Developing Data Driven Insights: The Energy Crisis

London Borough of Hounslow

ONE HOUNSLOW

Energy Efficiency and the Cost of Living: Insight

- Develop Insight
- Guide Service Areas
- Create an Accessible and Interactive Research Tool

Methods

Variables/Data:

- Energy Performance Certificate (EPC) Scores/Grades (*Department for Levelling UP, Housing and Communities*)

- Household Income (*CACI*)

- Acorn Segmentation Data (*CACI*)

Assess household variation across these three metrics
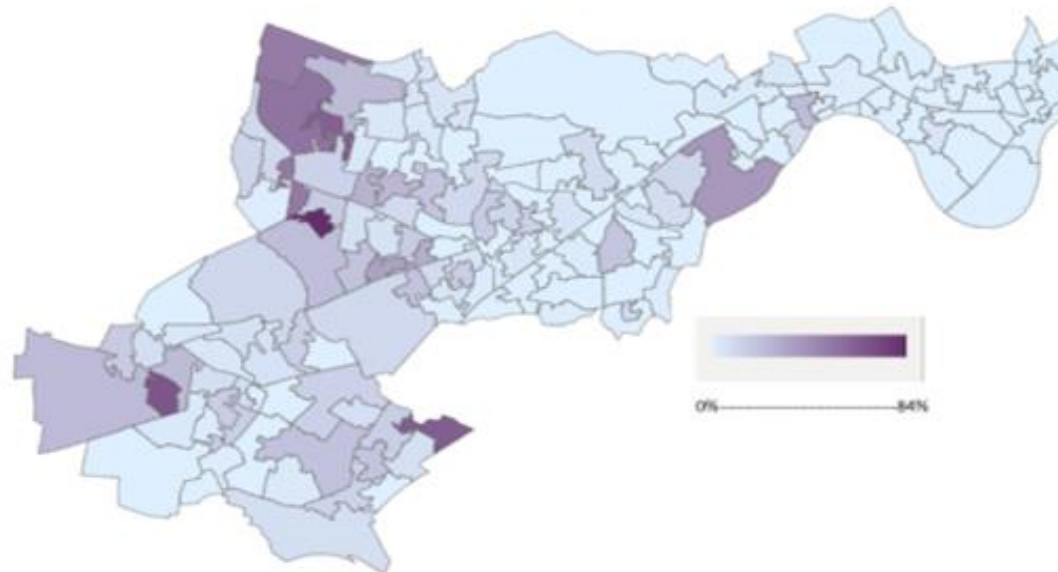
The Energy Crisis Matrix
Data

	A		B
	High Income and High Energy Efficiency		**High Income and Low Energy Efficiency**
	C		D
	Low Income and High Energy Efficiency		**Low Income and Low Energy Efficiency**

Category D = EPC < 69 + Annual Income < £25,000.

A	B
High Income and High Energy Efficiency	High Income and Low Energy Efficiency

C	D
Low Income and High Energy Efficiency	Low Income and Low Energy Efficiency
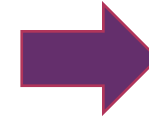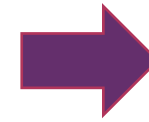
Categories A, B, C, D
May Require Varying Individual Policy Responses

Energy inefficient households with median income less than £25,000 (%)

0% ———— 84%

ONE HOUNSLOW
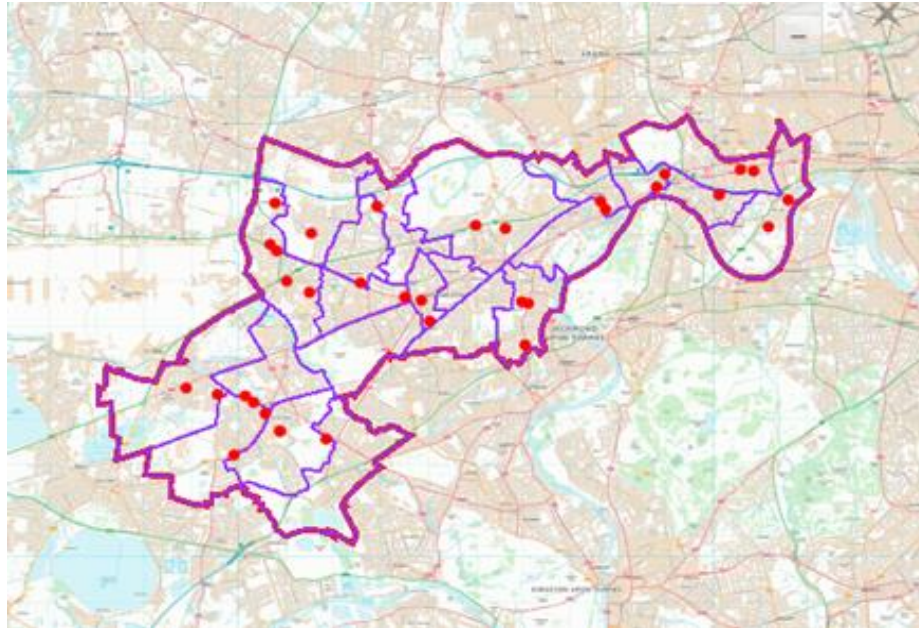
EPC + Income Data

A	B
High Income and High Energy Efficiency	High Income and Low Energy Efficiency

C	D
Low Income and High Energy Efficiency	Low Income and Low Energy Efficiency

Energy inefficient households with median income less than £25,000 (%)

0%   84%
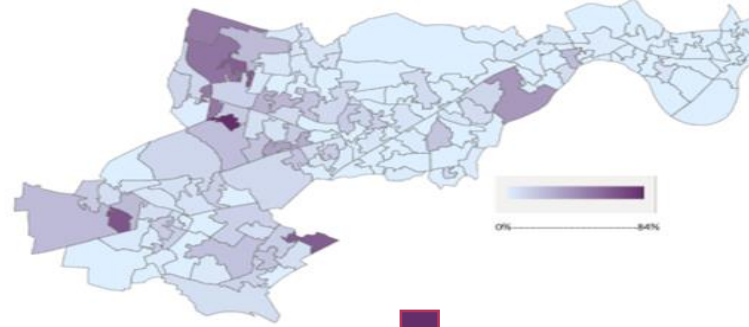
Hounslow's Warm Spaces Initiative

30 Warm Spaces across the borough, so people have somewhere to go if they cannot afford to keep the heating on at home.

The Online Report & Interactive Element
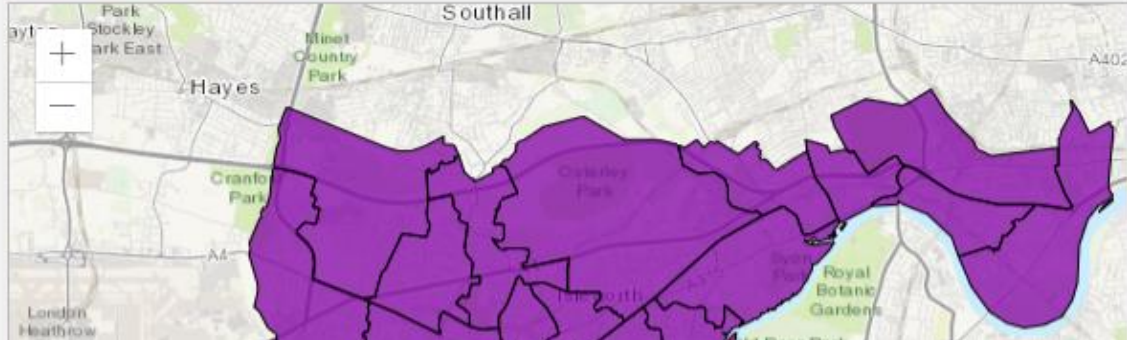

Energy Efficiency and the Cost of Living: Insight

https://stats.hounslow.gov.uk/

Quadrant D

Figure 3.2

Energy inefficient households with median income less than £25,000 (%)

0% ———————— 84%

Acorn_Description	Wellbeing_Acorn_Description	Sum of Households
Larger family homes, multi-ethnic areas	Happy families	8976
Deprived and ethnically diverse in flats	Despondent diversity	5519
Mixed metropolitan areas	Gym & juices	5068
Owner occupied terraces, average income	Borderline behaviours	4517
Social rented flats, families and single parents	Perilous futures	4510
Educated young people in flats and tenements	Borderline behaviours	4329
First time buyers in small, modern homes	Regular revellers	4222
Larger family homes, multi-ethnic areas	Gym & juices	4199
Younger professionals in smaller flats	Gym & juices	3651
Larger family homes, multi-ethnic areas	Healthy, wealthy & wine	2963

LSOA Name (2022)	EPC Score (LSOA Average)	EPC Score (Potential LSOA Average)	Potential Savings per Household (Percent)	Percentage of Households with EPC Score below 69	Percentage of Households where Income is below £25K Annually
Hounslow 001A	66	76	23%	58%	9%
Hounslow 001B	60	71	27%	73%	10%
Hounslow 001C	55	68	27%	86%	7%
Hounslow 001D	60	72	27%	77%	9%
Hounslow 001E	61	72	28%	68%	10%
Hounslow 004A	70	79	25%	40%	33%
Hounslow 004B	60	80	32%	81%	37%
Hounslow 004C	61	78	30%	74%	31%

Wards (2022)

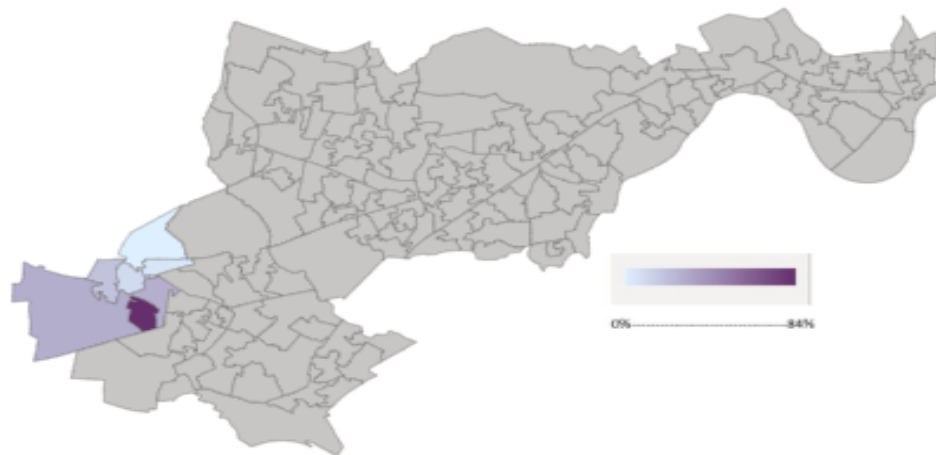|			
Bedfont	Chiswick Gunnersbury	Cranford	Hanworth Park
Brentford East	Chiswick Homefields	Feltham North	Hanworth Village
Brentford West	Chiswick Riverside	Feltham West	Heston Central

Quadrant D

Figure 3.2

Energy inefficient households with median income less than £25,000 (%)

0% ———— 84%

Acorn_Description	Wellbeing_Acorn_Description	Sum of Households
Social rented flats, families and single parents	Perilous futures	640
Families in right-to-buy estates	Rooted routines	558
Established suburbs, older families	Everything in moderation	381
Suburban semis, conventional attitudes	Everything in moderation	347
Larger family homes, multi-ethnic areas	Happy families	268
Low cost flats in suburban areas	Borderline behaviours	226
Smaller houses and starter homes	Everything in moderation	206
Young families in low cost private flats	Borderline behaviours	177
Social rented flats, families and single parents	Struggling smokers	173
Singles and young families, some	Perilous futures	167

LSOA Name (2022)	EPC Score (LSOA Average)	EPC Score (Potential LSOA Average)	Potential Savings per Household (Percent)	Percentage of Households with EPC Score below 69	Percentage of Households where Income is below £25K Annually
Hounslow 022A	63	78	28%	68%	31%
Hounslow 022B	60	77	31%	80%	32%
Hounslow 023A	65	79	29%	60%	43%
Hounslow 023B	72	80	21%	38%	42%
Hounslow 023C	70	81	25%	37%	43%
Hounslow 023D	65	80	29%	63%	45%
Hounslow 023E	68	79	19%	46%	63%

Wards (2022)

Bedfont	Chiswick Gunnersbury	Cranford	Hanworth Park
Brentford East	Chiswick Homefields	Feltham North	Hanworth Village
Brentford West	Chiswick Riverside	Feltham West	Heston Central

UPRN | Address – Fuzzy Matching Practices

Data Preparation Stage (R)

EPC Dataset Schema Transitions → LBH LLPG Dataset (Gold Standard)

IG & DPIA Scope

EPC Schema Match RUN for Future Data Automation Process
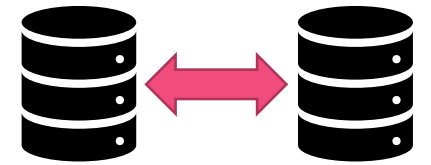
Remove Unwanted Variables

Data Cleansing Steps

Sort by Inspection Date

Duplication Checks (Dataset & UPRN Variable)

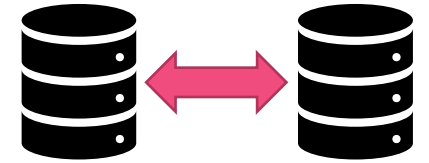Remove Duplications (for i.e. same property with 3 inspection dates)

Split EPC Dataset (With and W/O UPRN)

Ensure IG is aware & informed of data processing!

LLPG = Local Land and Property Gazetteer | EPC = Energy Performance Certificate

London Borough of Hounslow

UPRN | Address – Fuzzy Matching Practices

Data Preparation Stage (R)

```
199  dfepc$local_authority <- NULL    # Remove unwanted cols in dataset
200  dfepc$county <- NULL
201  dfepc$uprn_source <- NULL
202  dfepc$local_authority_label <- NULL
203  dfepc$lmk_key <- NULL
204
205
206  # Sort date order (Benchmark: inspection_date variable) - to keep the latest data on duplicated UPRN's
207
208  dfepc$inspection_date [10]
209  dfepc$inspection_date <- as.Date(dfepc$inspection_date, format = "%Y-%m-%d")
210  dfepc$inspection_date
211
212  dfepc <- dfepc[order(dfepc$inspection_date, decreasing = TRUE),]
213  head(dfepc)
214  View(dfepc) # To View entire dataset - Checking Date Order
215
```

This is an example snapshot of a R Script!

London Borough of Hounslow

UPRN | Address – Fuzzy Matching Practices

Data Transformation Stage (R)

Split EPC Dataset (With and W/O UPRN)

1. Using distinct function to filter out EPC Addresses with Unique UPRN's (dfA)

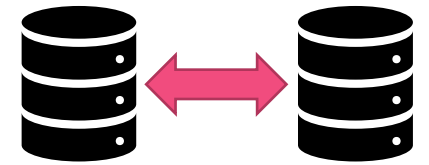2. Using is.na function to filter out EPC addresses with NO UPRN's (dfB)

Create a new variable by Merging "address" & "postcode" → merged_full_address

If dfB <40K rows, fuzzy matching can be performed well using Power Query in Excel

Set Threshold to .90 for Good Matching Results (>80%)

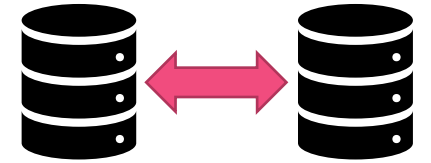If dfB >40K rows, apply Data Science Distance Based Methods for i.e. JW or LD

Where min distance b/w strings (X and Y) shows Matched Address Strength (Score)

Use 64 bit Excel rather than 32 bit Excel for a faster data processing.

LD = Levenshtein Distance | JW = Jaro Winkler | EPC = Energy Performance Certificate | df = Dataframe

London Borough of Hounslow

UPRN | Address – Fuzzy Matching Practices

Data Transformation Stage (R)

```
321   # ================================================================
322
323   # Small dataset <40K, fuzzy matching can be performed well using Power Query in Excel
324
325   # Use Template M Programming code for quick fuzzy matching: APPLIED STEPS: Left Outer Merge with .90 Threshold
326   # --------------------------------------------------------------------
327   # let
328   # Source = Table.FuzzyNestedJoin(dfepc_b, {"merged_full_address"}, refined_dfllpg, {"llpg_postal_address"}, "refined_dfl
329   #"Expanded refined_dfllpg" = Table.ExpandTableColumn(Source, "refined_dfllpg", {"llpg_postal_address", "llpg_easting", "
330   #"Removed Duplicates" = Table.Distinct(#"Expanded refined_dfllpg", {"merged_full_address"}),
331   #"Reordered Columns" = Table.ReorderColumns(#"Removed Duplicates",{"refined_dfllpg.llpg_postal_address", "merged_full_ad
332   #"Renamed Columns" = Table.RenameColumns(#"Reordered Columns",{{"refined_dfllpg.llpg_easting", "llpg_easting"}, {"refine
333   #"Removed Columns" = Table.RemoveColumns(#"Renamed Columns",{"lsoa name"}),
334   #"Renamed Columns1" = Table.RenameColumns(#"Removed Columns",{{"refined_dfllpg.llpg_postal_address", "llpg_postal_addres
335   #"Reordered Columns1" = Table.ReorderColumns(#"Renamed Columns1",{"address", "postcode", "address1", "address2", "addres
336   # in
337   #"Reordered Columns1"
338   # --------------------------------------------------------------------
```
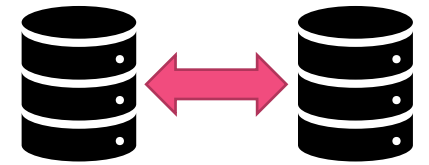
This is an example snapshot of a M code in R Script!

London Borough of Hounslow

UPRN | Address – Fuzzy Matching Practices

Data Transformation Stage (R)

```
387  dflev_llpg <- dfllpg # LLPG Target Data
388  dflev_epc <- dfepc_b # Incoming EPC Data (W/O UPRN's)
389
390
391  # String 1 = llpg_postal_address
392  # String 2 = merged_full_address
393
394  # Merge address and postcode variables into a single variable in dflev_epc
395  dflev_epc$merged_full_address <- paste(dflev_epc$address, dflev_epc$postcode, sep = " ")
396  View(dflev_epc)
397
398
399  # Create a function to calculate Levenshtein distance between two strings
400  lev_distance <- function(string1, string2) {
401     stringdist(string1, string2, method = "lv")
402  }
403
404  # Apply above LV function to both datasets
405  distances <- outer(dflev_llpg$llpg_postal_address, dflev_epc$merged_full_address, Vectorize(lev_distance))
406  results <- data.frame(dflev_llpg$llpg_postal_address, dflev_epc$merged_full_address, distances)
```

This is an example snapshot of a R Script!

Questions: