Analyzing Demographic and Socioeconomic Predictors of COVID-19 Mortality in England

Contents

[Abstract 1](#_Toc187093828)

[Introduction 2](#_Toc187093829)

[Research Questions 2](#_Toc187093830)

[Literature Review 4](#_Toc187093831)

[Methodology 6](#_Toc187093832)

[Conclusion and Discussion 18](#_Toc187093833)

[APPENDIX 19](#_Toc187093834)

[SQL Queries 19](#_Toc187093835)

[References 20](#_Toc187093836)

# Abstract

This study investigates the relationship between demographic and socioeconomic factors and COVID-19 mortality rates in England, focusing on variables such as age, method of travel to work, highest qualification, religion, and ethnic group. Using census data from the Office for National Statistics and COVID-19 death records, multiple statistical techniques were applied, including factor analysis and multiple linear regression modeling. The final regression model identified **Remote\_jobs, Taxi usage, Bicycle usage, and Children population** as significant predictors of COVID-19 mortality rates. Higher taxi usage and a greater proportion of children were associated with increased mortality, while remote work availability and bicycle usage were linked to lower death rates. The model explained a modest portion of the variance but highlighted key risk factors that can inform public health policies. These findings emphasize the need for targeted interventions and transport policies to mitigate future health risks.

# Introduction

COVID-19, caused by the SARS-CoV-2 virus, began in Wuhan, China, in late 2019 and quickly became a worldwide pandemic. It put a huge strain on hospitals and hurt the economy. The virus spreads mainly through droplets from coughs and sneezes, and it takes about 6.4 days for symptoms to appear. People usually get fever, cough, shortness of breath, muscle pain, and tiredness. While the majority of patients tend to have a mild illness, a minority of patients develop severe hypoxia requiring hospitalization and mechanical ventilation (Ochani *et al.*, 2021).

The COVID-19 pandemic has presented an unprecedented global health crisis, with England been one of the major countries grappling with its profound repercussions. The sheer scale of the pandemic has exposed pre-existing health disparities, shining a light on systemic loopholes that render certain populations more vulnerable to severe outcomes. In England, these disparities have been starkly reflected in the number of COVID-related deaths, which have varied significantly across regions, socioeconomic strata, and demographic groups. This study examines the complex interaction of Age, Method of travel to work, Highest Qualification, Religion, and Ethnic Group as factors influencing COVID-19-related deaths in England. Through data exploration and analysis, we aim to uncover relationships between these variables and COVID-19 fatalities. This approach seeks to explore the distinct vulnerabilities experienced by various segments of the population during the pandemic.

## Research Questions

* How does Age influence the likelihood of COVID-19 deaths in England?
* What ways does religion contribute to the risk of COVID-19 related deaths considering the impact of cultural practices, religious gatherings, and community dynamics within specific religious groups?
* What are the factors contributing to ethnic group differences in COVID-19 impacts, and how can this understanding inform public health strategies and interventions to address disproportionate impacts on certain communities?
* Does the primary method of travel to work (e.g., public transport, private vehicles, cycling, walking) correlate with higher rates of COVID-19-related deaths in England?
* Do individuals with lower levels of formal education experience higher mortality rates compared to those with advanced qualifications in England?
* How will these variables interact to influence the risk of COVID-19 related death? For example, does the combination of race and mode of transport exacerbate vulnerability compared to these factors individually.

These questions will guide the analysis and offer a framework for understanding the relationship between these variables COVID-19 mortality.

# Literature Review

The COVID-19 pandemic has exposed and exacerbated disparities in health outcomes across different demographic and socio-economic groups. This section reviews existing literature on the roles of age, mode of transport, highest qualification, religion, and race in influencing COVID-19-related mortality.

**Age**

Age is a critical determinant of vulnerability to COVID-19. Older adults, particularly those above 65 years, are more susceptible to severe outcomes and higher mortality due to weaker immune systems and the presence of comorbidities (Verity et al., 2020). In contrast, younger populations generally experience lower mortality rates, though the risks increase for those with underlying health conditions (Davies et al., 2020).

**Method of travel to work**

Method of travel to work significantly influence the likelihood of exposure to COVID-19. Public transport has been associated with increased risk due to crowding and inadequate ventilation, while private vehicles and active modes, such as walking or cycling, offer safer alternatives (Tirachini and Cats, 2020). These disparities often intersect with socio-economic status, as lower-income individuals are more likely to rely on public transport.

**Highest Level of Qualification**

Educational attainment affects health outcomes during pandemics through multiple pathways. Lower levels of education are linked to poorer health literacy, limited access to resources, and higher rates of comorbidities, which contribute to increased mortality risk (Marmot et al., 2020). Conversely, individuals with higher qualifications often have better access to healthcare and are more likely to adhere to public health measures (Falkingham et al., 2020).

**Religion**

Religion shapes health outcomes in complex ways. Religious beliefs and practices can affect health behaviors and adherence to public health measures during a pandemic. For example, while communal religious gatherings have been identified as potential sources of virus transmission, some religious communities have played a vital role in providing social and psychological support (Pew Research Center, 2020). Furthermore, religious beliefs have influenced COVID-19 vaccination acceptance and the willingness to adhere to preventative measures. Studies suggest that religious groups, especially those with strong communal ties, may face challenges in adjusting to social distancing guidelines.

**Ethnic Group**

Ethnic Group are significant predictors of health disparities during the pandemic. Minority groups often face disproportionate risks due to systemic inequalities in housing, healthcare access, and employment conditions (Bhala et al., 2020). Structural racism and historical downgrading have compounded these vulnerabilities, resulting in higher mortality rates among racial minorities (Devakumar et al., 2020).

**Hypothesis**

H1: Age is positively correlated with COVID-19-related mortality, with older age groups experiencing higher mortality rates.

H2: Reliance on public transport is associated with an increased risk of COVID-19-related deaths compared to private transport or active travel modes.

H3: Lower educational attainment correlates with higher COVID-19-related mortality.

H4: COVID-19-related mortality varies significantly across religious groups due to differences in practices and health behaviors.

H5: Racial minorities experience higher COVID-19-related mortality due to systemic health and socio-economic inequities.

# Methodology

This project adopts a comprehensive methodology, utilizing statistical techniques such as regression modeling to examine how various factors influenced COVID-19 mortality rates in England. The aim is to provide useful insights that can help policymakers, healthcare workers, and public health planners create targeted solutions. These solutions will address the specific challenges faced by different groups and help build a fairer and stronger healthcare system for future health crises.

For the analysis, the R programming language, accessed through RStudio, has been chosen due to its open-source nature and extensive array of statistical libraries and packages. By leveraging data analysis, this project will employ diverse analytical models selected based on their performance, accuracy, and predefined parameters. The objective is to pinpoint key variables that effectively highlight the primary risks associated with COVID-19 mortality in England.

**Data Acquisition**

For this project the total “Covid19 Deaths” in England was provided to us in CSV format from the UEL Moodle(moodle.uel.ac.uk), while The Office for National Statistics (NOMIS) was used to download the relevant census data in CSV format by selecting 2021 census, proceeding to the table finder and choosing your relevant variables, while also selecting “local authorities: district/unitary (prior to April 2023”) for “Type of area”, also selecting the format type to be CSV and also including the area codes then proceeding to download the data.

**Data Preparation**

The Covid-19 Deaths which had the values for total death in England had to be updated as it had district codes and names of the 2018 version, while the 2021 census has the 2023 version whereby some district names and codes have been merged to form new ones. To make the data form different sources compatible, some rows in the Covid-19 deaths had to be merged and changed accordingly to make the data a total of 296 districts in England. After data reorganization and simplification on Excel, some SQL Queries was done using SQL lite to update the Covid-19 deaths district code and names, verify that the names and codes are accurate and also merge all the Variables.

The census data for Ages after collection was then grouped into broader categories to simplify the data and reduce dimensionality which would help in the exploratory data analysis and modelling. The categories merged are “Children”, “Pre-Teens”, “Adults”, “Middle Aged” and “Elderly”.

All available data was then standardized to reflect per thousand of the population and this was done using the total population per district gotten from the data containing all Ages using the VLOOKUP function in Excel to call and match the population data for each district using the La\_code.

Using the SQL lite to run some [SQL queries](#_SQL_Queries) to retrieve and combine the data, Covid-19 deaths, Age, Religion, Ethnic Group, Highest level of Qualification and Method of travel to work. The final data included all 37 variables from 5 themes with the total death in the Covid-19 deaths. This was then exported as a CSV file to operate in R studio.

**Data Exploration and Analysis**

The Final Standardized Data Set of Covid-19 Deaths with all the variables was imported into R and the “Data.Frame” function was able to select the dependent and independent variables from the columns in the data.

**Boxplots, Histograms and Q-Q Plots:**

* A Boxplot is a visual summary of a dataset's distribution, displaying its central tendency, variability, and potential outliers. It consists of a rectangular box representing the interquartile range (IQR), spanning from the first quartile (Q1) to the third quartile (Q3), with a line inside the box marking the median. Whiskers extend from the box to the smallest and largest data points within 1.5 × IQR from Q1 and Q3, while outliers beyond this range are plotted individually. Boxplots are advantageous for quickly identifying the spread, skewness, and outliers in a dataset, and they are particularly useful for comparing distributions across multiple groups.
* A **Histogram** is a bar chart that visualizes the frequency distribution of a dataset by dividing the data range into intervals called bins. The height of each bar represents the number of observations within its corresponding bin, allowing the shape, central tendency, and spread of the data to be observed. Histograms are especially beneficial for identifying the overall distribution of data, such as whether it is normal, skewed, or multimodal. They are intuitive, effective for large datasets, and allow for insights into data density and gaps.
* A **Q-Q plot** is a scatter plot used to compare the quantiles of a sample dataset with the quantiles of a theoretical reference distribution, such as a normal distribution. It plots the sample quantiles on the y-axis against the theoretical quantiles on the x-axis, with a diagonal reference line indicating perfect alignment between the two distributions. Deviations from the line highlight discrepancies such as skewness, heavy tails, or other departures from the assumed distribution. Q-Q plots are powerful tools for assessing whether data follow a specific distribution, validating assumptions in statistical models, and detecting subtle deviations in data.

For all the variables in the dataset a normality test and central tendency of variables was carried out using Boxplots, Histogram and Q-Q plots.

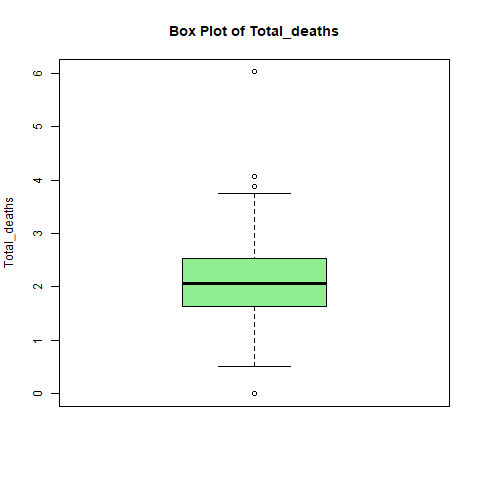


Figure 1: Boxplot of the dependent variable

The boxplot shows the distribution of **Total\_deaths**, with most values concentrated within the interquartile range (IQR). The median is slightly closer to the upper quartile, suggesting a mild skewness toward lower values. The whiskers indicate the typical range of the data, while a few outliers above and below the whiskers highlight unusual cases with significantly higher or lower total deaths. Overall, the plot summarizes the central tendency, variability, and presence of extreme values in the dataset.

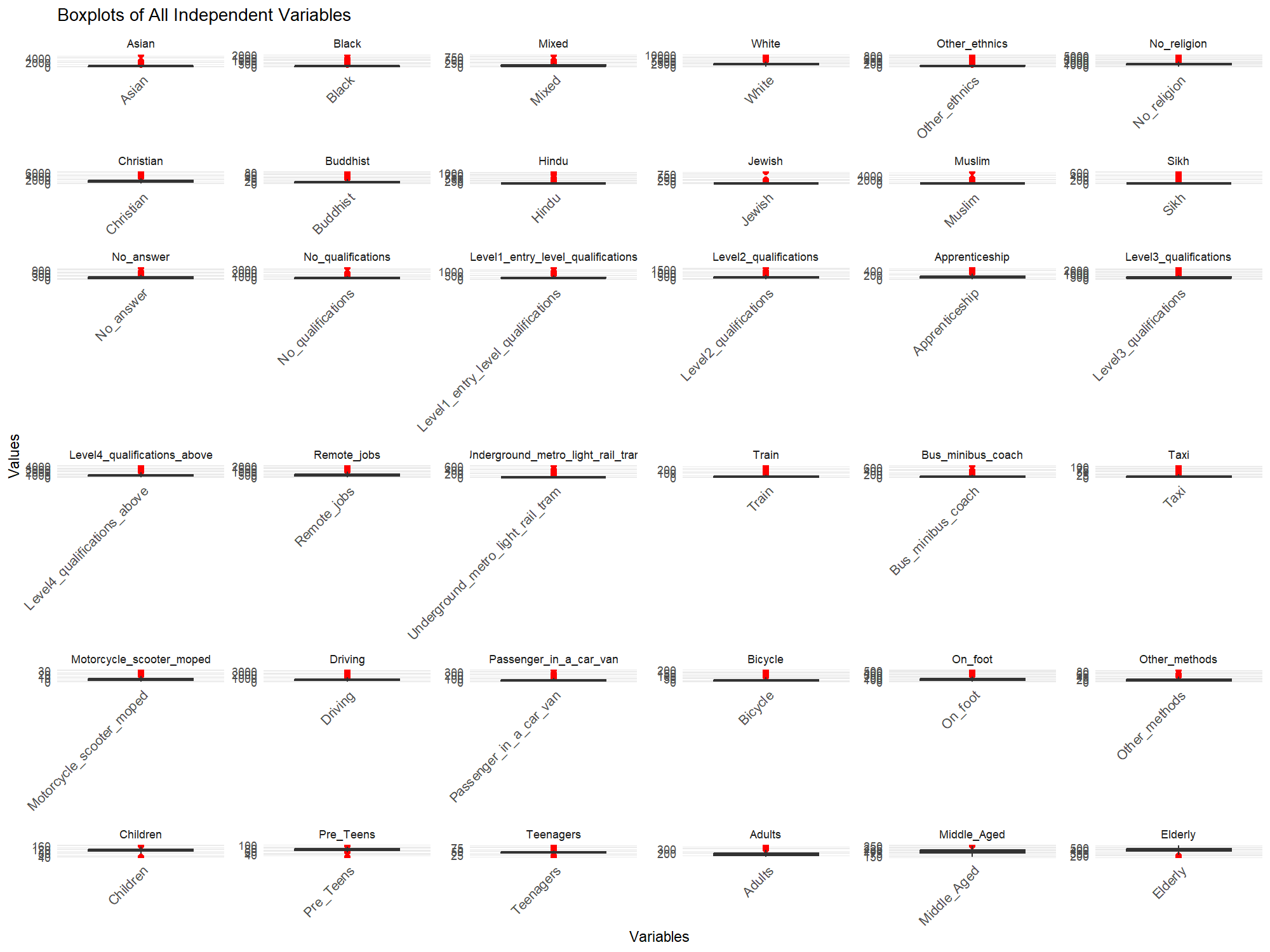


Figure 2 Boxplot of all the independent variables.

The boxplots summarize the distribution of all independent variables in the dataset. Key observations include:

* **Outliers**: Many variables show outliers, such as Asian, Remote\_jobs, and Children.
* **Skewness**: Several variables are skewed, indicated by asymmetric boxplots (e.g., No\_religion, Driving).
* **Variability**: Some variables (e.g., Christian, Adults, Middle\_Aged) show low variability, while others (e.g., Driving, No\_religion) have high variability.

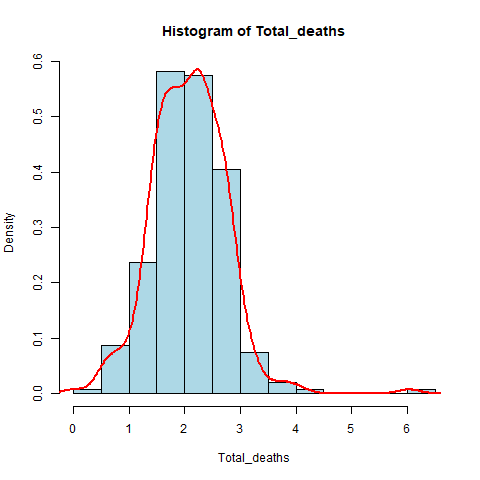


Figure 3: Histogram showing the Density of normal distribution of Total\_deaths.

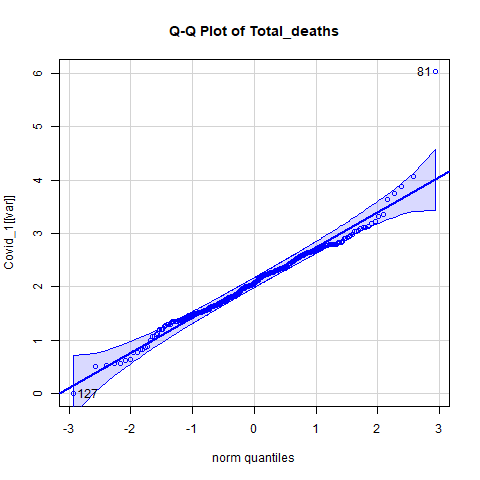


Figure 4: Q-Q plot of Total\_deaths

**KOLMOGOROV-SMIRNOV AND SHAPIRO TEST FOR NORMALITY**

The Kolmogorov-Smirnov (K-S) Test and Shapiro-Wilk Test assess whether a dataset follows a normal distribution. A p-value below 0.05 leads to rejecting the null hypothesis, indicating the data does not follow a normal distribution. Conversely, a p-value above 0.05 suggests we fail to reject the null hypothesis, implying the data likely comes from a normally distributed population. In this instance, with a p-value of 0.551, we do not reject the null hypothesis and conclude that the data is likely drawn from a normally distributed population.

|  |
| --- |
| Variable KS\_p\_value KS\_Normal SW\_p\_value SW\_Normal  1 Asian 4.386225e-23 No 7.526706e-28 No  2 Black 7.584262e-25 No 1.341456e-27 No  3 Mixed 7.945340e-14 No 6.829363e-23 No  4 White 1.389456e-09 No 6.002839e-20 No  5 Other\_ethnics 1.032394e-20 No 1.916331e-26 No  6 No\_religion 3.288499e-08 No 3.575608e-19 No  7 Christian 1.408722e-08 No 7.293751e-19 No  8 Buddhist 6.739443e-11 No 6.213638e-21 No  9 Hindu 4.713675e-28 No 8.041448e-30 No  10 Jewish 7.184948e-43 No 1.112223e-33 No  11 Muslim 1.545434e-28 No 1.894818e-29 No  12 Sikh 5.381472e-34 No 5.911971e-31 No  13 No\_answer 6.116580e-08 No 2.448411e-19 No  14 No\_qualifications 6.349641e-08 No 3.568366e-21 No  15 Level1\_entry\_level\_qualifications 5.820663e-07 No 1.189537e-19 No  16 Level2\_qualifications 2.354814e-07 No 6.063405e-20 No  17 Apprenticeship 9.577314e-09 No 1.678273e-19 No  18 Level3\_qualifications 1.269206e-07 No 9.011876e-21 No  19 Level4\_qualifications\_above 1.466927e-07 No 3.371287e-18 No  20 Remote\_jobs 4.311853e-06 No 1.504666e-17 No  21 Underground\_metro\_light\_rail\_tram 9.279413e-40 No 2.838255e-30 No  22 Train 1.930367e-16 No 4.391385e-25 No  23 Bus\_minibus\_coach 4.205374e-15 No 3.718357e-24 No  24 Taxi 4.273718e-14 No 1.498626e-24 No  25 Motorcycle\_scooter\_moped 2.352837e-09 No 1.763669e-19 No  26 Driving 5.157530e-10 No 1.246571e-19 No  27 Passenger\_in\_a\_car\_van 3.159819e-10 No 9.030536e-20 No  28 Bicycle 2.823171e-10 No 7.105959e-22 No  29 On\_foot 2.524342e-06 No 1.935876e-19 No  30 Other\_methods 5.373014e-08 No 2.098694e-18 No  31 Children 9.526494e-01 Yes 3.515333e-03 No  32 Pre\_Teens 1.353152e-01 Yes 2.309163e-07 No  33 Teenagers 2.087461e-11 No 1.304996e-19 No  34 Adults 5.580359e-07 No 1.962071e-16 No  35 Middle\_Aged 1.871418e-01 Yes 2.224928e-03 No  36 Elderly 5.793021e-01 Yes 6.461375e-02 Yes |

Most of the variables in the dataset do not follow a normal distribution, as indicated by p-values less than 0.05 in both the Kolmogorov-Smirnov (K-S) and Shapiro-Wilk (S-W) tests. Only a few variables, such as "Children," "Middle\_Aged," and "Elderly," show p-values greater than 0.05, suggesting they may be normally distributed. The Shapiro-Wilk test is more sensitive than the K-S test, identifying normality deviations in more variables. Overall, non-parametric methods or transformations may be necessary for most variables, while a few can be analyzed using parametric methods.

**Spearman Correlation**

The dependent variable was defined and a spearman correlation test was conducted to check the correlation between the independent variables and the dependent variables.

|  |
| --- |
| Variable Correlation  1 Asian 0.016164705  2 Black 0.008545485  3 Mixed -0.085422286  4 White -0.049147569  5 Other\_ethnics -0.033507517  6 No\_religion -0.125220072  7 Christian 0.002983457  8 Buddhist -0.193240543  9 Hindu -0.031923282  10 Jewish -0.155498061  11 Muslim 0.057790380  12 Sikh 0.014328119  13 No\_answer -0.140942907  14 No\_qualifications 0.154714424  15 Level1\_entry\_level\_qualifications 0.088210749  16 Level2\_qualifications 0.052253193  17 Apprenticeship 0.099112702  18 Level3\_qualifications -0.058860643  19 Level4\_qualifications\_above -0.193862037  20 Remote\_jobs -0.218693012  21 Underground\_metro\_light\_rail\_tram 0.108003020  22 Train 0.097535339  23 Bus\_minibus\_coach 0.033825444  24 Taxi 0.235376158  25 Motorcycle\_scooter\_moped -0.159034981  26 Driving 0.094325138  27 Passenger\_in\_a\_car\_van 0.161150660  28 Bicycle -0.295341531  29 On\_foot -0.135836654  30 Other\_methods -0.021117642  31 Children 0.202961342  32 Pre\_Teens 0.202283558  33 Teenagers 0.004108949  34 Adults -0.029085662  35 Middle\_Aged -0.097140472  36 Elderly 0.071256510 |

The table shows a mix of weak, moderate, and strong correlations. Variables like **Children** (0.203) and **Pre\_Teens** (0.202) have strong positive correlations, while **Bicycle** (-0.295) and **Remote\_jobs** (-0.219) show notable negative correlations. Most variables, such as **Asian** (0.016) and **Black** (0.009), have weak correlations, indicating little relationship.



Figure 5 : correlation matrix plot

**FACTOR ANALYSIS**

**This** is a statistical method used to identify underlying relationships between observed variables by grouping them into fewer, unobserved variables called **factors**. It helps reduce the dimensionality of data, simplifying complex datasets by identifying patterns or structures in the data. There are two main types which are Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). Dimensionality reduction techniques simplify datasets by preserving key features and relationships, while removing less important information. This makes the data easier to manage, improves analysis efficiency, and offers clearer insights into underlying patterns.

**KMO Test**

Variables with a correlation greater than 0.2 with the dependent variable were chosen for principal component analysis. Before conducting a KMO test on these variables, partial correlation will be applied, ensuring there are no missing or excluded values.

|  |
| --- |
| Kaiser-Meyer-Olkin factor adequacy  Call: KMO(r = pca\_data)  Overall MSA = 0.57  MSA for each item =  Remote\_jobs Taxi Bicycle Children Pre\_Teens  0.60 0.84 0.52 0.54 0.51 |

The overall Kaiser-Meyer-Olkin (KMO) measure is 0.57, indicating moderate adequacy for factor analysis. Individual MSAs show that **Taxi** (0.84) is highly suitable, while **Remote\_jobs** (0.60) is adequate. Variables like **Bicycle** (0.52), **Children** (0.54), and **Pre\_Teens** (0.51) have marginal adequacy, suggesting they may not be ideal for factor analysis.

**Eigen Values**

|  |
| --- |
| [1] 2.2776730 1.7567076 0.5555737 0.2978266 0.1122192 |



Figure 6: Cumulative Scree Plot of Eigen Values

From the Cumulative Scree Plot above, it appears that the cumulative variance explained begins to level off around 3 factors. This implies that selecting 3 factors would capture a significant portion of the variance in my dataset while minimizing the inclusion of less informative factors.

**PRINCIPAL COMPONENTS ANALYSIS**

|  |
| --- |
| Principal Components Analysis  Call: principal(r = pca\_data, nfactors = 4, rotate = "varimax", scores = TRUE)  Standardized loadings (pattern matrix) based upon correlation matrix  RC1 RC2 RC3 RC4 h2 u2 com  Remote\_jobs 0.07 0.37 0.22 0.90 1.00 4.2e-04 1.5  Taxi 0.23 0.14 0.94 0.20 1.00 5.7e-05 1.3  Bicycle -0.09 0.92 0.14 0.34 1.00 4.0e-03 1.3  Children 0.95 0.06 0.18 0.05 0.95 5.2e-02 1.1  Pre\_Teens 0.95 -0.16 0.12 0.03 0.94 5.5e-02 1.1  RC1 RC2 RC3 RC4  SS loadings 1.88 1.04 1.01 0.97  Proportion Var 0.38 0.21 0.20 0.19  Cumulative Var 0.38 0.58 0.78 0.98  Proportion Explained 0.38 0.21 0.21 0.20  Cumulative Proportion 0.38 0.60 0.80 1.00  Mean item complexity = 1.3  Test of the hypothesis that 4 components are sufficient.  The root mean square of the residuals (RMSR) is 0.02  with the empirical chi square 2 with prob < NA  Fit based upon off diagonal values = 1 |

The PCA results display factor loadings for four components (RC1, RC2, RC3, RC4) derived from the correlation matrix**. Children** and **Pre\_Teens** load strongly on RC1 (0.95), **Bicycle** on RC2 (0.92), **Taxi** on RC3 (0.94), and **Remote\_jobs** on RC4 (0.90). Together, the components explain **98%** of the variance, with RC1 contributing **38%**. The model shows a good fit, indicated by a low **RMSR of 0.02** and a non-significant chi-square test, suggesting the four components effectively capture the data structure.

**Multiple linear Regression Modelling**

A Multiple Linear regression model was applied using the same filtered variables used to conduct the principal components analysis.

|  |
| --- |
| Call:  lm(formula = dependent\_var ~ ., data = final\_data)  Residuals:  Min 1Q Median 3Q Max  -2.1760 -0.3741 0.0010 0.3667 3.7452  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 1.5462206 0.3038728 5.088 6.49e-07 \*\*\*  Remote\_jobs -0.0003433 0.0001635 -2.099 0.036635 \*  Taxi 0.0108934 0.0032303 3.372 0.000846 \*\*\*  Bicycle -0.0050964 0.0016304 -3.126 0.001952 \*\*  Children 0.0064695 0.0027787 2.328 0.020583 \*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 0.6143 on 291 degrees of freedom  Multiple R-squared: 0.1522, Adjusted R-squared: 0.1406  F-statistic: 13.06 on 4 and 291 DF, p-value: 8.493e-10 |

Out of the initial 4 variables, the regression model results indicate that Remote\_jobs, Taxi, Bicycle, and Children all have significant impact on the dependent variable. While Taxi shows the strongest positive influence, while Bicycle and Remote\_jobs have negative effects. The model accounts for 14.06% of the adjusted R-squared (R2 = 0.141) and demonstrates a good fit, supported by a significant F-statistic of 13.06 and a p-value of <8.493e-10. That is the overall model is statistically significant.

Checking for Multicollinearity among variables:

|  |
| --- |
| Remote\_jobs Taxi Bicycle Children  2.126547 1.460632 1.895015 1.230570  > sqrt(vif(model\_1)) > 2 # if > 2 vif too high  Remote\_jobs Taxi Bicycle Children  FALSE FALSE FALSE FALSE |

All the independent variables have VIF values not > 2 (indicated by “FALSE”), therefore no significant multicollinearity in the model. This suggest that the predictors are highly correlated, and their regression coefficients are reliable.



Figure 7: Residuals vs Leverage plot of optimal regression model

The "Residuals vs Leverage" plot shows that most points have low leverage and residuals close to 0, indicating a good fit. However, points 81, 127, and 25 stand out as potential outliers or influential observations. These may indicate unusual values in predictors like Remote\_jobs, Taxi, Bicycle, or Children, and should be investigated further to ensure model reliability.

# Conclusion and Discussion

This study aimed to identify key risk factors for COVID-19 deaths using 2021 census data from the Office for National Statistics. Data was imported into SQLite, merged, and analyzed in R. After an initial exploration to select variables and check normality, factor analysis reduced the independent variables to key components. Multiple regression models were tested to examine relationships with COVID-19 death rates. The final model was chosen based on accuracy, significant variables, low internal correlations, and well-distributed residuals, highlighting the complex and multi-factorial nature of these risk factors.

The study concludes that, although other variables should not be dismissed the variables such as Remote\_job, Children, Bicycle, and Taxi standout as the most significant contributors to Covid-19 deaths in England.

**Taxi**: It involves close, enclosed contact between passengers and drivers, making them high-risk spaces for virus transmission. Limited ventilation and frequent use by multiple passengers could have contributed to spreading the virus. This result highlights the role of taxis in facilitating transmission, especially in urban areas where taxis are commonly used for short-distance travel. This positive coefficient suggests that greater use of taxis is associated with an increase in the COVID-19 mortality rate.

**Bicycle:** This represents the percentage of people using bicycle as a method of transport to work. It was highly significant in all the model with a strong negative correlation coefficient, indicating that higher transportation usage is associated with fewer Covid-19 deaths. A possible reason is that cycling may promote better health and boost the immune system which reduces vulnerability to severe Covid-19 outcomes.

**Children:** While children were generally less affected by severe COVID-19, they could act as asymptomatic carriers, bringing the virus into households. Families with more children may have faced greater difficulties in maintaining strict isolation and social distancing, thereby increasing the risk of transmission to vulnerable adults**.** This positive coefficient suggests that regions with a higher proportion of children had slightly higher COVID-19 mortality rates.

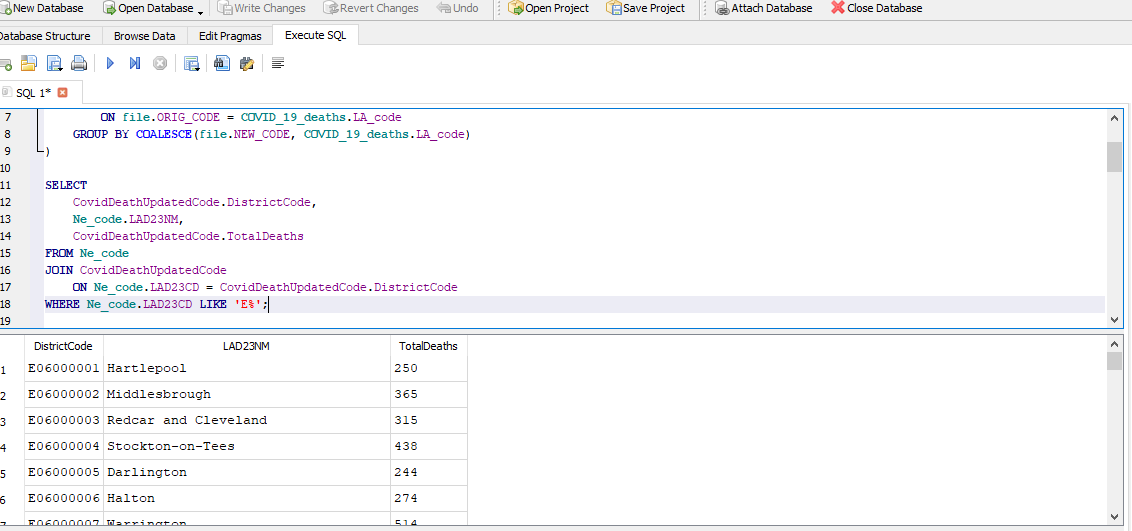
**Remote\_jobs:** It reduced the need for commuting and in-person interactions, lowering the transmission risk. People working from home were less likely to be exposed to the virus compared to those who had to travel or work in physical settings. This negative coefficient indicates that an increase in the availability of remote jobs is associated with a slight decrease in the COVID-19 mortality rate.

While the model explains a modest portion of the variation in COVID-19 mortality rates, it underscores the importance of transport behavior and demographic factors in the spread of the virus. These insights can inform future strategies for minimizing transmission, particularly through transport policy and remote working incentives.

# APPENDIX

# SQL Queries

The SQL queries used in merging all the independent variable with the Covid-19 data downloaded from NOMIS (all after standardization to per thousand population), also updating the district names and codes while creating a subset table in the SQL are included below;



# References

Bhala, N., Curry, G., Martineau, A. R., Agyemang, C., and Bhopal, R. (2020) ‘Sharpening the global focus on ethnicity and race in the time of COVID-19’, *The Lancet*, 395(10238), pp. 1673-1676.

Bowleg, L. (2020) ‘We’re not all in this together: On COVID-19, intersectionality, and structural inequality’, *American Journal of Public Health*, 110(7), pp. 917-918.

Davies, N. G., Klepac, P., Liu, Y., Prem, K., Jit, M., and Eggo, R. M. (2020) ‘Age-dependent effects in the transmission and control of COVID-19 epidemics’, *Nature Medicine*, 26(8), pp. 1205-1211.

Devakumar, D., Shannon, G., Bhopal, S. S., and Abubakar, I. (2020) ‘Racism and discrimination in COVID-19 responses’, *The Lancet*, 395(10231), p. 1194.

Falkingham, J., Evandrou, M., Qin, M., and Vlachantoni, A. (2020) ‘Inequalities in healthcare use among older people during the COVID-19 pandemic: A study of the English Longitudinal Study of Ageing’, *BMJ Open*, 10(11).

Katikireddi, S. V., Hainey, K. J., and Vasileiou, E. (2021) ‘The impact of COVID-19 on different population subgroups’, *Public Health Reviews*, 42.

Kramer, S. M. (2020) ‘Religion and COVID-19: Impact and lessons learned’, *Journal of Religion and Health*, 59(5), pp. 2073-2084.

Marmot, M., Allen, J., Goldblatt, P., Boyce, T., and McNeish, D. (2020) ‘Building back fairer: The COVID-19 Marmot review’, *The Health Foundation*. Available at: https://www.health.org.uk/publications/reports/building-back-fairer-the-covid-19-marmot-review (Accessed: 6 January 2025).

Pew Research Center (2020) ‘How religion influences COVID-19 responses around the world’. Available at: [https://www.pewresearch.org](https://www.pewresearch.org" \t "_new) (Accessed: 6 January 2025).

Tirachini, A. and Cats, O. (2020) ‘COVID-19 and public transportation: Current assessment, prospects, and research needs’, *Journal of Public Transportation*, 22(1), pp. 1-21.

Verity, R., Okell, L. C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., and Ferguson, N. M. (2020) ‘Estimates of the severity of coronavirus disease 2019: A model-based analysis’, *The Lancet Infectious Diseases*, 20(6), pp. 669-677.

Wildman, W. J. (2020) ‘Religious ecology and the COVID-19 pandemic’, *Zygon*, 55(3), pp. 611-623.