Group 10

Project Title: ARMORY CONTROL IN USA

Course: IE6600 Computation & Visualization

- DEEKSHITH REDDY CHAGAMREDDY
- AKHIL VARDURU
- VEDA UPASAN PEDAGADI

INTRODUCTION

- The project deals with analyzing the data to understand the relationship between crime incarceration rates and crime rates in order to analyze and answer the question of whether or not it improved public safety. The dataset thus consists of data on crimes committed and prisoner countin every 50 state and the interested crime types to analyze are violent crime total, murder manslaughter, and aggravated assaults. The important and significant features would help in determining the crime estimated within the state and help predict the whether or not the state changed reporting system had an affect with comparison to the previous years in United States
- Limiting access to weapons, according to proponents of gun control, makes it more difficult for criminals to get them and use them to breach the law.
- Stronger gun prohibitions result in increased complexity in the illegal activity and a drop in crime.
- This self-defense or "good man with a gun" effect suggests that more weapons would be used, increasing the cost of criminal action.
- The State Weapon Regulations initiative aims to provide scientists with the data they need to evaluate the efficacy of various gun control measures.

Users Stories

General Person

There are many who are interesting in keeping guns or using for animal hunting or safety purpose.

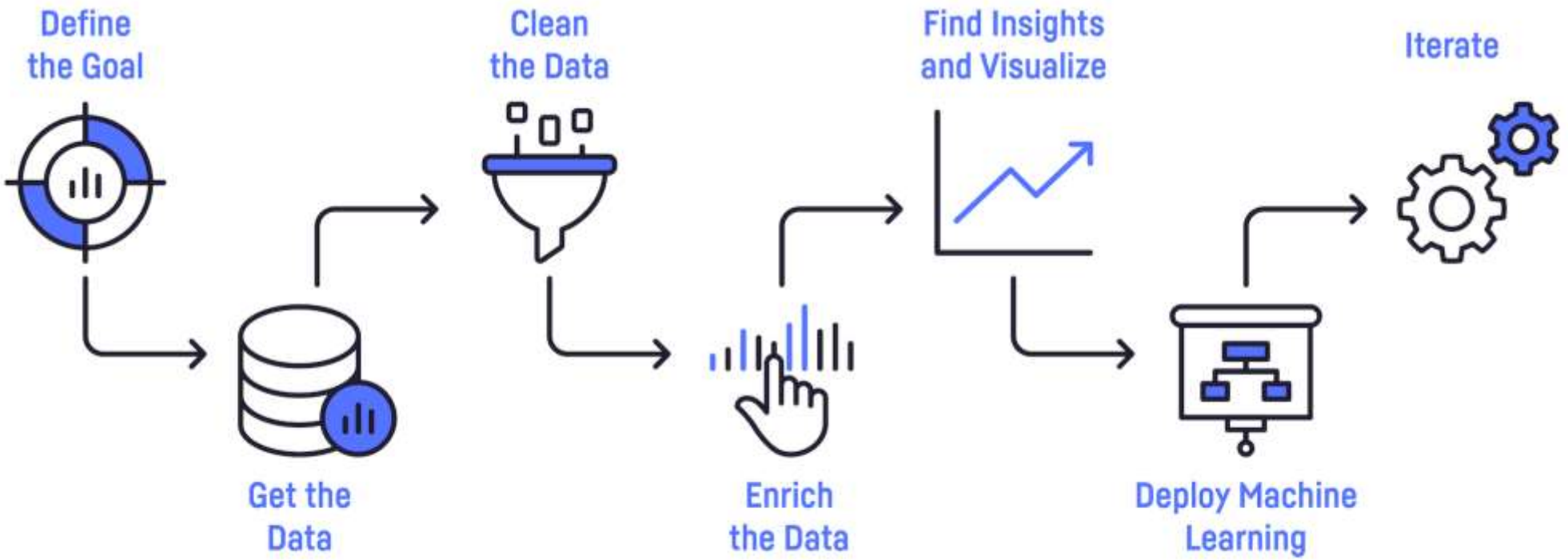
Social Person

Gun owners consider their weapons to be tools that may be used for good or harm. This explains why many gun owners believe that legislation should center on "keeping weapons out of the wrong hands."

Government

Background checks on private gun transactions and restrictions on the sale of firearms to those who are mentally ill or on no-fly or watch lists are both widely supported by both gun owners and non-owners in both parties

- Implementation



References

- Goyal, C. (2021b, July 1). How to Detect and Remove Outliers | Outlier Detection And Removal.
  - Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/05/feature-engineering- how-to-detect-and-remove-outliers-with-python-code/
  - Kumar, A. (2022a, January 23). Python - Replace Missing Values with Mean, Median & Mode. Data Analytics. https://vitalflux.com/pandas-impute-missing-values-mean-median-mode/
  - Visualizing distributions of data — seaborn 0.11.2 documentation. (2021). Seaborn. https://seaborn.pydata.org/tutorial/distributions.html
  - sklearn.linear_model.LinearRegression. (2022). Scikit-Learn. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html
- Decision Tree Regression. (2022). Scikit-Learn.
  - https://scikit-learn.org/stable/auto_examples/tree/plot_tree_regression.html
  - sklearn.ensemble.RandomForestRegressor. (2022). Scikit-Learn. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html
  - D. (2019b, October 10). Regression Accuracy Check in Python (MAE, MSE, RMSE, R- Squared). Data Tech Notes. https://www.datatechnotes.com/2019/10/accuracy-check-in- python-mae-mse-rmse-r.html

Exploratory Data Analysis

- The exploratory data analysis is performed on the crime and incarceration data of United States for 50 states from year 2001 to 2016 which contains **17 data field values** and **816 rows of data** having categorical, numerical, and boolean type data. The different parameters of the dataset are state, year, prisoner count, crimes estimated, violent crime total, rape legacy, and other various types of crime data.

```
Column Names:
 Index(['jurisdiction', 'includes_jails', 'year', 'prisoner_count',
       'crime_reporting_change', 'crimes_estimated', 'state_population',
       'violent_crime_total', 'murder_manslaughter', 'rape_legacy',
       'rape_revised', 'robbery', 'agg_assault', 'property_crime_total',
       'burglary', 'larceny', 'vehicle_theft'],
      dtype='object')
```

Loading the dataset

```
In [3]: crimes_data = pd.read_csv("/Users/deekshith/Desktop/crime_and_incarceration_by_state.csv")
        crimes_data
```

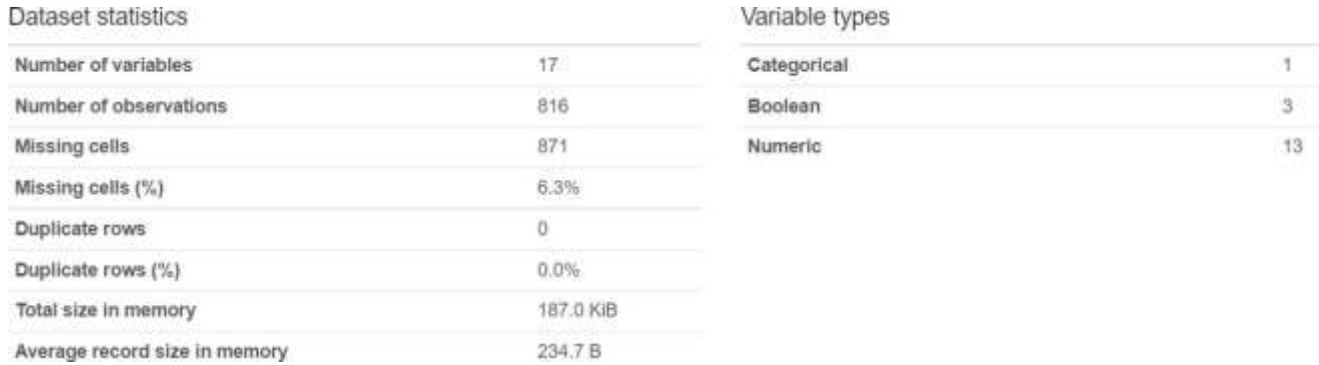
Out[3]:

	jurisdiction	includes_jails	year	prisoner_count	crime_reporting_change	crimes_estimated	state_population	violent_crime_total	murder_manslaughter
0	FEDERAL	False	2001	149852	NaN	NaN	NaN	NaN	NaN
1	ALABAMA	False	2001	24741	False	False	4468912.0	19582.0	379.0
2	ALASKA	True	2001	4570	False	False	633630.0	3735.0	39.0
3	ARIZONA	False	2001	27710	False	False	5306966.0	28675.0	400.0
4	ARKANSAS	False	2001	11489	False	False	2694698.0	12190.0	148.0
...	...	...	...	...	...	...	...	...	...
811	VIRGINIA	False	2016	29882	False	False	8414380.0	18495.0	482.0
812	WASHINGTON	False	2016	17228	False	False	7280934.0	22101.0	195.0
813	WEST VIRGINIA	False	2016	5899	False	False	1828637.0	6633.0	85.0
814	WISCONSIN	False	2016	23163	False	False	5772917.0	17716.0	232.0
815	WYOMING	False	2016	2352	False	False	584910.0	1431.0	20.0

816 rows × 17 columns

Data Profiling

- The data profiling report gives a detailed overview of the dataset that we are trying to analyze. The dataset statistics shown in the below figure helps understand the different statistics of the data with respect to missing data, duplicate data values, variable types, and number of observations.
- Here, we can observe that there are **871 missing values** which is **6.3% missing cells** in the dataset having **no duplicate rows**, which has 13 numerical data, 1 categorical data, and 3 boolean data values.



Dataset statistics

Number of variables	17
Number of observations	816
Missing cells	871
Missing cells (%)	6.3%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	187.0 KiB
Average record size in memory	234.7 B

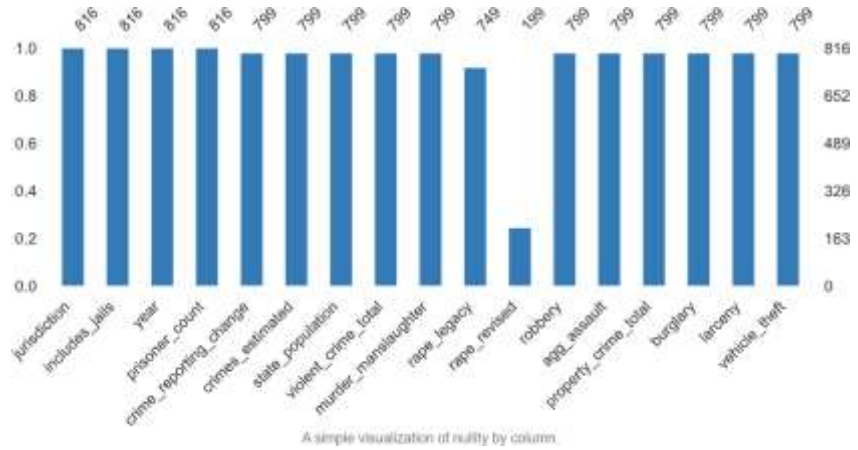
Variable types

Categorical	1
Boolean	3
Numeric	13

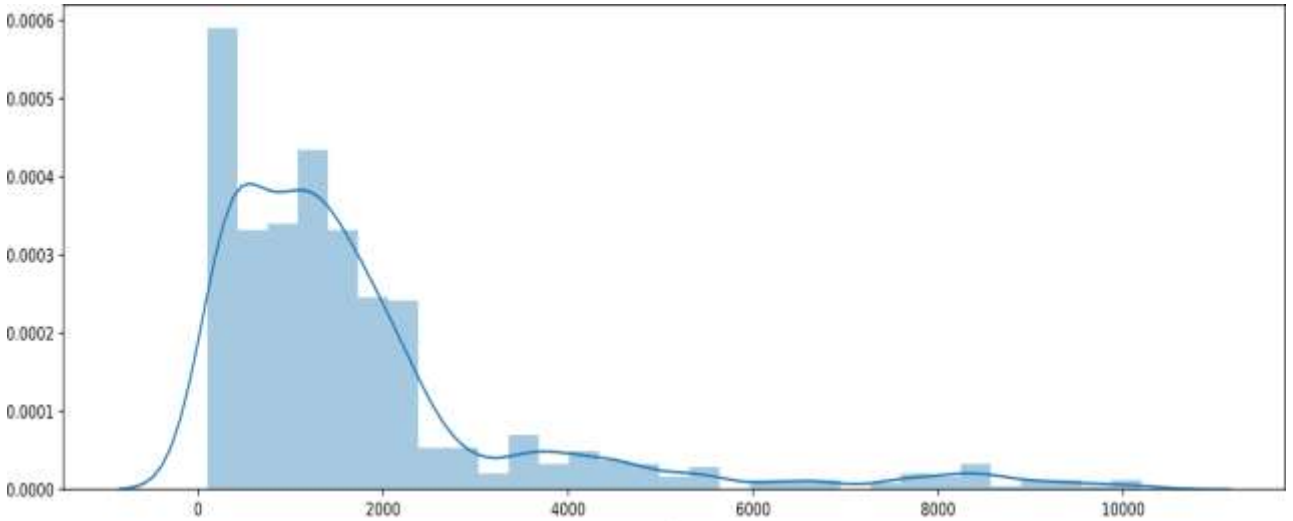
Data Cleaning

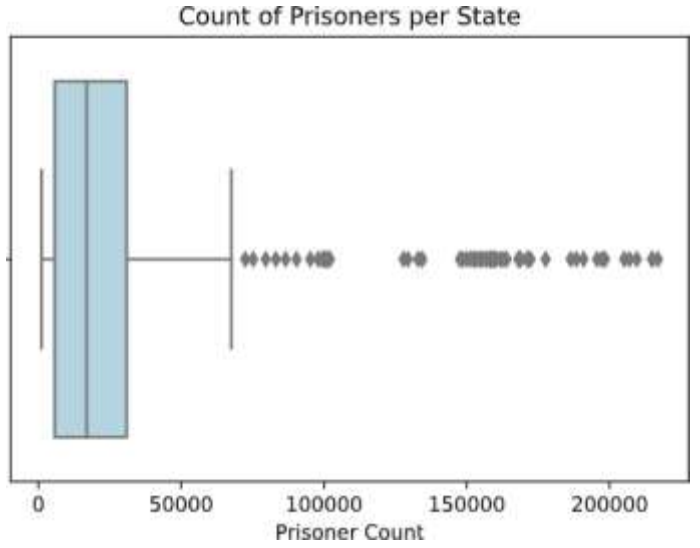
- The data cleaning is an important step to be performed in the analysis process and from the above dataset statistics we observe that there are **871 missing values** which we need to consider for cleaning.
- In order to determine which field values consists of the missing values, the count of missing values for each column is displayed. Apart from this, the data profiling report generates the missing values plot which gives a better understanding of the missing data

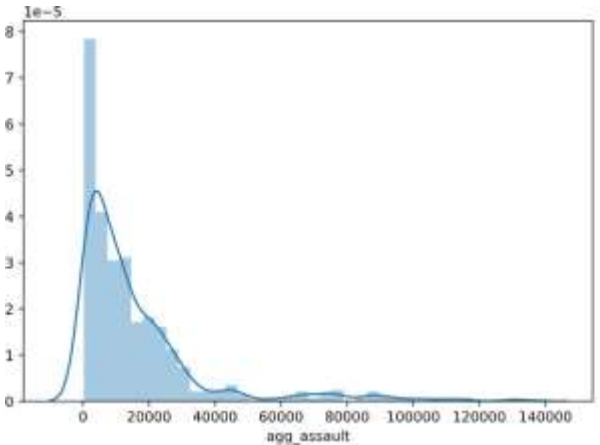
jurisdiction	0
includes_jails	0
year	0
prisoner_count	0
crime_reporting_change	17
crimes_estimated	17
state_population	17
violent_crime_total	17
murder_manslaughter	17
rape_legacy	67
rape_revised	617
robbery	17
agg_assault	17
property_crime_total	17
burglary	17
larceny	17
vehicle_theft	17

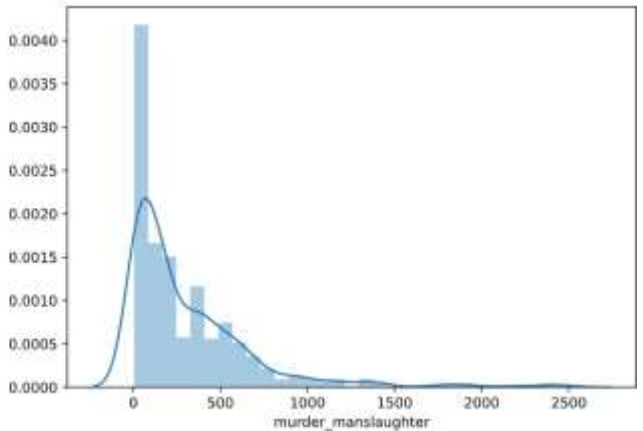


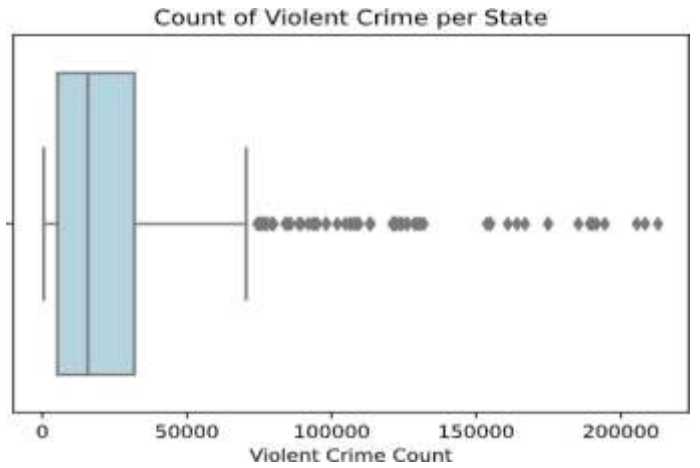
- The data is right skewed and hence the median operation would be the best to perform for imputation of the missing data. If the data was normally distributed, the mean or median could have been used but since data is skewed, the median approach is used to fill in for the missing values.





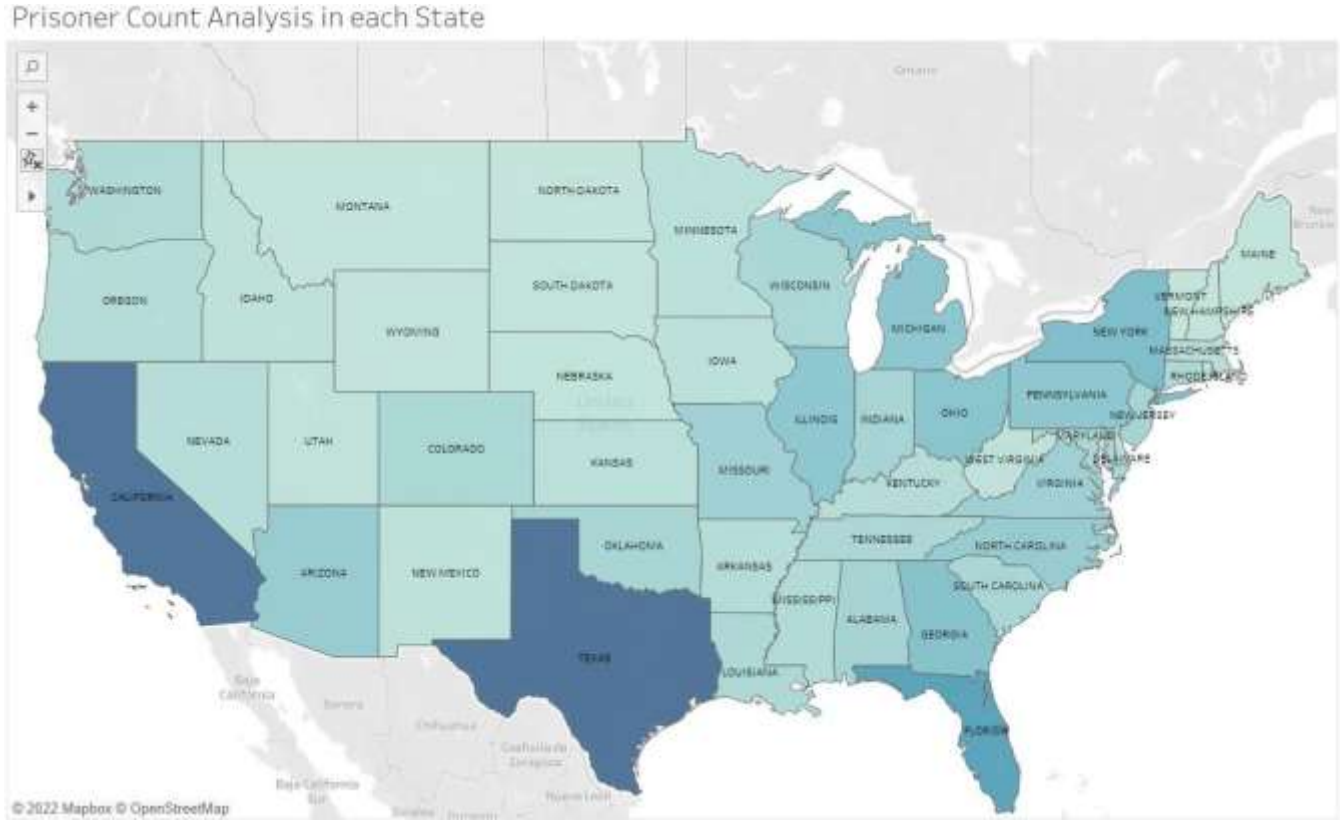




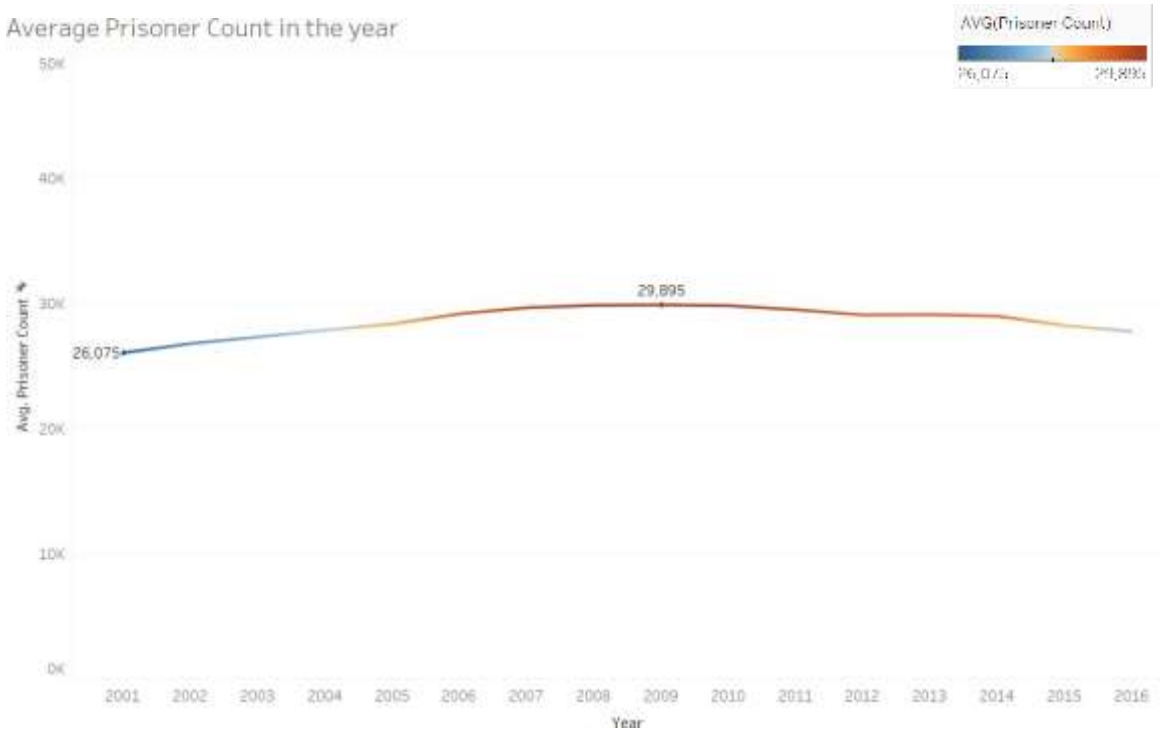


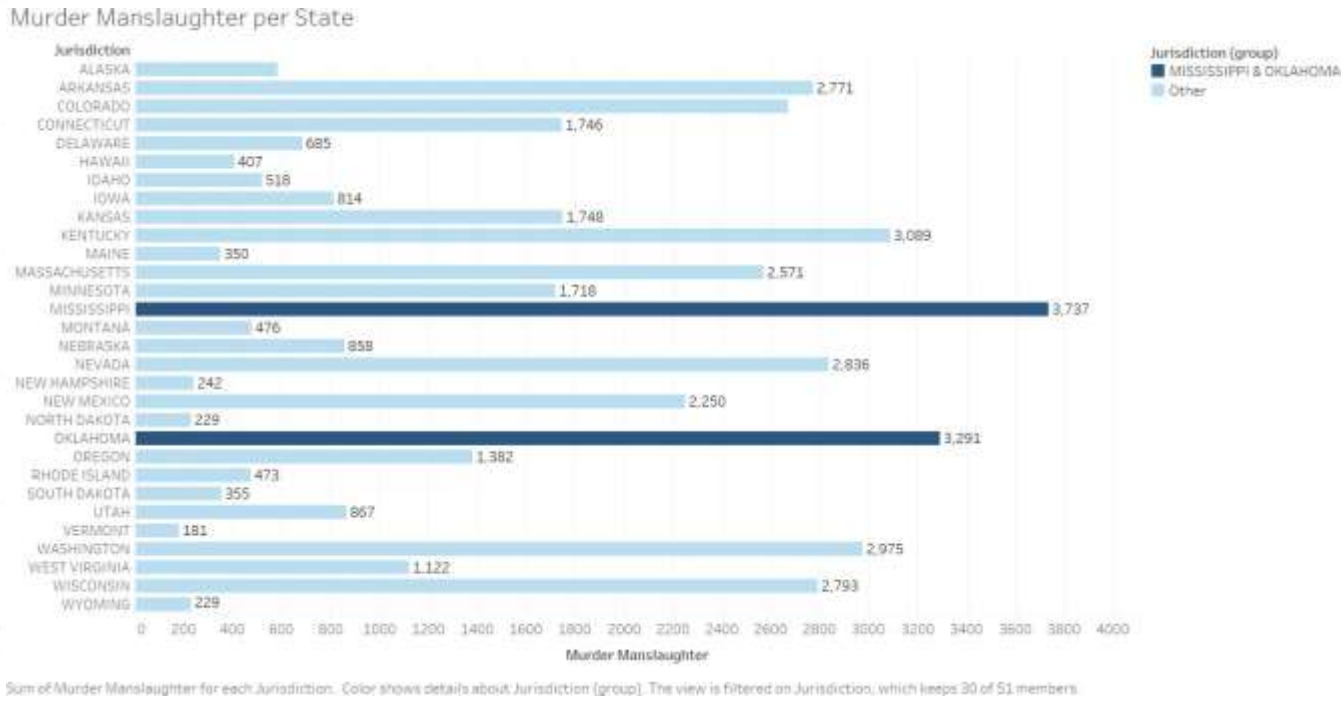
Data Visualization

- Data visualization helps in analyzing and presenting the data in an effective way through the graphical representations. The crime and incarceration dataset contains various parameters for which the visualizations would help in gaining some insights from the data and also helping to analyzing the relationship between the parameters.

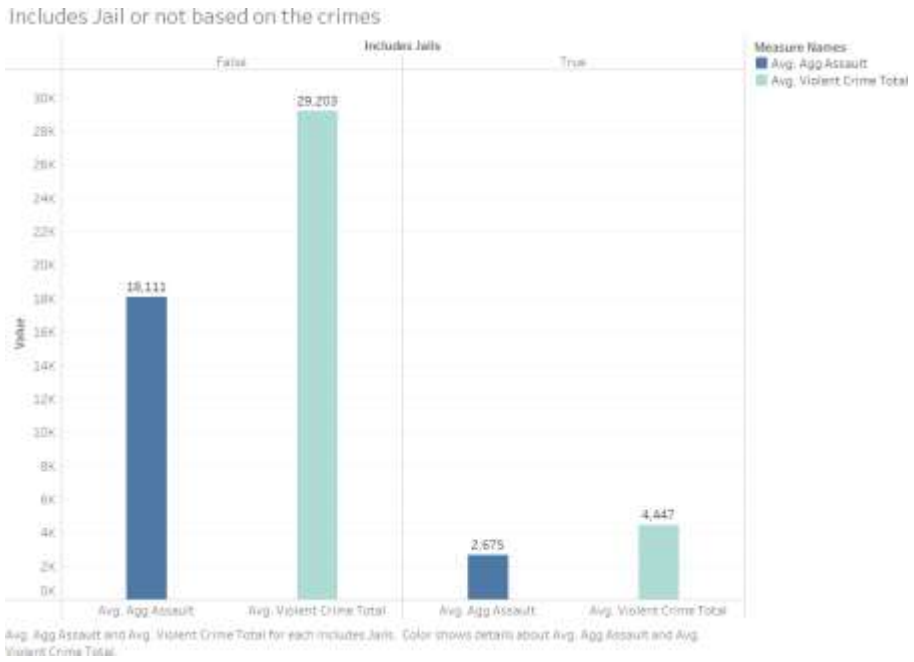


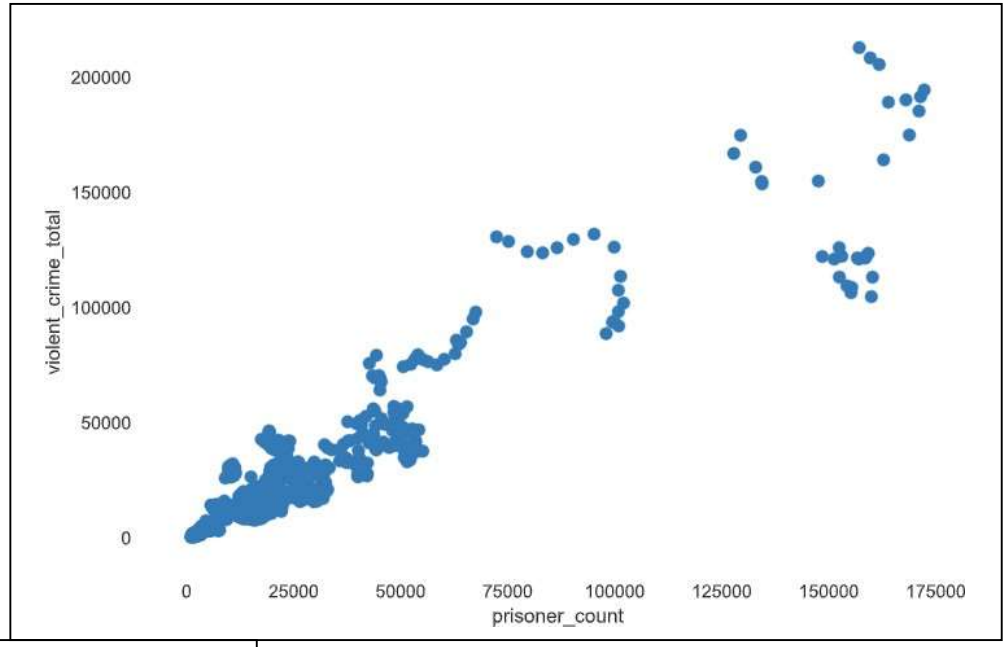
- As observed, the highest average prisoner count of 29,895 is in the year 2009 and the lowest average prisoner count of 26,075 is in the year 2001. Also, after the year 2009 the average count of prisoners decreased until 2016.

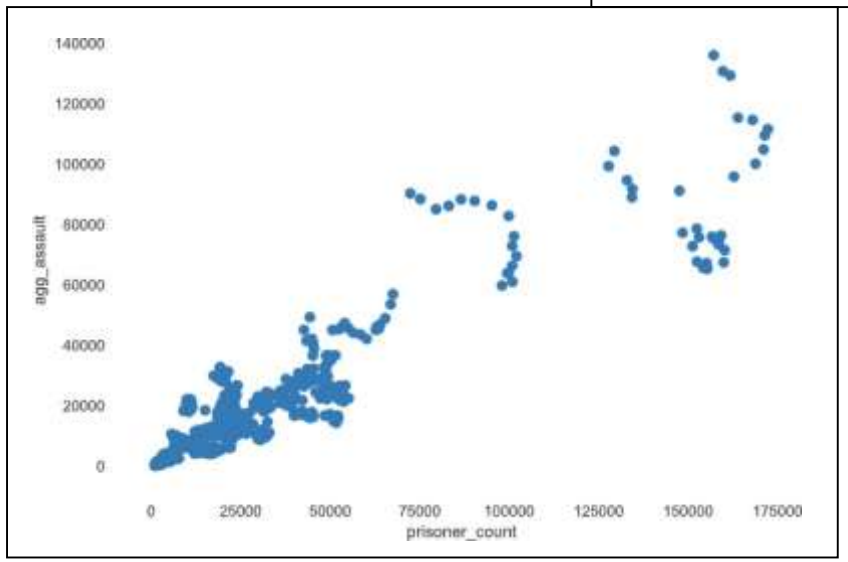




- Next visual of whether the crime include jail or not is based on the two crime types, namely, average violent crime total and average aggravated assault crime type







Label Encoding

- Since the dataset consists of categorical data which needs to be considered as the features for model building, the independent variables need to be label encoded in order to have numerical data passed to the model. Thus, the features such as ‘includes_jails’, and ‘jurisdiction’ are label encoded to numerical form which can be considered as features to training of the model

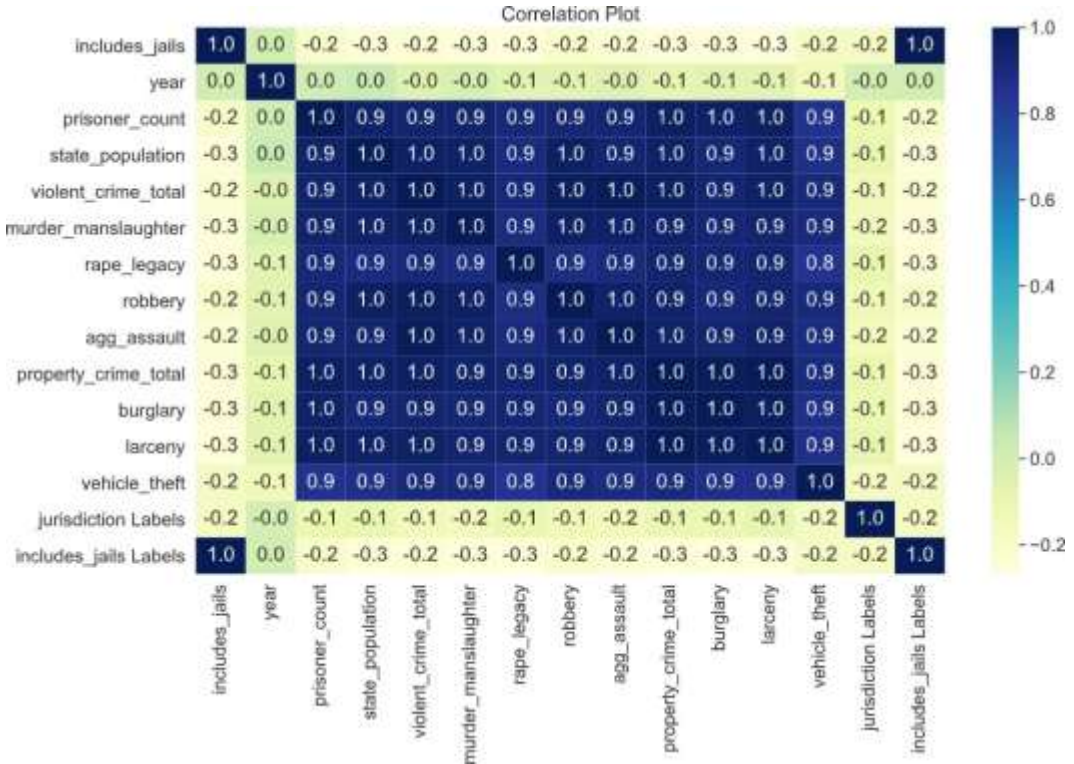
crimes_estimated	state_population	violent_crime_total	murder_manslaughter	rape_legacy	robbery	agg_assault	property_crime_total	burglary	larceny	vehicle_theft	jurisdiction Labels	includes_jails Labels
False	4468912.0	19582.0	379.0	1369.0	5584.0	12250.0	173253.0	40642.0	119992.0	12619.0	0	0
False	633630.0	3735.0	39.0	501.0	514.0	2681.0	23160.0	3847.0	16695.0	2618.0	1	1
False	5306966.0	28675.0	400.0	1518.0	8868.0	17889.0	293874.0	54821.0	186850.0	52203.0	2	0
False	2694698.0	12190.0	148.0	892.0	2181.0	8969.0	99106.0	22196.0	69590.0	7320.0	3	0
False	34600463.0	212867.0	2206.0	9960.0	64614.0	136087.0	1134189.0	232273.0	697739.0	204177.0	4	0
False	4430989.0	15492.0	158.0	1930.0	3555.0	9849.0	170887.0	28533.0	121360.0	20994.0	5	0
False	3434602.0	11492.0	105.0	639.0	4183.0	6565.0	95299.0	17159.0	65762.0	12378.0	6	1
False	796599.0	4868.0	23.0	420.0	1156.0	3269.0	27399.0	5144.0	19476.0	2779.0	7	1
False	16373330.0	130713.0	874.0	6641.0	32867.0	90331.0	782517.0	176052.0	516548.0	89917.0	8	0
False	8405677.0	41671.0	598.0	2180.0	14402.0	24491.0	347872.0	71799.0	238484.0	37589.0	9	0

Selection & Extraction

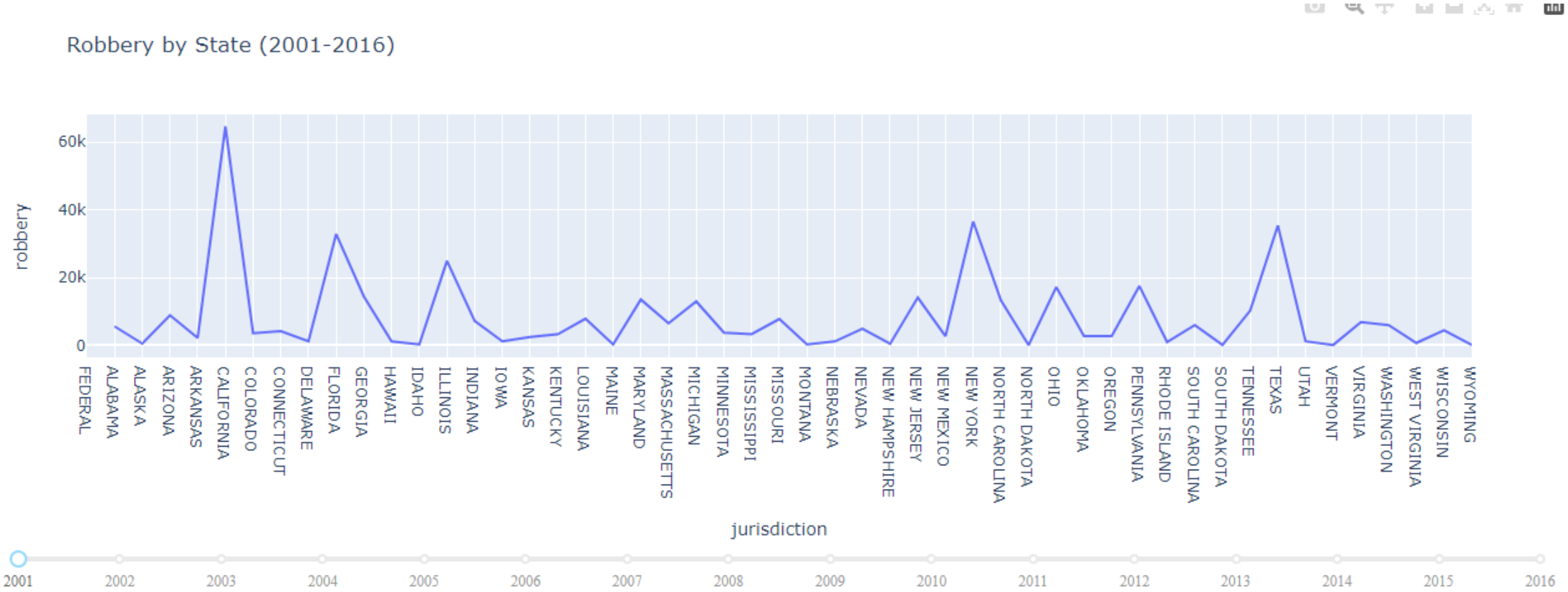
The automatic feature selection and extraction is performed in order to extract the important and significant features from the dataset which would help in the prediction of the prisoner count. The features that were selected are **'state_population', 'jurisdiction', 'includes_jails', 'crime_reporting_change', and 'crimes_estimated'**. These features would help in the prediction of prisoner count in each state and to also understand the various crimes that are affecting the prisoner count in each state

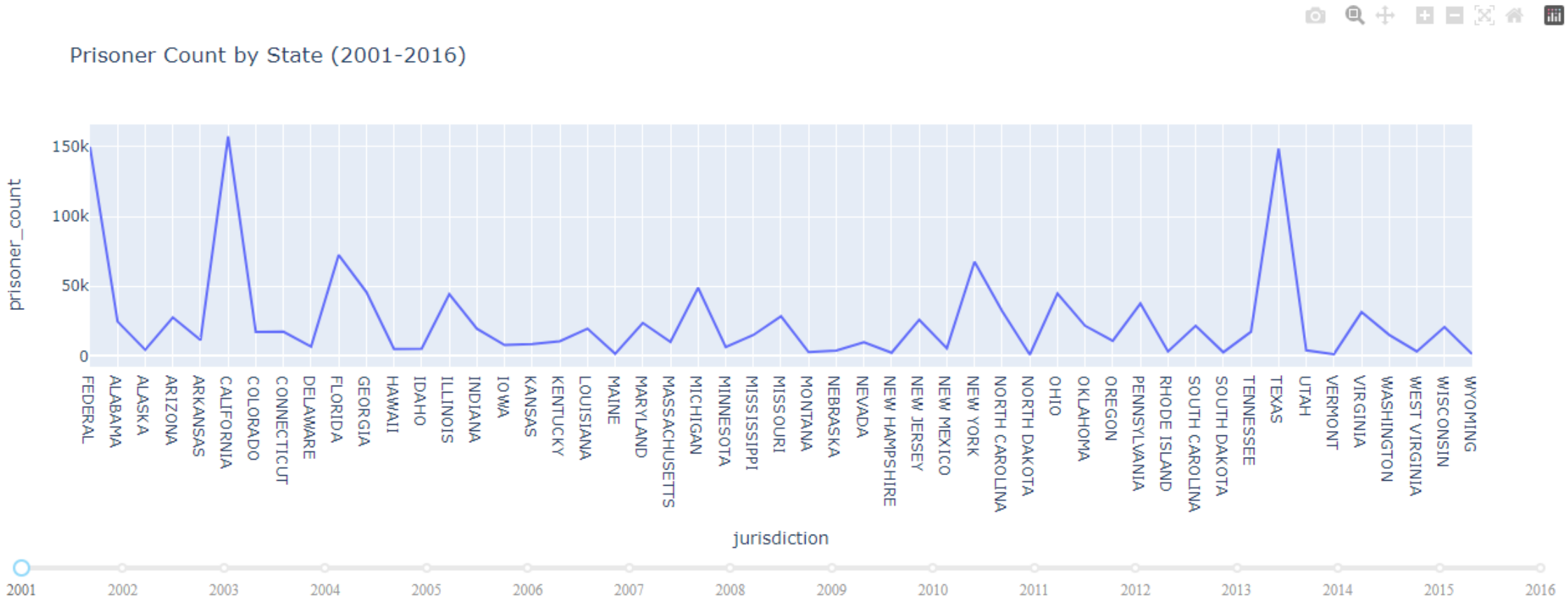
Correlation Plot

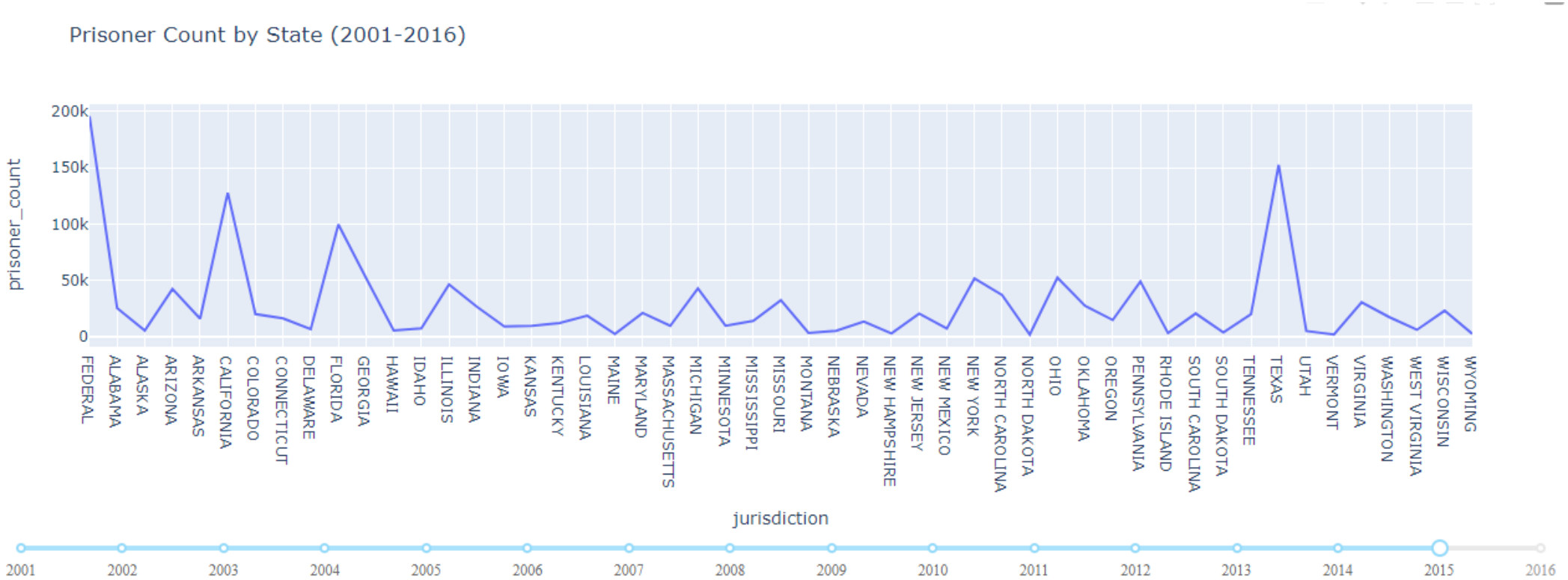
- The correlation plot helps to understand the correlation between each of the independent variables of the dataset and to plot the correlation values of the parameters
- From the correlation plot it is observed that since there are different crime variables having the same value, there is a **high collinearity** that is existing between the variables with a correlation value of **0.9**



Web App


Robbery by State (2001-2016)
robbery
60k
40k
20k
0
FEDERAL
ALABAMA
ALASKA
ARIZONA
ARKANSAS
CALIFORNIA
COLORADO
CONNECTICUT
DELAWARE
FLORIDA
GEORGIA
HAWAII
IDAHO
ILLINOIS
INDIANA
IOWA
KANSAS
KENTUCKY
LOUISIANA
MAINE
MARYLAND
MASSACHUSETTS
MICHIGAN
MINNESOTA
MISSISSIPPI
MISSOURI
MONTANA
NEBRASKA
NEVADA
NEW HAMPSHIRE
NEW JERSEY
NEW MEXICO
NEW YORK
NORTH CAROLINA
NORTH DAKOTA
OHIO
OKLAHOMA
OREGON
PENNSYLVANIA
RHODE ISLAND
SOUTH CAROLINA
SOUTH DAKOTA
TENNESSEE
TEXAS
UTAH
VERMONT
VIRGINIA
WASHINGTON
WEST VIRGINIA
WISCONSIN
WYOMING
jurisdiction
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016


Prisoner Count by State (2001-2016)
prisoner_count
150k
100k
50k
0
FEDERAL
ALABAMA
ALASKA
ARIZONA
ARKANSAS
CALIFORNIA
COLORADO
CONNECTICUT
DELAWARE
FLORIDA
GEORGIA
HAWAII
IDAHO
ILLINOIS
INDIANA
IOWA
KANSAS
KENTUCKY
LOUISIANA
MAINE
MARYLAND
MASSACHUSETTS
MICHIGAN
MINNESOTA
MISSISSIPPI
MISSOURI
MONTANA
NEBRASKA
NEVADA
NEW HAMPSHIRE
NEW JERSEY
NEW MEXICO
NEW YORK
NORTH CAROLINA
NORTH DAKOTA
OHIO
OKLAHOMA
OREGON
PENNSYLVANIA
RHODE ISLAND
SOUTH CAROLINA
SOUTH DAKOTA
TENNESSEE
TEXAS
UTAH
VERMONT
VIRGINIA
WASHINGTON
WEST VIRGINIA
WISCONSIN
WYOMING
jurisdiction
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016



[click here for Web App]

Conclusion

- This application is developed to meet the goals of the users. It helps in looking at the laws for controlling weapons and learn about crime and incarceration in the US to address lawmakers about the legalization of firearms that cannot be made available to everyone.


THANK YOU!