| Student Number | 23045206 |
|---|---|

# Comparative study of Socioeconomic, Demographic, and Housing factors impact on fire and false alarm rates over time.

## 1. Introduction:

Numerous research were conducted to comprehend socio-economic factors impact on fire rate. (Hastie and Searle, 2016) analysed the accidental dwelling fires distribution across different socio-economic and demographic groups in the UK. Other research addressed the relationship between fire incidents and human behaviour. (Thompson, Galea and Hulse, 2018) examined existing literature to understand human behaviour during dwelling fires. They discussed the differences in socio-economic and demographic factors impact on human behaviour during dwelling fires versus other environments. Furthermore, they highlighted the importance of studying dwellings fire incidents due to the high frequency and significant consequences of such fires. A vast body of research studies COVID-19 impact on fire injuries. However, little research investigates COVID-19 impact on dwelling fire incidents. (Suzuki and Manzello, 2022) used LFB data to investigate COVID-19 lockdown impact on dwelling fires, finding increase in incidents during lockdown which was associated with behaviour changes. Research on false fire alarms is limited, with few papers investigating false alarms as distinct incidents type. Therefore, addressing this issue is crucial, given the significant financial burden false alarms impose on the UK economy each year, estimated to cost over £1 billion annually ('False fire alarms continue to cost the UK | IFSEC Insider', 2023). (Corcoran *et al.*, 2007) is one of few studies that investigated false alarm incidents examining socio-economic factors impact on malicious false alarms and different fire incident types. Using spatial analysis, they created incident rate maps and kernel density estimation surfaces. They also applied Poisson and negative binomial regression models to examine relationships between fire incidents and socio-economic variables. Their findings indicated that false alarm incidents were higher in deprived areas and significantly associated with educational attainment and car ownership.

Previous papers studied socio-economic, demographic factors or COVID-19 impact separately on fire incidents at various scales using several methods, including statistical and spatial analysis, to detect the most influential factors driving fire risk. Most of these studies used multiple regression models, and some implemented advanced techniques to explore fire incidents dynamics and their relationships with socio-economic variables. However, few studies examined those factors combined impact on fire incidents. Furthermore, comparative analyses of changes in socio-economic factors affecting fire incidents over time in the same geographical area were rare.

This work aims to investigate commonly and rarely studied socio-economic, demographic and housing variables on dwelling fire and false alarm incidents, using multiple regressions. A comparative study between 2011 and 2021 analyses patterns and changes over time. The analysis highlights how socio-economic inequalities impact dwelling fire and false alarm incidents, identifying at-risk populations. These insights can inform targeted fire prevention and response strategies, aiding LFB in campaign targeting.

## 2. Research Question:

a. Do socio-economic and demographic factors influence London wards fire and false alarm rates in 2011 and 2021?
b. Do socio-economic and demographic factors driving fire and false alarm rates show different behaviour for 2011 versus 2021?

**Hypothesis:**
Null Hypothesis (H0): Socio-economic, demographic and housing factors have no significant effect on fire and false alarm rates.

Alternative Hypothesis (H1): Socio-economic, demographic and housing factors have significant effect on fire and false alarm rates.

## 3. Data:

LFB dataset includes several variables, such as year, date, incident group, property category, location, and property type. This analysis focuses on dwelling property category while fires and false alarms incident groups.

**Fire and False Alarm Rates:**
The study scale is on wards-level as Census data most granular level is wards. To calculate fire and false alarm rates, data was aggregated to calculate wards incident counts. Following (Hastie and Searle, 2016), households number was used as the denominator. For dwelling fires, the population at risk is more accurately represented by dwellings number, as the risk of fire is associated with dwellings, not inhabitants number.

(Corcoran, Higgs and Higginson, 2011) advised to use Census data in analysing fire incidents, which provides detailed demographic, socioeconomic, and housing information at small spatial scales, enabling comparisons across London and identifying local inequalities and trends. 2011 and 2021 Census datasets are used. As census data provides variables of too granular classification, aggregation was performed to generate a high-level and hand-crafted classification for this analysis. Then Census data were aggregated on wards level to generate percentages for independent variables.

**Lockdown:**
One more feature was added to represent lockdown in 2021 based on UK national lockdown dates, then fire and false alarm rates and independent variables percentages were calculated for lockdown and no lockdown of 2021 only.

| **Housing Characteristics** | | |
|---|---|---|
| Variable Name | Description | Hand-crafted Category |
| Central Heating | No central heat, Central heat (oil, solid fuel, electric...) which were later aggregated to represent central heat. | No central heat, central heat. |
| Accommodation Type | Simi-detached, Detached, Flats, terraced, flats, flats in converted house, flats in commercial building all of which were aggregated into categories. | Not detached, detached, flats. |
| House Tenure | Owned, owned with a mortgage, shared ownership, rented from local authority, other social rented, private landlord or letting agency, other private rented, rent free | Owned, social rented, privet rented, Other |
| Occupancy Rating | This data represents whether households have an adequate number of bedrooms based on the composition and needs of the household members. | Occupancy good fit, occupancy bad fit, occupancy fit. |

**Table 1:**Housing Variables

| **Demographic Characteristics** | | |
|---|---|---|
| Variable Name | Description | Hand-crafted Category |
| Ethnicity | Different ethnicities of White, Black, Asian, Mixed which were later aggregated. | White, Asian, Black, Mixed, Other |
| Language Proficiency | Data represents the degree to which residents speak English (native, well, not very well, etc…) | Speak English, doesn’t speak English. |

**Table 2:**Demographic Variables

| Socioeconomic Factors | | |
|---|---|---|
| Variable Name | Description | Hand-crafted Category |
| Deprivation Dimension | Represent the deprivation dimensions of a household (1,2,3 or 4 dimensions). | Deprived, severe deprived, medium deprived |
| Economic Activity | Represent residents' economic activity (active, inactive, unemployed, retired, etc..) | Economically active/ inactive, unemployed |
| Crime Rate | Represent all types of crimes and their respective counts within wards. | Crime rate. |
| Number of Cars | The number of cars in a household is a proxy for poverty. | No cars, one car, two or more cars. |
| Education | Represents educational qualification (no qualification, level 1,2, 3...). | Low, medium, high, other education. |

**Table 3:**Socioeconomic Variables

# 4. Methodology

### Multicollinearity

Census data have high correlation in nature, hence using Pearson Correlation and VIF to handle multicollinearity is crucial to produce a robust model that truly explains relationships between dependent and independent variables.

$$VIF = 1 / (1 - R_j^2) \tag{1}$$

VIF is not enough to address insignificant variables. Using aggressive alpha values for lasso might remove some important features, hence a two-layer approach were lasso applied first to get the most significant ones with lenient alpha followed by a stepwise approach to produce the best model for each dependent variable.

### Model

Multiple linear regression model for fire rate was developed for each year separately to measure independent variables effect and changes between the two years. The same approach was used to study false alarm rate.

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \tag{2}$$

### Residuals Investigation

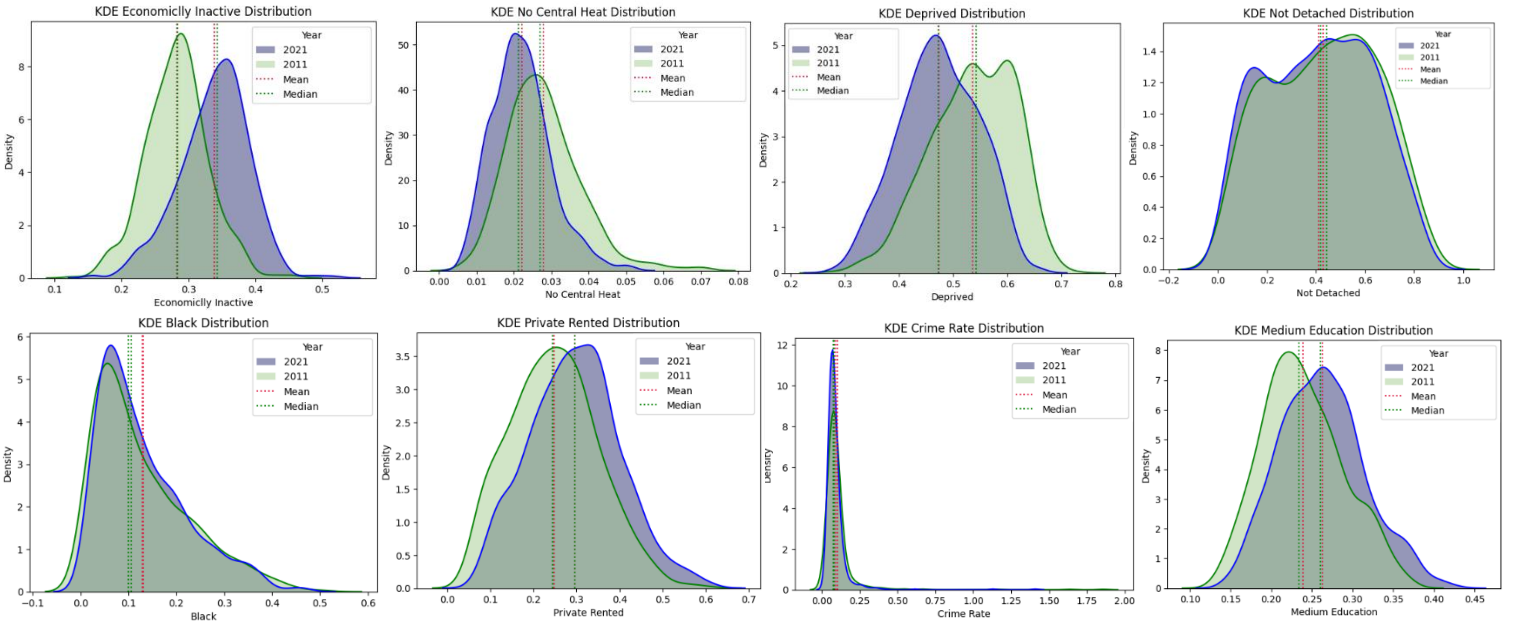
A combination of plots like Q-Q and residuals versus fitted plots and statistical tests (**Table-8**) were used to check models' adherence to linear regression assumptions.

### Independent Variables PCA versus Hand-crafted Classification

Following (Corcoran *et al.*, 2007), Principal Component Analysis is a useful approach to construct laten and higher classification variables from lower classification level Census data provide. An experiment was conducted to compare between hand-crafted classification variables against PCA features (one and two components). Results showed models' performances of hand-crafted variables were better (around 20% increase in R-square) and more robust against statistical fit tests.

# 5. Results:

## Independent Variables Statistic:




## Multicollinearity:

**Figure 1:**Top Variables KDE Plots

| Variable | 2011 VIF | 2021 VIF |
|---|---|---|
| Unemployed | 6.77 | 6.03 |
| Severe Deprived | Dropped | 5.45 |
| Deprived | 6.66 | Dropped |
| Black | 5.11 | 3.34 |
| Doesn't Speak English | 4.99 | 5.47 |
| Private Rented | 4.87 | 4.55 |
| Not Detached | 4.58 | 5.24 |
| Medium Education | 4.06 | 3.25 |
| One Car | 3.74 | 3.25 |
| Asian | 3.59 | 3.46 |
| Mixed | 3.39 | 3.25 |
| Economically Inactive | 2.95 | 3.22 |
| Other Ethnicity | 2.2 | 2.3 |
| No Central Heat | 2.12 | 3.33 |
| Detached | 2.04 | 2.41 |
| Crime Rate | 1.35 | 1.45 |
| Lockdown | Not Relevant | 1 |

**Table 4:**VIF Values

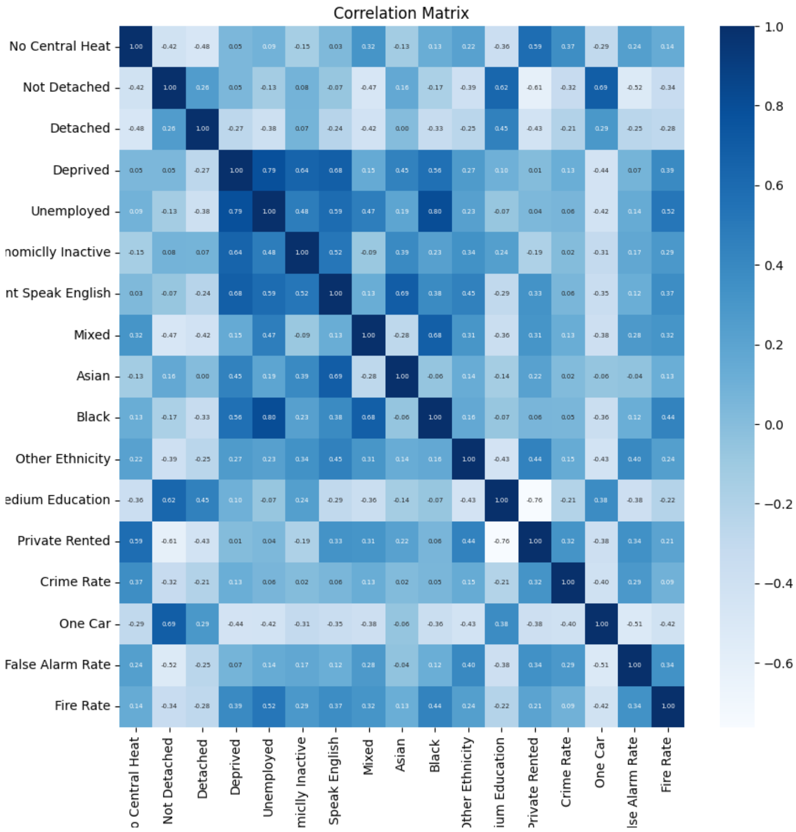



**Figure 2:**Correlation Matrix

VIF threshold used to drop highly correlated features is 7. **Table-4** shows while features meeting VIF threshold remained the same between the two years apart from deprivation (severe deprived versus deprived), VIF values ranks changed (e.g. Black or Not Detached) suggesting relationships shift between independent variables across the years.

## Multiple Regression Models

Fire rate models' R-squared of 2011 and 2021 are 0.877 and 0.772 respectively meaning around 87.7% and 77.2% of fire rate variability can be explained by independent variables. Furthermore, false alarm rate models' R-squared of 2011 and 2021 are 0.8 and 0.797 respectively meaning around 80% and 79.7% of false alarm rate variability can be explained by independent variables. From **table-5** R-squared and adjusted R-squared are very close proving that models are not penalised for unnecessary complexity. Also, models' F-statistics values are high indicating that models are statistically significant.

| Model Name | R-Squared | Adjusted R-Squared | F-Statistics | Model Df |
|---|---|---|---|---|
| 2011 Fire Rate | 0.877 | 0.876 | 687 | 7 |
| 2021 Fire Rate | 0.772 | 0.770 | 570.3 | 8 |
| 2011 False Alarm Rate | 0.8 | 0.798 | 298 | 9 |
| 2021 False Alarm Rate | 0.797 | 0.796 | 530.1 | 10 |

**Table 5:**Models Statistics

Features P-values of fire and false alarm rates models are less than 0.05 meaning all features are statistically significant and can reject the null hypothesis concluding that included features have statistically significant relationship with dependent variables.

| Variable | 2011 Coeff | 2021 Coeff | 2011 P>\|t\| | 2021 P>\|t\| | 2011 t-stat | 2021 t-stat |
|---|---|---|---|---|---|---|
| Not Detached | -0.0015 | -0.0004 | 0.000 | 0.000 | -8.399 | -4.524 |
| Detached | -0.0014 | N/A | 0.003 | N/A | -2.953 | N/A |
| Deprived | 0.0014 | N/A | 0.020 | N/A | 2.339 | N/A |
| Severe Deprived | N/A | 0.0051 | N/A | 0.000 | | 5.985 |
| Economically Inactive | 0.0041 | 0.0012 | 0.000 | 0.000 | 4.429 | 5.265 |
| Asian | 0.0005 | N/A | 0.032 | N/A | 2.152 | N/A |
| Black | 0.0027 | 0.0009 | 0.000 | 0.000 | 6.489 | 5.588 |
| Private Rented | N/A | 0.0004 | N/A | 0.001 | | 3.435 |
| Crime Rate | N/A | 0.0004 | N/A | 0.024 | | 2.265 |
| One Car | 0.0011 | 0.0006 | 0.011 | 0.001 | 2.535 | 3.386 |
| Lockdown | N/A | -0.0006 | N/A | 0.000 | | -26.136 |

**Table 6**:Fire Rate Models Coefficients

**Table-6** shows both fire rate models share some features (e.g. Not Detached and Economically Inactive), however big shifts exist in independent variables effect between 2011 and 2021 (e.g. deprivation effect becomes more important in 2021 versus 2011). Also, lockdown effect has a statistically significant relationship with fire rate for 2021.

| Variable | 2011 Coeff | 2021 Coeff | 2011 P>\|t\| | 2021 P>\|t\| | 2011 t-stat | 2021 t-stat |
|---|---|---|---|---|---|---|
| No Central Heat | N/A | 0.0650 | N/A | 0.000 | N/A | 4.076 |
| Not Detached | -0.0058 | -0.0046 | 0.000 | 0.000 | -6.350 | -10.148 |
| Detached | -0.0046 | -0.0068 | 0.033 | 0.000 | -2.138 | -5.187 |
| Economically Inactive | 0.0346 | 0.0185 | 0.000 | 0.000 | 9.037 | 13.035 |
| Doesn't Speak English | -0.0356 | -0.0317 | 0.000 | 0.000 | -5.429 | -8.883 |
| Black | N/A | 0.0019 | N/A | 0.038 | N/A | 2.078 |
| Mixed | 0.0196 | N/A | 0.035 | N/A | 2.108 | N/A |
| Other Ethnicity | 0.0233 | 0.0129 | 0.001 | 0.000 | 3.237 | 4.766 |
| Private Rented | 0.004 | 0.0021 | 0.009 | 0.019 | 2.631 | 2.345 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Medium Education | -0.0101 | N/A | 0.023 | N/A | -2.277 | N/A |
| Crime Rate | 0.0041 | 0.0055 | 0.001 | 0.000 | 3.316 | 5.660 |
| Lockdown | N/A | -0.0042 | N/A | 0.000 | N/A | -29.947 |

**Table 7:**False Alarm Rate Models Coefficients

**Table-7** shows both false alarm models share similar features coefficients (e.g. Not Detached or Doesn't Speak English), however big shifts exist in independent variables effect between 2011 and 2021 (e.g. ethnicity effect changed from black to mixed in 2021 versus 2011). Also, lockdown effect has a statistically significant relationship with false alarm rate for 2021.

**Table-8** shows both fire rate models pass most tests except Jarque Bera because residuals don't 100% follow a normal distribution, however they still meet linear relationship, equal variance and independent errors tests. Equally, false alarm rate models fail to meet normal distribution Jarque Bera test. However, although they fail independence errors Ljung Box test, they pass Durbin-Watson test (values around 2) meaning there is some evidence of residuals independence errors.

| **Model** | **Rainbow** | **Durbin-Watson** | **Ljung Box** | **Jarque Bera** | **Goldfeld-Quandt** |
|---|---|---|---|---|---|
| 2011 Fire Rate | 0.22 | 1.84 | 0.22 | 0 | 0.16 |
| 2021 Fire Rate | 0.53 | 1.96 | 0.43 | 0 | 0.47 |
| 2011 False Alarm Rate | 1 | 1.78 | 0.004 | 0 | 0.4 |
| 2021 False Alarm Rate | 0.97 | 1.88 | 0.036 | 0 | 0.39 |

**Table 8:**Residuals Statistical Tests

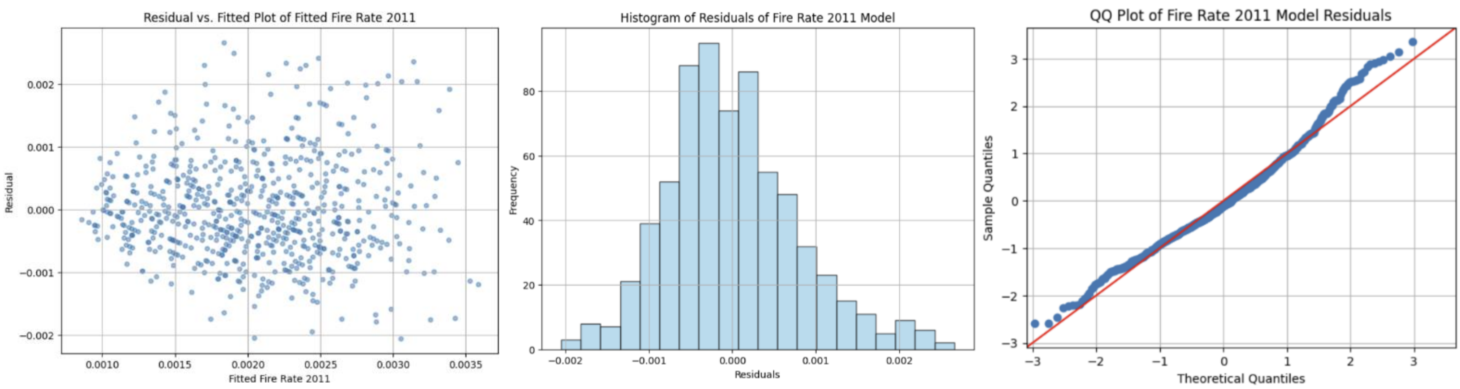



**Figure 3:**Fire Rate Residuals

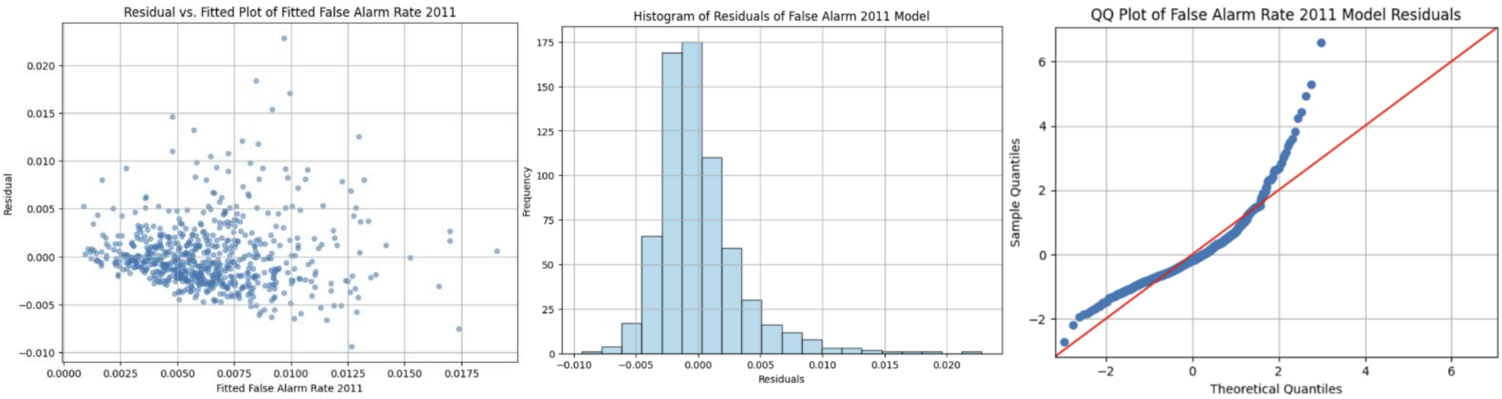



**Figure 4:**False Alarm Residuals

Figure-3 shows fire rates models' plots having a slight tail in the residuals distribution agreeing with statistical tests results. Similarly, figure-4 shows false alarm rates models having right skewness which aligns with statistical tests that residuals don't follow normal distribution.

## 6. Discussion

Fire rate models show that different socio-economic, demographic and housing variables are statistically significant. Top three positive coefficients for both years are economically inactive, black ethnicity and deprivation variables (medium deprivation for 2011 versus sever for 2021), while the only negative coefficient for both years is owning a house (detached or simi-detached). This shows that although London fire rate dropped from 0.2% in 2011 to 0.14% in 2021, it still has considerable inequality in dwelling fires distribution across different societal areas affecting groups like black ethnicity and deprived people confirming (Hastie and Searle, 2016) results.

Additionally, lockdown coefficient is negative for fire rates meaning that from January till early April 2021, fire rate was lower compared to rest of the year. Although this seems contradicting with (Suzuki and Manzello, 2022), however in their study they showed how Covid first wave and each wave first week had the most increase of fire rate. They presented for Covid third wave (from end of December 2020 till March) that fire rate increased at first then dropped which aligns with these results.

Finally, although false alarm increased over time from 0.62% to 0.84%, top three positive coefficients for 2011 are economically inactive, ethnicity and crime rate which remined important for 2021 with an additional impact of no central heat variable. This aligns with (Corcoran *et al.*, 2007) results that deprivation has a positive impact on false alarm (represented as economically inactive variable). Like fire rate models, lockdown showed negative coefficient meaning that false alarm rates were lower during lockdown period.

## 7. Conclusion

This study shows how socio-economic, demographic and housing variables have statistically significant relationship with both fire and false alarm rates using multiple linear regressions models. It proves that even with the UK government work to reduce fire rate, inequality is still present in fires and false alarms distributions assuring the need for further efforts. Additionally, it shows structural changes like Covid have an impact on people behaviours and eventually translated to changes in fire rates.

| | |
|---|---|
| Word Count | 1750 |

# References:

Corcoran, J. *et al.* (2007) 'The use of spatial analytical techniques to explore patterns of fire incidence: A South Wales case study', *Computers, Environment and Urban Systems*, 31(6), pp. 623–647. Available at: https://doi.org/10.1016/j.compenvurbsys.2007.01.002.

Corcoran, J., Higgs, G. and Higginson, A. (2011) 'Fire incidence in metropolitan areas: A comparative study of Brisbane (Australia) and Cardiff (United Kingdom)', *Applied Geography*, 31(1), pp. 65–75. Available at: https://doi.org/10.1016/j.apgeog.2010.02.003.

'False fire alarms continue to cost the UK | IFSEC Insider' (2023) *IFSEC Insider | Security and Fire News and Resources*, 21 September. Available at: https://www.ifsecglobal.com/fire-alarm/false-fire-alarms-continue-to-cost-the-uk/ (Accessed: 13 January 2025).

Hastie, C. and Searle, R. (2016) 'Socio-economic and demographic predictors of accidental dwelling fire rates', *Fire Safety Journal*, 84, pp. 50–56. Available at: https://doi.org/10.1016/j.firesaf.2016.07.002.

Suzuki, S. and Manzello, S.L. (2022) 'Effect of three lockdowns in London: Case study for residential fires', *Public Health in Practice*, 3, p. 100264. Available at: https://doi.org/10.1016/j.puhip.2022.100264.

Thompson, O.F., Galea, E.R. and Hulse, L.M. (2018) 'A review of the literature on human behaviour in dwelling fires', *Safety Science*, 109, pp. 303–312. Available at: https://doi.org/10.1016/j.ssci.2018.06.016.