

# Appendix

December 12, 2019

## 1 References

- 1 Robertson, C. (2019, 25 April). Crime Is Down, Yet U.S. Incarceration Rates Are Still Among the Highest in the World. *The New York Times*, <https://www.nytimes.com/2008/04/23/world/americas/23iht-23prison.12253738.html>. Accessed 12 December, 2019.
- 2 Liptak, A. (2008, 23 April). U.S. prison population dwarfs that of other nations. *The New York Times*, <https://www.nytimes.com/2008/04/23/world/americas/23iht-23prison.12253738.html>. Accessed 12 December, 2019.
- 3 Bureau of Justice Statistics. Data Collection: Annual Survey of Jails (ASJ). *Bureau of Justice Statistics*, <https://www.bjs.gov/index.cfm?ty=dcdetail&iid=261>. Accessed 18 November, 2019.
- 4 Bureau of Justice Statistics. Data Collection: Census of Jail Inmates. *Bureau of Justice Statistics*, <https://www.bjs.gov/index.cfm?ty=dcdetail&iid=404>. Accessed 18 November, 2019.
- 5 Bureau of Justice Statistics. Data Collection: Census of State and Federal Adult Correctional Facilities (CSFACF). *Bureau of Justice Statistics*, <https://www.bjs.gov/index.cfm?ty=dcdetail&iid=255>. Accessed 18 November, 2019.
- 6 Lansing, M. Sentenced Inmates in Correctional Facilities. *Data.ct.gov*, <https://catalog.data.gov/dataset/sentenced-inmates-in-correctional-facilities>. Accessed 2 December 2019.
- 7 Peter H. Westfall (2014) Kurtosis as Peakedness, 1905–2014. R.I.P., *The American Statistician*, 68:3, 191-195, DOI: 10.1080/00031305.2014.917055. <https://www.tandfonline.com/doi/full/10.1080/00031305.2014.917055?scroll=top&needAccess=true>
- 8 The Legacy Museum: From Enslavement to Mass Incarceration. *ejl.org*. <https://museumandmemorial.ejl.org/museum>. Accessed 12 December, 2019.

## 2 Data Preparation

### 2.1 Cleaning

The following is some code that was used to clean and organize data that was retrieved from the BJS sources.

```
In [5]: import re
import pandas as pd
import pickle
import json
import numpy as np
```

```

import scipy.linalg as la
import statsmodels.api as sm
from sklearn import linear_model, model_selection, metrics
import sklearn
import plotly.graph_objs as go
import matplotlib.pyplot as plt
import sys
np.set_printoptions(threshold=sys.maxsize)
import warnings
warnings.filterwarnings('ignore')

```

```

In [ ]: # regex things that I want
        comma = re.compile(r',')
        perc = re.compile(r'%')
        make_perc = re.compile(r'(.*)\(%.*%\)\s')
        # make_perc = re.compile(r'hello')
        perc = re.compile(r'(.*)\(\s*([0-9.]*)(%\s*)\s*(.?)')
        fix_dash = re.compile(r'\s*-\s*([A-z\s]*)')
        comma = re.compile(r',')
        u = re.compile(r'(u)([0-9]*)')
        letter = re.compile(r'([\(\)A-z]?)([0-9]*)([\(\)A-z]?)'

```

```

In [ ]: data = pd.read_csv('male_incarceration_data.csv')
        cols = data.columns
        for col in cols[3:]:
            items = data[col]
            for i,item in enumerate(items):
                new_item = float(comma.sub(',', item))
                data[col][i] = new_item
        data.to_csv('male_incarceration_data.csv')

```

```

In [ ]: data = pd.read_csv('incarceration_by_race.csv')
        cols = data.columns
        for col in cols[3:]:
            items = data[col]
            # print(col, items)
            for i,item in enumerate(items):
                # print(type(item))
                if type(item) != int:
                    new_item = float(comma.sub(',', item))
                else:
                    new_item = item
                data[col][i] = new_item
        data.to_csv('incarceration_by_race.csv')

```

```

In [ ]: data = pd.read_csv('female_incarceration_data.csv')
        cols = data.columns
        for col in cols[3:]:
            items = data[col]

```

```

#     print(col, items)
    for i,item in enumerate(items):
#         print(type(item))
        if type(item) != int:
            new_item = float(comma.sub(',', item))
        else:
            new_item = item
            data[col][i] = new_item
data.to_csv('female_incarceration_data.csv')

In [ ]: data = pd.read_csv('incarceration_count.csv')
cols = data.columns
for col in cols[2:]:
    items = data[col]
#     print(col, items)
    for i,item in enumerate(items):
#         print(type(item))
        if type(item) != int and type(item) != float:
#             print(type(item) != float)
            new_item = float(comma.sub(',', item))
        else:
            new_item = item
            data[col][i] = new_item
data.to_csv('incarceration_counts.csv')

```

## 2.2 Feature Engineering

Here is code that I used to engineer different features that I used for regression

```

In [ ]: df = pd.read_csv('Sentenced_Inmates_in_Correctional_Facilities.csv')
df.columns

In [ ]: df = pd.read_csv('Sentenced_Inmates_in_Correctional_Facilities.csv')
df.RACE.replace(
    to_replace={
        'WHITE' : 'WHITE', 'BLACK' : 'BLACK', 'ASIAN' : 'ASIAN'
    }, inplace=True
)
df.to_csv('individuals.csv')

In [ ]: df['LATEST ADMISSION DATE'] = pd.to_datetime(df['LATEST ADMISSION DATE'])
# df['END SENTENCE DATE'] = pd.to_datetime(df['END SENTENCE DATE'])
mask = df['END SENTENCE DATE'] is np.nan
esd = 'END SENTENCE DATE'
nesd = 'N END SENTENCE DATE'
lad = 'LATEST ADMISSION DATE'
sd = 'SENTENCE DAYS'
new_end = np.empty_like(df[esd])
for i in range(len(new_end)):

```

```

to_add = float(df.iloc[i][sd])
if (to_add / 365) > 300:
    to_add = 300. * 365
new_end[i] = df.iloc[i][lad] + pd.DateOffset(days=to_add)
df[esd] = new_end
df.to_csv('fixed_date.csv')

```

```
In [ ]: df = pd.read_csv('individuals.csv').sample(500000)
```

```

offence = pd.get_dummies(df.OFFENSE)
race = pd.get_dummies(df.RACE)
race.drop('AMER IND', axis=1, inplace=True)
race = pd.get_dummies(df.RACE)
race.drop('AMER IND', axis=1, inplace=True)

```

```
gender = pd.get_dummies(df.GENDER).drop('F', axis=1)
```

*# because of bad encoding I need to clean these variables*

```

t1 = offence['53A021'] + offence['53A021']
t2 = offence['ALTERNATE PENALTY FOR DRUG SALES'] + offence['ALTERNATE PENALTY FOR DRUG SALES']
t3 = offence['APPLICATION TO PURCHASE A FIREARM'] + offence['APPLICATION TO PURCHASE A FIREARM']
t4 = offence['CARRY PIST/RVOLV W/O PERMIT'] + offence['CARRY PIST/RVOLV W/O PERMIT']
t5 = offence['CARRYING OR SALE OF DANGEROUS WEAPON'] + offence['CARRYING OR SALE OF DANGEROUS WEAPON']
t6 = offence['CHEATING AT GAMBLING'] + offence['CHEATING AT GAMBLING']
t7 = offence['CONSPIRACY'] + offence['CONSPIRACY']
t8 = offence['CREDIT CARD THEFT'] + offence['CREDIT CARD THEFT']
t9 = offence['CRIMINAL ATTEMPT'] + offence['CRIMINAL ATTEMPT']
t10 = offence['CRIMINAL LIABILITY FOR ANOTHER PERSON'] + offence['CRIMINAL LIABILITY FOR ANOTHER PERSON']
t11 = offence['CRUELTY TO ANIMALS'] + offence['CRUELTY TO ANIMALS']
t12 = offence['CRUELTY TO PERSONS'] + offence['CRUELTY TO PERSONS']
t13 = offence['DRIVING WHILE LICENSE SUSP FOR DWI'] + offence['DRIVING WHILE LICENSE SUSP FOR DWI']
t14 = offence['DRUGS NEAR PROHIBITED PLACE'] + offence['DRUGS NEAR PROHIBITED PLACE']
t15 = offence['ENTICING A MINOR'] + offence['ENTICING A MINOR']
t16 = offence['ESCAPE FROM CUSTODY'] + offence['ESCAPE FROM CUSTODY']
t17 = offence['FAILURE TO OBEY AN OFFICER'] + offence['FAILURE TO OBEY AN OFFICER']
t18 = offence['FALSELY REPORTING INCIDENT'] + offence['FALSELY REPORTING INCIDENT']
t19 = offence['HOME INVASION'] + offence['HOME INVASION']
t20 = offence['ILLEGAL USE OF CREDIT CARD'] + offence['ILLEGAL USE OF CREDIT CARD']
t21 = offence['INTERFERENCE WITH SEARCH'] + offence['INTERFERENCE WITH SEARCH']
t22 = offence['INTERSTATE CUSTODY COMPACT'] + offence['INTERSTATE CUSTODY COMPACT']
t23 = offence['ISSUING A BAD CHECK'] + offence['ISSUING A BAD CHECK']
t24 = offence['LARCENY DEFINED'] + offence['LARCENY DEFINED']
t25 = offence['OPERATE UNDER INFLU OF LIQ OR DRUGS'] + offence['OPERATE UNDER INFLU OF LIQ OR DRUGS']
t26 = offence['POSS OF ASSAULT WEAPON PROHIBITED'] + offence['POSS OF ASSAULT WEAPON PROHIBITED']
t27 = offence['POSSESS OF LT 4 OZ MJ OR CONTROLLED SUB'] + offence['POSSESS OF LT 4 OZ MJ OR CONTROLLED SUB']
t28 = offence['POSSESSION OF NARCOTICS'] + offence['POSSESSION OF NARCOTICS']

```



```

offence['SALE, CARRY & BRAND OF FACSIMILE'] = t35
offence['STRANGULATION 3RD DEGREE'] = t36
offence['USING MOTOR VEHICLE W/O PERMISSION'] = t37
offence['VIOLATION OF CONDITIONS OF RELEASE'] = t38
offence['VIOLATION OF PROBATION OR COND DISCHG'] = t39
offence['VOYEURISM'] = t40
offence['YOUTHFUL OFFENDER'] = t41

```

```

In [ ]: # now we drop all the excess categories
        # AND one of the categories to fix linear dependence

```

```

offence.drop(
    [
        '53A021', # this is the dependence
        'ALTERNATE PENALTY FOR DRUG SALES',
        'APPLICATION TO PURCHASE A FIREARM',
        'CARRY PIST/RVOLV W/O PERMIT',
        'CARRYING OR SALE OF DANGEROUS WEAPON',
        'CHEATING AT GAMBLING',
        'CONSPIRACY',
        'CREDIT CARD THEFT',
        'CRIMINAL ATTEMPT',
        'CRIMINAL LIABILITY FOR ANOTHER PERSON',
        'CRUELTY TO ANIMALS',
        'CRUELTY TO PERSONS',
        'DRIVING WHILE LICENSE SUSP FOR DWI',
        'DRUGS NEAR PROIBITED PLACE',
        'ENTICING A MINOR',
        'ESCAPE FROM CUSTODY',
        'FAILURE TO OBEY AN OFFICER',
        'FALSELY REPORTING INCIDENT',
        'HOME INVASION',
        'ILLEGAL USE OF CREDIT CARD',
        'INTERFERENCE WITH SEARCH',
        'INTERSTATE CUSTODY COMPACT',
        'ISSUING A BAD CHECK',
        'LARCENY DEFINED',
        'OPERATE UNDER INFLU OF LIQ OR DRUGS',
        'POSS OF ASSAULT WEAPON PROHIBITED',
        'POSSESS OF LT 4 OZ MJ OR CONTROLD SUB',
        'POSSESSION OF NARCOTICS',
        'PROHIB ACTS RE: CONTROLLED SUBSTANCES',
        'PROHIBITED ACTIVITIES',
        'REGULATION OF RESTRICTED SUBSTANCES',
        'ROBBERY BY CARJACKING',
        'ROBBERY, FIRST DEGREE',
        'ROBBERY, FIRST DEGREE',
        'SALE OF HEROIN, COC BY NON-DEPENDENT',
        'SALE, CARRY & BRAND OF FACSIMILE',
    ],

```

```

        'STRANGULATION 3RD DEGREE',
        'USING MOTOR VEHICLE W/O PERMISSION',
        'VIOLATION OF CONDITIONS OF RELEASE',
        'VIOLATION OF PROBATION OR COND DISCHG',
        'VOYEURISM',
        'YOUTHFUL OFFENDER'
    ], axis=1, inplace=True
)

one = pd.DataFrame(np.ones_like(df.AGE))
reg_df = pd.concat(
    [
        df['IDENTIFIER'],
        df['LATEST ADMISSION DATE'],
        df['AGE'],
        df['SENTENCE DAYS'],
        offence,
        gender,
        race,
        one
    ], axis=1
)

reg_df.to_csv('regression_df.csv')

```

```
In [ ]: df = pd.read_csv('individuals.csv').sample(3000000)
```

```

race = pd.get_dummies(df.RACE)
race.drop('AMER IND', axis=1, inplace=True)

one = pd.DataFrame(np.ones_like(df.AGE))
reg_df = pd.concat(
    [
        df['IDENTIFIER'],
        df['LATEST ADMISSION DATE'],
        df['AGE'],
        df['SENTENCE DAYS'],
        gender,
        race,
        one
    ], axis=1
)
reg_df.to_csv('race_regression_df.csv')

```

### 3 Extras

Here I have inserted the full results from the regressions that I ran.

### 3.1 Full regression

```
In [6]: regr_df = pd.read_csv('small_regression_df.csv')
        regr_df.dropna(inplace=True)
        sentence = regr_df['SENTENCE DAYS']
        regr_df.drop(
            ['SENTENCE DAYS', 'Unnamed: 0', 'IDENTIFIER', 'LATEST ADMISSION DATE'], axis=1, inplace=True
        )
        X_train, X_test, y_train, y_test = model_selection.train_test_split(
            regr_df.astype(float), sentence, test_size=.3
        )
        res_gen = sm.OLS(y_train, X_train).fit()
        print(res_gen.summary())
```

#### OLS Regression Results

=====						
Dep. Variable:	SENTENCE DAYS	R-squared:	0.582			
Model:	OLS	Adj. R-squared:	0.578			
Method:	Least Squares	F-statistic:	162.4			
Date:	Thu, 12 Dec 2019	Prob (F-statistic):	0.00			
Time:	14:50:27	Log-Likelihood:	-2.6340e+05			
No. Observations:	22824	AIC:	5.272e+05			
Df Residuals:	22629	BIC:	5.288e+05			
Df Model:	194					
Covariance Type:	nonrobust					
=====						
		coef	std err	t	P> t	[0.025
-----						
AGE		256.1859	15.438	16.595	0.000	225.99
1ST DEGREE MURDER, OTHER HOMICIDES	F	3.573e+05	2.16e+04	16.511	0.000	3.15e+05
53-202A		2474.7037	2.79e+04	0.089	0.929	-5.23e+04
53A189C		4.107e-10	4.2e-10	0.978	0.328	-4.13e-10
ABUSE OF PERSONS-1ST DEGREE	CF	-6927.1594	2.16e+04	-0.320	0.749	-4.93e+04
ABUSE OF PERSONS-2ND DEGREE	DF	-9.721e-11	1.54e-10	-0.630	0.528	-3.99e-11
AGGAVTD SEX ASSLT-VCTM < 13YR	AF	6149.0934	1.91e+04	0.322	0.747	-3.12e+04
ALLOWING PERSON UNDER 16 TO OPERATE MV	M	1.496e-09	3.22e-10	4.652	0.000	8.66e-10
ALTERING OR REMOVING IDENT NUMBER	F	2979.1333	1.44e+04	0.206	0.836	-2.53e+04
ALTERNATE PENALTY FOR DRUG SALES		4273.5572	2.79e+04	0.153	0.878	-5.05e+04
APPLICATION TO PURCHASE A FIREARM		-811.3957	2.79e+04	-0.029	0.977	-5.55e+04
ARSON MURDER	F	2.233e+04	2.16e+04	1.032	0.302	-2.01e+04
ARSON, FIRST DEGREE	AF	1.59e+04	1.29e+04	1.231	0.218	-9410.8
ARSON, SECOND DEGREE	BF	1150.3351	1.35e+04	0.085	0.932	-2.53e+04
ARSON, THIRD DEGREE	CF	-36.4595	1.61e+04	-0.002	0.998	-3.16e+04
ASSAULT 2ND DEGREE WITH A FIREARM	DF	-1086.0567	1.44e+04	-0.075	0.940	-2.94e+04
ASSAULT 2ND WITH MV WHILE INTOXICATED	DF	-1236.9713	1.3e+04	-0.095	0.924	-2.67e+04
ASSAULT 2ND, VICTIM 60 OR OVER	DF	-1450.2232	1.43e+04	-0.101	0.919	-2.95e+04
ASSAULT 3RD DEGREE, VICTIM OVER 59	AM	-652.9615	1.34e+04	-0.049	0.961	-2.68e+04
ASSAULT ON POLICE OR FIRE OFFICER	CF	1298.8095	1.26e+04	0.103	0.918	-2.35e+04
ASSAULT, 1ST VICTIM 60 OR OVER	BF	2941.4608	1.32e+04	0.224	0.823	-2.28e+04



ASSAULT, FIRST DEGREE	BF	2937.3466	1.25e+04	0.235	0.814	-2.16e+0
ASSAULT, SECOND DEGREE	DF	-194.6543	1.26e+04	-0.015	0.988	-2.48e+0
ASSAULT, THIRD DEGREE	AM	-1130.0424	1.26e+04	-0.090	0.928	-2.58e+0
BREACH OF PEACE	BM	-2820.6217	1.31e+04	-0.216	0.829	-2.85e+0
BRIBERY OF WITNESS	DF	-2.421e-10	1.24e-10	-1.958	0.050	-4.84e+0
BURGLARY, FIRST DEGREE	BF	2180.1980	1.26e+04	0.174	0.862	-2.24e+0
BURGLARY, FIRST DEGREE: 53A101A(2)	BF	388.4405	1.4e+04	0.028	0.978	-2.7e+0
BURGLARY, SECOND DEGREE	CF	951.2981	1.26e+04	0.075	0.940	-2.38e+0
BURGLARY, SECOND DEGREE WITH FIREARM	CF	2714.9037	1.61e+04	0.168	0.866	-2.89e+0
BURGLARY, THIRD DEGREE	DF	-837.2618	1.25e+04	-0.067	0.947	-2.54e+0
BURGLARY, THIRD DEGREE WITH FIREARM	DF	3702.1911	2.79e+04	0.133	0.895	-5.1e+0
CAPITAL FELONY	F	3.536e+05	1.27e+04	27.863	0.000	3.29e+0
CARRY PIST/RVOLV W/O PERMIT		1552.4495	1.29e+04	0.120	0.904	-2.37e+0
CARRYING OR SALE OF DANGEROUS WEAPON		-1502.4240	1.33e+04	-0.113	0.910	-2.75e+0
CARRYING WEAPON IN A MOTOR VEHICLE	F	1947.7223	1.29e+04	0.151	0.880	-2.33e+0
CARRYING WEAPONS WITHOUT A PERMIT	F	2301.0846	1.26e+04	0.182	0.855	-2.25e+0
CHEATING AT GAMBLING		-1444.2007	1.91e+04	-0.076	0.940	-3.89e+0
COERCION		-2.258e-10	1.46e-10	-1.547	0.122	-5.12e-1
COMMIT SEXUAL OFF IN ANOTHER JURIS	DF	1233.4832	1.57e+04	0.079	0.937	-2.95e+0
CONSPIRACY		3470.6699	1.25e+04	0.277	0.782	-2.11e+0
CONTEMPT OF COURT	M	-362.6651	1.53e+04	-0.024	0.981	-3.04e+0
CONVEY UNAUTHZD ITEMS INTO INST	DF	-2407.6837	1.77e+04	-0.136	0.892	-3.7e+0
CREDIT CARD THEFT		-539.3108	1.77e+04	-0.031	0.976	-3.52e+0
CRIM DAMAGE PROPERTY 1ST DEG	DF	1.068e-10	2.69e-10	0.397	0.691	-4.2e-1
CRIM VIOL CIVIL PROTECTION ORDER	DF	2220.4474	1.77e+04	0.126	0.900	-3.24e+0
CRIM VIOL OF PROTECTIVE ORDER	AM	1.141e-11	1.33e-10	0.086	0.932	-2.5e-1
CRIM VIOL OF PROTECTIVE ORDER	DF	-1069.2360	1.25e+04	-0.085	0.932	-2.56e+0
CRIM VIOL RES ORDER	AM	-1601.5699	1.3e+04	-0.123	0.902	-2.71e+0
CRIM VIOL ST CRIM RESTR ORDER	DF	-1953.9716	1.27e+04	-0.154	0.878	-2.69e+0
CRIMINAL ATTEMPT		1810.4962	1.25e+04	0.145	0.885	-2.27e+0
CRIMINAL CONTEMPT	M	-4530.0232	1.68e+04	-0.270	0.787	-3.74e+0
CRIMINAL IMPERSONATION	BM	-4068.6268	2.79e+04	-0.146	0.884	-5.88e+0
CRIMINAL LIABILITY FOR ANOTHER PERSON		4.211e+04	1.26e+04	3.346	0.001	1.74e+0
CRIMINAL MISCHIEF, FIRST DEGREE	DF	376.1072	1.41e+04	0.027	0.979	-2.72e+0
CRIMINAL MISCHIEF, SECOND DEGREE	AM	-1507.6840	1.42e+04	-0.106	0.915	-2.93e+0
CRIMINAL MISCHIEF, THIRD DEGREE	BM	88.5849	1.77e+04	0.005	0.996	-3.45e+0
CRIMINAL POSS BODY ARMOR	AM	-693.0370	2.79e+04	-0.025	0.980	-5.54e+0
CRIMINAL POSS OF PISTOL/REVOLVER	DF	844.4109	1.26e+04	0.067	0.947	-2.38e+0
CRIMINAL POSSESSION OF A PISTOL	DF	968.2757	1.26e+04	0.077	0.939	-2.36e+0
CRIMINAL TRESPASS, FIRST DEGREE	AM	-3179.3497	1.31e+04	-0.242	0.809	-2.9e+0
CRIMINAL TRESPASS, SECOND DEGREE	BM	1953.0521	1.91e+04	0.102	0.918	-3.54e+0
CRIMINAL TRESPASS, THIRD DEGREE	CM	-8.981e-11	2.83e-10	-0.317	0.751	-6.45e-1
CRIMINAL TROVER, 2ND DEGREE	AM	1.053e-10	2.88e-10	0.365	0.715	-4.6e-1
CRIMINAL USE OF A FIREARM	DF	-1.206e-10	5.46e-11	-2.209	0.027	-2.28e-1
CRIMINAL VIOL. PROTECTION ORDER	AM	-1.441e-10	2.41e-10	-0.597	0.550	-6.17e-1
CRIMINALLY NEGLIGENT HOMICIDE	AM	1212.4548	1.68e+04	0.072	0.942	-3.17e+0
CRUELTY TO ANIMALS		37.9630	1.77e+04	0.002	0.998	-3.46e+0
CRUELTY TO PERSONS		353.9600	1.77e+04	0.020	0.984	-3.43e+0

DISORDERLY CONDUCT	CM	-1988.3754	1.4e+04	-0.142	0.887	-2.94e+
DISSEMINATING VOYEURISTIC MATERIAL	DF	-2.714e-11	5.55e-11	-0.489	0.625	-1.36e-
DRIVING WHILE LICENSE SUSP FOR DWI		-2924.8132	1.29e+04	-0.227	0.821	-2.82e+
DRUG PARAPHERNALIA IN DRUG FACTORY	F	-52.0274	1.77e+04	-0.003	0.998	-3.47e+
DRUGS NEAR PROIBITED PLACE		9480.1008	2.79e+04	0.339	0.734	-4.53e+
DUI .02 BAC AND UNDER 21 YRS		-7.223e-10	1.25e-09	-0.576	0.565	-3.18e-
EMPLOYING MINOR IN OBSCENE PERFORMNCE	AF	-253.1694	1.46e+04	-0.017	0.986	-2.88e+
ENTICING A MINOR		97.2867	1.48e+04	0.007	0.995	-2.89e+
ESCAPE FROM CUSTODY		2551.7198	2.16e+04	0.118	0.906	-3.98e+
ESCAPE, FIRST DEGREE	CF	272.1017	1.29e+04	0.021	0.983	-2.5e+
EVADING RESPONSIBILITY	M	424.6694	1.31e+04	0.033	0.974	-2.52e+
FAILURE REGISTER SEXUAL PURPOSE	DF	6.599e-11	1.41e-10	0.469	0.639	-2.1e-
FAILURE TO APPEAR, FIRST DEGREE	DF	-1154.7023	1.28e+04	-0.090	0.928	-2.63e+
FAILURE TO APPEAR, SECOND DEGREE	AM	-926.6051	1.3e+04	-0.071	0.943	-2.65e+
FAILURE TO FURNISH TAX RETURN		7.94e-11	1.41e-10	0.563	0.573	-1.97e-
FAILURE TO OBEY AN OFFICER		-61.0930	1.43e+04	-0.004	0.997	-2.81e+
FAILURE TO PAY OR PLEAD	M	-5591.7585	2.79e+04	-0.200	0.841	-6.03e+
FALSE STATEMENT, 2ND DEGREE	AM	3.848e-11	5.13e-11	0.750	0.453	-6.21e-
FALSELY REPORTING AN INCIDENT 2ND	AM	-2.475e-11	3.37e-11	-0.734	0.463	-9.09e-
FALSELY REPORTING INCIDENT		866.3319	1.91e+04	0.045	0.964	-3.65e+
FEDERAL OFFENSE		-5.513e-12	2.06e-11	-0.267	0.789	-4.59e-
FELONY MURDER	AF	2.795e+04	1.26e+04	2.227	0.026	3349.5
FINES FOR VIOLATIONS		-1.083e-11	3.33e-11	-0.325	0.745	-7.61e-
FLR KEEP NARC IN ORG CONTAINER	F	-1158.5817	2.16e+04	-0.054	0.957	-4.36e+
FORGERY, FIRST DEGREE	CF	2500.9181	1.57e+04	0.159	0.873	-2.83e+
FORGERY, SECOND DEGREE	DF	210.8111	1.31e+04	0.016	0.987	-2.54e+
FORGERY, THIRD DEGREE	BM	-294.2558	1.5e+04	-0.020	0.984	-2.97e+
HARASSMENT 1ST DEGREE	DF	-2999.7833	1.53e+04	-0.196	0.845	-3.3e+
HARASSMENT, 2ND DEGREE	CM	-8608.7172	2.16e+04	-0.398	0.691	-5.1e+
HINDER PROSECUTION DEFINED	F	1812.8970	1.91e+04	0.095	0.924	-3.56e+
HINDER PROSECUTION, FIRST DEGREE	DF	948.9291	1.44e+04	0.066	0.948	-2.73e+
HINDERING PROSECUTION 2ND DEG		-4.174e-11	1.82e-11	-2.294	0.022	-7.74e-
HOME INVASION		1320.1544	1.92e+04	0.069	0.945	-3.62e+
HOME INVASION	AF	4213.4629	1.26e+04	0.334	0.738	-2.05e+
IDENTITY THEFT 1ST DEGREE	BF	-452.3815	1.57e+04	-0.029	0.977	-3.11e+
IDENTITY THEFT 2ND DEGREE	CF	1090.7198	2.16e+04	0.050	0.960	-4.13e+
IDENTITY THEFT 3RD DEGREE	DF	94.9545	1.34e+04	0.007	0.994	-2.62e+
IL POS LRG CAP MAG OBT AF 4/13	DF	4.899e-11	2.46e-11	1.989	0.047	7.16e-
ILL CARRY FIREARM-UI DRUGS/ALC		2.935e-11	2.03e-11	1.443	0.149	-1.05e-
ILL OP MV WO IGNITION DEVICE	CM	-3036.0253	1.41e+04	-0.216	0.829	-3.06e+
ILL POSSESS CHILD PORN 2ND DEG	F	-4367.6658	1.29e+04	-0.340	0.734	-2.96e+
ILL POSSESS CHILD PORN 3RD DEG	DF	-2212.5658	1.3e+04	-0.170	0.865	-2.78e+
ILLEGAL GAIN FROM CREDIT CARD	AM	2516.3800	1.77e+04	0.142	0.887	-3.21e+
ILLEGAL MANUFACTUAL OR SALE OF DRUGS		8.225e-12	1.72e-11	0.478	0.633	-2.55e-
ILLEGAL POSSES OF SHOP LIFTING DEVICE	AM	-6191.3002	2.16e+04	-0.286	0.775	-4.86e+
ILLEGAL POSSESSION OF EXPLOSIVES	F	-7218.9735	2.79e+04	-0.258	0.796	-6.2e+
ILLEGAL REFUSAL OF DNA TESTING		-2.629e-11	2.15e-11	-1.224	0.221	-6.84e-
ILLEGAL USE OF CREDIT CARD		919.6072	1.53e+04	0.060	0.952	-2.91e+

IMPERSONATION OF POLICE OFFICER	DF	-2312.5597	1.77e+04	-0.131	0.896	-3.69e+
IMPORTING CHILD PORNOGRAPHY	F	-3.758e-11	1.6e-11	-2.346	0.019	-6.9e-
INCITING INJURY TO PERSONS OR PROPTY	CF	-1673.5290	1.61e+04	-0.104	0.917	-3.33e+
INDECENT ASSAULT	F	6.654e-12	1.35e-11	0.493	0.622	-1.98e-
INJUNCTION AND RESTRAINING ORDER		-2.137e-11	1.69e-11	-1.263	0.207	-5.45e-
INJURY OR RISK OF INJURY TO MINOR	F	-312.1629	1.26e+04	-0.025	0.980	-2.49e+
INJURY/RISK OF INJURY TO MINOR- SEXUAL C	C	-870.4994	1.26e+04	-0.069	0.945	-2.55e+
INSURANCE FRAUD	DF	6.231e-11	2.15e-11	2.902	0.004	2.02e-
INTERFERENCE WITH SEARCH		-1067.5817	1.91e+04	-0.056	0.955	-3.85e+
INTERFERING WITH AN EMERGENCY CALL	AM	-42.8380	2.16e+04	-0.002	0.998	-4.24e+
INTERFERING WITH AN OFFICER	AM	-614.8651	1.27e+04	-0.048	0.961	-2.55e+
INTERSTATE CUSTODY COMPACT		-6437.9414	2.79e+04	-0.230	0.818	-6.12e+
INTIMIDATE DUE TO BIAS 3RD DEG	AM	-2.407e-12	1.42e-11	-0.169	0.866	-3.03e-
INTIMIDATING A WITNESS	CF	1164.9057	1.91e+04	0.061	0.951	-3.62e+
INTIMIDATION BSD BIGOTRY/BIAS	DF	-1.198e-11	1.36e-11	-0.880	0.379	-3.87e-
ISSUING A BAD CHECK		-2294.1394	2.79e+04	-0.082	0.935	-5.7e+
KIDNAPPING 1ST DEG W/ FIREARM	AF	1.498e+04	1.44e+04	1.038	0.299	-1.33e+
KIDNAPPING 2ND DEG W/ FIREARM	BF	5809.3431	1.61e+04	0.360	0.719	-2.58e+
KIDNAPPING, 2ND DEGREE	BF	2289.4163	1.29e+04	0.178	0.859	-2.29e+
KIDNAPPING, FIRST DEGREE	AF	3.038e+04	1.27e+04	2.393	0.017	5501.1
LARCENY DEFINED		2032.8193	2.16e+04	0.094	0.925	-4.04e+
LARCENY, FIFTH DEGREE	BM	-327.1816	1.34e+04	-0.024	0.981	-2.66e+
LARCENY, FIRST DEGREE	BF	-501.7151	1.26e+04	-0.040	0.968	-2.52e+
LARCENY, FOURTH DEGREE	AM	359.5995	1.31e+04	0.027	0.978	-2.54e+
LARCENY, SECOND DEGREE	CF	-185.7433	1.26e+04	-0.015	0.988	-2.49e+
LARCENY, SIXTH DEGREE	CM	-3191.9601	1.26e+04	-0.253	0.800	-2.79e+
LARCENY, THIRD DEGREE	DF	-1130.2126	1.26e+04	-0.090	0.928	-2.58e+
LEAV CHILD UNSUPERV IN PUBL/MOT VEH	C	1979.6363	2.16e+04	0.091	0.927	-4.04e+
LIQUOR SALES TO MINORS		-1.061e-11	7.71e-12	-1.376	0.169	-2.57e-
MANSLAUGHTER 2ND WITH MV (INTOX)	CF	563.0916	1.27e+04	0.044	0.965	-2.43e+
MANSLAUGHTER 2ND, W/FIREARM	CF	3516.7989	1.43e+04	0.246	0.806	-2.45e+
MANSLAUGHTER, FIRST DEGREE	BF	4595.7474	1.26e+04	0.366	0.714	-2e+
MANSLAUGHTER, FIRST DEGREE, FIREARM	BF	7504.5626	1.26e+04	0.597	0.550	-1.71e+
MANSLAUGHTER, SECOND DEGREE	CF	2454.3073	1.29e+04	0.190	0.849	-2.29e+
MANUF OR POSSESSION OF BURGLAR TOOLS	AM	1.053e-11	7.41e-12	1.420	0.156	-4e-
MANUFACTURE OF BOMBS	BF	2360.0304	1.91e+04	0.124	0.902	-3.5e+
MISCONDUCT WITH A MOTOR VEHICLE	DF	-1391.2893	1.36e+04	-0.102	0.919	-2.81e+
MURDER	F	3.571e+05	1.48e+04	24.143	0.000	3.28e+
MURDER	AF	2.997e+04	1.25e+04	2.395	0.017	5446.6
NEGLG STORAGE OF A FIREARM	DF	2037.7614	1.91e+04	0.107	0.915	-3.54e+
NEGLIGENT HOMICIDE W/ MOTOR VEHICLE	BM	2455.3800	1.91e+04	0.129	0.898	-3.49e+
OBSCENITY AS TO MINORS	F	1.772e-11	6.83e-12	2.593	0.010	4.32e-
OPERATE UNDER INFLU OF LIQ OR DRUGS		-3785.9369	1.26e+04	-0.302	0.763	-2.84e+
OPERATE UNDER INFLU OF LIQ OR DRUGS	M	2.207e-12	1.04e-11	0.211	0.833	-1.83e-
OPERATING MV W/O A LICENSE		4.098e-11	5.27e-12	7.776	0.000	3.06e-
OPERATING MV W/O INSURANCE	CM	-4.178e-12	7.14e-12	-0.585	0.559	-1.82e-
OPERATING UNDER SUSPENDED LIC OR REG.	M	-2496.2037	1.36e+04	-0.184	0.854	-2.91e+
ORGANIZED RETAIL THEFT	F	8.649e-12	7.7e-12	1.123	0.262	-6.45e-

OUI MV W/CHILD PSNGR	AM	-3104.8831	2.79e+04	-0.111	0.911	-5.78e+
PATRONIZE PROSTITUTE FM MV	AM	-5.63e-12	5.81e-12	-0.969	0.333	-1.7e-
PATRONIZING A PROSTITUTE	AM	8.697e-13	9.18e-12	0.095	0.925	-1.71e-
PENALTY FAIL TO SUBMIT FINGERPRINTS		-8.069e-13	4.75e-12	-0.170	0.865	-1.01e-
PENALTY-EMPLOYER VIOL SEC. OF 31-71		-3.87e-12	7.47e-12	-0.518	0.605	-1.85e-
PERMITTING PROSTITUTION	AM	-4.941e-12	7.31e-12	-0.676	0.499	-1.93e-
POSS OF ASSAULT WEAPON PROHIBITED		1718.4215	1.68e+04	0.103	0.918	-3.11e+
POSS OF WEAPON IN A CCI	BF	-1482.0827	1.57e+04	-0.095	0.925	-3.22e+
POSS WEAPON ON SCHOOL GROUNDS	DF	9.453e-13	8.46e-12	0.112	0.911	-1.56e-
POSSESS OF HALLUC OR 4+ OZ MARIJUANA	F	-587.0711	1.91e+04	-0.031	0.975	-3.8e+
POSSESS OF LT 4 OZ MJ OR CONTROLD SUB		1593.8222	2.79e+04	0.057	0.954	-5.32e+
POSSESSING CHILD PORNOGRAPH	F	-2738.4402	1.27e+04	-0.216	0.829	-2.76e+
POSSESSION OF NARCOTICS		-290.3926	1.26e+04	-0.023	0.982	-2.5e+
POSSESSION OF SHOTGUN OR SILENCER	DF	-471.4484	1.68e+04	-0.028	0.978	-3.33e+
PRACTICE OF LAW BY NON-ATTORNEYS	M	6.963e-13	1.22e-11	0.057	0.954	-2.32e-
PRIMARY OFFENSE NOT DETERMINED		9.864e-14	4.67e-12	0.021	0.983	-9.05e-
PROHIB ACTS RE: CONTROLLED SUBSTANCES		1771.5430	2.79e+04	0.063	0.949	-5.3e+
PROHIB ACTS RE: DRUG PARAPHERNALIA		-1615.5817	2.79e+04	-0.058	0.954	-5.64e+
PROHIBITED ACTIVITIES		3249.9515	1.4e+04	0.233	0.816	-2.41e+
PROMOTING MINOR IN OBSCENE PERFORMNCE	BF	-804.2158	1.48e+04	-0.054	0.957	-2.98e+
PROMOTING PROSTITUTION, FIRST DEGREE	BF	1740.7741	1.77e+04	0.099	0.922	-3.29e+
PROMOTING PROSTITUTION, SECOND DEGREE	CF	-2435.2679	1.57e+04	-0.155	0.876	-3.31e+
PROMOTING PROSTITUTION, THIRD DEGREE	DF	-1731.1555	2.79e+04	-0.062	0.951	-5.65e+
PROSTITUTION	AM	-2602.9753	2.79e+04	-0.093	0.926	-5.74e+
PUBLIC INDECENCY	BM	-4277.6399	2.16e+04	-0.198	0.843	-4.67e+
RAPE, SECOND DEGREE	CF	-1.047e-13	2.26e-12	-0.046	0.963	-4.54e-
RECKLESS BURNING	DF	1831.2680	1.5e+04	0.122	0.903	-2.76e+
RECKLESS DRIVING	M	-4.407e-13	3.92e-12	-0.112	0.911	-8.13e-
RECKLESS ENDANGERMENT, FIRST DEGREE	AM	-151.4686	1.29e+04	-0.012	0.991	-2.55e+
RECKLESS ENDANGERMENT, SECOND DEGREE	BM	-1008.6525	1.57e+04	-0.064	0.949	-3.18e+
REG PERSON WHO COMM OFF AGAINST MINOR	DF	-698.6306	1.37e+04	-0.051	0.959	-2.76e+
REG PERSON WHO COMM SEX VIOL OFF	DF	-3399.1900	1.4e+04	-0.243	0.808	-3.08e+
REGULATION OF RESTRICTED SUBSTANCES		-3173.4409	1.91e+04	-0.166	0.868	-4.06e+
RIGGING SPORT EVENT	DF	4.404e-13	2.65e-12	0.166	0.868	-4.75e-
RIOT, SECOND DEGREE	BM	2.985e-13	2.54e-12	0.118	0.906	-4.67e-
ROBBERY BY CARJACKING		-652.7679	1.48e+04	-0.044	0.965	-2.96e+
ROBBERY, FIRST DEGREE	BF	2374.8850	1.25e+04	0.190	0.849	-2.22e+
ROBBERY, SECOND DEGREE	CF	1339.8515	1.26e+04	0.107	0.915	-2.33e+
ROBBERY, THIRD DEGREE	DF	-357.6969	1.27e+04	-0.028	0.978	-2.52e+
SALE OF CONTROLLED SUBSTANCE	F	-127.2492	1.26e+04	-0.010	0.992	-2.48e+
SALE OF HALLUCIGEN/NARCOTIC SUBSTANCE	F	104.3447	1.25e+04	0.008	0.993	-2.44e+
SALE OF HEROIN, COC BY NON-DEPENDENT		165.2617	1.31e+04	0.013	0.990	-2.55e+
SALE OF NARC/AMPHET BY NON-DEPENDENT	F	1144.0585	1.26e+04	0.091	0.927	-2.35e+
SALE, CARRY & BRAND OF FACSIMILE		7015.5208	2.79e+04	0.251	0.802	-4.77e+
SALE/POSS/TRAN OF ASLT WEAPON	F	-1282.6898	1.91e+04	-0.067	0.946	-3.87e+
SELLING WEAPONS TO ALIENS PROHIBITED	F	736.3564	1.42e+04	0.052	0.959	-2.7e+
SEX ASLT, SPOUSE/COHAB RELATIONSHIP	BF	287.1059	1.33e+04	0.022	0.983	-2.57e+
SEXUAL ASSAULT 1ST DEGREE-AGGRVTD	F	5424.8557	1.29e+04	0.419	0.675	-2e+

SEXUAL ASSAULT, 3RD DEG W/ FIREARM	CF	0	0	nan	nan	
SEXUAL ASSAULT, FIRST DEGREE	F	3195.7634	1.25e+04	0.255	0.799	-2.13e+
SEXUAL ASSAULT, FOURTH DEGREE	AM	-1020.2827	1.27e+04	-0.080	0.936	-2.6e+
SEXUAL ASSAULT, SECOND DEGREE	F	658.5671	1.26e+04	0.052	0.958	-2.4e+
SEXUAL ASSAULT, THIRD DEGREE	DF	-983.8750	1.29e+04	-0.076	0.939	-2.63e+
SPECIAL PAROLE		0	0	nan	nan	
STALKING, 1ST DEGREE	DF	0	0	nan	nan	
STALKING, 2ND DEGREE	AM	-6962.7875	2.79e+04	-0.249	0.803	-6.17e+
STALKING, 3RD DEGREE	BM	0	0	nan	nan	
STEALING A FIREARM	DF	2072.4066	1.31e+04	0.158	0.875	-2.37e+
STRANGULATION 1ST DEGREE	CF	-1565.0762	1.38e+04	-0.113	0.910	-2.86e+
STRANGULATION 2ND DEGREE	DF	-577.1669	1.28e+04	-0.045	0.964	-2.56e+
STRANGULATION 3RD DEGREE		-499.2108	1.4e+04	-0.036	0.971	-2.79e+
TAMPERING WITH OR FABRICATING EVIDNC	DF	-99.4102	1.41e+04	-0.007	0.994	-2.77e+
TAMPERING WITH WITNESS	DF	-1740.5876	1.42e+04	-0.123	0.902	-2.95e+
THEFT OF UTILITY SERVICE FOR PROFIT	DF	0	0	nan	nan	
THREATENING	AM	-486.1205	1.28e+04	-0.038	0.970	-2.56e+
THREATENING 1ST DEGREE	DF	1184.6918	1.41e+04	0.084	0.933	-2.64e+
TRAFFICKING IN PERSONS	BF	0	0	nan	nan	
UNLAWFUL DISCHARGE OF FIREARMS	M	0	0	nan	nan	
UNLAWFUL REPRODUCTION OF CREDIT CARD	DF	-5570.1142	2.79e+04	-0.199	0.842	-6.03e+
UNLAWFUL RESTRAINT, FIRST DEGREE	DF	-242.0405	1.28e+04	-0.019	0.985	-2.53e+
UNLAWFUL RESTRAINT, SECOND DEGREE	AM	1751.5927	1.42e+04	0.124	0.902	-2.6e+
USE OF MACHINE GUN IN A CRIME DELETED	F	-787.9770	1.91e+04	-0.041	0.967	-3.82e+
USING MACHINE GUN IN CRIME	F	1940.3282	1.61e+04	0.120	0.904	-2.97e+
USING MOTOR VEHICLE W/O PERMISSION		-1645.3945	1.48e+04	-0.111	0.911	-3.06e+
VIO DEADLY WEAPON REGIST REQS	DF	1733.7807	1.91e+04	0.091	0.928	-3.57e+
VIOLATE COND OF RELEASE-2ND DEG	AM	-120.1933	1.42e+04	-0.008	0.993	-2.79e+
VIOLATION OF CONDITIONS OF RELEASE		-81.3957	2.79e+04	-0.003	0.998	-5.48e+
VIOLATION OF PROBATION OR COND DISCHG		-679.3522	1.25e+04	-0.054	0.957	-2.52e+
VIOLATION, HOME IMPROVE REGISTRATION		0	0	nan	nan	
VOYEURISM		-1.168e+04	2.79e+04	-0.418	0.676	-6.64e+
YOUTHFUL OFFENDER		3517.0867	1.29e+04	0.272	0.785	-2.18e+
M		1858.6646	725.493	2.562	0.010	436.6
AMER IND		-3443.2723	3234.671	-1.064	0.287	-9783.4
ASIAN		-1613.6784	2500.746	-0.645	0.519	-6515.3
BLACK		-1385.9135	422.088	-3.283	0.001	-2213.2
HISPANIC		-813.7306	455.424	-1.787	0.074	-1706.3
0		-8441.0325	1.25e+04	-0.674	0.500	-3.3e+

```

=====
Omnibus:                 37727.512   Durbin-Watson:                2.005
Prob(Omnibus):            0.000   Jarque-Bera (JB):            24649468.400
Skew:                     11.294   Prob(JB):                     0.00
Kurtosis:                 162.403   Cond. No.                     1.27e+19
=====

```

#### Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The smallest eigenvalue is 2.15e-31. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

### 3.2 Race Only Regression

```
In [7]: race_regr_df = pd.read_csv('race_regression_df.csv')
        race_regr_df.dropna(inplace=True)
        sentence = race_regr_df['SENTENCE DAYS']
        race_regr_df.drop(
            ['SENTENCE DAYS', 'Unnamed: 0'], axis=1, inplace=True
        )
        X_train, X_test, y_train, y_test = model_selection.train_test_split(
            race_regr_df.astype(float), sentence, test_size=.3
        )
        res_gen = sm.OLS(y_train, X_train).fit()
        print(res_gen.summary())
```

```

                        OLS Regression Results
=====
Dep. Variable:          SENTENCE DAYS      R-squared:                0.036
Model:                  OLS                Adj. R-squared:          0.036
Method:                 Least Squares       F-statistic:              6083.
Date:                  Thu, 12 Dec 2019     Prob (F-statistic):       0.00
Time:                  14:52:13             Log-Likelihood:           -9.7050e+06
No. Observations:      809642              AIC:                     1.941e+07
Df Residuals:          809636              BIC:                     1.941e+07
Df Model:               5
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
AGE	645.7792	3.728	173.207	0.000	638.472	653.087
ASIAN	1883.5618	1065.757	1.767	0.077	-205.287	3972.411
BLACK	7252.2937	860.786	8.425	0.000	5565.182	8939.405
HISPANIC	5606.3145	862.281	6.502	0.000	3916.272	7296.357
WHITE	4858.4437	861.595	5.639	0.000	3169.747	6547.141
0	-2.202e+04	870.209	-25.302	0.000	-2.37e+04	-2.03e+04

```

=====
Omnibus:                 1125252.206      Durbin-Watson:              2.001
Prob(Omnibus):           0.000          Jarque-Bera (JB):           193455625.499
Skew:                    8.498          Prob(JB):                   0.00
Kurtosis:                76.795          Cond. No.                   1.77e+03
=====

```

Warnings:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 1.77e+03. This might indicate that there are

strong multicollinearity or other numerical problems.

```
In [11]: # df = pd.read_csv('individuals.csv')
        for thing in regr_df.columns:
            print(thing)
```

```
AGE
1ST DEGREE MURDER, OTHER HOMICIDES      F
53-202A
53A189C
ABUSE OF PERSONS-1ST DEGREE              CF
ABUSE OF PERSONS-2ND DEGREE              DF
AGGAVTD SEX ASSLT-VCTM < 13YR           AF
ALLOWING PERSON UNDER 16 TO OPERATE MV  M
ALTERING OR REMOVING IDENT NUMBER        F
ALTERNATE PENALTY FOR DRUG SALES
APPLICATION TO PURCHASE A FIREARM
ARSON MURDER                             F
ARSON, FIRST DEGREE                      AF
ARSON, SECOND DEGREE                     BF
ARSON, THIRD DEGREE                      CF
ASSAULT 2ND DEGREE WITH A FIREARM        DF
ASSAULT 2ND WITH MV WHILE INTOXICATED    DF
ASSAULT 2ND, VICTIM 60 OR OVER            DF
ASSAULT 3RD DEGREE, VICTIM OVER 59       AM
ASSAULT ON POLICE OR FIRE OFFICER        CF
ASSAULT, 1ST VICTIM 60 OR OVER           BF
ASSAULT, FIRST DEGREE                    BF
ASSAULT, SECOND DEGREE                   DF
ASSAULT, THIRD DEGREE                    AM
BREACH OF PEACE                          BM
BRIBERY OF WITNESS                       DF
BURGLARY, FIRST DEGREE                   BF
BURGLARY, FIRST DEGREE: 53A101A(2)       BF
BURGLARY, SECOND DEGREE                  CF
BURGLARY, SECOND DEGREE WITH FIREARM     CF
BURGLARY, THIRD DEGREE                   DF
BURGLARY, THIRD DEGREE WITH FIREARM      DF
CAPITAL FELONY                           F
CARRY PIST/RVOLV W/O PERMIT
CARRYING OR SALE OF DANGEROUS WEAPON
CARRYING WEAPON IN A MOTOR VEHICLE       F
CARRYING WEAPONS WITHOUT A PERMIT        F
CHEATING AT GAMBLING
COERCION
COMMIT SEXUAL OFF IN ANOTHER JURIS       DF
CONSPIRACY
```

CONTEMPT OF COURT	M
CONVEY UNAUTHZD ITEMS INTO INST	DF
CREDIT CARD THEFT	
CRIM DAMAGE PROPERTY 1ST DEG	DF
CRIM VIOL CIVIL PROTECTION ORDER	DF
CRIM VIOL OF PROTECTIVE ORDER	AM
CRIM VIOL OF PROTECTIVE ORDER	DF
CRIM VIOL RES ORDER	AM
CRIM VIOL ST CRIM RESTR ORDER	DF
CRIMINAL ATTEMPT	
CRIMINAL CONTEMPT	M
CRIMINAL IMPERSONATION	BM
CRIMINAL LIABILITY FOR ANOTHER PERSON	
CRIMINAL MISCHIEF, FIRST DEGREE	DF
CRIMINAL MISCHIEF, SECOND DEGREE	AM
CRIMINAL MISCHIEF, THIRD DEGREE	BM
CRIMINAL POSS BODY ARMOR	AM
CRIMINAL POSS OF PISTOL/REVOLVER	DF
CRIMINAL POSSESSION OF A PISTOL	DF
CRIMINAL TRESPASS, FIRST DEGREE	AM
CRIMINAL TRESPASS, SECOND DEGREE	BM
CRIMINAL TRESPASS, THIRD DEGREE	CM
CRIMINAL TROVER, 2ND DEGREE	AM
CRIMINAL USE OF A FIREARM	DF
CRIMINAL VIOL. PROTECTION ORDER	AM
CRIMINALLY NEGLIGENT HOMICIDE	AM
CRUELTY TO ANIMALS	
CRUELTY TO PERSONS	
DISORDERLY CONDUCT	CM
DISSEMINATING VOYEURISTIC MATERIAL	DF
DRIVING WHILE LICENSE SUSP FOR DWI	
DRUG PARAPHERNALIA IN DRUG FACTORY	F
DRUGS NEAR PROIBITED PLACE	
DUI .02 BAC AND UNDER 21 YRS	
EMPLOYING MINOR IN OBSCENE PERFORMNCE AF	
ENTICING A MINOR	
ESCAPE FROM CUSTODY	
ESCAPE, FIRST DEGREE	CF
EVADING RESPONSIBILITY	M
FAILURE REGISTER SEXUAL PURPOSE	DF
FAILURE TO APPEAR, FIRST DEGREE	DF
FAILURE TO APPEAR, SECOND DEGREE	AM
FAILURE TO FURNISH TAX RETURN	
FAILURE TO OBEY AN OFFICER	
FAILURE TO PAY OR PLEAD	M
FALSE STATEMENT, 2ND DEGREE	AM
FALSELY REPORTING AN INCIDENT 2ND	AM
FALSELY REPORTING INCIDENT	



FEDERAL OFFENSE	
FELONY MURDER	AF
FINES FOR VIOLATIONS	
FLR KEEP NARC IN ORG CONTAINER	F
FORGERY, FIRST DEGREE	CF
FORGERY, SECOND DEGREE	DF
FORGERY, THIRD DEGREE	BM
HARASSMENT 1ST DEGREE	DF
HARASSMENT, 2ND DEGREE	CM
HINDER PROSECUTION DEFINED	F
HINDER PROSECUTION, FIRST DEGREE	DF
HINDERING PROSECUTION 2ND DEG	
HOME INVASION	
HOME INVASION	AF
IDENTITY THEFT 1ST DEGREE	BF
IDENTITY THEFT 2ND DEGREE	CF
IDENTITY THEFT 3RD DEGREE	DF
IL POS LRG CAP MAG OBT AF 4/13	DF
ILL CARRY FIREARM-UI DRUGS/ALC	
ILL OP MV WO IGNITION DEVICE	CM
ILL POSSESS CHILD PORN 2ND DEG	F
ILL POSSESS CHILD PORN 3RD DEG	DF
ILLEGAL GAIN FROM CREDIT CARD	AM
ILLEGAL MANUFACTUAL OR SALE OF DRUGS	
ILLEGAL POSSES OF SHOP LIFTING DEVICE	AM
ILLEGAL POSSESSION OF EXPLOSIVES	F
ILLEGAL REFUSAL OF DNA TESTING	
ILLEGAL USE OF CREDIT CARD	
IMPERSONATION OF POLICE OFFICER	DF
IMPORTING CHILD PORNOGRAPHY	F
INCITING INJURY TO PERSONS OR PROPTY	CF
INDECENT ASSAULT	F
INJUNCTION AND RESTRAINING ORDER	
INJURY OR RISK OF INJURY TO MINOR	F
INJURY/RISK OF INJURY TO MINOR- SEXUAL C	
INSURANCE FRAUD	DF
INTERFERENCE WITH SEARCH	
INTERFERING WITH AN EMERGENCY CALL	AM
INTERFERING WITH AN OFFICER	AM
INTERSTATE CUSTODY COMPACT	
INTIMIDATE DUE TO BIAS 3RD DEG	AM
INTIMIDATING A WITNESS	CF
INTIMIDATION BSD BIGOTRY/BIAS	DF
ISSUING A BAD CHECK	
KIDNAPPING 1ST DEG W/ FIREARM	AF
KIDNAPPING 2ND DEG W/ FIREARM	BF
KIDNAPPING, 2ND DEGREE	BF
KIDNAPPING, FIRST DEGREE	AF

LARCENY DEFINED	
LARCENY, FIFTH DEGREE	BM
LARCENY, FIRST DEGREE	BF
LARCENY, FOURTH DEGREE	AM
LARCENY, SECOND DEGREE	CF
LARCENY, SIXTH DEGREE	CM
LARCENY, THIRD DEGREE	DF
LEAV CHILD UNSUPERV IN PUBL/MOT VEH	C
LIQUOR SALES TO MINORS	
MANSLAUGHTER 2ND WITH MV (INTOX)	CF
MANSLAUGHTER 2ND, W/FIREARM	CF
MANSLAUGHTER, FIRST DEGREE	BF
MANSLAUGHTER, FIRST DEGREE, FIREARM	BF
MANSLAUGHTER, SECOND DEGREE	CF
MANUF OR POSSESSION OF BURGLAR TOOLS	AM
MANUFACTURE OF BOMBS	BF
MISCONDUCT WITH A MOTOR VEHICLE	DF
MURDER	F
MURDER	AF
NEGLG STORAGE OF A FIREARM	DF
NEGLIGENT HOMICIDE W/ MOTOR VEHICLE	BM
OBSCENITY AS TO MINORS	F
OPERATE UNDER INFLU OF LIQ OR DRUGS	
OPERATE UNDER INFLU OF LIQ OR DRUGS	M
OPERATING MV W/O A LICENSE	
OPERATING MV W/O INSURANCE	CM
OPERATING UNDER SUSPENDED LIC OR REG.	M
ORGANIZED RETAIL THEFT	F
OUI MV W/CHILD PSNGR	AM
PATRONIZE PROSTITUTE FM MV	AM
PATRONIZING A PROSTITUTE	AM
PENALTY FAIL TO SUBMIT FINGERPRINTS	
PENALTY-EMPLOYER VIOL SEC. OF 31-71	
PERMITTING PROSTITUTION	AM
POSS OF ASSAULT WEAPON PROHIBITED	
POSS OF WEAPON IN A CCI	BF
POSS WEAPON ON SCHOOL GROUNDS	DF
POSSESS OF HALLUC OR 4+ OZ MARIJUANA	F
POSSESS OF LT 4 OZ MJ OR CONTROLD SUB	
POSSESSING CHILD PORNOGRAPH	F
POSSESSION OF NARCOTICS	
POSSESSION OF SHOTGUN OR SILENCER	DF
PRACTICE OF LAW BY NON-ATTORNEYS	M
PRIMARY OFFENSE NOT DETERMINED	
PROHIB ACTS RE: CONTROLLED SUBSTANCES	
PROHIB ACTS RE: DRUG PARAPHERNALIA	
PROHIBITED ACTIVITIES	
PROMOTING MINOR IN OBSCENE PERFORMNCE	BF

PROMOTING PROSTITUTION, FIRST DEGREE	BF
PROMOTING PROSTITUTION, SECOND DEGREE	CF
PROMOTING PROSTITUTION, THIRD DEGREE	DF
PROSTITUTION	AM
PUBLIC INDECENCY	BM
RAPE, SECOND DEGREE	CF
RECKLESS BURNING	DF
RECKLESS DRIVING	M
RECKLESS ENDANGERMENT, FIRST DEGREE	AM
RECKLESS ENDANGERMENT, SECOND DEGREE	BM
REG PERSON WHO COMM OFF AGAINST MINOR	DF
REG PERSON WHO COMM SEX VIOL OFF	DF
REGULATION OF RESTRICTED SUBSTANCES	
RIGGING SPORT EVENT	DF
RIOT, SECOND DEGREE	BM
ROBBERY BY CARJACKING	
ROBBERY, FIRST DEGREE	BF
ROBBERY, SECOND DEGREE	CF
ROBBERY, THIRD DEGREE	DF
SALE OF CONTROLLED SUBSTANCE	F
SALE OF HALLUCIGEN/NARCOTIC SUBSTANCE	F
SALE OF HEROIN, COC BY NON-DEPENDENT	
SALE OF NARC/AMPHET BY NON-DEPENDENT	F
SALE, CARRY & BRAND OF FACSIMILE	
SALE/POSS/TRAN OF ASLT WEAPON	F
SELLING WEAPONS TO ALIENS PROHIBITED	F
SEX ASLT, SPOUSE/COHAB RELATIONSHIP	BF
SEXUAL ASSAULT 1ST DEGREE-AGGRVTD	F
SEXUAL ASSAULT, 3RD DEG W/ FIREARM	CF
SEXUAL ASSAULT, FIRST DEGREE	F
SEXUAL ASSAULT, FOURTH DEGREE	AM
SEXUAL ASSAULT, SECOND DEGREE	F
SEXUAL ASSAULT, THIRD DEGREE	DF
SPECIAL PAROLE	
STALKING, 1ST DEGREE	DF
STALKING, 2ND DEGREE	AM
STALKING, 3RD DEGREE	BM
STEALING A FIREARM	DF
STRANGULATION 1ST DEGREE	CF
STRANGULATION 2ND DEGREE	DF
STRANGULATION 3RD DEGREE	
TAMPERING WITH OR FABRICATING EVIDNC	DF
TAMPERING WITH WITNESS	DF
THEFT OF UTILITY SERVICE FOR PROFIT	DF
THREATENING	AM
THREATENING 1ST DEGREE	DF
TRAFFICKING IN PERSONS	BF
UNLAWFUL DISCHARGE OF FIREARMS	M

```

UNLAWFUL REPRODUCTION OF CREDIT CARD  DF
UNLAWFUL RESTRAINT, FIRST DEGREE      DF
UNLAWFUL RESTRAINT, SECOND DEGREE     AM
USE OF MACHINE GUN IN A CRIME  DELETED F
USING MACHINE GUN IN CRIME             F
USING MOTOR VEHICLE W/O PERMISSION
VIO DEADLY WEAPON REGIST REQS         DF
VIOLATE COND OF RELEASE-2ND DEG       AM
VIOLATION OF CONDITIONS OF RELEASE
VIOLATION OF PROBATION OR COND DISCHG
VIOLATION, HOME IMPROVE REGISTRATION
VOYEURISM
YOUTHFUL OFFENDER
M
AMER IND
ASIAN
BLACK
HISPANIC
0

```

## 4 Code Documentation

Here are the code blocks, in order that I removed from the final report pdf.

```

In [ ]: import pandas as pd
        import numpy as np
        import scipy.linalg as la
        import scipy.stats as stats
        import statsmodels.api as sm
        from sklearn import linear_model, model_selection, metrics
        import sklearn
        import plotly.graph_objs as go
        import matplotlib.pyplot as plt
        import pprint
        import warnings
        warnings.filterwarnings('ignore')

In [ ]: # load in bjs data
        incar = pd.read_csv('incarceration_trends.csv')
        race = pd.read_csv('incarceration_by_race.csv')
        crime = pd.read_csv('crime_data.csv')

In [ ]: # print samples from the data sets
        print(incar[['Year', 'State prisons', 'Population']].sample(3))

In [ ]: race['White_rate']
        cols = ['Geography', 'Total', 'White', 'Black', 'White_rate', 'Black_rate']
        print(race[cols].sample(2))

```

```

In [ ]: cols = ['Year', 'Violent crime', 'Murder', 'Rape', 'Robbery', 'Assault']
        print(crime[cols].sample(3))

In [ ]: # load in connecticut data sets
        inmates = pd.read_csv('individuals.csv')
        race_regr_df = pd.read_csv('race_regression_df.csv')

In [ ]: # print samples
        cols = ['LATEST ADMISSION DATE', 'AGE', 'RACE', 'SENTENCE DAYS']
        print(inmates[cols].sample(3))

In [ ]: # disclose sample sizes for races
        races = {'BLACK': 'Blacks', 'WHITE': 'Whites',
                  'HISPANIC': 'Hispanic', 'AMER IND': 'American Indian',
                  'ASIAN': 'Asian'
                  }
        for rac in races.keys():
            mask = inmates.RACE == rac
            print(f'Sample size for {races[rac]}: {len(inmates[mask])}')

In [ ]: # plot the incarceration trends against crime
        fig, ax = plt.subplots(figsize=(10,5))
        ax.plot(
            incar['Year'].values,
            incar['State prisons'].values / (incar.Population*10**4),
            label='State Prisons', color='tab:blue')

        ax.plot(
            incar['Year'].values,
            incar['Local jails'].values / (incar.Population*10**4),
            label='Local Jails', color='tab:orange')
        ax2 = ax.twinx()
        ax2.plot(
            crime.Year, crime['Violent crime'],
            label='violent crime', color='tab:red'
        )
        ax2.set_ylabel('rate per 100,000')
        ax2.legend()
        ax.legend()
        ax.set_title('Incarceration and Crime Trends Over Time')
        ax.set_xlabel('Year')
        ax.set_ylabel('rate per 100,000')
        plt.show()

In [ ]: # plot histogram of sentence lengths
        fig, ax = plt.subplots(1, figsize=(10,5))
        ax.set_xlim(0,10000)
        mask = inmates['SENTENCE DAYS'] < 10000
        ax.hist(

```

```

        inmates['SENTENCE_DAYS'][mask].values,
        bins=200
    )
    plt.show()

In [ ]: # plot histograms of incarceration rates by race
fig, ax = plt.subplots(2,2, figsize=(10,7))
ax[0,0].hist(race['White_rate'])
ax[0,0].set_title('White Incarceration Rates')
ax[0,0].set_xlabel('rate per 100,000')

ax[0,1].hist(race['Black_rate'], color='tab:orange')
ax[0,1].set_title('Black Incarceration Rates')
ax[0,1].set_xlabel('rate per 100,000')

ax[1,0].hist(race['Hisp_rate'], color='tab:green')
ax[1,0].set_title('Hispanic Incarceration Rates')
ax[1,0].set_xlabel('rate per 100,000')

ax[1,1].hist(race['Asian_rate'], color='tab:gray')
ax[1,1].set_title('Asian Incarceration Rates')
ax[1,1].set_xlabel('rate per 100,000')
fig.tight_layout()
plt.show()

In [ ]: # plot other stats related to sentence lengths by race
races = inmates.groupby(by='RACE')
print(
    'Median sentence lengths:\n {}'.format(races['SENTENCE_DAYS'].median())
)
print('\n')
print(
    'Standard deviation of sentence lengths:\n {}'.format(races['SENTENCE_DAYS'].std())
)

In [ ]: # plot boxplot of sentence lengths
fig, ax = plt.subplots(1, figsize=(10,7))
ax.set_xlim(-50,20000)
bp = inmates.boxplot(
    ['SENTENCE_DAYS'], by=['RACE'],
    vert=False, grid=False,
    ax=ax
)
ax.set_title('distribution of sentence lengths')
ax.set_xlabel('sentence length in days')
plt.show()

In [ ]: # plot histograms of sentence lengths by race
fig, ax = plt.subplots(3,2, figsize=(8,5))

```

```

indexer = {
    (0,0): 'BLACK',
    (0,1): 'WHITE',
    (1,0): 'ASIAN',
    (1,1): 'HISPANIC',
    (2,0): 'AMER IND'
}
for key in indexer.keys():
    ax[key].set_xlim(-50,20000)
    mask = races.get_group(indexer[key])['SENTENCE DAYS'] < 20000
    ax[key].hist(
        races.get_group(indexer[key])['SENTENCE DAYS'][mask],
        bins=50
    )
    ax[key].set_xlabel('distribution of sentence length in days')
    ax[key].set_title(indexer[key])

fig.tight_layout()
plt.show()

```

In [ ]: *# plot the kurtosis of sentence lengths by race*

```

out = '{} Kurtosis: {}'
kurts = {
    'White':
        stats.kurtosis(races.get_group('WHITE')['SENTENCE DAYS']),
    'Black':
        stats.kurtosis(races.get_group('BLACK')['SENTENCE DAYS']),
    'Hispanic':
        stats.kurtosis(races.get_group('HISPANIC')['SENTENCE DAYS']),
    'American Indian':
        stats.kurtosis(races.get_group('AMER IND')['SENTENCE DAYS']),
    'Asian':
        stats.kurtosis(races.get_group('ASIAN')['SENTENCE DAYS']),
}

for key in kurts.keys():
    print(
        out.format(key, kurts[key])
    )

```

In [ ]: *# explore results of regression*

```

race_regr_df.dropna(inplace=True)
sentence = race_regr_df['SENTENCE DAYS']
race_regr_df.drop(
    ['SENTENCE DAYS', 'Unnamed: 0'], axis=1, inplace=True
)
X_train, X_test, y_train, y_test = model_selection.train_test_split(
    race_regr_df.astype(float), sentence, test_size=.3
)

```

```

)
res_gen = sm.OLS(y_train, X_train).fit()
print('Regression Results:\n')
print(f'\tMethod: Least Squares')
print(f'\tR-squared value: {res_gen.rsquared}')
print(f'\tAIC: {res_gen.aic}')
print(f'\tBIC: {res_gen.bic}')

In [ ]: # plot sentence lengths of people that are convicted of
# criminal liability
race_mask1 = inmates.RACE != 'AMER IND'
race_mask2 = inmates.RACE != 'ASIAN'
race_mask = np.logical_and(race_mask1, race_mask2)

mask = inmates[race_mask]['OFFENSE'] == 'CRIMINAL LIABILITY FOR ANOTHER PERSON'

fig, ax = plt.subplots(1, figsize=(5,5))
ax.set_xlim(-50,80000)
bp = inmates[race_mask][mask].boxplot(
    ['SENTENCE DAYS'], by=['RACE'],
    vert=False, grid=False,
    ax=ax
)
ax.set_title('Distribution of Sentence Lengths for CRIMINAL LIABILITY FOR ANOTHER PERSON')
ax.set_xlabel('sentence length in days')
plt.show()

In [ ]: # plot kurtosis related to criminal liability
mask = inmates[race_mask]['OFFENSE'] == 'CRIMINAL LIABILITY FOR ANOTHER PERSON'

out = '{} Kurtosis for\n\t{}: {}'.format(race_mask, mask)
kurts = {
    'White':
        stats.kurtosis(
            races.get_group('WHITE')['SENTENCE DAYS'][race_mask][mask]
        ),
    'Black':
        stats.kurtosis(
            races.get_group('BLACK')['SENTENCE DAYS'][race_mask][mask]
        ),
    'Hispanic':
        stats.kurtosis(
            races.get_group('HISPANIC')['SENTENCE DAYS'][race_mask][mask]
        )
}
# 'American Indian':
# stats.kurtosis(races.get_group('AMER IND')['SENTENCE DAYS'][mask]),
# 'Asian':

```



```

#     stats.kurtosis(races.get_group('ASIAN')['SENTENCE DAYS'][mask]),
}

for key in kurts.keys():
    print(
        out.format(key, 'CRIMINAL LIABILITY FOR ANOTHER PERSON', kurts[key])
    )

In [ ]: # plot a sample of offenses
race_mask1 = inmates.RACE != 'AMER IND'
race_mask2 = inmates.RACE != 'ASIAN'
race_mask = np.logical_and(race_mask1, race_mask2)
offenses = [
    'MURDER', 'AF',
    'SALE OF CONTROLLED SUBSTANCE', 'F',
    'SALE OF HALLUCIGEN/NARCOTIC SUBSTANCE', 'F',
    'ROBBERY, FIRST DEGREE', 'BF'
    # 'CRIMINAL LIABILITY FOR ANOTHER PERSON'
]
for off in offenses:

    mask = inmates[race_mask]['OFFENSE'] == off
    # inmates[mask]['RACE'].sample(10)
    fig, ax = plt.subplots(1, figsize=(5,5))
    lim = np.percentile(
        inmates['SENTENCE DAYS'][race_mask][mask],
        99
    )
    ax.set_xlim(-50, lim)
    bp = inmates[race_mask][mask].boxplot(
        ['SENTENCE DAYS'], by=['RACE'],
        vert=False, grid=False,
        ax=ax
    )
    ax.set_title(f'distribution of sentence lengths for {off}')
    ax.set_xlabel('sentence length in days')
    plt.show()
    out = '{} Kurtosis: {}'.format(off, kurts[off])
    kurts[off] = {
        'White':
            stats.kurtosis(
                races.get_group('WHITE')['SENTENCE DAYS'][race_mask][mask]
            ),
        'Black':
            stats.kurtosis(
                races.get_group('BLACK')['SENTENCE DAYS'][race_mask][mask]
            ),
        'Hispanic':

```

```

stats.kurtosis(
    races.get_group('HISPANIC')['SENTENCE DAYS'][race_mask][mask]
)
# 'American Indian':
# stats.kurtosis(races.get_group('AMER IND')['SENTENCE DAYS'][mask]),
# 'Asian':
# stats.kurtosis(races.get_group('ASIAN')['SENTENCE DAYS'][mask]),
}

for key in kurts.keys():
    print(
        out.format(key, kurts[key])
    )

```