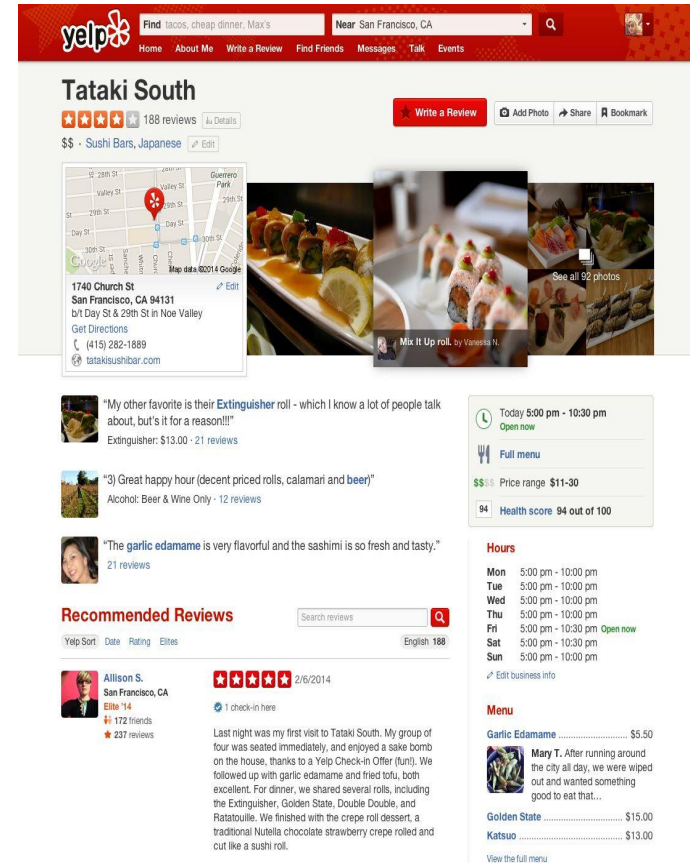


Restaurant Analysis Dashboard-Yelp Data

Professor : Dr. Najam Hassan

Abstract

- The Yelp Open Dataset provides information on a variety of businesses such as restaurants, bars, hotels, and coffee shops.
- There are millions of user reviews, but these are not usable in their current state.
- By analyzing customer review data and star ratings, we will suggest businesses to improve in areas such as food quality, ambiance, and parking.
- Our analysis will help increase customer loyalty and retention by providing insights into areas for improvement.



The screenshot shows the Yelp profile for Tataki South, a Japanese restaurant in San Francisco. The page includes a map, star ratings (4.5 stars), address (1740 Church St), phone number, and a list of reviews. A recommended review from Allison S. is highlighted, describing a recent visit with a group of four.

Tataki South
 4.5 stars 188 reviews
 \$\$ • Sushi Bars, Japanese
 1740 Church St
 San Francisco, CA 94131
 b/t Day St & 29th St in Noe Valley
 Get Directions
 (415) 282-1889
 tatakisushibar.com

Recommended Reviews

Allison S.
 San Francisco, CA
 Elite '14
 172 friends
 237 reviews

2/6/2014
 1 check-in here

Last night was my first visit to Tataki South. My group of four was seated immediately, and enjoyed a sake bomb on the house, thanks to a Yelp Check-in Offer (fun!). We followed up with garlic edamame and fried tofu, both excellent. For dinner, we shared several rolls, including the Extinguisher, Golden State, Double Double, and Ratatouille. We finished with the crepe roll dessert, a traditional Nutella chocolate strawberry crepe rolled and cut like a sushi roll.

Hours

Day	Hours
Mon	5:00 pm - 10:00 pm
Tue	5:00 pm - 10:00 pm
Wed	5:00 pm - 10:00 pm
Thu	5:00 pm - 10:00 pm
Fri	5:00 pm - 10:30 pm
Sat	5:00 pm - 10:30 pm
Sun	5:00 pm - 10:00 pm

Menu

Garlic Edamame	\$5.50
Golden State	\$15.00
Katsuo	\$13.00

Data source

The Yelp dataset is a subset of

1. Businesses
2. User reviews
3. user profile

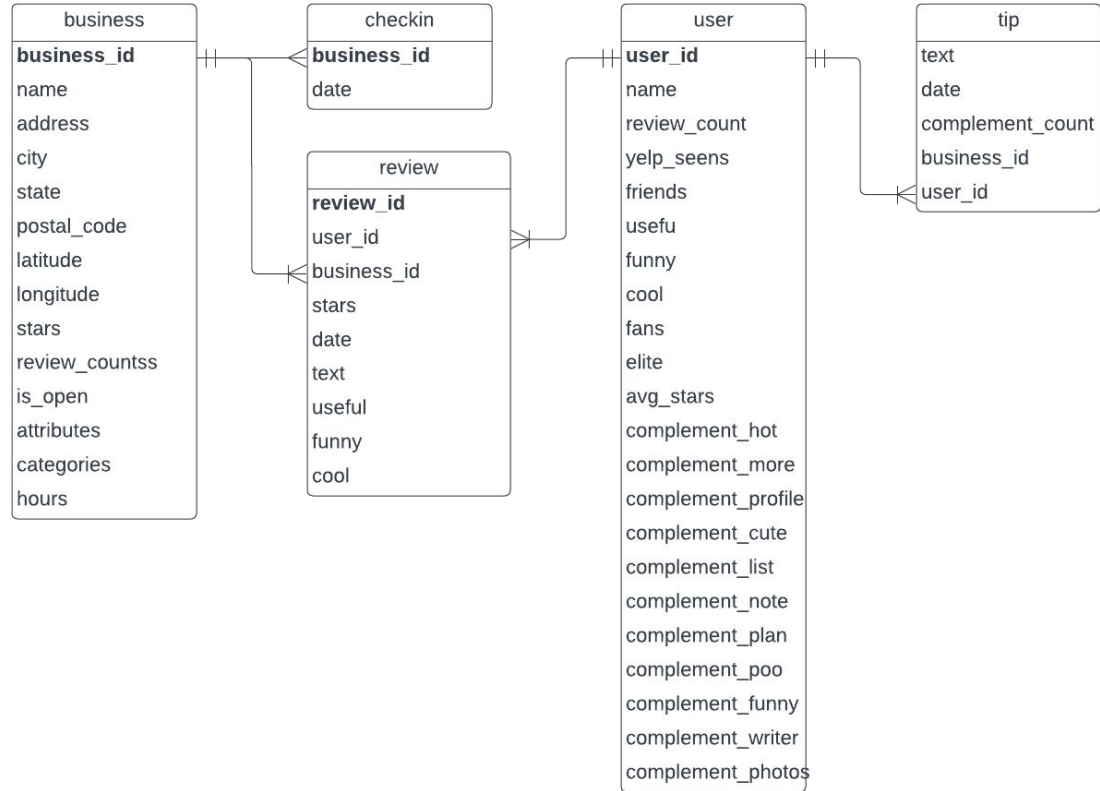
Data Attributes

1. 47 unique dimensions
2. 5 Tables.
3. More than **6,990,280** user reviews.
4. More than **100,000** businesses.

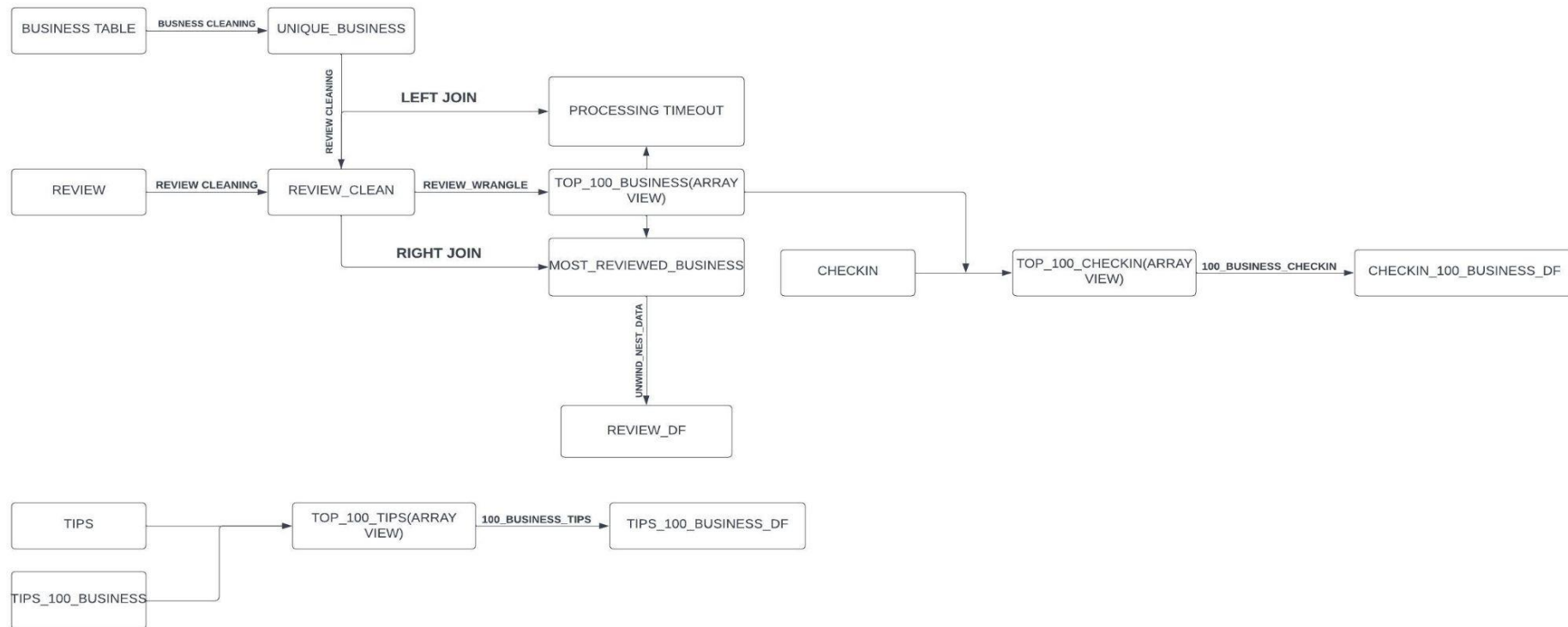
Dataset Size : 8.65 GB

Data source: <https://www.yelp.com/dataset>

ERD



Data Preprocessing



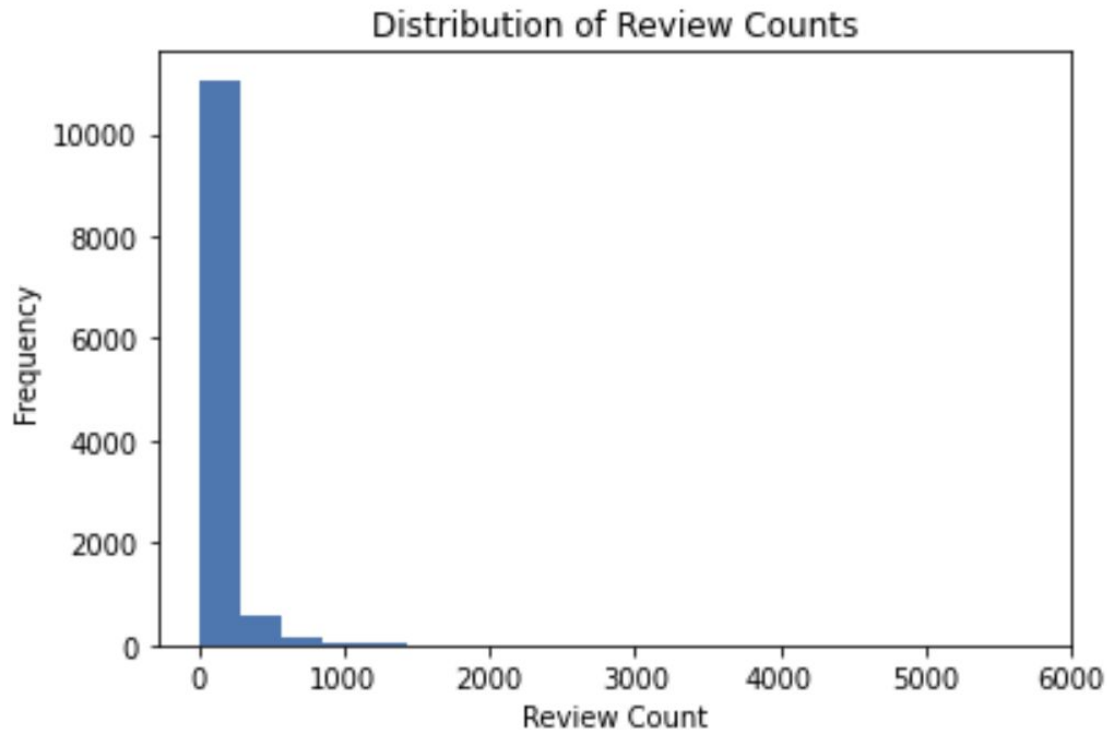
code-walkthrough

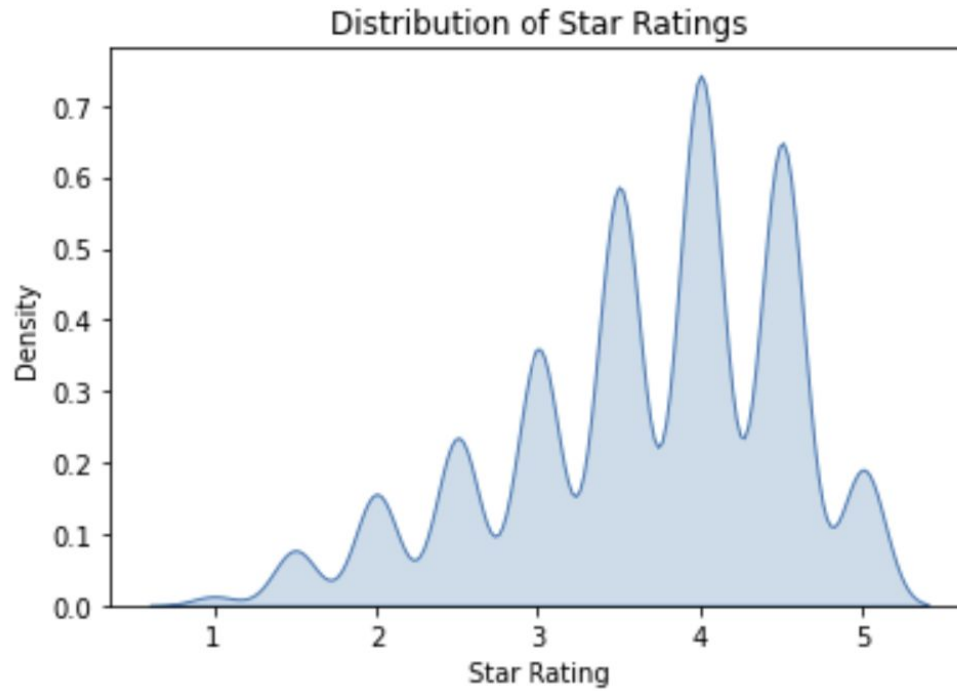
Challenges Faced

1. Data Loading was not straightforward.
2. Spark was unable to process large amount of text data.
3. MongoDB was much better at cleaning large number of data.
4. Sentiment Extraction was very time taking.
5. Exporting data out of Databricks community edition is a hassle.

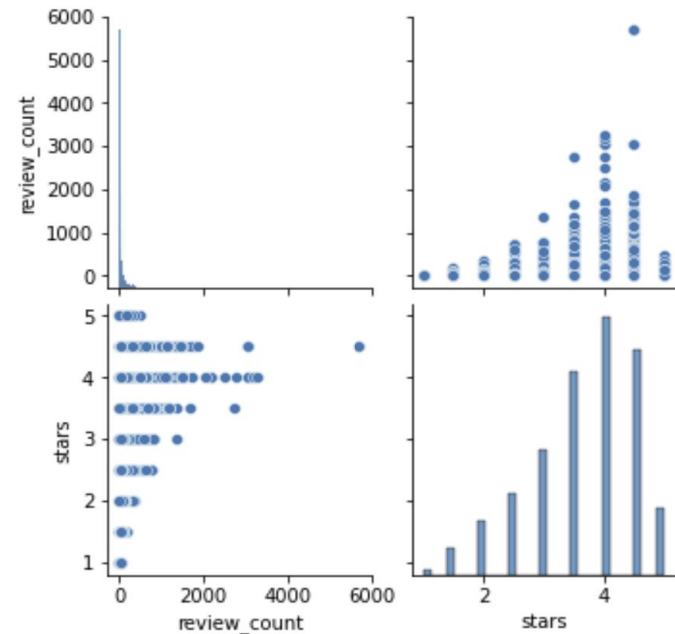
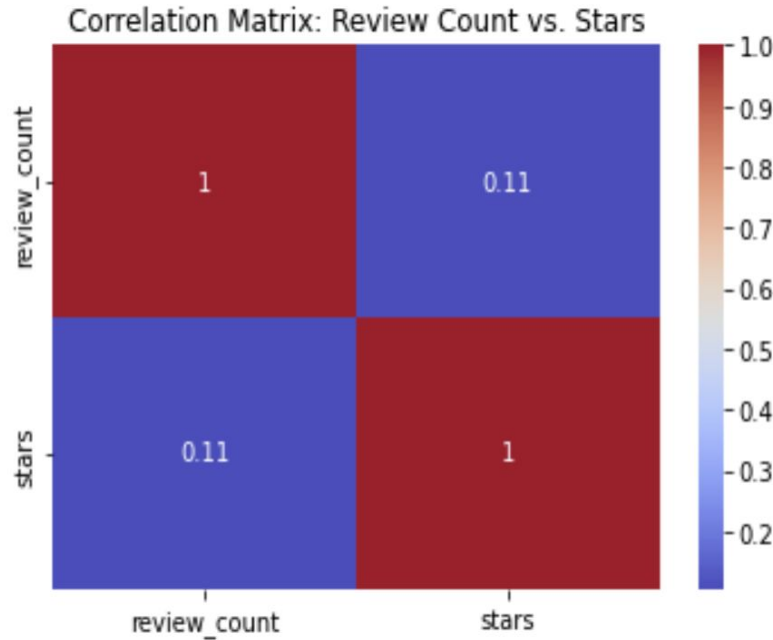
Solution: Larger Cluster

EDA

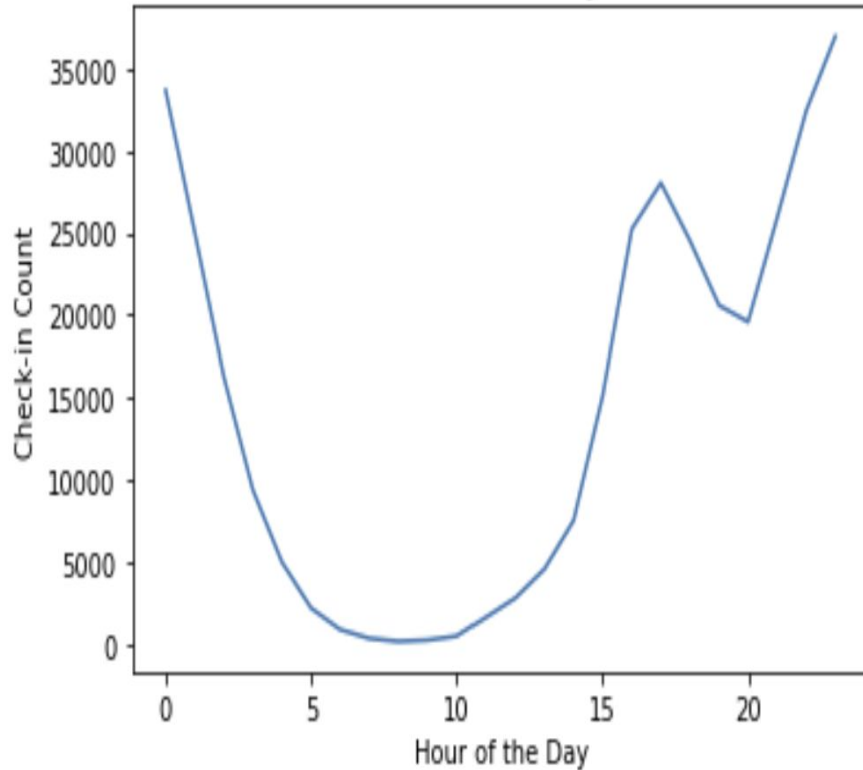




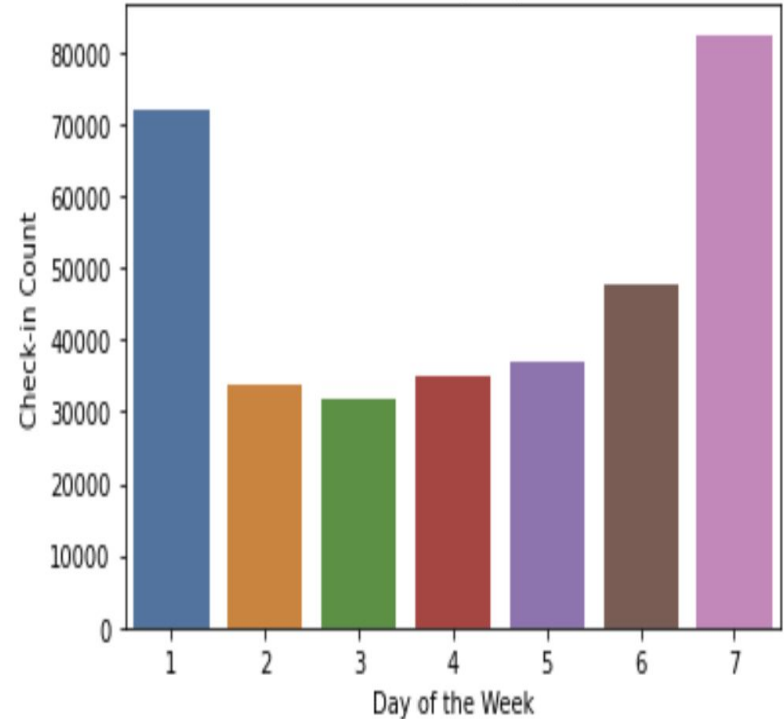
Correlation between Review Count and Stars: 0.10596399644465825



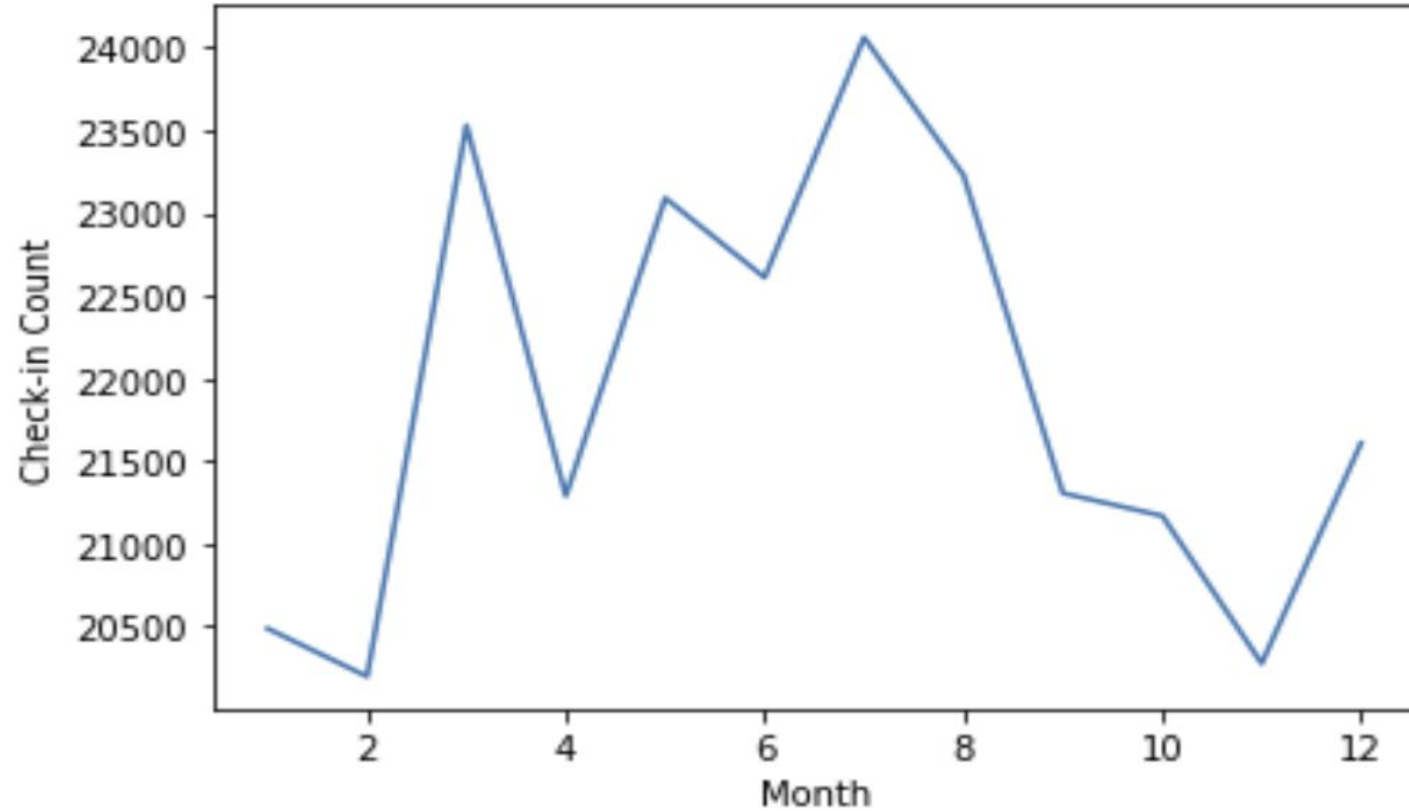
Check-in Patterns by Hour

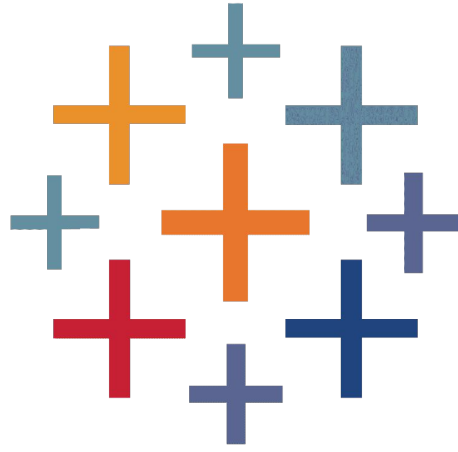


Check-in Patterns by Day of the Week



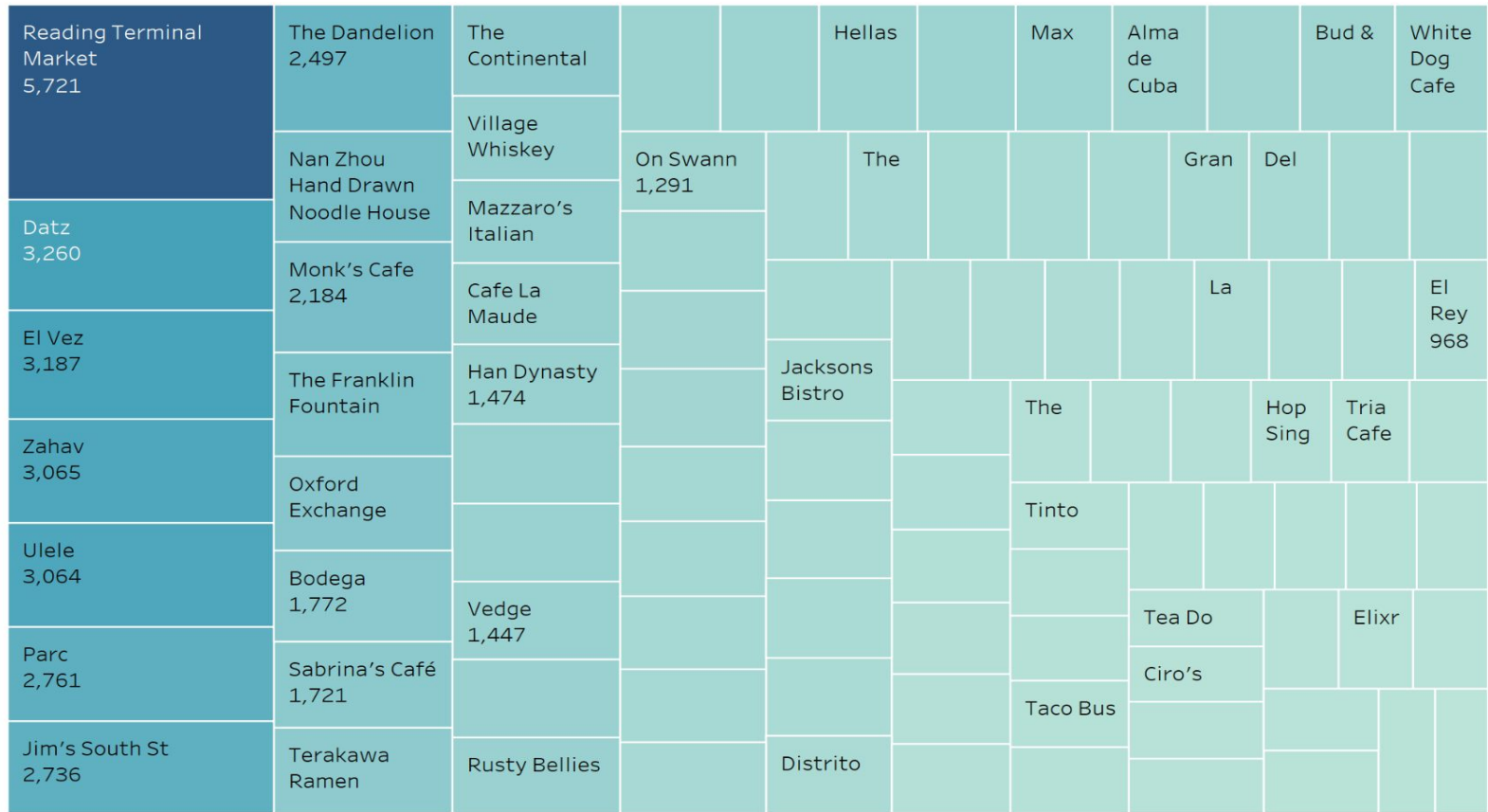
Seasonal Trends in Check-in Data





+ a b | e a u

REVIEW COUNT PER RESTAURANT



STAR RATING PER RESTAURANT

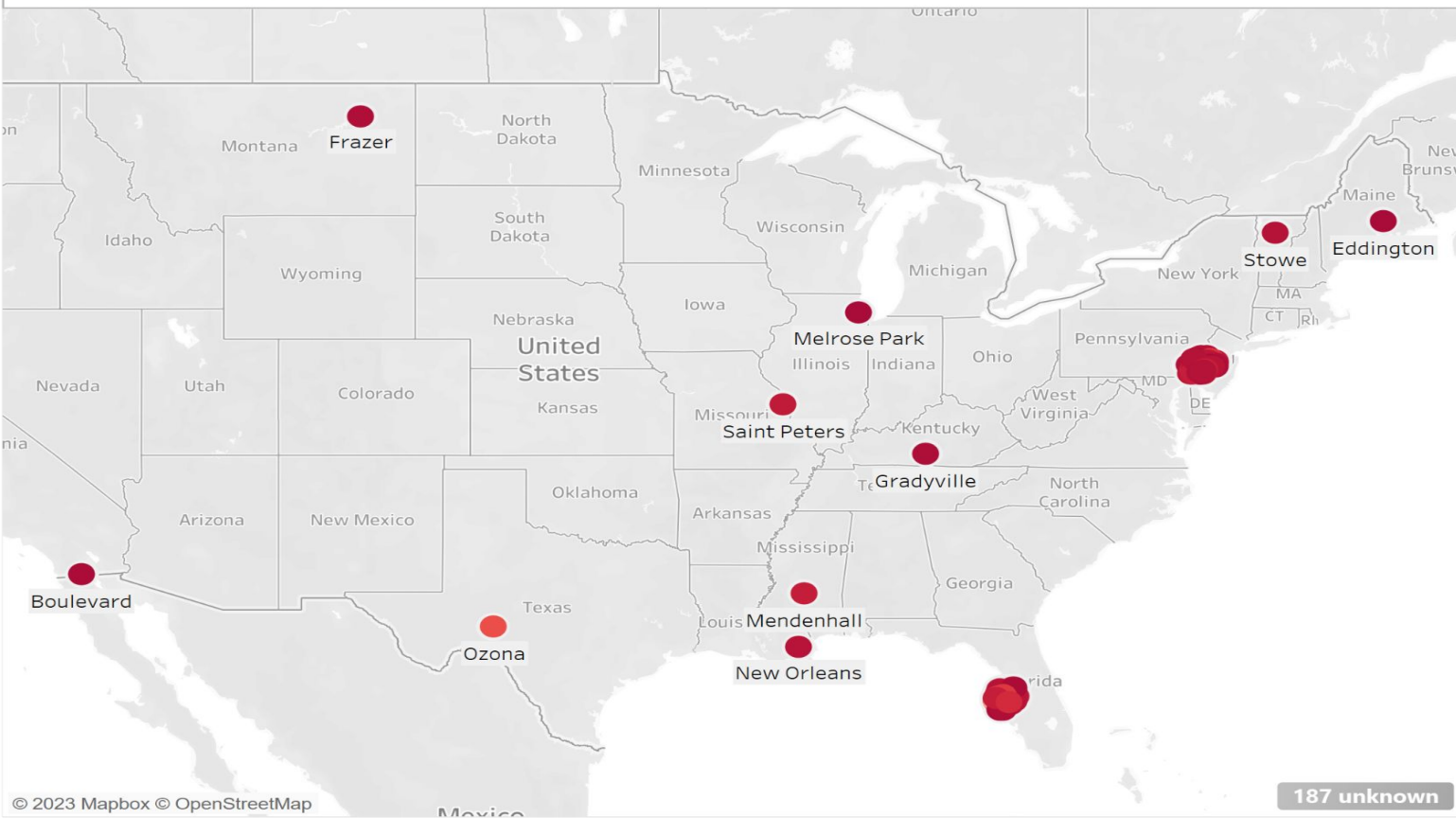
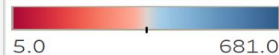
Armature Works 4.500	Elixir Coffee Roasters 4.500	Noble Crust 4.500	Terakawa Ramen 4.500	Besito	Bud &	Café Soho 4.000	Cantina Dos	Charlie Was a Sinner 4.000	Circles	Ciro's 4.000			Datz 4.000
Beiler's Bakery 4.500	Engine No 9 4.500	Ocean Prime 4.500	The Mill Restaurant 4.500	Del Frisco's Double Eagle			Hop Sing	Ichicoro Ramen 4.000	Khyber Pass Pub 4.000	La Teresita Cafe 4.000		Monk's Cafe 4.000	Nan Zhou Hand Drawn Noodle
Bodega 4.500	Gran Caffè L'Aquila 4.500	On Swann 4.500	The Stein & Vine 4.500										
Cafe La Maude 4.500	Green Lemon 4.500	Oxford Exchange 4.500	Tria Cafe Rittenhouse 4.500	Ella's Americana			Tampa Bay	Tea Do 4.000	The	The	The	Tinto 4.000	
Chubby Cattle 4.500	Hellas Restaurant 4.500	Reading Terminal Market	Vedge 4.500										Frankford Hall 4.000
Cigar City Brewing 4.500	IndeBlue Modern Indian Food & Spirits	Rusty Bellies 4.500	Vernick Food & Drink 4.500	Frenchy's Saltwater	Parc 4.000	Ulele 4.000	Caddy's	Cantina Los	Casa Tina 3.500	Cuba Libre	Devil's Alley Bar & Grille		
Craft Street Kitchen - Oldsmar	Lolis Mexican Cravings 4.500	Shrimpy's Blues Bistro 4.500	Zahav 4.500									Good Dog Bar 4.000	Sabrina's Café 4.000
Double Knot 4.500	Mazzaro's Italian Market 4.500	Snapper's Sea Grill 4.500	Alma de Cuba 4.000	Han Dynasty 4.000		edison: food+drink lab	El Rey 3.500	Max Brenner -	Taco Bus 3.500				
Eddie & Sam's NY Pizza 4.500	Mr. Dunderbak's 4.500	Suraya 4.500	BellaBrava 4.000								Harp & Crown 4.000		

AVG(Stars)



AVG(REVIEW COUNTS) PER CITY

AVG(Review Count)



LEAST RATED RESTAURANT

SEMINOLE HARD ROCK HOTEL & CASINO

pizza Sentiment: 1.667	music Sentiment: 1.286	pasta Sentiment: 1.000	price Sentiment: 1.000	rice Sentiment: 1.000	soup Sentiment: 1.000	
	service Sentiment: 1.154					
food Sentiment: 1.385	atmosphere Sentiment: 1.000	sushi Sentiment: 1.000		quality Sentiment: 0.750	lighting Sentiment: 0.667	lobster
staff Sentiment: 1.353	beer Sentiment: 1.000	waiter Sentiment: 1.000		location Sentiment: 0.500		menu Sentiment: 0.500
steak Sentiment: 1.333	budget Sentiment: 1.000	cost Sentiment: 0.750		manager Sentiment: 0.500		parking Sentiment: 0.362

AVERAGELY RATED RESTAURANT

HARP & CROWN

burger Sentiment: 2.000	ambience Sentiment: 1.917	cuisine Sentiment: 1.500	service	pizza	vibe	gluten	host	
cost Sentiment: 2.000	bartender Sentiment: 1.846	lobster Sentiment: 1.500						
knowledge Sentiment: 2.000	decor Sentiment: 1.838	location Sentiment: 1.500	texture Sentiment: 1.333	dish	flavor	allergy	chips	curry
sandwich Sentiment: 2.000	atmosphere Sentiment: 1.818	music Sentiment: 1.500	food Sentiment: 1.297					
sushi Sentiment: 2.000	beer Sentiment: 1.533	parking Sentiment: 1.500	taste Sentiment: 1.286	fries Sentiment: 1.000	rice	waiter	hostess	
vegan Sentiment: 2.000	staff Sentiment: 1.522	price Sentiment: 1.500	value Sentiment: 1.286	portion Sentiment: 1.000				
		salad Sentiment: 1.500	menu Sentiment: 1.276	presentation Sentiment: 1.000	quality Sentiment:	pasta		
				ribs Sentiment: 1.000	lighting Sentiment:	steak Sentiment:		
					crab Sentiment:	attitude		

HIGHEST RATED RESTAURANT

OXFORD EXCHANGE

allergy Sentiment: 2.000	pizza Sentiment: 2.000	ambiance Sentiment: 1.800	staff Sentiment: 1.619	sandwich Sentiment: 1.611	menu Sentiment: 1.553	food Sentiment: 1.535	portion	
chips Sentiment: 2.000	seasoning Sentiment: 2.000	atmosphere Sentiment: 1.729						
curry Sentiment: 2.000	soup Sentiment: 2.000	service Sentiment: 1.675	recipe Sentiment: 1.500		burger	cost	lobster	texture
			rice Sentiment: 1.500					
distance Sentiment: 2.000	waiter Sentiment: 2.000	flavor Sentiment: 1.667	taste Sentiment: 1.500	waitress Sentiment: 1.333			price	
				location Sentiment: 1.182				
ingredients Sentiment: 2.000	vibe Sentiment: 1.938	gluten Sentiment: 1.667	fries Sentiment: 1.467	steak Sentiment: 1.167		dish Sentiment:		
				attitude Sentiment:		host Sentiment:		
lighting Sentiment: 2.000	decor Sentiment: 1.875	presentation Sentiment: 1.667	salad Sentiment: 1.393	attitude Sentiment:		quality		noise

Conclusion

1. No correlation between number of reviews and rating (popular restaurant need not be the best).
2. The Best restaurants cater to food allergy and offer great service.
3. Both averagely and best rated restaurants had a poor manager sentiment (needs further investigation).
4. People love Pizza and Burger.
5. Sampled restaurants had poor sentiment towards the staff i.e (Waitress, Waiter, Manager).

