

CORNELL UNIVERSITY

**NBA 6550: INTRO TO SQL AND TABLEAU**

Prof. KASHISH ARORA

**YELP PROJECT**

**Team #7**

Heesun Chang (HC483)

Patchara Suensilpong (PS972)

Sebastian Koh (YK784)

Si Park (SP2477)

FALL 2020

ITHACA, NEW YORK

4 DECEMBER 2020

## Intro

The purpose of this report is to detail how we analyzed the Yelp dataset and extracted meaningful insights that would help future restaurant entrepreneurs decide on the most optimized location and type of food business. Particularly, this report will focus on what attributes of starting a restaurant are most important for future restaurant owners. After such analysis, the report will conclude with the recommendation for future business plans.

## Overview of the Yelp Dataset and Data filtering

The Yelp dataset contains data related to businesses using Yelp, reviews recorded by customers, users' information as well as other information collected from the platform. This dataset includes a total of 6 distinct tables: 'business', 'checkin', 'photos', 'review', 'tip' and 'user' table. In these 6 tables, it carries data of 28,450 businesses, 26,881 check-ins, 48,862 reviews, 125,642 photos, 9,523 tips and 13,261 users. Within the business table in the Yelp dataset, there are several types of business aside from the restaurants. [Query 0] 861 business ids are in the Arts & Entertainment category. To omit the miscategorized problem, we will use "restaurant" and "food" as the filter words for our subsequent studies, which contain a total of 10368 business ids.

## Data Analysis

### *1. Locations and Competitors Analysis*

Las Vegas is the most populated city, with most businesses operating and most reviews recorded by the users and customers. [Query 1] This can imply that Las Vegas is the most populated area by customers, but can also merely mean there is a large amount of active Yelp users in this city. Assuming that Las Vegas is the most populated area, the area with postal code 89109 is identified to be the most businesses currently operating and most reviews recorded compared to other areas within Las Vegas. [Query 2] Then, to see where the high rated reviews were recorded, we listed out cities with the most number of businesses with only 5 -star rated reviews, and 5 & 4-star rated reviews. Top 5 cities with the most number of businesses and reviews were Las Vegas, Phoenix, Toronto, Scottsdale and Charlotte, considering all 1 to 5-star rated reviews. When we only looked at 5-star rated reviews, the list remained the same. [Query 3] However, with only 5 and 4-star rated reviews, the list slightly changed: Pittsburgh replacing Charlotte. [Query 4] With the fact that Las Vegas is the most populated city with various food businesses and where the users leave the most reviews, we also studied which area in Las Vegas is the good location for the new business to attract customers. As a result, 'The Strip', the stretch of Las Vegas Boulevard Avenue, was identified to be the area with the most populations, and the 'Southeast' area was the second in the list. [Query 5] Though it is important to find a location where there is a high floating population where new businesses can lure customers into the store, it is also important to be aware of other competitors in the area. By looking at food businesses that received more than 100 reviews and located in 'The Strip' area, 'Lip Smacking Foodie Tours' was the only business with the 5.0 rating, but had relatively less reviews recorded, thus cannot be concluded to be the best restaurant in this area. But still, there were about 190 businesses with the rating higher than 4.0, with a large number of reviews recorded. From this, we can conclude that 'The Strip' area is highly populated, but also expected to be a very competitive area for the new businesses to locate. [Query 6]

### *2. Restaurant Type Analysis*

Once we know that Las Vegas is a high potential area to open a new restaurant, we want to explore what type of food that new restaurant should sell. We found that each restaurant can have more than one category [Query 7] . We will use each keyword to search and aggregate the table. Focusing on the top nine nationality foods (Mexican, Taiwanese, American, Japanese, Thai, Vietnamese, Italian, Korean, Indian), We found that

American food restaurants have the highest total check-in frequency 503,475, followed by Japanese and Mexican restaurants [Query 8]. From the reviewing perspective, Mexican restaurants have the lowest average review star, 3.60 stars [Query 9]. These make Mexican restaurants one of the most interesting choices because they have high demand, and it will be easier for a new business owner to open this type of restaurant and get a high review rank. We dug deep down into Mexican type restaurant data by putting each Mexican restaurant into the review star brackets [Query 10]. We found that Mexican restaurants with stars from 4.0 to 4.5 trend to have higher check-in than other brackets. This indicates the review score target for new restaurants. Finally, according to the check-in rate for the periods during the week, the restaurant can be opened only at the high peak demand period to optimize the cost and revenue [Query 11]. We found that the demand patterns are the same throughout the week, with higher demand from Friday to Sunday. Therefore, we suggest a new Mexican restaurant to set their open hours from 17:00 to 03:00 for both weekday and weekend.

### *3. Review Analysis on Customer Reviews*

Moving on, we explored the customer reviews in Yelp and the tips comments to learn about the key topics that customers are most interested in. The word cloud [Tableau Dashboard] provided an overview of the most commonly used words and suggested a possibility that customers are most concerned about the service and staff of the restaurant. Besides that, the most common food that appeared in the reviews are chicken and pizza, which suggests that people could be interested in these types of food. With this information, we looked at the distribution of the star ratings of restaurants which received a 'service/staff' related comment. It shows that a total of 1877 restaurants received at least one 'service/staff' related comment [Query 12] and most of the service related comments are given to restaurants with high ratings [Query 13]. This indicated that customers looked into 'good service' when patronizing high rating restaurants and we should inculcate this in our new business culture to acquire customers' satisfaction. Apart from that, we also looked into the distribution of the chicken related reviews and the ratings associated with the restaurant that received the review. Bearing in mind that the most common food that people comment about is chicken, we learnt that the ratings associated with those reviews are high (4 stars) [Query 14]. This indicated that if our new business was to sell chicken, there will be competition and we need to place greater emphasis on the quality of the chicken dishes. Similarly, Pizza is the 2nd most commonly commented food and their reviews are generally of high ratings (4 stars) [Query 15]. While we learn that a significant number of people in Las Vegas eat Pizza, there might be strong competition in selling pizza. Hence, if we were to venture into selling pizza, we need to ensure high quality of our food in order to attract customers.

### **Conclusion**

The Yelp dataset shows that Las Vegas, specifically, 'The Strip' area, has the highest proportion of customers among the cities and would probably be a suitable location to start a new restaurant. However, new businesses should also be aware that the competition is high in this area with many food businesses having ratings above 4.0. The check-in dataset also revealed that Mexican restaurants have higher check-in frequency and lowest review star, signifying a potential cuisine that we can venture into. Lastly, we obtained insights from Yelp review that customers are most concerned about the service of the restaurant and the most commonly mentioned food are chicken and pizza. These insights suggest important areas we should focus on to ensure maximum customer satisfaction. We provided a preliminary analysis on the possible strategies that new businesses could adopt; however, more data would have to be collected on this alternative so that our analysis and recommendation would be more accurate for potential business stakeholders.

## [ Appendices - SQL Queries ]

	categories	Freq
1	Arts & Entertainment	861
2	Beauty & Spas	1434
3	Restaurant & Food	10368
4	Shopping	2210

**Query 0 - List of categories aside from the restaurant and food service**

```
select 'Beauty & Spas' as categories , count(distinct categories) as Freq
from business
where categories like '%Beauty & Spas%'
union
select 'Shopping' as categories , count(distinct categories) as Freq
from business
where categories like '%Shopping%'
union
select 'Arts & Entertainment' as categories , count(distinct categories) as Freq
from business
where categories like '%Arts & Entertainment%'
union
select 'Restaurant & Food' as categories , count(distinct categories) as Freq
from business
where (categories like '%Restaurant%') or (categories like '%Food%');
```

	Location_City	NumberOfBusiness
1	Las Vegas	9129
2	Phoenix	3245
3	Toronto	3074
4	Scottsdale	1827
5	Charlotte	1488
6	Pittsburgh	1444
7	Montréal	991
8	Tempe	952
9	Henderson	916
10	Chandler	729
11	Mesa	729
12	Cleveland	690
13	Madison	666
14	Gilbert	568
15	Glendale	403
16	Mississauga	320

**Query 1 - List of Cities with most number of businesses operating, finding populated area in a broader scale**

```
select business.city as Location_City, count () as NumberOfBusiness
from review JOIN business using (business_id)
where categories like '%restaurant%' or categories like '%food%'
```

group by city  
order by NumberOfBusiness desc;

	Postal_Code	Number_Of_Business_In_The_Area
1	89109	3087
2	85251	703
3	89102	595
4	85281	577
5	89119	563
6	89123	414
7	89101	408
8	89103	397
9	89146	390
10	85016	378
11	15222	349
12	85004	326
13	89052	309
14	89169	289
15	53703	277
16	85018	269

**Query 2 - List of Postal Code with most number of businesses operating, finding populated area in a smaller scale**

select business.postal\_code as Postal\_Code, count () as Number\_Of\_Business\_In\_The\_Area  
from review JOIN food\_business1 using (business\_id)  
group by postal\_code  
order by Number\_Of\_Business\_In\_The\_Area desc;

	City	NumberOfBusinesses
1	Las Vegas	3732
2	Phoenix	1434
3	Toronto	911
4	Scottsdale	836
5	Charlotte	586

**Query 3 - List of cities with most number of businesses (only considering 5 star ratings reviews)**

select City, count() as NumberOfBusinesses  
from food\_business1 JOIN review using (business\_id)  
where review.stars = 5  
group by city  
order by NumberOfBusinesses desc;

	City	NumberOfBusinesses
1	Las Vegas	5982
2	Phoenix	2234
3	Toronto	1935
4	Scottsdale	1205
5	Pittsburgh	1008

**Query 4 - List of cities with most number of businesses (only considering 5 & 4 star ratings reviews)**

select city, count() as freq  
from food\_business1 JOIN review using (business\_id)

where review.stars >=4  
group by city  
order by freq desc;

	Area_Name	City	NumberOfBusinesses
1	NULL	Chandler	8910
2	The Strip	Las Vegas	583
3	Southeast	Las Vegas	343
4	Westside	Las Vegas	318
5	Ville-Marie	Montréal	316
6	Downtown	Pittsburgh	284
7	Spring Valley	Las Vegas	277
8	Chinatown	Las Vegas	229
9	Eastside	Las Vegas	222
10	Scarborough	Toronto	172

**Query 5 - List of specific Areas with most number of businesses**

select neighborhood as Area\_Name, City, count() as NumberOfBusinesses  
from food\_business1  
group by neighborhood  
order by NumberOfBusinesses desc;

	Business_Name	Ratings	Number_Of_Reviews	Types_Categories
1	Lip Smacking Foodie Tours	5.0	179	Food Tours,Restaurants,Event Planning & Services,Food,Hotels,Hi
2	Chica	4.5	264	Restaurants,Wine Bars,Mexican,Bars,Nightlife,Latin American
3	L'Atelier de Joël Robuchon	4.5	974	French,Restaurants
4	Honolulu Cookie Company	4.5	180	Bakeries,Desserts,Food
5	Amorino Gelato	4.5	178	Ice Cream & Frozen Yogurt,Specialty Food,Desserts,Food,Gelato,I
6	Picasso	4.5	789	Nightlife,Bars,Spanish,Restaurants,French
7	Luke's Lobster Las Vegas	4.5	368	Restaurants,Seafood
8	Fat Tuesday	4.5	226	American (New),Bars,Restaurants,Nightlife
9	Bruxie	4.5	907	Breakfast & Brunch,Food,Desserts,Restaurants,Waffles,Sandwich
10	SUGARCANE Raw Bar Grill - Las Vegas	4.5	214	American (Traditional),Tapas/Small Plates,Nightlife,Desserts,Wine
11	Le Cirque	4.5	642	French,Restaurants
12	Fat Tuesday	4.5	383	Beer Wine & Spirits,Nightlife,Food Bars

**Query 6 - Competitors List in Area 'The Strip' in Las Vegas**

select name as Business\_Name, stars as Ratings, review\_count as Number\_Of\_Reviews, categories as  
Types\_Categories  
from food\_business1  
where neighborhood like 'The Strip' and (review\_count >100)  
order by stars desc;

	categories
1	American (New),Cocktail Bars,Bars,Restaurants,Seafood,Nightlife,Mexican,Tacos
2	American (New),Food,Mexican,Fast Food,Restaurants
3	American (New),Restaurants,Fast Food,Mexican
4	American (New),Restaurants,Mexican
5	American (New),Restaurants,Mexican,Italian,Breakfast & Brunch
6	American (New),Specialty Food,Ethnic Food,Food,Restaurants,Mexican
7	American (Traditional),Mexican,Fast Food,Breakfast & Brunch,Restaurants
8	American (Traditional),Restaurants,Indian,Mexican
9	American (Traditional),Restaurants,Mexican
10	American (Traditional),Rotisserie Chicken,Chicken Shop,Mexican,Restaurants
11	American (Traditional).Seafood.Nightlife.Restaurants.Mexican.Bars

### Query 7 - Sub categories of each restaurant in Las Vegas

select categories

from business

where (categories like '%restaurant%' or categories like '%food%') and categories like '%Mexican%'  
group by categories;

	Country	Total
1	American	503475
2	Japanese	146179
3	Mexican	131874
4	Italian	85717
5	Thai	52764
6	Vietnamese	43697
7	Korean	38857
8	Taiwanese	31771
9	Indian	12043

### Query 8 - Total check-in for each type of restaurant

create temporary table T1 as

select R.city as city, R.categories as categories,

sum("time.Sunday.0:00") as "1",

sum("time.Sunday.2:00") as "2",

sum("time.Sunday.3:00") as "3",

sum("time.Sunday.6:00") as "4",

sum("time.Sunday.14:00") as "5",

sum("time.Sunday.16:00") as "6",

sum("time.Sunday.17:00") as "7",

sum("time.Sunday.18:00") as "8",

sum("time.Sunday.19:00") as "9",

sum("time.Sunday.20:00") as "10",

sum("time.Sunday.21:00") as "11",

sum("time.Sunday.23:00") as "12",

sum("time.Monday.4:00") as "13",

sum("time.Monday.11:00") as "14",

sum("time.Monday.12:00") as "15",

sum("time.Monday.14:00") as "16",

sum("time.Monday.15:00") as "17",

sum("time.Monday.17:00") as "18",

sum("time.Monday.18:00") as "19",

sum("time.Monday.19:00") as "20",

sum("time.Monday.20:00") as "21",

sum("time.Monday.23:00") as "22",

sum("time.Tuesday.4:00") as "23",

sum("time.Tuesday.12:00") as "24",

sum("time.Tuesday.13:00") as "25",

sum("time.Tuesday.15:00") as "26",

sum("time.Tuesday.16:00") as "27",

sum("time.Tuesday.17:00") as "28",

sum("time.Tuesday.18:00") as "29",

sum("time.Tuesday.20:00") as "30",

sum("time.Tuesday.21:00") as "31",

sum("time.Tuesday.23:00") as "32",

sum("time.Wednesday.0:00") as "33",

sum("time.Wednesday.1:00") as "34",

sum("time.Wednesday.2:00") as "35",

sum("time.Wednesday.6:00") as "36",

sum("time.Wednesday.11:00") as "37",

sum("time.Wednesday.13:00") as "38",

sum("time.Wednesday.14:00") as "39",

sum("time.Wednesday.17:00") as "40",

sum("time.Wednesday.18:00") as "41",

sum("time.Wednesday.19:00") as "42",

sum("time.Wednesday.20:00") as "43",

sum("time.Wednesday.21:00") as "44",

```

sum("time.Thursday.1:00") as "45",
sum("time.Thursday.2:00") as "46",
sum("time.Thursday.4:00") as "47",
sum("time.Thursday.13:00") as "48",
sum("time.Thursday.15:00") as "49",
sum("time.Thursday.18:00") as "50",
sum("time.Thursday.19:00") as "51",
sum("time.Thursday.20:00") as "52",
sum("time.Thursday.21:00") as "53",
sum("time.Thursday.22:00") as "54",
sum("time.Thursday.23:00") as "55",
sum("time.Friday.0:00") as "56",
sum("time.Friday.3:00") as "57",
sum("time.Friday.10:00") as "58",
sum("time.Friday.14:00") as "59",
sum("time.Friday.15:00") as "60",
sum("time.Friday.16:00") as "61",
sum("time.Friday.17:00") as "62",
sum("time.Friday.18:00") as "63",
sum("time.Friday.19:00") as "64",
sum("time.Friday.21:00") as "65",
sum("time.Friday.22:00") as "66",
sum("time.Friday.23:00") as "67",
sum("time.Saturday.0:00") as "68",
sum("time.Saturday.1:00") as "69",
sum("time.Saturday.2:00") as "70",
sum("time.Saturday.10:00") as "71",
sum("time.Saturday.12:00") as "72",
sum("time.Saturday.13:00") as "73",
sum("time.Saturday.14:00") as "74",
sum("time.Saturday.15:00") as "75",
sum("time.Saturday.16:00") as "76",
sum("time.Saturday.17:00") as "77",
sum("time.Saturday.18:00") as "78",
sum("time.Saturday.21:00") as "79",
sum("time.Saturday.23:00") as "80"
from checkin as L , business as R
where L.business_id like R.business_id and R.city = 'Las Vegas'
group by R.categories;

```

```

create temporary table T2 as
select categories,
sum("1"+"2"+"3"+"4"+"5"+"6"+"7"+"8"+"9"+"10"
+"11"+"12"+"13"+"14"+"15"+"16"+"17"+"18"+"19"+"20"
+"21"+"22"+"23"+"24"+"25"+"26"+"27"+"28"+"29"+"30"
+"31"+"32"+"33"+"34"+"35"+"36"+"37"+"38"+"39"+"40"
+"41"+"42"+"43"+"44"+"45"+"46"+"47"+"48"+"49"+"50"
+"51"+"52"+"53"+"54"+"55"+"56"+"57"+"58"+"59"+"60"
+"61"+"62"+"63"+"64"+"65"+"66"+"67"+"68"+"69"+"70"

```

```

+"71"+"72"+"73"+"74"+"75"+"76"+"77"+"78"+"79"+"80") as
total
from T1
where (categories like '%restaurant%' or categories like
'%food%')
group by categories;

```

```

create temporary table T3 as
select "Mexican" as Country, sum(total) as Total
from T2
where categories like '%Mexican%'
union
select "Taiwanese" as Country, sum(total) as Total
from T2
where categories like '%Taiwanese%'
union
select "American" as Country, sum(total) as Total
from T2
where categories like '%American%'
union
select "Japanese" as Country, sum(total) as Total
from T2
where categories like '%Japanese%'
union
select "Thai" as Country, sum(total) as Total
from T2
where categories like '%Thai%'
union
select "Vietnamese" as Country, sum(total) as Total
from T2
where categories like '%Vietnamese%'
union
select "Italian" as Country, sum(total) as Total
from T2
where categories like '%Italian%'
union
select "Korean" as Country, sum(total) as Total
from T2
where categories like '%Korean%'
union
select "Indian" as Country, sum(total) as Total
from T2
where categories like '%Indian%'

```

```

select *
from T3
order by Total desc;

```



	categories	freq_review	freq_business	avgstar
1	American	2695	687	3.75807050092764
2	Mexican	788	291	3.60406091370558
3	Japanese	737	215	3.88738127544098
4	Italian	645	212	3.74728682170543
5	Thai	267	81	3.9250936329588
6	Korean	195	54	3.87692307692308
7	Vietnamese	172	51	3.72674418604651
8	Indian	91	29	3.97802197802198
9	Taiwanese	65	18	3.78461538461538

**Query 9** - The information of review from there different type of restaurant

```

create temporary table T4 as
select "Mexican" as categories , count(distinct R.review_id) as
freq_review, count(distinct L.business_id) as freq_business ,
avg(R.stars) as avgstar
from business as L , review as R
where L.business_id = R.business_id and L.city = 'Las Vegas'
and L.categories like '%Mexican%'
union
select "Taiwanese" as categories , count(distinct R.review_id) as
freq_review, count(distinct L.business_id) as freq_business ,
avg(R.stars) as avgstar
from business as L , review as R
where L.business_id = R.business_id and L.city = 'Las Vegas'
and L.categories like '%Taiwanese%'
union
select "Japanese" as categories , count(distinct R.review_id) as
freq_review, count(distinct L.business_id) as freq_business ,
avg(R.stars) as avgstar
from business as L , review as R
where L.business_id = R.business_id and L.city = 'Las Vegas'
and L.categories like '%Japanese%'
union
select "Thai" as categories , count(distinct R.review_id) as
freq_review, count(distinct L.business_id) as freq_business ,
avg(R.stars) as avgstar
from business as L , review as R
where L.business_id = R.business_id and L.city = 'Las Vegas'
and L.categories like '%Thai%'
union
select "Vietnamese" as categories , count(distinct R.review_id) as
freq_review, count(distinct L.business_id) as freq_business ,
avg(R.stars) as avgstar
from business as L , review as R
where L.business_id = R.business_id and L.city = 'Las Vegas'
and L.categories like '%Vietnamese%'

```

```

union
select "Italian" as categories , count(distinct R.review_id) as
freq_review, count(distinct L.business_id) as freq_business ,
avg(R.stars) as avgstar
from business as L , review as R
where L.business_id = R.business_id and L.city = 'Las Vegas'
and L.categories like '%Italian%'
union
select "Korean" as categories , count(distinct R.review_id) as
freq_review, count(distinct L.business_id) as freq_business ,
avg(R.stars) as avgstar
from business as L , review as R
where L.business_id = R.business_id and L.city = 'Las Vegas'
and L.categories like '%Korean%'
union
select "Indian" as categories , count(distinct R.review_id) as
freq_review, count(distinct L.business_id) as freq_business ,
avg(R.stars) as avgstar
from business as L , review as R
where L.business_id = R.business_id and L.city = 'Las Vegas'
and L.categories like '%Indian%'
union
select "American" as categories , count(distinct R.review_id) as
freq_review, count(distinct L.business_id) as freq_business ,
avg(R.stars) as avgstar
from business as L , review as R
where L.business_id = R.business_id and L.city = 'Las Vegas'
and L.categories like '%American%'
;

select *
from T4
order by freq_review desc;

```

	stars	avgcheckin
1	1.5	72.0
2	2.0	173.6666666666667
3	2.5	229.51724137931
4	3.0	263.036363636364
5	3.5	414.146341463415
6	4.0	709.915492957746
7	4.5	612.225
8	5.0	63.6666666666667

### Query 10 - The relationship between stars and check-in frequency

create temporary table T5 as

```
select R.business_id as business_id,
sum("time.Sunday.0:00") as "1",
sum("time.Sunday.2:00") as "2",
sum("time.Sunday.3:00") as "3",
sum("time.Sunday.6:00") as "4",
sum("time.Sunday.14:00") as "5",
sum("time.Sunday.16:00") as "6",
sum("time.Sunday.17:00") as "7",
sum("time.Sunday.18:00") as "8",
sum("time.Sunday.19:00") as "9",
sum("time.Sunday.20:00") as "10",
sum("time.Sunday.21:00") as "11",
sum("time.Sunday.23:00") as "12",
```

```
sum("time.Monday.4:00") as "13",
sum("time.Monday.11:00") as "14",
sum("time.Monday.12:00") as "15",
sum("time.Monday.14:00") as "16",
sum("time.Monday.15:00") as "17",
sum("time.Monday.17:00") as "18",
sum("time.Monday.18:00") as "19",
sum("time.Monday.19:00") as "20",
sum("time.Monday.20:00") as "21",
sum("time.Monday.23:00") as "22",
```

```
sum("time.Tuesday.4:00") as "23",
sum("time.Tuesday.12:00") as "24",
sum("time.Tuesday.13:00") as "25",
sum("time.Tuesday.15:00") as "26",
sum("time.Tuesday.16:00") as "27",
sum("time.Tuesday.17:00") as "28",
sum("time.Tuesday.18:00") as "29",
sum("time.Tuesday.20:00") as "30",
sum("time.Tuesday.21:00") as "31",
sum("time.Tuesday.23:00") as "32",
```

```
sum("time.Wednesday.0:00") as "33",
sum("time.Wednesday.1:00") as "34",
sum("time.Wednesday.2:00") as "35",
sum("time.Wednesday.6:00") as "36",
sum("time.Wednesday.11:00") as "37",
sum("time.Wednesday.13:00") as "38",
sum("time.Wednesday.14:00") as "39",
```

```
sum("time.Wednesday.17:00") as "40",
sum("time.Wednesday.18:00") as "41",
sum("time.Wednesday.19:00") as "42",
sum("time.Wednesday.20:00") as "43",
sum("time.Wednesday.21:00") as "44",
```

```
sum("time.Thursday.1:00") as "45",
sum("time.Thursday.2:00") as "46",
sum("time.Thursday.4:00") as "47",
sum("time.Thursday.13:00") as "48",
sum("time.Thursday.15:00") as "49",
sum("time.Thursday.18:00") as "50",
sum("time.Thursday.19:00") as "51",
sum("time.Thursday.20:00") as "52",
sum("time.Thursday.21:00") as "53",
sum("time.Thursday.22:00") as "54",
sum("time.Thursday.23:00") as "55",
```

```
sum("time.Friday.0:00") as "56",
sum("time.Friday.3:00") as "57",
sum("time.Friday.10:00") as "58",
sum("time.Friday.14:00") as "59",
sum("time.Friday.15:00") as "60",
sum("time.Friday.16:00") as "61",
sum("time.Friday.17:00") as "62",
sum("time.Friday.18:00") as "63",
sum("time.Friday.19:00") as "64",
sum("time.Friday.21:00") as "65",
sum("time.Friday.22:00") as "66",
sum("time.Friday.23:00") as "67",
```

```
sum("time.Saturday.0:00") as "68",
sum("time.Saturday.1:00") as "69",
sum("time.Saturday.2:00") as "70",
sum("time.Saturday.10:00") as "71",
sum("time.Saturday.12:00") as "72",
sum("time.Saturday.13:00") as "73",
sum("time.Saturday.14:00") as "74",
sum("time.Saturday.15:00") as "75",
sum("time.Saturday.16:00") as "76",
sum("time.Saturday.17:00") as "77",
sum("time.Saturday.18:00") as "78",
sum("time.Saturday.21:00") as "79",
sum("time.Saturday.23:00") as "80"
```

```

from checkin as L , business as R
where L.business_id like R.business_id and R.city = 'Las Vegas'
and R.categories like '%Mexican%'
group by L.business_id;

```

```

create temporary table T6 as
select business_id,
sum("1"+"2"+"3"+"4"+"5"+"6"+"7"+"8"+"9"+"10"
+"11"+"12"+"13"+"14"+"15"+"16"+"17"+"18"+"19"+"20"
+"21"+"22"+"23"+"24"+"25"+"26"+"27"+"28"+"29"+"30"
+"31"+"32"+"33"+"34"+"35"+"36"+"37"+"38"+"39"+"40"
+"41"+"42"+"43"+"44"+"45"+"46"+"47"+"48"+"49"+"50"
+"51"+"52"+"53"+"54"+"55"+"56"+"57"+"58"+"59"+"60"
+"61"+"62"+"63"+"64"+"65"+"66"+"67"+"68"+"69"+"70"
+"71"+"72"+"73"+"74"+"75"+"76"+"77"+"78"+"79"+"80") as
total
from T5
group by business_id;

```

```

-- the relation between rating and checkin rate
select L.stars, Avg(R.total) as avgcheckin
from business as L, T6 as R
where L.business_id = R.business_id and L.stars = 5.0
union
select L.stars, Avg(R.total) as avgcheckin
from business as L, T6 as R

```

```

where L.business_id = R.business_id and L.stars = 4.5
union
select L.stars, Avg(R.total) as avgcheckin
from business as L, T6 as R
where L.business_id = R.business_id and L.stars = 4.0
union
select L.stars, Avg(R.total) as avgcheckin
from business as L, T6 as R
where L.business_id = R.business_id and L.stars = 3.5
union
select L.stars, Avg(R.total) as avgcheckin
from business as L, T6 as R
where L.business_id = R.business_id and L.stars = 3.0
union
select L.stars, Avg(R.total) as avgcheckin
from business as L, T6 as R
where L.business_id = R.business_id and L.stars = 2.5
union
select L.stars, Avg(R.total) as avgcheckin
from business as L, T6 as R
where L.business_id = R.business_id and L.stars = 2.0
union
select L.stars, Avg(R.total) as avgcheckin
from business as L, T6 as R
where L.business_id = R.business_id and L.stars = 1.5;

```



**Query 11 -** Check-in rate for each period during the week of Mexican restaurants.

```

create temporary table T7 as
select "Mexican" as Country,
sum("time.Sunday.0:00") as "Sunday.00:00",
sum("time.Sunday.2:00") as "Sunday.02:00",
sum("time.Sunday.3:00") as "Sunday.03:00",
sum("time.Sunday.6:00") as "Sunday.06:00",
sum("time.Sunday.14:00") as "Sunday.14:00",
sum("time.Sunday.16:00") as "Sunday.16:00",
sum("time.Sunday.17:00") as "Sunday.17:00",
sum("time.Sunday.18:00") as "Sunday.18:00",
sum("time.Sunday.19:00") as "Sunday.19:00",

```

```

sum("time.Sunday.20:00") as "Sunday.20:00",
sum("time.Sunday.21:00") as "Sunday.21:00",
sum("time.Sunday.23:00") as "Sunday.23:00",
sum("time.Monday.4:00") as "Monday.04:00",
sum("time.Monday.11:00") as "Monday.11:00",
sum("time.Monday.12:00") as "Monday.12:00",
sum("time.Monday.14:00") as "Monday.14:00",
sum("time.Monday.15:00") as "Monday.15:00",
sum("time.Monday.17:00") as "Monday.17:00",
sum("time.Monday.18:00") as "Monday.18:00",
sum("time.Monday.19:00") as "Monday.19:00",

```

```

sum("time.Monday.20:00") as "Monday.20:00",
sum("time.Monday.23:00") as "Monday.23:00",
sum("time.Tuesday.4:00") as "Tuesday.04:00",
sum("time.Tuesday.12:00") as "Tuesday.12:00",
sum("time.Tuesday.13:00") as "Tuesday.13:00",
sum("time.Tuesday.15:00") as "Tuesday.15:00",
sum("time.Tuesday.16:00") as "Tuesday.16:00",
sum("time.Tuesday.17:00") as "Tuesday.17:00",
sum("time.Tuesday.18:00") as "Tuesday.18:00",
sum("time.Tuesday.20:00") as "Tuesday.20:00",
sum("time.Tuesday.21:00") as "Tuesday.21:00",
sum("time.Tuesday.23:00") as "Tuesday.23:00",
sum("time.Wednesday.0:00") as "Wednesday.00:00",
sum("time.Wednesday.1:00") as "Wednesday.01:00",
sum("time.Wednesday.2:00") as "Wednesday.02:00",
sum("time.Wednesday.6:00") as "Wednesday.06:00",
sum("time.Wednesday.11:00") as "Wednesday.11:00",
sum("time.Wednesday.13:00") as "Wednesday.13:00",
sum("time.Wednesday.14:00") as "Wednesday.14:00",
sum("time.Wednesday.17:00") as "Wednesday.17:00",
sum("time.Wednesday.18:00") as "Wednesday.18:00",
sum("time.Wednesday.19:00") as "Wednesday.19:00",
sum("time.Wednesday.20:00") as "Wednesday.20:00",
sum("time.Wednesday.21:00") as "Wednesday.21:00",
sum("time.Thursday.1:00") as "Thursday.01:00",
sum("time.Thursday.2:00") as "Thursday.02:00",
sum("time.Thursday.4:00") as "Thursday.04:00",
sum("time.Thursday.13:00") as "Thursday.13:00",
sum("time.Thursday.15:00") as "Thursday.15:00",
sum("time.Thursday.18:00") as "Thursday.18:00",
sum("time.Thursday.19:00") as "Thursday.19:00",
sum("time.Thursday.20:00") as "Thursday.20:00",

```

```

sum("time.Thursday.21:00") as "Thursday.21:00",
sum("time.Thursday.22:00") as "Thursday.22:00",
sum("time.Thursday.23:00") as "Thursday.23:00",
sum("time.Friday.0:00") as "Friday.00:00",
sum("time.Friday.3:00") as "Friday.03:00",
sum("time.Friday.10:00") as "Friday.10:00",
sum("time.Friday.14:00") as "Friday.14:00",
sum("time.Friday.15:00") as "Friday.15:00",
sum("time.Friday.16:00") as "Friday.16:00",
sum("time.Friday.17:00") as "Friday.17:00",
sum("time.Friday.18:00") as "Friday.18:00",
sum("time.Friday.19:00") as "Friday.19:00",
sum("time.Friday.21:00") as "Friday.21:00",
sum("time.Friday.22:00") as "Friday.22:00",
sum("time.Friday.23:00") as "Friday.23:00",
sum("time.Saturday.0:00") as "Saturday.00:00",
sum("time.Saturday.1:00") as "Saturday.01:00",
sum("time.Saturday.2:00") as "Saturday.02:00",
sum("time.Saturday.10:00") as "Saturday.10:00",
sum("time.Saturday.12:00") as "Saturday.12:00",
sum("time.Saturday.13:00") as "Saturday.13:00",
sum("time.Saturday.14:00") as "Saturday.14:00",
sum("time.Saturday.15:00") as "Saturday.15:00",
sum("time.Saturday.16:00") as "Saturday.16:00",
sum("time.Saturday.17:00") as "Saturday.17:00",
sum("time.Saturday.18:00") as "Saturday.18:00",
sum("time.Saturday.21:00") as "Saturday.21:00",
sum("time.Saturday.23:00") as "Saturday.23:00"
from checkin as L , business as R
where L.business_id like R.business_id and R.city = 'Las Vegas'
and categories like "%Mexican%"

```

business_id	name	stars	categories
WRbBmXMHmdWokeGp21OoeA	Mamma Olivia	5	Restaurants,Italian,Pizza
VbaI-AZnu2w6NtIKfdjDg	Desert Wind Coffee Roasters	5	Coffee Roasteries,Food,Internet Cafes,Coffee & Tea
3p5Ur_cdrphurO6m1HMP9A	J Karaoke Bar	5	Bars,American (New),Karaoke,Nightlife,Asian Fusion,Cocktail Bars,Restaurants
IhNA5EZ3XnBHmuuVnWdIwA	Brew Tea Bar	5	Cafes,Tea Rooms,Food,Bubble Tea,Restaurants,Desserts
L1-1P3acJc4gEFvWwjXcNQ	Meraki Greek Grill	5	Restaurants,Greek
N8RvK4XrKaHYXninxg9Q	Renaissance Catering	5	Food,Beer,Wine & Spirits,Event Planning & Services,Specialty Food,Cate
hFT3HuxVZJb1uYEsYfp-0A	Blacqat Ultra Hookah Lounge	5	Adult Entertainment,Lounges,Hookah Bars,Restaurants,Coffee & Tea,B
MxGlt2k5ra81sOFLo1dBDg	Original Meal Prep Vegas	5	Food Delivery Services,Food
EkYbGk9L5KdCfL3ocZ7ETQ	Vegas Discount Nutrition Sup...	5	Shopping,Specialty Food,Health Markets,Food,Vitamins & Supplements,Health & Medical,...
GzBIAJAMHD9iVZK8BT6XMw	Tacos Huevos	5	Breakfast & Brunch,Mexican,Restaurants,Desserts,Tacos,Soup,Food
3ZJWwhoH-Kiu1wRFTvHvW9Q	Pie Me Over	5	Do-It-Yourself Food,Restaurants,Food,Desserts,American (Traditional),Bakeries

Column: categories  
Data type:  
Table: business  
ROWID: []

### Query 12 - Restaurants in Las Vegas that have been associated with service related comments

select distinct business\_id, name, stars, categories--For companies with tip reviews on service, what are their star ratings?

from (

select tip.business\_id, text, business.stars, name, city, categories

from tip

inner join business on tip.business\_id = business.business\_id

union all

select review.business\_id, text, business.stars, name, city, categories

```

from review
inner join business on review.business_id = business.business_id
)
where (text like '%service%' or text like '%staff%') and city like '%Vegas%' and (categories like
'%restaurant%' or categories like '%food%')
order by stars desc
;

```

	rating	NumberOfBusiness
1	5	29
2	4.5	753
3	4	1704
4	3.5	1042
5	3	436
6	2.5	200
7	2	68
8	1.5	15

***Query 13 - Ratings of Restaurant associated with service related comments***

```

select stars as rating, count(*) as NumberOfBusiness
  from (
    select tip.business_id, text, business.stars, name, city, categories
    from tip
    inner join business on tip.business_id = business.business_id
    union all
    select review.business_id, text, business.stars, name, city, categories
    from review
    inner join business on review.business_id = business.business_id
  )
where (text like '%service%' or text like '%staff%') and city like '%Vegas%' and (categories like
'%restaurant%' or categories like '%food%')
group by stars
order by stars desc;

```

	rating	No_of_Chicken_related_Review
1	5	3
2	4.5	215
3	4	442
4	3.5	309
5	3	110
6	2.5	40
7	2	25
8	1.5	5

***Query 14 - Ratings of Restaurant associated with chicken related comments***

```

select stars as ratings, count(*) as No_of_Chicken_related_Review
  from (
    select tip.business_id, text, business.stars, name, city, categories
    from tip
    inner join business on tip.business_id = business.business_id
    union all
    select review.business_id, text, business.stars, name, city, categories
    from review
    inner join business on review.business_id = business.business_id
  )
where text like '%chicken%' and city like '%Vegas%' and (categories like '%restaurant%' or categories
like '%food%')
group by stars
order by stars desc;

```

	ratings	No_of_Pizza_related_Review
1	4.5	64
2	4	248
3	3.5	143
4	3	52
5	2.5	26
6	2	16
7	1.5	2

***Query 15 - Ratings of Restaurant associated with pizza related comments***

```

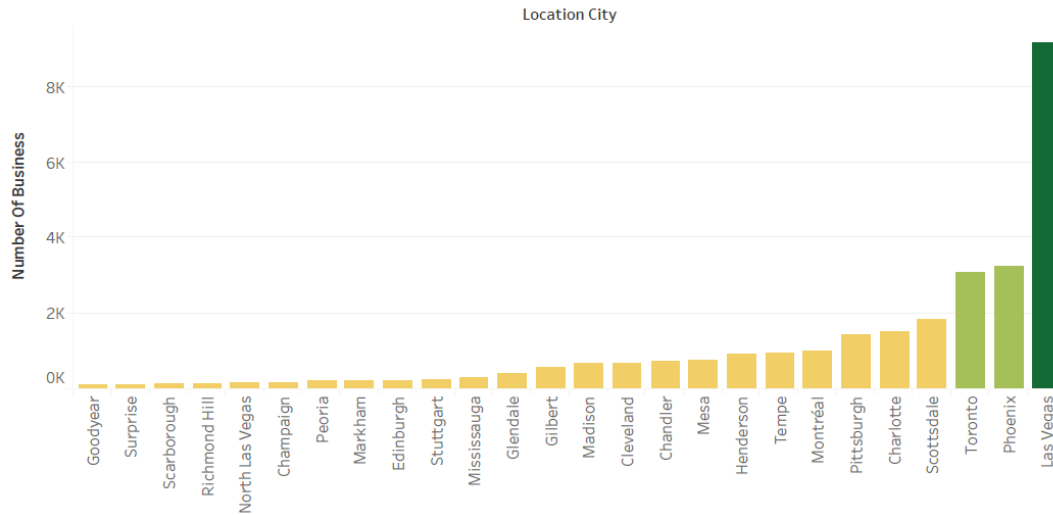
select stars as ratings, count(*) as No_of_Pizza_related_Review
  from (
    select tip.business_id, text, business.stars, name, city, categories
    from tip
    inner join business on tip.business_id = business.business_id
    union all
    select review.business_id, text, business.stars, name, city, categories
    from review
    inner join business on review.business_id = business.business_id
  )
where text like '%pizza%' and city like '%Vegas%' and (categories like '%restaurant%' or categories like
'%food%')
group by stars
order by stars desc;

```

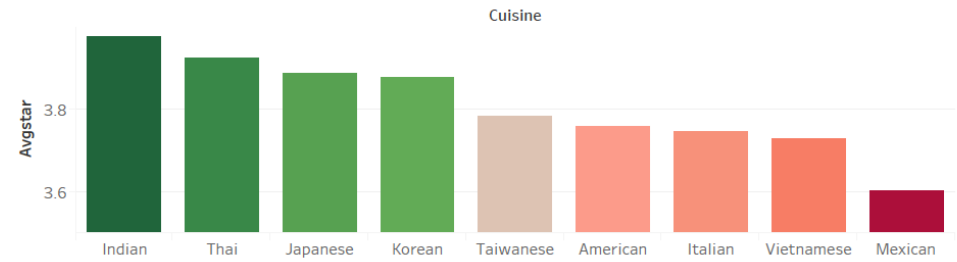
## Yelp Dataset Summary

HEESUN CHANG (HC483), PATCHARA SUENSILPONG (PS972)  
SEBASTIAN KOH (YK784), SI PARK (SP2477)

### Cities With the Most Restaurant Businesses



### Cuisines Ranked by Average Rating



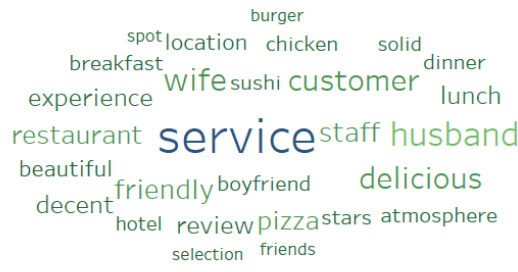
### Most Popular Cuisines



### Most Popular Words in Yelp Tips



### Most Popular Words in Yelp Reviews



### Check-in Rate for Mexican Restaurants

