A screenshot of a computer

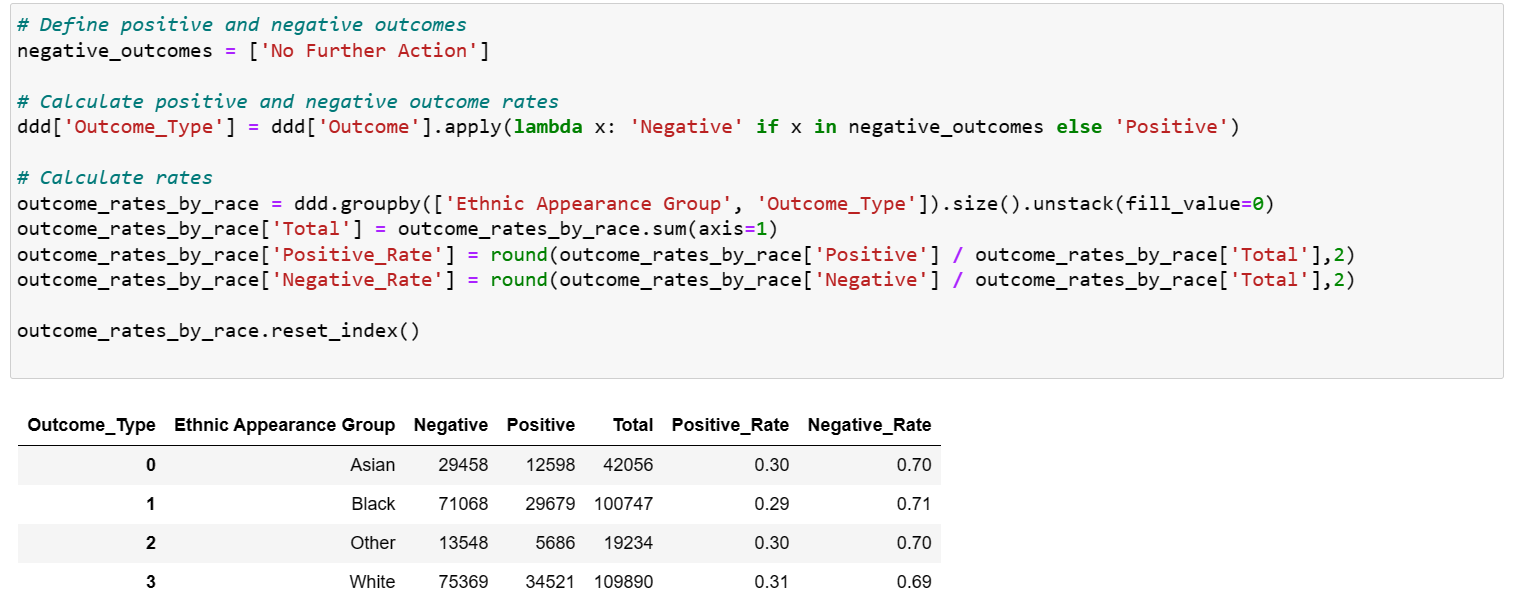
Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated



# AUTHOR

VIVID IBORI

# Project Overview

Objective

The primary objective of this project is to analyse data related to stop and search incidents to uncover patterns, trends, and insights. The goal is to provide actionable recommendations to improve the effectiveness of stop and search practices while ensuring fairness and minimizing bias.

Stop and search is a policing practice used to prevent and detect crime by allowing officers to search individuals they suspect of carrying illegal items, such as drugs, weapons, or stolen property. This practice has generated significant debate regarding its effectiveness, fairness, and impact on communities.

In many countries, stop and search practices are governed by specific legal frameworks and guidelines to ensure they are conducted lawfully and fairly. In the UK, for example, the practice is regulated by the Police and Criminal Evidence Act 1984 (PACE) and its subsequent amendments.

While some community members appreciate the visible police efforts to tackle crime, others feel targeted and harassed. The psychological impact on individuals and communities can be significant, contributing to feelings of alienation and mistrust.

## Scope

This project involves collecting, cleaning, and analysing stop and search data from the Metropolitan Police London Datastore. The focus will be on understanding the demographics, outcomes, and geographic distribution of stop and search incidents within the last 24 months, from May 2022 to April 2024. The dataset has been refreshed every month since 2017.

## Key Questions

* What are the demographic characteristics of individuals subject to stop and search?
* What are the outcomes of these searches (e.g., arrests, no further action)?
* What is the geographical distribution of the stops.
* Is there evidence of bias or disproportionate targeting of certain groups?

### Data Collection

* Source: Stop and search data from the law enforcement agency's database (Metro police London Datastore).
* Period: The analysis will cover data from the past twenty-four months.
* Variables: Age, gender, ethnicity/race, reason for the stop, outcome, location, date.

### Methodology

Data Cleaning.

* Handle missing values and inconsistencies.

The dataset originally contained over 308,000 records. After removing less than 3% of records that had missing values, the dataset was reduced to over 301,000 records.

* Normalize data (categorical variables).

In the dataset, variables such as the reason for the stop needed to be consolidated into fewer categories, following recommendations from the London Metropolitan Police's official website. Instances where different drugs were the reason for the stop were grouped under the ASB category. Stops related to weapons were categorized as weapons, while crimes such as theft and stolen property were classified under Key Crimes. Additionally, the MPS Area variable was standardized into Central, West, East, North, and South.

* Removing Outliers.

Outliers in any analysis can significantly impact the results and interpretation of the data, as well as its quality and integrity. Upon checking for outliers in this dataset, I found records with ages such as -1, 1, 2, 4, 5, 6, 7, 124, and 145. These values are likely errors from data entry or mistakes by the officers, as they are implausible given that the data pertains to vehicle stops and searches. Consequently, these outliers needed to be removed.

Using British law as a reference, I set the minimum age threshold to 16, even though the minimum age for obtaining a driver's license in the UK is 17, because there is no upper age limit for driving as long as individuals are deemed fit by their GP (doctor). Therefore, I did not include a maximum age threshold.

After establishing this threshold, I applied the Interquartile Range (IQR) method to detect and remove outliers using a 1.5 boundary. Following this process, the dataset size was reduced to over 270,000 records, retaining approximately 90% of the original dataset. The distribution of ages in the cleaned dataset is as follows: minimum = 16, Q1 = 20, median (Q2) = 25, Q3 = 35, and maximum = 57.

With the outliers removed, it is reasonable to conclude that the dataset is now free of noise that could negatively impact the results of this project.

Exploratory Data Analysis (EDA):

* Descriptive statistics to summarize the data.
* Visualizations to identify patterns and trends.

Effectiveness Analysis:

* Calculate the success rate of searches leading to arrests or discoveries of illegal items.
* Compare against demographic and geographic variables.

Tools and Technologies

Data Processing: Python (Pandas, NumPy), Power Query Power BI

Visualization: Matplotlib, Seaborn, Power BI

# KPI’S

* Average Age = 28.
* Total Stops = 271,927.
* Positive Outcome = 82,484 (30.33%).
* Negative Outcome = 189,443 (69.67%).

# Answers to Key Questions

1. **What are the demographic characteristics of individuals subject to stop and search:**

* **Gender**: Over 90% (244,922) of the recorded individuals are males, 8.7% (23,566) are females, and 0.93% (2,532) have no response.
* **Race**: The race demographic is slightly dominated by Whites with 109,890 (40.41%) individuals, followed by Blacks with 100,747 (37.05%) individuals.
* **Age Bracket**: The age bracket with the highest record is 18-24, accounting for 35.77% of the total records, followed by the 25-34 age bracket, which makes up 26.7% of the total records.
* **Reason for Stop**: The most common reason for stops and searches is drug-related incidents (ASB), comprising 63.19% of the total records, followed by key crimes (stolen property) with 19.45% of the total data.

1. **What are the outcomes of these searches (e.g., arrests, no further action)?**

The outcome of this searches are categorised as Negative or positive outcomes, when the outcome is \*No Further Actions\* it is negative hence any other category is termed as a positive outcome situation like arrest and community service.

The positive outcome of the searches which is the main reason for stops is made up of 30.33% (82,484) of the total records and negative outcome which officers tends to minimise in order to not create fear and tension is 69.67% (189,443).

1. **What is the geographical distribution of the stops.**

The location for this stops and search records is London United Kingdom comprising of five regions (Central, North, West, East, South), all this region has a total of 32 borough or towns. The highest stops occurred in Westminster borough followed by Tower hamlets and Southwark all in the central region (Tree map in the report page), the central region as the highest incident of over 109,000 records.

1. **Is there evidence of bias or disproportionate targeting of certain groups?**

The analysis indicates a significant association between race and the likelihood of undergoing stop and search procedures, suggesting evidence of bias towards racial demographics beyond the White population. Specifically, demographic groups such as Black individuals, characterized by lower positive outcomes, exhibit higher frequencies of negative records. This observation implies that stops involving Black individuals may lack viable evidence, as evidenced by the comparatively lower proportion of positive outcomes.

In contrast, the White demographic demonstrates a higher incidence of positive outcomes and a lower frequency of negative records, indicative of stops conducted with more substantial evidence. This approach aims to mitigate potential intimidation and fear among road users. The provided visualization, Python table and Chi2 analysis further substantiate these findings.

# Recommendations

Although the consequences of failing to stop an individual who is committing a crime are more detrimental than falsely stopping an innocent person, it is also important to reduce the occurrences of falsely stopping a road user to prevent harassment and alleviate fear among them. Given the high proportion of negative search outcomes, it's crucial to address potential factors contributing to this imbalance. Here are some recommendations:

1. **Review and Refine Stop and Search Procedures**: Conduct a comprehensive review of current stop and search protocols to identify any biases or inefficiencies. Collaborate with law enforcement agencies, community leaders, and civil rights organizations to refine procedures and ensure they are fair, transparent, and effective.
2. **Training and Education**: Provide extensive training to law enforcement officers on unconscious bias, cultural sensitivity, and effective communication techniques. Emphasize the importance of treating all individuals with dignity and respect during stop and search encounters. Additionally, educate community members about their rights and how to effectively engage with law enforcement during these encounters.
3. **Community Engagement and Accountability**: Foster open dialogue and collaboration between law enforcement agencies and the communities they serve. Establish community oversight mechanisms, such as civilian review boards, to increase transparency and accountability in stop and search practices. Engage with community stakeholders to develop mutual trust and understanding.
4. **Regular Evaluation and Adjustment**: Continuously evaluate the effectiveness of implemented reforms and adjust policies and procedures as needed. Solicit feedback from both law enforcement and community members to ensure ongoing improvement and accountability.

By implementing these recommendations, the project can work towards reducing the high proportion of negative search outcomes and promoting fair and equitable stop and search practices that enhance public safety while respecting individual rights and dignity.