Big Data Analytics Symposium 2022

Yelp Data Analysis - NYC Eating Team

PRESENTED BY Junhua Liang, Virginia Wu, Xin Lu, Wenhao Li

05/05/2022

Yelp Data Analysis

Abstract

Feast is important for everyone in the world, especially in this era as the development of the society's productivity. People who travel to a new city usually want to have a taste of the best restaurant in this area, and those who want to repair their house's roof would also check the rating of the store. For merchant, they also want to open their restaurants in the best location in the city. Thus, it's important to have related analytics to help them. One direction to have has analytics is using people's reviews and their check-in information to get the pattern of consuming in the area and build related models to help recommend restaurants for people and find location for merchant. In this project, we will use Yelp's open datasets to visualize specific information and build model to for recommendations and rating predictions.

Platform(s) where the application runs: NYU Peel Cluster

Yelp Data Analysis

Motivation

Who are the users of this analytics?

People who wrote their reviews and provided feedback to the stores they went to will be the users of this analytics.

Who will benefit from this analytics?

People who need to go to a high rate restaurant.

People who use Yelp app, the system will provide the best recommended stores for them.

People who go to explore a new city and want to have a taste of the best stores.

Why is this analytics so important?

1. The analytics can retrieve people's consuming habits, and recommend the most perfect one for them.
2. The analytics can obtain consuming distribution in each city, and business man could open stores in the good location.

Yelp Data Analysis

Goodness

What steps were taken to access the ‘goodness’ of the analytics?

1. For the recommendation systems, our team verified the recommended items, compared it to the user’s past consuming items. If the recommended one is meaningful, then we could state that our analytics is good.
2. For the data visualization and the pattern we retrieve from it, our team will regard it as a good result if the pattern match normal people’s thinking.

Data Sources

Data Sources

Dataset name	Description	Size of Data
yelp_academic_dataset_user.json	Users' information, including user_id and metadata.	4.34 GB
yelp_academic_dataset_business.json	Restaurant's information, including business_id and metadata.	118.9 MB
yelp_academic_dataset_review.json	Review information on specific user and restaurant.	5.34 GB
yelp_academic_dataset_checkin.json	Restaurant's business ID and their check-in date.	287 MB
yelp_academic_dataset_tip.json	Tip's information on specific user and restaurant.	180.6 MB

Yelp Data Analysis

Data Sample: yelp_academic_dataset_user

average_stars	compliment_cool	compliment_cute	compliment_funny	compliment_hot	compliment_list	compliment_more	compliment_note	compliment_photos	compliment_plain	compliment_profile	compliment_writer	cool	elite	fans	friends	funny	name	review_count	useful	user_id	yelping_since
3.91	467	56	467	250	18	65	232	180	844	55	239	5994	2007	267	NSCy54eWehBJyZdG2...	1259	Walker	585	7217	qVz8ODYU5SZjKXVBg...	2007-01-25 16:47:26
3.74	3131	157	3131	1145	251	264	1847	1946	7054	184	1521	27281	2009,2010,2011,20...	3138	ueRPE0CX75ePGMqOF...	13066	Daniel	4333	43091	j14WgRoU_-2ZE1aw1...	2009-01-25 04:35:42
3.32	119	17	119	89	3	13	66	18	96	10	35	1003	2009,2010,2011,20...	52	LuO3Bn4f3rlhyHIaN...	1010	Steph	665	2086	2WnXYQFK0hXEoTxPt...	2008-07-25 10:41:00
4.27	26	6	26	24	2	4	12	9	16	1	10	299	2009,2010,2011	28	enx1vVPnfdNUdPho6...	330	Gwen	224	512	SZDeASXq7o05mMNLs...	2005-11-29 04:38:33
3.54	0	0	0	1	0	1	1	0	1	0	0	7		1	PBK4q9KEEBHhFvSXC...	15	Karen	79	29	hA5lMy-EnncsH4JoR...	2007-01-05 19:40:59
3.85	2543	361	2543	1713	147	163	1212	323	5696	191	815	11211	2006,2007,2008,20...	1357	xBDpTUbai0DXrvxCe...	9940	Jane	1221	14953	q_QQ5kBBwlCcbL1s4...	2005-03-14 20:26:35
2.75	0	0	0	0	0	0	0	0	1	0	0	0		1	HDAQ74AEznP-YsMk1...	1	Rob	12	6	cxuxXkcihfCbqt5By...	2009-02-24 03:09:06
3.73	12	0	12	4	0	7	8	0	6	2	5	143		23	y2GyxJF5VQWohxgw_...	102	Mike	358	399	E9kcWJdJUHuTKfQur...	2008-12-11 22:11:56
4.04	5	3	5	2	0	0	3	1	4	0	3	46		7	tOQDlz36rI__SOsbL...	40	Rachelle	40	109	lO1iq-f75hnPNZkTy...	2008-12-29 22:40:56
3.4	3	0	3	0	0	0	1	0	6	0	0	23		4	gy5fWeSv3Gamuq9Ox...	20	John	109	154	AUi8MPWJ0mLkMfwbu...	2010-01-07 18:32:04
4.0	0	0	0	0	0	0	0	0	0	0	0	1		1	Vq4Pc81l6MWTnc-h4...	0	Chris	4	1	iYzhPPqnrjJkg1JHZ...	2010-11-03 18:59:20
3.89	36	3	36	23	5	9	31	7	41	1	24	573	2009,2010,2011,2012	31	6tbXpUIU6upoeqWND...	487	Ryan	535	1130	xoZvMJPDW6Q9pDAXI...	2009-05-27 06:12:10
4.51	0	0	0	4	0	0	1	0	5	0	1	27		4	zkK6c9BcDyqreU0fq...	3	Charlene	37	63	vVukUtqoLF5BvH_Vt...	2011-01-29 17:18:59
3.08	0	0	0	0	0	0	0	0	0	0	0	0		0	f0mdrhyxVZ0TsFZD5...	3	Kenny	11	30	_crIokUeTCHVK_JVO...	2009-10-07 17:23:44
4.29	2	0	2	1	0	0	1	0	0	0	0	13		1	piejMEdRkGB7-1aL4...	3	Teresa	7	18	1McG5Rn_UDkmlkZOr...	2009-05-26 16:11:11
3.75	221	22	221	212	34	17	67	17	158	17	45	1297	2007,2008,2009,20...	44	j3MBGSLaXMlhLZNeA...	1138	Eugene	682	1819	SgiBkhXeqIKl1PlFp...	2006-08-25 16:47:25
4.15	2	0	2	0	0	1	6	0	2	0	1	19		1	hJiJzw6obCmbGAfwr...	2	Jennifer	25	29	fJZO_skqpnhk1kvom...	2008-07-14 16:01:36
3.84	66	0	66	15	0	1	12	12	33	0	1	29		9	EPBLDry-ObheloH-N...	29	Ronskee	37	56	x7YtLnBW2dUnrrpwa...	2010-05-06 00:40:56
4.11	808	29	808	1020	23	79	144	723	407	68	587	4149	2010,2011,2012,20...	131	dLts9bY66tXEFqYG0...	3714	Catherine	607	4573	QF1Kuhs8iwLWANNZx...	2009-04-27 20:25:54
3.6	0	0	0	1	0	0	2	0	4	0	1	44		4	Ynu2Z2L8Wv2fbTwQ_...	52	AJ	133	201	VcLRGCG_VbAo8MxOm...	2009-07-11 16:47:38

Yelp Data Analysis

Data Sample: yelp_academic_dataset_business

address	attributes	business_id	categories	city	hours	is_open	latitude	longitude	name	postal_code	review_count	stars	state
1616 Chapala St, ...	[,,,,,,,,,,, True...	Pns2l4eNsfO8kk83d...	Doctors, Traditio...	Santa Barbara	null	0	34.4266787	-119.7111968	Abby Rappoport, L...	93101	7	5.0	CA
87 Grasso Plaza S...	[,,,,,,,,, True,,...	mpf3x-BjTdTEA3yCZ...	Shipping Centers,...	Affton	[8:0-18:30, 0:0-0...	1	38.551126	-90.335695	The UPS Store	63123	15	3.0	MO
5255 E Broadway Blvd	[,,,,,,, True,, T...	tUFrWirKiKi_TAnsV...	Department Stores...	Tucson	[8:0-23:0, 8:0-22...	0	32.223236	-110.880452	Target	85711	22	3.5	AZ
935 Race St	[,, u'none',,,,, ...	MTSW4McQd7CbVtyjq...	Restaurants, Food...	Philadelphia	[7:0-21:0, 7:0-20...	1	39.9555052	-75.1555641	St Honore Pastries	19107	80	4.0	PA
101 Walnut St	[,,,,,,, True,, T...	mWMc6_wTdE0EUBKIG...	Brewpubs, Breweri...	Green Lane	[12:0-22:0,, 12:0...	1	40.3381827	-75.4716585	Perkiomen Valley ...	18054	13	4.5	PA
615 S Main St	[,, u'none', None...	CF33F8-E6oudUQ46H...	Burgers, Fast Foo...	Ashland City	[9:0-0:0, 0:0-0:0...	1	36.269593	-87.058943	Sonic Drive-In	37015	6	2.0	TN
8522 Eager Road, ...	[,,,,,,, True,, T...	n_0UpQx1hsNbnPUSl...	Sporting Goods, F...	Brentwood	[10:0-18:0, 0:0-0...	1	38.627695	-90.340465	Famous Footwear	63144	13	2.5	MO
400 Pasadena Ave S	null	qkRM_2X51Yqxk3btl...	Synagogues, Relig...	St. Petersburg	[9:0-17:0, 9:0-17...	1	27.76659	-82.732983	Temple Beth-El	33707	5	3.5	FL
8025 Mackenzie Rd	[,, u'full_bar', ...	k0hlBqXX-Bt0vf1op...	Pubs, Restaurants...	Affton	null	0	38.5651648	-90.3210868	Tsevi's Pub And G...	63123	19	3.0	MO
2312 Dickerson Pike	[,, u'none',,,,,,,...	bBDDEgkFA1Otx9Lfe...	Ice Cream & Froze...	Nashville	[6:0-16:0, 0:0-0:...	1	36.2081024	-86.7681696	Sonic Drive-In	37207	10	1.5	TN
21705 Village Lak...	[,,,,,,, True,, T...	UJsufbvfyfONHeWdv...	Department Stores...	Land O' Lakes	[9:30-21:30, 9:30...	1	28.1904587953	-82.4573802199	Marshalls	34639	6	3.5	FL
	[,, 'none', {'tou...	eEOYSgkmpB90uNA7l...	Vietnamese, Food,...	Tampa Bay	[11:0-14:0, 11:0-...	1	27.9552692	-82.4563199	Vietnamese Food T...	33602	10	4.0	FL
8901 US 31 S	[,, 'none', {'tou...	il_Ro8jwPlHresjw9...	American (Traditi...	Indianapolis	[6:0-22:0, 6:0-22...	1	39.6371332838	-86.127217412	Denny's	46227	28	2.5	IN
15 N Missouri Ave	[,,,,,,,,,,, True...	jaxMSoInw8Poo3XeM...	General Dentistry...	Clearwater	[, 7:30-15:30,,, ...	1	27.966235	-82.787412	Adams Dental	33755	10	5.0	FL
2575 E Bay Dr	[,, u'none', {'ro...	0bPLkL0QhhPO5kt1_...	Food, Delis, Ital...	Largo	[10:0-20:0, 10:0-...	0	27.9161159	-82.7604608	Zio's Italian Market	33771	100	4.5	FL
205 Race St	[,, 'full_bar', {...	MUTTqe8uqyMdBl186...	Sushi Bars, Resta...	Philadelphia	[13:30-23:0,, 13:...	1	39.953949	-75.1432262	Tuna Bar	19106	245	4.0	PA
625 N Stone Ave	[,,,,,,,,, True,,...	rBmpy_Y1UbBx8ggHl...	Automotive, Auto ...	Tucson	[8:0-17:0, 0:0-0:...	1	32.2298719	-110.9723419	Arizona Truck Out...	85705	10	4.5	AZ
712 Adams St	[,,,,,,, True,, T...	M0XSSHqrASOnhgbWD...	Vape Shops, Tobac...	New Orleans	[10:0-19:0, 10:0-...	1	29.9414679565	-90.129952757	Herb Import Co	70118	5	4.0	LA
1241 Airline Dr	null	8wGISYjYkE2tSqn3c...	Automotive, Car R...	Kenner	[8:0-17:0, 8:0-17...	1	29.981183	-90.2540123	Nifty Car Rental	70062	14	3.5	LA
1224 South St	[,, u'none', {'to...	ROeacJQwBeh05Rqg7...	Korean, Restaurants	Philadelphia	[11:30-20:30, 11:...	1	39.943223	-75.162568	BAP	19147	205	4.5	PA

NYU

Yelp Data Analysis

Data Sample: yelp_academic_dataset_tip

business_id	compliment_count	date	text	user_id
3uLgwr0qeCNMjKenH...	0	2012-05-18 02:17:21	Avengers time wit...	AGNUgVwnZUey3gcPC...
QoezRbYQncpRqyrLH...	0	2013-02-05 18:35:10	They have lots of...	NBN4MgHP9D3cw--Sn...
MYoRNLb5chwjQe3c_...	0	2013-08-18 00:56:08	It's open even wh...	-copOvldyKh1qr-vz...
hV-bABTK-glh5wj31...	0	2017-06-27 23:05:38	Very decent fried...	FjMQVZjSqY8syIO-5...
_uN0OudeJ3Zl_tf6n...	0	2012-10-06 19:43:09	Appetizers.. plat...	ld0AperBXk1h6Ubqm...
7Rm9Ba50bw23KTA8R...	0	2012-03-13 04:00:52	Chili Cup + Singl...	trf3Qcz8qvCDKXiTg...
kH-0iXqkL7b8UXNpg...	0	2013-12-03 23:42:15	Saturday, Dec 7th...	SMGAlRjyfuYu-c-22...
jtri188kuhe_AuEOJ...	0	2016-11-22 22:14:58	This is probably ...	YVBB9g23nuVJ0u44z...
xODBZmX4EmlVvbqtK...	0	2012-07-27 01:48:24	Tacos	VL12EhEdT4OWqGq0n...
pICJRcyqW1cF96Q3X...	0	2012-06-09 22:57:04	Starbucks substit...	4ay-fdVks5WMerYL_...
clwjLY7PdYJpe7IP9...	0	2014-06-17 01:20:14	Order the Tortill...	OttfcRxgRrYsTg9EV...
wLHodvVFLTgK3nl2X...	0	2017-03-23 22:01:41	Very good will de...	JsXhBw6MntzTJjH_U...
wUMuvdUeVZODZk7Tj...	0	2013-02-28 02:05:54	If the Hotlight i...	Y0JfJh4B-jrtGc_AH...
MDr7KLYSPkEonvGoj...	0	2011-07-20 21:52:57	Let's go Yankees!	MlnuJ7T14CE0JDK2Z...
aK6R2akvIK9ijw3Fv...	0	2014-06-12 17:34:20	Basically same fo...	ffWWVlmsrN5lZ6sjA...
EXYbKA1tocvOK_1tX...	0	2011-10-13 03:15:15	Don't go for dinn...	j2sEA3hiUcwHfq9Ml...
H9fkf4Xkj_j7Zxs1F...	0	2012-03-11 23:16:12	30 mins for take ...	jsaN4TDygu76AGTiB...
ReX09lhufLTAx19kr...	0	2013-06-10 20:18:41	Got the grilled c...	kjFgyrCvmVVGSlgWz...
c5nLy7YgXG-IIrOmq...	0	2016-04-23 02:44:03	This is the bomb ...	I6aRZ4sE7ixv0_2r3...
LJaR65ALpz261_dlV...	0	2012-06-02 14:39:28	Helping Mona find...	Ll5l4WTKPH7zWQWA6...

NYU

Yelp Data Analysis

Data Sample: yelp_academic_dataset_review

business_id	cool	date	funny	review_id	stars	text	useful	user_id
XQfwVwDr-v0ZS3_Cb...	0	2018-07-07 22:09:11	0	KU_O5udG6zpxOg-Vc...	3.0	If you decide to ...	0	mh_-eMZ6K5RLWhZyI...
7ATYjTIgM3jUlt4UM...	1	2012-01-03 15:28:18	0	BiTunyQ73aT9WBnpR...	5.0	I've taken a lot ...	1	OyoGAe7OKpv6SyGZT...
YjUWPpI6HXG530lwP...	0	2014-02-05 20:30:30	0	saUsX_uimxRlCVr67...	3.0	Family diner. Had...	0	8g_iMtfSiwikVnbP2...
kxX2SOes4o-D3ZQBk...	1	2015-01-04 00:01:03	0	AqPFMleE6RsU23_au...	5.0	Wow! Yummy, diff...	1	_7bHUi9Uuf5__HHc_...
e4Vwtrqf-wpJfwesg...	1	2017-01-14 20:54:15	0	Sx8TMOWLNuJBWer-0...	4.0	Cute interior and...	1	bcjbaE6dDog4jkNY9...
04UD14gamNjLY0IDY...	1	2015-09-23 23:10:31	2	JrIxlS1TzJ-iCu79u...	1.0	I am a long term ...	1	eUta8W_HdHMXPzLBB...
gmjsEdUsKpj9Xxu6p...	0	2015-01-03 23:21:18	2	6AxgBCNX_PNTOxmbR...	5.0	Loved this tour! ...	0	r3zeYsv1XFBRA4dJp...
LHSTtnW3YHCeUkRDG...	0	2015-08-07 02:29:16	0	_ZeMknuYdlQcUqng_...	5.0	Amazingly amazing...	2	yfFzsLmaWF2d4Sr0U...
B5XSoSG3SfvQGtKEG...	0	2016-03-30 22:46:33	1	ZKvDG2sBvHVdF5oBN...	3.0	This easter inste...	1	wSTuiTk-sKNdcFypr...
gebiRewfieSdtt17P...	0	2016-07-25 07:31:06	0	pUycOfUwM8vqX7KjR...	3.0	Had a party of 6 ...	0	59MxRhNVhU9MYndMk...
uMvVYRgGNXf5boolA...	0	2015-06-21 14:48:06	0	rGQRf8UafX7OTlMNN...	5.0	My experience wit...	2	1WHRWwQmZOZDAhp2Q...
EQ-TZ2eeD_E0BHuvo...	0	2015-08-19 14:31:45	0	l3Wk_mvAog6XANIuG...	4.0	Locals recommende...	0	ZbqSHbgCjzVAqaa7N...
lj-E32x9_FA7GmUrB...	0	2014-06-27 22:44:01	0	XW_LfMv0fV21l9c6x...	4.0	Love going here f...	0	9OAtfnWag-ajVxRbU...
RZtGWDLCAtuipwaZ-...	0	2009-10-14 19:57:14	0	8JFGBuHMoiNDyfcxu...	4.0	Good food--loved ...	0	smOvOajNG0lS4Pq7d...
otQS34_MymijPTdNB...	0	2011-10-27 17:12:05	2	UBp0zWyH60Hmw6Fsa...	4.0	The bun makes the...	0	4Uh27DgGzsp6PqrH9...
BVndHaLihEYbr76Z0...	0	2014-10-11 16:22:06	0	OAhBYw8IQ6wlfw1ow...	5.0	Great place for b...	0	1C2lxzUo1Hyye4RFI...
YtSqYv1Q_pOltsVPS...	0	2013-06-24 11:21:25	0	oyaMhzBSwfGgemSGu...	5.0	Tremendous servic...	0	Dd1jQj7S-BFGqRbAp...
rBdG_23USc7DletfZ...	0	2014-08-10 19:41:43	0	LnGZB0fjfgeVDVz5I...	4.0	The hubby and I h...	1	j2wlzrntrbKwyOcOi...
CLEWowfkj-wKYJlQD...	1	2016-03-07 00:02:18	0	u2vzZaOqJ2feRshaa...	5.0	I go to blow bar ...	2	NDZvyYHTUWWu-kqgQ...
eFvzHawVJofxSnD7T...	0	2014-11-12 15:30:27	0	Xs8Z8lmKkosqW5mw_...	5.0	My absolute favor...	0	IQsF3Rc6IgCzjVV9D...

Yelp Data Analysis

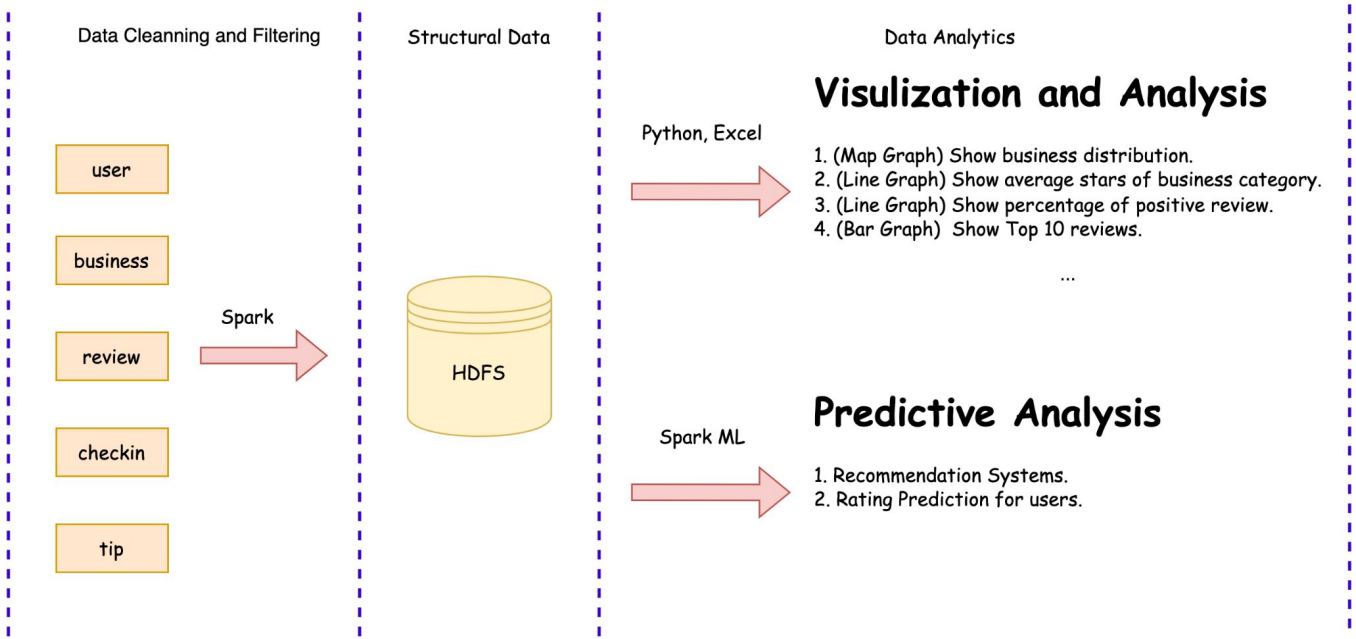
Data Sample: yelp_academic_dataset_checkin

business_id	date
---kPU91CF4Lq2-Wl...	2020-03-13 21:10:...
--0iUa4sNDFiZFrAd...	2010-09-13 21:43:...
--30_8IhuyMHbSOcN...	2013-06-14 23:29:...
--7PUidqRWpRSpXeb...	2011-02-15 17:12:...
--7jw19RH9JKXgFoh...	2014-04-21 20:42:...
--8IbOsAAxjKRoYsB...	2015-06-06 01:03:...
--9osgUCSDUWUkoTL...	2015-06-13 02:00:...
--ARBQr1WMsTWiwOK...	2014-12-12 00:44:...
--FWWsIwxRwuw9vIM...	2010-09-11 16:28:...
--FcbSxK1AoEtEAxO...	2017-08-18 19:43:...
--LC8cIrALInl2vyo...	2017-01-12 19:10:...
--MbOh2O1pATkXa7x...	2013-04-21 01:52:...
--N9yp3ZWqQIm7DqK...	2012-10-06 20:46:...
--O3ip9NpXTKD4oBS...	2010-04-17 21:07:...
--OS_I7dnABrXvRCC...	2018-05-11 18:23:36
--S43ruInmIsGrnnk...	2010-08-29 01:17:...
--SJXpAa0E-GCp2sm...	2014-04-06 22:23:...
--Sd93OFWITqDHifM...	2013-01-09 17:42:...
--ZVrH2X2QXBFdCil...	2010-08-12 18:21:...
--ZWv8kGlM2YL58uK...	2010-10-13 18:41:...

NYU

Yelp Data Analysis

Design Diagram



Yelp Data Analysis

Core Challenge 1

Data visualization:

Peel cluster doesn’t support tableau for data visualization. We export analytics results in csv format then use python and excel to visualize our results

Yelp Data Analysis

Core Challenge 2

Recommended Systems:

For the recommended system, the mainly core challenge is training the model using large volume of dataset (4G+). This scale of data requires more training iterations for our model. Plus, our model will have lower accuracy since and bias because of the lack of complete data. Thus, it's important to do the future engineering of the training set.

Yelp Data Analysis

Core Challenge 3

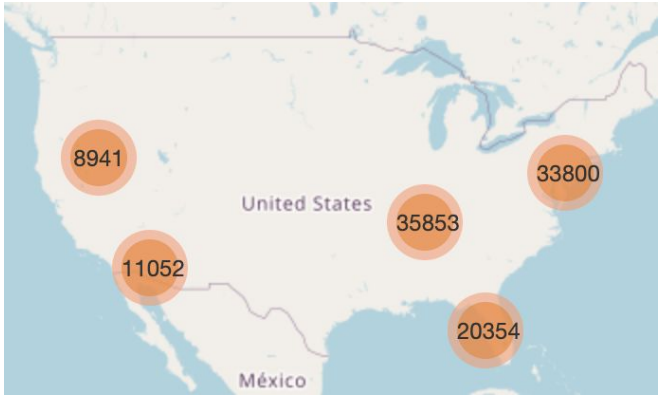
In friends relationship analysis:

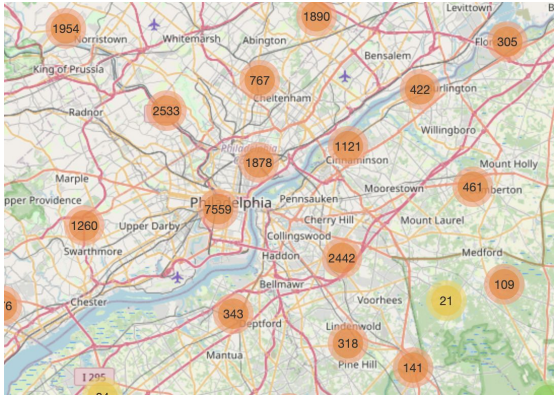
Each user has a list of his friends' id. However, some of those ids are not included in our primary key. In other words, we don't have data for that user. But while rendering the Force Atlas Graph, these data points are necessary. To solve this problem, we created a dummy point for that user and assigned it with 0 weight.

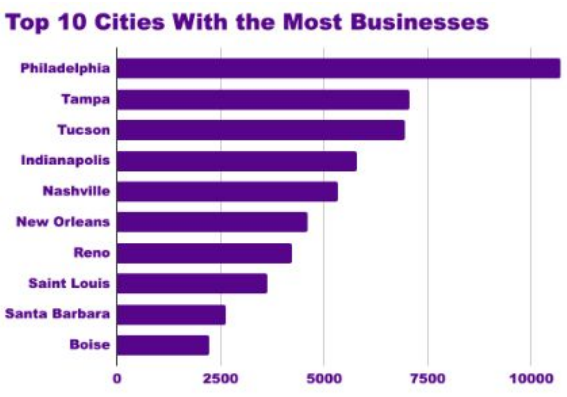
Yelp Data Analysis

Results - Overview:

We show our data of each business and number of business in different area on the map

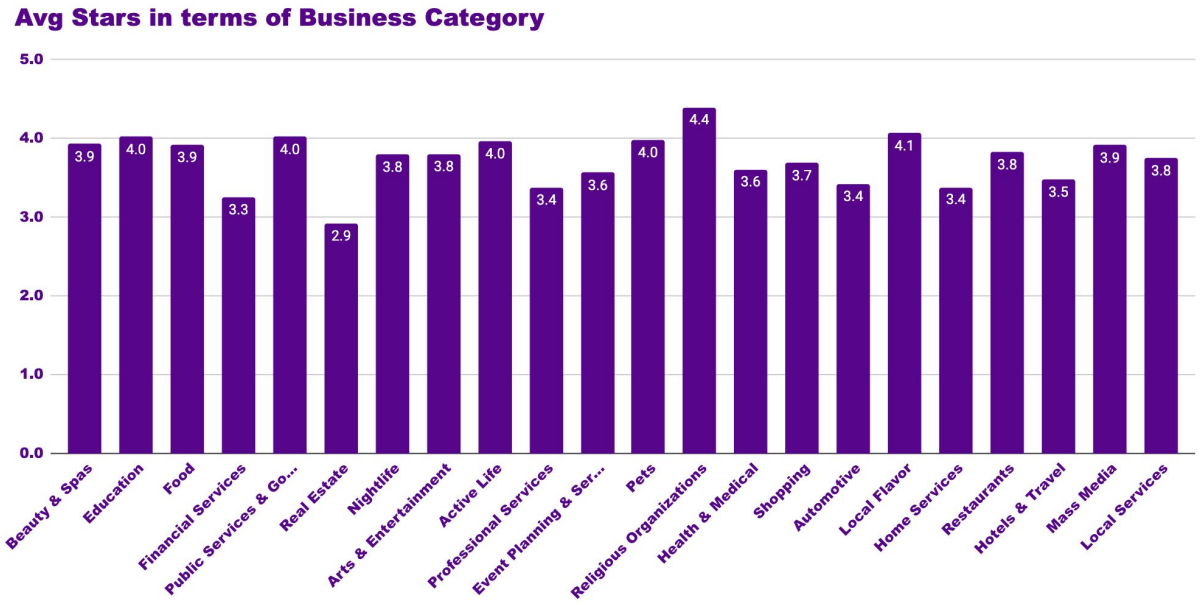






Yelp Data Analysis

Result: Average Stars of Business Category

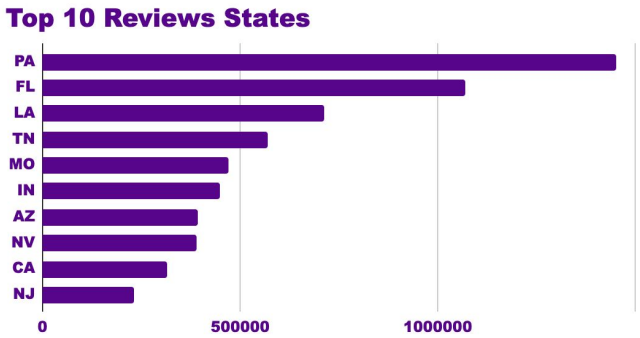


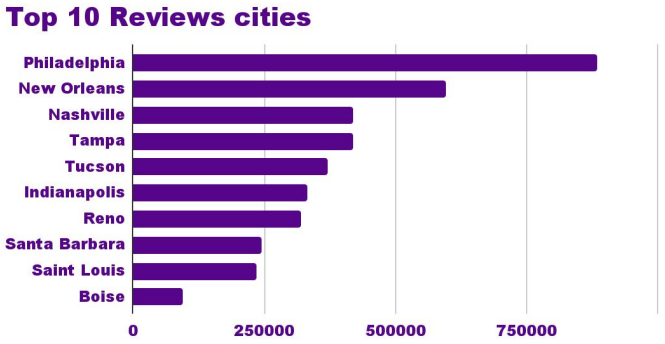
These figures show the Average Stars of Business Category.

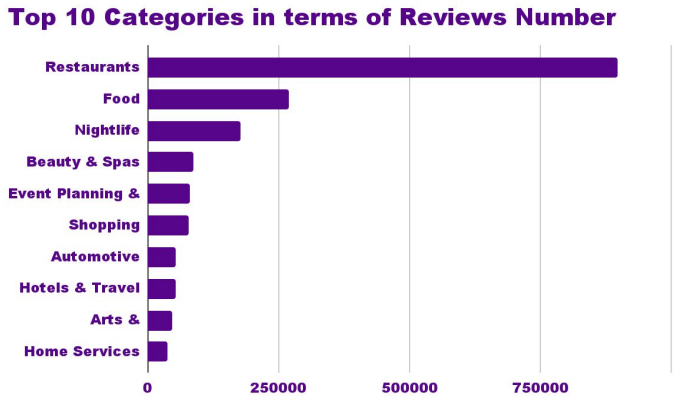
People scores higher on religious organizations.
The worst rating category is Real Estate.

Yelp Data Analysis

Result: Top 10 Review(State, City, Category)







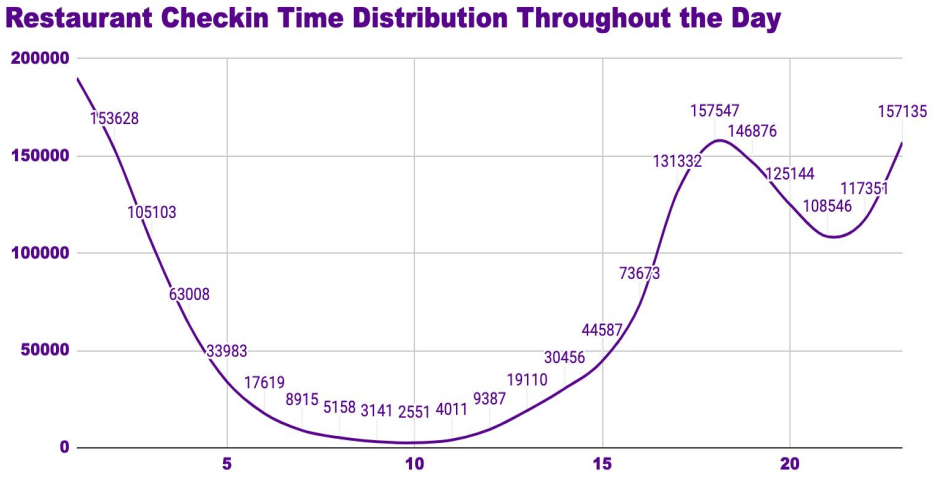
As we can see, restaurants is the most popular business category amount the dataset.

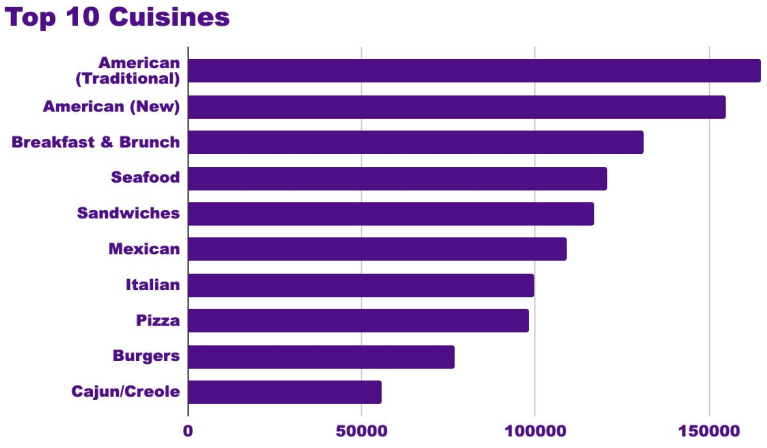
So we do some research on this category and get some interesting results

NYU

Yelp Data Analysis

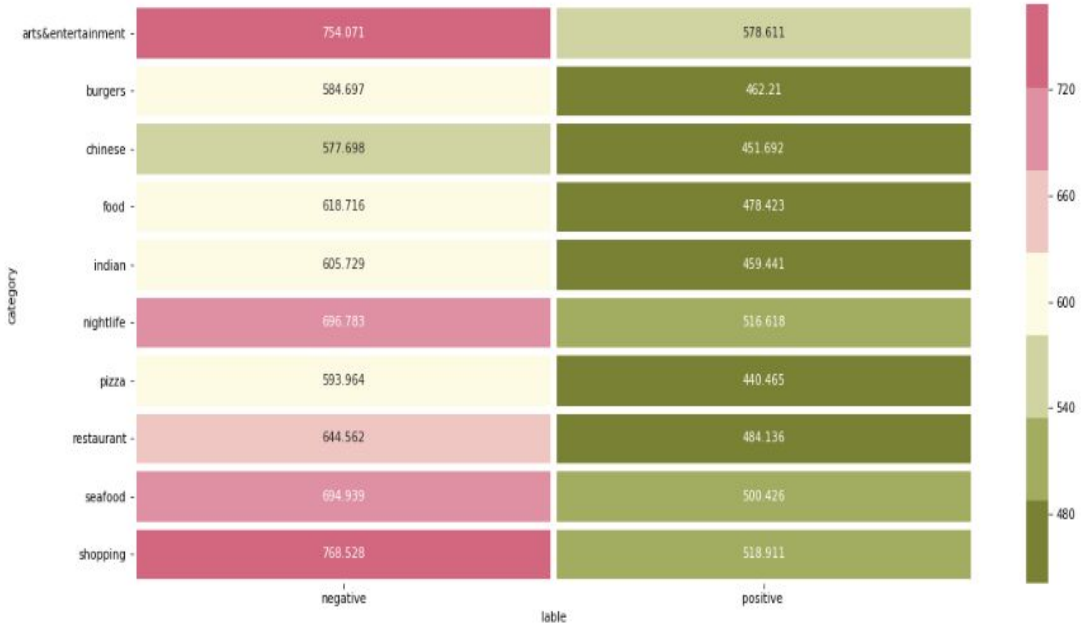
Results: Restaurant Overview Analyze





Yelp Data Analysis

Results - Review : length of positive and negative review



This graph compares the length of positive and negative reviews.

It shows that people were more likely to use more words to express dissatisfaction when they scored low, while they were more stingy with their evaluations when they scored high.

NYU

Yelp Data Analysis

Results - Review: the most used words

Indian Food Positive Keywords

Indian Food Negative Keywords

Chinese Food Positive Keywords

Chinese Food Negative Keywords

This is the word most used by users in positive and negative reviews of Indian Restaurant and Chinese Restaurant.

We can use this method to analyze people's evaluation of the restaurant and make improvements.

Yelp Data Analysis

Results: Recommended Systems

In this project, our team applied **Alternating Least Square (ALS)** model in Spark to train a recommended system.

After finishing the training, our team recommended 5 items for 10 randomly selected people.

Result is listed as follows (just show the first 4 people and their top 2 recommended services):

name	r1_name	r1_categories	r2_name	r2_categories
Benny	It' Sugar	Food, Candy Stores, **Specialty Food**	El Pastorcito	Restaurants, **Mexican**
Varusha	Not Just Crab	Restaurants, Seafood, Nightlife	Zitners Candy	Candy Stores
Crosby	June's Restaurant	Ice Cream & Frozen Yogurt	Pizza Hut	Italian, Chicken Wings
Daniel	City Grocers	Food, Grocery	Amore Pasta	Italian, Restaurants

Yelp Data Analysis

Results: Recommended Systems - Simple Test

Our team verified our recommended systems by comparing people's past consuming records and the recommended items.

name	p1_name	p1_categories	p2_name	p2_categories
Benny	The Seed	Food, Juice Bars, **Specialty Food**	The Cheesecake Factory - Reno	Restaurants, **Mexican**
Varusha	Roller Kingdom	Active Life, Skating Rinks		
Crosby	Polish Cottage	Polish, Restaurants	Pasco Kitchen & Lounge	American (Traditional)
Daniel	Pokey's	Pubs, Nightlife, Bars		

Yelp Data Analysis

Results: Friends relationship

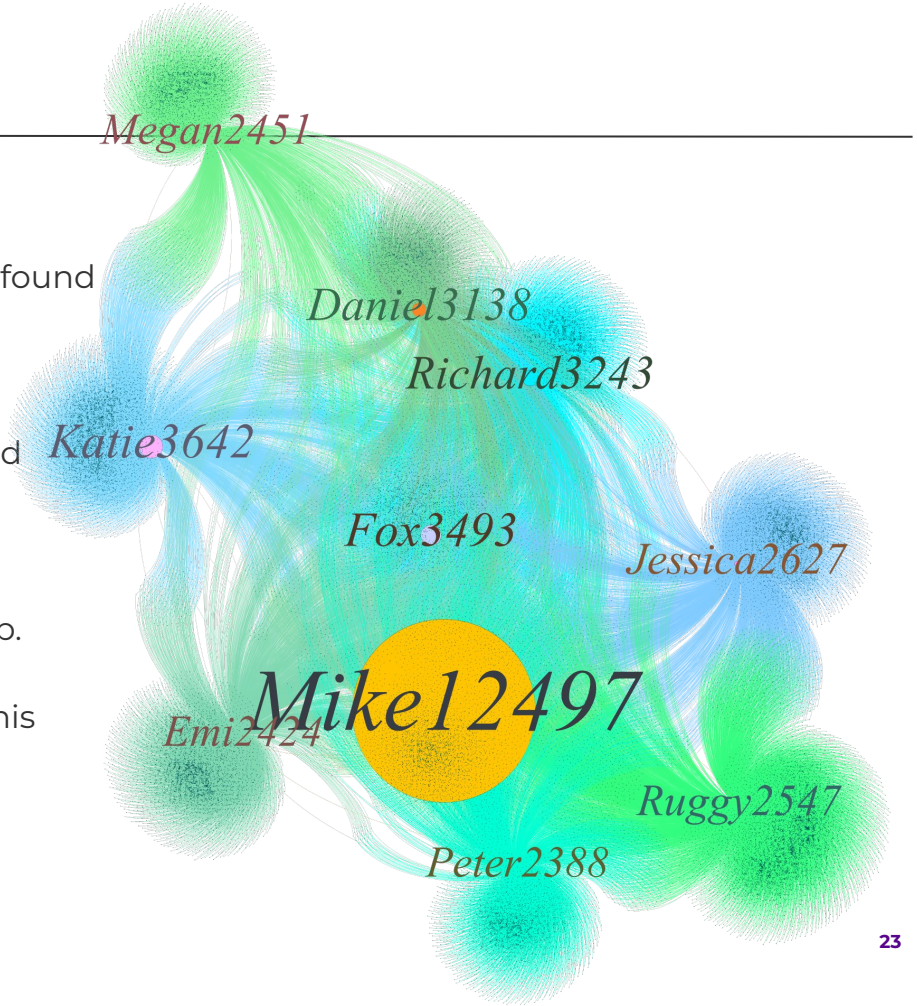
We analyzed yelp_academic_dataset_user.json and found the most popular users, aka influencers, who have the most followers.

The right hand side is a Force Atlas 2 Graph rendered with 57624 users data.

Each vertice represents for a user and an edge represents for a friend relationship.

For example, Mike has 12497 followers and most of his followers have also followed other influencers.

This was a 4K image before compressing.



NYU

Yelp Data Analysis

Obstacles

1. The datasets are really large compared to other teams, especially the review and user datasets. Thus, when running our program on the clusters, it may take long time to finishing each step, and it's not convenient to use the Zeppelin, because it's too slow to show the results.

2. The training set for the recommended system is too large to train for more iterations in Spark. It's easy to cause stack overflow when the number of iterations is greater than 25, this situation greatly affects the accuracy of the model. I think our model should have better performance if it had enough training iterations.

Yelp Data Analysis

Summary

Yelp's open datasets could actually explore many interesting patterns.

We analyze what makes a good restaurant and what concerns customers, and then make recommendations of the future improvement and profit growth. Building visualization on this dataset could provide people with potential stores to solve their requirements, sometimes helps them reduce much time to do the decision and exploration. Analyzing friends relationship helps for improving users experience of commercial advertising.

Hence, this data analytics could actually help people a lot.

Yelp Data Analysis

Acknowledgement

1. Thanks NYU Peel Cluster for providing us with machines to run our applications.
2. Thanks Professor Tang Yang for organizing and instructing this course.
3. Thanks Yelp for opening the Yelp datasets for us to perform our analytics.
4. Thanks everyone in this team for providing support for each other.
5. Thanks Tableau for providing us free student license.

Thank you!