The Turning Point:

Crime analysis between October 2019 – October 2020, exploring the impact of lockdown.

Mussa Yousef
MSc Data Science
School of Mathematics, Computer
Science & Engineering
City,University of London
London, United Kingdom
Email: mussa.yousef.1@city.ac.uk

Abstract — **This project investigates the impact of lockdown on crime in the heart of London. By scrutinising the statistics of crime and their respective outcome, we were able to understand the capacity of impact the lockdown has had. We partitioned policing into two perspectives, from the reactive stance where the police are being called out to crimes being committed. And the proactive viewpoint where the police are actively stop and searching individual who they suspect. This approach was the spirit of the report, we utilised feature engineering to create features which representing both aspects. Furthering our scope using correlation analysis and machine learning utilising Random Forests to investigate if the police have become harsher in their approach towards crime. The journey this report has taken us through has highlighted some insightful findings, we inspect how ethnicity, age and gender shape in the crime being committed.**

I. INTRODUCTION (*HEADING 1*)

Sirens followed by the blue and red lights is an iconic signal you're in the United Kingdom. The Metropolitan police was formed almost 200 years ago, since then constitutions, laws and regulations have changed and adapted with the time and governing bodies. Fast-forward to the current year, a year which was completely unpredictable. The unprecedented year has dismantled any strategic plans and forecasts for everyone, especially the public service sector. The Lockdown began on the 23rd of March 2020. By the end of March 2020 the value of goods and services produced in the UK also known as Gross Domestic Product (GDP), had dropped by 5.8%. Last time the GDP had dropped this severely was over the financial crisis 2008 [1]. United Nations have reported that crime rates frequently increase when an economic crisis develops[2], consequently this report will be reviewing the last year from an analytical lens scrutinising policing and crime. This report will be linking relationships and explore the impact this year will inevitably have for the future. London, one of the largest and most diverse cities in the world divided by 33 Local authorities, over 50 languages spoken and a population density of around 4,699 per sq. kilometre [3]. Crime certainly would be tailored to the borough and region however this project will be focusing within the inner London region, precisely City of London and bordering boroughs where crime would have spilled over.

II. ANALYTICAL QUESTIONS AND DATA

Government policing division releases data regularly every month, split for each district respectively. Hence, datasets were downloaded for each month from two categories of policing to enrich investigation and enable a fruitful analysis. Categories used:

- City of London Street Crime
- City of London Stop and Search

Street Crime features	Stop and Search features
Crime ID	Type (Type of search)
Month	Date
Reported by	Part of a policing operation
Falls within	policing operation
Longitude	Latitude
Latitude	Longitude
Location	Gender
LSOA code	Age Range
LSOA name	Self-defined ethnicity
Crime type	Officer-defined ethnicity
Last outcome category	Legislation
Context	Object of search
	Outcome
	Outcome linked to object of search
	Removal of more than just outer clothing

Table 1: Table of features of both datasets used

Street crime dataset was very rigid with 12 features which label information on each incident police was called out. Two features which were tailored to the report are 'Crime type' and 'Last outcome category' which described the type of crime committed and their respective outcome. Therefore, we would be able to match and examine the correlation between the crime committed and the outcome. Additionally, inferential statistic was explored, breaking down calculations into segments which improved our understanding of how the police respond to each type of crime. Using this methodology aligns with the first analytical question:

1. Looking at the crimes before the lockdown how have the statistics of crime committed, and outcomes of crimes been impacted?

This is a rather broad analytical question allowing a few assumptions to be made before analysis, which are, crime rates are seasonal [4] and the outcomes which have been included in the dataset are the only possible outcomes. The search dataset links to the second analytical question:

2. As a result of the lockdown period have the police become more lenient/harsher/ or effectively the same?

The stop and search datasets are broader in terms of dimensionality, with 18 features that included a description of each individual stopped over the year. The crime dataset describes crime from a responsive view, where the police were reactive to a call from a civilian. However, search dataset portrays the proactive aspect of how the police handled crime, therefore using both datasets would provide a wide scope for this investigation.

III. ANALYSIS

To be able to adequately explore the analytical questions raised in-depth, the following plan was devised:

1. Import crime & search datasets for each month and merge all 12 months.
2. Examine and cleanse both datasets respectively by removing invalid values or choosing the best way to impute data.

3. Initial characteristic analysis and statistical analysis on both data sets
4. Feature Engineering: Combining variables and converting categorical dataset into features for further investigations
5. Explore new features visually and investigate correlation analysis
6. Use geospatial to view, describe and analyse data
7. Using Machine learning to evaluate how policing and crime have been affected

As we found each dataset split into monthly reports, we begin by concatenating using outer join, keeping each month's information unchanged. Wrangling was then carried out, inspecting the information missing from the datasets. We found most were longitude and latitude coordinates missing which accumulated to 13% and 23% respectively for both stop and search and crime data, therefore considering this element is definitive we cannot use generalised statistics to find a mean/median to impute the missing data. Therefore, we decided to delete the rows which had omitted information on location retaining the rest.

Crime type	Counter	%scale c	%Cumulative percentage to tot
Anti-social behaviour	28298	0.268296	0.268296
Other theft	16840	0.159662	0.427958
Violence and sexual offences	13473	0.127739	0.555697
Theft from the person	11874	0.112579	0.668275
Drugs	4968	0.047102	0.715377
Vehicle crime	4879	0.046258	0.761636
Bicycle theft	4622	0.043822	0.805457
Burglary	4508	0.042741	0.848198
Public order	4175	0.039584	0.887782
Shoplifting	3951	0.03746	0.925242
Robbery	3692	0.035004	0.960246
Criminal damage and arson	3078	0.029183	0.989429
Other crime	645	0.006115	0.995544
Possession of weapons	470	0.004456	1

Table 2. Table representing the types of crimes committed and their respective frequency

Supercategory	Subcategories repectively merged into supercategory
Arrested	[Under investigation], [awaiting court] , [court result unavailable]
Domestic	[Formal action not in public interest], [local resolution], [Action to be taken by another organisation], [Further investigation is not in the public interest]
Warning given	[Status update unavailable], [Offender given penalty notice],[Offender given a caution],[Offender given a drugs possession warning]
No Further Action	[Investigation complete; no suspect identified], [unable to prosecute suspect]
Theft	[Other theft], [Bicycle theft], [Shoplifting]
Robbery	[Robbery], [Burglary]
Social_Infringement	[Anti-social behaviour],[Public order]
Drug,Weapon and other	[Drugs], [Other crime], [Possession of weapons]

Table 3. Table expressing the feature engineering process and merge of information to create more expressive features

When Boris Johnson made the decision to bring London to a lockdown, several temporary restrictions were placed on citizens. Restrictions was of three categories: Gathering, Movement and Business [10]. The Police were given right to enforce, by stopping citizens who have left their homes without a "reasonable excuse", citizens found guilty had breached anti-social behavior [11] which explains the accumulated 26% shown in (Table 2). To be able to completely understand both datasets we split our investigation into three traits: Understanding of the different type of crimes committed, the relationship between the crimes and outcomes, finally overview of both. We began by grouping crimes committed and outcomes of crimes into super-categories (Table 2). We find that the unusual number of anti-social behavior had a correlation to the number of warning and fines issued by the police (Fig 1).

Previous reports [7 Malleson 2016] considers the borough being a low residential and high ambient area where theft and robbery were 95% more likely to occur during weekdays compared to weekend. Dissecting the highlighted lockdown period in (Fig 1) we find a dramatic reduction in the number of theft and robbery, suggesting lockdown had significantly reduced these, in parallel we found a sharp decrease in no further actions, which suggests the police may have been under resourced to conclude cases. However, arrests had been steadily increasing prior to lockdown, while during and after lockdown we found this uprising abruptly. This suggests the police may have been pursuing prosecution more during lockdown.

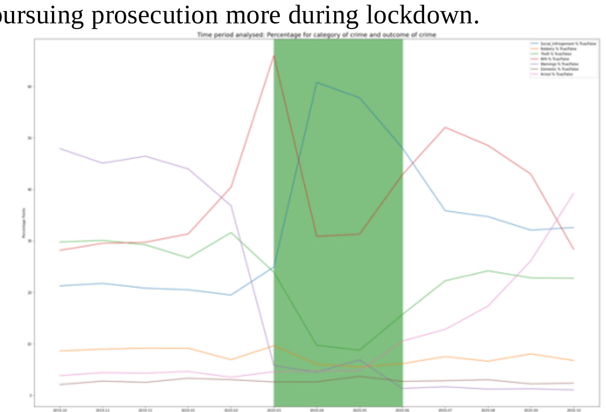

Fig 1. Line plot representing the types of crimes, outcomes and their respective frequency. Highlighted in green: - Lockdown period

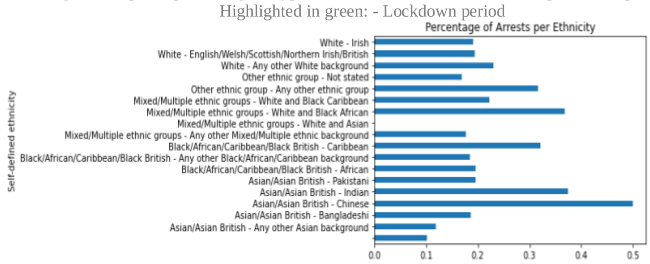

Fig 2. Bar chart showing the percentage of Arrests per self-defined Ethnicity

Stop and search data indicates the lowest records during the lockdown period (fig.3) at an average of 7 stops per day, followed by an upsurge to 30 stop and searches recorded during the following months after lockdown. Calculating the distribution of outcomes for stop and searches we find that 19% of all stop results in an arrest, when a breakdown of those being stopped using self-defined ethnicity (fig.2). we uncover that ethnic minorities, Asian, Black or mixed are almost twice more likely to be stopped, while they comprise of 31% of the data.

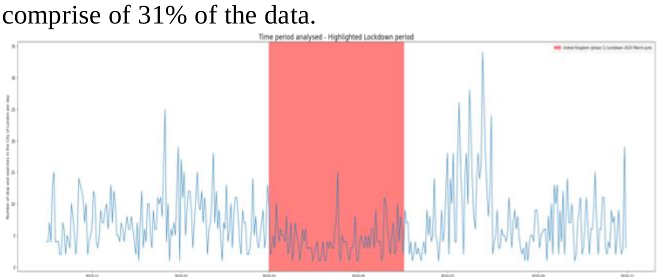

Fig 3. Continuous line graph representing Stop and Search records

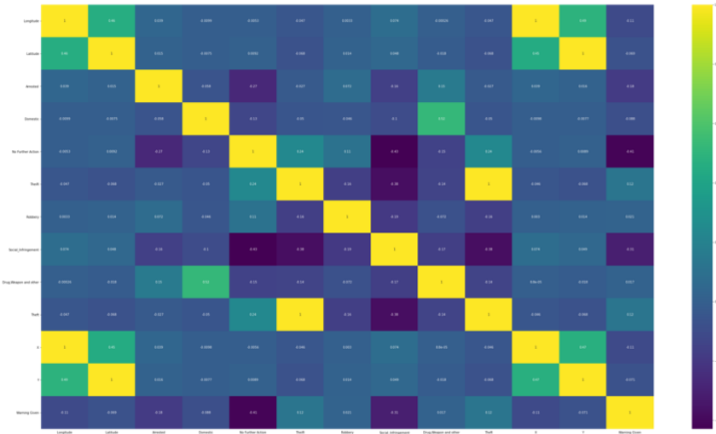

Fig 4. Heatmap illustrating the Correlation matrix of all features in the crime dataset

Establishing two perspectives, the proactive and reactive strategies of policing. We used correlation analysis (fig 4) to evaluate the relationship all categories have with the outcome of an arrest. Deploying the Pearson correlation analysis [9] where all relations are labelled over a spectrum between -1 to 1, corresponding negative and positive correlation. We find an individual is likely to be arrested if a drug, robbery or theft crime had been committed. However, when the police are called out or in pursuit of an anti-social behaviour, fireworks or criminal damage crime, we find these outcomes to have a negative relationship with arrests, which suggest the police are likely to conclude the crime with a 'no further action', warning or fine.

Furthermore, the matrix analysis really narrows the description to identify tags which increase the probability that a citizen stopped would be arrested. We find a positive correlation for the features: Male, Over 25, Black, White, Theft, Robbery, and Removal of outer clothing. Observing the geospatial image of the arrests of both datasets (Fig5) we find that although most arrests are concentrated within the City of London region, there are two distinct traits of arrests, there is a huge cluster of crime between the region of Liverpool street and Aldgate, which [7 Malleson 2016] had similar findings withing the ambient area, and the majority of arrests by the City of London Police outside the jurisdiction of the had been from being stopped.

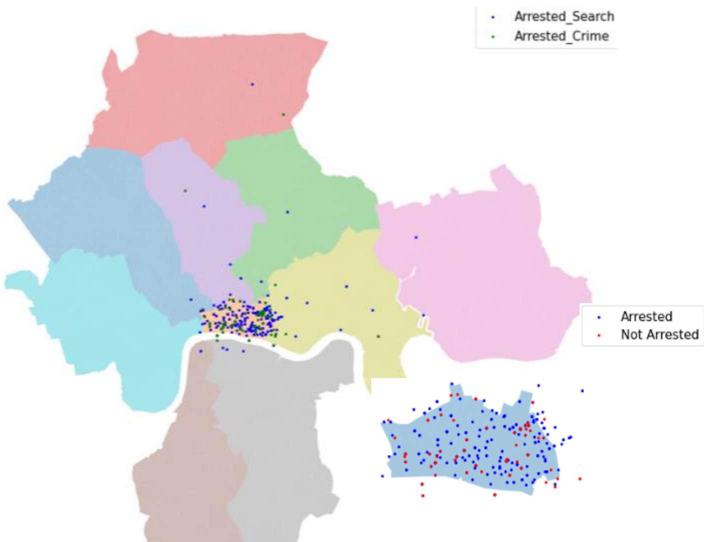

Fig 5. Geospatial image of inner London portraying the Arrests made from stop and search and called out crime

To be able to assess how lockdown had affected the attitude of the police towards crime, we exploited the use of Machine Learning (ML). We aimed to build a model that would assess the variables which influence arrests, thus using the arrest class as our target variable. This approach was inspired by previous work [12] where ML was used for crime analysis in Vancouver. The ensemble Decision Tree algorithm, Random Forest was used by splitting the data into train and test, where we trained the model using the months between October 2019 – July 2020, therefore testing the model on the data between August 2020 – Oct 2020. To Choose the number of trees in our model we used a baseline which was the number of features squared.

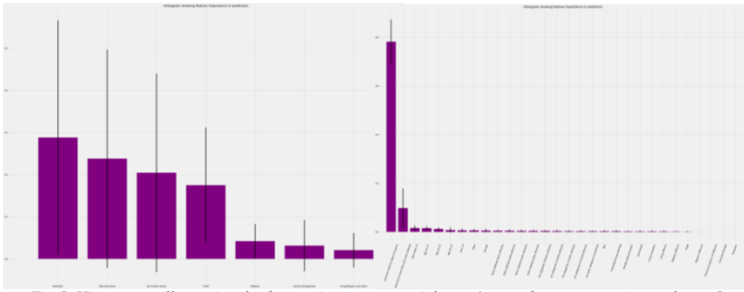

Fig 6. Histograms illustrating the feature importance weight on Arrests from street crime and stop & search dataset

Fig 7. Heatmap illustrating the confusion matrix of classifying polices on both street crime and stop and search respectively

Crime Dataset				
	Precision	Recall	F1-score	Support
Negative class	0.97	1	0.99	12681
Positive class	1	0.91	0.95	4239
Accuracy			0.98	16920
Macro avg	0.99	0.95	0.97	16920
Weighted avg	0.98	0.98	0.98	16920
Stop and search Dataset Predicting Arrests				
	Precision	Recall	F1-score	Support
Negative class	1	0.98	0.99	786
Positive class	0.94	1	0.97	187
accuracy			0.99	973
macro avg	0.97	0.99	0.98	973
weighted avg	0.99	0.99	0.99	973

Table 4. Classification report on both datasets assessing performance

The model analysis found that the features influencing both models were both negatively correlated with arrests. Although accuracy for both models was at 98% and 99% respectively, scrutinising both models we found a discrepancy with recall at 91% for crime and precision at 94% for stop and search. Therefore, the police were more prone to arrest individuals for crimes they wouldn't have before and during the lockdown. However, contrasting results were shown for stop and search as the police, decided to resolve matters without arresting individuals which during lockdown they may have arrested.

IV. Findings, Reflections and further work

To summarise, before lockdown Theft and Robbery was the highest crimes committed in the City of London by those who have self-defined as White between 18-24, while most arrests were of those Asians and black for Weapon and Drug offences. However, during lockdown, we found that the ethnicity of those being arrested were more evenly distributed. The police also saw a change in the type of crimes, which therefore saw adjustment in the way policing was carried out. We were able to create a random forest model which was able to predict arrests. We found that the police had made overall more arrests, for same crimes previously a fine or warning would had been issued, concluding the police had become slightly harsher after lockdown.

The data itself was very limited, there wasn't any significant numerical information in both datasets to allow deep correlation analysis. However, we were able to derive some relationship analysis utilising feature engineering, which enabled seeding of statistic which helped us quantify impact of lockdown on crime. However not enough to make ground-breaking discovery.

	No Further Action	Arrests	Warning	PCN	Court	Caution	Total	Court_per_stop	Arrest_Rate
Self-defined ethnicity									
N/A	95	11	2	0	0	0	108	0	0.101852
Asian/Asian British - Any other Asian background	68	10	2	0	2	0	82	0.02439	0.121951
	2	0	0	0	0	0	2	0	0
Asian/Asian British - Bangladeshi	87	24	4	1	2	2	120	0.016667	0.2
	13	1	0	0	0	0	14	0	0.071429
Asian/Asian British - Chinese	0	2	2	0	0	0	4	0	0.5
Asian/Asian British - Indian	14	8	1	0	0	0	23	0	0.347826
	0	1	0	0	0	0	1	0	1
Asian/Asian British - Pakistani	33	10	6	1	0	0	50	0	0.2
	1	0	0	0	0	0	1	0	0
Black/African/Caribbean/Black British - African	116	31	8	1	0	0	156	0	0.198718
	5	1	0	1	0	0	7	0	0.142857
Black/African/Caribbean/Black British - Any other Black/African/Caribbean	30	7	1	0	0	0	38	0	0.184211
Black/African/Caribbean/Black British - Caribbean	49	24	6	0	0	0	79	0	0.303797
	0	2	0	0	0	0	2	0	1
Mixed/Multiple ethnic groups - Any other Mixed/Multiple ethnic background	19	5	8	0	0	0	32	0	0.15625
	1	1	0	0	0	0	2	0	0.5
Mixed/Multiple ethnic groups - White and Asian	7	0	1	0	0	0	8	0	0
Mixed/Multiple ethnic groups - White and Black African	8	7	4	0	0	0	19	0	0.368421
Mixed/Multiple ethnic groups - White and Black Caribbean	11	4	2	0	0	0	17	0	0.235294
	1	0	0	0	0	0	1	0	0
Other ethnic group - Any other ethnic group	23	11	2	1	0	0	37	0	0.297297
	0	1	0	0	0	0	1	0	1
Other ethnic group - Not stated	678	139	24	2	3	3	852	0.003521	0.163146
	21	10	0	0	0	0	31	0	0.322581
White - Any other White background	0	0	0	0	0	0	1	0	0
	257	89	34	1	1	0	384	0.002604	0.231771
	7	2	0	0	0	0	9	0	0.222222
White - English/Welsh/Scottish/Northern Irish/British	344	92	36	3	2	2	483	0.004141	0.190476
	8	4	0	1	0	0	13	0	0.307692
White - Irish	17	4	0	0	0	0	21	0	0.190476

Table 4: Breakdown of the ethnicity of those stopped and searched

Furthermore, both datasets had complete contrasting structures, the crime dataset was rather large in rows however limited in features, whereas the stop and search dataset had great insight in attributes however limited in number of rows. Moreover, scrutinising the City of London, the borough is vastly different to the other boroughs. Therefore, in the future we would incorporate census data to be able to achieve a fuller analysis, breaking down crime and tailoring race and gender to demography and social class (Table 4). We would therefore be able to deploy clustering, utilising weighted statistics which would link features geographically and further policing practical strategies

We aimed to scrutinise the affect lockdown has had, therefore including the recent census which was in 2011 wouldn't have been relevant to an extent. A practical recommendation to the police would be to include further details on crime, such as age, employment and even information such as citizenship status, all of which could be collected anonymously. This would allow the policing force to be able to deploy resources efficiently, and furthering proactive strategies in minimising crime by understanding why and who are most likely to commit the type of crime.

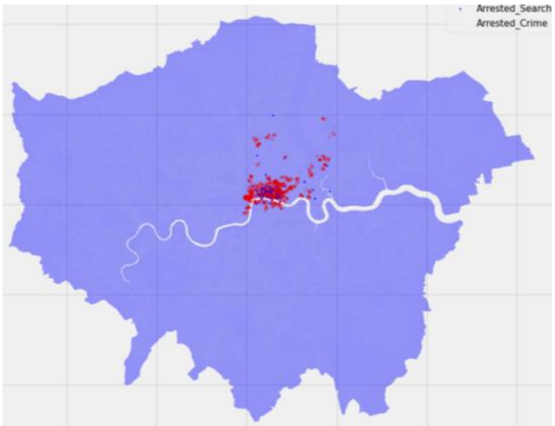

Fig 8. Geospatial image of London and arrests during the year

In retrospect, although we were able to use geospatial visualisation tools to examine the distribution of crime in the City of London, to be able to compare between boroughs as a whole is crucial in concluding exploration. Therefore, in the future we would extend our analysis to the entire London area (Fig 8), or various boroughs around London, then inspect narrow into the City of London. This would substantiate findings and add dimensionality to our research.

Ultimately, the lockdown has had several impacts on citizens of the United Kingdom all aspects. Inevitably crime has been disturbed while policing is rigorously tailoring their approach to criminals. We find the lockdown has provided several lessons for the police of city of London particularly in the attitude towards certain crimes. While our datasets used in this project had its limitations, this project has fueled further investigation. We find ourselves entering lockdown again in the United Kingdom, it's evident how detrimental the first lockdown has been, for future work we would be able to have a full painting of the year. Therefore, With the introduction of the census 2021, crime would be able to be investigated in respect to the demographical construct of the respective boroughs.

References

[1] Office for National Statistics – *"Coronavirus and the impact on output in the UK economy: March 2020"* - https://www.ons.gov.uk/economy/grossdomesticproductgdp/articles/coronavirusandtheimpactonoutputintheukeconomy/march2020

[2] United Nations- *"Economic crises may trigger rise in crime"* - http://www.unodc.org/unodc/en/frontpage/2012/February/economic-crises-can-trigger-rise-in-crime.html

[3] Project Britain – *"Learn about London": by Mandy Barrows* - https://www.projectbritain.com/london/facts.htm

[4] *Data Police UK – Data downloads*- https://data.police.uk/data/

[5] Statista – *"Crime rate per 1,000 population in the United Kingdom from 2002/03 to 2019/20, by country"* - https://www.statista.com/statistics/1030625/crime-rate-uk/

[6] World Population review – London Poputlation 2020 https://worldpopulationreview.com/world-cities/london-population

[7] Nick Malleson, Martin A. Andresen,

[Malleson 2016]

Exploring the impact of ambient population measures on London crime hotspots,
Journal of Criminal Justice,
Volume 46,
2016,
Pages 52-63,
ISSN 0047-2352,
https://doi.org/10.1016/j.jcrimjus.2016.03.002.
(http://www.sciencedirect.com/science/article/pii/S0047235216300198)

[8] Full Fact – *"Stop and search in England and wales"* - https://fullfact.org/crime/stop-and-search-england-and-wales/

[9] Question Pro – *"Pearson correlation coefficient: Intoduction, formula,calculation , and examples* - https://www.questionpro.com/blog/pearson-correlation-coefficient/

[10]House of Commons Library- *"Coronavirus: the lockdown laws"* - https://commonslibrary.parliament.uk/research-briefings/cbp-8875/

[11] The Converstation – *'Lockdown crime trends: why antisocial behaviour is up'* https://theconversation.com/lockdown-crime-trends-why-antisocial-behaviour-is-up-140479#:~:text=It%20is%20plausible%20that%20the%20lockdown%20resulted%20in,picture%20is%20not%20so%20clear%20for%20criminal%20offences.

[12] *S. Kim, P. Joshi, P. S. Kalsi and P. Taheri, "Crime Analysis Through Machine Learning,"* 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), *Vancouver, BC, 2018, pp. 415-420, doi: 10.1109/IEMCON.2018.8614828.*

Word Count

Abstract: **141/150 words**

Introduction: **237/300 words**

Analytical questions and data: **298/300**

Analysis: **1000/1000**

Findings, reflections and further work: **575/600**

Total: **2251/2350 words**