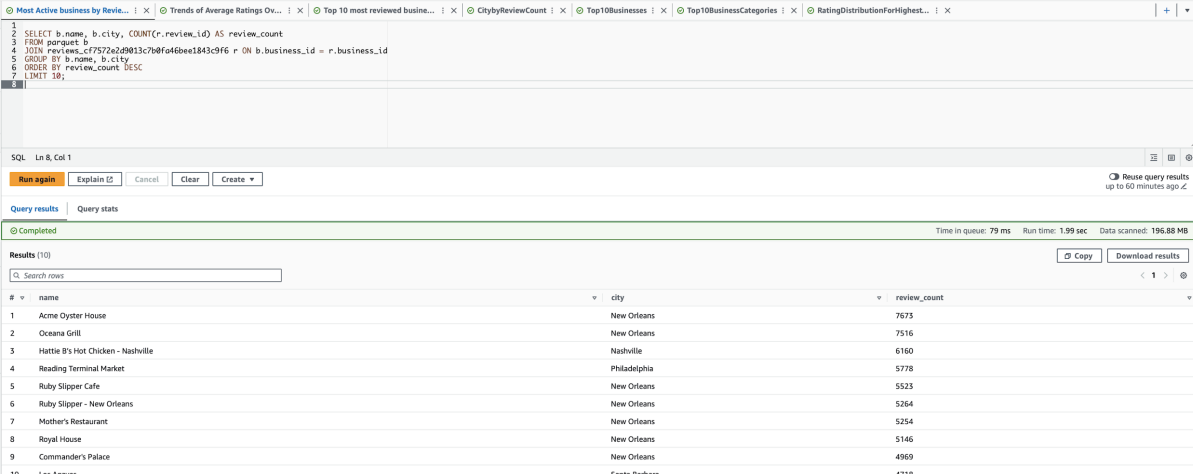


# Business Analysis Using Athena Queries

## 1. Most Active Business by Review Count

```
SELECT b.name, b.city, COUNT(r.review_id) AS review_count
FROM parquet b
JOIN reviews_cf7572e2d9013c7b0fa46bee1843c9f6 r ON b.business_id = r.business_id
GROUP BY b.name, b.city
ORDER BY review_count DESC
LIMIT 10;
```



SQL Ln 8, Col 1

Run again Explain Explain Explain Cancel Clear Create

Query results Query stats

Completed Time in queue: 79 ms Run time: 1.99 sec Data scanned: 196.88 MB

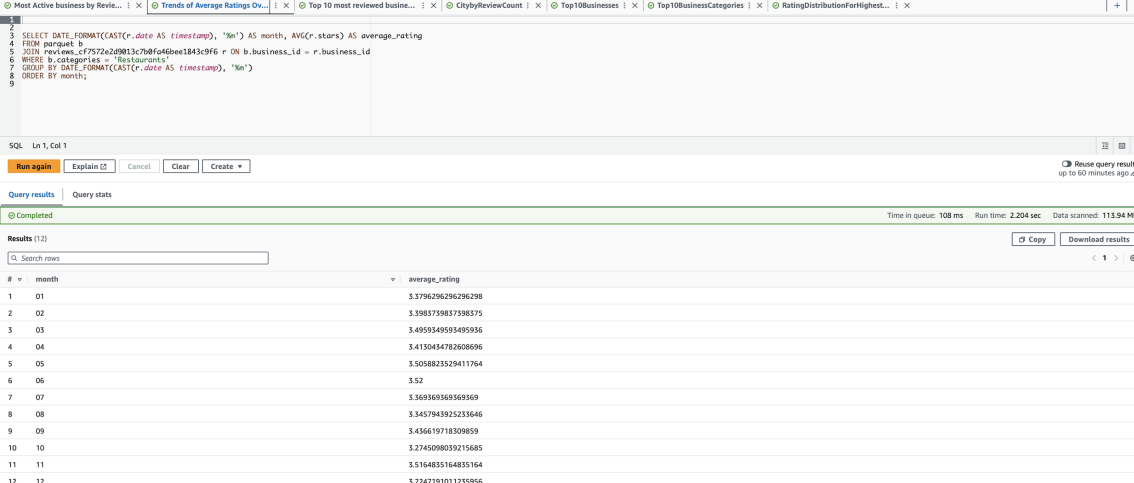
Results (10)

Search rows

#	name	city	review_count
1	Acme Oyster House	New Orleans	7673
2	Oceana Grill	New Orleans	7516
3	Hattie B's Hot Chicken - Nashville	Nashville	6160
4	Reading Terminal Market	Philadelphia	5778
5	Ruby Slipper Cafe	New Orleans	5523
6	Ruby Slipper - New Orleans	New Orleans	5264
7	Mother's Restaurant	New Orleans	5254
8	Royal House	New Orleans	5146
9	Commander's Palace	New Orleans	4969
10	Los Agaves	Santa Barbara	4718

## 2. Trends of Average Ratings Over Time for a Specific Category

```
SELECT DATE_FORMAT(CAST(r.date AS timestamp), '%m') AS month, AVG(r.stars) AS
average_rating
FROM parquet b
JOIN reviews_cf7572e2d9013c7b0fa46bee1843c9f6 r ON b.business_id = r.business_id
WHERE b.categories = 'Restaurants'
GROUP BY DATE_FORMAT(CAST(r.date AS timestamp), '%m')
ORDER BY month;
```



SQL Ln 1, Col 1

Run again Explain Explain Explain Cancel Clear Create

Query results Query stats

Completed Time in queue: 108 ms Run time: 2.204 sec Data scanned: 113.94 MB

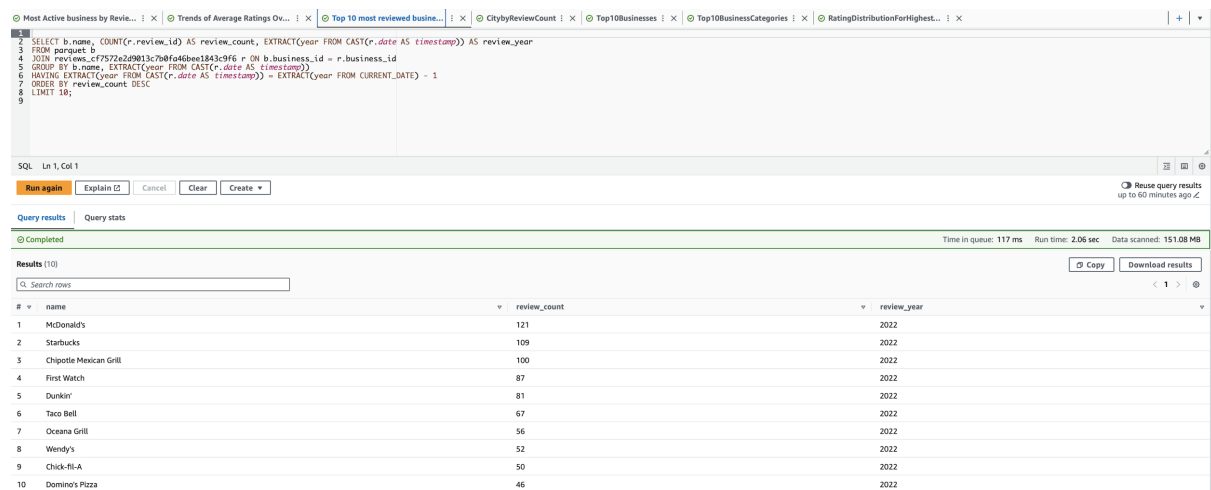
Results (12)

Search rows

#	month	average_rating
1	01	3.3796296296296298
2	02	3.3983739837398375
3	03	3.409349593495936
4	04	3.4130434782608696
5	05	3.5008823529411764
6	06	3.52
7	07	3.369369369369369
8	08	3.3457943925233646
9	09	3.436619718309859
10	10	3.2745098039215685
11	11	3.5164835164835164
12	12	3.2247191011235956

### 3. Top 10 Most Reviewed Businesses in the Previous Year

```
SELECT b.name, COUNT(r.review_id) AS review_count, EXTRACT(year FROM
CAST(r.date AS timestamp)) AS review_year
FROM parquet b
JOIN reviews_cf7572e2d9013c7b0fa46bee1843c9f6 r ON b.business_id = r.business_id
GROUP BY b.name, EXTRACT(year FROM CAST(r.date AS timestamp))
HAVING EXTRACT(year FROM CAST(r.date AS timestamp)) = EXTRACT(year FROM
CURRENT_DATE) - 1
ORDER BY review_count DESC
LIMIT 10;
```

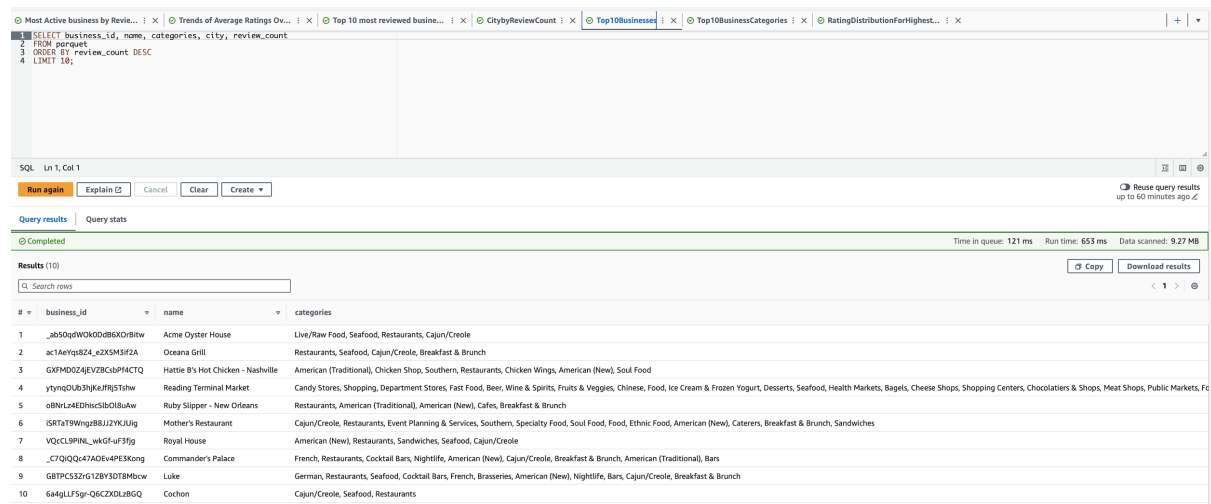


```
SELECT b.name, COUNT(r.review_id) AS review_count, EXTRACT(year FROM CAST(r.date AS timestamp)) AS review_year
FROM parquet b
JOIN reviews_cf7572e2d9013c7b0fa46bee1843c9f6 r ON b.business_id = r.business_id
GROUP BY b.name, EXTRACT(year FROM CAST(r.date AS timestamp))
HAVING EXTRACT(year FROM CAST(r.date AS timestamp)) = EXTRACT(year FROM CURRENT_DATE) - 1
ORDER BY review_count DESC
LIMIT 10;
```

#	name	review_count	review_year
1	McDonald's	121	2022
2	Starbucks	109	2022
3	Chipotle Mexican Grill	100	2022
4	First Watch	87	2022
5	Dunkin'	81	2022
6	Taco Bell	67	2022
7	Oceana Grill	56	2022
8	Wendy's	52	2022
9	Chick-Ri-A	50	2022
10	Dominos Pizza	46	2022

### 4. Top 10 Businesses

```
SELECT business_id, name, categories, city, review_count
FROM parquet
ORDER BY review_count DESC
LIMIT 10;
```

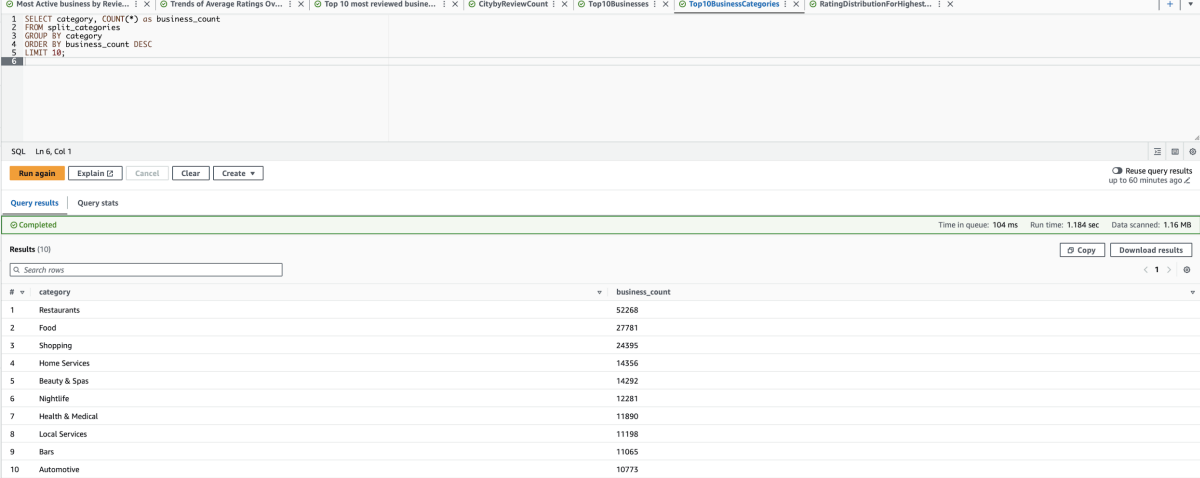


```
SELECT business_id, name, categories, city, review_count
FROM parquet
ORDER BY review_count DESC
LIMIT 10;
```

#	business_id	name	categories
1	_ab50qWOK0ddB6X0Bltw	Acme Oyster House	Live/Raw Food, Seafood, Restaurants, Cajun/Creole
2	ac1AeYq824_k2XSM3fZA	Oceana Grill	Restaurants, Seafood, Cajun/Creole, Breakfast & Brunch
3	GXFMD0Z4EVBZC8P4CQ	Hattie B's Hot Chicken - Nashville	American (Traditional), Chicken Shop, Southern, Restaurants, Chicken Wings, American (New), Soul Food
4	ytymQ0U3JhKcJfJfSThwr	Reading Terminal Market	Candy Stores, Shopping, Department Stores, Fast Food, Beer, Wine & Spirits, Fruits & Veggies, Chinese, Food, Ice Cream & Frozen Yogurt, Desserts, Seafood, Health Markets, Bagels, Cheese Shops, Shopping Centers, Chocolatiers & Shops, Meat Shops, Public Markets, Food Trucks
5	oBtML14EDh1c5t0l8uAw	Ruby Slipper - New Orleans	Restaurants, American (Traditional), American (New), Cafes, Breakfast & Brunch
6	i58Tat9Wng8BJZVYUJlg	Mother's Restaurant	Cajun/Creole, Restaurants, Event Planning & Services, Southern, Specialty Food, Soul Food, Food, Ethnic Food, American (New), Caterers, Breakfast & Brunch, Sandwiches
7	VQcL9PmL_wkG4-vF3fg	Royal House	American (New), Restaurants, Sandwiches, Seafood, Cajun/Creole
8	_C7QIQQ47A0Ev4PE3Kong	Commander's Palace	French, Restaurants, Cocktail Bars, Nightlife, American (New), Cajun/Creole, Breakfast & Brunch, American (Traditional), Bars
9	G8TPC32G1ZB93D8bcw	Luke	German, Restaurants, Seafood, Cocktail Bars, French, Brasseries, American (New), Nightlife, Bars, Cajun/Creole, Breakfast & Brunch
10	6a4gLLF5p-Q6CXDL8GQ	Cochon	Cajun/Creole, Seafood, Restaurants

## 5. Top 10 Business Categories

```
SELECT category, COUNT(*) AS business_count
FROM split_categories
GROUP BY category
ORDER BY business_count DESC
LIMIT 10;
```



The screenshot shows a SQL query editor with the following query:

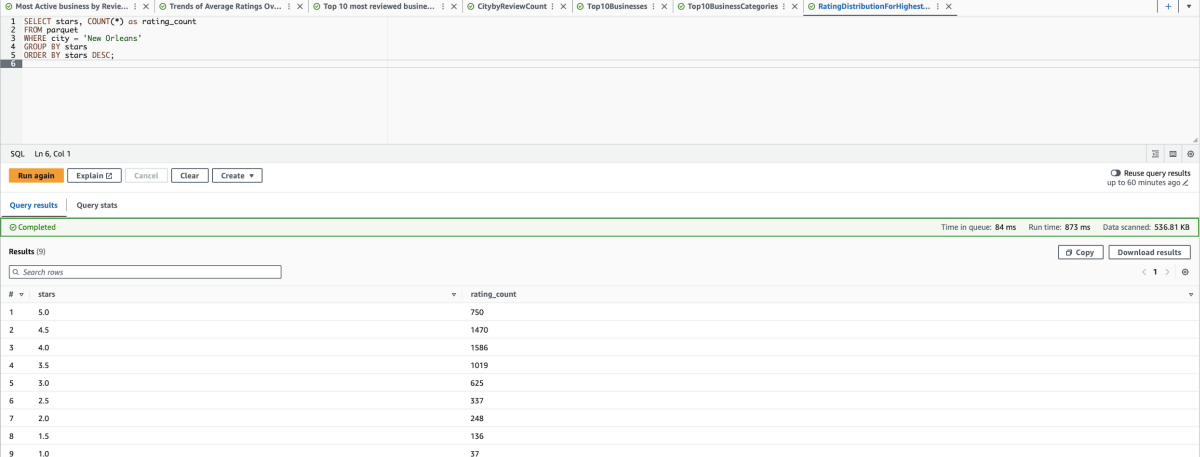
```
1 SELECT category, COUNT(*) AS business_count
2 FROM split_categories
3 GROUP BY category
4 ORDER BY business_count DESC
5 LIMIT 10;
```

The query results are displayed in a table with 10 rows and 2 columns: category and business\_count.

#	category	business_count
1	Restaurants	52268
2	Food	27781
3	Shopping	24395
4	Home Services	14356
5	Beauty & Spas	14292
6	Nightlife	12281
7	Health & Medical	11890
8	Local Services	11198
9	Bars	11065
10	Automotive	10773

## 6. Rating Distribution for Highest Rated City

```
SELECT stars, COUNT(*) AS rating_count
FROM parquet
WHERE city = 'New Orleans'
GROUP BY stars
ORDER BY stars DESC;
```



The screenshot shows a SQL query editor with the following query:

```
1 SELECT stars, COUNT(*) AS rating_count
2 FROM parquet
3 WHERE city = 'New Orleans'
4 GROUP BY stars
5 ORDER BY stars DESC;
```

The query results are displayed in a table with 9 rows and 2 columns: stars and rating\_count.

#	stars	rating_count
1	5.0	750
2	4.5	1470
3	4.0	1586
4	3.5	1019
5	3.0	625
6	2.5	337
7	2.0	248
8	1.5	136
9	1.0	37

## 7. Businesses with Most 5-star Reviews

```
SELECT b.business_id, b.name, COUNT(*) AS five_star_reviews
FROM parquet b
JOIN reviews_cf7572e2d9013c7b0fa46bee1843c9f6 r ON b.business_id = r.business_id
WHERE r.stars = 5.0
GROUP BY b.business_id, b.name
ORDER BY five_star_reviews DESC
LIMIT 10;
```

The screenshot shows a SQL query execution interface. The query is: `SELECT b.business_id, b.name, COUNT(*) AS five_star_reviews FROM parquet b JOIN reviews_cf7572e2d9013c7b0fa46bee1843c9f6 r ON b.business_id = r.business_id WHERE r.stars = 5.0 GROUP BY b.business_id, b.name ORDER BY five_star_reviews DESC LIMIT 10;` The results table has 10 rows, showing business IDs, names, and their five-star review counts.

#	business_id	name	five_star_reviews
1	yymq0UbShKaJfRjStshw	Reading Terminal Market	4014
2	ac1aTqptkZ4_a2XSMHf3A	Oceana Grill	4012
3	GXFMD0Z4jEVZCubPFACTQ	Hattie B's Hot Chicken - Nashville	3838
4	_ab50qdwKODb8XOrBtw	Acme Oyster House	3645
5	_CFQZQ47AOeV4PE3Kong	Commander's Palace	3095
6	oBNLz4EDhicsB0lBuAw	Ruby Slipper - New Orleans	3075
7	yP5qj3_ernofedVYTbInZA	Los Agaves	2693
8	L_BLMZ_1m2m80uL0ePig	Pappy's Smokehouse	2675
9	gTCBQ_1BkytW5ly3TEvg	Cochon Butcher	2537
10	GBTPCS3ZG12BY3DT8Mbcw	Luke	2365

## 8. City and State Total Reviews

```
SELECT city, state, SUM(review_count) AS total_reviews
FROM parquet
GROUP BY city, state
ORDER BY total_reviews DESC;
```

The screenshot shows a SQL query execution interface. The query is: `SELECT city, state, SUM(review_count) AS total_reviews FROM parquet GROUP BY city, state ORDER BY total_reviews DESC;` The results table has 24 rows, showing city, state, and total review counts.

#	city	state	total_reviews
1	Philadelphia	PA	936144
2	New Orleans	LA	621321
3	Nashville	TN	440971
4	Tampa	FL	439402
5	Tucson	AZ	387201
6	Indianapolis	IN	349210
7	Reno	NV	334519
8	Santa Barbara	CA	262848
9	Saint Louis	MO	244339
10	Boise	ID	101876
11	Edmonton	AB	98197
12	Clearwater	FL	84190
13	Saint Petersburg	FL	76201
14	Sparks	NV	69519
15	Metairie	LA	61941
16	St. Louis	MO	61256
17	Franklin	TN	54688
18	St. Petersburg	FL	52620
19	Goleta	CA	44126
20	Wilmington	DE	42969
21	Brandon	FL	41782
22	Carmel	IN	39897
23	Cherry Hill	NJ	38336
24	Meridian	ID	35972

## 9. Yearly Growth in Number of Businesses

```
SELECT EXTRACT(year FROM CAST(r.date AS timestamp)) AS year,  
COUNT(DISTINCT b.business_id) AS number_of_businesses  
FROM parquet b  
JOIN reviews_cf7572e2d9013c7b0fa46bee1843c9f6 r ON b.business_id = r.business_id  
GROUP BY EXTRACT(year FROM CAST(r.date AS timestamp))  
ORDER BY year;
```

The screenshot shows a SQL query editor with the following query:

```
1 SELECT EXTRACT(year FROM CAST(r.date AS timestamp)) AS year, COUNT(DISTINCT b.business_id) AS number_of_businesses  
2 FROM parquet b  
3 JOIN reviews_cf7572e2d9013c7b0fa46bee1843c9f6 r ON b.business_id = r.business_id  
4 GROUP BY EXTRACT(year FROM CAST(r.date AS timestamp))  
5 ORDER BY year;
```

Below the query editor, the results are displayed in a table with 18 rows and 2 columns: year and number\_of\_businesses. The table shows a steady increase in the number of businesses over the years.

#	year	number_of_businesses
1	2005	674
2	2006	2449
3	2007	6351
4	2008	14681
5	2009	19220
6	2010	29429
7	2011	41612
8	2012	51453
9	2013	63179
10	2014	74165
11	2015	85409
12	2016	92335
13	2017	96456
14	2018	100885
15	2019	101904
16	2020	89917
17	2021	87467
18	2022	19096

## 10. Average Star Rating by State

```
SELECT state, AVG(stars) AS average_rating  
FROM parquet  
GROUP BY state  
ORDER BY average_rating DESC;
```

1 SELECT state, AVG(stars) AS average\_rating  
2 FROM reviews  
3 GROUP BY state  
4 ORDER BY average\_rating DESC;

SQL: Ln 5, Col 1

Run query Run query results up to 60 minutes ago

Query results Query status

Completed Time in queue: 109 ms Run time: 661 ms Data scanned: 155.91 KB

Results (27)

Search rows

#	state	average_rating
1	MT	5.0
2	SD	4.5
3	UT	4.5
4	VT	4.5
5	HI	4.25
6	CD	4.0
7	MS	4.0
8	CA	3.99711649505286
9	NV	3.73799384674876
10	ID	3.70941397649425
11	LA	3.67991676909055
12	IL	3.61767880107174
13	AZ	3.59320428638627
14	IN	3.58893287734595
15	PA	3.57444273366924
16	WA	3.5744033699751
17	MO	3.54703461465287
18	MA	3.5
19	NJ	3.46005439193116
20	AB	3.44788482835648
21	AL	3.27032817940286
22	DE	3.33764899627496
23	TX	2.875
24	VI	2.5
25	ME	2.5
26	NC	2.0
27	MA	1.25

## Saved Queries

Amazon Athena > Query editor

Editor Recent queries Saved queries Settings

Workgroup primary

Saved queries (10)

Search saved queries

ID	Name	Query	Description
d5997254-a297-4872-b062-ae7a14d2ad	Top 10 Business Categories	SELECT category, COUNT(*) as business_count FROM ipdc_categories GROUP BY category	
89767a38-5be7-40b6-b3dc-ea77f509402	Most Active business by Review Count	S.Most Active business by Review Count SELECT to.name, to.city, COUNT(review_id) AS re	
94693775-d7b6-4cfe-b8cd-8ca0663a7709	Average Star Rating by state	SELECT state, AVG(stars) AS average_rating FROM parquet GROUP BY state ORDER BY av	
21d0913d-df12-42a0-8745-5ab1ff6c9bde	CityandStateTotalReviews	SELECT city, state, SUM(review_count) AS total_reviews FROM parquet GROUP BY city, st	
81268671-9291-4946-9784-4691101913d8	Top 10 most reviewed businesses in the previous year	SELECT to.name, COUNT(review_id) AS review_count, EXTRACT(year FROM cast(to.date	
2e091733-8071-4d7a-57d8-83a8f1f5509f	Trends of Average Ratings Over Time for a Specific Category	SELECT DATE_FORMAT(CAST(to.date AS timestamp), 'mm') AS month, AVG(stars) AS ave	
783994fc-8656-4d75-b0ee-d948703062	RatingofBusinessofagivenBusinessCity	SELECT sum, COUNT(*) as rating_count FROM parquet WHERE city = 'New Orleans' GRO	
7d59593e-c5f1-4375-a331-77611d0a688e	BusinesseswithMostReviews	SELECT to.business_id, to.name, COUNT(*) AS top_star_reviews FROM business_0237855	
70598a6dc7a0c463043dfe19a81887328	Yearly Growth in number of businesses	SELECT EXTRACT(year FROM cast(to.date AS timestamp)) AS year, COUNT(DISTINCT to.bu	
65b8c1b2-7bea-4796-b4e5-82f533a632cd	Top10Businesses	SELECT business_id, name, categories, is_open, city, review_count FROM business_022a7...	