MSc ESDA Title Page

UCL Candidate Code: **XFJF7**

Module Code: **BENV0092**

Module Title: **Energy Analytics in the Built Environment**

Coursework Title: **The Impact of Residential Building Characteristics on Building Energy Efficiency in Newham**

Module Leader: **Dr. Despina Manouseli**

Date: **15 January 2023**

Word Count: **2,118**

By submitting this document, you are agreeing to the Statement of Authorship:

I/We certify that the attached coursework exercise has been completed by me/us and that each and every quotation, diagram or other piece of exposition which is copied from or based upon the work of other has its source clearly acknowledged in the text at the place where it appears.

I/We certify that all field work and/or laboratory work has been carried out by me/us with no more assistance from the members of the department than has been specified.

I/We certify that all additional assistance which I/we have received is indicated and referenced in the report.

Please note that penalties will be applied to coursework which is submitted late, or which exceeds the maximum word count. Information about penalties can be found in your Course Handbook which is available on Moodle: https://moodle.ucl.ac.uk/course/view.php?id=29435&section=0

- ***Penalties for late submission***
- ***Penalties for going over the word count***

In the case of coursework that is submitted late and is also over length, then the greater of the two penalties shall apply. This includes research projects, dissertations and final reports.

I. Introduction

The UK's net-zero strategy is founded on a drive to reduce GHG emissions and delay/mitigate the exacerbation of climate change (BEIS, 2021). The strategies employed do not only target corporates and industrial activity but also residential activities – the Executive Summary begins with this local focus, "*From heating our homes to filling up our cars* […]" (BEIS, 2021). One strategy contemplates strengthening building regulations to increase buildings' energy efficiency, for instance by tightening minimum energy efficiency standards and tackling fuel poverty (BEIS, 2021). In particular, the Amended Buildings Regulations 2021 are expected to reduce 30% of buildings' carbon emissions compared to the 2013 Part L Standard (HM Government, 2022).

This essay studies the energy efficiency of residential dwellings and the features which affect it, including built forms, insulation, and household energy services (lighting, heating, and hot water). To ensure that the data analysis meets the localised/bottom-up strategy, this essay will take a targeted look at one London borough (Newham) to identify opportunities for strengthening building regulations. This essay utilises clustering techniques (Decision Trees, K-means clustering) to study the buildings' characteristics and energy services, and Ordinary Least Squares regression ("**OLS regression**") for the relationship between building characteristics and energy efficiency.

II. Literature Review

A. Residential GHG Emissions

Residential activities are a major UK GHG contributor (c.16%), especially since 2020 as the Covid-19 pandemic/associated lockdowns and colder temperatures increased GHG emissions by 7% in aggregate (BEIS, 2022). Even before 2020, residential GHG emissions were consistent, unlike other contributors which have been decreasing (Figure 1).

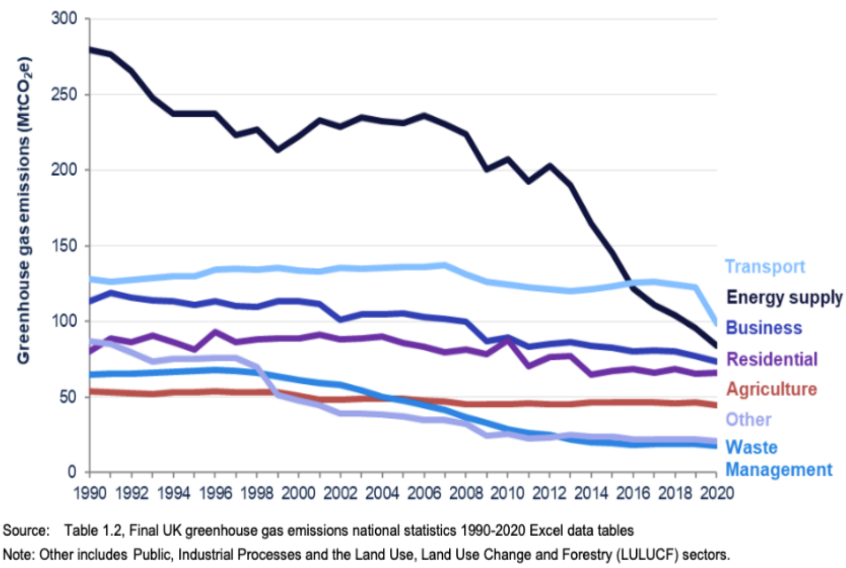

Figure 1 Territorial UK greenhouse gas emissions by sector (BEIS, 2022).

B. Energy Efficiency

Energy efficiency is key to reducing GHG emissions: the higher a dwelling's energy efficiency, the lower its energy requirements and therefore the lower its reliance on GHG-emitting energy production (Rosenow et al., 2018). Building energy efficiency is rated in standardised Energy Performance Certificates ("**EPCs**"), a concept stemming from EU legislation (Jenkins et al., 2017).

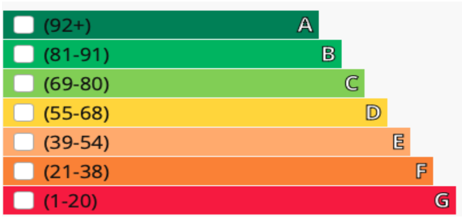

Figure 2 Energy Efficiency Ratings (epc.opendatacommunities.org, 2022).

In the UK, EPCs were first widely used as part of the Green Deal to assess loan eligibility, and are now considered in policies to reduce GHG emissions, including linking energy efficiency to stamp duty or solar panel feed-in-tariffs (Jenkins et al., 2017).

EPCs have been criticised for not being very robust: the standard assessment procedure used relies on assessors and has yielded variable/inconsistent results (Jenkins et al., 2017). The reduced data method for dwellings is particularly vulnerable as it uses more default assumptions and smaller samples (Crawley et al., 2019). Nevertheless, EPCs are frequently relied on (including in the net-zero strategy) as they are easy to comprehend and provide for simple but powerful benchmarks across different areas/jurisdictions (Jenkins et al., 2017).

C. Factors Impacting Buildings' Energy Efficiency

EPC ratings measure a matrix of factors, including built form, insulation, and service systems.

Firstly, built forms (property type, terracing) impact on efficiency, particularly due to the level of detachment and the implications this has on wall thickness (Chen et al., 2020).

Secondly, insulation can vary energy efficiency by c.85%, with material upgrades alone yielding 15% energy saving (Rosenow et al., 2018). The most effective insulators are location-/construction-driven, but typically wall and roof insulation are the most effective (Chen et al., 2020). Window-glazing is also a very effective insulator (Chen et al., 2020): hard-coat low-E triple-glazed windows can save 31% energy (Somasundaram et al., 2020).

Finally, household service systems (lighting, heating, and hot water) significantly impact on energy efficiency (Rosenow et al., 2018): thermal solar systems can reduce energy consumption for hot water by 78% (Chen et al., 2020).

III. Methodology

A. Dataset

The dataset for this study is from the Energy Performance of Building Data England and Wales published by the Department for Levelling Up, Housing & Communities (updated 22

November 2022), which includes EPCs issued up to 29 September 2022. Among the dataset, 123,802 residential buildings in the Newham borough of London were studied.

The dataset included several parameters relating to residential building characteristics, including built form, building type, insulation description, and emissions.

B. Data Analysis

This essay considers the relationship between built form, insulation, and service systems in the Newham residential dwellings and EPC rating.

To map the built form corresponding to each EPC rating, I cleaned the dataset and visualised it using a pie chart, bar chart and box plot. Using insulation and built types, I created a decision tree that classified each energy rating's building characteristics. I used clustering methods (unsupervised machine learning techniques) to visualise how buildings are grouped by the efficiency of different service systems (lighting, heating, and hot water). Furthermore, to see the relationship between EPC rating and insulation and each service systems, I used OLS regression.

1. *Decision Tree*

Decision tree methodology classifies large datasets with multiple covariates – ideally where there is no intricate parametric structure – and also addresses continued categorical variables (Song and Lu, 2015). Therefore, this method is suitable for categorising data in simple, mutually exclusive ways, and so was used to organise different built forms (including terracing and property types), and insulation characteristics (walls, windows, and roof).

Decision tree methodology classifies populations in hierarchical tree-systems, via root nodes, internal nodes, leaf nodes, and branches (Song and Lu, 2015). To ensure a clear classification in a decision tree, impurities need to be reduced (Song and Lu, 2015) – in this study, the Gini index was used.

2. *K-means Clustering*

K-means clustering forms groups by variables with similar characteristics, making clusters using calculated means (Sinaga and Yang, 2020). This was used to group the residential buildings by other energy service efficiency parameters. In particular, K-means++ was used to mine the data; the Euclidean distance was used to control the average's centre; and the steps were repeated until the convergence criteria were met (Sinaga and Yang, 2020; Na, Xumin and Yong, 2010).

As the variables were categorical, prior to clustering, the values were changed using One-Hot Encoding to be continuous to measure the distance between them. Moreover, the K-means clustering algorithm analyses the difference in calculating each variable's distance. To avoid any distortion, the scales were changed used the Standard scaler.

3. *OLS Regression*

OLS regression describes the relationship between independent variables and a dependent variable on linear regression and is very useful in showing how EPC ratings correlate to built forms, insulation, and energy services (Kilmer and Rodríguez, 2016). OLS Regression was used to find the relationship between built forms, insulation (of windows and walls), and energy services efficiency (of lighting, heating, and hot water).

IV. Analysis and Results

A. Preliminary Cleaning and Visualisation

Firstly, I categorised the built forms in Newham borough and compared them against their EPC rating. Figure 3 shows the percentage of each built form in Newham borough. The majority of building forms are terraces (72.4%).

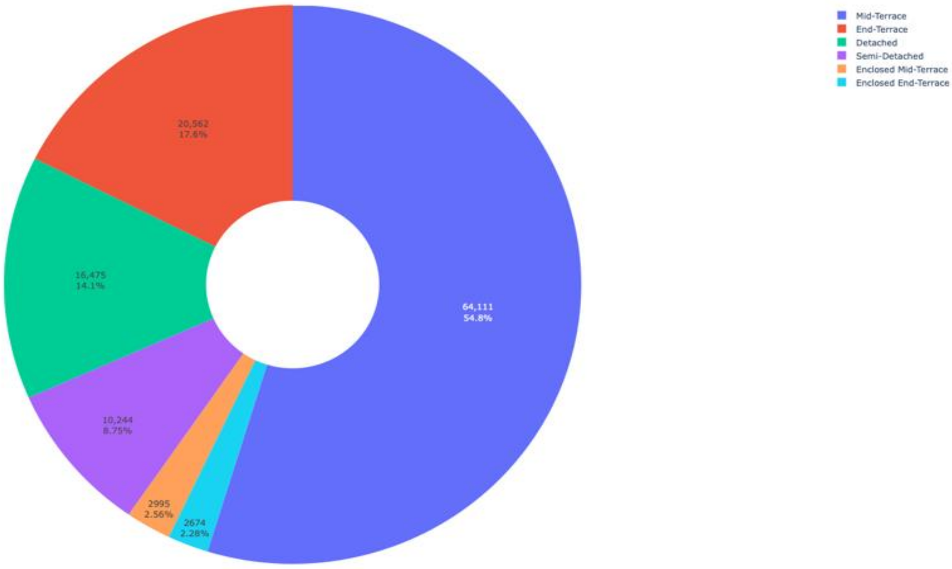

Figure 3 Domestic built forms (Newham).

As most built forms are terraces (especially mid-terraces) Figure 4 shows the largest group of residences comprises mid-terraces. Most residences have an EPC rating of D or C, with terrace forms being over-represented in these low rating: mid-terraces make up c.67% of D-rating dwellings, 53% of C-rating dwellings, but only 31% of B-rating dwellings.

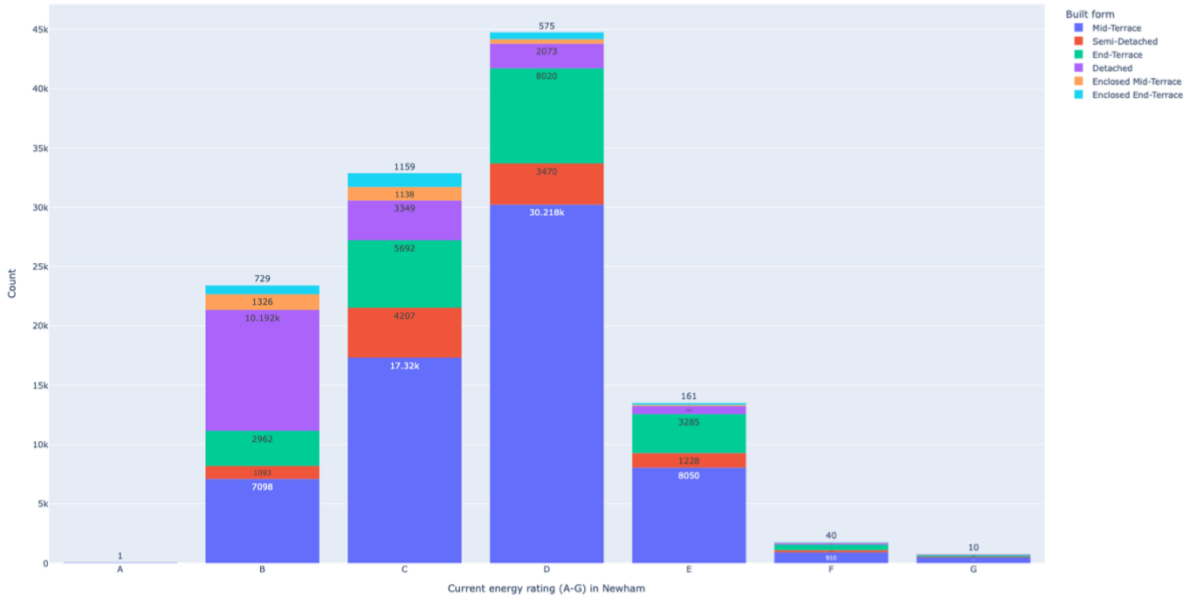

Figure 4 Energy rating by built form (Newham).

From Figure 5 it is also clear that flats are more energy efficient than other property types, especially when detached.

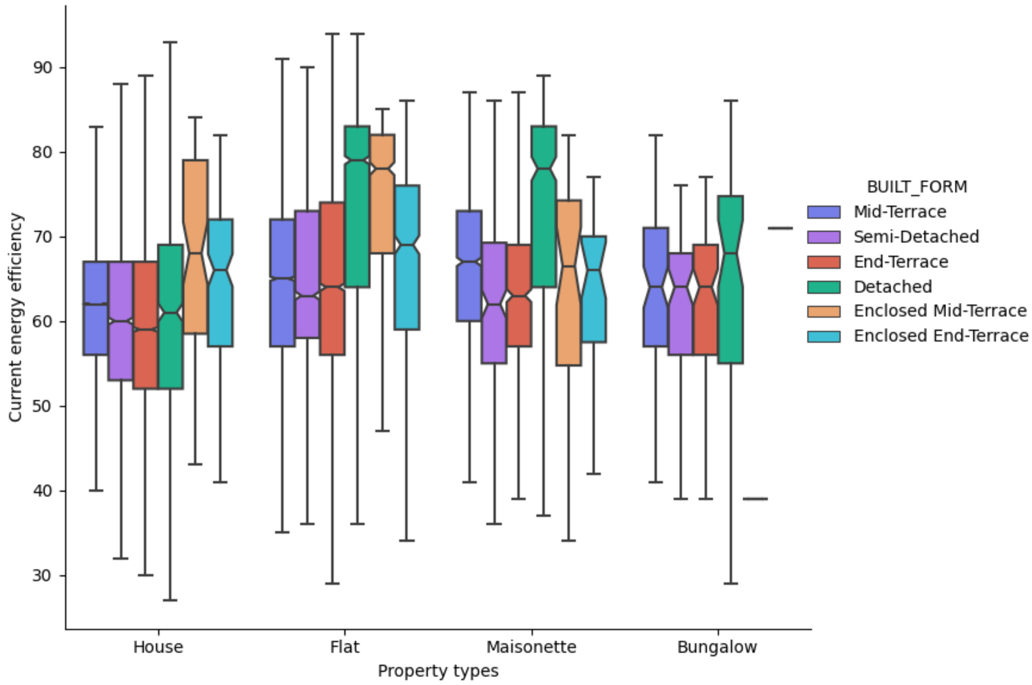

Figure 5 Energy rating by property types (Newham).

B. Decision Tree

The decision tree algorithm for built forms and insulation started with buildings as a root node, as this delivers the most information because of their preponderance in Newham. Due to the size of the decision tree, extracts have been given in Figure 6 and Figure 7, with the full tree appended.

The decision tree shows most buildings' EPC rating is C (in light green). According to the results, A-rated buildings are determined predominantly by property type (flat), whereas B and C ratings are determined mainly by wall insulation. For the lower EPC ratings (C-G), the key factor is window insulation: one group installed partial double-glazed windows and had an energy rating of D, whereas anther group did not and has an F-rating.

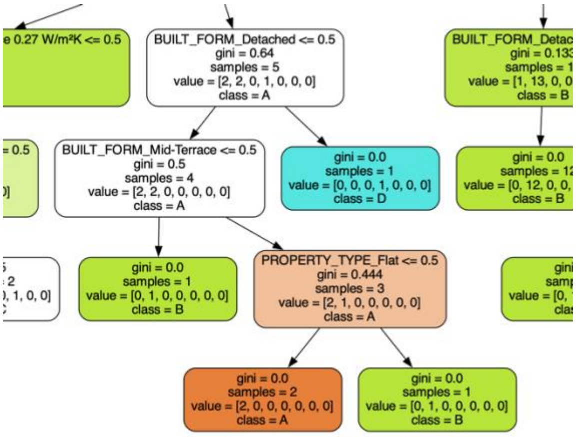

Figure 7 Examples of A and B ratings.

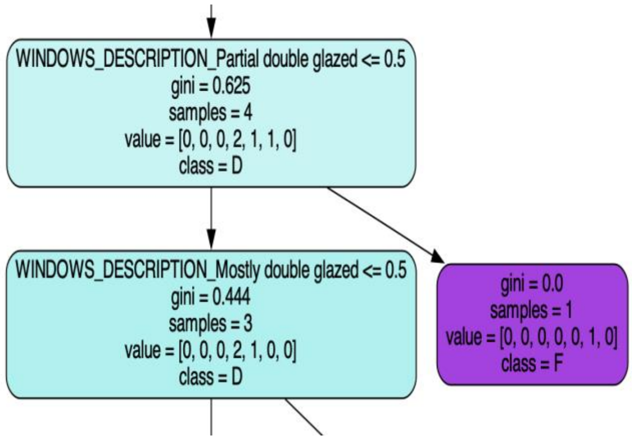

Figure 6 Examples of D and F ratings.

C. <u>K-means Clustering</u>

K-means clustering was used to cluster heating, hot water, and lighting efficiency (Figure 8 and Figure 9). The elbow point pointed when the cluster numbers were 8. Thus, I assumed that heating, water, and lighting efficiency clustered in 8 clusters. In Figure 8, the black line shows the optimal cluster numbers, and the green line shows the training time.

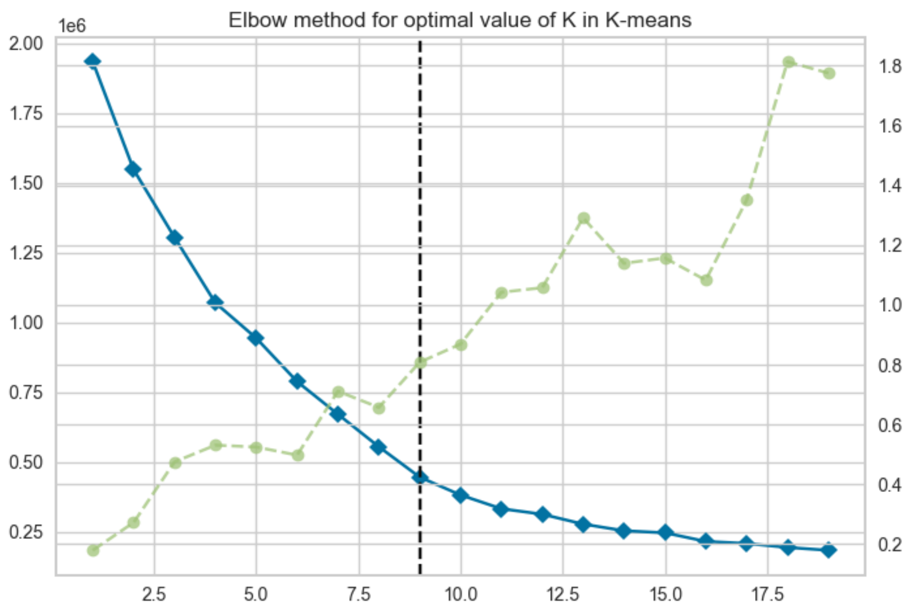

Figure 8 K-means clustering for heating, hot water, and lighting efficiency (Newham).

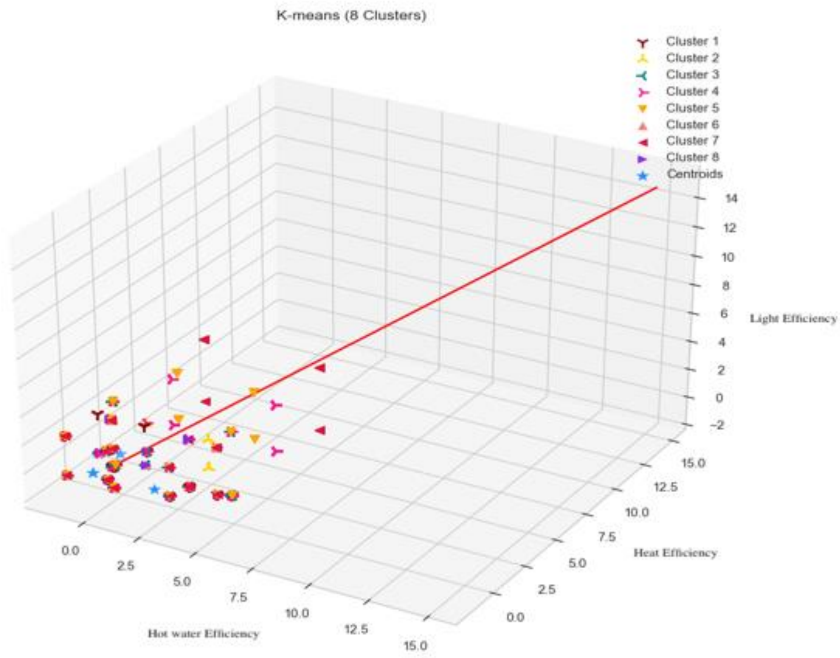

Figure 9 K means clustering for heating, hot water, and lighting efficiency (Newham) (8 clusters).

Moreover, when the clusters are 8, the silhouette score is c.0.70. It shows the clusters are well separated and dense. Moreover, when I generated the coefficient graph, all clusters tend to be uniform in size and all clusters are over the average of silhouette coefficient.

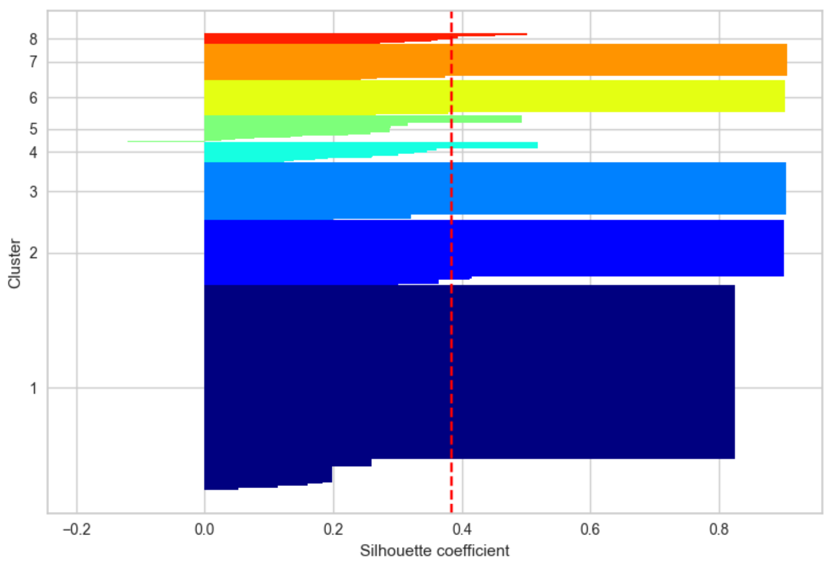

Figure 10 Silhouette Coefficient.

Based on these results, the buildings in cluster 7 tend to have stronger characteristics in lighting efficiency than other clusters. However, the majority of clusters do not show outstanding characteristics. For instance, it seems that most buildings' energy efficiency depends on heating efficiency.

D. <u>OLS Regression</u>

OLS Regression was used to see how this study's parameters impact on building energy efficiency/EPC rating. Linear regression with categorical data using dummy variables was used to show each variable's correlation.

1. *Wall energy efficiency*

The dataset divided the wall energy efficiency as Very good, Good, Average, Poor, and Very poor. I changed the character to A, B, C, D, and E and set A as a baseline variable. In the chart, baseline is the intercept.

```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
=================
Dep. Variable:     CURRENT_ENERGY_EFFICIENCY   R-squared:
0.512
Model:                                   OLS   Adj. R-squared:
0.512
Method:                        Least Squares   F-statistic:
3.230e+04
Date:                       Thu, 05 Jan 2023   Prob (F-statistic):
0.00
Time:                               14:32:43   Log-Likelihood:
-4.4759e+05
No. Observations:                     123212   AIC:
8.952e+05
Df Residuals:                         123207   BIC:
8.952e+05
Df Model:                                  4
Covariance Type:                   nonrobust
==============================================================================
==========
                 coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
----------
intercept     83.2367      0.060   1388.691      0.000      83.119
83.354
B             -9.7140      0.080   -121.108      0.000      -9.871
-9.557
C            -14.2173      0.127   -112.245      0.000     -14.466
-13.969
D            -19.9701      0.104   -191.880      0.000     -20.174
-19.766
E            -24.2567      0.072   -336.443      0.000     -24.398
-24.115
==============================================================================
==========
Omnibus:                    49949.808   Durbin-Watson:
1.952
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
312486.689
Skew:                          -1.840   Prob(JB):
0.00
Kurtosis:                       9.879   Cond. No.
6.72
==============================================================================
==========

Notes:
[1] Standard Errors assume that the covariance matrix of the errors
is correctly specified.
"""
```

Figure 11 Categorical data using dummy variables for wall & energy efficiency.

When wall insulation is very good (A), the average energy efficiency is 83.354.

The coefficient of B means the difference with the baseline (intercept). Thus, the wall insulation at a Good level of efficiency is 9.714 less than Very Good. In order to check that this result is statistically significant, this study confirmed that all variables are less than a p-value of 0.05 and so are statistically significant.

This study used the same methodology to study the efficiency of windows, other energy services, and property conditions.

2. *Windows*

```
<class 'statsmodels.iolib.summary.Summary'>
"""
                                OLS Regression Results
==================================================================
================
Dep. Variable:     CURRENT_ENERGY_EFFICIENCY   R-squared:
0.381
Model:                                   OLS   Adj. R-squared:
0.381
Method:                        Least Squares   F-statistic:
1.897e+04
Date:                       Thu, 05 Jan 2023   Prob (F-statistic):
0.00
Time:                               14:31:46   Log-Likelihood:
-4.6220e+05
No. Observations:                     123212   AIC:
9.244e+05
Df Residuals:                         123207   BIC:
9.245e+05
Df Model:                                  4
Covariance Type:                   nonrobust
==================================================================
==========
                 coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------
----------
intercept     83.7363      0.074   1127.546      0.000      83.591
83.882
B            -11.5700      0.096   -119.916      0.000     -11.759
-11.381
C            -20.2657      0.084   -240.003      0.000     -20.431
-20.100
D            -28.1562      0.182   -154.815      0.000     -28.513
-27.800
E            -27.3283      0.149   -183.983      0.000     -27.619
-27.037
==================================================================
==========
Omnibus:                    34349.388   Durbin-Watson:
1.946
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
130972.944
Skew:                          -1.357   Prob(JB):
0.00
Kurtosis:                       7.260   Cond. No.
8.23
==================================================================
==========

Notes:
[1] Standard Errors assume that the covariance matrix of the errors
is correctly specified.
"""
```

Figure 12 Categorical data using dummy variables for windows & energy efficiency.

3. *Heat system – Base line is very good – A*

```
<class 'statsmodels.iolib.summary.Summary'>
"""
                                OLS Regression Results
==================================================================
================
Dep. Variable:     CURRENT_ENERGY_EFFICIENCY   R-squared:
0.224
Model:                                   OLS   Adj. R-squared:
0.224
Method:                        Least Squares   F-statistic:
8903.
Date:                       Sat, 14 Jan 2023   Prob (F-statistic):
0.00
Time:                               22:07:28   Log-Likelihood:
-4.7588e+05
No. Observations:                     123146   AIC:
9.518e+05
Df Residuals:                         123141   BIC:
9.518e+05
Df Model:                                  4
Covariance Type:                   nonrobust
==================================================================
==========
                 coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------
----------
intercept     82.0060      0.087    942.233      0.000      81.835
82.177
B            -15.5536      0.095   -163.491      0.000     -15.740
-15.367
C            -19.1381      0.178   -107.609      0.000     -19.487
-18.790
D            -13.3048      0.232    -57.465      0.000     -13.759
-12.851
E            -25.5552      0.163   -156.507      0.000     -25.875
-25.235
==================================================================
==========
Omnibus:                    12698.539   Durbin-Watson:
1.902
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
23110.461
Skew:                          -0.705   Prob(JB):
0.00
Kurtosis:                       4.586   Cond. No.
9.82
==================================================================
==========

Notes:
[1] Standard Errors assume that the covariance matrix of the errors
is correctly specified.
"""
```

Figure 13 Categorical data using dummy variables for heat & energy efficiency.

4. *Hot water*

```
<class 'statsmodels.iolib.summary.Summary'>
"""
                                OLS Regression Results
====================================================================
=================
Dep. Variable:     CURRENT_ENERGY_EFFICIENCY   R-squared:
0.110
Model:                                   OLS   Adj. R-squared:
0.110
Method:                        Least Squares   F-statistic:
3821.
Date:                       Sat, 14 Jan 2023   Prob (F-statistic):
0.00
Time:                               22:05:19   Log-Likelihood:
-4.8432e+05
No. Observations:                     123146   AIC:
9.687e+05
Df Residuals:                         123141   BIC:
9.687e+05
Df Model:                                  4
Covariance Type:                   nonrobust
====================================================================
==========
                 coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------
----------
intercept     71.6181      0.046   1559.601      0.000      71.528
71.708
B             -7.8762      0.100    -78.410      0.000      -8.073
-7.679
C             -9.0066      0.120    -75.037      0.000      -9.242
-8.771
D            -10.0791      0.160    -63.044      0.000     -10.392
-9.766
E             -9.2675      0.119    -78.053      0.000      -9.500
-9.035
====================================================================
==========
Omnibus:                    20052.020   Durbin-Watson:
1.958
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
41446.951
Skew:                          -0.983   Prob(JB):
0.00
Kurtosis:                       5.053   Cond. No.
4.82
====================================================================
==========

Notes:
[1] Standard Errors assume that the covariance matrix of the errors
is correctly specified.
"""
```

Figure 14 Categorical data using dummy variables for hot water & energy efficiency.

5. *Lighting*

```
<class 'statsmodels.iolib.summary.Summary'>
"""
                                OLS Regression Results
====================================================================
=================
Dep. Variable:     CURRENT_ENERGY_EFFICIENCY   R-squared:
0.247
Model:                                   OLS   Adj. R-squared:
0.247
Method:                        Least Squares   F-statistic:
1.010e+04
Date:                       Sat, 14 Jan 2023   Prob (F-statistic):
0.00
Time:                               22:01:34   Log-Likelihood:
-4.7406e+05
No. Observations:                     123146   AIC:
9.481e+05
Df Residuals:                         123141   BIC:
9.482e+05
Df Model:                                  4
Covariance Type:                   nonrobust
====================================================================
==========
                 coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------
----------
intercept     81.6312      0.084    972.028      0.000      81.467
81.796
B            -14.5359      0.093   -156.683      0.000     -14.718
-14.354
C            -18.9172      0.137   -138.367      0.000     -19.185
-18.649
D            -20.1887      0.187   -108.195      0.000     -20.554
-19.823
E            -28.0495      0.170   -165.319      0.000     -28.382
-27.717
====================================================================
==========
Omnibus:                    12037.855   Durbin-Watson:
1.905
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
22437.780
Skew:                          -0.667   Prob(JB):
0.00
Kurtosis:                       4.610   Cond. No.
8.66
====================================================================
==========

Notes:
[1] Standard Errors assume that the covariance matrix of the errors
is correctly specified.
"""
```

Figure 17 Categorical data using dummy variables for lighting & energy efficiency.

6. *Property type – the baseline (intercept) is Bungalow.*

```
                            OLS Regression Results
==============================================================================
=================
Dep. Variable:     CURRENT_ENERGY_EFFICIENCY   R-squared:
0.045
Model:                                   OLS   Adj. R-squared:
0.045
Method:                        Least Squares   F-statistic:
1120.
Date:                       Thu, 05 Jan 2023   Prob (F-statistic):
0.00
Time:                               14:24:46   Log-Likelihood:
-2.7780e+05
No. Observations:                      71519   AIC:
5.556e+05
Df Residuals:                          71515   BIC:
5.556e+05
Df Model:                                  3
Covariance Type:                   nonrobust
==============================================================================
==========
                 coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
----------
intercept     61.7412      0.665     92.823      0.000      60.438
63.045
Flat           4.2418      0.670      6.329      0.000       2.928
5.555
House         -1.2638      0.667     -1.894      0.058      -2.572
0.044
Maisonette     3.7395      0.700      5.346      0.000       2.368
5.111
==============================================================================
==========
Omnibus:                    16638.724   Durbin-Watson:
0.275
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
56225.098
Skew:                          -1.165   Prob(JB):
0.00
Kurtosis:                       6.666   Cond. No.
38.0
==============================================================================
==========

Notes:
[1] Standard Errors assume that the covariance matrix of the errors
is correctly specified.
```

Figure 15 Categorical data using dummy variables for property type & energy efficiency.

7. *Built form – the baseline (intercept) is Detached.*

```
                            OLS Regression Results
====================================================================
=================
Dep. Variable:     CURRENT_ENERGY_EFFICIENCY   R-squared:
0.049
Model:                                   OLS   Adj. R-squared:
0.049
Method:                        Least Squares   F-statistic:
743.9
Date:                       Thu, 05 Jan 2023   Prob (F-statistic):
0.00
Time:                               14:25:52   Log-Likelihood:
-2.7763e+05
No. Observations:                      71519   AIC:
5.553e+05
Df Residuals:                          71513   BIC:
5.553e+05
Df Model:                                  5
Covariance Type:                   nonrobust
====================================================================
====================
                          coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------
--------------------
intercept              71.0248      0.171    415.648      0.000
70.690      71.360
Enclosed End-Terrace   -5.9849      0.454    -13.175      0.000
-6.875      -5.095
Enclosed Mid-Terrace    0.6167      0.415      1.487      0.137
-0.196       1.430
End-Terrace           -10.5940      0.201    -52.837      0.000
-10.987     -10.201
Mid-Terrace            -9.3162      0.179    -51.990      0.000
-9.667      -8.965
Semi-Detached          -9.4967      0.236    -40.203      0.000
-9.960      -9.034
====================================================================
==========
Omnibus:                    17046.894   Durbin-Watson:
0.279
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
61086.250
Skew:                          -1.174   Prob(JB):
0.00
Kurtosis:                       6.871   Cond. No.
14.6
====================================================================
==========

Notes:
[1] Standard Errors assume that the covariance matrix of the errors
is correctly specified.
```

Figure 16 Categorical data using dummy variables for built form & energy efficiency.

When the variables were compared to the intercept, the P-values were tested. However, as the P-value was greater than 0.05, the comparison between the baseline and other variables was not statistically significant. When each variable's correlation was compared, the confidence interval was tested using the statistical hypothesis test. On the statistical hypothesis test, the confidence intervals were all steady showing that the results are statistically significant.

V. Discussion and Conclusion

Discussion

This essay considered the relationship between domestic buildings' characteristics (built form, insulation, and service systems) and EPC ratings in the London borough of Newham.

This study has found that the majority of domestic buildings in Newham are mid-terraces with an EPC rating of C or D. This means that Newham has serious scope for improvement in terms of energy efficiency.

As it is difficult to 'correct' terracing, this essay considered other building characteristics, such as insulation and energy services (heating, water, and lighting efficiency) which are feasible to improve. It was found that the most important factors among these are wall insulating and window-glazing features, confirming the literature. Conversely, the analysis regarding energy services did not yield any meaningful results, so it is difficult to confirm the findings in the literature in that respect. It is therefore recommended that Newham council strategies focus on improvements to wall insulation and window-glazing.

Limitations

First, as EPC ratings rely on a vast matrix of factors, this study is inherently limited to the variables selected. Although this selection is based on variables the literature considers to be critical, other variables like building age and size are potentially equally important (Jenkins et al., 2017).

Second, Newham borough did not have a range of EPC ratings, with only one A-rating and significant C-D clustering. Any correlations are therefore susceptible to a lack of adequate sampling.

Third, certain data has been null: although roof and floor insulation were considered in preliminary stages, their decision trees were null/not clear and could not be used. Similarly, the k-means clustering of energy services did not produce distinguishable clusters.

Finally, EPC ratings are inherently prone to criticisms that they are not robustly measured (Crawley et al., 2019), which creates a contagion risk in analyses using EPC ratings.

Further Study

Future studies could more comprehensively examine all the EPC-rating factors, grouping them by hierarchy of impact according to the literature, and analysing them in that hierarchy, to minimise time spent on factors which are thought to be insignificant.

Another viable approach could be to compare one variable (e.g. built form) across each London borough, to see if there is a strong correlation between one variable and EPC ratings across different locations (and then repeat for each other variable).

In either case, the data collected must be presented with more granularity as to the different features in residential dwellings that are thought to impact on energy efficiency, as characteristics were often amalgamated unhelp.

Reference:

1. BEIS (2021). *Net Zero Strategy : Build Back Greener.* UK: Hm Government.
2. BEIS (2022). 2020 UK greenhouse gas emissions, Final Figures. [online] Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1051408/2020-final-greenhouse-gas-emissions-statistical-release.pdf.
3. Chen, S., Zhang, G., Xia, X., Setunge, S. and Shi, L. (2020). A review of internal and external influencing factors on energy efficiency design of buildings. *Energy and Buildings*, 216, p.109944. doi:10.1016/j.enbuild.2020.109944.
4. Crawley, J., Biddulph, P., Northrop, P.J., Wingfield, J., Oreszczyn, T. and Elwell, C. (2019). Quantifying the Measurement Error on England and Wales EPC Ratings. *Energies*, [online] 12(18), p.3523. doi:10.3390/en12183523.
5. Department for Business, Energy & Industrial Strategy (BEIS) (2022). Provisional UK greenhouse gas emissions national statistics 2021. [online] Available at: https://www.gov.uk/government/statistics/provisional-uk-greenhouse-gas-emissions-national-statistics-2021.
6. epc.opendatacommunities.org. (2022). *Energy Performance of Buildings Data England and Wales*. [online] Available at: https://epc.opendatacommunities.org/domestic/search [Accessed 10 Dec. 2022].
7. Gupta, B., Rawat, A., Jain, A., Arora, A. and Dhami, N. (2017). Analysis of Various Decision Tree Algorithms for Classification in Data Mining. *International Journal of Computer Applications*, 163(8), pp.15–19. doi:10.5120/ijca2017913660.
8. HM Government (2022). *The Building Regulations 2010. Approved document L1A, Approved Document L, Conservation of fuel and power, Volume 1: Dwellings*. 2021 edition – for use in England ed. [online] London: HM Government. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1099626/ADL1.pdf [Accessed 10 Dec. 2022].
9. Jenkins, D., Simpson, S. and Peacock, A. (2017). Investigating the consistency and quality of EPC ratings and assessments. *Energy*, [online] 138, pp.480–489. doi:10.1016/j.energy.2017.07.105.
10. Kilmer, J.T. and Rodríguez, R.L. (2016). Ordinary least squares regression is indicated for studies of allometry. *Journal of Evolutionary Biology*, 30(1), pp.4–12. doi:10.1111/jeb.12986.
11. Na, S., Xumin, L. and Yong, G. (2010). Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm. *2010 Third International Symposium on Intelligent Information Technology and Security Informatics*. doi:10.1109/iitsi.2010.74.
12. Rosenow, J., Guertler, P., Sorrell, S. and Eyre, N. (2018). The remaining potential for energy savings in UK households. *Energy Policy*, 121, pp.542–552. doi:10.1016/j.enpol.2018.06.033.
13. Sinaga, K.P. and Yang, M.-S. (2020). Unsupervised K-Means Clustering Algorithm. *IEEE Access*, 8, pp.80716–80727. doi:10.1109/access.2020.2988796.
14. Somasundaram, S., Chong, A., Wei, Z. and Thangavelu, S.R. (2020). Energy saving potential of low-e coating based retrofit double glazing for tropical climate. *Energy and Buildings*, 206, p.109570. doi:10.1016/j.enbuild.2019.109570.
15. Song, Y. and Lu, Y. (2015). Decision tree methods: applications for classification and prediction. *Shanghai Arch Psychiatry*, [online] 27(2), pp.130–135. doi:10.11919/j.issn.1002-0829.215044.