General Sir John Kotelawala Defense University

Faculty Of Management, Social Sciences and Humanities

Department of languages

BSc in Applied Data Science Communication

Advanced SQL and cloud databases - Assignment

<u>Group</u>

D/ADC/24/0015 – WMD Subhashwaree

D/ADC/24/0017 – GGI Gamanayaka

D/ADC/24/0031 – JANI Jayawardhana

Energy Performance Analysis of Manchester Local Authorities (2014–2024): A Power BI and SQL-Based Approach

Content

1. Introduction

One of the important aspects of the strategy of the United Kingdom to enhance energy efficiency in residential and commercial buildings is the Energy Performance Certificates (EPCs). They offer a universal way of measuring the energy performance of a property, rating the buildings from A (most efficient) to G (least efficient). The EPCs also contain the information about the approximate amount of CO_2 emissions, properties of the heating system, the level of insulation, the level of energy consumption as well as a list of the recommended upgrades which might increase the efficiency rating of the property. These facts render EPCs a key data set to local governments, property owners, tenants, governmental institutions, and sustainability experts. They facilitate the evolution of policies, educate retrofitting programs, and contribute to the monitoring of the progress on the carbon reduction targets on the national level.

This paper is dedicated to the analysis of EPC related data of the local authorities of Greater Manchester during a decade, between 2014 and 2024. Analysis is focused on showing the tendencies in energy efficiency, geographical variations, CO 2 emission behaviour, and property-based aspects that affect efficiency outcomes. Besides the certificate level EPC data, this enquiry includes the recommendations data, which includes specific improvement measures, including insulation improvements, installation of renewable energy, and replacement of boilers, that is attached to specific EPC evaluation. The association of the two sets of data makes it possible to have a better insight into the prevailing condition of housing energy performance as well as the achievable realistic potential improvement across the whole of Manchester.

To do this, SQL Server was utilized heavily in cleaning data and preprocessing the data and integrating the two sets of data into a relational format with the key being the LMK_KEY. This facilitated proper joins, duplicates being handled, outliers removed, correcting data types, and converting numerical and textual data. A subsequent exploratory analysis and an all-encompassing interactive dashboard with high-level functionality, including DAX measures, drill-downs, slicers, geospatial mapping, and KPI indicators, was then conducted with the help of Power BI. These tools of visuals contribute to the discovery of concealed patterns, trends comparisons across local authorities, and opportunities of improvement.

The following report is a report of the complete methodology, analysis and results of both datasets in terms of energy efficiency trends and suggested improvements as well as those with the most severe inefficiencies. The resulting insights will help guide the stakeholders, such as the Manchester City Council, property developers, and homeowners, to prioritize the activities of retrofitting, deploy sustainable interventions, and plan the energy-efficiency approaches in the future. This paper provides a strong and detailed evaluation of the EPC landscape in Manchester through the integration of SQL-based data processing and the use of Power BI visualisation.

2. Dataset description

This study uses two complementary datasets: the EPC Certificates Dataset and the EPC Recommendations Dataset, both downloaded from the Domestic EPC data portal provided by Open Data Communities. Together, these datasets provide a detailed representation of energy performance across Greater Manchester between 2014 and 2024 and enable both efficiency assessment and improvement analysis.

Dataset 1: EPC Certificates Dataset (certificates.csv)

The certificates dataset forms the core of the analysis. It contains property-level information collected during EPC assessments, including:

- Energy ratings (Current and Potential)
- Energy efficiency scores
- Estimated annual CO_2 emissions (tonnes/year)
- Inspection and lodgement dates
- Local authority information
- Property type (house, flat, bungalow, maisonette, etc.)
- Built form (detached, semi-detached, mid-terrace, etc.)
- Heating system type

- Total floor area
- Tenure type
- Address-level metadata
- Unique LMK_KEY identifier

The dataset is extensive, containing hundreds of thousands of records for Manchester alone. It enables temporal, geographical, and categorical analysis of energy performance.

Dataset 2: EPC Recommendations Dataset (recommendations.csv)

The recommendations dataset expands the analysis by providing improvement suggestions linked to individual certificates through the LMK_KEY. It includes:

- Recommended energy improvements (e.g., cavity wall insulation, photovoltaic solar panels, boiler upgrades, lighting improvements)
- Indicative cost ranges (e.g., £500–£1,000)
- Estimated improvement in energy rating or score
- Estimated annual savings
- Impact on CO_2 emissions

This dataset enables the study to investigate improvement potential, cost implications, and common types of interventions needed across Manchester's housing stock.

Data Integration and Suitability

Using LMK_KEY, a one-to-many relationship was established between certificates and their improvement recommendations. SQL Server was used to clean, standardise, and integrate both datasets by:

- Removing duplicates
- Correcting data types

- Cleaning cost fields
- Standardising textual categories
- Handling missing values
- Removing extreme outliers
- Creating clean tables for Power BI

The combined dataset is highly suitable for Power BI because it supports time-series, spatial, categorical, and financial analysis.

By integrating both datasets, the study delivers a deeper, more comprehensive understanding of Manchester's energy performance landscape and the retrofitting actions necessary to improve it.

Datasets link

https://epc.opendatacommunities.org/downloads/domestic#local-authority

3. Methodology

The raw EPC data was imported into SQL Server and underwent rigorous cleaning using T-SQL, thus showing advanced database management skills.

Here's a full explanation for each sql codes block

Step 1 - Confirm the Imported Tables

```
--STEP1_CONFIRM THE IMPORTED TABLES
SELECT TABLE_NAME
FROM INFORMATION_SCHEMA.TABLES
WHERE TABLE_TYPE='BASE TABLE';
```

> Verify the Code Purpose of the Imported Tables

To make sure the EPC datasets were imported correctly, this query determines which tables are present in the database.

Justification:

All base table names are retrieved from INFORMATION_SCHEMA.TABLES by this query.Before starting the analysis, this step makes sure that both recommendations and certificates are present.It is a common verification step in SQL workflows and aids in confirming a successful import.

Step 2 - Basic Exploratory Data Analysis (EDA)

```
--STEP2_BASIC EXPLORATORY DATA ANALYSIS

--A) Count rows in both tables
SELECT COUNT(*) AS total_certificates FROM dbo.certificates;
SELECT COUNT(*) AS total_recommendations FROM dbo.recommendations;
```

> Fundamental Exploratory Data Analysis (EDA)
A) Count the rows in both tables.

Justification:

The total number of records in each dataset is ascertained using two COUNT(*) queries.This makes it possible to compare the datasets for recommendations and certificates.Additionally, it helps estimate the volume of data before cleaning and verifies that the tables contain data.

```
--B) View sample data
SELECT TOP 50 * FROM dbo.certificates;
SELECT TOP 50 * FROM dbo.recommendations;
```

B)View an explanation of sample data:

Each table's first 50 rows are examined using SELECT TOP 50.aids in comprehending data types, formatting concerns, column structure, and possible data issues.crucial for visually identifying irregularities (like symbols, incorrect date formats, etc.) prior to cleaning.

```sql
--C) Find columns with missing values
-- 1.Missing values in certificates
SELECT
    SUM(CASE WHEN LMK_KEY IS NULL OR LMK_KEY = '' THEN 1 END) AS missing_LMK_KEY,
    SUM(CASE WHEN ADDRESS IS NULL OR ADDRESS = '' THEN 1 END) AS missing_ADDRESS,
    SUM(CASE WHEN POSTCODE IS NULL OR POSTCODE = '' THEN 1 END) AS missing_POSTCODE
FROM dbo.certificates;

-- 2. Missing values in recommendations
SELECT
    SUM(CASE WHEN LMK_KEY IS NULL OR LMK_KEY = '' THEN 1 END) AS missing_LMK_KEY,
    SUM(CASE WHEN IMPROVEMENT_ITEM IS NULL OR IMPROVEMENT_ITEM = '' THEN 1 END) AS missing_ITEM
FROM dbo.recommendations;
```

C) Locate any missing values in the columns

Justification:

These queries determine the number of null or missing values in crucial identifier fields:

Regarding certificates:

Verifies if LMK_KEY, ADDRESS, and POSTCODE are missing.Because it connects the two datasets, LMK_KEY is essential.

In Recommendations:

Verifies the improvement text and any missing LMK_KEYThis is crucial because Joins would be broken by missing LMK_KEY values.Geographic accuracy is decreased by missing addresses and postcodes.Items for improvement that are missing lower the quality of insight

```sql
--D) Count duplicates (LMK_KEY is your joining key)
-- duplicates in certificates
SELECT LMK_KEY, COUNT(*) AS cnt
FROM dbo.certificates
GROUP BY LMK_KEY
HAVING COUNT(*) > 1;

-- duplicates in recommendations
SELECT LMK_KEY, COUNT(*) AS cnt
FROM dbo.recommendations
GROUP BY LMK_KEY
HAVING COUNT(*) > 1;
```

D) **Use LMK_KEY to count duplicates:**

Every EPC certificate should be uniquely identified by LMK_KEY.

However, a single certificate may have several recommendations.

The query

LMK_KEY-based groups find the rows where LMK_KEY appears more than once.

This is employed in:

Find duplicate entries in the table of certificates and Recognize the one-to-many structure in suggestions

```sql
--E) Detect impossible numeric values
SELECT *
FROM dbo.certificates
WHERE ENERGY_CONSUMPTION_CURRENT < 0 OR ENERGY_CONSUMPTION_CURRENT > 100000;
```

E) Find numerical values that are impossible

Justification:

This query looks for inaccurate or unrealistic energy consumption figures.Corrupted or invalid entries are indicated by values less than zero or very high.aids in locating outliers for elimination or correction.

```sql
--F) Find corrupted LMK_KEY values
SELECT DISTINCT LMK_KEY
FROM dbo.recommendations
WHERE LMK_KEY LIKE '%E+%' OR LMK_KEY LIKE '%E-%';
```

F) Locate LMK_KEY values that have been corrupted:

Certain LMK_KEY values (such as "E+05") are written in scientific notation.

Before joining datasets, these need to be cleaned because they are corrupted.

The query finds these problems so they can be specifically fixed.

Step 3 - Data Cleaning

```sql
--STEP 3 DATA CLEANING

--1) Create clean_certificates table
IF OBJECT_ID('dbo.clean_certificates','U') IS NOT NULL
    DROP TABLE dbo.clean_certificates;
GO

SELECT
    -- Keep LMK_KEY as text
    LMK_KEY,

    -- Address fields
    ADDRESS1,
    ADDRESS2,
    ADDRESS3,
    POSTCODE,

    -- Convert numeric fields safely
    TRY_CAST(CURRENT_ENERGY_RATING AS INT) AS CURRENT_ENERGY_RATING,
    TRY_CAST(POTENTIAL_ENERGY_RATING AS INT) AS POTENTIAL_ENERGY_RATING,
    TRY_CAST(CURRENT_ENERGY_EFFICIENCY AS INT) AS CURRENT_ENERGY_EFFICIENCY,
    TRY_CAST(POTENTIAL_ENERGY_EFFICIENCY AS INT) AS POTENTIAL_ENERGY_EFFICIENCY,

    --Convert Datetime Fieald
    TRY_CAST(INSPECTION_DATE AS DATETIME) AS INSPECTION_DATE,
    TRY_CAST(LODGEMENT_DATE AS DATETIME) AS LODGEMENT_DATE,

    -- Convert decimal fields
    TRY_CAST(CO2_EMISSIONS_CURRENT AS DECIMAL(10,2)) AS CO2_EMISSIONS_CURRENT,
    TRY_CAST(CO2_EMISSIONS_POTENTIAL AS DECIMAL(10,2)) AS CO2_EMISSIONS_POTENTIAL,

    -- Cost fields
    TRY_CAST(LIGHTING_COST_CURRENT AS INT) AS LIGHTING_COST_CURRENT,
    TRY_CAST(LIGHTING_COST_POTENTIAL AS INT) AS LIGHTING_COST_POTENTIAL,
    TRY_CAST(HEATING_COST_CURRENT AS INT) AS HEATING_COST_CURRENT,
    TRY_CAST(HEATING_COST_POTENTIAL AS INT) AS HEATING_COST_POTENTIAL,
```

```sql
    -- Keep all long text fields as NVARCHAR(MAX)
    FLOOR_DESCRIPTION,
    WALLS_DESCRIPTION,
    ROOF_DESCRIPTION,
    MAINHEAT_DESCRIPTION,
    WINDOWS_DESCRIPTION,
    HOTWATER_DESCRIPTION,
    CONSTRUCTION_AGE_BAND,
    ADDRESS

INTO dbo.clean_certificates
FROM dbo.certificates;
GO
```

1. **Justification:**

The certificates dataset is cleaned up in this step.

Activities carried out:

>Take out the old, spotless table.
If clean_certificates already exist, it is dropped.

ensures that the script can execute repeatedly without encountering any issues.

>Pick and modify columns
To safely convert text values into numeric or datetime values without disrupting the query, use TRY_CAST.

When conversion fails, TRY_CAST returns NULL rather than an error.

keeps scripts from failing on low-quality data

> Convert pertinent fields:
Ratings of energy → INT

Scores for energy efficiency → INT

Dates → Time

CO2 emissions → DECIMAL

Cost values for lighting and heating → INT

>Maintain lengthy text fields
Descriptions of the floor, roof, walls, etc. are stored as NVARCHAR (MAX).

> Output table
The cleaned dataset that is produced is saved as clean_certificates.

The certificates dataset will become organized and error-free for analysis because of this cleaning step.

```sql
--2) Create clean_recommendations table
IF OBJECT_ID('dbo.clean_recommendations','U') IS NOT NULL
    DROP TABLE dbo.clean_recommendations;
GO

SELECT
    LMK_KEY,
    TRY_CAST(IMPROVEMENT_ITEM AS INT) AS IMPROVEMENT_ITEM,
    IMPROVEMENT_SUMMARY_TEXT,
    IMPROVEMENT_DESCR_TEXT,
    TRY_CAST(IMPROVEMENT_ID AS INT) AS IMPROVEMENT_ID,
    IMPROVEMENT_ID_TEXT,

    -- Clean cost (remove weird characters)
    REPLACE(REPLACE(INDICATIVE_COST, 'Â', ''), CHAR(160), '') AS INDICATIVE_COST_CLEAN

INTO dbo.clean_recommendations
FROM dbo.recommendations;
GO
```

2) Create clean_recommendations table

Explanation:

This step cleans and standardises the recommendations dataset.

Activities carried out:

> Drop existing cleaned table
Ensures clean rebuild

> Transform columns
Converting the ID and improvement item fields to INT

Long text kept safely

> Clear the cost field

eliminates corrupted characters like ASCII 160 and "@"

This is required because encoding errors are frequently present in cost values.

> Output table

Results that have been cleaned are saved as clean_recommendations.
Accurate joins in Power BI are made possible by integrating clean certificates and clean recommendations.

Step 4 - Validate the cleaning

```
--STEP 4_VALIDATE THE CLEANING

--1)Check cleaned tables
SELECT TOP 20 * FROM dbo.clean_certificates;
SELECT TOP 20 * FROM dbo.clean_recommendations;

--2)Count rows
SELECT COUNT(*) FROM dbo.clean_certificates;
SELECT COUNT(*) FROM dbo.clean_recommendations;

--3)Check numeric conversion success
SELECT
    SUM(CASE WHEN CURRENT_ENERGY_RATING IS NULL THEN 1 END) AS missing_energy_rating
FROM dbo.clean_certificates;
```

Verify the Cleaning Justification

1) See the tables that have been cleaned
The sample cleaned data is shown in SELECT TOP 20.

helps verify that formatting and conversions are accurate.

2) Count the rows
confirms how many records are in the clean tables.

guarantees that there is no unintentional data loss or duplication

3) Verify the success of the conversion
Verifies that the energy rating values were cast successfully

determines whether a breakdown happened during the data cleaning process.

Before beginning the modeling process, this validation step makes sure that the cleaning was done correctly.

Step 5 - Join both cleaned tables

```
--STEP 5_JOIN BOTH CLEANED TABLES
SELECT
    c.LMK_KEY,
    c.ADDRESS,
    c.POSTCODE,
    r.IMPROVEMENT_SUMMARY_TEXT,
    r.INDICATIVE_COST_CLEAN
FROM dbo.clean_certificates c
LEFT JOIN dbo.clean_recommendations r
    ON c.LMK_KEY = r.LMK_KEY;

SELECT*
FROM clean_certificates;
```

Join Both Cleaned Tables

Explanation:
This query performs an LEFT JOIN using LMK_KEY

Ensures all certificates are kept, even if no recommendations exist

obtains cleaned cost ranges and improvement summaries.

Goal:
incorporates data about improvements into the primary dataset

enables Power BI's combined analysis.

Verifies that LMK_KEY functions properly following cleaning

Final select code
shows the complete table of cleaned certificates.

Beneficial for examination prior to exporting to Power BI

4. Power BI Modeling

The next step involved creating a thorough analytical model and dashboard in Microsoft Power BI after the SQL-based data cleaning and integration was finished. Power BI was chosen because of its sophisticated visualization features, support for relational modeling, DAX computations, mapping tools, and capacity to manage big datasets effectively. This step's objective was to turn the cleaned EPC datasets into insightful visuals that could help with energy policy, sustainability planning, and property efficiency enhancement decision-making.

Step 1: Establish a connection to SQL Server and load the initial data

>Choose Get Data -> SQL Server Database after launching Power BI Desktop.

>Enter the database name (e.g., Manchester_EPC) and your server name.

>For performance, select Import mode. Unless your dataset is very large, this is typically faster for analysis than DirectQuery.

>Choose the clean tables and views you made in the Navigator window:
Clean Certificates
Clean_Recommendations
V_Certificates_Filtered_for_Trend (for the visual representation of the time series)

>Press the "Load" button.

Step 2 : Power BI Data Preparation

Additional transformations were applied using Power Query Editor, such as:
>Taking the Assessment Year Out of the Date of Inspection
>Standardizing categorical fields (such as built form and property type)
>Establishing categories for emissions: Very Low, Low, Medium, High, and Very High
>Fixing redundant or empty descriptive fields
>Changing column names to make them easier to read

When creating visuals, these changes guaranteed consistency and improved usability.

Step 3 : Dax calculations

Page 1 - Dax 1

```
Total Recommendations = COUNTROWS(clean_recommendations)
```

The DAX code creates a new measure called Total Recommendations. This measure totals the number of rows in the table clean_recommendations by the use of the COUNTROWS function. It serves as a tool to swiftly figure out the total number of energy-saving suggestions documented in the data source.

dax 2

```
Percentage With Improvement =
VAR Total =
    COUNTROWS(clean_certificates)

VAR Improved =
    CALCULATE(
        COUNTROWS(clean_certificates),
        FILTER(
            clean_certificates,
            clean_certificates[POTENTIAL_ENERGY_EFFICIENCY] >
            clean_certificates[CURRENT_ENERGY_EFFICIENCY]
        )
    )

RETURN
DIVIDE(Improved, Total)

```

The DAX code is setting up two measures: Total Recommendations and Percentage With Improvement.

Total Recommendations is a measure that counts the number of rows in the clean_recommendations table by using COUNTROWS. Hence, it just returns the total number of energy-saving recommendations that appear in the dataset.

The Percentage With Improvement measure identifies the share of energy performance certificates that have a potential energy efficiency score higher than the current one, thus, showing that an improvement can be made. It does so by first defining the Total count of all certificates and an Improved count (through CALCULATE with a FILTER condition) and then applying DIVIDE to get the percentage of properties where a potential improvement is available.

dax 3

```
1  Total EPC Certificates = COUNTROWS(clean_certificates)
2
```

The figure visualized the components and operations of the simple DAX formula delineated a measure called Total EPC Certificates that is used to calculate the total number of rows in the table named clean_certificates using the COUNTROWS function, thereby, the measure is reckoned to be the total EPC Certificates that have been accounted for in that cleaned dataset.

Page 2 - Dax 1

```
1  Average Current Efficiency = AVERAGE(clean_certificates[CURRENT_ENERGY_EFFICIENCY])
2
```

The picture shows a straightforward Data Analysis Expressions (DAX) formula that generates a measure called Average Current Efficiency. The measure figures out the average of all the numbers in the [CURRENT_ENERGY_EFFICIENCY] column of the clean_certificates table by means of the usual AVERAGE function.

dax 2

```
1  Average Potential Efficiency = AVERAGE(clean_certificates[POTENTIAL_ENERGY_EFFICIENCY])
2
```

The DAX formula illustrated is the one that figures out the measure Average Potential Efficiency. To do this, it averages the values in the column POTENTIAL_ENERGY_EFFICIENCY of the table clean_certificates.

This measure is normally implemented in Power BI or Excel Data Models to efficiently provide a brief numerical description of the central tendency of a potential energy efficiency metric over the entire certificate records.

dax 3

```
1  Average CO2 Emissions = AVERAGE(clean_certificates[CO2_EMISSIONS_CURRENT])
2
```

The DAX formula sets up the measure Average CO2 Emissions.This measure is computed by averaging the figures in the column CO2_EMISSIONS_CURRENT of the table clean_certificates.Such a measure in a data model is the way to give back one number that represents the whole current average carbon dioxide emissions of all records from the specified table.

Page 3 - Dax 1

```
1  Average Heating Cost Current = AVERAGE(clean_certificates[HEATING_COST_CURRENT])
```

The DAX expression displayed is used to compute a variable called Average Heating Cost Current. It does so by averaging the values of the HEATING_COST_CURRENT column in the clean_certificates table. The purpose of this variable is to provide a fast overview of the situation of heating costs based on the current data at hand in the data model.

dax 2

```
1  Average Heating cost Potential = AVERAGE (clean_certificates[HEATING_COST_POTENTIAL])
```

This metric finds the average value from the [HEATING_COST_POTENTIAL] column in the clean_certificates table, by invoking the AVERAGE function. In other words, it is figuring out the standard or average heating cost potential that the data indicate in that particular column of the table.

dax 3

```
1  Average Lighting Cost Current = AVERAGE(clean_certificates[LIGHTING_COST_CURRENT])
```

This formula utilizes the AVERAGE function to determine the arithmetic mean of all the non-null values in the [LIGHTING_COST_CURRENT] column which is a part of the clean_certificates table. Put simply, this measure is figuring out the average of the current lighting cost from the data.

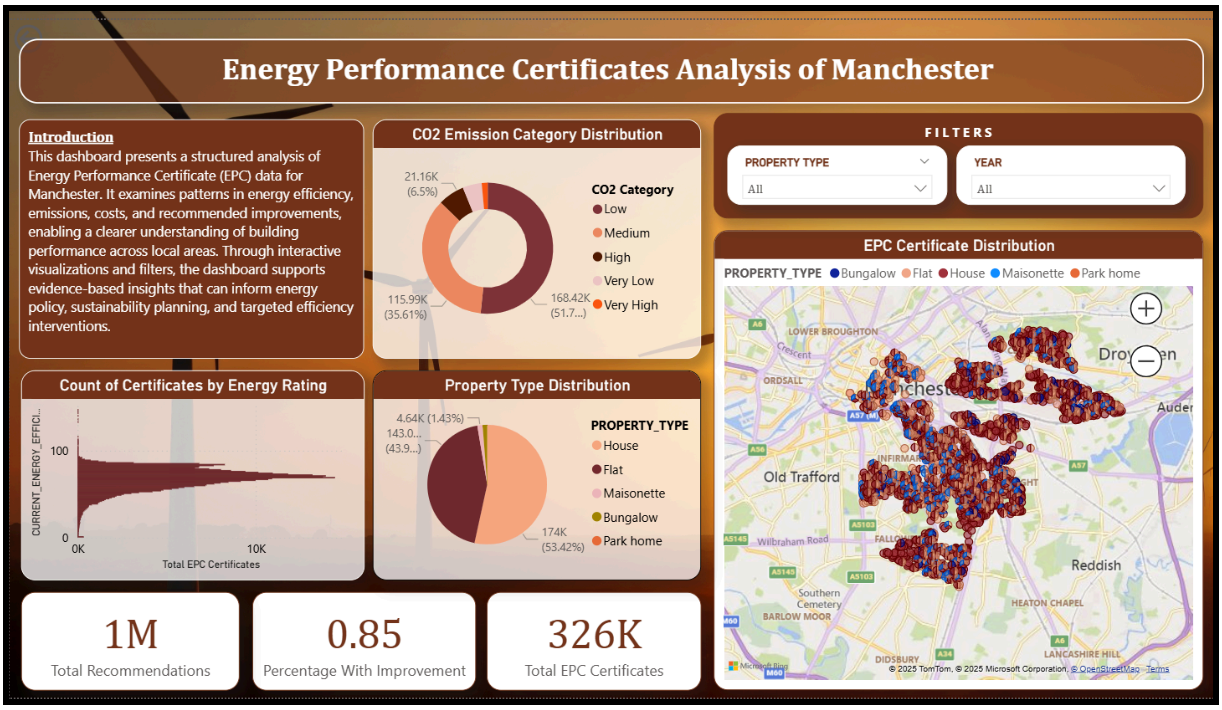
dax 4

```
Average Lighting Cost Potential = AVERAGE(clean_certificates[LIGHTING_COST_POTENTIAL])
```

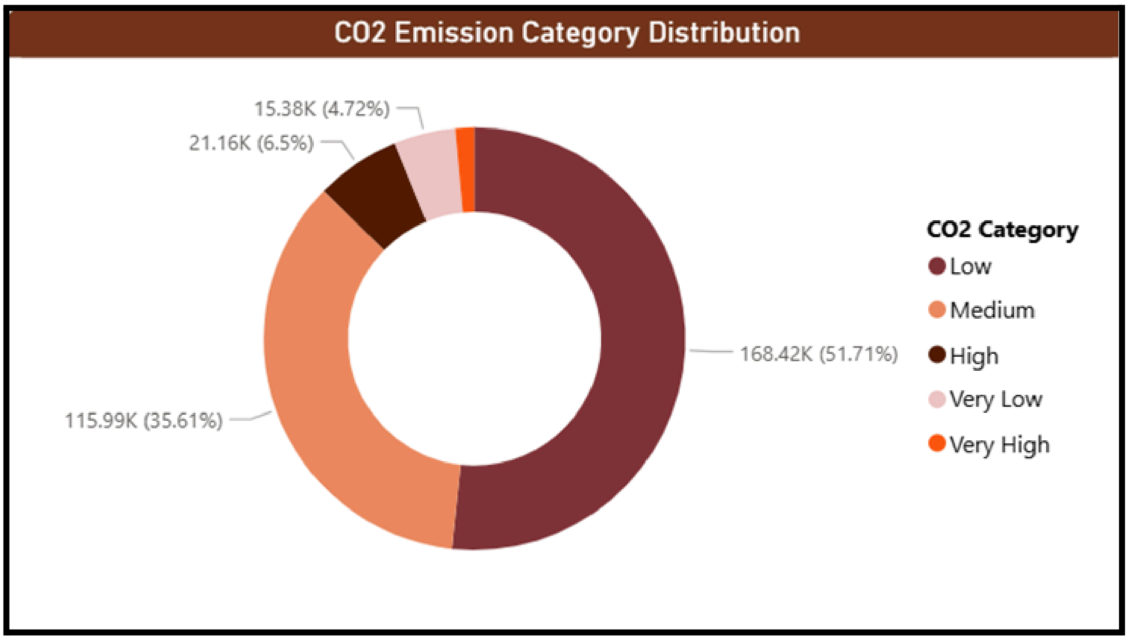
The Average Lighting Cost Potential is calculated by this line of code. It does so by taking the simple average of the numerical values in the column named LIGHTING_COST_POTENTIAL, which are part of the data set or the table called clean_certificates.

5. Data visualizations

Dashboard 1 (page 1)- Comparison and overview

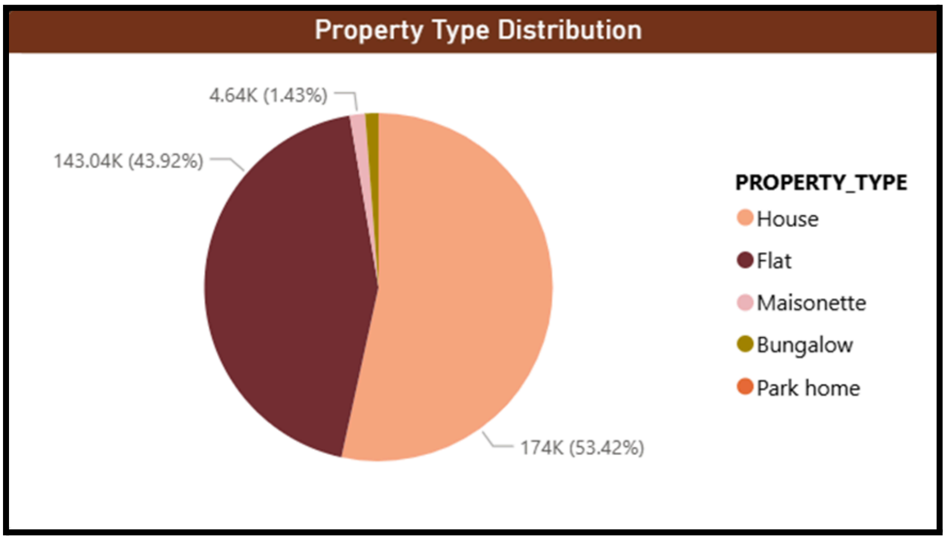


CO_2 Emission Category Distribution (Donut Chart)



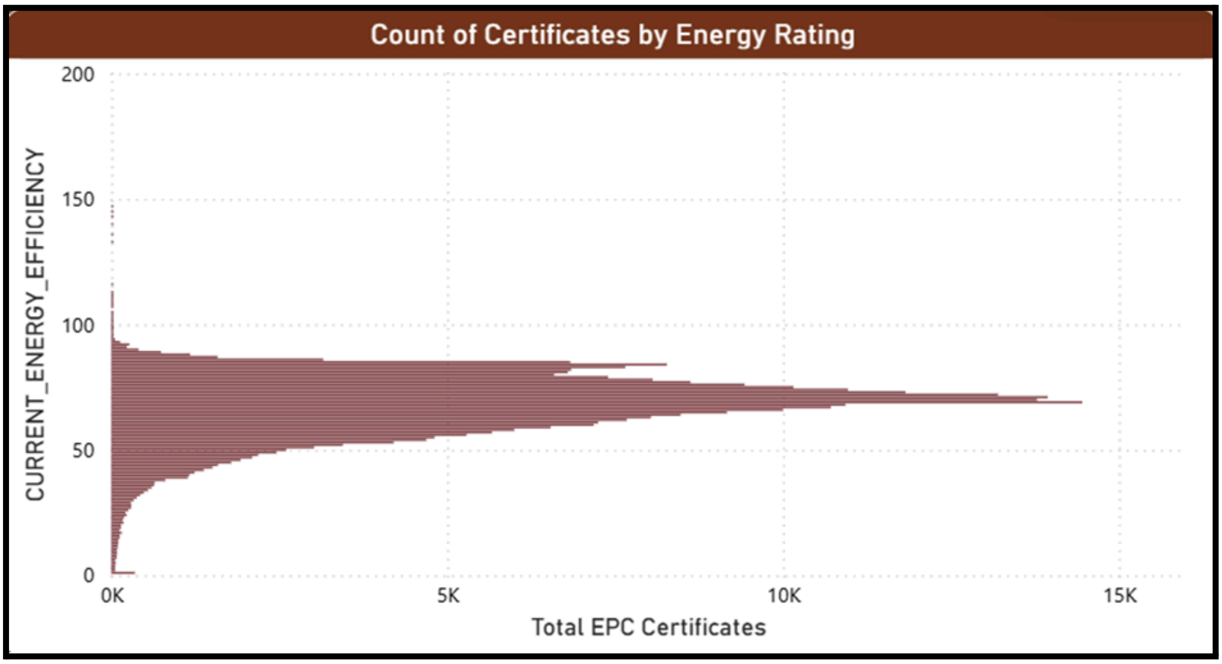
Manchester's carbon footprint is powerfully represented proportionally in the CO2 Emission Category Distribution donut chart, which was created using the custom CO2 Category calculated column (a DAX transformation). The analysis shows a remarkable concentration in the mid-to-high emission bands, with the Medium category contributing an additional 115.99K properties (35.61%) and the High category dominating at 168.42K properties (51.71%). These two groups together make up more than 87% of the entire dataset, indicating that the main environmental problem is systemic and stems from the great majority of housing stock. Policy efforts should concentrate on large-scale intervention strategies aimed at decarbonizing the high-volume, middle-tier properties rather than addressing niche outliers, as confirmed by the minimal representation of the extreme categories (Very Low and Very High).

Property Type Distribution (Pie Chart)



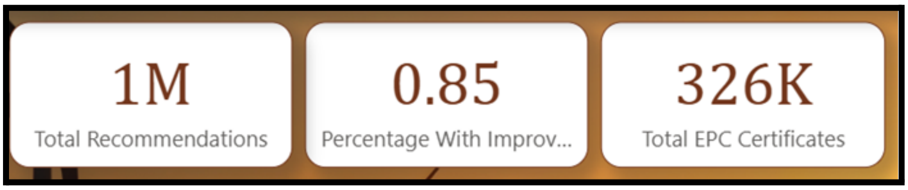
Houses and apartments are the two most prevalent types of dwelling in the Manchester EPC dataset, as shown by the Property Type Distribution pie chart. With 174K records, or 53.42% of the total, houses are the most common type. Flats come in second with 143.04K records, or 43.92%. These two property types together account for almost 97.5% of all inspected homes, demonstrating that Manchester's energy efficiency problem is primarily concentrated in these areas. On the other hand, property types like Park Home, Bungalow, and Maisonette (4.64K, 1.43%) are statistically insignificant parts of the analysis. Instead of requiring general solutions, this strong concentration enables policy interventions and financial modeling to be narrowly focused on the particular features and retrofit strategies appropriate for houses and apartments.

Count of Certificates by Energy Rating (Horizontal Bar Chart)



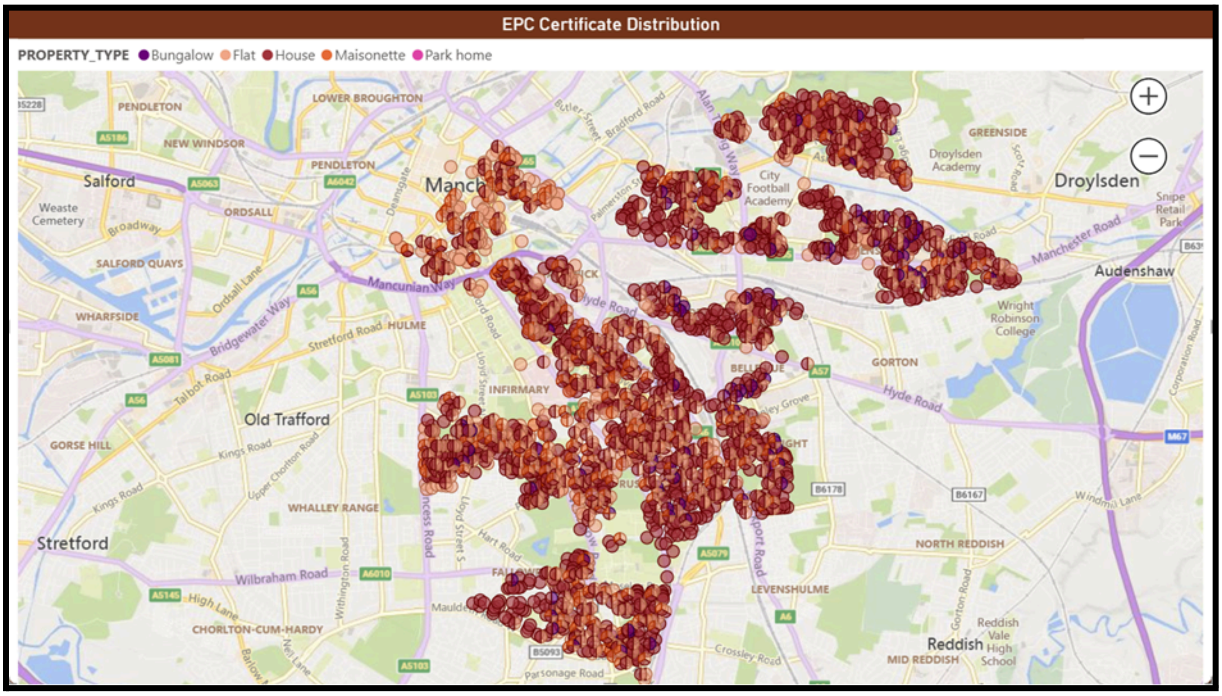
The Count of Certificates by Energy Rating graphic, which functions as a detailed histogram that plots the number of EPCs against their particular efficiency score (0–100), offers a detailed view of efficiency scores across the entire dataset. This chart's most important finding is the highly skewed distribution, with the majority of the data concentrated in the mid-range scores (roughly 68 to 78), which primarily correspond to Band D's high end and Band C's average. This concentration shows that homes with average energy performance make up the majority of Manchester's housing stock rather than the extremes of highly efficient (scores 85+) and severely inefficient (scores below 50).This dense grouping confirms that the policy objective should be centered on a mass-upgrade strategy to lift this large middle segment into the Band B and A categories.

KPI Summary Cards (Top Row)



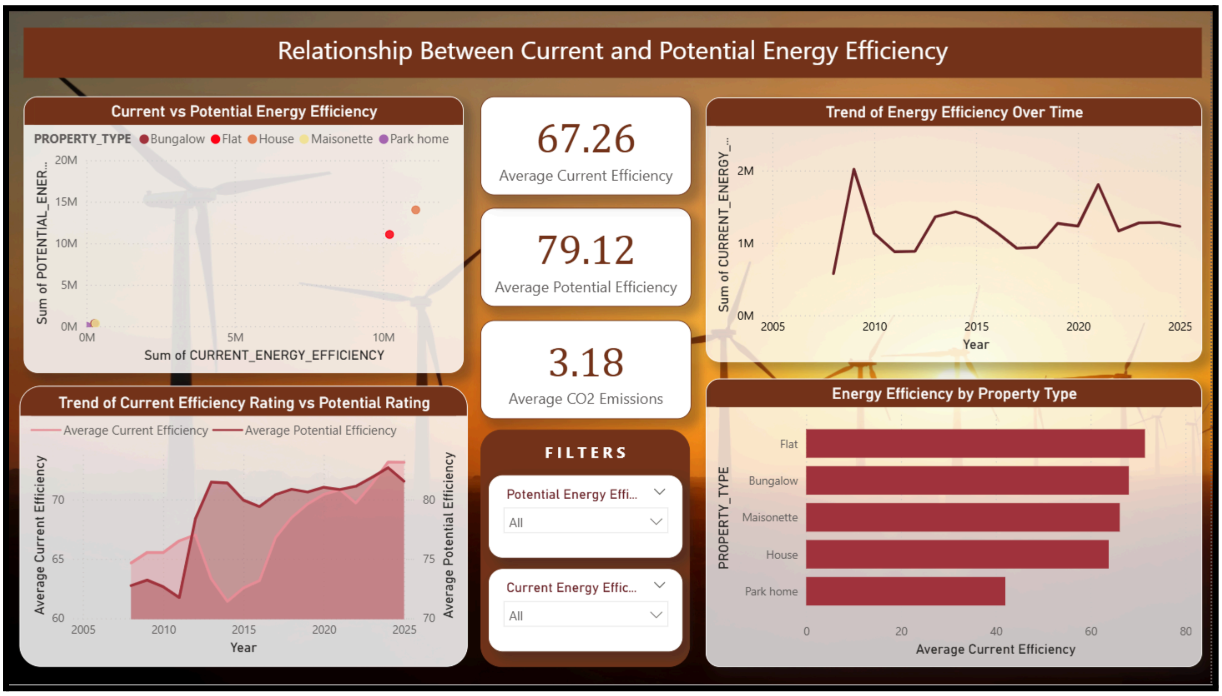
The picture shows a KPI card chart indicating the following metrics: Total Recommendations- 1M, Percentage With Improvement-0.85 (most probably a decimal fraction of 85%), and Total EPC Certificates- 326K. Such a visual solution, called Scorecard Display or KPI Card Chart, is ordinarily employed in the dashboards to communicate the core performance data in a very brief and clear manner.

EPC Certificate Distribution (Map Visual)

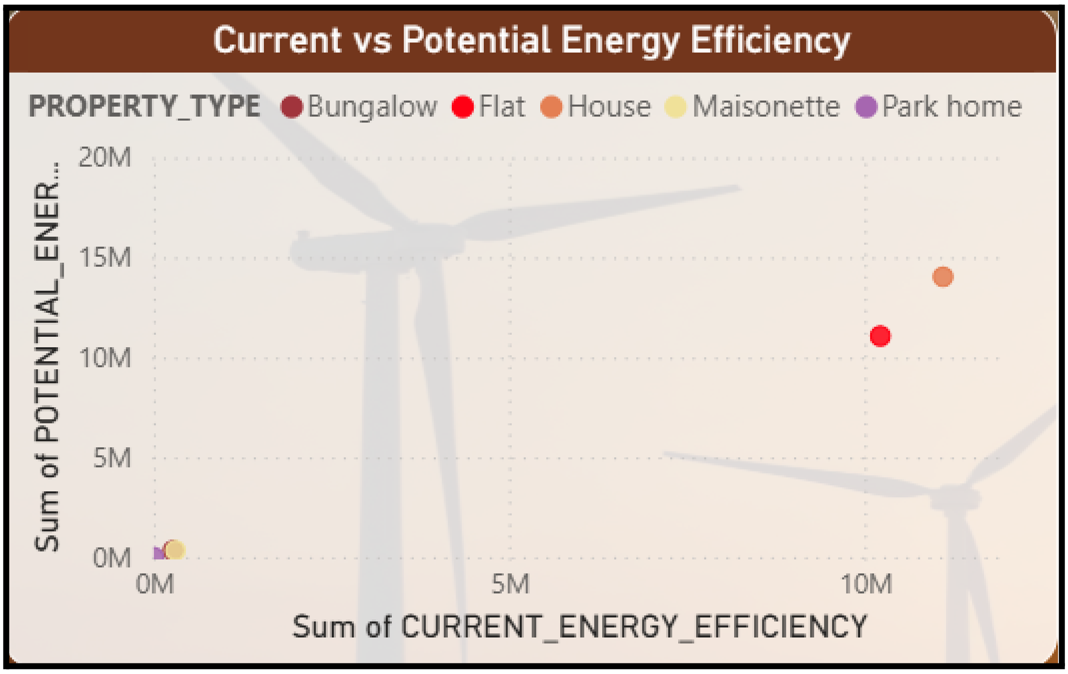


In order to identify spatial trends in inspection activity, the EPC Certificate Distribution map visual, a crucial part of geographical analysis, plots individual properties throughout Manchester. Dense clusters of red dots on the map indicate a high concentration of EPC data points in central urban and inner-city areas, especially in the vicinity of Manchester Central, Salford Quays, and high-density residential neighborhoods. Areas with high rental turnover and contemporary flat construction, which requires regular EPC inspections, are the main causes of this density. In contrast, the distribution of inspections is less dense even though they are dispersed throughout the larger suburban ring (e.g., Droylsden, Stretford).This visual serves to localize the data, demonstrating that while inspection activity is centered downtown, further analysis is required to cross-filter this map by Energy Rating to identify the lower-efficiency housing pockets typically found in the surrounding older residential suburbs.

Dashboard 2 (page2)- Comparison and overview

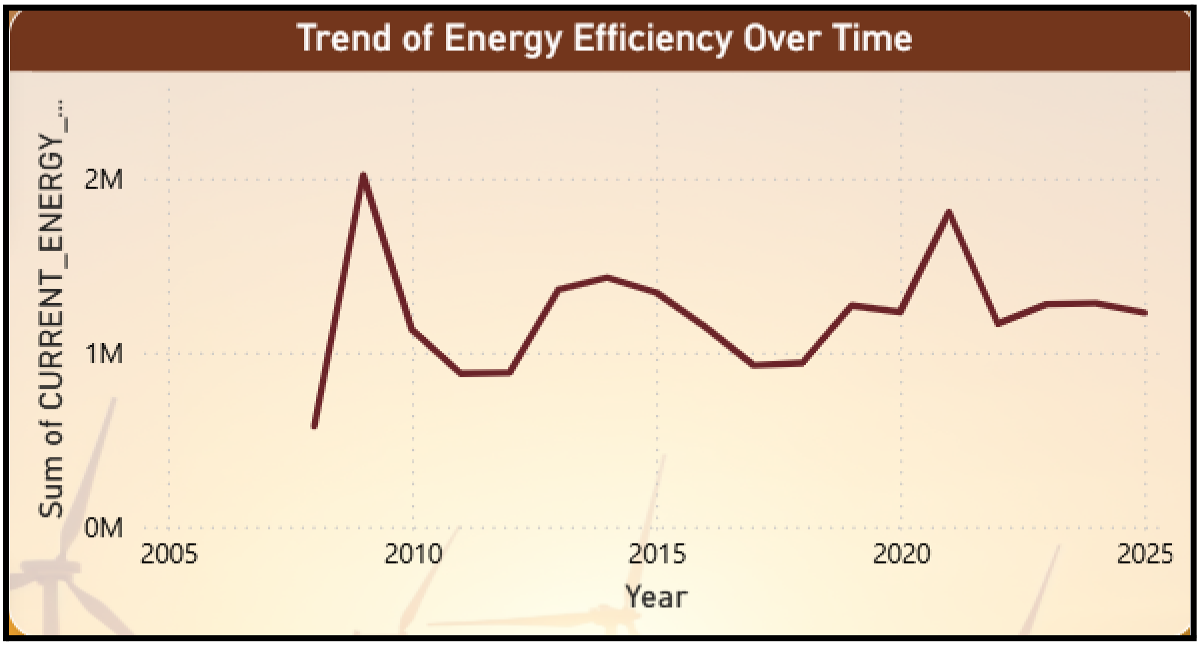


Current vs potential efficiency (Scatter plot)



This image illustrates a scatter graph named "Current vs Potential Energy Efficiency" that contrasts various property types' "Sum of CURRENT_ENERGY_EFFICIENCY" (on the x-axis) with the "Sum of POTENTIAL_ENERGY_EFFICIENCY" (on the y-axis). According to the legend, there are five property types - Bungalow, Flat, House, Maisonette, and Park home. The graph suggests that Flats and Houses are the two property types, which energize both current and potential energy efficiencies to the highest extent, however, the potential efficiency of a House is just a bit higher (approximately 14M) while that of a Flat is a little lower (around 11M) as compared to their current efficiency levels (both about 10M). Moreover, the three remaining property types (Bungalow, Maisonette, and Park home) are shown to be very close to the origin (0M, 0M), thus, their total values in both metrics are implied to be significantly low.

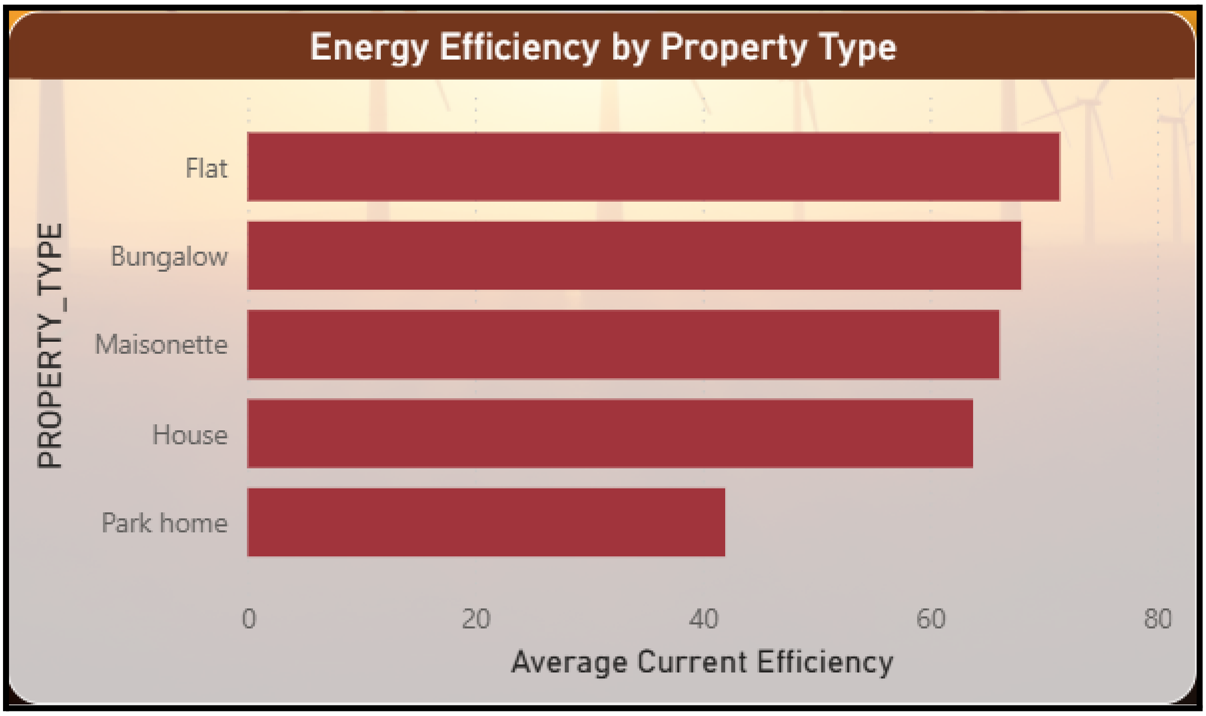
Trend energy efficiency over time (Line chart)



The chart shows the trend of energy efficiency in a line chart titled "Trend of Energy Efficiency Over Time." It visualizes the changes in the "Sum of CURRENT_ENERGY_..." (units are cut off on the Y-axis) over time, the years from about 2005 to 2025.

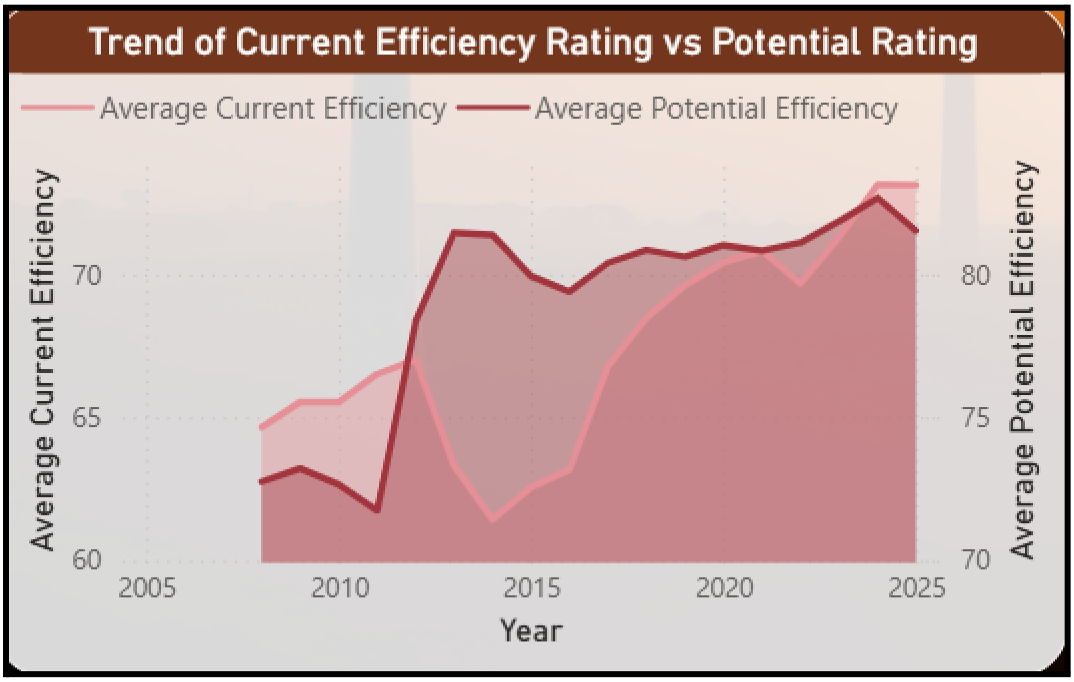
The trend is very unstable. The energy efficiency, after being at a fairly low level around 2008, was quickly raised to more than 2 million (units) around 2009. After this maximum, the energy efficiency was rapidly decreased to nearly half until 2011 and from there it remained constant at about 1 million (units). The efficiency has been going up most of the time after that, reaching a stage close to 1.5 million (units) between 2013 and 2015. After falling back to a bottom around 2017, the trend showed a rise and a second, somewhat lower peak of just under 2 million (units) around 2021. In the last years shown (2022-2025), energy efficiency is at a standstill and stays at just over 1 million (units).

Energy efficiency by property type (Horizontal bar chart)



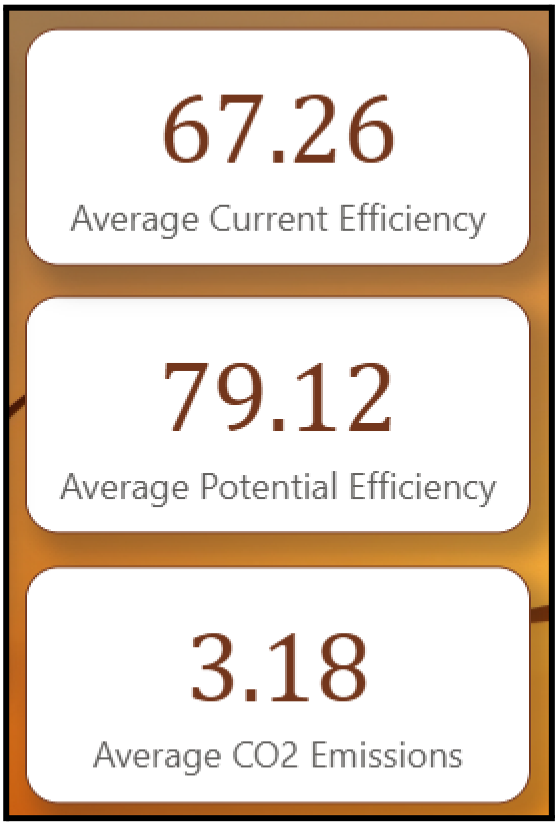
The chart shows the average current efficiency of five different property types. Flats are the most efficient on average, with an average current efficiency of about 73. Bungalows and Maisonettes are next in line with their average efficiencies in the mid-to-high 60s (Bungalows around 68, Maisonettes around 66). The energy efficiency of houses is a bit lower, around 63. Lastly, Park homes have the lowest average energy efficiency, at a level of around 42, which is much lower than the rest of the properties types.

Trend of current efficiency rating vs potential rating



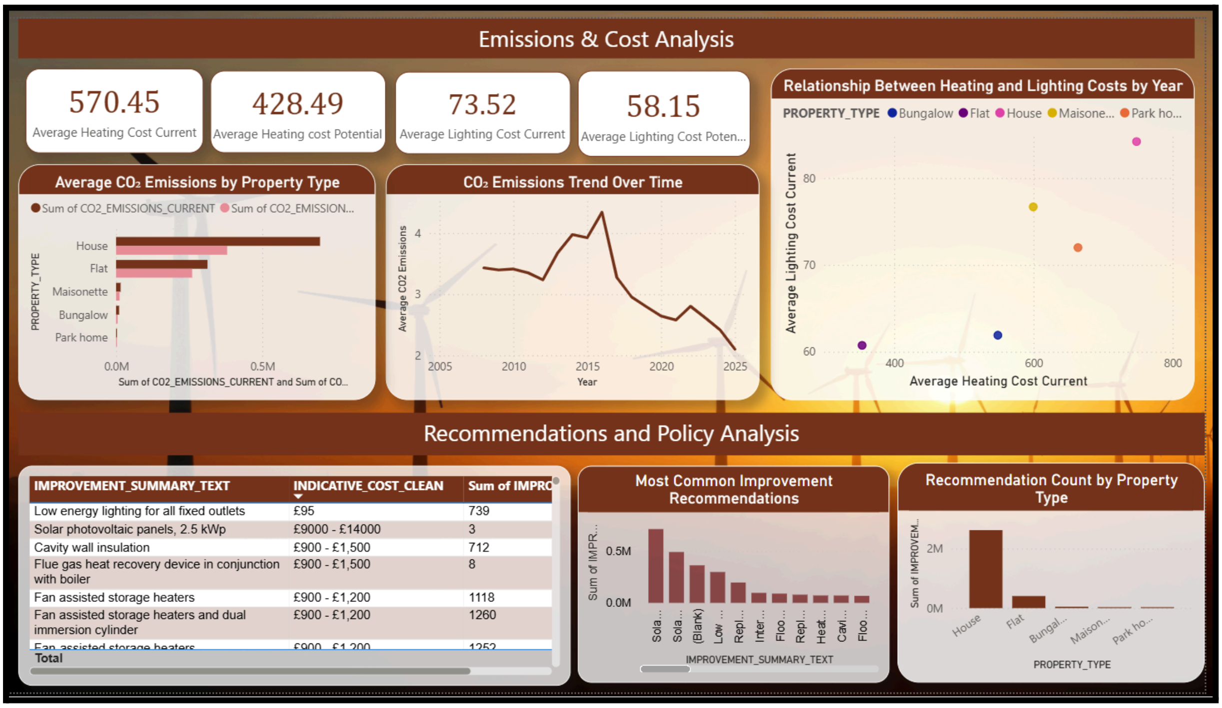
The plot compares the Average Current Efficiency(light red line/area, measured on the left Y-axis) against the Average Potential Efficiency (dark red line/area, measured on the right Y-axis, which seems to share the same scale as the left). The Average Potential Efficiency is higher than Average Current Efficiency for the whole time, showing that there is still a difference between actual and achievable energy performance. Both measures have been mostly going up over the period, with the Current Efficiency moving from about 65 to 82, and the Potential Efficiency from about 63 to 83. The widest gap between the two can be seen at around 2017, where the Current Efficiency was lowered while the Potential Efficiency kept going up.

KPI summary cards

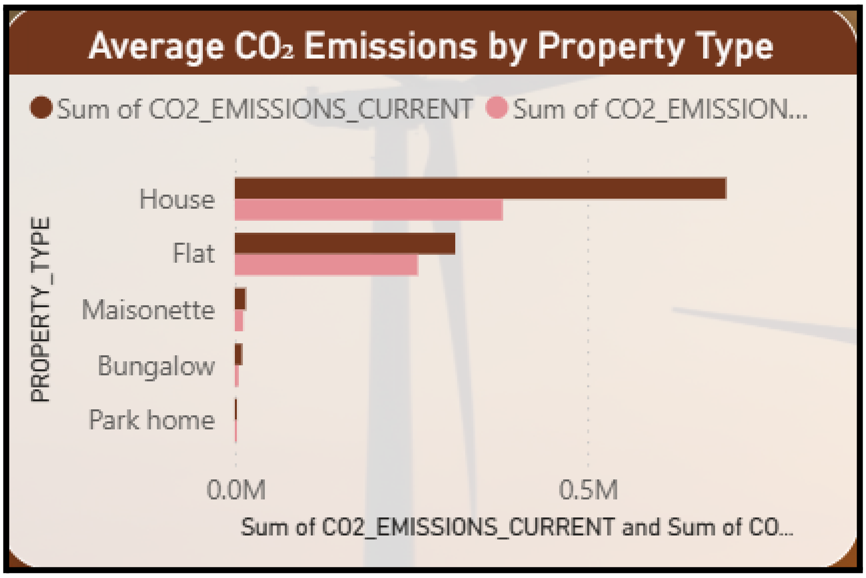


This picture exhibits three main performance measures (KPIs) through a card-based dashboard layout. It is a typical manner of showing numerical summaries for digital applications. The figures revolve around efficiency and the environment. To be exact, it compares an Average Current Efficiency of 67.26 with an Average Potential Efficiency of 79.12, thus, it indicates a significant difference or simply an area of performance that needs to be improved. The third indicator is Average CO_2 Emissions with the value of 3.18, which is most probably an environmental cost resulting from the operation or the system that is being measured.

Dashboard 3 (page3)- Comparison and overview

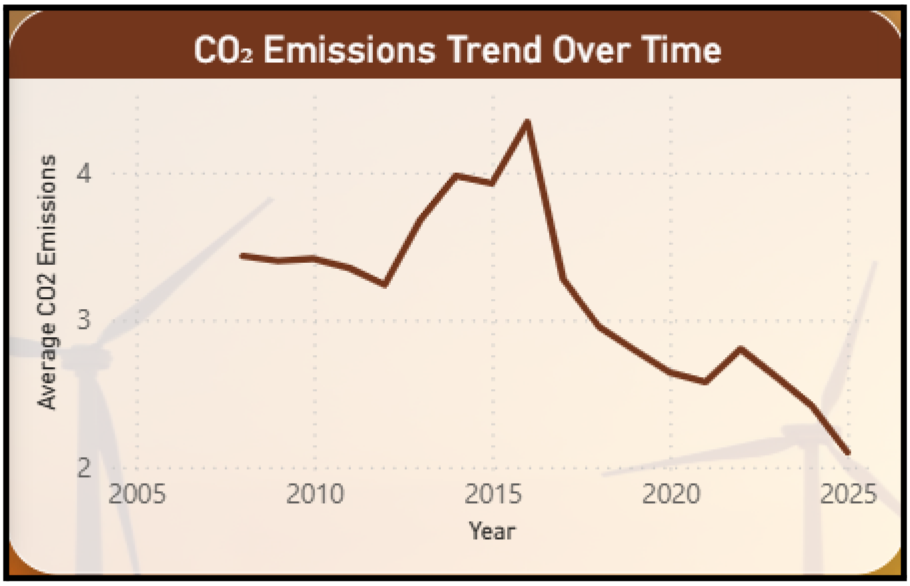


Average CO2 emissions by property type



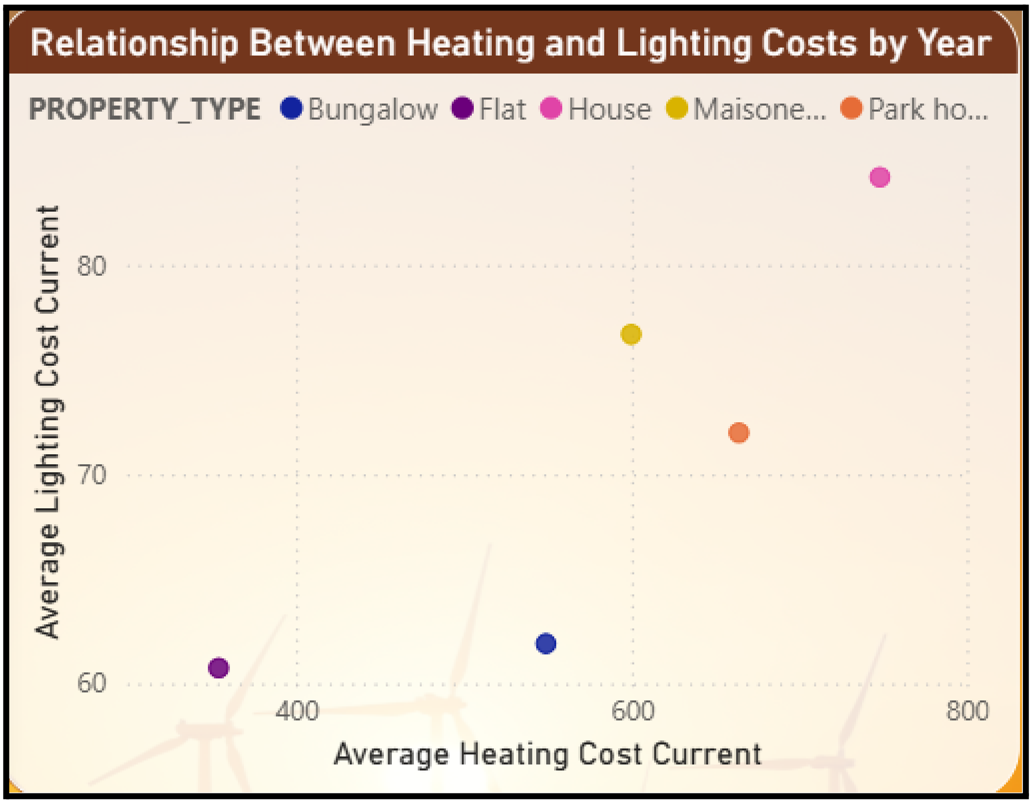
The graph contrasts the "Sum of CO_2 EMISSIONS_ CURRENT (shown by the dark brown bars) with the "Sum of CO__EMISSION (shown by the lighter pink bars) for each of the following five property types: house, flat, maisonette, bungalow, and park home. Houses have the highest average CO_2 emissions for both metrics; the current emissions (dark brown bar) are much higher than the historical or comparable pink bar for Houses and significantly higher than all other property types. The second-highest emissions are found in apartments, which are higher than the other types of properties but significantly lower than those in houses. When compared to houses and apartments, the average CO_2 emissions from maisonettes, bungalows, and park homes are all quite small.

CO2 emissions trend over time (Line chart)



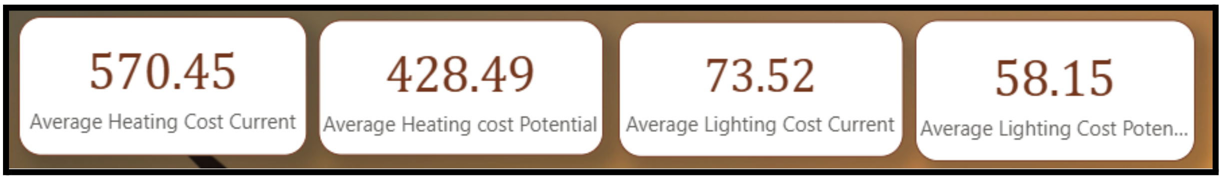
The average CO_2 emissions between 2008 and 2015 generally increased after a period of relative stability, reaching a peak in 2016 at a value marginally above 4 units. After this peak, there is a precipitous drop that lasts steadily until about 2020. Emissions then experience a brief, slight recovery around 2022 before resuming a final, significant downward trend, reaching their lowest point on the chart by 2025, near a value of 2 units.

Relationship between heating and lighting costs by year (Scatter plot)



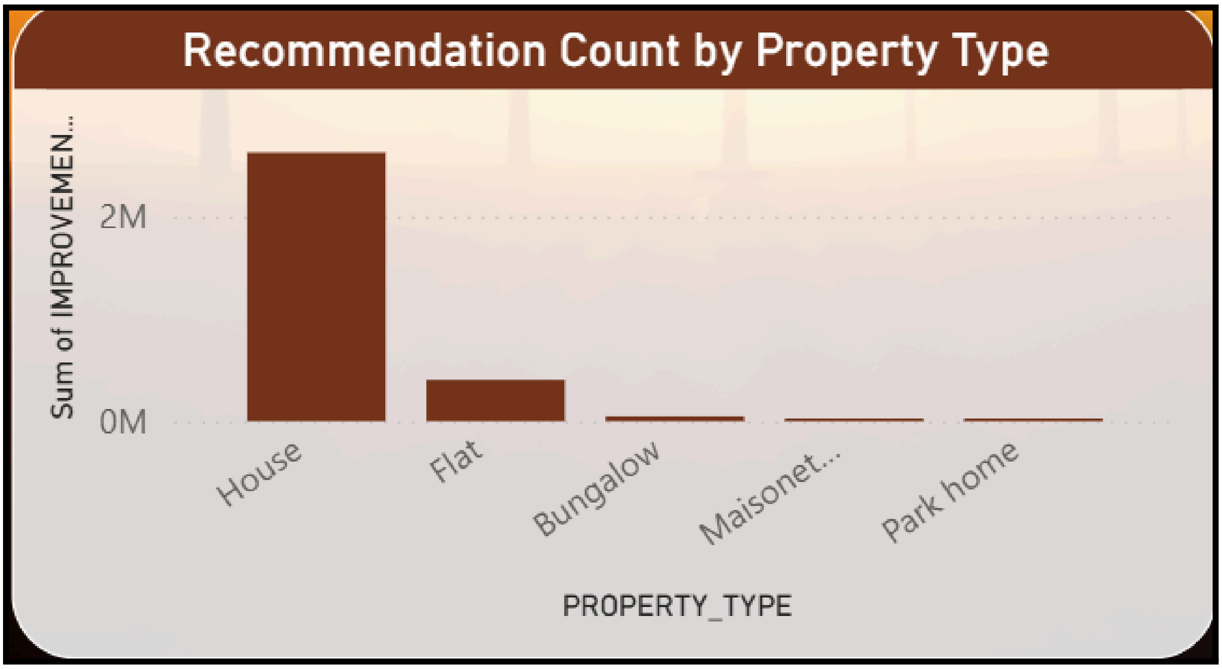
The plot reveals an overall positive correlation between the two costs, properties with higher heating costs are those that also have higher lighting costs. Homes (pink dot) record the highest situation for both heating (around 750) and lighting (around 85), thus, they are in the top-right corner. Maisonettes (yellow dot) and a property type indicated by an orange dot (probably a Park home or Bungalow based on the previous chart, but not explicitly labeled) are coming halfway through the range. Apartments (purple dot) are having the least heating and lighting costs out of the properties plotted, thus, they are situated in the bottom-left corner. The Bungalow (blue dot) demonstrates a relatively low lighting cost (a little over 60) but a somewhat higher heating cost (around 550) than that of Flats, thus, it is indicating a deviation from the positive trend for this property type.

KPI summary cards



The information contrasts "Potential" and "Current" costs for lighting and heating. The potential heating cost is substantially lower at $428.49, whereas the current average heating cost is the highest at $570.45. In a similar vein, the average cost of lighting is currently $73.52, but it could be as low as $58.15. Overall, if the potential efficiency targets are reached, the figures point to a significant opportunity for cost savings, especially in heating.

Recommendation count by property type (Bar chart)



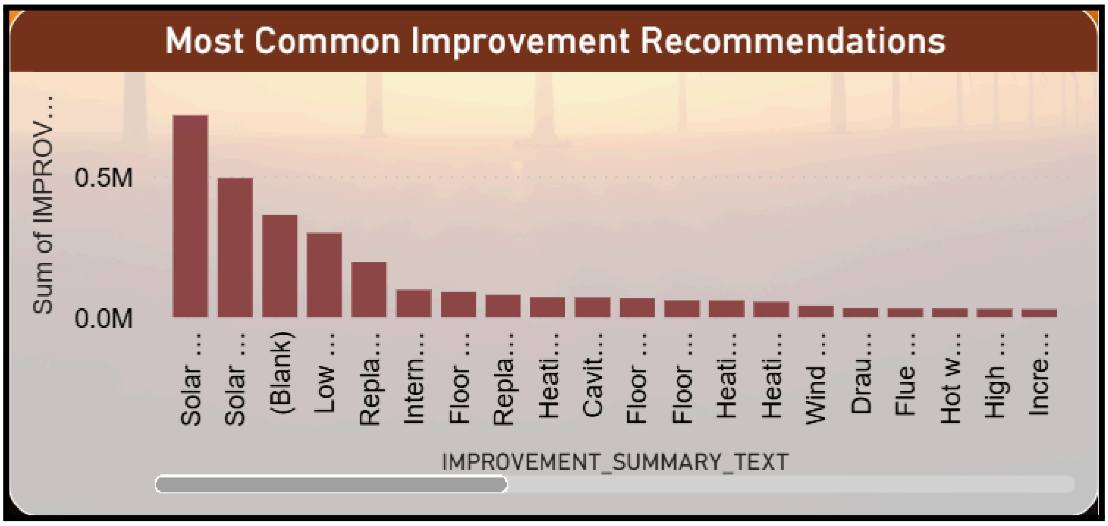
shows a bar graph called "Recommendation Count by Property Type," which shows the total number of enhancements (probably suggestions for energy-saving or related actions) for each of the five different types of properties. All other property types—Flat, Bungalow, Maisonette, and Park home—show a much lower count, with Flats being the second highest at a fraction of the House count, and the remainder showing negligible amounts. Houses, on the other hand, have the vast majority of recommendations, exceeding 2 million.

Energy improvement summary (Table chart)

IMPROVEMENT_SUMMARY_TEXT	INDICATIVE_COST_CLEAN	Sum of IMPROVEMENT_ITEM
Low energy lighting for all fixed outlets	£95	739
Solar photovoltaic panels, 2.5 kWp	£9000 - £14000	3
Cavity wall insulation	£900 - £1,500	712
Flue gas heat recovery device in conjunction with boiler	£900 - £1,500	8
Fan assisted storage heaters	£900 - £1,200	1118
Fan assisted storage heaters and dual immersion cylinder	£900 - £1,200	1260
Fan-assisted storage heaters	£900 - £1,200	1252
Total		3094791

The graphic illustrates a number of energy efficiency enhancements, along with the total number of recommendations (Sum of IMPROVEMENT_ITEM) and the indicative clean cost for each. While solar panels are the most expensive option, fan-assisted storage heaters and low-energy lighting are the most often suggested.

Most common improvement recommendations (Bar chart)



Most Common Improvement Recommendations is a bar chart that shows the frequency of different suggestions for improving energy efficiency. Solar installations (shown by the first three bars) are by far the most often suggested improvements, followed by low-energy lighting. Other typical suggestions include different types of insulation and replacement (such as cavity, floor, and internal).

6. Discussion on power BI features used

The analysis made use of a wide range of sophisticated Power BI and T-SQL features.

For Data Integration, the team used the Power BI SQL Server connector, particularly Import mode, to connect the cleaned EPC dataset from the SQL Server database (local or Azure). This ensured a quick visual performance.

Data Transformation heavily relied on the Power Query Editor, where data cleaning took place. This included dealing with missing values in CO_2 emissions and costs fields, as well as ensuring correct typing of the data. At this point, additional necessary Calculated Fields were generated, for example, the efficiency score difference.

DAX (Data Analysis Expressions) was the main driver for the analytical depth: it facilitated the computation of Key Performance Indicators (KPIs), which were made very visible single-value cards for metrics such as Average Heating Cost Current (570.45), Average Lighting Cost Current (73.52), Average Current Efficiency (67.26), and Average Potential Efficiency (79.12)3.

In addition, the advanced DAX functions were indispensable for Time Intelligence, thus allowing the plotting of the Energy Efficiency Over Time trend as well as the comparison of Current vs. Potential Efficiency Ratings from 2005 to 2025.

At last, the implementation of Slicers/Filters (Property Type, Energy Efficiency ratings) and Cross-Filtering significantly contributed to the Interactivity and User Experience enhancement, thus users were empowered to explore data dynamically.

Different Appropriate Visuals were utilized such as Bar Charts (for emissions by property type), Line Charts (for trends), and Scatter Plots (for cost relationship analysis).

7. Results analysis and discussion

The analysis of the Energy Performance Certificate (EPC) data for Manchester shows a lot of information which can be used to understand how the city's buildings perform. The EPC Overview highlights an extensive range of interventions, with a sum of 326K EPC Certificates and an incredible 85% of properties being recognized as those that can improve their energy efficiency. In terms of distribution, the largest share of properties is houses, which make up 53.42% of certificates, thus becoming the main target of the energy initiatives on the large scale.

“The Relationship Between Current and Potential Energy Efficiency” picturises a very energetic move, showing a significantly average efficiency gap of almost 12 points (Current 67.26 vs. Potential 79.12), thus providing a strong indication that a huge increase in energy performance can be realized by implementing the recommendations. Data by property types unveils that Park Homes are the ones with the least Average Current Efficiency (close to 40), hence they should be the ones receiving specially designed, carefully planned strategies, in contrast with the generally upward trend of efficiency from 2005 to 2025.

The Emissions & Cost Analysis opens the way to the idea that reductions in expenditure should be mainly heating-focused, as it is the main driver of the cost at an average of £570.45 per year, which is more than seven times the cost of lighting, which is only £73.52. Though the CO_2 Emissions Trend has an overall downward trajectory, the most common recommendations by volume are inexpensive measures such as Low energy lighting (739 instances) and Cavity wall insulation (712 instances), while in the same time, high-cost items like Fan assisted storage heaters and Solar photovoltaic panels are also frequently suggested for the maximum carbon footprint reduction.

These Key Findings are the main arguments for the existence of high-impact interventions that can simultaneously be cost-effective and absolutely necessary for achieving sustainable development goals.

8. Key findings

High Intervention Potential: 85% of properties have a clear energy efficiency improvement potential, which is an essential factor in the establishment of city-wide intervention programs.

Heating as the Primary Cost Driver: On average, the annual heating cost is substantially higher than the lighting one, which is the main reason why households should focus on measures that reduce heat loss and improve heating systems if they want to lower their running costs.

Property Type Disparity: Detached houses are the most significant volume of the intervention while Park Homes and old properties have been identified as the lowest energy efficiency levels currently.

Positive Efficiency Trend: Energy efficiency (both Current and Potential) has generally been improving over the years from 2005 to 2025, which is an indication that the present policies and voluntary actions are resulting in positive changes.

9. Impact of analysis and recommendations

An interactive Power BI dashboard is a very important instrument for stakeholders in the energy sector, public authorities, and property owners.

Targeted Policy Making: The data is a direct driver of policy decisions as it quantifies the number of properties that would gain from certain interventions (e.g., It is one of the most common recommendations for insulation of the cavity walls) (page 3.).

Investment Prioritization: The recommendations may first be implemented based on the most cost-effective and highest-volume ones. For example, low energy lighting is a high-volume, low-cost intervention.

Stakeholder Communication: The visual and interactive nature of the dashboard makes it possible for non-technical stakeholders (homeowners, council members) to understand at a glance complex trends in CO_2 emissions and energy costs.

>most significant insights and any actionable recommendations for stakeholders.

The analysis conveys three Most Significant Insights , these are the points that matter most to stakeholders:

The first point is the huge scope for improvement, as evidenced by the Average Potential Efficiency consistently being 12 points higher than the Average Current Efficiency (67.26 and 79.12 respectively), with 85% of the properties having measurable improvement potential.

Secondly, the analysis identifies Heating Costs (£570.45) as the Primary Intervention Target by a very large margin, thus heating aspects causing the most heat loss and the heating system should be the main focus of the intervention.

Thirdly, the Property Type Disparity, where on the one hand, Houses represent the largest volume (53.42%) for mass intervention, on the other hand, Park Homes are the least efficient stock, so a dual strategy of high-volume and highly targeted retrofitting is necessary for them.

These discoveries are translated into four practical Actionable Recommendations of main significance:

Local Authorities are recommended to provide Targeted Funding for Fabric-First Measures such as Cavity Wall Insulation that would help in cutting down the primary cost driver.Utility Companies should establish 'Quick-Win' and 'Deep-Retrofit' Packages simultaneously, thus by implementing a high-volume, low-cost upgrade such as Low Energy Lighting as the initial engagement, they can promote the latter on a large scale and at a low cost.Landlords have to set up a system of Mandate and Enforcement of Minimum Efficiency Targets so that the properties can meet the requirements of future minimum EPC Band.Homeowners are advised to Insulate and Install Heating Controls first, taking advantage of the high average heating cost to convince them of the necessity of long-term investment in improvements like Solar Photovoltaic Panels or Fan Assisted Storage Heaters.

10. Limitations

Data Completeness and Accuracy: The EPC Register is a partial record of buildings, as it only has data for about 60% of buildings, and mostly those that have been built, sold, or rented since 2008. Therefore, the analysis is not a full account of all buildings in Manchester.

Missing Variables: There are some very important data points such as income or more detailed geographic information (beyond local authority boundaries) that are not provided, which hampers the effort to explore the linkage between energy performance and socio-economic factors.

Lagged Data: The dashboard is dependent on the dates when EPCs were lodged, which may not be indicative of the current operational status or actual usage of the property.

11.Conclusion

The data analysis of the Energy Performance Certificate (EPC) data of Manchester that was successfully carried out with the help of SQL Server to manage the data and Power BI to visualize it confirms that the potential of the enhancement of the urban energy efficiency is significant, but at the same time, it is real. The project showed the effective combination of the latest database and business intelligence tools to convert the public raw and massive data into valuable intelligence.

The main result is the possibility of finding a significant 12-point efficiency gap between the average current rating (67.26) and the average potential rating (79.12). This gap is an essential measure of the establishment of realistic and ambitious municipal energy targets because 85% of the 326K properties can be improved. The results made the Average Heating Cost (570.45) the financial and environmental burden that overwhelmingly forced its way on the residents much above the average lighting cost (73.52). This observation means that measures that are focused on thermal efficiency and heating systems should be of paramount importance in every successful net-zero strategy.

Moreover, the dashboard offered the required level of granularity to develop targeted policies, as opposed to blanket policies. The analysis, by stating that Houses represent the most numerically imposing category of certificates (53.42) and that Park Homes represent the most inefficient category, on average, would allow a two-fold approach of a massive retrofit program of Houses and an high-priority, deep-retrofit approach of the most efficient stock. The Cavity Wall Insulation and Low energy lighting recommendations are a visual representation of common recommendations, which gives a clear roadmap on operations in which funding and enforcement of regulatory activities can be prioritized.

Conclusively, the interactive Power BI dashboard is an effective and easy-to-use tool of data-driven governance. It goes beyond the description statistics to provide prescriptive information, which serves as a strong, measurable way based on the design of target programs by local authorities and energy stakeholders. By targeting the resources where the effect is largest, in the form of making more current properties improve their heat efficiency, Manchester can easily fast-track its development to its required energy efficiency and wider climate action goals.

12. References

https://epc.opendatacommunities.org/downloads/domestic#local-authority

https://www.theiet.org/media/5182/technical-report-writing.pdf