

# **PROJECT MILESTONE 1**

## **1. PROJECT DETAIL:**

- **Name of the project:** Analysis of YELP Dataset
- **Team:** Kryptonite
- **Members:**
  - a. Siddhi Satish Jadhav(50442859)
  - b. Nisarg Negi(50442370)
  - c. Jagath Sai Narayana Kakaraparty(50442123)

## **2. PROBLEM STATEMENT:**

### **• DESCRIPTION OF THE DATA:**

Yelp is a local business directory service and review site with social networking features. It allows users to give ratings and review businesses. The review is usually short text consisting of a few lines with about a hundred words. Often, a review describes various dimensions of a business and the experience of the user with respect to those dimensions. The dataset is a subset of Yelp's businesses, reviews, and user data. It was originally put together for the Yelp Dataset Challenge which is a chance for students to conduct research or analysis on Yelp's data and share their discoveries. In the dataset, you'll find information about businesses across 11 metropolitan areas in four countries.

### **• DESCRIPTION OF PROBLEM:**

The objective is to design a relational database system that will use existing yelp data to provide insightful analytics and help existing business owners, and future business owners to make important decisions regarding new business or business expansion.

An exponential amount of data is created per day on the internet. Storing huge amounts of data and retrieving knowledge from it is a challenging task. Yelp is a website that publishes crowd-sourced reviews about local businesses. It provides an opportunity to business owners to improve their services and for users to choose the best business available.

### **• COMPARISON BETWEEN DB SYSTEMS AND EXCEL FILES:**

The original data set is in form of a JSON file, which is converted to CSV using python scripts in tabular form. Due to the following reasons, we have opted to store the data in a database system:

- a. The size of the dataset is 4GB which is difficult to store and access the data using CSV files. Whereas Postgres SQL has the capability to store large

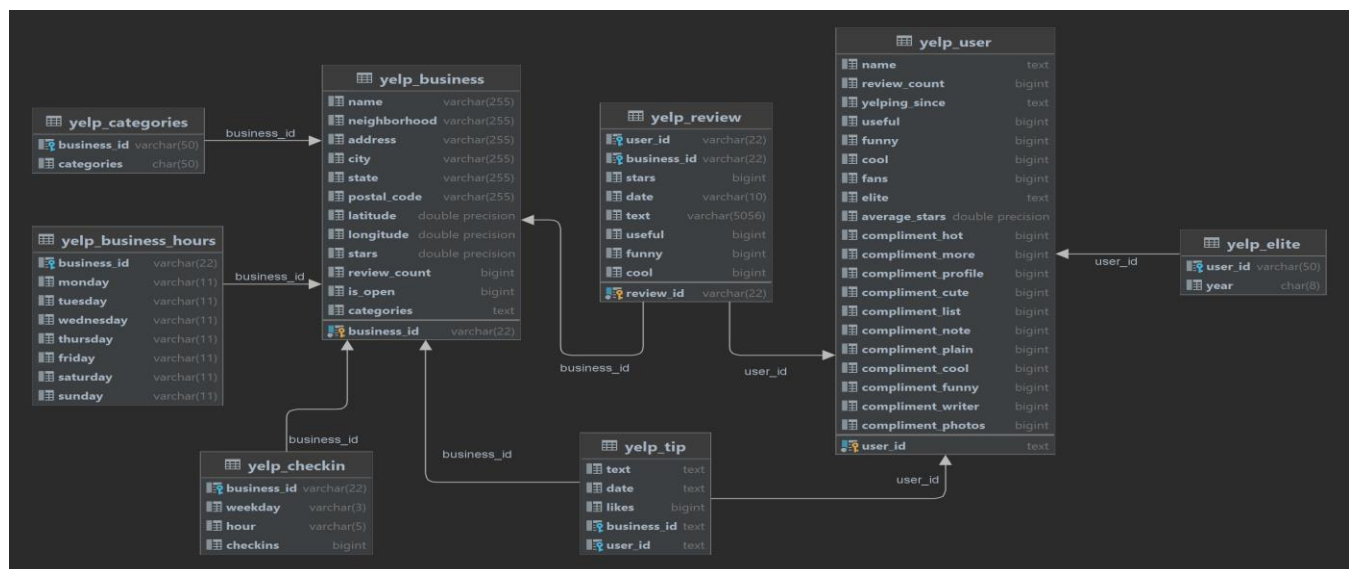
amounts of data. Hence for our case database system would be a better approach.

- b. The whole purpose of the project is to develop a relationship among the tables to obtain insights into the businesses. Postgre SQL helps in achieving this by means of primary keys and foreign keys. Whereas in the case of excel, a relational schema cannot be established.
- c. The yelp data has multiple groups that are there are multiple categories for each business. Using a database system, we can enlist the business based on the groups and perform further analysis. Such functionality is not present in excel.
- d. The main reason behind choosing a database base system over CSV for the current data is faster access as the size of the data is quite huge. Indexing can be performed on the database tables for faster lookup whereas CSV depends on the size of the ram to access the data.

### 3. TARGET USER:

- **User:** Users of the system will be owners who are willing to keep track of their business and make important decisions based on the visual plots such as changing the infrastructure or focusing on an area that will help them grow.
- **Real-life scenario description:** The objective of the dataset is to provide insightful analytics and help existing business owners, and future business owners to make important decisions regarding new business or business expansion.

### 4. E/R Diagram:



## 5. DATABASE IMPLEMENTATION:

### a. DATA SCHEMAS

Relation	Description	Relation with
Business	The table contains details about the business such as id, address location and etc.	Checkin, Business_hours. Categories Review
Business_hours	The table contains at what time will the business be open or close across the week.	Business
Categories	The table describes which domain the business belongs to, for instance, hotel, food and etc.	Business
Review	The table contains reviews shared by the user regarding the business.	Business User
User	The table contains details of the user such as id, name, year since they have registered and etc.	Review Tip Friend Elite_years
Friend	Yelp has the feature to add friends. The table contains id of people who are friends with the user.	User
Elite	Users who are writer and photographers are assigned the elite status. Hence the table has details when the user was made elite.	User
Tip	The table contains comments by the user on the particular business and the date and time.	User
Checkin	The table contains the number of checkin in the business on a particular week day.	Business

### b. RELATIONSHIP BETWEEN TABLES:

- The Checkin, Business\_hours, Categories, Review tables are linked to business table using business\_id.
- Review, Tip, Friend, Elite\_years are linked to user table using user\_id.
- Review table is linked to both user and business tables using user\_id and business\_id.

### c. ATTRIBUTES

Business Relation			
Name	Data type	Description	Constraints
Id	Varchar(22)	Unique id assigned to the business	No Null values
Name	Varchar(255)	Name of the business	No Null values
Neighbourhood	Varchar(255)	Which neighbourhood the business is located	Can contain null values
Address	Varchar(255)	Address of the business	No Null values
City	Varchar(255)	City the business located.	No Null values
State	Varchar(255)	State the business located	No Null values
Postal_Code	Varchar(255)	Postal code of the business	No Null values
Latitude	Float	Latitude of the location	No Null values
Longitude	Float	Longitude of the location	No Null values
Stars	Float	Rating of the business	No Null values
Review_count	Integer	Number of reviews the business has	No Null values
Is_open	Boolean	If business is open or not	No Null values

Business_hour Relation			
Name	Data type	Description	Constraints
Business_id	Varchar(22)	Unique id assigned to the business	No null values
Monday	Varchar(20)	Business timings on Monday	Contains null values
Tuesday	Varchar(20)	Business timings on Tuesday	Contains null values
Wednesday	Varchar(20)	Business timings on Wednesday	Contains null values
Thrusday	Varchar(20)	Business timings on Thrusday	Contains null values
Friday	Varchar(20)	Business timings on Friday	Contains null values
Saturday	Varchar(20)	Business timings on Saturday	Contains null values
Sunday	Varchar(20)	Business timings on Sunday	Contains null values

Categories Relation			
Name	Data type	Description	Constraints
Business_id	Varchar(22)	Unique id assigned to the business	No null values
Category	Varchar(225)	Category in which the business belongs to	No nuill values

Review Relation			
Name	Data type	Description	Constraints
Review_Id	Varchar(22)	Unique id for each review	No Null values
User_id	Varchar(22)	Unique id for each user	No Null values
Business_id	Varchar(22)	Unique id for each business	No Null values
Stars	Integer	Rating of each business by user	No Null values
Date	Datetime	Date of the review posted	No Null values
Text	Text	Review by the user	No Null values
Useful	Integer	User found the review useful	No Null values

Funny	Integer	User found the review funny	No Null values
Cool	Integer	User found the review cool	No Null values

Elite Relation			
Name	Data type	Description	Constraints
User_Id	Varchar(22)	Unique id for each user	No Null values
Year	char(4)	Year in which the user was made elite	Contains Null values

Tip Relation			
Name	Data type	Description	Constraints
User_Id	Varchar(22)	Unique id for each user	No Null values
Business_id	Varchar(22)	Id of the user who is friend of the current user	No Null values
Text	Text	Review left by user in the tip section	Contains null values
Date	DateTime	Date the review was given	Contains null values
Likes	Integer	Number of likes for the review	No null values, default value is 0

Checkin Relation			
Name	Data type	Description	Constraints
Business_Id	Varchar(22)	Unique id for each business	No Null values
Hour	Varchar(22)	Checkins at particular time	Contains null values
Checkins	Integer	Number of checkin on the particular date	No Null values
weekdays	Varchar(22)	Day of the week the checkins are counted	No Null values, default value 0

User Relation			
Name	Data type	Description	Constraints
user_id	Varchar(22)	Unique id for each user	No Null values
name	Varchar(32)	User's first name	No Null values
review_count	Integer	Number of reviews written	No Null values
yelping_since	Varchar(10)	Date when user joined Yelp	No Null values
useful	Integer	Number of useful votes sent by user	No Null values, default value 0
funny	Integer	Number of funny votes sent by user	No Null values, default value 0
cool	Integer	Number of cool votes sent by user	No Null values, default value 0
fans	Integer	Number of fans user has	No Null values, default value 0
elite	Varchar(76)	Years user was elite	Contains null values
average_stars	Float	Average rating of all reviews	No Null values

compliment_hot	Integer	Number of hot compliments received by the user	No Null values, default value 0
compliment_more	Integer	Number of more compliments received by the user	No Null values, default value 0
compliment_profile	Integer	Number of profile compliments received by the user	No Null values, default value 0
compliment_cute	Integer	Number of cute compliments received by the user	No Null values, default value 0
compliment_list	Integer	Number of list compliments received by the user	No Null values, default value 0
compliment_note	Integer	Number of note compliments received by the user	No Null values, default value 0
compliment_plain	Integer	Number of plain compliments received by the user	No Null values, default value 0
compliment_cool	Integer	Number of cool compliments received by the user	No Null values, default value 0
compliment_funny	Integer	Number of funny compliments received by the user	No Null values, default value 0
compliment_writer	Integer	Number of writer compliments received by the user	No Null values, default value 0
compliment_photos	Integer	Number of photos compliments received by the user	No Null values, default value 0

#### d. **PRIMARY AND FOREIGN KEYS**

- Business Relation
  - Primary key - business\_id
  - Foreign key - No Foreign keys
- Business\_hour Relation
  - Primary key - No Primary key
  - Foreign key - business\_id referenced from Business relation on Business.business\_id
- Categories Relation
  - Primary key - No Primary key
  - Foreign key - business\_id referenced from Business relation on Business.business\_id
- Review Relation
  - Primary key - review\_id
  - Foreign key - business\_id referenced from Business relation on Business.business\_id
  - user\_id referenced from User relation on User.user\_id
- Elite Relation

- Primary key - No Primary key
- Foreign key - user\_id referenced from User relation on User.user\_id

➤ Tip Relation

- Primary key - No Primary key
- Foreign key - business\_id referenced from Business relation on Business.business\_id  
user\_id referenced from User relation on User.user\_id

➤ Checkin Relation

- Primary key - No Primary key
- Foreign key - business\_id referenced from Business relation on Business.business\_id

➤ User Relation

- Primary key - user\_id
- Foreign key - No Foreign keys

➤ DELETE CASCADE constraint is used when we want the corresponding entries in the referencing relation to be deleted if the entries from the referenced relation are deleted.

We use the DELETE CASCADE constraint while adding foreign keys to the existing relations. The following relations with their respective foreign key are created with a DELETE CASCADE constraint.

- Business\_hour - Business\_hour.business\_id
- Categories – Categories.business\_id
- Review – Review.business\_id, Review.user\_id
- Elite – Elite.user\_id
- Tip – Tip.business\_id, Tip.user\_id
- Checkin – Checkin.business\_id

### e. RECORDS INSERTION

- User Table:

	business_name	name	neighborhood	address	city	state	postal_code	latitude	longitude
1	b2I2DXtZVnpUMC...	"Meineke	<null>	"2518 Ironw...	Sun Prair...	WI	53590	43.18508	-89.262047
2	Z-4KHBdbdwyC7P...	"William... Regent		"2701 Unive...	Madison	WI	53705	43.0731452	-89.4344018
3	zWB- kwaANIPMS...	"Banana ...		"32 W Towne...	Madison	WI	53719	43.0570896	-89.5092997
4	B5E7LEDH6AVDK8...	"Villa T... Sherman		"2302 Packe...	Madison	WI	53704	43.1233893	-89.3553901
5	NoxznJbWS4xD9f...	"The Bri... Williamson - M...		"701 E Wash...	Madison	WI	53703	43.0798171	-89.3762962
6	ewwSIX2mmvnpN0v...	"Spay Me...	<null>	"4475 Rober...	Madison	WI	53714	43.091157	-89.308487
7	AR6mrW089rq-ku...	"Munchie... Capitol		"111 N Broo...	Madison	WI	53703	43.0727667	-89.3901614
8	p0yt4Fon5r7833...	"Apple W...	<null>	"6313 McKee...	Fitchburg	WI	53719	43.0153668	-89.4816815
9	AuE8ETbL0gY1C...	"Walgree...	<null>	"275 Daviso...	Sun Prair...	WI	53590	43.1869254	-89.2288242
10	--LAZxfssKWd...	"Fitch's...	<null>	"5957 McKee...	Fitchburg	WI	53719	43.0148509	-89.4579482
11	SdqVhZDwk8c89E...	"J&P Fre...	<null>	"6634 Watts...	Madison	WI	53719	43.050634	-89.497717
12	in7ne7A3fhah-g...	"Stanton... Norman Acres		"4302 E Was...	Madison	WI	53704	43.127049421	-89.3107579753
13	ugdUuApH3pYxg...	"Mr. Tor... Elvehjem		"4527 Cotta...	Madison	WI	53716	43.083976745	-89.3033632212
14	gAy4LYpsScrj8P...	"Toast C...	<null>	"2429 Hwy 1...	Fort Mill	SC	29708	35.0472868	-80.9905588
15	7gquCdaFoHZCcL...	"Bubbly ...	<null>	"8439 Charl...	Fort Mill	SC	29707	34.9570617825	-80.8539120108
16	nSD-uxCmb745bH...	"Taco Be...	<null>	"2453 HWY 1...	Tega Cay	SC	29708	35.04967	-80.99063
17	d4GB80eVWVdfL5...	"Casual ...	<null>	"Tega Cay 6...	Tega Cay	SC	29708	35.024312	-81.0278546
18	Mmh4w2g2bsAkdS...	"Red Bow...	<null>	"845 Stockb...	Fort Mill	SC	29708	35.0630611	-80.993741

- Business\_hours table:

[illegible]



- Tip Table:

	text	date	likes	business_id	user_id
1	Great breakfast large portions and f...	2015-08-12		jH19V2I9fIslnNhDzPmdkA	ZcLKXikTHY0nYt5VYR05sg
2	Nice place. Great staff. A fixture ...	2014-06-20		dAa0hB2yrnHzVmsCkN4YvQ	oaYhjQ8bh18ZhU0bpyzSuw
3	Happy hour 5-7 Monday - Friday	2016-10-12		dAa0hB2yrnHzVmsCkN4YvQ	u1Q8Nyyj7jCUR8M83SUMoRQ
4	Parking is a premium, keep circling,...	2017-01-28		ESz03Av0b1_TzK0iQzbQYQ	u1Q8Nyyj7jCUR8M83SUMoRQ
5	Homemade pasta is the best in the ar...	2017-02-25		k7WRPbDd7rztjHcG6kEj1w	u1Q8Nyyj7jCUR8M83SUMoRQ
6	Excellent service, staff is dressed ...	2017-04-08		k7WRPbDd7rztjHcG6kEj1w	u1Q8Nyyj7jCUR8M83SUMoRQ
7	Come early on Sunday's to avoid the ...	2016-07-03		SqW3igh1_Png336Vib5DUA	u1Q8Nyyj7jCUR8M83SUMoRQ
8	Love their soup!	2016-01-07		KNpcPGqD0RDdvtekXd348w	u1Q8Nyyj7jCUR8M83SUMoRQ
9	Soups are fantastic!	2016-05-22		KNpcPGqD0RDdvtekXd348w	u1Q8Nyyj7jCUR8M83SUMoRQ
10	Thursday night is \$5 burger night	2016-06-09		KNpcPGqD0RDdvtekXd348w	u1Q8Nyyj7jCUR8M83SUMoRQ
11	Very good for lunch, bread and soup ...	2016-06-01		8qN0I6Q1-rJrvWWD5Btz6w	u1Q8Nyyj7jCUR8M83SUMoRQ
12	Tuesday spaghetti special all day \$...	2016-06-14		8qN0I6Q1-rJrvWWD5Btz6w	u1Q8Nyyj7jCUR8M83SUMoRQ
13	Sunday \$.55 bone-in wings Monday \$.5...	2016-08-22		--ujyvoQ1wVoBgMYtAdiLA	u1Q8Nyyj7jCUR8M83SUMoRQ
14	There isn't a bad seat in the house!	2017-01-21		FCUVjQf762no86Uzcbv1Tg	u1Q8Nyyj7jCUR8M83SUMoRQ
15	Nice happy hour m-f 5-7pm and th-fr ...	2016-06-13		TL_jT2a0bRMac5_YW656Pg	u1Q8Nyyj7jCUR8M83SUMoRQ
16	ALways have something new on the int...	2016-06-30		TL_jT2a0bRMac5_YW656Pg	u1Q8Nyyj7jCUR8M83SUMoRQ
17	Brisket sandwich not bad but ask for...	2015-10-28		c5NHHW0sNm7eaQBwvQJkkw	u1Q8Nyyj7jCUR8M83SUMoRQ
18	Double wide does offer free charging...	2016-02-11		zJwm3DThV4WUIkyNv68Ahw	u1Q8Nyyj7jCUR8M83SUMoRQ

## Checkin Table:

	business_id	weekday	hour	checkins
1	L30YVwagsz4Fuxpili4HUA	Thu	4:00	1
2	4Ast5npg50uPqUvLuY3yYw	Wed	20:00	2
3	6SWrq8hoojInUHeAcJBjCCg	Tue	23:00	1
4	_aX7bPzYQdrIVUJp6_-0zg	Sat	20:00	2
5	8dXY98cHf3HsWK72Azo1mw	Tue	23:00	1
6	ZPhQyU09MY5xNEcTqZ0rTQ	Thu	21:00	1
7	cWZgq05H-qYuCz-BjDUB7Q	Sun	1:00	1
8	XxnQKE9jT00bTE3LZ5N0tQ	Tue	13:00	1
9	iUavFbiZ44R2qj1r8eZ-QA	Tue	22:00	1
10	L9WB06gjzErC1bDyudzT6A	Mon	17:00	1
11	HAXeSAqr06j4EWFjiIbWEg	Tue	3:00	2
12	ByFMv3p5X1aNeZhU61rDcA	Sun	22:00	67
13	hJX8WoV-KbFVHojZ6i83Cg	Fri	19:00	1
14	-95mbLJsa0CxXhpaNL4LVa	Wed	14:00	2
15	-f2cr0yPvKENed2_-wmL9Uw	Mon	20:00	3
16	ad0LjmuXhU6BxTSKve5LZA	Tue	20:00	5
17	WpHj-1SBjdQ6og8isLVuPw	Tue	14:00	2
18	k0XZ8ks5u2t0Pkd-_BqITg	Tue	3:00	6

## Categories Table:

	business_id	categories
1	FYWN1wneV18bWNgQjJ26Ng	Dentists
2	FYWN1wneV18bWNgQjJ26Ng	General Dentistry
3	FYWN1wneV18bWNgQjJ26Ng	Health & Medical
4	FYWN1wneV18bWNgQjJ26Ng	Oral Surgeons
5	FYWN1wneV18bWNgQjJ26Ng	Cosmetic Dentists
6	FYWN1wneV18bWNgQjJ26Ng	Orthodontists
7	He-67vWjzVUysIKrfNbPUQ	Hair Stylists
8	He-67vWjzVUysIKrfNbPUQ	Hair Salons
9	He-67vWjzVUysIKrfNbPUQ	Men's Hair Salons
10	He-67vWjzVUysIKrfNbPUQ	Blow Dry/Out Services
11	He-67vWjzVUysIKrfNbPUQ	Hair Extensions
12	He-67vWjzVUysIKrfNbPUQ	Beauty & Spas
13	KQPW8LFf1y5BT2MxiSZ3QA	Departments of Motor Vehicles
14	KQPW8LFf1y5BT2MxiSZ3QA	Public Services & Government
15	8DSHNS-LuFqpEWIp0HxijA	Sporting Goods
16	8DSHNS-LuFqpEWIp0HxijA	Shopping
17	Pf0CPjBrLqAnz__NXj9h_w	American (New)
18	Pf0CPjBrLqAnz__NXj9h_w	Nightlife

## Review table:

	review_id	user_id	business_id	stars	date	text	useful	funny	cool
1	vkVSCC7x1jrrAI4UG...	bv2nC15qv5vroFi...	AEx2SYEUJmTxvVB...	5	2016-05-28	Super simple place but amazi...	0	0	0
2	n6QzIU0bkYshz4dz2...	bv2nC15qv5vroFi...	VR66pWIda3SfvPC...	5	2016-05-28	Small unassuming place that ...	0	0	0
3	MV3CcKScW05U5LVFF6ok0g	nC15qv5vroFi...	CKC0-MOWMHqoeWf6...	5	2016-05-28	Lester's is located in a bea...	0	0	0
4	IXv0zsEHYtiJi0CAR...	bv2nC15qv5vroFi...	ACFtxLv8pGrxHm...	4	2016-05-28	Love coming here. Yes the pl...	0	0	0
5	L_9BTb55X0GdtThi6...	bv2nC15qv5vroFi...	s2I_Ni76bjJNK9y...	4	2016-05-28	Had their chocolate almond c...	0	0	0
6	HRPm3vEZ_F-33TYVT...	_4iM0XbXZ1p10NG...	8QWPLVQ60-0ExqX...	5	2014-09-24	Cycle Pub Las Vegas was a bl...	1	0	0
7	ymAU680ZfqcFTBS0i...	u0LXt3Uea_GidxR...	9_C6hHMz8698M9-...	4	2012-05-11	Who would have guess that yo...	0	0	2
8	8UIishPU092hXtScS...	u0LXt3Uea_GidxR...	gkCorLgPyQLsptT...	4	2015-10-27	Always drove past this coffe...	1	0	0
9	w41ZS9shnepf03uEyn...	u0LXt3Uea_GidxR...	5r6-69C4YlbC7Zi...	3	2013-02-09	Not bad!! Love that there is...	1	0	0
10	WF_QTN3p-thD74hqp...	u0LXt3Uea_GidxR...	f0F_o2JPUB8R16y...	5	2016-04-06	Love this place! =Peggy is g...	3	0	0
11	PisUSmvaUWB00qv5K...	u0LXt3Uea_GidxR...	z8oIoCT1cXz7gZP...	4	2013-05-01	This is currently my parents...	1	0	0
12	PdZ_uFjbbkjtM3SCY...	u0LXt3Uea_GidxR...	XWTPNfksXoUL-Lf...	3	2011-09-28	Server was a little rude. =0...	5	0	1
13	x5oV6wm9_Pb1Q66jk...	u0LXt3Uea_GidxR...	13nKUHH-uEUxVZy...	1	2011-02-16	I thought Tidy's Flowers had...	9	0	1
14	LsoSqIrr0bQvWpHvs...	u0LXt3Uea_GidxR...	RtUvSW0_UZ8V3Wp...	3	2012-12-03	Wanted to check out this pla...	2	1	1
15	23eqwL2CwZKADWfD...	u0LXt3Uea_GidxR...	Aov96CM4FZAXeZv...	5	2010-07-16	This place is awesome! Defin...	2	0	1
16	FunI9om-ak5oMIiJm...	u0LXt3Uea_GidxR...	0W4lkcLzZThpx3V...	4	2011-09-28	a must stop when you are in ...	0	0	0
17	FKu4iU62EmWT6GZXP...	u0LXt3Uea_GidxR...	fdnNZMHk1NP7ZhL...	1	2012-10-23	I too have been trying to bo...	0	0	0

f. EXAMPLE OF RELATION WITH SAMPLE QUERIES

- select business\_id, hour  
from yelp\_checkin  
where weekday = 'Wed'  
and checkins > 4;

	business_id	hour
1	Rs1P2TXhuieeVQCsmInh8A	2:00
2	h5VZ0MqRt4vEUUG8oH6eEA	21:00
3	S0Pi1SpjaiPesEgrhbcPjQ	16:00
4	rZrhB-arwsF0LFVXzKvB6A	2:00
5	ybiJRruF5Cv5s26PPvPp7Q	3:00
6	K2Kx4NNKLfR1BnXL0gUb0A	15:00
7	5diHqpk_JSC4p8tIo4APow	5:00
8	97K4904NnvB_zSEVGHDc-A	0:00
9	Wxxvi3LZbHNIWJ-ZimtNA	10:00
10	hbaDM4I7PF7jtMM9oUx0kA	23:00
11	L0Wn1xi9PfG2kxrUuC21nA	21:00
12	gL0c1Fzl8yuKGq7B1TWj_A	22:00
13	QG9hp2nhcgIAIat88yH6aw	1:00
14	1ThbwFVBL9YA2o0M11i07g	16:00
15	64h5mGa2cJqLgoEwRfpDUQ	2:00
16	vXEZ-r6fah-5Fjt3a6c-Gw	22:00
17	GTgBY1nHrHHpP_wyeSU5Nw	14:00

- select max(review\_count) as Max\_review\_count  
from yelp\_business  
group by city ;

	max_review_count
1	63
2	4
3	4
4	6
5	15
6	270
7	9
8	3
9	87
10	5
11	3
12	3
13	6
14	164
15	50
16	68
17	16
18	9

- select business\_id, address, city from yelp\_business where business\_id in (select business\_id from yelp\_categories group by business\_id, categories);

	business_id	address	city
1	80AQRZr2D2uwYd50D0fLcg	"1171 Rue Sainte-Catherine E"	Montréal
2	2o-3auwmj6hHp9g1hgskuQ	"9440 W Sahara Ave, Ste 115"	Las Vegas
3	7c6VhzyyZc2NqfEx738tBQ	"3850 E Baseline Rd, Ste 111"	Mesa
4	16A01-VWTAkV72huVT409g	"2695 S Decatur Blvd"	Las Vegas
5	2E557i7ilCcneb-r-sDPAG	" "	Scottsdale
6	1Dp_zUe2P4vfPMJ8ZmR_pA	"19841 N 27th Ave, Ste 200"	Phoenix
7	1gMDCGRDP8KDvve-krHdmQ	"16 Waterloo Place"	Edinburgh
8	_ioKQnsGb_omiQaq0wYfAQ	"14045 N 7th St, Ste 3"	Phoenix
9	3j0wQL11Nt97avX7JRbrBQ	"2727 S Decatur Blvd"	Las Vegas
10	2aNbZrFzXBzSirexFDr5A	"1007 N 7th St"	Phoenix
11	3myjLbd7ggEVw09TYI5Tqw	"760A rue Rachel Est"	Montréal
12	2T7PzBsvLkqvXcP1C-YXVw	"7075 W Bell Rd, Ste 10"	Glendale
13	56go8Iox7D7SUzSwsuJ9Hw	"7615 Colony Rd, Ste 105"	Charlotte
14	2pNdmJjM7D-hgZX-uRcC6A	"4922 Sherbrook Ouest"	Montréal
15	5nAVwSp9WPHLvrP082Jrsw	"2900 Eglinton Avenue E"	Toronto
16	7bubnFWcymrUoBMvQ7W-1w	"327 Bathurst Street"	Toronto
17	7Ve39RJ0XBR8j1IPLQ127Q	" "	Scottsdale
18	1aXt6uHPrCxn0pbpdZa1hA	"5031 Wagon Trail Ave, Ste 107"	Las Vegas

- select name, review\_count, year from yelp\_user natural join yelp\_elite ;

	name	review_count	year
1	James	9	None
2	Lesley	3	None
3	Michael	22	None
4	Ben	203	2010
5	Ben	203	2011
6	Ben	203	2009
7	Ben	203	2012
8	Ben	203	2013
9	David	15	None
10	Jeff	11	None
11	Steven	1	None
12	Heather	64	2015
13	Heather	64	2016
14	Elizabeth	1	None
15	Chad	2	None
16	Eric	5	None
17	Onion	8	None

- `select user_id, name,  
review_count from  
yelp_user order by  
review_count desc limit  
10;`

	user_id	name	review_count
1	8k3a0-mPeyhbR5HUucA5aA	Victor	11954
2	RtGqdDBvvBCjcu5dUqwFzA	Shila	11323
3	P5bUL3Engv-2z6kKohB6qQ	Kim	9788
4	hWDybu_KvYLSdEFzGrniTw	Bruce	8353
5	8RcEwGrFIgkt9WQ35E6SnQ	George	7764
6	Xwnf20FKuikiHcSpcEbpKQ	Kenneth	6653
7	nmdkHL2JKFx55T3nq5VziA	Nijole	6314
8	CxD0IDnH8gp9KXzpBHJYXw	Jennifer	5868
9	HFECrzYDpgbS5EmTBtj2zQ	Eric	5344
10	KS1MQHYwIfD0462PE61IBw	Rob	4634

## 6. REFERENCE

### a. Dataset:

- <https://www.kaggle.com/code/jagangupta/what-s-in-a-review-yelp-ratings-eda>
- <https://www.yelp.com/dataset>

### b. Code:

- [https://github.com/sai2499/Kryptonite\\_Yelp\\_Dataset](https://github.com/sai2499/Kryptonite_Yelp_Dataset)

### c. Other references:

- <https://www.yelp.com/dataset/documentation/main>
- <https://github.com/Yelp/dataset-examples>
- <https://towardsdatascience.com/converting-yelp-dataset-to-csv-using-pandas-2a4c8f03bd88>