Isolating the most popular restaurant from the Correlation Matrix

- Extract the correlation values between the top five restaurants with all other restaurants from corr_matrix.

Collaborative Filtering

Recommend highly correlated restaurants

- Village Whiskey - Bars, Nightlife, Whiskey Bars, Burgers, Restau...
 : ["Chan's Halal Restaurant", 'Moy Fong Restaurant', "Sister Muhammad's Kitchen"]
- Honey's Sit-N-Eat - Southern, Restaurants, American (Traditional)
 : ['16th Street Seafood', "George's Sandwich Shop", 'Quality Pizza']
- HipCityVeg - Burgers, Vegetarian, Restaurants, Vegan, Sandwiches, Breakfast & Brunch,...
 : ['Fresh Gulf Seafood', 'Ricky Cafe']
- Penang - Malaysian, Asian Fusion, Cambodian, Chinese, R..
 : ["Applebee's Grill + Bar", 'Mayfair Pizza', 'New England Pizza', "Potito's Italian American Pastries", 'Wawa Food Markets']
- El Camino Real - Mexican, Tex-Mex, Restaurants, Barbeque
 : ['Bonks Bar', "Ed's Pizza House", 'El Camino Real', 'Little Kitchen', 'Los Potrillos', "Ron's Caribbean Cuisine", "Sorrento's Pizza", 'The Freshworks of Port Richmond', "Wendy's", "Wilson's Market"]

Evaluation and Analysis

Content-based Filtering

Performance of the KNN

```
Score on training set: 0.8226975681177017
Score on test set: 0.7997428938723039
```

Actual Selected Features

Restaurant features that tend to be preferred by target users	the characteristics of recommended restaurants
Features of the user's preference: <ul style="list-style-type: none"> - RestaurantsDelivery - OutdoorSeating - BusinessAcceptsCreditCards - RestaurantsPriceRange2 - RestaurantsTakeOut - WiFi - Alcohol - RestaurantsAttire - GoodForKids - RestaurantsTableService - RestaurantsGoodForGroups - WheelchairAccessible - NoiseLevel - lot - casual - lunch - brunch - breakfast - American (New) - Breakfast & Brunch - Coffee & Tea - Food - Juice Bars & Smoothies - Sandwiches 	RestaurantsDelivery_True 5 OutdoorSeating_True 4 BusinessAcceptsCreditCards_True 5 BikeParking_True 3 RestaurantsPriceRange2_2 5 RestaurantsTakeOut_True 5 WiFi_free 3 Alcohol_none 4 RestaurantsAttire_casual 4 GoodForKids_True 5 RestaurantsTableService_True 5 RestaurantsGoodForGroups_True 5 WheelchairAccessible_True 3 NoiseLevel_u'average' 4 lot_True 5 casual_True 4 lunch_True 5 brunch_True 4 breakfast_True 5 Breakfast & Brunch 5 Coffee & Tea 4 Food 4 Juice Bars & Smoothies 3 Restaurants 3 Sandwiches 4 dtype: int64

Evaluation and Analysis

Collaborative Filtering

Method 1. Hit Rate

1. Gets all of the items that you've preferred or clicked on for a particular user.
2. Percentage of N items recommended to the user → Leave-One-Out Cross-Validation that indicates whether the user actually evaluated any items
3. After learning the recommendation model with the remaining items, extract the Top K recommendation list
4. If there is an item on the K recommendation list that I removed earlier, it's a hit, or it's not a hit.

Hit Rate: 0.010752688172043012

However, due to the nature of the dataset, the Hit Rate is judged to be low depending on the text nature.

Evaluation and Analysis

Collaborative Filtering

Recommend restaurants to each user from the entire dataset.

```
User Recommendations:
User _7bHu9Uuf5_HHc_Q8guQ: ['1 Stop Pizza', 'New Lee's Chinese Restaurant', 'New Napoli Famous Pizza']
User kSM0JwJXuEUqzfmuFnck4A: ['George's Sandwich Shop', 'Honey's Sit-N-Eat', 'Quality Pizza']
User mqBiACmaHflm4eh_Ofp16Q: ['2 in One Cafe', 'Pasqually's Pizza']
User Z-xgVb4nM42943m2wbBkFw: ['3J's Food Market', 'Dunkin''', 'Federal Donuts', 'Franny Lou's Porch', 'McDonald's']
User 2SEoXb6r6hPKrl9V9VzBgA: []
User tofMojWZJKqs0mhE7khbRA: []
User cMvau1VAjuko_ah0Bck1g: []
User DVModL-MS3_03qMw0Dxa1A: ['China House']
User rmYqxdycak7SL9vY7xE21A: ['Giovanni's Pizzeria']
User MpIZMLdDjZRu5BzGyEjEOw: ['Ants Pants Café', 'Fabio's Chicken Restaurant Takeout', 'Trattoria Carina', 'Vetri Cucina']
User 3fTw9hmzbMK2Q3Q-a4Hw: ['Mayfair Pizza', 'New England Pizza', 'Penang', 'Potito's Italian American Pastries']
User JXB91quCyabpRX0fwP-RDA: ['Mike's Breakfast']
User LOVARg-JZc7Guw_fy2EgHQ: ['Loews Philadelphia Hotel', 'Saladworks']
User Uzzwl_rQkjA11iRzd-e-ZA: []
User e4qo1KDdmKgPyC8DA97EkQ: []
User DFyRD_pEpEhbNzSMH_q5Mg: ['Axis Pizza', 'Hyon's Seafood']
User F7j1AubXAJl7BdLoQ54QQ: []
User NYTkPfR8dIZMKuak8iOIEg: []
User J8i_t7Dw432k_YODtxNfQ: ['BAP', 'Cucina Zapata', 'Jasmine Rice - Rittenhouse', "Mike's Hot Dog Cart"]
User aTlICJkPxwhCeEojjVRyg: ['Bai Wei']
User cUoKFJ4PDsZOTkxmEt-HeQ: []
User 6cjXa37M_QFvrkdqNnmEGg: ['Beck's Cajun Cafe', 'Taco Bell']
User c4rJccGZ0gR1-C65qbehQ: []
User s1Cyd2IS0dlczPsCSP87A: ['Ed's Pizza House', 'El Camino Real', 'Little Kitchen', 'Los Potrillos', 'Ron's Caribbean Cuisine']
User 8lzJun8DssNGjxc4qIZIfA: []
User epoI_-7_JjGAfVZHctImA: []
User BbhS3WUmTuZi7RGvLMiDug: ['Smiley's Cafe', 'Urban Eatery']
User 0F1ic9kbHPkrinSSQ_ActQ: []
User j84sng6u1a60lNPBGaWZRA: []
User ny63PhlBidEZzEg50tdZUw: ['Chef Joseph Poon', 'Crazy Sushi', 'Minas Pizza', "Pina's Pizza", 'Pizzeria Vetrini']
User YUUaNbQdj8SEK-59FTjtfc: ['Mike's Hot Dog Cart']
User fLWRFDbyYeZeV4ShQCsMnA: ['Burger King']
User SA--iAwtfIfeR4WEU8-6omg: ['Minas Pizza']
User EMMKdK1cGrLS008wgSYHjq: []
User 4d5oNl9oHazgXmXyuvTQw: []
User fbVlaq9ASXggQyGTNN-hzg: []
User 9VJ5zcQ26ff6i-uFKVbPaA: ['Gachi Sushi and Noodle']

User 4TSPgc-rnJsEhZibbbhwhkw: ['Tokai Sushi Hibachi Steakhouses']
User L5CIViS8g309rjfN_IJrsA: ['Mike's Breakfast', 'Three 12 Sport Bar and Lounge']
User S6DjAyTFjooafp5KuMY1MA: ['Tomo Sushi & Ramen']
User UBVi6kQTEscWxopDviMozw: []
User DCQ9w5F1MS-BNWRPjN-DFg: ['Chef Joseph Poon', 'Kumo Hibachi Sushi', "Pina's Pizza", 'Pizzeria Vetri']
User yerQSk2GKkG8kocvbwItpg: ['Golden China Wok Restaurant']
User xg3H0i3SMh-c_p_Z_shVog: []
User JuHHlydQ7RxjWmtZpSEZvA: ['Ants Pants Café', 'DK Sushi', "Fabio's Chicken Restaurant Takeout", 'Vetri Cucina']
User us1stewu4_gmuBoLcUUWg: ['Oteri's Italian Bakery | North 5th St', 'Ramble Pizza', 'Tria Cafe Wash West']
User kH15up-Q1WfVrgBpFMMRCA: []
User I0ezkVvCcBnavJf2solHja: []
User itOuvBF5oHLGQELQ5mnF4A: ['3J's Food Market', 'Dunkin''', 'Federal Donuts', 'Franny Lou's Porch', 'McDonald's']
User T9w31gVwdy8ZMN0kLaWSSQ: ['Cedar Park Cafe', 'Happy Joy Family Restaurant', 'New Lee's Chinese Restaurant', 'New Napoli Famous Pizza', 'New Tavern Bar']
User kTpjyDUh1981R10FVuNreA: ['Breakfast Boutique Chestnut Hill', 'Kumo Hibachi Sushi', 'Pearl of East']
User bHPpSsGbFM9BxwiG30YqTA: ['Urban Village Brewing Co']
User YUY3_A6NUZTYnGQRsX9gpg: ['Ants Pants Café', 'Trattoria Carina', 'Vetri Cucina']
User HhMnurW1J7jMANSAgisL_Q: []
User YmD-AVT67fmYRGxnIRSPA: ['Copabana', 'Queen Sheba', 'Tasties']
User A3TLdhgsYKybx3SsDdffza: ['Buena Onda', 'Jong Ka Jib', 'Kumo Hibachi Sushi']
User KgN-TlzBr-n_4pX9C7slw: ['Chan's Halal Restaurant', 'Moy Fong Restaurant', 'Sister Muhammad's Kitchen', 'Village Whiskey']
User ie9VaHvAaQgjMJP3Gn-AQQ: ['Pearl of East']
User iSxfxxndt8ljpIJ2EjyhrQ: ['Wawa']
User zu-e06_BM_TdkAZEKMrIw: ['Penang', 'Wawa Food Markets']
User y9B_A0ZCE0S03ut40kUt0Q: ['Bonks Bar', "Ed's Pizza House", 'El Camino Real', 'Little Kitchen', 'Los Potrillos']
User xCkdy6W5-Wix3LCrrqfNqw: ['Rangoli: Vibrant Flavors of India']
User AGqdhuVNL-KIpQw1vLU8cw: ['Bonks Bar', "Ed's Pizza House", 'El Camino Real', 'Little Kitchen', 'Los Potrillos']
User N7NnC6-q00-Gq2n_7YjkNg: ['Foodery']
User uCmP9M1PEXU9c77_sdUOTA: ['Jasmine Rice - Rittenhouse']
User trv2a6m_08_83ay2-CgoEQ: ['Lee How Fook', 'Square 1682']
User WYvnQmNKT401kj6f1vTEAQ: ['Dagwood's Pub', 'Domino's Pizza', 'Guavaberry Foods & Drinks', 'Happy Panda', 'Khyber Pass Pub']
User Ohu0VDjvejnucgyFXex26g: ['Dagwood's Pub', 'Khyber Pass Pub', 'Spice Finch', 'The Palace of Indian', 'Zaika']
User Jyq3AFGfi0kVhFpjvwHMqA: ['Zorba's Taverna']
User py6wYGuBEt5MwzGn1dbmNw: []
```

Evaluation and Analysis

Collaborative Filtering

Method 2.

Find user_id with the most rows

가장 많은 행을 가지는 user_id: 0DB3Irpf_ETVXu_Ou9vPow

1. Check the applicable table ->

	user_id	business_id	stars	\
243	0DB3Irpf_ETVXu_Ou9vPow	eFvzHawVJofxSnD7TgbZtg	5	
873	0DB3Irpf_ETVXu_Ou9vPow	EtKSTHV5Qx_Q7Aur9o4kQQ	4	
1263	0DB3Irpf_ETVXu_Ou9vPow	VJEzpfLs_Jnzgqh5A_FVTg	3	
2100	0DB3Irpf_ETVXu_Ou9vPow	Dv6RfXLYe1atjgz3Xf4GGw	4	
2873	0DB3Irpf_ETVXu_Ou9vPow	eaDZ1SuVS0EY67Ke6pRP6Q	4	
3651	0DB3Irpf_ETVXu_Ou9vPow	-OIUunijjcq_ZzyyQhPPFQ	5	
6235	0DB3Irpf_ETVXu_Ou9vPow	C_EtrXTygRX5RTUOKtO6Dg	4	
7825	0DB3Irpf_ETVXu_Ou9vPow	U30ggGzFpXvc2NZYwOW3qg	5	
8197	0DB3Irpf_ETVXu_Ou9vPow	PYUI10JVksGUbCrteU68bw	5	
10012	0DB3Irpf_ETVXu_Ou9vPow	lRbHFOIFuusN2WOR_ypQ_A	5	
10253	0DB3Irpf_ETVXu_Ou9vPow	14ZGwnDyydXdSBsLXpSUrA	5	
10684	0DB3Irpf_ETVXu_Ou9vPow	b5WL0Yxfawjs03_g2kgujA	3	
10856	0DB3Irpf_ETVXu_Ou9vPow	uVJkqQ1U9MH7hBorrwGhBg	3	
12171	0DB3Irpf_ETVXu_Ou9vPow	F8yoZE3NWnImNAphO347gQ	5	
12862	0DB3Irpf_ETVXu_Ou9vPow	D0121AqtT74dUZ3kvXNnCa	4	
13179	0DB3Irpf_ETVXu_Ou9vPow	ROeacJQw6eh05Rqg7F6TCg	4	
13934	0DB3Irpf_ETVXu_Ou9vPow	zwTm0j4B_OVPMTMYijQiKg	4	
14295	0DB3Irpf_ETVXu_Ou9vPow	MF5-JPr0auUkY_WLTTVUaA	5	
15460	0DB3Irpf_ETVXu_Ou9vPow	eqNutOMnPqKa9QDAY6KFew	5	
15538	0DB3Irpf_ETVXu_Ou9vPow	IH_ZeeTh13jQURh31d1Kw	4	

	date	name	address
243	2018-03-18 22:08:06	Good Karma Cafe	928 Pine St
873	2016-09-14 01:21:40	Village Whiskey	118 S 20th St
1263	2017-06-01 02:06:57	Jasmine Rice - Rittenhouse	306 S 16th St
2100	2017-03-29 23:53:50	HipCityVeg	127 S 18th St
2873	2017-09-10 02:31:38	Penang	117 N 10th St
3651	2018-05-09 20:41:28	Zorba's Taverna	2230 Fairmount Ave
6235	2015-11-08 18:06:36	Copabana	4000 Spruce St
7825	2015-11-07 17:06:42	Mood Cafe	4618 Baltimore Ave
8197	2016-09-11 19:12:00	Bourbon & Branch	705 N 2nd St
10012	2016-06-25 18:30:37	Ants Pants Café	2212 S St
10253	2016-07-18 23:30:18	Vientiane Café	4728 Baltimore Ave
10684	2015-11-08 17:52:22	Allegro Pizza	3942 Spruce St
10856	2015-11-07 03:48:07	Wishbone	4034 Walnut St
12171	2018-06-09 03:10:24	Marrakesh	517 S Leithgow St
12862	2017-10-16 22:18:28	Golden Triangle	610 S 5th St
13179	2016-12-20 01:35:26	BAP	1224 South St
13934	2015-11-08 17:29:29	Queen Sheba	4511 Baltimore Ave
14295	2018-02-11 23:17:39	Tacodelphia	427 S Broad St
15460	2018-07-08 15:15:29	J'aime French Bakery	212 S 12th St
15538	2018-06-26 14:33:30	Bui's	38th St & Spruce St

Evaluation and Analysis

Collaborative Filtering

2. Import restaurants evaluated by the user
3. Find other restaurants except those already evaluated by the user
4. Finding restaurants that other users like
5. Output the list of recommended restaurants

Recommended Restaurants for User 0DB3Irpf_ETVXu_0u9vPow:		
	name	address
0	Zaika	2481 Grant Ave
65	Rittenhouse Grill	1701 Locust St
291	The Coventry Deli	2000 Market St
316	Square 1682	121 S 17th St
1278	Chase's Hop Shop	7235 Rising Sun Ave
...
16254	Junes Breakfast	6227 Lancaster Ave
16255	Bai Wei	1038 Race St
16256	Little Caesars Pizza	200 W Oregon Ave
16257	Hyon's Seafood	1320 S 21st St
16258	Urban Eatery	3400 Lancaster Ave

Just recommend restaurants which have high stars from other users to a specific user.

Comparison to previous projects

With the same data in the past

Encode and prepare data: Encode business and user IDs using Sklearn's Label Encoder and store unique users, restaurants, minimum ratings, and maximum ratings.

Data segmentation: Split data into training and test sets.

Factor definition for a model: Create a variable to store the number of factors per user/restaurant for the collaborative filtering model.

Save users and restaurants: Make separate arrangements for users and restaurants input to train and test sets

Embedded representation: uses embedding to represent users and restaurants in the data.

Comparison to previous projects

With the same data in the past

Model Enhanced: Adds bias to each embedding; runs the output of dot products through the sigmoid layer and scales the results using the minimum and maximum ratings of the data to improve model performance.

Prediction from Test Dataset: Use the generated model to predict the test dataset and compare the actual stars to predict the model performance.

Calculate cosine similarity: Use cosine similarity to measure cosine similarity between two restaurants.

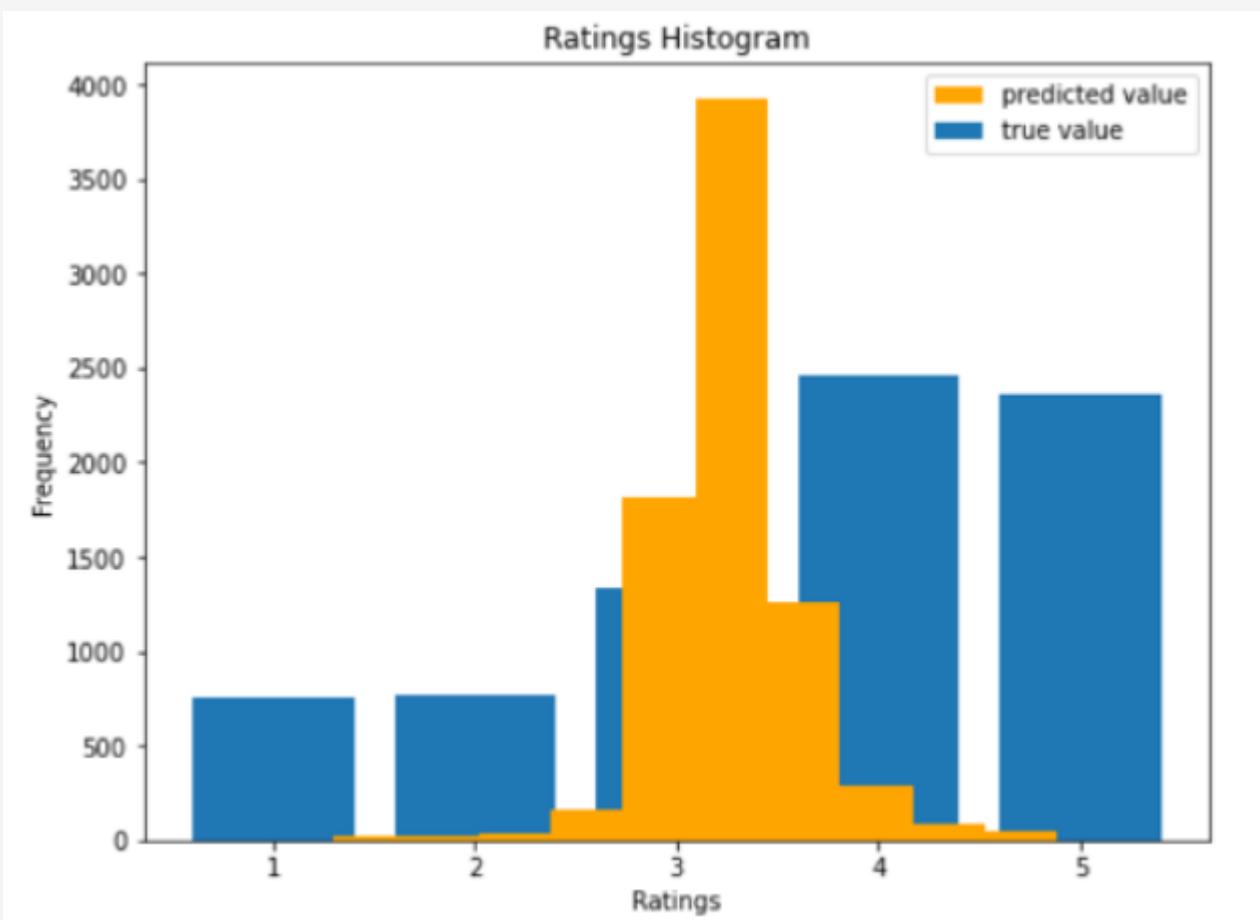
Embedding Normalization: Normalizes the embedding so that the dot product between two embeddings reflects cosine similarity.

Recommendation function: Create a function to calculate the cosine similarity of the target restaurant and other restaurants and return the ticket with the result.

Comparison to previous projects

How different is it compared to ours

1. Comparing the actual stars and predictions.



Comparison to previous projects

How different is it compared to ours

2. Calculates the cosine similarity between the target and the rest of the other restaurants

	similar_rest	cosine
0	Levetto	[0.11052720741649735]
1	Scaddabush Italian Kitchen & Bar	[0.025430804155849432]
2	Patria	[0.21501770082707605]
3	Faema Caffe	[-0.011985546689240609]
4	Faryab Kabab	[-0.04505609454747686]

	similar_rest	cos
17	Wvrst	1.000000
23	KaKa All You Can Eat	0.394581
76	Le Gourmand	0.378359
756	Wild Wing	0.374532
771	Sushi Mugen	0.360950
915	Subway	0.357526
928	Extreme Pita	0.356498
476	Greedy Duck & Noodles	0.354541
751	Elephant & Castle	0.350917
860	Adam's Halal Pub	0.348824

Learning Experience

Difficulties encountered and how you have solved them

- The datasize is big, but it took too long to upload and download.
- It is a dataset that needs to be converted from json to csv, but many errors occurred in the process.
There was a reference that it is convenient to use python 2 version, but sagemaker does not support it.
→ It was solved through modifications such as replacing the function, changing the mode, and changing the processing to fit the python 3 version.
- photo.json file does not open
→ We think it's a file that I don't need much, so we exclude it.
- review.json file also encountered memory error during conversion.
→ Use larger instance, error again. We reduced size.

Learning Experience

If you had more time or if you have to do the project all over again, what would you do or do differently or spend more time on.

- When the recommendation system was turned around, only the data set related to the Philadelphia city, which occupies the largest part of the data set, but it is regrettable that the recommendation system could have been constructed based on more cities if it had been given a little more time.
- It seems that I prepared in a hurry because the 12-week class and the schedule overlap or deviate. If you have more time, it would be nice to implement a recommendation system while feeling the strengths of SageMaker more.

Learning Experience

What you have learned doing the project (individually)

- DaHyeon : I was able to apply the content of the recommendation system learned in class to the project, enabling a deeper understanding beyond a simple grasp of the concepts. Moreover, as utilizing recommendation systems involves incorporating other models, it provided an opportunity to study additional concepts.
- DoYeon : It was a great experience to learn the recommendation system used in the content we always encounter in the major class and implement it with the dataset we wanted. At the same time, it was nice to review various concepts to incorporate the machine learning model.
- HyeWon : It was interesting to directly implement a recommendation system model that we can easily access in our daily lives and use a restaurant dataset called yelp, an actual dataset.
- TaeWan : I've always been deeply interested in recommendation systems using generative AI and GNNs. Implementing such a vast recommendation system has been an exhilarating experience, and I believe I can swiftly apply this to the service I am currently developing.

Teamwork Data

- LeeDaHyeon (25%) : Configuration Content-based Filtering
- KooDoYeon (25%) : Production of PPT and presentation
- YoonHyeWon (25%) : Configuration Collaborative Filtering
- ChoTaeWan (25%) : SageMaker setting and data preprocessing

References

- <https://www.yelp.com/dataset>
- <https://github.com/Yelp/dataset-examples>
- <https://brunch.co.kr/@mapthecity/32>
- Yelp-Review-Analysis/Group 21 - Yelp Review Analysis.ipynb at master · zoehuang7/Yelp-Review-Analysis · GitHub
- <https://www.kaggle.com/code/jagangupta/what-s-in-a-review-yelp-ratings-eda>
- Business Review Analysis from Yelp Dataset | Kaggle
- <https://velog.io/@dlskawns/%EC%B6%94%EC%B2%9C%EC%8B%9C%EC%8A%A4%ED%85%9C-%EC%9D%B4%EB%A1%A0-%EC%BB%A8%ED%85%90%EC%B8%A0-%EA%B8%B0%EB%B0%98%EC%B6%94%EC%B2%9CCB-%ED%98%91%EC%97%85%ED%95%84%ED%84%B0%EB%A7%81CF>
- <https://www.kaggle.com/code/zolboo/recommender-systems-knn-svd-nn-keras>
- <https://github.com/adarsh18raj/Yelp-Restaurant-Recommendation-System/tree/master>
- <https://github.com/karantyagi/Restaurant-Recommendations-with-Yelp>



Thank You :)