GENERAL SIR JOHN KOTELAWALA DEFENCE UNIVERSITY

FACULTY OF MANAGEMENT SOCIAL SCIENCES & HUMANITIES

DEPARTMENT OF LANGUAGES

BACHELOR OF SCIENCE IN APPLIED DATA SCIENCE AND COMMUNICATION

INTAKE 41

LB 2114: ADVANCED SQL AND CLOUD DATABASES

ASSIGNMENT

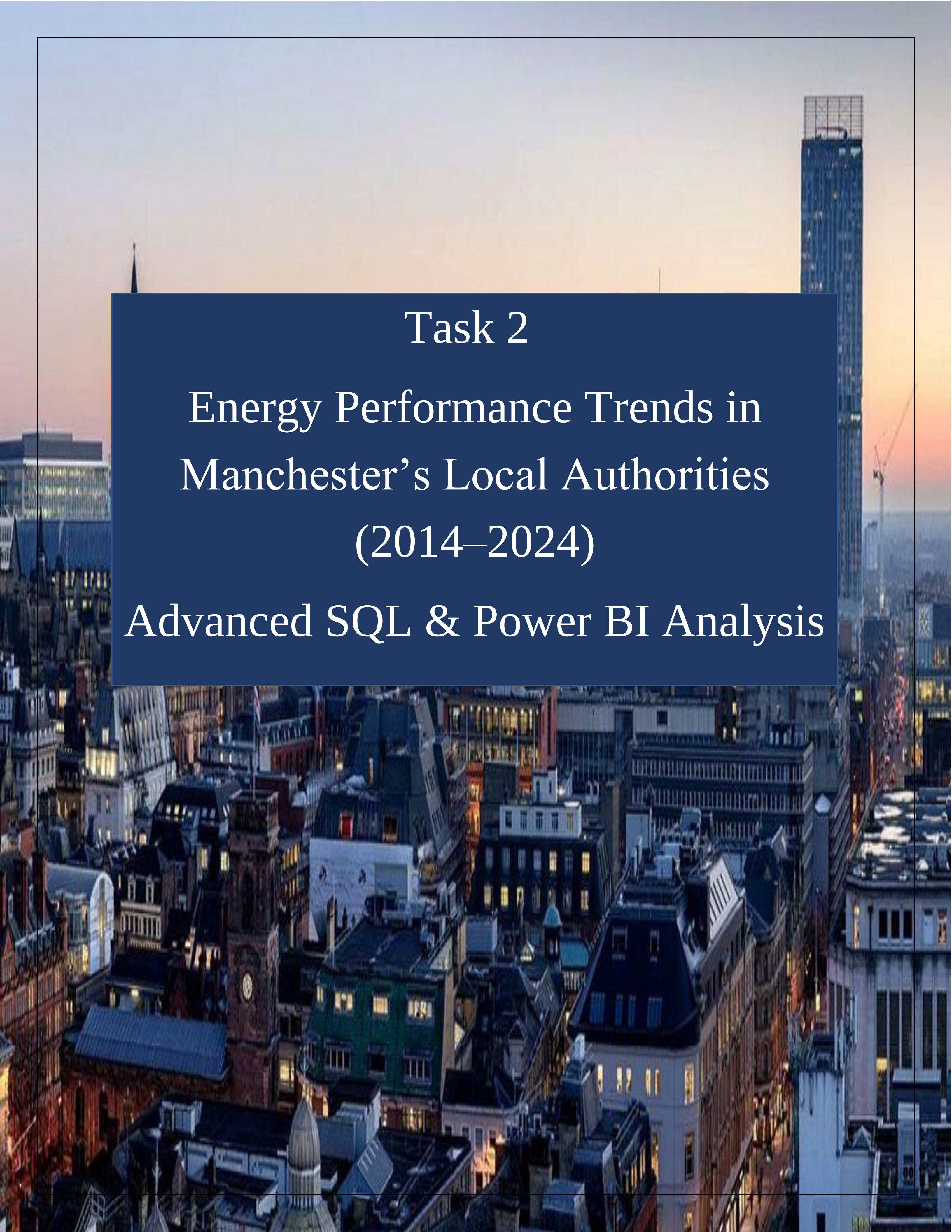
Group Members

D/ADC/24/0003 – Thesanya Lamahewa

D/ADC/24/0013 – Samoda De Silva

D/ADC/24/0028 – Thisari Perera

D/ADC/24/0048 – Dilakna Godagamage



Task 2

Energy Performance Trends in Manchester's Local Authorities (2014–2024)

Advanced SQL & Power BI Analysis

Content

1 Introduction 1

1.1 Background 1

1.2 Purpose of the report 1

1.3 Scope 1

1.4 Data Sources 2

2 SQL Data Processing 3

2.1 Importing Data to SQL server 3

2.2 Data cleaning with CTEs and Views 5

2.3 Challenges during importing the data 5

3 Exploratory Data Analysis 6

3.1 Create a view for cleaned data 6

3.2 Stored Procedure for Automated Cleaning 7

3.3 Use of system functions and Aggregates 8

3.4 Aggregate Summary by Property Type 11

4 Power BI Data Analysis & Visualization 12

4.1 Importing the Dataset to Power BI 12

4.2 DAX measures for Data Analyzing 13

4.3 Power BI Data Visualization 21

4.4 Explanation of Dashboard Visuals 22

4.5 The Connection of the SQL Server with the Power BI 27

4.6 Interpretation of the Dashboard 28

4.7 Application of Advanced Power BI Features 28

4.8 Key Insights and actionable recommendations 29

5 Conclusion 29

1 Introduction

1.1 Background

The energy performance certificates (EPC) are used to give details about the efficiency of energy of properties used in the UK, both residential and commercial. They display the current and potential energy rating of a property, A (most energy efficient) to G (least energy efficient). The use of the EPC data can be beneficial in researching the trends in the energy efficiency and in supporting the environmental policies as well as the housing standards. The local authorities of Manchester have experienced a great development during the last ten years and therefore an EPC analysis is useful in comprehending how property performance and sustainability objectives have changed.

1.2 Purpose of the report

This report is aimed at analyzing and visualizing EPC data of local authorities in Manchester, with the help of SQL advanced methods and using Power BI dashboards.
This report aims to:

- Prepare EPC datasets with SQL
- Determine energy performance trends 2014-2024
- Compare price of efficiency rates amongst local authorities

1.3 Scope

The current analysis is devoted to the EPC records that are to be issued within the Manchester Local Authority and in the period between 2014 and 2024.

The scope includes:

- Domestic Property only – Important variables included EPC rating, the property type, the age at which it was constructed, the postcode and the energy performance indicators.

1.4 Data Sources

The report uses publicly available EPC datasets provide through:

- UK Government Open Data (EPC Register) https://epc.opendatacommunities.org/downloads/domestic#local-authority

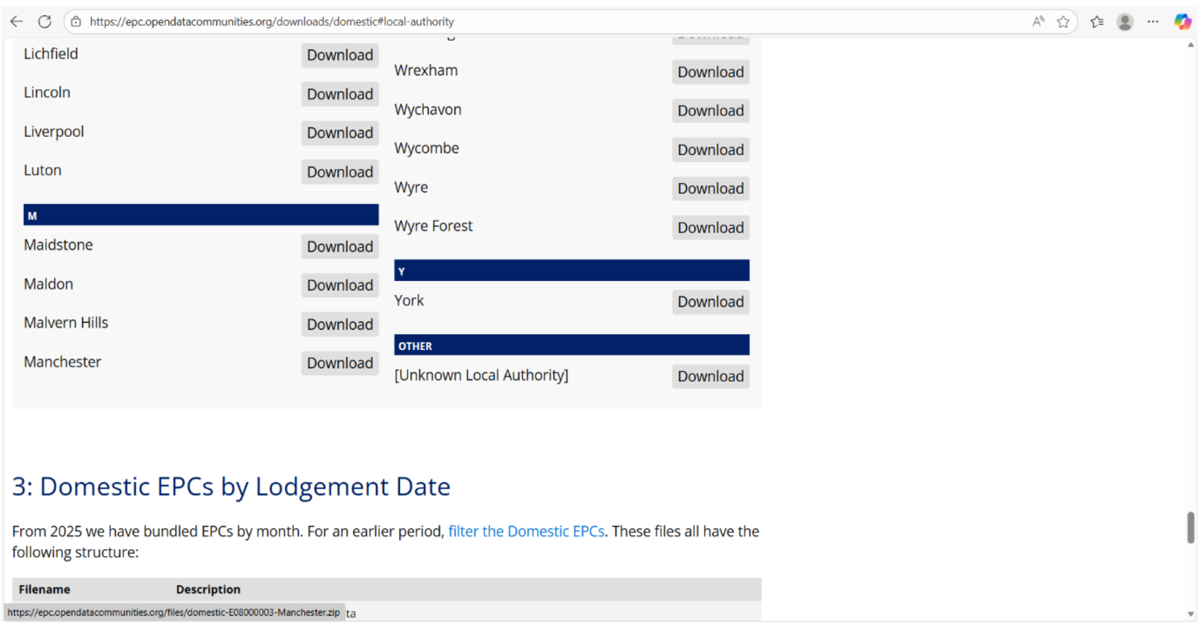


Figure 1: Dataset

2 SQL Data Processing

2.1 Importing Data to SQL server

All three EPC CSV datasets were successfully imported into SQL Server, forming the basis for subsequent data cleaning and transformation.

A new database was created in SQL Server to store the EPC datasets. Then the three CSV files were imported into the database using the SQL Server Import Wizard. During this process, the column mappings were checked to ensure that all fields matched correctly, and the data was successfully loaded in to staging tables. This completed the initial setup required for further cleaning, transformation, and analysis.

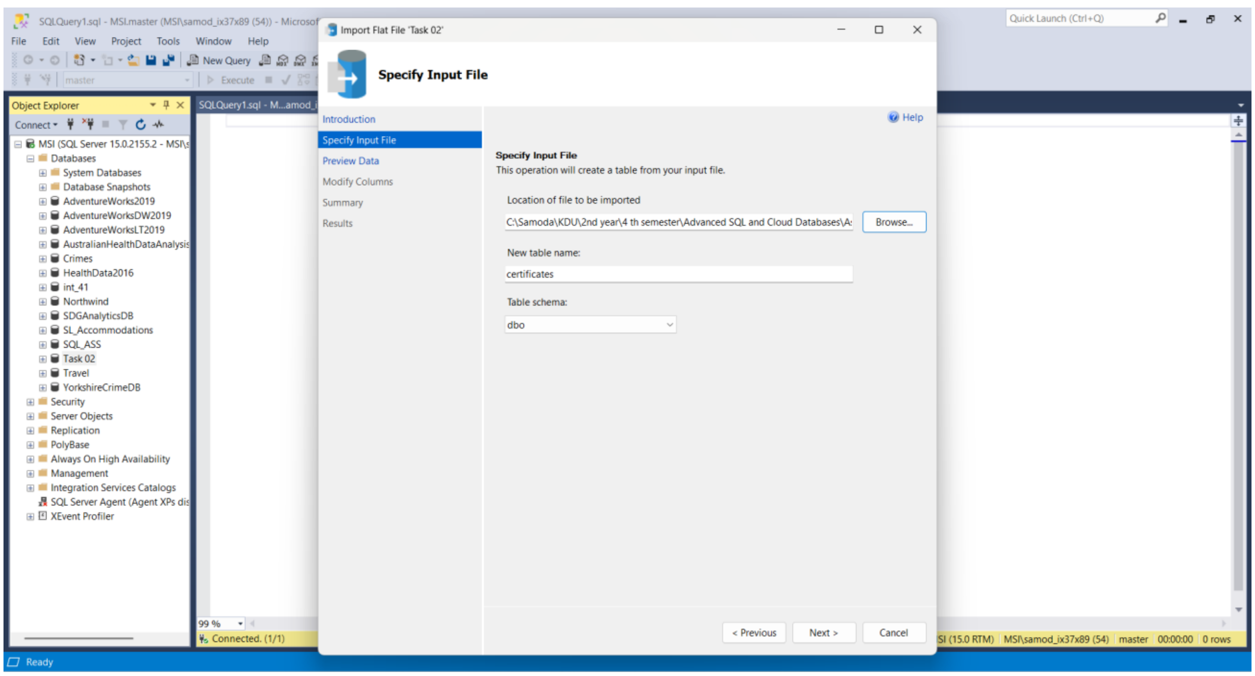


Figure 2: Specify Input File

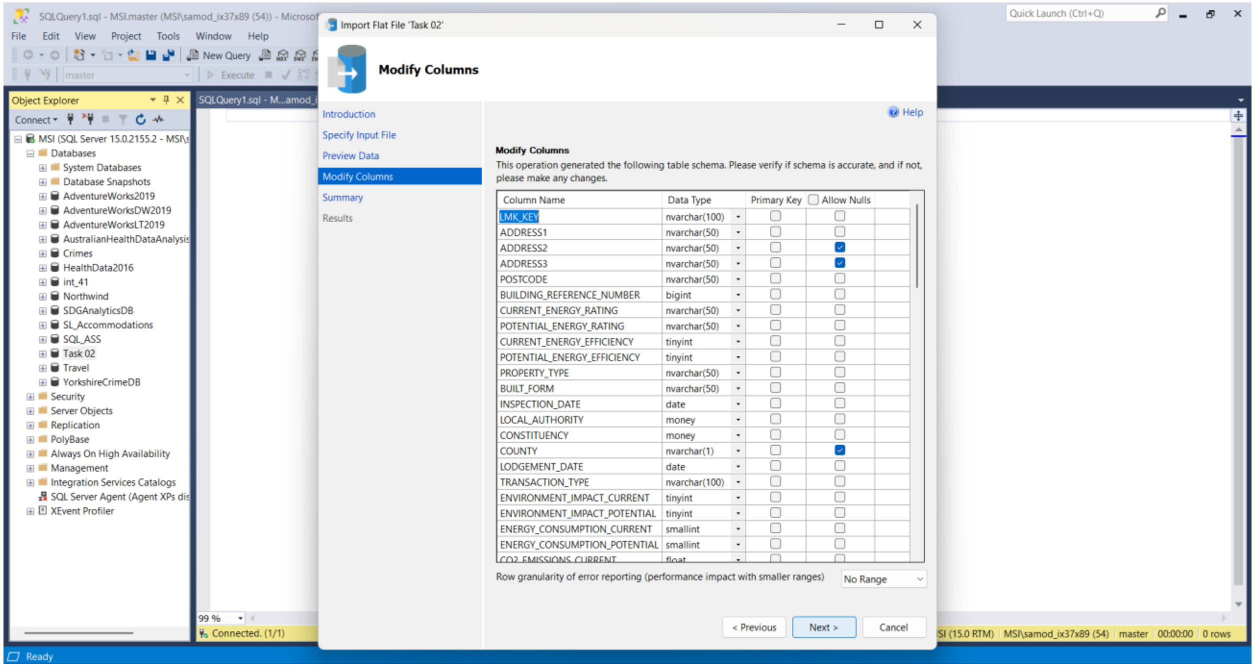


Figure 3: Modify Columns

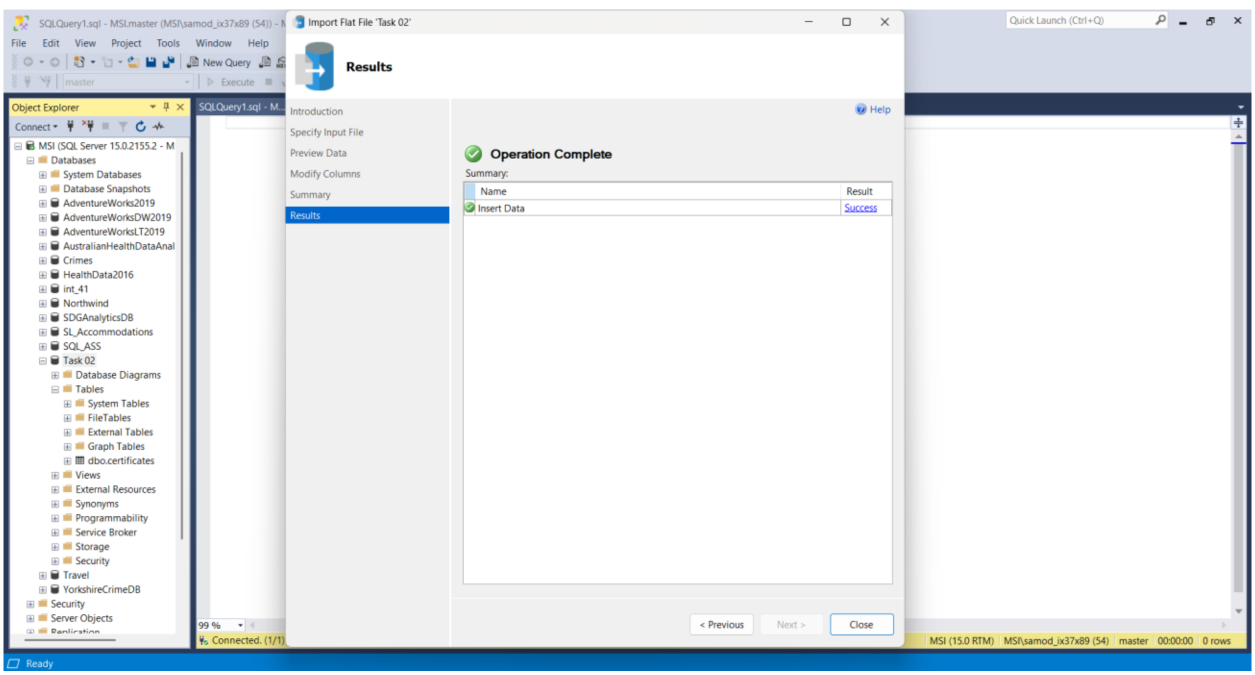


Figure 4: Operation complete

2.2 Data cleaning with CTEs and Views

The process of data cleaning was conducted with the help of Common Table Expressions (CTEs) and SQL Views that facilitated the process of arranging and making the transformation. Missing values, invalid records and the standardization of key fields were dealt with using CTEs and Views were generated to hold the cleaned and formatted datasets to facilitate easier querying and analysis in the Power BI.

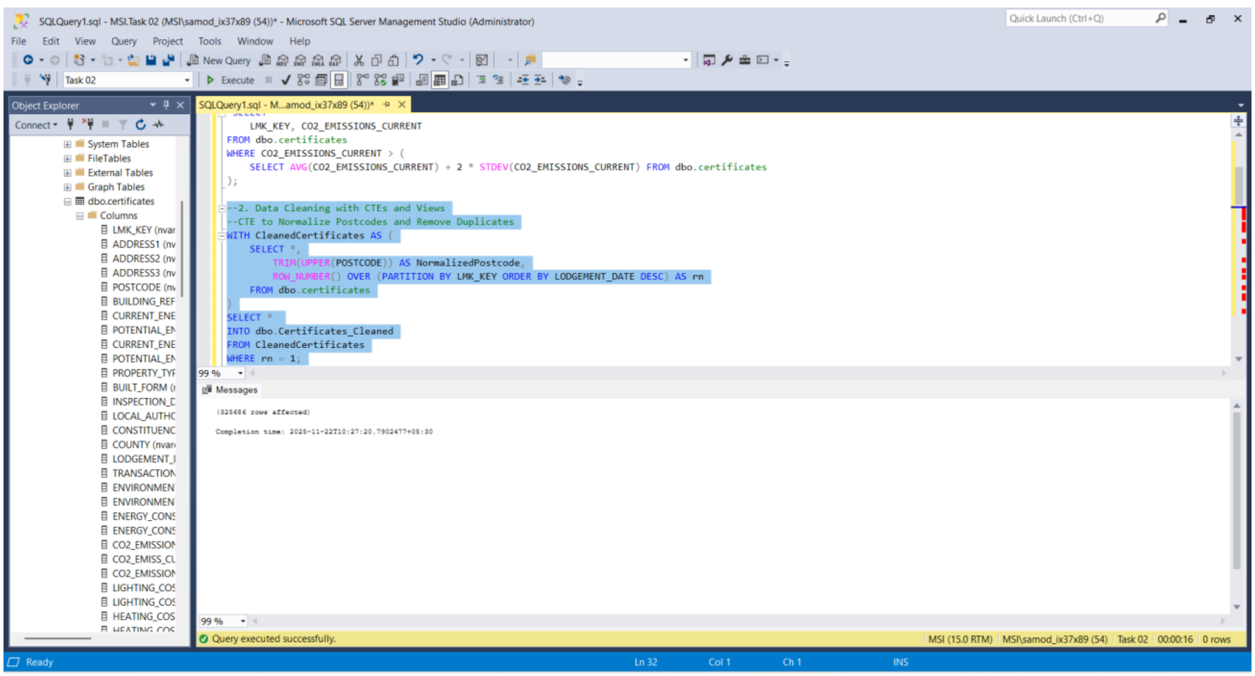


Figure5: Data Cleaning with CTE View

2.3 Challenges during importing the data

When handling EPC Manchester data, a number of challenges were encountered during importation and cleaning. There were irregular formatting, missing values and non- numeric entries in numeric key fields for an example, energy score and floor area in the raw files. These problems had to be carefully validated, invalid rows were to be eliminated and data types were to be standardized to make the data accurate.

3 Exploratory Data Analysis

3.1 Create a view for cleaned data

Check for missing values

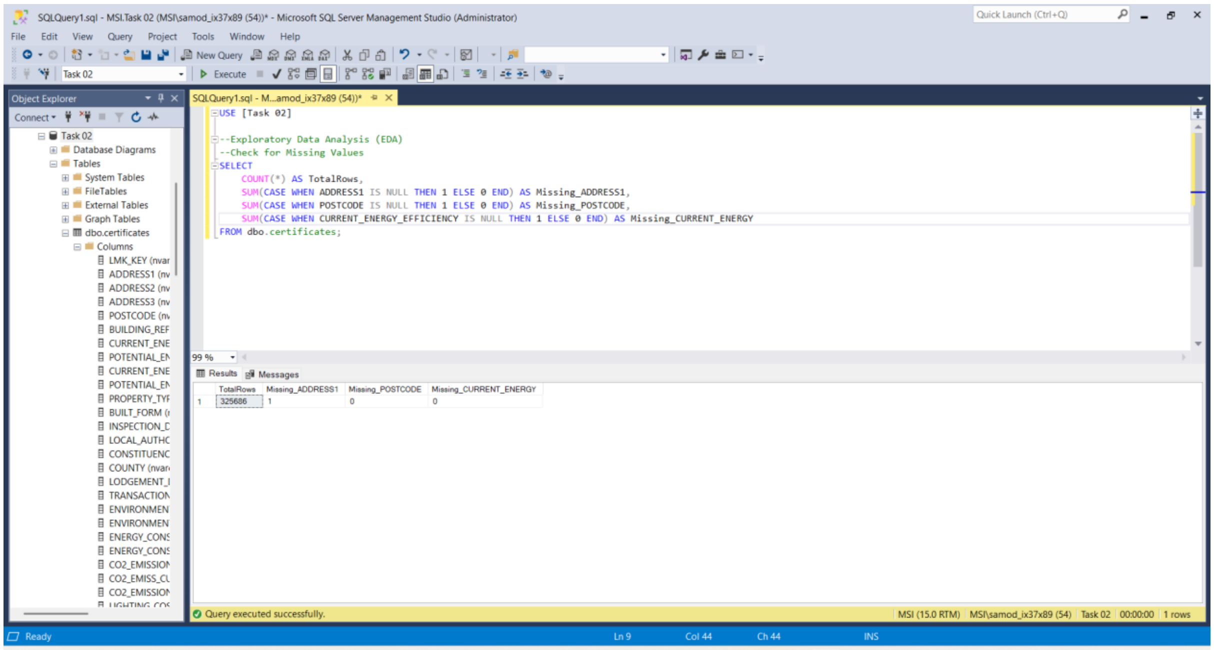
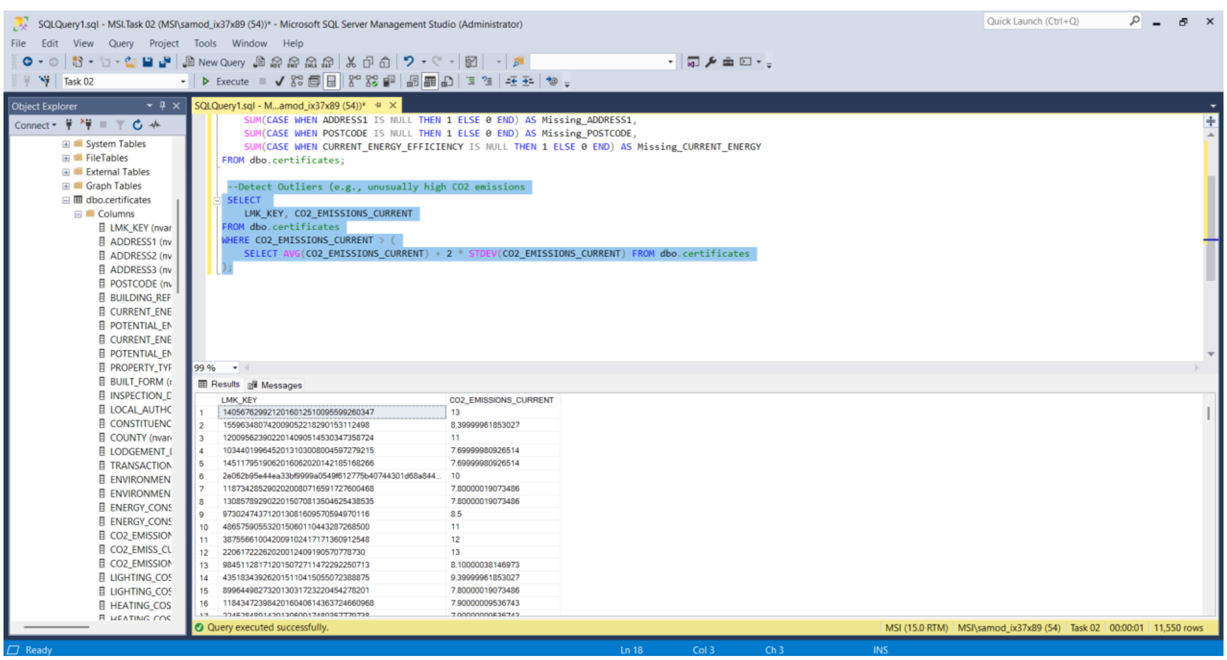


Figure 6: Checking Missing Values

Detect Outliers



Creating a View for Cleaned Data

The Cleaned EPC Manchester data was visualized to give a simplified and reliable version of the data. It enables end users to view the corrected and standardized data without making any changes in the original tables. This simplifies analysis, provides uniformity in the reports, and facilitates effective data accessibility to proceed with additional Power BI visualization.

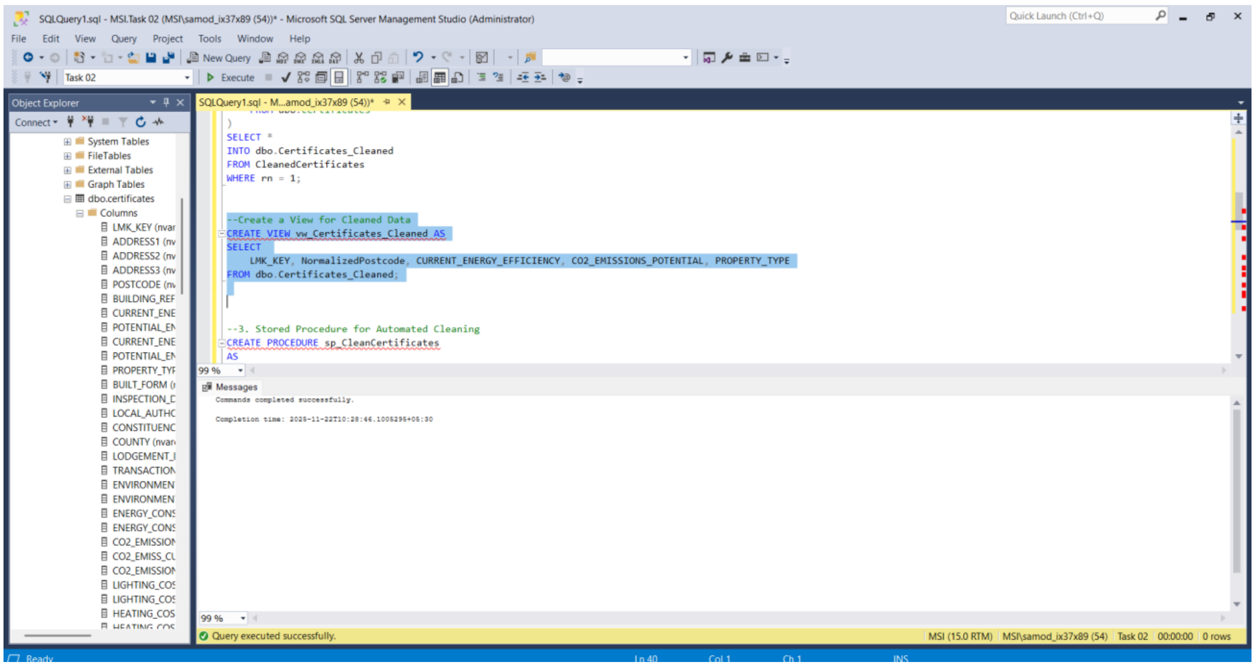


Figure 7: Create a view for cleaned data

3.2 Stored Procedure for Automated Cleaning

A Stored procedure was written to automatize the process of data cleaning, which is consistent and efficient when importing new EPC data. The process normalizes format, eliminates invalid or duplicate entries, manages missing data, and automatically updates the refined data, which saves manual effort and enhances the reliability of the data.

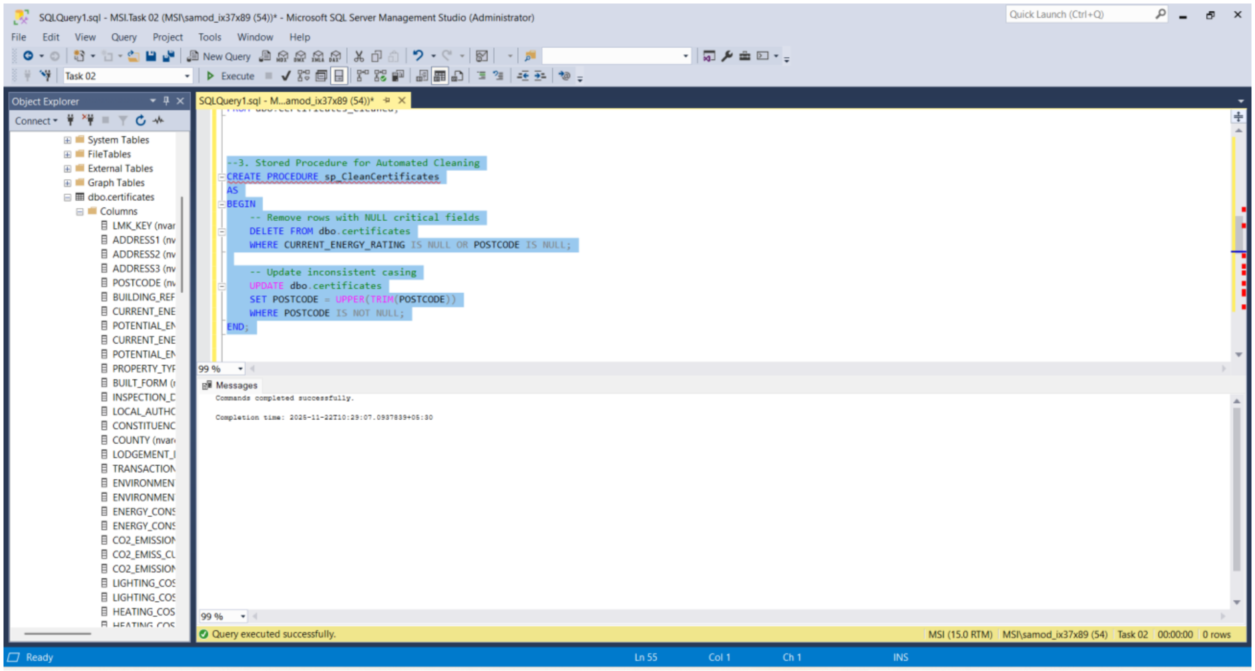


Figure 8: Stored Procedure for automated cleaning

3.3 Use of system functions and Aggregates

Major elements of the EPC data were summarized and analyzed using system functions and aggregate functions. System functions assisted in formatting dates, Standardization of therapeutic texts and extraction of particular data values and aggregate functions were used to compute rating distributions, discovering trends over times and comparing energy performance across the local authorities of Manchester as an example.

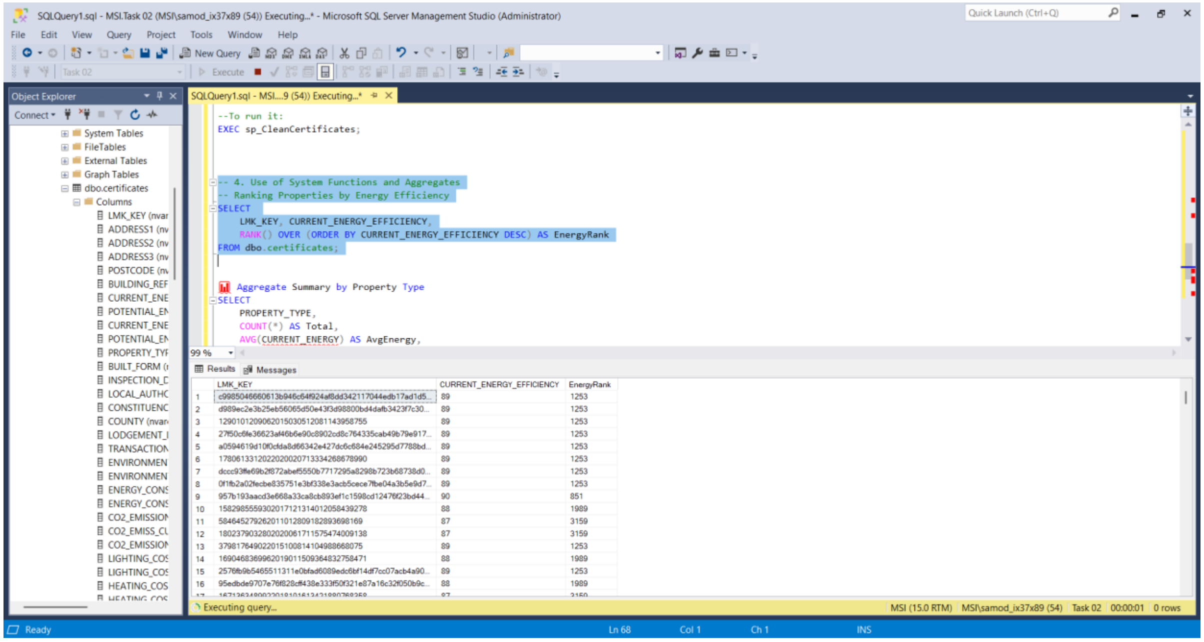


Figure 9: Use of system functions and aggregations

Using ISNUMERIC () to identify Non-numeric entries

The ISNUMERIC () is applied to assess whether a value could be treated as a number. This aids in the cleaning of data during the cleaning process since it allows one to detect rows that contain text or symbols or wrongly typed in values in fields which are only supposed to hold numeric values. Identifying nonnumeric values, it is possible to filter out or amend them in order to guarantee the accuracy of the data and preserve the quality of the cleaned EPC Manchester dataset.

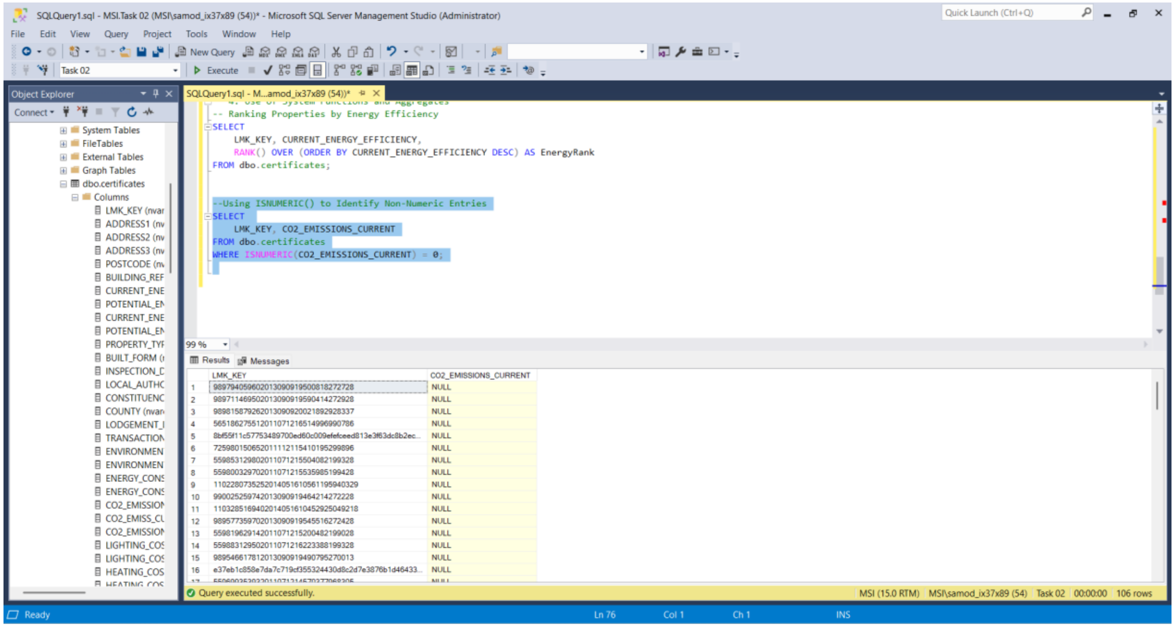


Figure 10: Using Isnumeric ()

Using TRY_CAST () to safely convert and filter valid numbers

The TRY_CAST () function was adopted to convert the values safely into the numeric types eliminating invalid entries and only admitting the valid numbers into the analysis.

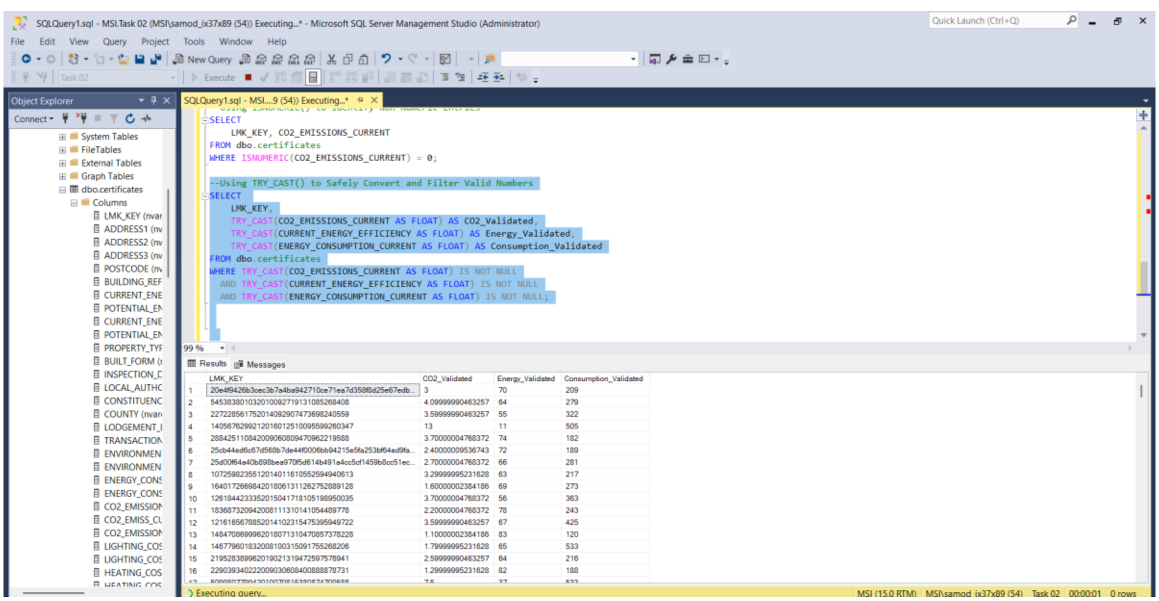


Figure 11: Using TRY_CAST ()

Create a Cleaned View with Validated Numeric Fields

An create view was made to records that had valid numeric values were incorporated in the final dataset. The invalid or improperly formatted numbers were filtered out using the check functions like ISNUMERIC () or TRY_CAST. The view offers a reliable analysis ready presentation of the EPC Manchester data and assist in preserving the accuracy of the data and ensure all calculations, visualizations and reporting use confirmed numeric fields.

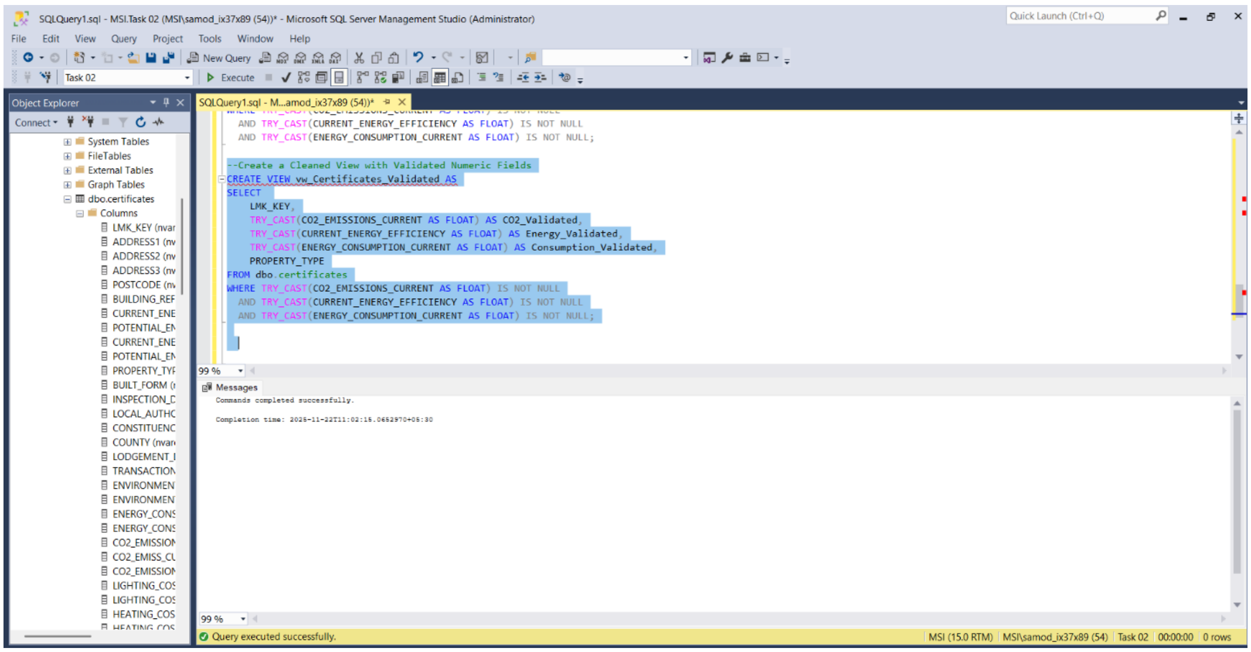


Figure 12: Create a Cleaned View

3.4 Aggregate Summary by Property Type

The number of averages of totals of or other indicators of interest summarized by each property type. This allows comparing and analyzing trends, performance level, or distribution within various categories fast to make informed decisions and based on facts.

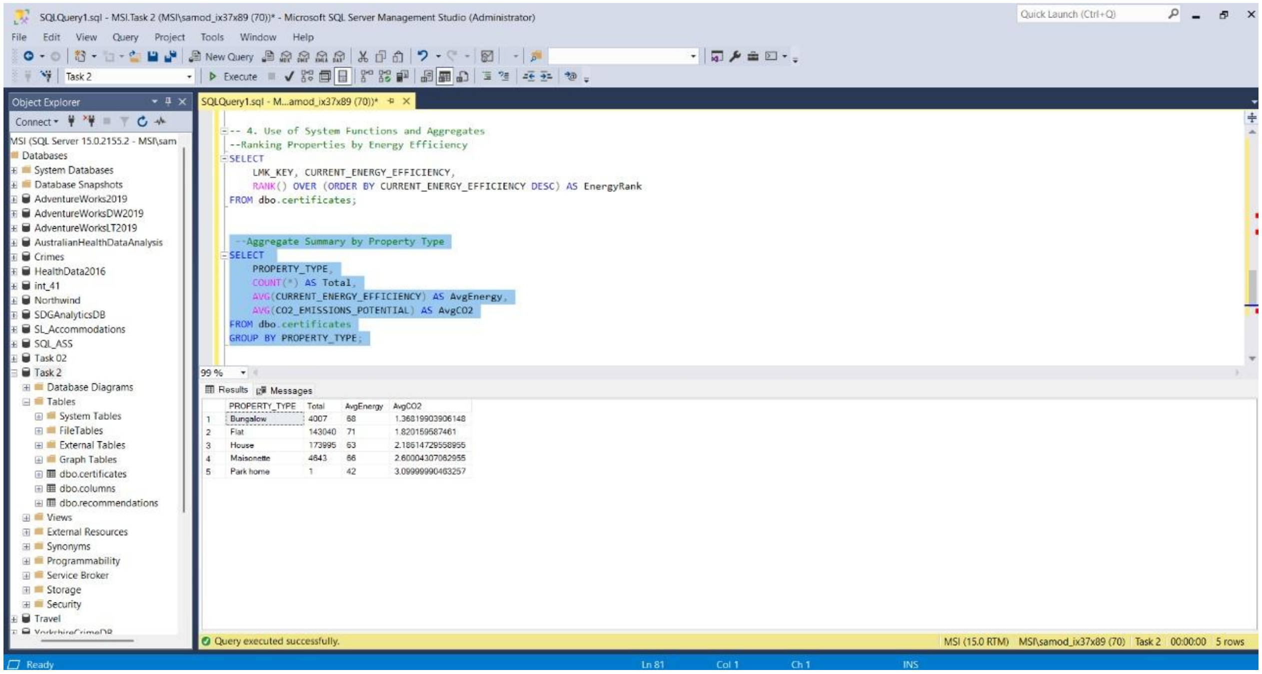


Figure 12: Aggregate Summary

4 Power BI Data Analysis & Visualization

4.1 Importing the Dataset to Power BI

Once the data was cleaned and ready in SQL Server, the processed data was added to power BI with the inbuilt SQL Server connector. Power BI loaded the cleaned tables by connecting directly to the database and was able to model and visualize the data.

The Certificates, Columns, and Recommendations CSV files were successfully imported into Power BI using a structured data-loading method, ensuring that all datasets were properly transferred and fully accessible for additional cleaning, modelling and analysis.

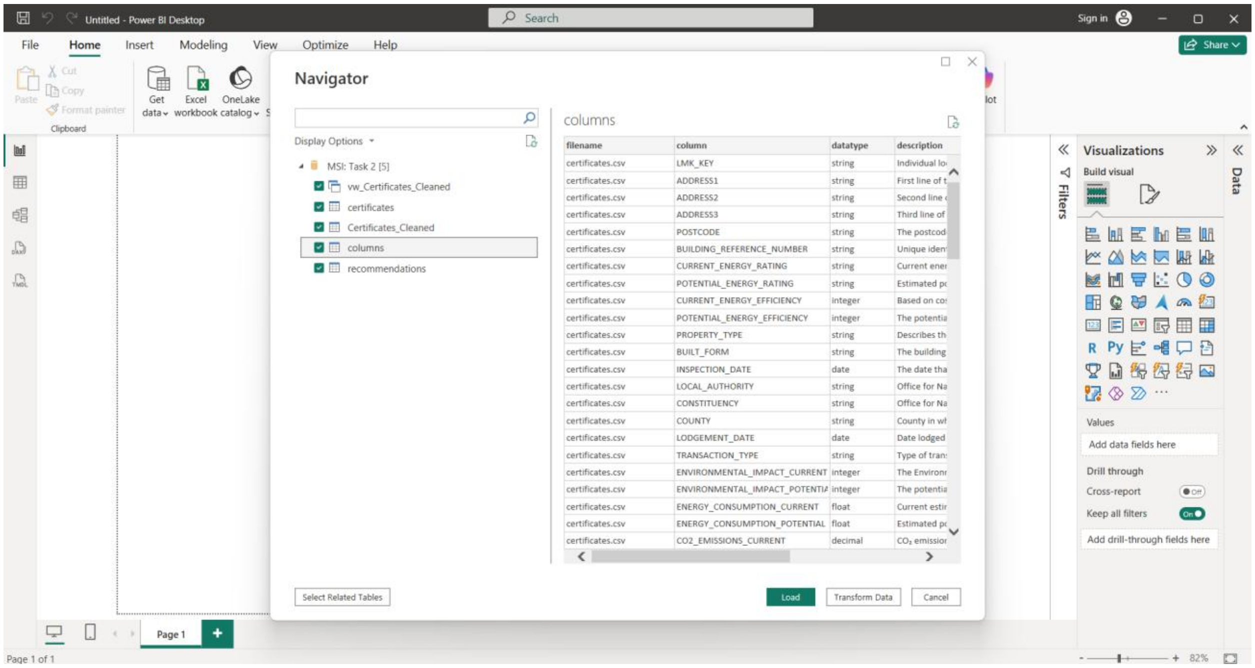


Figure 13: Importing Data to Power BI

4.2 DAX measures for Data Analyzing

The measures were developed in Power BI to have dynamic calculations like totals, averages and performance comparisons with the help of DAX, the dashboard is capable of generating correct real time insights in cases of user selection and filtering. These actions assist in the conversion of raw data to meaningful metrics which enhances the depth of analysis and clarity of the visualizations in general.

Power Query Editor

Power Query Editor cleans, transforms, and prepares raw data and then loads it in the model. It enables to eliminate the errors, divide or combine columns, alter data types, filter rows and carry out the transformation without coding. All the actions are automatically registered, which makes the cleaning process visible and repeatable. Reliance on the Power Query Editor makes the EPC Manchester data is arranged well, correct, and can be analyzed in the dashboard.

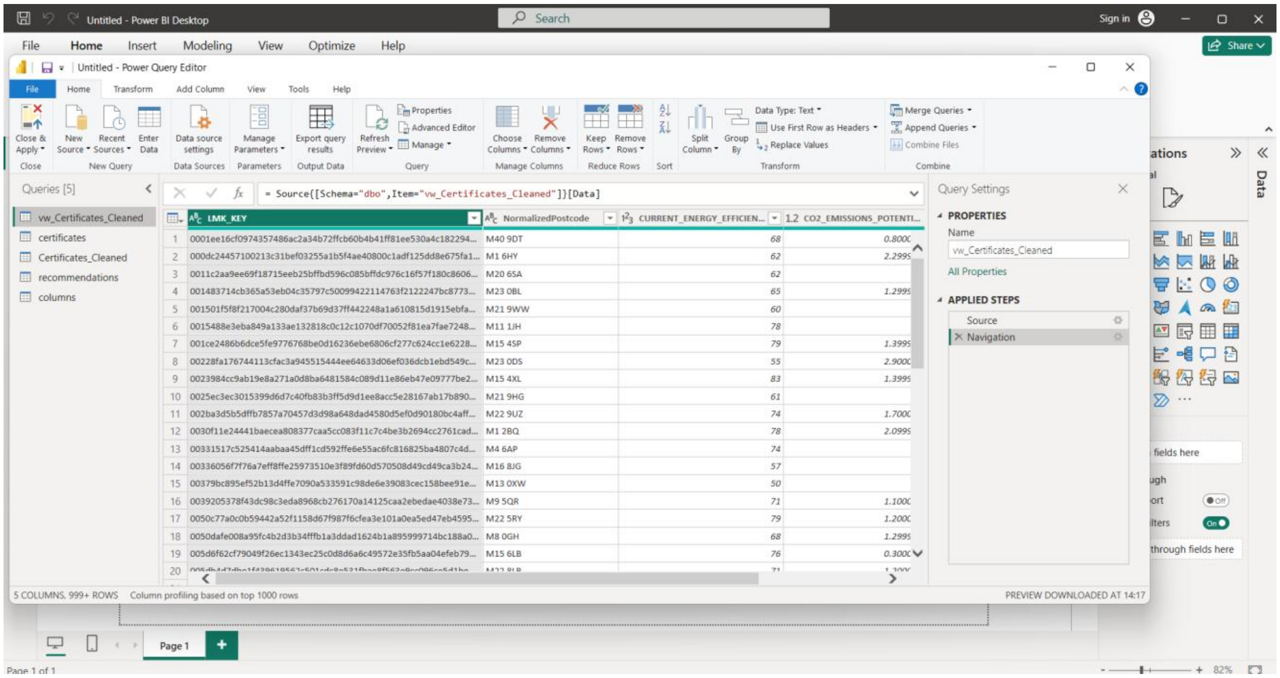


Figure 14: Power Query Editor

AVERAGE DAX FUNCTION

The mean value of chosen fields was computed with the AVERAGE DAX measure that gave a clear picture of the average level of performance.

This measure is used to compute the average of a numeric column. It helps in interpretation of key trends in the EPC Manchester data, including the mean energy rating, mean Co2 emissions or mean potential improvement score. This measure will assist in the identification of overall levels of performance and the comparison of energy efficiency in various groups of properties as a result of summarizing large volumes of data in to one representative value.

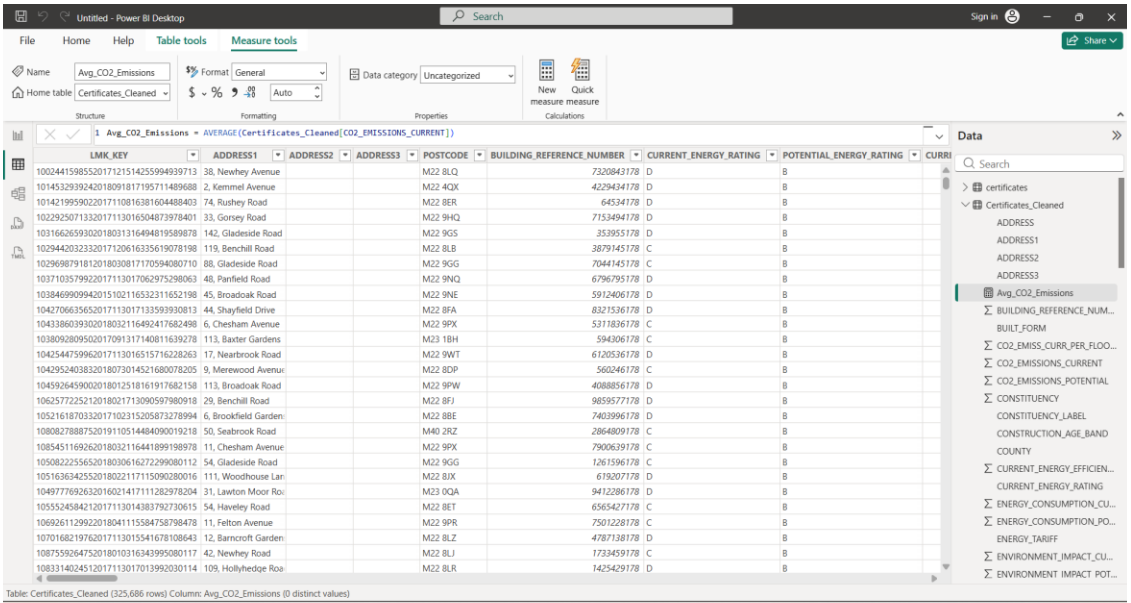


Figure 15: AVERAGE DAX FUNCTION

SWITCH DAX FUNCTION

The assigning labels or categories according to certain conditions was performed by the SWITCH DAX function. Simplifies complex nested IF statements and simplifies the logic to manage which, through the simplification of these statements, aids in creating more accurate and clearer classifications in the dashboard.

This measure classifies or tags information according to certain requirements. In the EPC Manchester dataset, as an example a switch-based measurement can simply give each property an energy efficiency band based on its numeric EPC score. The measure compares every row to specific ranges and the relevant category is provided and enables the dashboard to group the properties, compare performance and view the results in a meaningful manner.

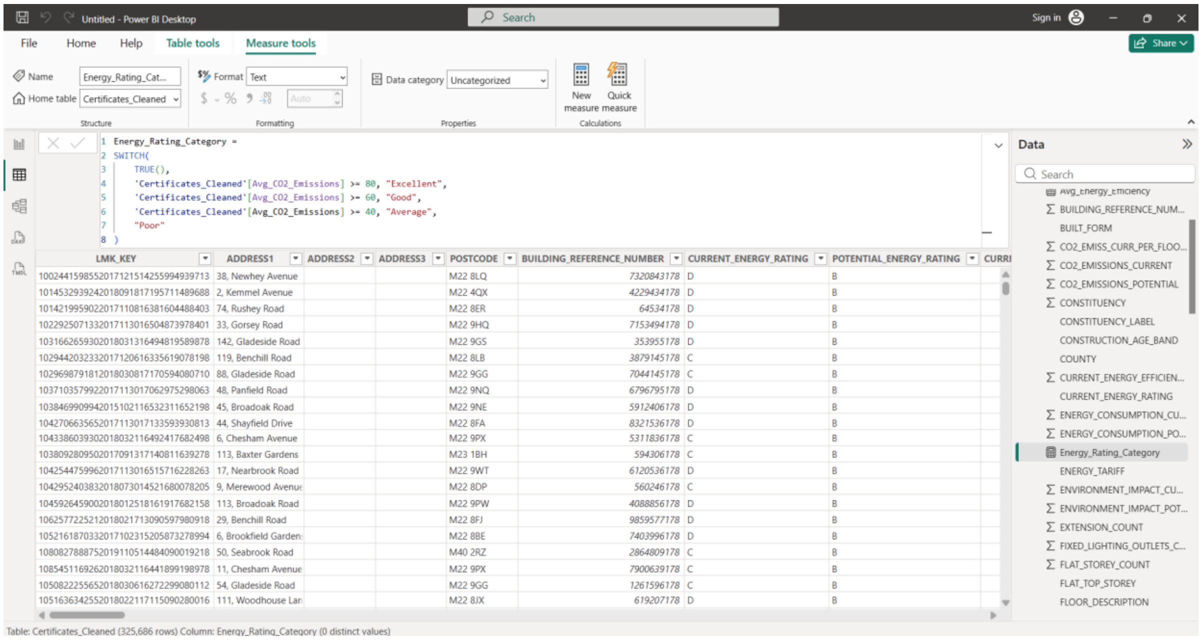


Figure 16: SWITCH DAX FUNCTION

DIVIDE DAX FUNCTION

To safely calculate ratios and percentages while avoiding division by zero errors, the DIVIDE DAX function was used. When compared to the conventional division operator, it yields cleaner, more dependable results, increasing the precision of dashboard metrics.

This measure can divide two numbers and in case the denominator maybe zero. Ratios that can be calculated using this based on DIVIDE in the EPC Manchester dataset include average Co2 emissions per property or energy improvement per unit. This makes there are no errors when dividing by zero and results are accurate thus reliable comparisons and calculations in the dashboard.

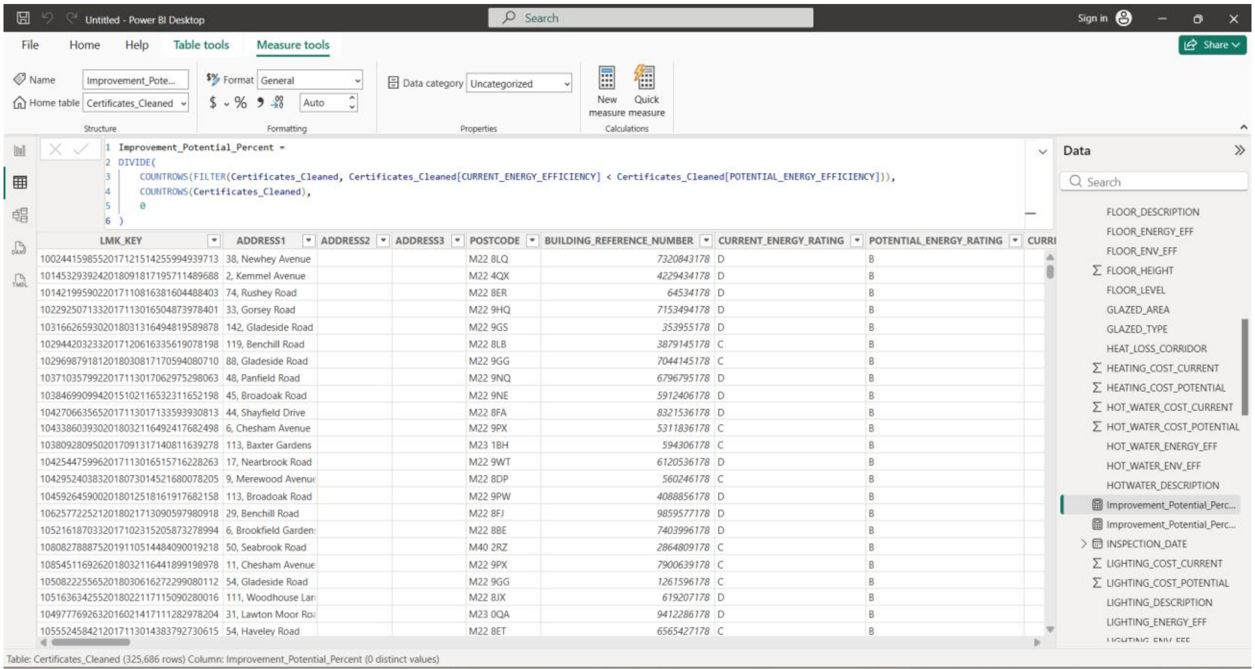


Figure 17: DIVIDE DAX FUNCTION

This picture shows how a calculated column called Co2_Reduction_Potential was created in the Certificates_Cleaned table in Power BI Desktop.

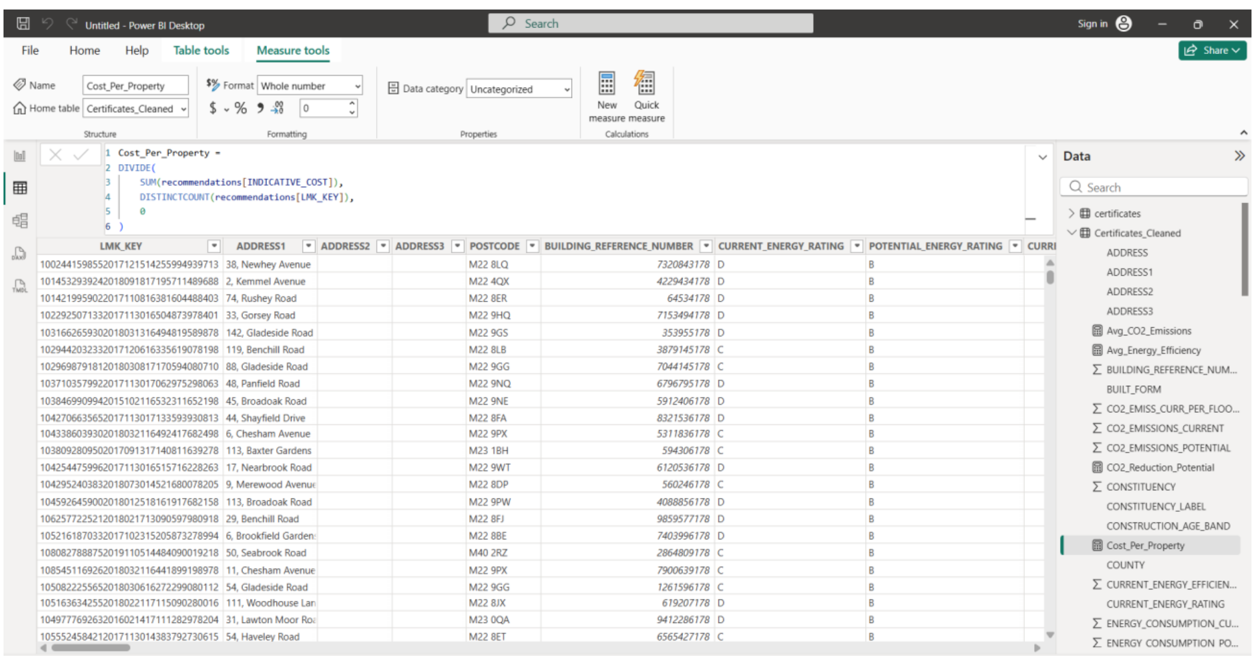


Figure 18: Co2_Reduction_Potential

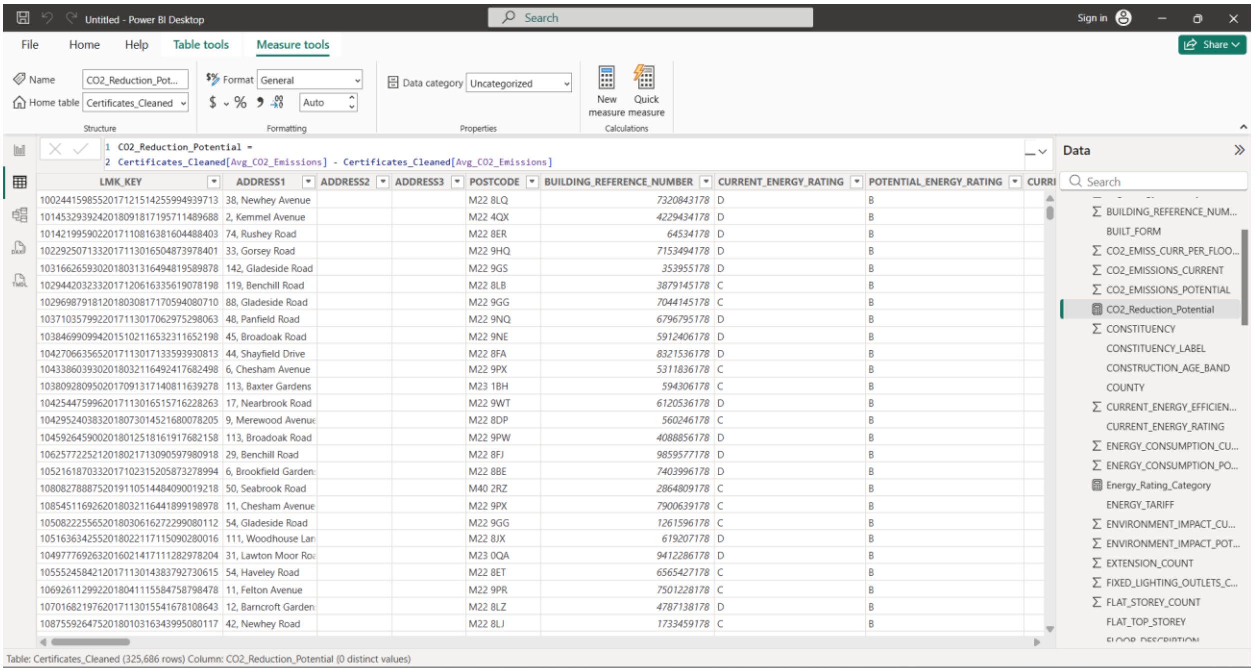


Figure 19: Cleaned Co2_Reduction_Potential

COUNT_ROW DAX FUNCTION

The number of records in a table or filtered dataset was determined using the COUNT_ROW DAX function, this metric aids in measuring the amount of data and supports dashboard insights.

This measure titled Solar_Heating_Count counts the amount of property certificates with the Solar Water Heating Flag set to “Yes”. This measure helps in measuring the adoption of solar water-heating systems and allows comparison across regions, time periods, or other dimensions in the EPC Manchester analysis by limiting the cleaned dataset for solar equipped properties and counting each valid record.

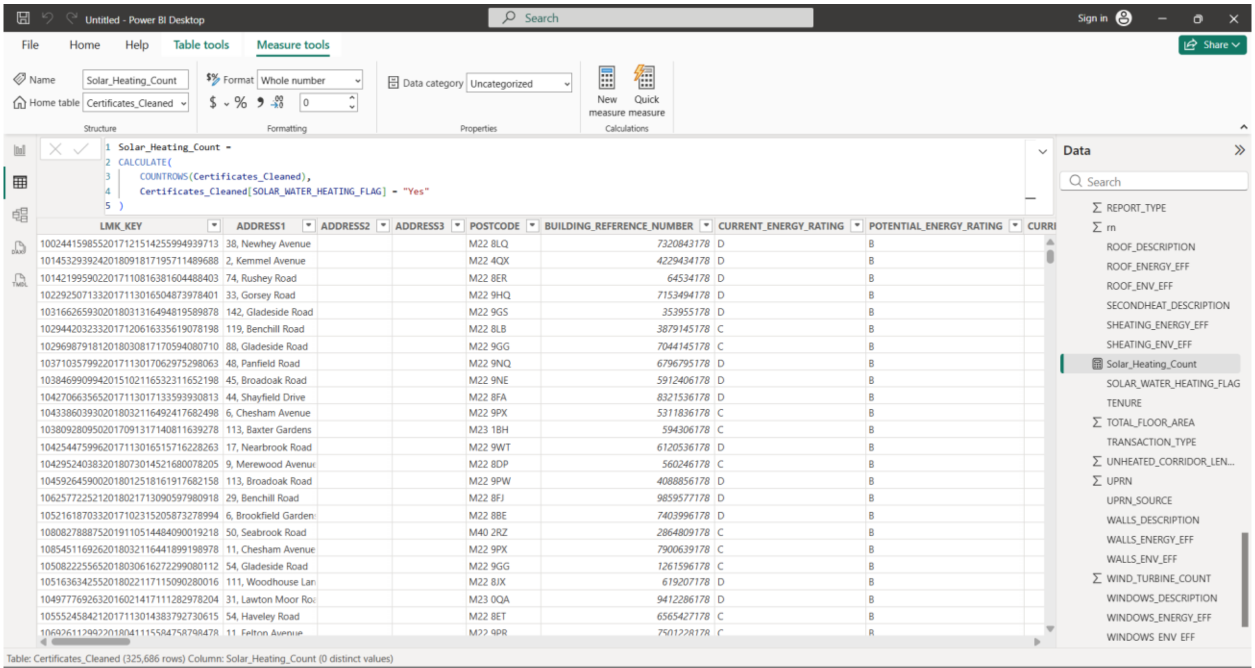


Figure 19: COUNT_ROW DAX FUNCTION

DISTINCTCOUNT DAX FUNCTION

The number of distinct values in a column was counted using the DISTINCTCOUNT DAX function. The DISTINCTCOUNT function, which reduces duplicates before finding the total number of unique values in the selected column, is applied in the EPC Manchester dataset to measure this.

It makes it easier to identify how many distinct properties, certificates, or categories actually exist and presents a clear and more accurate picture of the data without repetition.

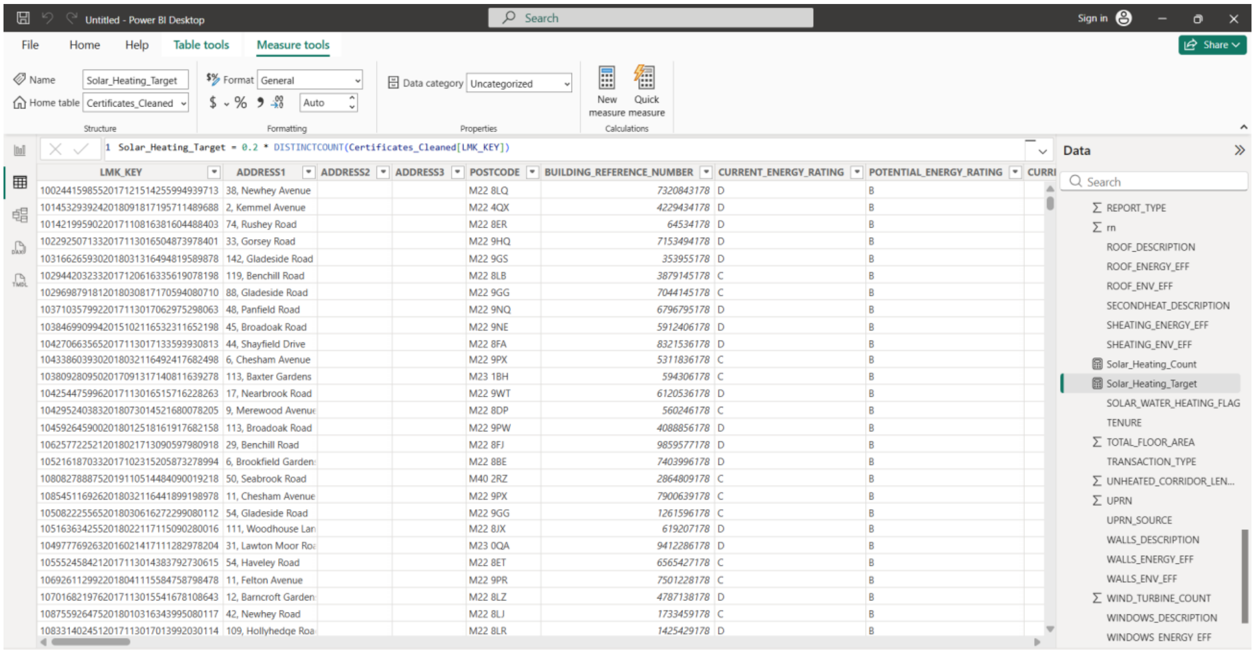


Figure 20: DISTINCTCOUNT DAX FUNCTION

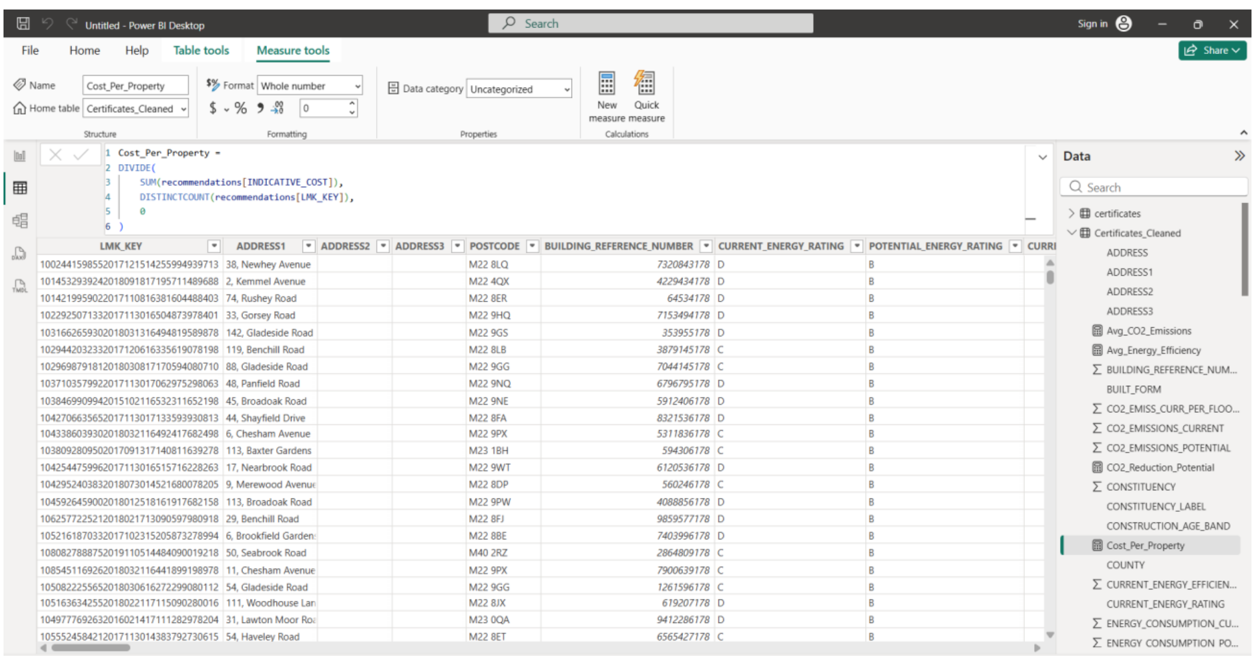


Figure 21: DISTINCTCOUNT DAX FUNCTION 2

4.3 Power BI Data Visualization

Key insights from the cleaned EPC Manchester dataset were visually presented in the Power BI dashboard through the use of interactive charts, cards, and slicers. By emphasizing significant trends, comparisons, and patterns among Manchester properties, each visual aids user in rapidly comprehending energy performance data.

Clearer and more informed decision-making is supported by the dashboard's interactive features, which include drilldowns and filters that make it simple to delve deeper into the data and explore particular areas.

FINAL DASHBOARD

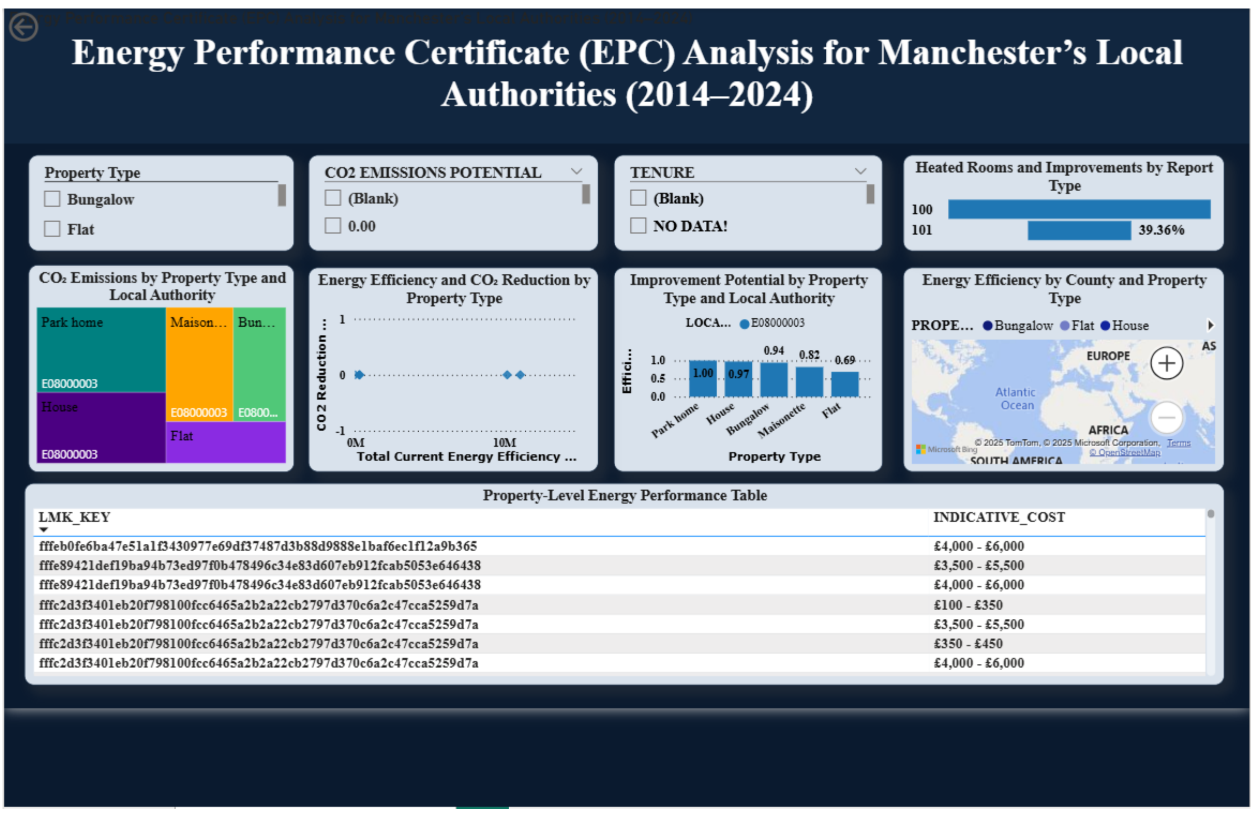


Figure 22: Dashboard

4.4 Explanation of Dashboard Visuals

Filled Map

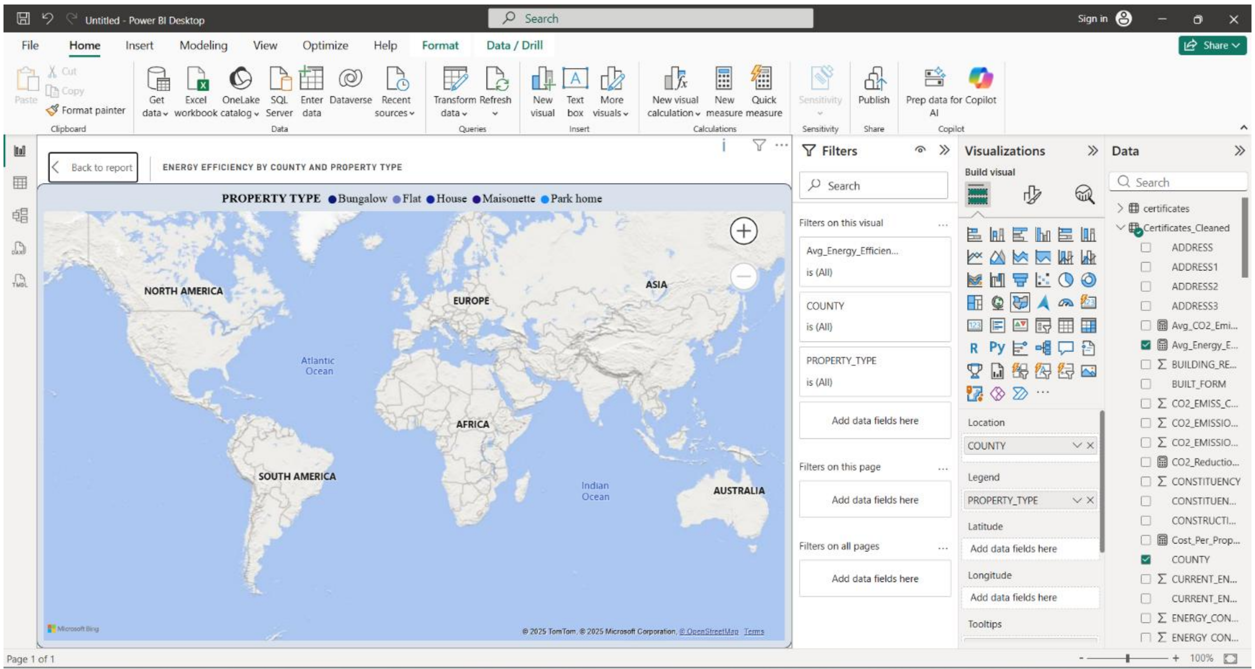


Figure 23: Filled Map

A data foundation is created using the certificates_cleaned table, which contains property level energy ratings. A crucial dax calculation that calculates the maximum savings per property is the Co2 Reduction Potential. By aggregating this metric and displaying it on a Filled Map that is divided by COUNTY and PROPERTY_TYPE, the analysis then directly identifies the specific geographic regions and housing types with the highest potential for reduction efforts.

Funnel

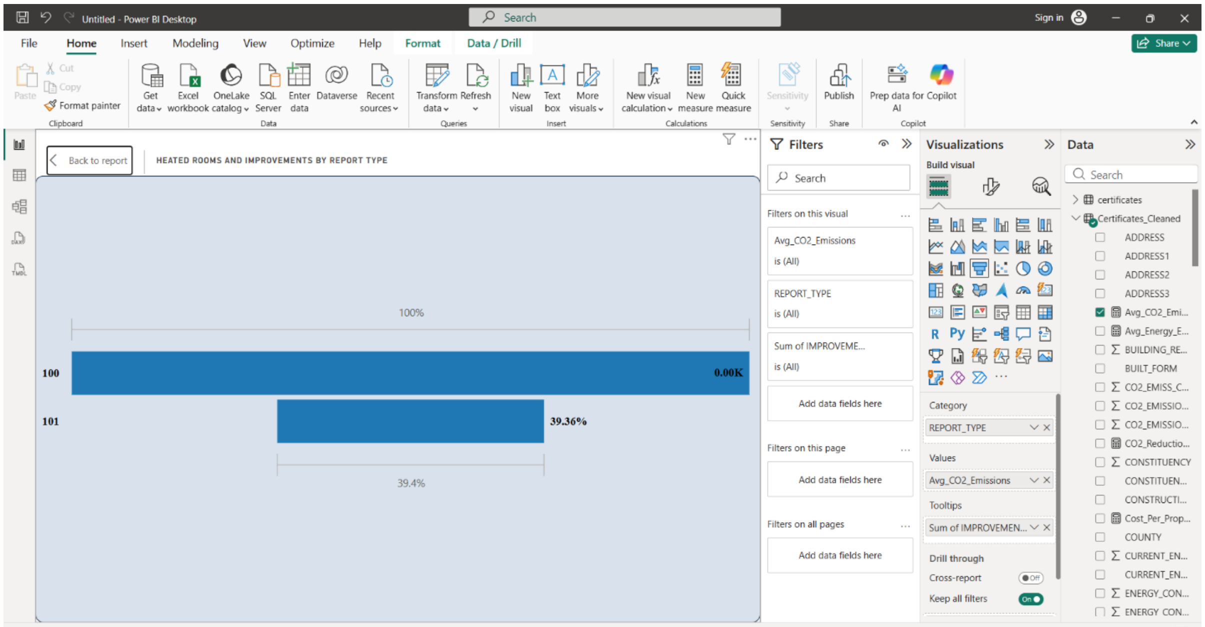
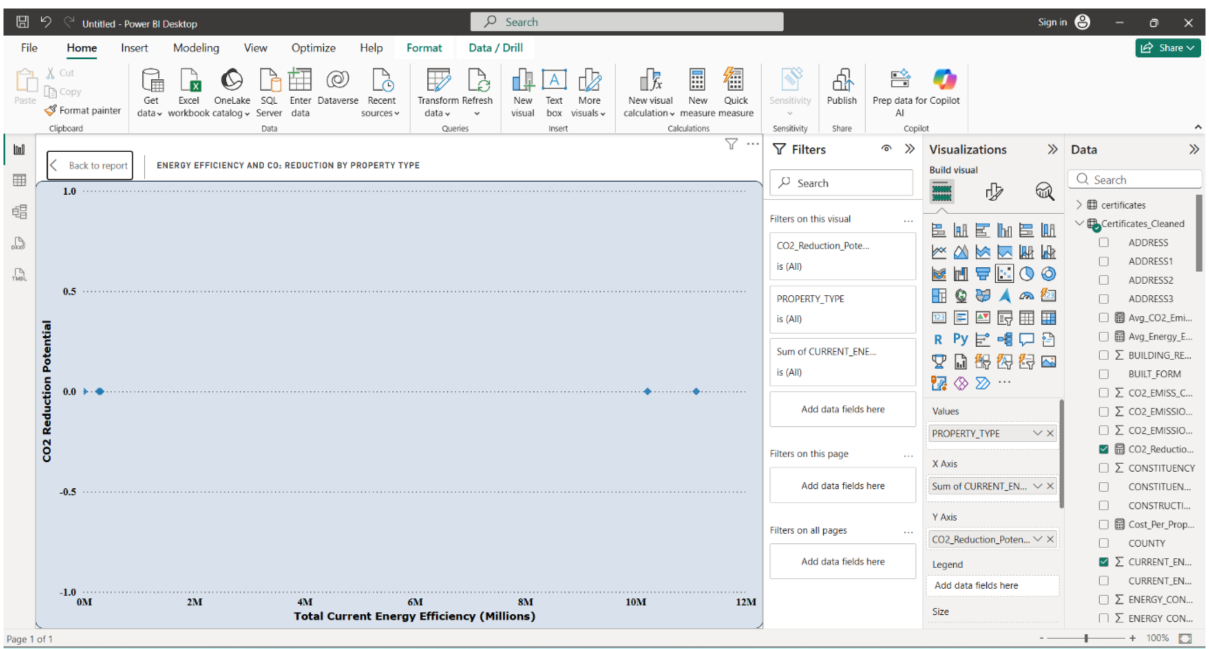


Figure 24: Funnel

This graph is used as a visual aid for qualification or conversion. Using the raw data to the report it reveals that various report categories (defined by REPORT_TYPE) add to the overall improvements. When the criteria move from the general baseline (stage 100) to the specific group (stage 101), the graphic emphasizes a significant limitation or feature.

Scatter Chart



The overall conclusion is that, despite significant differences in the total energy efficiency reported across property types, the immediate remaining reduction opportunity for these aggregated groups appears marginal.

Slicers

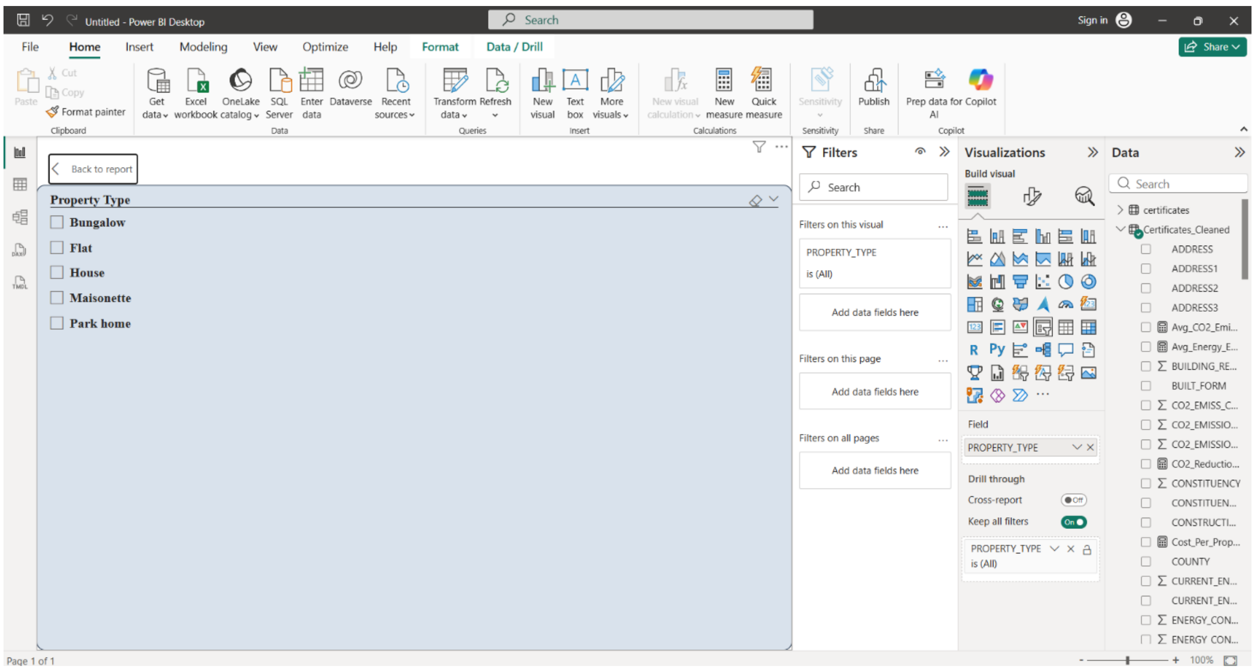


Figure 26: Property Slicer

The guidance of this slicer, users can narrow down the dashboard according to different types of properties found in the EPC Manchester dataset, such as Park Home, Bungalow, Flat, House, and Maisonette. By choosing one or more categories, users may concentrate the analysis on particular asset types. This makes it possible to get specific insights into energy performance and other important statistics connected to each kind of property.

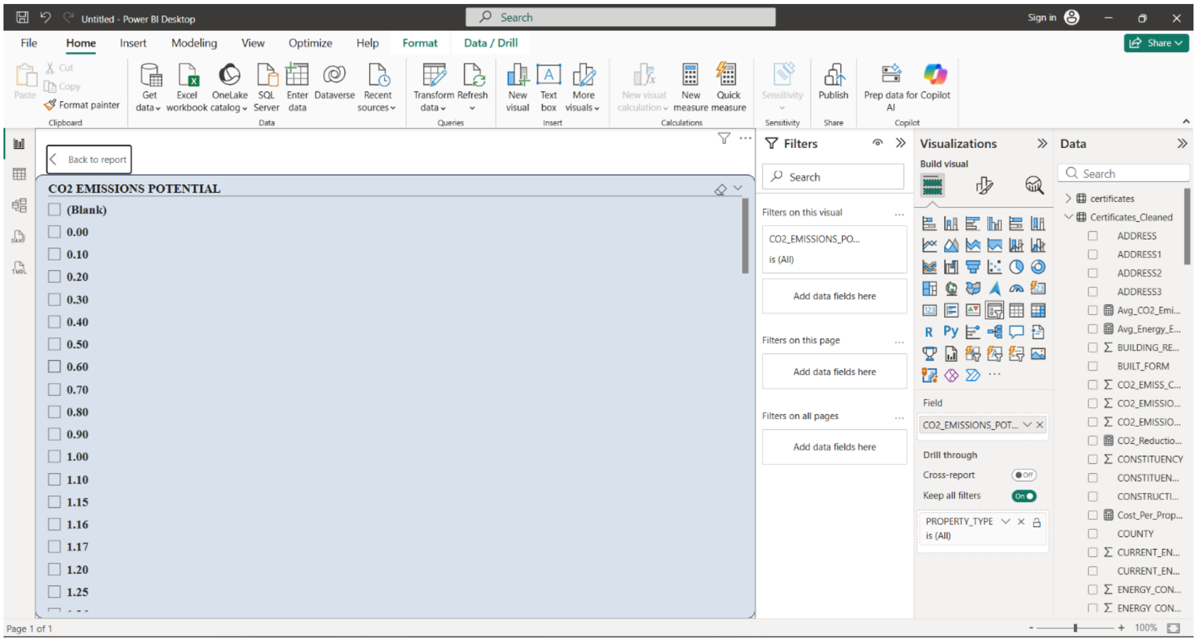


Figure 27: Co2 Emissions Slicer

To filter the dataset using the estimated Co2 emissions value of each property, can use the Co2 Emissions potential slicer. It makes environmental analysis much more comprehensive by assisting in comparing the degree of energy efficiency in different ranges of emissions. This is an instrument that enables decision makers, analysts and homeowners to easily calculate the properties that have a higher or lesser environmental impact.

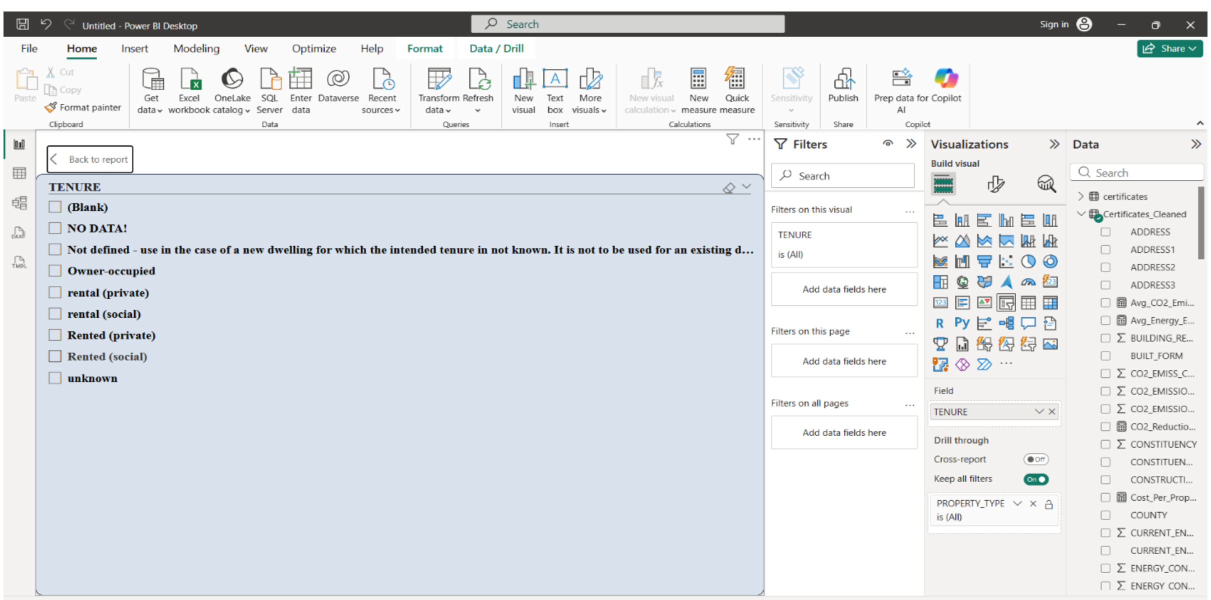


Figure 28: Tenure Slicer

The Tenure Slicer offers filtering options according to each property's occupancy status, including owner -occupied, private/social rental, and unknown categories. This enables users to investigate how different living arrangements affect a property's energy performance. By exposing trends related to ownership, rental markets and unclear tenure types, it also aids in socioeconomic analysis.

Table

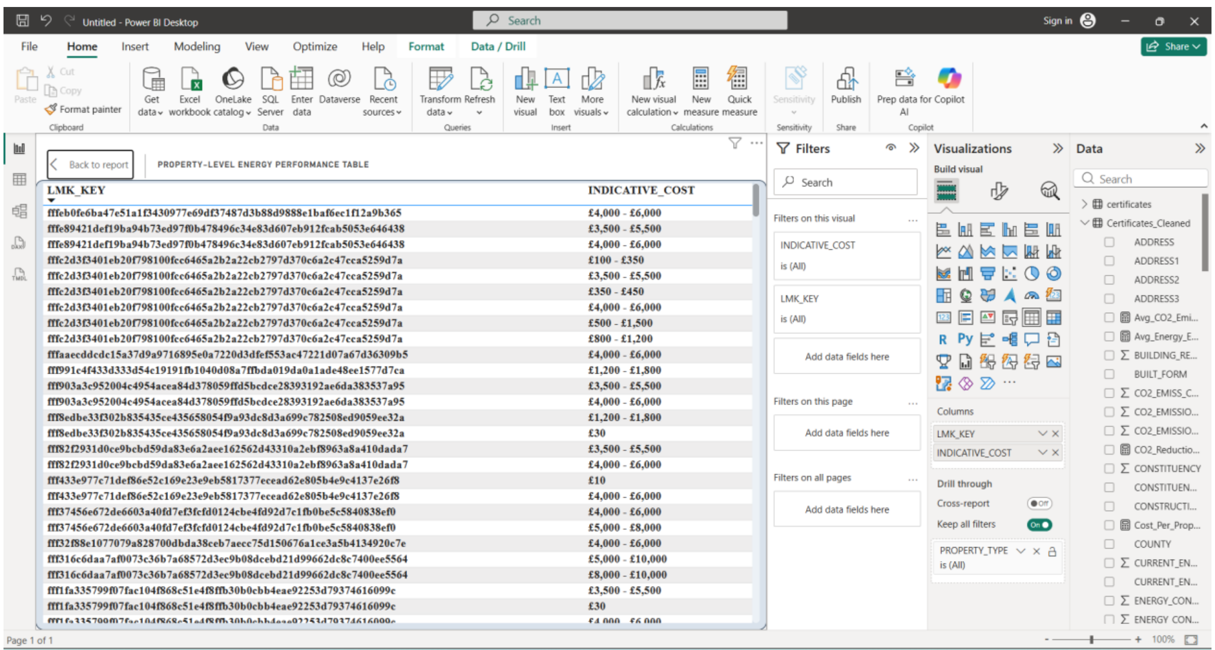


Figure 29: Table

This table shows the possible cost for energy efficiency upgrades by connecting each property's LMK_KEY to its estimated INDICATIVE COST range. It provides a financial viewpoint that aids in evaluating both the potential for Co2 reduction and cost effectiveness.

Tree Map

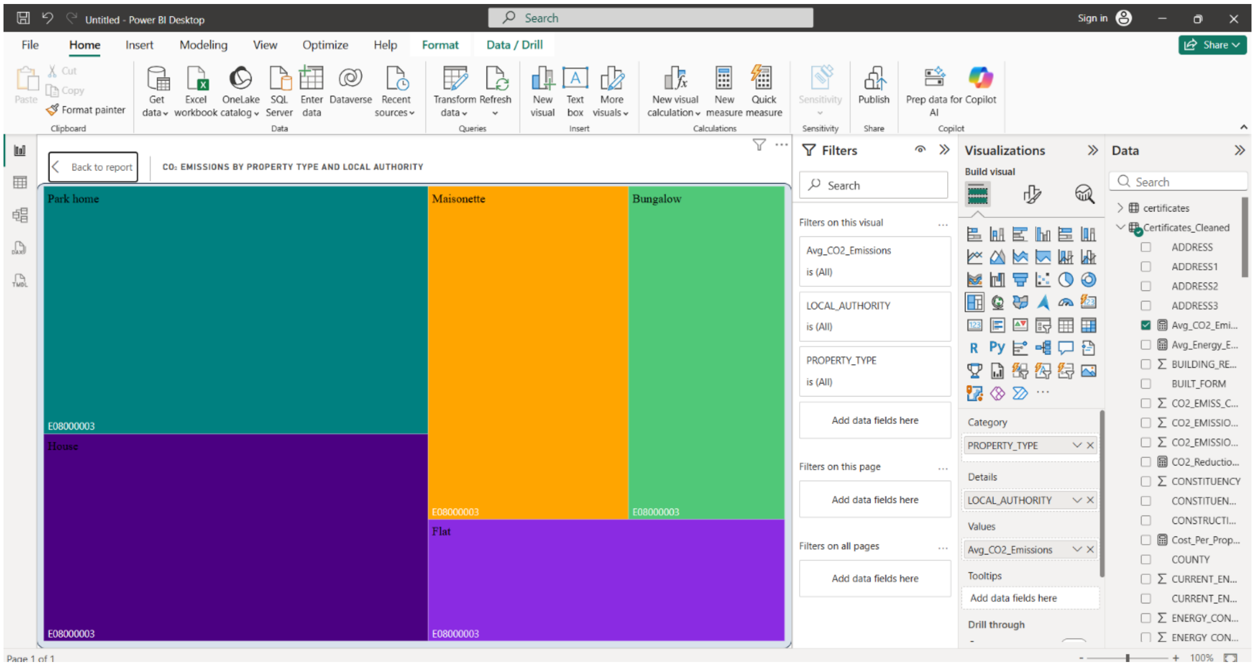


Figure 30: Tree Map

By clearly displaying which property segments generates the highest emissions, this tree map helps in resource prioritization. It recognizes important goals for energy efficiency measures and directs the most efficient use of funds based on the estimated INDICATIVE COST.

4.5 The Connection of the SQL Server with the Power BI

It is easier to analyze the data as validated and it minimized the manual cleaning in Power BI, Provided the steady and updated data by automatic refreshing and enhanced the overall performance.

4.6 Interpretation of the Dashboard

This dashboard allows seeing a comprehensive overview of trends in Energy Performance Certificate (EPC) costs in local authorities throughout Manchester in 2014 – 2024, showing the role of various types of property in the energy consumption and carbon dioxide emissions. The visualizations demonstrate that some types of properties (flats and maisonettes) are more associated with a greater number of emissions in particular local authority areas, whereas the scatter and bar graphs demonstrate significant differences in the current energy efficiency and the potential of its improvement. The interactive filters enable the user to inquire the patterns by property type, tenure and Co2 emissions, giving information on where the targeted inventions may offer the most significant carbon reduction benefits. In general, the dashboard is a visible tool to detect performance gaps and make data-driven decisions to improve the housing conditions to be more energy efficient throughout the region.

4.7 Application of Advanced Power BI Features

Application of advanced Power BI were used to improve the analysis of EPC Manchester Data. DAX was used to create measures and calculated columns to derive insights as average energy scores, comparison of emissions and rating classification. The data was cleaned and structured using Power query transformations thus the errors were removed, the fields were split, and the formats standardized. Interactive capabilities that included slicers, drill through pages were also added to enable users to take a closer look at the performance of the property.

4.8 Key Insights and actionable recommendations

According to the EPC Manchester analysis, a sizable percentage of properties are in the mid-to-lower energy efficiency bands, suggesting that there is a lot of room for improvement. Due in large part to ineffective insulation and heating systems, older properties and some types of properties consistently have lower EPC ratings. Stakeholders are urged to give priority to specific retrofit projects like enhancing roof and wall insulation, switching to energy-efficient heating systems, and encouraging the use of renewable technologies. While landlords and homeowners can use the results to lower energy costs, increase compliance, and boost long-term property value, policymakers and local authorities can use these insights to create incentive programs and awareness campaigns.

5 Conclusion

This comparison of EPC records on local authorities within Manchester between 2014 to 2024 reveals that there are evident differences in the property-level energy efficiency, Co2 emissions and potential improvements. The dashboard indicates that although certain types of properties perform fairly, others are highly effective in the domain of carbon reduced with the help of specific upgrades and policy solutions. The visualization will be an effective decision-making tool as it can be used to identify which areas the company should focus on to implement sustainability strategies in the future by filtering, geographical mapping, and performance indicators. In general, the results support the necessity of the data-driven planning to enhance the housing energy standards, decrease emissions, and facilitate the long term environmental and energy effectiveness objectives of Manchester.