IOD Capstone Project
East London Metropolitan Police - Usage of Force

METROPOLITAN POLICE

Joel Domingo

Introduction

Use of force has increasingly come into the spotlight of mainstream media outlets globally over the past 5 years. It has had major negative implications on sentiment and behavior of police and public alike. Positive public sentiment on police performance has declined, proportional to police retaliation against crime, and in turn triggering further aggression against police, creating a cycle of steadily increasing aggression and safety risk.

This concept has been dubbed 'The Ferguson Effect', the idea that increased public criticism and distrust of police following the 2014 shooting of Michael Brown in Ferguson Missouri.

One particular case which illustrates this cycle is the case of George Floyd. For 8 minutes and 46 seconds, U.S Police officer Derek Chauvin used overwhelming force on an unarmed African American civilian resulting in fatality, causing protests (some violence) around the globe and creating pressure on police departments and politicians to reform law-enforcement tactics.

This presents an opportunity for police departments around the globe to investigate and explore the factors which lead to these behaviours. If done correctly, police agencies and local governments can use this data to develop a system as a learning loop for constant performance improvement, and become more responsive to public needs and mindful of the impact of policing efforts.

According to studies shown by the University of Massachusetts and Arizona State University in 2010, 'law enforcement with some college education use force much less often than those with no higher education'. Particularly, at the time of the study, 'only 1% of police forces within the United States have education requirements for those looking to join'. Ultimately, studies have shown that use of excessive force is attributed to a question or preparedness, and thus, we will be exploring ways in which data can increase factors of preparedness in police forces.

In 2019, a report indicated that UK police out-of-court settlements totalled £30 Million between 2015 and 2019. The payments ranged from small sums for loss for 'embarrassment and humiliation' through to six-figure settlements paid for wrongful arrest, according to records revealed under the Freedom of Information Act. Beyond this, the Metropolitan Police were responsible for £7.9m to settle 479 claims categorised for being 'malfeasance'.

The issues this report aims to solve is to find potential measures which decrease risk to officer and civilian safety by increasing police department preparedness to respond to incidents in order to:

1. Reduce likelihood of settlement payout from government funds for malfeasance and police brutality lawsuits.
2. Reduce the usage of police force to a point of 'last resort'.

This report will to develop an analytically-based system in such an area, with the purpose of influencing strategies that mitigates the subjectivity which influences violence-oriented performance of officers in public incidents and arrests.

Specifically, we will be looking at data provided by the London Datastore, operated by the UK government, which reports historical figures contained in incident reports dating from April 2020 to February 2021, and contains features such as dates and time figures, demographic data on incidents and subjects, and officer-related data, and use-of-force. The report will illustrate trends in numerical and categorical data which will be used to develop a model that can predict the likelihood and probability that an officer will require the usage of force to arrest or detain a subject prior to encounters. Furthermore, analysis of such features will be used to develop recommendations to police department protocols and resource allocation which promote the safety of officers and civilians alike. Despite the specificity of the data, the model of analysis is predicted to be adjusted for use around the world.

Model Expectations:

Inputs:	Categorical variables regarding subject, location, weapons on scene etc.
Functions:	Provide whether use of force will be necessary when responding to an incident, as well as the probability/likelihood that it will be required.
Output:	**Boolean value:** 1 - Force usage expected 2 - No force expected **Probability:** If no force expected: 1 - Probability of no force expected. If force expected: Probability of 1.
Usability:	**Allocate** resources and deploy officers suitable for incident.
Example Input:	Incident on street/highway. Male subject ages between 18-34, no apparent disability, identified with knife, located in Lewisham.
Example Output:	1 - Force Expected 0.75 probability of force expected
Example Recommendations	Deploy experienced partnered officers trained and equipped with taser and safety equipment.

Stakeholders

Stakeholder	Estimated project impact on stakeholder	Reason
Police Department (London)	High	Project reflects on their local statistics on which analysis is performed to create our models. Models are relevant to this particular population.
Police Departments (around the world)	High	Models can be developed using this report and catered towards different local police departments around the world to address issues of police usage of force.
Civilians	Medium	The safety of civilians is always of high importance, however it is dependent on the police force presiding over them to implement the recommendations, and the effects of which will show in civilian behaviour over time.
Officers	High	Our models and analytics will directly affect officers on the front line, and aim to address issues to their safety and conduct.
Government	Low	Changes in police conduct as a result of our analysis can manifest itself in millions of dollars in savings from reduction in lawsuits regarding abuse/assault, malfeasance and excessive use of force.

Data Overview

The data will be used to answer questions regarding the frequency in which force is used in effecting arrests in the East-London districts, as well as the factors involved in this distribution. Furthermore, analysis on features will reveal any contingencies which may be implemented to police department protocols and procedures to reduce risk to officer safety.

Source of data: (https://data.london.gov.uk/dataset/use-of-force)

Author: Metropolitan Police Service

Description: Collection of data-points which reflect incident reports submitted by the London Metropolitan Police Service. The data contains information on incidents presiding in the borough's of East-London.

Size: Data contains 147,895 incident reports, each with 271 features.

Reliability: Data sourced from the UK Government can be deemed reliable.

Quality: The raw data consists of a high volume of categorical data, with minimal missing values.

Time-relevance: Data contains information between the dates 01/04/2020 - 31/01/2021.

Generation: Data was generated from standardized police incident reports.

Availability: Figures are updated and reconciled each month from the start of each FY to the latest completed month.

Core Features (condensed):

Incident Date	Assaulted with weapon	Effectiveness of tactic	Subject ethnicity
Incident Time	Impact Factors	Level of CED usage	Subject disabilities
Incident Location	Reasons for force	Firearm aimed	Staff Injury
Borough (district)	Main duty of officer	Firearm fired	Staff medical provided
Primary Conduct of subject	Single-crewed or partnered	Subject age	Subject injury
Assaulted by Subject	Trained in CED (taser)	Subject Gender	Subject medical offered/provided
Threatened with weapon	Tactics in arrest	Subject ethnicity	Outcomes

Data Science Process

Feature Engineering and Numeration:
Some features of the dataset contain 'string' variables with different classes. For our model to be effective, it must be converted to numeric classes, each with different levels of severity. This applies to the following features:

- Primary Conduct of Subject
- Tactics

Primary Conduct of Subject:
Incident reports specify 6 classes of resistance levels. These are as following:

- Complaint
- Verbal resistance / gestures
- Passive resistance
- Active resistance
- Aggressive resistance
- Serious or aggravated resistance

Therefore, each class will be assigned a number, with each subsequent number indicating a higher level of aggression.

Class	Class Number
Compliant	0
Verbal resistance / gestures	1
Passive resistance	2
Active resistance	3
Aggressive resistance	4
Serious or aggravated resistance	5

<u>Levels of Force - Tactics</u>

Numerating this feature is essential in order to determine if level of force was used. In the original dataset, the following tactics are used.

Compliant handcuffing	CED (Taser) red-dotted	Firearm aimed
Tactical communications	Limb/body restraints	Batton drawn
Non-compliant handcuffing	Other/improvised	Spit guard
Unarmed skills (including pressure points, strikes, restraints, take-downs)	CED (Taser) drawn	Baton used
Ground restraint	CED (Taser) fired	Dog deployed
Irritant spray	Dog bite	Shield
CED (Taser) arced	CED (Taser) arced	CED (Taser) drive stun
CED (Taser) angle drive stun	AEP aimed	Firearm fired

It is paramount to distinguish as to what counts as force. Therefore, it will be distinguished by identifying anything other than 'Compliant handcuffing' and 'Tactical communications' as 'use of force'.

Tactic Used	Class
Compliant Handcuffing	0
Tactical Communications	0
Other	1

Subject Age

The original dataset has the following ranges of age classes:

1 - 10
11 - 17
18 - 34
35 - 49
50 - 64
65 +

To be viable for our models, we require them be in numeric form (classed).

Age	Class Number
1 - 10	1
11 - 17	2
18 - 34	3
35 - 49	4
50 - 64	5
65 +	6

Feature Engineering - Use of Force

Using the tactic classes, we are able to distinguish whether force by our definition, was used in a specific incident. This column will be defined as the sum of all tactics, and if the sum is greater than or equal to 1, then force was indeed used in that incident.

Example 1 - No force used

	Tactic 1	Tactic 2	Tactic 3	Was Force Required?
Binary	0	0	0	0
Translation	Tactical Communications	Compliant handcuffing	null	No force required.

Example 2 - Force was required

	Tactic 1	Tactic 2	Tactic 3	Was Force Required?
Binary	0	1	0	1
Translation	Tactical Communications	Unarmed skills	Compliant handcuffing	Force required.

Creating Dummy Variables
Due to the fact that a large portion of our features are categorical, dummy variables are essential to create a numerical model. The following features require dummy variables.

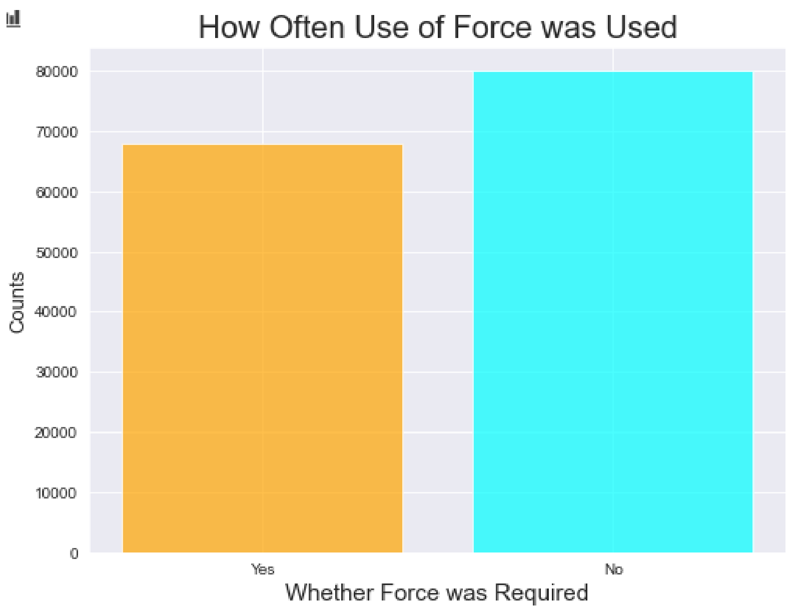
- Officer main duty
- Subject Gender

Insights

Frequency of Force Usage

For the purpose of analysis, it is important to look at the distribution of which officers were required to use force in effecting arrests.Results show that by our definition of force, 45.94% of officers were required to use force.

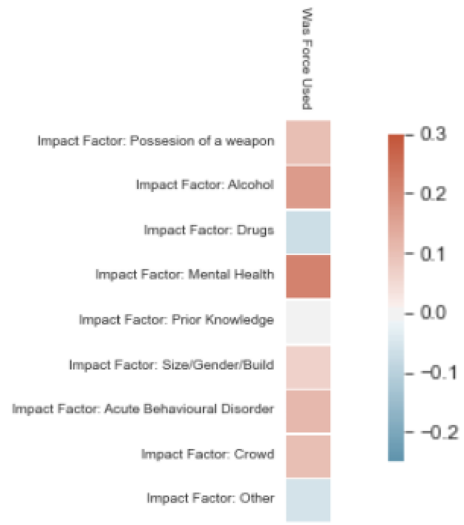
	Yes	No
Officer required force	67,950	79,945



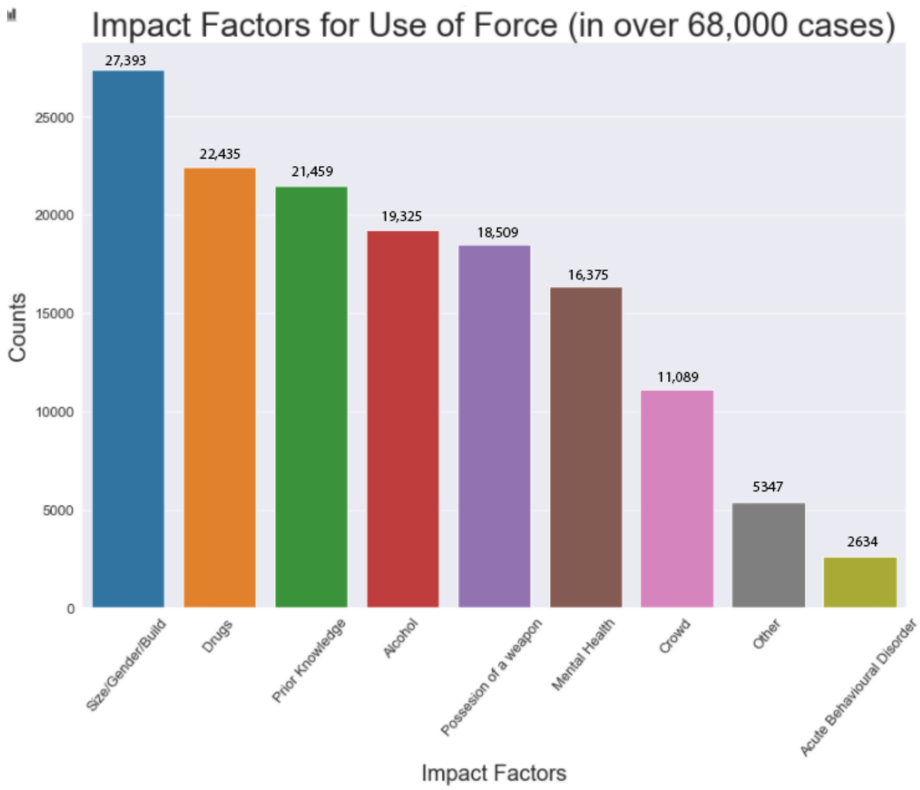
Impact Factors for Force Usage (subject-oriented)

It is important to note that multiple factors can influence an officers decision-making when opting to use force. The following is a list of all impact factors which incident reports allow an officer to specify.

Factor	Count
Possession of Weapon	18,509
Alcohol	19,235
Drugs	22,435
Mental Health	16,375
Prior Knowledge	21,459
Size/Gender/Build	27,393
Acute Behavioural Disorder	2635
Crowd	11,089
Other	5347



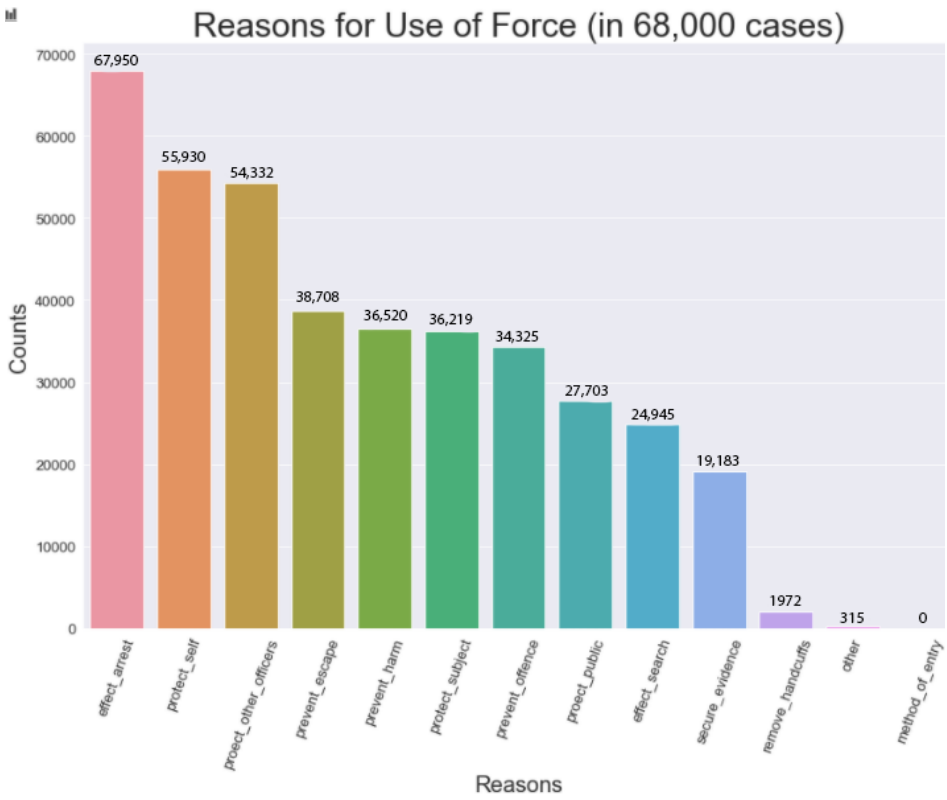
Looking at correlation, we see that the factors which most correlate to an outcome of force include ***mental health***, ***alcohol*** and ***behavioural disorders***. However, out of the 67,950 officers which required force usage, we see that size/gender/build, drugs and prior knowledge account for the most frequent factors.



Reasons to Use Force (Officers)

It is important to note that effecting an arrest accounts for all records, and that there can be multiple reasons for an officer to decide to use force.

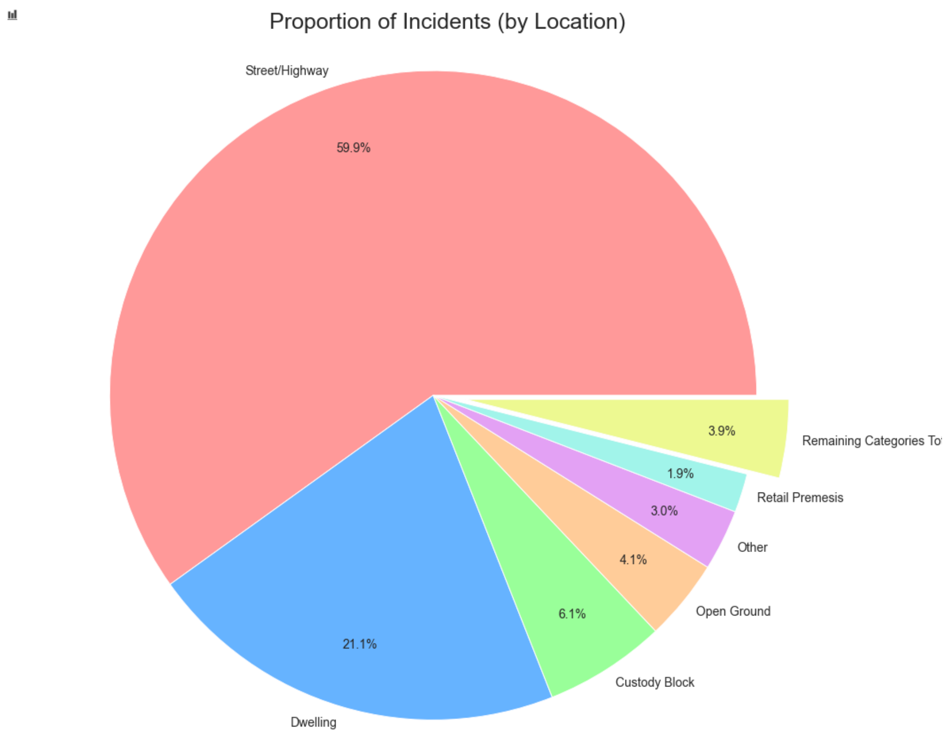
Reason	Counts
Effect arrest	67,950
Protect self	55,930
Protect other officers	54,332
Prevent escape	38,708
Prevent harm	36,520
Protect subject	36,219
Prevent offence	34,325
Protect public	27,703
Effect search	24,945
Secure evidence	19,183
Remove handcuffs	1972
Other	315
Method of entry	0

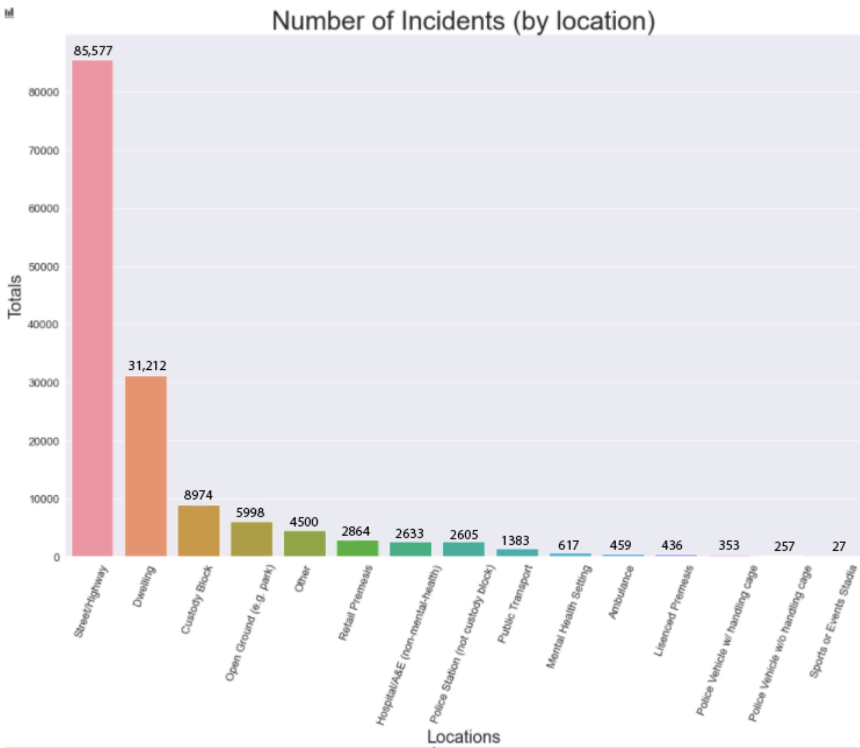


Vicinity of Arrests (total)

The following visualizations illustrate the distribution of locations in which arrests are made. The highest risk location for officers to attend to seem to be street/highway incidents, with 59.9% of incidents

Location	Counts
Street/Highway	88,577
Public Transport	1383
Retail Premises	2864
Licensed Premises	436
Sports or Event Stadia	27
Hospital (non-mental health setting)	2633
Mental Health setting	617
Police vehicle with prisoner handling cage	353
Police vehicle without prisoner handling cage	257
Dwelling	31,212
Police station	2605
Custody Block	8974
Ambulance	459
Other	4500

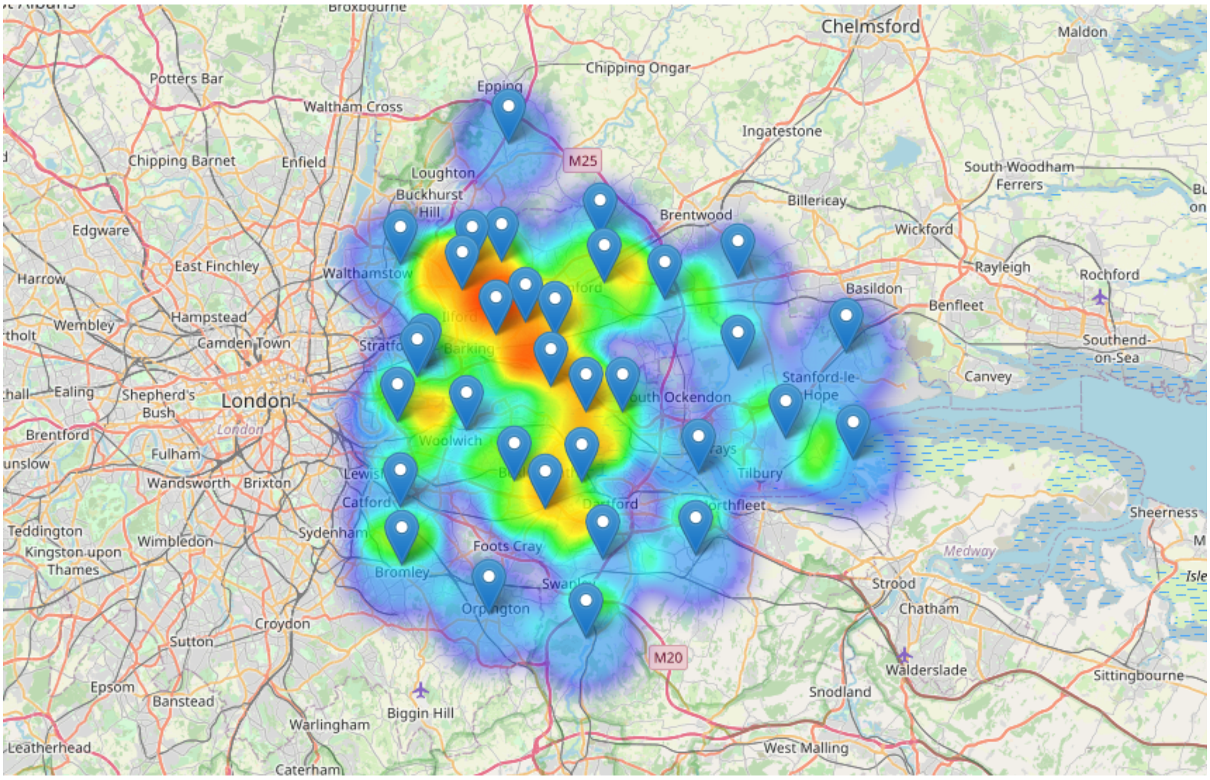


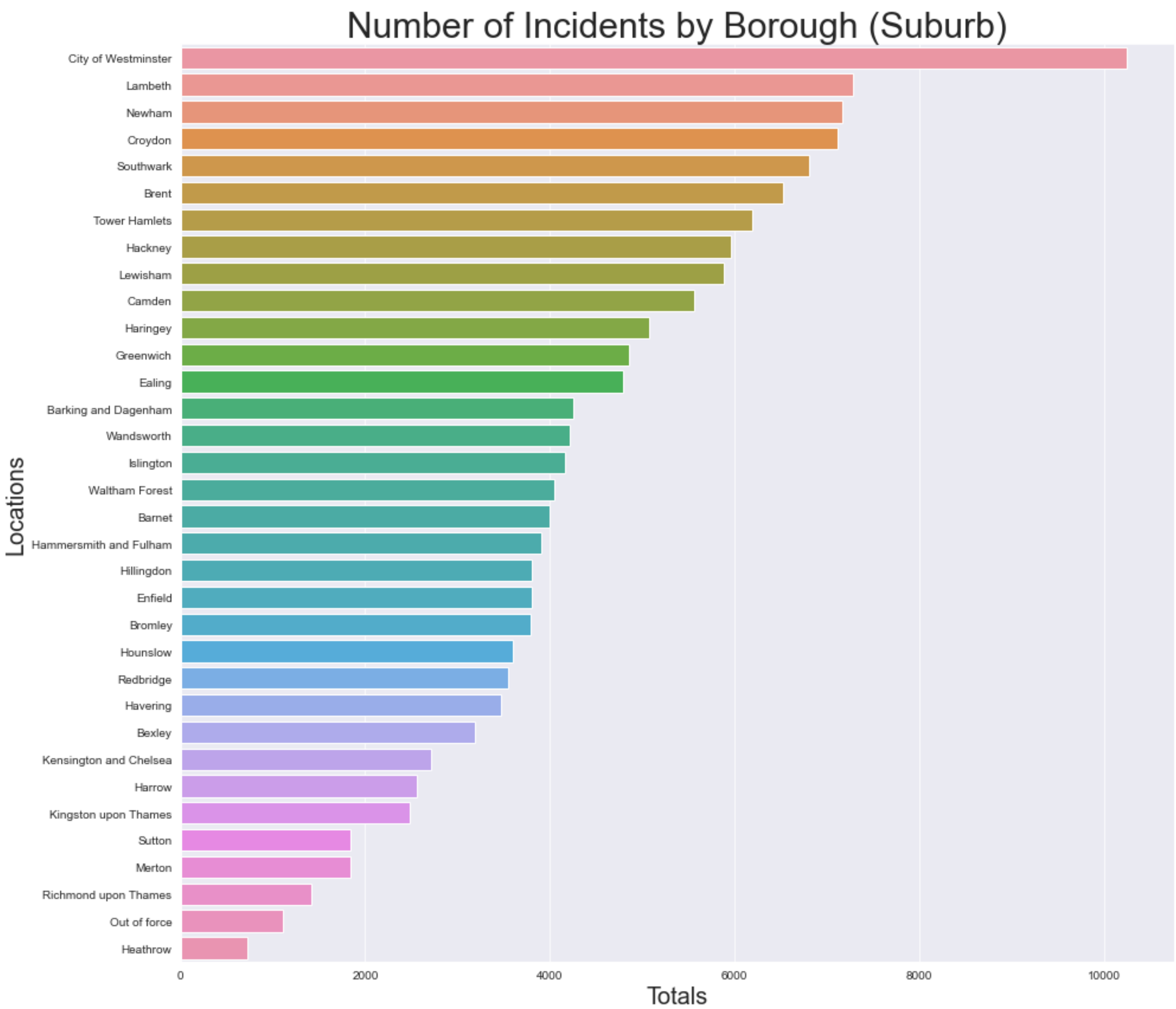


Location of arrests (Borough)

The following tables and graphs illustrate the distribution of incidents across Boroughs in East London. We see that the majority of incidents occur in the Boroughs of Westminster, Lambeth, Newham and Croydon.

	Incidents	longitude	latitude
City of Westminster	10239	51.5145	0.1595
Lambeth	7274	51.4571	0.1231
Newham	7160	51.5255	0.0352
Croydon	7119	51.3762	0.0982
Southwark	6802	51.4881	0.0763
Brent	6528	51.5673	0.2711
Tower Hamlets	6192	51.5203	0.0293
Hackney	5960	51.5734	0.0724
Lewisham	5884	51.4415	0.0117
Camden	5560	51.5455	0.1628
Haringey	5071	51.5906	0.1110
Greenwich	4855	51.4934	0.0098
Ealing	4797	51.5250	0.3414
Barking and Dagenham	4258	51.5541	0.1340
Wandsworth	4210	51.4568	0.1897
Islington	4164	51.5465	0.1058
Waltham Forest	4044	51.5886	0.0118
Barnet	3999	51.6050	0.2076
Hammersmith and Fulham	3900	51.4990	0.2291
Hillingdon	3804	51.5352	0.4481
Enfield	3802	51.6623	0.1181
Bromley	3788	51.4060	0.0132
Hounslow	3601	51.4828	0.3882
Redbridge	3541	51.5886	0.0824
Havering	3464	51.5779	0.2121
Bexley	3192	51.4399	0.1543
Kensington and Chelsea	2707	51.4991	0.1938
Harrow	2563	51.5806	0.3420
Kingston upon Thames	2487	51.4123	0.3007
Sutton	1844	51.3614	0.1940
Merton	1839	51.4098	0.2108
Richmond upon Thames	1412	51.4613	0.3037
Out of force	1107	51.7520	1.2577
Heathrow	728	51.4700	0.4543

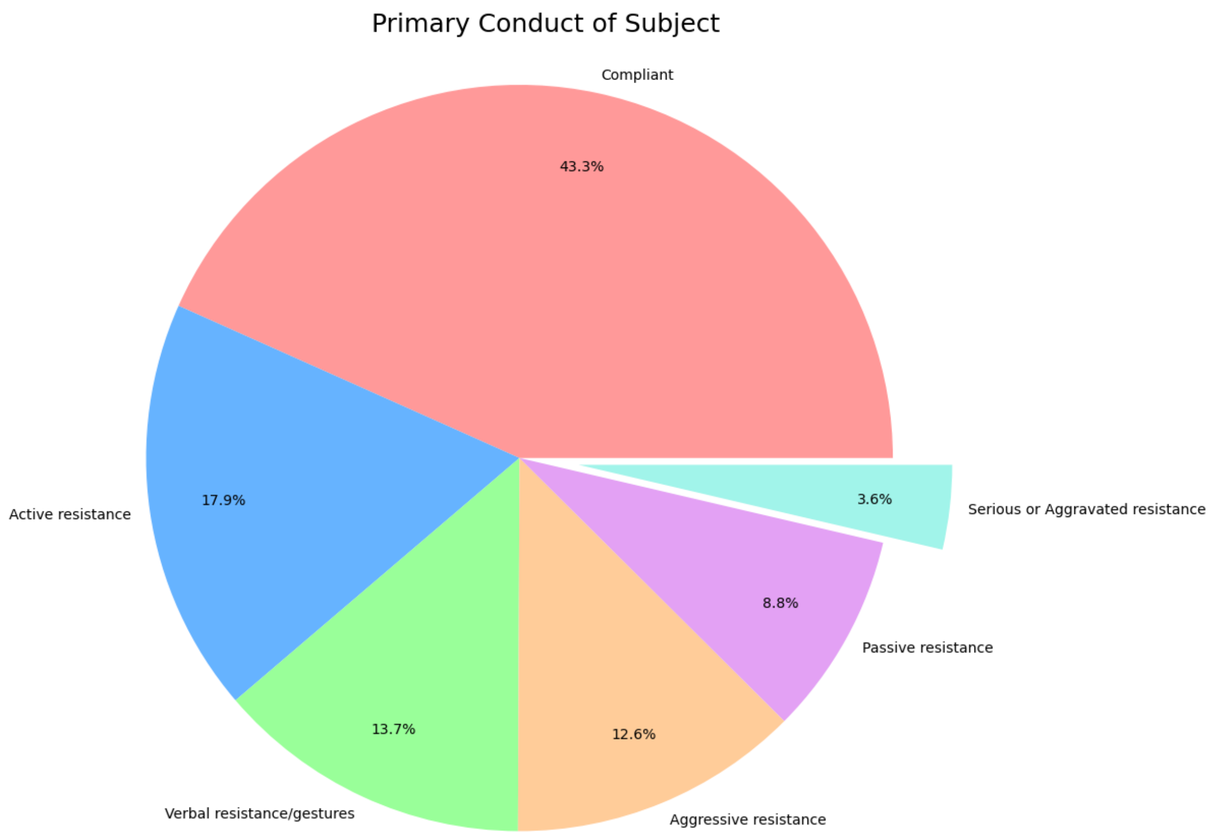



Number of Incidents by Borough (Suburb)
City of Westminster
Lambeth
Newham
Croydon
Southwark
Brent
Tower Hamlets
Hackney
Lewisham
Camden
Haringey
Greenwich
Ealing
Barking and Dagenham
Wandsworth
Islington
Waltham Forest
Barnet
Hammersmith and Fulham
Hillingdon
Enfield
Bromley
Hounslow
Redbridge
Havering
Bexley
Kensington and Chelsea
Harrow
Kingston upon Thames
Sutton
Merton
Richmond upon Thames
Out of force
Heathrow
Locations
0
2000
4000
6000
8000
10000
Totals

Primary Conduct of Subjects

The following table and chart illustrate the distribution of resistance levels across all reported incidents. It is good to note that 43.3% of officers experience initial compliance. The aim is to increase this proportion and reduce the number of aggressive and aggravated resistance.

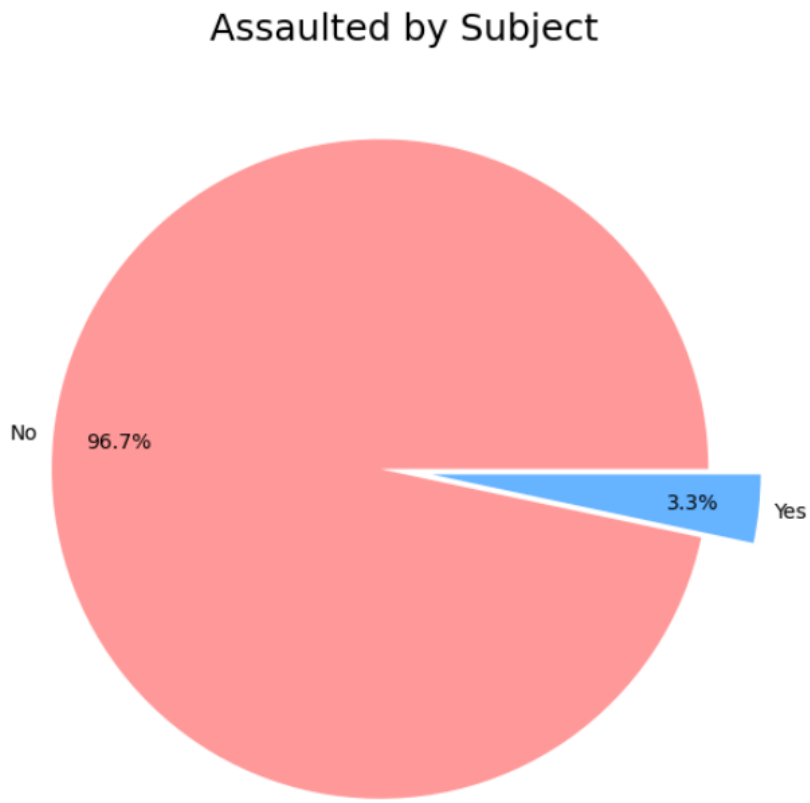
Resistance Level	Counts
Compliant	64,058
Active resistance	26,525
Verbal resistance/gestures	20,243
Aggressive resistance	18,655
Passive resistance	13,017
Serious or aggravated resistance	5397



Officers Assaulted by Subject

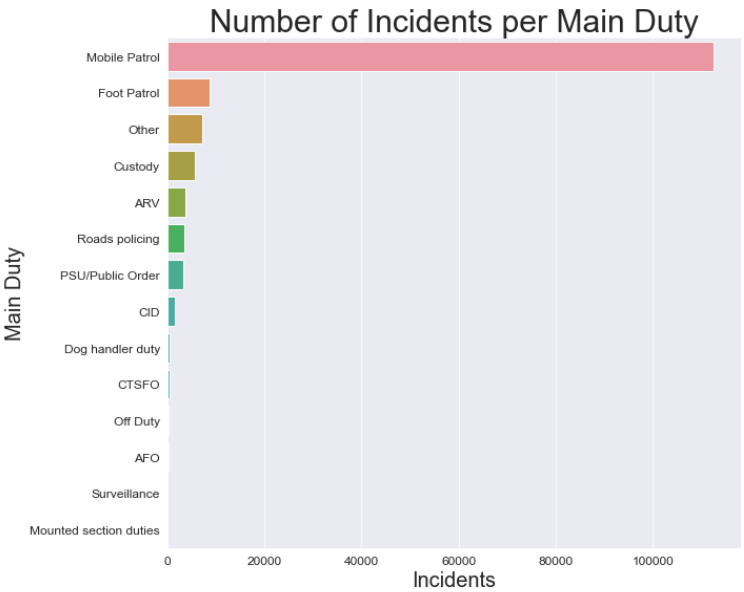
The following chart illustrates that only 3.3% of officers experience assault when responding to an incident.

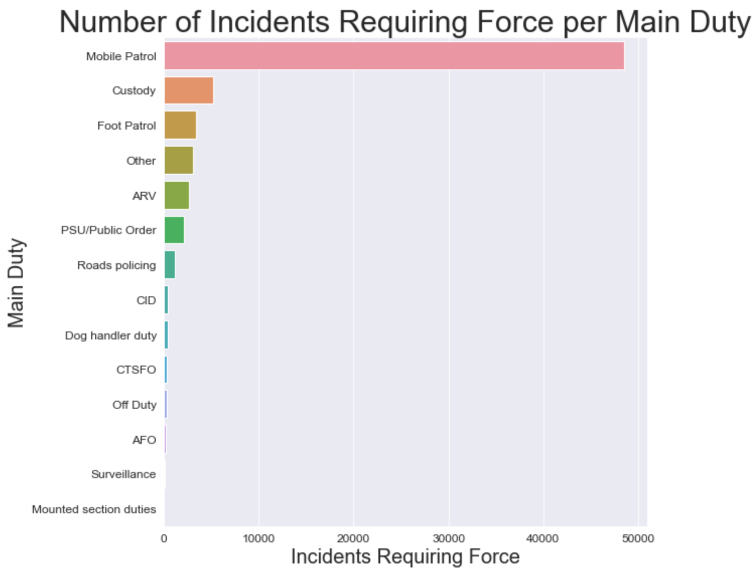
Assaulted	Counts
No	142,968
Yes	4927

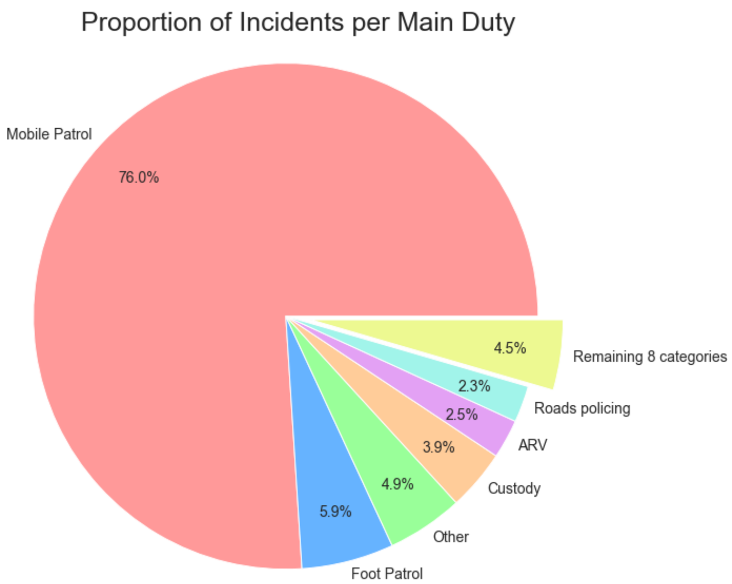


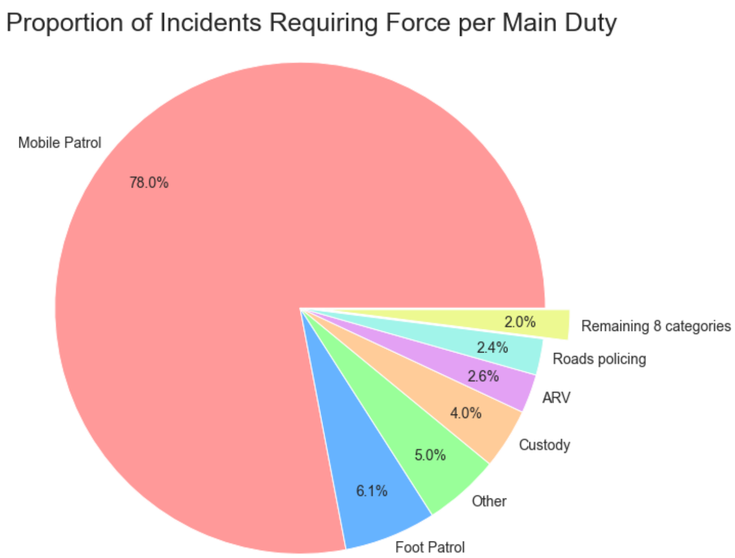
Analysis of Police Duties

If we specifically look at the distribution of all incidents per main duty of the officer, we see that the incidents are disproportionately leaning towards mobile patrol duties. Narrowing that down to only the number of incidents which require force, we see that the distribution remains the same. Despite this, mobile patrol officers seem to be of the highest risk duty for officers in London.





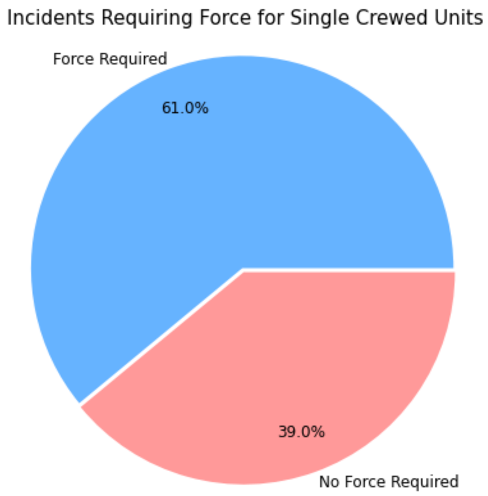


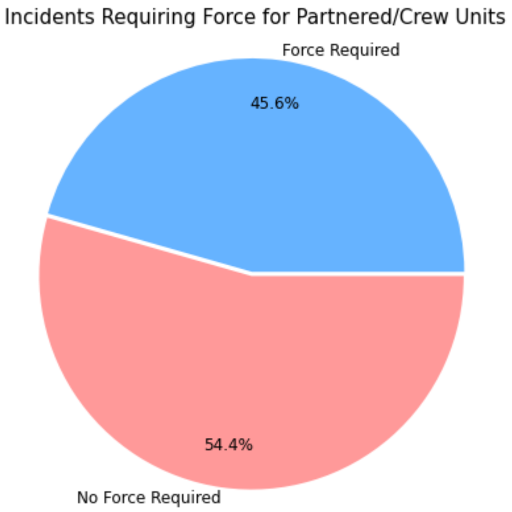


Single Crew vs Partnered/Squad Distribution

Our analysis of deployment revealed that only 2.3% of those are single-crewed officers. However, data indicates that at least 61% of those incidents require use of force. This can be attributed to the higher risk of being individually deployed. This contrasts to squad or partnered units, in which only 45% of incidents for partnered units require the use of force. This means that single units are 1.35x more likely required to use force when being deployed.

	Force required	**No force required**	**Total**
Single Crew	2065	1322	**3387**
Squad	65,885	78,623	**144,508**

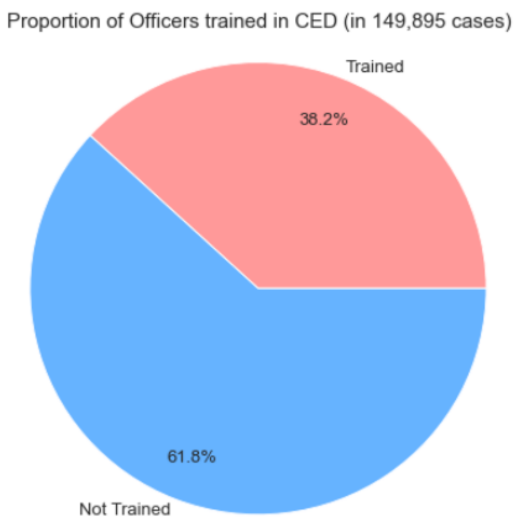




Officers Trained in CED (Taser)

The following charts illustrate the distribution of officers trained in CED usage. The data shows us that only 38.2% of officers are properly trained in taser usage.

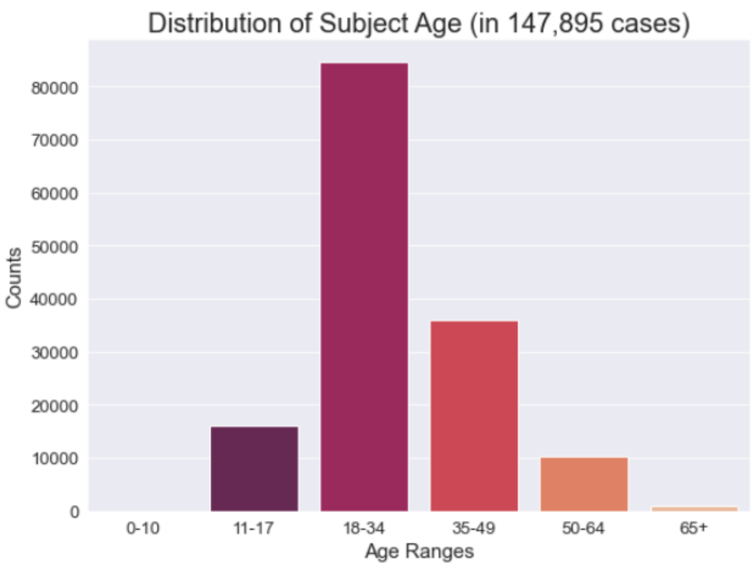
Trained in Taser	Counts
Trained	56,450
Not Trained	91,445

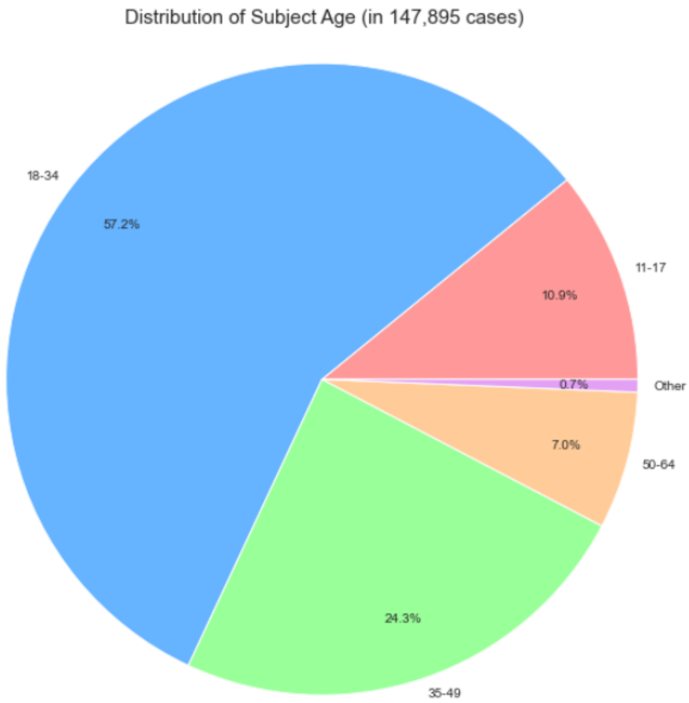


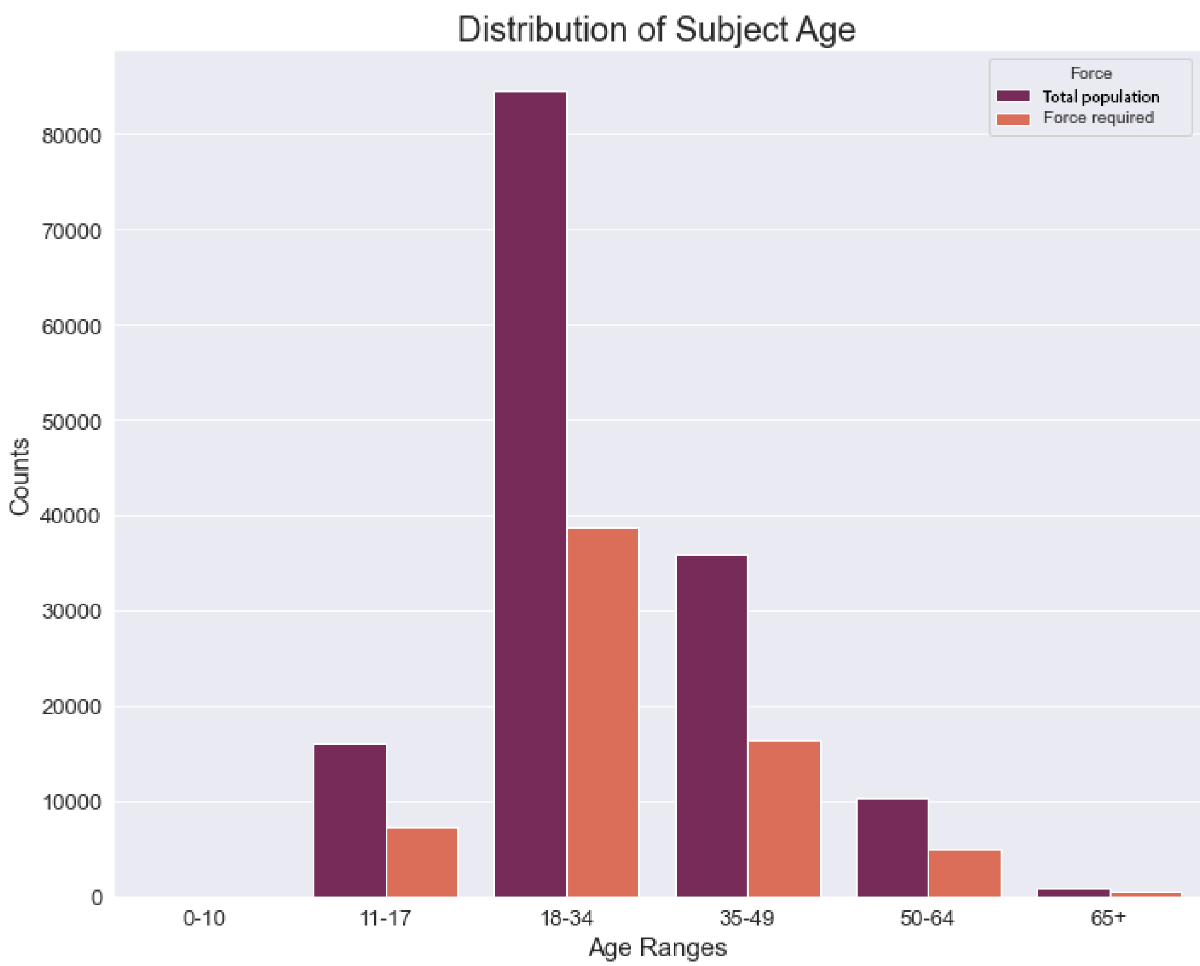
Subject Age Analysis

In all reported incidents, at least 57.2% of subjects are in the ages between 18-34, with 35-49 being the second largest group at 24.3%. When looking at distribution of age and scale it for only cases which have required force, we see that the distribution proportions remain the same.

Age Range	Distribution across population	Distribution which required force	Percentage of total population of age which require force
0-10	46	42	91%
11-17	16,094	7321	45.5%
18-34	84,545	38,767	45.9%
35-49	35,886	16,410	45.7%
50-64	10,401	4897	47%
65+	923	513	55.6%
Total	**147,895**	**67,950**	**46%**


Distribution of Subject Age (in 147,895 cases)
Counts
80000
70000
60000
50000
40000
30000
20000
10000
0
0-10
11-17
18-34
35-49
50-64
65+
Age Ranges

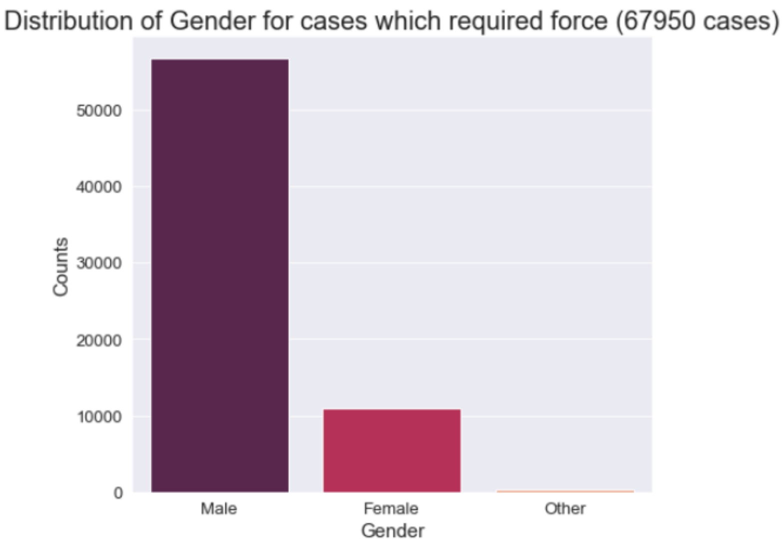

Distribution of Subject Age (in 147,895 cases)
18-34
57.2%
11-17
10.9%
Other
0.7%
7.0%
50-64
24.3%
35-49

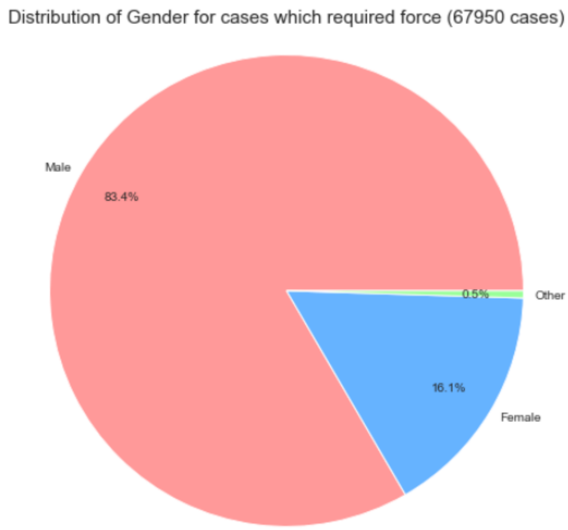

Distribution of Subject Age
Force
Total population
Force required
Counts
80000
70000
60000
50000
40000
30000
20000
10000
0
0-10
11-17
18-34
35-49
50-64
65+
Age Ranges

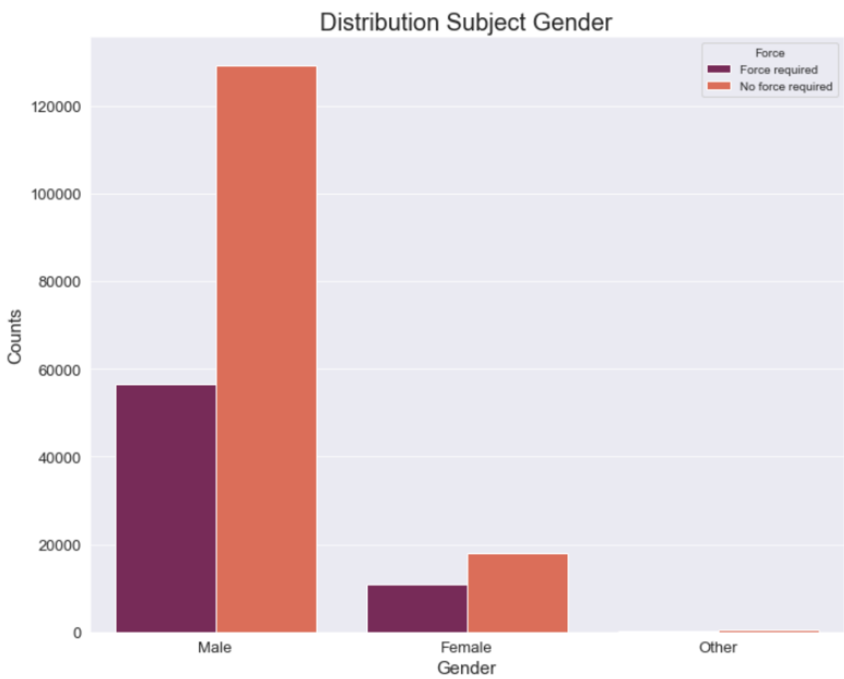
Distribution of Gender

In cases which require force (67,960), majority are males, which make up 83.4% of cases, while females only make up 16.1%, and 0.5% consist of others.

Gender	Counts	Percentage of force-required cases
Male	56,646	83.4%
Female	10,968	16.1%
Other	336	0.5%
Total	**67,950**	**100%**



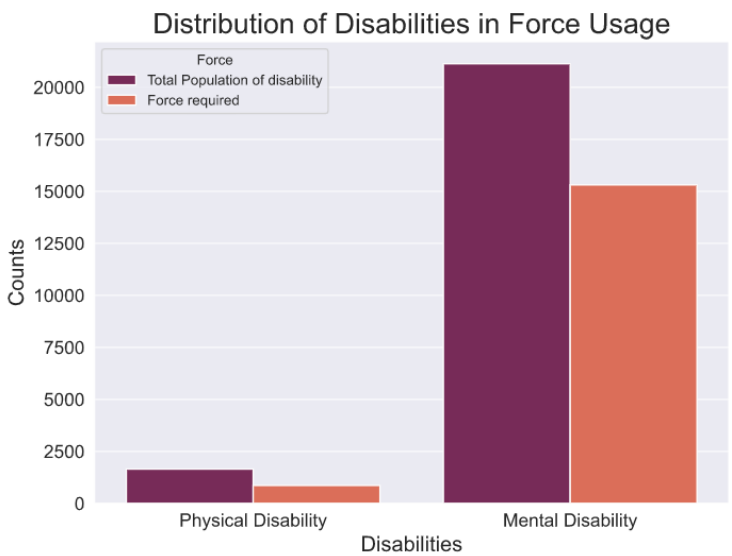


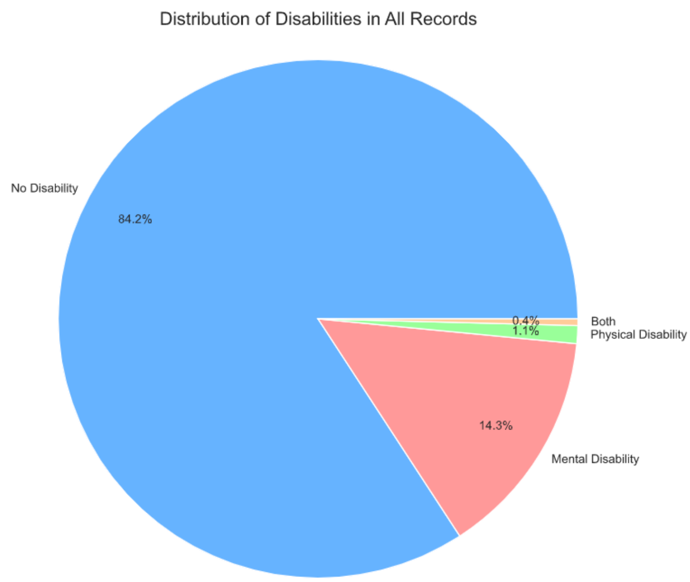


Distribution of Disabilities

Our data shows that of all cases, 84.2% have no known disability, however, 14.3% have a mental disability, while only 1.1% have a physical disability, with 0.4% having both. When we look at force usage, officer usage of force is at an alarmingly large rate of the mental disability population. This shows lack of preparation of officers dealing with individuals with mental health issues.

	Total population	**Proportion which require use of force**	**Percentage of population**
No Disability	124,501	51,775	45.6%
Physical Disability	1651	864	52.3%
Mental Disability	21,133	15,311	72.5%

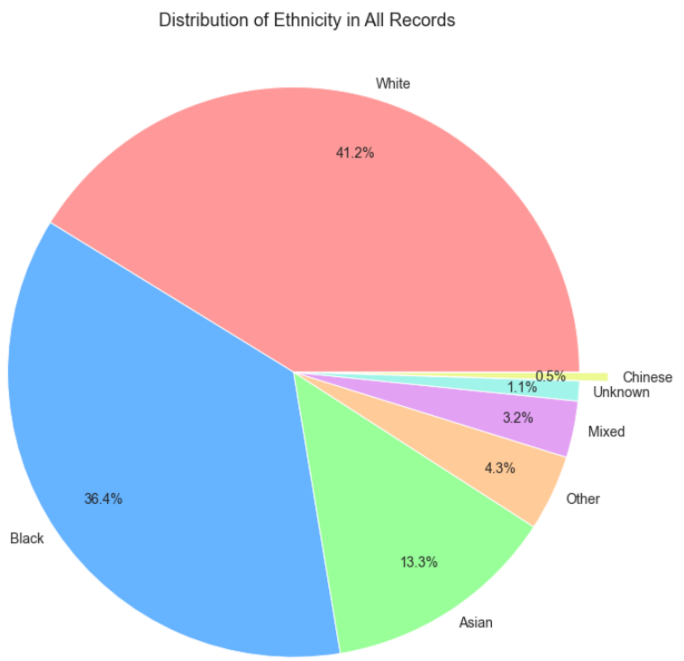


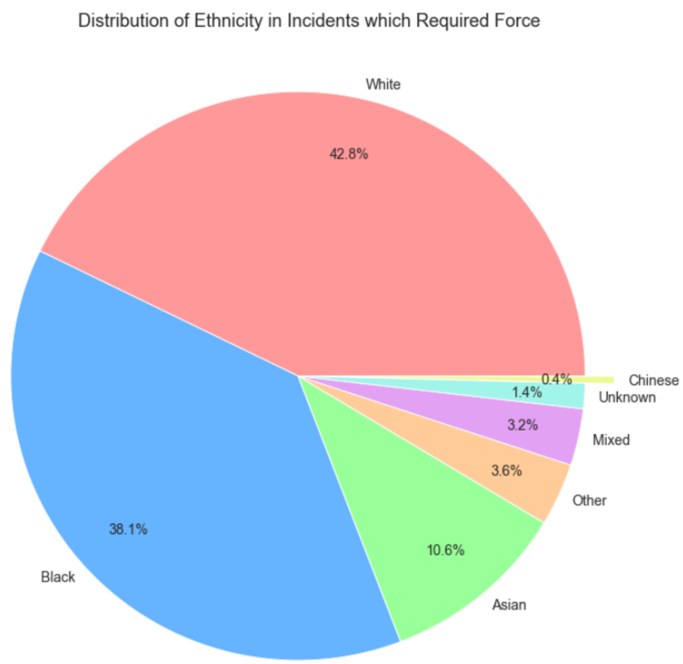


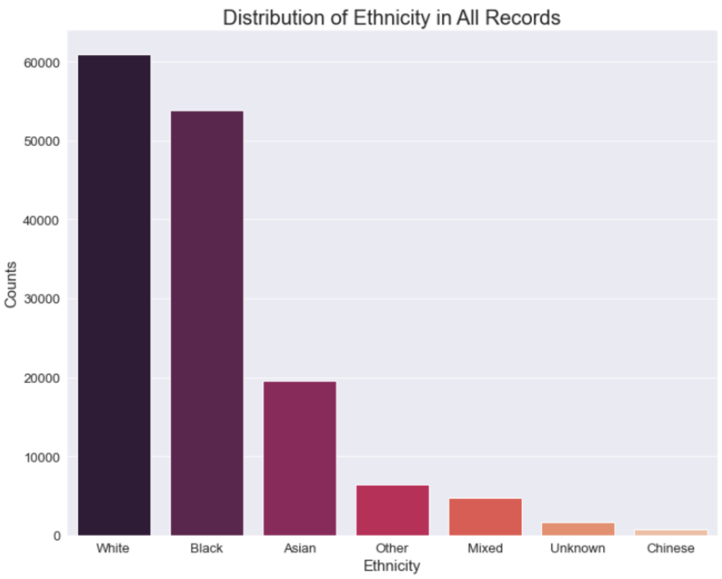
Ethnic Distribution

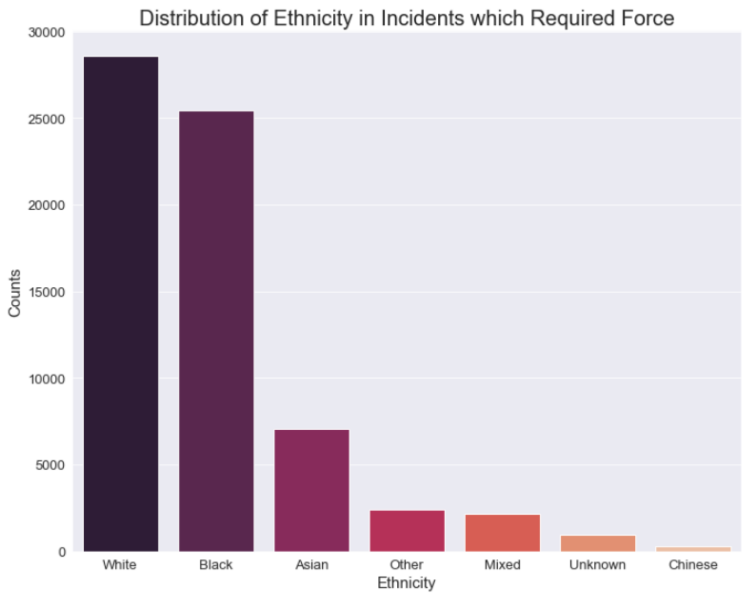
When looking at the distribution of ethnicities for total incidents, people of white and black descent seem to be the most frequent. However, this does not reflect the ethnic population of London, being a predominantly white demographic. When comparing total ethnic populations to incidents which have required force, we see that the trend follows the same distribution.

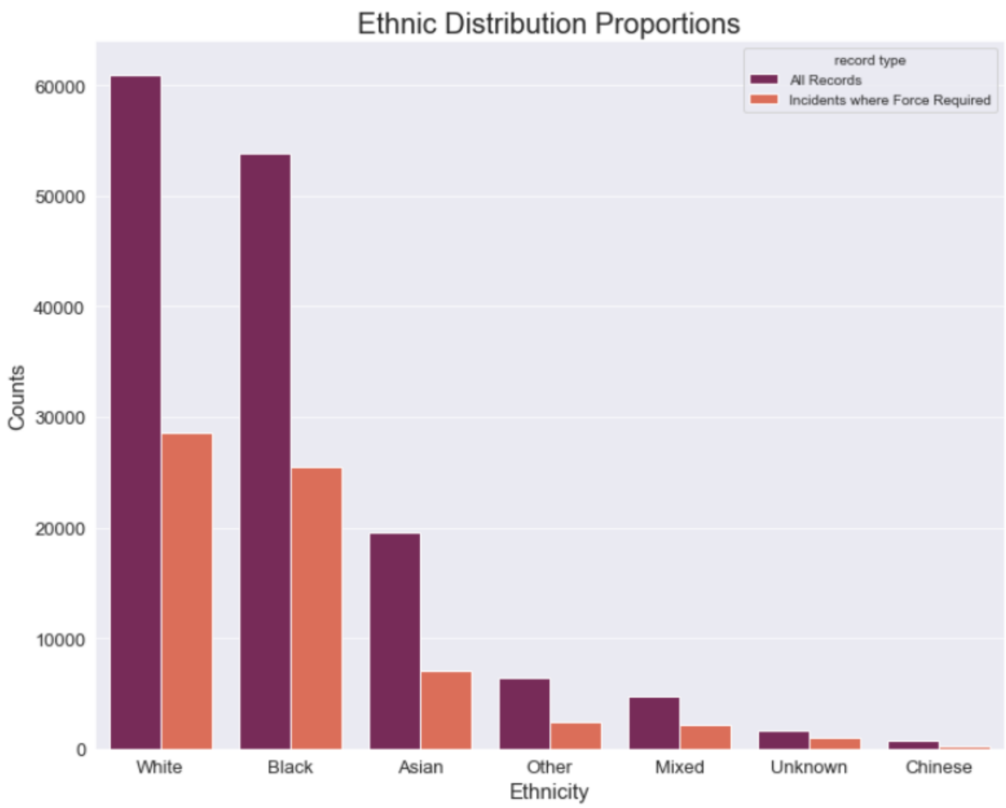
	Counts for total population	Counts which required force	Percentage of total ethnic population
White	60,937	27,578	47%
Black	53,853	25,441	47%
Asian	19,607	7052	36%
Other	6359	2378	37%
Mixed	4741	2157	45%
Unknown	1666	951	58%
Chinese	732	273	37%






Distribution of Ethnicity in All Records
Counts
Ethnicity
White
Black
Asian
Other
Mixed
Unknown
Chinese


Distribution of Ethnicity in Incidents which Required Force
Counts
Ethnicity
White
Black
Asian
Other
Mixed
Unknown
Chinese


Ethnic Distribution Proportions
record type
All Records
Incidents where Force Required
Counts
Ethnicity
White
Black
Asian
Other
Mixed
Unknown
Chinese

Modelling

Main Features and Target Variable

For our model, the aim is to predict whether a particular subject(s) will require the use of force, to give some expectation to officers who are called to the scene. For this particular model, we will require only features which the officer/police triage would be made available to via public reports. These include:

- Date
- Incident Location
- The main duty of the officer
- Number of officers sent (single-crew / partnered or crew)
- Subject age
- Physical Disabilities
- Mental Disabilities
- Subject Gender
- Subject Ethnicity
- Borough

Our target variable will be **whether use of force is expected or not.** This variable/feature was engineered by distinguishing tactics used over the course of the incident, with tactical communications and compliant handcuffing not under the umbrella of use of force. Anything above said tactics will be considered use of force.

Training Method

The training method used will be a train test split of 20% with a random state of 7 for reproducibility. Due to the large volume of data (150,000 records) and the amount of features used, the data will be randomly sampled to 20,000 records to maintain a reliable cpu workload and compute time. Furthermore, we will explore training with and without dimensionality reduction, via PCA (Principal Component Analysis).

Breakdown of Models Used

NOTE: Any defined pipelines for evaluation will be available through the jupyter notebooks.

Baseline Model Results

<table>
<tr><th>Model and Accuracy Score</th><th>Classification Report</th></tr>
<tr><td>Logistic Regression
Accuracy: 0.55</td><td><pre>Classification Report:
              precision   recall f1-score   support

           0   0.55   1.00   0.71   2199
           1   0.00   0.00   0.00   1801

    accuracy   0.55   4000
   macro avg   0.27   0.50   0.35   4000
weighted avg   0.30   0.55   0.39   4000</pre></td></tr>
<tr><td>K-Nearest Neighbors
Accuracy: 0.6

We performed GridSearchCV to find the optimal number of K. In this case, it yielded a parameter result of 2.</td><td><pre>Classification Report:
              precision   recall f1-score   support

           0   0.59   0.84   0.70   2199
           1   0.61   0.30   0.40   1801

    accuracy   0.60   4000
   macro avg   0.60   0.57   0.55   4000
weighted avg   0.60   0.60   0.56   4000</pre></td></tr>
<tr><td>Support Vector Classifier
Accuracy: 0.55</td><td><pre>Classification Report:
              precision   recall f1-score   support

           0   0.55   1.00   0.71   2199
           1   0.00   0.00   0.00   1801

    accuracy   0.55   4000
   macro avg   0.27   0.50   0.35   4000
weighted avg   0.30   0.55   0.39   4000</pre></td></tr>
<tr><td>Decision Tree Classifier
Accuracy: 0.67

We performed a GridSearch Cross Validation function to find the optimal number of max leaf nodes. In this case, results yielded a parameter of 25.</td><td><pre>Classification Report:
              precision   recall f1-score   support

           0   0.65   0.90   0.75   2199
           1   0.76   0.40   0.53   1801

    accuracy   0.67   4000
   macro avg   0.70   0.65   0.64   4000
weighted avg   0.70   0.67   0.65   4000</pre></td></tr>
</table>

Naive Bayes (Bernoulli)
Accuracy: 0.67

```
Classification Report:
              precision    recall  f1-score   support

           0       0.65      0.84      0.73      2199
           1       0.70      0.46      0.55      1801

    accuracy                           0.67      4000
   macro avg       0.67      0.65      0.64      4000
weighted avg       0.67      0.67      0.65      4000
```

Random Forest Classifier
Accuracy: 0.63

```
Classification Report:
              precision    recall  f1-score   support

           0       0.66      0.67      0.67      2199
           1       0.59      0.59      0.59      1801

    accuracy                           0.63      4000
   macro avg       0.63      0.63      0.63      4000
weighted avg       0.63      0.63      0.63      4000
```

Bagging Classifier Results

Bagging Classifier	Classification Report

Logistic Regression Bagging

Accuracy: 0.55

```
Classification Report:
              precision    recall  f1-score   support

           0       0.55      1.00      0.71      2199
           1       0.00      0.00      0.00      1801

    accuracy                           0.55      4000
   macro avg       0.27      0.50      0.35      4000
weighted avg       0.30      0.55      0.39      4000
```

K-Nearest Neighbors Bagging (K = 2)

Accuracy: 0.59

```
Classification Report:
              precision    recall  f1-score   support

           0       0.61      0.69      0.65      2199
           1       0.55      0.47      0.51      1801

    accuracy                           0.59      4000
   macro avg       0.58      0.58      0.58      4000
weighted avg       0.58      0.59      0.58      4000
```

Support Vector Classifier Bagging

Accuracy: 0.55

Classification Report:

	precision	recall	f1-score	support
0	0.61	0.69	0.65	2199
1	0.55	0.47	0.51	1801
accuracy			0.59	4000
macro avg	0.58	0.58	0.58	4000
weighted avg	0.58	0.59	0.58	4000

Decision Tree Classifier Bagging

Accuracy: 0.67

Classification Report:

	precision	recall	f1-score	support
0	0.65	0.89	0.75	2199
1	0.75	0.41	0.53	1801
accuracy			0.67	4000
macro avg	0.70	0.65	0.64	4000
weighted avg	0.69	0.67	0.65	4000

Naive Bayes (Bernoulli) Bagging

Accuracy: 0.67

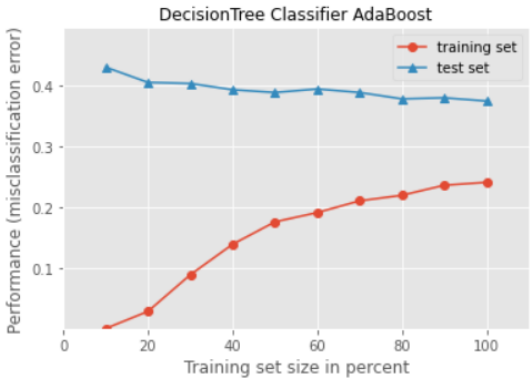
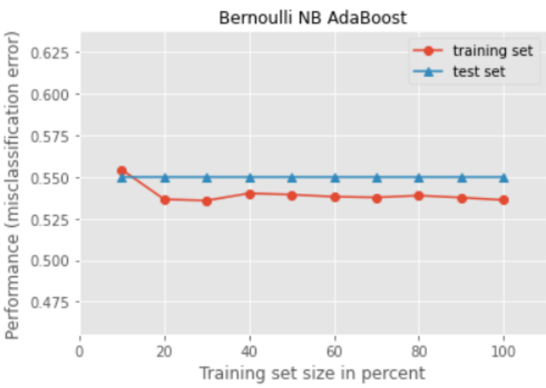
Classification Report:

	precision	recall	f1-score	support
0	0.65	0.89	0.75	2199
1	0.75	0.41	0.53	1801
accuracy			0.67	4000
macro avg	0.70	0.65	0.64	4000
weighted avg	0.69	0.67	0.65	4000

<u>Boosting Techniques</u>

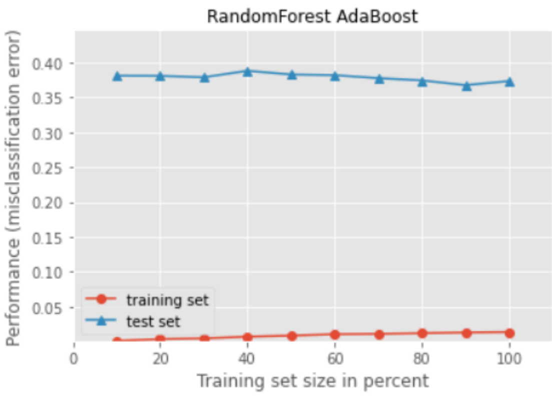
Techniques used:

- AdaBoost Classifier
- GradientBoost Classifier
- XGBoost Classifier

Model and Accuracy	**Learning Curve**	**Classification Report**
Logistic Regression AdaBoost **Accuracy: 0.55**	 	Classification Report: precision recall f1-score support 0 0.55 1.00 0.71 2199 1 0.00 0.00 0.00 1801 accuracy 0.55 4000 macro avg 0.27 0.50 0.35 4000 weighted avg 0.30 0.55 0.39 4000
Decision Tree AdaBoost **Accuracy: 0.55**	 	Classification Report: precision recall f1-score support 0 0.64 0.72 0.68 2199 1 0.60 0.51 0.55 1801 accuracy 0.63 4000 macro avg 0.62 0.62 0.62 4000 weighted avg 0.62 0.63 0.62 4000
Naive Bayes (Bernoulli) AdaBoost **Accuracy: 0.45**	 	Classification Report: precision recall f1-score support 0 0.00 0.00 0.00 2199 1 0.45 1.00 0.62 1801 accuracy 0.45 4000 macro avg 0.23 0.50 0.31 4000 weighted avg 0.20 0.45 0.28 4000

Random Forest AdaBoost

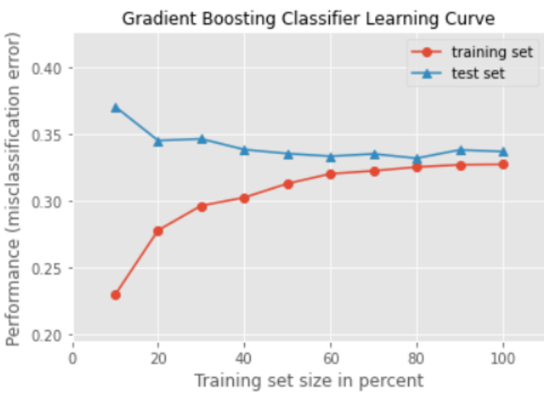
Accuracy: 0.63




```
Classification Report:

              precision    recall  f1-score   support

           0       0.66      0.67      0.67      2199
           1       0.59      0.57      0.58      1801

    accuracy                           0.63      4000
   macro avg       0.62      0.62      0.62      4000
weighted avg       0.63      0.63      0.63      4000
```

Gradient Boost Classifier

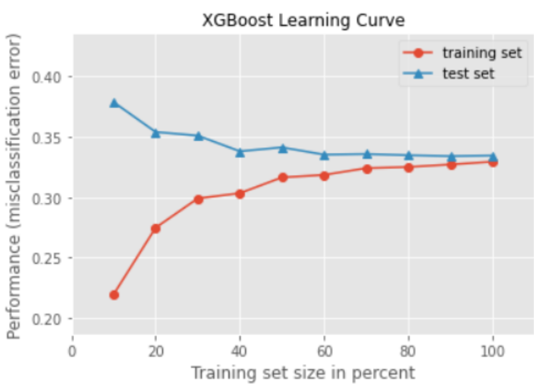
Accuracy: 0.68




```
Classification Report:

              precision    recall  f1-score   support

           0       0.65      0.89      0.75      2199
           1       0.75      0.41      0.54      1801

    accuracy                           0.68      4000
   macro avg       0.70      0.65      0.64      4000
weighted avg       0.70      0.68      0.65      4000
```

XGBoost Classifier (max depth = 1)

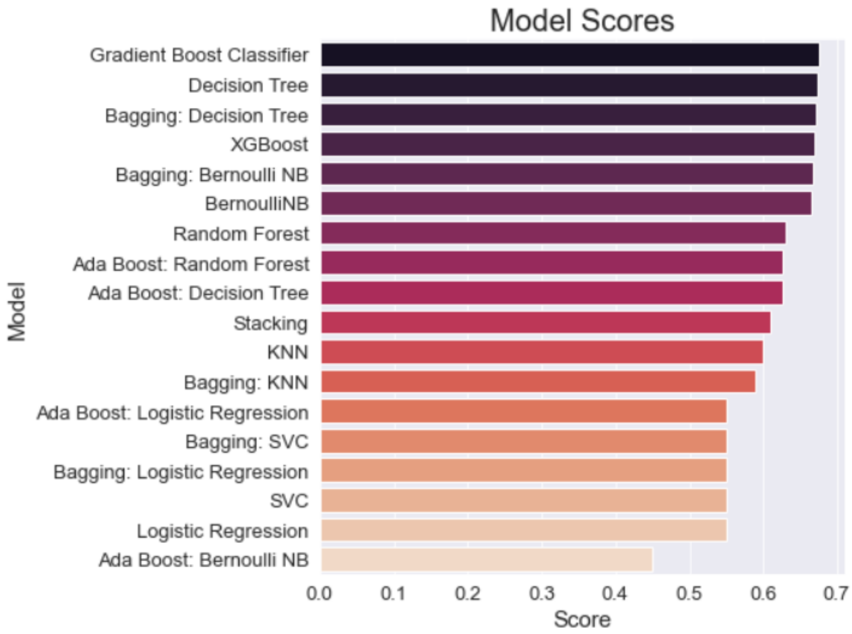
Accuracy: 0.67




```
Classification Report:

              precision    recall  f1-score   support

           0       0.65      0.88      0.75      2199
           1       0.74      0.41      0.53      1801

    accuracy                           0.67      4000
   macro avg       0.70      0.65      0.64      4000
weighted avg       0.69      0.67      0.65      4000
```

Summary of Baseline Results: Without Dimensionality Reduction

	Scores
Gradient Boost Classifier	0.676
Decision Tree	0.674
Bagging: Decision Tree	0.672
XGBoost	0.670
Bagging: Bernoulli NB	0.668
BernoulliNB	0.666
Random Forest	0.630
Ada Boost: Random Forest	0.627
Ada Boost: Decision Tree	0.626
Stacking	0.609
KNN	0.600
Bagging: KNN	0.590
Ada Boost: Logistic Regression	0.550
Bagging: SVC	0.550
Bagging: Logistic Regression	0.550
SVC	0.550
Logistic Regression	0.550
Ada Boost: Bernoulli NB	0.450



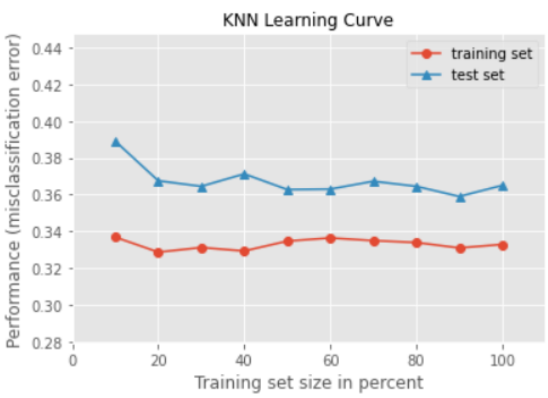
Outcomes:

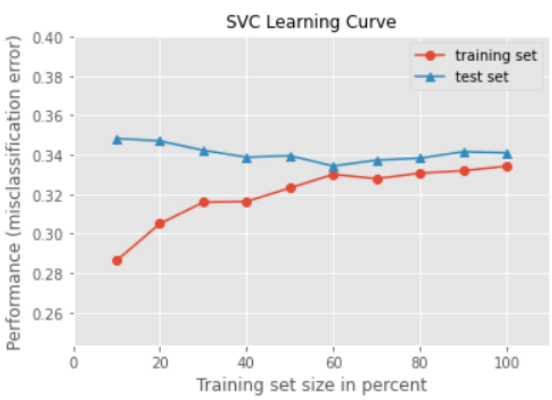
- Overall, the Decision Tree Classifier had the best baseline model improvement on our target accuracy of 45%.
- Ensemble techniques such as random forest, bagging and boosting techniques did little to improve on them.
- Models which involved SVC and Random Forest took approximately 30 minutes to train the model and predict.
- Using GridSearchCV on certain model parameters took 10 minutes to compute.
- Using Gradient Boost provided us with our best ensemble and overall method.
- We can potentially increase the accuracy of our model by experimenting via Principal Component dimensionality reduction.
- However, aiming for a target accuracy of over 75% is very optimistic, due to the nature of our data and the problem we are trying to solve. Predicting human behaviour and risk of certain actions is highly subjective and circumstantial. As a result, any significant increase in our models ability to predict, above the capability of target result of 45% can be considered a success.

Results with Principal Components = 20

<table>
<tr><th>Model and Accuracy</th><th>Learning Curve</th><th>Classification Report</th></tr>
<tr>
<td>Logistic Regression Classifier

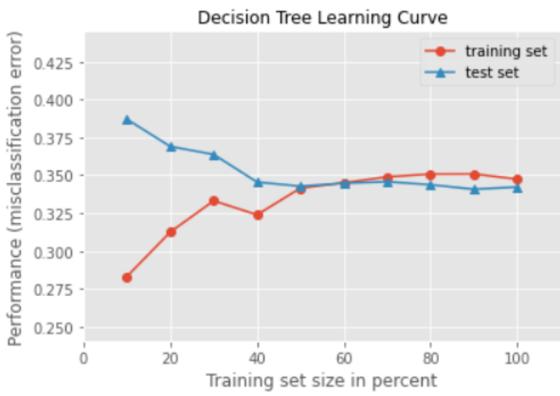
Accuracy: 0.67</td>
<td>
</td>
<td>Classification Report:
<table>
<tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr>
<tr><td>0</td><td>0.64</td><td>0.89</td><td>0.74</td><td>2199</td></tr>
<tr><td>1</td><td>0.74</td><td>0.40</td><td>0.52</td><td>1801</td></tr>
<tr><td>accuracy</td><td></td><td></td><td>0.67</td><td>4000</td></tr>
<tr><td>macro avg</td><td>0.69</td><td>0.64</td><td>0.63</td><td>4000</td></tr>
<tr><td>weighted avg</td><td>0.69</td><td>0.67</td><td>0.64</td><td>4000</td></tr>
</table></td>
</tr>
<tr>
<td>K-Neighbors Classifier (GSCV K = 17)

Accuracy: 0.64</td>
<td>
</td>
<td>Classification Report:
<table>
<tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr>
<tr><td>0</td><td>0.64</td><td>0.79</td><td>0.70</td><td>2199</td></tr>
<tr><td>1</td><td>0.63</td><td>0.45</td><td>0.53</td><td>1801</td></tr>
<tr><td>accuracy</td><td></td><td></td><td>0.64</td><td>4000</td></tr>
<tr><td>macro avg</td><td>0.63</td><td>0.62</td><td>0.61</td><td>4000</td></tr>
<tr><td>weighted avg</td><td>0.63</td><td>0.64</td><td>0.62</td><td>4000</td></tr>
</table></td>
</tr>
<tr>
<td>Support Vector Classifier

Accuracy: 0.66</td>
<td>
</td>
<td>Classification Report:
<table>
<tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr>
<tr><td>0</td><td>0.64</td><td>0.85</td><td>0.73</td><td>2199</td></tr>
<tr><td>1</td><td>0.70</td><td>0.43</td><td>0.53</td><td>1801</td></tr>
<tr><td>accuracy</td><td></td><td></td><td>0.66</td><td>4000</td></tr>
<tr><td>macro avg</td><td>0.67</td><td>0.64</td><td>0.63</td><td>4000</td></tr>
<tr><td>weighted avg</td><td>0.67</td><td>0.66</td><td>0.64</td><td>4000</td></tr>
</table></td>
</tr>
</table>

Decision Tree Classifier (GSCV max leaf nodes = 21)

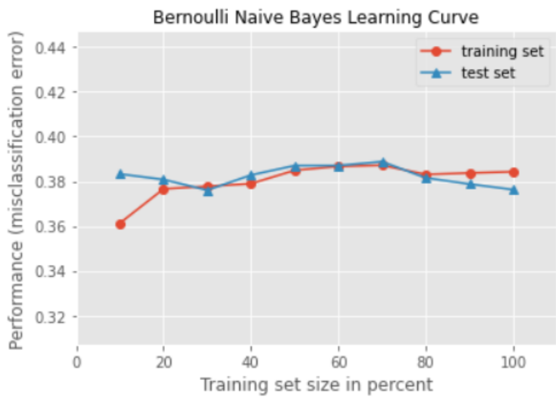
Accuracy: 0.66




```
Classification Report:

              precision    recall  f1-score   support

           0       0.64      0.85      0.73      2199
           1       0.69      0.43      0.53      1801

    accuracy                           0.66      4000
   macro avg       0.67      0.64      0.63      4000
weighted avg       0.67      0.66      0.64      4000
```

Naive Bayes (Bernoulli) Classifier

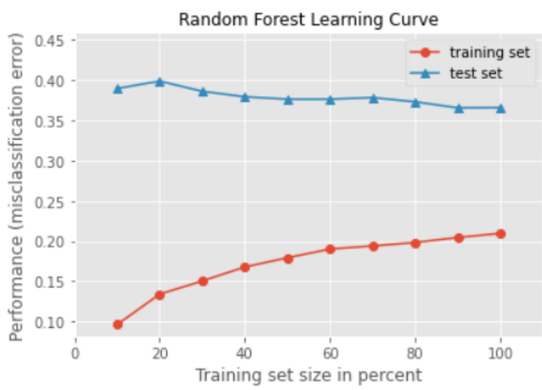
Accuracy: 0.62




```
Classification Report:

              precision    recall  f1-score   support

           0       0.63      0.76      0.69      2199
           1       0.61      0.46      0.52      1801

    accuracy                           0.62      4000
   macro avg       0.62      0.61      0.61      4000
weighted avg       0.62      0.62      0.61      4000
```

Random Forest Classifier

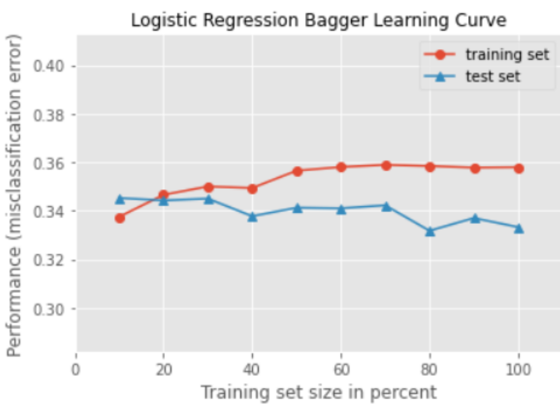
Accuracy: 0.63




```
Classification Report:

              precision    recall  f1-score   support

           0       0.65      0.72      0.69      2199
           1       0.61      0.52      0.56      1801

    accuracy                           0.63      4000
   macro avg       0.63      0.62      0.62      4000
weighted avg       0.63      0.63      0.63      4000
```

Logistic Regression Bagging

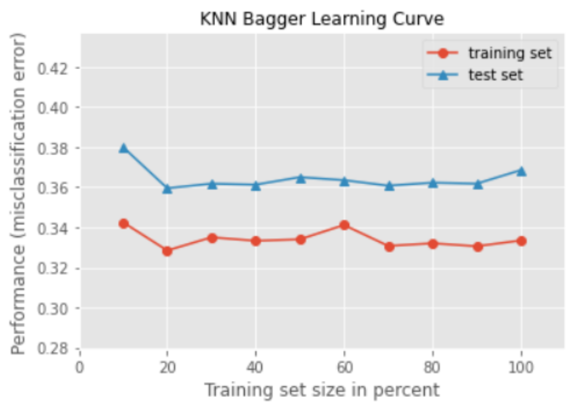
Accuracy: 0.67




```
Classification Report:

              precision    recall  f1-score   support

           0       0.64      0.88      0.74      2199
           1       0.73      0.40      0.52      1801

    accuracy                           0.66      4000
   macro avg       0.69      0.64      0.63      4000
weighted avg       0.68      0.66      0.64      4000
```

K-Nearest Neighbors Bagging

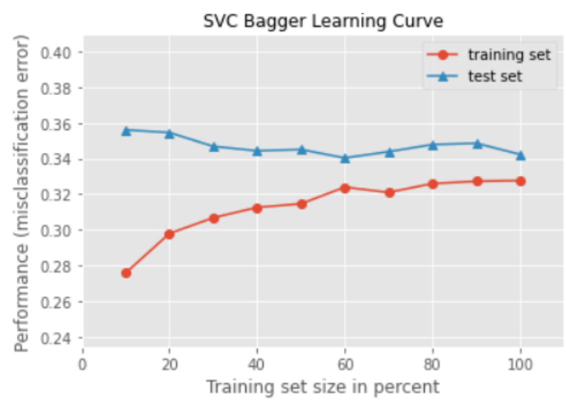
Accuracy: 0.63




```
Classification Report:

              precision    recall  f1-score   support

           0       0.65      0.76      0.70      2199
           1       0.63      0.49      0.55      1801

    accuracy                           0.64      4000
   macro avg       0.64      0.63      0.62      4000
weighted avg       0.64      0.64      0.63      4000
```

Support Vector Classifier Bagging

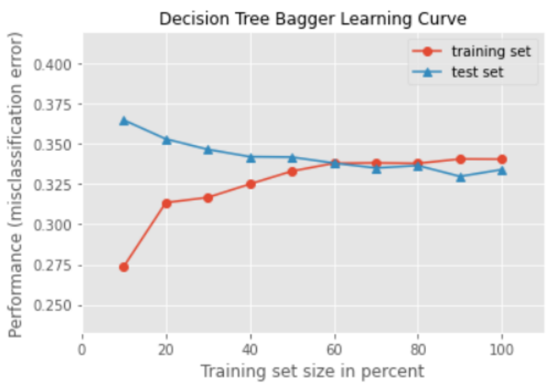
Accuracy: 0.66




```
Classification Report:

              precision    recall  f1-score   support

           0       0.65      0.81      0.72      2199
           1       0.68      0.47      0.55      1801

    accuracy                           0.66      4000
   macro avg       0.66      0.64      0.64      4000
weighted avg       0.66      0.66      0.65      4000
```

Decision Tree Bagging

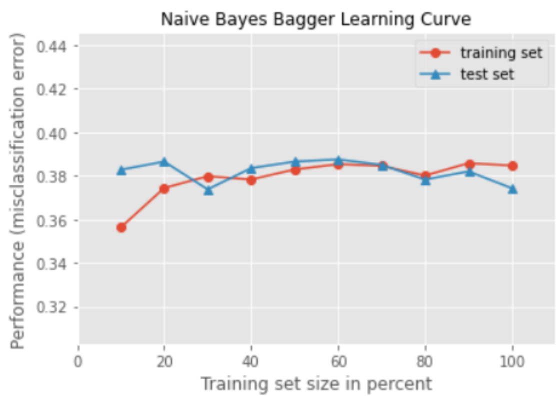
Accuracy: 0.67




```
Classification Report:

              precision    recall  f1-score   support

           0       0.65      0.87      0.74      2199
           1       0.72      0.42      0.53      1801

    accuracy                           0.67      4000
   macro avg       0.68      0.64      0.64      4000
weighted avg       0.68      0.67      0.65      4000
```

Naive Bayes (Bernoulli) Bagging

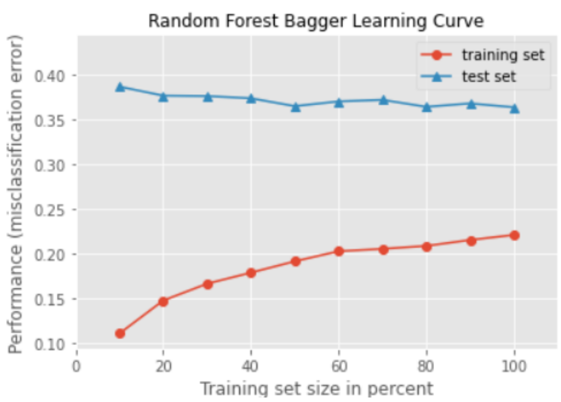
Accuracy: 0.63




```
Classification Report:

              precision    recall  f1-score   support

           0       0.63      0.77      0.69      2199
           1       0.61      0.45      0.52      1801

    accuracy                           0.62      4000
   macro avg       0.62      0.61      0.60      4000
weighted avg       0.62      0.62      0.61      4000
```

Random Forest Bagging (GSCV n_estimators = 85)

Accuracy: 0.64




```
Classification Report:

              precision    recall  f1-score   support

           0       0.65      0.74      0.69      2199
           1       0.62      0.52      0.56      1801

    accuracy                           0.64      4000
   macro avg       0.63      0.63      0.63      4000
weighted avg       0.64      0.64      0.63      4000
```

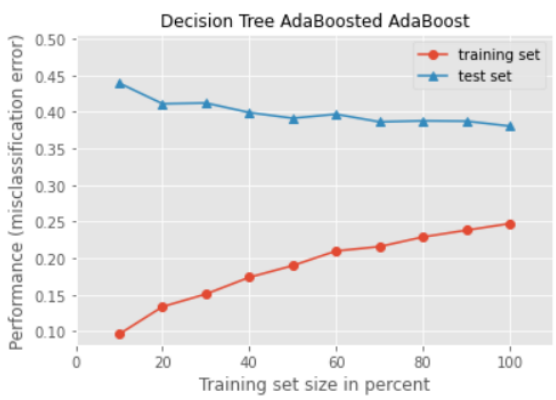
Logistic Regression AdaBoost

Accuracy: 0.66


```
Classification Report:

              precision    recall  f1-score   support

           0       0.64      0.87      0.74      2199
           1       0.72      0.41      0.52      1801

    accuracy                           0.66      4000
   macro avg       0.68      0.64      0.63      4000
weighted avg       0.68      0.66      0.64      4000
```

Decision Tree AdaBoost

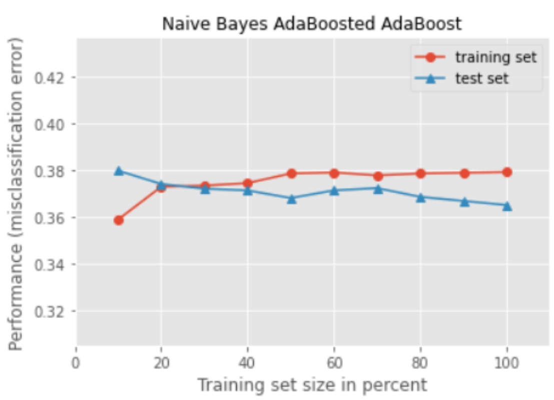
Accuracy: 0.66




```
Classification Report:

              precision    recall  f1-score   support

           0       0.63      0.73      0.68      2199
           1       0.60      0.48      0.53      1801

    accuracy                           0.62      4000
   macro avg       0.61      0.61      0.61      4000
weighted avg       0.62      0.62      0.61      4000
```

Naive Bayes (Bernoulli) AdaBoost

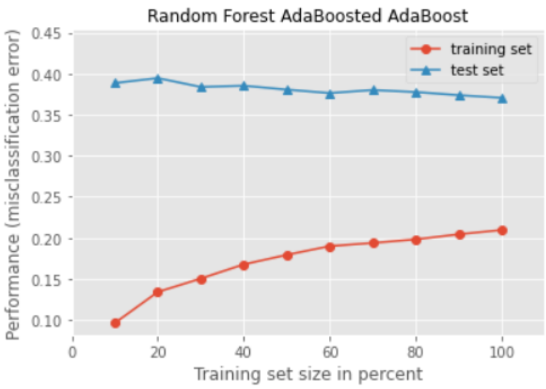
Accuracy: 0.64




```
Classification Report:

              precision    recall  f1-score   support

           0       0.63      0.80      0.71      2199
           1       0.64      0.43      0.51      1801

    accuracy                           0.64      4000
   macro avg       0.64      0.62      0.61      4000
weighted avg       0.64      0.64      0.62      4000
```

Random Forest AdaBoost

Accuracy: 0.63

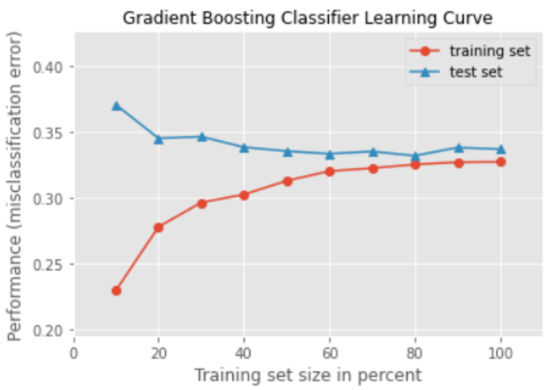


Classification Report:

	precision	recall	f1-score	support
0	0.65	0.71	0.68	2199
1	0.60	0.53	0.56	1801
accuracy			0.63	4000
macro avg	0.62	0.62	0.62	4000
weighted avg	0.63	0.63	0.63	4000

Gradient Boosting

Accuracy: 0.66



Classification Report:

	precision	recall	f1-score	support
0	0.65	0.86	0.74	2199
1	0.71	0.42	0.53	1801
accuracy			0.66	4000
macro avg	0.68	0.64	0.63	4000
weighted avg	0.68	0.66	0.64	4000

XGBoost (max_depth = 2, learning rate = 0.3)

Accuracy: 0.67

Classification Report:

	precision	recall	f1-score	support
0	0.65	0.85	0.74	2199
1	0.70	0.45	0.55	1801
accuracy			0.67	4000
macro avg	0.68	0.65	0.64	4000
weighted avg	0.67	0.67	0.65	4000

Stacking Classifier (estimators = Logistic Regression, Decision Tree & Bernoulli NB)

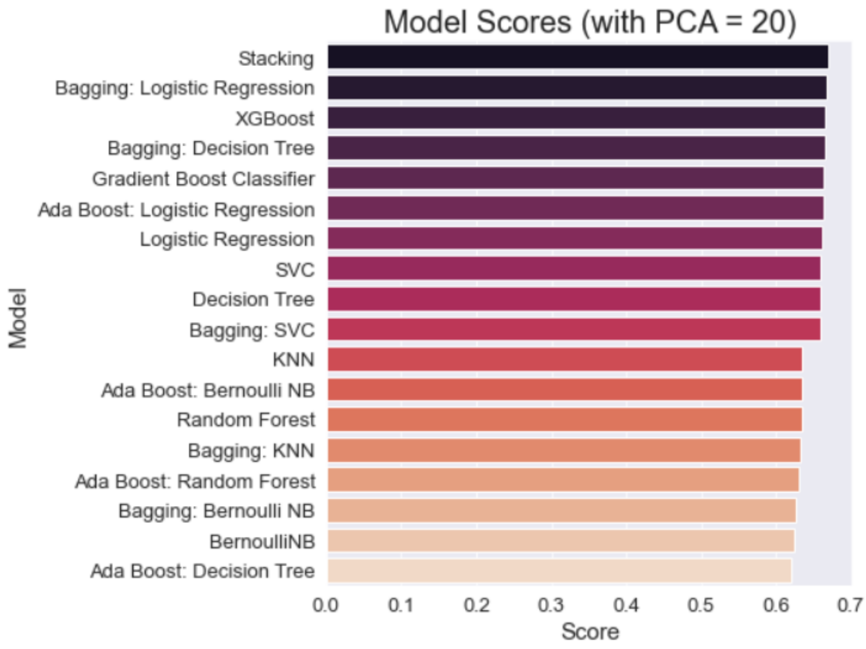
Accuracy: 0.67

Classification Report:

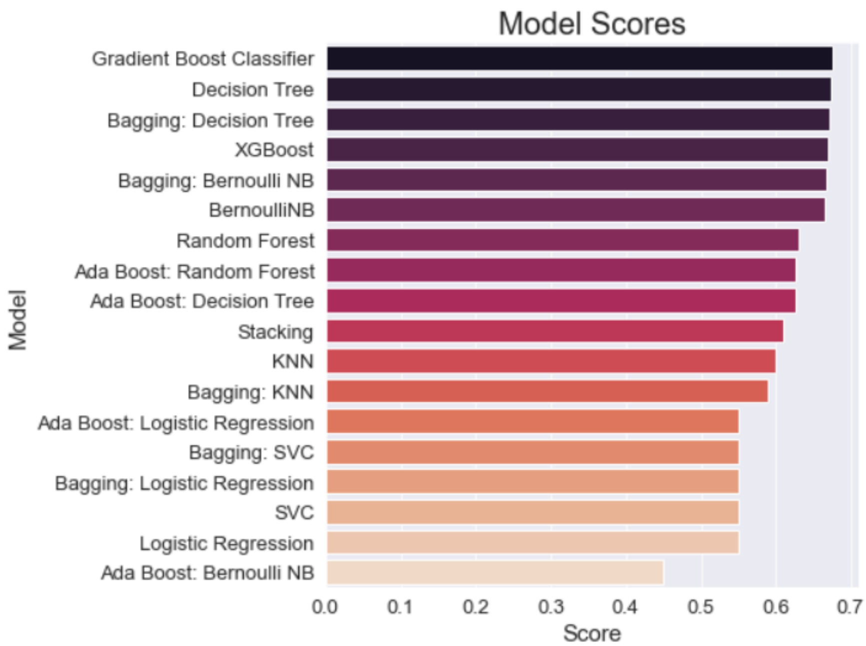
	precision	recall	f1-score	support
0	0.65	0.87	0.74	2199
1	0.73	0.43	0.54	1801
accuracy			0.67	4000
macro avg	0.69	0.65	0.64	4000
weighted avg	0.68	0.67	0.65	4000

Comparison of Results: With Principal Components = 20 vs No Reduced Dimensions

	Scores
Stacking	0.669
Bagging: Logistic Regression	0.667
XGBoost	0.666
Bagging: Decision Tree	0.666
Gradient Boost Classifier	0.663
Ada Boost: Logistic Regression	0.662
Logistic Regression	0.660
SVC	0.659
Decision Tree	0.658
Bagging: SVC	0.658
KNN	0.635
Ada Boost: Bernoulli NB	0.635
Random Forest	0.634
Bagging: KNN	0.632
Ada Boost: Random Forest	0.630
Bagging: Bernoulli NB	0.626
BernoulliNB	0.624
Ada Boost: Decision Tree	0.620



	Scores
Gradient Boost Classifier	0.676
Decision Tree	0.674
Bagging: Decision Tree	0.672
XGBoost	0.670
Bagging: Bernoulli NB	0.668
BernoulliNB	0.666
Random Forest	0.630
Ada Boost: Random Forest	0.627
Ada Boost: Decision Tree	0.626
Stacking	0.609
KNN	0.600
Bagging: KNN	0.590
Ada Boost: Logistic Regression	0.550
Bagging: SVC	0.550
Bagging: Logistic Regression	0.550
SVC	0.550
Logistic Regression	0.550
Ada Boost: Bernoulli NB	0.450



Outcomes:

- To recall, our model takes in all the factors involved such as subject demographic, location, officers involved etc. and predicts whether force will be required.
- We can see that conducting dimensionality reduction improved the average result of our models, revealing more homogeneous results.

- However, the top models were unable to exceed results without dimensionality reduction. Therefore, we will choose the top models from our trials without model scores, and experiment on a full dataset (147,895 records vs 20,000 records).

Final Model on Full Dataset (147,895 records)

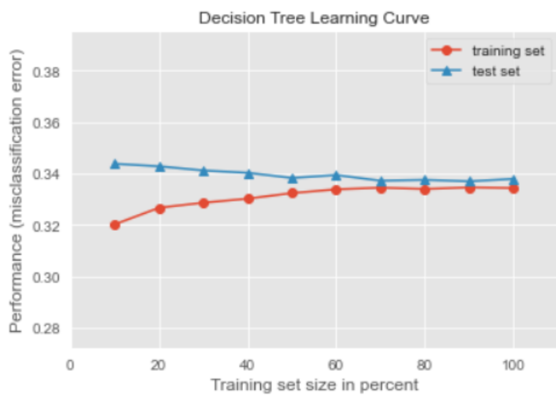
For the final model, we will be examining the model using our top models from the testing phase:

- Decision Tree Classifier
- Decision Tree Bagging Classifier
- Gradient Boosting Classifier

For training the model, we will again use a train test split of 20%, which is in actual volume is nearly double the records for our previous 20,000 record tests.

Parameters for Decision Tree:
Due to the change in volume for our dataset, finding the best parameters for this training set will be required. In this case, GridsearchCV reported an optimal max leaf node value of 95.

<table>
<tr><th>Model and Accuracy</th><th>Learning Curve</th><th>Classification Report</th></tr>
<tr><td>Decision Tree Classifier

Accuracy: 0.66</td><td>
</td><td><pre>
Classification Report:

              precision   recall f1-score   support

           0   0.64   0.87   0.73   15998
           1   0.73   0.42   0.53   13581

    accuracy   0.66   29579
   macro avg   0.68   0.64   0.63   29579
weighted avg   0.68   0.66   0.64   29579
</pre></td></tr>
<tr><td>Decision Tree Bagging Classifier

Accuracy: 0.66</td><td>
</td><td><pre>
Classification Report:

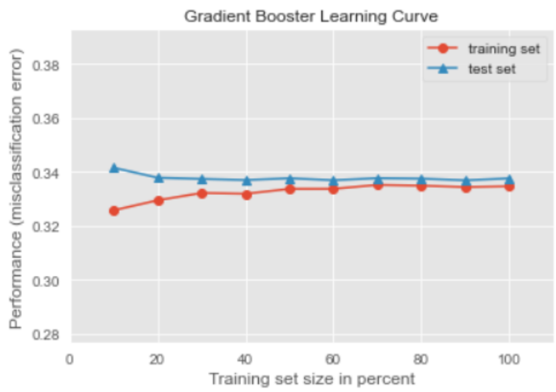
              precision   recall f1-score   support

           0   0.64   0.87   0.74   15998
           1   0.73   0.42   0.54   13581

    accuracy   0.66   29579
   macro avg   0.68   0.64   0.64   29579
weighted avg   0.68   0.66   0.64   29579
</pre></td></tr>
</table>

Gradient Boost Classifier:

Accuracy: 0.66




```
Classification Report:

              precision    recall  f1-score   support

           0       0.64      0.88      0.74     15998
           1       0.75      0.40      0.52     13581

    accuracy                           0.66     29579
   macro avg       0.69      0.64      0.63     29579
weighted avg       0.69      0.66      0.64     29579
```

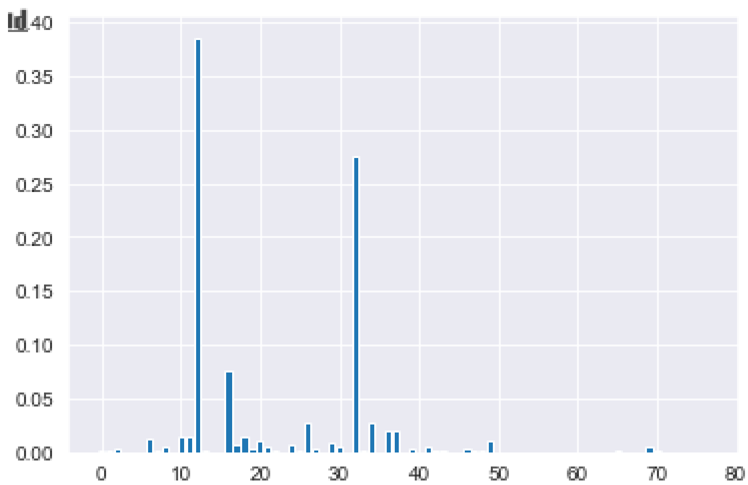
<u>Final Results:</u>

Model	**Scores**
Decision Tree	0.662
Decision Tree Bagger	0.664
Gradient Booster	0.662

Feature importance:

From our final models, we see that the following features had the largest impact on our model, seen in the graph below, with the x axis displaying the feature number, and the y axis showing the coefficient of that feature.:

Feature	**Coefficient**
Location: Dwelling	0.38587
Subject Age:	0.27518



Final Outcomes and Conclusions of Model

- In our final models, we see that in the learning curve, each model was able to converge at the optimal balance between variance and bias, converging at the best possible result given the ambiguous nature of our dataset, achieving scores of 0.66, consistent with the optimal scores for our training models.

- Within the final models, the most important features appeared to be whether or not the location is at a person's dwelling, as well as the subject's age, and seemed to be consistent between the models.

- The processing time of each model took no longer than 2 minutes, contrasting to our experimental phase, in which SVC, KNN and Random Forest models took a considerable amount of computing time and power. Considering the size of the dataset, we remain optimistic on this model regarding training time, when adapted to other datasets.

- After testing, we can confidently use this model to predict our target variable in various instances of a use case, predicting an expected outcome, along with its probabilities.

Summary of Statistical Analysis of Data:

The following are the main insights for our data **analysis** process.

- 45.94% of officers are required to use force, meaning the majority of officers are capable of settling disputes using tactical communication and compliant handcuffing.

- Multiple factors influence an officer's decision to use force, with the most frequent being a subjects build and drug use. However, in cases including factors of mental health, alcohol and behavioural disorders, the rate of force usage increases.

- Officers use force primarily as a mechanism for protection of themselves, other officers and to prevent escape.

- Most incidents occur on the street or on highways.

- The highest incident rates are present in Westminster, Lambeth, Newham and Croydon.

- More than half (56%) of subjects are not primarily compliant with police, and exert some form of resistance.

- 3.3% of officers are assaulted by the subject.

- The highest-risk main duty for an officer are mobile patrols, experiencing 78% of incidents which require force.

- Only 2.3% of deployed officers are single-crewed. However, they are 1.35x more likely required to use force compared to partnered or crewed units.

- Only 38.2% of officers are properly trained in taser usage.

- The largest age demographic for subjects is the 18-34yo range. Trends for use of force regarding subject age remains proportional to the population distribution, except for 0-10. It is alarming that 91% of minors under 10 require usage of force.

- In cases which require force, 83.4% of those are male.

- While 84.2% of all cases have no known disability, while 15.4% of those have either a mental or physical disability. Officers have an alarmingly large rate of force usage on the mental health population over any other demographic.

Recommendations

Suggestions for London Police Agencies

- Implement further or more extensive training in tactical communications. By implementing such a course, officers will be better trained to settle disputes through their ability to mitigate incitement of violence and handling aggression in non-violent methods.

- Police departments to implement education requirements prior to successful recruitment.

- Implement further training/education on dealing with mentally-ill individuals without resorting to violence.

- Use models to **appropriately allocate resources**. If an incident is located in a high risk area, with high risk subject conditions, deploy experienced officers with support of 1 or more officers, equipped with proper safety equipment.

- Officers who are more experienced to be shifted on mobile patrol duty, or be the primary candidates sent to incidents located in high-risk boroughs and locations such as streets/highways to mitigate risk of extended altercations.

- Develop protocols which remove single-crew deployments. Single-crewed officers are of highest risk in experiencing aggression.

- Implement training on settling disputes with minors in non-violent solutions. Furthermore, implement protocols to send officers with experience with children to such incidents. A rate of 91% force usage on minors is damaging to the subject, officer and negatively influences public opinion.

- Suggest implementation of police-department/community initiatives to reduce the divide between the public and officers. As a recommendation, communities with African descent should be a priority due to their statistically high proportion of incidents (violent and non-violent) with officers compared to other ethnic backgrounds in London.

- Implement suggested models and distribute to all London police departments for use in real time.

Implementation of Model

Suggestions on how to implement the final models.

The goal of this model is to provide officers with details on whether or not to use force when responding to any deployment. When responding to a situation, an officer must be prepared for any situation, and be trained to appropriately follow protocols which maintain the safety of himself, the subject and the general public. These are complex factors to consider, and thus this model provides a certain level of expectation to assist the officer in preparation, as well as the triage officer in allocating additional resources to the scene.

The final model consists of 77 potential features, and due to the time-series nature of officer deployment, it is important to keep in mind as such. Therefore, it is important that a triage officer be able to use the model with a certain level of ease and efficiency.

Our proposed implementation of the model requires a simple and easy-to-use graphical user interface (GUI) software with the following features to accommodate for such requirements:

- Each input variable must have a default. These defaults are essential for such a reason that a triage officer may not receive all required feature information across an emergency call. By having a default for each variable, not only does this reduce input time, but it gives the officer a more accurate probability by giving value to any assumptions made, compared to having null values or no information on other variables.

- Such defaults will be demographically related. For example, if subject demographics are unknown, assumptions will be able to be drawn from the location and Borough of the incident, such as the average demographic or modal value. Again, each variable input instance is dependent on the call given to triage officers.

- The output of such a software will report back whether to expect to use force, as well as the probability of that outcome.

Conclusions

After performing numeric analysis on our data, many observations have appeared to be statistically significant enough to make recommendations based on such analysis. These recommendations may pertain to either the police department protocols, resource allocation and training suggestions on how to properly deal with subjects and scenarios outlined by our data.

The models developed create a practical way to foresee and diminish harm to both civilian and officer populations, by informing police agencies on appropriate resource deployment suggestions.

Ultimately the goal of this process is to find ways to reduce risk to both officers and the public regarding safety, by reducing usage of force to a point of last resort, and in turn increase police sentiment and inversely treatment of subjects in the post-Ferguson era of policing. It is expected that through implementations of potential recommendations, these goals are possible over time, and provides a tangible way for police departments, not just in London, but around the world, to enact positive change by being more prepared in resources and increased capacity to serve and protect. Overall, the optimistic effects of these combined implementations may see a reduction in police brutality, public opinion, and be tangibly evident through a steady reduction over time in lawsuit settlements for malfeasance and wrongful arrests.

References

Baker, K., 2019. *Should all frontline police officers use Tasers?*. [online] BBC News. Available at: <https://www.bbc.com/news/uk-49405999> [Accessed 5 March 2021].

McLay, C., 2018. *Sentiment Analysis: The Missing Link in Police Performance Management Systems*. [online] Policechiefmagazine.org. Available at: <https://www.policechiefmagazine.org/sentiment-analysis/> [Accessed 4 March 2021].

Peeples, L., 2020. *What the data say about police brutality and racial bias — and which reforms might work*. [online] Nature.com. Available at: <https://www.nature.com/articles/d41586-020-01846-z> [Accessed 6 March 2021].

Phys.org. 2020. *Police officers' views before and after Ferguson counter accuracy of 'Ferguson Effect'*. [online] Available at: <https://phys.org/news/2020-03-police-officers-views-ferguson-counter.html> [Accessed 6 March 2021].

Quinn, B., 2019. *UK police out-of-court settlements total £30m in four years*. [online] the Guardian. Available at: <https://www.theguardian.com/uk-news/2019/dec/26/uk-police-out-of-court-settlements-total-30m-in-four-years> [Accessed 6 March 2021].

Shaw, D., 2019. *Rise of knife crime in England and Wales*. [online] BBC News. Available at: <https://www.bbc.com/news/uk-42749089> [Accessed 6 March 2021].

Dataset:
Metropolitan Police Dept. 2021, *Use of Force Dataset*, UK Government. <https://data.london.gov.uk/dataset/use-of-force>

Code and Notebooks can be found at: https://github.com/jdomingo117/use_of_force

Notebooks included:

- *Conduct and assault analysis*
- *Location analysis*
- *Location mapping*
- *Main duty analysis*
- *Taser analysis*
- *Feature engineering notebook*
- *Final model notebook*
- *Impact factors for use of force*
- *Model feature selection*
- *Reasons for force analysis*
- *Single crew analysis*
- *Subject age & gender analysis*
- *Subject disabilities analysis*
- *Subject ethnicity analysis*
- *Preliminary model training notebooks*

Libraries used in code:

- *Numpy*
- *Seaborn*
- *Matplotlib*
- *Pandas*
- *SkLearn*
- *XGBoost*
- *Mlxtend*
- *Datetime*

Algorithms used in code:

- *Cross val score*
- *Train test split*
- *GridsearchCV*
- *Decision Tree Classifier*
- *Logistic Regression*
- *KNeighborsClassifier*
- *Support Vector Classifier*
- *Naive Bayes Classifier*
- *Bagging Classifier*
- *AdaBoost Classifier*
- *Random Forest Classifier*
- *Gradient Boosting Classifier*
- *XG boost Classifier*
- *Stacking Classifier*