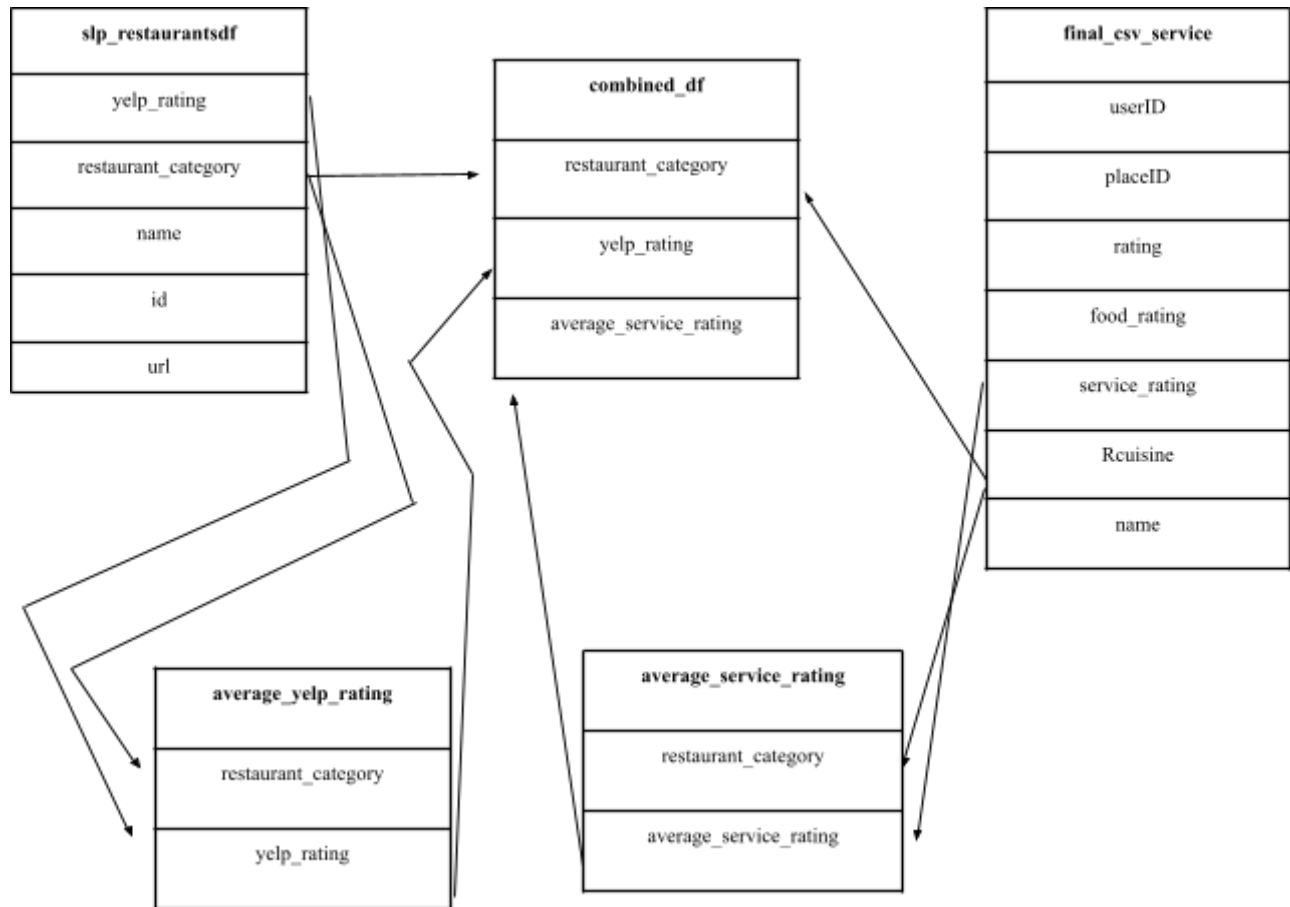# Final Project Report

Data

- List of Data Sources
  - Yelp API
    - Data Available: Reviews about different restaurants that can be filtered down to different locations and types of restaurants using keywords
  - https://www.kaggle.com/datasets/uciml/restaurant-data-with-consumer-ratings?resource=download&select=geoplaces2.csv
    - Data Available: CSV files that contain information that all corresponds back to a placeID. Information includes restaurant name, location, cuisine, service rating, etc.
- Data Used
  - Yelp API: I used the reviews about different restaurants by using yelps' 'search_query' function to search restaurants in a specific location. I also used the 'keyword' argument to filter the restaurants by cuisine.
  - CSV Data: I imported three different types of CSV files that all contained different information about restaurants. One contained information about the service reviews, another about the cuisine type, and another about the name of the restaurant. I combined all of the CSV files into one so that I could access all of the data from them with one ID.
- Cleaning Challenges
  - I found that when I was using the Yelp 'search_query', it had a limit on how many reviews it would return and it made it difficult for me to gather all the reviews for my data.

- RDB Schema:

slp_restaurantsdf

| yelp_rating |
| restaurant_category |
| name |
| id |
| url |

combined_df

| restaurant_category |
| yelp_rating |
| average_service_rating |

final_csv_service

| userID |
| placeID |
| rating |
| food_rating |
| service_rating |
| Rcuisine |
| name |

average_yelp_rating

| restaurant_category |
| yelp_rating |

average_service_rating

| restaurant_category |
| average_service_rating |

- Queries:
  - Query Case #1:

Query #1: Someone is looking for a restaurant that has a Yelp Review above 4.

```
#returns all rows from all columns where parameters are met
qs='''
SELECT * FROM total4_restaurants WHERE yelp_rating > 4
'''
run_query(qs)
```

	name	id	yelp_rating
0	La Bocana II	c7ePlnc5U8XoRSjNdwGPUw	5
1	Café Pacífico	_ep1A1tl70xn8Tfgr6AQdw	5
2	Mariscos Atlántico	dRCPnmmkTYOeMlmqN_IC4Q	5
3	Chocolatas el Empanada	C6wwn2rLCCSDg2B0aHhsWw	5
4	La Legendaria	ZajgrE7HHbloT1i7-VlKmA	5
...	...	...	...
227	El Pozole	ScWzHns4pkMaMxCo7Eox0A	5
228	Akopa	4v6FDum0fylcaEqL_mJpyg	5
229	Comedor Huasteco el Huapanogo	3FqjEBnTlphAcBxlM0aaXA	5
230	Rafa's Tacos	GM-G5s2T19mMlDcVsIps9A	5
231	Las Cazuelas	4Kt0zhI535U0Vra-_4gQZQ	5

232 rows × 3 columns

○ Query Case #2:

Query #2: Someone wants to know what Fast Food restaurants have a service rating higher than 1.

```
#returns all rows in columns except cuisine type where parameters are met
qt='''
SELECT name,service_rating,placeID FROM service_table2 WHERE restaurant_category = 'Fast_Food' AND service_rating > 1
'''
run_query(qt)
```

	name	service_rating	placeID
0	tortas hawai	2	132717
1	Subway	2	135021
2	Subway	2	135021
3	pizza clasica	2	135043
4	pizza clasica	2	135043
...	...	...	...
151	Tortas Locas Hipocampo	2	135085
152	Tortas Locas Hipocampo	2	135085
153	Tortas Locas Hipocampo	2	135085
154	Mcdonalds Parque Tangamanga	2	135086
155	Mcdonalds Parque Tangamanga	2	135086

156 rows × 3 columns

○ Query Case #3:

Query #3: Someone wants to know the service rating of all the Mexican Restaurants to pick the best one.

```
#returns all rows in columns except cuisine type where parameters are met
qk='''
SELECT name, service_rating, placeID FROM service_table2 WHERE restaurant_category = 'Mexican'
'''
run_query(qk)
```

	name	service_rating	placeID
0	Gorditas Dona Tota	2	132584
1	Gorditas Dona Tota	1	132584
2	Gorditas Dona Tota	2	132584
3	Gorditas Dona Tota	0	132584
4	Gorditas Dona Tota	1	132584
...	...	...	...
1423	El Rinc�n de San Francisco	2	135106
1424	El Rinc�n de San Francisco	0	135106
1425	El Rinc�n de San Francisco	0	135106
1426	El Rinc�n de San Francisco	2	135106
1427	El Rinc�n de San Francisco	1	135106

1428 rows × 3 columns

○ Query Case #4:

```
#returns all rows from all columns where parameters are met
ql='''
SELECT * FROM total4_restaurants WHERE name = 'La Fogatita'
'''

run_query(ql)
```

	name	id	yelp_rating
0	La Fogatita	kAuXfCQCthMuyNQxTptKhQ	2
1	La Fogatita	kAuXfCQCthMuyNQxTptKhQ	5
2	La Fogatita	kAuXfCQCthMuyNQxTptKhQ	2
3	La Fogatita	kAuXfCQCthMuyNQxTptKhQ	2
4	La Fogatita	kAuXfCQCthMuyNQxTptKhQ	5
5	La Fogatita	kAuXfCQCthMuyNQxTptKhQ	5
6	La Fogatita	kAuXfCQCthMuyNQxTptKhQ	5
7	La Fogatita	kAuXfCQCthMuyNQxTptKhQ	5

○   Query Case #5:

Query #5: Someone wants to know what restaurants that start with 'La' have a Yelp Review Rating higher than 3.

```
#returns all rows in all columns except id where parameters are met
qb='''SELECT name, yelp_rating FROM total4_restaurants WHERE name LIKE 'La%' AND yelp_rating > 3'''
run_query(qb)
```
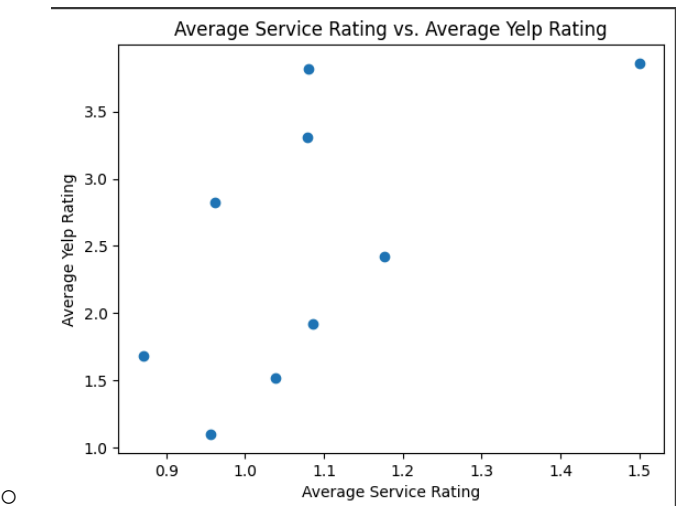
	name	yelp_rating
0	La Bocana	4
1	La Bocana II	5
2	La Chiquita del Caribe	4
3	La Legendaria	5
4	La Bamba	4
...	...	...
115	La Flama	5
116	La Cabaña Del Buen Sabor	5
117	La Fragua Steak Taco	4
118	La Cabaña de Pecos	4
119	Las Cazuelas	5

120 rows × 2 columns

○

- Plots
  - Average Service Rating vs. Average Yelp Rating
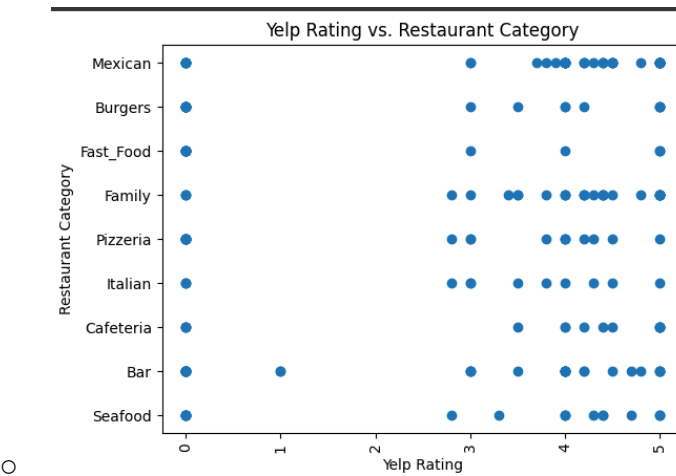

    Average Service Rating vs. Average Yelp Rating

  - Yelp Rating on Restaurant Categories

    Yelp Rating vs. Restaurant Category

  - Service Rating on Restaurant Categories

    Service Rating vs. Cuisine Type