

Analytics in Business – Problem Set 2

The Metropolitan Police Stop and Search Initiatives Effectiveness Analysis

Group L

Josiane Samira Ulrike Dufour 01388388

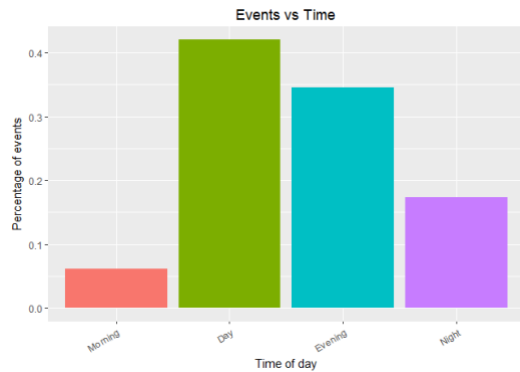
Otto Godwin 01422474

Zana-Ljubica Krsticevic 01423239

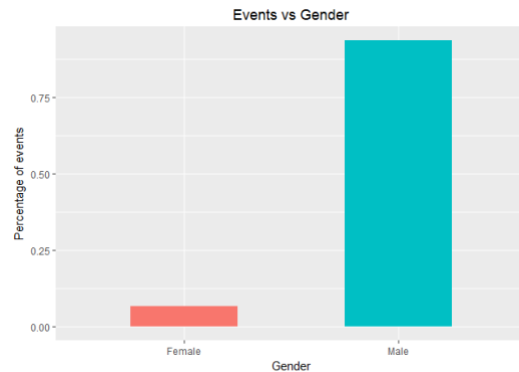
Seongmin Lee 01247436

Shuyu Zhou 01382817

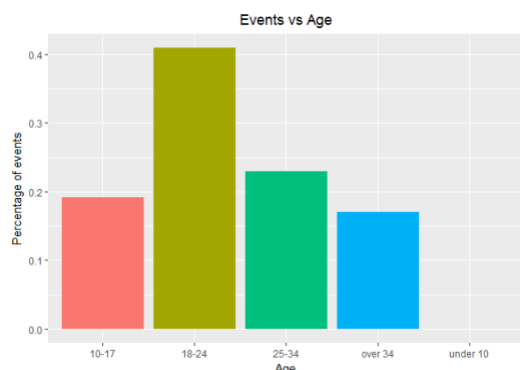
A. Describe



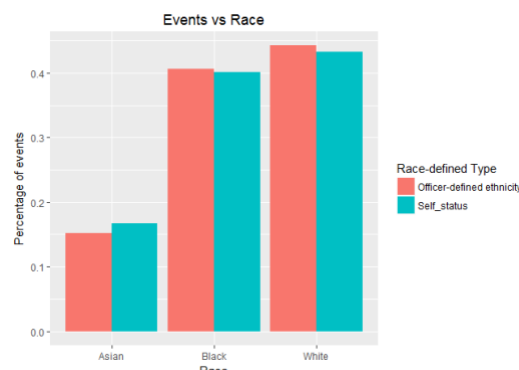
- 1) Most stop-and-search events are happening between 10:00 and 23:00
- 2) The peak time slot is daytime (10:00-17:00), which includes 42% events
- 3) Also, about 34% (the second largest) events are in the evening
- 4) Night events are not as little as expected. The percentage is about 17% and more than that in morning (maybe challenge the belief people hold)
- 5) Only about 6% stop-and-search events happen in the morning



The number of stopped male is about 15 times higher than that of female, which, generally...makes sense...



- 1) The most common age range for stop-and-search events is 18-24, which represents 39% of all people stopped and is significantly higher than any other group
- 2) The percentage for 10-17, 25-34 and over 34 age groups are individually 18%, 22% and 16%
- 3) It is really unusual for people who are under 10 to be stopped by police (only 0.0025%)



- 1) Most people who are stopped are white, and the percentage is about 44%
- 2) The percentage of black people stopped is about 40%
- 2) About 20% of people stopped are Asians
- 2) The officer-defined and self-defined race for people is roughly the same. However, in some cases, officers mid-defined Asian individuals

B. Evaluate: Harassment, or Efficiency?

In July 2016, 32.58% of stop and searches lead to an outcome of 'something'. Therefore, the Metropolitan police wasn't doing particularly good at looking for a crime where it occurs.

The finding 'something' rate does, however, vary significantly depending on the reason that a person was searched (object of search). In 34% of cases where people were searched on suspicion of possessing drugs police subsequently found 'something'. In contrast, in only 19% of cases where people were searched because a police officer suspected that they held firearms, the outcome was 'something'.

To improve the 'success' ratio, we could include the data about London boroughs based on ethnic distribution and combine that with the number of successful searches in each specific borough. This would help us identify if neighbourhoods were being targeted fairly and potentially improve the ratio of successful searches. We could also check whether boroughs that have more reported crimes/known crime networks are being searched more or if stop and searches happen without a clear strategy. To tackle these questions, we would need to perform regression analysis to see if these new variables are significant in predicting the outcome of the stop and search event. If we also had the data about people who were previously searched and the outcome of each of these events, it could also help us understand if the wrong people are being targeted, and if a more focused approach would yield better results.

What are searches targeted at?

Object	Proportion of searches in total numbers	Proportion of searches in percentage	Ratio of 'successful' searches
Drugs	7625	64%	34.37%
Criminal damage	4186	35%	29.72%
Firearms	131	0.01%	19.08%
Threatening or harming	2	0.00%	50%
Total	11944		32.58%

C. Explain: Racial Fairness?

In the regression model, we didn't find a strong relationship between the dichotomous variable and the racial background. Holding other factors fixed and using Asians as the base group, we found that if the race of people who get stopped is black or white, the odds are only about 0.03 and 0.04. That means if people are black or white, the probability of 'something' versus 'nothing' increases only by 3% and 4%, compared to Asians.

Furthermore, the p-values are insignificant (0.6 and 0.5, respectively). Therefore, we believe there is no obvious indication of racial discrimination when it comes to the outcome of the stop and search event. However, when looking at the probability of being stopped and searched, there are significant differences in terms of racial background. The black population had the highest rate of stop and search at 4.42 per 1,000. The rate for Asian people was 1.14 per 1,000, and it was 1 per 1,000 for white people.

Gender and time of the day also don't seem to have a strong influence on the dependent variable (with the exception of more crimes occurring after midnight). However, we did get statistically significant results for age in all 3 groups.

D. Predict: Does PredPol really work?

- a) The assumption put forward in the exhibit posits that an increased police presence would lead to an increased number of stop and search events where something is found, with successful stop and search initiatives leading to a decreasing number of events found in areas where there are frequent stop and search events. However, the graph shows no clear relationship between the neighbourhoods with highest ratios of found events and number of stop and search events - this would be confirmed by a larger number of red points higher on the y-axis, as well as in the top right part of the graph, and more yellow points on the lower part of the y-axis, as well as the bottom left part of the graph. However, the red points (indicating the areas where the ratio of found events compared to total search events is highest) are mostly clustered around the bottom part of the graph, in areas where there are less stop and search events, with moderate ratios (points in orange) randomly spread out throughout the graph.
- b) One way to assess whether increased police presence leads to fewer crimes by neighbourhood would be to compare neighbourhoods with similar crime rates, with one group acting as control with typical police presence, and the other with increased police presence, and assess the crime rates over time. If the neighbourhoods with increased police presence see their crime rates (or higher incidence of 'found' events for each stop and search event) drop as compared to their corresponding control neighbourhood, then it can be assumed that increased

police presence does help decrease crime rates. In addition to this panel data, collecting statistics on how many times an individual was stopped in the past, and how many times they were stopped and nothing was found or something was found, would permit analysis on whether stop and search events are targeting innocent people, and whether crimes are being perpetrated by the same few individuals over and over again, in which case it would guide policing efforts to narrow down their search to only relevant and known criminals. Furthermore, an analysis of whether an individual is stopped as part of a police investigation will enable to compare the efficiency of targeted crime-stopping tactics vs random stop and search events.

- c) To track the effectiveness of police presence over time, and in addition to the data already being collected in the stop and search dataset, we would need the following new variables:

Stopped_Something	variable with counts of occurrences for stopped before and something found
Stopped_Nothing	variable with counts of occurrences for stopped before and nothing found
Reason	specific reason why an individual was stopped during a stop-and-search process

Given the solution suggested in part b, a few key regressions that would help derive insight into the effectiveness of stop-and-search initiatives would be the following:

- Outcome as a function of Stopped_Something ~ Reason, Type, Part of a policing operation, Gender, Age, Ethnicity, and Object of search
- Outcome as a function of Stopped_Nothing ~ Reason, Type, Part of a policing operation, Gender, Age, Ethnicity, and Object of search
- Outcome as a function of Stopped_Nothing ~ Stopped_Nothing, Reason, Part of a policing operation

E. Automate: Startup pitch

Violence is a public and mental health issue, and stop and search initiatives are not the solution. The key to its eradication lies in IDENTIFYING, INTERRUPTING, and EDUCATING community members, providing a holistic approach to fighting crime by involving everyone concerned. First, the majority of crime is committed by a few individuals. By analyzing data related to crime within and between communities for repeat offenders, these individuals can be targeted for the right reasons, and offered proper support through job training and mental health services, with more severe consequences for any further criminal activity.

Next, identifying the most strongly connected members of the community by collecting data on the frequency of their interactions with others, and training the most influential members to be interrupters - trusted members of the community who will hold a momentous role: to stop violence from spreading, by being the eyes and ears of the community, and being a respected and trusted person, everyone can feel safe turning to for advice and guidance. Interrupters are well-aware of what issues their fellow community members face, having had to face them themselves. Coupled with support from the police, these efforts will reduce violence within communities, leading to dramatic decreases in crime rates.

Appendix: Method used in part C

In our stop and search dataset, there are two measures of ethnic groups: officer-defined ethnicity which is police's classification of people according to their appearance, and self-defined ethnicity which is defined by the people themselves. Although self-reported ethnicity should be preferred, in this case, we combined both measures as we believe 'ethnic appearance' has a huge influence on racial fairness.

The four ethnic appearance categories are estimated as follows:

White	White - Any other White ethnic background (W9)
	White - White British (W1)
Black	Black or Black British - African (B2)
	Black or Black British - Any other Black ethnic background (B9)
	Black or Black British - Caribbean (B1)
	Mixed - White and Black African (M2)
	Mixed - White and Black Caribbean (M1)
Asian	Asian = Asian or Asian British - Any other Asian ethnic background (A9)
	Asian or Asian British - Bangladeshi (A3)
	Asian or Asian British - Indian (A1)
	Asian or Asian British - Pakistani (A2)
	Chinese or other ethnic group - Chinese (O1)
Other	Mixed - Any other Mixed ethnic background (M9)
	Chinese or other ethnic group - Any other ethnic group (O9)
	Not Stated (NS)

The black population had the highest rate of stop and search at 4.42 per 1,000. The rate for Asian people was 1.14 per 1,000, and it was 1 per 1,000 for white people. If the self-defined ethnicity is included, the rate of stop and search for white population drops to 0.84 per 1,000, and the rates for the black and Asian populations to 3.65 per 1,000 and 0.93 per 1,000 respectively.

Part(c) - explain

Table 1: Regression Results

	<i>Dependent variable:</i>
	Something
time_of_dayNoon(10:00-17:00)	−0.016 (0.087)
time_of_dayEvening(17:00-23:00)	−0.105 (0.089)
time_of_dayMidnight(23:00-5:00)	−0.316*** (0.096)
‘Age range‘18-24	0.705*** (0.061)
‘Age range‘25-34	0.743*** (0.067)
‘Age range‘over 34	0.430*** (0.073)
‘Age range‘under 10	−10.332 (113.719)
GenderMale	−0.071 (0.082)
‘Officer-defined ethnicity‘Black	0.032 (0.062)
‘Officer-defined ethnicity‘White	0.040 (0.062)
Constant	−1.093*** (0.136)
Observations	10,959
Log Likelihood	−6,900.306
Akaike Inf. Crit.	13,822.610
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01