Project 4 - Machine Learning with Crime Data

By: Brett Squires, Samuel Haynes, Sabar Mouri & Grace Bolin

Predictive Modeling on Crime Data in Kansas City, Missouri

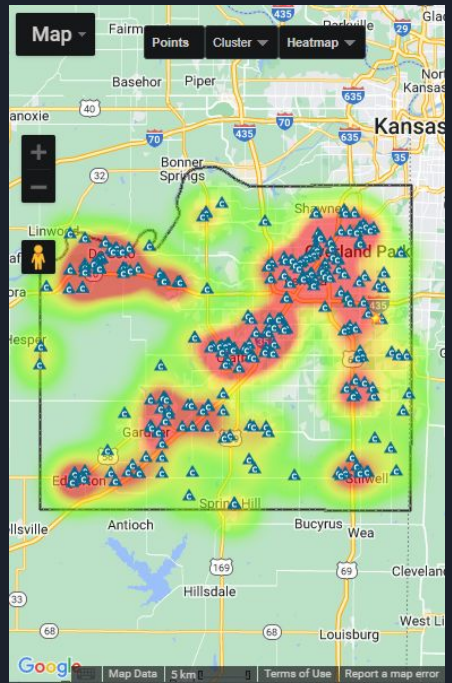
Summary: Our group developed a predictive model for identifying a class type/description of a crime that is most likely to occur based on variables such as location, zip code, domestic violence involvement, patrol district, and firearm usage. This information and tool could be valuable by enhancing community safety and resource allocation. Forecasting a type of crime based on these factors can help law enforcement agencies to strategically deploy resources to areas most at risk, proactively deterring crime. This data could also help urban planners and policymakers to make informed decisions to allocate resources for social programs that target root causes of crime making a safer and more secure community for everyone.

Our questions:
Where does crime happen most often in Kansas City, MO?
What type of crime does the model predict will happen in certain areas based on our variables?

Data Source: Search & Browse Crime | Page 1 of 3 | Open Data KC | data.kcmo.org

Dataset: Johnson County Sheriff's Office

ID	EventType	Class	UCRDescri	IncidentD	TheftDesc	Location	PremiseD	APID	UCRCode	Agency	ReptDate	ReptTime	Y	X
D2300308(	crimes	Property	Counterfe	FORGERY; POSSESS V		100 BLOCK	Jail / Prisc	2.00E+17	250	JCSO	9/4/2023	832	38.88317	-94.8212
D2300307-	crimes	Society	Weapon L	UNLAWFUL DISCHAR		24800 BLO	Single Res	2.00E+17	520	JCSO	9/3/2023	1948	38.89399	-94.8735
D2300308]	crimes	Society	All Other (	PROBATION VIOLATI		100 BLOCK	Jail / Prisc	2.00E+17	90Z	JCSO	9/3/2023	1440	38.88317	-94.8212

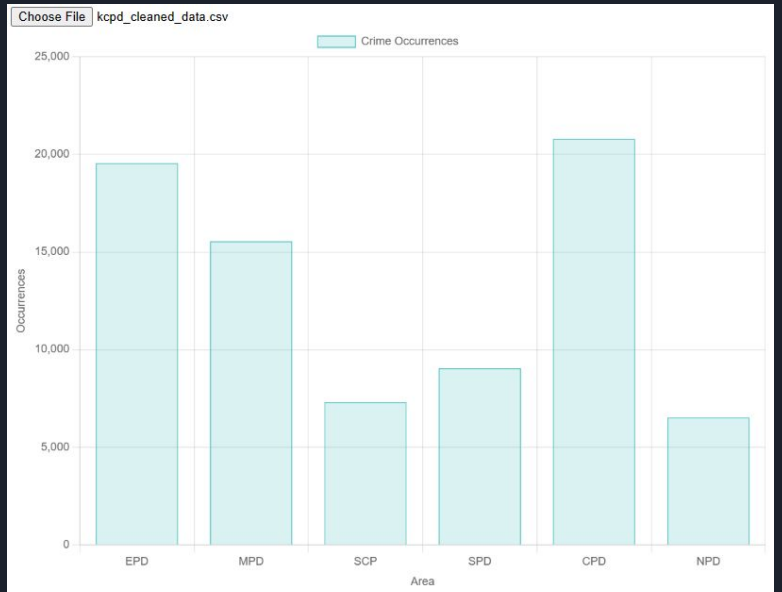


- Heatmap of Crime in Johnson County
- Lat and Long needed converted to Zip Codes
- Reverse Geocoding with Geopy
- Rate limitations and errors

Dataset: KCPD Crime Data 2021

Report_No	Reported_Date	From_Date	From_Time	Offense	IBRS	Description	Beat	Address	Zip Code	Area	DVFlag	Involvement	Firearm Used Flag
KC21055624	8/21/2021	8/21/2021	19:45	Murder	09A	Murder	332	3500 AGNES AVE	64128	EPD	FALSE	VIC	TRUE
KC21012401	2/24/2021	2/24/2021	11:35	Stealing from Building/Residence	23D	Theft From Building	234	8800 STATE LINE RD	64114	MPD	FALSE	SUS	FALSE
KC21017210	3/17/2021	3/17/2021	15:35	Assault (Aggravated)	13A	Aggravated Assault	631	NE PARVIN RD and I 29 HWY	64116	SCP	FALSE	ARR CHA	TRUE
KC21010791	2/17/2021	2/11/2021	20:30	Stolen Auto	240	Motor Vehicle Theft	211	4200 E 53RD ST	64130	MPD	FALSE	SUS	FALSE
KC21012025	2/22/2021	2/22/2021	17:55	Assault (Aggravated)	13A	Aggravated Assault	222	E 52ND ST and EUCLID AVE	64130	MPD	FALSE	VIC	FALSE
KC21001020	1/5/2021	1/5/2021	19:40	Domestic Violence Assault (Non-A	13B	Simple Assault	534	5100 LONGVIEW RD	64134	SPD	TRUE	CHA SUS	FALSE
KC21003742	1/17/2021	1/17/2021	21:30	Assault (Aggravated)	13A	Aggravated Assault	541	6400 E 87TH ST	64138	SPD	FALSE	SUS	TRUE
KC21009520	2/11/2021	2/1/2021	20:00	Stolen Auto	240	Motor Vehicle Theft	542	9400 NEWTON DR	64134	SPD	FALSE	VIC	FALSE
KC21004380	1/20/2021	1/20/2021	16:44	Assault (Aggravated)	13A	Aggravated Assault	322	E 12TH ST and HARDESTY AVE	64127	EPD	FALSE	VIC	TRUE
KC21005430	1/25/2021	4/13/2020	12:00	Forgery	250	Counterfeiting / Forgery	332	3200 PROSPECT AVE	64128	EPD	FALSE	SUS	FALSE
KC21008041	2/5/2021	12/19/2020	12:00	Identity Theft	26F	Identity Theft	641	8100 N LAWN AVE	64119	SCP	FALSE	CMP VIC	FALSE
KC21005267	1/24/2021	1/24/2021	17:27	Domestic Violence Assault (Non-A	13B	Simple Assault	231	7600 WYANDOTTE ST	64114	MPD	TRUE	ARR CHA SUS	FALSE
KC21011417	2/20/2021	2/20/2021	1:38	Robbery (Armed Street)	120	Robbery	134	700 W 44TH ST	64111	CPD	FALSE	SUS	TRUE
KC21011702	2/21/2021	2/20/2021	17:00	Stolen Auto	240	Motor Vehicle Theft	636	4500 NE WORLDS OF FUN DR	64161	SCP	FALSE	VIC	FALSE
KC21012637	2/25/2021	2/4/2021	11:45	Identity Theft	26F	Identity Theft	123	3000 PASEO	64109	CPD	FALSE	VIC	FALSE
KC21010718	2/17/2021	2/17/2021	0:04	Murder	09A	Murder	543	11300 SYCAMORE TER	64134	SPD	FALSE	VIC DEC	TRUE
KC21012022	2/22/2021	2/22/2021	17:40	Stealing â€“ Shoplift	23C	Shoplifting	113	00 E 13TH ST	64106	CPD	FALSE	VIC	FALSE
KC21006497	1/29/2021	1/29/2021	15:45	Stealing â€“ Shoplift	23C	Shoplifting	345	11600 E US 40 HWY E	64133	EPD	FALSE	VIC	FALSE
KC21015438	3/9/2021	3/9/2021	13:09	Soliciting Prostitution	40A	Prostitution	345	4000 BLUE RIDGE CTOF	64133	EPD	FALSE	SUS	FALSE
KC21009523	2/11/2021	2/7/2021	18:00	Stolen Auto	240	Motor Vehicle Theft	323	3200 E 21ST ST	64127	EPD	FALSE	SUS	FALSE
KC21013729	3/2/2021	3/1/2021	21:30	Stealing from Auto (Theft from Au	23F	Theft From Motor Vehicle	423	9200 N JARBOE ST	64155	NPD	FALSE	VIC	FALSE
KC21009495	2/11/2021	2/3/2021	10:56	Embezzlement	270	Embezzlement	234	8500 STATE LINE RD	64114	MPD	FALSE	VIC	FALSE
KC21010861	2/17/2021	2/17/2021	15:52	Assault (Non-Aggravated)	13B	Simple Assault	143	3900 TROOST AVE	64110	CPD	FALSE	VIC	FALSE
KC21013274	2/28/2021	2/28/2021	5:19	Murder	09A	Murder	411	100 NW HARLEM RD	64116	NPD	FALSE	ARR CHA SUS	TRUE
KC21009313	2/10/2021	2/10/2020	14:10	Stealing â€“ Shoplift	23C	Shoplifting	345	11600 E US 40 HWY E	64133	EPD	FALSE	VIC	FALSE
KC21006619	1/30/2021	1/30/2021	2:57	Murder	13A	Aggravated Assault	322	E 19TH ST and CYPRESS AVE	64127	EPD	FALSE	VIC	TRUE
KC21012281	2/23/2021	2/23/2021	18:30	Stolen Auto	240	Motor Vehicle Theft	641	9900 NE BARRY RD	64158	SCP	FALSE	VIC	FALSE
KC21011079	2/18/2021	2/18/2021	16:30	Patronizing Prostitution	40A	Prostitution	345	4000 BLUE RIDGE CTOF	64133	EPD	FALSE	ARR CHA SUS	FALSE
KC21009888	2/12/2021	2/12/2021	22:55	Domestic Violence Assault (Non-A	13B	Simple Assault	142	3900 PARK AVE	64130	CPD	FALSE	VIC SUS	FALSE
KC21005366	1/25/2021	1/22/2021	15:30	Stolen Auto	240	Motor Vehicle Theft	221	E 49TH ST and WALNUT ST	64112	MPD	FALSE	VIC	FALSE
KC21010754	2/17/2021	2/15/2021	18:00	Stolen Auto	240	Motor Vehicle Theft	342	8900 E US 40 HWY	64129	EPD	FALSE	VIC	FALSE
KC21013898	3/2/2021	3/2/2021	18:30	Assault (Aggravated)	13A	Aggravated Assault	532	I 435 HWY NE and US 71 HWY	64137	SPD	FALSE	SUS	TRUE

Visualization of the Data



AREAS - DIVISION OFFENSE OCCURRED IN

EPD - EAST

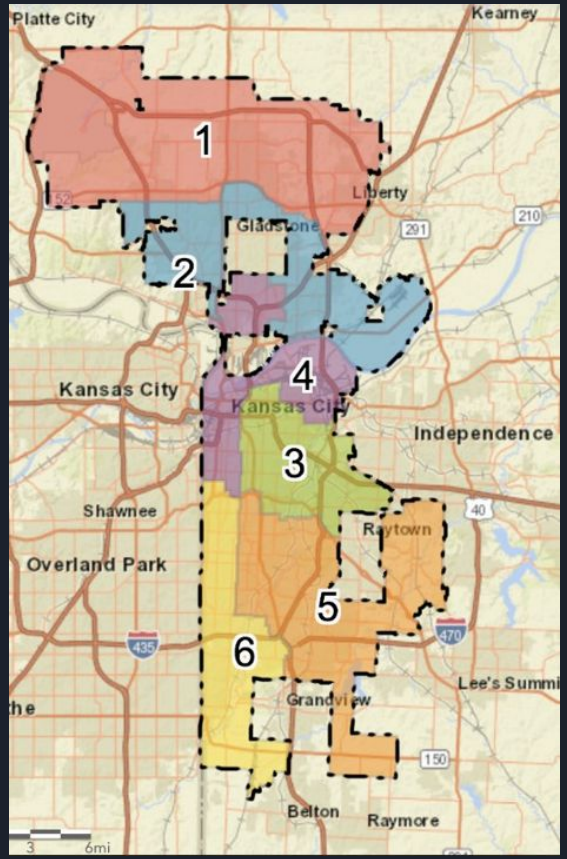
MPD - METRO

SCP - SHOAL CREEK

SPD - SOUTH

CPD - CENTRAL

NPD - NORTH



Map Sources: KCMO City Council 1st District in-district candidates weigh in on key issues (kshb.com)

2017annual.pdf (kcpd.org)

Cleaning the data

```
path = 'KCPD_Crime_Data_2021.csv'
dtypes = {
    'Beat': 'int',
    'Zip Code': 'int',
    'Area': 'string',
    'DVFlag': 'bool',
    'Firearm Used Flag': 'bool',
    'Description': 'string'
}
```

- Removed unnecessary columns in Excel
- Declare data types

Importing data with SQLalchemy

```
#create database for table import
import psycopg2
import json
#change database creds to postgres db
with open('create_database_creds.json') as O:
    config = json.load(O)
username = config['user']
password = config['password']
db_host = config['host']
db_port = config['port']
db_ = 'crime_data_db'
conn = psycopg2.connect(
    host=db_host,
    port=db_port,
    user=username,
    password=password,
)
conn.autocommit = True
cursor = conn.cursor()
#create database to house table
cursor.execute(f"CREATE DATABASE {db_}")
cursor.close()
conn.close()
```

- Connect to Postgres with Psycopg2
- Database automatically created

SQL Database

Query | Query History

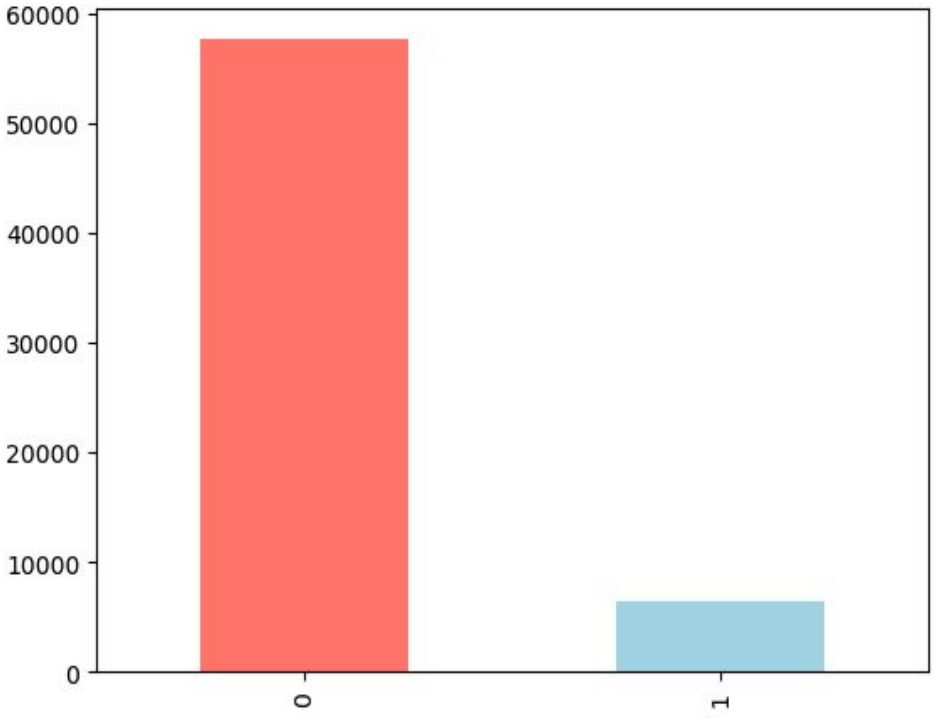
```
1  SELECT * FROM crime_data
```

Data Output | Messages | Notifications

	Description text	Beat text	Zip Code text	Area text	DVFlag boolean	Firearm Used Flag boolean
1	Murder	332	64128	EPD	false	true
2	Theft From Building	234	64114	MPD	false	false
3	Aggravated Assault	631	64116	SCP	false	true

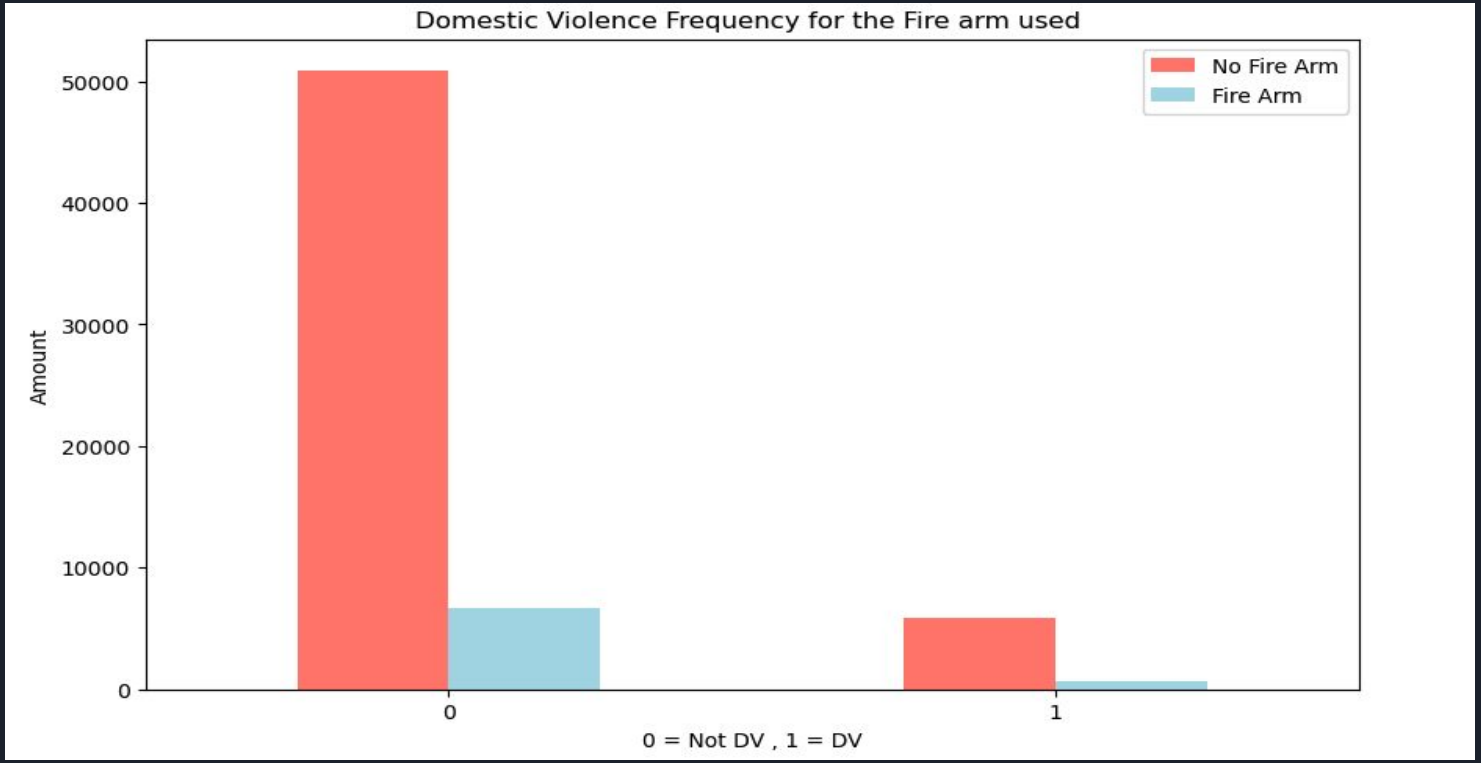
Plotting Domestic Violence Flag

In [43]:
```
# Ploting to understand the variable
df["DVFlag"].value_counts().plot(kind = "bar", color = ["salmon", "lightblue"]);
```



Plot of Crosstab

Firearm Used Flag	0	1
DVFlag		
0	50913	6685
1	5885	610



Model Setup (Data)

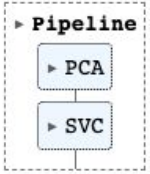
- Categorical Data
- pd.get_dummies
  - Many rows of data

	Area_CPD	Area_EPD	Area_MPD	Area_NPD	Area_SCP	Area_SPD	Zip Code_50023	Zip Code_6106	Zip Code_61408	Zip Code_61441	...	Zip Code_64165	Zip Code_64166
0	0	1	0	0	0	0	0	0	0	0	...	0	0
1	0	0	1	0	0	0	0	0	0	0	...	0	0
2	0	0	0	0	1	0	0	0	0	0	...	0	0
4	0	0	1	0	0	0	0	0	0	0	...	0	0
5	0	0	1	0	0	0	0	0	0	0	...	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
92121	1	0	0	0	0	0	0	0	0	0	...	0	0
92122	1	0	0	0	0	0	0	0	0	0	...	0	0
92123	1	0	0	0	0	0	0	0	0	0	...	0	0
92124	0	1	0	0	0	0	0	0	0	0	...	0	0
92125	1	0	0	0	0	0	0	0	0	0	...	0	0

Model Setup (Predictions)

- Used the PCA and SVC pipeline
  - Standardization
- Categorical NB/Naive Bayes
- Predicted the class of crime
  - Others (Cont,.)
- More research than coding

```
from sklearn.pipeline import Pipeline
from sklearn.svm import SVC
from sklearn.decomposition import PCA
estimators = [('reduce_dim', PCA()), ('clf', SVC())]
pipe = Pipeline(estimators)
pipe
Pipeline(steps=[('reduce_dim', PCA()), ('clf', SVC())])
```




```
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import CategoricalNB

from sklearn.naive_bayes import CategoricalNB
clf = CategoricalNB(force_alpha=True)
clf.fit(X_train, y_train)
```

```
CategoricalNB
CategoricalNB(force_alpha=True)
```

Initial Attempts

```
6877            All Other Offenses
45050                       Murder
34694          Theft From Building
65139               Simple Assault
78704           Aggravated Assault
                    ...
65557                       Robbery
74961               Simple Assault
60879    Theft From Motor Vehicle
62252             All Other Larceny
3999     Theft From Motor Vehicle
Name: Description, Length: 51274, dtype: object
```

```python
from sklearn.metrics import (
    accuracy_score,
    confusion_matrix,
    ConfusionMatrixDisplay,
    f1_score,
)

y_pred = gnb.predict(X_test)
accuray = accuracy_score(y_pred, y_test)
f1 = f1_score(y_pred, y_test, average="weighted")

print("Accuracy:", accuray)
print("F1 Score:", f1)
```

```
Accuracy: 0.06316532970082236
F1 Score: 0.046057757115858404
```

64126	0.97	1.00	0.99	190
64127	1.00	1.00	1.00	680
64128	0.99	1.00	1.00	551
64129	0.99	1.00	1.00	349
64130	1.00	1.00	1.00	843
64131	1.00	1.00	1.00	509
64132	0.99	1.00	0.99	586
64133	1.00	1.00	1.00	681
64134	0.99	1.00	1.00	502
64136	1.00	0.98	0.99	46
64137	1.00	1.00	1.00	222
64138	1.00	1.00	1.00	273
64139	0.92	1.00	0.96	22
64145	1.00	1.00	1.00	110
64146	1.00	1.00	1.00	16
64147	0.99	1.00	0.99	67
64149	0.00	0.00	0.00	2
64151	1.00	0.99	0.99	196
64152	0.99	1.00	0.99	75
64153	0.99	1.00	1.00	224
64154	1.00	1.00	1.00	227
64155	1.00	1.00	1.00	207
64156	0.92	1.00	0.96	49
64157	0.99	0.98	0.99	115
64158	0.96	0.99	0.98	80
64159	0.00	0.00	0.00	1
64161	0.99	1.00	1.00	101
64163	1.00	0.33	0.50	3
64165	0.00	0.00	0.00	5
64166	0.00	0.00	0.00	1
64217	0.00	0.00	0.00	1
64218	0.00	0.00	0.00	1
64318	0.00	0.00	0.00	1
65129	0.00	0.00	0.00	1
66126	0.00	0.00	0.00	1
accuracy			1.00	12819
macro avg	0.74	0.72	0.73	12819
weighted avg	0.99	1.00	1.00	12819

Further Analysis

- Changed prediction criteria
- Predicted other variables
  - Yielder better results
- Zip code, police department, patrol area
- Continued to train for our initial goal of predicting crime

64166	0.00	0.00	0.00	1
64217	0.00	0.00	0.00	1
64218	0.00	0.00	0.00	1
64318	0.00	0.00	0.00	1
65129	0.00	0.00	0.00	1
66126	0.00	0.00	0.00	1
accuracy			0.71	12819
macro avg	0.43	0.44	0.43	12819
weighted avg	0.72	0.71	0.71	12819

Accuracy: 0.7125360792573524
F1 Score: 0.717079427012908

CPD	1.00	1.00	1.00	3416
EPD	1.00	1.00	1.00	3167
MPD	1.00	1.00	1.00	2492
NPD	1.00	1.00	1.00	1094
SCP	1.00	1.00	1.00	1188
SPD	1.00	1.00	1.00	1462
accuracy			1.00	12819
macro avg	1.00	1.00	1.00	12819
weighted avg	1.00	1.00	1.00	12819

Accuracy: 0.9981277790779312
F1 Score: 0.9981278483264036

Final Score

Accuracy: 0.9777777777777777
F1 Score: 0.9778106995884774

	precision	recall	f1-score	support
Aggravated Assault	1.00	1.00	1.00	1684
All Other Larceny	1.00	1.00	1.00	641
All Other Offenses	1.00	1.00	1.00	707
Animal Cruelty	0.00	0.00	0.00	3
Arson	0.96	1.00	0.98	44
Assisting or Promoting Prostitution	0.00	0.00	0.00	1
Bad Checks	0.00	0.00	0.00	1
Disorderly Conduct	1.00	1.00	1.00	228
Driving Under the Influence	0.99	1.00	0.99	200
Drug Equipment Violations	1.00	1.00	1.00	44
Embezzlement	0.98	1.00	0.99	60
Family Offenses, Nonviolent	1.00	1.00	1.00	107
Forcible Fondling	1.00	1.00	1.00	58
Forcible Rape	1.00	1.00	1.00	112
Forcible Sodomy	1.00	1.00	1.00	63
Identity Theft	1.00	1.00	1.00	109
Impersonation	1.00	1.00	1.00	3
Incest	0.00	0.00	0.00	1
Intimidation	1.00	1.00	1.00	362
Liquor Law Violations	1.00	0.75	0.86	4
Motor Vehicle Theft	1.00	1.00	1.00	1261
Murder	1.00	1.00	1.00	64
Pocket-Picking	1.00	1.00	1.00	7
Prostitution	1.00	1.00	1.00	9
Purchasing Prostitution	1.00	0.50	0.67	6
Purse-Snatching	1.00	0.92	0.96	13
Robbery	1.00	1.00	1.00	608
Shoplifting	1.00	1.00	1.00	843
Simple Assault	1.00	1.00	1.00	2202
Statutory Rape	1.00	1.00	1.00	20
Stolen Property Offenses	1.00	1.00	1.00	233
Theft From Building	1.00	1.00	1.00	412
Theft From Coin-Operated Machine	1.00	0.50	0.67	2
Theft From Motor Vehicle	1.00	1.00	1.00	1199
Theft of Vehicle Parts and Accessories	1.00	1.00	1.00	688
Trespass of Real Property	1.00	1.00	1.00	618
Weapon Law Violations	1.00	1.00	1.00	131
Welfare Fraud	0.00	0.00	0.00	2
Wire Fraud	1.00	1.00	1.00	69
accuracy			1.00	12819
macro avg	0.87	0.84	0.85	12819
weighted avg	1.00	1.00	1.00	12819

Q&A