2021-I_mcpp_taller_7_Valentina_Cuenca

April 16, 2021

1 Taller 7

Métodos Computacionales para Políticas Públicas - URosario

Entrega: viernes 16-abr-2021 11:59 PM

Valentina Cuenca

norma.cuenca@urosario.edu.co

1.1 Instrucciones:

- Guarde una copia de este *Jupyter Notebook* en su computador, idealmente en una carpeta destinada al material del curso.
- Modifique el nombre del archivo del *notebook*, agregando al final un guión inferior y su nombre y apellido, separados estos últimos por otro guión inferior. Por ejemplo, mi *notebook* se llamaría: mcpp_taller7_santiago_matallana
- Marque el *notebook* con su nombre y e-mail en el bloque verde arriba. Reemplace el texto "[Su nombre acá]" con su nombre y apellido. Similar para su e-mail.
- Desarrolle la totalidad del taller sobre este *notebook*, insertando las celdas que sea necesario debajo de cada pregunta. Haga buen uso de las celdas para código y de las celdas tipo *markdown* según el caso.
- Recuerde salvar periódicamente sus avances.
- Cuando termine el taller:
  1. Descárguelo en PDF. Si tiene algún problema con la conversión, descárguelo en HTML.
  2. Suba todos los archivos a su repositorio en GitHub, en una carpeta destinada exclusivamente para este taller, antes de la fecha y hora límites.

(Todos los ejercicios tienen el mismo valor.)

En este taller exploraremos los datos de crimen de Chicago.

Descargue los datos de crimen del Chicago Data Portal solo para el año 2015 (https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2/data).

1.1.1 1.

Calcule el número de crímenes en cada Community Area en 2015. Haga un gráfico de barras que lo ilustre.

[68]:
```python
import pandas as pd
import matplotlib.pylab as plt
import numpy as np

crimes = pd.read_csv('Crimes_2015.csv')
crimes
```

[68]:
```
              ID Case Number                    Date                Block  \
0       10224738    HY411648  09/05/2015 01:30:00 PM       043XX S WOOD ST
1       10224739    HY411615  09/04/2015 11:30:00 AM   008XX N CENTRAL AVE
2       10224740    HY411595  09/05/2015 12:45:00 PM     035XX W BARRY AVE
3       10224741    HY411610  09/05/2015 01:00:00 PM   0000X N LARAMIE AVE
4       10224742    HY411435  09/05/2015 10:55:00 AM   082XX S LOOMIS BLVD
…            …           …                     …                   …
264510  12328217    JE184320  07/19/2015 01:29:00 PM       003XX W ERIE ST
264511  12328224    JE184321  03/18/2015 08:15:00 PM       003XX W ERIE ST
264512  12329490    JE186214  09/23/2015 12:00:00 PM  044XX S MICHIGAN AVE
264513  10337750    HY528847  12/07/2015 01:45:00 PM     068XX S TRIPP AVE
264514  12335256    JE193406  05/26/2015 12:00:00 AM   037XX W GIDDINGS ST

        IUCR                Primary Type  \
0       0486                     BATTERY
1       0870                       THEFT
2       2023                   NARCOTICS
3       0560                     ASSAULT
4       0610                    BURGLARY
…        …                         …
264510  1153          DECEPTIVE PRACTICE
264511  1153          DECEPTIVE PRACTICE
264512  1153          DECEPTIVE PRACTICE
264513  0265     CRIMINAL SEXUAL ASSAULT
264514  1752  OFFENSE INVOLVING CHILDREN

                                      Description  \
0                          DOMESTIC BATTERY SIMPLE
1                                   POCKET-PICKING
2                             POSS: HEROIN(BRN/TAN)
3                                           SIMPLE
4                                   FORCIBLE ENTRY
…                                              …
264510           FINANCIAL IDENTITY THEFT OVER $ 300
264511           FINANCIAL IDENTITY THEFT OVER $ 300
264512           FINANCIAL IDENTITY THEFT OVER $ 300
264513                           AGGRAVATED - OTHER
264514  AGGRAVATED CRIMINAL SEXUAL ABUSE BY FAMILY MEMBER

       Location Description  Arrest  Domestic  …  Ward  Community Area  \
```

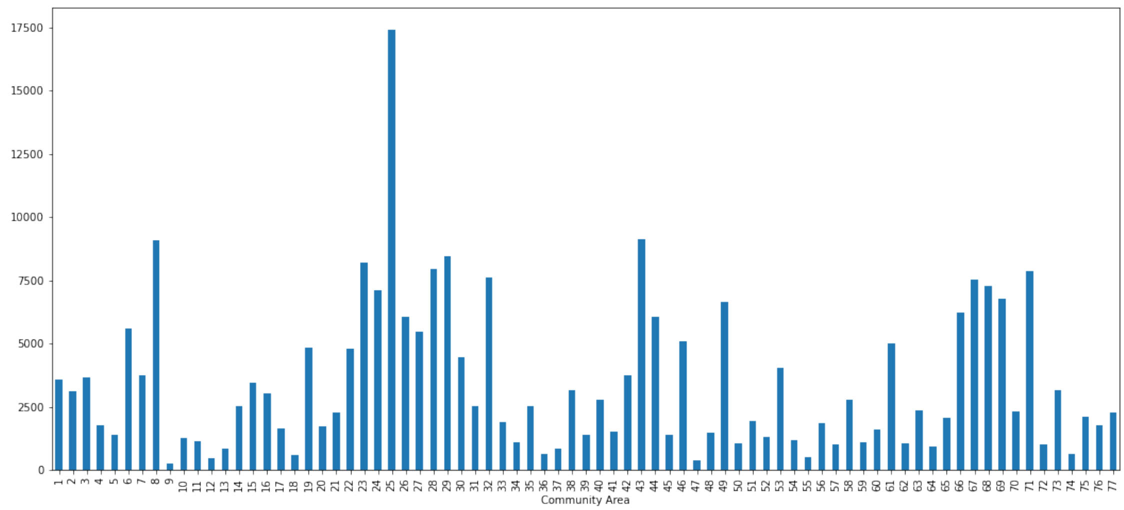
```
0                 RESIDENCE   False      True  …  12.0             61
1                   CTA BUS   False     False  …  29.0             25
2                  SIDEWALK    True     False  …  35.0             21
3                 APARTMENT   False      True  …  28.0             25
4                 RESIDENCE   False     False  …  21.0             71
…                       …      …        …   …   …              …
264510                  NaN   False     False  …  42.0              8
264511                  NaN   False     False  …  42.0              8
264512     DEPARTMENT STORE   False     False  …   3.0             38
264513            RESIDENCE   False     False  …  13.0             65
264514            RESIDENCE   False      True  …  35.0             14

        FBI Code  X Coordinate Y Coordinate  Year             Updated On  \
0            08B     1165074.0    1875917.0  2015  02/10/2018 03:50:01 PM
1             06     1138875.0    1904869.0  2015  02/10/2018 03:50:01 PM
2             18     1152037.0    1920384.0  2015  02/10/2018 03:50:01 PM
3            08A     1141706.0    1900086.0  2015  02/10/2018 03:50:01 PM
4             05     1168430.0    1850165.0  2015  02/10/2018 03:50:01 PM
…            …            …           …   …                    …
264510        11           NaN          NaN  2015  03/31/2021 04:59:47 PM
264511        11           NaN          NaN  2015  03/31/2021 04:59:47 PM
264512        11           NaN          NaN  2015  04/02/2021 05:04:22 PM
264513        02     1149218.0    1858851.0  2015  04/07/2021 05:10:28 PM
264514        17           NaN          NaN  2015  04/09/2021 05:06:35 PM

         Latitude  Longitude                      Location
0       41.815117 -87.670000  (41.815117282, -87.669999562)
1       41.895080 -87.765400  (41.895080471, -87.765400451)
2       41.937406 -87.716650  (41.937405765, -87.716649687)
3       41.881903 -87.755121  (41.881903443, -87.755121152)
4       41.744379 -87.658431  (41.744378879, -87.658430635)
…            …         …                            …
264510        NaN        NaN                            NaN
264511        NaN        NaN                            NaN
264512        NaN        NaN                            NaN
264513  41.768607 -87.728603  (41.768606685, -87.728602694)
264514        NaN        NaN                            NaN

[264515 rows x 22 columns]
```

[12]:
```
crimes_by_Comunity = crimes.groupby("Community Area")
crimes_by_Comunity['ID'].agg('count')
```

[12]:
```
Community Area
1     3594
2     3119
3     3666
```

```
4      1771
5      1395
       …
73     3177
74      619
75     2091
76     1752
77     2265
Name: ID, Length: 77, dtype: int64
```

[13]:
```
plt.rcParams["figure.figsize"] = [18.0, 8.0]
crimes_by_Comunity['ID'].agg('count').plot(kind='bar');
```



1.1.2 2.

Ordene las Community Areas de acuerdo con el número de crímenes. ¿Qué Community Area (por nombre, idealmente) presenta el mayor número de crímenes? ¿El menor?

[14]:
```
cimesC=crimes_by_Comunity['ID'].agg('count').to_frame()
cimesC.sort_values("ID")
```

[14]:

	ID
Community Area	
9	258
47	389
12	450
55	516
18	588
…	…
23	8205

```
29                8435
8                 9085
43                9121
25               17419

[77 rows x 1 columns]
```

[15]:
```
cimesC.sort_values("ID", ascending=False)
```

[15]:
```
                     ID
Community Area
25               17419
43                9121
8                 9085
29                8435
23                8205
…                …
18                 588
55                 516
12                 450
47                 389
9                  258

[77 rows x 1 columns]
```

El que presenta el mayor número de crimenes es el 25 y el menos es el 9.

1.1.3 3.

Cree una tabla cuyas filas sean días del año (yyyy-mm-dd) y las columnas las 77 Community Areas. En cada campo de la tabla deberá haber el correspondiente número de crímenes. Seleccione algunas Community Areas que le llamen la atención y haga un gráfico de serie de tiempo.

Pista: El siguiente código puede serle útil.

[16]:
```
import datetime as dt
crimes['Date'] = pd.to_datetime(crimes['Date']).dt.date
```

[17]:
```
crimes_by_date = crimes.groupby([ "Community Area",'Date'])
fecha=crimes_by_date['ID'].agg('count')
fecha
```

[17]:
```
Community Area  Date
1               2015-01-01    18
                2015-01-02     5
                2015-01-03     7
                2015-01-04    12
                2015-01-05     6
                              ..
```

```
77              2015-12-27     2
                2015-12-28     9
                2015-12-29     6
                2015-12-30     4
                2015-12-31     4
Name: ID, Length: 26913, dtype: int64
```

[18]:

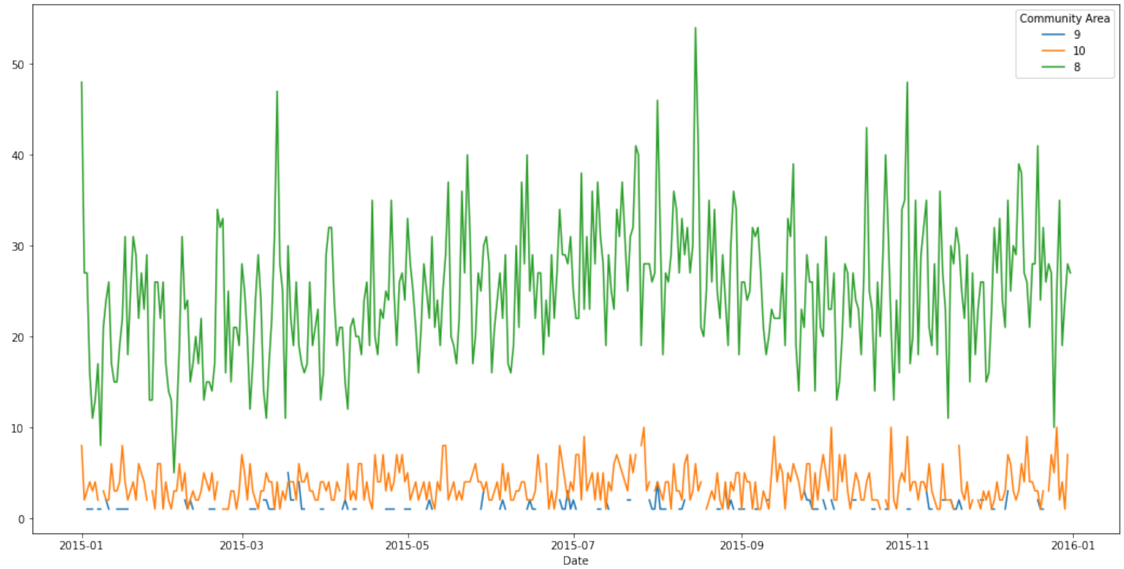
```python
community_date_timeseries = fecha.unstack('Community Area')
community_date_timeseries
```

[18]:

```
Community Area     1     2     3     4    5     6     7     8    9    10  …  \
Date                                                                     …
2015-01-01      18.0  14.0  15.0   6.0  5.0  25.0  15.0  48.0  1.0  8.0  …
2015-01-02       5.0   9.0   8.0   3.0  2.0  10.0   9.0  27.0  NaN  2.0  …
2015-01-03       7.0  11.0   9.0   7.0  4.0   6.0  11.0  27.0  1.0  3.0  …
2015-01-04      12.0   7.0   9.0  10.0  3.0  15.0   5.0  16.0  1.0  4.0  …
2015-01-05       6.0   7.0   5.0   4.0  5.0  15.0   7.0  11.0  1.0  3.0  …
…                …    …     …    …   …     …    …     …   …    …  …
2015-12-27      13.0   8.0   6.0   4.0  1.0  16.0  10.0  35.0  NaN  2.0  …
2015-12-28       7.0   8.0   6.0   2.0  2.0  14.0   9.0  19.0  NaN  4.0  …
2015-12-29       6.0   8.0  14.0   8.0  4.0  10.0   5.0  24.0  NaN  1.0  …
2015-12-30       5.0   9.0   8.0   4.0  1.0  12.0  17.0  28.0  1.0  7.0  …
2015-12-31       8.0   4.0   9.0   4.0  3.0  19.0   5.0  27.0  NaN  NaN  …

Community Area    68    69    70    71   72    73   74   75   76    77
Date
2015-01-01      33.0  31.0  10.0  53.0  3.0  10.0  2.0  9.0  7.0  11.0
2015-01-02      12.0  22.0   6.0  17.0  1.0  11.0  1.0  3.0  6.0   5.0
2015-01-03      23.0  12.0   8.0  18.0  NaN   8.0  1.0  7.0  3.0   3.0
2015-01-04      13.0  15.0   9.0  12.0  1.0   5.0  NaN  1.0  6.0   1.0
2015-01-05      16.0  12.0   8.0  17.0  NaN   5.0  2.0  2.0  7.0   5.0
…                …    …     …    …   …     …    …   …   …     …
2015-12-27      13.0  19.0   4.0  26.0  2.0   8.0  2.0  1.0  4.0   2.0
2015-12-28      12.0  23.0   9.0  14.0  2.0   6.0  2.0  2.0  3.0   9.0
2015-12-29      19.0  16.0   7.0  19.0  NaN   8.0  3.0  3.0  3.0   6.0
2015-12-30      11.0  23.0   6.0  14.0  2.0   9.0  1.0  7.0  4.0   4.0
2015-12-31      19.0  17.0   4.0  19.0  1.0   8.0  1.0  3.0  2.0   4.0

[365 rows x 77 columns]
```

[25]:

```python
plt.rcParams["figure.figsize"] = [18.0, 9.0]
community_date_timeseries[[9,10,8]].plot();
```



1.1.4 4.

Descargue la base de datos de información socioeconómica (https://data.cityofchicago.org/Health-Human-Services/Census-Data-Selected-socioeconomic-indicators-in-C/kn9c-c2s2).

Cree una tabla que agregue el número de crímenes por Community Area. Una esa tabla con la de datos socioeconómicos y cree un "scatter plot" de número de crímenes vs ingreso per cápita. Explique la relación en palabras.

[34]:
```python
soec = pd.read_csv("Socioeconomico.csv")
soec
```

[34]:
```
    Community Area Number COMMUNITY AREA NAME  PERCENT OF HOUSING CROWDED  \
0                     1.0         Rogers Park                         7.7
1                     2.0         West Ridge                          7.8
2                     3.0             Uptown                          3.8
3                     4.0       Lincoln Square                        3.4
4                     5.0        North Center                         0.3
..                    …                   …                           …
73                   74.0     Mount Greenwood                         1.0
74                   75.0         Morgan Park                         0.8
75                   76.0             O'Hare                          3.6
76                   77.0          Edgewater                          4.1
77                    NaN            CHICAGO                          4.7

    PERCENT HOUSEHOLDS BELOW POVERTY  PERCENT AGED 16+ UNEMPLOYED  \
0                               23.6                          8.7
1                               17.2                          8.8
2                               24.0                          8.9
```

```
3                                  10.9                          8.2
4                                   7.5                          5.2
..                                  …                           …
73                                  3.4                          8.7
74                                 13.2                         15.0
75                                 15.4                          7.1
76                                 18.2                          9.2
77                                 19.7                         12.9

    PERCENT AGED 25+ WITHOUT HIGH SCHOOL DIPLOMA  \
0                                           18.2
1                                           20.8
2                                           11.8
3                                           13.4
4                                            4.5
..                                           …
73                                           4.3
74                                          10.8
75                                          10.9
76                                           9.7
77                                          19.5

    PERCENT AGED UNDER 18 OR OVER 64  PER CAPITA INCOME   HARDSHIP INDEX
0                               27.5              23939             39.0
1                               38.5              23040             46.0
2                               22.2              35787             20.0
3                               25.5              37524             17.0
4                               26.2              57123              6.0
..                               …                  …              …
73                              36.8              34381             16.0
74                              40.3              27149             30.0
75                              30.3              25828             24.0
76                              23.8              33385             19.0
77                              33.5              28202              NaN

[78 rows x 9 columns]
```

[42]:
```
crimes_by_community = crimes.groupby('Community Area')
community_crime_count = crimes_by_community['ID'].agg('count')
community_crime_count=community_crime_count.to_frame()
community_crime_count = community_crime_count.reset_index()
community_crime_count
```

[42]:
```
    Community Area    ID
0                1  3594
1                2  3119
2                3  3666
```

```
3                4  1771
4                5  1395
..             …   …
72              73  3177
73              74   619
74              75  2091
75              76  1752
76              77  2265

[77 rows x 2 columns]
```

[52]:
```python
soec_crimenes=pd.merge(soec, community_crime_count, how='left',␣
 ↪left_index=True, right_index=True)
soec_crimenes=soec_crimenes.rename(columns={'ID':'Crimenes'})
```

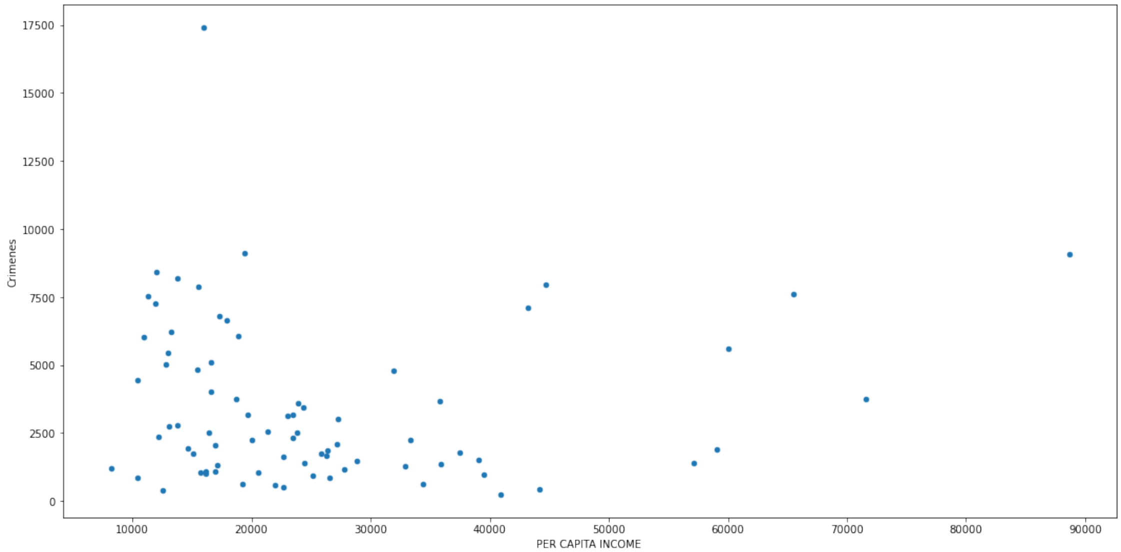
[54]:
```python
soec_crimenes
```

[54]:
```
    Community Area Number COMMUNITY AREA NAME  PERCENT OF HOUSING CROWDED  \
0                     1.0         Rogers Park                         7.7
1                     2.0         West Ridge                          7.8
2                     3.0             Uptown                          3.8
3                     4.0      Lincoln Square                         3.4
4                     5.0        North Center                         0.3
..                    …                 …                           …
73                   74.0      Mount Greenwood                        1.0
74                   75.0         Morgan Park                         0.8
75                   76.0             O'Hare                          3.6
76                   77.0          Edgewater                          4.1
77                    NaN            CHICAGO                          4.7

    PERCENT HOUSEHOLDS BELOW POVERTY  PERCENT AGED 16+ UNEMPLOYED  \
0                               23.6                          8.7
1                               17.2                          8.8
2                               24.0                          8.9
3                               10.9                          8.2
4                                7.5                          5.2
..                               …                           …
73                               3.4                          8.7
74                              13.2                         15.0
75                              15.4                          7.1
76                              18.2                          9.2
77                              19.7                         12.9

    PERCENT AGED 25+ WITHOUT HIGH SCHOOL DIPLOMA  \
0                                           18.2
1                                           20.8
2                                           11.8
```

```
3                                            13.4
4                                             4.5
..                                             …
73                                            4.3
74                                           10.8
75                                           10.9
76                                            9.7
77                                           19.5

    PERCENT AGED UNDER 18 OR OVER 64  PER CAPITA INCOME   HARDSHIP INDEX  \
0                               27.5              23939             39.0
1                               38.5              23040             46.0
2                               22.2              35787             20.0
3                               25.5              37524             17.0
4                               26.2              57123              6.0
..                               …                  …              …
73                              36.8              34381             16.0
74                              40.3              27149             30.0
75                              30.3              25828             24.0
76                              23.8              33385             19.0
77                              33.5              28202              NaN

    Community Area  Crimenes
0              1.0    3594.0
1              2.0    3119.0
2              3.0    3666.0
3              4.0    1771.0
4              5.0    1395.0
..             …       …
73            74.0     619.0
74            75.0    2091.0
75            76.0    1752.0
76            77.0    2265.0
77             NaN       NaN

[78 rows x 11 columns]
```

[69]:
```
crimenes_pib=soec_crimenes[["Community Area",'PER CAPITA INCOME ',"Crimenes"]]
```

[73]:
```
crimenes_pib.plot.scatter (x = 'PER CAPITA INCOME ', y = "Crimenes");
```



Podemos darnos cuenta que para las areas que tienen un alto PIB per-capita son pocas, pero tienen un crimenes considerables, y los que tienen un PIB per-capita mas bajo, no tienen crimenes tan altos, por lo que se puede deducir una relacion positiva no muy fuerte.