

# Yelp Restaurant Review Analysis

**Created By:**

Madhura Jadhav

MSIS

## Table of Contents

1.	Introduction .....	3
2.	Data Sourcing .....	3
3.	Data Preprocessing.....	3
4.	Exploratory Data Analysis .....	5
	4.1 Review Length Analysis .....	5
	4.2 Sentiment based Word Clouds .....	5
	4.3 Top 5 Aspects - Adjectives .....	8
5.	Conclusion.....	8

## 1. Introduction

For any successful business, customer feedback on their scope of services is the biggest driver to thrive in the competitive market. A business (or type of business) can leverage such text reviews from customers to derive useful insights and then take actions to make improvements and hence, enhance their quality.

For this project, exploration and analysis of Yelp reviews is performed for Indian-cuisine restaurants, using Python and Tableau. The main purpose of our project is to conduct thorough analysis on Indian restaurants in Boston and nearby area based on their customer reviews and try to figure out answers for questions like what makes a good Indian restaurant? What concerns customers?

Following 3 questions are explored as a part of this project:

1. **Review Length Analysis:** Explore review data to gain insights about customer behavior based on the length of the review they are writing.
2. **Sentiment based Word Clouds:** Determine Positive and Negative sentiments for the reviews and develop word clouds to highlight the important words mentioned in good and bad reviews.
3. **Top 5 Aspects:** Discover the top aspects from reviews along with the popular adjectives associated with them.

## 2. Data Sourcing

The dataset is obtained from 'Yelp' website in JSON format. For this project, following two files are utilized:

- **Business** - Contains business data including location data, attributes, and categories.
- **Review** - Contains full review text data including the *business\_id* the review is written for.

Other way of obtaining data was through Yelp API. However, the restrictions of limited calls/day as well as just 3 reviews per call for a specific requested business ID made it difficult, to accumulate adequate number of reviews for multiple businesses. Also, some of the review data had only JPG (image) links and some were duplicate responses. Hence, it was determined to go with dataset in JSON format available on Yelp website.

## 3. Data Preprocessing

To select the dataset for the following scope of interest, *business* data is filtered for location, category, and attribute:

- **Location** – Boston and its neighborhood (Boston, Somerville, Cambridge, Jamaica Plain, Brookline, Allston, Quincy)
- **Category** - Restaurant
- **Attribute** – Indian Cuisine

Based on star ratings, the filtered restaurants are graded as Top, Mediocre and Bottom and 10 restaurants for each grade were selected. For this purpose, list of 62 filtered restaurants is sorted based on number of stars and then the number of reviews. Post sorting, top 10 restaurants are graded as "Top", middle 10 as "Mediocre" and bottom 10 as "Bottom" and are being used in analysis phase.

"Grade" column, as shown below, is created for this purpose.

Unnamed: 0		business_id	name	address	city	state	postal_code	latitude	longitude	stars	review_count	is_open	attributes	categories	hours	Grade
0	35008	qsjwHTPpUof24ID8l-GvuQ	Punjab Cafe	653 Southern Artery	Quincy	MA	2169	42.252078	-70.988062	4.5	552	1	{'Caters': 'True', 'RestaurantsAttire': 'casu...	Pakistani, Indian, Restaurants	{'Monday': '11:30-20:0', 'Tuesday': '11:30-20:0'}	Top
1	81149	Q-_5cf30RqSxqD-_rjNFxA	Guru The Caterer	1295 Broadway	Somerville	MA	2144	42.405809	-71.130903	4.5	438	1	{'WheelchairAccessible': 'True', 'RestaurantsT...	Seafood, Indian, Restaurants, Halal, Event Pla...	{'Monday': '16:0-21:0', 'Tuesday': '16:0-21:0'}	Top
2	5572	CAfckJLcWHOaDRcQNiNzQ	Himalayan Kitchen	40 Bow St	Somerville	MA	2143	42.381526	-71.097896	4.5	205	1	{'Caters': 'True', 'HasTV': 'True', 'Restauran...	Restaurants, Pizza, Halal, American (New), Him...	{'Monday': '0:0-0:0', 'Tuesday': '11:30-22:0', ...}	Top
3	119519	njjacBMGySzM0sBsYYxUhQ	Punjabi Dhaba	225 Hampshire St	Cambridge	MA	2139	42.373883	-71.100787	4.0	935	1	{'Corkage': 'False', 'RestaurantsReservations'...	Pakistani, Restaurants, Indian, Caterers, Even...	{'Monday': '12:0-22:30', 'Tuesday': '12:0-22:30'}	Top
19	42195	ZA6kZBzb0qZvWf1T0AMGmA	Ghazal Fine Indian Cuisine	711 Centre St	Jamaica Plain	MA	2130	42.311367	-71.114610	3.5	209	0	{'RestaurantsTakeOut': 'True', 'OutdoorSeating...	Pakistani, Restaurants, Indian	{'Tuesday': '11:30-15:0', 'Wednesday': '11:30-15:0'}	Mediocre
20	126875	Lz8FSSRB_r9X8RI2YEy97w	Curry House	1335 Beacon St	Brookline	MA	2446	42.341787	-71.121644	3.0	199	1	{'RestaurantsTakeOut': 'True', 'OutdoorSeating...	Restaurants, Indian, Seafood, Vegetarian, Hala...	{'Monday': '0:0-0:0', 'Tuesday': '16:30-21:0', ...}	Bottom

For these 30 restaurants, *text* data from *reviews* file for respective *business\_ids* is extracted, as shown below.

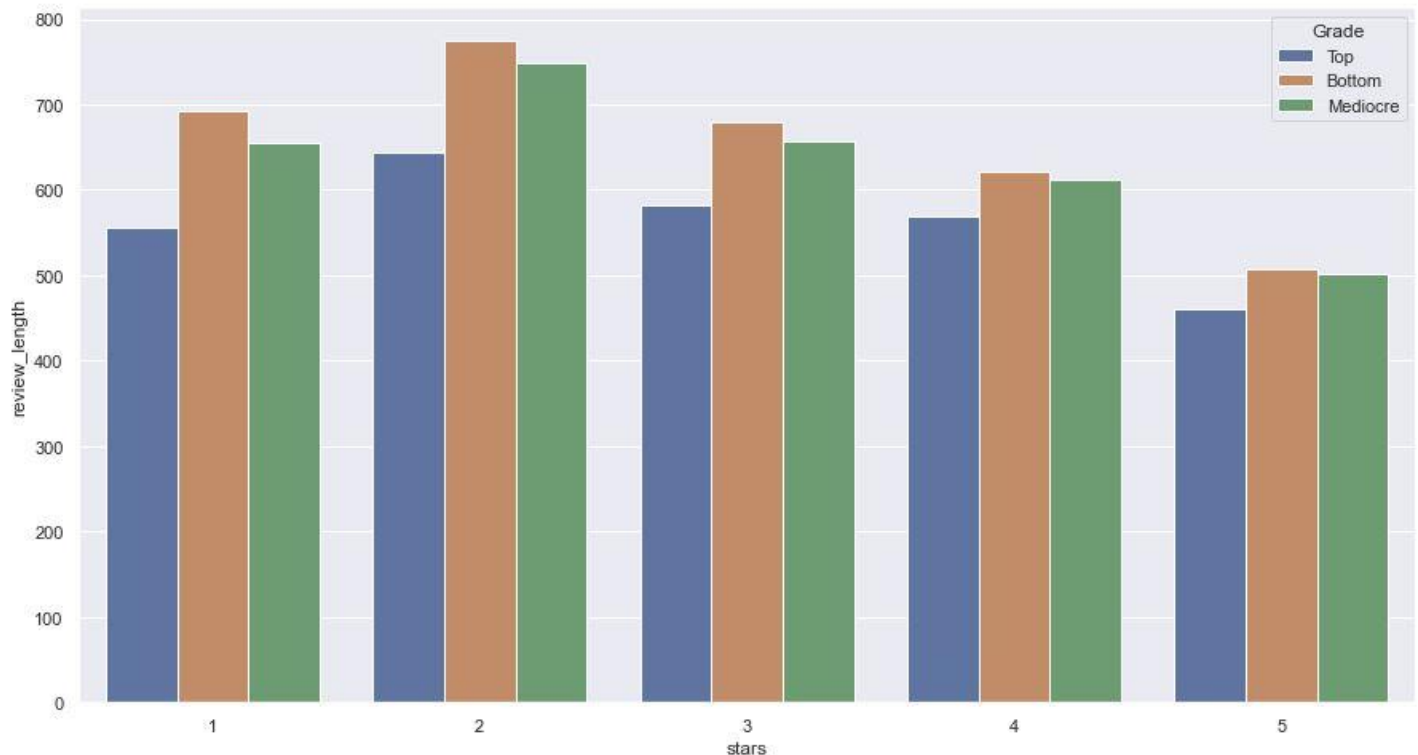
Unnamed: 0				review_id	user_id	business_id	stars	useful	funny	cool	text	date
0	0	Do3JEdVlqxiSoV_q0i0GHA	HpHkpOyCX0gcFuvbKYHpCg	C--wlpXJ4j1y01G_raHdbA	5	1	0	1	Wow. Thank you, Yelp. I'm SO glad I found th...	2007-12-08 03:53:22		
1	1	SjyigQPHeo5DEMT3Y0xXyQ	p8yQsVA51dzkc9cecDpvrw	HxegWRjhi7m73mXRI8qQlg	4	3	0	2	I mostly eat here at lunch, for the tremendous...	2007-05-17 13:48:57		
2	2	ckqQkgN_6E99JyrrpgCHlg	k1Qwwh_WqTbGby6rZNdg9Q	HxegWRjhi7m73mXRI8qQlg	5	5	1	2	As a Nepali person, I get asked the difference...	2016-03-17 14:21:35		
3	3	P5zpEmzurVIsQ394CsN5RQ	_97mTpKkuLB5x0UMladd7A	4UjU7F_EX3lgUtkzN8Bbrw	5	1	0	0	Ordered from here yesterday and loved their ch...	2016-02-11 23:36:46		
4	4	u1E-LQaRBcdhcjHGT7kkjw	S3nCmgSpJmsThVwAcf9Opw	4UjU7F_EX3lgUtkzN8Bbrw	4	1	0	1	I ordered a large order from here via foodler ...	2017-04-19 17:13:44		
...	...	...	...	...	...	...	...	...	...	...		
10345	10345	QzVsrfWUilS_sX6QnKzHZgg	BovkrYDODcINKe71da2qvw	tQbYsSvTAIV6uzZFmvdI2A	1	1	0	0	I'm sorry, They were nice enough but the food ...	2017-12-12 17:54:30		
10346	10346	2cPQnl9h6fzQCgLAAttg18A	0gCtnERzLtxa-0QRxXgPeg	L9S4WSPE3r1Bg9KglITqpA	3	2	0	0	We stopped here during our road trip last week...	2020-10-20 15:32:48		

Therefore, dataset for analysis has 10,350 reviews for 30 restaurants categorized in 3 different grades.

## 4. Exploratory Data Analysis

### 4.1 Review Length Analysis

For this analysis, review data is explored, to gain insights about customer behavior based on the length of each review.



Average Review Length per Star Rating for each Grade of the Restaurant

Above bar graph compares average review length against star ratings (1, 2, 3, 4, 5) for each grade of the restaurants (Top, Mediocre, Bottom).

**Insight:** The graph implies that a good dining experience might not require many words to describe. People tend to put down short reviews, for example: “Definitely recommend!” or “Could not be happier than my decision.” But the 1 or 2-star reviews tend to have longer text and probably contains more complaints.

### 4.2 Sentiment based Word Clouds

For this analysis, sentiments for the reviews are extracted and then word clouds are developed to highlight the important words mentioned in positive and negative reviews.

Firstly, text mining techniques are applied to clean the textual reviews data. Reviews’ text is converted to lower case, extra white spaces are stripped, special characters and punctuations are removed and lastly lemmatized the text. Additionally, removed NLTK library default English stop words and self-defined domain specific stop words based on the context. For example, words like “restaurant”, “indian” and “boston” are added in custom stop word list as they were bound to occur in almost every review due to the filtered data.

Secondly, on this cleaned data, calculated sentiment score for each review using TextBlob and categorized each review into positive or negative sentiment.

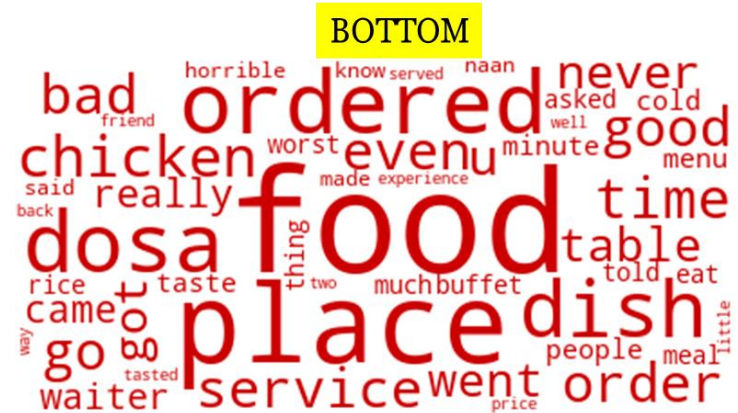
Lastly, word clouds for each grade of the restaurant were created to find insights.



### Positive Sentiment Word Clouds across different Grades

**Insight:** From the above word clouds for Positive sentiment reviews of Top, Mediocre and Bottom grade, it is clear that 2 most talked about words are 'food' and 'place', which is expected as we are talking about restaurants.





### Negative Sentiment Word Clouds across different Grades

**Insight:** From the above word clouds for Negative sentiment reviews of Top, Mediocre and Bottom grade, similar to positive reviews, 'food' and 'place' is still a priority for customers but in negative reviews customers are showing concerns about 'service' and also being more specific about items like 'chicken' and 'Dosa'.

### 4.3 Top 5 Aspects - Adjectives

For this analysis, extracted the top 5 aspects from each grade of the restaurant reviews, along with the top 5 popular adjectives associated with those aspects, by leveraging concepts learned in aspect-based analysis.

TOP	MEDIOCRE	BOTTOM
<pre>[[ 'place', ['good', 'great', 'best', 'favorite', 'amazing']], [ 'food', ['good', 'best', 'great', 'delicious', 'amazing']], [ 'chicken', ['good', 'delicious', 'great', 'best', 'spicy']], [ 'lunch', ['good', 'great', 'special', 'best', 'excellent']], [ 'time', ['good', 'many', 'great', 'several', 'second']]</pre> <pre>{ 'place': [[('good', 94), ('great', 82), ('best', 66), ('favorite', 58), ('amazing', 37)]], 'food': [[('good', 262), ('best', 226), ('great', 206), ('delicious', 109), ('amazing', 81)]], 'chicken': [[('good', 48), ('delicious', 25), ('great', 22), ('best', 17), ('spicy', 15)]], 'lunch': [[('good', 26), ('great', 23), ('special', 18), ('best', 10), ('excellent', 8)]], 'time': [[('good', 32), ('many', 26), ('great', 26), ('several', 20), ('second', 17)]]}</pre>	<pre>[[ 'lunch', ['good', 'great', 'favorite', 'old', 'best']], [ 'place', ['good', 'great', 'favorite', 'best', 'amazing']], [ 'food', ['good', 'great', 'best', 'delicious', 'fast']], [ 'time', ['several', 'many', 'great', 'long', 'good']], [ 'buffet', ['good', 'great', 'decent', 'delicious', 'tasty']]</pre> <pre>{ 'lunch': [[('good', 34), ('great', 18), ('favorite', 7), ('old', 6), ('best', 6)]], 'place': [[('good', 45), ('great', 44), ('favorite', 27), ('best', 18), ('amazing', 17)]], 'food': [[('good', 119), ('great', 82), ('best', 60), ('delicious', 42), ('fast', 35)]], 'time': [[('several', 21), ('many', 12), ('great', 11), ('long', 11), ('good', 11)]], 'buffet': [[('good', 42), ('great', 20), ('decent', 10), ('delicious', 6), ('tasty', 5)]]}</pre>	<pre>[[ 'food', ['good', 'great', 'chinese', 'bad', 'worst']], [ 'place', ['great', 'good', 'worst', 'best', 'bad']], [ 'lunch', ['good', 'great', 'small', 'whole', 'tasty']], [ 'time', ['long', 'good', 'second', 'rude', 'several']], [ 'buffet', ['good', 'small', 'much', 'great', 'extra']]</pre> <pre>{ 'food': [[('good', 80), ('great', 42), ('chinese', 26), ('bad', 20), ('worst', 20)]], 'place': [[('great', 27), ('good', 21), ('worst', 13), ('best', 10), ('bad', 8)]], 'lunch': [[('good', 14), ('great', 8), ('small', 5), ('whole', 4), ('tasty', 4)]], 'time': [[('long', 12), ('good', 9), ('second', 8), ('rude', 6), ('several', 5)]], 'buffet': [[('good', 13), ('small', 5), ('much', 5), ('great', 5), ('extra', 4)]]}</pre>

### Most Talked About Aspects for Each Grade of Restaurants

For each grade of restaurants, firstly, extracted top 5 aspects. This tells what customers are talking about.

Secondly, for these aspects, looked up 5 most common adjectives used to describe them. Also, calculated how many times each of these adjectives is used for the respective aspects, in order to understand the tone of the customer when talking about a particular aspect for the restaurants.

## 5. Conclusion

Based on the analysis, it was observed that:

- Customers are likely to write longer reviews to rant about a bad experience than to rave about a good one. It is probably because customers tend to write specifics in bad reviews for example, describing specific dish on the menu, which makes the negative reviews wordier.
- Most talked about aspect/word/topic in a restaurant review, agnostic of good or bad, is “food” and “place”. It implies an obvious observation that, in a restaurant, customer cares the most about the type of quality of the food and the ambience or hygiene of the place (area of restaurant).