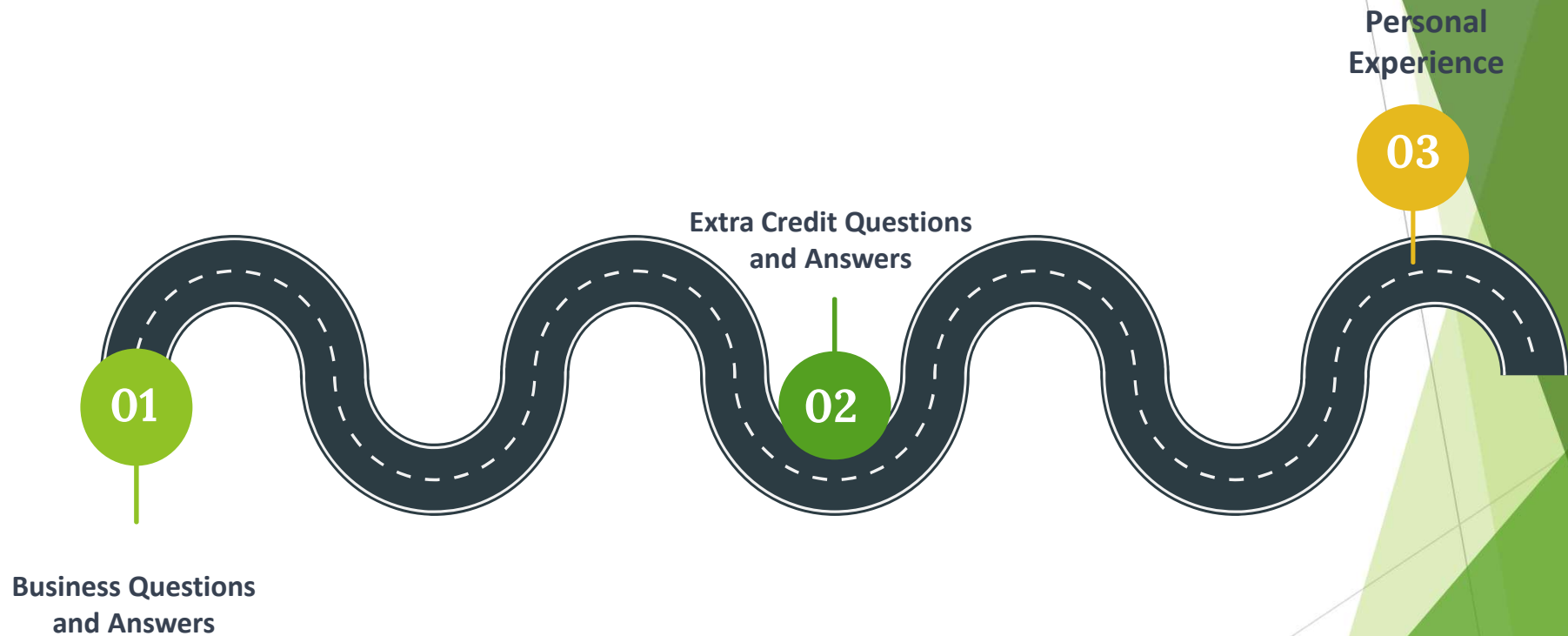


The background features abstract, overlapping green geometric shapes, primarily triangles and polygons, in various shades of green, creating a modern and dynamic visual effect.

# Data Management for Information Technology (MSIT 3860) Final Project

-Nandini Vimala Motupalli

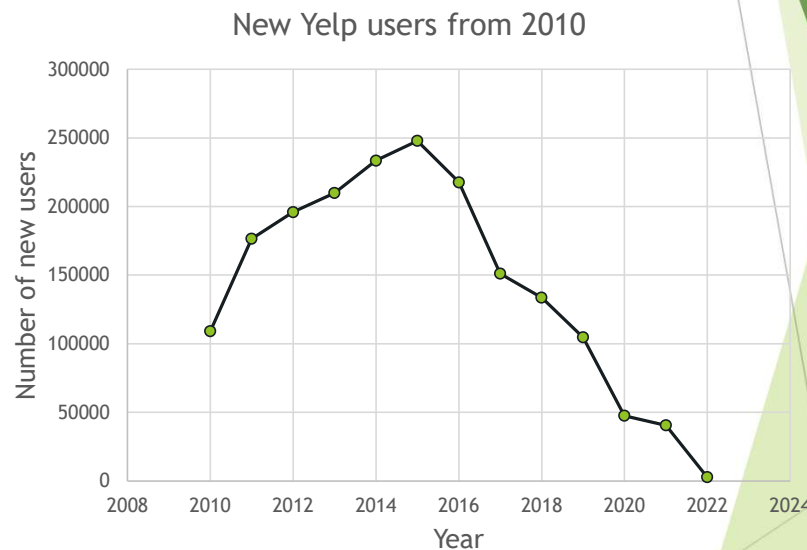
# Outline of the Presentation



## Yelp Users since 2010

### ► 1. How many users have joined Yelp since 2010?

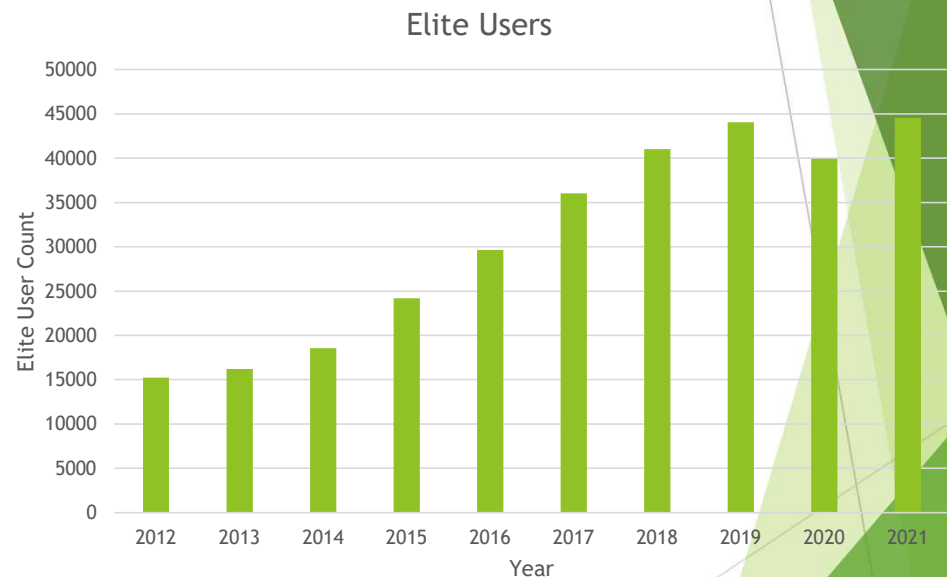
- The number of users who have joined Yelp since 2010 is 1870099.
- Based on the graph, it can be seen that the peak year for new users was 2015 after which there is a downward trend.
- The registrations reached an all time low in 2022.



## Number of Elite Users

2. How many users were elite in each of the 10 years from 2012 through 2021? Does it look like the number of elite users is increasing, decreasing, or staying about the same?

- The number of elite users kept increasing from 2012 to 2019.
- It was then followed by a period of slight decline in 2020 following which it increased again to the previous count of higher number of users.



## Most 5-star reviews

- 3. Which of our users has the most 5-star reviews of all time? Give us the person's name, when they joined Yelp, how many fans they have, how many funny, useful, and cool ratings they've gotten. Please also gives us 3-5 examples of recent 5-star reviews they have written.

"Love love love this place! Great food and always two meals with the portions. Service is always friendly too!"



"Jerry did an excellent job! My car is running great. He especially cares about, and I believe even offers discounts for the University students. He answers any questions you may have and is super helpful."

"Such yummy food and great customer service! I also had some delicious hot chai. The restaurant was empty and quiet, most of their customers were take away, so we got our food fairly quick. Everything we ordered was delicious highly recommend!"

Name	Fox
Joined Yelp	5/26/2009 11:33
Number of fans	3493
Funny Ratings	185823
Useful Ratings	206296
Cool Ratings	195814

## Relationship between business hours and reviews

4. We are wondering if there is any relationship between a business's hours of operation and the reviews they receive. Can you give us a spreadsheet giving the following information for every business we have in our database: business id, business category, and total hours of operation per week (i.e., the total number of hours each business is open across the entire week).

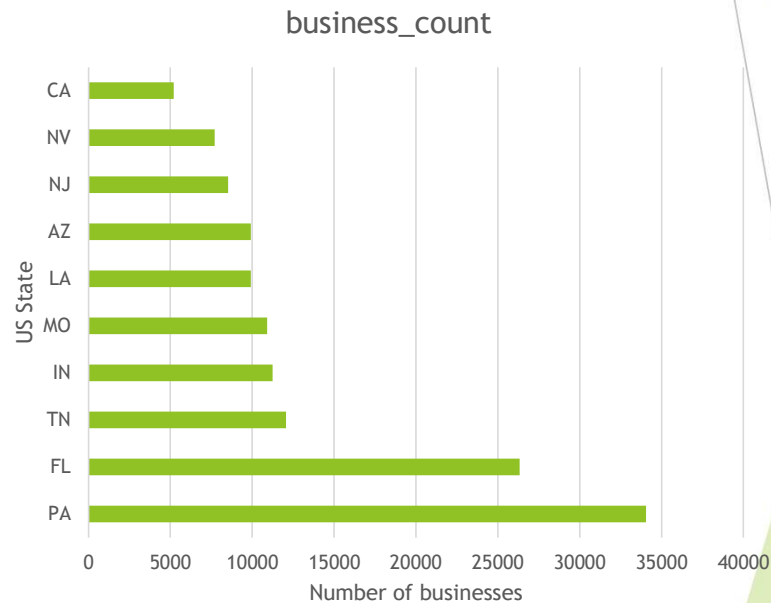
There are about 594858 rows in the table generated for this query. To answer this question here is the sample data for three businesses.

business_id	category	total_hours_per_week
kPU91CF4Lq2-WIRu9Lw	Food	52
kPU91CF4Lq2-WIRu9Lw	Food Trucks	52
kPU91CF4Lq2-WIRu9Lw	Latin American	52
kPU91CF4Lq2-WIRu9Lw	Restaurants	52
kPU91CF4Lq2-WIRu9Lw	Seafood	52
lZuj_WCGnDG6n0emlRg	Department Stores	91
lZuj_WCGnDG6n0emlRg	Discount Store	91
lZuj_WCGnDG6n0emlRg	Fashion	91
lZuj_WCGnDG6n0emlRg	Home & Garden	91
lZuj_WCGnDG6n0emlRg	Home Decor	91
lZuj_WCGnDG6n0emlRg	Men's Clothing	91
lZuj_WCGnDG6n0emlRg	Shopping	91
lZuj_WCGnDG6n0emlRg	Women's Clothing	91
--7jw19RH9JKXgFohspgQw	Dentists	42
--7jw19RH9JKXgFohspgQw	General Dentistry	42
--7jw19RH9JKXgFohspgQw	Health & Medical	42
--7jw19RH9JKXgFohspgQw	Orthodontists	42
--7jw19RH9JKXgFohspgQw	Pediatric Dentists	42

## States with the most businesses

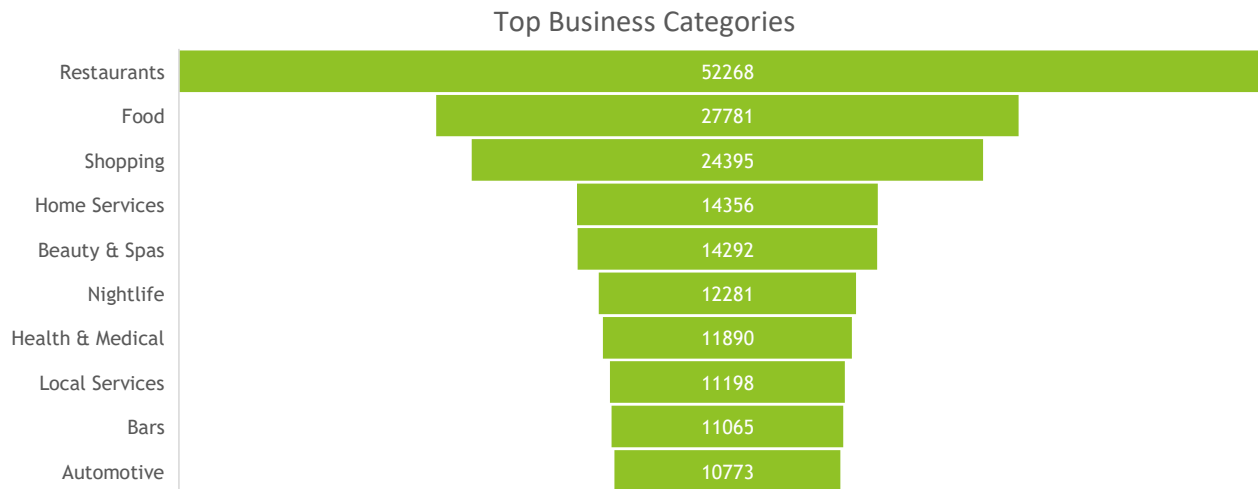
5. Which US states have the most businesses in our database? Give us the top 10 states.

- Pennsylvania and Florida have the highest number of businesses in the US.
- California despite being densely populated has the least number of businesses in the top 10 states.



# Top Ten Business Categories

6. What are our top ten business categories? In other words, which 10 categories have the most businesses assigned to them?

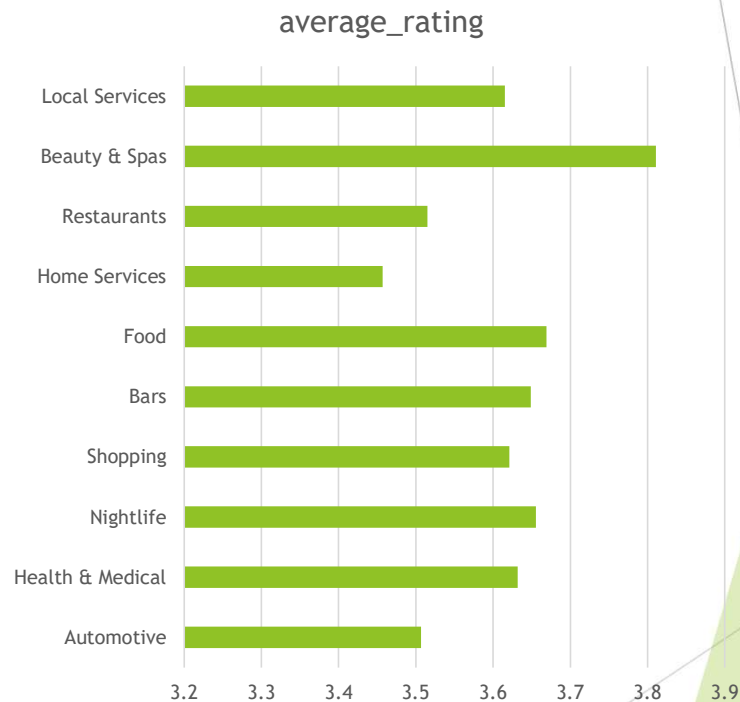




## Average Business Ratings

7. What is the average rating of the businesses in each of those top ten categories?

- The Beauty & Spas category stands out with the highest average rating, indicating overall satisfaction.
- Automotive and Home Services categories show relatively lower ratings. This suggests that there are areas of improvement which will help them increase their average rating.



# Funny and unfunny reviews

8. We're wondering what makes users tag a Restaurant review as "funny". Can you give us 5 examples of the funniest Restaurant reviews and 5 examples of the least funny?

## Funniest Reviews

review_text	funny_count
Went there for a birthday dinner and had reservations for 9 people (and 3 additional people)	792
Went to dinner here with the wife and the in-laws. Extremely solid all around. Great service, classy atmosphere, and extremely delicious food.	
One word of caution.... don't get your heart set on anything on the menu until you hear the specials. There were several the night we went and they all sounded great. I had the venison and it was outstanding. The rest of my party had seafood dishes and everyone was impressed.	
The wine list was solid as well--Cab heavy like most places, but had some good selections at acceptable price points in other areas as well.	
Dessert was all winners too. It's pricey, but it should be, because it's worth it.	378
Went to Brick's last night and it was amazing! A true Reno gem that I highly recommend	375
I don't know why I used to think Bricks was a French restaurant. Maybe it was at one point	370
Bricks is easily the nicest restaurant in Reno! Such a beautiful space and classy food. They gave our group the 5 star treatment! We ordered a bunch of appetizers to share - the best were the Barbecue Peppered Prawns and Baked Mushroom Caps. I think I ate two orders of the mushrooms myself! They were served in an Escargot plate and were stuffed with Escargot and Blue Cheese.	
Next I had a cup of their soup special which I can't remember if it was pumpkin or butternut squash - I just remember it was so delicious I forgot to share with my husband! (oops) We were supposed to be splitting the soup and the Spinach Salad. The best part of the Spinach Salad was the warm goat cheese - so melty and creamy! My entree was the Fettuccine Pescatore with huge chunks of lobster plus scallops and shrimp.	
Definitely a memorable meal with good friends in the Biggest Little City in the World!	369

## Least funny reviews

review_text	funny_count
Don't let the name of this restaurant fool you. This is my second time here. This is a heart attack in a box, which usually means great tasting food. It does here. Fried chicken, Shrimp. Gizzards, Chicken Livers, Fish "and steak fingers too," as the commercial jingle goes. (A jingle you'll never get out of your head once you hear it.) All succulent and flavorful. We love the fries, slaw and garlic toast too. EVERYTHING is a joy to eat, even if you have to pay later. The original family recipes from 1953 are still followed. This location is their first location in Tucson. #1, but there are now 8 locations in Tucson. This place is #1 indeed.	1
Morton's Steakhouse simply isn't worth it. Everything about my dinner there was lackluster. But the most important question is, how was the steak? My 8oz filet was overcooked (medium instead of medium-rare), burnt/blackened on top, and lacking in flavor. My date's slightly smaller filet was properly cooked (medium rare), but was similarly lacking in flavor. The chocolate cake dessert was okay, but I've had better at local diners. To make matters worse, our server did not introduce himself and was not pleasant towards us. There are so many wonderful steakhouses in Philly. Morton's just doesn't happen to be one of them.	1
Terrible place. Terrible service. Terrible customers. Terrible employees. Shut it down.	1
Stumbled across this unassuming-looking little joint, but very glad I did. Granted, I've only eaten there once.	

We'd also like you to look at a larger set of funny and unfunny reviews and tell us if you see any words or phrases that are commonly found in the funniest reviews. (We know the last part is qualitative but tell us anything you see that may be useful.)

After analysing 1000 funny reviews and using a frequency analysis tool these were the words which are most commonly present in the funniest reviews and their frequency.

- ▶ great: 520
- ▶ time: 517
- ▶ restaurant: 505
- ▶ chicken: 491
- ▶ delicious: 368
- ▶ cheese: 351
- ▶ nice: 314

## Relation between tips and compliments

9. We think the compliments that tips receive are mostly based on how long the tip is. Can you compare the average length of the tip text for the 100 most-complimented tips with the average length of the 100 least-complimented tips and tell us if that seems to be true? (Hint: you will need to use computed properties to answer this question).

- ▶ Average length of tip text for the 100 most-complimented tips: 146.63 characters
- ▶ Average length of tip text for the 100 least-complimented tips (49.55 characters)
- ▶ There seems to be a positive correlation between the length of the tip text and the number of compliments it receives. However, further analysis of more data is required to confirm this relationship.

## What factors influence reviews?

10. We are trying to figure out whether restaurant reviews are driven mostly by price range, how many hours the restaurant is open, or the days they are open. Can you please give us a spreadsheet with the data we need to answer that question? (Note from Professor Augustyn: You don't actually have to hand in a spreadsheet...just give me a table with 10 rows of sample data returned by your query.)

We would need the average hours per week that the restaurant is open, the days that it is open and the number of reviews as well so that we can identify that drives the reviews the most.

business_id	business_name	business_category	avg_hours_per_week	days_open	num_reviews
---kPU91CF4Lq2-WIRu9Lw	Frankie's Raw Bar	Restaurants	8.666666667	{Friday,Monday,Saturday,Thursday,Tuesday,Wednesday}	24
--8IbOsAAxjKR0YsBFL-PA	The Original Italian Pie	Restaurants	10.78571429	{Friday,Monday,Saturday,Sunday,Thursday,Tuesday,Wednesday}	27
--epgcb7xHGwJ-4PUeSLAw	Manhattan Bagel	Restaurants	7	{Friday,Monday,Saturday,Sunday,Thursday,Tuesday,Wednesday}	38
--hF_3v1JmU9nlu4zfXJ8Q	Green District Salads - Monument Circle	Restaurants	4.4	{Friday,Monday,Thursday,Tuesday,Wednesday}	15
--lqlzK-ZVTtgiQM63XgQ	Wendy's	Restaurants	-10	{Friday,Monday,Saturday,Sunday,Thursday,Tuesday,Wednesday}	16
--qLiYw2ErSmvVwumb2kdw	C.C.'s Kitchen	Restaurants	-1	{Friday,Monday,Saturday,Thursday,Wednesday}	5
--ZVrH2X2QXBfDCilbirsw	Chris's Sandwich Shop	Restaurants	10	{Friday,Monday,Saturday,Sunday,Thursday,Tuesday,Wednesday}	36
-_0w82rV27I51N3ABkCI0A	Prima Pizza Taqueria Mexicana	Restaurants	-10.85714286	{Friday,Monday,Saturday,Sunday,Thursday,Tuesday,Wednesday}	129
-_1ctmwdtSpjfpFuOpHXzw	Fat Boy's Pizza	Restaurants	-9.142857143	{Friday,Monday,Saturday,Sunday,Thursday,Tuesday,Wednesday}	7
-_arFpHzoPHvh5p-YKw5fQ	Krispy Krab 2	Restaurants	9.714285714	{Friday,Monday,Saturday,Sunday,Thursday,Tuesday,Wednesday}	69

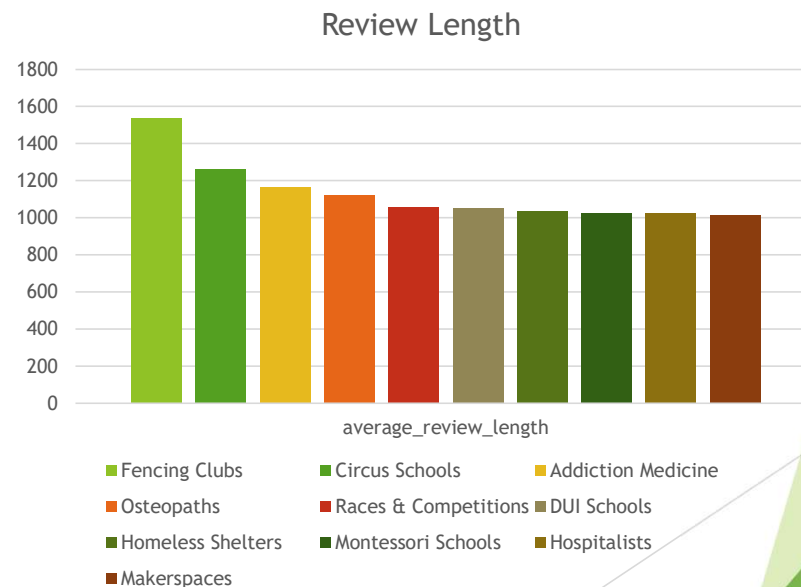
# Extra Credit Questions

## Categories with the longest reviews

1. Which business categories have the highest average review length?

Fencing clubs have the highest average review length with approximately 1538.2 words in a review.

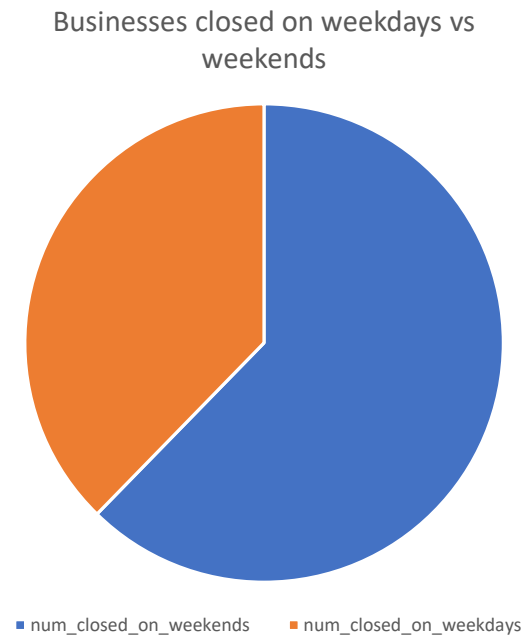
In the top 10 categories, there is a similar review length starting from races & competitions to makerspaces.



# Business closed on weekends vs weekdays

2. Which businesses operate solely during the weekdays and which operate only during the weekends?

- 38,949 businesses are closed during weekends.
- 23552 businesses only operate during the weekends and are closed the rest of the week.



# Personal Experience

1. What was your biggest challenge in working with this data? How did you overcome it?

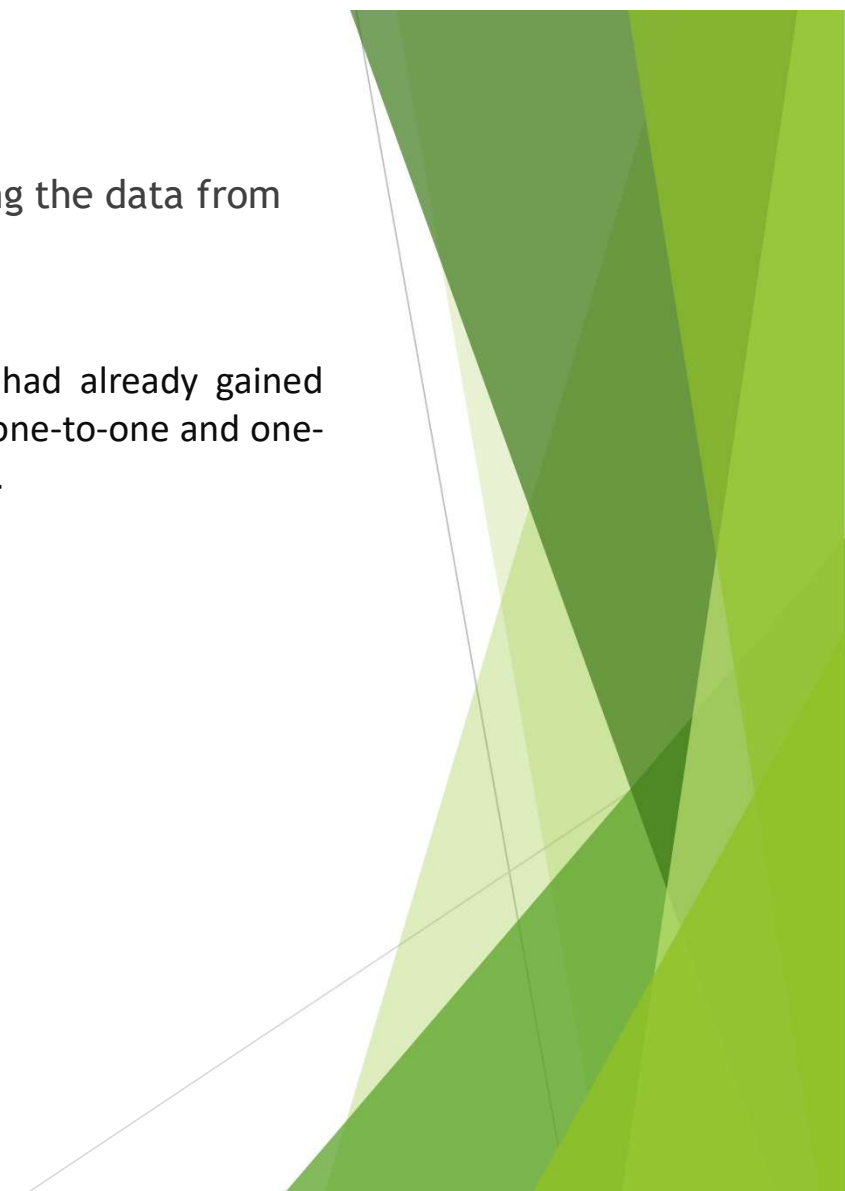
The biggest challenge I encountered while working with this data was the encoding mismatch. Whenever I attempted to load the file, I encountered an encoding error because pgAdmin uses WIN1252 format, whereas our files need to be UTF8 encoded. I managed to overcome this obstacle by making a slight modification during the file loading process, ensuring that the encoding was converted to UTF8. Here's the command:

```
\copy businesses FROM 'C:\Users\Nandini\OneDrive\Desktop\businesses.csv' WITH  
(FORMAT csv, DELIMITER ',', HEADER false, ENCODING 'UTF8');
```



2. What was the easiest part of the project for you (downloading the data from Moodle doesn't count!)? What made it easy?

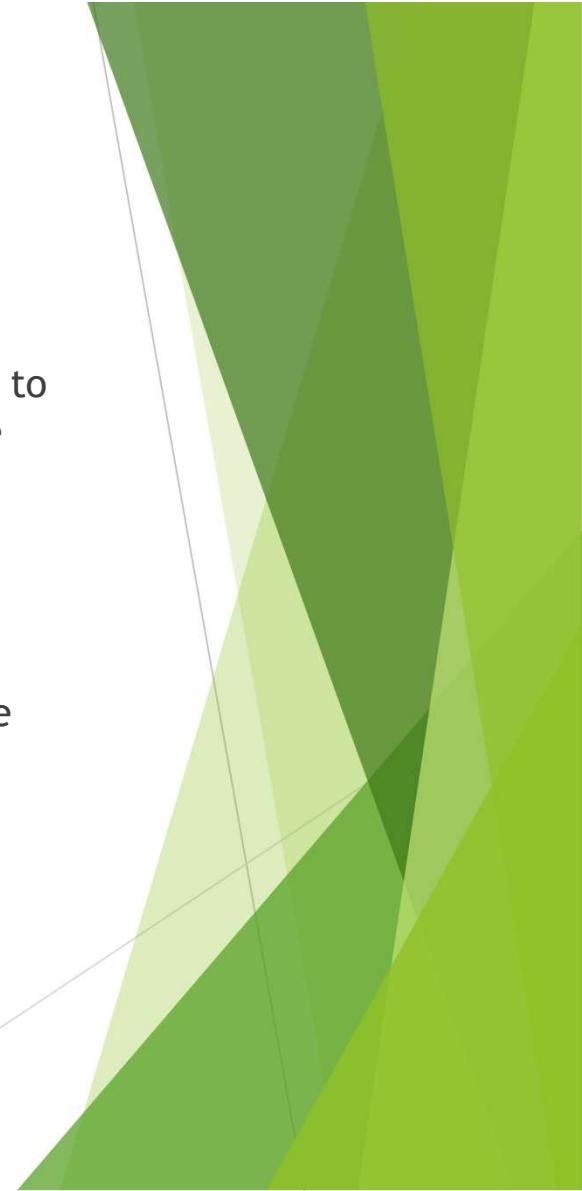
The easiest part of the project for me was creating the ERD. We had already gained experience with the Spotify ERD diagram and had practiced mapping one-to-one and one-to-many relationships, which made it straightforward to draw this ERD.



### 3. Was there anything that surprised you about the data?

I was surprised by the sheer volume of the data, particularly the review data. With 30 review files containing millions of rows each, it was a massive dataset to work with. Similarly, the number of attributes for each business was also more than my initial expectation.

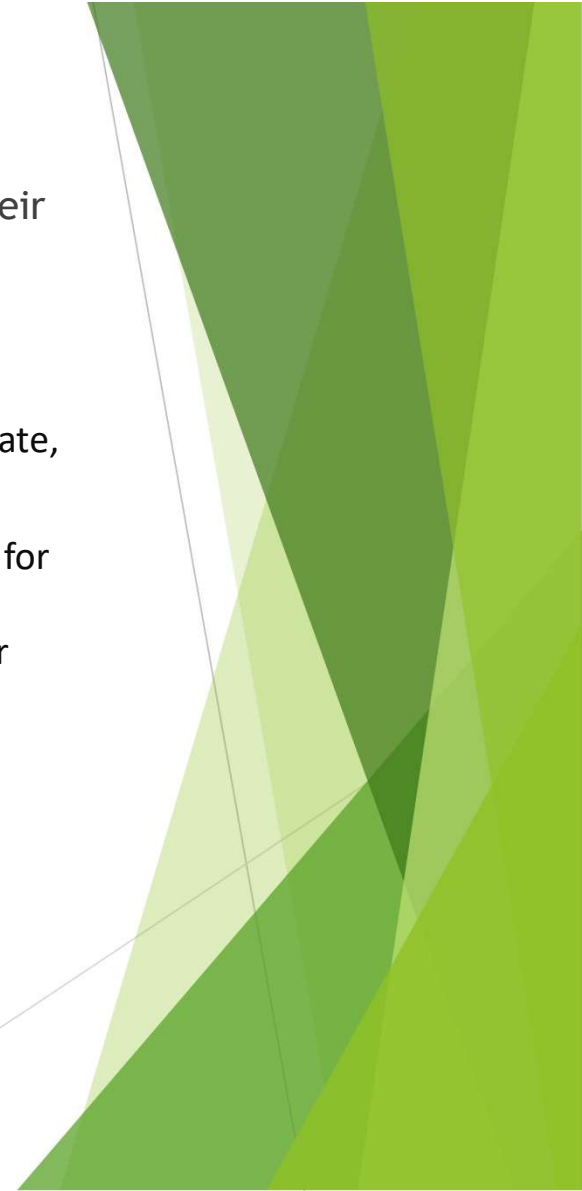
Another surprising aspect was how well-organized the data was, and how effectively it was interconnected. All the data ultimately led back to the businesses table, which shows that this is a well-structured relational database design.



4. Is there anything you wonder about how Yelp's real data might look, how their database systems are designed, or what they do with this kind of data?

Maintaining data integrity across their platform would be crucial for Yelp. This involves ensuring that information about businesses, reviews, and users is accurate and up-to-date, even as it evolves over time. I wonder how that is being done from their end.

Another aspect that I wonder about is the data analysis. Yelp likely leverages their data for various purposes, including understanding user behavior, identifying trends, and improving recommendations. Exploring how they analyze and derive insights from their data is something that I would like to get to know.



5. Are there any changes you would recommend to Yelp to make this data easier to work with?

Developing user-friendly query and analysis tools that allow users to explore Yelp's data, run custom queries, and visualize results can help users who work with data derive insights and make data-driven decisions more effectively.

Having proper data quality assurance processes, including data validation, cleansing, and monitoring, can help maintain the accuracy, completeness, and consistency of Yelp's data, making it more reliable for analysis and decision-making.

