Project background

Research from Cornell University shows that 27% of new restaurants close in the first year, while nearly 60% close in three years. A key obstacle to success is yelp's lack of exposure. Good restaurants, customers will visit regularly, will have a higher number of scores and comments, which will improve the restaurant's position in the search ranking, and reserve buffer space for any bad luck. This project will discuss the status of New York restaurants

Analysis method

Yelp's Restaurant ranking is globally recognized and is one of the best examples of using public experience and comments. The answer we want to seek is: what are the key characteristics that affect the success of restaurant management? We believe that from the official data related to scoring provided by yelp, we will be able to identify important features that can be identified.

These key characteristics can be the inherent attributes of business operation, such as opening time, noisy environment, or subjective factors of customers. Through data visualization and prediction model, we explore the key features of yelp to achieve high scores.

Our goal is to identify the key attributes and features that can score high in yelp. Combined with data visualization, the impact attributes and characteristics of high business score are identified. Natural language processing is used to extract valuable information from user evaluation.

- Map: analyze the location of successful restaurants

- Business status: analyze whether restaurants in different blocks are open

- Price and rating: analysis of restaurants with different prices and their rating

Data preparation
Our data is a collection of foursqure and yelp from restaurants in New York.
The data includes the following items：

	categories	coordinate	display_ph	distance	id		image_url	is_closed	location	name	phone	price	rating	review_cou	transaction	url	cat
0	[{'alias': 'sha	{'latitude': 4	(212) 233-8	1066.509	joes-shangl	https://s3-r	FALSE	{'address1':	Joe's Shang	1.21E+10	$$		4	4696	[]	https://www
1	[{'alias': 'chi	{'latitude': 4	(212) 791-1	1020.196	tasty-hand-	https://s3-r	FALSE	{'address1':	Tasty Hand	1.21E+10	$		4	1280	['delivery', '	https://www
2	[{'alias': 'chi	{'latitude': 4	(212) 989-6	4183.124	buddakan-r	https://s3-r	FALSE	{'address1':	Buddakan	1.21E+10	$$$		4	3278	['restaurant	https://www
3	[{'alias': 'sha	{'latitude': 4	(718) 321-3	14909.39	nan-xiang-)	https://s3-r	FALSE	{'address1':	Nan Xiang)	1.72E+10	$		4	2664	[]	https://www

Because there are other special characters in the latitude and longitude information of the data, we need to clean them and readjust them:

{'latitude': 40.7146691442891, 'longitude': -7...
{'latitude': 40.7142, 'longitude': -73.99787}
{'latitude': 40.7422275387194, 'longitude': -7...