

Statistical Analysis of Central London Crime Data

Ravi Dahiya | Khairul Omar

Motivation

We wanted to analysing the crime data in London and investigate the claims made in media such as there is a correlation between seasons. We proceeded to come up with our own hypothesis tests to investigate them.

Hypothesis under investigation

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

Methodology

Data import and data cleansing

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

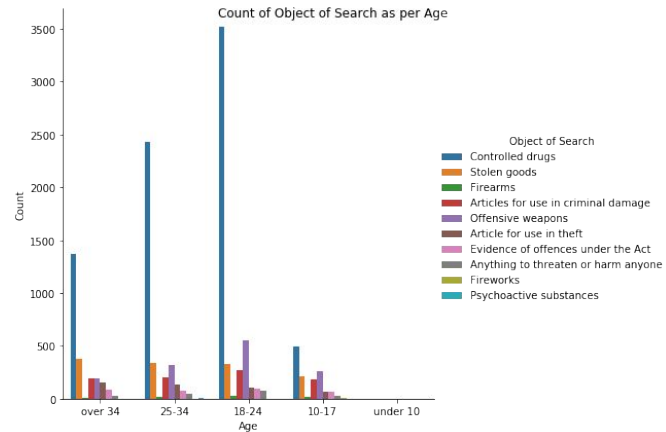
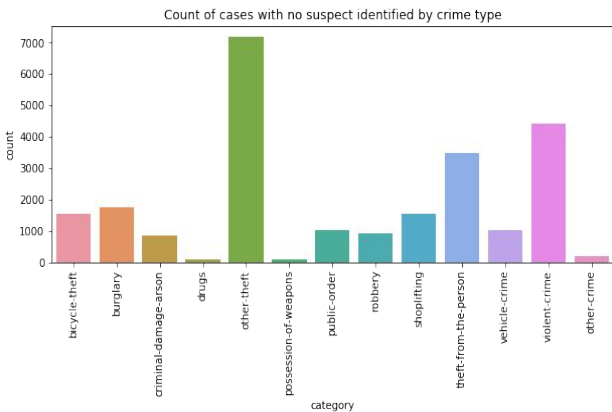
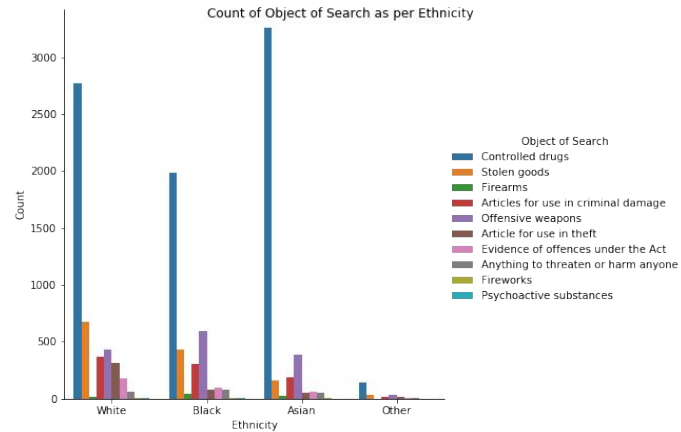
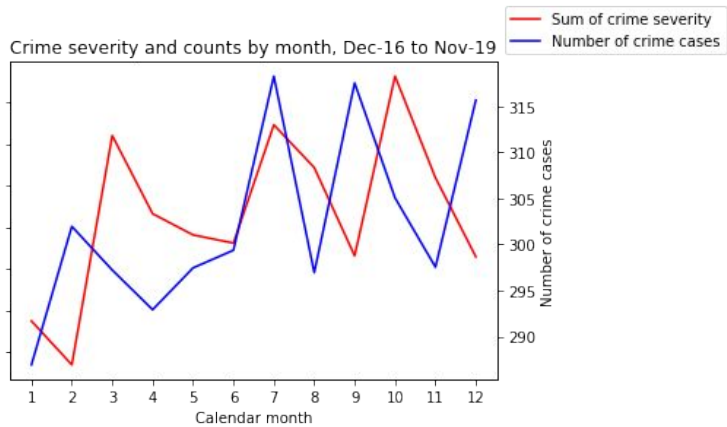
EDA & subsetting for hypothesis

- Combine categorical information into larger groups: this was carried out for ethnicity information.
- Assign weights to categorical data for more meaningful analysis: this was carried out for crime category using crime severity indices published by ONS.
- Study the variation of crime level by month and possible drivers behind them, assisted by a line plot on the number of cases and the severity of cases.
- Study the status of crime investigations to see how many cases were not resolved.
- Study the breakdown of stop and search by ethnicity via a bar plot to see if the pattern is similar to what is claimed by the media.

Hypothesis tests

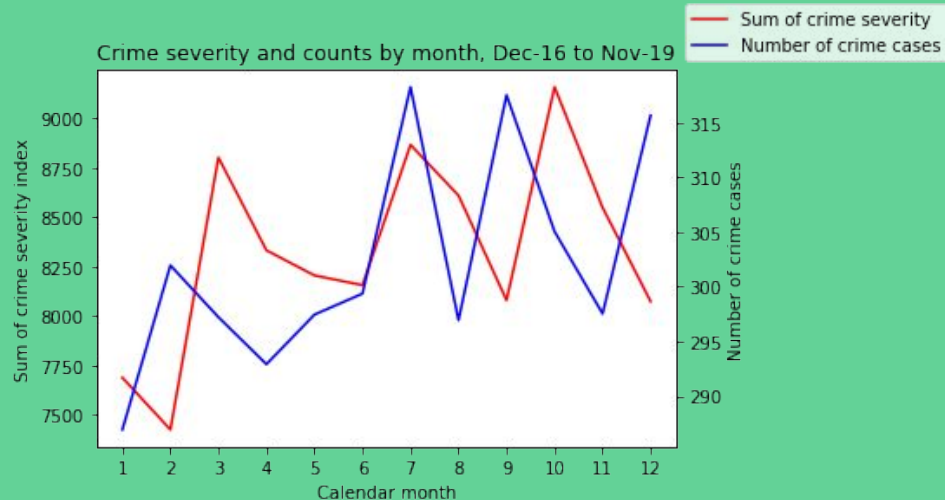
- Formulate null and alternative hypotheses to be tested.
- Select the appropriate statistical test for each hypothesis based on the availability and type of data in hand.
- Using 95% confidence level, reject null hypothesis when p-value is lower than 0.05 for a one-side test.

Data Visualisation



Hypothesis-1

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



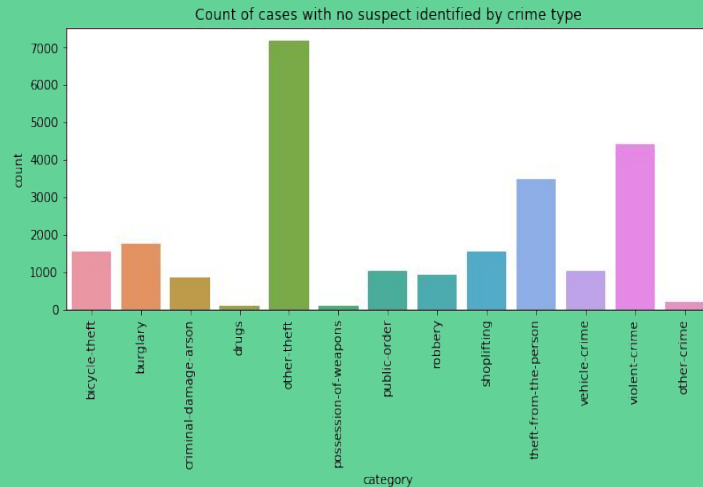
Result

Our study defined crime level as the total number of crimes weighted by their severity index. By categorizing month into the four seasons, conducted one-way ANOVA test.

p- value from the ANOVA test is 0.089, we fail to reject the null hypothesis and hence, there is no significant statistical difference in the crime activity between different seasons.¶

Hypothesis-2

- 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.



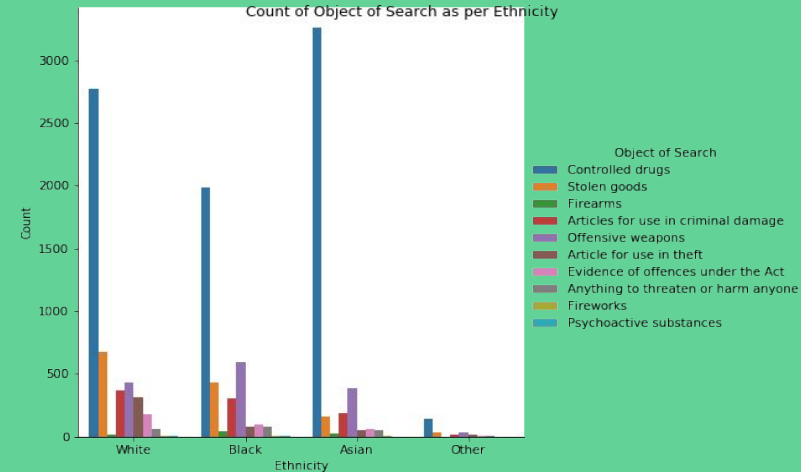
Result

Serious crimes are defined based on the severity index published ONS, which includes violent crimes, robbery, burglary and arson. We used Chi-square test with equally likely outcome.

With p-value of practically zero, 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

- 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



Result

For this hypothesis, we used Chi-square test where the expected outcome is based on UK census data for the demographic breakdown.

With p-values very close to zero, 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