DB_Connection

October 23, 2020

```
[2]: import pandas as pd
     import geopandas as gpd
     from geopandas import GeoDataFrame
     import numpy as np
     import matplotlib.pyplot as plt
     import psycopg2
     import sqlalchemy
     import contextily as ctx
     %matplotlib inline
     pd.set_option('display.max_columns', 500)
```

```
[3]: %load_ext sql
     %sql postgresql://ds4a_t53:DS4ACatastro@ds4av1.cd21iciz5zqe.us-east-2.rds.
      ↪amazonaws.com/ofertascatastro
```

For copy the spatial data to RDS (First create extensions postgis and postgis topology in the database, note: postgis_raster is not supported in RDS postgis install) we use the shp2pgsql help source where -s is the spatial projection -I is the path to the shapefile

```
shp2pgsql -s 3116 ofertas.shp data.ofertas_catastro | psql -h ds4av1.cd21iciz5zqe.us-east-2.rd:
```

this can take a long

0.1 The type of the data

```
[4]: %%sql
     select column_name, data_type from information_schema.columns
     where table_name = 'ofertas_catastro';
```

```
 * postgresql://ds4a_t53:***@ds4av1.cd21iciz5zqe.us-
east-2.rds.amazonaws.com/ofertascatastro
71 rows affected.
```

```
[4]: [('gid', 'integer'),
     ('lotcodigo', 'character varying'),
     ('oft_codigo', 'character varying'),
     ('lotcodig', 'character varying'),
     ('nombre_barrio', 'character varying'),
```

```
('oft_tipo_inmueble', 'character varying'),
('oft_tipo_norma_juridica', 'character varying'),
('oic_area_terreno', 'numeric'),
('oic_area_construccion', 'numeric'),
('oic_valor_adm', 'numeric'),
('oic_valor_anexos', 'numeric'),
('oia_tiene_ascensor', 'character varying'),
('oia_cant_garajes', 'double precision'),
('vfventa2020', 'numeric'),
('codigo_con', 'numeric'),
('codigo_res', 'numeric'),
('x', 'numeric'),
('y', 'numeric'),
('d_park', 'numeric'),
('d_highway', 'numeric'),
('d_bikeway', 'numeric'),
('d_ssf', 'numeric'),
('d_mus', 'numeric'),
('d_lib', 'numeric'),
('d_sitp', 'numeric'),
('d_tm', 'numeric'),
('d_p_tm', 'numeric'),
('d_gy', 'numeric'),
('d_ies', 'numeric'),
('d_bom', 'numeric'),
('d_col', 'numeric'),
('d_ips', 'numeric'),
('dm_bikeway', 'numeric'),
('dm_bom', 'numeric'),
('dm_col', 'numeric'),
('dm_gy', 'numeric'),
('dm_highway', 'numeric'),
('dm_ies', 'numeric'),
('dm_ips', 'numeric'),
('dm_lib', 'numeric'),
('dm_mus', 'numeric'),
('dm_park', 'numeric'),
('dm_p_tm', 'numeric'),
('dm_sitp', 'numeric'),
('dm_ssf', 'numeric'),
('dm_tm', 'numeric'),
('loccodigo', 'character varying'),
('locnombre', 'character varying'),
('z', 'numeric'),
('barmanpre', 'character varying'),
('preczhf', 'character varying'),
('suelo', 'character varying'),
```

```
('actividad', 'character varying'),
('tratamiento_urb', 'character varying'),
('topografia', 'character varying'),
('serpub', 'character varying'),
('serpub_tipo', 'character varying'),
('serpub_especifico', 'character varying'),
('via', 'character varying'),
('clase_via', 'character varying'),
('estado_via', 'character varying'),
('influencia_via', 'character varying'),
('actividad_economica', 'character varying'),
('actividad_economica_tipo', 'character varying'),
('tipo_segun_actividad', 'character varying'),
('manzana_id', 'character varying'),
('cp_terr_ar', 'character varying'),
('avaluo_com', 'numeric'),
('avaluo_cat', 'numeric'),
('log(vfvent', 'numeric'),
('geom', 'USER-DEFINED')]
```

0.2 The first record

[5]:
```sql
%%sql
SELECT * FROM data.ofertas_catastro LIMIT 1;
```

```
 * postgresql://ds4a_t53:***@ds4av1.cd21iciz5zqe.us-
east-2.rds.amazonaws.com/ofertascatastro
1 rows affected.
```

[5]: [(1, '009114036002', '338319', '009114036002', 'GILMAR', 'Apartamento', 'PH',
Decimal('1.000000000000000'), Decimal('154.300000000000011'),
Decimal('400000.000000000000000'), Decimal('0E-15'), 'NO', 2.0,
Decimal('838553782.39999997615814'), Decimal('1.000000000000000'),
Decimal('9001.000000000000000'), Decimal('1001028.997588216909207'),
Decimal('1016411.384014175971970'), Decimal('0E-15'),
Decimal('31.022020375038409'), Decimal('132.624585534118836'),
Decimal('102.587984691571350'), Decimal('2797.923570148704130'),
Decimal('1356.907786958068755'), Decimal('251.030171051320423'),
Decimal('1631.847033111359906'), Decimal('2591.530972494082562'),
Decimal('8781.765578473681671'), Decimal('1878.591433783431057'),
Decimal('1306.720761672033632'), Decimal('267.330342037544881'),
Decimal('632.441884708359200'), Decimal('589.145548007686102'),
Decimal('1856.402508952535982'), Decimal('545.193318791890533'),
Decimal('10944.909890410533990'), Decimal('111.565197915998837'),
Decimal('2492.107310746735948'), Decimal('733.827240501773645'),
Decimal('1558.177474543395419'), Decimal('4239.753002801080584'),
Decimal('0E-15'), Decimal('4185.200310278198231'),

Decimal('432.420139045859685'), Decimal('192.138505446132427'),
Decimal('2866.639841571150555'), '11', 'SUBA', Decimal('2564.825714115544088'),
'9114036002', '6612125153114', '6', '61', '21', '2', '51', '5', '1', '531', '5',
'3', '1', '14', '1', '4', '9114036', 'P', Decimal('2833333.000000000000000'),
Decimal('2173704.000000000000000'), Decimal('20.547189278375107'), '01060000202C
0C0000010000000103000000010000000F10000004BAF849B518D2E410736F10FD6042F41AB593DDD
248D2E4182EEB1C302042F4142D54630E08C2E41DBCA468423042F41B … (7466 characters
truncated) … D2E41B60238D6D6042F41F3D203C4518D2E4125C2C8D3D6042F410B2A5EC2518D
2E41F5328ECAD6042F418C4D13BA518D2E41615DADA1D6042F414BAF849B518D2E410736F10FD604
2F41')]

0.3 Count by type of property

```
%%sql
SELECT oft_tipo_inmueble,COUNT(*) from
data.ofertas_catastro
GROUP BY oft_tipo_inmueble;
```

 * postgresql://ds4a_t53:***@ds4av1.cd21iciz5zqe.us-
east-2.rds.amazonaws.com/ofertascatastro
27 rows affected.

[6]: [('Institucional', 1),
  ('Casa bodega', 4),
  ('Bodega comercial', 4),
  ('Dotacional', 38),
  ('Edificio comercial', 15),
  ('Apartamento', 50553),
  ('Residencias-hotel', 1),
  ('Inmueble comercial', 188),
  ('Depósito', 4),
  ('Parqueadero - lava autos', 8),
  ('Consultorio', 31),
  ('Casa uso mixto', 3982),
  ('Deposito', 187),
  ('Casa campestre', 1),
  ('Apartaestudio', 327),
  ('Garaje', 101),
  ('Edificio', 619),
  ('Casa lote', 331),
  ('Local', 1570),
  ('Casa', 23799),
  ('Casa residencial', 2),
  ('Parqueadero', 301),
  ('Lote', 1835),
  ('Casa comercial', 4),
  ('Industria', 17),

```
    ('Bodega', 1423),
    ('Oficina', 1626)]
```

0.4 The most expensive and cheapest properties

```
[8]: %%sql
     SELECT  oft_codigo, round(vfventa2020,2) as␣
      ↪vfventa2020,nombre_barrio,oft_tipo_inmueble  FROM data.ofertas_catastro
     ORDER BY vfventa2020 DESC
     LIMIT 5;
```

 * postgresql://ds4a_t53:***@ds4av1.cd21iciz5zqe.us-
east-2.rds.amazonaws.com/ofertascatastro
5 rows affected.

```
[8]: [('413843', Decimal('576834000000.00'), 'JULIO FLOREZ', 'Apartamento'),
      ('414529', Decimal('546997000000.00'), 'LOS CEDROS', 'Apartamento'),
      ('417368', Decimal('517161000000.00'), 'CEDRITOS', 'Apartamento'),
      ('404904', Decimal('397816000000.00'), 'TINTALITO', 'Casa'),
      ('513891', Decimal('344015000000.00'), 'EL RUBI', 'Casa')]
```

```
[9]: %%sql
     SELECT oft_codigo,  round(vfventa2020,2) as␣
      ↪vfventa2020,nombre_barrio,oft_tipo_inmueble FROM data.ofertas_catastro
     ORDER BY vfventa2020 ASC
     LIMIT 5;
```

 * postgresql://ds4a_t53:***@ds4av1.cd21iciz5zqe.us-
east-2.rds.amazonaws.com/ofertascatastro
5 rows affected.

```
[9]: [('525071', Decimal('215000.00'), 'BOSQUE CALDERON', 'Apartamento'),
      ('387982', Decimal('988998.96'), 'BRITALIA', 'Deposito'),
      ('379614', Decimal('988998.96'), 'BRITALIA', 'Deposito'),
      ('380046', Decimal('1047668.39'), 'SAN FELIPE DE USME', 'Lote'),
      ('388414', Decimal('1047668.39'), 'SAN FELIPE DE USME', 'Lote')]
```

0.5 Average price of properties by locality

```
[10]: %%sql
      SELECT locnombre,round(avg(vfventa2020),2) AS Valor_promedio_venta_2020
      FROM data.ofertas_catastro
      GROUP BY locnombre
      ORDER BY Valor_promedio_venta_2020 DESC;
```

 * postgresql://ds4a_t53:***@ds4av1.cd21iciz5zqe.us-
east-2.rds.amazonaws.com/ofertascatastro

```
19 rows affected.
```

```
[10]: [('CHAPINERO', Decimal('971844773.08')),
       ('PUENTE ARANDA', Decimal('906433187.27')),
       ('LOS MARTIRES', Decimal('796023917.87')),
       ('BARRIOS UNIDOS', Decimal('753405307.58')),
       ('USAQUEN', Decimal('710576011.04')),
       ('TEUSAQUILLO', Decimal('704608931.19')),
       ('CANDELARIA', Decimal('675672741.48')),
       ('ANTONIO NARIÑO', Decimal('653074796.95')),
       ('SUBA', Decimal('603577641.35')),
       ('FONTIBON', Decimal('597613360.89')),
       ('KENNEDY', Decimal('488032893.81')),
       ('SANTA FE', Decimal('483836949.48')),
       ('TUNJUELITO', Decimal('452325264.87')),
       ('ENGATIVA', Decimal('437033708.43')),
       ('RAFAEL URIBE URIBE', Decimal('292961149.55')),
       ('BOSA', Decimal('254594196.71')),
       ('SAN CRISTOBAL', Decimal('247088869.14')),
       ('CIUDAD BOLIVAR', Decimal('229002322.45')),
       ('USME', Decimal('225119234.69'))]
```

0.6 The correlation between two types of distances to the same places (must be correlated)

```
[11]: %%sql
      SELECT corr(d_park, dm_park) FROM
      data.ofertas_catastro
```

```
 * postgresql://ds4a_t53:***@ds4av1.cd21iciz5zqe.us-
east-2.rds.amazonaws.com/ofertascatastro
1 rows affected.
```

```
[11]: [(0.880087119113674,)]
```

0.7 Visualize the data

```
[12]: con = psycopg2.
       ↪connect(database='ofertascatastro',user='ds4a_t53',password='DS4ACatastro',host='ds4av1.
       ↪cd21iciz5zqe.us-east-2.rds.amazonaws.com')
      sql_ofertas = "SELECT * FROM data.ofertas_catastro"
      ofertas = gpd.GeoDataFrame.from_postgis(sql_ofertas, con, geom_col='geom')
```

```
[13]: ofertas_bm = ofertas.to_crs(epsg=3857)
```

```
[14]: ax = ofertas_bm.plot(figsize=(12, 10),  linewidth=0.2, column='vfventa2020',
       ↪scheme='Quantiles', k=5, cmap='GnBu')
```

```
ctx.add_basemap(ax, zoom=12)
```