Module 5

Use Case 1

LESSON 1

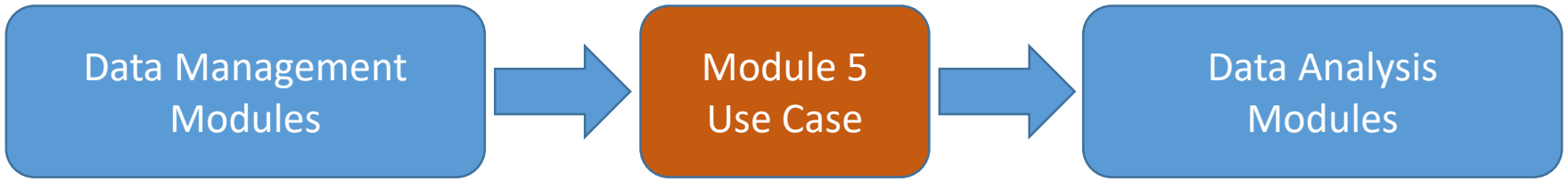
José Manuel García Nieto – University of Málaga

Table of contents

Introduction

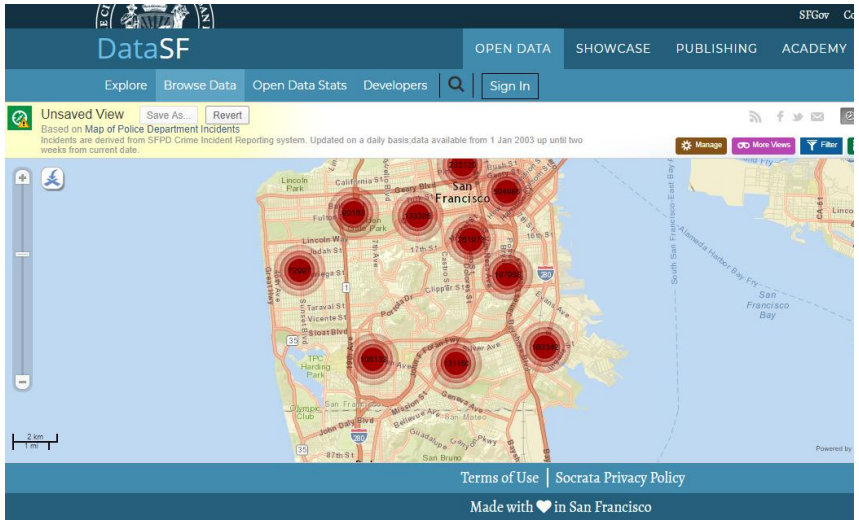
- At this point, our main *motivation* is to have a first experience by tackling with real-world problems and managing real-world data

- This is a *transition module*. We apply acquired knowledge in past modules, we use it in practice, and we promote new skills to be used in future modules



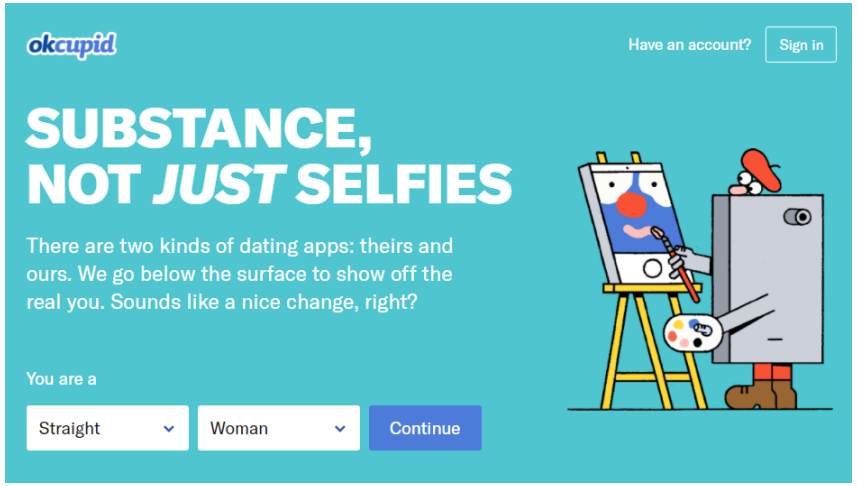
Introduction

- Use Case 1.1: San Francisco City Police Department (SFPD) Incidents
  - Incidents are derived from SFPD Crime Incident Reporting system
  - Updated on a daily basis data available from 1 Jan 2003 up until two weeks from current date
  - Of course: Open Data



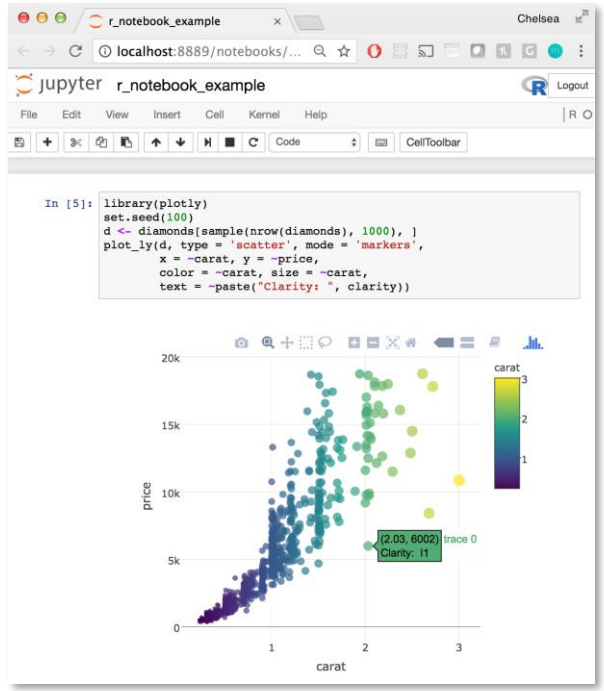
Introduction

- Use Case 1.2: OKCupid profile dataset
  - Analysis of a public dataset of almost 60,000 online dating profiles
  - Dataset published in the *Journal of Statistics Education, Volume 23, Number 2 (2015) by Albert Y. Kim et al.,*
  - Collection and distribution explicitly allowed by OkCupid president and co-founder Christian Rudder
  - Using these data is therefore ethically and legally acceptable. This is in contrast to another recent release of a different OkCupid profile dataset, which was collected without permission and without anonymizing the data (more on the ethical issues in this Wired article)



Work Environment

- Work style: Notebook
  - A Notebook is an open-source web application that allows you to create and share documents that contain *live code*, *equations, visualizations and narrative text*.
  - Use cases include: *data cleaning and transformation*, numerical simulation, *statistical modelling, data visualization*, machine learning, and much more
- Programming language: Python 3.6

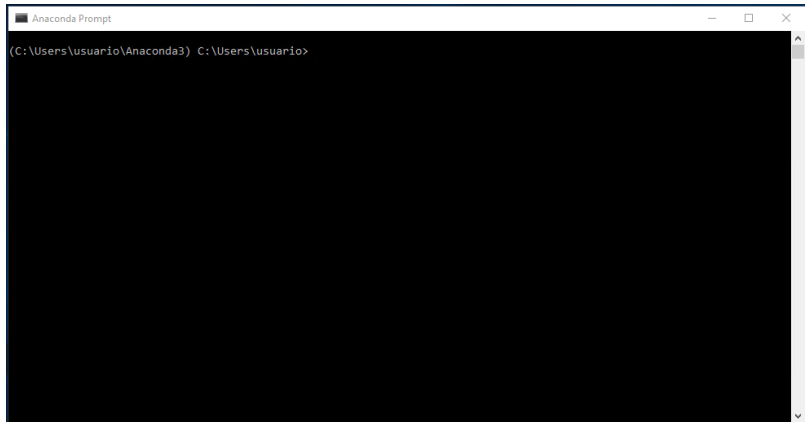


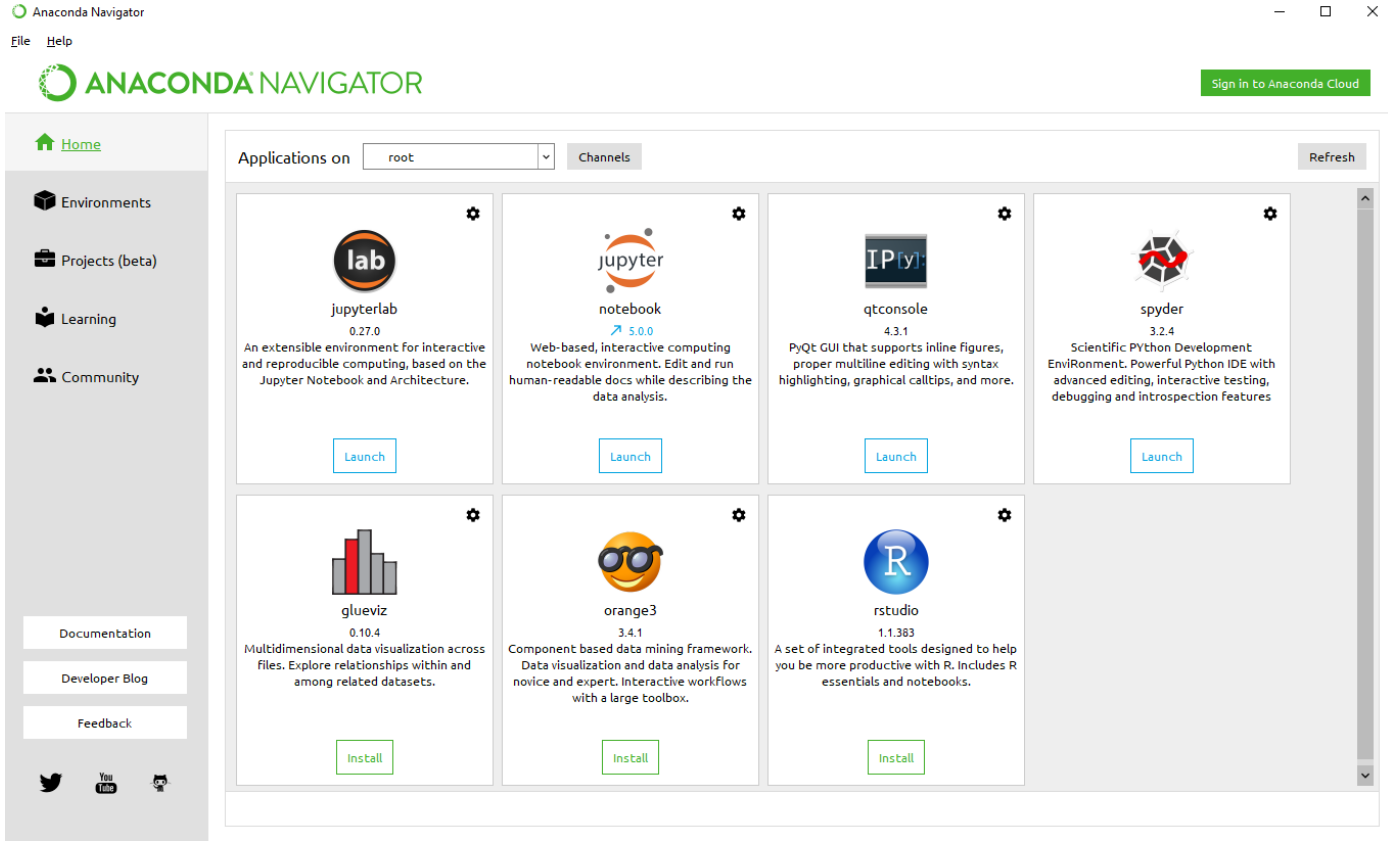
Work Environment

- We want to organize and visualize the narrative of data, but we also look for managing data in depth
- Working platform: Anaconda
  - Spyder: Scientific python development environment
  - Jupyter: web-based, interactive computing notebook
- Data integration: MongoBD

Work Environment

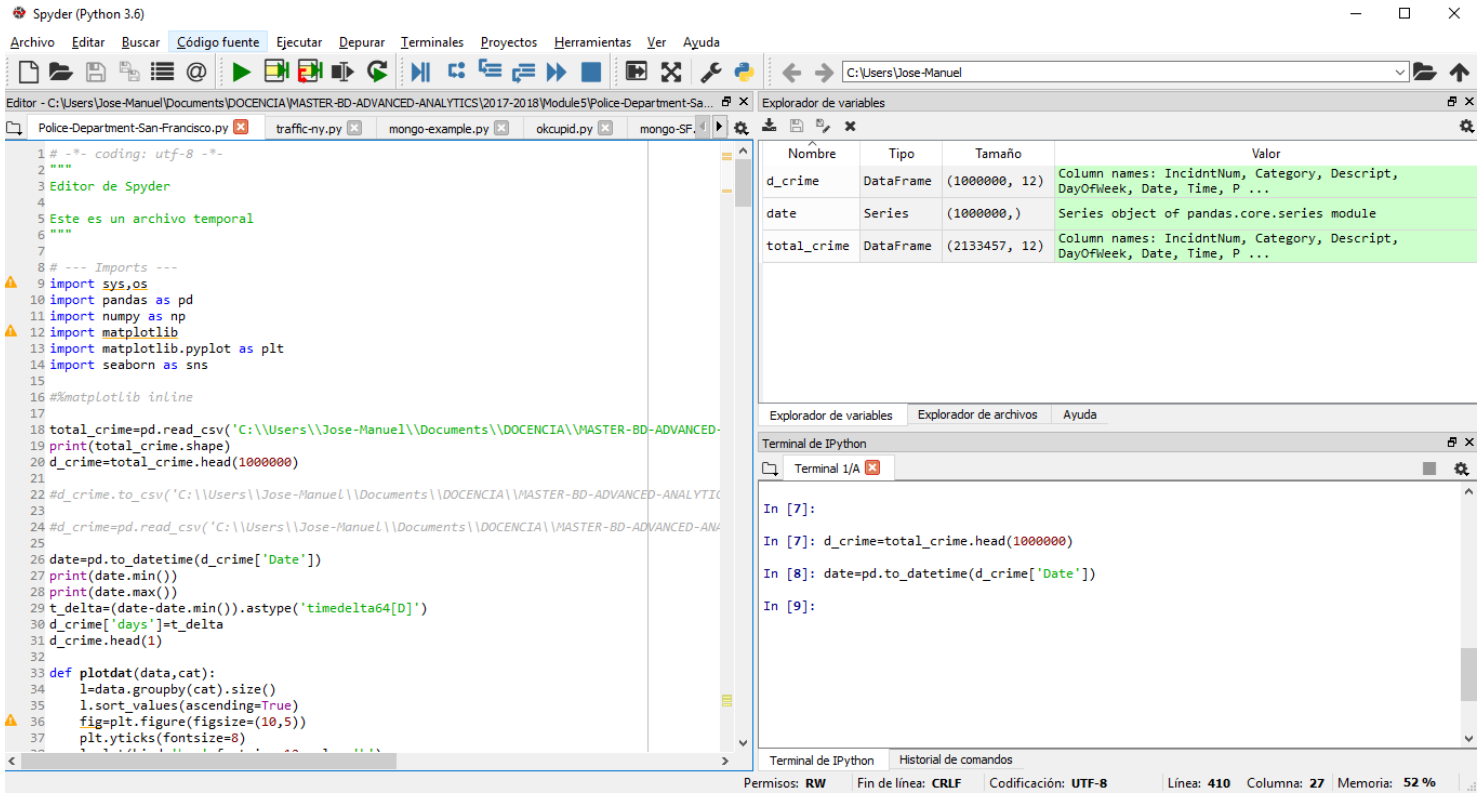
- Working platform: *Anaconda Navigator*
  - Projects
  - Cloud deployment
  - Environments
  - Terminal (*Anaconda Prompt*)





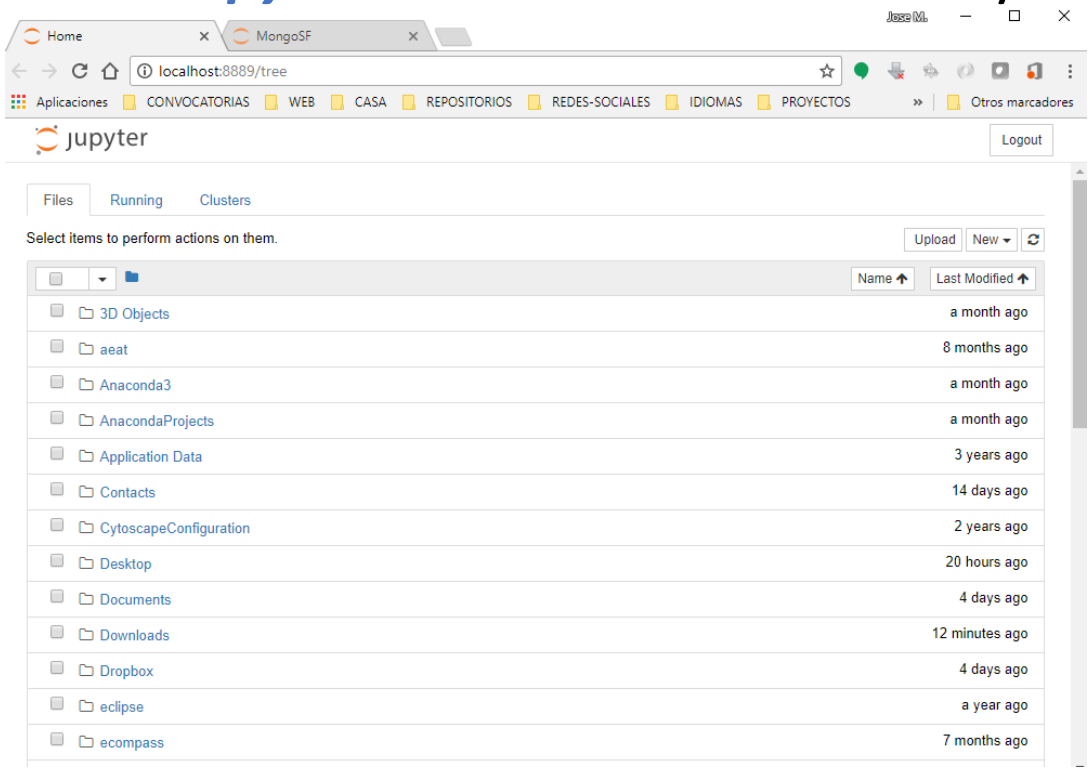
Work Environment

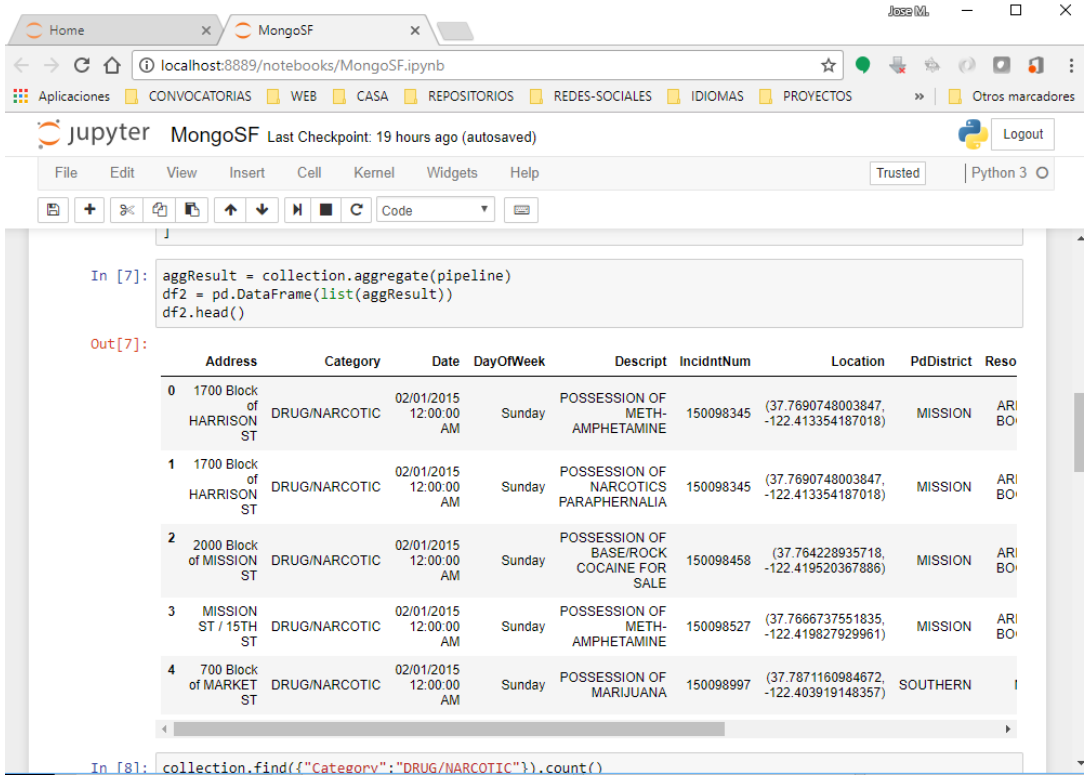
- Working platform: *Anaconda Navigator*
  - Spyder: Development IDE for Python



Work Environment

- Working platform: *Anaconda Navigator*
  - ***Jupyter***: Notebook for data story composition and publication





	Address	Category	Date	DayOfWeek	Descript	IncidntNum	Location	PdDistrict	Reso
0	1700 Block of HARRISON ST	DRUG/NARCOTIC	02/01/2015 12:00:00 AM	Sunday	POSSESSION OF METH-AMPHETAMINE	150098345	(37.7690748003847, -122.413354187018)	MISSION	ARI BO
1	1700 Block of HARRISON ST	DRUG/NARCOTIC	02/01/2015 12:00:00 AM	Sunday	POSSESSION OF NARCOTICS PARAPHERNALIA	150098345	(37.7690748003847, -122.413354187018)	MISSION	ARI BO
2	2000 Block of MISSION ST	DRUG/NARCOTIC	02/01/2015 12:00:00 AM	Sunday	POSSESSION OF BASE/ROCK COCAINE FOR SALE	150098458	(37.764228935718, -122.419520367886)	MISSION	ARI BO
3	MISSION ST / 15TH ST	DRUG/NARCOTIC	02/01/2015 12:00:00 AM	Sunday	POSSESSION OF METH-AMPHETAMINE	150098527	(37.7666737551835, -122.419827929961)	MISSION	ARI BO
4	700 Block of MARKET ST	DRUG/NARCOTIC	02/01/2015 12:00:00 AM	Sunday	POSSESSION OF MARIJUANA	150098997	(37.7871160984672, -122.403919148357)	SOUTHERN	

Work Environment

- Data Integration: *Mongo DB*
  - ***mongod.*** Available to connect, import, load, export, find, etc.
    - Required path
      - [Windows] C:\data\db
      - [Linux] /data/db

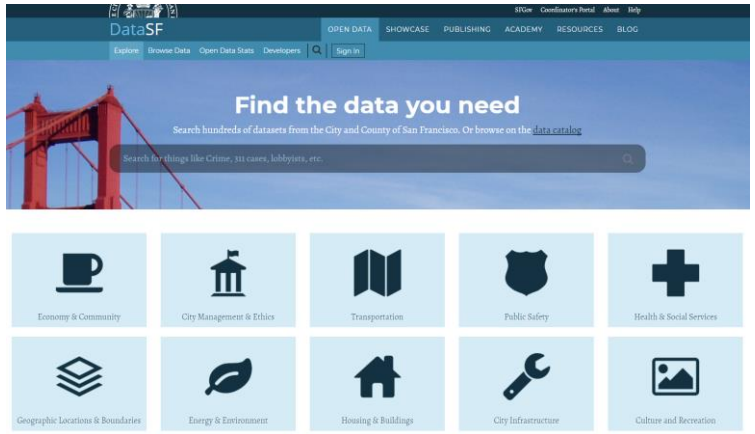
  - ***Pymongo.*** Python distribution containing tools for working with MongoDB, and is the recommended way to work with MongoDB from Python

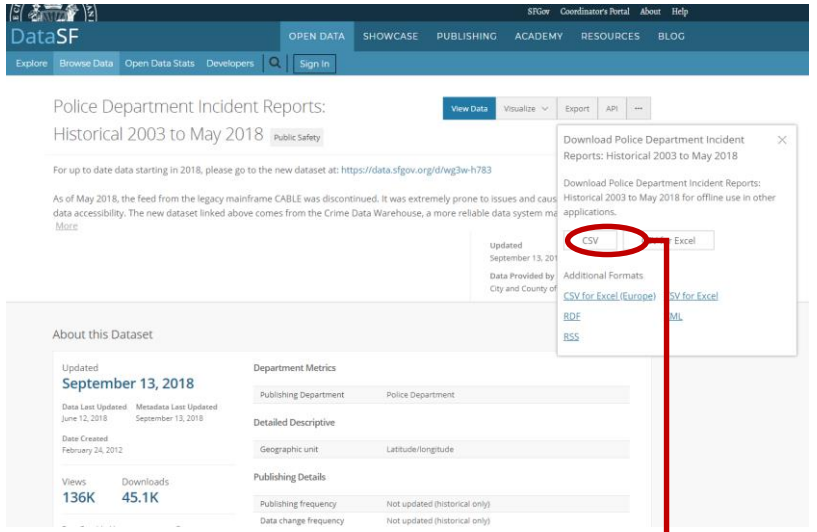
    - From Anaconda Prompt > **pip install pymongo**

https://api.mongodb.com/python/current/

Resources, Repositories and Sites

- Use Case 1.1: San Francisco City Police Department (SFPD) Incidents
  - Incidents are derived from SFPD Crime Incident Reporting system
  - Updated on a daily basis data available from 1 Jan 2003 up until two weeks from current date
  - Open Data: https://datasf.org/opendata/

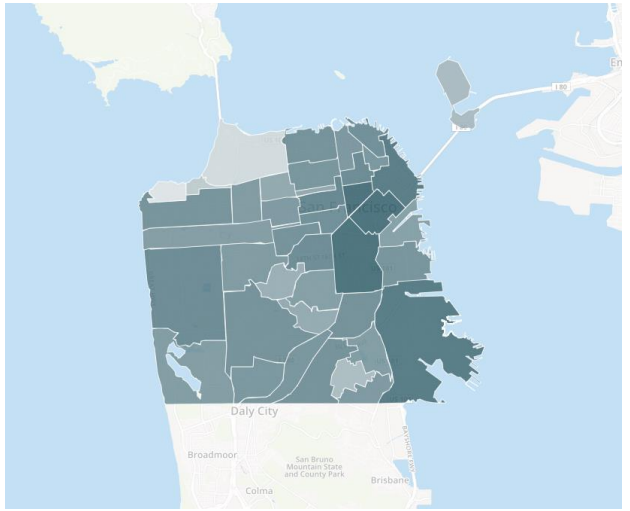




https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-Historical-2003/tmnf-yvry
https://data.sfgov.org/api/views/tmnf-yvry/rows.csv

Use Case Experimentation

- Use Case 1.1: SFPD Incidents
  - Dataset



Showing 2,148,658 out of 2,148,658 Incidents

IncidntNum	Category	Descript	DayOfWeek	Date	Time	PdDistrict	Resolution	Address	X	Y	Location
170905043	ARSON	ARSON	Wednesday	11/01/2017	09:11	MISSION	ARREST, BOOKED	800 Block of CAPP ST	-122.41749942997069	37.753084671409006	(37.753084671409°, -122.417499429971°)
170940425	ARSON	ATTEMPTED ARSON	Saturday	11/18/2017	04:00	NORTHERN	NONE	1500 Block of CALIFORNIA ST	-122.4198666939337	37.79062691003078	(37.790626910031°, -122.419866693934°)
170941928	ARSON	ATTEMPTED ARSON	Sunday	11/19/2017	15:15	MISSION	ARREST, BOOKED	2400 Block of MISSION ST	-122.41902526775176	37.75761346717424	(37.757613467174°, -122.419025267752°)
170905043	ARSON	ATTEMPTED ARSON	Wednesday	11/01/2017	09:11	MISSION	ARREST, BOOKED	800 Block of CAPP ST	-122.41749942997069	37.753084671409006	(37.753084671409°, -122.417499429971°)
170919721	ARSON	ARSON	Saturday	11/11/2017	06:43	TENDERLOIN	ARREST, BOOKED	100 Block of LEAVENWORTH ST	-122.41389315595086	37.782346520673045	(37.782346520673°, -122.413893155951°)
170925671	ARSON	ARSON	Monday	11/13/2017	15:05	TENDERLOIN	NONE	0 Block of GROVE ST	-122.41564248785478	37.77883231587925	(37.778832315879°, -122.415642487855°)
170919652	ARSON	ARSON OF A VEHICLE	Saturday	11/11/2017	04:25	MISSION	NONE	800 Block of TREAT AV	-122.41343796973081	37.75654520637429	(37.756545206374°, -122.413437969731°)
170929673	ARSON	ARSON OF AN INHABITE	Tuesday	11/14/2017	23:00	PARK	ARREST, BOOKED	1300 Block of GROVE ST	-122.43882822160607	37.775763233387636	(37.775763233388°, -122.438828221606°)
170927075	ARSON	ARSON OF A VEHICLE	Monday	11/13/2017	22:30	BAYVIEW	NONE	2800 Block of INGALLS ST	-122.39108756350615	37.7213138677367	(37.721313867737°, -122.391087563506°)
170913585	ARSON	ARSON OF A VEHICLE	Thursday	11/09/2017	01:00	INGLESIDE	NONE	1600 Block of GENEVA AV	-122.42796564550363	37.711581641573176	(37.711581641573°, -122.427965645504°)
170906223	ARSON	ARSON	Saturday	11/04/2017	04:07	TENDERLOIN	NONE	100 Block of LEAVENWORTH ST	-122.41389315595086	37.782346520673045	(37.782346520673°, -122.413893155951°)

Use Case Experimentation

- Use Case 1.1: SFPD Incidents.
  - Step 1. Import required packages
    - `import os # provides a way of using operating system dependent functionality. To execute shell commands`
    - `import pandas as pd # high-performance, easy-to-use data structures and data analysis tools`
    - `import numpy as np # fundamental package for scientific computing`
    - `import matplotlib.pyplot as plt # 2D plotting library which produces publication quality figures in a variety of hardcopy formats`
    - `import seaborn as sns # high-level interface for drawing attractive statistical graphics`
    - `from IPython.core.display import display, HTML # Top-level display functions for displaying object in different formats`

Use Case Experimentation

- Use Case 1.1: SFPD Incidents Step 1
  - Step 1. Import required packages
    - `import re # provides regular expression matching operations`
    - `import json # easy API to manage JSON files and encoding basic Python object hierarchies`
    - `import folium # builds on the data wrangling strengths of the Python ecosystem and the mapping strengths of the Leaflet.js library`
    - `import pymongo # Python distribution containing API tools for working with MongoDB`
    - `from pymongo import MongoClient, GEO2D # to obtain a MongoClient instance`

Use Case Experimentation

- Use Case 1.1: SFPD Incidents
  - Step 2. Perform a first data exploration
    - Import data
    - Slice data
    - Clean and transform special attributes: dates, indexes, etc
      - Convert to datedime and time delta

- 0 2015-02-04
- 1 2016-03-05

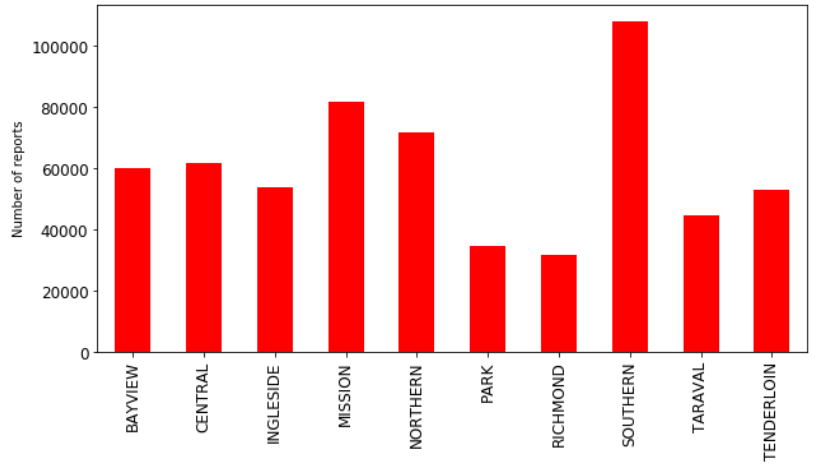
df2 - DataFrame

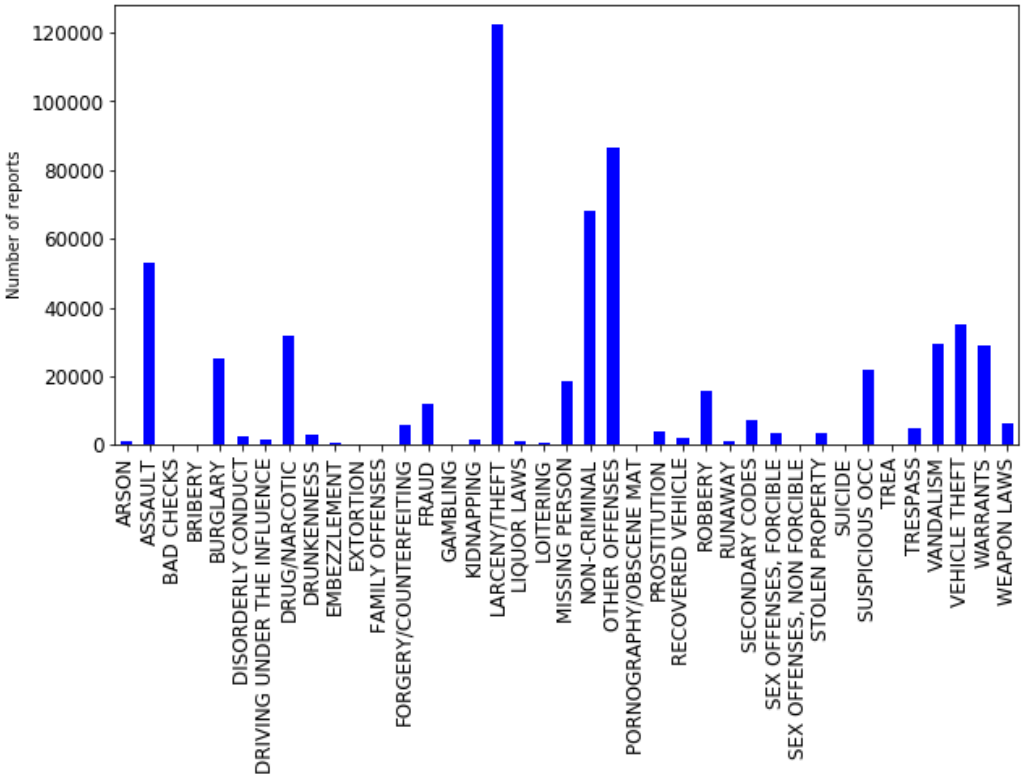
Index	Address	Category	Date	DayOfWeek	Descript	IncidntNum	Location	PdDistrict	Resolution	Time	X	Y	_id	days
0	1700 Block of HARRISON ST	DRUG/NARCOTIC	02/01/2015 12:00:00 AM	Sunday	POSSESSION OF METH-AMPHETA…	150098345	(37.76907480…	MISSION	ARREST, BOOKED	14:00	-122.413	37.7691	5a2e486574ba…	4414
1	1700 Block of HARRISON ST	DRUG/NARCOTIC	02/01/2015 12:00:00 AM	Sunday	POSSESSION OF NARCOTICS PA…	150098345	(37.76907480…	MISSION	ARREST, BOOKED	14:00	-122.413	37.7691	5a2e486574ba…	4414
2	2000 Block of MISSION ST	DRUG/NARCOTIC	02/01/2015 12:00:00 AM	Sunday	POSSESSION OF BASE/ROCK CO…	150098458	(37.76422893…	MISSION	ARREST, BOOKED	16:56	-122.42	37.7642	5a2e486574ba…	4414
3	MISSION ST / 15TH ST	DRUG/NARCOTIC	02/01/2015 12:00:00 AM	Sunday	POSSESSION OF METH-AMPHETA…	150098527	(37.76667375…	MISSION	ARREST, BOOKED	17:02	-122.42	37.7667	5a2e486574ba…	4414
4	700 Block of MARKET ST	DRUG/NARCOTIC	02/01/2015 12:00:00 AM	Sunday	POSSESSION OF MARIJUANA	150098997	(37.78711609…	SOUTHERN	NONE	20:35	-122.404	37.7871	5a2e486574ba…	4414
5	POLK ST / BUSH ST	DRUG/NARCOTIC	02/01/2015 12:00:00 AM	Sunday	POSSESSION OF NARCOTICS PA…	150099268	(37.78869895…	NORTHERN	ARREST, BOOKED	22:34	-122.42	37.7887	5a2e486574ba…	4414
6	0 Block of TURK ST	DRUG/NARCOTIC	02/01/2015 12:00:00 AM	Sunday	POSSESSION OF CONTROLLED S…	150099412	(37.78338623…	TENDERLOIN	ARREST, BOOKED	23:34	-122.41	37.7834	5a2e486574ba…	4414
7	100 Block of LEAVENWORTH …	DRUG/NARCOTIC	02/02/2015 12:00:00 AM	Monday	POSSESSION OF CONTROLLED S…	150099575	(37.78232912…	TENDERLOIN	ARREST, BOOKED	02:23	-122.414	37.7823	5a2e486574ba…	4415
8	KEZAR DR / MARTIN LUTHE…	DRUG/NARCOTIC	02/02/2015 12:00:00 AM	Monday	POSSESSION OF NARCOTICS PA…	150099810	(37.76719994…	PARK	ARREST, BOOKED	08:17	-122.459	37.7672	5a2e486574ba…	4415
9	0 Block of MCALLISTER ST	DRUG/NARCOTIC	02/02/2015 12:00:00 AM	Monday	POSSESSION OF NARCOTICS PA…	150101255	(37.78111921…	TENDERLOIN	ARREST, BOOKED	16:20	-122.413	37.7811	5a2e486574ba…	4415
10	2000 Block of MISSION ST	DRUG/NARCOTIC	02/02/2015 12:00:00 AM	Monday	LOITERING WHERE NARCOT…	150102350	(37.76422893…	MISSION	ARREST, BOOKED	21:43	-122.42	37.7642	5a2e486574ba…	4415
11	800 Block of BRYANT ST	DRUG/NARCOTIC	02/03/2015 12:00:00 AM	Tuesday	POSSESSION OF METH-AMPHETA…	150102968	(37.77542070…	SOUTHERN	NONE	06:00	-122.403	37.7754	5a2e486574ba…	4416
12	0 Block of WALTER ST	DRUG/NARCOTIC	02/03/2015 12:00:00 AM	Tuesday	POSSESSION OF NARCOTICS PA…	150103239	(37.76837670…	PARK	ARREST, BOOKED	08:50	-122.432	37.7684	5a2e486574ba…	4416
13	2100 Block of	DRUG/NARCOTIC	01/29/2015	Thursday	POSSESSION OF	150103411	(37.74746131…	TARAVAL	NONE	09:00	-122.481	37.7475	5a2e486574ba…	4411

Formato | Redimensionar | Color de fondo | Min/max de columna | OK | Cancel

Use Case Experimentation

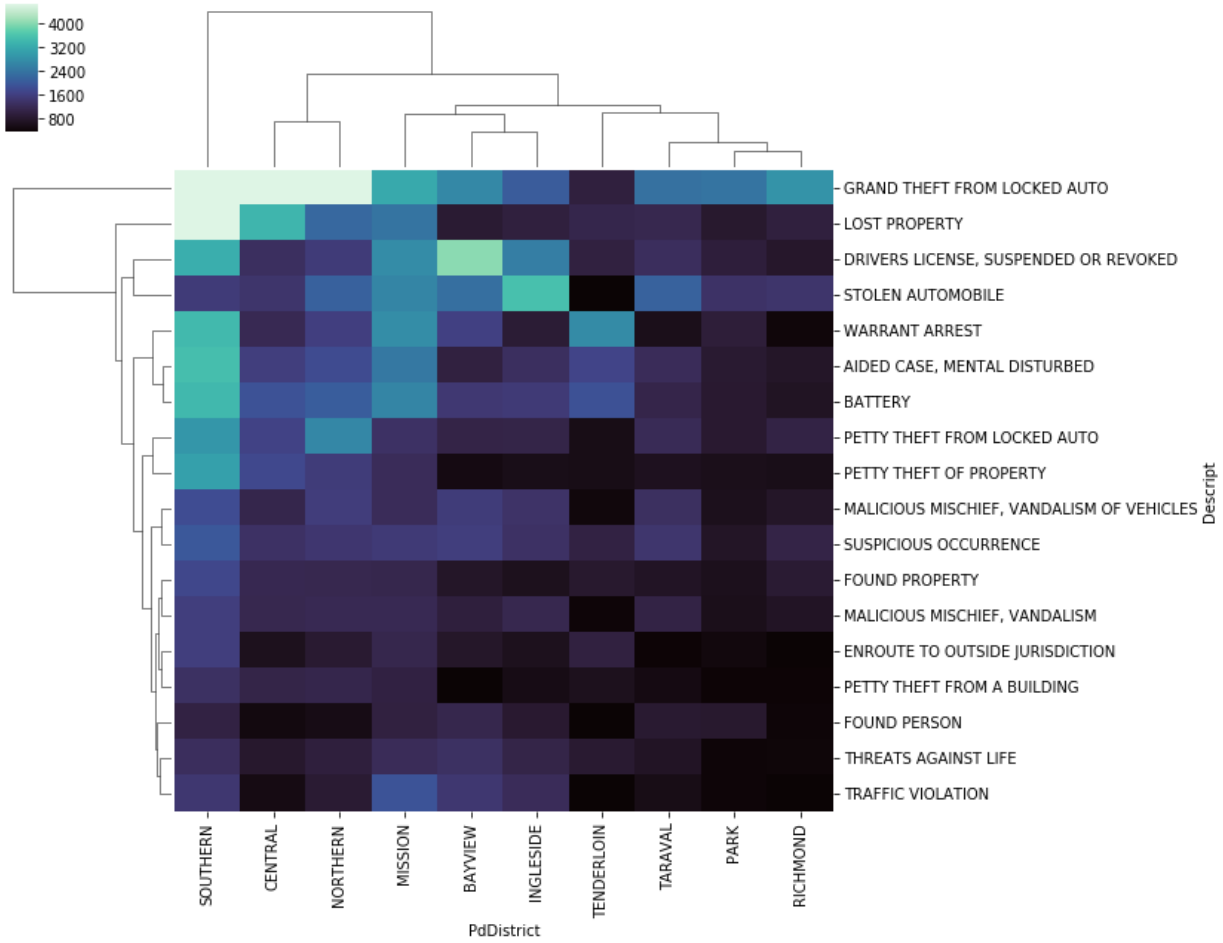
- Use Case 1.1: SFPD Incidents
  - Step 2. Perform a first data exploration
    - Preliminary plots





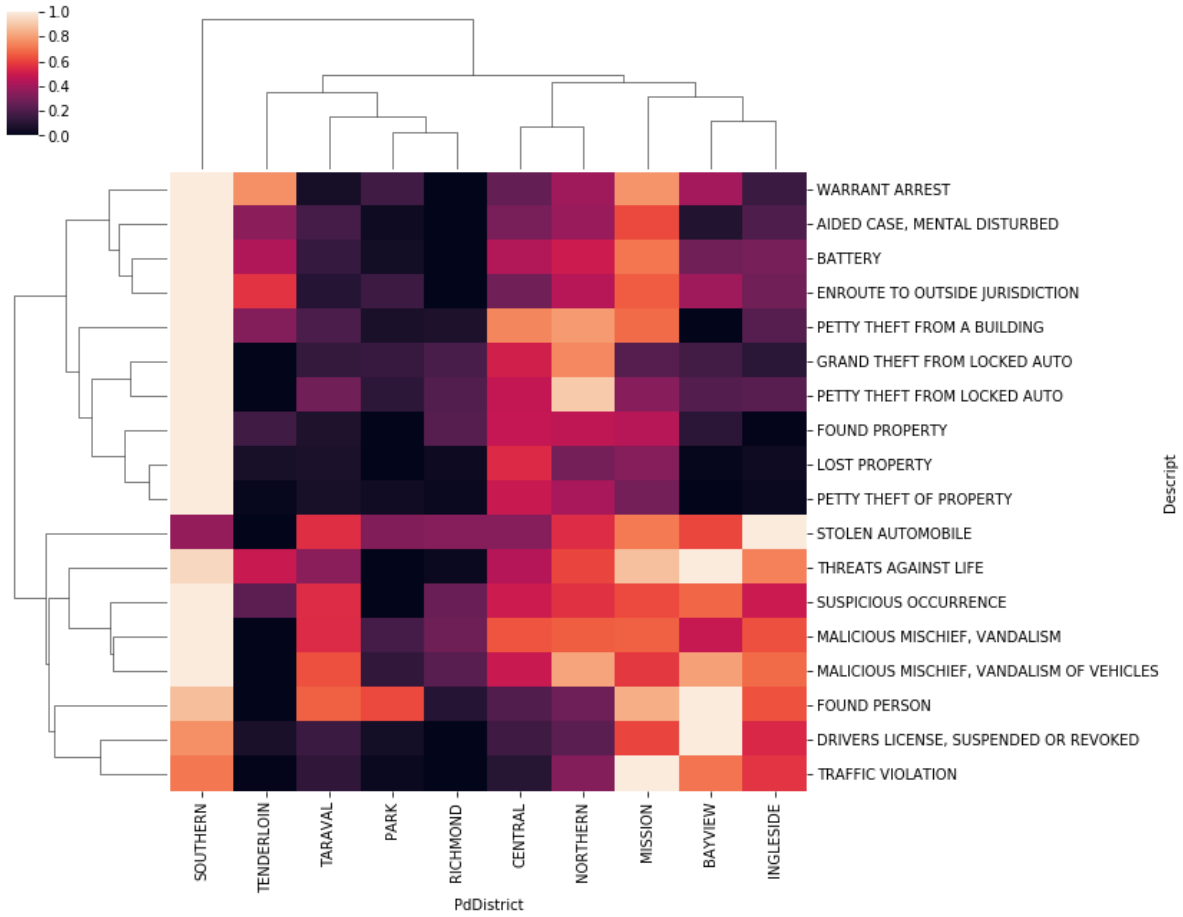
Use Case Experimentation

- Use Case 1.1: SFPD Incidents
  - Step 2. Perform a first data exploration
    - Group different attributes
    - Filter and slice outliers
    - Heatmaps and clusters



Use Case Experimentation

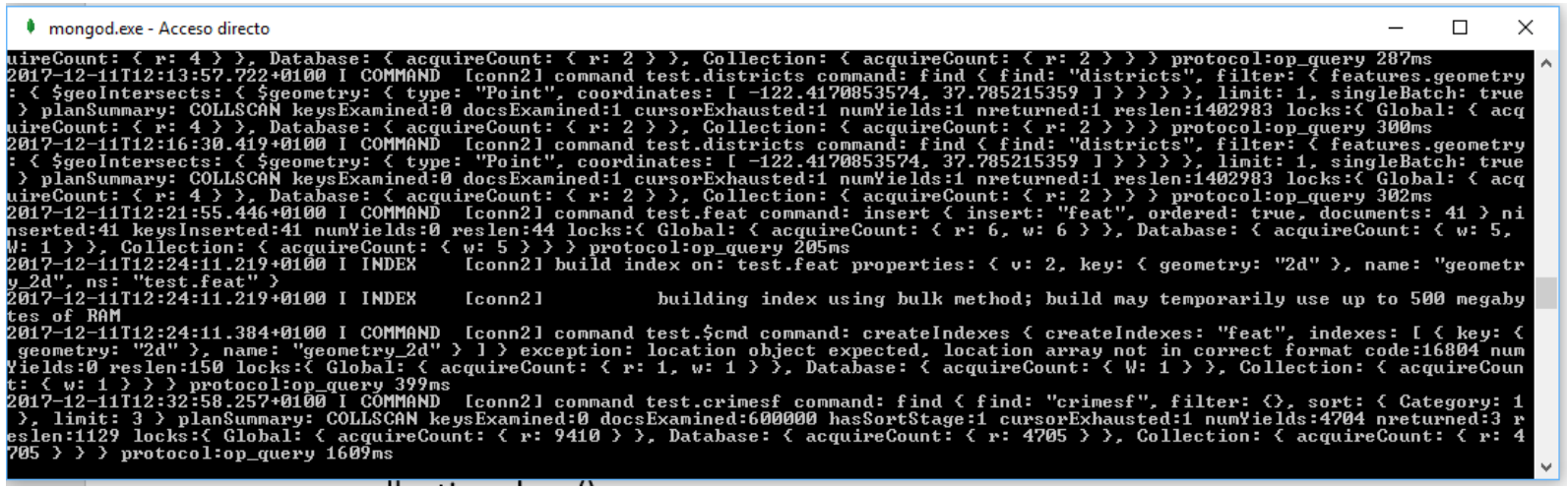
- Use Case 1.1: SFPD Incidents
  - Step 2. Perform a first data exploration
    - Normalize



Use Case Experimentation

- Use Case 1.1: SFPD Incidents
  - Step 3. Mongo Database connection and load

```
print('Mongo version', pymongo.__version__)
client = MongoClient('localhost', 27017)
db = client.test
collection = db.crimesf
collection.drop()
```



Use Case Experimentation

- Use Case 1.1: SFPD Incidents
  - Step 3. Mongo Database connection and load

```
records = json.loads(d_crime.to_json(orient='records'))
collection.delete_many({})
collection.insert_many(records)
```

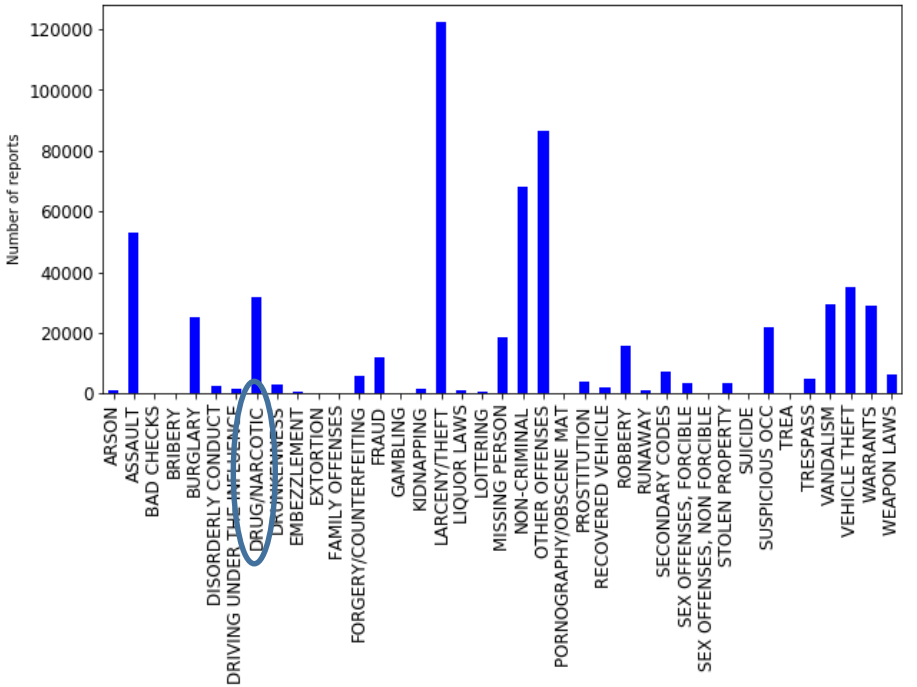
  - Check if you can access the data from the MongoDB

```
In [193]: cursor = collection.find().sort('Category',pymongo.ASCENDING).limit(3)

In [194]: for doc in cursor:
     ...:     print(doc)
     ...:
     ...:
{'_id': ObjectId('5a2e486574ba00207859c1f1'), 'IncidntNum': 140009459, 'Category': 'ARSON', 'Descript': 'ARSON',
'DayOfWeek': 'Saturday', 'Date': '01/04/2014 12:00:00 AM', 'Time': '03:52', 'PdDistrict': 'NORTHERN', 'Resolution':
'ARREST, BOOKED', 'Address': 'SACRAMENTO ST / POLK ST', 'X': -122.4208746324, 'Y': 37.7914943052, 'Location':
'(37.7914943051906, -122.420874632415)', 'days': 4021.0}
{'_id': ObjectId('5a2e486574ba00207859c2b3'), 'IncidntNum': 150100081, 'Category': 'ARSON', 'Descript': 'ARSON OF AN
INHABITED DWELLING', 'DayOfWeek': 'Monday', 'Date': '02/02/2015 12:00:00 AM', 'Time': '10:05', 'PdDistrict': 'MISSION',
'Resolution': 'ARREST, BOOKED', 'Address': '500 Block of VALENCIA ST', 'X': -122.4218764885, 'Y': 37.7640888945,
'Location': '(37.7640888944532, -122.421876488492)', 'days': 4415.0}
{'_id': ObjectId('5a2e486574ba00207859c2f2'), 'IncidntNum': 150100677, 'Category': 'ARSON', 'Descript': 'ARSON OF A
VEHICLE', 'DayOfWeek': 'Monday', 'Date': '02/02/2015 12:00:00 AM', 'Time': '12:56', 'PdDistrict': 'MISSION', 'Resolution':
'NONE', 'Address': '200 Block of SHOTWELL ST', 'X': -122.4163748358, 'Y': 37.7644225166, 'Location': '(37.7644225165568,
-122.416374835778)', 'days': 4415.0}
```

Use Case Experimentation

- Use Case 1.1: SFPD Incidents
  - Step 3. Perform specific queries on MongoBD to generate focus datasets,
  - e.g. {"Category":"DRUG/NARCOTIC"}



Use Case Experimentation

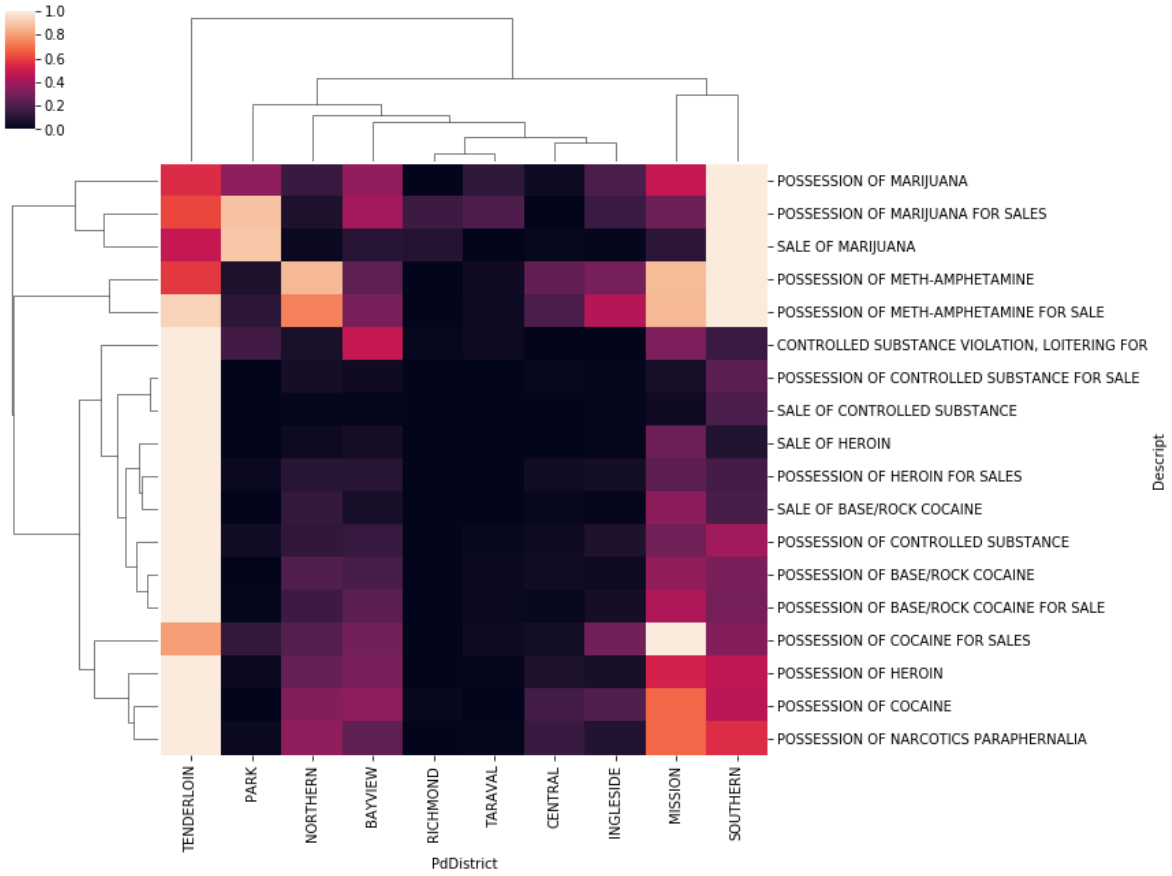
- Use Case 1.1: SFPD Incidents
  - Step 3. Filter specific data (DRUG/NARCOTIC)
  - Organize incidents' descriptions versus Districts where they were detected
  - Have a look on most common “Descriptions”

```
In [195]: c=df2['Descript'].value_counts()

In [196]: c.sort_values(ascending=False)
Out[196]:
POSSESSION OF NARCOTICS PARAPHERNALIA                6140
POSSESSION OF BASE/ROCK COCAINE                      3516
POSSESSION OF MARIJUANA                              3463
POSSESSION OF METH-AMPHETAMINE                       2631
SALE OF BASE/ROCK COCAINE                            2245
POSSESSION OF BASE/ROCK COCAINE FOR SALE             1946
POSSESSION OF MARIJUANA FOR SALES                    1620
POSSESSION OF HEROIN                                 1236
POSSESSION OF CONTROLLED SUBSTANCE                   1213
POSSESSION OF COCAINE                                 917
POSSESSION OF METH-AMPHETAMINE FOR SALE               781
SALE OF MARIJUANA                                     752
POSSESSION OF CONTROLLED SUBSTANCE FOR SALE           639
POSSESSION OF HEROIN FOR SALES                        533
SALE OF CONTROLLED SUBSTANCE                          418
POSSESSION OF COCAINE FOR SALES                       347
SALE OF HEROIN                                        306
CONTROLLED SUBSTANCE VIOLATION, LOITERING FOR         272
MAINTAINING PREMISE WHERE NARCOTICS ARE SOLD/USED     245
LOITERING WHERE NARCOTICS ARE SOLD/USED               202
TRANSPORTATION OF MARIJUANA                           201
PLANTING/CULTIVATING MARIJUANA                        164
UNDER INFLUENCE OF DRUGS IN A PUBLIC PLACE            163
SALE OF COCAINE                                       151
TRANSPORTATION OF METH-AMPHETAMINE                    134
TRANSPORTATION OF COCAINE                             127
POSSESSION OF AMPHETAMINE FOR SALES                   122
SALE OF METH-AMPHETAMINE                              103
UNDER THE INFLUENCE OF CONTROLLED SUBSTANCES          103
```

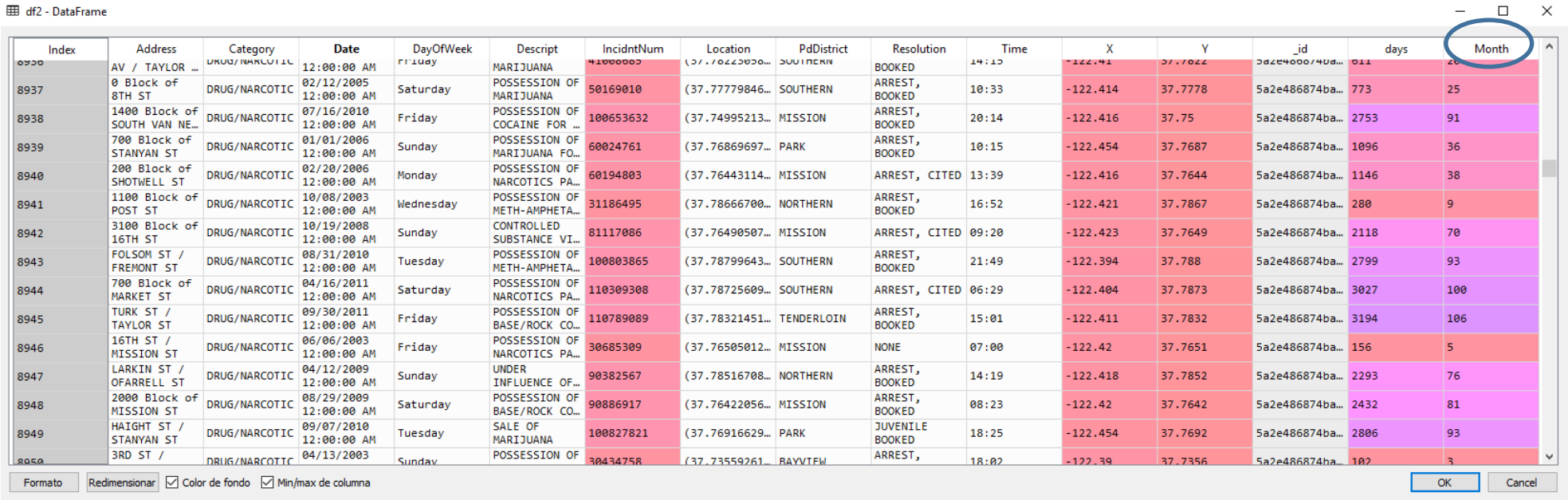
Use Case Experimentation

- Use Case 1.1: SFPD Incidents
  - Step 3. Filter specific data (DRUG/NARCOTIC)
  - Normalize and cluster



Use Case Experimentation

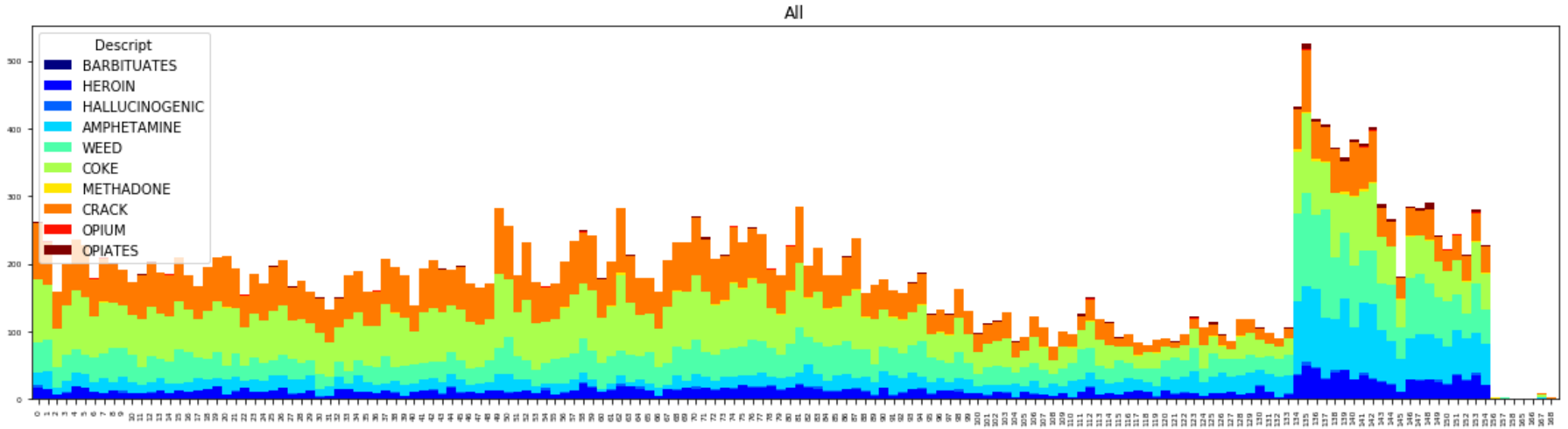
- Use Case 1.1: SFPD Incidents
  - Step 4. Time Series Analysis
    - Obtain new column with corresponding months
    - Group by month and incident description



Index	Address	Category	Date	DayOfWeek	Descript	IncidntNum	Location	PdDistrict	Resolution	Time	X	Y	_id	days	Month
8936	AV / TAYLOR …	DRUG/NARCOTIC	12:00:00 AM	Friday	MARIJUANA	41006865	(37.78223656…	SOUTHERN	BOOKED	14:15	-122.41	37.7822	5a2e486874ba…	611	20
8937	0 Block of 8TH ST	DRUG/NARCOTIC	02/12/2005 12:00:00 AM	Saturday	POSSESSION OF MARIJUANA	50169010	(37.77779846…	SOUTHERN	ARREST, BOOKED	10:33	-122.414	37.7778	5a2e486874ba…	773	25
8938	1400 Block of SOUTH VAN NE…	DRUG/NARCOTIC	07/16/2010 12:00:00 AM	Friday	POSSESSION OF COCAINE FOR …	100653632	(37.74995213…	MISSION	ARREST, BOOKED	20:14	-122.416	37.75	5a2e486874ba…	2753	91
8939	700 Block of STANYAN ST	DRUG/NARCOTIC	01/01/2006 12:00:00 AM	Sunday	POSSESSION OF MARIJUANA FO…	60024761	(37.76869697…	PARK	ARREST, BOOKED	10:15	-122.454	37.7687	5a2e486874ba…	1096	36
8940	200 Block of SHOTWELL ST	DRUG/NARCOTIC	02/20/2006 12:00:00 AM	Monday	POSSESSION OF NARCOTICS PA…	60194803	(37.76443114…	MISSION	ARREST, CITED	13:39	-122.416	37.7644	5a2e486874ba…	1146	38
8941	1100 Block of POST ST	DRUG/NARCOTIC	10/08/2003 12:00:00 AM	Wednesday	POSSESSION OF METH-AMPHETA…	31186495	(37.78666700…	NORTHERN	ARREST, BOOKED	16:52	-122.421	37.7867	5a2e486874ba…	280	9
8942	3100 Block of 16TH ST	DRUG/NARCOTIC	10/19/2008 12:00:00 AM	Sunday	CONTROLLED SUBSTANCE VI…	81117086	(37.76490507…	MISSION	ARREST, CITED	09:20	-122.423	37.7649	5a2e486874ba…	2118	70
8943	FOLSOM ST / FREMONT ST	DRUG/NARCOTIC	08/31/2010 12:00:00 AM	Tuesday	POSSESSION OF METH-AMPHETA…	100803865	(37.78799643…	SOUTHERN	ARREST, BOOKED	21:49	-122.394	37.788	5a2e486874ba…	2799	93
8944	700 Block of MARKET ST	DRUG/NARCOTIC	04/16/2011 12:00:00 AM	Saturday	POSSESSION OF NARCOTICS PA…	110309308	(37.78725609…	SOUTHERN	ARREST, CITED	06:29	-122.404	37.7873	5a2e486874ba…	3027	100
8945	TURK ST / TAYLOR ST	DRUG/NARCOTIC	09/30/2011 12:00:00 AM	Friday	POSSESSION OF BASE/ROCK CO…	110789089	(37.78321451…	TENDERLOIN	ARREST, BOOKED	15:01	-122.411	37.7832	5a2e486874ba…	3194	106
8946	16TH ST / MISSION ST	DRUG/NARCOTIC	06/06/2003 12:00:00 AM	Friday	POSSESSION OF NARCOTICS PA…	30685309	(37.76505012…	MISSION	NONE	07:00	-122.42	37.7651	5a2e486874ba…	156	5
8947	LARKIN ST / OFARRELL ST	DRUG/NARCOTIC	04/12/2009 12:00:00 AM	Sunday	UNDER INFLUENCE OF…	90382567	(37.78516708…	NORTHERN	ARREST, BOOKED	14:19	-122.418	37.7852	5a2e486874ba…	2293	76
8948	2000 Block of MISSION ST	DRUG/NARCOTIC	08/29/2009 12:00:00 AM	Saturday	POSSESSION OF BASE/ROCK CO…	90886917	(37.76422056…	MISSION	ARREST, BOOKED	08:23	-122.42	37.7642	5a2e486874ba…	2432	81
8949	HAIGHT ST / STANYAN ST	DRUG/NARCOTIC	09/07/2010 12:00:00 AM	Tuesday	SALE OF MARIJUANA	100827821	(37.76916629…	PARK	JUVENILE BOOKED	18:25	-122.454	37.7692	5a2e486874ba…	2806	93
8950	3RD ST /	DRUG/NARCOTIC	04/13/2003	Sunday	POSSESSION OF	30434758	(37.73559261…	BAYVIEW	ARREST,	18:02	-122.39	37.7356	5a2e486874ba…	102	3

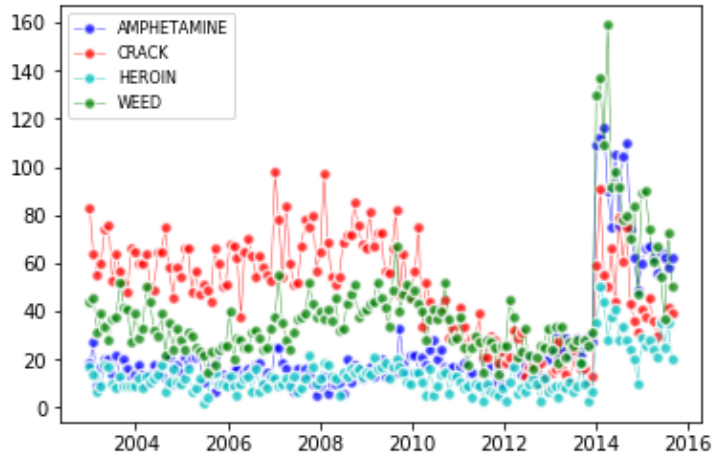
Use Case Experimentation

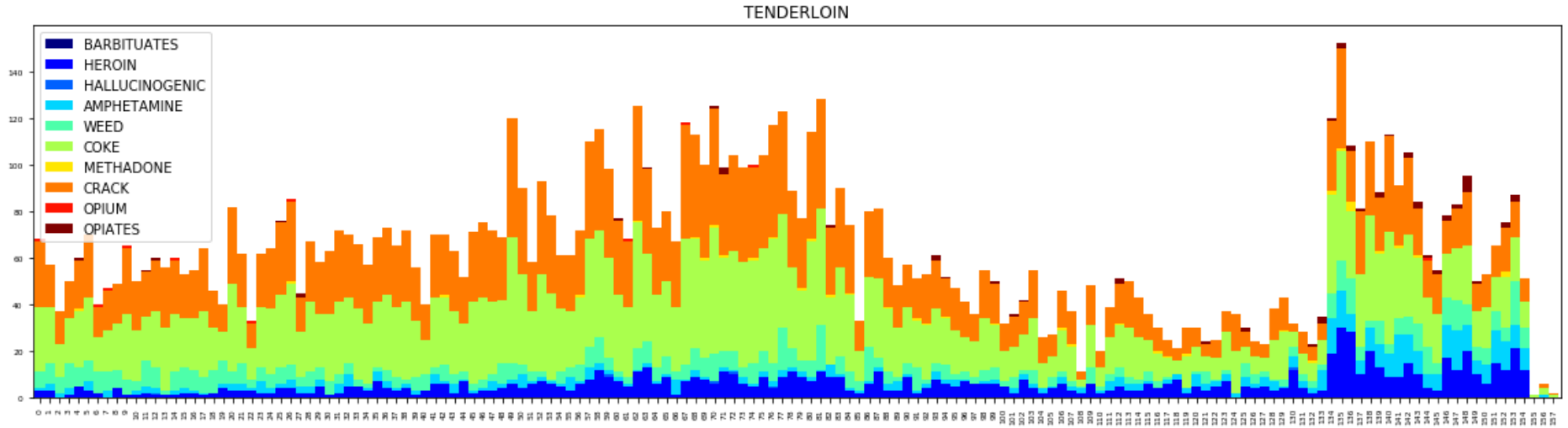
- Use Case 1.1: SFPD Incidents
  - Step 4. Time Series Analysis
    - Find incident's descriptions, in "DRUG/NARCOTIC“ category, which are related to certain word patters
    - For example: “BARBITUATES”, “COCAINE”, “METHADONE”, etc.



Use Case Experimentation

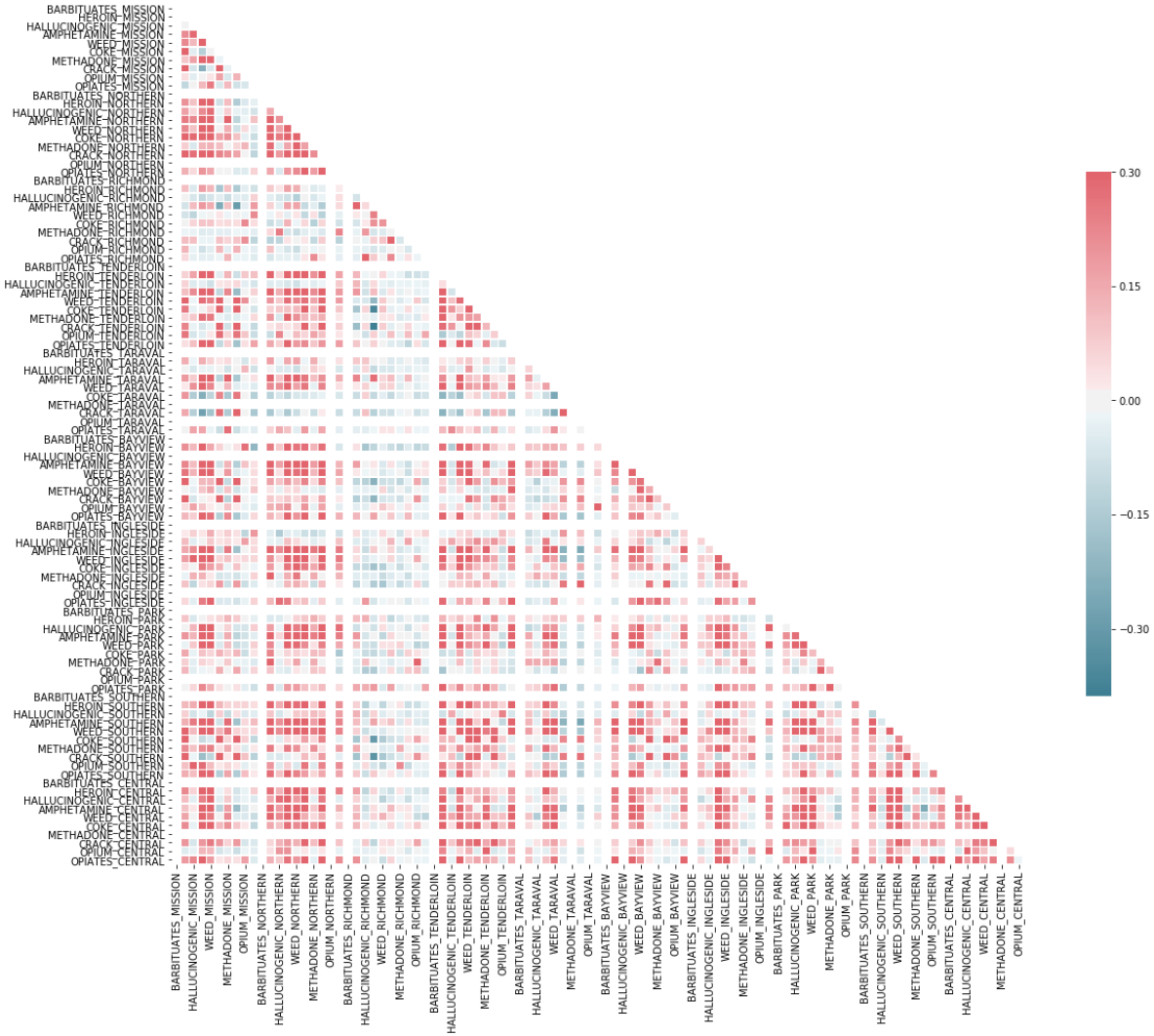
- Use Case 1.1: SFPD Incidents
  - Step 4. Time Series Analysis
    - Focus on real dates and Districts





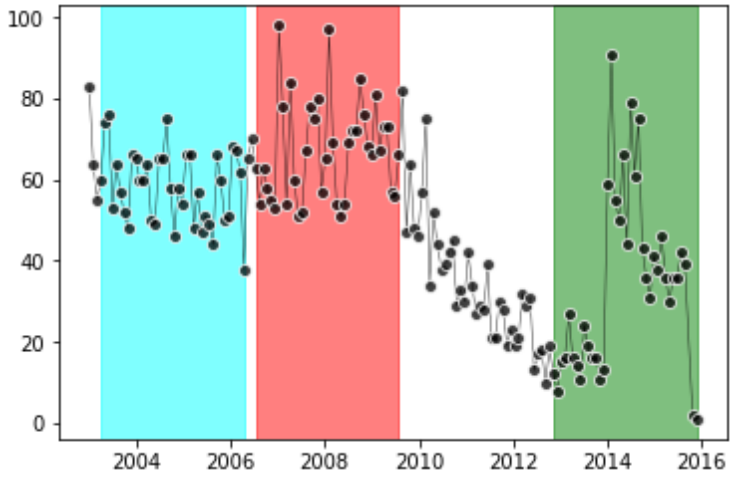
Use Case Experimentation

- Use Case 1.1: SFPD Incidents
  - Step 5. Correlation Analysis
    - Focus on Districts and Incident's descriptions



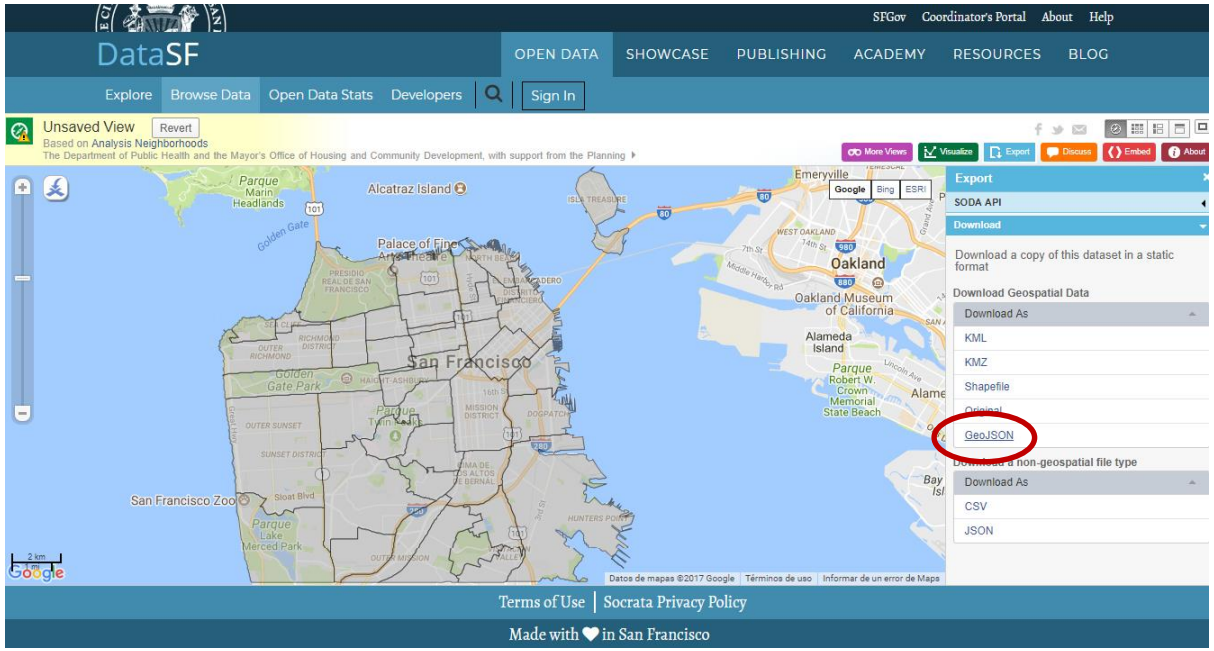
Use Case Experimentation

- Use Case 1.1: SFPD Incidents
  - Step 6. Mapping relationships
    - Let's isolate incident descriptions-related records, e.g. crack



Use Case Experimentation

- Use Case 1.1: SFPD Incidents
  - Step 6. Mapping relationships
    - Let's isolate incident descriptions-related records, e.g. crack
    - Generate separate MongoBD collections to perform geospatial querying
    - Before, store in MongoBD geospatial districts (GeoJson)



https://data.sfgov.org/Geographic-Locations-and-Boundaries/Analysis-Neighborhoods/p5b7-5n3h

Use Case Experimentation

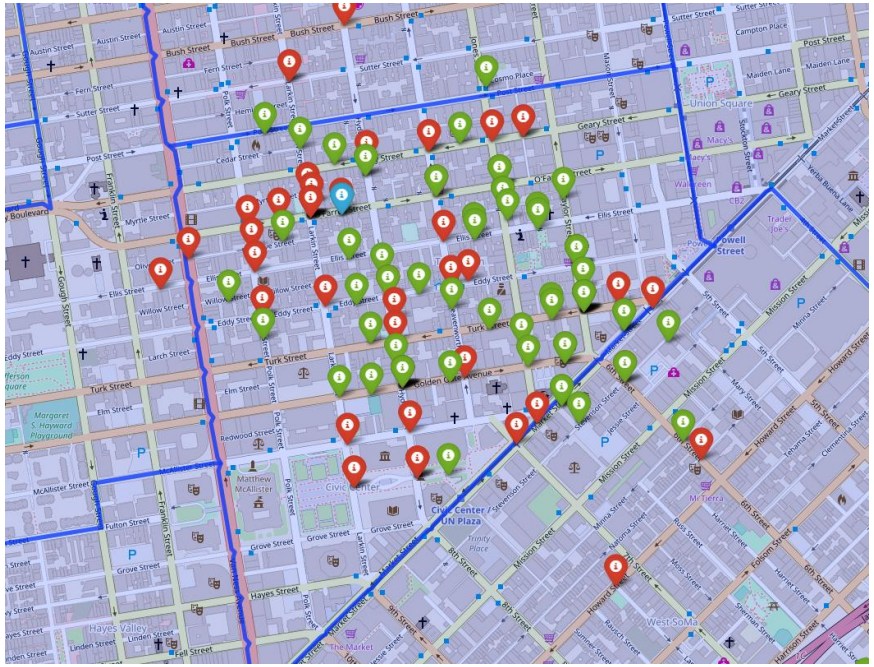
- Use Case 1.1: SFPD Incidents
  - Step 6. Mapping relationships
    - Geospatial querying in MongoBD

```
query_feat = {"geometry":
                { "$geoIntersects":
                { "$geometry":
                    {"type": "Point",
                     "coordinates": [cursor2["longitude"],cursor2["latitude"]]
                    }
                }
            }
        }

for doc in collection_features.find(query_feat):
    print(doc)
```

Use Case Experimentation

- Use Case 1.1: SFPD Incidents
  - Step 7. Printing data in Maps (Folium package)



folium

Python Data, Leaflet.js Maps

https://github.com/python-visualization/folium

Discussions and Conclusions

- ✓ Motivation covered
- ✓ A first approach to real-world
- ✓ Exploring fine-grain data
- ✓ Involving open data extraction, cleaning, consolidation, transformation, enrichment, analysis and visualization
- ✓ Consolidate acquired knowledge and introduce new one
  - ✓ Bringing past and next modules

Module 5

Use Case 1

LESSON 1

José Manuel García Nieto – University of Málaga