

Milestone 4: Presenting Your Findings

...

May 15, 2022
Choo Chang Hong

Content

Overview

Review of Questions to Answer / Hypotheses / Approach

Discuss Technical Challenges

Entity Relationship Diagram (ERD)

Initial Findings

Deeper Analysis

Hypotheses Results

Overview

CoffeeKing is a new startup coffee company providing a unique and novel experience to their customers. They want to appeal to a wide variety of clientele. The analysis will use dataset from Yelp, a crowd-sourced reviews about businesses to provide insights such as location selection and operating hours to CoffeeKing.

Questions to Answer

1. Where should CoffeeKing open their shop?
2. What is the preferred operating hour for CoffeeKing?
3. What are the key features CoffeeKing should provide to attract customers?

Initial Hypothesis

1. Coffee shop will have higher reviews count and rating in urban than in rural city.
2. Coffee shop operating from morning until afternoon generally.
3. A coffee shop should have high quality coffee, foods, and pleasant environment.

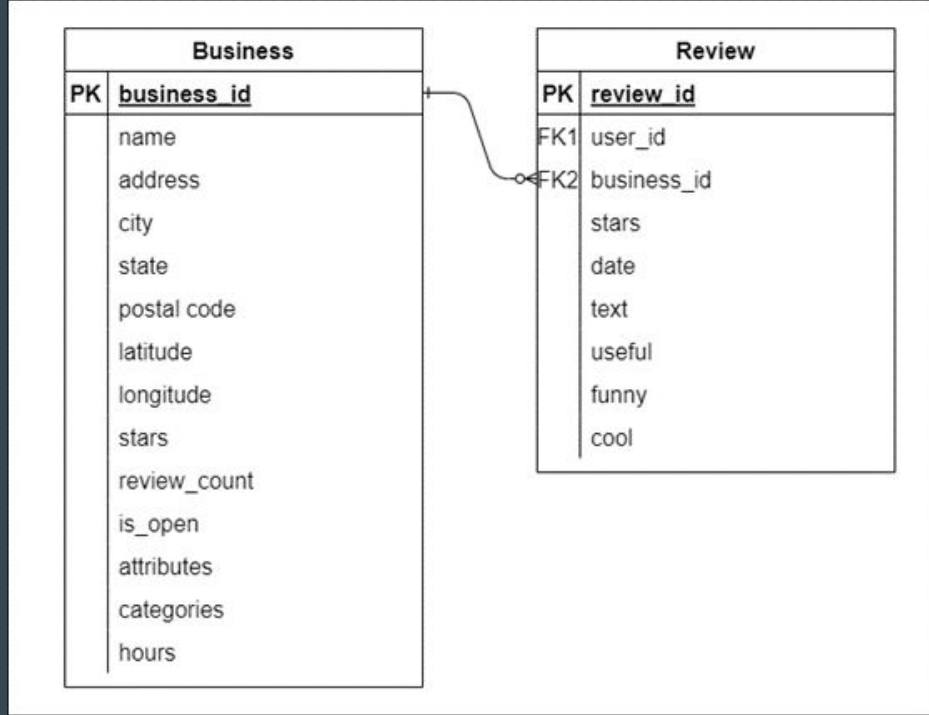
Data Analysis Approach

1. I will be looking primarily at features such as location and operating hour, to explore whether there is a relationship between the features and the review ratings.
2. Look at word counts in reviews to understand the context of different ratings.

Technical Challenges

1. Encounter challenges in reading the json data into Jupyter notebook as the Yelp data is huge. Therefore, data was pre-processing before read into Jupyter.
2. Limitation of pandasql to perform the SQL query on the dataset but manageable.

Entity Relationship Diagram



The initial dataset having 5 tables, and only this 2 tables are needed for the analysis.

Initial Findings

Found that initial hypotheses were evident in the data, however, a few key data points needed to be investigated further.

Query result on the right shows urban city having higher reviews than rural city.

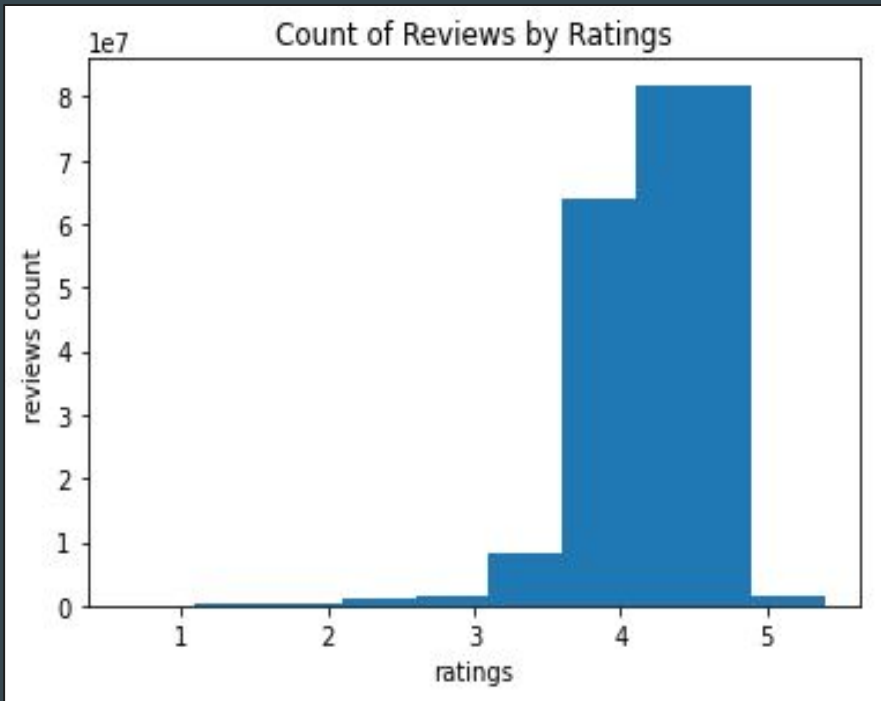
	city	total_review
0	Philadelphia	85155
1	Tampa	56429
2	New Orleans	41923
3	Nashville	31388
4	St. Louis	24925
5	Fishers	23054
6	Santa Barbara	20251
7	Sparks	19637
8	Tucson	19514
9	Pennsville	9441
10	Boise	8815
11	Edmonton	6287
12	Claymont	2523
13	Belleville	2292

Initial Findings (part 2)

	hours	total_review
0	{'Monday': '0:0-0:0', 'Tuesday': '0:0-0:0', 'Wednesday': '0:0-0:0', 'Thursday': '0:0-0:0', 'Friday': '0:0-0:0', 'Saturday': '0:0-0:0', 'Sunday': '0:0-0:0'}	18048
1	{'Monday': '8:0-18:0', 'Tuesday': '8:0-18:0', 'Wednesday': '8:0-18:0', 'Thursday': '8:0-18:0', 'Friday': '8:0-18:0', 'Saturday': '8:0-18:0', 'Sunday': '8:0-18:0'}	6647
2	{'Monday': '7:0-14:0', 'Tuesday': '7:0-14:0', 'Wednesday': '7:0-14:0', 'Thursday': '7:0-14:0', 'Friday': '7:0-14:0', 'Saturday': '7:0-14:0', 'Sunday': '7:0-14:0'}	4038
3	{'Monday': '7:0-15:0', 'Tuesday': '7:0-15:0', 'Wednesday': '7:0-15:0', 'Thursday': '7:0-15:0', 'Friday': '7:0-15:0', 'Saturday': '7:0-15:0', 'Sunday': '7:0-15:0'}	3676
4	{'Monday': '8:0-14:0', 'Tuesday': '8:0-14:0', 'Wednesday': '8:0-14:0', 'Thursday': '8:0-14:0', 'Friday': '8:0-14:0', 'Saturday': '8:0-14:0', 'Sunday': '8:0-14:0'}	3225

Data shows that coffee shop with higher reviews count generally operate from morning to afternoon.

Deeper Analysis



The graph shows that there is a left-skewed distribution, with most review ratings being at 4 to 4.5.

Deeper Analysis (part 2)

	city	avg_rating	total_review
0	Bryn Mawr	3.089744	624
1	Brandon	3.356757	925
2	Plant City	3.471795	585
3	Brentwood	3.477364	1502
4	Gretna	3.479032	620
...
61	Media	4.185567	1164
62	Saint Petersburg	4.317308	7904
63	Clayton	4.356338	710
64	Madeira Beach	4.389573	1055
65	St. Pete Beach	4.705357	672

The query is conditioned so that only cities with more than 500 reviews will be extracted.

Data shown that Bryn Mawr is the city that has lowest average rating of 3.09.

Deeper Analysis (part 3)

	hours	avg_rating	review_count
0	{'Monday': '8:0-18:0', 'Tuesday': '8:0-18:0', 'Wednesday': '8:0-18:0', 'Thursday': '8:0-18:0', 'Friday': '8:0-18:0', 'Saturday': '8:0-18:0', 'Sunday': '8:0-18:0'}	4.440349	33318333
1	{'Monday': '0:0-0:0', 'Tuesday': '7:0-15:0', 'Wednesday': '7:0-15:0', 'Thursday': '7:0-15:0', 'Friday': '7:0-17:0', 'Saturday': '7:0-17:0', 'Sunday': '7:0-15:0'}	4.000000	7316736
2	{'Monday': '7:0-21:0', 'Tuesday': '7:0-21:0', 'Wednesday': '7:0-21:0', 'Thursday': '7:0-21:0', 'Friday': '7:0-21:0', 'Saturday': '7:0-21:0', 'Sunday': '7:0-21:0'}	4.172897	4480866
3	{'Monday': '12:0-0:0', 'Tuesday': '11:0-23:45', 'Wednesday': '12:0-0:0', 'Thursday': '12:0-0:0', 'Friday': '12:0-0:0', 'Saturday': '12:0-0:0', 'Sunday': '12:0-0:0'}	4.000000	4330200
4	{'Monday': '8:0-14:0', 'Tuesday': '8:0-14:0', 'Wednesday': '8:0-14:0', 'Thursday': '8:0-14:0', 'Friday': '8:0-14:0', 'Saturday': '8:0-14:0', 'Sunday': '8:0-14:0'}	4.090698	4288555

The query extracted operating hours with review count and average review ratings.

Data shown that operating hour from 6am - 6pm daily has the highest review count and average rating.

Deeper Analysis (part 4)



The word count for review rating more than or equal to 4 is represented visually on the left hand side.

The height of each word in the picture is an indication of frequency of occurrence of the word in the entire review text.

Deeper Analysis (part 5)



The word count for review rating **less than 3** is represented visually on the left hand side.

The height of each word in the picture is an indication of frequency of occurrence of the word in the entire review text.

Final Findings (Results of Hypotheses)

1. Where should CoffeeKing open their shop?

I suggest that we choose a location that has lower review rating for other coffee companies. This is because location with numerous coffee companies come with huge competition, and we can strike in the suggested location.

The results suggested Bryn Mawr, Brandon, and Plant City.

Final Findings (part 2)

2. What is the preferred operating hour for CoffeeKing?

The result suggested that 7 days a week, operate from 6am to 6pm is the suitable operating hour for CoffeeKing.

Final Findings (part 3)

3. What are the key features CoffeeKing should provide to attract customers?

The result suggested that a high rating coffee shop should have good coffee, delicious food, and excellent service.

Besides, CoffeeKing should avoid lengthy wait time and service delay as found out from the word occurrence in the low rating reviews.