The aim of the project is to create the Geographic Information System (GIS) analysis. The database used in this project is a sql database postgreSQL. PostGIS package for the postgreSQL database was used. PostGIS is an open source software package that increases the capabilities of the database management system by adding support for spatial data. Here, spatial data refers to any kind of data that contains geographic or location information, i.e latitude, longitude, zip code, PostGIS allows you to store, manipulate and query the spatial data such as points, lines, and other dimensions within a database.

Execution of practical part

Dataset – The dataset required for the analysis of the spatial data was downloaded from https://simplemaps.com/data/world-cities

	A	B	C	D	E	F	G	H	I	J
1	city	city_ascii	lat	lng	country	iso2	iso3	admin_name	capital	population
2	Tokyo	Tokyo	35.6897	139.692	Japan	JP	JPN	Tōkyō	primary	37732000
3	Jakarta	Jakarta	-6.175	106.828	Indonesia	ID	IDN	Jakarta	primary	33756000
4	Delhi	Delhi	28.61	77.23	India	IN	IND	Delhi	admin	32226000
5	Guangzhou	Guangzh	23.13	113.26	China	CN	CHN	Guangdong	admin	26940000
6	Mumbai	Mumbai	19.0761	72.8775	India	IN	IND	Mahārāshtra	admin	24973000
7	Manila	Manila	14.5958	120.977	Philippine	PH	PHL	Manila	primary	24922000
8	Shanghai	Shanghai	31.1667	121.467	China	CN	CHN	Shanghai	admin	24073000
9	São Paulo	Sao Paulo	-23.55	-46.6333	Brazil	BR	BRA	São Paulo	admin	23086000
10	Seoul	Seoul	37.56	126.99	South Ko	KR	KOR	Seoul	primary	23016000
11	Mexico City	Mexico Ci	19.4333	-99.1333	Mexico	MX	MEX	Ciudad de Mé	primary	21804000
12	Cairo	Cairo	30.0444	31.2358	Egypt	EG	EGY	Al Qāhirah	primary	20296000
13	New York	New York	40.6943	-73.9249	United St	US	USA	New York		18972871
14	Dhaka	Dhaka	23.7639	90.3889	Banglade	BD	BGD	Dhaka	primary	18627000
15	Beijing	Beijing	39.904	116.408	China	CN	CHN	Beijing	primary	18522000
16	Kolkāta	Kolkata	22.5675	88.37	India	IN	IND	West Bengal	admin	18502000
17	Bangkok	Bangkok	13.7525	100.494	Thailand	TH	THA	Krung Thep M	primary	18007000
18	Shenzhen	Shenzhen	22.535	114.054	China	CN	CHN	Guangdong	minor	17619000
19	Moscow	Moscow	55.7558	37.6178	Russia	RU	RUS	Moskva	primary	17332000
20	Buenos Aires	Buenos A	-34.5997	-58.3819	Argentina	AR	ARG	Buenos Aires	primary	16710000
21	Lagos	Lagos	6.455	3.3841	Nigeria	NG	NGA	Lagos	minor	16637000
22	Istanbul	Istanbul	41.0136	28.955	Turkey	TR	TUR	İstanbul	admin	16079000
23	Karachi	Karachi	24.86	67.01	Pakistan	PK	PAK	Sindh	admin	15738000
24	Bangalore	Bangalore	12.9789	77.5917	India	IN	IND	Karnātaka	admin	15386000
25	Ho Chi Minh City	Ho Chi Mi	10.7756	106.702	Vietnam	VN	VNM	Hồ Chí Minh	admin	15136000
26	Ōsaka	Osaka	34.6939	135.502	Japan	JP	JPN	Ōsaka	admin	15126000
27	Chengdu	Chengdu	30.66	104.063	China	CN	CHN	Sichuan	admin	14645000
28	Tehran	Tehran	35.6892	51.3889	Iran	IR	IRN	Tehrān	primary	14148000
29	Kinshasa	Kinshasa	-4.325	15.3222	Congo (K	CD	COD	Kinshasa	primary	12836000
30	Rio de Janeiro	Rio de Jan	-22.9111	-43.2056	Brazil	BR	BRA	Rio de Janeiro	admin	12592000
31	Chennai	Chennai	13.0825	80.275	India	IN	IND	Tamil Nādu	admin	12395000
32	Xi'an	Xi'an	34.2667	108.9	China	CN	CHN	Shaanxi	admin	12328000
33	Lahore	Lahore	31.5497	74.3436	Pakistan	PK	PAK	Punjab	admin	12306000
34	Chongqing	Chongqin	29.55	106.507	China	CN	CHN	Chongqing	admin	12135000
35	Los Angeles	Los Ange	34.1141	-118.407	United St	US	USA	California		12121244
36	Baoding	Baoding	38.8671	115.485	China	CN	CHN	Hebei		11860000
37	London	London	51.5072	-0.1275	United Ki	GB	GBR	London, City	primary	11262000
38	Paris	Paris	48.8567	2.3522	France	FR	FRA	Île-de-France	primary	11060000
39	Linyi	Linyi	35.1041	118.35	China	CN	CHN	Shandong		11018365
40	Dongguan	Dongguan	23.0475	113.749	China	CN	CHN	Guangdong	minor	10646000

Dataset size – 44,690 tuples of world cities along with their informations like latitude, longitude and other information

Database creation and importing the data into table

create extension postgis;

select * from pg_catalog.pg_extension;

```
CREATE TABLE gis_location (
    id SERIAL PRIMARY KEY,
    country VARCHAR(50),
    city VARCHAR(50),
    lat NUMERIC(9, 6),
    lon NUMERIC(9, 6),
    population INT
);
```

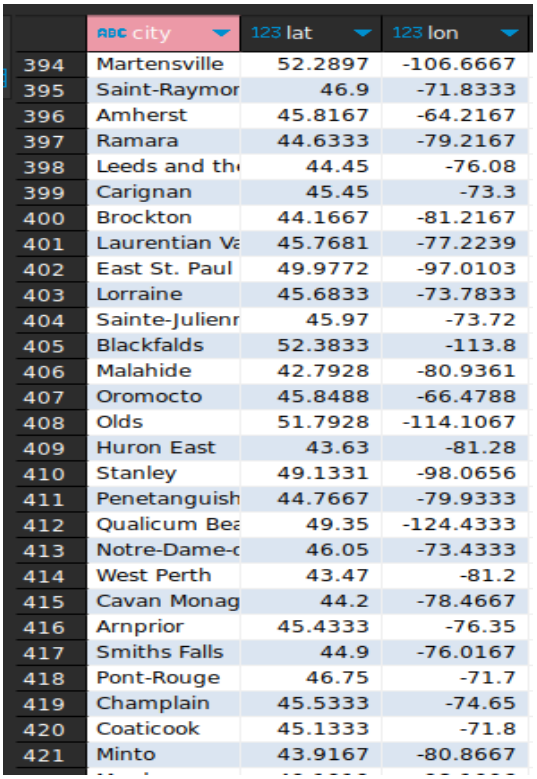
select table_name, column_name, data_type from information_schema.columns where table_name='gis_location';

copy gis_location(city, lat, lon, country, population) from '/tmp/worldcities.csv' delimiter ',' csv header;

1. Retrieve Locations of specific features

-- Retrive cities in a specific country

```
SELECT city, lat, lon FROM gis_location WHERE country = 'Canada';
```



	city	lat	lon
394	Martensville	52.2897	-106.6667
395	Saint-Raymor	46.9	-71.8333
396	Amherst	45.8167	-64.2167
397	Ramara	44.6333	-79.2167
398	Leeds and th	44.45	-76.08
399	Carignan	45.45	-73.3
400	Brockton	44.1667	-81.2167
401	Laurentian Va	45.7681	-77.2239
402	East St. Paul	49.9772	-97.0103
403	Lorraine	45.6833	-73.7833
404	Sainte-Julienr	45.97	-73.72
405	Blackfalds	52.3833	-113.8
406	Malahide	42.7928	-80.9361
407	Oromocto	45.8488	-66.4788
408	Olds	51.7928	-114.1067
409	Huron East	43.63	-81.28
410	Stanley	49.1331	-98.0656
411	Penetanguish	44.7667	-79.9333
412	Qualicum Bea	49.35	-124.4333
413	Notre-Dame-d	46.05	-73.4333
414	West Perth	43.47	-81.2
415	Cavan Monag	44.2	-78.4667
416	Arnprior	45.4333	-76.35
417	Smiths Falls	44.9	-76.0167
418	Pont-Rouge	46.75	-71.7
419	Champlain	45.5333	-74.65
420	Coaticook	45.1333	-71.8
421	Minto	43.9167	-80.8667

-- Retrieve the city with the highest population in each country

```
SELECT country, city, lat, lon, population
FROM gis_location
WHERE (country, population) IN (
SELECT country, MAX(population)
FROM gis_location
GROUP BY country
);
```

	country	city	lat	lon	population
1	Japan	Tokyo	35.6897	139.6922	37,732,000
2	Indonesia	Jakarta	-6.175	106.8275	33,756,000
3	India	Delhi	28.61	77.23	32,226,000
4	China	Guangzhou	23.13	113.26	26,940,000
5	Philippines	Manila	14.5958	120.9772	24,922,000
6	Brazil	São Paulo	-23.55	-46.6333	23,086,000
7	South Korea	Seoul	37.56	126.99	23,016,000
8	Mexico	Mexico City	19.4333	-99.1333	21,804,000
9	Egypt	Cairo	30.0444	31.2358	20,296,000
10	United States	New York	40.6943	-73.9249	18,972,871
11	Bangladesh	Dhaka	23.7639	90.3889	18,627,000
12	Thailand	Bangkok	13.7525	100.4942	18,007,000
13	Russia	Moscow	55.7558	37.6178	17,332,000
14	Argentina	Buenos Aires	-34.5997	-58.3819	16,710,000
15	Nigeria	Lagos	6.455	3.3841	16,637,000
16	Turkey	Istanbul	41.0136	28.955	16,079,000
17	Pakistan	Karachi	24.86	67.01	15,738,000
18	Vietnam	Ho Chi Minh City	10.7756	106.7019	15,136,000
19	Iran	Tehran	35.6892	51.3889	14,148,000
20	Congo (Kinshasa)	Kinshasa	-4.325	15.3222	12,836,000
21	United Kingdom	London	51.5072	-0.1275	11,262,000
22	France	Paris	48.8567	2.3522	11,060,000
23	Peru	Lima	-12.06	-77.0375	10,320,000
24	Taiwan	Taipei	25.0375	121.5625	9,078,000
25	Angola	Luanda	-8.8383	13.2344	9,051,000
26	Malaysia	Kuala Lumpur	3.1478	101.6953	8,911,000
27	South Africa	Johannesburg	-26.2044	28.0456	8,000,000
28	Colombia	Bogotá	4.7111	-74.0722	7,968,095
29	Tanzania	Dar es Salaam	-6.8161	39.2803	7,962,000
30	Sudan	Khartoum	15.5006	32.56	7,869,000
31	Hong Kong	Hong Kong	22.3	114.2	7,450,000
32	Saudi Arabia	Riyadh	24.6333	46.7167	7,237,000
33	Chile	Santiago	-33.4372	-70.6506	7,171,000
34	Spain	Madrid	40.4169	-3.7033	6,211,000
35	Iraq	Baghdad	33.3153	44.3661	6,183,000

-- Retrieve all cities within a certain distance of a specific point

```
SELECT city, lat, lon
FROM gis_location
WHERE ST_DistanceSphere(ST_SetSRID(ST_MakePoint(-73.9857, 40.7484), 4326),
ST_SetSRID(ST_MakePoint(lon, lat), 4326)) < 10000;
```

	city	lat	lon
1	New York	40.6943	-73.9249
2	Manhattan	40.7834	-73.9662
3	Jersey City	40.7184	-74.0686
4	Union City	40.7675	-74.0323
5	North Bergen	40.7938	-74.0242
6	Hoboken	40.7452	-74.0281
7	West New York	40.7857	-74.0094
8	Cliffside Park	40.8222	-73.9879
9	Weehawken	40.7677	-74.0168
10	Fairview	40.8182	-74.0022
11	Edgewater	40.8237	-73.974
12	Guttenberg	40.7928	-74.0049
13	Secaucus	40.781	-74.0659
14	Ridgefield	40.8313	-74.0147

2. Calculate Distance between points

-- distance of all the cities from a given points

```
SELECT city, ST_DistanceSphere(
        ST_SetSRID(ST_MakePoint(-73.9857, 40.7484), 4326),
        ST_SetSRID(ST_MakePoint(lon, lat), 4326)
) AS distance FROM gis_location;
```

	city	distance
1	Tokyo	10,846,034.78578309
2	Jakarta	16,169,870.76406535
3	Delhi	11,751,614.77884382
4	Guangzhou	12,872,733.29044069
5	Mumbai	12,533,776.44776157
6	Manila	13,670,523.62997991
7	Shanghai	11,862,073.97159141
8	São Paulo	7,688,346.50011519
9	Seoul	11,049,696.90430059
10	Mexico City	3,363,093.25356618
11	Cairo	9,018,451.74022576
12	New York	7,902.01754337
13	Dhaka	12,661,996.33104959
14	Beijing	10,985,446.40264448
15	Kolkāta	12,742,114.96721073
16	Bangkok	13,928,216.8306749
17	Shenzhen	12,929,352.69225842
18	Moscow	7,506,067.37114486
19	Buenos Aires	8,529,230.37004706
20	Lagos	8,476,670.14129785
21	Istanbul	8,064,099.4871853
22	Karachi	11,681,638.91234147
23	Bangalore	13,362,874.92911388

-- Distance between all cities in the gis_location table and a specific city (e.g. "New York") using the ST_DistanceSphere function:

```
SELECT city, ST_DistanceSphere(
        ST_SetSRID(ST_MakePoint(lon, lat), 4326),
        (SELECT ST_SetSRID(ST_MakePoint(lon, lat), 4326)
FROM gis_location WHERE city = 'New York')
) AS distance
FROM gis_location;
```

	city	distance
1	Tokyo	10,853,720.92969282
2	Jakarta	16,176,007.39630472
3	Delhi	11,754,766.17004931
4	Guangzhou	12,879,356.55871147
5	Mumbai	12,535,885.76923773
6	Manila	13,677,790.42909905
7	Shanghai	11,869,134.69281118
8	São Paulo	7,680,667.8431247
9	Seoul	11,056,930.81658883
10	Mexico City	3,363,536.46464962
11	Cairo	9,017,330.69601316
12	New York	0
13	Dhaka	12,666,404.28327636
14	Beijing	10,992,118.62296218
15	Kolkāta	12,746,257.35489525
16	Bangkok	13,933,602.79365184
17	Shenzhen	12,936,041.62434043
18	Moscow	7,508,123.22675136
19	Buenos Aires	8,522,208.2501931
20	Lagos	8,471,202.79170612
21	Istanbul	8,063,978.98602853
22	Karachi	11,683,456.05966915
23	Bangalore	13,365,197.27028753
24	Ho Chi Minh City	14,291,551.74756059
25	Ōsaka	11,107,806.99498189
26	Chengdu	12,078,164.38409228
27	Tehran	9,851,999.1750269
28	Kinshasa	10,257,600.78403938
29	Rio de Janeiro	7,752,892.34253677
30	Chennai	13,471,543.25182434
31	Xi'an	11,674,797.38895025
32	Lahore	11,344,963.43075164
33	Chongqing	12,204,167.01706981

-- the distance between all cities in the gis_location table that have a population greater than a certain value (e.g. 1 million) using the ST_DistanceSpheroid function:

```
SELECT city, ST_DistanceSpheroid(
        ST_SetSRID(ST_MakePoint(lon, lat), 4326),
        ST_SetSRID(ST_MakePoint(-118.2437, 34.0522), 4326),
'       SPHEROID["WGS 84",6378137,298.257223563]'
) AS distance
FROM gis_location
WHERE population > 1000000;
```

	ABC city	123 distance
1	Tokyo	8,834,495.846061641
2	Jakarta	14,458,171.339643758
3	Delhi	12,880,930.644858183
4	Guangzhou	11,665,230.325877277
5	Mumbai	14,008,377.246283054
6	Manila	11,760,295.867007777
7	Shanghai	10,460,364.658213947
8	São Paulo	9,894,918.69447497
9	Seoul	9,606,357.229531825
10	Mexico City	2,489,267.5837784093
11	Cairo	12,222,554.45823784
12	New York	3,951,396.295522072
13	Dhaka	12,929,435.911078753
14	Beijing	10,083,913.441523997
15	Kolkāta	13,137,979.494691048
16	Bangkok	13,320,539.765721096
17	Shenzhen	11,659,038.875442483
18	Moscow	9,792,779.626499515
19	Buenos Aires	9,831,110.18309535
20	Lagos	12,427,222.807161748
21	Istanbul	11,042,299.558278311
22	Karachi	13,460,505.312393032
23	Bangalore	14,538,198.78248167
24	Ho Chi Minh City	13,150,425.928541033
25	Ōsaka	9,209,065.95126458
26	Chengdu	11,583,499.572713893
27	Tehran	12,208,980.839957852
28	Kinshasa	14,210,507.912796102
29	Rio de Janeiro	10,131,060.584135354
30	Chennai	14,434,566.773720143
31	Xi'an	10,991,812.31368954

3. Calculate Areas of Interest

-- Find all cities within a certain distance of a given point:

```
SELECT city, ST_Distance(
  ST_SetSRID(ST_MakePoint(lon, lat), 4326),
  ST_SetSRID(ST_MakePoint(-118.2437, 34.0522), 4326)
) AS distance
FROM gis_location
WHERE ST_DWithin(
  ST_SetSRID(ST_MakePoint(lon, lat), 4326),
  ST_SetSRID(ST_MakePoint(-118.2437, 34.0522), 4326),
  50000
)
ORDER BY distance;
```

	ABC city	123 distance
1	Huntington Park	0.0770833315
2	East Los Angeles	0.0771318352
3	Florence-Graham	0.0840059522
4	Maywood	0.0847405452
5	Walnut Park	0.086757651
6	Bell	0.0963411646
7	South Pasadena	0.104118058
8	Commerce	0.1074804633
9	Cudahy	0.1078114094
10	Monterey Park	0.1111281243
11	Alhambra	0.1127762386
12	View Park-Windso	0.1189791999
13	South Gate	0.1190271398
14	Bell Gardens	0.1234767185
15	Westmont	0.1248433418
16	Glendale	0.1297370417
17	Willowbrook	0.1315496104
18	West Hollywood	0.1331857725
19	Lynwood	0.1349357254
20	Montebello	0.137874218
21	Inglewood	0.1388518995
22	West Athens	0.1418303564
23	San Marino	0.1481832987
24	West Rancho Don	0.148551035
25	Pasadena	0.1500611209

-- Find all cities within a given bounding box:

```
SELECT city, lat, lon
FROM gis_location
WHERE lon BETWEEN -118.5 AND -117.5
AND lat BETWEEN 33.5 AND 34.5;
```

	city	lat	lon
1	Los Angeles	34.1141	-118.4068
2	Mission Viejo	33.6096	-117.6551
3	Long Beach	33.7977	-118.167
4	Anaheim	33.839	-117.8574
5	Santa Ana	33.7367	-117.8819
6	Irvine	33.6772	-117.7738
7	Santa Clarita	34.4175	-118.4964
8	Huntington Beach	33.696	-118.0018
9	Glendale	34.1819	-118.2468
10	Ontario	34.0393	-117.6064
11	Rancho Cucamonga	34.1247	-117.5667
12	Garden Grove	33.7787	-117.9601
13	Corona	33.8616	-117.5649
14	Pomona	34.0585	-117.7626
15	Torrance	33.8346	-118.3416
16	Fullerton	33.8841	-117.9279
17	Orange	33.8038	-117.8218
18	Pasadena	34.1597	-118.139
19	East Los Angeles	34.0326	-118.1691
20	Downey	33.9379	-118.1311
21	Costa Mesa	33.6667	-117.9135
22	El Monte	34.0739	-118.0291

-- Find all cities within a given polygon:

```
SELECT city, lat, lon
FROM gis_location
WHERE ST_Contains(
  ST_SetSRID(
    ST_MakePolygon(
      ST_GeomFromText('LINESTRING(-118.2437 34.0522, -118.3 34.05, -118.2 33.9, -118.2437 34.0522)')
    ),
    4326
  ),
  ST_SetSRID(ST_MakePoint(lon, lat), 4326));
```

	city	lat	lon
1	Florence-Graham	33.9682	-118.2447
2	Walnut Park	33.9682	-118.222

-- Find all cities with a population above a certain threshold within a certain distance of a given point:

```
SELECT city, population, ST_Distance(
  ST_SetSRID(ST_MakePoint(lon, lat), 4326),
  ST_SetSRID(ST_MakePoint(-118.2437, 34.0522), 4326)
) AS distance
FROM gis_location
WHERE population > 100000
AND ST_DWithin(
  ST_SetSRID(ST_MakePoint(lon, lat), 4326),
  ST_SetSRID(ST_MakePoint(-118.2437, 34.0522), 4326),
  50000
)ORDER BY distance;
```

	city	population	distance
1	East Los Angeles	118,964	0.0771318352
2	Glendale	196,512	0.1297370417
3	Inglewood	108,206	0.1388518995
4	Pasadena	138,771	0.1500611209
5	Burbank	107,364	0.1574246804
6	Downey	114,293	0.1604470318
7	Los Angeles	12,121,244	0.1744511966
8	El Monte	110,144	0.2156943439
9	Norwalk	103,330	0.216722703
10	Torrance	147,156	0.2386088221
11	Long Beach	466,565	0.2658065838
12	West Covina	109,396	0.3324163805
13	Fullerton	142,964	0.3577530573
14	Garden Grove	172,708	0.3939939213
15	Huntington Beach	198,735	0.4305740935
16	Anaheim	348,204	0.441227753
17	Santa Clarita	275,230	0.4441861997
18	Santa Ana	313,818	0.4800411337
19	Pomona	151,592	0.4811412475
20	Orange	139,195	0.4895938827
21	Costa Mesa	112,148	0.5075847614
22	Simi Valley	127,158	0.5487867528

4. Sorting and Limit Executions

-- Retrieving GIS locations for specific features and sorting by population in descending order:

```
SELECT *
FROM gis_location
WHERE country = 'United States'
ORDER BY population DESC;
```

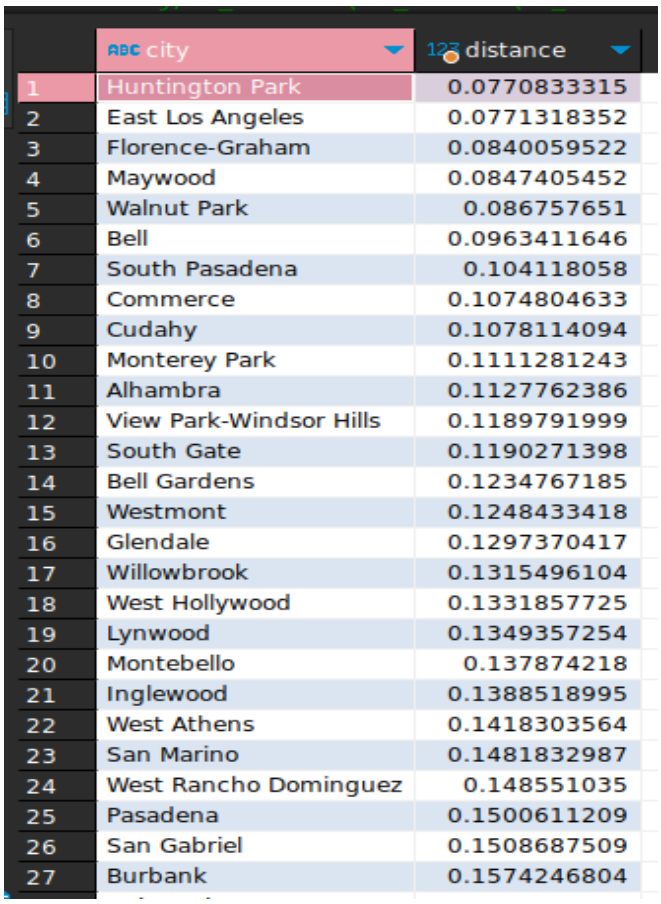
	id	country	city	lat	lon	population
1	12	United States	New York	40.6943	-73.9249	18,972,871
2	34	United States	Los Angeles	34.1141	-118.4068	12,121,244
3	56	United States	Chicago	41.8375	-87.6866	8,595,181
4	101	United States	Miami	25.784	-80.2101	5,711,945
5	105	United States	Dallas	32.7935	-96.7667	5,668,165
6	106	United States	Houston	29.786	-95.3885	5,650,910
7	116	United States	Philadelphia	40.0077	-75.1339	5,512,873
8	135	United States	Atlanta	33.7628	-84.422	5,046,555
9	150	United States	Washington	38.9047	-77.0163	4,810,669
10	181	United States	Boston	42.3188	-71.0852	4,208,580
11	188	United States	Phoenix	33.5722	-112.0892	4,047,095
12	220	United States	Detroit	42.3834	-83.1024	3,522,856
13	224	United States	Seattle	47.6211	-122.3244	3,438,221
14	238	United States	San Francisco	37.7558	-122.4449	3,290,197
15	254	United States	San Diego	32.8313	-117.1222	3,084,174
16	287	United States	Minneapolis	44.9635	-93.2678	2,856,952
17	292	United States	Brooklyn	40.6501	-73.9496	2,736,074
18	304	United States	Tampa	27.9945	-82.4447	2,683,956
19	309	United States	Denver	39.762	-104.8758	2,650,725
20	353	United States	Queens	40.7498	-73.7976	2,405,464
21	380	United States	Baltimore	39.3051	-76.6144	2,205,092
22	387	United States	Las Vegas	36.2333	-115.2654	2,150,373
23	398	United States	St. Louis	38.6359	-90.2451	2,092,481
24	403	United States	Portland	45.5371	-122.65	2,036,875
25	409	United States	Riverside	33.9381	-117.3949	2,022,285
26	425	United States	Orlando	28.4773	-81.337	1,927,699

```
-- Retrieving the top 10 most populous cities in the USA:
SELECT *
FROM gis_location
WHERE country = 'United States'
ORDER BY population DESC
LIMIT 10;
```

	id	country	city	lat	lon	population
1	12	United States	New York	40.6943	-73.9249	18,972,871
2	34	United States	Los Angeles	34.1141	-118.4068	12,121,244
3	56	United States	Chicago	41.8375	-87.6866	8,595,181
4	101	United States	Miami	25.784	-80.2101	5,711,945
5	105	United States	Dallas	32.7935	-96.7667	5,668,165
6	106	United States	Houston	29.786	-95.3885	5,650,910
7	116	United States	Philadelphia	40.0077	-75.1339	5,512,873
8	135	United States	Atlanta	33.7628	-84.422	5,046,555
9	150	United States	Washington	38.9047	-77.0163	4,810,669
10	181	United States	Boston	42.3188	-71.0852	4,208,580

-- Retrieving GIS locations within a certain distance from a point, and sorting by distance in ascending order:

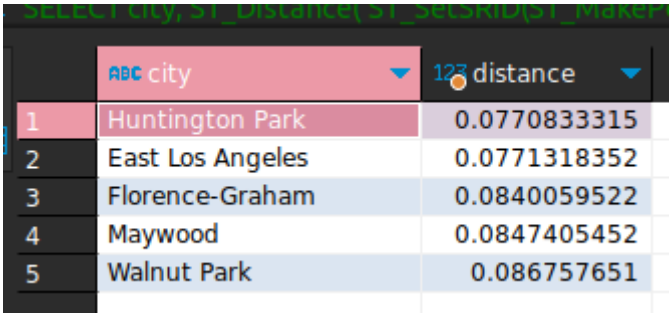
```
SELECT city, ST_Distance(
ST_SetSRID(ST_MakePoint(-118.2437, 34.0522), 4326),
ST_SetSRID(ST_MakePoint(gis_location.lon, gis_location.lat), 4326)
) AS distance
FROM gis_location
WHERE country = 'United States' AND ST_Distance(
    ST_SetSRID(ST_MakePoint(-118.2437, 34.0522), 4326),
    ST_SetSRID(ST_MakePoint(gis_location.lon, gis_location.lat), 4326)
) < 500000
ORDER BY distance ASC;
```



	city	distance
1	Huntington Park	0.0770833315
2	East Los Angeles	0.0771318352
3	Florence-Graham	0.0840059522
4	Maywood	0.0847405452
5	Walnut Park	0.086757651
6	Bell	0.0963411646
7	South Pasadena	0.104118058
8	Commerce	0.1074804633
9	Cudahy	0.1078114094
10	Monterey Park	0.1111281243
11	Alhambra	0.1127762386
12	View Park-Windsor Hills	0.1189791999
13	South Gate	0.1190271398
14	Bell Gardens	0.1234767185
15	Westmont	0.1248433418
16	Glendale	0.1297370417
17	Willowbrook	0.1315496104
18	West Hollywood	0.1331857725
19	Lynwood	0.1349357254
20	Montebello	0.137874218
21	Inglewood	0.1388518995
22	West Athens	0.1418303564
23	San Marino	0.1481832987
24	West Rancho Dominguez	0.148551035
25	Pasadena	0.1500611209
26	San Gabriel	0.1508687509
27	Burbank	0.1574246804

-- Retrieving the 5 closest cities to a specific point:

```
        SELECT city, ST_Distance(
    ST_SetSRID(ST_MakePoint(-118.2437, 34.0522), 4326),
    ST_SetSRID(ST_MakePoint(gis_location.lon, gis_location.lat), 4326)
        ) AS distance
        FROM gis_location
        WHERE country = 'United States' AND ST_Distance(
            ST_SetSRID(ST_MakePoint(-118.2437, 34.0522), 4326),
            ST_SetSRID(ST_MakePoint(gis_location.lon, gis_location.lat), 4326)
        ) < 500000
        ORDER BY distance ASC
        LIMIT 5;
```



	city	distance
1	Huntington Park	0.0770833315
2	East Los Angeles	0.0771318352
3	Florence-Graham	0.0840059522
4	Maywood	0.0847405452
5	Walnut Park	0.086757651

5. Optimize the queries to speed up execution time

-- creating spatial index on lon and lat

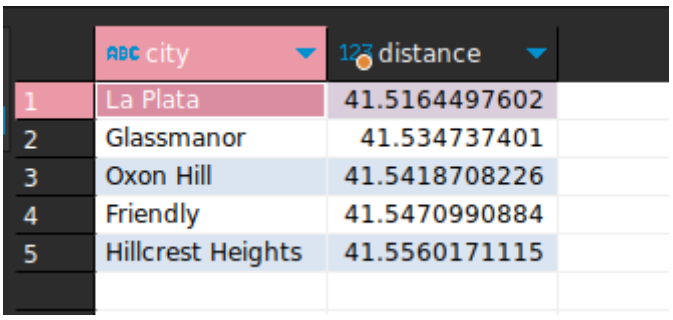
```
CREATE INDEX gis_location_geom_idx ON gis_location USING GIST
(ST_SetSRID(ST_MakePoint(lon, lat), 4326));

SELECT city, ST_Distance(
   ST_SetSRID(ST_MakePoint(-118.2437, 34.0522), 4326),
   ST_SetSRID(ST_MakePoint(gis_location.lon, gis_location.lat), 4326)
) AS distance
FROM gis_location
WHERE country = 'United States' AND
   ST_DWithin(
      ST_SetSRID(ST_MakePoint(-118.2437, 34.0522), 4326),
         ST_SetSRID(ST_MakePoint(gis_location.lon, gis_location.lat), 4326),
      500000
   )ORDER BY distance ASC LIMIT 5;
```

	city	distance
1	Huntington Park	0.0770833315
2	East Los Angeles	0.0771318352
3	Florence-Graham	0.0840059522
4	Maywood	0.0847405452
5	Walnut Park	0.086757651

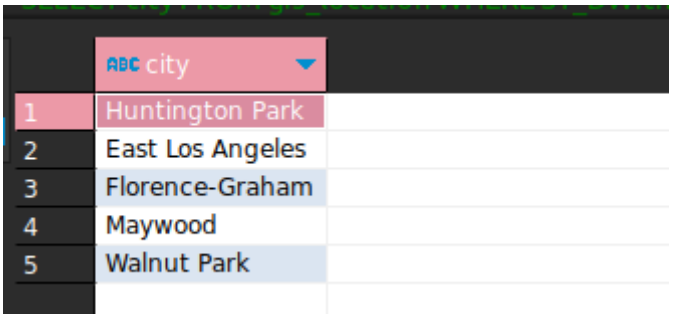
-- Use a smaller bounding box to limit the search space:

```
SELECT city, ST_Distance(
   ST_SetSRID(ST_MakePoint(-118.2437, 34.0522), 4326),
   ST_SetSRID(ST_MakePoint(gis_location.lon, gis_location.lat), 4326)
) AS distance
FROM gis_location
WHERE country = 'United States' AND
   lon BETWEEN -77 AND -76 AND
   lat BETWEEN 38 AND 39 AND
   ST_DWithin(
      ST_SetSRID(ST_MakePoint(-118.2437, 34.0522), 4326),
      ST_SetSRID(ST_MakePoint(gis_location.lon, gis_location.lat), 4326),
      500000
   )
ORDER BY distance ASC
LIMIT 5;
```



	city	distance
1	La Plata	41.5164497602
2	Glassmanor	41.534737401
3	Oxon Hill	41.5418708226
4	Friendly	41.5470990884
5	Hillcrest Heights	41.5560171115

```
-- Use the ST_DWithin function instead of ST_Distance:
        SELECT city
        FROM gis_location
        WHERE ST_DWithin(
            ST_SetSRID(ST_MakePoint(-118.2437, 34.0522), 4326),
            ST_SetSRID(ST_MakePoint(gis_location.lon, gis_location.lat), 4326),
            500000
        ) AND country = 'United States'
        ORDER BY ST_Distance(
            ST_SetSRID(ST_MakePoint(-118.2437, 34.0522), 4326),
            ST_SetSRID(ST_MakePoint(gis_location.lon, gis_location.lat), 4326)
        ) ASC
        LIMIT 5;
```



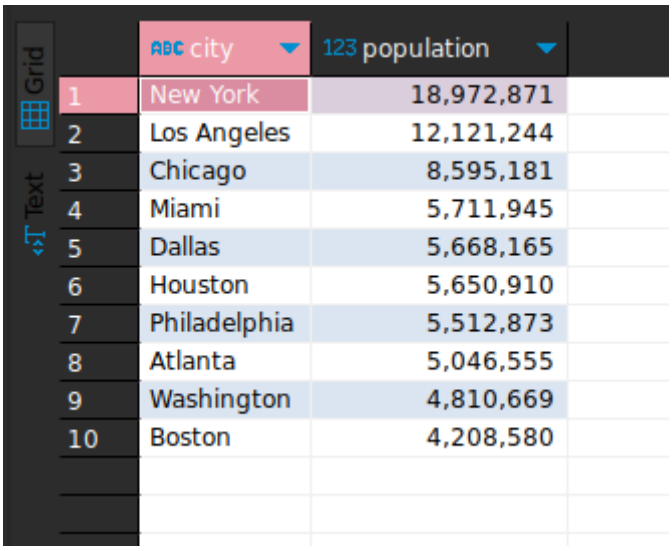
	city
1	Huntington Park
2	East Los Angeles
3	Florence-Graham
4	Maywood
5	Walnut Park

6. N-Optimization of queries

-- create spatial index

```
CREATE INDEX gis_location_geom_idx ON gis_location USING GIST
(ST_SetSRID(ST_MakePoint(lon, lat), 4326));

-- Use filters to narrow down the search space:
SELECT city, population
FROM gis_location
WHERE country = 'United States' AND population > 1000000
ORDER BY population DESC
LIMIT 10;
```



	city	population
1	New York	18,972,871
2	Los Angeles	12,121,244
3	Chicago	8,595,181
4	Miami	5,711,945
5	Dallas	5,668,165
6	Houston	5,650,910
7	Philadelphia	5,512,873
8	Atlanta	5,046,555
9	Washington	4,810,669
10	Boston	4,208,580

-- Use subqueries to break down complex queries:

```
SELECT city, population
FROM (
  SELECT city, population,
    ST_Distance(
      ST_SetSRID(ST_MakePoint(lon, lat), 4326),
      ST_SetSRID(ST_MakePoint(-118.2437, 34.0522), 4326)
    ) AS distance
  FROM gis_location
  WHERE country = 'United States'
) AS subquery
WHERE distance < 50000
ORDER BY population DESC
LIMIT 10;
```

	city	population
1	New York	18,972,871
2	Los Angeles	12,121,244
3	Chicago	8,595,181
4	Miami	5,711,945
5	Dallas	5,668,165
6	Houston	5,650,910
7	Philadelphia	5,512,873
8	Atlanta	5,046,555
9	Washington	4,810,669
10	Boston	4,208,580

-- Use aggregates to summarize data:

```
SELECT country, COUNT(*) AS num_cities, AVG(population) AS avg_population
FROM gis_location
GROUP BY country
ORDER BY num_cities DESC;
```

--

	country	num_cities	avg_population
1	India	5,933	73,896.2518593644
2	United States	5,393	70,489.2217689598
3	Brazil	2,903	69,992.1191870479
4	Germany	1,733	38,733.8817080208
5	Philippines	1,578	85,651.3095994914
6	China	1,562	911,694.4935815148
7	Italy	1,363	31,478.2787967718
8	United Kingdom	1,305	52,661.6666666667
9	Japan	1,253	145,342.3599361532
10	Russia	1,211	88,871.764657308
11	France	1,140	37,235.9105263158
12	Mexico	940	122,164.5265957447
13	Madagascar	822	24,439.9914841849
14	Spain	754	58,268.8368700265
15	Colombia	714	74,565.462184874
16	Turkey	614	159,623.6970684039
17	Morocco	578	49,368.7871972318
18	Argentina	467	98,245.8158458244
19	Iran	464	140,330.6255411255
20	Poland	456	47,533.6030701754
21	Canada	427	85,370.8149882904
22	Ukraine	419	65,544.2124105012
23	Belgium	388	29,894.118556701

Use LIMIT and OFFSET to paginate results:

```
select * from gis_location order by population desc limit 20 offset 40;
```

	id	country	city	lat	lon
1	11,297	Malawi	Balaka	-14.9889	34.9591
2	11,270	Serbia	Tutin	42.9875	20.3256
3	11,267	Laos	Xai	20.6914	101.9861
4	11,218	Cabo Verde	João Teves	15.068	-23.589
5	11,190	India	Dispur	26.15	91.77
6	11,152	The Gambia	Kanifing	13.45	-16.6667
7	11,009	Mauritania	Tevragh Zeina	18.0994	-15.9761
8	10,381	Vietnam	Đà Nẵng	16.0748	108.224
9	8,296	Taiwan	Banqiao	25.0143	121.4672
10	910	South Georgia An	King Edward	-54.2833	-36.5
11	909	Christmas Island	Flying Fish Cc	-10.4167	105.7167
12	908	Pitcairn Islands	Adamstown	-25.0667	-130.0833
13	907	Svalbard	Longyearbyer	78.2167	15.6333
14	906	Norfolk Island	Kingston	-29.0569	167.9617
15	905	Wallis and Futuna	Mata-Utu	-13.2825	-176.1736
16	904	Anguilla	The Valley	18.2167	-63.05
17	903	Saint Pierre and M	Saint-Pierre	46.7811	-56.1764
18	901	Saint Barthelemy	Gustavia	17.8958	-62.8508
19	900	Nauru	Yaren	-0.5477	166.9209
20	11,508	Kenya	Siaya	0.0667	34.2833

-- Use EXPLAIN to analyze query performance:

```
EXPLAIN ANALYZE SELECT city, population
FROM gis_location
WHERE country = 'USA' AND population > 1000000
ORDER BY population DESC LIMIT 10;
```

	QUERY PLAN
1	Limit (cost=1069.38..1069.38 rows=1 width=14) (actual time=3.370..3.371 rows=0 loops=1)
2	-> Sort (cost=1069.38..1069.38 rows=1 width=14) (actual time=3.369..3.369 rows=0 loops=
3	Sort Key: population DESC
4	Sort Method: quicksort Memory: 25kB
5	-> Seq Scan on gis_location (cost=0.00..1069.37 rows=1 width=14) (actual time=3.364..
6	Filter: ((population > 1000000) AND ((country)::text = 'USA'::text))
7	Rows Removed by Filter: 44691
8	Planning Time: 0.089 ms
9	Execution Time: 3.386 ms