



Sentiment and Data Analysis On Restaurant Reviews Final Project

Group 2
COMP_SCI-5540-0003
Fall 2022



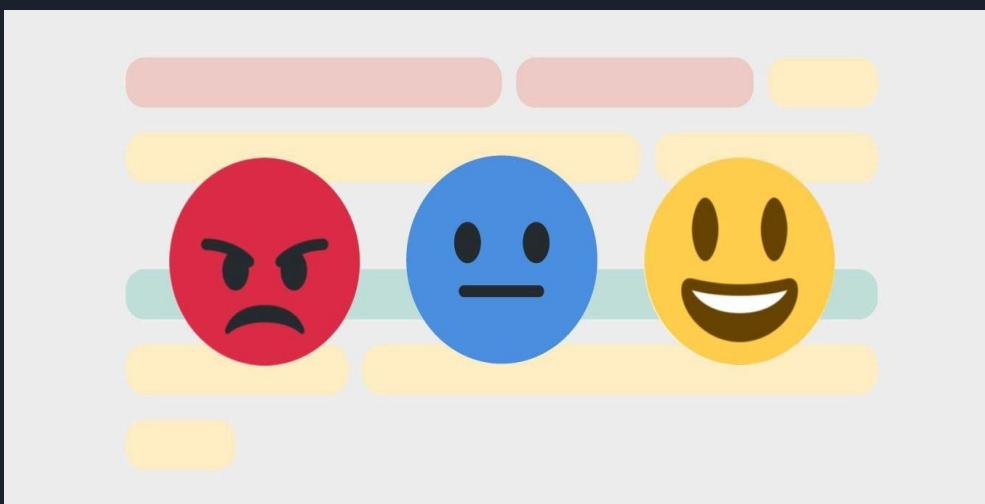
Project Recap

Dataset: TripAdvisor reviews for London restaurants

- 942.33 MB File
- 1,000,000 restaurant reviews
- Notable columns:
review_full,
rating_review,
restaurant_name



- 1) Develop a model that determines whether London restaurant reviews are positive or negative via Sentiment Analysis techniques.
- 2) Provide guidance to London restaurant goers by means of data analysis and visualization.





Previously Shown Data Cleaning for Data Analysis

1) Remove Inaccurately Named Restaurants

	ABC restaurant_name	Actions
1	2012,London_England,https://www.tripadvisor.com/Restaurant_Review-g186338-d2241866-Reviews-or420-Luna_Nuova-London_England	
2	pizza as good as I hoped. Will 100% recommend and go back.10/10,December 5	
3	uncomplicated Naples style pizza served in a cosy atmosphere by friendly staff. Warm recommendations go out to this little slice of Naples,	
4	Fantastic Service	
5	super old-style fries	
6	drinks and food came quickly. Only then i did like was the Pizza came on a wooden plate 😊,February 9	
7	2019,London_England,https://www.tripadvisor.com/Restaurant_Review-g186338-d13113158-Reviews-or150-Dishoom-London_England.htm	
8	professional manner and swiftly taken to our table. Service was good and we were presented with both the wine list and the menu. We did h	
9	Staff was polite and service is...,We ordered paneer masala dosa	
10	the food was unforgettable and the dirty martini I'd the best I have ever had. I highly recommend you go and have the tasting menu. ,Octobe	
11	courteous and wanted to go extra mile .Great...,Okra starter was so delicious that I finished in my first few minutes . Head Chef Sameer	
12	as we were there late afternoon we were not allowed the main menu. Staff were friendly helping us choose from the menu and letting us tas	
13	fantastic portions and fantastic price.What makes it even better though is the lovely family who run it - they are so friendly and attentive!,Fe	
14	,After a three month wait	
15	by Gordon Ramsey	
16	but added value is the attentiveness and obvious care taken and the trick of attention to detail without fussiness so that we all had a great t	
17	the food and drinks lovey and the vibe in tier 2 London this weekend spot on. Good luck to all venues like this in difficult times ★,October 2	
18	2020,London_England,https://www.tripadvisor.com/Restaurant_Review-g186338-d16733167-Reviews-or10-Kanishka_by_Atul_Kochhar-Lo	



2) Remove Null Restaurant Names

```
df = df.dropna()  
df.show(truncate = False, n = 5)
```

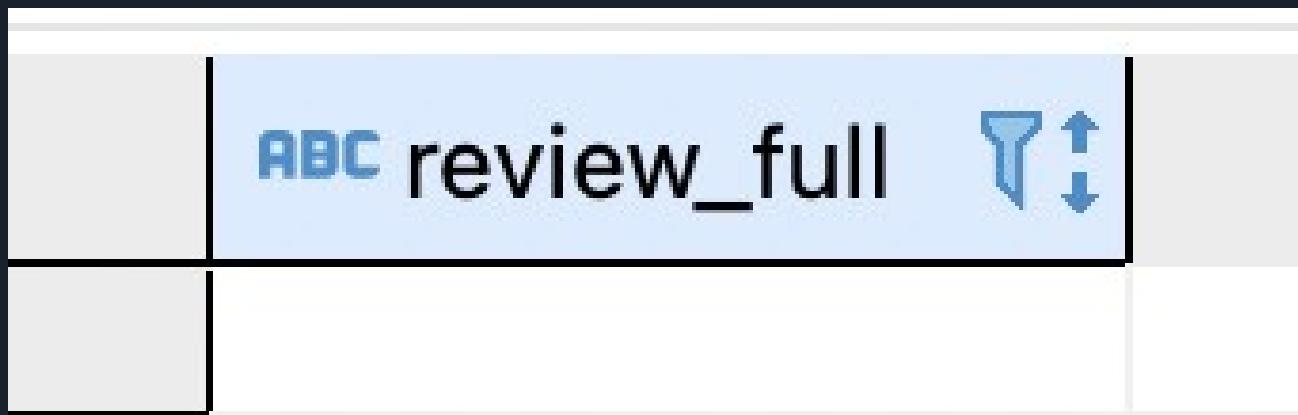
	ABC restaurant_name	↑↓
1	[NULL]	
2	[NULL]	
3	[NULL]	
4	[NULL]	
5	[NULL]	
6	[NULL]	
7	[NULL]	
8	[NULL]	
9	[NULL]	
10	[NULL]	
11	[NULL]	
12	[NULL]	
13	[NULL]	
14	[NULL]	
15	[NULL]	
16	[NULL]	
17	[NULL]	
18	[NULL]	



3) Remove Inaccurate City Names

	ABC city	Tools
1	London_England	
2	Good Food. Could be Great., You can't go wrong	

4) Check for Null Reviews



5) Check Review Rating Review Column

	123 rating_review	↑↓
1	5	
2	1	
3	4	
4	3	
5	2	



Previously Shown Data Cleaning for Sentiment Analysis

6) Drop Irrelevant Columns



	restaurant_name	rating_review	sample	review_full
	Cocotte_Notting_Hill	5.0	Positive	I was away for couple of months and I am so happy t
	Cocotte_Notting_Hill	5.0	Positive	I recently ordered over the phone from Cocotte to d
	Cocotte_Notting_Hill	5.0	Positive	Came here because we missed our table at Farmacy ad
	Cocotte_Notting_Hill	5.0	Positive	My first time in cocotte and was amazed by how frie
	Cocotte_Notting_Hill	5.0	Positive	A healthy-homemade dishes using farm-grown ingredie

only showing top 5 rows

punct

lower

restaurant_name	rating_review sample	review_full
Cocotte_Notting_Hill 5.0	Positive i was away for couple of months and i am so happy to be back for my	
Cocotte_Notting_Hill 5.0	Positive i recently ordered over the phone from cocotte to collect for a gat	
Cocotte_Notting_Hill 5.0	Positive came here because we missed our table at farmacy across the road th	
Cocotte_Notting_Hill 5.0	Positive my first time in cocotte and was amazed by how friendly this member	
Cocotte_Notting_Hill 5.0	Positive a healthyhomemade dishes using farmgrown ingredients really lovely	

only showing top 5 rows

stop



restaurant_name	rating_review sample	review_full
Cocotte_Notting_Hill 5.0	Positive [away, couple, months, happy, back, regular, cocotte, meal, collecting	
Cocotte_Notting_Hill 5.0	Positive [recently, ordered, phone, cocotte, collect, gathering, home, calling,	
Cocotte_Notting_Hill 5.0	Positive [came, missed, table, farmacy, across, road, waitress, accommodated, u	
Cocotte_Notting_Hill 5.0	Positive [first, time, cocotte, amazed, friendly, member, front, house, staff,	
Cocotte_Notting_Hill 5.0	Positive [healthyhomemade, dishes, using, farmgrown, ingredients, really, lovel	

only showing top 5 rows

Lemmatization

restaurant_name	rating_review	sample	review_full
Cocotte_Notting_Hill	5.0	Positive	away, couple, month, happy, back, regular, cocotte, meal, collect
Cocotte_Notting_Hill	5.0	Positive	recently, ordered, phone, cocotte, collect, gathering, home, cal
Cocotte_Notting_Hill	5.0	Positive	came, missed, table, farmacy, across, road, waitress, accommodat
Cocotte_Notting_Hill	5.0	Positive	first, time, cocotte, amazed, friendly, member, front, house, st
Cocotte_Notting_Hill	5.0	Positive	healthyhomemade, dish, using, farmgrown, ingredient, really, lov

only showing top 5 rows



New Cleaning and Modeling

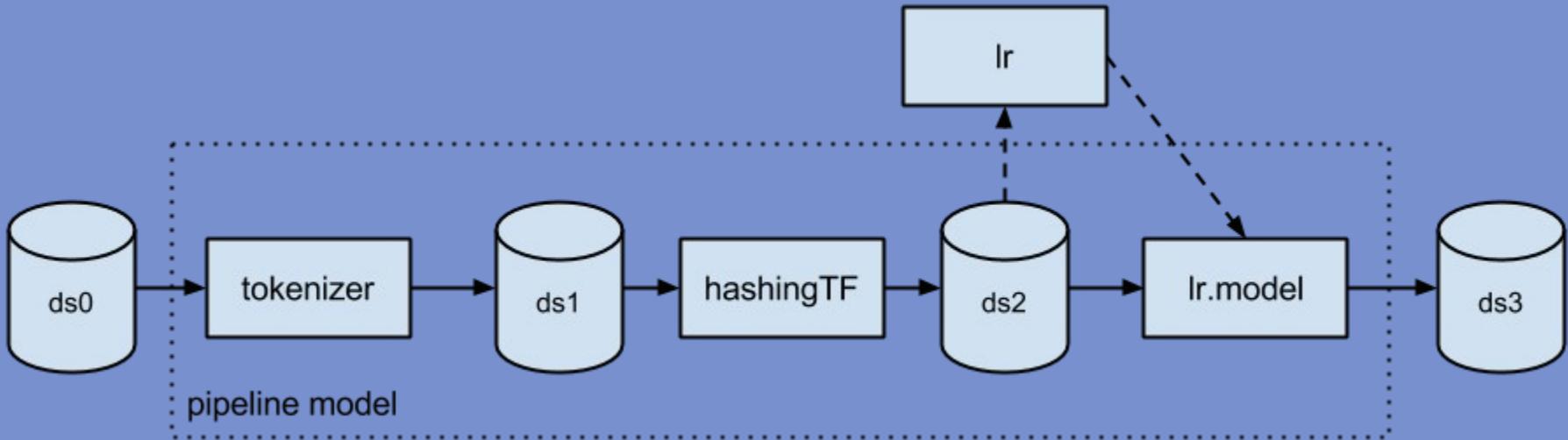
Sentiment Analysis Cleaning: Delete all rows with less than 5 restaurant reviews

```
+-----+  
|   restaurant_name|count(review_full)|  
+-----+  
|Merlot_Garden_Bar...|      5|  
|      Chotto_Matte|      5|  
|      Buns_and_Buns|      7|  
|Sweetsmile_Bakery...|      8|  
|Roxie_Steak_Earls...|      8|  
+-----+  
only showing top 5 rows
```

```
breaded chicken strips over cooked fried in old oil..." , 1  
Merlot_Garden_Bar_Restaurant , 5
```

Note: This approach not only allows for greater analysis opportunities, but also ensures that any inaccurately named column is removed. This step is slightly different than what was done for the data analysis cleaning/visualization not relevant to sentiment analysis, as this was a later discovery. For the data analysis in this project, we also wanted greater than 50 reviews for most cases and didn't want to delete single accurate reviews.

Future Analysis Steps



Our attempt

restaurant_name	review_full	rating_review	prediction
280_Degrees_Afric...	bar, man, one, co...	1.0	1.0
280_Degrees_Afric...	called, , degree,...	1.0	1.0
280_Degrees_Afric...	get, empty, resta...	1.0	1.0
280_Degrees_Afric...	god, bad, bad, ba...	1.0	1.0
280_Degrees_Afric...	ordered, food, on...	1.0	1.0
280_Degrees_Afric...	receiving, groupo...	1.0	2.0
280_Degrees_Afric...	visited, restaura...	1.0	2.0
280_Degrees_Afric...	watson, puff, puf...	1.0	1.0
280_Degrees_Afric...	arrived, hr, took...	2.0	2.0
280_Degrees_Afric...	booked, table, , ...	2.0	2.0
280_Degrees_Afric...	food, alright, ni...	2.0	3.0
280_Degrees_Afric...	restaurant, stand...	2.0	2.0
280_Degrees_Afric...	went, try, someth...	2.0	2.0
280_Degrees_Afric...	went, voucher, st...	2.0	2.0
280_Degrees_Afric...	bought, voucher, ...	3.0	3.0
280_Degrees_Afric...	came, deal, knowi...	3.0	3.0
280_Degrees_Afric...	day, ago, pm, aft...	3.0	3.0
280_Degrees_Afric...	dinner, recently,...	3.0	3.0
280_Degrees_Afric...	eatery, standard,...	3.0	3.0
280_Degrees_Afric...	groupon, worth, t...	3.0	3.0

Our attempt

restaurant_name	review_full	rating_review	prediction
Cahoots_Underground	first, ever, , st...	5.0	5.0
The_Monsoon	visit, monsoon, r...	5.0	5.0
Gaucho_Tower_Bridge	one, best, locati...	4.0	5.0
Park_Terrace_Rest...	park, terrace, ro...	4.0	4.0
Thali_Pickles_Ind...	tasty, food, serv...	5.0	5.0
L_Escargot_Restau...	last, time, ate, ...	3.0	3.0
Pollen_Street_Social	disappointing, oc...	3.0	3.0
Sketch_Lecture_Ro...	visited, restaura...	5.0	5.0
The_Hereford_Arms	researched, place...	5.0	5.0
Cosmoba	reservation, cosm...	2.0	4.0
Franco_Manca_Balham	luck, heart, balh...	5.0	5.0
GONG_Bar	, visited, drink,...	5.0	5.0
Bentley_s_Oyster_...	service, good, so...	3.0	3.0
Hunan	need, book, well,...	5.0	5.0
Mother_Mash	meal, potion, big...	4.0	4.0
Balthazar	great, restaurant...	5.0	5.0
Heist_Bank	food, nice, every...	4.0	4.0
O'Neill_s_Kings_C...	classic, oneills,...	3.0	4.0
SUSHISAMBA_Heron_...	husband, went, su...	5.0	5.0

Model Accuracy

restaurant_name	review_full	off_sent	pred_sent
The_Park_Room	whats, like, deli...	Positive	Positive
Dishoom	found, restaurant...	Neutral	Positive
Baltic_Restaurant...	boyfriend, took, ...	Positive	Positive
Hankies_Cafe	first, time, hank...	Positive	Positive
Sale_e_Pepe	every, time, come...	Positive	Positive
ASK_Italian	problem, getting,...	Positive	Positive
Cote_Brasserie_Soho	visited, meal, th...	Positive	Positive
Electric_Diner	love, nothing, gr...	Negative	Negative
Hache_Shoreditch	food, staff, amaz...	Positive	Positive
Spaghetti_House_W...	pleasant, atmosph...	Positive	Positive
Veeraswamy	wanted, go, somew...	Positive	Positive
Vietfood	lunch, venue, rec...	Neutral	Neutral
Franco_Manca	want, enjoy, best...	Positive	Positive
Honey_and_Smoke	friend, introduce...	Positive	Positive
The_Wellington	find, well, run, ...	Positive	Positive
Cote_Brasserie_Sl...	eaten, hear, plac...	Positive	Positive



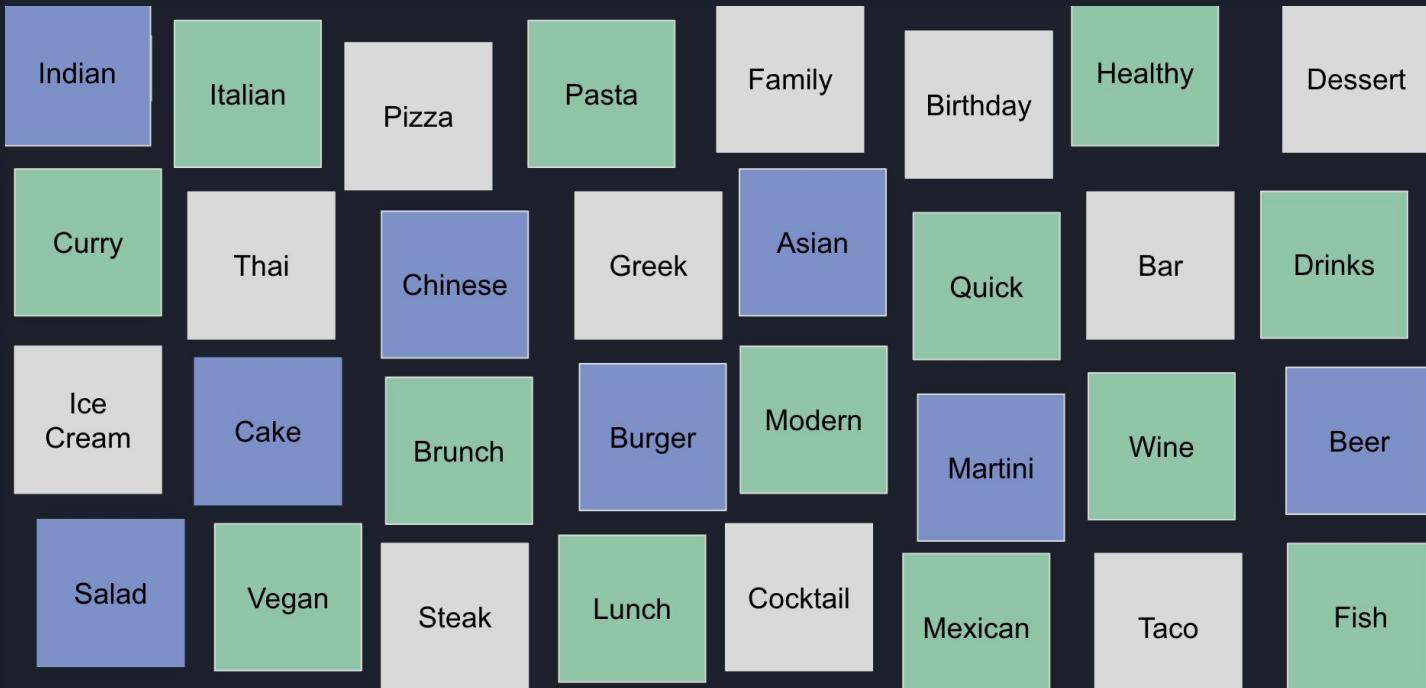
Data Analysis

Data Analysis Recap

Visualize distribution of rankings for each restaurant

Visualize # of reviews in each ranking

Extract data on various cuisines, types of food, or restaurant types





Extract data on
various cuisines,
types of food, or
restaurant types

The Original Query

Tacos

Enchiladas

avg(rating_review)	restaurant_name
5.0	Aqua_Nueva
5.0	Anglo
5.0	Busaba_Westfield_...
5.0	Berners_Tavern
5.0	AW_Restaurant
5.0	The_Blacksmith_s...
5.0	The_Good_Earth
4.9	Paladar
4.8	The_Fire_Station
4.8	Donnelly_s_Restau...
4.777777777777778	Farmacy
4.6923076923076925	May_Fair_Kitchen
4.676829268292683	Leicester_Square_...
4.6666666666666667	Yuu_Kitchen
4.615384615384615	COYA_Angel_Court
4.6	Dirty_Bones_Shore...
4.5	Momo
4.473684210526316	SUSHISAMBA_Heron_...
4.353535353535354	Jinjuu_Soho
4.317073170731708	COYA

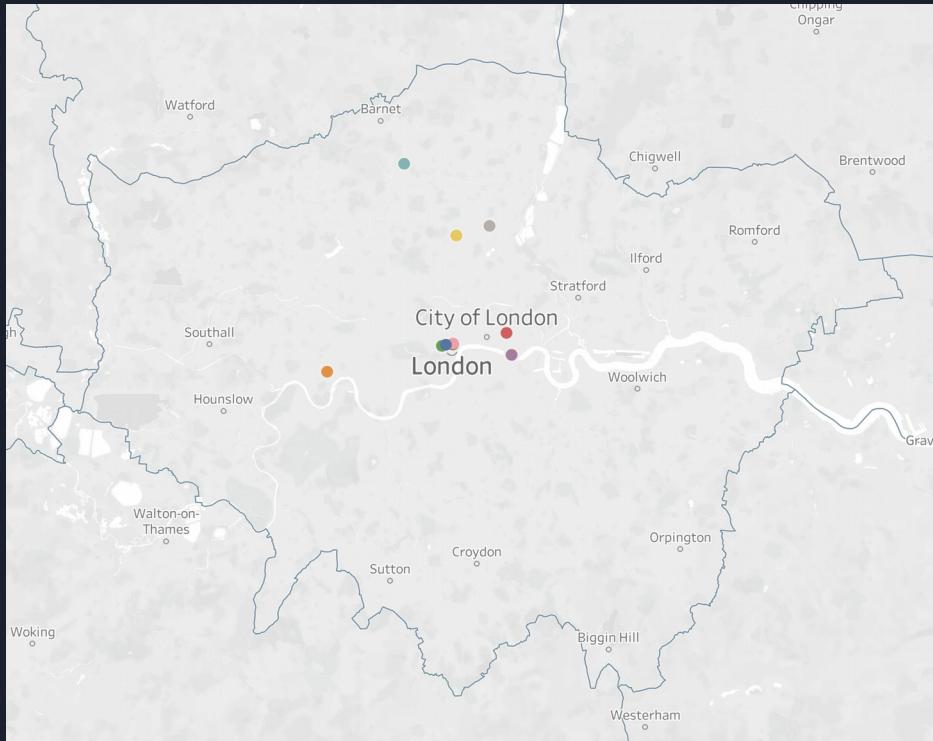
only showing top 20 rows



Solution: Require
>50 restaurant
reviews to be eligible

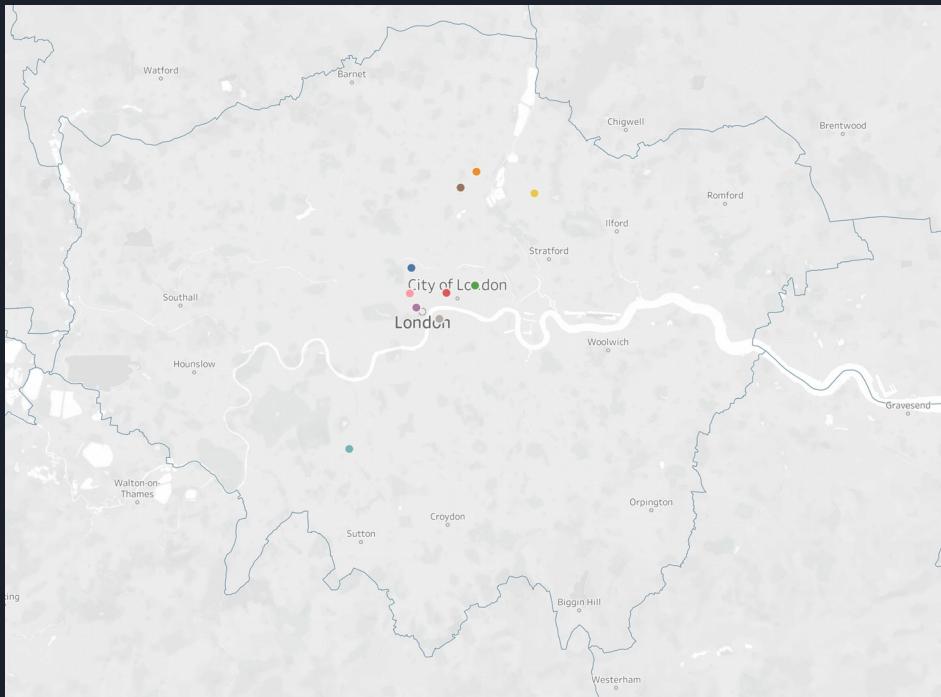
Steak

	123 avg(rating_review) ⚡️	RBC restaurant_name
1	4.6794871795	Smith_s
2	4.6376811594	Laurent_at_Cafe_Royal
3	4.6046948357	Steak_Co_St_Martins_Ln
4	4.5907335907	Blacklock_Soho
5	4.55	Melange
6	4.5471698113	Chez_Elles_Bistroquet
7	4.4793713163	Buenos_Aires_Argentine_Steakhouse
8	4.4493464052	Steak_Co_Garrick_Street_Covent_Gard
9	4.4234693878	The_Banc
10	4.2878787879	El_Vaquero_Whetstone



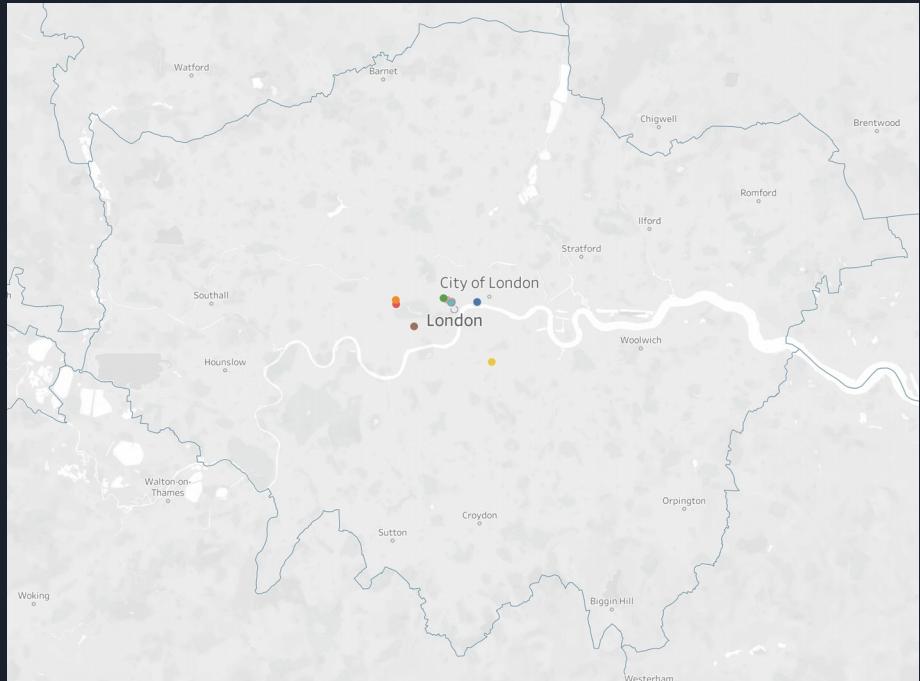
Burger

123	avg(rating_review)	⬇️	ABC	restaurant_name	⬆️
	4.9305555556			Bruce_Burgers	
	4.4736842105			Hard_Rock_Cafe_Piccadilly_Circus	
	4.3375			Dip_Flip	
	4.3269230769			Dirty_Bones_Shoreditch	
	4.3080357143			The_Banc	
	4.2567567568			Comptoir_Gascon	
	4.2333333333			Eat_17	
	4.1965811966			Band_of_Burgers	
	4.0948275862			The_Fire_Station	
	4.0530973451			Mac_And_Wild_Fitzrovia	



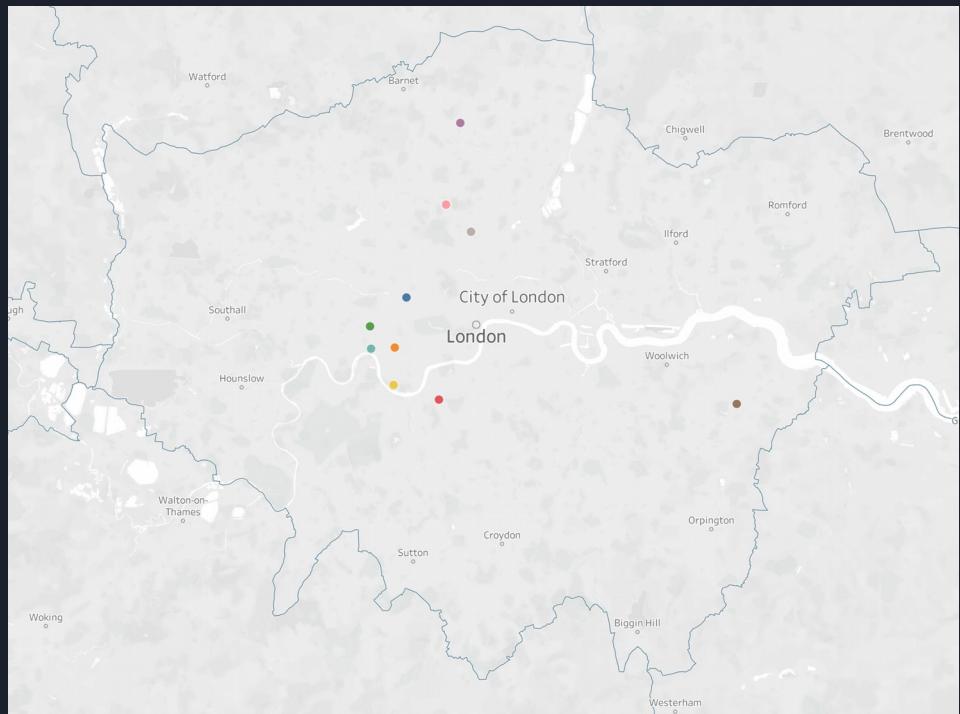
Chinese

123 avg(rating_review) ⚡	ABC restaurant_name ⚡
4.2767857143	Silk_Road
4.2180451128	Opium_Cocktail_Dim_Sum_Parlour
4.1730769231	The_Good_Earth
4.1398601399	Chinese_Cricket_Club
4.0497925311	Mandarin_Kitchen
4.0388349515	The_Duck_Rice
4.0153256705	Tao_Tao_Ju
3.9945799458	Wan_Chai_Corner
3.993485342	Gold_Mine
3.5466666667	Ping_Pong_Soho



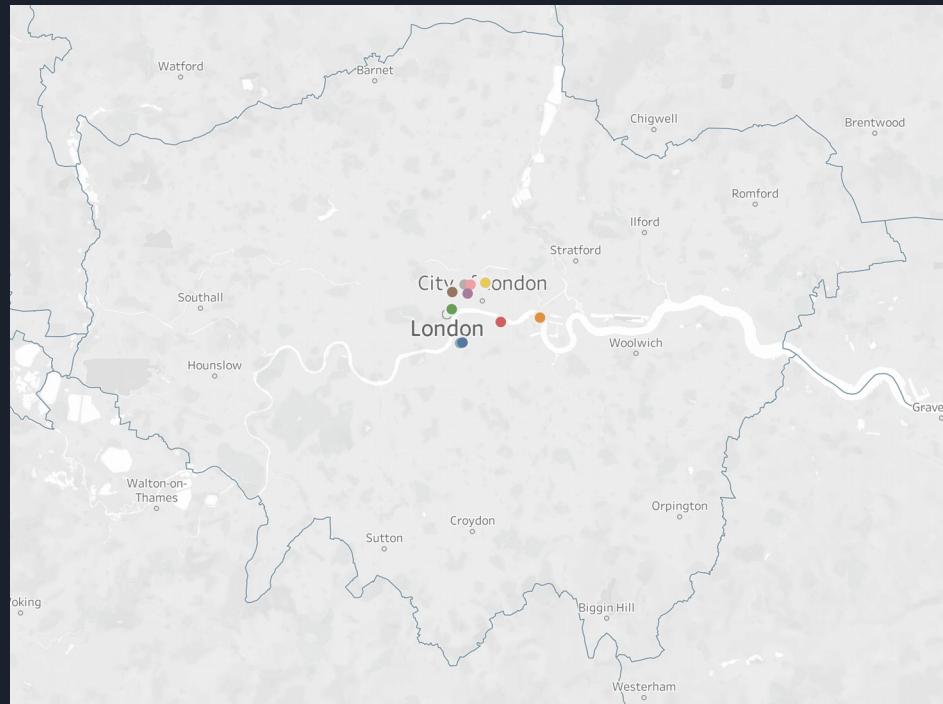
Indian

123	avg(rating_review)	ABC	restaurant_name
	4.8367346939		Kamasutra_Indian_Restaurant
	4.8131868132		Shahi_Pakwaan
	4.810483871		The_Sitara
	4.7413793103		Nepal_Authentic_Dining
	4.734939759		The_Jasmine_by_Spice_Nouv.
	4.7198177677		Indian_Express_West_Kensing
	4.7045454545		Dhaba_49
	4.6948356808		Pure_Indian_Cooking
	4.6427503737		Indian_Moment
	4.6078431373		Radhuni_Southgate



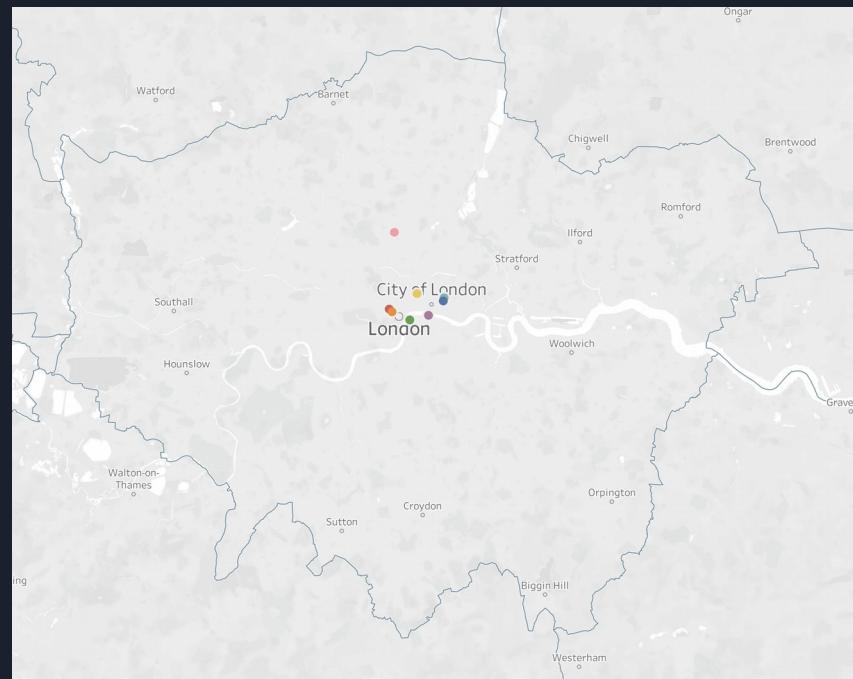
Pub

123	avg(rating_review)	restaurant_name
4.6617647059	The_Peasant	
4.5905797101	The_Blacksmith_s_Arms	
4.577092511	The_Wilmington	
4.5405405405	The_One_Tun	
4.5	The_Duchy_Arms	
4.4867256637	The_Black_Prince_Pub	
4.4297520661	The_Nell_Gwynne_Tavern	
4.4137931034	The_Old_Fountain	
4.3931623932	The_Queen_s_Larder	
4.3423423423	The_Dean_Swift	



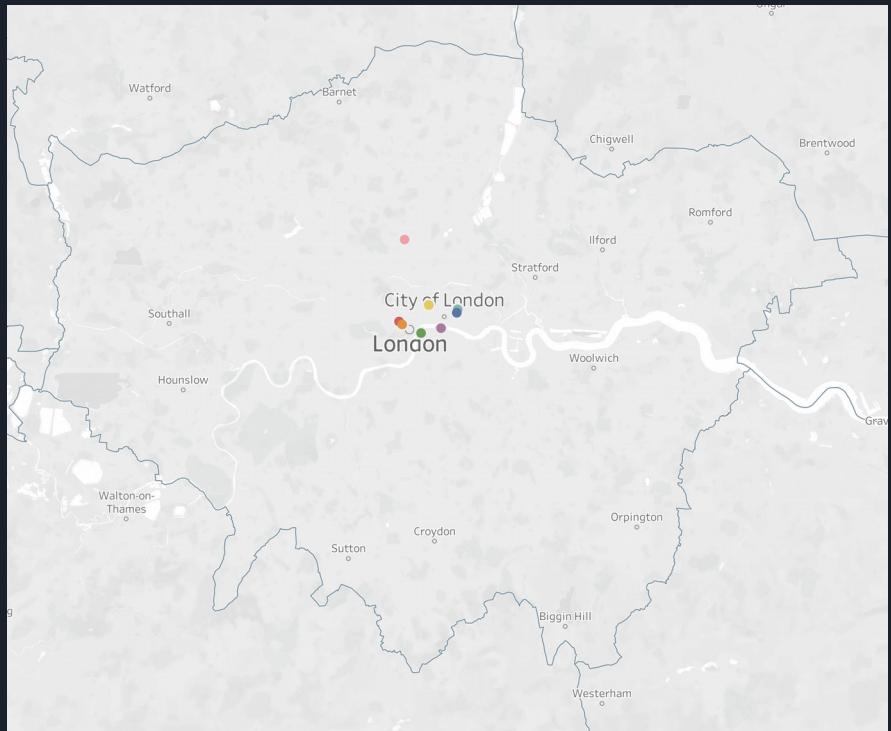
Live Music

123 avg(rating_review)	T↓	ABC restaurant_name	T↑
4.798611111		The_Sitara	
4.7358490566		Cahoots_Underground	
4.6170212766		Galvin_La_Chapelle	
4.3846153846		Brasserie_Zedel	
4.3076923077		Boisdale_of_Bishopsgate	
4.2586206897		The_Real_Greek_Bankside	
4.145631068		The_Piano_Works_Farringdon	
4.0396039604		The_Archduke	



Greek

123 avg(rating_review)	ABC restaurant_name
4.7467532468	Liman_Restaurant
4.7142857143	CERU
4.568627451	Kibele_Restaurant_Bar
4.56	Ishtar
4.5277777778	Sappho_Meze_Bar
4.3220338983	Genzo_Greek
4.2678571429	Wine_and_Mousaka
4.256097561	The_Life_Goddess_Store_Stre
4.2222222222	Al_Santo_Restaurant
4.1730769231	Likya_Ocakbasi_Restaurant



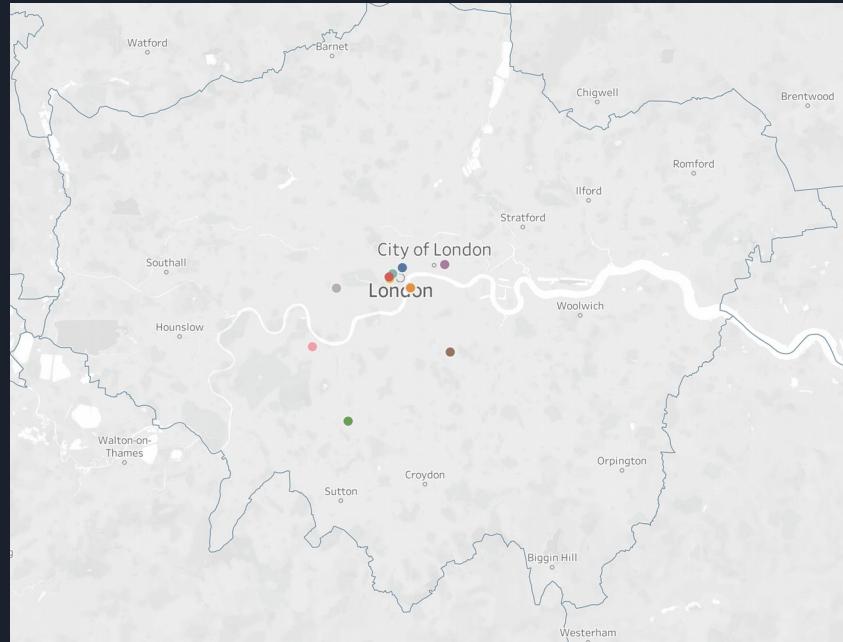
Fried Chicken

123	avg(rating_review)	ABC	restaurant_name
	4.3267973856		BAO_Soho
	4.1		Chick_n_Sours
	4.0642857143		On_the_Bab
	4.0511627907		Jinjuu_Soho



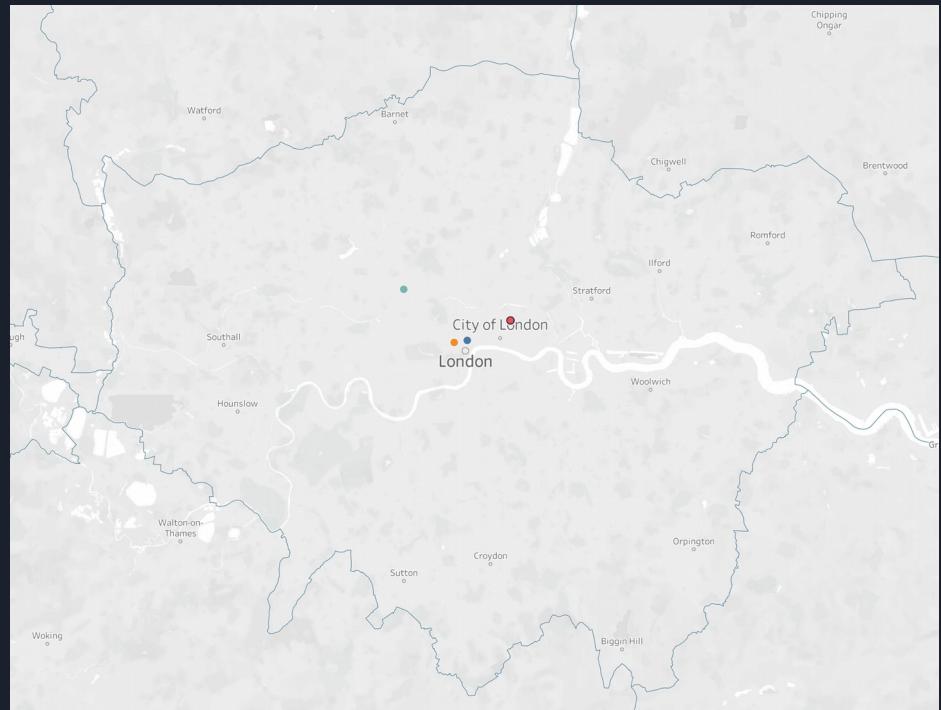
Sushi

avg(rating_review)	restaurant_name
4.7969639469	Rock_Star_Sushi_Bar
4.7804878049	Laurent_at_Cafe_Royal
4.6721311475	Tomoe
4.2678571429	Yama_Momo
4.1374045802	Yashin_Sushi
4.0735930736	Ichi_Sushi_and_Sashimi_Bar
4.0335811945	SUSHISAMBA_Heron_Tower
4	Flesh_Buns_Covent_Garden
3.9578947368	Itsu
3.9400921659	Sake_no_Hana



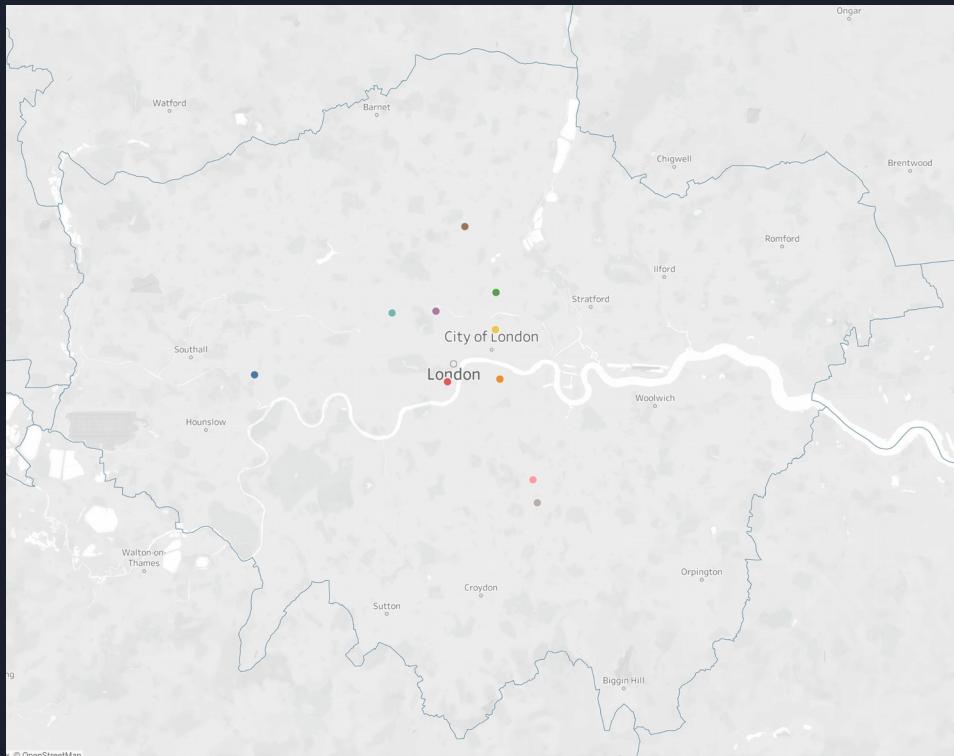
Korean

123	avg(rating_review)	ABC	restaurant_name
	4.7719298246		The_Petite_Coree
	4.2097902098		Flesh_Buns_Covent_Garden
	4.0759162304		Jinjuu_Soho
	3.9162995595		On_the_Bab



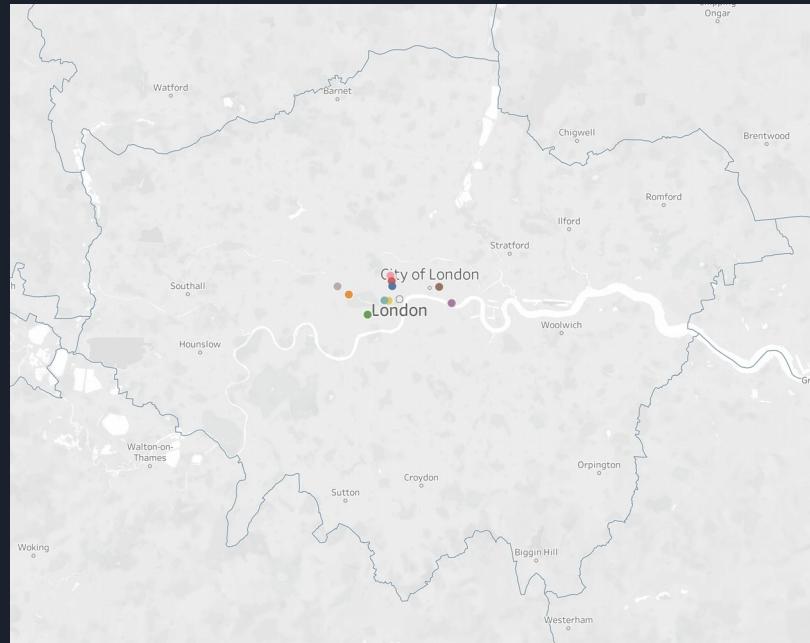
Thai

123	avg(rating_review)	ABC	restaurant_name
	4.7383419689		Awesome_Thai_Barnes
	4.6428571429		Yum_Sa
	4.5555555556		Kaosarn
	4.51		The_Blacksmith_s_Arms
	4.4736842105		Jasmine_Thai_Restaurant
	4.41		Poppy_s
	4.392		The_Kings_Arms
	4.3307086614		The_Lemon_Tree
	4.1972789116		Monkey_and_Me
	4.1847507331		The_Churchill_Arms



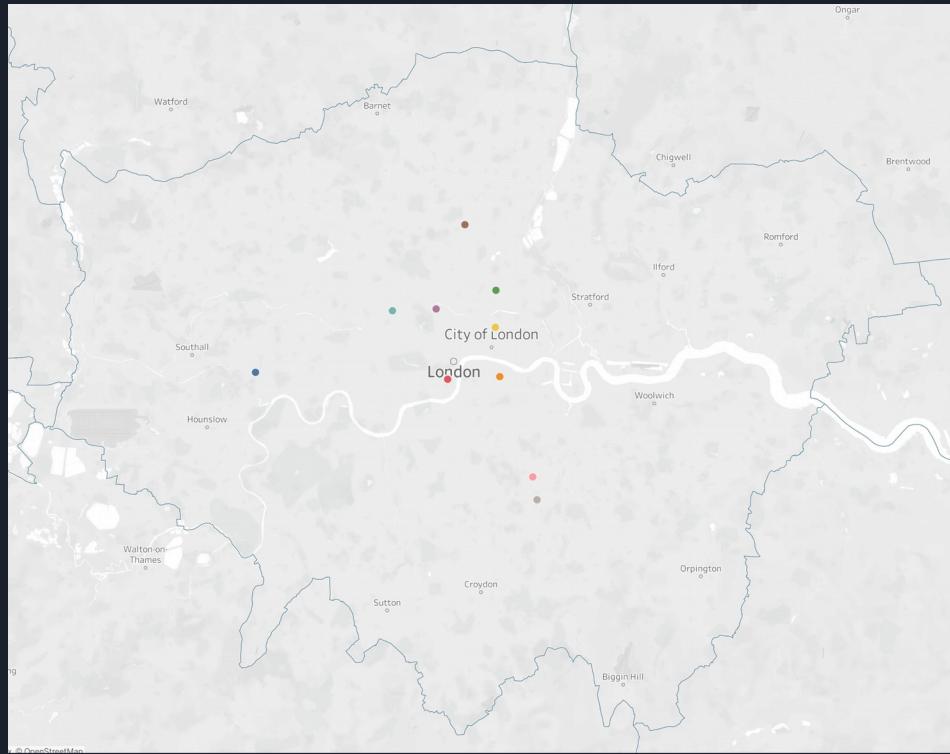
Lobster

123	avg(rating_review)	ABC	restaurant_name
	4.6949152542		Smith_s
	4.6910569106		Ormer_Mayfair
	4.6825396825		The_Ledbury
	4.5344827586		Seven_Park_Place_By_William_Drabble
	4.5		Mere
	4.4736842105		SUSHISAMBA_Heron_Tower
	4.2250489237		Steak_And_Lobster_Warren_Street
	4.2205882353		Scalini
	4.1050955414		Mandarin_Kitchen
	4.037037037		Berners_Tavern



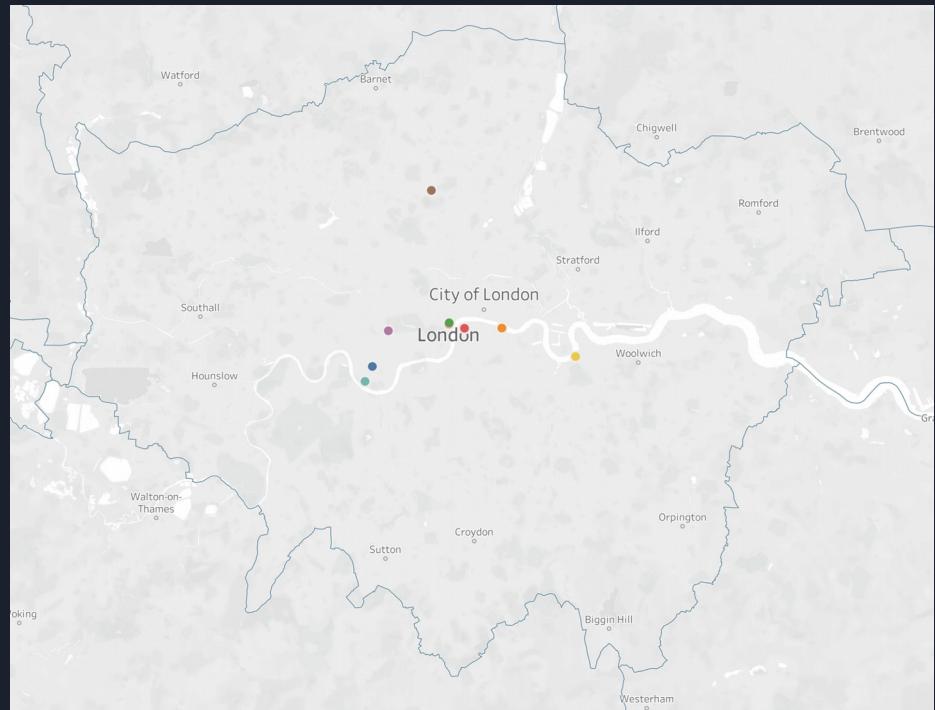
Italian

123 avg(rating_review)	ABC restaurant_name
4.9411764706	Da_Moreno
4.8831168831	Pizza_Union_Hoxton
4.8442622951	Fine_Foods
4.8415841584	Mio_Restaurant
4.8363636364	Sfizio_Italian_Tapas
4.8121546961	Trattoria_Raffaele
4.8050314465	Oi_Vita_Pizzeria
4.805	Morso_Abbey_Road
4.7727272727	Tomo_Pizzeria_Restaurant
4.7703927492	Purezza_Camden



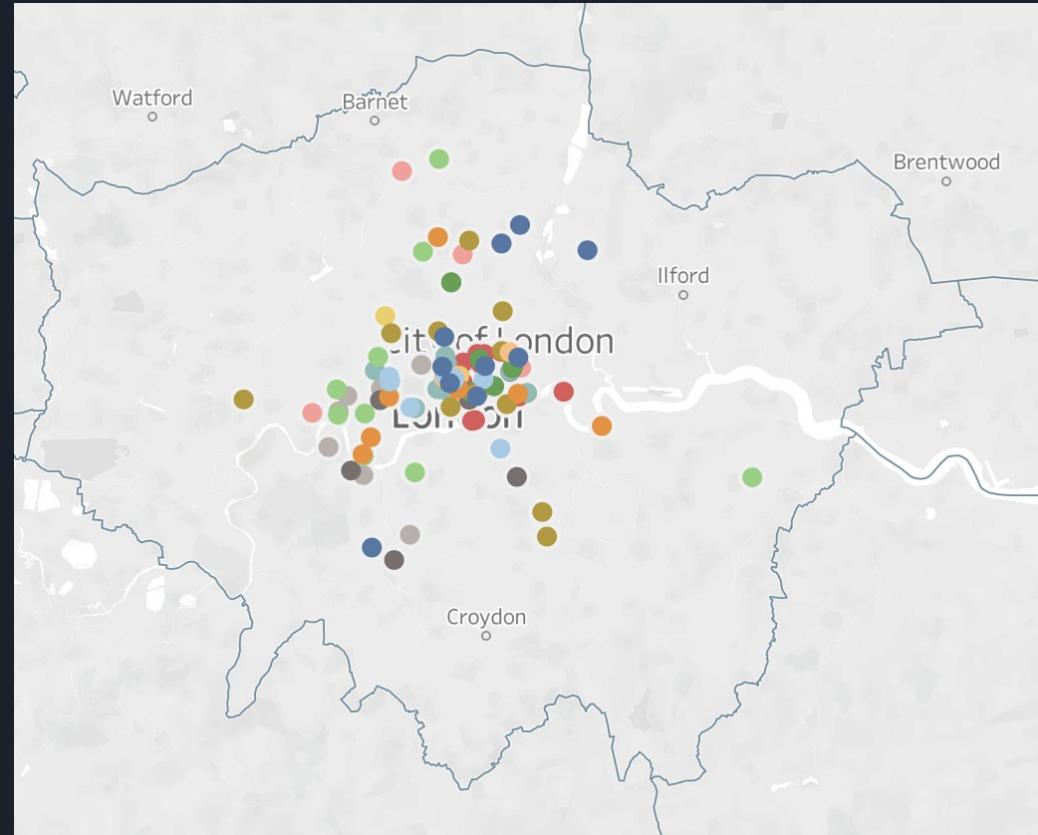
Fish and Chips

123	avg(rating_review)	restaurant_name
4.8343949045		<u>Fishers_Restaurant_Takeaway</u>
4.7204724409		Baileys_Fish_and_Chips
4.6179775281		Fishcotheque
4.1090909091		Toff_s
4.0925925926		The_Goat
3.9885057471		The_Cutty_Sark
3.9655172414		The_Old_Shades_London_Pub_Dining
3.8		Butlers_Wharf_Chop_House
3.7976190476		Garfunkel_s



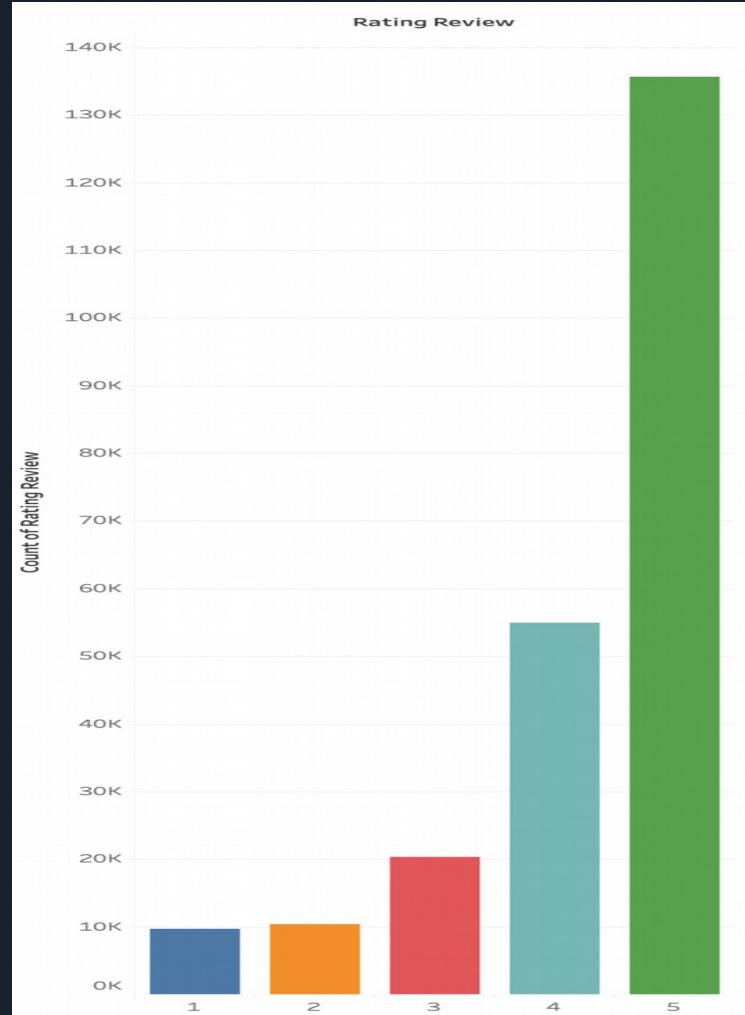
Compiled

Type
burger
chinese
fishandchips
friedchicken
greek
indian
italian
korean
livemusic
lobster
pub
steak
sushi
thai



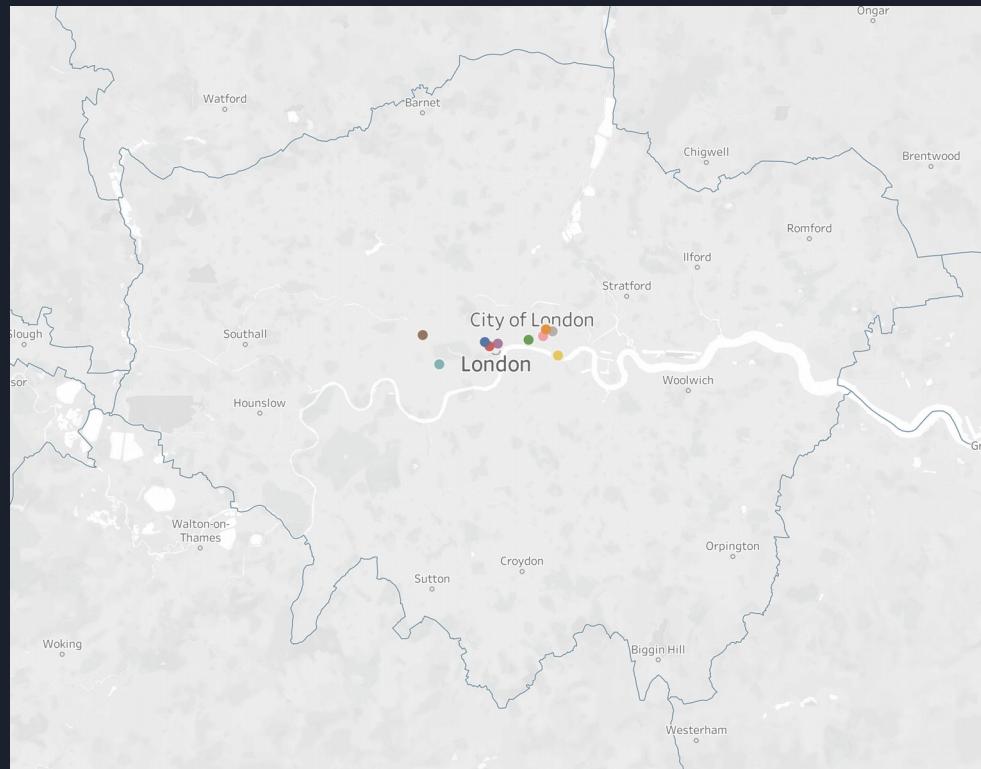
Visualize # of reviews in each ranking

123 count(rating_review)	123 rating_review
135,646	5
54,993	4
20,258	3
10,426	2
9,692	1



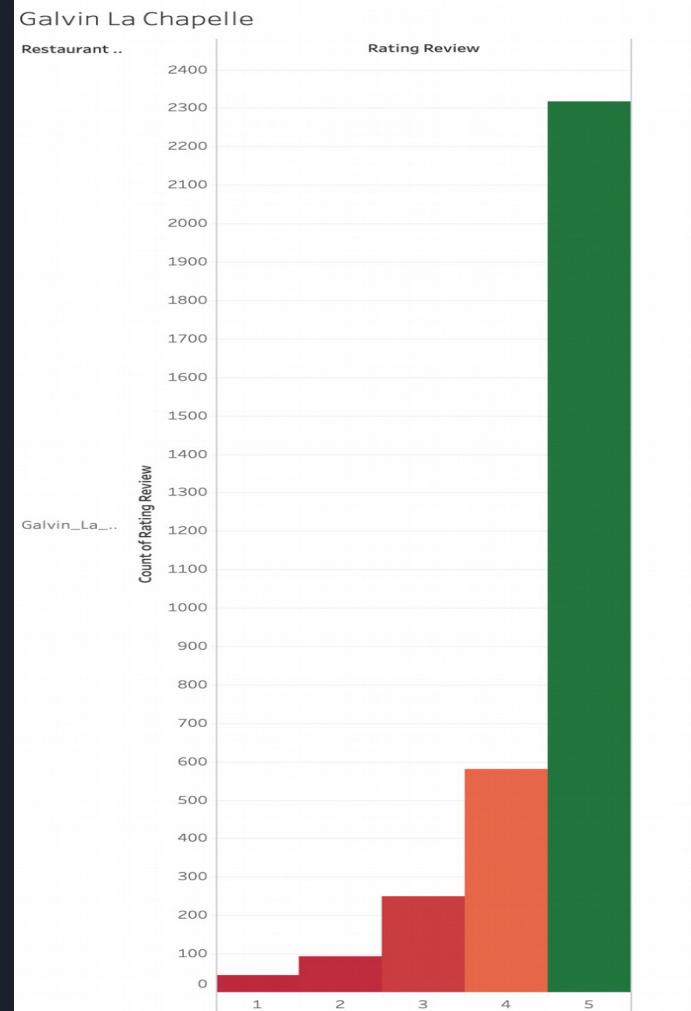
Most 5's

restaurant_name	count(rating_review)	rating_review
Galvin La Chapelle	2,318	5
The_Ledbury	2,182	5
SUSHISAMBA_Heron_Tower	1,994	5
Cahoots_Underground	1,573	5
Little_India	1,471	5
The_Monsoon	1,393	5
Hard_Rock_Cafe_Piccadilly_Circus	1,370	5
Smith_s	1,309	5
Steak_Co_Garrick_Street_Covent_Garden	1,205	5
Madison	1,181	5



Visualize distribution of rankings for each restaurant

123 count(rating_review) 	123 rating_review 	ABC restaurant_name 
2,318	5	Galvin_La_Chapelle
581	4	Galvin_La_Chapelle
250	3	Galvin_La_Chapelle
93	2	Galvin_La_Chapelle
44	1	Galvin_La_Chapelle



Future Practical Applications

Provide
Guidance to
Restaurant
Industry on Best
Location for
Certain Foods

Provide
Guidance
Regarding a
City's
Preferences or
Needs

Provide Deep
Insight to Tourists
or City Residents
on Where to Eat
and What to
Order

Project Limitations

Beautiful Soup for Website Scraping Didn't Work on TripAdvisor

```
def addresses(url):
    BeautifulSoup(requests.get(url).text, \
                  'html.parser').find_all('a', {'class': 'AYHFM'})

df.withColumn('address', lambda x: addresses(x[11])))
```