

The background image shows a modern office environment. On the left, a woman with glasses and a blue top is gesturing while talking to a woman in a dark patterned top. In the center, a man with a beard is seen from the back. On the right, a man in a blue shirt and a woman in a white shirt are sitting at a desk, looking at a large monitor displaying data. The room has wood-paneled walls, black pendant lights, and black office chairs.

Data Analysis for “Yelp Academic Dataset”

Mohammad Shkeer

Nov 1st, 2023



Agenda

Client/dataset choice

Import & Clean

Questions to Answer

Initial Hypothesis

Data Analysis Approach

Introduction

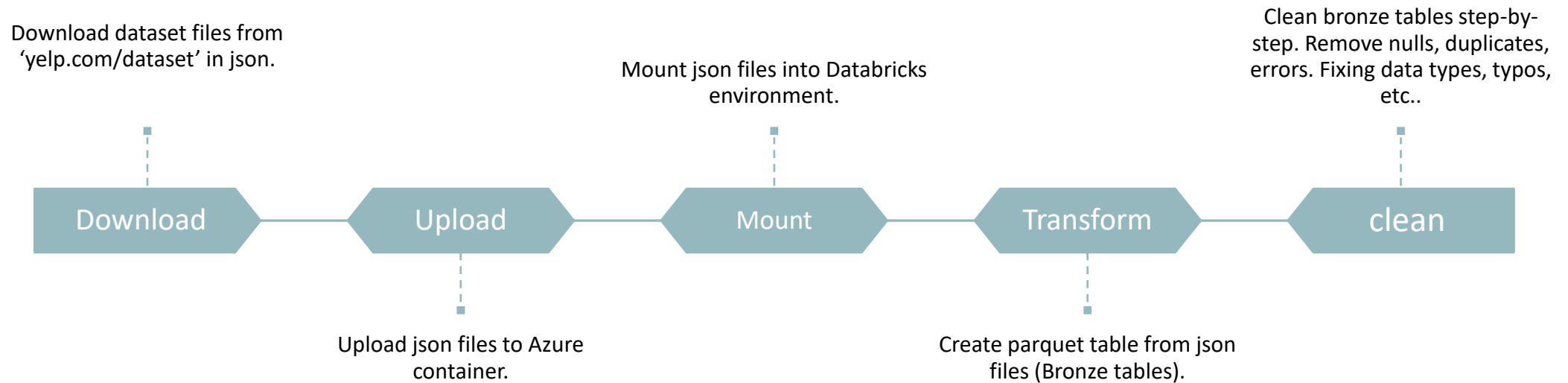
It is hard to work on a real project with multiple datasets and many data, especially when it's the FIRST project! I've challenged myself to dedicate as much time as necessary to learning the skills that will allow me to succeed on this assignment.



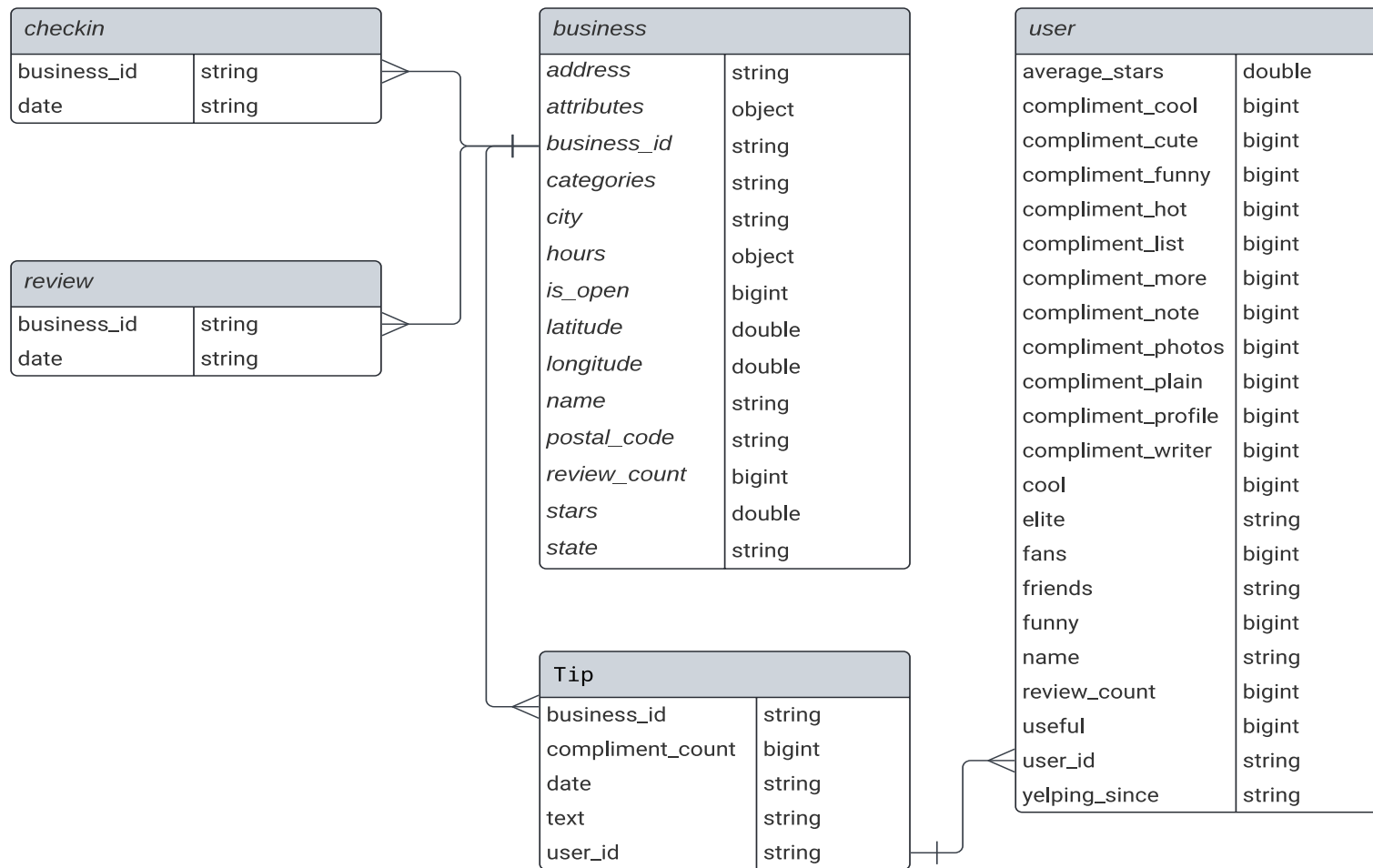
The primary reason for choosing the Yelp Academic Dataset was to utilize multiple skills and solve challenging problems. Additionally, the dataset presented an opportunity to apply machine learning for more in-depth feature analysis.



Import And Clean Data Process

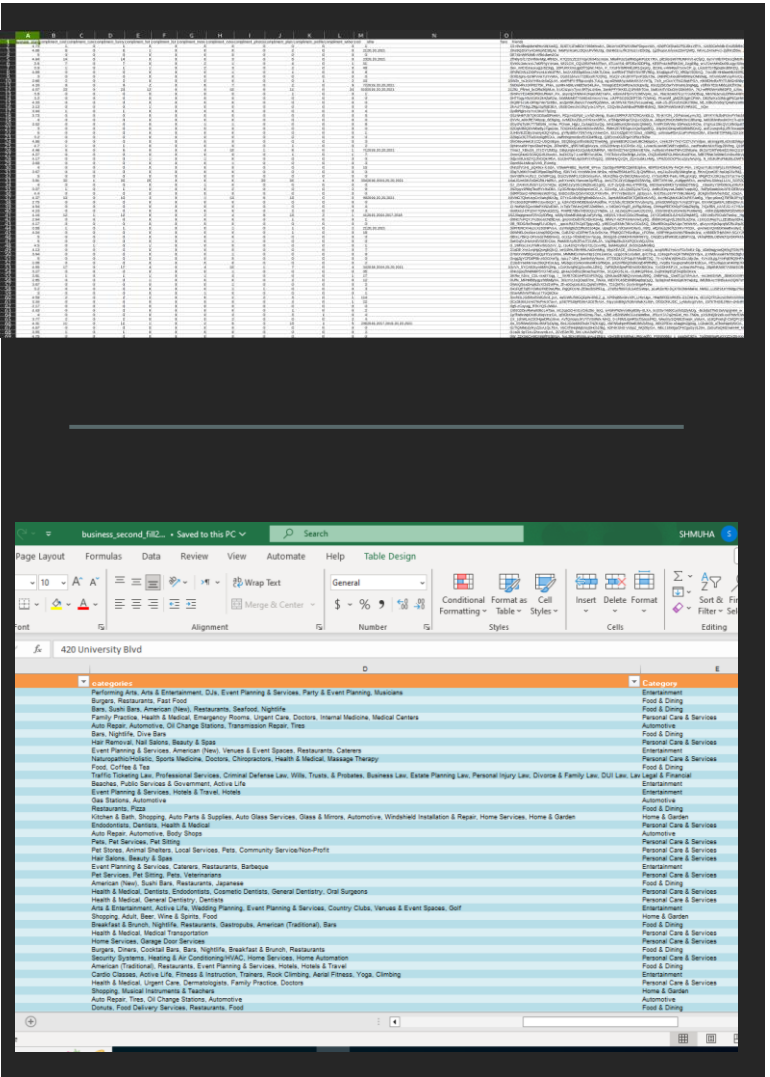


Entity Relationship Diagram



	A	B
1	business_id	date
2	BvSriuXZeaGAcFQExhRjg	2016-03-08 15:46:51, 2019-02-27 16:17:31
3	BvTRVSpjU5wvP2IEK2aqR	2015-04-02 22:20:23, 2015-06-10 22:10:54
4	BWV3138772-w6pF3nLQ	2012-02-19 02:58:17, 2012-10-20 00:14:10, 2013-04-09 21:49:22, 2014-12-10 01:59:16
5	BvYUdaSW4RLzEXtMvMzg	2011-11-06 21:53:23, 2011-12-21 16:47:14, 2011-04-14 23:28:13, 2011-03-07 15:13:28
6	BvZy_CaHplGmGmGmGmGm	2012-03-25 16:36:26, 2012-10-19 16:34:29, 2010-12-18 16:11:43, 2010-12-26 16:32:02, 2011-01-05 23:48:46, 2011-03-01 23:20:22, 2011-03-12 16:19:27, 2011-03-13 17:02:52, 2011-06-03 17:57:13, 2011-06-03 19:32:19, 2011-06-03 19:32:19
7	BvYgk14pm70uAbWfWv3	2015-03-21 18:04:37, 2015-05-25 14:39:30, 2015-10-23 19:28:41, 2017-10-10 13:25:25, 2019-01-10 23:34:54, 2019-02-06 23:29:10, 2019-02-07 23:04:57
8	BvYgk5Sme103v3v3v3v3	2018-08-25 01:53:33, 2018-08-25 14:20:45, 2018-10-16 18:40:33, 2018-10-10 17:47:21, 2018-11-03 19:41:49, 2018-11-07 18:08:55, 2018-11-09 17:46:00, 2018-11-28 20:04:33, 2018-12-03 23:06:52, 2018-12-06 19:19:22, 2018-12-19 16:46:01, 2018-12-20 15:50:59, 2019-02-01 14:58:45, 2019-02-01 14:58:45
9	BvYh5dC4V7L5YfL0S2GCG6A	2010-07-30 10:56:44, 2011-02-17 18:49:25, 2011-03-18 16:50:04, 2011-06-14 23:30:53, 2011-11-14 01:33:00, 2011-12-05 01:05:22, 2012-02-16 23:47:42, 2012-04-16 23:51:09, 2012-05-21 17:00:07, 2012-05-21 23:33:35, 2012-11-13 22:22:52, 2012-11-18 21:42:37, 2012-12-15 18:15:33, 2012-12-15 18:15:33
10	BvZj09gNOQU9GfCUMGAA	2012-03-20 15:46:28, 2015-09-16 23:10:48
11	BvIEKvHYEdb_6rc5S5	2011-02-12 00:33:48, 2012-01-03 14:13:02, 2012-01-04 20:07:29, 2012-01-09 16:40:34, 2012-01-11 21:31:21, 2012-02-10 21:08:54, 2012-02-13 14:05:52, 2012-02-13 16:10:04, 2012-02-13 22:12:07, 2012-02-15 14:48:26, 2012-02-16 16:35:15, 2012-02-16 16:35:15
12	BvYm9fGFC5HkYj1k4WAH	2011-11-26 00:34:59, 2013-05-18 15:56:59, 2015-04-15 14:56:15, 2015-07-02 15:15:45, 2015-08-30 23:35:13, 2015-08-30 22:09:52, 2015-09-28 14:05:05, 2015-12-02 19:38:27, 2015-12-03 21:35:16, 2016-05-13 18:16:24, 2016-12-29 21:32:16, 2017-06-27 16:55:09, 2018-03-24 18:57:55, 2018-03-24 18:57:55
13	BvYmXWqK5G5E5037gfw	2012-02-10 00:06:32, 2012-09-22 23:15:00, 2012-12-13 17:47:57, 2010-12-31 19:46:45, 2011-02-02 17:52:12, 2011-02-08 17:53:05, 2011-03-03 18:51:17, 2011-03-13 21:51:12, 2011-03-15 16:46:56, 2011-03-17 17:41:31, 2011-03-23 17:55:21, 2011-03-26 00:57:18, 2011-04-01 17:18:39, 2011-04-01 17:18:39
14	BvYmXWqK5G5E5037gfw	2011-05-13 12:39:26, 2011-05-13 12:39:26, 2011-06-21 17:21:32, 2011-08-07 07:52:51, 2011-09-02 02:29:36, 2011-09-24 01:16:59, 2011-10-12 02:19:37, 2011-12-03 17:49:46, 2011-12-03 20:35:16, 2011-12-04 02:43:28, 2012-01-30 20:25:25, 2012-02-29 12:23:23, 2012-03-03 14:26:45, 2012-03-03 14:26:45
15	BvYmXWqK5G5E5037gfw	2010-08-15 00:17:43, 2010-09-01 23:20:01, 2011-04-22 07:27:32, 2010-12-22 02:43:18, 2010-12-24 03:19:38, 2010-12-24 03:19:38, 2011-04-02 17:15:37, 2011-01-30 19:48:36, 2011-04-03 22:11:02, 2011-04-13 22:06:06, 2011-04-16 20:06:06, 2011-04-24 18:09:16, 2011-04-26 02:04:07, 2011-04-26 02:04:07
16	Bvu_tRuPQQG5SdWdy_FvIA	2019-03-28 23:22:34, 2019-03-29 22:54:42, 2019-04-05 23:02:19, 2019-05-17 23:58:23, 2019-06-29 00:03:16, 2019-07-13 18:42:29, 2019-08-17 23:02:47, 2019-09-10 05:03:08, 2019-10-12 23:39:48, 2019-12-14 18:13:48, 2020-06-03 05:05:29, 2020-06-03 23:03:28, 2020-06-03 22:21:25, 2020-06-03 22:21:25
17	BvxtJpJgX2B43yWdy_FvIA	2012-03-22 00:18:09, 2012-08-23 23:29:47, 2013-01-07 00:48:50, 2013-01-25 01:48:43, 2013-01-14 01:01:24, 2015-03-02 03:10:06, 2015-03-18 11:47:50, 2015-03-19 22:57:45, 2015-09-18 14:03:33, 2016-03-13 16:07:39, 2016-06-15 15:53:36, 2016-07-01 14:29:03, 2016-07-01 14:29:03
18	Bw-FkJBZC7YfgdKzGGCUsL	2011-03-17 21:28:15, 2011-11-09 22:56:46, 2011-12-13 14:20:47, 2012-02-21 20:57:22
19	Bw00b784A1eumhR0XRLg	2011-12-27 03:28:53, 2011-12-27 16:02:36, 2012-02-28 06:06:33, 2013-01-20 16:06:35, 2013-

	B	C	D
1	compliment	count	date
2			text
3	0	2016-01-30 23:59:00	Strong drinks fun dive bar
4	0	2018-09-16 15:29:56	Best breakfast! Wait is a little long but it is worth the wait!
5	0	2014-06-20 02:51:27	The food and service are the best in NOLA! Thank you
6	0	2021-05-08 15:00:47	They cold with it
7	0	2021-05-10 00:33:30	Amazing Customer Service! Highly recommend!!!
8	0	2021-05-22 07:45:07	Really good smoothie
9	0	2021-07-01 22:13:32	Coming out of lock down. Could not think of a better spot to have lunch with a buddy. Always a great for lunch or breakfast. Service and Food is always good.
10	0	2019-06-25 12:56:18	Soooooooooo goooooood
11	0	2021-08-15 23:55:13	Overpriced chain water ice. Tranzilliz is much better.
12	0	2021-07-16 23:10:28	Close to August 2021..
13	0	2020-10-14 01:45:19	Always so nice to me there! Especially sunny! Always a pleasure going here
14	0	2021-10-29 04:02:37	Waitress Amanda had an attitude, beer was served not chilled.
15	0	2018-05-18 02:21:48	The best chicken wings!!! Great friendly & welcoming service.
16	0	2017-07-06 18:50:18	Had a chicken cutlet sandwich it was rubbery
17	0	2018-06-18 23:40:23	Make a reservation
18	0	2018-06-26 22:24:08	Always a great time food is prepared with love and interesting.
19	0	2013-01-11 14:34:09	They should have 5 stars. Food is good and service is exceptional
20	0	2018-07-19 20:44:16	They are awesome.
21	0	2019-07-27 23:54:50	Jason dried off my car at car wash today and did an incredible job! Recommend them!
22	0	2011-06-18 14:58:00	Green lantern w Jeff
23	0	2011-10-15 02:15:53	Pork with pineapple & Shrimp tacos are amazing!
24	0	2018-09-10 21:40:16	Cheesesteaks are so juicy! Cooked onions were delicious!!
25	0	2016-04-09 19:40:24	We just had breakfast at the Nook. It was very good!! The service was excellent
26	0	2018-10-01 21:53:23	The restaurant is now serving slop and overpriced
27	0	2018-08-12 03:14:49	Bakery and del!!!!
28	0	2018-12-01 23:41:41	Yum!
29	0	2018-12-07 16:10:30	Great customer service and great pricing as well as quick turnaround times!
30	0	2016-06-26 21:47:06	You can sample only ONE flavor. Choose wisely
31	0	2016-08-21 16:17:35	It's a wonderful family atmosphere. Food is amazingly good! And prices are affordable and the amount of food you get is well worth it!
32	0	2018-12-07 19:40:54	My favorite person in this place is Katie, the studio manager and her baby Betsy!!
33	0	2019-01-16 22:32:17	Great tasting food! The food is always fresh and the restaurant is clean.
34	0	2018-02-26 21:37:04	Iced coffee is GROSS. strong and bitter and tastes old :(
35	0	2018-12-20 06:45:08	Happy hour everyday from 2pm to 6pm except on Mons and Tues it's all day. If it's your birthday month bring 4 friends on wednesdays and your entree is free!
36	0	2019-03-03 23:30:50	Great hot pot place with fast service and delicious broth!
37	0	2019-03-11 03:29:51	Shrimp and grits are amazing here!!
38	0	2015-04-19 20:50:59	Food is overrated and overpriced...better off at Popeyes!
39	0	2018-06-08 22:34:12	Far cheaper prices than the one on Golf Links and nicer employees
40	0	2019-06-20 22:47:46	Just amazing... service was great... food was delicious and timing was killer!
41	0	2019-08-11 18:47:29	The poutine and eggs were outstanding! So good!
42	0	2019-08-14 02:17:38	My favorite ! Great everything!
43	0	2014-12-01 19:11:51	Get your drink in a to go cup to enjoy while shopping in the market!
44	0	2019-10-13 20:26:14	Gelatos and ice cream were great! Creative and tasty.
45	0	2019-07-20 22:29:42	Beautiful view, nice servers but the food & drinks...not the best.
46	0	2010-10-20 18:21:51	One of my favorite brunch places!





Questions to Answer

Issues we're trying to solve

1- What are the fundamental features a restaurant must have to achieve success (high rating)?

2- What are the leading attributes customers look for in restaurants?

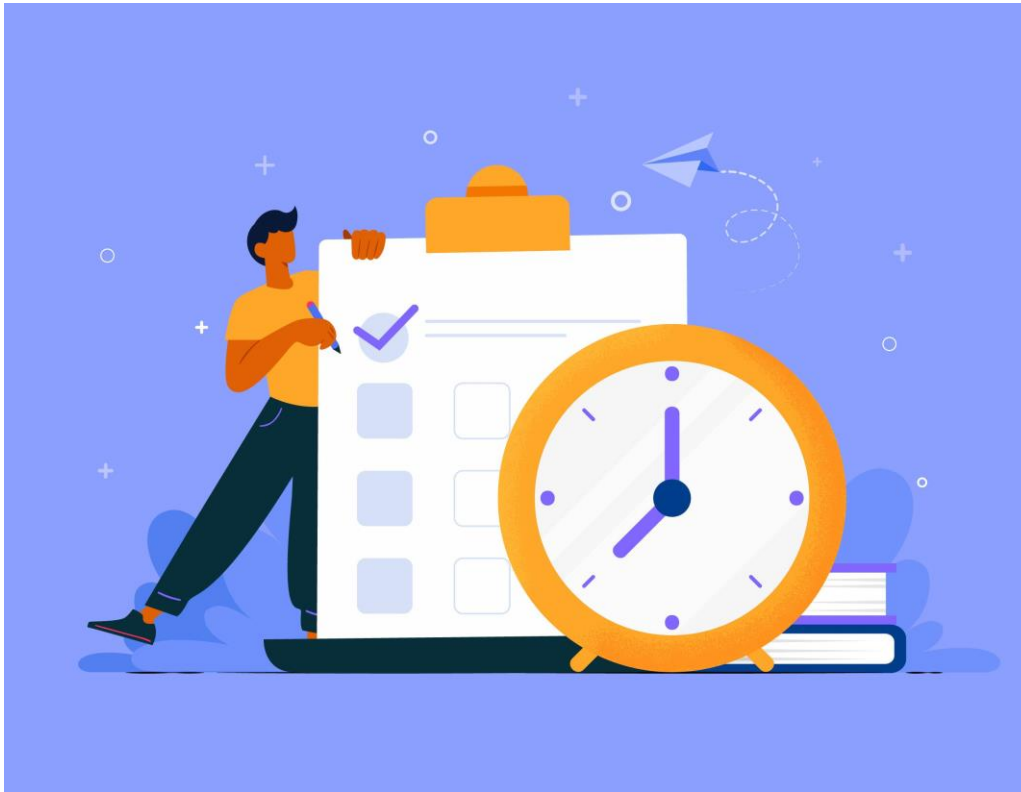
3- What are the specific types of food that successful restaurants specialize in?





Initial Hypothesis

What I'm believing to be true



1- Pay with credit card, Wi-Fi, long open hours and suitable for kids are essential features in successful restaurant.



2- Affordable prices, Good location, Delicious taste, and Good service are attributes customers looking for.



3- Most successful restaurants specialize in burgers and pizza.





Data Analysis Approach

The process to get answers

Analysis Process

Filter out non-restaurant businesses and clean data of errors, nulls, and inconsistencies.



Using Python and ML to discover trends in high-rated restaurants' features.

Using Python and machine learning techniques to identify common consumer advice and areas for business improvement.

Compiling a dataset consisting of the 1,000 highest-ranked restaurants for the purpose of conducting feature analysis.

Cleaning and processing a restaurant review dataset to identify common customer complaints and preferences and analyze patterns in the data.



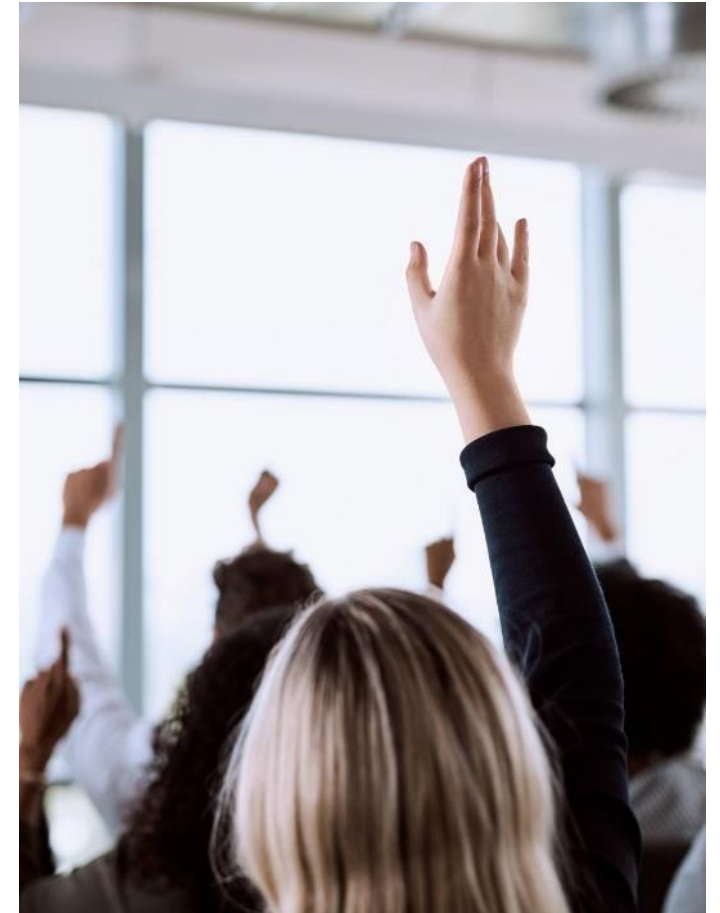
The way to get started is to quit talking and begin doing.

Walt Disney



Summary

This project is going to be both educational and helpful. I plan to apply the techniques and skills that I have learned either from this course or by myself to solve certain problems. I believe that this project will be a great addition to my portfolio.

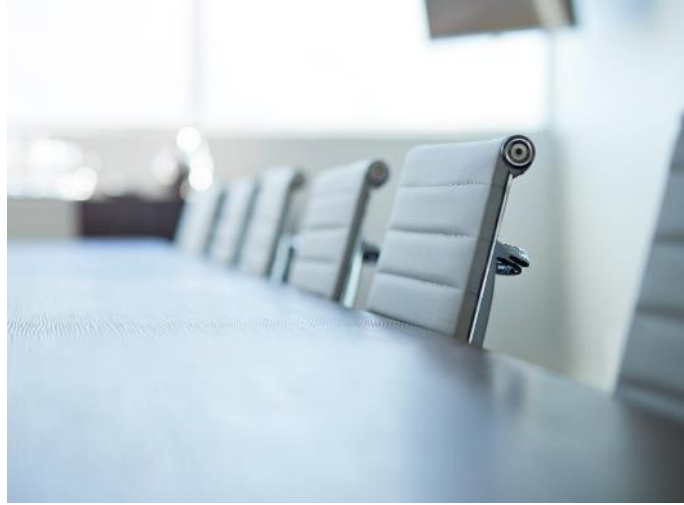


The background image shows a modern office environment with wood-paneled walls. In the foreground, a group of people are seated around a table, engaged in a meeting. A woman with glasses and a blue top is gesturing while speaking to a woman in a dark patterned top. Another person is partially visible behind them. In the background, a man in a blue shirt and a woman in a white shirt are seated at a desk, looking at a large monitor displaying data. The scene is lit with warm, natural light, and there are two black pendant lights hanging from the ceiling.

Presenting Analysis Findings

Mohammad Shkeer

Dec 10th, 2023



Agenda

Initial Findings

Deeper Analysis

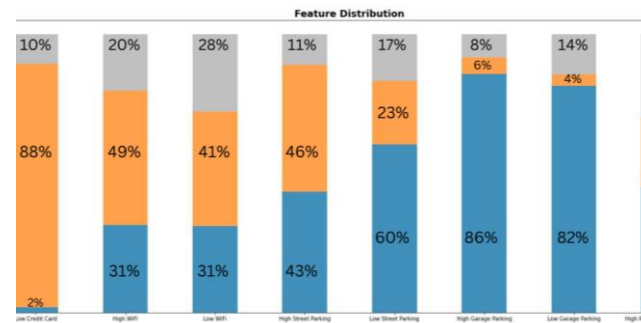
Hypotheses results

Recommendations and actions

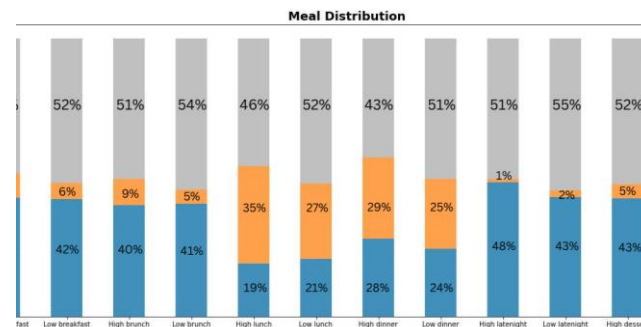
Appendix

Initial Findings

Premier examination of the visualizations.

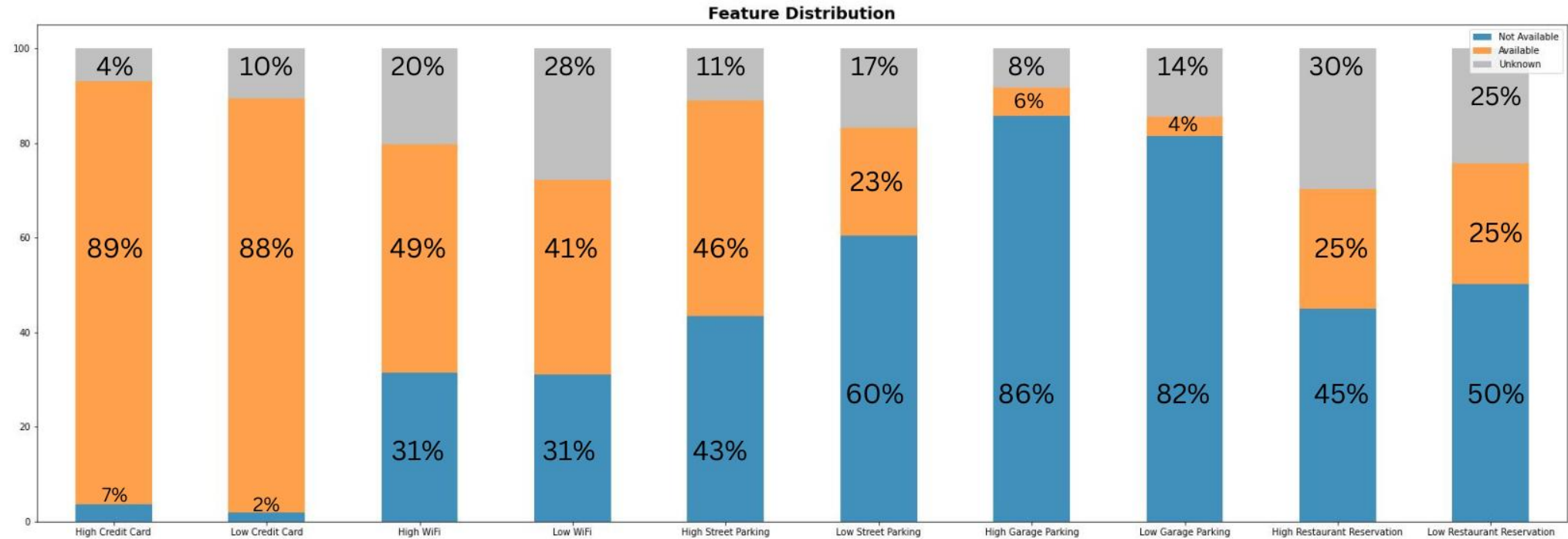


between High and low-rated restaurants.



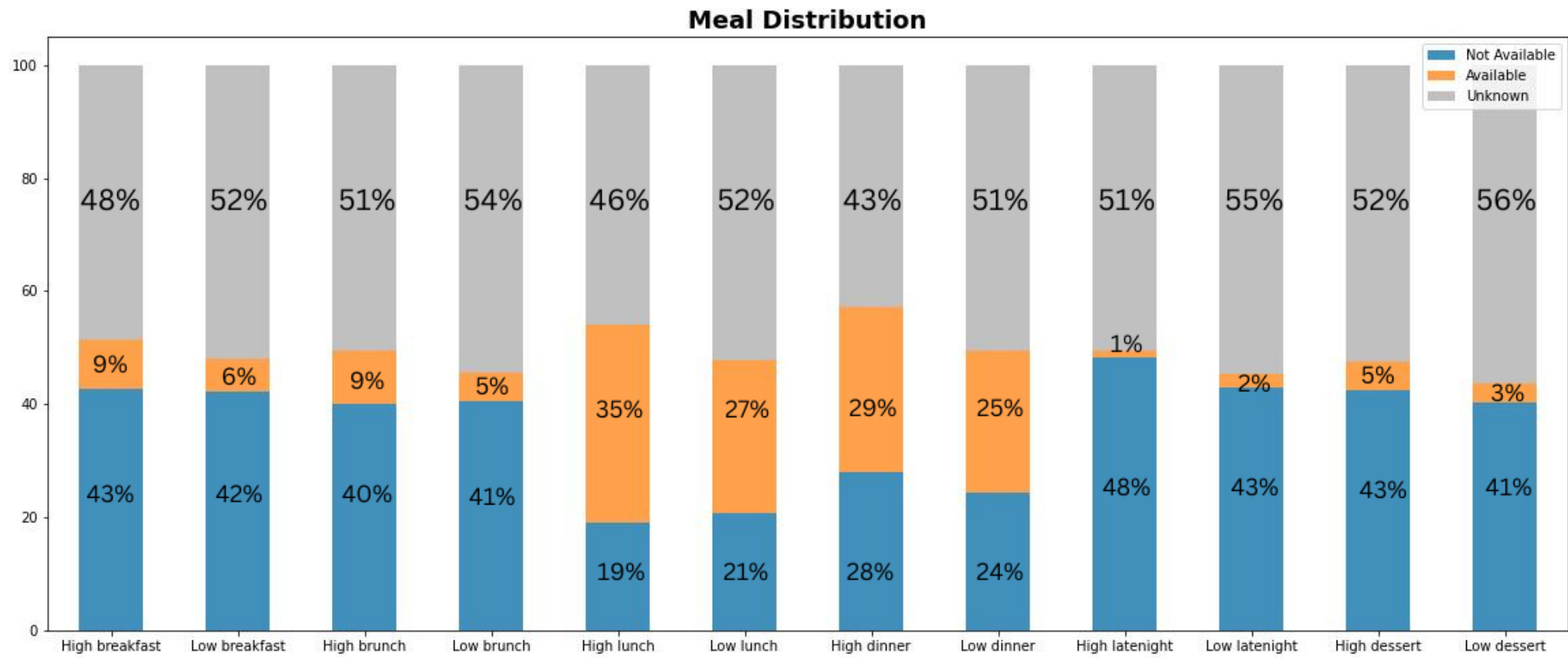
omparison between High and low-rated restaurants.





Showing a comparison between High and low-rated restaurants.

There is a positive correlation between restaurant ratings and the availability of features such as Wi-Fi, credit card payments, and parking.

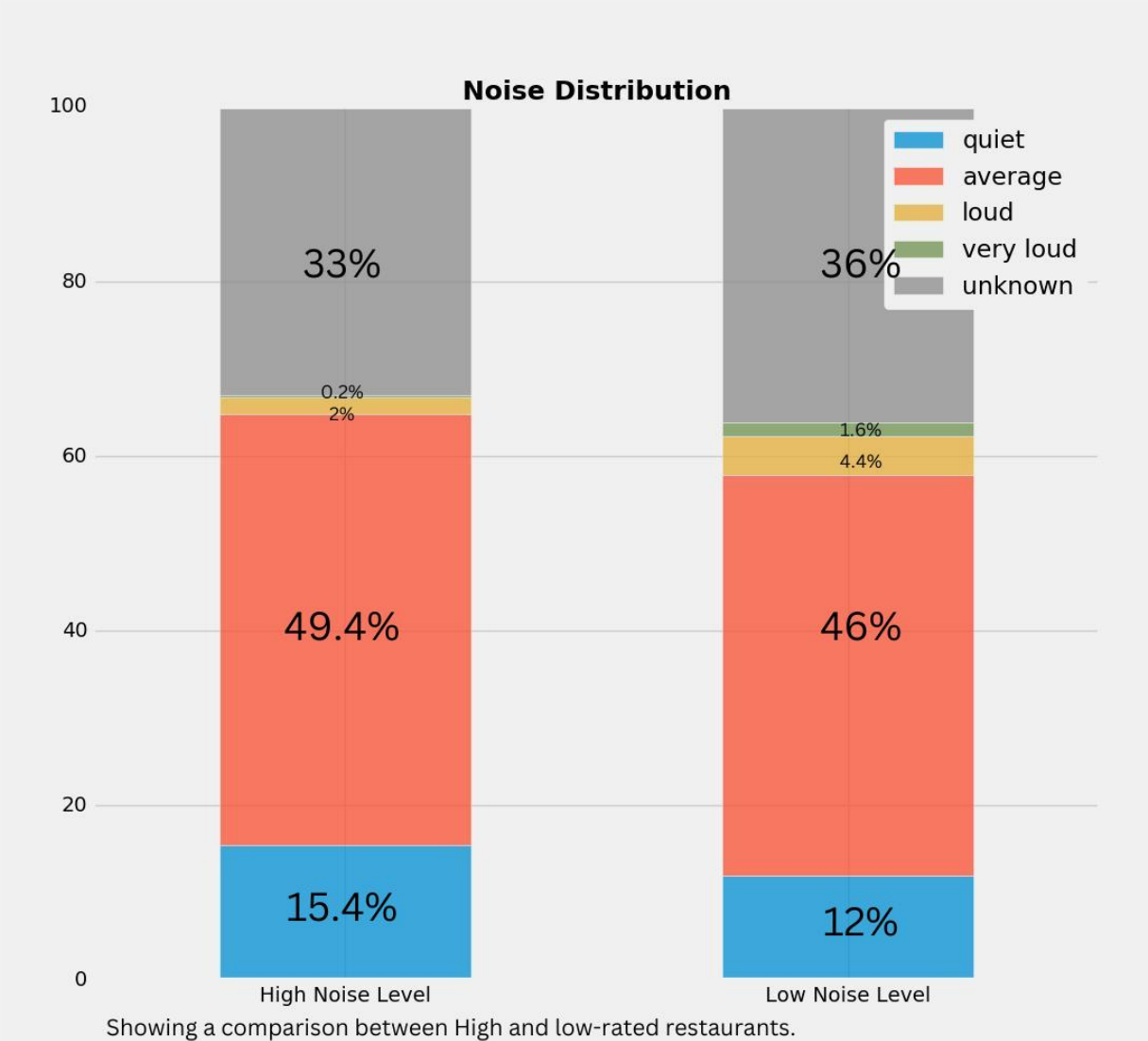


Showing a comparison between High and low-rated restaurants.

Restaurants with high ratings are more likely to serve breakfast, brunch, lunch, dinner, and desserts.

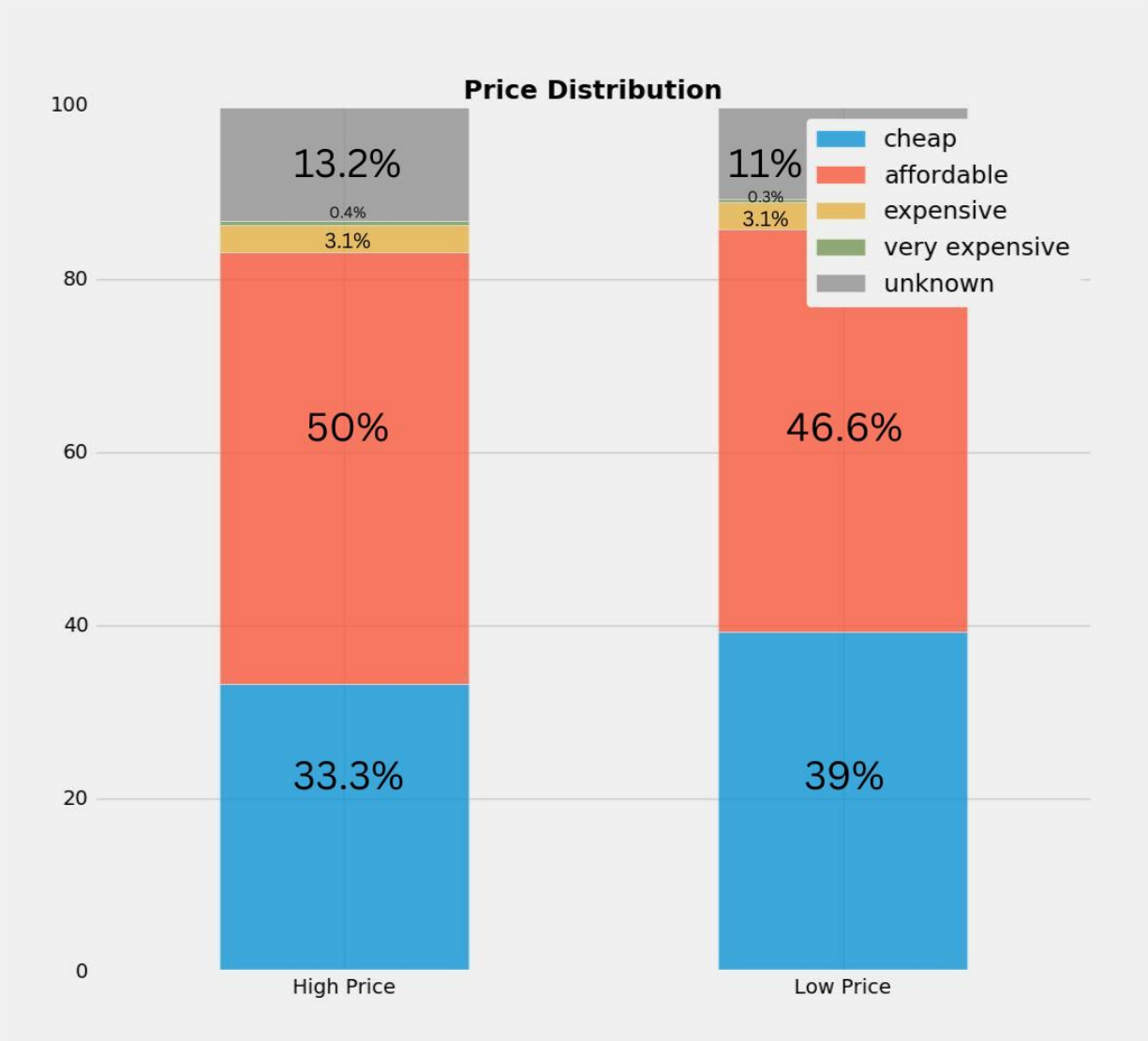
Noise Level

As expected, low-rated restaurants tend to be louder.



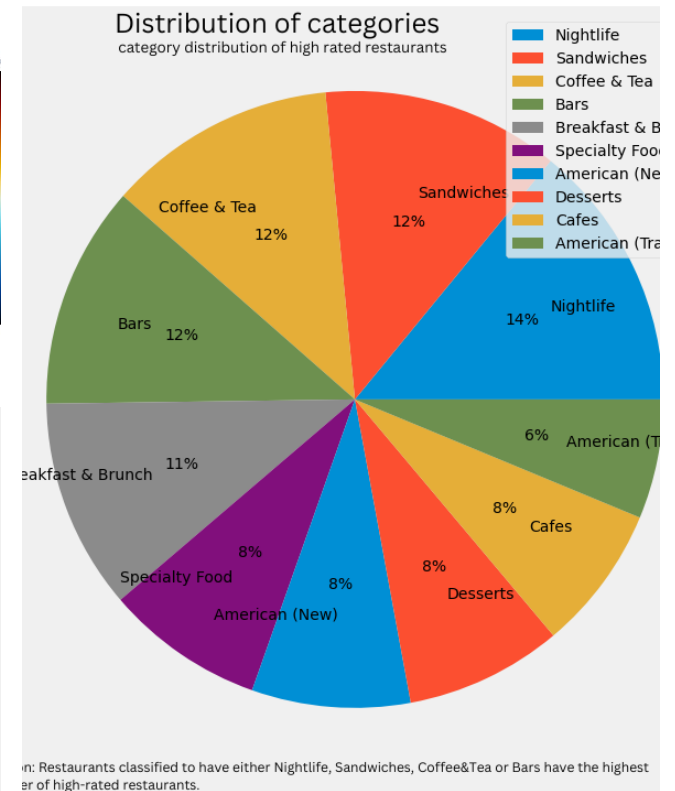
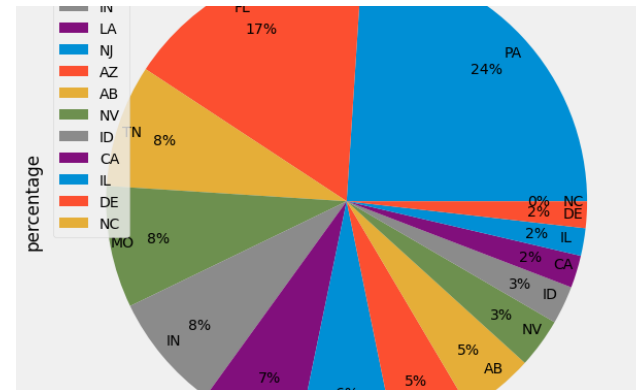
Price

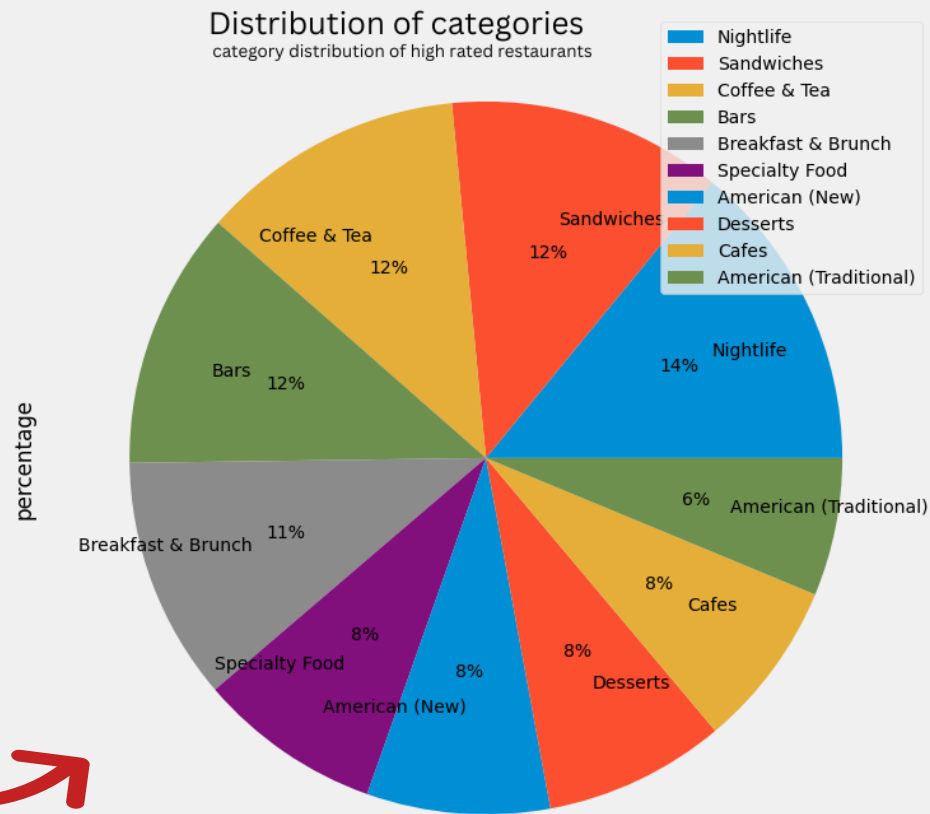
Surprisingly, there wasn't a big difference in food costs between high and low-rated restaurants.



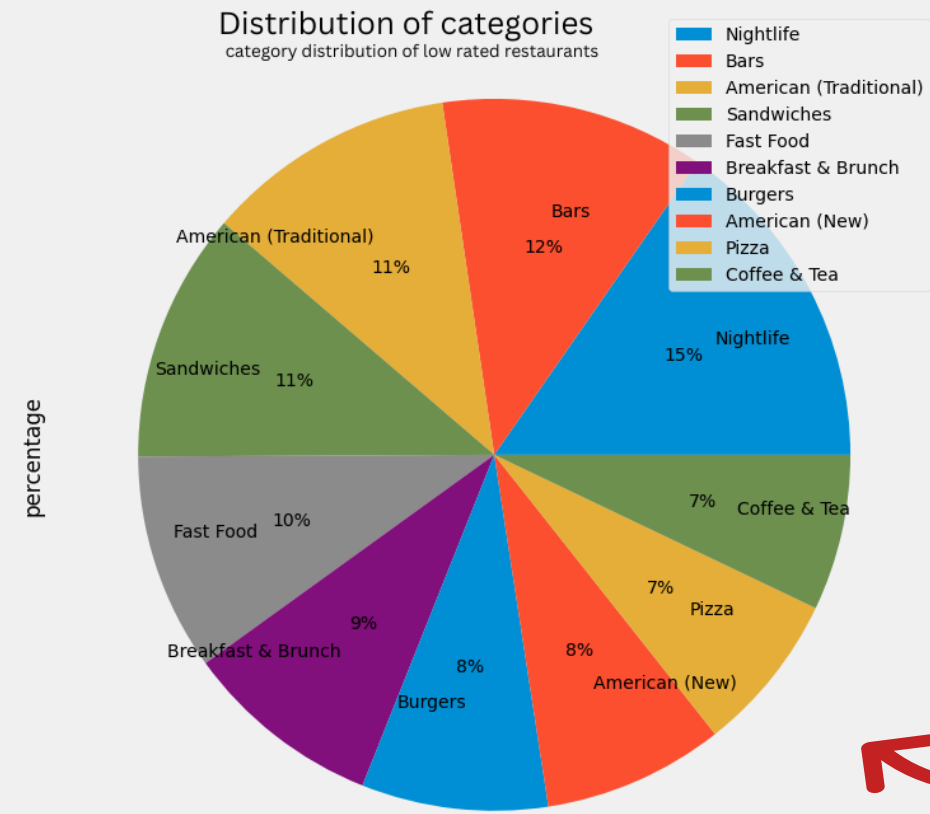
Deeper Analysis

Premier examination of the visualizations.



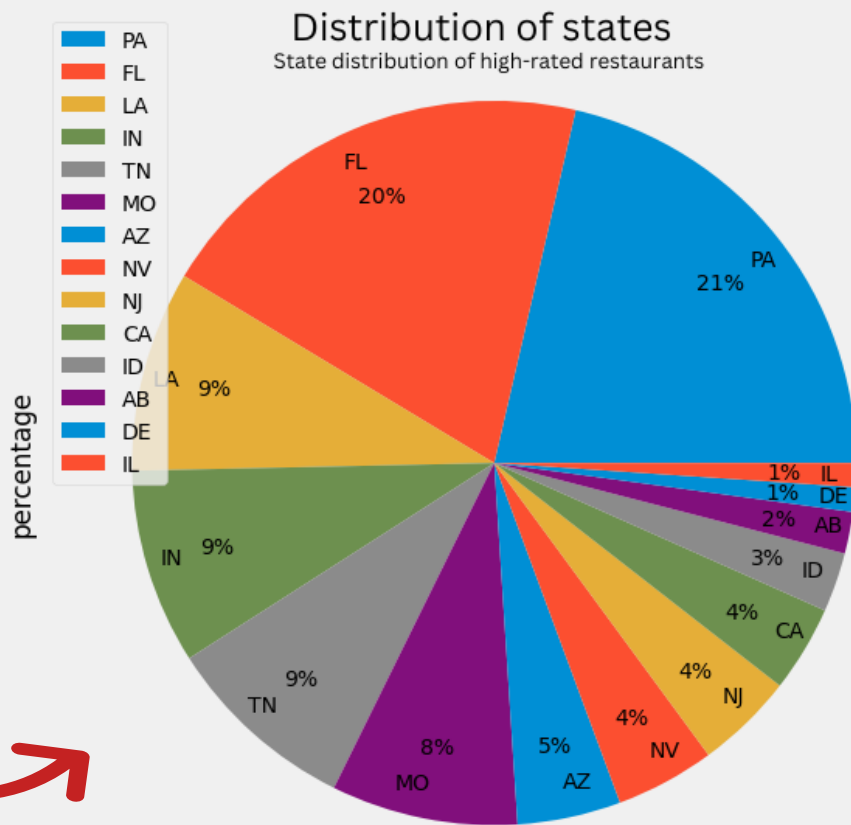


caption: Restaurants classified to have either Nightlife, Sandwiches, Coffee&Tea or Bars have the highest number of high-rated restaurants.



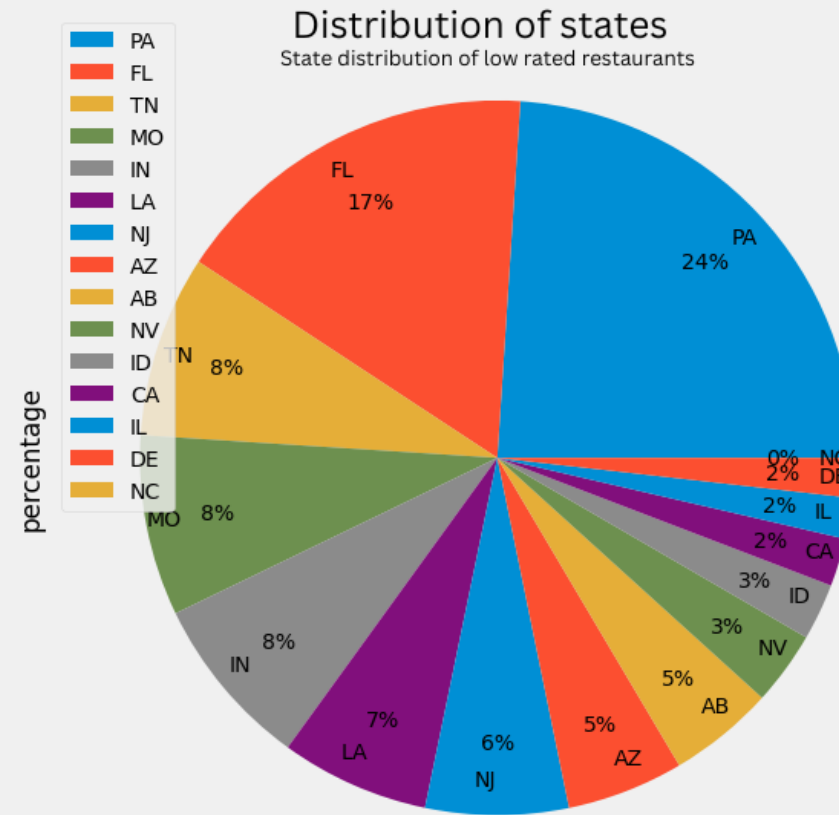
caption: Restaurants classified to have either Nightlife, Bars, American-Traditional or sandwiches have the highest number of low-rated restaurants.

Highly rated restaurants tend to offer specialty dishes and desserts, whereas low-rated restaurants typically serve more traditional and everyday food.



High-Rated

caption: Pennsylvania and Florida have the highest number of high-rated restaurants.



Low-Rated

caption: Pennsylvania and Florida have the highest number of low-rated restaurants.

The ranking of states by prevalence was not that different in both groups.
But what about cities?



High-Rated



Low-Rated

New Orleans and Santa Barbara have the second highest concentration of highly-rated restaurants, respectively, following Philadelphia.



Hypotheses results

What turned out to be true and what did not?

1- Pay with credit card, Wi-Fi, long open hours and suitable for kids are essential features in successful restaurant.

- This hypothesis was proven correct: restaurants with more features are more likely to have high reviews.

2- Affordable prices, Good location, Delicious taste, and Good service are attributes customers looking for.

- There is no correlation between affordable prices and ratings, and "good location" is subjective. Therefore, the hypothesis is incorrect.

3- Most successful restaurants specialize in burgers and pizza.

- It turns out that the opposite is true. Low-rated restaurants often serve traditional, everyday food.





Recommendations and actions

What steps should be taken?

1- Besides providing good quality food and taste, it is equally important to ensure customer comfort. To enhance the customer experience, you can provide additional features such as Wi-Fi, parking facilities, credit card payment options, and kids-friendly areas.

2- Offering meals for all times of the day- breakfast, brunch, lunch, and dinner.

3- The market is abundant with traditional food. It is crucial to serve specialty dishes and desserts.

4- Try to sustain a quiet environment, playing music in the background might be relaxing and a good solution.



Appendix


- Data: All the visualizations, insights, and recommendations are derived from the "Processed data".
- Processed Data: A filtered subset of the "Yelp Academic Dataset" meeting certain criteria.
- Filter criteria: I have applied machine learning to extract food-related businesses from a pool of over 150,000 businesses. To further refine my search, I have filtered out businesses that lack values in either the "attribute" or "hours" columns. For high-rated restaurants, I have set the criteria as a rating of 4.5 or higher and a review count of 30 or more. Any restaurant that doesn't meet these criteria is considered a low-rated restaurant.





Thank you

 Mohammad Shkeer

 Mohammad.shkir1@gmail.com