Analisis-Cassandra

April 24, 2018

0.1 Cassandra: python driver y consultas/gráficas

```
In [231]: %config IPCompleter.greedy=True
          %matplotlib inline
          from cassandra.cluster import Cluster
          import pandas as pd
          import numpy as np
          import matplotlib.pyplot as plt
```

0.1.1 Configuración de la sesión, conectores y funciones auxiliares

Para permitir realizar conexiones y consultas de forma genérica hemos implementado un conjunto de funciones auxiliares.

- La función *get_session* encapsula la conexión sobre nuestro "keyspace" generando el objeto Session característico del conector de cassandra.

- La función *q* es una abreviatura de *query* y permite acondicionar la realización de consultas y parametrizar valores sobre ellas. El resultado es un objeto tipo *data frame*, por lo cual estaría totalmente adaptado al entorno y sin necesidad de ser característico para el tipo de conector o base de datos.

```
session = get_session("espaciodatos")
q(session, "select * from incidents where incidentId = {id}", id = 10)
```

```
In [376]: def get_session(keyspace):
              """Obtiene el conector con la sesión actual al keyspace indicado."""
              return (Cluster(['127.0.0.1']).connect(keyspace))

          def q(session, query, **kwargs):
              """Función auxiliar para encapsular las queries producidas por cassandra en format
              return pd.DataFrame([row for row in session.execute(query.format(**kwargs))])

          conn = get_session("incidents")
```

Funciones auxiliares Las siguientes funciones auxiliares son necesarias para agilizar el tratamiento de las consultas sobre todo en la parte de parametrización.

```
In [345]: def nowOrdate(date=None):
              """Función auxiliar para permitir un valor por defecto en el timestamp de las quer
              si no se introduce un valor válido. Por defecto devuelve la fecha actual como time
              """
              return "dateof(now())" if (date is None) else "'" + date+" 00:00:00'"

          def eqOrInIntegers(obj=None):
              """Función que obtiene el operator = o in junto con los valores
              dependiendo si el objeto es una lista o no."""
              return None if obj is None else "in ("+",".join(str(x) for x in obj) + ")" if isin

          def eqOrIn(obj=None):
              """Función que obtiene el operator = o in junto con los valores
              dependiendo si el objeto es una lista o no."""
              return None if obj is None else "in ('"+"','".join(str(x) for x in obj) + "')" if
```

Testeando las funciones auxiliares

```
In [301]: [nowOrdate("2014-01-01"), nowOrdate()]
```

```
Out[301]: ["'2014-01-01 00:00:00'", 'dateof(now())']
```

```
In [306]: [eqOrIn("categoria1"), eqOrIn(["calle1","calle2"]) ,eqOrIn()]
```

```
Out[306]: ["='categoria1'", "in ('calle1','calle2')", None]
```

```
In [346]: [eqOrInIntegers(2015),eqOrInIntegers([2015,2016])]
```

```
Out[346]: ['=2015', 'in (2015,2016)']
```

0.2 Vistas

A continuación mostraremos todas las funciones necesarias para obtener cada una de las vistas de las consultas analizadas en la fase I:

-

0.2.1 Obtener incidencias (general)

- Obtener total de actividad criminal *(getIncidents)*
- Búsqueda de incidencia *(getIncident)*
- Búsqueda de incidencias en un rango/periodo de tiempo *(getIncidentsByDate)*

```
In [315]: def getIncidents(session, limit = 100):
              return q(session, "select * from incidents.overall limit {limit}",limit=limit)

          def getIncident(session, incidentId=None):
              return q(session, "select * from incidents.overall where incidentId={id}",id=incid
          def getIncidentsByDate(session, limit = 100, since="2014-01-01", to=None, year=2015):
              print(nowOrdate(since), nowOrdate(to))
              return q(session, "select * from incidents.overall " +
                                "where time >= {since} and time <= {to} and year = {year} allow
                               year=year,since=nowOrdate(since), to=nowOrdate(to), limit = limit
```

Ejemplo: Obtener la actividad criminal denominado con el identificador 140009459

```
In [232]: getIncident(conn, incidentId = 140009459)
```

```
Out[232]:    year                time  incidentid           subid  \
          0  2014 2014-01-04 03:52:00   140009459  14000945926030

                             address category  day dayoftheweek description  district  \
          0  SACRAMENTO ST / POLK ST    ARSON    4     Saturday       ARSON  NORTHERN

             hour                               location  month      resolution  \
          0     3  (37.7914943051906, -122.420874632415)      1  ARREST, BOOKED

                             x                  y
          0  -122.420874632415  37.7914943051906
```

Ejemplo: Obtener la actividad criminal en general (límite de 5 filas)

```
In [317]: getIncidents(conn, limit=5)
```

```
Out[317]:    year                time  incidentid           subid  \
          0  2014 2014-01-04 03:52:00   140009459  14000945926030
          1  2015 2015-02-04 20:27:00   150098577  15009857763010
          2  2015 2015-02-01 20:35:00   150098997  15009899716010
          3  2015 2015-02-01 20:35:00   150098997  15009899703414
          4  2015 2015-02-01 20:30:00   150098969  15009896903074

                             address       category  day dayoftheweek  \
          0  SACRAMENTO ST / POLK ST          ARSON    4     Saturday
          1   800 Block of BRYANT ST       WARRANTS    4    Wednesday
          2   700 Block of MARKET ST  DRUG/NARCOTIC    1       Sunday
          3   700 Block of MARKET ST        ROBBERY    1       Sunday
          4      PALOU AV / QUINT ST        ROBBERY    1       Sunday

                                                   description  district  hour  \
          0                                              ARSON  NORTHERN     3
          1                                     WARRANT ARREST  SOUTHERN    20
          2                            POSSESSION OF MARIJUANA  SOUTHERN    20
          3  ATTEMPTED ROBBERY ON THE STREET WITH BODILY FORCE  SOUTHERN    20
          4                              ROBBERY, BODILY FORCE   BAYVIEW    20

                                      location  month      resolution  \
0  (37.7914943051906, -122.420874632415)      1  ARREST, BOOKED
1   (37.775420706711, -122.403404791479)      2  ARREST, BOOKED
2  (37.7871160984672, -122.403919148357)      2            NONE
3  (37.7871160984672, -122.403919148357)      2            NONE
4  (37.7371564713337, -122.396516726913)      2            NONE

                  x                 y
0  -122.420874632415  37.7914943051906
1  -122.403404791479   37.775420706711
2  -122.403919148357  37.7871160984672
3  -122.403919148357  37.7871160984672
4  -122.396516726913  37.7371564713337
```

Ejemplo: Obtener la actividad criminal desde el 2 Enero del 2015 (límite de 5 filas)

```
In [316]: getIncidentsByDate(conn, since="2015-01-02", limit = 5)
```

```
'2015-01-02 00:00:00' dateof(now())
```

```
Out[316]:     year                time  incidentid           subid  \
          0   2015 2015-02-04 20:27:00   150098577  15009857763010
          1   2015 2015-02-01 20:35:00   150098997  15009899716010
          2   2015 2015-02-01 20:35:00   150098997  15009899703414
          3   2015 2015-02-01 20:30:00   150098969  15009896903074
          4   2015 2015-02-01 20:26:00   150098975  15009897564020
          5   2015 2015-02-01 20:15:00   150098953  15009895304134
          6   2015 2015-02-01 20:00:00   150098947  15009894728160
          7   2015 2015-02-01 20:00:00   150098947  15009894727170
          8   2015 2015-02-01 20:00:00   150098947  15009894726210
          9   2015 2015-02-01 20:00:00   150098947  15009894719090
          10  2015 2015-02-01 19:53:00   150098919  15009891965015
          11  2015 2015-02-01 19:53:00   150098919  15009891915030
          12  2015 2015-02-01 19:53:00   150098919  15009891907041
          13  2015 2015-02-01 19:50:00   150098890  15009889004014
          14  2015 2015-02-01 19:46:00   150098840  15009884065016
          15  2015 2015-02-01 19:35:00   150098925  15009892562050
          16  2015 2015-02-01 19:15:00   150098599  15009859965016
          17  2015 2015-02-01 19:06:00   150098765  15009876564085
          18  2015 2015-02-01 19:06:00   150098765  15009876504134
          19  2015 2015-02-01 19:00:00   150098981  15009898106303
          20  2015 2015-02-01 18:39:00   150098674  15009867465060
          21  2015 2015-02-01 18:39:00   150098674  15009867465016
          22  2015 2015-02-01 18:39:00   150098674  15009867465010
          23  2015 2015-02-01 18:38:00   150098630  15009863064010
          24  2015 2015-02-01 18:37:00   150098680  15009868064085
          25  2015 2015-02-01 18:23:00   150098624  15009862406243
          26  2015 2015-02-01 18:17:00   150098787  15009878764070
          27  2015 2015-02-01 18:17:00   150098787  15009878715200
```

```
28  2015 2015-02-01 18:09:00   150098602  15009860227130
29  2015 2015-02-01 18:09:00   150098602  15009860226170
..   ...                 ...         ...             ...
68  2015 2015-02-01 14:30:00   150098709  15009870906244
69  2015 2015-02-01 14:30:00   150098395  15009839506243
70  2015 2015-02-01 14:00:00   150098721  15009872172000
71  2015 2015-02-01 14:00:00   150098345  15009834563010
72  2015 2015-02-01 14:00:00   150098345  15009834516710
73  2015 2015-02-01 14:00:00   150098345  15009834516650
74  2015 2015-02-01 14:00:00   150098345  15009834506362
75  2015 2015-02-01 12:45:00   150098373  15009837372000
76  2015 2015-02-01 12:45:00   150098373  15009837307021
77  2015 2015-02-01 12:00:00   150098903  15009890306374
78  2015 2015-02-01 11:45:00   150098464  15009846407021
79  2015 2015-02-01 09:30:00   150098884  15009888406244
80  2015 2015-02-01 08:00:00   150098715  15009871507021
81  2015 2015-01-31 23:30:00   150098696  15009869668020
82  2015 2015-01-31 21:00:00   150098248  15009824828150
83  2015 2015-01-31 21:00:00   150098248  15009824815200
84  2015 2015-01-31 17:20:00   150098759  15009875906243
85  2015 2015-01-31 17:00:00   150098260  15009826006362
86  2015 2015-01-31 16:09:00   150098254  15009825405053
87  2015 2015-01-31 01:15:00   150098583  15009858371000
88  2015 2015-01-30 22:00:00   150098470  15009847006244
89  2015 2015-01-30 21:30:00   150098812  15009881206223
90  2015 2015-01-30 12:30:00   150098834  15009883405073
91  2015 2015-01-29 12:00:00   150098931  15009893174000
92  2015 2015-01-29 03:00:00   150098771  15009877103014
93  2015 2015-01-28 16:00:00   150098743  15009874307021
94  2015 2015-01-27 19:00:00   150098226  15009822628160
95  2015 2015-01-19 14:00:00   150060275  15006027571000
96  2015 2015-01-05 00:01:00   150098652  15009865228135
97  2015 2015-01-05 00:01:00   150098652  15009865215201

                                address                       category  day  \
0                800 Block of BRYANT ST                       WARRANTS    4
1                700 Block of MARKET ST                  DRUG/NARCOTIC    1
2                700 Block of MARKET ST                        ROBBERY    1
3                   PALOU AV / QUINT ST                        ROBBERY    1
4               1300 Block of REVERE AV                   NON-CRIMINAL    1
5           1000 Block of GOETTINGEN ST                        ASSAULT    1
6                 3400 Block of 16TH ST                      VANDALISM    1
7                 3400 Block of 16TH ST                 OTHER OFFENSES    1
8                 3400 Block of 16TH ST                        ASSAULT    1
9                 3400 Block of 16TH ST                    DRUNKENNESS    1
10       MCALLISTER ST / LEAVENWORTH ST                 OTHER OFFENSES    1
11       MCALLISTER ST / LEAVENWORTH ST                 OTHER OFFENSES    1
12       MCALLISTER ST / LEAVENWORTH ST              RECOVERED VEHICLE    1
```

```
13                100 Block of MASON ST                      ASSAULT    1
14                 TARAVAL ST / 32ND AV               OTHER OFFENSES    1
15               6300 Block of GEARY BL                     WARRANTS    1
16                  14TH ST / FOLSOM ST               OTHER OFFENSES    1
17               MISSION ST / DUBOCE AV               SUSPICIOUS OCC    1
18               MISSION ST / DUBOCE AV                      ASSAULT    1
19                2900 Block of 16TH ST                LARCENY/THEFT    1
20              CALIFORNIA ST / POLK ST  DRIVING UNDER THE INFLUENCE    1
21              CALIFORNIA ST / POLK ST               OTHER OFFENSES    1
22              CALIFORNIA ST / POLK ST               OTHER OFFENSES    1
23              1000 Block of FOLSOM ST                 NON-CRIMINAL    1
24             HOWARD ST / LAFAYETTE ST               SUSPICIOUS OCC    1
25                  300 Block of 9TH ST                LARCENY/THEFT    1
26               BACON ST / HAMILTON ST               SUSPICIOUS OCC    1
27               BACON ST / HAMILTON ST              SECONDARY CODES    1
28              FRANKLIN ST / SUTTER ST               OTHER OFFENSES    1
29              FRANKLIN ST / SUTTER ST               OTHER OFFENSES    1
..                                  ...                          ...  ...
68                1300 Block of 16TH AV                LARCENY/THEFT    1
69  MARTIN LUTHER KING JR DR / KEZAR DR                LARCENY/THEFT    1
70                 41ST AV / LINCOLN WY                 NON-CRIMINAL    1
71            1700 Block of HARRISON ST                     WARRANTS    1
72            1700 Block of HARRISON ST                DRUG/NARCOTIC    1
73            1700 Block of HARRISON ST                DRUG/NARCOTIC    1
74            1700 Block of HARRISON ST                LARCENY/THEFT    1
75                FRANKLIN ST / TURK ST                 NON-CRIMINAL    1
76                FRANKLIN ST / TURK ST                VEHICLE THEFT    1
77                2900 Block of 24TH ST                LARCENY/THEFT    1
78                GROVE ST / VANNESS AV                VEHICLE THEFT    1
79               2400 Block of JONES ST                LARCENY/THEFT    1
80          GOLDEN GATE AV / STANYAN ST                VEHICLE THEFT    1
81               600 Block of SILVER AV               OTHER OFFENSES   31
82             700 Block of KIRKWOOD AV                    VANDALISM   31
83             700 Block of KIRKWOOD AV              SECONDARY CODES   31
84                 400 Block of 10TH ST                LARCENY/THEFT   31
85                800 Block of GEARY ST                LARCENY/THEFT   31
86             200 Block of STOCKTON ST                     BURGLARY   31
87           800 Block of MONTGOMERY ST                 NON-CRIMINAL   31
88              0 Block of CRESTLINE DR                LARCENY/THEFT   30
89                     6TH AV / ANZA ST                LARCENY/THEFT   30
90                2600 Block of 18TH ST                     BURGLARY   30
91                 600 Block of CLAY ST               MISSING PERSON   29
92                 16TH ST / MISSION ST                      ROBBERY   29
93                 LINCOLN WY / 38TH AV                VEHICLE THEFT   28
94               LOMBARD ST / LAGUNA ST                    VANDALISM   27
95                18TH ST / VALENCIA ST                 NON-CRIMINAL   19
96               700 Block of POWELL ST               OTHER OFFENSES    5
97               700 Block of POWELL ST                      ASSAULT    5
```

```
    dayoftheweek                                        description  \
0      Wednesday                                     WARRANT ARREST   
1         Sunday                            POSSESSION OF MARIJUANA   
2         Sunday  ATTEMPTED ROBBERY ON THE STREET WITH BODILY FORCE   
3         Sunday                              ROBBERY, BODILY FORCE   
4         Sunday                       AIDED CASE, MENTAL DISTURBED   
5         Sunday                                            BATTERY   
6         Sunday          MALICIOUS MISCHIEF, VANDALISM OF VEHICLES   
7         Sunday                                   RESISTING ARREST   
8         Sunday              THREATENING SCHOOL OR PUBLIC EMPLOYEE   
9         Sunday       UNDER INFLUENCE OF ALCOHOL IN A PUBLIC PLACE   
10        Sunday                                  TRAFFIC VIOLATION   
11        Sunday           CONTRIBUTING TO THE DELINQUENCY OF MINOR   
12        Sunday                           VEHICLE, RECOVERED, AUTO   
13        Sunday               AGGRAVATED ASSAULT WITH BODILY FORCE   
14        Sunday              DRIVERS LICENSE, SUSPENDED OR REVOKED   
15        Sunday                    ENROUTE TO OUTSIDE JURISDICTION   
16        Sunday              DRIVERS LICENSE, SUSPENDED OR REVOKED   
17        Sunday                            INVESTIGATIVE DETENTION   
18        Sunday                                            BATTERY   
19        Sunday                        PETTY THEFT FROM A BUILDING   
20        Sunday         DRIVING WHILE UNDER THE INFLUENCE OF DRUGS   
21        Sunday              DRIVERS LICENSE, SUSPENDED OR REVOKED   
22        Sunday                           TRAFFIC VIOLATION ARREST   
23        Sunday                               AIDED CASE, DOG BITE   
24        Sunday                            INVESTIGATIVE DETENTION   
25        Sunday                       PETTY THEFT FROM LOCKED AUTO   
26        Sunday                              SUSPICIOUS OCCURRENCE   
27        Sunday                                  DOMESTIC VIOLENCE   
28        Sunday                       POSSESSION OF BURGLARY TOOLS   
29        Sunday                                PROBATION VIOLATION   
..           ...                                                ...   
68        Sunday                       GRAND THEFT FROM LOCKED AUTO   
69        Sunday                       PETTY THEFT FROM LOCKED AUTO   
70        Sunday                                     FOUND PROPERTY   
71        Sunday                                     WARRANT ARREST   
72        Sunday              POSSESSION OF NARCOTICS PARAPHERNALIA   
73        Sunday                     POSSESSION OF METH-AMPHETAMINE   
74        Sunday                            PETTY THEFT SHOPLIFTING   
75        Sunday                                     FOUND PROPERTY   
76        Sunday                                  STOLEN AUTOMOBILE   
77        Sunday                            GRAND THEFT OF PROPERTY   
78        Sunday                                  STOLEN AUTOMOBILE   
79        Sunday                       GRAND THEFT FROM LOCKED AUTO   
80        Sunday                                  STOLEN AUTOMOBILE   
81      Saturday                        MISCELLANEOUS INVESTIGATION   
82      Saturday                      MALICIOUS MISCHIEF, VANDALISM   
```

```
83     Saturday                                  DOMESTIC VIOLENCE
84     Saturday                       PETTY THEFT FROM LOCKED AUTO
85     Saturday                            PETTY THEFT SHOPLIFTING
86     Saturday                  BURGLARY OF STORE, UNLAWFUL ENTRY
87     Saturday                                      LOST PROPERTY
88       Friday                       GRAND THEFT FROM LOCKED AUTO
89       Friday                     PETTY THEFT FROM UNLOCKED AUTO
90       Friday                           BURGLARY, UNLAWFUL ENTRY
91     Thursday                                      MISSING ADULT
92     Thursday                   ROBBERY ON THE STREET, STRONGARM
93    Wednesday                                  STOLEN AUTOMOBILE
94      Tuesday          MALICIOUS MISCHIEF, VANDALISM OF VEHICLES
95       Monday                                      LOST PROPERTY
96       Monday                              HARASSING PHONE CALLS
97       Monday                                           STALKING

      district  hour                                 location  month  \
0     SOUTHERN    20   (37.775420706711, -122.403404791479)      2
1     SOUTHERN    20  (37.7871160984672, -122.403919148357)      2
2     SOUTHERN    20  (37.7871160984672, -122.403919148357)      2
3      BAYVIEW    20  (37.7371564713337, -122.396516726913)      2
4      BAYVIEW    20   (37.728979731984, -122.385545453301)      2
5    INGLESIDE    20  (37.7190699683467, -122.403249637246)      2
6      MISSION    20  (37.7643418581632, -122.430494795393)      2
7      MISSION    20  (37.7643418581632, -122.430494795393)      2
8      MISSION    20  (37.7643418581632, -122.430494795393)      2
9      MISSION    20  (37.7643418581632, -122.430494795393)      2
10  TENDERLOIN    19  (37.7809258336852, -122.413679376888)      2
11  TENDERLOIN    19  (37.7809258336852, -122.413679376888)      2
12  TENDERLOIN    19  (37.7809258336852, -122.413679376888)      2
13  TENDERLOIN    19  (37.7848337315521, -122.409336821846)      2
14     TARAVAL    19  (37.7424363113581, -122.489620429341)      2
15    RICHMOND    19  (37.7800894126441, -122.487429256289)      2
16     MISSION    19   (37.7685360123583, -122.41561633832)      2
17    SOUTHERN    19  (37.7701099298175, -122.420010175609)      2
18    SOUTHERN    19  (37.7701099298175, -122.420010175609)      2
19     MISSION    19   (37.7650244301204, -122.41920245941)      2
20    NORTHERN    18  (37.7905770710537, -122.420691680792)      2
21    NORTHERN    18  (37.7905770710537, -122.420691680792)      2
22    NORTHERN    18  (37.7905770710537, -122.420691680792)      2
23    SOUTHERN    18   (37.778254811523, -122.405834032593)      2
24    SOUTHERN    18  (37.7724556440219, -122.416305723264)      2
25    SOUTHERN    18   (37.773130623238, -122.410919016011)      2
26     BAYVIEW    18  (37.7260849566228, -122.409528258798)      2
27     BAYVIEW    18  (37.7260849566228, -122.409528258798)      2
28    NORTHERN    18  (37.7873381901532, -122.423404604063)      2
29    NORTHERN    18  (37.7873381901532, -122.423404604063)      2
..         ...   ...                                    ...    ...
```

```
68     TARAVAL    14   (37.7627511939927, -122.47391580547)       2
69        PARK    14  (37.7671999403456, -122.458638758608)       2
70     TARAVAL    14  (37.7644036412279, -122.500828300284)       2
71     MISSION    14  (37.7690748003847, -122.413354187018)       2
72     MISSION    14  (37.7690748003847, -122.413354187018)       2
73     MISSION    14  (37.7690748003847, -122.413354187018)       2
74     MISSION    14  (37.7690748003847, -122.413354187018)       2
75    NORTHERN    12   (37.7817528767881, -122.42227702764)       2
76    NORTHERN    12   (37.7817528767881, -122.42227702764)       2
77     MISSION    12   (37.752766726673, -122.410519781929)       2
78    NORTHERN    11  (37.7782510832412, -122.419884939457)       2
79     CENTRAL     9   (37.803594011382, -122.416489414289)       2
80    RICHMOND     8   (37.777023633257, -122.455125552189)       2
81   INGLESIDE    23  (37.7288459128524, -122.423207985156)       1
82     BAYVIEW    21   (37.729203356539, -122.374019331833)       1
83     BAYVIEW    21   (37.729203356539, -122.374019331833)       1
84    SOUTHERN    17  (37.7709130566165, -122.410541166987)       1
85     CENTRAL    17  (37.7862578545865, -122.417295322526)       1
86     CENTRAL    16   (37.7878092959561, -122.40656817787)       1
87     CENTRAL     1   (37.7969934198176, -122.40354422656)       1
88        PARK    22  (37.7501301863303, -122.446483988175)       1
89    RICHMOND    21  (37.7790877048755, -122.464147081977)       1
90     MISSION    12    (37.761791195706, -122.40867573409)       1
91     CENTRAL    12   (37.794611817992, -122.403994862674)       1
92     MISSION     3  (37.7650501214668, -122.419671780296)       1
93     TARAVAL    16   (37.7645433165084, -122.49761402653)       1
94    NORTHERN    19  (37.8004687042875, -122.431118543788)       1
95     MISSION    14   (37.7617007179518, -122.42158168137)       1
96     CENTRAL     0  (37.7915661006349, -122.409144009353)       1
97     CENTRAL     0  (37.7915661006349, -122.409144009353)       1

         resolution                  x                 y
0    ARREST, BOOKED  -122.403404791479   37.775420706711
1              NONE  -122.403919148357  37.7871160984672
2              NONE  -122.403919148357  37.7871160984672
3              NONE  -122.396516726913  37.7371564713337
4              NONE  -122.385545453301   37.728979731984
5    ARREST, BOOKED  -122.403249637246  37.7190699683467
6    ARREST, BOOKED  -122.430494795393  37.7643418581632
7    ARREST, BOOKED  -122.430494795393  37.7643418581632
8    ARREST, BOOKED  -122.430494795393  37.7643418581632
9    ARREST, BOOKED  -122.430494795393  37.7643418581632
10  JUVENILE BOOKED  -122.413679376888  37.7809258336852
11  JUVENILE BOOKED  -122.413679376888  37.7809258336852
12  JUVENILE BOOKED  -122.413679376888  37.7809258336852
13             NONE  -122.409336821846  37.7848337315521
14   ARREST, BOOKED  -122.489620429341  37.7424363113581
15   ARREST, BOOKED  -122.487429256289  37.7800894126441
```

```
16    ARREST, CITED   -122.41561633832  37.7685360123583
17             NONE  -122.420010175609  37.7701099298175
18             NONE  -122.420010175609  37.7701099298175
19             NONE   -122.41920245941  37.7650244301204
20   ARREST, BOOKED  -122.420691680792  37.7905770710537
21   ARREST, BOOKED  -122.420691680792  37.7905770710537
22   ARREST, BOOKED  -122.420691680792  37.7905770710537
23             NONE  -122.405834032593   37.778254811523
24             NONE  -122.416305723264  37.7724556440219
25             NONE  -122.410919016011   37.773130623238
26             NONE  -122.409528258798  37.7260849566228
27             NONE  -122.409528258798  37.7260849566228
28   ARREST, BOOKED  -122.423404604063  37.7873381901532
29   ARREST, BOOKED  -122.423404604063  37.7873381901532
..              ...                ...               ...
68             NONE   -122.47391580547  37.7627511939927
69             NONE  -122.458638758608  37.7671999403456
70             NONE  -122.500828300284  37.7644036412279
71   ARREST, BOOKED  -122.413354187018  37.7690748003847
72   ARREST, BOOKED  -122.413354187018  37.7690748003847
73   ARREST, BOOKED  -122.413354187018  37.7690748003847
74   ARREST, BOOKED  -122.413354187018  37.7690748003847
75             NONE   -122.42227702764  37.7817528767881
76             NONE   -122.42227702764  37.7817528767881
77             NONE  -122.410519781929   37.752766726673
78             NONE  -122.419884939457  37.7782510832412
79             NONE  -122.416489414289   37.803594011382
80        UNFOUNDED  -122.455125552189   37.777023633257
81             NONE  -122.423207985156  37.7288459128524
82             NONE  -122.374019331833   37.729203356539
83             NONE  -122.374019331833   37.729203356539
84             NONE  -122.410541166987  37.7709130566165
85             NONE  -122.417295322526  37.7862578545865
86             NONE   -122.40656817787  37.7878092959561
87             NONE   -122.40354422656  37.7969934198176
88             NONE  -122.446483988175  37.7501301863303
89             NONE  -122.464147081977  37.7790877048755
90             NONE   -122.40867573409   37.761791195706
91             NONE  -122.403994862674   37.794611817992
92             NONE  -122.419671780296  37.7650501214668
93             NONE   -122.49761402653  37.7645433165084
94             NONE  -122.431118543788  37.8004687042875
95             NONE   -122.42158168137  37.7617007179518
96             NONE  -122.409144009353  37.7915661006349
97             NONE  -122.409144009353  37.7915661006349

[98 rows x 16 columns]
```

-

0.2.2 Obtener actividad criminal por zona

- Obtener total de actividad criminal por distrito *(getCountByDistrict)*
- Obtener actividad criminal agrupado por distrito/año *(getByDistrict)*

```
In [355]: def getCountByDistrict(session, limit = 100, year=2015):
              return q(session, "select district, count(*) " +
                                "from incidents.bydistrict " +
                                "where year {eqyear} " +
                                "group by district limit {limit}; ", eqyear = eqOrInIntegers(yea

          def getByDistrict(session, limit = 100, year=2015, district=None):
              return q(session, "select district, year, incidentid, category, time, location " +
                                "from incidents.bydistrict where year={year} " +
                                (" " if district is None else "and district={district} ") +
                                "limit {limit}", year=year, limit=limit, district= district)
```

Ejemplos: Obtener numero de incidencias por distrito (en un año determinado)

```
In [341]: getCountByDistrict(conn, limit=10)

Out[341]:      district  count
          0        PARK      2
          1     MISSION     12
          2    NORTHERN      7
          3     TARAVAL      5
          4   INGLESIDE      2
          5  TENDERLOIN      7
          6     CENTRAL      9
          7    SOUTHERN      8
          8     BAYVIEW      9
          9    RICHMOND      6
```

```
In [358]: getCountByDistrict(conn, year=[2014], limit=10)

Out[358]:    district  count
          0  NORTHERN      1
```

Ejemplos: Obtener incidencias agrupadas por distrito (en un año determinado)

```
In [360]: getByDistrict(conn, limit=10)

Out[360]:   district  year  incidentid        category                 time  \
          0     PARK  2015   150098395   LARCENY/THEFT 2015-02-01 14:30:00
          1     PARK  2015   150098470   LARCENY/THEFT 2015-01-30 22:00:00
          2  MISSION  2015   150098947       VANDALISM 2015-02-01 20:00:00
          3  MISSION  2015   150098599  OTHER OFFENSES 2015-02-01 19:15:00
```

```
4  MISSION  2015   150098981   LARCENY/THEFT 2015-02-01 19:00:00
5  MISSION  2015   150098527     WEAPON LAWS 2015-02-01 17:02:00
6  MISSION  2015   150098458  OTHER OFFENSES 2015-02-01 16:56:00
7  MISSION  2015   150098367         ROBBERY 2015-02-01 16:20:00
8  MISSION  2015   150098856    NON-CRIMINAL 2015-02-01 15:30:00
9  MISSION  2015   150098345        WARRANTS 2015-02-01 14:00:00

                             location
0  (37.7671999403456, -122.458638758608)
1  (37.7501301863303, -122.446483988175)
2  (37.7643418581632, -122.430494795393)
3   (37.7685360123583, -122.41561633832)
4   (37.7650244301204, -122.41920245941)
5  (37.7666737551835, -122.419827929961)
6   (37.764228935718, -122.419520367886)
7  (37.7651107322703, -122.432198022433)
8  (37.7564864109309, -122.406539115148)
9  (37.7690748003847, -122.413354187018)
```

```
In [259]: getCountByDistrict(conn, limit=10)
```

```
Out[259]:      district  count
          0        PARK      2
          1     MISSION     12
          2    NORTHERN      7
          3     TARAVAL      5
          4   INGLESIDE      2
          5  TENDERLOIN      7
          6     CENTRAL      9
          7    SOUTHERN      8
          8     BAYVIEW      9
          9    RICHMOND      6
```

-

0.2.3 Obtener actividad criminal por tipo de delito

- Obtener total de actividad criminal por categoria *(getCountByCategory)*
- Obtener actividad criminal agrupado por category/año *(getRangeByCategory)* permitiendo consultar un rango de fechas.

```
In [403]: def getCountByCategory(session, limit = 100, year=2015):
              return q(session, "select category, year, count(*) " +
                                "from incidents.bycategory " +
                                "where year = {year} " +
                                "group by category limit {limit}; ", year = year, limit = limit)

          def getRangeByCategory(session, limit = 100, year=2015, since=None, to=None, category=
              return q(session, "select category, year, incidentid, category, time, location " +
                                "from incidents.bycategory " +
                                "where year = {year} and " +
                                ("" if category is None else "category {eqCategory} and ") +
                                "time >= {since} and time <= {to} limit {limit} allow filtering
                       year = year, limit = limit, since=nowOrdate(since), to=nowOrdate(to), eqC
```

Ejemplo: Obtener total de actividad criminal agrupado por cateogoria

```
In [404]: getCountByCategory(conn, limit = 10)
```

```
Out[404]:                       category  year  count
          0                  WEAPON LAWS  2015      2
          1                     BURGLARY  2015      2
          2                DRUG/NARCOTIC  2015      4
          3                  DRUNKENNESS  2015      1
          4               SUSPICIOUS OCC  2015      3
          5                LARCENY/THEFT  2015     14
          6                     WARRANTS  2015      4
          7              SECONDARY CODES  2015      3
          8  DRIVING UNDER THE INFLUENCE  2015      1
          9               MISSING PERSON  2015      2
```

Ejemplo: Obtener total de actividad criminal agrupado por categoria a partir del 1 Enero de 2015

```
In [208]: getRangeByCategory(conn, limit = 10, since= "2015-02-01")
```

```
Out[208]:         category  year  incidentid       category_                time  \
          0     WEAPON LAWS  2015   150098420     WEAPON LAWS 2015-02-01 17:10:00
          1     WEAPON LAWS  2015   150098527     WEAPON LAWS 2015-02-01 17:02:00
          2   DRUG/NARCOTIC  2015   150098997   DRUG/NARCOTIC 2015-02-01 20:35:00
          3   DRUG/NARCOTIC  2015   150098527   DRUG/NARCOTIC 2015-02-01 17:02:00
          4   DRUG/NARCOTIC  2015   150098458   DRUG/NARCOTIC 2015-02-01 16:56:00
          5   DRUG/NARCOTIC  2015   150098345   DRUG/NARCOTIC 2015-02-01 14:00:00
          6     DRUNKENNESS  2015   150098947     DRUNKENNESS 2015-02-01 20:00:00
          7  SUSPICIOUS OCC  2015   150098765  SUSPICIOUS OCC 2015-02-01 19:06:00
          8  SUSPICIOUS OCC  2015   150098680  SUSPICIOUS OCC 2015-02-01 18:37:00
          9  SUSPICIOUS OCC  2015   150098787  SUSPICIOUS OCC 2015-02-01 18:17:00

                                      location
          0   (37.784696907904, -122.413609328985)
          1  (37.7666737551835, -122.419827929961)
          2  (37.7871160984672, -122.403919148357)
          3  (37.7666737551835, -122.419827929961)
          4   (37.764228935718, -122.419520367886)
          5  (37.7690748003847, -122.413354187018)
          6  (37.7643418581632, -122.430494795393)
          7  (37.7701099298175, -122.420010175609)
```

```
8  (37.7724556440219, -122.416305723264)
9  (37.7260849566228, -122.409528258798)
```

```
In [206]: getRangeByCategory(conn, limit = 100, since= "2015-02-01", category=["WEAPON LAWS","DR
```

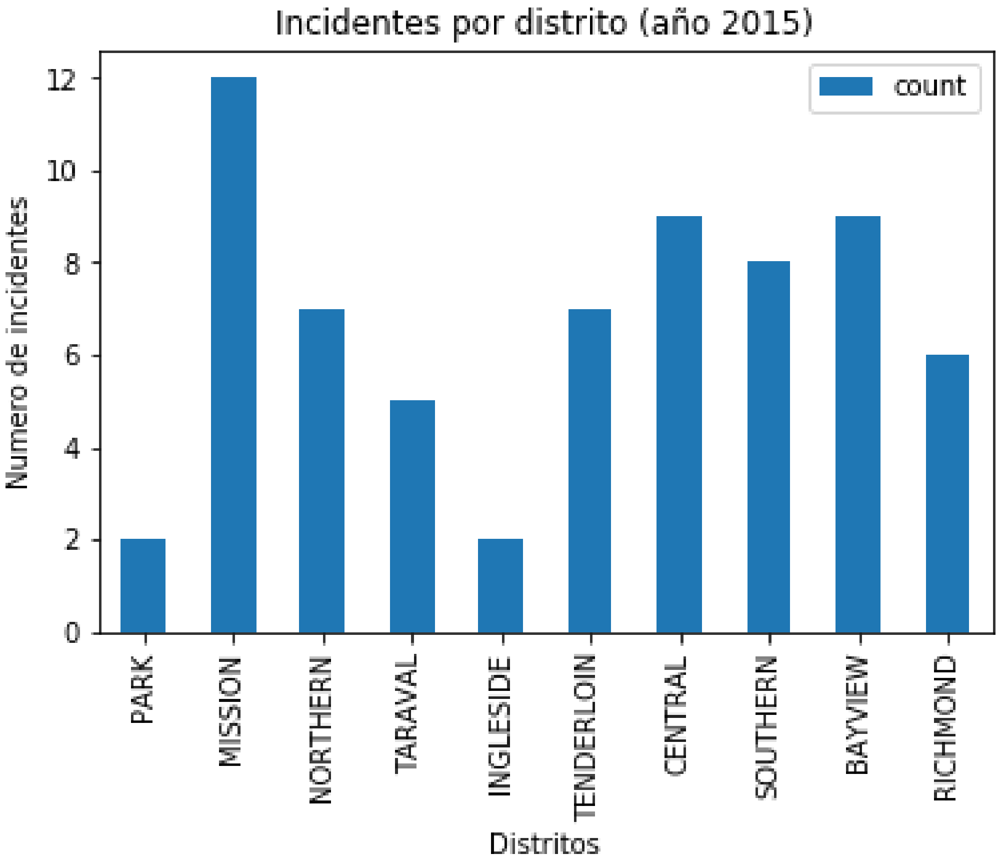
```
Out[206]:        category  year  incidentid    category_                 time  \
          0  DRUNKENNESS  2015   150098947  DRUNKENNESS 2015-02-01 20:00:00
          1  WEAPON LAWS  2015   150098420  WEAPON LAWS 2015-02-01 17:10:00
          2  WEAPON LAWS  2015   150098527  WEAPON LAWS 2015-02-01 17:02:00

                                     location
          0  (37.7643418581632, -122.430494795393)
          1   (37.784696907904, -122.413609328985)
          2  (37.7666737551835, -122.419827929961)
```

0.2.4 Representaciones gráficas

- Representación de la frecuencia de delitos por distritos para un año

```
In [419]: df = getCountByDistrict(conn, limit=10)
          df.set_index(['district'])
          graph = df.plot(kind='bar')
          graph.set_xlabel("Distritos")
          graph.set_ylabel("Numero de incidentes")
          graph.set_title('Incidentes por distrito (año 2015)')
          graph.set_xticklabels(df['district'])
```

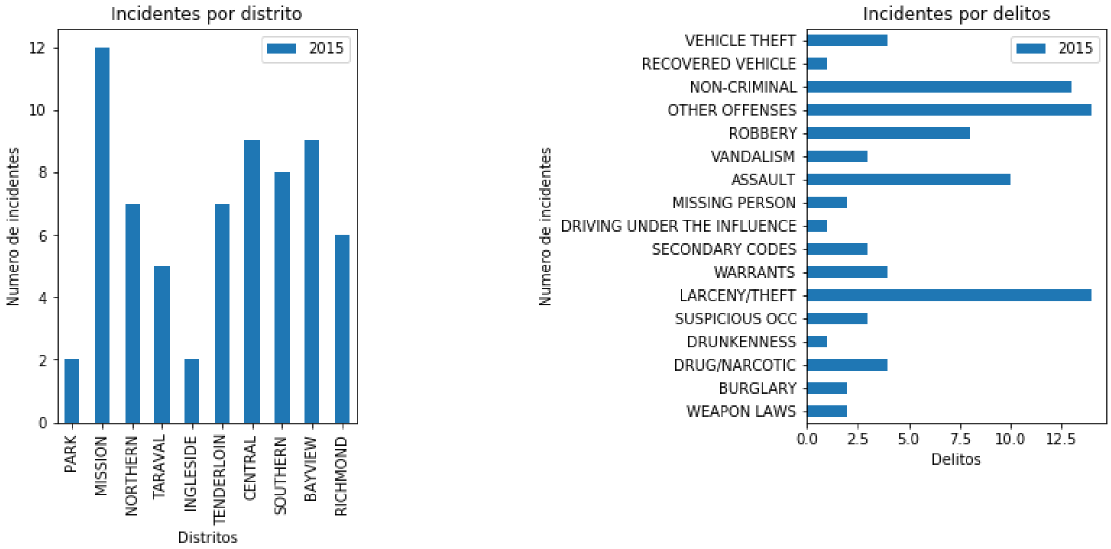
```
Out[419]: [Text(0,0,'PARK'),
           Text(0,0,'MISSION'),
           Text(0,0,'NORTHERN'),
           Text(0,0,'TARAVAL'),
           Text(0,0,'INGLESIDE'),
           Text(0,0,'TENDERLOIN'),
           Text(0,0,'CENTRAL'),
           Text(0,0,'SOUTHERN'),
           Text(0,0,'BAYVIEW'),
           Text(0,0,'RICHMOND')]
```



- Representación de la frecuencia de delitos por zona/tipo de delito

```
In [484]: from pylab import rcParams
          rcParams['figure.figsize'] = 13, 5
          fig, (ax1, ax2) = plt.subplots(1, 2)


          df = getCountByDistrict(conn, limit=10)
          graph1 = df.plot(kind='bar', ax=ax1)
          graph1.set_xlabel("Distritos")
          graph1.set_ylabel("Numero de incidentes")
          graph1.set_title('Incidentes por distrito')
          graph1.set_xticklabels(df['district'])
          ax1.legend(['2015','2016','2017'])

          df = getCountByCategory(conn)[['category','count']]
          graph2 = df.plot(kind='barh', ax=ax2)
          graph2.set_xlabel("Delitos")
          graph2.set_ylabel("Numero de incidentes")
          graph2.set_title('Incidentes por delitos')
```

```
graph2.set_yticklabels(df['category'])
ax2.legend(['2015','2016','2017'])

fig.subplots_adjust(wspace=1.5)
```



Cerramos la conexión con la base de datos

```
In [485]: conn.shutdown()
```

Álvaro López