Statistical Analysis of Central London Crime Data

Ravi Dahiya | Khairul Omar

Motivation

We wish to analyse crime data in London and investigate the claims made in the media such as low conviction rate and concerns on possible racial profiling.

Our data comes from UK Police API covering 3-year period for cases recorded within 1-mile radius of Moorgate between Dec 2016 and Nov 2019.

We proceeded to come up with our own hypothesis tests to investigate them.

Hypotheses under investigation

1. **Is there a seasonality effect between crime level and month?**
2. **Is the suspect for serious crimes less likely to be identified?**
3. **Are ethnic minorities more likely to be stopped and searched?**

Methodology

Data import and data cleansing

- Download crime data from UK Police API interface and perform data exploration & cleaning

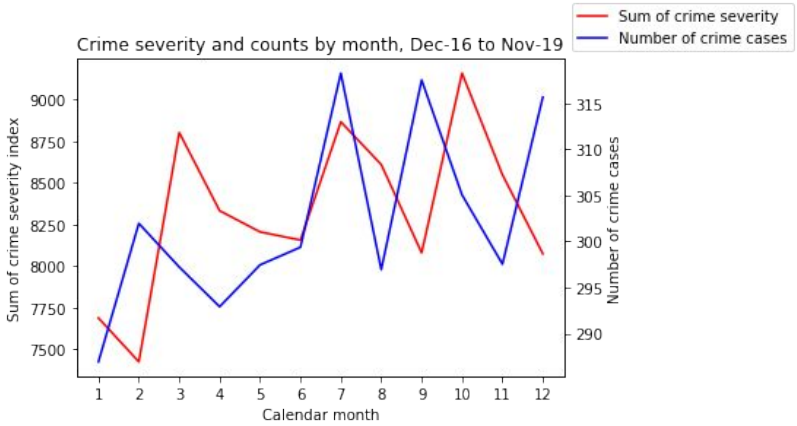
EDA & subsetting for hypothesis

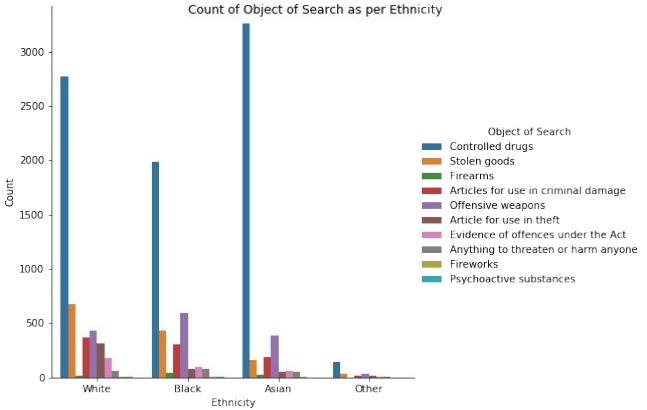
- Combine categorical information into larger groups: this was carried out for ethnicity information.
- Assign ONS indices as weights to categorical data for more meaningful analysis
- Study the variation of crime level by month and possible drivers behind them
- Study the status of crime investigations to see how many cases were not resolved.
- Study the breakdown of stop and search by ethnicity

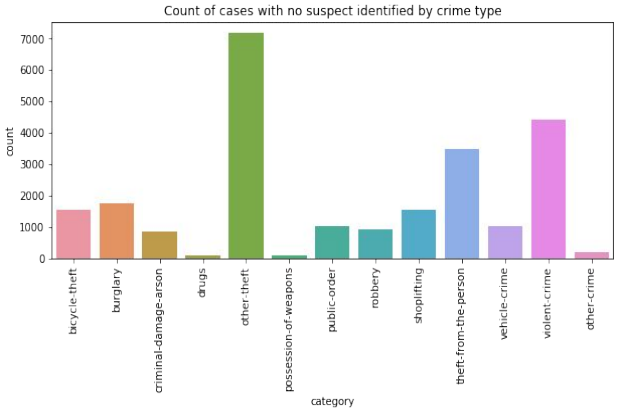
Hypothesis tests

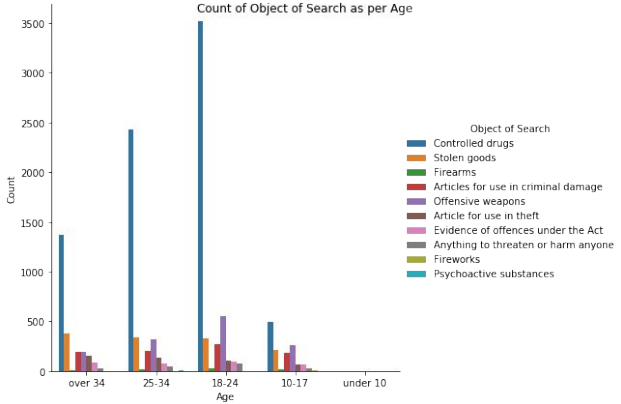
- Formulate null and alternative hypotheses to be tested.
- Select the appropriate statistical test for each hypothesis based on the availability and type of data.
- Using 95% confidence level, reject null hypothesis when p-value is lower than 0.05 for a one-side test.

Data Exploration





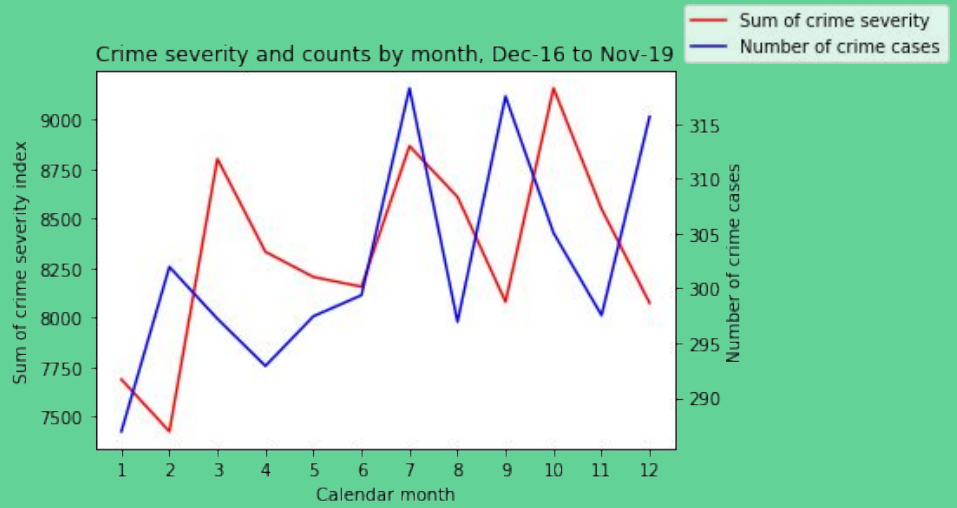




Hypothesis 1

Severity of criminal activities

- H0: There is no difference in the severity of criminal activities among different seasons
- H1: Season does affect the severity of criminal activities



Result (1)

Our study defines crime level as the total number of crimes weighted by their severity index published by Office of National Statistics (ONS).

By categorizing month into the four seasons, conducted one-way ANOVA test.

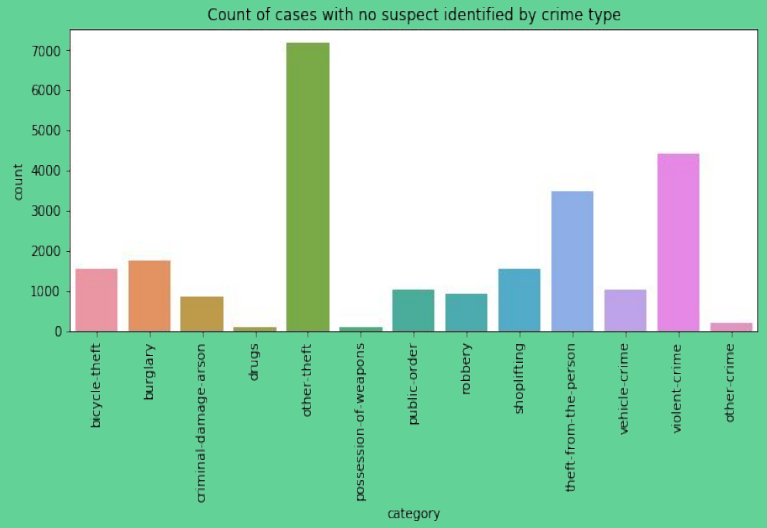
p-value result is 0.089: we fail to reject the null hypothesis.

Therefore, there is no significant statistical difference in the crime level around Moorgate in 2016-2019 between different seasons.

Hypothesis 2

Identification of suspects of violent crimes

- H0: Suspect of violent crimes are equally likely to be identified after investigation is completed as other crimes.
- H1: Suspect of violent crimes are less likely to be identified after investigation is completed.



Number of cases broken down by investigation status

	Serious crimes	Less serious crimes
Suspect not identified	7,943	16,166
Suspect identified	6,161	28,298

Result (2)

Serious crimes are defined based on ONS crime severity index of over 400, which includes violent crimes, robbery, burglary and arson.

We used **Chi-square test with equally likely outcome**.

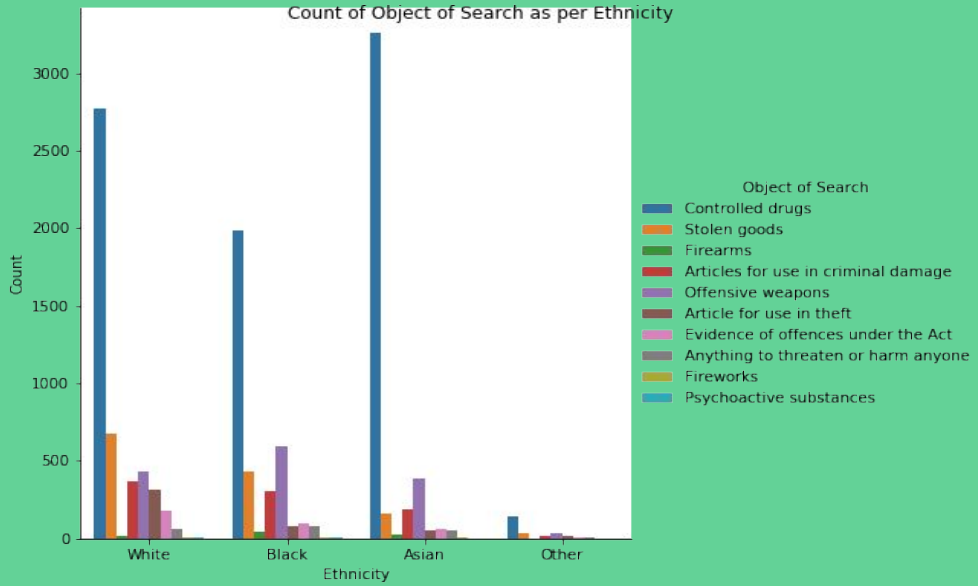
p-value result is practically zero: we can therefore reject null hypothesis.

We can conclude that it is statistically significant that the suspect for serious crimes is less likely to be identified compared to lesser crimes committed around Moorgate based on 2016-2019 data.

Hypothesis 3

Stop & search by ethnicity

- H0: Ethnic minorities are not more likely to be stopped and searched than the general population
- H1: Ethnic minorities are more likely to be stopped and searched than the general population



Stop and search cases by ethnicity

	Stop and search cases	London population share
White	37%	60%
Asian	32%	18%
Black	28%	13%
Others	3%	9%

Result (3)

If there is no bias in a perfect society, the number of stop and search made towards each ethnicity should match the overall demographic breakdown. 2011 UK Census data is used to represent the population.

We used **Chi-square test with expected values** from the census for this test.

p-value result is very close to zero, we can therefore reject null hypothesis.

We can conclude that it is statistically significant that ethnic minorities are more likely to be stopped and searched around Moorgate based on 2016-2019 data.

Q&A