

APRENTIZAJE COMPUTACIONAL

PRÁCTICA 3

KAGGLE

CRIMES IN CHICAGO

ADRIÁN NIETO - 1569312

- **DATASET**
- **ANÁLISIS DEL DATASET**
- **PREPROCESSING**
- **CLASIFICACIÓN**
- **REGRESIÓN**

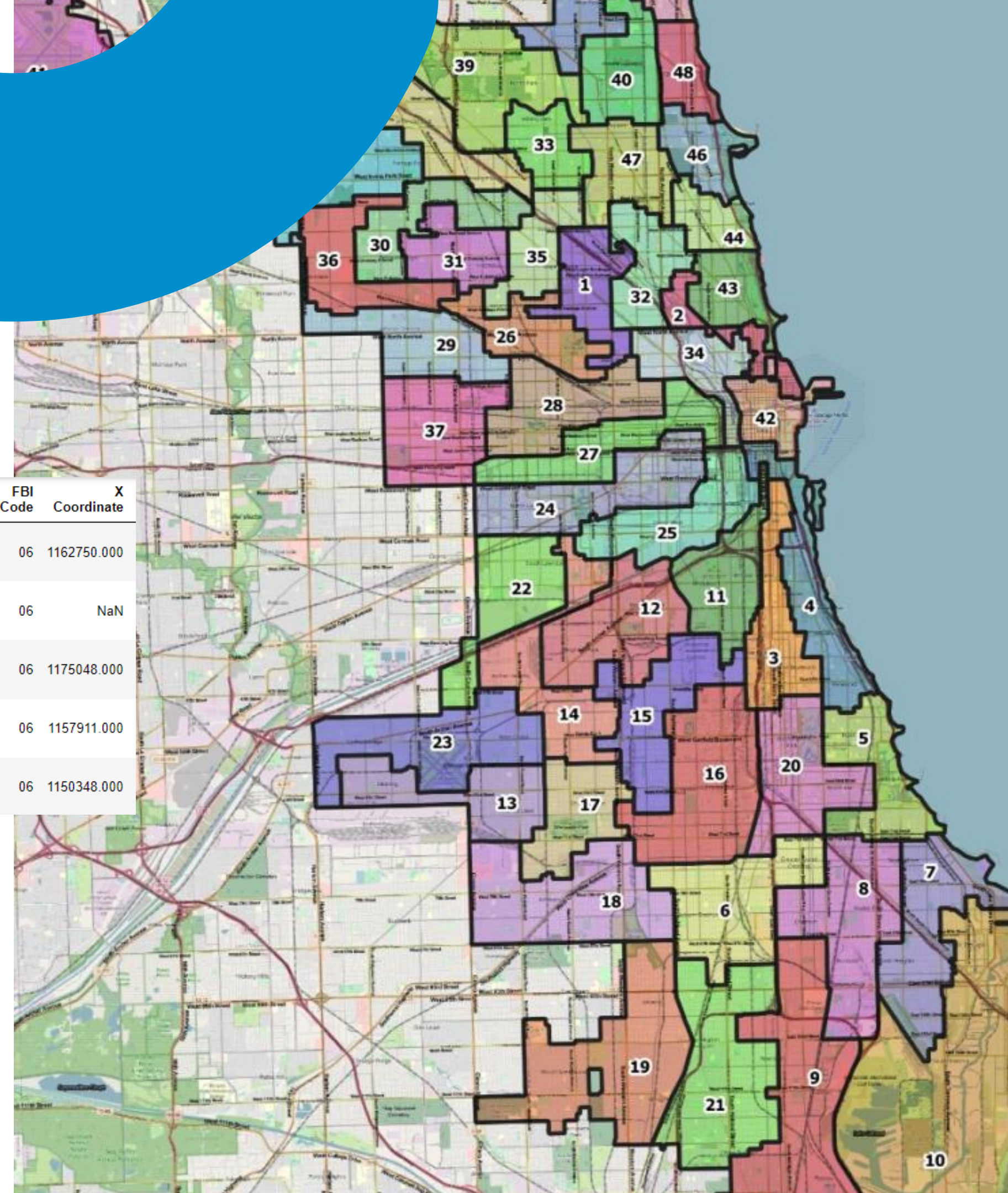


DATASET

- Recopilación de datos sobre los crímenes en Chicago en el periodo de 2001 a 2017.

ID	Case Number	Date	Block	IUCR	Primary Type	Description	Location Description	Arrest	Domestic	Beat	District	Ward	Community Area	FBI Code	X Coordinate
4835774	HM440784	01/01/2001 12:00:00 AM	022XX W 80TH ST	0842	THEFT	AGG: FINANCIAL ID THEFT	OTHER	False	False	835	8.000	18.000	70.000	06	1162750.000
4838005	HM450002	08/01/2004 12:00:00 PM	036XX N NOTTINGHAM AVE	0841	THEFT	FINANCIAL ID THEFT:\$300 &UNDER	RESIDENCE	False	False	1632	16.000	36.000	17.000	06	NaN
4722424	HM325827	10/01/2003 12:00:00 AM	002XX W 22ND PL	0842	THEFT	AGG: FINANCIAL ID THEFT	RESIDENCE	False	False	2111	9.000	25.000	34.000	06	1175048.000
4724079	HM328405	08/03/2003 12:01:00 AM	051XX S FRANCISCO AVE	0840	THEFT	FINANCIAL ID THEFT: OVER \$300	RESIDENCE	False	False	911	9.000	14.000	63.000	06	1157911.000
4724354	HM328186	07/01/2003 12:01:00 AM	022XX N AVERS AVE	0841	THEFT	FINANCIAL ID THEFT:\$300 &UNDER	APARTMENT	False	False	2525	25.000	30.000	22.000	06	1150348.000

Y Coordinate	Year	Updated On	Latitude	Longitude	Location
1851604.000	2001.000	07/10/2006 03:52:16 AM	41.748	-87.679	(41.748448075, -87.679202816)
NaN	2004.000	08/17/2015 03:03:40 PM	NaN	NaN	NaN
1889495.000	2003.000	04/15/2016 08:55:02 AM	41.852	-87.633	(41.852159338, -87.633007505)
1870458.000	2003.000	04/15/2016 08:55:02 AM	41.800	-87.696	(41.8002858, -87.69642307)
1914512.000	2003.000	04/15/2016 08:55:02 AM	41.921	-87.723	(41.921325664, -87.723010704)





ANÁLISIS DEL DATASET

Case Number: Identificador único para cada crimen.

Primary Type: Tipo de crimen cometido.

Location Description: Descripción de la ubicación del incidente.

Arrest: Bool que indica si hubo arresto o no.

Domestic: Bool que indica si el crimen ocurrió en un domicilio o no.

Beat: Barrio donde ocurrió el incidente.

District: Indica el distrito policial donde ocurrió el incidente.

Ward: Distrito del ayuntamiento donde ocurrió el incidente.

Community Area: Indica el área de la comunidad.

Year: Año en el que se produjo el crimen.



PREPROCESSING

Unnamed: 0	0
ID	0
Case Number	7
Date	0
Block	0
IUCR	0
Primary Type	0
Description	0
Location Description	1990
Arrest	0
Domestic	0
Beat	0
District	91
Ward	700224
Community Area	702091
FBI Code	0
X Coordinate	105573
Y Coordinate	105573
Year	0
Updated On	0
Latitude	105573
Longitude	105574
Location	105574
dtype: int64	
(7941282, 23)	

10,24% de valores NaN



ID	0
Case Number	0
Date	0
Block	0
IUCR	0
Primary Type	0
Description	0
Location Description	0
Arrest	0
Domestic	0
Beat	0
District	0
Ward	0
Community Area	0
FBI Code	0
X Coordinate	0
Y Coordinate	0
Year	0
Updated On	0
Latitude	0
Longitude	0
Location	0
dtype: int64	
(7145213, 22)	



PREPROCESSING

1.666.884 valores duplicados

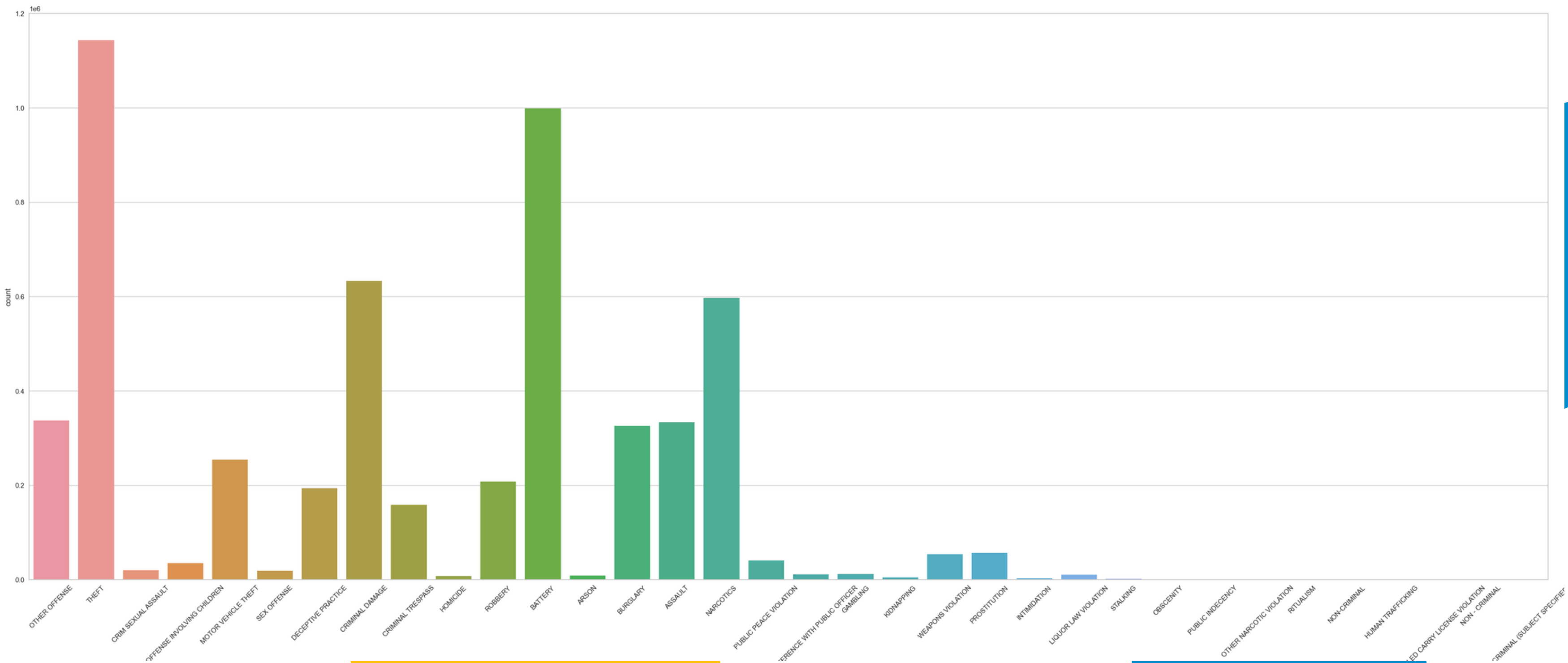
Un total de **31,014%** de valores eliminados entre Nan y valores duplicados

	ID	Date	Primary Type	Location Description	Arrest	Domestic	Beat	District	Ward	Community Area	X Coordinate	Y Coordinate	Year	Latitude	Longitude
0	4676906	2003-03-01 00:00:00	0	0	False	True	935	9.000	11.000	61.000	1173974.000	1876757.000	2003.000	41.817	-87.637
1	4677901	2003-05-01 01:00:00	1	0	False	False	2233	22.000	34.000	49.000	1174948.000	1831051.000	2003.000	41.692	-87.635
2	4679521	2003-03-15 00:00:00	0	1	False	False	2222	22.000	21.000	73.000	1169911.000	1844832.000	2003.000	41.730	-87.653
3	4680124	2003-01-01 00:00:00	1	0	False	False	1134	11.000	24.000	29.000	1154521.000	1895755.000	2003.000	41.870	-87.708
4	4682244	2003-06-05 00:00:00	1	0	False	False	2223	22.000	21.000	73.000	1173946.000	1842903.000	2003.000	41.724	-87.638

Dataset resultante: +5 millones de filas



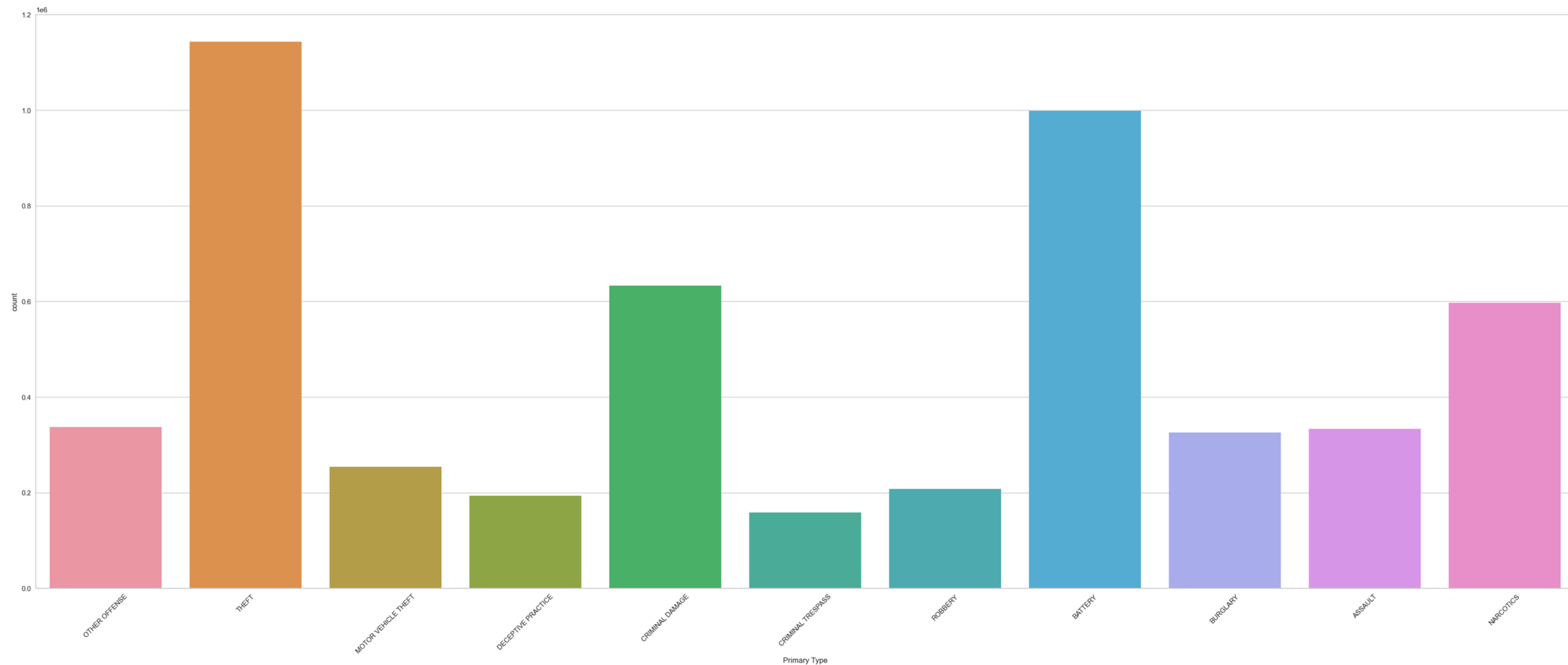
PREPROCESSING - VISUALIZACIÓN DE LOS DATOS





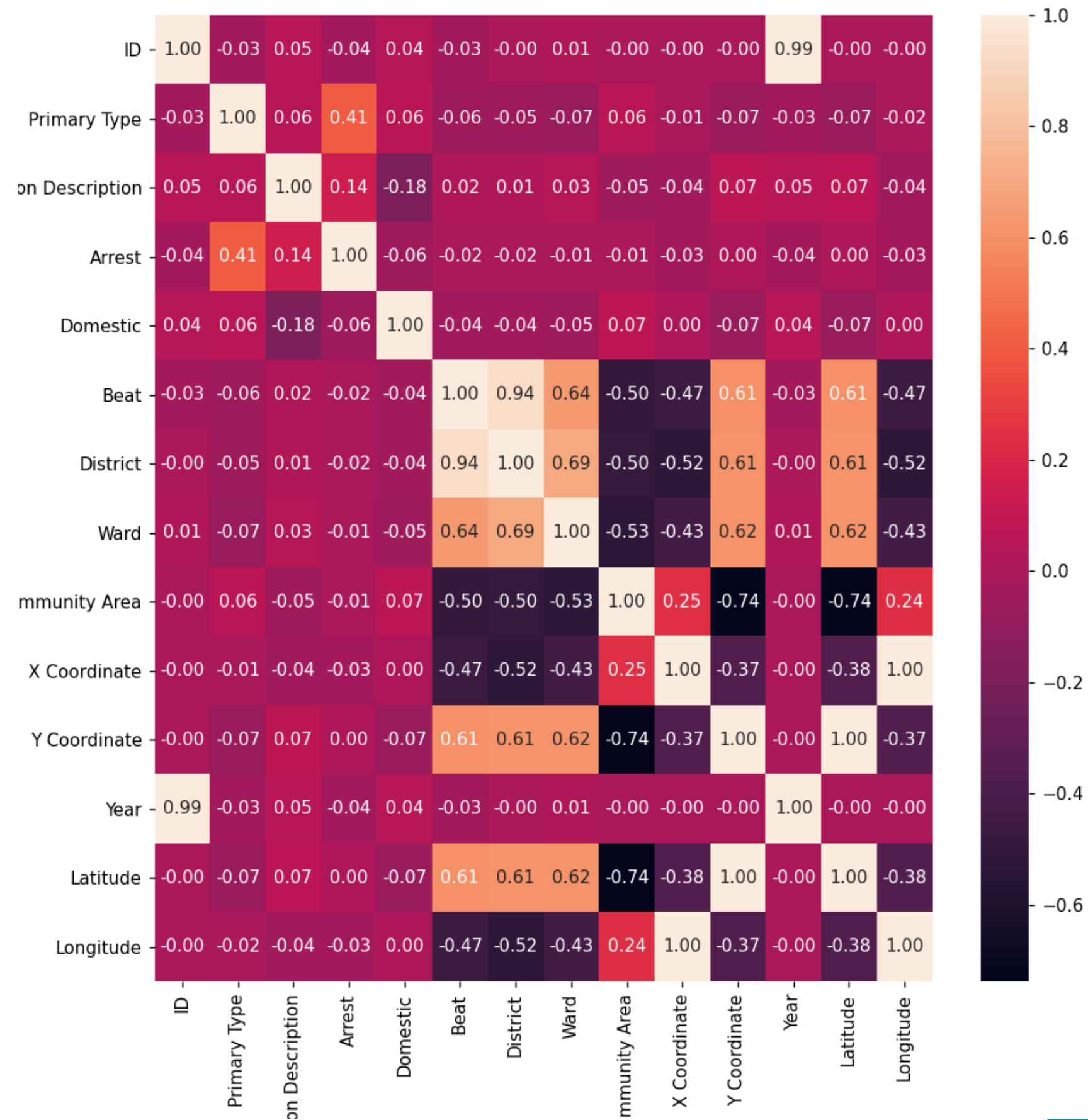
PREPROCESSING - VISUALIZACIÓN DE LOS DATOS

Tipos de crimen con ocurrencia > 100k



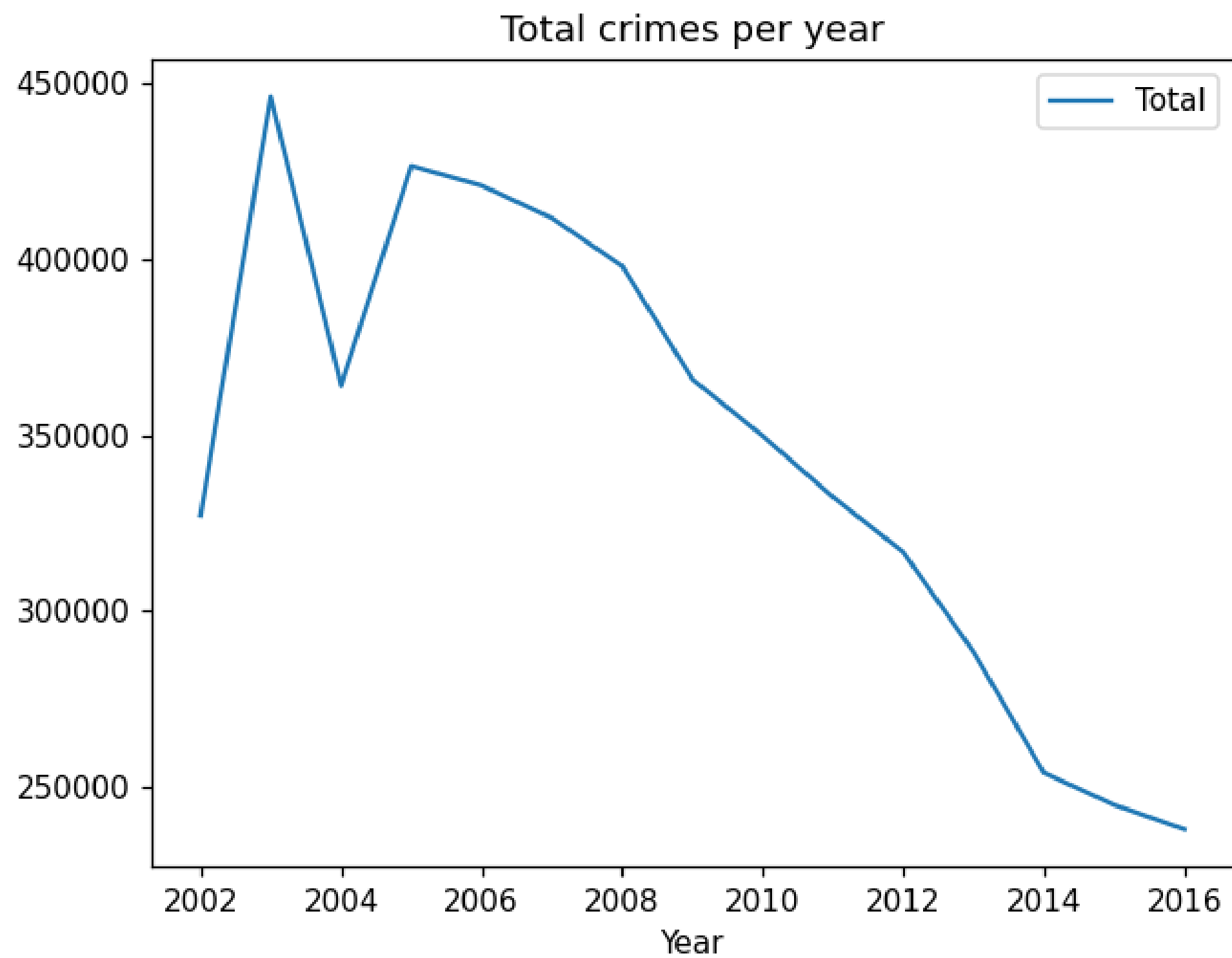


PREPROCESSING - NORMALIZACIÓN DE LOS DATOS





PREPROCESSING - VISUALIZACIÓN DE LOS DATOS

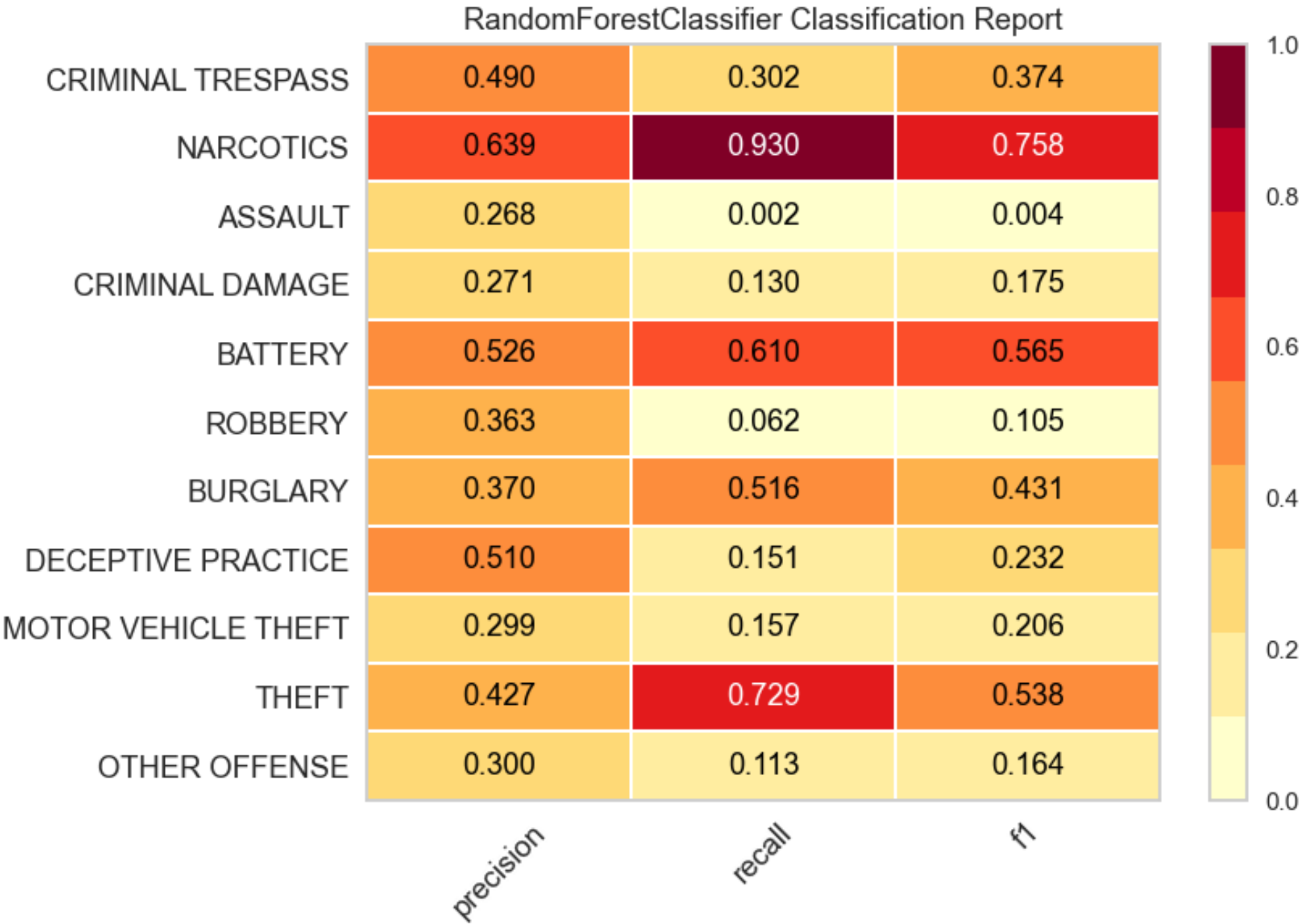




CLASIFICACIÓN

Random Forest Classifier

Accuracy: 46,5%

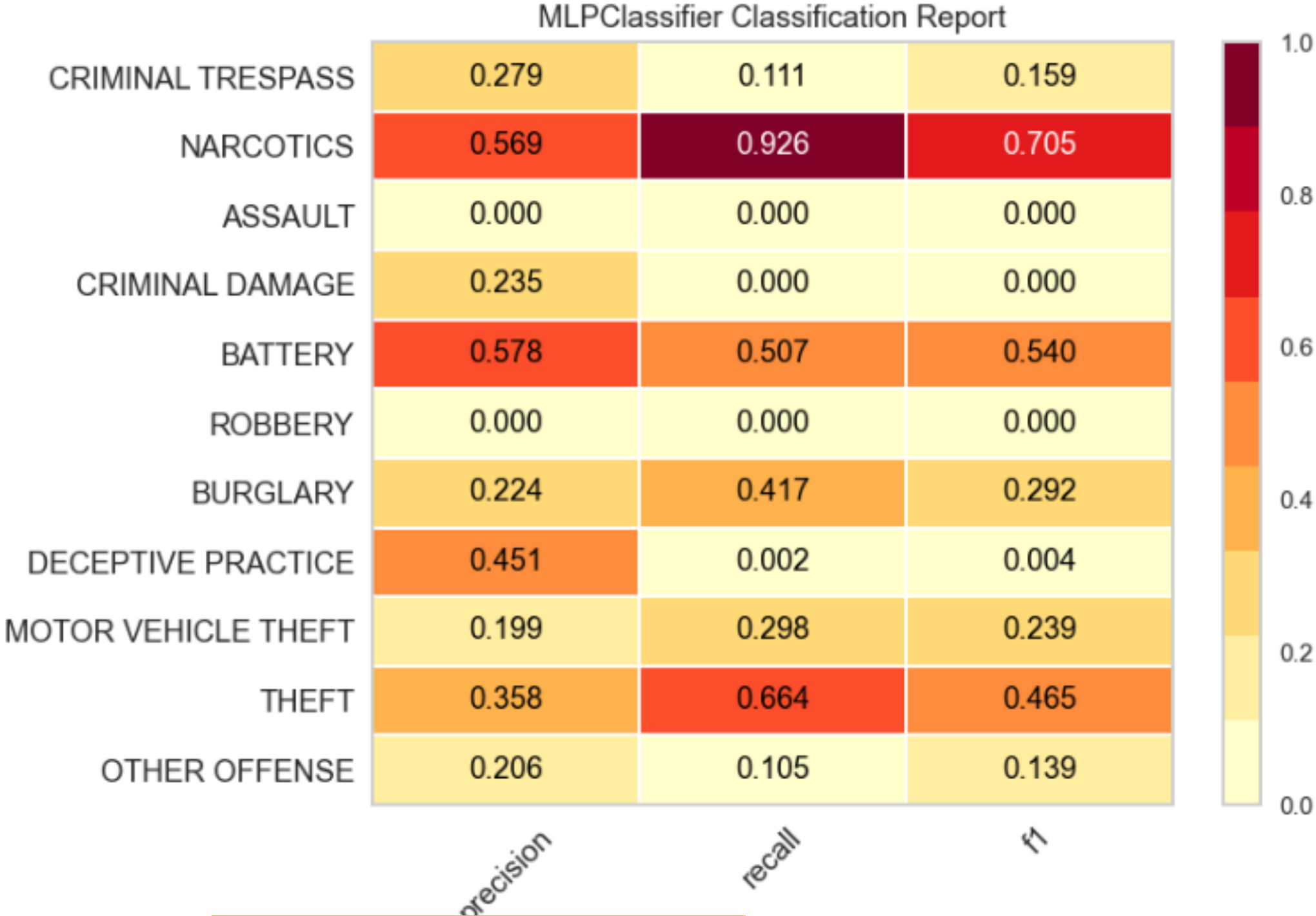




CLASIFICACIÓN

Multi-Layer Perceptron Classifier

Accuracy: 40,12%





CLASIFICACIÓN

KNN Classifier

Accuracy: 37,58%

KNeighborsClassifier Classification Report				<div><div></div><div>1.0</div><div>0.8</div><div>0.6</div><div>0.4</div><div>0.2</div><div>0.0</div></div>
CRIMINAL TRESPASS	0.422	0.199	0.271	
NARCOTICS	0.709	0.723	0.716	
ASSAULT	0.126	0.034	0.053	
CRIMINAL DAMAGE	0.221	0.119	0.155	
BATTERY	0.472	0.435	0.453	
ROBBERY	0.193	0.133	0.157	
BURGLARY	0.311	0.281	0.295	
DECEPTIVE PRACTICE	0.252	0.215	0.232	
MOTOR VEHICLE THEFT	0.184	0.185	0.184	
THEFT	0.398	0.581	0.473	
OTHER OFFENSE	0.149	0.279	0.194	
	precision	recall	f1	



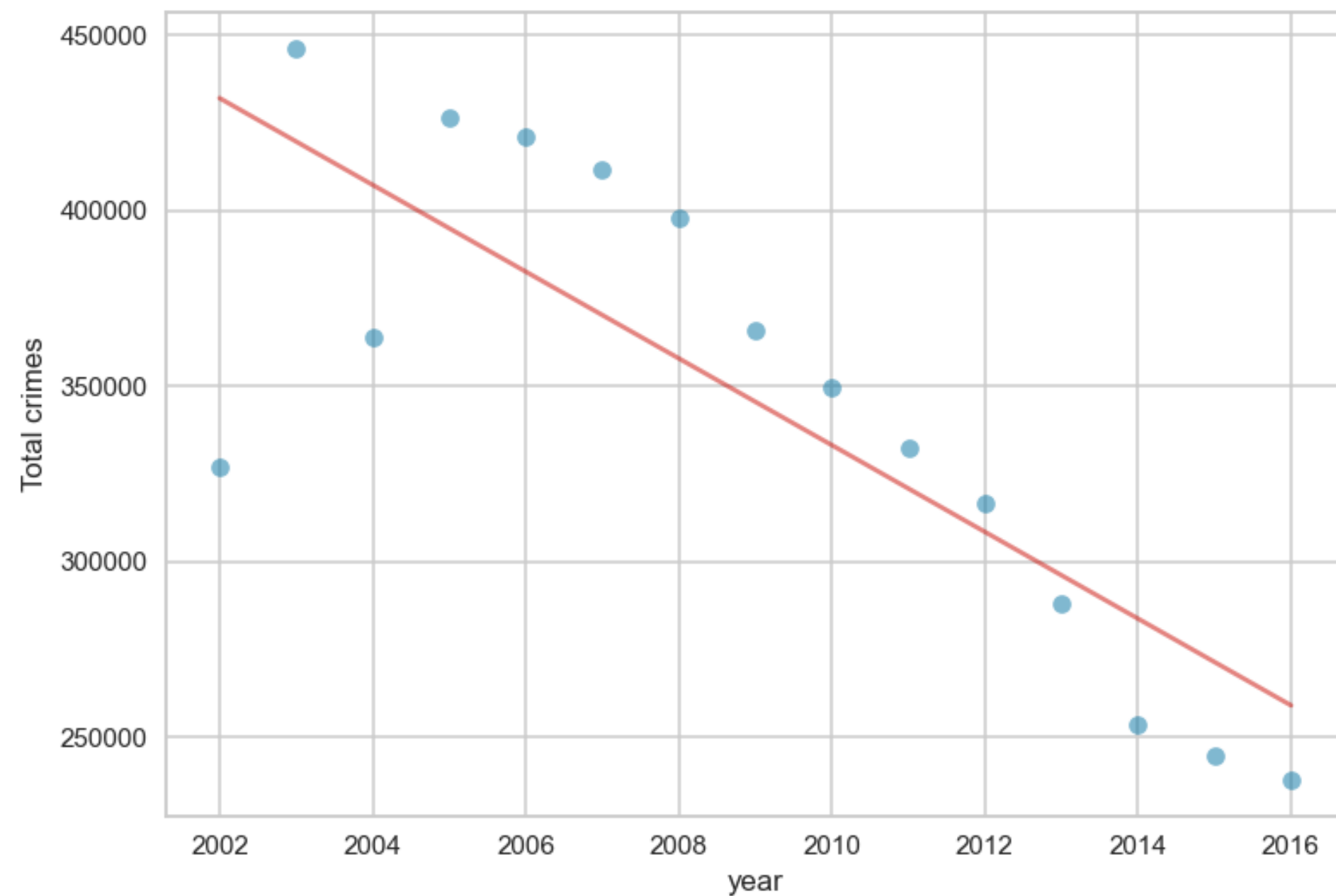
REGRESIÓN

Linear Regression

MSE: 1491641528.5362532

R^2 : 0.47585501834335653

VAR: 4337498268.0622225





**GRACIAS POR
VUESTRA ATENCIÓN**

The background features two large, decorative curved shapes. A blue shape is in the top right corner, and a yellow shape is in the bottom left corner.

**GITHUB: [HTTPS://GITHUB.COM/ADRIAN-
NIETO/PRACTICA3-APC-KAGGLE.GIT](https://github.com/Adrian-Nieto/practica3-APC-kaggle.git)**