Urban Culture Through Taste

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

Datasets

1. Business
2. Review
3. Tip
4. User
5. Check-in

Dataset - Business

	business_id	name	address	city	state	postal_code	latitude	longitude	stars	review_count	is_open	attributes	categories	hours
0	Pns2l4eNsfO8kk83dixA6A	Abby Rappoport, LAC, CMQ	1616 Chapala St, Ste 2	Santa Barbara	CA	93101	34.426679	-119.711197	5.0	7	0	{'ByAppointmentOnly': 'True'}	Doctors, Traditional Chinese Medicine, Naturop...	None
1	mpf3x-BjTdTEA3yCZrAYPw	The UPS Store	87 Grasso Plaza Shopping Center	Affton	MO	63123	38.551126	-90.335695	3.0	15	1	{'BusinessAcceptsCreditCards': 'True'}	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	85711	32.223236	-110.880452	3.5	22	0	{'BikeParking': 'True', 'BusinessAcceptsCredit...	Department Stores, Shopping, Fashion, Home & G...	{'Monday': '8:0-22:0', 'Tuesday': '8:0-22:0', ...
													Restaurants,	{'Monday':

1. Business_id
2. Name
3. Address
4. City
5. State
6. Postal_code
7. Latitude

1. Longitude
2. Stars
3. Review_count
4. Is_open
5. Attributes
6. Categories
7. hours

Dataset - Review

df_reviews

	review_id	user_id	business_id	stars	useful	funny	cool	text	date
0	KU_O5udG6zpxOg-VcAEodg	mh_-eMZ6K5RLWhZyISBhwA	XQfwVwDr-v0ZS3_CbbE5Xw	3	0	0	0	If you decide to eat here, just be aware it is...	2018-07-07 22:09:11
1	BiTunyQ73aT9WBnpR9DZGw	OyoGAe7OKpv6SyGZT5g77Q	7ATYjTIgM3jUlt4UM3IypQ	5	1	0	1	I've taken a lot of spin classes over the year...	2012-01-03 15:28:18
2	saUsX_uimxRlCVr67Z4Jig	8g_iMtfSiwikVnbP2etR0A	YjUWPpI6HXG530lwP-fb2A	3	0	0	0	Family diner. Had the buffet. Eclectic assortm...	2014-02-05 20:30:30
3	AqPFMleE6RsU23_auESxiA	_7bHUi9Uuf5__HHc_Q8guQ	kxX2SOes4o-D3ZQBkiMRfA	5	1	0	1	Wow! Yummy, different, delicious. Our favo...	2015-01-04 00:01:03
4	Sx8TMOWLNuJBWer-0pcmoA	bcjbaE6dDog4jkNY91ncLQ	e4Vwtrqf-wpJfwesgvdgxQ	4	1	0	1	Cute interior and owner (?) gave us tour of up...	2017-01-14 20:54:15
...	...	...	...	...	...	...	...	...	...
6990275	H0RIamZu0B0Ei0P4aeh3sQ	qskILQ3k0I_qcCMI-k6_QQ	jals67o91gcrD4DC81Vk6w	5	1	2	1	Latest addition to services from ICCU is Apple...	2014-12-17 21:45:20
6990276	shTPgbgdwTHSuU67mGCmZQ	Zo0th2m8Ez4gLSbHftiQvg	2vLksaMmSEcGbjl5gywpZA	5	2	1	2	This spot offers a great, affordable east week...	2021-03-31 16:55:10
6990277	YNfNhgZlaaCO5Q_YJR4rEw	mm6E4FbCMwJmb7kPDZ5v2Q	R1khUUxidqfaJmcpmGd4aw	4	1	0	0	This Home Depot won me over when I needed to g...	2019-12-30 03:56:30
6990278	i-I4ZOhoX70Nw5H0FwrQUA	YwAMC-jvZ1fvEUum6QkEkw	Rr9kKArrMhSLVE9a53q-aA	5	1	0	0	For when I'm feeling ignoring my calorie-...	2022-01-19 18:59:27
6990279	RwcKOdEuLRHNJe4M9-qpqg	6JehEvdoCvZPJ_XIxnzlIw	VAeEXLbEcl9Emt9KGYq9aA	3	10	3	7	Located in the 'Walking District' in Nashville...	2018-01-02 22:50:47

6990280 rows × 9 columns

1. Review_id
2. User_id
3. Business_id
4. Stars
5. Useful
6. Funny
7. Cool
8. Text
9. date

Dataset - Tip

```
df_tip
```

	user_id	business_id	text	date	compliment_count
0	AGNUgVwnZUey3gcPCJ76iw	3uLgwr0qeCNMjKenHJwPGQ	Avengers time with the ladies.	2012-05-18 02:17:21	0
1	NBN4MgHP9D3cw--SnauTkA	QoezRbYQncpRqyrLH6Iqjg	They have lots of good deserts and tasty cuban...	2013-02-05 18:35:10	0
2	-copOvldyKh1qr-vzkDEvw	MYoRNLb5chwjQe3c_k37Gg	It's open even when you think it isn't	2013-08-18 00:56:08	0
3	FjMQVZjSqY8syIO-53KFKw	hV-bABTK-glh5wj31ps_Jw	Very decent fried chicken	2017-06-27 23:05:38	0
4	Id0AperBXk1h6UbqmM80zw	_uN0OudeJ3Zl_tf6nxg5ww	Appetizers.. platter special for lunch	2012-10-06 19:43:09	0
...	...	...	...	...	...
908910	eYodOTF8pkqKPzHkcxZs-Q	3lHTewuKFt5IImbXJoFeDQ	Disappointed in one of your managers.	2021-09-11 19:18:57	0
908911	1uxtQAuJ2T5Xwa_wp7kUnA	OaGf0Dp56ARhQwIDT90w_g	Great food and service.	2021-10-30 11:54:36	0
908912	v48Spe6WEpqehsF2xQADpg	hYnMeAO77RGyTtlzUSKYzQ	Love their Cubans!!	2021-11-05 13:18:56	0
908913	ckqKGM2hl7l9Chp5IpAhkw	s2eyoTuJrcP7l_XyjdhUHQ	Great pizza great price	2021-11-20 16:11:44	0
908914	4tF1CWdMxvvwpUIgGsDygA	_cb1Vg1NlWry8UA0jyuXnQ	Food is good value but a bit hot!	2021-12-07 22:30:00	0

908915 rows × 5 columns

1. User_id
2. Business_id
3. Text
4. Date
5. compliment_count

Dataset - User

```
In [32]: df_user
```

	user_id	name	review_count	yelping_since	useful	funny	cool	elite
0	qVc8ODYU5SZjKXVBgXdI7w	Walker	585	2007-01-25 16:47:26	7217	1259	5994	2007
1	j14WgRoU_-2ZE1aw1dXrJg	Daniel	4333	2009-01-25 04:35:42	43091	13066	27281	2009,2010,2011,2012,2013,2014,2015,2016,2017,2...
2	2WnXYQFK0hXEoTxPtV2zvg	Steph	665	2008-07-25 10:41:00	2086	1010	1003	2009,2010,2011,2012,2013
3	SZDeASXq7o05mMNLshsdIA	Gwen	224	2005-11-29 04:38:33	512	330	299	2009,2010,2011
4	hA5lMy-EnncsH4JoR-hFGQ	Karen	79	2007-01-05 19:40:59	29	15	7	
...	...	...	...	...	...	...	...	...
1987892	fB3jbHi3m0L2KgGOxBv6uw	Jerrold	23	2015-01-06 00:31:31	7	0	0	

```
df_user.loc[:,'friends' : 'compliment_cute']
```

	friends	fans	average_stars	compliment_hot	compliment_more	compliment_profile	compliment_cute
0	NSCy54eWehBJyZdG2iE84w, pe42u7DcCH2Qml81NX-8qA...	267	3.91	250	65	55	56
1	ueRPE0CX75ePGMqOFVj6lQ, 52oH4DrRvzzl8wh5UXyU0A...	3138	3.74	1145	264	184	157
2	LuO3Bn4f3rlhyHIaNfTlnA, j9B4XdHUhDfTKVecyWQgyA...	52	3.32	89	13	10	17
3	enx1vVPnfdNUdPho6PH_wg, 4wOcvMLtU6a9Lslggq74Vg...	28	4.27	24	4	1	6
4	PBK4q9KEEBHhFvSXCUirlw, 3FWPpM7KU1gXeOM_ZbYMbA...	1	3.54	1	1	0	0
...	...	...	...	...	...	...	...
1987892	None	0	4.92	0	0	0	0
1987893	None	0	5.00	0	0	0	0
1987894	None	0	2.00	0	0	0	0
1987895	None	0	3.00	0	0	0	0
1987896	None	0	5.00	0	0	0	0

1987897 rows × 7 columns

1. 'User_id'
2. 'name'
3. 'Review_count'
4. 'Yelping_since'
5. 'Useful'
6. 'Funny'
7. 'Cool'
8. 'Elite'
9. 'Friends'
10. 'Fans'
11. 'Average_stars'
12. 'Compliment_hot'
13. 'Compliment_more'
14. 'Compliment_profile'
15. 'Compliment_cute'

Dataset - User

	compliment_list	compliment_note	compliment_plain	compliment_cool	compliment_funny	compliment_writer	compliment_photos
0	18	232	844	467	467	239	180
1	251	1847	7054	3131	3131	1521	1946
2	3	66	96	119	119	35	18
3	2	12	16	26	26	10	9
4	0	1	1	0	0	0	0
...	...	...	...	...	...	...	...
1987892	0	0	0	0	0	0	0
1987893	0	0	0	0	0	0	0
1987894	0	0	0	0	0	0	0
1987895	0	0	0	0	0	0	0
1987896	0	0	0	0	0	0	0

1987897 rows × 7 columns

1. Compliment_list
2. Compliment_note
3. Compliment_plain
4. Compliment_cool
5. Compliment_funny
6. Compliment_writer
7. compliment_photos

Dataset - Check-in

	df_checkin	

	business_id	date
0	---kPU91CF4Lq2-WlRu9Lw	2020-03-13 21:10:56, 2020-06-02 22:18:06, 2020...
1	--0iUa4sNDFiZFrAdIWhZQ	2010-09-13 21:43:09, 2011-05-04 23:08:15, 2011...
2	--30_8IhuyMHbSOcNWd6DQ	2013-06-14 23:29:17, 2014-08-13 23:20:22
3	--7PUidqRWpRSpXebiyxTg	2011-02-15 17:12:00, 2011-07-28 02:46:10, 2012...
4	--7jw19RH9JKXgFohspgQw	2014-04-21 20:42:11, 2014-04-28 21:04:46, 2014...
...	...	...
131925	zznJox6-nmXlGYNWgTDwQQ	2013-03-23 16:22:47, 2013-04-07 02:03:12, 2013...
131926	zznZqH9CiAznbkV6fXyHWA	2021-06-12 01:16:12
131927	zzu6_r3DxBJuXcjnOYVdTw	2011-05-24 01:35:13, 2012-01-01 23:44:33, 2012...
131928	zzw66H6hVjXQEt0Js3Mo4A	2016-12-03 23:33:26, 2018-12-02 19:08:45
131929	zzyx5x0Z7xXWWvWnZFuxlQ	2015-01-06 17:51:53

131930 rows × 2 columns

1. Business_id
2. Date

Project Scope

```
df_reviews.text[25]
```

'Went for lunch. Beef brisket sandwich was awesome.   So juicy and tender.   Pulled pork was was just as good!'

- Review - 'text' column
- Food entity - Beef brisket sandwich, pulled pork
- Adjective - juicy, tender, awesome, good
- Time entity - lunch

```
df_reviews.text[50]
```

'We started with the artichoke dip, big mistake. It was cold and the chips were stale. Mentioned to our waitress, but nothing came of it. We ordered a large pizza, thin crust which after adding toppings it was a bit pricey for pizza. Pizza was ok, not great. Except for not being our advocate our server was good. I have been here before for Cardinals game, it was better then. Will try another venue next time.'

- Food entity - artichoke dip, chips, pizza
- Adjective - cold, stale, thin, large, pricey, ok, not great

Project Scope

```
df_reviews.text[45]
```

'Comfortable bed, good breakfast, fast internet, great location, and low price. Overall I was really impressed with this place. However, if you stay on the top floors you may be waiting up to 15 minutes for an elevator in the morning.'

```
df_reviews.text[65]
```

'I stopped here because it was by the hotel I was staying in. The guys working at the shop were super nice and friendly. I really like the buy one get one deal they do. I also ordered some hot wings and they were pretty good. They had a ton of sauce on them which is awesome. Next time I come back to town I will be sure to stop here for some more pizza!'
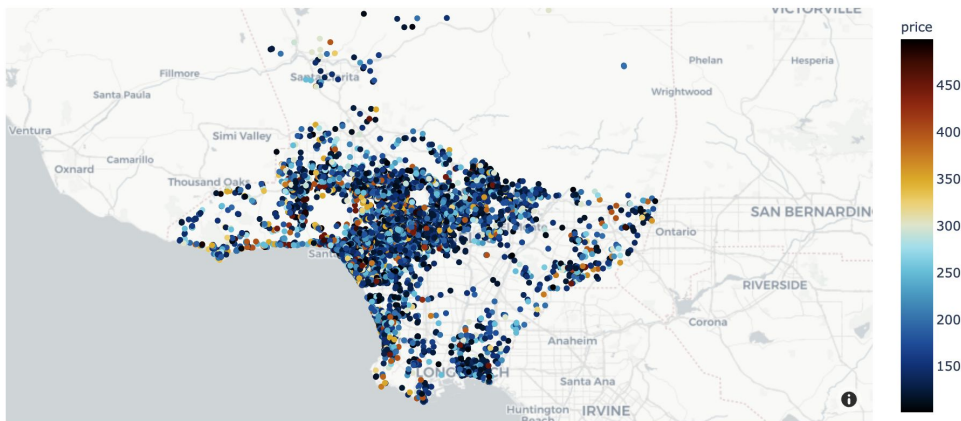
- Not a restaurant
- Ignore reviews like the first one
- Consider 2nd review
- Categorize reviews if you find a food entity in them

Part 1

- Tokenization
- Stemming & Lemmatization
- POS tagging
  - Adjective - taste, sentiment, texture
- Entity Extraction
  - Cuisine
  - Food
  - Price
- Plotting
  - Plot maps regionwise

Part 1

```python
# Go to mapbox and get your own token
# Alan's Mapbox token (reference only): pk.eyJ1IjoiYWxhbnR6dGFiiwiYSI6ImNqa2VqdEFnejB1dTEzcHBxczBhaGZhZz2gifQ.moVxmxTkE
px.set_mapbox_access_token ='pk.eyJ1IjoiYWxhbnR6dGFiiwiYSI6ImNqa2VqdEFnejB1dTEzcHBxczBhaGZhZz2gifQ.moVxmxTkE53tZDfBFr3u'
fig = px.scatter_mapbox(data_frame=df,
                        lat="latitude",
                        lon="longitude",
                        color="price",
                        color_continuous_scale=px.colors.cyclical.IceFire,
                        #center = {"lat": 34.10, "lon": -118.25}
                        )
fig.update_layout(mapbox_style="carto-positron",\
                  mapbox_zoom=8, mapbox_center = {"lat": 34.10, "lon": -118.25})
```


Plot maps based on,

- Cuisine
  - Eg. Plot all the places where you can get chinese food. This can include fusion restaurants as well. Color - rating
- Taste adjective
  - Plot places where you can get crispy food. Includes fried chicken, Vada, Tacos, etc. Color - cuisine type
- Price
  - Eg. Plot places where you can get affordable crispy food.

Part 1 - Schedule

WBS NUMBER	TASK TITLE	TASK OWNER	START DATE	DUE DATE	DURATION	PCT OF TASK COMPLETE
1	Project Conception and Initiation					
2	Data preprocessing					
2.1	Tokenization	G1 - Grp of 1	5/16/22	5/18/22	3	0%
2.2	Stemming & Lemmatization	G1 - Grp of 1	5/18/22	5/19/22	2	0%
2.3	Internal meeting 4	Group	5/19/22	5/19/22	1	0%
2.4	POS tagging - adjectives, adverbs	G2 - Grp of 2	5/20/22	5/21/22	2	0%
2.5	Classify adj, adv - taste, texture, sentiment	G2 - Grp of 2	5/22/22	5/23/22	2	0%
2.6	Entity extraction - cuisine, food, price	G3 - Grp of 2	5/20/22	5/21/22	2	0%
2.7	Merge data - entities, adj, adv, rating, location, lat, long	G3 - Grp of 2	5/22/22	5/23/22	2	0%
2.8	Internal meeting 5	Group	5/23/22	5/23/22	1	0%
4	Visualize Data					
4.1	Plot geography - cuisine, taste, price	G2 - Grp of 2	5/29/22	5/31/22	3	0%
4.2	Urban insights from the plot	G2 - Grp of 2	5/31/22	6/1/22	2	0%

Part 2

- Text similarity
    - soft , tender, buttery - all the same
    - Text classification
    - Generate cuisine similarity score
- Plotting
    - Generate heat map based on cuisine similarity

Part 2

Food entity

- List all the foods that are soupy.
- Text classification (AutoGluon)
- Includes ramen, consomme, pani puri, rasam, etc.
- Get the cuisine type of the extracted foods.
- Food similarity score

Part 2 - Heat Map

	Korean	South Indian	Chinese	Malaysian	Italian	Thai	Pakistani	North Indian
Korean	100%	50%	70%	40%	20%	40%	10%	20%
South Indian	50%	100%	50%	60%	5%	30%	40%	50%
Chinese	70%	50%	100%	35%	10%	45%	20%	25%
Malaysian	40%	60%	35%	100%	25%	75%	40%	55%
Italian	30%	5%	10%	25%	100%	20%	25%	30%
Thai	40%	30%	45%	75%	20%	100%	30%	40%
Pakistani	10%	40%	20%	40%	25%	30%	100%	70%
North Indian	20%	50%	25%	55%	30%	40%	70%	100%

Part 2 - Schedule

WBS NUMBER	TASK TITLE	TASK OWNER	START DATE	DUE DATE	DURATION	PCT OF TASK COMPLETE
1	Project Conception and Initiation					
3	Generate ML model for text similarity score					
3.1	Model 1	Aniketh	5/23/22	5/28/22	6	0%
3.2	Model 2	Swetha	5/23/22	5/28/22	6	0%
3.3	Model 3	Summer	5/23/22	5/28/22	6	0%
3.4	Model 4	Sejoon	5/23/22	5/28/22	6	0%
3.5	Model 5	Shilpi	5/23/22	5/28/22	6	0%
3.6	Internal meeting 6	Group	5/28/22	5/28/22	1	0%
4.3	Heat map based on similarity score	G2 - Grp of 2	5/29/22	5/31/22	3	0%
4.4	Urban insights from heat map	G2 - Grp of 2	5/31/22	6/1/22		0%

Project schedule

- [Schedule](#)

Small chunks

```
(base) swetha@Swethas-MBP yelp_dataset % ll
zsh: command not found: ll
(base) swetha@Swethas-MBP yelp_dataset % ls
Dataset_User_Agreement.pdf              yelp_academic_dataset_checkin.json      yelp_academic_dataset_tip.json
yelp_academic_dataset_business.json     yelp_academic_dataset_review.json       yelp_academic_dataset_user.json
(base) swetha@Swethas-MBP yelp_dataset % ls -al
total 18185216
drwx------@  8 swetha  staff         256 May 11 18:32 .
drwx------@ 68 swetha  staff        2176 May 15 21:34 ..
-rw-r--r--@  1 swetha  staff       80358 Feb 15 14:03 Dataset_User_Agreement.pdf
-rw-r--r--@  1 swetha  staff   118863795 Jan 19 14:35 yelp_academic_dataset_business.json
-rw-r--r--@  1 swetha  staff   286958945 Jan 19 14:39 yelp_academic_dataset_checkin.json
-rw-r--r--@  1 swetha  staff  5341868833 Jan 19 14:51 yelp_academic_dataset_review.json
-rw-r--r--@  1 swetha  staff   180604475 Jan 19 14:40 yelp_academic_dataset_tip.json
-rw-r--r--@  1 swetha  staff  3363329011 Jan 19 14:39 yelp_academic_dataset_user.json
(base) swetha@Swethas-MBP yelp_dataset % less yelp_academic_dataset_review.json
(base) swetha@Swethas-MBP yelp_dataset % wc yelp_academic_dataset_review.json
 6990280 735427116 5341868833 yelp_academic_dataset_review.json
(base) swetha@Swethas-MBP yelp_dataset %
(base) swetha@Swethas-MBP yelp_dataset % wc -l yelp_academic_dataset_review.json
 6990280 yelp_academic_dataset_review.json
(base) swetha@Swethas-MBP yelp_dataset % less yelp_academic_dataset_review.json
(base) swetha@Swethas-MBP yelp_dataset % less yelp_academic_dataset_business.json
(base) swetha@Swethas-MBP yelp_dataset % less yelp_academic_dataset_business.json
(base) swetha@Swethas-MBP yelp_dataset % vi yelp_academic_dataset_business.json
(base) swetha@Swethas-MBP yelp_dataset % vi businesss.json
(base) swetha@Swethas-MBP yelp_dataset % vi businesss.json
(base) swetha@Swethas-MBP yelp_dataset % vi businesss.json
(base) swetha@Swethas-MBP yelp_dataset % less y
(base) swetha@Swethas-MBP yelp_dataset % vi yelp_academic_dataset_review.json
vi r
(base) swetha@Swethas-MBP yelp_dataset % less reviews.json
(base) swetha@Swethas-MBP yelp_dataset % less yelp_academic_dataset_review.json
```

```
(base) swetha@Swethas-MBP yelp_dataset % vi reviews.json
```

References

1. [The Geography of Taste: Using Yelp to Study Urban Culture](#)

2. [Yelp-review-text-mining](#)