**Grid Emissions Per Borough:**

<https://data.london.gov.uk/dataset/london-atmospheric-emissions-inventory--laei--2016>

This dataset presents the total gas emissions of NOx, PM10, PM2.5 and CO2 in tonnes per year, for London boroughs in the years 2010, 2013, and 2016 **(London Datastore, a, 2021)**. This data was used to indicate air pollution per borough, informed by studies conducted by **Jorgenson et al. (2021) and Afoakwah et al. (2020),** as outlined in our literature review**.** We decided to use the 2016 values, as they were the most recent, and subsequently, the most representative of our contemporary context.

The dataset is available as an Excel file (XLSX format), and is thereby structured, although this requires a Microsoft licence - proprietary software - to be read. Nonetheless, the data may be converted to CSV, and consequently read by Python without the aid of other packages. The data is linked to other datasets on the site for research, indicating that it is of high-quality.

Furthermore, this data was collected by **TFL Air Quality and the Greater London Authority, and was** acquired by us from the **London Datastore (a, 2021)**. It is thus open data - freely accessible for public (re-)use, and may be combined with other information, as we do in our research project, so long as its source is attributed.

*Contains public sector information licensed under the Open Government Licence v2.0.*

---------

**Land Area and Population Density of Each Borough:**

<https://data.london.gov.uk/dataset/land-area-and-population-density-ward-and-borough>

This dataset presents the total land area and population density of each London borough in square kilometres, as of 2018 **(London Datastore, b, 2021)**. This dataset was used to scale the total gas emissions and noise pollution data, to take into consideration that the size of boroughs may affect their gas emissions and number of noise complaints.

The dataset is available as both CSV and Excel (XLSX format) files, and is thereby machine-readable. Furthermore, the data is comprehensive, structured and clean. This data was collected by **the Greater London Authority, and was** acquired by us from the **London Datastore (b, 2021)**. It is under Crown Copyright (Open Government Licence), and thus open data - freely accessible for public (re-)use, and may be combined with other information, as we do in our research, so long as its source is attributed.

*Contains public sector information licensed under the Open Government Licence v2.0.*

------

**Land Use of Each Borough:**

<https://data.london.gov.uk/dataset/land-use-ward>

We have collected this dataset to derive the total proportion of land in a London borough that is green space - one of the variables chosen to represent environmental factors, informed by **Salgado et al.'s (2020) research.**

The dataset is available as both CSV and Excel (XLSX format) files, and is subsequently machine-readable, indicating it is of good quality. Furthermore, the dataset is clearly structured, and comprehensive, with no gaps in data collected. One key limitation, however, is that this data is from 2005, and thus may not be an accurate representation of current land use and green space per London borough. Nonetheless, it is still the most recent dataset of its kind available, and was useful for our investigation.

This data was collected by **the Ministry of Housing, Communities and Local Government, and was** acquired by us from the **London Datastore (c, 2021)**. It is open data - freely accessible for public (re-)use, and may be combined with other information, so long as its source is attributed.

*Contains public sector information licensed under the Open Government Licence v2.0.*

-------

**Public Health Outcomes Framework Indicators:**

<https://data.london.gov.uk/dataset/public-health-outcomes-framework-indicators>

This dataset presents the total number of noise complaints in a borough, or local authority, per 1,000 people (**London Datastore, d, 2021)**. We used this data to indicate noise pollution - one of the variables chosen to represent environmental factors as a whole - informed by **Murphy and King’s (2014)** research. We further used this dataset for two indicators of health status - self-reported well-being score for low happiness, and self-reported well-being score for anxiety amongst individuals across boroughs.

The dataset is available as an Excel file (XLSX format), which requires proprietary software - Microsoft - to be read. However, the data may be converted to a CSV format, and thus read by Python. Furthermore, this dataset is clearly structured, and collates values for many different, linked Public Health indicators together, besides noise complaints and self-reported well-being scores, indicating it is of good-quality for research. However, one significant limitation of this dataset is that the the values for the City of London and Hackney (for the noise complaints information) have been combined (**London Datastore d, 2021)**. Consequently, as outlined in our methodology, we had to derive different data points for Hackney and the City of London individually, which may have introduced inaccuracies. Nevertheless, this dataset is the highest quality one of its kind available, and was useful for our investigation.

Comparatively, another dataset we considered for the noise pollution variable was by **Cirrus Research (2017).** However, whilst this is more comprehensive and presents individual values for all boroughs, the data was only available as an infographic - closed data - and needed to be typed manually in Excel. Moreover, whilst they acknowledge the data was gathered from government sites via Freedom of Information requests, the appropriate links and references are not given (**Cirrus Research 2017)**. Hence, this dataset was discarded, in favour of the one we selected.

Indeed, the dataset we selected was collected by **Public Health England, and** acquired by us from the **London Datastore (d, 2021)**. It is open data - freely accessible for public (re-)use, and may be combined with other information, so long as its source is attributed.

*Contains public sector information licensed under the Open Government Licence v2.0.*

-----------

**Trust for London: Life Expectancy per Borough**

<https://www.trustforlondon.org.uk/data/life-expectancy-borough/>

This dataset presents the life expectancy and healthy life expectancy of both males and females across the 32 official London boroughs (not including City of London), measured between 2017 to 2019 **(Trust for London, a, 2021)**. Both life expectancy and healthy life expectancy were used as indicators to represent overall health status, informed by **Demakakos et al. (2008) and Braveman and Gottlieb’s (2014**) research.

From the website, the dataset is available to download as a CSV file, and thereby may be read by Python without the aid of other packages, indicating it is of good quality. Moreover, the data is comprehensive and clearly presented. This data was collected by the ONS - **Office for National Statistics (2021)** - and was acquired by us from the **Trust for London (a, 2021).** It is under Crown Copyright (Open Government Licence), and is thus open data - freely accessible for public (re-)use, and may be used so long as its source is attributed.

*Contains public sector information licensed under the Open Government Licence v3.0.*

---------------

**City of London Health Profile**

<https://democracy.cityoflondon.gov.uk/documents/s84263/Appendix%202-%20City%20of%20London%20Health%20Profile%202017%203.pdf>

We have collected this dataset in order to obtain values for the life expectancy of men and women in the City of London, as this was not included in the **Trust for London (a, 2021)** dataset. The data is from 2017 **(Public Health England, 2017)**. Limiting the quality of the data is the fact that it is stored as a PDF file, and thus the values had to be manually typed on Excel. Nonetheless, as there were only two values, and no other datasets were readily available, this source proved valuable for our research.

This data was collected and obtained from **Public Health England (2017).** It is thus open data - freely accessible for public (re-)use, and may be combined with other information, as we do in our investigation, so long as its source is attributed.

*Contains public sector information licensed under the Open Government Licence v3.0.*

-------

**Earnings per Head per Borough**

<https://data.london.gov.uk/dataset/earnings-place-residence-borough>

This dataset presents the total earnings per head per borough. We have collected this dataset as one of three indicators used to represent the income variable as a whole, informed by the **Greater London Authority (2008),** as detailed in our literature review. We are using the values for total weekly earnings per head in 2018, as this is recent (before the pandemic however, to avoid accounting changes for that), and includes data for the City of London.

The dataset is available as an Excel file (XLSX format), and is machine-readable, though this requires a Microsoft licence - proprietary software. Nonetheless, the data may be converted to CSV, and consequently read by Python.

This data was collected by theOffice for National Statistics (ONS)**,** and was acquired by us from the **London Datastore (e, 2021)**. It is under Crown Copyright (Open Government Licence), and is thus open data - freely accessible for public (re-)use, and may be combined with other information, so long as its source is attributed.

*Contains public sector information licensed under the Open Government Licence v2.0.*

--------

**Cost of Living**

<https://www.numbeo.com/cost-of-living/>

We used this website to obtain values for the cost of living in different London boroughs **(Numbeo, 2022)**. One limitation of this data is that the website is not structured, and Alexia had to manually search for boroughs, then copy and paste values onto an Excel file, which was time-consuming. Furthermore, data for all boroughs was not available on this site, so certain estimations had to be completed - as will be discussed when mentioning our methodological limitations. Moreover, this website is crowd-sourced, and anyone can contribute, which significantly restricts the reliability of data. Nonetheless, despite these clear limitations, we thought it would be informative to include cost of living as an indicator for income in our investigation, and this data was freely accessible online. In terms of licensing for academic purposes, a link back to Numbeo.com is sufficient.

--------

**London Borough Profiles**

<https://data.london.gov.uk/dataset/london-borough-profiles>

This dataset includes multiple demographic, social, and economic indicators across all London boroughs or local authority districts (terms used interchangeably here), including the City of London (**London Datastore, f, 2021)**. We have used this dataset to collect values for the percentage of Black, Asian and Minority Ethnic (BAME) groups in each borough as of 2016, in addition to the new 2015/2016 migrant rates, in order to calculate our Migration and Race Score - further discussed in our methodology section - as one of the three socioeconomic variables. Furthermore, we have also used this dataset for the 2015 employment rates per borough (one of the indicators used to represent income - a socioeconomic variable we have chosen), in addition to the 2013/2014 GCSE attainment per borough. This is the percentage of students who obtained 5 or more A\*-C GCSE grades, including Mathematics and English, and is one of the indicators used to represent the socioeconomic variable of educational attainment.

The dataset is available as both CSV and Excel (XLSX format) files, and is thereby machine-readable. Furthermore, the data is comprehensive, and clearly structured.This data was collected by theGreater London Authority**,** and was acquired by us from the **London Datastore (e, 2021)**. It is under Crown Copyright (Open Government Licence), and is thus open data - freely accessible for public (re-)use, and may be combined with other information, so long as its source is attributed.

*Contains public sector information licensed under the Open Government Licence v2.0.*

----------

**A Level Results by Borough**

<https://data.london.gov.uk/dataset/level-results-gender-and-place-residence-borough>

This dataset was used for the Average Point Score per student (in terms of A Level/Level 3 attainment) across London boroughs, including the City of London (**London Datastore, g, 2021)**. The dataset includes values for the school years 2004/2005 to 2013/2014, the latter of which we chose to include, do its recency. We have collected this information as one of the indicators to represent the socioeconomic variable of educational attainment.

The dataset is solely available as an Excel (XLSX format) file, and hence requires a Microsoft licence to be read. By converting it to a CSV file, the data may further be read by Python.This data was collected by theDepartment of Education,and was acquired by us from the **London Datastore (g, 2021)**. It is under Crown Copyright (Open Government Licence), and is thus open data - freely accessible for public (re-)use, and may be combined with other information, so long as its source is fully attributed.

*Contains public sector information licensed under the Open Government Licence v2.0.*

----------

**School Type per Borough**

<https://data.london.gov.uk/dataset/schools-and-pupils-type-school-borough>

This dataset presents the proportion of state schools and independent schools in each borough, including City of London, from the years 2011 to 2019 (**London Datastore, h, 2021)**. We used this as an indicator of education, one of our chosen socioeconomic variables.

The dataset is available as an Excel (XLSX format) file, and hence requires a Microsoft licence to be read, though it may also be read by Python, by converting it to a CSV file.This data was collected by theDepartment of Education,and was acquired by us from the **London Datastore (h, 2021)**. It is open data - freely accessible for public (re-)use, and may be combined with other information, so long as its source is attributed.

*Contains public sector information licensed under the Open Government Licence v2.0.*

-------------

**2011 Census: Health and Social Care**

<https://data.london.gov.uk/dataset/2011-census-health-care>

This dataset presents the percentage of healthy residents in each borough, including City of London (**London Datastore, i, 2021)**. We used this data to estimate the healthy life expectancy of City of London residents, as outlined in our methodology section.

The dataset is available as an Excel (XLSX format) file, and hence requires a Microsoft licence to be read, though it may also be read by Python, by converting it to a CSV file.This data was collected as part of the Census Information Scheme,and was acquired by us from the **London Datastore (i, 2021)**. It is open data - freely accessible for public (re-)use, and may be combined with other information, so long as its source is attributed.

*Contains public sector information licensed under the Open Government Licence v2.0.*

---------

**General Discussion of the Data**

The datasets that we have obtained were mostly collected by public, government agencies such as Public Health England, the Greater London Authority, and the Office for National Statistics, and are available from the London Datastore. Subsequently, in addition to being freely accessible, the datasets are less likely to be biassed than privately published information, and collection methodologies are outlined.

However, it is important to note that government data still has potential for bias, and may be used to serve a political agenda. This is difficult to verify, as government sources are often the only investigations conducted into the variables we have chosen. This limits our scope to cross-reference data, and ensure that it is reliable.

Nevertheless, overall we are happy with the quality of our data sources, and by addressing limitations, we hope our analysis and any conclusions drawn are more contextualised.