BAYES
BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON

Analysis of the risk exposure of police forces across England and Wales

This report is submitted as part of the requirements for the award of the

Master of Science in Business Analytics

in Cass Business School, City University of London

YEAR OF SUBMISSION
September 2021

Thi Hong Ngoc Nguyen | ID: 20052999
Academic supervisor: Dr. Rosalba Radice

Abstract

The police profession is considered a high-risk occupation (Edwards and Kotera, 2020). Police officers can be suffered both physically and mentally. This paper will review the potential risks from previous research and paper to police officers and compare risk levels between 43 territorial police forces to analyse risk exposure by regions. All the risk factors discussed in the paper are available on public resources. The K-mean method with Euclidean distance is chosen to classify police forces into similar risk level groups. The provision of insurance is used to evaluate the adequacy of risk cover in police forces.

In summary, 43 police forces are divided into three groups: high-risk, moderate-risk, and low-risk group. City of London force is found creating its own cluster. Furthermore, some police forces are considered as having relatively low insurance provision per officer rate, which may not cover the risk that police officers are facing.

Acknowledge

I would like to formally thank my academic supervisor, Dr Rosalba Radice, and industrial supervisors for their guidance and support throughout the duration of the project. I would also like to thank my teammates for their contribution to data collection and analysis.

Preliminary pages

Table of Contents

List of tables

List of figures

Introduction

In England and Wales, policing structure includes 43 territorial police forces, the Ministry of Defence Police, the Civil Nuclear Police, and the British Transport Police (Brown, 2021). Apart from London, in each geographic police force, the Police and Crime Commissioner has responsibility for whole policing: serves to deliver policing services to the members of the public, hold the police to account, and set the force's direction (Association of Police and Crime Commissioner, 2021). In London, the Mayor's Office for Policing and Crime (MOPAC) is responsible for ensuring the governance of the Metropolitan Police Service, and the City of London Corporation has the responsibility of overseeing the City of London Police (Association of Police and Crime Commissioner, 2021).

According to the Association of Police and Crime Commissioners (2020), "police officer" is a term referring to the total officers within the police force, including from the most junior to the Chief Constable. Their responsibilities are protecting all public and their properties, preventing crime and alleviating the fear of crime, and enhancing the quality of life for civilians in order to maintain law and order in the community (Police Foundation, 1996).

Due to the nature of the career, the police occupation is considered a high-risk profession (Edwards and Kotera, 2020). They may be suffered mentally and physically, which is recorded as "assault with injury" and "assault without injury" in police recorded crime statistics series. To deal with the danger from the nature of career, each Police and Crime Commissioner self-insures by setting insurance provisions. This insurance provision includes all types of liabilities, such as public and motor insurance. It is reviewed and reset yearly to cover any claim from that year and the previous years. The insurance provision reflects the level of insurance claims in each police force, which indicates the risk that police officers have to deal with.

This paper will analyse the relevant variables collected from public resource to answer the questions: What do risk factors indicate the current risk situation across England and Wales? How to classify forces' risk levels by regions? And are insurable risks well cover?

Literature review

Because of the virtue of the authority that society expects on police officers, the police profession is responsible for all encounters with civilians, which lead to the confrontation with danger such as criminals, people with mental impairment or under the influence of drugs and alcohol (the Police Foundation and the Policy Studies Institute, 1996). In the year to March 2019, on average, two out of five police officers on duty were found being assaulted (Dodd, 2020).

In the Crime Survey for England and Wales findings, police occupation listed under the protective service professions is found the most high-risk profession in 2014/15 (Worksmart, 2021). Furthermore, this survey also shows that the risk rates of the police service and security guards professions are higher than the average ones over the last number of years (Worksmart, 2021). Several researchers also listed police as a stressful and hazardous career (Edwards and Kotera, 2020). The result of a stressful working environment associated with the high-risk nature of the career may put police officers at greater risk of evolving psychological issues and illnesses (Habersaat et al, 2015). From 2012 to 2017, the rate of police officers absent from work due to mental health-related illnesses significantly increased up to 47% (Police Firearms Officers Association, 2017). In 2016, a study implemented on 17000 police officers by the Police Federation revealed that mental well-being among officers was significantly poorer than the general population, especially, 39% of participants need support for mental illnesses (Houdmount and Elliott-Davies, 2016). Moreover, a study conducted with 1531 police officers in the United Kingdom reveals that one in five officers is suffering from post-traumatic or complex post-traumatic stress disorder (Miller et al., 2021).

Since 2010, 19% of police funding was cut off, which caused the loss of 45000 police officers and staff in England (Dodd, 2019). Under the significant strain of the budget, some crimes were considered a “failure of police’s response” due to the change of police’s priorities (Dearden, 2018). Consequently, some crime was recorded as considerably increasing, which indicated the higher risk for police officers (Institute for Government, 2019).

Methodology

Data sources

All the data used in this paper is collected from public resources. The data includes crime data, the number of police officers by the police force, the number of assaults on police officers, and the provision of insurance.

- Crime data in this paper are collected from police recorded crime statistics. Police recorded crime data may give a helpful insight into the type of crime that reflects police priorities and activities (Office for National Statistics, 2021). This can explain the different structures of statistics and data between police recorded crime and Crime Survey for England and Wales. For instance, police record crime statistics do not contain "non-notifiable" offences such as parking offences but provide good detail about homicide (Crime in England and Wales, 2021). Although police recorded crime is affected by the change in recording practices, data used in this report is consistent in recording practices as no change happens from 2015 to 2021. Police recorded crime data are summarised from the report of 43 territorial police forces; therefore, crime recorded is published in detail of the number of crimes in each force, which help break down the crime level across England and Wales. With the consideration of the aims of this paper, police recorded crime data is chosen over the Crime Survey for England and Wales.
- The number of police officers by police force areas in this paper has used the statistics in September from 2017 to 2020 from National Statistics.
- The number of assaults on police officers is collected from the Office for National Statistics source in police recorded crime series. Since the change on the crime classification for "assault with injury on a constable" was introduced in 2017, the following statistics published are not compatible with previous years. Therefore, the data used in this paper is only collected from 2017 to 2020.
- Insurance provision is defined as the money set aside to cover the police force's own risk (e.g. employee liability, motor vehicle liability, and public liability) when an obligation to pay is recognised with an uncertain amount and time. Insurance provisions in this paper are collected from public websites, mainly the police force's web pages.

Only 27 police forces are mentioned in this paper as the information cannot be found from public resources.

Clustering

Variables and Method

With the purpose of classifying police forces based on risk levels, 16 variables which are may indicate the risk levels are used in this paper: ratio of crime per police officer from 2015/16 to 2019/20 (5 variables), the ratio of officers per mile from 2015/16 to 2019/20 (5 variables), rate of assaults with and without injuries per officer from 2017/18 to 2019/20 (6 variables). Due to the wide range of mean and variance between features, the dataset is scaled to avoid the dominance of any variables. Besides, it is noted that three missing values in terms of assaults in Greater Manchester in 2019/20 are filled by the mean value.

The aim of the research is to cluster all the police forces, no observation is considered as outliers and removed from the data. There is no minimum number of observations in a cluster, therefore, one observation can get its own cluster. Euclidean distance is used because it is graphically straightforward and has well-constrained computational costs. The purpose of this clustering is to focus on similar risk levels between police forces, therefore, K-means is used because it emphasises homogeneity rather than the separation between groups (Hennig, 2015). Besides, as discussed before, all 43 observations are used, which leads to unavoidable noise. K-means is able to outperform hierarchical clustering on datasets that contain outliers and noises (Punj & Stewart, 1983). Furthermore, the dataset in this paper is not considered a high-dimensional dataset (43 observations and 17 variables). Therefore, it is unlikely that curse of dimensionality, which caused the negative impact on clustering, will occur (Google Developers, 2021).

Choose the number of clusters

In order to choose the best number of clusters *(k)*, the Elbow method and seven validation measures are performed by *clValid* function in the clValid package: the average proportion of non-overlap (APN), the average distance (AD), the average distance between mean (ADM), the figure of mean (FOM) for stability measure, the connectivity, Silhouette Width, and Dunn index for the internal measure.

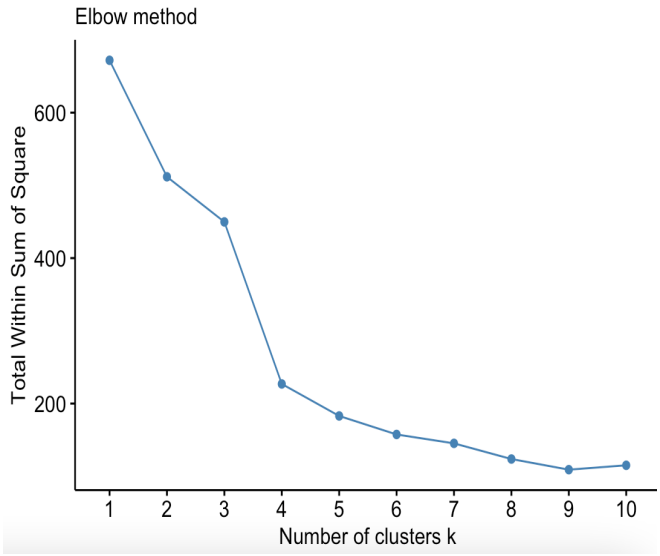

Figure 1: Elbow method

Measurement	Score	The optimal number of cluster (k)
APN	0.0562	3
AD	2.0322	10
ADM	0.2544	3
FOM	0.4762	10
Connectivity	19.9516	3
Dunn	0.3520	10
Silhouette	0.2618	3

Table 1: Result of validation measures (clValid function)

From Figure 1, the elbow point is 3 because, after that value, improvement in distortion shows a significant decrease. Besides, 4 out of 7 validation tests also recommend k equals 3 is the most optimal value (see Table 1). Hence, the number of clusters is chosen as 3.

Data Analysis

The number of police officers

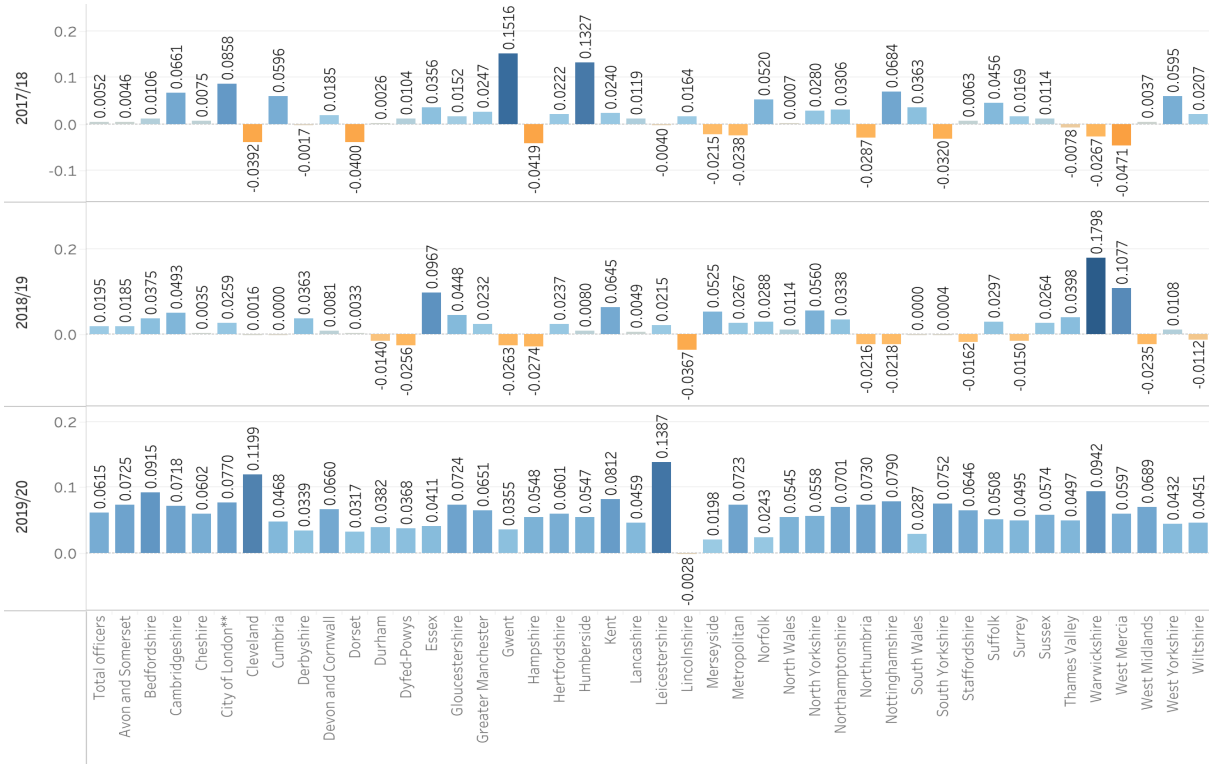

Figure 2: The change of number police officers compared to the previous year

The number of police officers in most police forces changed less than 10% in both 2017/18 and 2018/19, then raised around 6% in the year 2019/20. From September 2016 to 2020, there was an increase in the total number of police officers. Especially in 2019/20, all forces, excepting Lincolnshire, experienced an increase in the number of officers (see Figure 2). Only a few police forces witnessed a decrease in police officers in 2017/18 and 2018/19. Hampshire, Lincolnshire, and Northumbria police are the only forces experiencing the decrease in police offices two successive years.

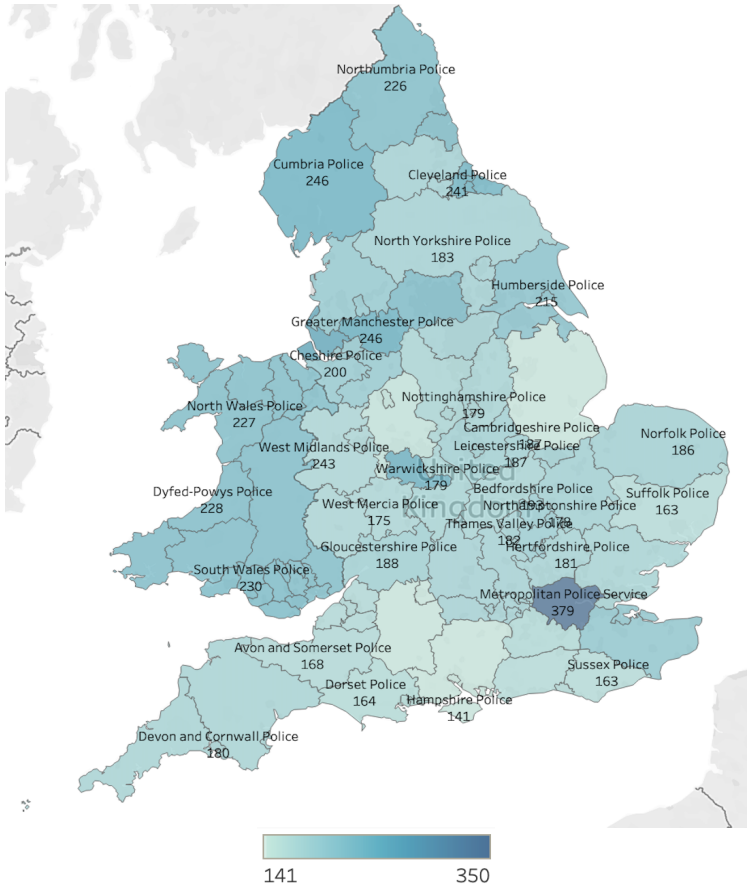

Figure 3: Number of Police Officers over 100,000 head of Population, September 2020
Note: City of London Police and Metropolitan Police Service have been merged

Based on data collected from Police Service Strength (Allen and Zayed, 2021), it is found that:

- The highest ratio of police officers per 100,000 head of civilians is 379 in the Metropolitan Police force. In fact, London has a dense population and has to face a higher risk of crime, which requires a high number of police officers.
- Generally, Figure 3 shows that major population areas in England and Wales have a higher ratio of police officers over 100,000 head of population, such as Merseyside and Metropolitan police forces.
- The majority of police forces in Wales and North England have a higher ratio of police officers to civilians than in the rest of the forces, except for Metropolitan.

Police assault

As mentioned above, police officers have to confront several risks, which can lead to assaults with or without injuries. In fact, on average, in every 20 minutes, there is a police officer assaulted (The Times, 2018). Besides, two out of five police officers were attacked on duty in the year to March 2019 (The Guardians, 2020). Overall, these may imply the frequent occurrence of violence against police officers.

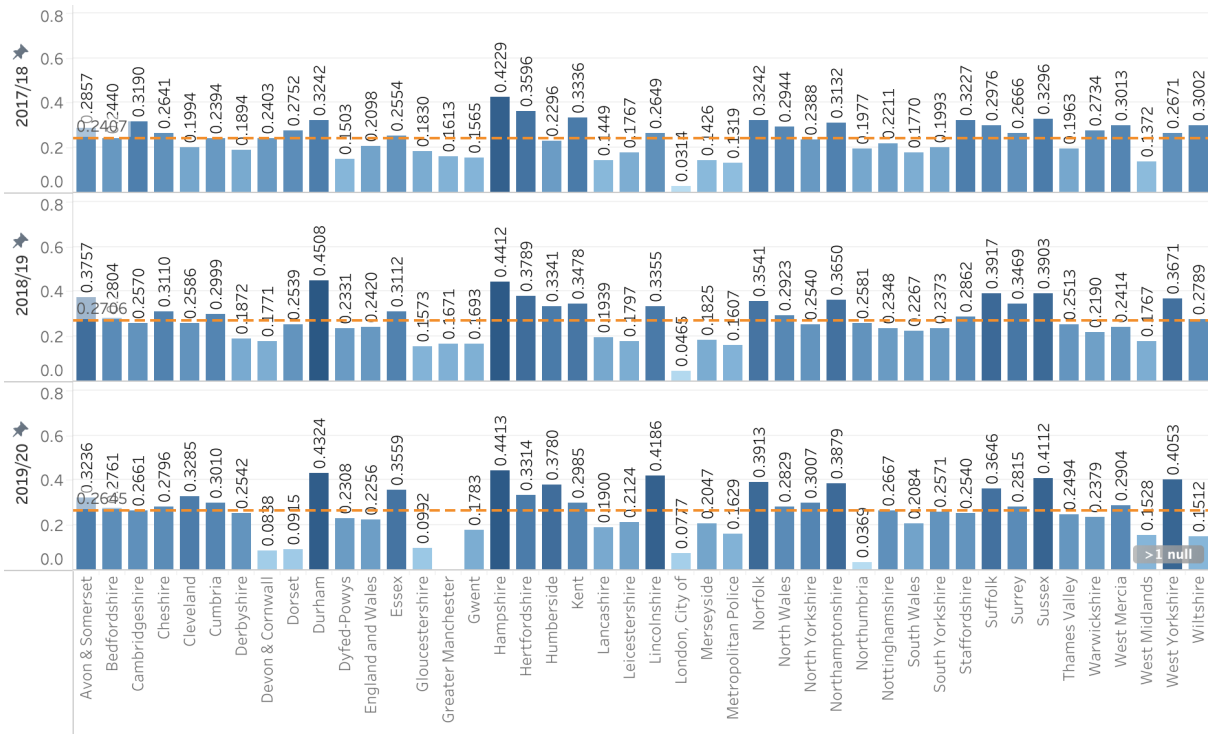

Figure 4: Rate of total assaults recorded by police over police officers

(Note: figure of Greater Manchester in 2019/20 is 0 due to missing data)

The average rate of total assaults over police officers increased by 12,4% from 2017/18 to 2018/19, then slightly decreased by 2.25% in 2019/20. The decline of the average rate may be explained by the impact of COVID-19 when people experienced quarantine and social distancing.

Apart from Devon and Cornwall, Wiltshire, Dorset, and Gloucestershire, where total assault rates are recorded as constant decrease over three years, other forces have witnessed an increase or fluctuation in the development of this rate.

Looking closely at figure 4, Durham, Hampshire, Hertfordshire, Norfolk, Northamptonshire, Suffolk, and Sussex police forces consistently rank the top group having the highest risk of

being assaulted. In 2017, the rate of total assaults per police officer in Hampshire was the highest rate and the only one over 0.36. However, in the following years, the number of police forces having this rate over 0.36 increased, up to 8 and 10 forces in 2019/20 and 2019/20 respectively. The increase of police forces having high assault rates, plus the decrease of the average assault rate across England and Wales, may imply the increase and the expansion of the risk of being assaulted in some particular areas. For instance, in 2017/18, West Yorkshire and Essex police forces had a moderate rate of being assaulted which is similar to the average rate of being assaulted across England and Wales (0.2406). However, in 2019/20, both rates in two police forces significantly raised to over 0.36.

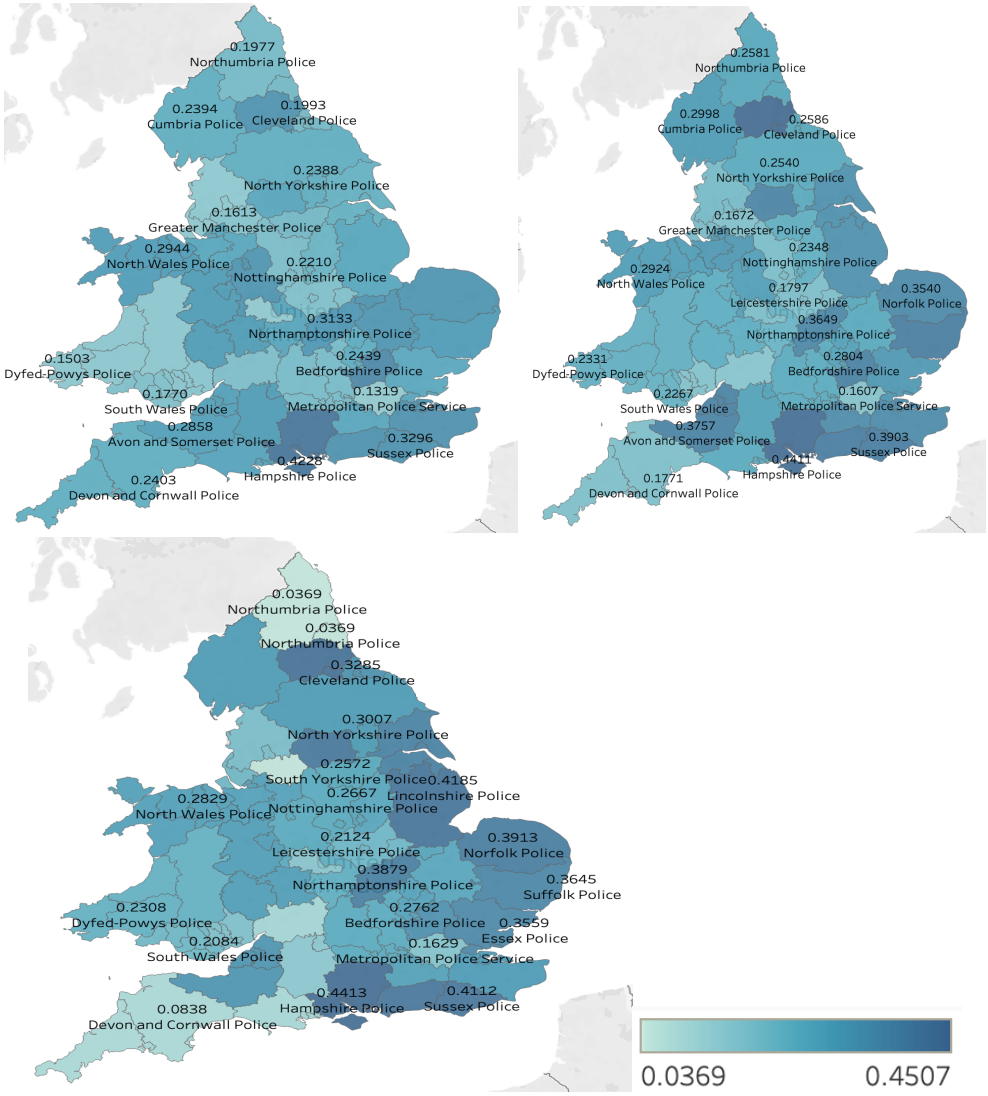

Figure 5: Rate of total assaults recorded per police officers in 2018(a - top left); 2019 (b - top right); 2020 (c - bottom)

As can be seen from Figure 5a,b, in general, the rate of total assaults over police officers in most police forces rose from 2017/18 to 2018/19. Compared to Figure 5a, Figure 5b has more dark colour tones, which presents the increase of the risk of being assaulted across England and Wales. However, there are some significant differences in Figure 5c. The assault rate fell below 0.09 in Northumbria and South West (except Avon and Somerset), while Sussex and Lincolnshire have a reserve trend.

It can be summarised that from 2018 to 2020, the risk of being assaulted raised in police forces located in East and South East England while showing a trend of decrease in the Northumbria and South West of England (except for Avon and Somerset). However, the impact of covid should be considered as one of the factors that affect the trend and the prediction of the assault rate.

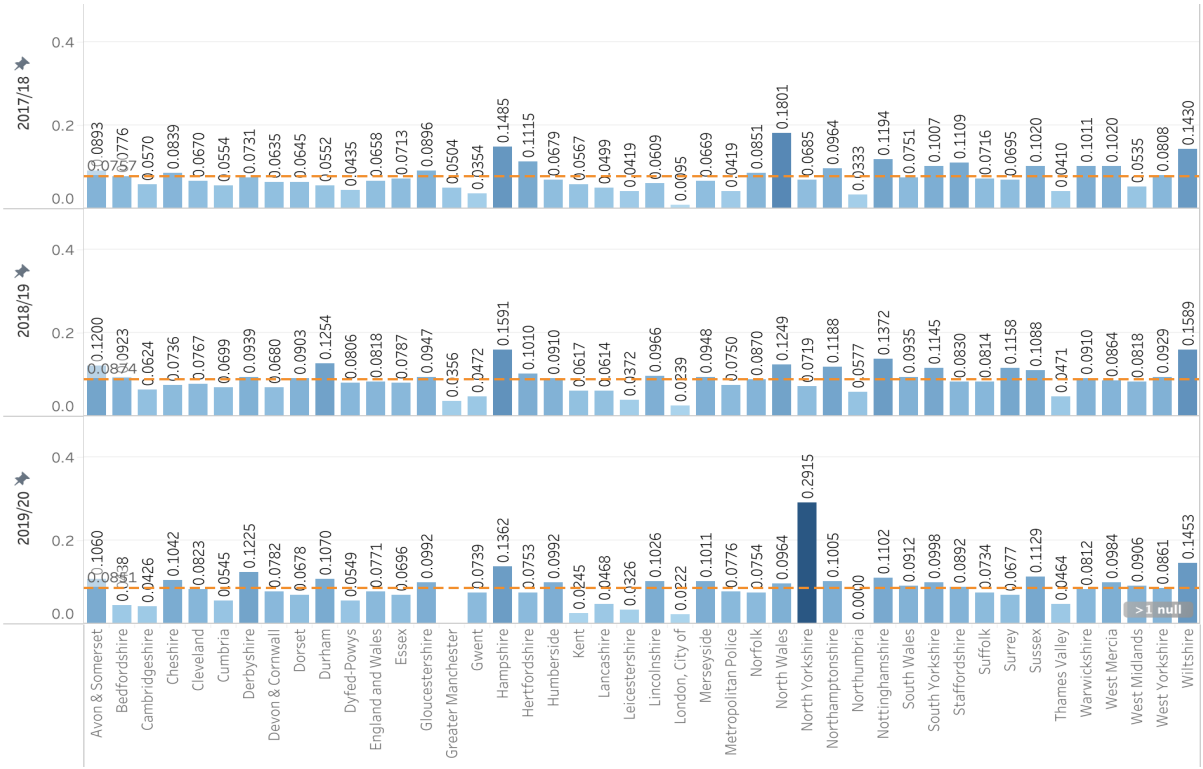

Figure 6: Rate of assaults with injuries over the number of police officers

(Note: figure of Greater Manchester in 2019/20 is 0 due to missing data)

Figure 6 presents the risk of being assaulted with injuries which help to give a closer look at the more serious level of being assaulted. From 2017/18 to 2018/19, the average ratio of assaults with injury increased by 15.46%, then slightly decreased by 2.63% in 2019/20. Hampshire had high rates of being assaulted with injuries, approximately double the average rate in all three years. It should be considered that the rate of assaults with injuries in North Wales policies strongly fell from 2017/18 to 2019/20, while North Yorkshire police witnessed a boom of this rate, from 0.06825 in 2017/18 to 0.2915 in 2019/20. Police officers working in the rest of the police forces are confronted with a similar risk of being assaulted with injury (less than 0.1) from 2017 to 2020.

Although the rates of total assault per officer in some forces, such as Kent, are generally high, police officers in there are likely to deal with less risk of being assaulted with injury than those in other forces. Therefore, the possibility of being assaulted and the level of assault are generally different across England and Wales.

Crime

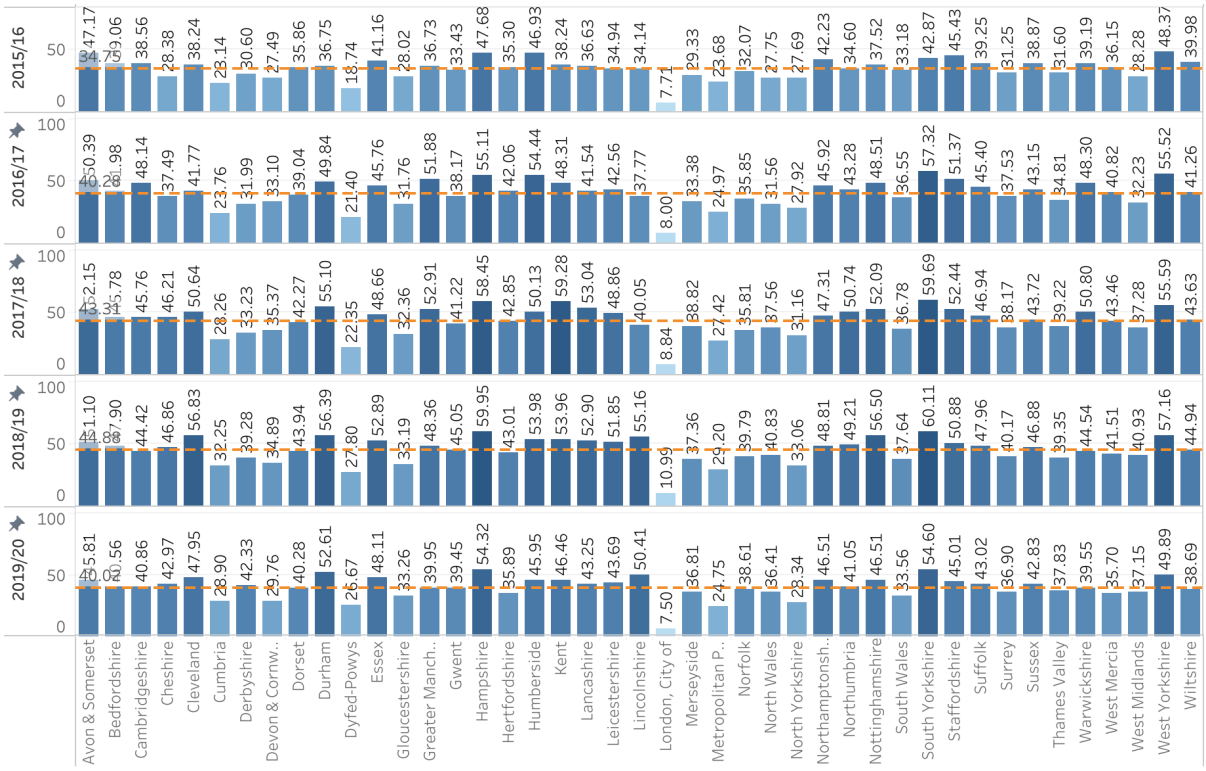

Figure 7: Rate of total crimes per police officer

From 2015/16 to 2018/19, the average rate of crime per police officer considerably increased up to 29%, from 34.75 to 44.88 crimes per police, then suddenly decreased by 11% in 2019/20 compared to the previous year (Figure 7). From 2016/17, on average, a police officer has to deal with over 40 crimes a year, and the number of crimes increased year by year. The high workload in a high-risk work environment may lead to a higher risk of being assaulted and mentally ill.

Closely looking at Figure 7, Hampshire, Durham, Humbershire, South Yorkshire, Staffordshire, and West Yorkshire are always in the top group, having high rates of crime per officer over five years. The highest rate recorded is 60.11 in South Yorkshire in 2018/19. In contrast, police officers in the City of London are likely to deal with a lower number of crimes in a year over five years. In 2018/19, the crime rate per office was below 11, which is one out sixth of the highest rate in the last five years. However, it could be explained that the land size covered by the City of London is only 1.1 miles, which is a significantly small area.

It is remarkable that Hampshire is on the top ranking of both the rate of being assaulted and the rate of total crimes, implying that officers in the Hampshire police force are likely to confront more risk than others.

From 2016/17 to 2019/20, the change of total crime across England and Wales compared to the previous year was 15.3%, 8.6%, 4,7%, and -6.5% respectively, which shows a decrease in the development of crimes. On the other hand, there is an opposite change in the total number of police officers. The change of the number of police officers compared to the year-earlier are -0.8%, 0.4%, 1.9%, and 6.2% respectively in the same time period. This may imply the inverse relationship between the number of crimes and police officers. One of the reasons is mentioned in the literature review: The decrease in the number of police during this time led to the shortage of the police workforce. As a consequence, some types of crime which were not well covered significantly increased.

Insurance provision

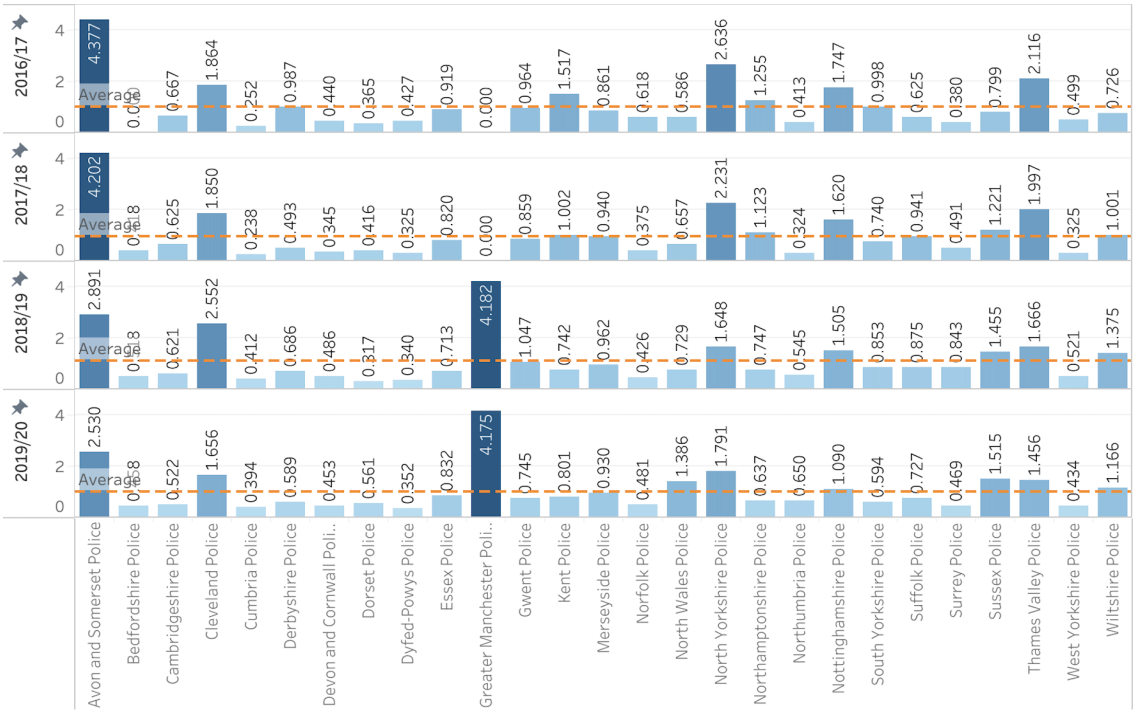

Figure 8a: Rate of Insurance Provision per Police Officer

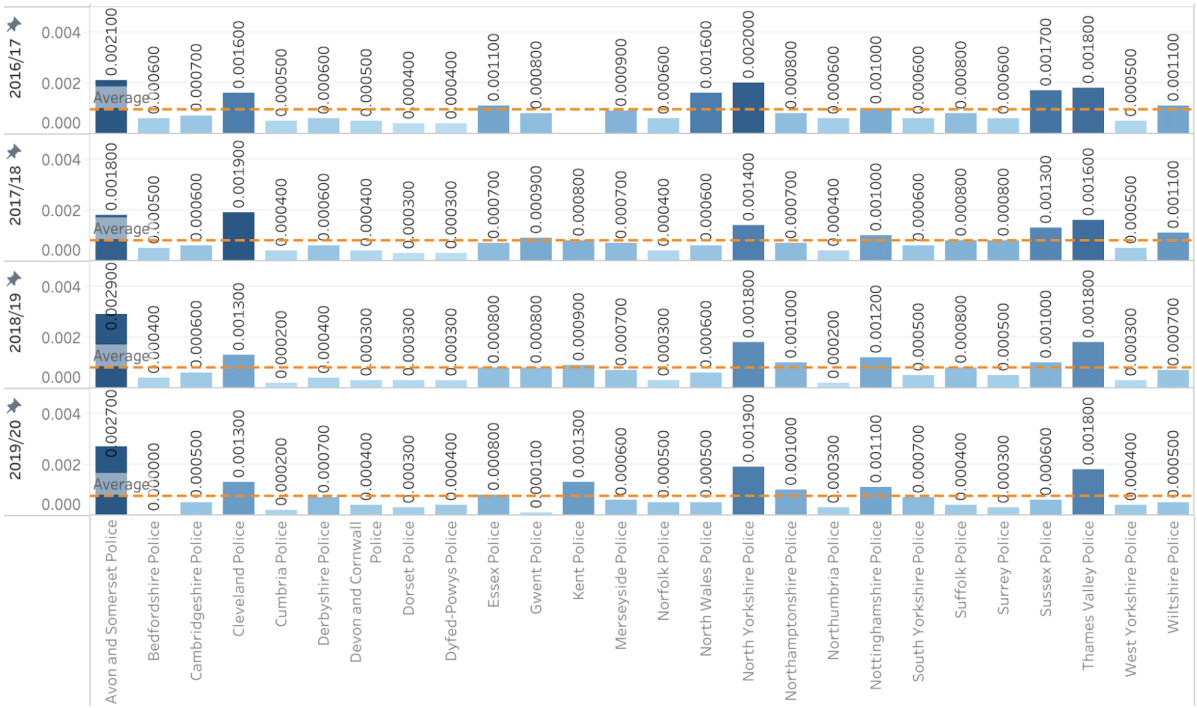

Figure 8b: Rate of Insurance Provision over Total Liabilities.

Note: insurance provisions for Greater Manchester in 2016/17 and 2017/18 are missing values

Ratios of insurance provision over total liabilities and insurance provision over the number of police officers are varied across England and Wales. Figures 8b and 8a respectively show that the range of average rate of insurance provision over total liabilities between 0.7% and 0.9%, and an amount from £947 to £1099 from insurance provision is set for each police officer across 27 police forces from 2016 to 2020.

The majority of police forces set the rate of insurance provision over total liabilities below 0.1%, which implies that the similar estimation of claims impacts on total force's liabilities.

The rate of insurance provision over total liabilities has a relatively similar shape to that of insurance provision per police office, which indicates the direct proportion relationship between the total liabilities and the number of police officers.

It is remarked that Avon and Somerset, Thames Valley, and North Yorkshire have significantly higher rates of insurance provision over total liabilities and insurance provision over police officers than others. It is reasonable because these two forces also have a high rate of assaults per officer and crimes per officer. This may imply the relationship between risk and insurance provision in police forces.

Clustering

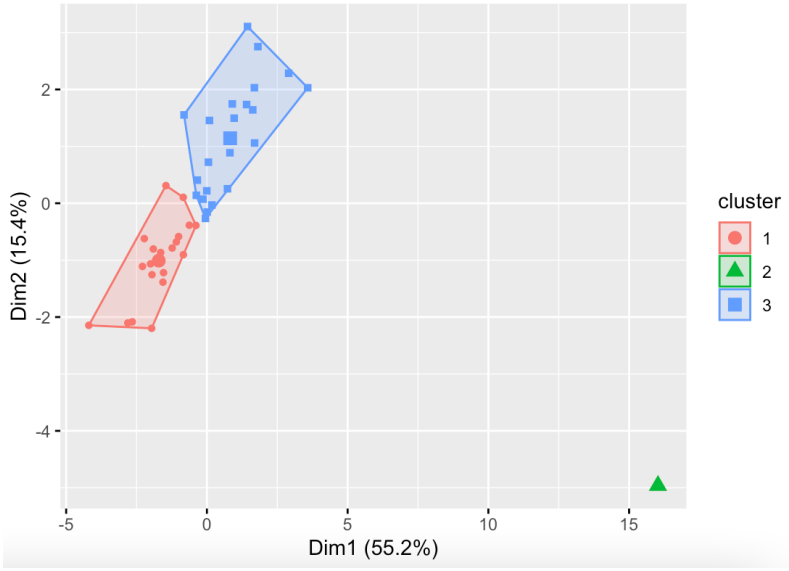

Figure 9: Plot of clusters when k = 3 in two dimensions

Cluster	Number of police forces	Police force
1	20	Cleveland, Durham, Humberside, South Yorkshire, West Yorkshire, Lincolnshire, Northamptonshire, Nottinghamshire, Staffordshire, Warwickshire, Bedfordshire, Cambridgeshire, Essex, Hertfordshire, Norfolk, Suffolk, Hampshire, Kent, Sussex, Avon & Somerset
2	1	London, City of
3	22	Northumbria, Cheshire, Cumbria, Greater Manchester, Lancashire, Merseyside, North Yorkshire, Derbyshire, Leicestershire, West Mercia, West Midlands, Metropolitan Police, Surrey, Thames Valley, Devon & Cornwall, Dorset, Gloucestershire, Wiltshire, Dyfed-Powys, Gwent, North Wales, South Wales

Table 2: Clustering result

As shown from Figure 9, which presents observations in two dimensions that capture 70.6% variance of the whole dataset in total, cluster 1 and 3 are close to each other, while both are quite far away from cluster 2. Because of the relatively high variance explained in the two first dimensions, principal components analysis is considered a good tool to explain each cluster's meaning further.

As can be seen from (Appendix 1), PC1 is dominated by variables related to the number of police officers per mile and crime rate in all five years, with similar absolute values. All values of variables related to crime per officer rate are negative while all values of variables related to the number of police officers are positive, this presents the two inverse relationships: variables related to the number of police officers and PC1, and variables related to the number of police officers and variables related to crime per officer rate. Besides, variables in terms of assault without injury rate are also presented by PC1 but do not dominate. Features in terms of assault without injury rate in 2020 and number of police officers per mile in five years and, followed by features in terms of crime rate, dominate PC2, which indicates all inverse relationships between variables and PC2. Besides, PC2 is also present assault with an injury rate in 2019/20. All values of features in terms of assault with injury rate in PC2 are positive, which means that an increase in these variables will proportionally increase in PC2. PC1 and PC2 capture over 70.6%, which can be considered a relatively good proportion. This can be explained by the highly correlated features in this dataset.

As shown from Appendix 2, cluster 1 has the highest average rate of crime per office and average rate of assault with and without injury and the lowest average rate of the number of officers per mile during the time. In contrast, cluster 2 has the lowest average crime rate and assault with and without injury rate and the highest number of officers per mile in all the years. It can be explained that there is only one observation (City of London) in cluster 2 and the land cover by this force is significantly small.

In general, police officers in cluster 1 are the most likely to deal with a high number of crimes and be assaulted, while those in cluster 3 are less likely to be assaulted and face a high number of crimes. Officers in cluster 2 are least likely to be dangerous because of the considerable small crime rate and assault rate. Therefore, officers in clusters 1,3, and 2 are classified as the high risk, moderate risk, and low risk group respectively.

It is expected that forces in the high-risk group will have a greater insurance provision and insurance provision per officer rate. However, Essex, Northamptonshire, Cambridge, Suffolk,

Norfolk, and Bedfordshire have a low rate of insurance provision per officer below the average rate. Especially, the rate of insurance provision per officer in Essex, Northamptonshire, Suffolk, and Cambridge force kept decreasing over the last four years, indicating the inadequacy of the risk level that police offices there are facing. It is possible that provision amount does not well cover current potential risks.

In contrast, although in the moderate-risk group, Thames Valley and North Yorkshire forces are on the top having a high rate of insurance provision per officer. However, because insurance provision also covers potential other liabilities such as motor vehicles, a specific proportion of each type of liability in insurance provision is needed to analyse further. Besides, the missing insurance provision data of 16 police force also trigger the difficulty in comparing police forces insurance provision between and within clusters.

Cluster	Jaccard mean	The number of dissolved times	The number of recovered times
1	0.8033027	7	68
2	0.6100000	39	61
3	0.7429693	10	51

Table 3: Stability measure *(Clusterboot function)*

Stability is one of the important elements of the cluster validation. The '*clusterboost'* function in R is used to compute the Jaccard bootstrap mean (see Table 4) to measure clusters' stability. Jaccard mean values of two out three clusters are considered relatively high (roughly and over 0.75), which means that these clusters are quite stable. Nevertheless, the number of recovered times of one cluster of them is only 51, which is considered relatively low. Furthermore, the Jaccard means value of the rest cluster is only 0.61, indicating the patterns between clusters (Hennig, 2007). The number of dissolved and recovered times for this cluster is 39 and 61 respectively, which indicates that this cluster is highly unstable. Therefore, two out of three clusters may be considered as not stable.

Limitation

The number of assault statistics that are recorded by the police is likely to underestimate the actual number of assaults. Several officers believe that being assaulted is normal and is an unavoidable part of the police job, so they do not report to crime records (Office for National Statistic, 2019). The risk of being assaulted, especially without injury, is likely to be higher than the evaluation in the paper. Besides, as mentioned before, crime statistics from police recorded crime do not cover all types of crimes, which means that the actual crimes police officers are dealing with are higher ratios in the previous discussion.

The result of the stability test for clustering indicates that one of the clusters is likely to be unstable, which may indicate that clusters are not validated. One of the reasons can be the small number of variables in the model, which does not provide enough information. Hence, more features relating to risks are needed to improve the stability and validation of clusters. Besides, due to the lack of information available from public resources and time-consuming in collecting data, the data used in this paper are limited to illustrate the risk that police officers are facing. For instance, more risk factors, such as insurance premium, the number of police officers and vehicles involved in car accidents by regions, may help identify the risk exposure and the ability to cover risks in police forces. Also, this paper's lack of insurance provision information in 16 police forces leads to the uncompleted comparison of claim effect levels between forces.

Conclusion

In order to react to the increase in the number of crimes in recent years, the number of police increased in general, but at different rates across regions. If the year 2019/20 is treated at a different level due to the impact of COVID-19, both the number of crimes and assaults recorded by police show a trend of increasing, indicating the risk development to police officers, especially in East and South East England. The effect of insurance claims is seen as similar in the majority of police forces because of the similar rate of insurance provision over total liabilities. Except for the City of London force which is considered as low risk group, other forces can be divided into high risk and moderate risk groups. Although the incomplete insurance provision in this situation is not allowed to illustrate a full picture of police forces's self-insurance, the existence of inadequacy of insurance provision and risk level is found in some forces. It is possible that the insurable risks are not well covered in some forces.

Reference

1. Allen, G. and Zayed, Y., 2021. *Police Service Strength*. [online] Researchbriefings.files.parliament.uk. Available at: <https://researchbriefings.files.parliament.uk/documents/SN00634/SN00634.pdf> [Accessed 17 August 2021].
2. Association of Police and Crime Commissioners, 2020. *Is there a difference between a Police Officer and a Police constable? - a Freedom of Information request to Association of Police and Crime Commissioners*. [online] WhatDoTheyKnow. Available at: <https://www.whatdotheyknow.com/request/is_there_a_difference_between_a> [Accessed 14 August 2021].
3. Association of Police and Crime Commissioner, 2021. *Role of the PCC*. [online] Apccs.police.uk. Available at: <https://www.apccs.police.uk/role-of-the-pcc/> [Accessed 15 August 2021].
4. Brown, J., 2021. *Policing in the UK*. [online] House of Commons Library. Available at: <https://commonslibrary.parliament.uk/research-briefings/cbp-8582/> [Accessed 9 August 2021].
5. Edwards, A. and Kotera, Y., 2020. Mental Health in the UK Police Force: a Qualitative Investigation into the Stigma with Mental Illness. *International Journal of Mental Health and Addiction*, [online] 19(4), pp.1116-1134. Available at: <https://link.springer.com/article/10.1007/s11469-019-00214-x> [Accessed 6 August 2021].
6. Dearden, L., 2018. *'Significant strain' on police budgets has made crime response worse, Home Office official admits*. [online] The Independent. Available at: <https://www.independent.co.uk/news/uk/crime/police-budget-cuts-home-office-conservatives-austerity-crime-rise-violence-a8578171.html> [Accessed 6 August 2021].
7. Dodd, V., 2019. *Criminals going unpunished because of cuts, says police chief*. [online] the Guardian. Available at: <https://www.theguardian.com/uk-news/2019/may/02/criminals-going-unpunished-because-of-cuts-says-police-chief> [Accessed 10 August 2021].
8. Dodd, V., 2020. *Two out of five police officers assaulted on duty in year to March 2019*. [online] the Guardian. Available at: <https://www.theguardian.com/uk-

news/2020/sep/02/two-out-of-five-police-officers-assaulted-on-duty-in-year-to-march-2019> [Accessed 14 August 2021].

9. Habersaat, S., Geiger, A., Abdellaoui, S. and Wolf, J., 2015. Health in police officers: Role of risk factor clusters and police divisions. *Social Science & Medicine*, 143, pp.213-222.
10. Google Developers, 2021. *k-Means Advantages and Disadvantages | Clustering in Machine Learning*. [online] Google Developers. Available at: <https://developers.google.com/machine-learning/clustering/algorithm/advantages-disadvantages> [Accessed 15 July 2021].
11. Houdmont, J. and Elliott-Davies, M., 2016. *Officer Demand, Capacity and Welfare Survey Descriptive Statistics Summary Report MENTAL HEALTH & WELLBEING*. [online] Polfed.org. Available at: <https://www.polfed.org/media/14070/welfaresurveymentalhealth-summaryreport-25-01-2017-v1.pdf> [Accessed 8 August 2021].
12. Hennig, C., 2007. Cluster-wise assessment of cluster stability. *Computational Statistics & Data Analysis*, 52(1), pp.258-271.
13. Hennig, C. 2015. Clustering strategy and method selection. arXiv [Online]. Available at: https://arxiv.org/abs/1503.02059 [Accessed 15 August 2021].
14. Sky News. 2019. *How dangerous is it to be a UK police officer?*. [online] Available at: <https://news.sky.com/story/how-dangerous-is-it-to-be-a-uk-police-officer-11786665> [Accessed 14 August 2021].
15. Miller, J., Brewin, C., Soffia, M., Elliott-Davies, M., Burchell, B. and Peart, A., 2021. The development of a UK police traumatic events checklist. *The Police Journal: Theory, Practice and Principles*, [online] pp.0032258X2110025. Available at: <http://The development of a UK police traumatic events checklist> [Accessed 2 August 2021].
16. Office for National Statistics, 2021. *Crime in England and Wales QMI - Office for National Statistics*. [online] Ons.gov.uk. Available at: <https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/methodologies/crimeinenglandandwalesqmi> [Accessed 14 August 2021].
17. Police Foundation, 1996. *THE ROLE AND RESPONSIBILITIES OF THE POLICE*. [online] Police-foundation.org.uk. Available at: <https://www.police-foundation.org.uk/2017/wp-content/uploads/2017/06/roles_and_resp.pdf> [Accessed 11 August 2021].

18. Police Firearms Officers Association, 2021. *Police Mental Health sickness up by 47%.* [online] PFOA. Available at: <https://www.pfoa.co.uk/blog/police-mental-health-sickness-up-by-47> [Accessed 8 August 2021].
19. Punj, G. and Stewart, D. W. ,1983. Cluster analysis in marketing research: Review and suggestions for application. Journal of Marketing Research, 20(2), 134-148 [Online]. Available at: https://doi.org/10.2307/3151680 [Accessed 12 August 2021].
20. The Institute for Government, 2019. *Police*. [online] The Institute for Government. Available at: <https://www.instituteforgovernment.org.uk/publication/performance-tracker-2019/police> [Accessed 12 August 2021].
21. The Times, 2018. *Police under attack*. [online] Thetimes.co.uk. Available at: <https://www.thetimes.co.uk/article/police-under-attack-gcdwv3qjs> [Accessed 20 August 2021].
22. University of Southampton, 2021. *Police Officer and Civil Roles in the Police | Economic, Social and Political Sciences | University of Southampton*. [online] Southampton.ac.uk. Available at: <https://www.southampton.ac.uk/socsci/undergraduate/career_path_police.page> [Accessed 15 August 2021].
23. Worksmart, 2021. *Which jobs carry the highest risk of workplace violence? | WorkSmart: The career coach that works for everyone*. [online] Worksmart.org.uk. Available at: <https://worksmart.org.uk/health-advice/illnesses-and-injuries/violence-and-bullying/violence/which-jobs-carry-highest-risk> [Accessed 14 August 2021].

Appendix 1

PCA loading analysis

	PC1	PC2
X2015.16.crime.rate	-0.28619692	-0.20154898
X2016.17.crime.rate	-0.27841850	-0.22063572
X2017.18.crime.rate	-0.27843318	-0.21951871
X2018.19.crime.rate	-0.28645957	-0.21696427
X2019.20.crime.rate	-0.29589187	-0.18419290
X2017.18.Assault.with.injury.rate	-0.16399235	0.05530295
X2017.18.assault.without.injury.rate	-0.20781788	-0.22565657
X2018.19.assault.with.injury.rate	-0.17292842	0.03533615
X2018.19.assault.without.injury.rate	-0.20782770	-0.29189978
X2019.20.assault.with.injury.rate	-0.07077958	0.22720656
X2019.20.assault.without.injury.rate	-0.16801365	-0.36822039
X2015.16.no.of.officers.per.mile	0.28920323	-0.29998609
X2016.17.no.of.officers.per.mile	0.28908987	-0.30039107
X2017.18.no.of.officers.per.mile	0.28860155	-0.30188757
X2018.19.no.of.officers.per.mile	0.28857689	-0.30198127
X2019.20.no.of.officers.per.mile	0.28851813	-0.30217735

Appendix 2:

The average rate based on groups

Variables	Cluster 1	Cluster 2	Cluster 3
X2015.16.crime.rate	40.451679	7.714572	30.793572
X2016.17.crime.rate	47.062688	8.139053	35.406609
X2017.18.crime.rate	49.662873	8.839967	39.105080
X2018.19.crime.rate	51.41131	10.99044	40.48122
X2019.20.crime.rate	45.773050	7.503113	36.260421
X2017.18.Assault.with.injury.rate	0.086539522	0.009541985	0.069367809
X2017.18.assault.without.injury.rate	0.20278601	0.02181025	0.13813570
X2018.19.assault.with.injury.rate	0.09897252	0.02389804	0.08001055
X2018.19.assault.without.injury.rate	0.23186016	0.02257037	0.14722944
X2019.20.assault.with.injury.rate	0.08590038	0.02219126	0.08526762
X2019.20.assault.without.injury.rate	0.25542654	0.05547816	0.11221468
X2015.16.no.of.officers.per.mile	2.018686	625.027273	5.771625
X2016.17.no.of.officers.per.mile	2.039596	614.545455	5.653063
X2017.18.no.of.officers.per.mile	2.072363	666.909091	5.614078
X2018.19.no.of.officers.per.mile	2.112649	684.727273	5.707208
X2019.20.no.of.officers.per.mile	2.258623	737.390909	6.050393

Appendix 3

```
# import library
library(ggplot2)
library(gridExtra)
library(readr)
library(ISLR)
library(fpc)
library(dplyr)
library(faraway)
library(cluster)
library(factoextra)
library(tidyverse)
library(NbClust)
library(mvoutlier)
library("readxl")
library(clValid)
library(kohonen)
library(mclust)
library(mvoutlier)
# import data
data <- read_excel("Ngoc_Nguyen_Rosalba.xlxs", sheet = "R-input")
# transform input to dataframe type
data <- data.frame(data)
# explore dataset
colSums(is.na(data))
sum(duplicated(data))
dim(data)
# get the necessary data
data[1:43,2:17] <- sapply(data[1:43,2:17], as.numeric) # transfrom data to numeric
apply(data[1:43,2:17] , 2, var)     # check range of variance of features
apply(data[1:43,2:17] , 2, mean)    # check range of mean of features
data_1 <- data[1:43,2:17]%>% as.data.frame()
# overview the dataset used
colnames(data_1)
colSums(is.na(data_1))
##########################scaled data###################################
data.sc <- scale(data_1) %>% data.frame()           #scale data
# using total within sum of squares to choose k
set.seed(4)
fviz_nbclust(data.sc, kmeans, method = "wss") +labs(subtitle = "Elbow method")
# may choose k=3,4
# transform data to matrix
data_final <- as.matrix(data.sc)
```

```
# use other tests
valid_test <- clValid(data_final, c(3:10),
                      clMethods = c("kmeans"),
                      validation = c("internal", "stability")
)
summary(valid_test)
## may choose k=3

set.seed(4)
k_means_3_sc <- kmeans(data.sc,3) ## clustering with k=3
fviz_cluster(k_means_3_sc, geom = "point", data = data.sc) + ggtitle("k = 3")
k_means_3_sc$cluster

# check stability of clusters

boot_kmean_3_sc <-clusterboot(data=data.sc,bscompare=T,
                          multipleboot=F,bootmethod="boot", B=100, clustermethod=kmeansCBI,
                          count=T, showplots=T, krange=3, seed = 4)

boot_kmean_3_sc

################################## PCA ######################################

pr.out= prcomp(data.sc, scale=TRUE)
biplot(pr.out, scale = 0)
# Compute variance explained
pr.out$sdev
pr.var = pr.out$sdev ^2
pr.var
pve = pr.var/sum(pr.var)
pve

pr.out$rotation
# summary information between cluster
data_sc_kmean <- data_1%>%mutate(cluster=k_means_3_sc$cluster)
summaries_info <- data_sc_kmean%>%group_by(cluster)%>%summarise_all(funs(mean(.)))

summary_table <-summaries_info%>%as.data.frame()
view(summary_table)
```