

Initial review of the data

September 26, 2020

```
[2]: import esda
import pandas as pd
import geopandas as gpd
from geopandas import GeoDataFrame
import libpysal as lps
import numpy as np
import matplotlib.pyplot as plt
from shapely.geometry import Point
import folium
import branca
import psycopg2
import sqlalchemy
import re
%matplotlib inline
pd.set_option('display.max_columns', 500)
```

1 Valuation per block

```
[8]: pd.set_option('display.float_format', lambda x: '%.3f' % x)
```

```
[3]: Avaluo_cat_man = gpd.read_file('/home/rcamilo/Documents/DS4A/ds4a/initialData/
↳Reto MINTIC/Avaluo Catastral Manzana/AvaluoCatastralManzana/Avaluo_Manzana.
↳shp')
# Avaluo_com_man = gpd.read_file('/mnt/d/Documentos/DS4A/initialData/Avaluo_
↳Comercial Manzana/AvaluoComercialManzana/Avaluo_Manzana.shp')
```

```
[4]: Avaluo_cat_man.head()
```

```
[4]:
```

	OBJECTID	MANZANA_ID	CP_TERR_AR	GRUPOP_TER	AVALUO_COM	AVALUO_CAT	\
0	1	009259086	N	RESIDENCIAL	951874.0	733096.0	
1	2	001355027	N	RESIDENCIAL	973208.0	666239.0	
2	3	001355010	N	RESIDENCIAL	1100000.0	818602.0	
3	4	001355012	N	RESIDENCIAL	687104.0	469428.0	
4	5	001355021	N	RESIDENCIAL	697258.0	474940.0	

OBSERVACIO \

```

0 Este valor corresponde a la mediana y puede di...
1 Este valor corresponde a la mediana y puede di...
2 Este valor corresponde a la mediana y puede di...
3 Este valor corresponde a la mediana y puede di...
4 Este valor corresponde a la mediana y puede di...

```

	GLOBALID	SHAPE_Leng	SHAPE_Area \
0	{70BABE94-C17D-48FC-ADD0-BEF0283A45B4}	0.002464	1.213815e-07
1	{73AA8E80-CC4E-4DF5-887F-E1620A2356E4}	0.000684	2.818937e-08
2	{ED6BDEF8-72FB-40F3-B4B5-A979800F5B47}	0.000875	2.915943e-08
3	{2EBCB819-9A4E-4F41-9A8B-0578AE28045A}	0.001018	3.175301e-08
4	{F25CC9EA-C4FA-49D6-9878-5EF2EBEE73FA}	0.002131	9.812038e-08

	geometry
0	POLYGON ((-74.12100 4.74751, -74.12101 4.74747...
1	POLYGON ((-74.09776 4.55222, -74.09776 4.55221...
2	POLYGON ((-74.09745 4.55194, -74.09747 4.55190...
3	POLYGON ((-74.09723 4.55258, -74.09718 4.55258...
4	POLYGON ((-74.09668 4.55296, -74.09673 4.55288...

```
[5]: Avaluo_cat_man.CP_TERR_AR.unique()
```

```
[5]: array(['N', 'P', None], dtype=object)
```

```
[9]: Avaluo_cat_man.describe()
```

	OBJECTID	AVALUO_COM	AVALUO_CAT	SHAPE_Leng	SHAPE_Area
count	43847.000	43847.000	43847.000	43847.000	43847.000
mean	21924.000	1517346.309	1124711.015	0.003	0.000
std	12657.683	1296343.439	1006170.267	0.003	0.000
min	1.000	0.000	0.000	0.000	0.000
25%	10962.500	652376.500	487016.500	0.002	0.000
50%	21924.000	1400000.000	991554.000	0.002	0.000
75%	32885.500	2076806.000	1508997.000	0.003	0.000
max	43847.000	56041687.000	43399762.000	0.204	0.001

```
[10]: Avaluo_cat_man.dtypes
```

```

[10]: OBJECTID      int64
      MANZANA_ID    object
      CP_TERR_AR     object
      GRUPOP_TER     object
      AVALUO_COM     float64
      AVALUO_CAT     float64
      OBSERVACIO     object
      GLOBALID       object
      SHAPE_Leng     float64

```

```
SHAPE_Area      float64
geometry        geometry
dtype: object
```

```
[11]: Avaluo_cat_man.isnull().any()
```

```
[11]: OBJECTID      False
MANZANA_ID      True
CP_TERR_AR      True
GRUPOP_TER      True
AVALUO_COM      False
AVALUO_CAT      False
OBSERVACIO      False
GLOBALID      False
SHAPE_Leng      False
SHAPE_Area      False
geometry        False
dtype: bool
```

```
[14]: len(Avaluo_cat_man.index)
```

```
[14]: 43847
```

```
[16]: Avaluo_cat_man['REL'] = Avaluo_cat_man['AVALUO_CAT']/
↳Avaluo_cat_man['AVALUO_COM']
Avaluo_cat_man['REL'].mean()
```

```
[16]: 0.7426563618120285
```

The cadastral valuation is the 74% of the commercial valuation which means, according to references and cadastral standards, that the cadastral assessment is up to date

2 Reference Value

```
[12]: Valor_referencia = gpd.read_file('/home/rcamilo/Documents/DS4A/ds4a/initialData/
↳Reto MINTIC/Valor Referencia del Suelo/Valor_Ref/Valor_Ref.shp')
```

```
[13]: Valor_referencia.head()
```

```
[13]:  MANCODIGO      V_REF      AÑO  SHAPE_AREA  SHAPE_LEN  \
0  009109014  1420000.000  2016-01-01      0.000      0.003
1  009111022  2200000.000  2016-01-01      0.000      0.006
2  009111025  2100000.000  2016-01-01      0.000      0.003
3  009113030   800000.000  2016-01-01      0.000      0.022
4  009115041  2200000.000  2016-01-01      0.000      0.005
```

```
geometry
```

```

0 POLYGON ((-74.07808 4.70739, -74.07808 4.70726...
1 POLYGON ((-74.06901 4.70363, -74.06933 4.70355...
2 POLYGON ((-74.06507 4.70105, -74.06507 4.70105...
3 POLYGON ((-74.06706 4.76573, -74.06760 4.76499...
4 POLYGON ((-74.07292 4.72923, -74.07292 4.72895...

```

```
[15]: len(Valor_referencia.index)
```

```
[15]: 351159
```

```
[18]: Valor_referencia.isnull().any()
```

```

[18]: MANCODIGO      False
      V_REF          False
      AÑO            False
      SHAPE_AREA     False
      SHAPE_LEN      False
      geometry       False
      dtype: bool

```

```
[19]: Valor_referencia.dtypes
```

```

[19]: MANCODIGO      object
      V_REF          float64
      AÑO            object
      SHAPE_AREA     float64
      SHAPE_LEN      float64
      geometry       geometry
      dtype: object

```

3 Offers Clean

```

[12]: %reload_ext sql
      %sql postgresql://postgres:152622@localhost/MapaReferencia

```

```

[75]: ofertas = pd.read_csv('/home/rcamilo/Documents/DS4A/ds4a/initialData/Reto_
      ↪MINTIC/Ofertas/Ofertar.csv', sep=';', decimal=',')
      ofertas.head()

```

```

[75]:
      Nombre Barrio  Codigo Barrio  Codigo Manzana  Codigo Predio  \
0      ESTACION CENTRAL          6204            17            15
1  COOPERATIVA DE SUB-OFICIALES          6501            3            10
2      ESTACION CENTRAL          6204            4             4
3      ARBORIZADORA ALTA          2568           10           14
4      ARBORIZADORA ALTA          2568           10           12

```

	Codigo Construccion	Codigo Resto	Fecha De Cargue \
0	0	0	2019-09-30T00:00:00Z
1	0	0	2019-09-30T00:00:00Z
2	1	6011	2019-09-30T00:00:00Z
3	0	0	2019-09-30T00:00:00Z
4	0	0	2019-09-30T00:00:00Z

	Direccion	Codigo Oferta	Tipo Oferta	Tipo Inmueble	Fuente \
0	KR 33 17 27	500000.0	VENTA	Bodega	CAMPO
1	KR 71F 12 02	500001.0	VENTA	Casa uso mixto	CAMPO
2	CL 14 30 17 AP 611	500002.0	ARRIENDO	Apartamento	CAMPO
3	CL 74 SUR 38A 60	500003.0	ARRIENDO PARCIAL	Casa uso mixto	CAMPO
4	CL 74 SUR 38A 52	500004.0	VENTA	Casa uso mixto	CAMPO

	Chip	Area Terreno	Area Construida	Vr Inicial Venta \
0	AAA0251LMRJ	2060.80	2306	7.000000e+09
1	AAA0080RXPP	100.10	211,3	0.000000e+00
2	AAA0164RCHY	18.04	50,44	0.000000e+00
3	AAA0028RFAF	42.00	42	0.000000e+00
4	AAA0028REYN	42.00	130,2	1.500000e+08

	Vr Final Venta	Vr Inicial Arriendo	Vr Final Arriendo
0	6.880300e+09	0.0	0.0
1	0.000000e+00	0.0	0.0
2	0.000000e+00	0.0	0.0
3	0.000000e+00	1200000.0	1178000.0
4	1.453350e+08	0.0	0.0

```
[76]: ofertas['Chip'].value_counts()
```

```
[76]: AAA01640RAF    12
      AAA0159CCBS    10
      AAA0242TUNX     8
      AAA0227PPSY     6
      AAA0218CLKL     6
      ..
      AAA0065ZMOE     1
      AAA0055KERU     1
      AAA0105HBJH     1
      AAA0142ENZM     1
      AAA0042WEYX     1
      Name: Chip, Length: 17897, dtype: int64
```

```
[16]: nombres = [x.replace(' ','_') for x in list(ofertas)]
      ofertas.columns= nombres

      def clean_code(code,n):
```

```

if len(str(code)) < n:
    return ('0'*(n-len(str(code))))+str(code)
else:
    return str(code)

ofertas['Codigo_Barrio'] = ofertas['Codigo_Barrio'].apply(lambda x:
↳clean_code(x,6))
ofertas['Codigo_Manzana'] = ofertas['Codigo_Manzana'].apply(lambda x:
↳clean_code(x,3))
ofertas['Codigo_Predio'] = ofertas['Codigo_Predio'].apply(lambda x:
↳clean_code(x,3))
ofertas['Codigo_Construccion'] = ofertas['Codigo_Construccion'].apply(lambda x:
↳clean_code(x,3))
ofertas['Codigo_Resto'] = ofertas['Codigo_Resto'].apply(lambda x:
↳clean_code(x,5))
ofertas['Area_Construida'] = ofertas['Area_Construida'].apply(lambda x:
↳float(re.sub("-", "0", re.sub(".", ".", x))))
ofertas['Codigo_Oferta'] = ofertas['Codigo_Oferta'].apply(lambda x: str(int(x)))
ofertas.head()

```

```

[16]:
      Nombre_Barrio Codigo_Barrio Codigo_Manzana Codigo_Predio \
0      ESTACION CENTRAL      006204      017      015
1  COOPERATIVA DE SUB-OFICIALES      006501      003      010
2      ESTACION CENTRAL      006204      004      004
3      ARBORIZADORA ALTA      002568      010      014
4      ARBORIZADORA ALTA      002568      010      012

      Codigo_Construccion Codigo_Resto      Fecha_De_Cargue      Direccion \
0      000      00000  2019-09-30T00:00:00Z      KR 33 17 27
1      000      00000  2019-09-30T00:00:00Z      KR 71F 12 02
2      001      06011  2019-09-30T00:00:00Z  CL 14 30 17 AP 611
3      000      00000  2019-09-30T00:00:00Z  CL 74 SUR 38A 60
4      000      00000  2019-09-30T00:00:00Z  CL 74 SUR 38A 52

      Codigo_Oferta      Tipo_Oferta      Tipo_Inmueble Fuente      Chip \
0      500000      VENTA      Bodega CAMPO  AAA0251LMRJ
1      500001      VENTA      Casa uso mixto CAMPO  AAA0080RXPP
2      500002      ARRIENDO      Apartamento CAMPO  AAA0164RCHY
3      500003  ARRIENDO PARCIAL      Casa uso mixto CAMPO  AAA0028RFAF
4      500004      VENTA      Casa uso mixto CAMPO  AAA0028REYN

      Area_Terreno      Area_Construida      Vr_Inicial_Venta      Vr_Final_Venta \
0      2060.80      2306.00      7.000000e+09      6.880300e+09
1      100.10      211.30      0.000000e+00      0.000000e+00
2      18.04      50.44      0.000000e+00      0.000000e+00
3      42.00      42.00      0.000000e+00      0.000000e+00

```

4	42.00	130.20	1.500000e+08	1.453350e+08
---	-------	--------	--------------	--------------

	Vr_Inicial_Arriendo	Vr_Final_Arriendo
0	0.0	0.0
1	0.0	0.0
2	0.0	0.0
3	1200000.0	1178000.0
4	0.0	0.0

```
[47]: ofertas.dtypes
```

```
[47]: Nombre_Barrio      object
Codigo_Barrio          object
Codigo_Manzana         object
Codigo_Predio          object
Codigo_Construccion    object
Codigo_Resto           object
Fecha_De_Cargue        object
Direccion              object
Codigo_Oferta          object
Tipo_Oferta            object
Tipo_Inmueble          object
Fuente                 object
Chip                   object
Area_Terreno           float64
Area_Construida        float64
Vr_Inicial_Venta        float64
Vr_Final_Venta         float64
Vr_Inicial_Arriendo    float64
Vr_Final_Arriendo      float64
dtype: object
```

```
[54]: ofertas.to_csv('ofertas.csv',encoding='utf8',index=False)
```

```
[55]: list(pd.read_csv('ofertas.csv'))
```

```
[55]: ['Nombre_Barrio',
'Codigo_Barrio',
'Codigo_Manzana',
'Codigo_Predio',
'Codigo_Construccion',
'Codigo_Resto',
'Fecha_De_Cargue',
'Direccion',
'Codigo_Oferta',
'Tipo_Oferta',
'Tipo_Inmueble',
```

```
'Fuente',
'Chip',
'Area_Terreno',
'Area_Construida',
'Vr_Inicial_Venta',
'Vr_Final_Venta',
'Vr_Inicial_Arriendo',
'Vr_Final_Arriendo']
```

[48]: `%%sql`

```
CREATE TABLE DATA.ofertas(
Nombre_Barrio TEXT,
Codigo_Barrio VARCHAR(6),
Codigo_Manzana VARCHAR(3),
Codigo_Predio VARCHAR(3),
Codigo_Construccion VARCHAR(3),
Codigo_Resto VARCHAR(5),
Fecha_De_Cargue DATE,
Direccion TEXT,
Codigo_Oferta TEXT,
Tipo_Oferta TEXT,
Tipo_Inmueble TEXT,
Fuente TEXT,
Chip TEXT,
Area_Terreno FLOAT,
Area_Construida FLOAT,
Vr_Inicial_Venta FLOAT,
Vr_Final_Venta FLOAT,
Vr_Inicial_Arriendo FLOAT,
Vr_Final_Arriendo FLOAT
);
```

```
* postgresql://postgres:***@localhost/MapaReferencia
Done.
```

[48]: `[]`

```
psql -h localhost -U postgres -d MapaReferencia -c "\copy DATA.ofertas from 'ofertas.csv' with
```

[23]: `ofertas.dtypes`

```
[23]: Nombre Barrio          object
Codigo Barrio              int64
Codigo Manzana             int64
Codigo Predio              int64
Codigo Construccion        int64
```



```

Codigo Resto          int64
Fecha De Cargue       object
Direccion             object
Codigo Oferta         float64
Tipo Oferta           object
Tipo Inmueble          object
Fuente                object
Chip                  object
Area Terreno          float64
Area Construida       object
Vr Inicial Venta      float64
Vr Final Venta        float64
Vr Inicial Arriendo   float64
Vr Final Arriendo     float64
dtype: object

```

```

[78]: %%sql
SELECT COUNT(DISTINCT(prechip)) FROM data.predios_ofertas
;

```

```

* postgresql://postgres:***@localhost/MapaReferencia
1 rows affected.

```

```

[78]: [(17871,)]

```

```

[62]: %%sql
CREATE VIEW data.predios_ofertas as(
SELECT * FROM area_catastral.predio
WHERE prechip IN(SELECT chip FROM DATA.ofertas));

```

```

* postgresql://postgres:***@localhost/MapaReferencia
Done.

```

```

[62]: []

```

```

[88]: %%sql
SELECT * FROM data.predios_ofertas
WHERE prechip = 'AAA0000COWF' LIMIT 5;

```

```

* postgresql://postgres:***@localhost/MapaReferencia
4 rows affected.

```

```

[88]: [(1506, '001101', 'LAS BRISAS', '09', '02', '000', '00000', 'AAA0000COWF', '7S 6
14', '110010111040100090002000000000', 6, datetime.date(1997, 12, 31), 'N', 'CL
7 SUR 6 54', 'D', 'S', 'CL 7 SUR 6 64', Decimal('1622.00000000'),
Decimal('908.850000000'), '6222125152141', '06', 1957, 1998, 2020, '013',
Decimal('598.910000000'), 'N', 'S', 'N', 'A', datetime.date(2010, 11, 16), 22,
'111', '122', '132', '141', '212', '222', '233', '241', '000', '000', '332',

```

```
'000', '000', '000', '000', '000', '000', '000', 'C', '001101009002'),
(1507, '001101', 'LAS BRISAS', '09', '02', '000', '00000', 'AAA0000COWF', '7S 6
14', '110010111040100090002000000000', 6, datetime.date(1997, 12, 31), 'N', 'CL
7 SUR 6 54', 'D', 'S', 'CL 7 SUR 6 44', Decimal('1622.00000000'),
Decimal('908.850000000'), '6222125152141', '06', 1957, 1998, 2020, '013',
Decimal('598.910000000'), 'N', 'S', 'N', 'A', datetime.date(2010, 11, 16), 22,
'111', '122', '132', '141', '212', '222', '233', '241', '000', '000', '332',
'000', '000', '000', '000', '000', '000', '000', 'C', '001101009002'),
(1508, '001101', 'LAS BRISAS', '09', '02', '000', '00000', 'AAA0000COWF', '7S 6
14', '110010111040100090002000000000', 6, datetime.date(1997, 12, 31), 'N', 'CL
7 SUR 6 54', 'D', 'S', 'CL 7 SUR 6 64', Decimal('1622.00000000'),
Decimal('908.850000000'), '6222125152141', '06', 2004, 1998, 2020, '013',
Decimal('309.940000000'), 'N', 'S', 'N', 'B', datetime.date(2010, 11, 16), 51,
'114', '125', '132', '143', '212', '222', '233', '243', '000', '000', '332',
'000', '000', '000', '000', '000', '000', '000', 'C', '001101009002'),
(1509, '001101', 'LAS BRISAS', '09', '02', '000', '00000', 'AAA0000COWF', '7S 6
14', '110010111040100090002000000000', 6, datetime.date(1997, 12, 31), 'N', 'CL
7 SUR 6 54', 'D', 'S', 'CL 7 SUR 6 44', Decimal('1622.00000000'),
Decimal('908.850000000'), '6222125152141', '06', 2004, 1998, 2020, '013',
Decimal('309.940000000'), 'N', 'S', 'N', 'B', datetime.date(2010, 11, 16), 51,
'114', '125', '132', '143', '212', '222', '233', '243', '000', '000', '332',
'000', '000', '000', '000', '000', '000', '000', 'C', '001101009002')]
```

```
[ ]: %%sql
SELECT prechip, prevetustz, avg(prepuntaje)
FROM data.predios_ofertas
group by (prechip, prevetustz)
```

```
[82]: %%sql
SELECT * FROM data.ofertas
WHERE chip = 'AAA0251LMRJ' LIMIT 5;
```

```
* postgresql://postgres:***@localhost/MapaReferencia
1 rows affected.
```

```
[82]: [('ESTACION CENTRAL', '006204', '017', '015', '000', '00000',
datetime.date(2019, 9, 30), 'KR 33 17 27', '500000', 'VENTA', 'Bodega', 'CAMPO',
'AAA0251LMRJ', 2060.8, 2306.0, 7000000000.0, 6880300000.0, 0.0, 0.0)]
```

There is duplicated data of each chip and age of the parcel (pending quest to entity

4 Mapa de referencia

```
[17]: con = psycopg2.
↳connect(database='MapaReferencia',user='postgres',password='152622',host='localhost')
cursor = con.cursor()
```

```
[72]: cursor.execute("""select schema_name
from information_schema.schemata;""")
schemas = []
for s in cursor.fetchall():
    schemas.append(s)
```

```
[22]: schemas
```

```
[22]: [('pg_toast',),
      ('pg_temp_1',),
      ('pg_toast_temp_1',),
      ('pg_catalog',),
      ('information_schema',),
      ('topology',),
      ('area_catastral',),
      ('area_planeamiento',),
      ('corriente_agua',),
      ('elevacion',),
      ('entidad_territorial',),
      ('nomenclatura',),
      ('poblacion_vivienda',),
      ('punto_control',),
      ('superficie_agua',),
      ('toponimo',),
      ('transporte_terrestre',),
      ('zona_hidrografica',),
      ('public',),
      ('spatial',),
      ('data',)]
```

```
[75]: dictfinal = {}
for s in schemas[6:]:
    schema = s[0]
    cursor.execute(f"""SELECT table_name FROM information_schema.tables
WHERE table_schema = '{schema}'""")
    temp_list = []
    for table in cursor.fetchall():
        sql = f'SELECT * FROM {schema}.{table[0]}'
        temp_list.append(table[0])
    dictfinal[schema] = temp_list
```

```
[76]: dictfinal
```

```
[76]: {'area_catastral': ['lote', 'predio', 'manz', 'scat', 'uso', 'cons'],
      'area_planeamiento': ['aurb', 'suel', 'upla'],
      'corriente_agua': ['cdag'],
      'elevacion': ['cniv'],
```

```

'entidad_territorial': ['corr', 'cpob', 'loca', 'muni'],
'nomenclatura': ['cpos', 'pdom'],
'poblacion_vivienda': ['esoc'],
'punto_control': ['pgeo'],
'superficie_agua': ['cagu'],
'toponimo': ['ngeo'],
'transporte_terrestre': ['aitr',
    'ande',
    'estado',
    'cicl',
    'calz',
    'puen',
    'psitp',
    'rbic',
    'sepa',
    'ntra',
    'mvi'],
'zona_hidrografica': ['cuenca',
    'microcuenca',
    'subcuenca',
    'subzonahidrografica'],
'public': ['d_preclasepr',
    'geography_columns',
    'geometry_columns',
    'spatial_ref_sys',
    'd_pretdsi',
    'd_pretpropie',
    'd_predestino',
    'd_premdirecc',
    'd_rbiclase',
    'd_rbisentido',
    'd_rbitsuperf',
    'd_sepmosfera',
    'd_aitestado',
    'd_aitfuente',
    'd_aittransit',
    'd_aitzsitp',
    'd_andmosfera',
    'd_calestado',
    'd_calfuncion',
    'd_caltsuperf',
    'd_cictsuperf',
    'd_conmejora',
    'd_contsemis',
    'd_lotdispers',
    'd_lotdistrit',
    'd_corriente_agua',

```

```

'd_cnitipo',
'd_cpotipo',
'd_esostrato',
'd_cuerpo_agua',
'd_mvisvia',
'd_mvitypo',
'd_ntrmodo',
'd_ngeestado',
'd_ntrmtransp',
'd_ntrtypo',
'd_ucoviviend',
'sbt_scatipo',
'sbt_pdotipo',
'sbt_mvitycla',
'd_uconph',
'd_ucoph',
'd_usotuso',
'sbt_suecsuelo',
'sbt_upltypo',
'sbt_puetipo',
'd_aurtplano',
'd_ngeclasifi',
'sbt_rbilocaliz'],
'spatial': [],
'data': ['ofertas', 'predios_ofertas']]

```

```

[17]: #For read all the data but is unnecessary, takes a lot, DONT RUN
#para leer todas las tablas creo que no es necesario y se demora mucho y
→ algunas tablas no tienen geom, creo que no es necesario
# dictfinal = {}
# for s in schemas[6:7]:
#     schema = s[0]
#     cursor.execute(f"""SELECT table_name FROM information_schema.tables
# WHERE table_schema = '{schema}'""")
#     temp_dict = {}
#     for table in cursor.fetchall():
#         sql = f'SELECT * FROM {schema}.{table[0]}'
#         temp_dict[table] = gpd.GeoDataFrame.from_postgis(sql, con,
→ geom_col='geom')
#     dictfinal[schema] = temp_dict

```

```

[37]: cursor.execute("""SELECT COUNT(*) FROM entidad_territorial.loca ;""")
cursor.fetchall()

```

```

[37]: [(20,)]

```

```
[18]: sql_sectCat = 'SELECT * FROM area_catastral.scat'
      sql_zhs = 'SELECT * FROM zona_hidrografica.subcuenca'
      sql_lote = 'SELECT * FROM area_catastral.lote'
      # without limit, crashes
      sql_predio = 'SELECT * FROM area_catastral.predio LIMIT 100'
```

```
[8]: SectorCatastral = gpd.GeoDataFrame.from_postgis(sql_sectCat, con,
      ↪geom_col='geom')
      ZonaHidrograficaSubcuenca = gpd.GeoDataFrame.from_postgis(sql_zhs, con,
      ↪geom_col='geom')
      Lote = gpd.GeoDataFrame.from_postgis(sql_lote, con, geom_col='geom')
      Predio = pd.read_sql_query(sql_predio, con)
```

```
[19]: Lote = gpd.GeoDataFrame.from_postgis(sql_lote, con, geom_col='geom')
```

```
[5]: Predio.head()
```

```
[5]:
```

	gid	precbarrio	prenbarrio	precmanz	precpredio	precons	precresto	\
0	1	001101	LAS BRISAS	04	35	000	00000	
1	2	001101	LAS BRISAS	05	01	000	00000	
2	3	001101	LAS BRISAS	05	02	000	00000	
3	4	001101	LAS BRISAS	05	03	000	00000	
4	5	001101	LAS BRISAS	05	04	000	00000	

		prechip	precedcata		prenupre	...	prebmobili	\
0	AAA0000BAMR	3S 5 21	110010111040100040035000000000	...	None			
1	AAA0000BANX	4S 5 26	110010111040100050001000000000	...	None			
2	AAA0000BAOM	4S 5 27	110010111040100050002000000000	...	332			
3	AAA0000BAPA	4S 5 28	110010111040100050003000000000	...	None			
4	AAA0000BARJ	4S 5 29	110010111040100050004000000000	...	331			

		prebcons	prectamano	precenchap	precmobili	preconse	precindus	preacercha	\
0	None	411	None	None	None	None	None	None	
1	None	000	None	None	None	000	000	000	
2	341	412	422	432	441	000	000	000	
3	341	412	422	431	441	None	None	None	
4	341	412	422	431	441	None	None	None	

		preclcons	barmanpre
0	R	001101004035	
1	R	001101005001	
2	R	001101005002	
3	R	001101005003	
4	R	001101005004	

[5 rows x 52 columns]

```
[92]: Lote.head()
```

```
[92]:   gid    lotcodigo lotdispers lotildispe  lotupredia manzcodigo  \
0    1    009108093009          N          None          1    009108093
1    2    009108093010          N          None          1    009108093
2    3    009222023018          N          None          1    009222023
3    4    005675053034          N          None          1    005675053
4    5    005675053036          N          None          1    005675053

                                geom  lotdistrit
0  MULTIPOLYGON (((-74.07234 4.72298, -74.07235 4...          0
1  MULTIPOLYGON (((-74.07228 4.72295, -74.07229 4...          0
2  MULTIPOLYGON (((-74.08605 4.72798, -74.08605 4...          0
3  MULTIPOLYGON (((-74.13056 4.70580, -74.13057 4...          0
4  MULTIPOLYGON (((-74.13046 4.70576, -74.13048 4...          0
```

```
[10]: SectorCatastral.head()
```

```
[10]:   gid scacodigo  scatipo          scanombre  \
0    1    002598        0  EL MOCHUELO II URBANO
1    2    004524        0          SAN PABLO BOSA
2    3    001344        0          LOS SOCHES
3    4    002630        0  EL PORTAL DEL DIVINO
4    7    005103        0          SAN MIGUEL

                                geom
0  MULTIPOLYGON (((-74.12983 4.53681, -74.12938 4...
1  MULTIPOLYGON (((-74.19447 4.60992, -74.19288 4...
2  MULTIPOLYGON (((-74.08600 4.50096, -74.08693 4...
3  MULTIPOLYGON (((-74.09576 4.48804, -74.09586 4...
4  MULTIPOLYGON (((-74.07701 4.65949, -74.07734 4...
```

```
[38]: ZonaHidrograficaSubcuenca.head()
```

```
[38]:   gid  codch    nomch  codsch    nomsch  \
0    1  2120.0  Río Bogotá  212005.0  Embalse del Muña
1    2  2120.0  Río Bogotá  212002.0    Río Calandaima
2    3  2120.0  Río Bogotá  212001.0    Río Bajo Bogotá
3    4  2120.0  Río Bogotá  212006.0    Río Soacha
4    5  2120.0  Río Bogotá  212007.0    Río Tunjuelo

                                geom
0  MULTIPOLYGON (((-74.24142 4.54149, -74.23972 4...
1  MULTIPOLYGON (((-74.49380 4.54855, -74.49311 4...
2  MULTIPOLYGON (((-74.61330 4.58091, -74.61265 4...
3  MULTIPOLYGON (((-74.17585 4.50547, -74.17563 4...
4  MULTIPOLYGON (((-74.17568 4.50378, -74.17564 4...
```

4.1 Join Parcels with Offers to locate them

```
[5]: ofertas.isnull().any()
```

```
[5]: Nombre Barrio      False
Codigo Barrio          False
Codigo Manzana         False
Codigo Predio          False
Codigo Construccion    False
Codigo Resto           False
Fecha De Cargue        False
Direccion              False
Codigo Oferta          False
Tipo Oferta            False
Tipo Inmueble          False
Fuente                 False
Chip                  False
Area Terreno           False
Area Construida        False
Vr Inicial Venta       False
Vr Final Venta         False
Vr Inicial Arriendo    False
Vr Final Arriendo      False
dtype: bool
```

```
[14]: ofertas[(ofertas['Vr Final Venta'] == '0') & (ofertas['Vr Inicial Venta'] == '0')]
```

```
[14]:
```

	Nombre Barrio	Codigo Barrio	Codigo Manzana	\
1	COOPERATIVA DE SUB-OFICIALES	6501	3	
2	ESTACION CENTRAL	6204	4	
3	ARBORIZADORA ALTA	2568	10	
5	SAN PABLO BOSA	4524	14	
7	BOLONIA	2605	29	
...	
18781	SANTA MATILDE	4209	7	
18782	SANTA MATILDE	4209	2	
18783	BOCHICA	4211	22	
18784	BOCHICA	4211	22	
18785	BOCHICA	4211	14	

	Codigo Predio	Codigo Construccion	Codigo Resto	Fecha De Cargue	\
1	10	0	0	2019-09-30T00:00:00Z	
2	4	1	6011	2019-09-30T00:00:00Z	
3	14	0	0	2019-09-30T00:00:00Z	
5	26	0	0	2019-09-30T00:00:00Z	
7	9	0	0	2019-09-30T00:00:00Z	
...	

18781	13	1	3001	2020-08-04T00:00:00Z
18782	2	1	3001	2020-08-04T00:00:00Z
18783	13	0	0	2020-08-04T00:00:00Z
18784	11	0	0	2020-08-04T00:00:00Z
18785	6	0	0	2020-08-04T00:00:00Z

	Direccion	Codigo Oferta	Tipo Oferta	Tipo Inmueble \
1	KR 71F 12 02	500001	VENTA	Casa uso mixto
2	CL 14 30 17 AP 611	500002	ARRIENDO	Apartamento
3	CL 74 SUR 38A 60	500003	ARRIENDO PARCIAL	Casa uso mixto
5	CL 67 SUR 78D 15	500005	ARRIENDO PARCIAL	Casa
7	CL 81C SUR 4 15 ESTE	500007	ARRIENDO PARCIAL	Casa uso mixto
...
18781	KR 31C 1 41 AP 301	518892	VENTA	Apartamento
18782	KR 31B 1 06 AP 301	518893	VENTA	Apartamento
18783	KR 35 2 70	518894	VENTA	Casa
18784	KR 35 2 48	518895	ARRIENDO PARCIAL	Casa
18785	KR 35A 1H 32	518896	VENTA	Casa

	Fuente	Chip	Area Terreno	Area Construida \
1	CAMPO	AAA0080RXPP	100,1	211,3
2	CAMPO	AAA0164RCHY	18,04	50,44
3	CAMPO	AAA0028RFAF	42	42
5	CAMPO	AAA0046HECN	184	347,5
7	CAMPO	AAA0144UWUH	72,3	72,3
...
18781	SIN REGISTRO	AAA0271EMTO	30,76	58,88
18782	SIN REGISTRO	AAA0215YCEA	14,63	34,44
18783	SIN REGISTRO	AAA0037LKAW	220,1	438,01
18784	SIN REGISTRO	AAA0037LJYX	137,2	432,88
18785	SIN REGISTRO	AAA0037KYYX	133	398,9

	Vr Inicial Venta	Vr Final Venta	Vr Inicial Arriendo	Vr Final Arriendo
1	0	0	0	0
2	0	0	0	0
3	0	0	1200000	1178000
5	0	0	5,00E+05	491000
7	0	0	380000	373000
...
18781	0	0	0	0
18782	0	0	0	0
18783	0	0	0	0
18784	0	0	0	0
18785	0	0	0	0

[9089 rows x 19 columns]

```
[15]: ofertas[(ofertas['Vr Final Arriendo'] == '0') & (ofertas['Vr Inicial Arriendo']_
↪== '0')]
```

```
[15]:
```

	Nombre Barrio	Codigo Barrio	Codigo Manzana	\
0	ESTACION CENTRAL	6204	17	
1	COOPERATIVA DE SUB-OFICIALES	6501	3	
2	ESTACION CENTRAL	6204	4	
4	ARBORIZADORA ALTA	2568	10	
6	BOLONIA	2605	29	
...	
18781	SANTA MATILDE	4209	7	
18782	SANTA MATILDE	4209	2	
18783	BOCHICA	4211	22	
18784	BOCHICA	4211	22	
18785	BOCHICA	4211	14	

	Codigo Predio	Codigo Construcccion	Codigo Resto	Fecha De Cargue	\
0	15	0	0	2019-09-30T00:00:00Z	
1	10	0	0	2019-09-30T00:00:00Z	
2	4	1	6011	2019-09-30T00:00:00Z	
4	12	0	0	2019-09-30T00:00:00Z	
6	9	0	0	2019-09-30T00:00:00Z	
...	
18781	13	1	3001	2020-08-04T00:00:00Z	
18782	2	1	3001	2020-08-04T00:00:00Z	
18783	13	0	0	2020-08-04T00:00:00Z	
18784	11	0	0	2020-08-04T00:00:00Z	
18785	6	0	0	2020-08-04T00:00:00Z	

	Direccion	Codigo Oferta	Tipo Oferta	Tipo Inmueble	\
0	KR 33 17 27	5,00E+05	VENTA	Bodega	
1	KR 71F 12 02	500001	VENTA	Casa uso mixto	
2	CL 14 30 17 AP 611	500002	ARRIENDO	Apartamento	
4	CL 74 SUR 38A 52	500004	VENTA	Casa uso mixto	
6	CL 81C SUR 4 15 ESTE	500006	VENTA	Casa uso mixto	
...	
18781	KR 31C 1 41 AP 301	518892	VENTA	Apartamento	
18782	KR 31B 1 06 AP 301	518893	VENTA	Apartamento	
18783	KR 35 2 70	518894	VENTA	Casa	
18784	KR 35 2 48	518895	ARRIENDO PARCIAL	Casa	
18785	KR 35A 1H 32	518896	VENTA	Casa	

	Fuente	Chip	Area Terreno	Area Construida	\
0	CAMPO	AAA0251LMRJ	2060,8	2306	
1	CAMPO	AAA0080RXPP	100,1	211,3	
2	CAMPO	AAA0164RCHY	18,04	50,44	
4	CAMPO	AAA0028REYN	42	130,2	

6	CAMPO	AAA0144UWUH	72,3	72,3
...
18781	SIN REGISTRO	AAA0271EMTO	30,76	58,88
18782	SIN REGISTRO	AAA0215YCEA	14,63	34,44
18783	SIN REGISTRO	AAA0037LKAW	220,1	438,01
18784	SIN REGISTRO	AAA0037LJYX	137,2	432,88
18785	SIN REGISTRO	AAA0037KYYX	133	398,9

	Vr Inicial Venta	Vr Final Venta	Vr Inicial Arriendo	Vr Final Arriendo
0	7,00E+09	6880300000	0	0
1	0	0	0	0
2	0	0	0	0
4	1,50E+08	145335000	0	0
6	5,00E+07	48345000	0	0
...
18781	0	0	0	0
18782	0	0	0	0
18783	0	0	0	0
18784	0	0	0	0
18785	0	0	0	0

[13271 rows x 19 columns]

```
[22]: ofertas['lotcodigo'] =
      ↳ofertas['Codigo_Barrío'] +ofertas['Codigo_Manzana'] +ofertas['Codigo_Predio']
```

```
[95]: ofertas.head()
```

```
[95]:
```

	Nombre_Barrío	Codigo_Barrío	Codigo_Manzana	Codigo_Predio	\
0	ESTACION CENTRAL	006204	017	015	
1	COOPERATIVA DE SUB-OFICIALES	006501	003	010	
2	ESTACION CENTRAL	006204	004	004	
3	ARBORIZADORA ALTA	002568	010	014	
4	ARBORIZADORA ALTA	002568	010	012	

	Codigo_Construccion	Codigo_Resto	Fecha_De_Cargue	Direccion	\
0	000	00000	2019-09-30T00:00:00Z	KR 33 17 27	
1	000	00000	2019-09-30T00:00:00Z	KR 71F 12 02	
2	001	06011	2019-09-30T00:00:00Z	CL 14 30 17 AP 611	
3	000	00000	2019-09-30T00:00:00Z	CL 74 SUR 38A 60	
4	000	00000	2019-09-30T00:00:00Z	CL 74 SUR 38A 52	

	Codigo_Oferta	Tipo_Oferta	Tipo_Inmueble	Fuente	Chip	\
0	500000	VENTA	Bodega	CAMPO	AAA0251LMRJ	
1	500001	VENTA	Casa uso mixto	CAMPO	AAA0080RXPP	
2	500002	ARRIENDO	Apartamento	CAMPO	AAA0164RCHY	
3	500003	ARRIENDO PARCIAL	Casa uso mixto	CAMPO	AAA0028RFAF	

```
4          500004          VENTA Casa uso mixto CAMPO AAA0028REYN
```

	Area_Terreno	Area_Construida	Vr_Inicial_Venta	Vr_Final_Venta	\
0	2060.80	2306.00	7.000000e+09	6.880300e+09	
1	100.10	211.30	0.000000e+00	0.000000e+00	
2	18.04	50.44	0.000000e+00	0.000000e+00	
3	42.00	42.00	0.000000e+00	0.000000e+00	
4	42.00	130.20	1.500000e+08	1.453350e+08	

	Vr_Inicial_Arriendo	Vr_Final_Arriendo	lotcodigo
0	0.0	0.0	006204017015
1	0.0	0.0	006501003010
2	0.0	0.0	006204004004
3	1200000.0	1178000.0	002568010014
4	0.0	0.0	002568010012

```
[96]: Lote[Lote.lotcodigo=='006501003010']
```

```
[96]:          gid      lotcodigo lotdispers lotildispe  lotupredia manzcodigo  \
826640  826561  006501003010          N      None          1  006501003

                                geom  lotdistrit
826640  MULTIPOLYGON (((-74.13047 4.64354, -74.13052 4...          0
```

```
[97]: Lotes_ofertas = pd.merge(Lote,ofertas,on='lotcodigo',how='right')
```

```
[ ]: Lotes_ofertas.to_file("/home/rcamilo/Documents/DS4A_53/project/transform/
↳geocode/ofertas_lote/ofertas_lote.shp")
```

```
[58]: Lotes_ofertas.crs
```

```
[58]: <Geographic 2D CRS: EPSG:4686>
Name: MAGNA-SIRGAS
Axis Info [ellipsoidal]:
- Lat[north]: Geodetic latitude (degree)
- Lon[east]: Geodetic longitude (degree)
Area of Use:
- name: Colombia
- bounds: (-84.77, -4.23, -66.87, 15.51)
Datum: Marco Geocentrico Nacional de Referencia
- Ellipsoid: GRS 1980
- Prime Meridian: Greenwich
```

```
[20]: Lote_3116 = Lote.to_crs("EPSG:3116")
```

```
[24]: Lotes_ofertas_3116['X'] = Lotes_ofertas_3116.centroid.x
Lotes_ofertas_3116['Y'] = Lotes_ofertas_3116.centroid.y
```

```
[23]: Lotes_ofertas_3116 = pd.merge(Lote_3116,ofertas,on='lotcodigo',how='right')
```

```
[25]: Lotes_ofertas_3116.to_file("/home/rcamilo/Documents/DS4A_53/project/transform/
↳geocode/ofertas_lote_3116/ofertas_lote.shp")
```

```
[32]: Lotes_ofertas_3116.to_file("/home/rcamilo/Documents/DS4A_53/project/transform/
↳geocode/ofertas_lote_3116/ofertasPolygon.geojson", driver='GeoJSON')
```

```
[31]: Lotes_ofertas_3116.head()
```

```
[31]:      gid      lotcodigo lotdispers lotildispe lotupredia manzcodigo \
0   52608.0  006204017015          N         None          1.0  006204017
1   826561.0  006501003010          N         None          1.0  006501003
2   763098.0  006204004004          N         None         67.0  006204004
3   763098.0  006204004004          N         None         67.0  006204004
4   763098.0  006204004004          N         None         67.0  006204004
```

```
      geom      lotdistrib \
0  MULTIPOLYGON (((998313.698 1002469.281, 998311...      0.0
1  MULTIPOLYGON (((994123.577 1005234.898, 994117...      0.0
2  MULTIPOLYGON (((998440.059 1002066.050, 998424...      0.0
3  MULTIPOLYGON (((998440.059 1002066.050, 998424...      0.0
4  MULTIPOLYGON (((998440.059 1002066.050, 998424...      0.0
```

```
      Nombre_Barrío Código_Barrío ... Fuente      Chip \
0      ESTACION CENTRAL      006204 ... CAMPO  AAA0251LMRJ
1  COOPERATIVA DE SUB-OFICIALES      006501 ... CAMPO  AAA0080RXPP
2      ESTACION CENTRAL      006204 ... CAMPO  AAA0164RCHY
3      ESTACION CENTRAL      006204 ... CAMPO  AAA0164RBWW
4      ESTACION CENTRAL      006204 ... CAMPO  AAA0164RBNX
```

```
      Área_Terreno Área_Construida Vr_Inicial_Venta Vr_Final_Venta \
0      2060.80      2306.00      7.000000e+09      6.880300e+09
1      100.10      211.30      0.000000e+00      0.000000e+00
2      18.04      50.44      0.000000e+00      0.000000e+00
3      18.06      50.50      0.000000e+00      0.000000e+00
4      17.38      48.58      1.800000e+08      1.745100e+08
```

```
      Vr_Inicial_Arriendo Vr_Final_Arriendo      X      Y
0      0.0      0.0  998325.432468  1.002503e+06
1      0.0      0.0  994116.848415  1.005239e+06
2      0.0      0.0  998415.474873  1.002071e+06
3      1300000.0      1300000.0  998415.474873  1.002071e+06
4      0.0      0.0  998415.474873  1.002071e+06
```

```
[5 rows x 29 columns]
```

```
[26]: Lotes_ofertas_3116_punto = Lotes_ofertas_3116.copy()

[27]: Lotes_ofertas_3116_punto['geom'] = Lotes_ofertas_3116_punto.centroid

[28]: Lotes_ofertas_3116_punto.to_file("/home/rcamilo/Documents/DS4A_53/project/
↳transform/geocode/ofertas_lote_3116_point/ofertas_point.shp")

[33]: Lotes_ofertas_3116_punto.to_file("/home/rcamilo/Documents/DS4A_53/project/
↳transform/geocode/ofertas_lote_3116_point/ofertasPoint.geojson",
↳driver='GeoJSON')

[30]: len(ofertas.index),len(Lotes_ofertas_3116.index),len(Lotes_ofertas_3116_punto.
↳index)
```

```
[30]: (18786, 18786, 18786)
```

```
[35]: Lotes_ofertas_3116_punto.head()
```

```
[35]:
```

	gid	lotcodigo	lotdispers	lotildispe	lotupredia	manzcodigo	\
0	52608.0	006204017015	N	None	1.0	006204017	
1	826561.0	006501003010	N	None	1.0	006501003	
2	763098.0	006204004004	N	None	67.0	006204004	
3	763098.0	006204004004	N	None	67.0	006204004	
4	763098.0	006204004004	N	None	67.0	006204004	

		geom	lotdistrit	Nombre_Barrio	\
0	POINT (998325.432 1002503.318)	0.0	ESTACION CENTRAL		
1	POINT (994116.848 1005239.151)	0.0	COOPERATIVA DE SUB-OFICIALES		
2	POINT (998415.475 1002070.740)	0.0	ESTACION CENTRAL		
3	POINT (998415.475 1002070.740)	0.0	ESTACION CENTRAL		
4	POINT (998415.475 1002070.740)	0.0	ESTACION CENTRAL		

	Codigo_Barrio	...	Fuente	Chip	Area_Terreno	Area_Construida	\
0	006204	...	CAMPO	AAA0251LMRJ	2060.80	2306.00	
1	006501	...	CAMPO	AAA0080RXPP	100.10	211.30	
2	006204	...	CAMPO	AAA0164RCHY	18.04	50.44	
3	006204	...	CAMPO	AAA0164RBWW	18.06	50.50	
4	006204	...	CAMPO	AAA0164RBNX	17.38	48.58	

	Vr_Inicial_Venta	Vr_Final_Venta	Vr_Inicial_Arriendo	Vr_Final_Arriendo	\
0	7.000000e+09	6.880300e+09	0.0	0.0	
1	0.000000e+00	0.000000e+00	0.0	0.0	
2	0.000000e+00	0.000000e+00	0.0	0.0	
3	0.000000e+00	0.000000e+00	1300000.0	1300000.0	
4	1.800000e+08	1.745100e+08	0.0	0.0	

X Y

```

0  998325.432468  1.002503e+06
1  994116.848415  1.005239e+06
2  998415.474873  1.002071e+06
3  998415.474873  1.002071e+06
4  998415.474873  1.002071e+06

```

[5 rows x 29 columns]

```
[38]: lote_oferta_table = Lotes_ofertas_3116_punto.drop(columns=['geom'])
lote_oferta_table.to_csv('ofertas.csv',encoding='utf8',index=False)
```

```
[73]: ofertas['Chip'].value_counts()
```

```

[73]: AAA01640RAF      12
      AAA0159CCBS      10
      AAA0242TUNX       8
      AAA0227PPSY       6
      AAA0218CLKL       6
      ..
      AAA0065ZMOE       1
      AAA0055KERU       1
      AAA0105HBJH       1
      AAA0142ENZM       1
      AAA0042WEYX       1
      Name: Chip, Length: 17897, dtype: int64

```

```
[69]: lote_oferta_table[lote_oferta_table['Chip']=='AAA01640RAF']
```

```

[69]:      gid      lotcodigo lotdispers lotildispe  lotupredia manzcodigo  \
6582  595956.0  008501050012      N      None      524.0  008501050
6583  595956.0  008501050012      N      None      524.0  008501050
6584  595956.0  008501050012      N      None      524.0  008501050
6587  595956.0  008501050012      N      None      524.0  008501050
6591  595956.0  008501050012      N      None      524.0  008501050
6592  595956.0  008501050012      N      None      524.0  008501050
6593  595956.0  008501050012      N      None      524.0  008501050
6594  595956.0  008501050012      N      None      524.0  008501050
6596  595956.0  008501050012      N      None      524.0  008501050
6599  595956.0  008501050012      N      None      524.0  008501050
6600  595956.0  008501050012      N      None      524.0  008501050
6601  595956.0  008501050012      N      None      524.0  008501050

      lotdistrit Nombre_Barrío Código_Barrío Código_Manzana Código_Predio  \
6582      0.0      EL TOBERIN      008501      050      012
6583      0.0      EL TOBERIN      008501      050      012
6584      0.0      EL TOBERIN      008501      050      012
6587      0.0      EL TOBERIN      008501      050      012

```

6591	0.0	EL TOBERIN	008501	050	012
6592	0.0	EL TOBERIN	008501	050	012
6593	0.0	EL TOBERIN	008501	050	012
6594	0.0	EL TOBERIN	008501	050	012
6596	0.0	EL TOBERIN	008501	050	012
6599	0.0	EL TOBERIN	008501	050	012
6600	0.0	EL TOBERIN	008501	050	012
6601	0.0	EL TOBERIN	008501	050	012

	Codigo_Construccion	Codigo_Resto	Fecha_De_Cargue \
6582	001	01001	2019-12-16T00:00:00Z
6583	001	01001	2019-12-17T00:00:00Z
6584	001	01001	2019-12-17T00:00:00Z
6587	001	01001	2019-12-17T00:00:00Z
6591	001	01001	2019-12-17T00:00:00Z
6592	001	01001	2019-12-17T00:00:00Z
6593	001	01001	2019-12-17T00:00:00Z
6594	001	01001	2019-12-17T00:00:00Z
6596	001	01001	2019-12-17T00:00:00Z
6599	001	01001	2019-12-17T00:00:00Z
6600	001	01001	2019-12-17T00:00:00Z
6601	001	01001	2019-12-17T00:00:00Z

	Direccion	Codigo_Oferta	Tipo_Oferta	Tipo_Inmueble \
6582	CL 169 16C 10 BQ 1 AP 101	506602	VENTA, ARRIENDO	Apartamento
6583	CL 169 16C 10 BQ 1 AP 101	506625	VENTA	Apartamento
6584	CL 169 16C 10 BQ 1 AP 101	506626	VENTA	Apartamento
6587	CL 169 16C 10 BQ 1 AP 101	506629	ARRIENDO	Apartamento
6591	CL 169 16C 10 BQ 1 AP 101	506640	VENTA	Apartamento
6592	CL 169 16C 10 BQ 1 AP 101	506641	VENTA	Apartamento
6593	CL 169 16C 10 BQ 1 AP 101	506642	ARRIENDO	Apartamento
6594	CL 169 16C 10 BQ 1 AP 101	506643	VENTA	Apartamento
6596	CL 169 16C 10 BQ 1 AP 101	506645	VENTA	Apartamento
6599	CL 169 16C 10 BQ 1 AP 101	506648	VENTA	Apartamento
6600	CL 169 16C 10 BQ 1 AP 101	506649	VENTA	Apartamento
6601	CL 169 16C 10 BQ 1 AP 101	506651	VENTA	Apartamento

	Fuente	Chip	Area_Terreno	Area_Construida	Vr_Inicial_Venta \
6582	CAMPO	AAA0164ORAF	23.69	57.32	220000000.0
6583	CAMPO	AAA0164ORAF	23.69	57.32	200000000.0
6584	CAMPO	AAA0164ORAF	23.69	57.32	0.0
6587	CAMPO	AAA0164ORAF	23.69	57.32	0.0
6591	CAMPO	AAA0164ORAF	23.69	57.32	0.0
6592	CAMPO	AAA0164ORAF	23.69	57.32	0.0
6593	CAMPO	AAA0164ORAF	23.69	57.32	0.0
6594	CAMPO	AAA0164ORAF	23.69	57.32	0.0
6596	CAMPO	AAA0164ORAF	23.69	57.32	0.0

6599	CAMPO	AAA0164ORAF	23.69	57.32	0.0
6600	CAMPO	AAA0164ORAF	23.69	57.32	0.0
6601	CAMPO	AAA0164ORAF	23.69	57.32	0.0

	Vr_Final_Venta	Vr_Inicial_Arriendo	Vr_Final_Arriendo	X \
6582	220000000.0	900000.0	900000.0	1.004362e+06
6583	193980000.0	0.0	0.0	1.004362e+06
6584	0.0	0.0	0.0	1.004362e+06
6587	0.0	0.0	0.0	1.004362e+06
6591	0.0	0.0	0.0	1.004362e+06
6592	0.0	0.0	0.0	1.004362e+06
6593	0.0	0.0	0.0	1.004362e+06
6594	0.0	0.0	0.0	1.004362e+06
6596	0.0	0.0	0.0	1.004362e+06
6599	0.0	0.0	0.0	1.004362e+06
6600	0.0	0.0	0.0	1.004362e+06
6601	0.0	0.0	0.0	1.004362e+06

	Y
6582	1.016775e+06
6583	1.016775e+06
6584	1.016775e+06
6587	1.016775e+06
6591	1.016775e+06
6592	1.016775e+06
6593	1.016775e+06
6594	1.016775e+06
6596	1.016775e+06
6599	1.016775e+06
6600	1.016775e+06
6601	1.016775e+06

```
[70]: ofertas_no_dup = lote_oferta_table.drop_duplicates(subset=['Chip',
↳, 'Area_Terreno', 'Area_Construida',
↳, 'Vr_Inicial_Venta', 'Vr_Final_Venta', 'Vr_Inicial_Arriendo',
↳, 'Vr_Final_Arriendo'])
```

```
[71]: ofertas_no_dup['Chip'].value_counts()[ofertas_no_dup['Chip'].value_counts()>1]
```

```
[71]: AAA0158MTMR    4
      AAA0227PPSY    4
      AAA0083CHNN    4
      AAA0004ALEP    3
      AAA0074JZOM    3
      ..
      AAA0150JEJZ    2
      AAA0017EDLF    2
```

```

AAA0062REKC      2
AAA0095BTUZ      2
AAA0040JTYX      2
Name: Chip, Length: 555, dtype: int64

```

```
[72]: ofertas_no_dup[ofertas_no_dup['Chip'] == 'AAA0083CHNN']
```

```

[72]:      gid      lotcodigo lotdispers lotildispe  lotupredia manzcodigo \
13760  610429.0  007102011004          N      None          1.0  007102011
13761  610429.0  007102011004          N      None          1.0  007102011
13762  610429.0  007102011004          N      None          1.0  007102011
13763  610429.0  007102011004          N      None          1.0  007102011

      lotdistrit  Nombre_Barrio  Codigo_Barrio  Codigo_Manzana  Codigo_Predio \
13760          0.0  SANTA TERESITA          007102          011          004
13761          0.0  SANTA TERESITA          007102          011          004
13762          0.0  SANTA TERESITA          007102          011          004
13763          0.0  SANTA TERESITA          007102          011          004

      Codigo_Construccion  Codigo_Resto      Fecha_De_Cargue  Direccion \
13760                  000          00000  2020-03-07T00:00:00Z  CL 43 15 54
13761                  000          00000  2020-03-07T00:00:00Z  CL 43 15 54
13762                  000          00000  2020-03-07T00:00:00Z  CL 43 15 54
13763                  000          00000  2020-03-07T00:00:00Z  CL 43 15 54

      Codigo_Oferta      Tipo_Oferta  Tipo_Inmueble  Fuente      Chip \
13760          512630  ARRIENDO PARCIAL      Edificio  CAMPO  AAA0083CHNN
13761          512632  ARRIENDO PARCIAL      Edificio  CAMPO  AAA0083CHNN
13762          512634  ARRIENDO PARCIAL      Edificio  CAMPO  AAA0083CHNN
13763          512635  ARRIENDO PARCIAL      Edificio  CAMPO  AAA0083CHNN

      Area_Terreno  Area_Construida  Vr_Inicial_Venta  Vr_Final_Venta \
13760          194.2          634.15          0.0          0.0
13761          194.2          634.15          0.0          0.0
13762          194.2          634.15          0.0          0.0
13763          194.2          634.15          0.0          0.0

      Vr_Inicial_Arriendo  Vr_Final_Arriendo      X      Y
13760          1300000.0          1276000.0  1.000863e+06  1.003890e+06
13761          1500000.0          1472000.0  1.000863e+06  1.003890e+06
13762          2800000.0          2749000.0  1.000863e+06  1.003890e+06
13763          1000000.0          982000.0  1.000863e+06  1.003890e+06

```

```

[88]: ofertas_no_dup = ofertas_no_dup[(ofertas_no_dup.Vr_Inicial_Venta +
↳ofertas_no_dup.Vr_Final_Venta + ofertas_no_dup.Vr_Inicial_Arriendo+
↳ofertas_no_dup.Vr_Final_Arriendo ) != 0]

```

```
[89]: ofertas_no_dup['Chip'].value_counts()[ofertas_no_dup['Chip'].value_counts()>1]
```

```
[89]: AAA0083CHNN      4
      AAA0158MTMR      4
      AAA0150JXDM      3
      AAA0244OSOE      3
      AAA0048MCXR      3
      ..
      AAA01630YPP      2
      AAA0001YFKL      2
      AAA0060ATAW      2
      AAA0009AAFZ      2
      AAA0186AUBS      2
      Name: Chip, Length: 414, dtype: int64
```

```
[90]: ofertas_no_dup[ofertas_no_dup['Chip'] == 'AAA0083CHNN']
```

```
[90]:      gid      lotcodigo lotdispers lotildispe  lotupredia manzcodigo \
13760  610429.0  007102011004      N      None      1.0  007102011
13761  610429.0  007102011004      N      None      1.0  007102011
13762  610429.0  007102011004      N      None      1.0  007102011
13763  610429.0  007102011004      N      None      1.0  007102011

      lotdistrit  Nombre_Barrio  Codigo_Barrio  Codigo_Manzana  Codigo_Predio \
13760      0.0  SANTA TERESITA      007102      011      004
13761      0.0  SANTA TERESITA      007102      011      004
13762      0.0  SANTA TERESITA      007102      011      004
13763      0.0  SANTA TERESITA      007102      011      004

      Codigo_Construccion  Codigo_Resto      Fecha_De_Cargue  Direccion \
13760      000      00000  2020-03-07T00:00:00Z  CL 43 15 54
13761      000      00000  2020-03-07T00:00:00Z  CL 43 15 54
13762      000      00000  2020-03-07T00:00:00Z  CL 43 15 54
13763      000      00000  2020-03-07T00:00:00Z  CL 43 15 54

      Codigo_Oferta      Tipo_Oferta  Tipo_Inmueble  Fuente      Chip \
13760      512630  ARRIENDO PARCIAL      Edificio  CAMPO  AAA0083CHNN
13761      512632  ARRIENDO PARCIAL      Edificio  CAMPO  AAA0083CHNN
13762      512634  ARRIENDO PARCIAL      Edificio  CAMPO  AAA0083CHNN
13763      512635  ARRIENDO PARCIAL      Edificio  CAMPO  AAA0083CHNN

      Area_Terreno  Area_Construida  Vr_Inicial_Venta  Vr_Final_Venta \
13760      194.2      634.15      0.0      0.0
13761      194.2      634.15      0.0      0.0
13762      194.2      634.15      0.0      0.0
13763      194.2      634.15      0.0      0.0
```

	Vr_Inicial_Arriendo	Vr_Final_Arriendo	X	Y
13760	1300000.0	1276000.0	1.000863e+06	1.003890e+06
13761	1500000.0	1472000.0	1.000863e+06	1.003890e+06
13762	2800000.0	2749000.0	1.000863e+06	1.003890e+06
13763	1000000.0	982000.0	1.000863e+06	1.003890e+06

```
[94]: ofertas_no_dup['temp'] = ofertas_no_dup.Vr_Inicial_Venta + ofertas_no_dup.
      ↪Vr_Final_Venta + ofertas_no_dup.Vr_Inicial_Arriendo+ ofertas_no_dup.
      ↪Vr_Final_Arriendo
```

/home/rcamilo/anaconda3/envs/geospatial/lib/python3.6/site-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
 """Entry point for launching an IPython kernel.

```
[100]: test_without_duplicates = ofertas_no_dup.sort_values('temp').
      ↪drop_duplicates(subset=['Chip'], keep='last')
```

```
[103]: test_without_duplicates[test_without_duplicates['Chip'] == 'AAA0083CHNN']
```

```
[103]:      gid      lotcodigo lotdispers lotildispe  lotupredia manzcodigo \
13762  610429.0  007102011004          N          None          1.0  007102011

      lotdistrit  Nombre_Barrío Código_Barrío Código_Manzana Código_Predio \
13762          0.0  SANTA TERESITA          007102          011          004

      Código_Construcción Código_Resto      Fecha_De_Cargue  Dirección \
13762          000          00000  2020-03-07T00:00:00Z  CL 43 15 54

      Código_Oferta      Tipo_Oferta Tipo_Inmueble Fuente      Chip \
13762          512634  ARRIENDO PARCIAL      Edificio  CAMPO  AAA0083CHNN

      Área_Terreno  Área_Construida  Vr_Inicial_Venta  Vr_Final_Venta \
13762          194.2          634.15          0.0          0.0

      Vr_Inicial_Arriendo  Vr_Final_Arriendo          X          Y \
13762          2800000.0          2749000.0  1.000863e+06  1.003890e+06

      temp
13762  5549000.0
```

```
[104]: test_without_duplicates[test_without_duplicates['Chip'] == 'AAA0164ORAF']
```

```

[104]:      gid      lotcodigo lotdispers lotildispe lotupredia manzcodigo \
6582  595956.0  008501050012      N      None      524.0  008501050

      lotdistrit Nombre_Barrio Codigo_Barrio Codigo_Manzana Codigo_Predio \
6582      0.0      EL TOBERIN      008501      050      012

      Codigo_Construccion Codigo_Resto      Fecha_De_Cargue \
6582      001      01001  2019-12-16T00:00:00Z

      Direccion Codigo_Oferta      Tipo_Oferta Tipo_Inmueble \
6582  CL 169 16C 10 BQ 1 AP 101      506602  VENTA, ARRIENDO  Apartamento

      Fuente      Chip  Area_Terreno  Area_Construida  Vr_Inicial_Venta \
6582  CAMPO  AAA01640RAF      23.69      57.32      220000000.0

      Vr_Final_Venta  Vr_Inicial_Arriendo  Vr_Final_Arriendo      X \
6582      220000000.0      900000.0      900000.0  1.004362e+06

      Y      temp
6582  1.016775e+06  441800000.0

```

4.1.1 There are duplicated in the data (pending for quest to entity)