

Project Proposal

- 1) Things we are trying to prove:
 - a) Committing certain crimes can have a greater potential of going to jail when compared to other crimes committed.
 - i) We predict that when a person is pulled over and found with a controlled substance they have a higher probability of ending in jail. The way we measure success is by comparing the results with all other types and see which crime led to the most arrest.
 - b) What gender commits the most crimes overall, male or female and of those crimes committed which gender has a higher percentage of ending up in getting arrested.
 - i) We predict that more males get stopped and searched compared to females. The way we measure if the prediction was a success is by analyzing how many males get stopped and searched compared to females.
 - c) Is there any sort of discrimination when it comes to the police stopping people of certain races?
 - i) We predict that there *is* some sort of discrimination when it comes to the police stopping people of certain races. The way we measure success is by comparing the races, then checking which race got pulled over the most.
 - d) What age range is more likely to get pulled over?
 - i) We predict that people in their 20's get pulled over the most. The way we measure if our prediction was successful is by checking which age group has the higher percentage of being pulled over.
 - e) What's the percentage of outcomes committed that resulted in no further action?
 - i) We predict that crimes that's harder to prove will result in no action being taken. The way we measure if our prediction is success is by comparing crimes that had action taken with crimes that did not. We will show the result in a pie chart.
 - f) What are the top objects found on a search?
 - i) In regards to the objects that have been found, what are the cases where extra clothing was needs to be taken off to find the objects searched
 - g) How many crimes are committed within a certain time(year, Weeks) ?
 - i) We predict that different seasons will result in higher crimes than others. The way we measure if our prediction is success is by comparing the top five crimes and measuring their action throughout one year. We will also look into crimes that are committed within a given week. We predict more crimes happen on the weekends compared to the weekdays.
 - h) Plot a map for the data set using Longitude and Latitude.

- i) Plotting the map will allow to see where crimes are located as well as the amount in an area
- i) Which crimes are the safest to commit?
 - i) Many times when cops record a crime, they record that they let off the person who committed the crime with a caution, warning, or penalty. We consider these the “safer” crimes to commit. We predict that crimes like noise-complaints, and shoplifting are likely safer to commit and are the least likely to lead to arrest when caught.
- j) Which crimes are the most successful?
 - i) Many criminals manage to get away with their crimes. For example, there are cases wherein a suspect is unable to be prosecuted, or an investigation is completed but officers were still unable to find a suspect. The most successful crimes are the ones where the criminal gets away, and faces no penalty. We predict crimes involving theft are the most successful in this regard.
- 2) How we plan to collect our data:
 - a) The dataset we are using is extremely large. We will be picking and choosing a certain amount of files we want to include in our project. We plan to manipulate the data using a dataframe.
- 3) Challenges we anticipate:
 - a) Communicating
 - b) Regretting the section of the dataset we chose to manipulate
- 4) Why we picked what we picked:
 - a) We picked this dataset to analyze because we want to see how much of the crime committed in the UK ends in an arrest. Based on this data the police force in the UK can see what leads to more arrest and focus their resources more to that type of stop and search crime.

Progress Report

1) Introduction to the data:

- a) The data is provided in several folders. One folder for each month for the years 2017-2020. Inside each of those folders there are multiple CSV files pertaining to “stop and search” in the United Kingdom. Each individual file is about 4.6MB. We have combined all these files into one CSV and the size of that file is 253MB.

This file contains all the attributes of the other files:

- i) Type
 - ii) Date
 - iii) Part of a policing operation
 - iv) Latitude
 - v) Longitude
 - vi) Gender
 - vii) Age range
 - viii) Self-defined ethnicity
 - ix) Officer-defined ethnicity
 - x) Legislation
 - xi) Object of search
 - xii) Outcome
 - xiii) Outcome linked to object of search
 - xiv) Removal of more than just outer clothing
- b) <https://drive.google.com/file/d/1FvU0kskpMWYimx5XJXzjfRDhT0cWlnwv/view?usp=sharing>
<https://github.com/Bryanna178/418-Final-Project>

	Type	Date	Latitude	Longitude	Gender	Age range	Self-defined ethnicity	Officer-defined ethnicity	Legislation	Object of search	Outcome	Removal of more than just outer clothing
0	Person search	2017-03-01 11:10:54+00:00	51.486273	-2.521011	Female	over 34	NaN	NaN	Misuse of Drugs Act 1971 (section 23)	Controlled drugs	Nothing found - no further action	False
1	Person search	2017-03-01 11:32:44+00:00	51.187611	-2.546681	Male	10-17	White - White British (W1)	White	Police and Criminal Evidence Act 1984 (section 1)	Article for use in theft	Nothing found - no further action	False
2	Person search	2017-03-01 12:10:53+00:00	51.346507	-2.971149	Male	over 34	NaN	NaN	Misuse of Drugs Act 1971 (section 23)	Controlled drugs	Nothing found - no further action	False
3	Person search	2017-03-01 12:10:53+00:00	51.346507	-2.971149	Male	over 34	NaN	NaN	Misuse of Drugs Act 1971 (section 23)	Controlled drugs	Nothing found - no further action	False
4	Person search	2017-03-01 14:02:17+00:00	50.941519	-2.640688	Male	over 34	White - White British (W1)	White	Misuse of Drugs Act 1971 (section 23)	Controlled drugs	Nothing found - no further action	False
c)

2) Data Collection Process:

- a) How we collected our data

- i) We combined all the data into one CSV file for everyone to manipulate.
- b) How we cleaned our data
 - i) We cleaned the data by removing the columns:
 - (1) Part of a policing operation
 - (2) Policing operation
 - (3) Outcome linked to object of search
 - ii) Convert 'Date' column into datetime format
- c) Difficulties in the beginning steps:
 - i) The files were separate and needed to be combined in order for anyone to start manipulating the data.
- 3) A summary of challenges and observations made so far:
 - a) There are quite a few NaN values that we need to keep in mind when manipulating the data.
- 4) What we plan to do:
 - a) Plan out how to prove the hypothesis we mentioned we wanted to prove in the project proposal.
 - b)
- 5) Group member duties:
 - a) Since a cleaned dataset file has been made, everyone will have their own notebook to showcase what they have done with the dataset.
 - b) Richard
 - i) Committing certain crimes can have a greater potential of going to jail when compared to other crimes committed.
 - ii) What gender commits the most crimes overall, male or female and of those crimes committed which gender has a higher percentage of ending up in getting arrested.
 - c) Bryanna
 - i) Is there any sort of discrimination when it comes to the police stopping people of certain races?
 - ii) What age range is more likely to get pulled over?
 - d) Quincy
 - i) What's the percentage of outcomes committed that resulted in no further action?
 - ii) What are the top objects found on a search?
 - e) Ivan
 - i) How many crimes are committed within a certain time(year, Weeks) ?
 - ii) Plot a map for the data set using Longitude and Latitude.
 - f) Aashish
 - i) Which crimes are the safest to commit?
 - ii) Which crimes are the most successful?