1. Download the Yelp Dataset Challenge data from this folder or from the [website](http://www.yelp.com/dataset_challenge). Read the tabs called **Dataset and Documentation and agree to the Terms of Use**.

Yelp wants your help solving the following problem: add a feature to the Yelp website listing the businesses of one type nearest to a given business of another type -- for example, the doctors nearest to a given restaurant -- in order of increasing distance.

Solve as much of the problem as you can, and sketch how your part of the solution would be combined with other parts to make a complete solution.

… OR PICK YOUR OWN QUESTION TO ANSWER! Follow the same instructions.

yelp\_xchai=> select count(\*) from business;

 count

--------

 156639

(1 row)

There are 156639 businesses.

yelp\_xchai=> select count(\*) from (select category, count(\*) from category group by 1 order by 2 desc) as hi;

 count

-------

  1231

(1 row)

There are 1231 different types of business.

**Problem 1: list the doctors nearest to the restaurant John's Chinese BBQ Restaurant (latitude: 43.8409 longitude: -79.3996), in order of increasing distance.**

First, I try to create a function in psql to calculate the distance between each doctor business and John’s Chinese BBQ Restaurant. Here’s the code:

CREATE OR REPLACE FUNCTION John\_Chinese\_BBQ(\_category char)

RETURNS TABLE (id char, name char, distance double precision) AS

$func$

DECLARE

latitude double precision;

longitude double precision;

BEGIN

FOR latitude, longitude IN

SELECT a.latitude, a.longitude

FROM business\_full a

WHERE a.category = \_category

LOOP

distance := sqrt((latitude - 43.8409)^2 + (longitude + 79.3996)^2);

RETURN NEXT;

END LOOP;

END

$func$ LANGUAGE plpgsql STABLE;

The function works successfully to return a set of distance when you feed in the type of business you want. However, it doesn’t provide the associated business name or other information. Here’s example of output:

yelp\_xchai=> select John\_Chinese\_BBQ('Doctors') order by 1 limit 5;

    john\_chinese\_bbq

------------------------

 (,,0.0102883429180714)

 (,,0.0305235974288641)

 (,,0.033998382314461)

 (,,0.0385519130524013)

 (,,0.0413859879669377)

(5 rows)

So this is not a good solution.

Then I rewrote the function. And it worked!!! Here’s the code:

CREATE OR REPLACE FUNCTION distance\_John\_Chinese\_BBQ(\_category char)

RETURNS SETOF business\_full AS

$$

SELECT id, name, address, city, state, postal\_code,latitude, longitude, category, sqrt((latitude - 43.8409)^2 + (longitude + 79.3996)^2) as distance

FROM business\_full

WHERE category = \_category

ORDER BY distance

$$ LANGUAGE 'sql';

So the answer for the list of the doctors nearest to John’s Chinese BBQ Restaurant in order of increasing distance is as following:

yelp\_xchai=# select distance\_John\_Chinese\_BBQ('Doctors') limit 5;

                         distance\_john\_chinese\_bbq

------------------------------------------------------------------------------------

 (3yL8BU4jmmvq5OtqbypJoA,"Matlis Medical, Urgent Walk In Clinic, Sports Medicine","8200 Bayview Avenue",Thornhill,ON,"L3T 2S2",43.8318,-79.4044,Doctors,0.0102883429180714)

 (EfDE7XvY2-6-6g76o0dMpQ,"SkinVitality Medical Clinic","8865 Yonge St","Richmond Hill",ON,"L4C 6Z1",43.8421,-79.4301,Doctors,0.0305235974288641)

 (ZEeA\_2Aa-6yqZG-gM5j9HQ,"Honsberger Physio +","675 Cochrane Drive, Suite 105",Markham,ON,"L3R 0B8",43.8467,-79.3661,Doctors,0.033998382314461)

 (w1F4ketex-BMHz0xesFjUw,"All About Health Medical Clinic","1070 Major MacKenzie Drive E","Richmond Hill",ON,"L4S 1P3",43.8773,-79.4123,Doctors,0.0385519130524013)

 (fE\_gCGRk7pSBG-5hxJwVxA,"Pinnacle Health Sciences Centre","2 - 9651 Yonge Street","Richmond Hill",ON,"L4C 1V7",43.8633,-79.4344,Doctors,0.0413859879669377)

(5 rows)

The names of those businesses are marked in blue and the distance is marked in pink.

With the function distance\_John\_Chinese\_BBQ(\_category char), we can actually find all the lists of different types of businesses nearest to John’s Chinese BBQ Restaurant in order of increasing distance. Just feed in different types. Following are some examples.

**Problem 1i: list the shopping nearest to the restaurant John's Chinese BBQ Restaurant (latitude: 43.8409 longitude: -79.3996)**

yelp\_xchai=# select distance\_John\_Chinese\_BBQ('Shopping') limit 5;

                              distance\_john\_chinese\_bbq

------------------------------------------------------------------------------------

 (p9-gSAaWMVFHRwu2FS6jIw,Dollarama,"8190 Bayview Avenue",Markham,ON,"L3T 2S2",43.8316,-79.4045,Shopping,0.0105118980208072)

 (iWIEg3DeZLvszGXKddo\_kA,Solutions,"295 High Tech Road","Richmond Hill",ON,"L4B 0A3",43.8425,-79.4116,Shopping,0.0121061967603381)

 (oE7IUTNjTbgODqtcfG\_3qg,"Flowers by Sylvia","100 West Beaver Creek Road","Richmond Hill",ON,"L4B 1H4",43.8515,-79.3933,Shopping,0.0123308556069807)

 (wKrnJ8AYAix5lJCtnpT8fA,"Urban Barn","295 High Tech Road, Unit 1","Richmond Hill",ON,"L4B 0A3",43.8427,-79.412,Shopping,0.0125299640861416)

 (4LTSMkK4P2SAqsyCv88lUA,"Bombay & Co Inc.","295 High Tech Road","Richmond Hill",ON,"L4B 0A3",43.843,-79.4124,Shopping,0.0129711217710724)

(5 rows)

**Problem 1ii: list the Convenience Stores nearest to the restaurant John's Chinese BBQ Restaurant (latitude: 43.8409 longitude: -79.3996)**

yelp\_xchai=# select distance\_John\_Chinese\_BBQ('Convenience Stores') limit 5;

                     distance\_john\_chinese\_bbq

------------------------------------------------------------------------------------

 (5eBNGReFI2fT\_Nc1wPezsA,"Shoppers Drug Mart","4990 Yonge St","North York",ON,"M2N 0A1",43.7658,-79.4126,"Convenience Stores",0.076216861651472)

 (A7h6Gg94Xa3F3hcscceyKg,"Shoppers Drug Mart","946 Lawrence Avenue E, Unit 2","North York",ON,"M3C 3M9",43.7372,-79.3458,"Convenience Stores",0.116825211320161)

 (kTD8tQ7IcxRPghEoJ7p\_xA,"Shell Select","9270 McCowan Road",Markham,ON,"L6C 2L1",43.8857,-79.289,"Convenience Stores",0.119328957089222)

 (UYqXTafScQuzaAZNlBe1yA,7-Eleven,"7389 McCowan Road, Units 1&2",Markham,ON,"L3S 3H8",43.8407,-79.2785,"Convenience Stores",0.121100165152666)

 (a9y5V1o5Dbewb-y0sDnkmA,"Jane & Major Mac Esso","3100 Major MacKenzie Drive W",Vaughan,ON,"L6A 1S1",43.8499,-79.5374,"Convenience Stores",0.138093591451594)

(5 rows)

If you want to find the solution for another business, for example, Healing Hands Esthetics (latitude: 35.9791, longitude: -114.83), which is Skin Care business. We need to adjust the function a little bit.

CREATE OR REPLACE FUNCTION healing\_hands\_esthetics(\_category char)

RETURNS SETOF business\_full AS

$$

SELECT id, name, address, city, state, postal\_code,latitude, longitude, category, sqrt((latitude - 35.9791)^2 + (longitude - (-114.83))^2) as distance

FROM business\_full

WHERE category = \_category

ORDER BY distance

$$ LANGUAGE 'sql';

You need to change the function name and the latitude and longitude for the function. And here’s an example.

**Problem 2: list the doctors nearest to Healing Hands Esthetics (latitude: 35.9791, longitude: -114.83**), **in order of increasing distance.**

yelp\_xchai=> select healing\_hands\_esthetics('Doctors') limit 5;

                                                                 healing\_hands\_esthetics

----------------------------------------------------------------------------------------------------------------------------------------------------------

 (6vGdF1fzsMKVm84VgCSkUQ,"Smith Warren L MD","895 Adams Blvd","Boulder City",NV,89005,35.9681,-114.843,Doctors,0.0170293863659322)

 (\_tj19b\_1VZ\_IDvaH8pFXWA,"Family Doctors of Boulder City","895 Adams Blvd","Boulder City",NV,89005,35.9681,-114.843,Doctors,0.0170293863659322)

 (c3X8LSEJfZsu55iNaipTWQ,"Clear Vision Eye Centers - Boulder City","1627 Nevada Hwy","Boulder City",NV,89005,35.9687,-114.857,Doctors,0.0289337173553649)

 (BUX5BUQEkcHAni6IGlLycQ,"Woodson S Linda, MD","2800 N Tenaya Way, Ste 203",Henderson,NV,89128,36.0323,-114.963,Doctors,0.143245383869773)

 (SoJ4ozoFY2yvnk0GWZUbaA,"Noel Alonzo, LMT FMT - Spiritt","250 E Horizon Dr",Henderson,NV,89015,36.0129,-114.983,Doctors,0.156688991317201)

(5 rows)

**Problem 2i: list the Water Heater Installation/Repair**

**nearest to Healing Hands Esthetics (latitude: 35.9791, longitude: -114.83), in order of increasing distance.**

yelp\_xchai=> select healing\_hands\_esthetics('Water Heater Installation/Repair') limit 5;

                                                                                   healing\_hands\_esthetics

--------------------------------------------------------------------------------------------------------------------------------------------------------------

--------------------------------

 (7s39xlfRI4AJqOkjexPRMA,"Innovative Plumbing Pros","661 Powder Brook Dr",Henderson,NV,89002,36.0053,-114.946,"Water Heater Installation/Repair",0.118921991237953)

 (XFUBFmF-y6Aaxqq-Sp-5Hg,"Benjamin Franklin Plumbing","247 Elliot Rd, Ste 5",Henderson,NV,89011,36.0666,-115.016,"Water Heater Installation/Repair",0.20555352

1010958)

 (8F3PP5h89NqXkd4PxUCbvw,"The Honest Plumber Heating & Air","150 N Gibson Rd, Ste A",Henderson,NV,89014,36.039,-115.029,"Water Heater Installation/Repair",0.2 07819657395539)

 (RnxgXS5iAxVvP9gHgvagGw,"One Hour Air Conditioning & Heating","739 W Sunset Rd",Henderson,NV,89011,36.0632,-115.024,"Water Heater Installation/Repair",0.2114

44579027226)

 (yhGcivj55RrxbXUlk9TPNg,"TempRite Air Conditioning, Heating & Plumbing","1111 Mary Crest Rd, Ste O",Henderson,NV,89074,36.0375,-115.035,"Water Heater Install

ation/Repair",0.21315618686775)

(5 rows)

The distance function can find the distance of all different types of business to a specific given business. The only disadvantage of the function is that you have to adjust the latitude and longitude of the given business in the function when you need to search for a different business.

The following *feature* function doesn’t require you to adjust the latitude and longitude every time you look for a different business. You only need to pass the id of the business and the category of the other business.

CREATE OR REPLACE FUNCTION feature(\_id char, \_category char)

RETURNS SETOF business\_full AS

$$

SELECT id, name, address, city, state, postal\_code,latitude, longitude, category,

sqrt((latitude - (select latitude from business\_full where id = \_id))^2 + (longitude - (select longitude from business\_full where id = \_id))^2) as distance

FROM business\_full

WHERE category = \_category

ORDER BY distance

$$ LANGUAGE 'sql';

For example, the business id of John's Chinese BBQ Restaurant is ‘--6MefnULPED\_I942VcFNA’ and we are looking for the nearest doctors.

yelp\_xchai=# select feature('--6MefnULPED\_I942VcFNA', 'Doctors') limit 5;

                                        feature

------------------------------------------------------------------------------------

 (3yL8BU4jmmvq5OtqbypJoA,"Matlis Medical, Urgent Walk In Clinic, Sports Medicine","8200 Bayview Avenue",Thornhill,ON,"L3T 2S2",43.8318,-79.4044,Doctors,0.0102883429180714)

 (EfDE7XvY2-6-6g76o0dMpQ,"SkinVitality Medical Clinic","8865 Yonge St","Richmond Hill",ON,"L4C 6Z1",43.8421,-79.4301,Doctors,0.0305235974288641)

 (ZEeA\_2Aa-6yqZG-gM5j9HQ,"Honsberger Physio +","675 Cochrane Drive, Suite 105",Markham,ON,"L3R 0B8",43.8467,-79.3661,Doctors,0.033998382314461)

 (w1F4ketex-BMHz0xesFjUw,"All About Health Medical Clinic","1070 Major MacKenzie Drive E","Richmond Hill",ON,"L4S 1P3",43.8773,-79.4123,Doctors,0.0385519130524013)

 (fE\_gCGRk7pSBG-5hxJwVxA,"Pinnacle Health Sciences Centre","2 - 9651 Yonge Street","Richmond Hill",ON,"L4C 1V7",43.8633,-79.4344,Doctors,0.0413859879669377)

(5 rows)