# Projeto LabRedes de Conhecimento

## San Francisco Crime Classification


João Paulo de Melo

## Sobre o tema

De 1934 a 1963, San Francisco foi famoso por abrigar alguns dos criminosos mais notórios do mundo na incontornável ilha de Alcatraz.

O conjunto de dados da competição fornece quase 12 anos de relatórios sobre crimes em todos os bairros de São Francisco. Com o tempo e o local, devemos prever a categoria do crime que ocorreu.

Outro desafio proposto foi exibir estes dados de forma visual, já que temos coordenadas.

# Features

| Descript | DayOfWeek | PdDistrict | Resolution | Address | X | Y | Category |
|---|---|---|---|---|---|---|---|
| WARRANT ARREST | Wednesday | NORTHERN | ARREST, BOOKED | OAK ST / LAGUNA ST | -122.425892 | 37.774599 | WARRANTS |
| TRAFFIC VIOLATION ARREST | Wednesday | NORTHERN | ARREST, BOOKED | OAK ST / LAGUNA ST | -122.425892 | 37.774599 | OTHER OFFENSES |
| TRAFFIC VIOLATION ARREST | Wednesday | NORTHERN | ARREST, BOOKED | VANNESS AV / GREENWICH ST | -122.424363 | 37.800414 | OTHER OFFENSES |
| GRAND THEFT FROM LOCKED AUTO | Wednesday | NORTHERN | NONE | 1500 Block of LOMBARD ST | -122.426995 | 37.800873 | LARCENY/THEFT |
| GRAND THEFT FROM LOCKED AUTO | Wednesday | PARK | NONE | 100 Block of BRODERICK ST | -122.438738 | 37.77154 | LARCENY/THEFT |

# Describe

```
In [172]: df_test.describe(include='all')
```

Out[172]:

| | Id | DayOfWeek | PdDistrict | Address | X | Y |
|---|---|---|---|---|---|---|
| **count** | 884262.000000 | 884262 | 884262 | 884262 | 884262.000000 | 884262.000000 |
| **unique** | NaN | 7 | 10 | 23184 | NaN | NaN |
| **top** | NaN | Friday | SOUTHERN | 800 Block of BRYANT ST | NaN | NaN |
| **freq** | NaN | 134703 | 157456 | 26984 | NaN | NaN |
| **mean** | 442130.500000 | NaN | NaN | NaN | -122.422693 | 37.771476 |
| **std** | 255264.596206 | NaN | NaN | NaN | 0.030985 | 0.484824 |
| **min** | 0.000000 | NaN | NaN | NaN | -122.513642 | 37.707879 |
| **25%** | 221065.250000 | NaN | NaN | NaN | -122.433069 | 37.752374 |
| **50%** | 442130.500000 | NaN | NaN | NaN | -122.416517 | 37.775421 |
| **75%** | 663195.750000 | NaN | NaN | NaN | -122.406959 | 37.784353 |
| **max** | 884261.000000 | NaN | NaN | NaN | -120.500000 | 90.000000 |

```
In [173]: df_train.describe(include='all')
```

Out[173]:

| | Category | Descript | DayOfWeek | PdDistrict | Resolution | Address | X | Y |
|---|---|---|---|---|---|---|---|---|
| **count** | 878049 | 878049 | 878049 | 878049 | 878049 | 878049 | 878049.000000 | 878049.000000 |
| **unique** | 39 | 879 | 7 | 10 | 17 | 23228 | NaN | NaN |
| **top** | LARCENY/THEFT | GRAND THEFT FROM LOCKED AUTO | Friday | SOUTHERN | NONE | 800 Block of BRYANT ST | NaN | NaN |
| **freq** | 174900 | 60022 | 133734 | 157182 | 526790 | 26533 | NaN | NaN |
| **mean** | NaN | NaN | NaN | NaN | NaN | NaN | -122.422616 | 37.771020 |
| **std** | NaN | NaN | NaN | NaN | NaN | NaN | 0.030354 | 0.456893 |
| **min** | NaN | NaN | NaN | NaN | NaN | NaN | -122.513642 | 37.707879 |
| **25%** | NaN | NaN | NaN | NaN | NaN | NaN | -122.432952 | 37.752427 |
| **50%** | NaN | NaN | NaN | NaN | NaN | NaN | -122.416420 | 37.775421 |
| **75%** | NaN | NaN | NaN | NaN | NaN | NaN | -122.406959 | 37.784369 |
| **max** | NaN | NaN | NaN | NaN | NaN | NaN | -120.500000 | 90.000000 |

# Mapeamento de crimes

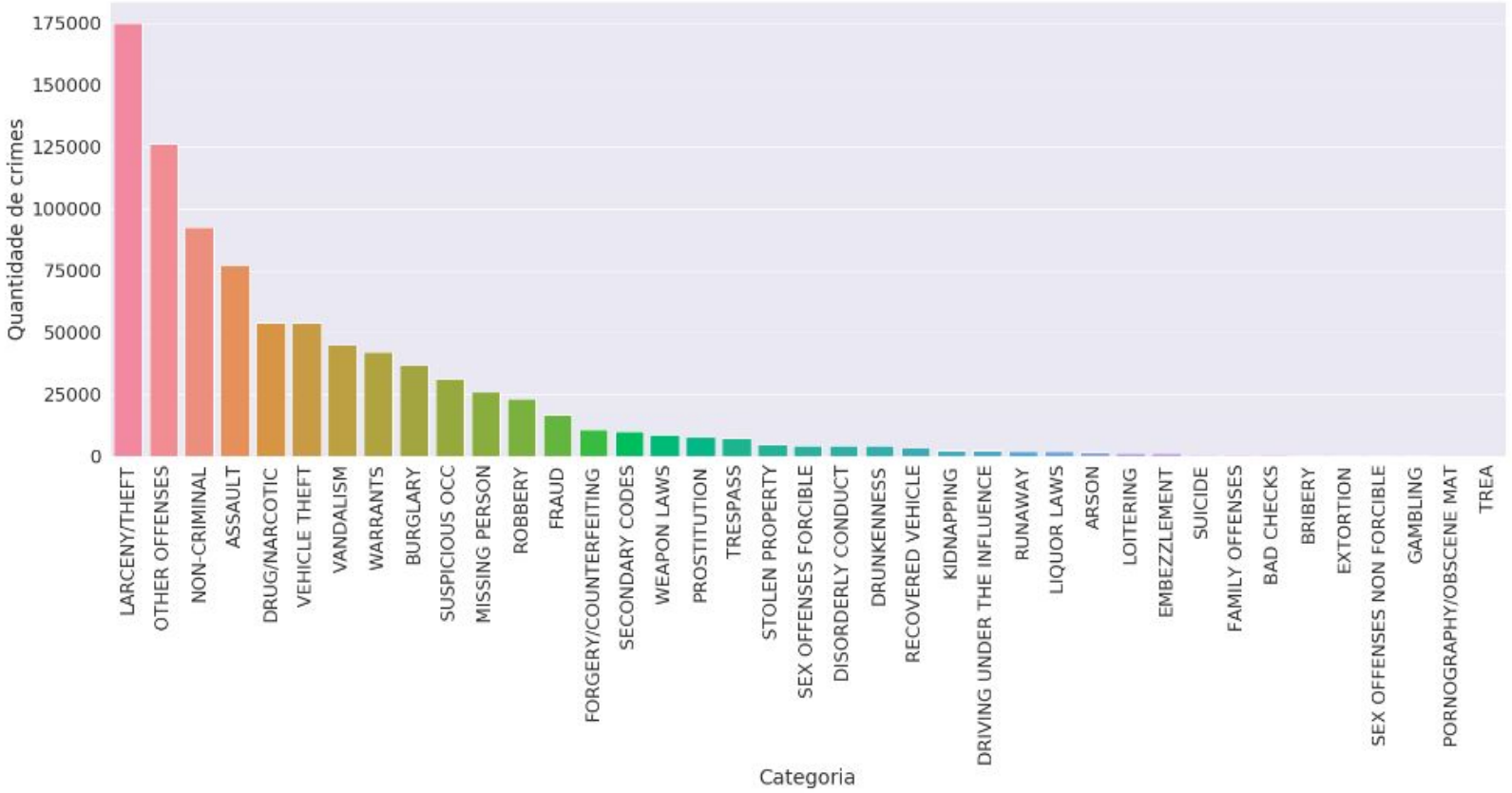
In [3]:
```
categories = df_train["Category"].unique()
category_map = {}
for key, crime in enumerate(categories):
    category_map[key] = crime

print(category_map)
print("\nQuantidade de crimes:",len(categories))
```

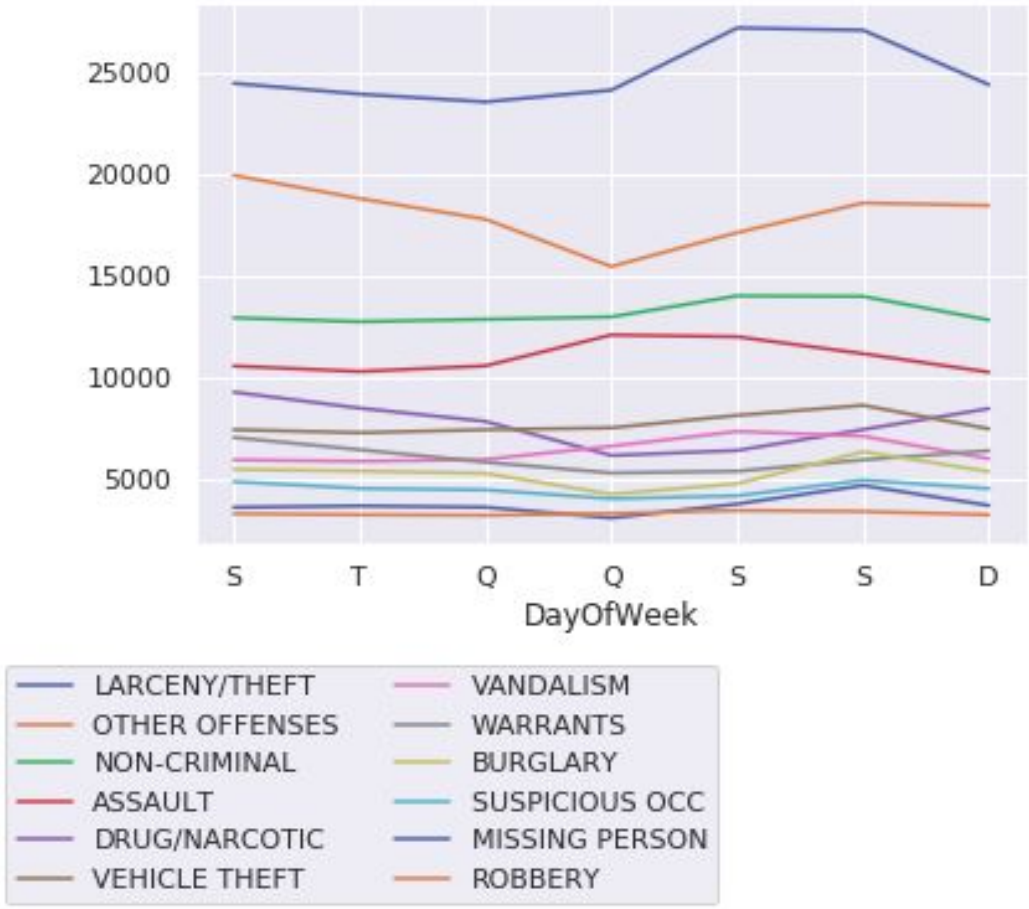
```
{0: 'WARRANTS', 1: 'OTHER OFFENSES', 2: 'LARCENY/THEFT', 3: 'VEHICLE THEFT', 4: 'VANDALISM', 5: 'NON-CRIMINAL', 6:
'ROBBERY', 7: 'ASSAULT', 8: 'WEAPON LAWS', 9: 'BURGLARY', 10: 'SUSPICIOUS OCC', 11: 'DRUNKENNESS', 12: 'FORGERY/COU
NTERFEITING', 13: 'DRUG/NARCOTIC', 14: 'STOLEN PROPERTY', 15: 'SECONDARY CODES', 16: 'TRESPASS', 17: 'MISSING PERSO
N', 18: 'FRAUD', 19: 'KIDNAPPING', 20: 'RUNAWAY', 21: 'DRIVING UNDER THE INFLUENCE', 22: 'SEX OFFENSES FORCIBLE', 2
3: 'PROSTITUTION', 24: 'DISORDERLY CONDUCT', 25: 'ARSON', 26: 'FAMILY OFFENSES', 27: 'LIQUOR LAWS', 28: 'BRIBERY',
29: 'EMBEZZLEMENT', 30: 'SUICIDE', 31: 'LOITERING', 32: 'SEX OFFENSES NON FORCIBLE', 33: 'EXTORTION', 34: 'GAMBLIN
G', 35: 'BAD CHECKS', 36: 'TREA', 37: 'RECOVERED VEHICLE', 38: 'PORNOGRAPHY/OBSCENE MAT'}

Quantidade de crimes: 39
```

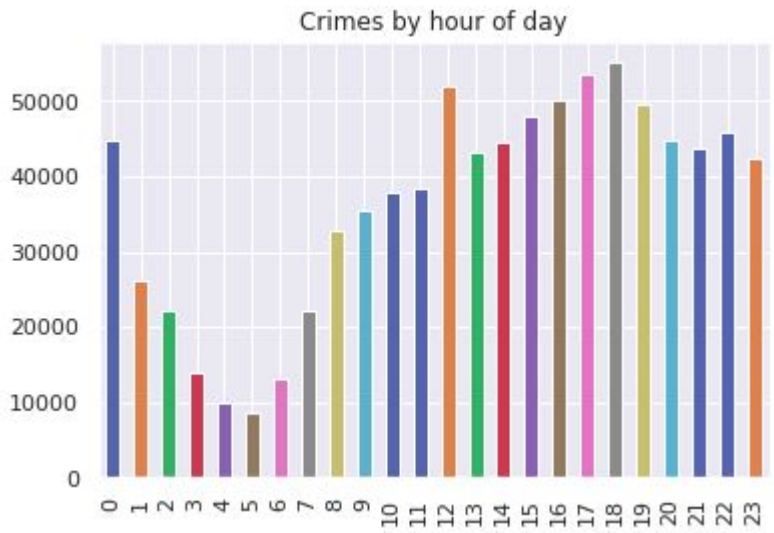
# Quantidade de crimes por categoria

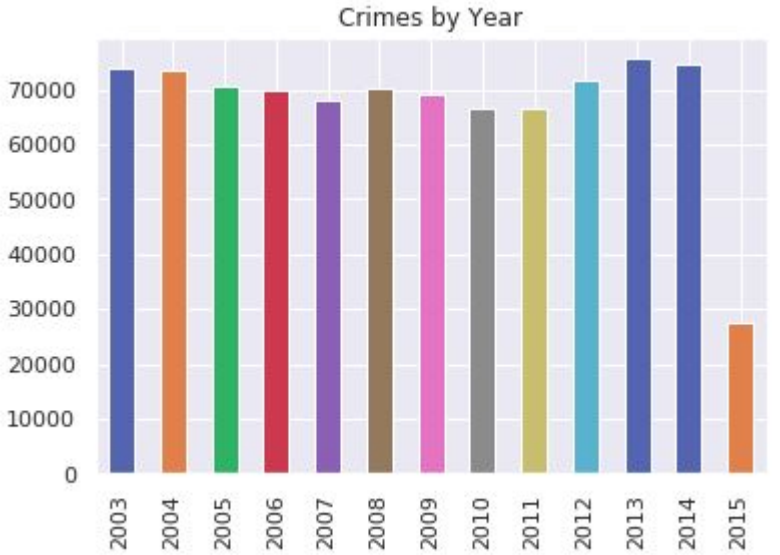


# Crimes por dia da semana

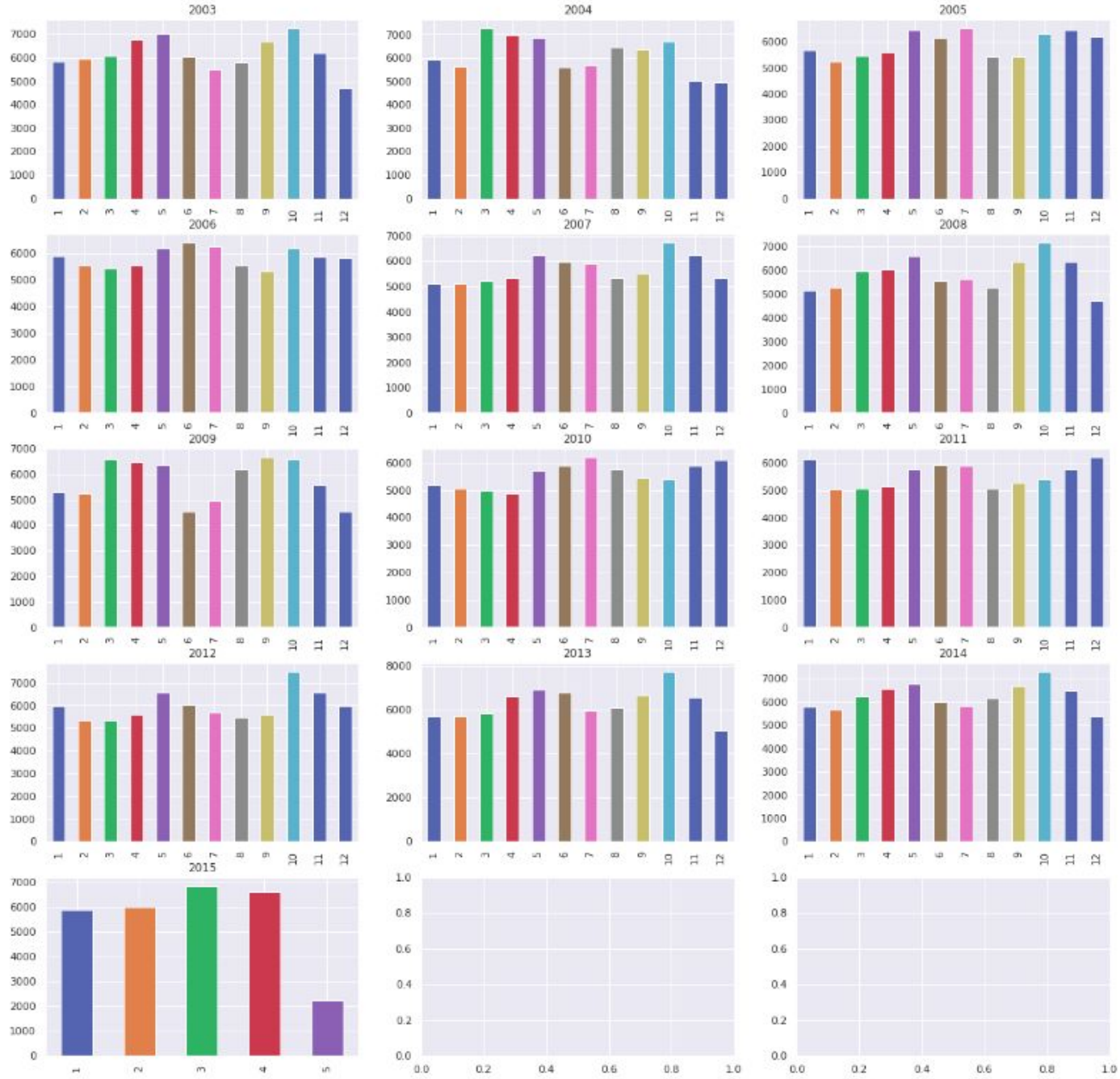


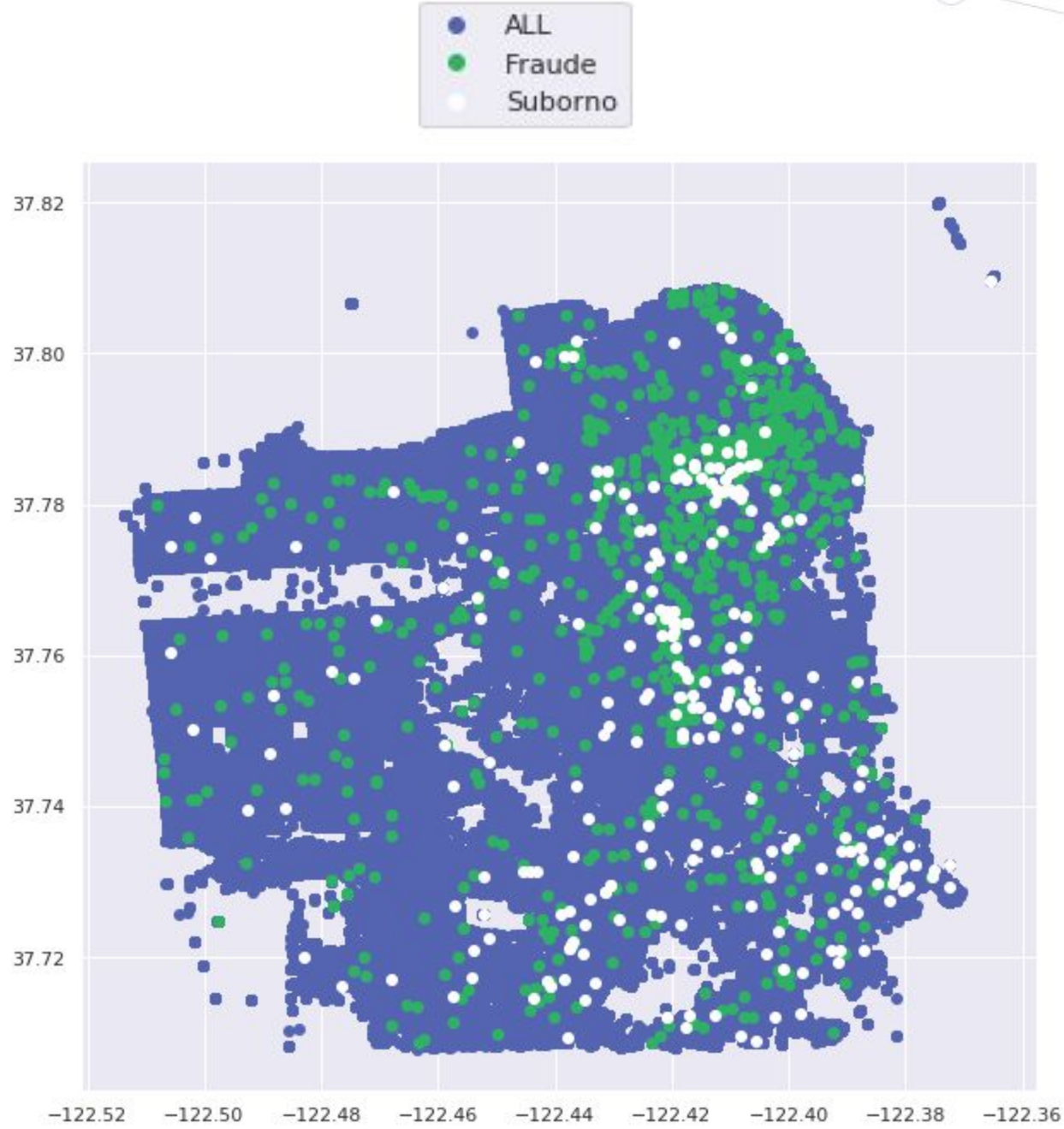
# Quantidade por tempo

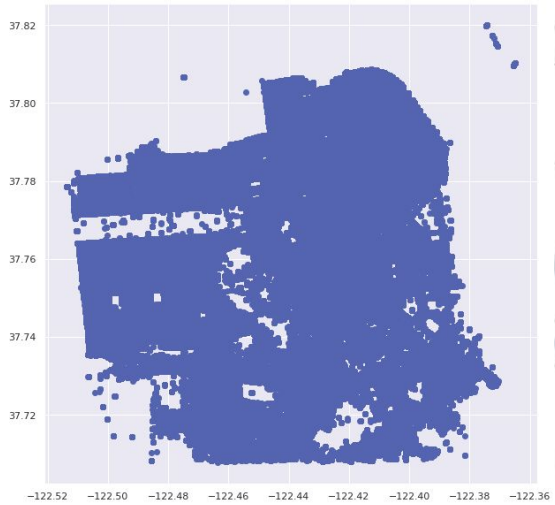


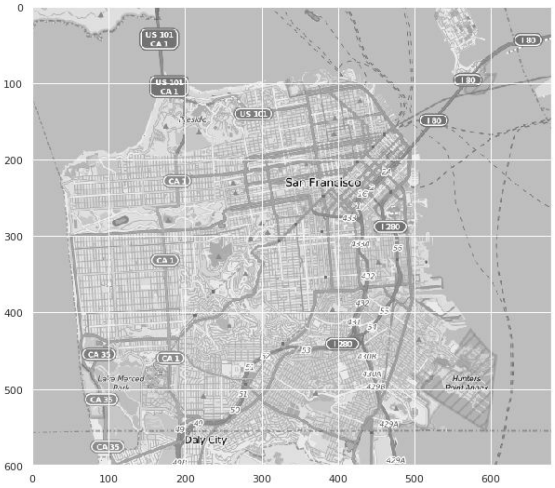



Crimes by month an each year




ALL
Fraude
Suborno
37.82
37.80
37.78
37.76
37.74
37.72
−122.52
−122.50
−122.48
−122.46
−122.44
−122.42
−122.40
−122.38
−122.36


37.82
37.80
37.78
37.76
37.74
37.72
−122.52
−122.50
−122.48
−122.46
−122.44
−122.42
−122.40
−122.38
−122.36


0
100
200
300
400
500
600
San Francisco
Daly City