General Sir John Kotelawala Defense University

Faculty of Management, Social Sciences, and Humanities

Department of Language

BSc in Applied Data Science Communication

Advanced SQL and Cloud Databases

Year 2 Semester 2

Group Assignment 02

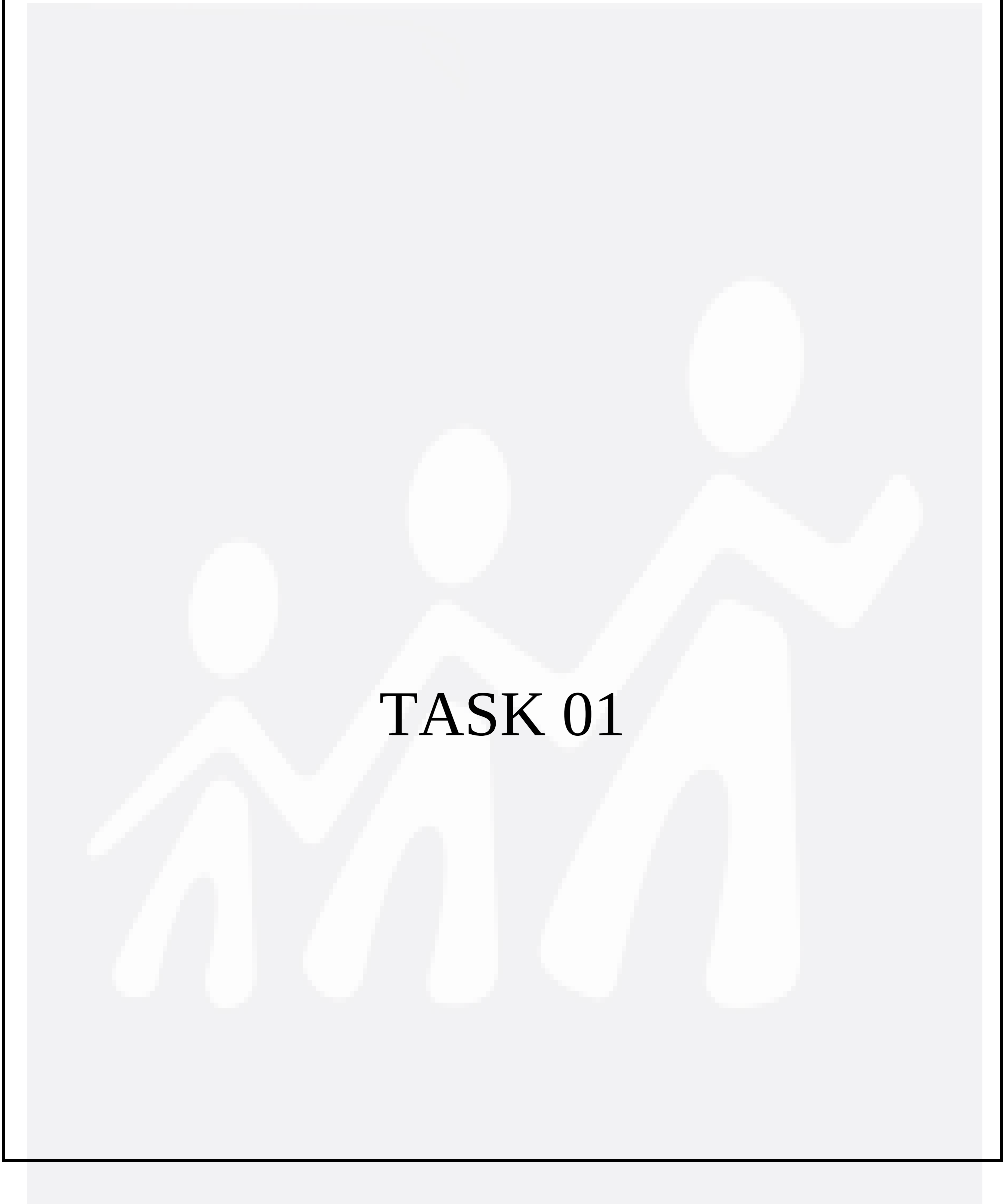
13.10.2024

D/ADC/23/0012 - S.A.R. Maleesha

D/ADC/23/0032 - W.A.K.N Wedagedara

D/ADC/23/0033 - K.G.K Chani

D/ADC/23/0039 - H.W Kaweesha



TASK 01

CONTENT

1.1 Introduction..03

1.2 Downloading the dataset…………………………………………………03

1.3 Exploring the dataset..04

1.4 Importing the dataset and implementation of SQL...................................06

1.5 Implementation in Microsoft Report Builder..10

- Household size differences between urban and rural areas
- Impact of birth weight on child health
- Exploring the impact of household economic strain on child health
- The link between food security and child health
- The Influence of Parental Education on Child Development
- Impacting of Alcohol and Smoking

1.6 Conclusion…………………………………………………………26

1.1 Introduction

A long-term initiative called Child Well-Being Monitor was created to assess and examine childhood poverty in lower-income nations. The project makes use of Power BI, SQL Server, and Report Builder to extract valuable insights about the well-being of children from data and present it visually. We highlight the Young Lives study, a ground-breaking research project that has been monitoring 12,000 children's lives in Ethiopia, India, Peru, and Vietnam since 2001. Enhancing knowledge of childhood poverty and its impact on health, education, and general well-being is the goal of this project, which is being conducted by the Department of International Development at the University of Oxford and funded by the Department for International Development.

We will examine statistics from Ethiopia and India in this paper, with an emphasis on four main areas: social well-being, health, education, and economic standing. We want to employ Power BI Report Builder for report development and SQL Server for data administration in order to produce comprehensive insights and visualizations that emphasize the main variables influencing child poverty. With the use of these technologies, we will be able to create a user-friendly reporting system that will guide the development of intervention plans and legislation to lessen child poverty in emerging nations.

With the use of SQL Server and Power BI, we will demonstrate our expertise in database administration, data analysis, and report creation, providing useful information on child poverty in the two chosen nations.

1.2 Downloading the dataset

We downloaded the dataset from the following link:

https://beta.ukdataservice.ac.uk/datacatalogue/series/series?id=2000060#!/access

We started by creating a username and registering for the UK Data Service. We downloaded the "International Study of Child Poverty: Rounds 1-5 Constructed Files, 2002-2016" dataset after obtaining access. We were able to utilize the login and password we had created upon registration to log in once we gained access. The dataset was then successfully uploaded to our account after this. We also finished the dataset download and submission process by submitting the abstract for our investigation.

1.3 Exploring the dataset

'Constructed files' are supplied to assist researchers utilizing the Young Lives data in addition to the raw information. These created files contain aggregated subsets of characteristics from various rounds of the Young Lives child and household surveys, which were carried out between 2002 and 2016. These data files are an updated version of previously archived data that encompass several rounds and guarantee uniformity in the round-to-round definitions of the variables. To keep things uniform, certain variables have been changed but the majority have remained the same.

The CSV format was created for these files using a particular procedure. Every file with tabs was initially opened with a text editor such as Notepad and saved in text format. The text files were imported into different Excel worksheets and then saved as CSV files by using Microsoft Excel's features. The tabular data from the datasets may now be more easily accessed and used in CSV format thanks to this conversion process.
Key details about the sort of data contained in each dataset are provided by the name, description, and number of rows and columns.

File name	Description	No of Rows
all_countries_math_irt_scores.csv	Includes children's math proficiency ratings from several different nations. Data is grouped by group, survey round, and nation.	32,632 rows

all_countries_ppvt_irt_scores.csv	Includes results from the Peabody Picture Vocabulary Test (PPVT), which measures verbal and vocabulary skills. Scores from multiple survey rounds, countries, and linguistic groups are included in the dataset.	34,010 rows
all_countries_reading_irt_scores.csv	Gives children's reading proficiency scores across national boundaries using information from many survey rounds and linguistic groupings.	17451 rows
et_in_pe_sibling_ppvt_irt_scores.csv	Focuses on the PPVT results of siblings from Peru, Ethiopia, and India, emphasizing the differences in the siblings' ability and language scores.	6164 rows
ethiopia_constructed.csv	A large dataset with precise socioeconomic, educational, and household demographic information from Ethiopia. This contains details about the family history, composition of the home, and child development.	14995 rows
india_constructed.csv	This dataset, which is similar to the Ethiopia built file, focuses on household and child characteristics in India. It contains information on test results, household assets, education, and literacy.	15097 rows
india_sibling_math_irt_scores.csv	Includes Indian siblings' math test results, enabling study of performance and educational variations between siblings.	3023 rows
peru_constructed.csv	Comprehensive socioeconomic and child development data from Peru that includes test results, household variables, and education.	13830 rows
vietnam_constructed.csv	Rich dataset on Vietnamese families' circumstances and children's learning	15,000 rows

	outcomes, with an emphasis on home and school features.	

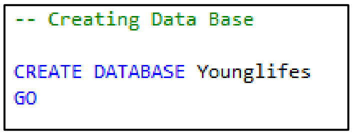
In order to give researchers context for the Young Lives data and to reflect the pertinent questions posed to households and children in subsequent survey rounds, additional variables were added (Azubuike & Briones, 2016). These variables are arranged in the created files into four primary categories: panel information, child characteristics, household characteristics, and identity and location factors. Among the categories are

- Identification and locating variables
- Panel details
- Child characteristics
- Household characteristics

For our case study in the report, we have selected the **india_constructed** and **vietnam_constructed** tables.

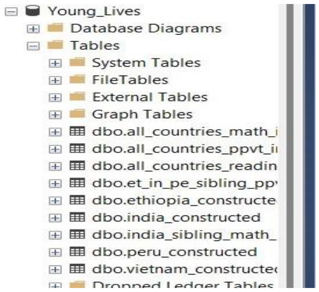
1.4 Importing the dataset and implementation of SQL

Once we downloaded the dataset, we saved it as a csv file format. Then we created a database called “Young lifes” in SQL server studio.



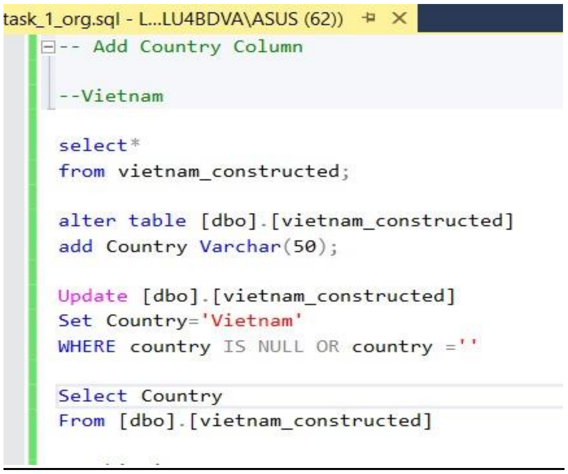
```
-- Creating Data Base

CREATE DATABASE Younglifes
GO
```

Data Base



```
⊟ Young_Lives
  ⊞ Database Diagrams
  ⊟ Tables
    ⊞ System Tables
    ⊞ FileTables
    ⊞ External Tables
    ⊞ Graph Tables
    ⊞ dbo.all_countries_math_i
    ⊞ dbo.all_countries_ppvt_ii
    ⊞ dbo.all_countries_readin
    ⊞ dbo.et_in_pe_sibling_ppv
    ⊞ dbo.ethiopia_constructe
    ⊞ dbo.india_constructed
    ⊞ dbo.india_sibling_math_
    ⊞ dbo.peru_constructed
    ⊞ dbo.vietnam_constructed
    ⊞ Dropped Ledger Tables
```

Subsequently, we utilized the ALTER TABEL statement to incorporate a new column named 'Country' into the current database, designating it as VARCHAR(50). Next, we changed this new column so that for any entry in which the country field included either an empty string or NULL, 'Vietnam' was assigned as the country name. Ultimately, we obtained the most recent national data to confirm the modifications.

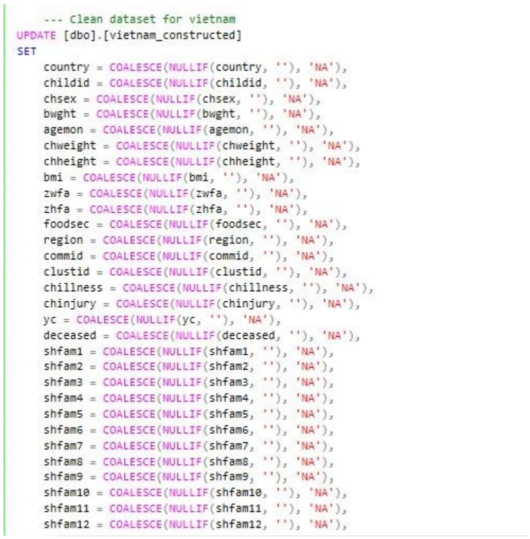


task_1_org.sql - L...LU4BDVA\ASUS (62))

```
-- Add Country Column

--Vietnam

select*
from vietnam_constructed;

alter table [dbo].[vietnam_constructed]
add Country Varchar(50);

Update [dbo].[vietnam_constructed]
Set Country='Vietnam'
WHERE country IS NULL OR country =''

Select Country
From [dbo].[vietnam_constructed]
```

** Although instructions were to utilize Power BI’s data transformation tools to clean the dataset, we encountered technical challenges in establishing a seamless connection between the SQL database and Power BI. As a result, we opted to perform the data cleaning process directly using SQL**

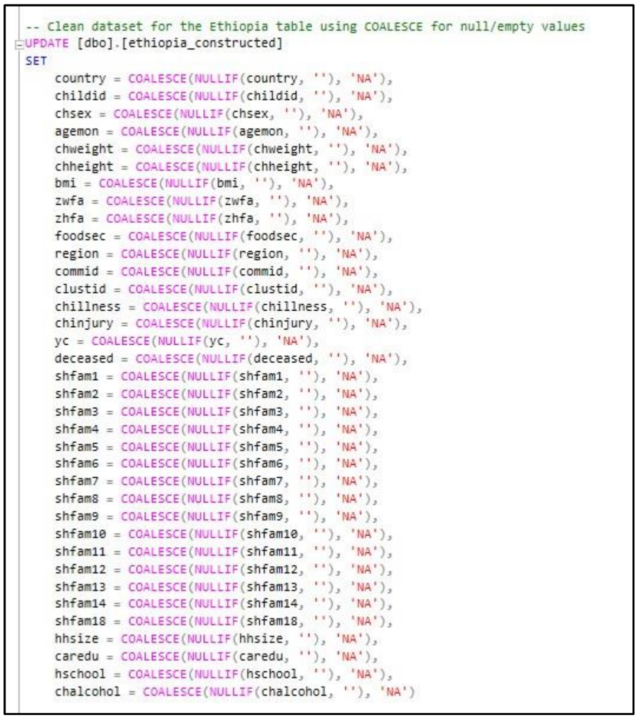
We discovered inconsistent and missing data in several columns when we cleaned the dataset. We resolved this issue by using SQL Server Management Studio rather than doing the preprocessing in Excel. To ensure data consistency throughout the table, we used the COALESCE function to replace NULL or empty entries in each column with 'NA'.



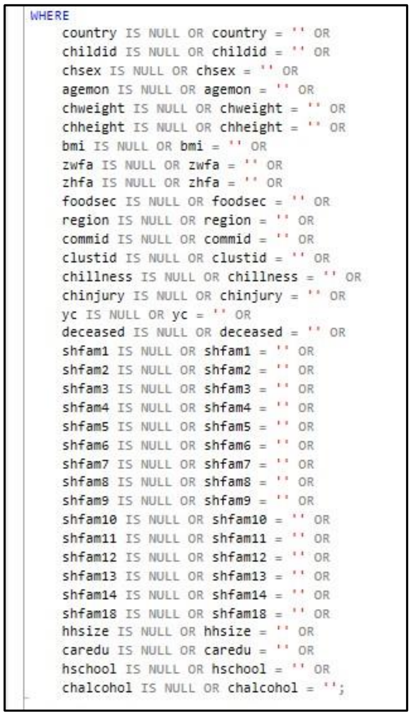
```
    --- Clean dataset for vietnam
UPDATE [dbo].[vietnam_constructed]
SET
    country = COALESCE(NULLIF(country, ''), 'NA'),
    childid = COALESCE(NULLIF(childid, ''), 'NA'),
    chsex = COALESCE(NULLIF(chsex, ''), 'NA'),
    bwght = COALESCE(NULLIF(bwght, ''), 'NA'),
    agemon = COALESCE(NULLIF(agemon, ''), 'NA'),
    chweight = COALESCE(NULLIF(chweight, ''), 'NA'),
    chheight = COALESCE(NULLIF(chheight, ''), 'NA'),
    bmi = COALESCE(NULLIF(bmi, ''), 'NA'),
    zwfa = COALESCE(NULLIF(zwfa, ''), 'NA'),
    zhfa = COALESCE(NULLIF(zhfa, ''), 'NA'),
    foodsec = COALESCE(NULLIF(foodsec, ''), 'NA'),
    region = COALESCE(NULLIF(region, ''), 'NA'),
    commid = COALESCE(NULLIF(commid, ''), 'NA'),
    clustid = COALESCE(NULLIF(clustid, ''), 'NA'),
    chillness = COALESCE(NULLIF(chillness, ''), 'NA'),
    chinjury = COALESCE(NULLIF(chinjury, ''), 'NA'),
    yc = COALESCE(NULLIF(yc, ''), 'NA'),
    deceased = COALESCE(NULLIF(deceased, ''), 'NA'),
    shfam1 = COALESCE(NULLIF(shfam1, ''), 'NA'),
    shfam2 = COALESCE(NULLIF(shfam2, ''), 'NA'),
    shfam3 = COALESCE(NULLIF(shfam3, ''), 'NA'),
    shfam4 = COALESCE(NULLIF(shfam4, ''), 'NA'),
    shfam5 = COALESCE(NULLIF(shfam5, ''), 'NA'),
    shfam6 = COALESCE(NULLIF(shfam6, ''), 'NA'),
    shfam7 = COALESCE(NULLIF(shfam7, ''), 'NA'),
    shfam8 = COALESCE(NULLIF(shfam8, ''), 'NA'),
    shfam9 = COALESCE(NULLIF(shfam9, ''), 'NA'),
    shfam10 = COALESCE(NULLIF(shfam10, ''), 'NA'),
    shfam11 = COALESCE(NULLIF(shfam11, ''), 'NA'),
    shfam12 = COALESCE(NULLIF(shfam12, ''), 'NA'),
```

```
    --- Clean dataset for vietnam
UPDATE [dbo].[vietnam_constructed]
SET
    country = COALESCE(NULLIF(country, ''), 'NA'),
    childid = COALESCE(NULLIF(childid, ''), 'NA'),
    chsex = COALESCE(NULLIF(chsex, ''), 'NA'),
    bwght = COALESCE(NULLIF(bwght, ''), 'NA'),
    agemon = COALESCE(NULLIF(agemon, ''), 'NA'),
    chweight = COALESCE(NULLIF(chweight, ''), 'NA'),
    chheight = COALESCE(NULLIF(chheight, ''), 'NA'),
    bmi = COALESCE(NULLIF(bmi, ''), 'NA'),
    zwfa = COALESCE(NULLIF(zwfa, ''), 'NA'),
    zhfa = COALESCE(NULLIF(zhfa, ''), 'NA'),
    foodsec = COALESCE(NULLIF(foodsec, ''), 'NA'),
    region = COALESCE(NULLIF(region, ''), 'NA'),
    commid = COALESCE(NULLIF(commid, ''), 'NA'),
    clustid = COALESCE(NULLIF(clustid, ''), 'NA'),
    chillness = COALESCE(NULLIF(chillness, ''), 'NA'),
    chinjury = COALESCE(NULLIF(chinjury, ''), 'NA'),
    yc = COALESCE(NULLIF(yc, ''), 'NA'),
    deceased = COALESCE(NULLIF(deceased, ''), 'NA'),
    shfam1 = COALESCE(NULLIF(shfam1, ''), 'NA'),
    shfam2 = COALESCE(NULLIF(shfam2, ''), 'NA'),
    shfam3 = COALESCE(NULLIF(shfam3, ''), 'NA'),
    shfam4 = COALESCE(NULLIF(shfam4, ''), 'NA'),
    shfam5 = COALESCE(NULLIF(shfam5, ''), 'NA'),
    shfam6 = COALESCE(NULLIF(shfam6, ''), 'NA'),
    shfam7 = COALESCE(NULLIF(shfam7, ''), 'NA'),
    shfam8 = COALESCE(NULLIF(shfam8, ''), 'NA'),
    shfam9 = COALESCE(NULLIF(shfam9, ''), 'NA'),
    shfam10 = COALESCE(NULLIF(shfam10, ''), 'NA'),
    shfam11 = COALESCE(NULLIF(shfam11, ''), 'NA'),
    shfam12 = COALESCE(NULLIF(shfam12, ''), 'NA'),
```

Delete Null Values from tables.



```sql
-- Clean dataset for the Ethiopia table using COALESCE for null/empty values
UPDATE [dbo].[ethiopia_constructed]
SET
    country = COALESCE(NULLIF(country, ''), 'NA'),
    childid = COALESCE(NULLIF(childid, ''), 'NA'),
    chsex = COALESCE(NULLIF(chsex, ''), 'NA'),
    agemon = COALESCE(NULLIF(agemon, ''), 'NA'),
    chweight = COALESCE(NULLIF(chweight, ''), 'NA'),
    chheight = COALESCE(NULLIF(chheight, ''), 'NA'),
    bmi = COALESCE(NULLIF(bmi, ''), 'NA'),
    zwfa = COALESCE(NULLIF(zwfa, ''), 'NA'),
    zhfa = COALESCE(NULLIF(zhfa, ''), 'NA'),
    foodsec = COALESCE(NULLIF(foodsec, ''), 'NA'),
    region = COALESCE(NULLIF(region, ''), 'NA'),
    commid = COALESCE(NULLIF(commid, ''), 'NA'),
    clustid = COALESCE(NULLIF(clustid, ''), 'NA'),
    chillness = COALESCE(NULLIF(chillness, ''), 'NA'),
    chinjury = COALESCE(NULLIF(chinjury, ''), 'NA'),
    yc = COALESCE(NULLIF(yc, ''), 'NA'),
    deceased = COALESCE(NULLIF(deceased, ''), 'NA'),
    shfam1 = COALESCE(NULLIF(shfam1, ''), 'NA'),
    shfam2 = COALESCE(NULLIF(shfam2, ''), 'NA'),
    shfam3 = COALESCE(NULLIF(shfam3, ''), 'NA'),
    shfam4 = COALESCE(NULLIF(shfam4, ''), 'NA'),
    shfam5 = COALESCE(NULLIF(shfam5, ''), 'NA'),
    shfam6 = COALESCE(NULLIF(shfam6, ''), 'NA'),
    shfam7 = COALESCE(NULLIF(shfam7, ''), 'NA'),
    shfam8 = COALESCE(NULLIF(shfam8, ''), 'NA'),
    shfam9 = COALESCE(NULLIF(shfam9, ''), 'NA'),
    shfam10 = COALESCE(NULLIF(shfam10, ''), 'NA'),
    shfam11 = COALESCE(NULLIF(shfam11, ''), 'NA'),
    shfam12 = COALESCE(NULLIF(shfam12, ''), 'NA'),
    shfam13 = COALESCE(NULLIF(shfam13, ''), 'NA'),
    shfam14 = COALESCE(NULLIF(shfam14, ''), 'NA'),
    shfam18 = COALESCE(NULLIF(shfam18, ''), 'NA'),
    hhsize = COALESCE(NULLIF(hhsize, ''), 'NA'),
    caredu = COALESCE(NULLIF(caredu, ''), 'NA'),
    hschool = COALESCE(NULLIF(hschool, ''), 'NA'),
    chalcohol = COALESCE(NULLIF(chalcohol, ''), 'NA')
```

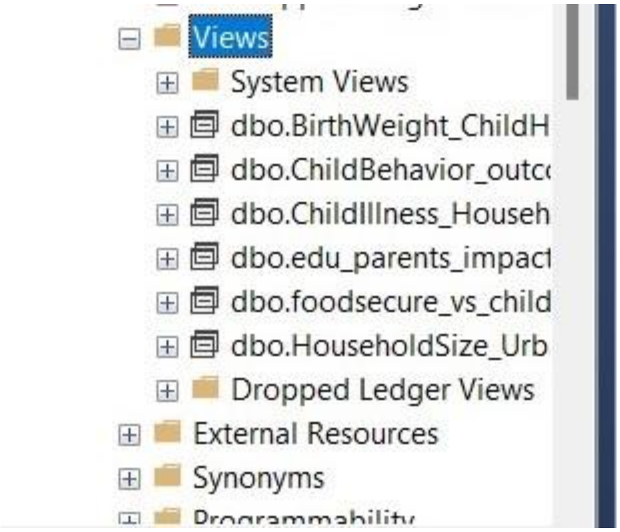


```sql
WHERE
    country IS NULL OR country = '' OR
    childid IS NULL OR childid = '' OR
    chsex IS NULL OR chsex = '' OR
    agemon IS NULL OR agemon = '' OR
    chweight IS NULL OR chweight = '' OR
    chheight IS NULL OR chheight = '' OR
    bmi IS NULL OR bmi = '' OR
    zwfa IS NULL OR zwfa = '' OR
    zhfa IS NULL OR zhfa = '' OR
    foodsec IS NULL OR foodsec = '' OR
    region IS NULL OR region = '' OR
    commid IS NULL OR commid = '' OR
    clustid IS NULL OR clustid = '' OR
    chillness IS NULL OR chillness = '' OR
    chinjury IS NULL OR chinjury = '' OR
    yc IS NULL OR yc = '' OR
    deceased IS NULL OR deceased = '' OR
    shfam1 IS NULL OR shfam1 = '' OR
    shfam2 IS NULL OR shfam2 = '' OR
    shfam3 IS NULL OR shfam3 = '' OR
    shfam4 IS NULL OR shfam4 = '' OR
    shfam5 IS NULL OR shfam5 = '' OR
    shfam6 IS NULL OR shfam6 = '' OR
    shfam7 IS NULL OR shfam7 = '' OR
    shfam8 IS NULL OR shfam8 = '' OR
    shfam9 IS NULL OR shfam9 = '' OR
    shfam10 IS NULL OR shfam10 = '' OR
    shfam11 IS NULL OR shfam11 = '' OR
    shfam12 IS NULL OR shfam12 = '' OR
    shfam13 IS NULL OR shfam13 = '' OR
    shfam14 IS NULL OR shfam14 = '' OR
    shfam18 IS NULL OR shfam18 = '' OR
    hhsize IS NULL OR hhsize = '' OR
    caredu IS NULL OR caredu = '' OR
    hschool IS NULL OR hschool = '' OR
    chalcohol IS NULL OR chalcohol = '';
```

Six fresh concepts pertaining to the Young Life initiative were produced by the filtering that was done the following day. Every concept was explored by choosing the appropriate, pertinent columns.

- The link between food security and child health
- Household size differences between urban and rural areas
- Impact of birth weight on child health
- Exploring the impact of household economic strain on child health
- The Influence of Parental Education on Child Development
- Impacting of Alcohol and Smoking

In this way, we have developed six perspectives based on the six previously indicated themes.



1.5 Implementation in Microsoft Report Builder

Create an Embedded Data Source in Report Builder

1. Open Report Builder

   Open the Report Builder software on your desktop.

   The New Report or Dataset dialog box will show up when it opens.

2. Launch a New Report

   Make sure that New Report is chosen in the dialog box's left pane.

   Click on Table or Matrix Wizard in the right pane. This will launch a fresh report creation interface.

3. Select a Dataset

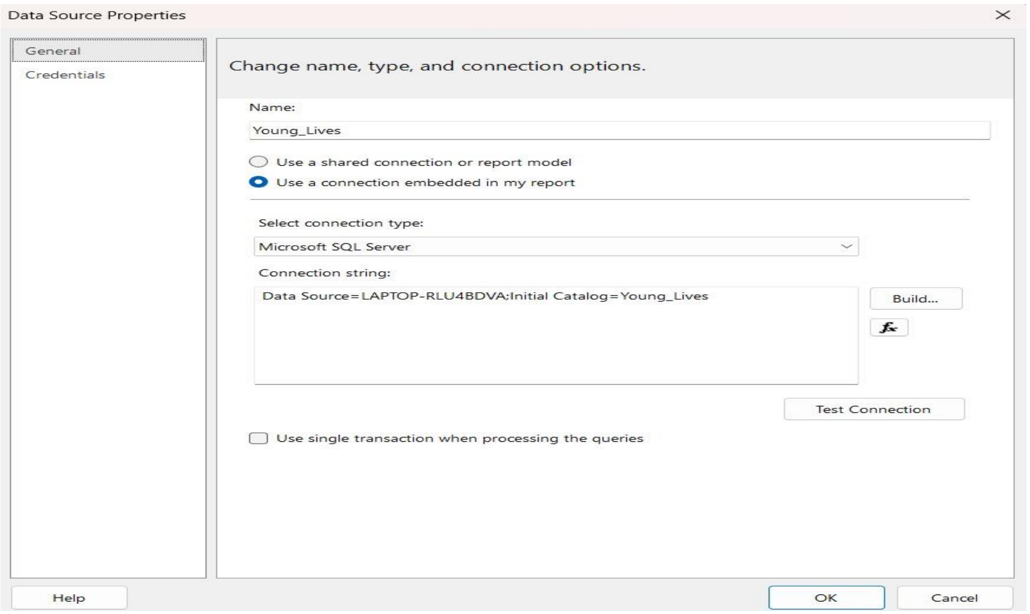
   Select the Create a dataset option on the Choose a dataset screen.

   To continue, click Next. This will take you to the screen where you can select a connection to a data source.

4. Establish a New Source of Data

   Select New. In order to configure a new data source for the report, this will open the Data Source Properties dialog box.

   Identify and Set Up



Name: Young_Lives

○ Use a shared connection or report model

● Use a connection embedded in my report

Select connection type: Microsoft SQL Server

Connection string: Data Source=LAPTOP-RLU4BDVA;Initial Catalog=Young_Lives

□ Use single transaction when processing the queries

5. Construct a Connection String:

   To generate a new connection string, click Build in the Connection string area.

6. Indicate the SQL Server Instance:

   Enter the name of your SQL Server instance in the Server Name field or choose it from the drop-down menu.

7. Choose the Database:

   Choose the required database (in this case, "younglifes") from the drop-down option.

8. Check the Link:

To make sure the connection to the data source is operating correctly, click Test Connection.

If everything goes well, the message "Connection created successfully" will appear.

9. Verify Connection:

   To save the connection, click OK.

   After selecting your new data source, you will be returned to the Select a connection to a data source screen. To continue, click Next.

TO CREATE A QUERY

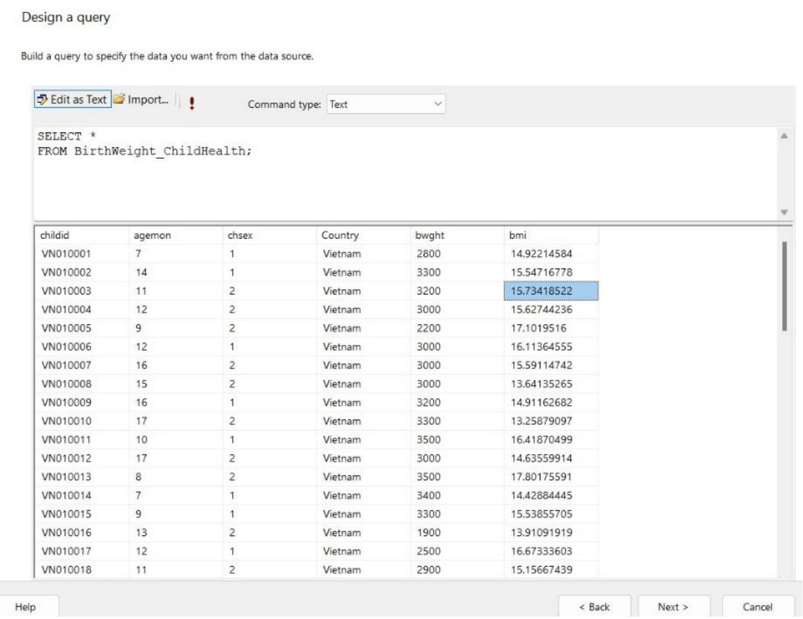
1. Design a Query:

   On the Design a query page, the relational query designer is open by default. For this tutorial, switch to the text-based query designer.

2. Switch to Text-based Query:

   Click Edit as Text to open the text-based query designer, displaying a query pane and a results pane.

3. Input SQL Query:



Design a query

Build a query to specify the data you want from the data source.

Command type: Text

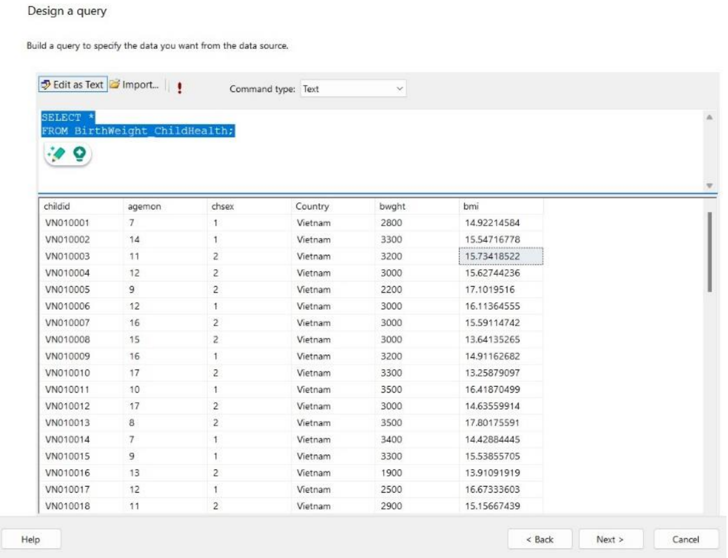
```
SELECT *
FROM BirthWeight_ChildHealth;
```

childid	agemon	chsex	Country	bwght	bmi
VN010001	7	1	Vietnam	2800	14.92214584
VN010002	14	1	Vietnam	3300	15.54716778
VN010003	11	2	Vietnam	3200	15.73418522
VN010004	12	2	Vietnam	3000	15.62744236
VN010005	9	2	Vietnam	2200	17.1019516
VN010006	12	1	Vietnam	3000	16.11364555
VN010007	16	2	Vietnam	3000	15.59114742
VN010008	15	2	Vietnam	3000	13.64135265
VN010009	16	1	Vietnam	3200	14.91162682
VN010010	17	2	Vietnam	3300	13.25879097
VN010011	10	1	Vietnam	3500	16.41870499
VN010012	17	2	Vietnam	3000	14.63559914
VN010013	8	2	Vietnam	3500	17.80175591
VN010014	7	1	Vietnam	3400	14.42884445
VN010015	9	1	Vietnam	3300	15.53855705
VN010016	13	2	Vietnam	1900	13.91091919
VN010017	12	1	Vietnam	2500	16.67333603
VN010018	11	2	Vietnam	2900	15.15667439

4. Run the Query:

   On the query designer toolbar, click Run (!) to execute the query.

   The results pane will display the data fetched from the BirthWeight_ChildHealth table.



Design a query

Build a query to specify the data you want from the data source.

```
SELECT *
FROM BirthWeight_ChildHealth;
```

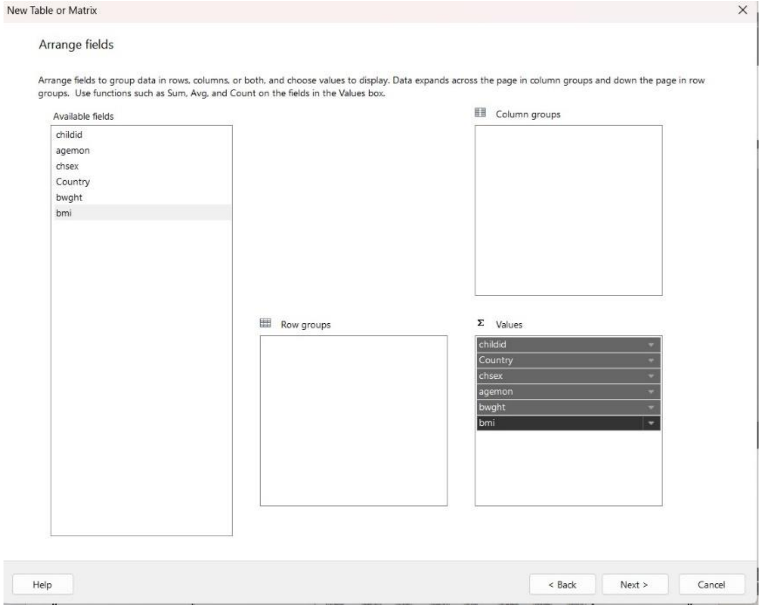
childid	agemon	chsex	Country	bwght	bmi
VN010001	7	1	Vietnam	2800	14.92214584
VN010002	14	1	Vietnam	3300	15.54716778
VN010003	11	2	Vietnam	3200	15.73418522
VN010004	12	2	Vietnam	3000	15.62744236
VN010005	9	2	Vietnam	2200	17.1019516
VN010006	12	1	Vietnam	3000	16.11364555
VN010007	16	2	Vietnam	3000	15.59114742
VN010008	15	2	Vietnam	3000	13.64135265
VN010009	16	1	Vietnam	3200	14.91162682
VN010010	17	2	Vietnam	3300	13.25879097
VN010011	10	1	Vietnam	3500	16.41870499
VN010012	17	2	Vietnam	3000	14.63559914
VN010013	8	2	Vietnam	3500	17.80175591
VN010014	7	1	Vietnam	3400	14.42884445
VN010015	9	1	Vietnam	3300	15.53855705
VN010016	13	2	Vietnam	1900	13.91091919
VN010017	12	1	Vietnam	2500	16.67333603
VN010018	11	2	Vietnam	2900	15.15667439

5. Proceed to the Next Step:

   Click Next to move on to the grouping and layout phase.

6. Organize Data into Values:

   Drag all the columns into the Values section of the wizard to include them in your report.



New Table or Matrix

Arrange fields

Arrange fields to group data in rows, columns, or both, and choose values to display. Data expands across the page in column groups and down the page in row groups. Use functions such as Sum, Avg, and Count on the fields in the Values box.

Available fields

- childid
- agemon
- chsex
- Country
- bwght
- bmi

Column groups

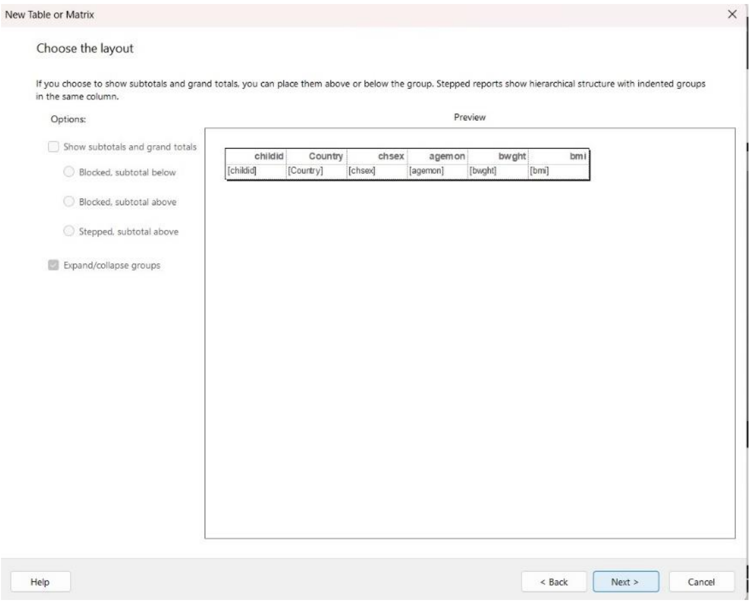
Row groups

Values

- childid
- Country
- chsex
- agemon
- bwght
- bmi

7. Continue to Preview:

Click Next to proceed. Click Next again to preview the table structure, and then click Finish to complete the setup.



New Table or Matrix

Choose the layout

If you choose to show subtotals and grand totals, you can place them above or below the group. Stepped reports show hierarchical structure with indented groups in the same column.

Options:

□ Show subtotals and grand totals

○ Blocked, subtotal below

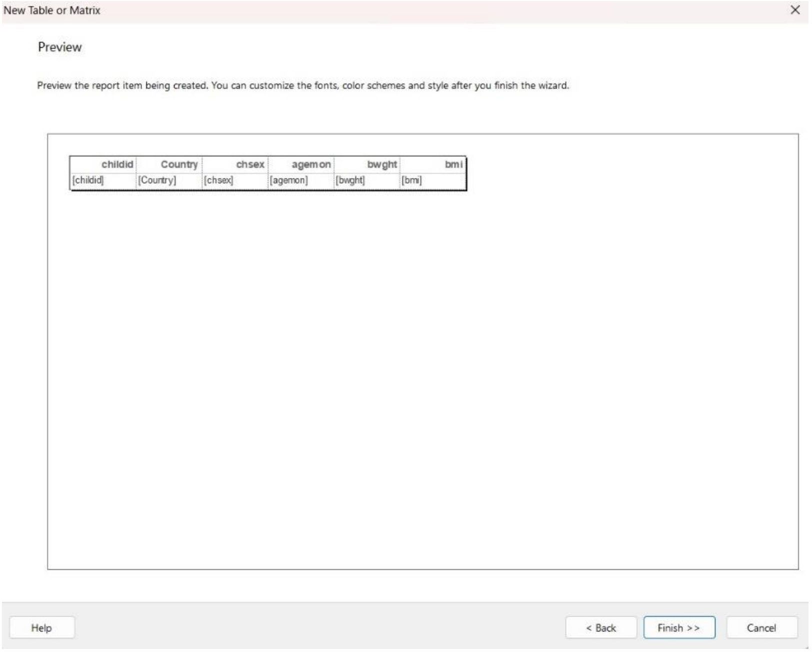
○ Blocked, subtotal above

○ Stepped, subtotal above

☑ Expand/collapse groups

Preview

childid	Country	chsex	agemon	bwght	bmi
[childid]	[Country]	[chsex]	[agemon]	[bwght]	[bmi]


New Table or Matrix

Preview

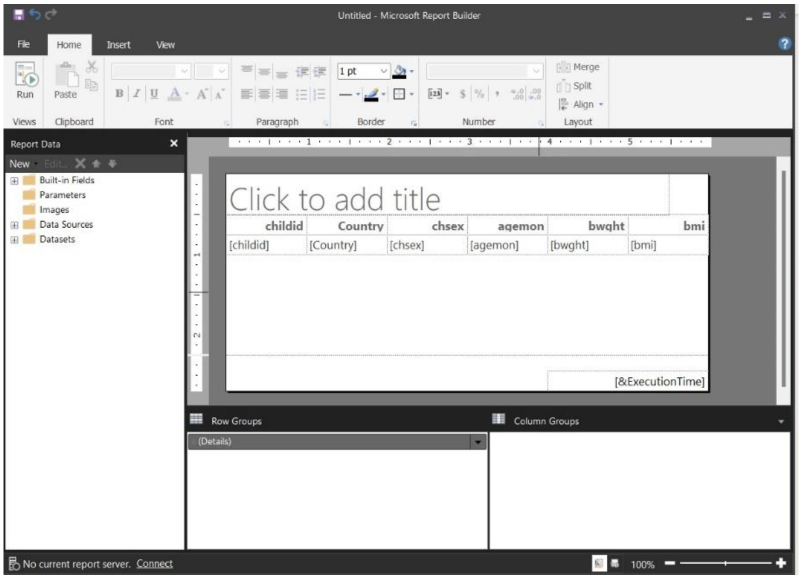
Preview the report item being created. You can customize the fonts, color schemes and style after you finish the wizard.

childid	Country	chsex	agemon	bwght	bmi
[childid]	[Country]	[chsex]	[agemon]	[bwght]	[bmi]

Finalizing and Previewing the Report

8. Add the Table to Design Surface:

   After finishing the wizard, the table is added to the design surface of the report.


Untitled - Microsoft Report Builder

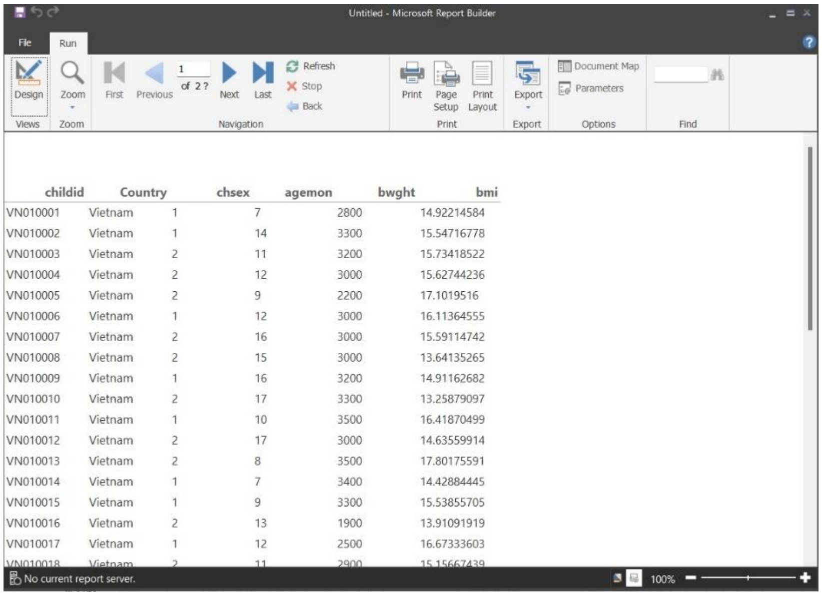
Click to add title

childid	Country	chsex	agemon	bwght	bmi
[childid]	[Country]	[chsex]	[agemon]	[bwght]	[bmi]

[&ExecutionTime]

9. Preview the Report:

   Click Run to preview your report. This allows you to see how the data will be displayed in the final report.



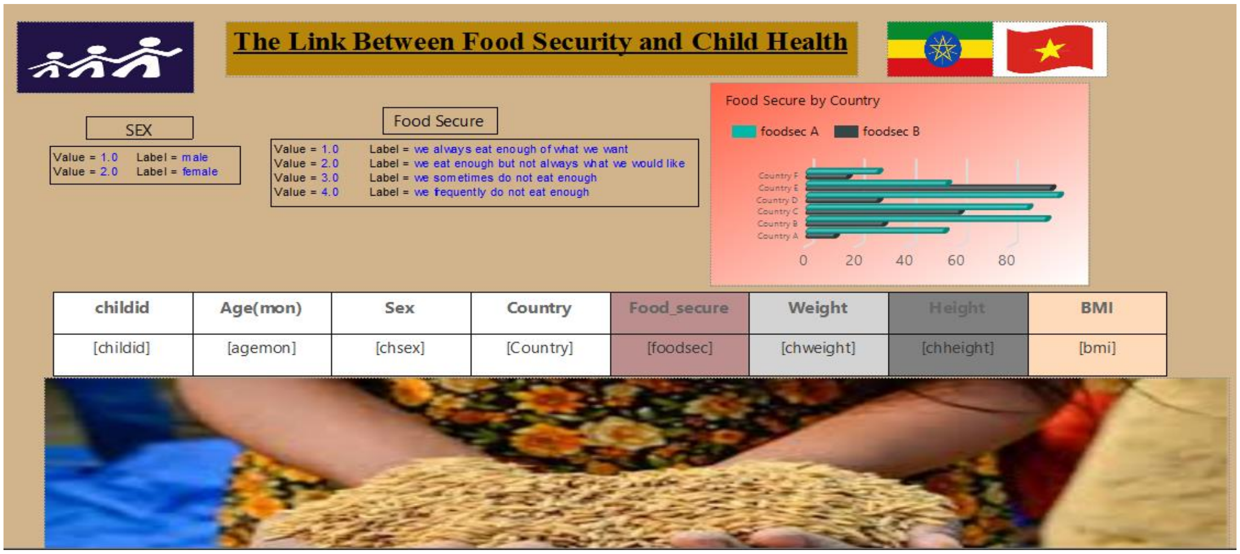
childid	Country	chsex	agemon	bwght	bmi
VN010001	Vietnam	1	7	2800	14.92214584
VN010002	Vietnam	1	14	3300	15.54716778
VN010003	Vietnam	2	11	3200	15.73418522
VN010004	Vietnam	2	12	3000	15.62744236
VN010005	Vietnam	2	9	2200	17.1019516
VN010006	Vietnam	1	12	3000	16.11364555
VN010007	Vietnam	2	16	3000	15.59114742
VN010008	Vietnam	2	15	3000	13.64135265
VN010009	Vietnam	1	16	3200	14.91162682
VN010010	Vietnam	2	17	3300	13.25879097
VN010011	Vietnam	1	10	3500	16.41870499
VN010012	Vietnam	2	17	3000	14.63559914
VN010013	Vietnam	2	8	3500	17.80175591
VN010014	Vietnam	1	7	3400	14.42884445
VN010015	Vietnam	1	9	3300	15.53855705
VN010016	Vietnam	2	13	1900	13.91091919
VN010017	Vietnam	1	12	2500	16.67333603
VN010018	Vietnam	2	11	2900	15.15667439

Creating reports using Microsoft Report Builder

❖ Report 1: The link between food security and child health

In Vietnam and Ethiopia, food security has a significant impact on the health of children. Children from households experiencing food insecurity are more likely to suffer from malnutrition in Ethiopia, where food shortages and drought are prevalent. Stunted growth compromised immune systems, and heightened susceptibility to infections are frequently the results of this. Long-term cognitive and physical problems may arise from inadequate nutrition throughout critical developmental phases.

Food insecurity is a serious problem in Vietnam as well, especially in rural regions. Low-income children might not have as much access to a variety of nutrient-dense foods, which could result in vitamin and mineral deficits. Because undernourished children frequently find it difficult to focus in class, this has a detrimental impact on both academic performance and physical health. Improving the general health and development of children in these nations requires ensuring food security.



The Link Between Food Security and Child Health

SEX

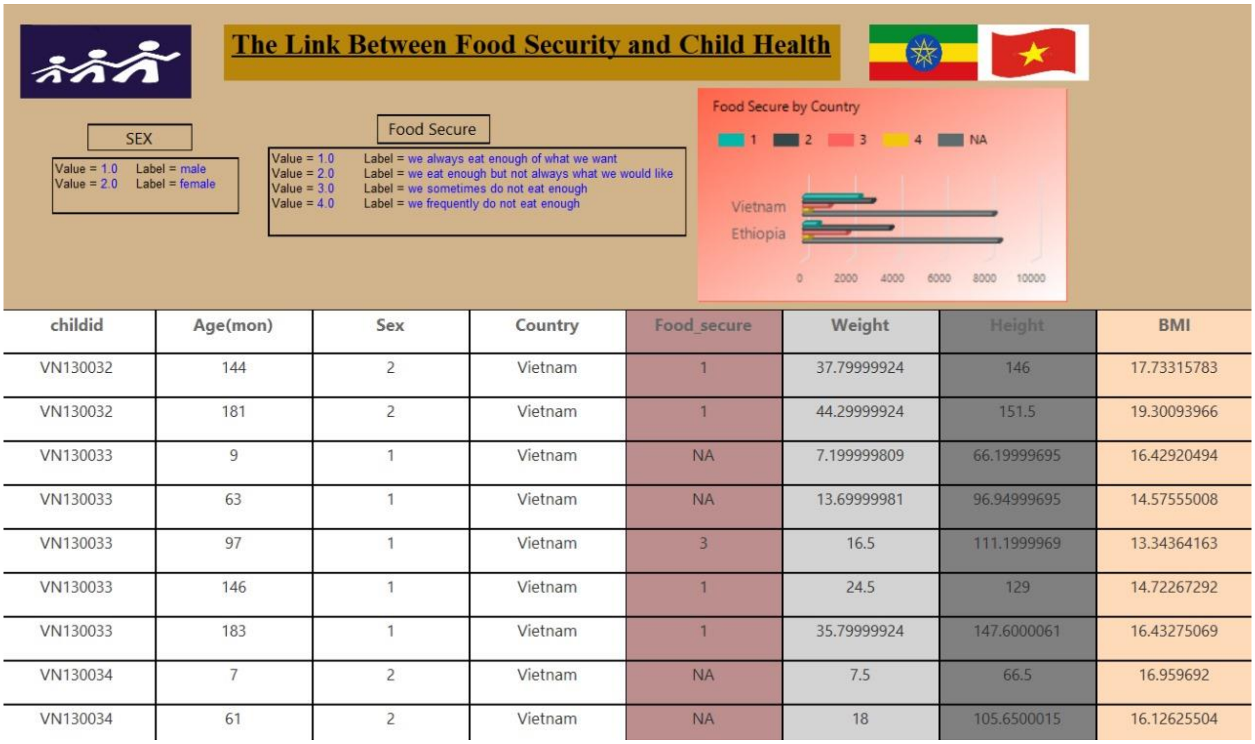
Value = 1.0 Label = male
Value = 2.0 Label = female

Food Secure

Value = 1.0 Label = we always eat enough of what we want
Value = 2.0 Label = we eat enough but not always what we would like
Value = 3.0 Label = we sometimes do not eat enough
Value = 4.0 Label = we frequently do not eat enough

childid	Age(mon)	Sex	Country	Food_secure	Weight	Height	BMI
[childid]	[agemon]	[chsex]	[Country]	[foodsec]	[chweight]	[chheight]	[bmi]

Final Report



The Link Between Food Security and Child Health

SEX

Value = 1.0 Label = male
Value = 2.0 Label = female

Food Secure

Value = 1.0 Label = we always eat enough of what we want
Value = 2.0 Label = we eat enough but not always what we would like
Value = 3.0 Label = we sometimes do not eat enough
Value = 4.0 Label = we frequently do not eat enough

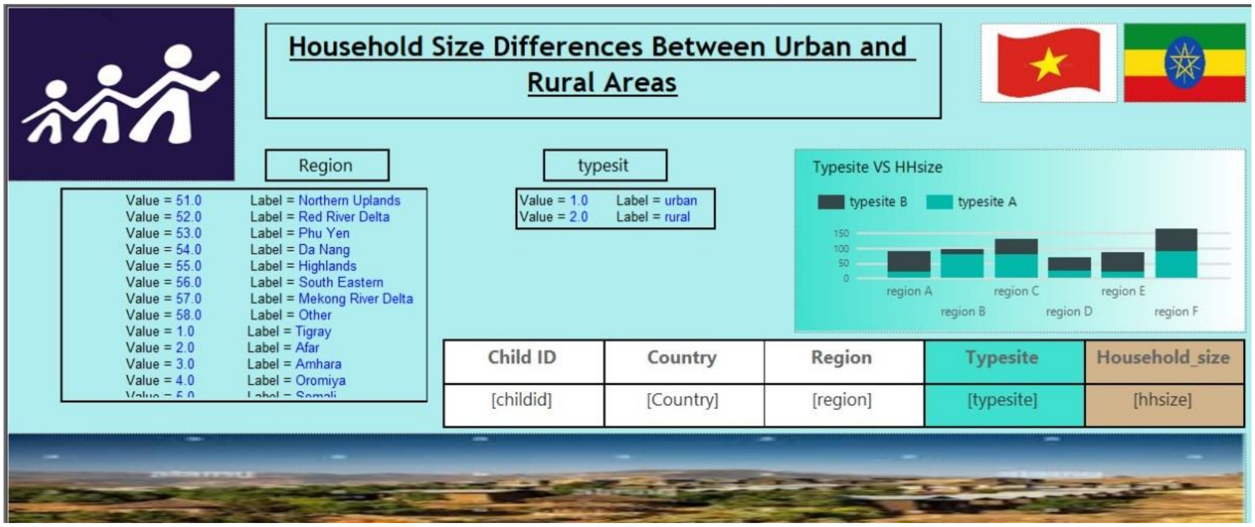
childid	Age(mon)	Sex	Country	Food_secure	Weight	Height	BMI
VN130032	144	2	Vietnam	1	37.79999924	146	17.73315783
VN130032	181	2	Vietnam	1	44.29999924	151.5	19.30093966
VN130033	9	1	Vietnam	NA	7.199999809	66.19999695	16.42920494
VN130033	63	1	Vietnam	NA	13.69999981	96.94999695	14.57555008
VN130033	97	1	Vietnam	3	16.5	111.1999969	13.34364163
VN130033	146	1	Vietnam	1	24.5	129	14.72267292
VN130033	183	1	Vietnam	1	35.79999924	147.6000061	16.43275069
VN130034	7	2	Vietnam	NA	7.5	66.5	16.959692
VN130034	61	2	Vietnam	NA	18	105.6500015	16.12625504

❖ Report 2: Household size differences between urban and rural areas

Ethiopian and Vietnamese urban-rural household size disparities are a reflection of a number of social, economic, and cultural issues. Households are typically larger in rural areas of both countries, frequently as a result of the traditional emphasis on agriculture. In many communities, having more children is viewed as advantageous for labor-intensive job, and more family members participate in farming activities. Additionally, increased birth rates in remote areas may result from less access to family planning and healthcare services.

On the other hand, household sizes are often lower in urban areas like Vietnam and Ethiopia. Reduced birth rates, a move toward nuclear families, and easier access to healthcare and education are all linked to urbanization. Economic considerations also come into play because larger families are discouraged by the higher cost of metropolitan life. Additionally, urban families frequently place a higher value on work and education than on agriculture, which results in different dynamics inside the home.

In conclusion, the disparity between rural areas' reliance on agriculture and urban areas' access to healthcare, education, and economic possibilities is what drives the variations in household sizes between Ethiopia and Vietnam.



Household Size Differences Between Urban and Rural Areas

Region

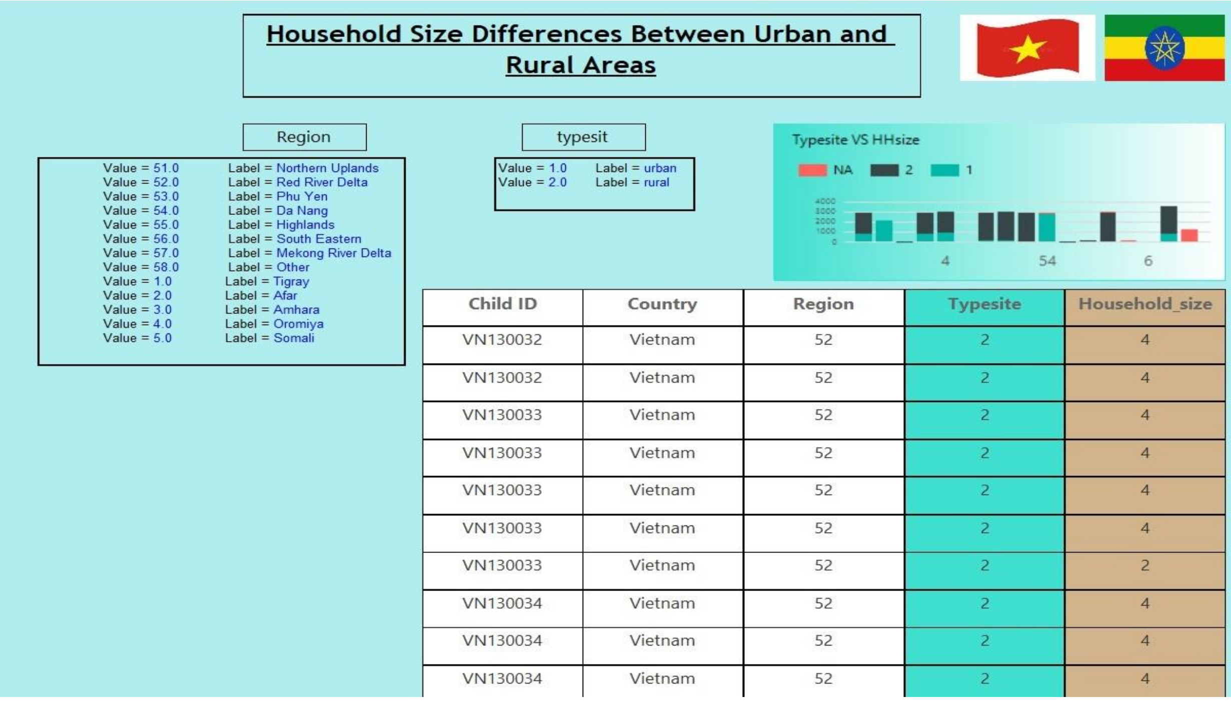
Value = 51.0	Label = Northern Uplands
Value = 52.0	Label = Red River Delta
Value = 53.0	Label = Phu Yen
Value = 54.0	Label = Da Nang
Value = 55.0	Label = Highlands
Value = 56.0	Label = South Eastern
Value = 57.0	Label = Mekong River Delta
Value = 58.0	Label = Other
Value = 1.0	Label = Tigray
Value = 2.0	Label = Afar
Value = 3.0	Label = Amhara
Value = 4.0	Label = Oromiya
Value = 5.0	Label = Somali

typesit

Value = 1.0	Label = urban
Value = 2.0	Label = rural

Child ID	Country	Region	Typesite	Household_size
[childid]	[Country]	[region]	[typesite]	[hhsize]

Final Report



Household Size Differences Between Urban and Rural Areas

Region

Value = 51.0	Label = Northern Uplands
Value = 52.0	Label = Red River Delta
Value = 53.0	Label = Phu Yen
Value = 54.0	Label = Da Nang
Value = 55.0	Label = Highlands
Value = 56.0	Label = South Eastern
Value = 57.0	Label = Mekong River Delta
Value = 58.0	Label = Other
Value = 1.0	Label = Tigray
Value = 2.0	Label = Afar
Value = 3.0	Label = Amhara
Value = 4.0	Label = Oromiya
Value = 5.0	Label = Somali

typesit

Value = 1.0	Label = urban
Value = 2.0	Label = rural

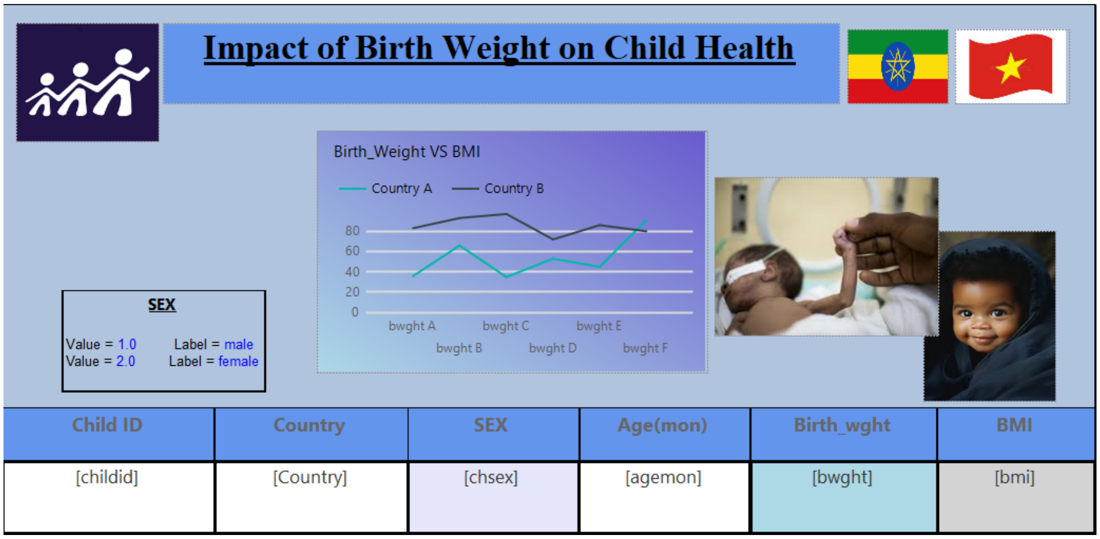
Child ID	Country	Region	Typesite	Household_size
VN130032	Vietnam	52	2	4
VN130032	Vietnam	52	2	4
VN130033	Vietnam	52	2	4
VN130033	Vietnam	52	2	4
VN130033	Vietnam	52	2	4
VN130033	Vietnam	52	2	4
VN130033	Vietnam	52	2	2
VN130034	Vietnam	52	2	4
VN130034	Vietnam	52	2	4
VN130034	Vietnam	52	2	4

❖ Report 3: Impact of birth weight on child health

Given that low birth weight is a key predictor of future health issues, birth weight has a substantial impact on children's health in Ethiopia and Vietnam. Low birth weight babies are more prone to suffer from hunger, poor growth, and developmental problems in both nations. Poor maternal nutrition, restricted access to healthcare, and socioeconomic challenges that impact prenatal care are some of the contributing causes.

Low birth weight is frequently associated with high infant death rates in Ethiopia as well as long-term health problems like delayed cognitive development and infection susceptibility. Similar to this, low birth weight babies in Vietnam might have trouble developing to their full potential both intellectually and physically, which could affect their academic performance and future productivity.

Improving maternal nutrition and healthcare is essential to resolving these problems in both nations.



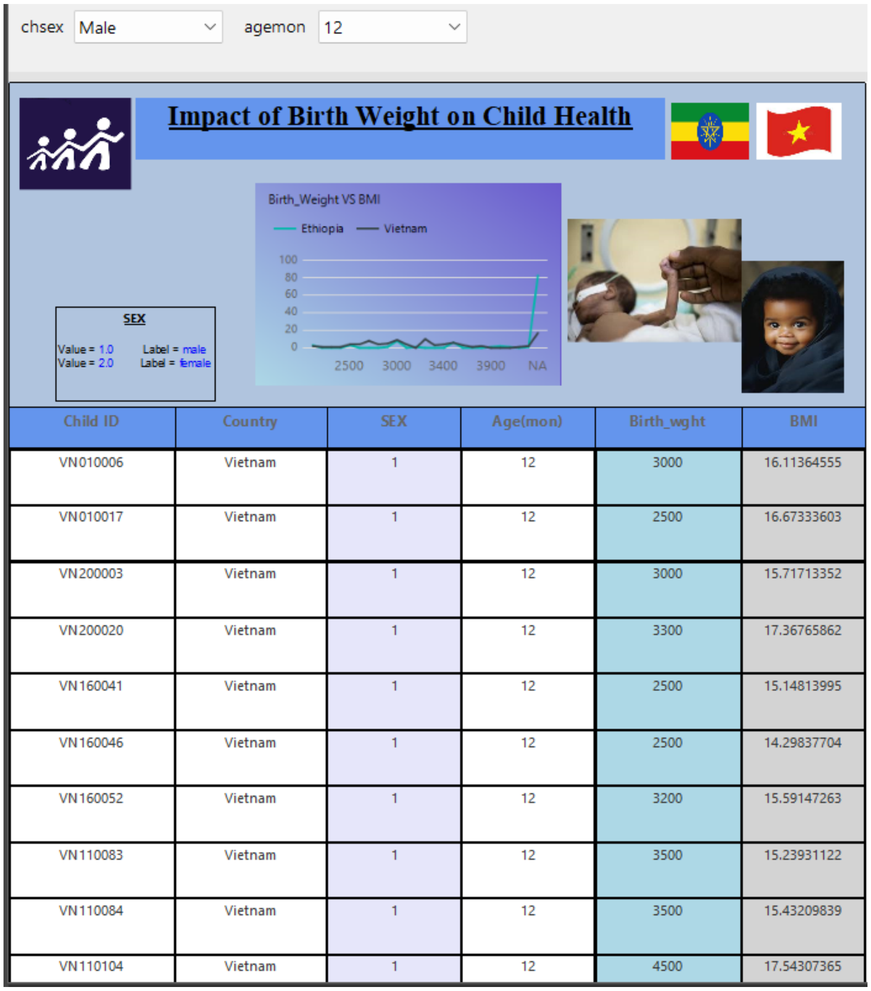
Impact of Birth Weight on Child Health

SEX

Value = 1.0 Label = male
Value = 2.0 Label = female

Child ID	Country	SEX	Age(mon)	Birth_wght	BMI
[childid]	[Country]	[chsex]	[agemon]	[bwght]	[bmi]

Final Report



chsex Male

agemon 12

Impact of Birth Weight on Child Health

SEX

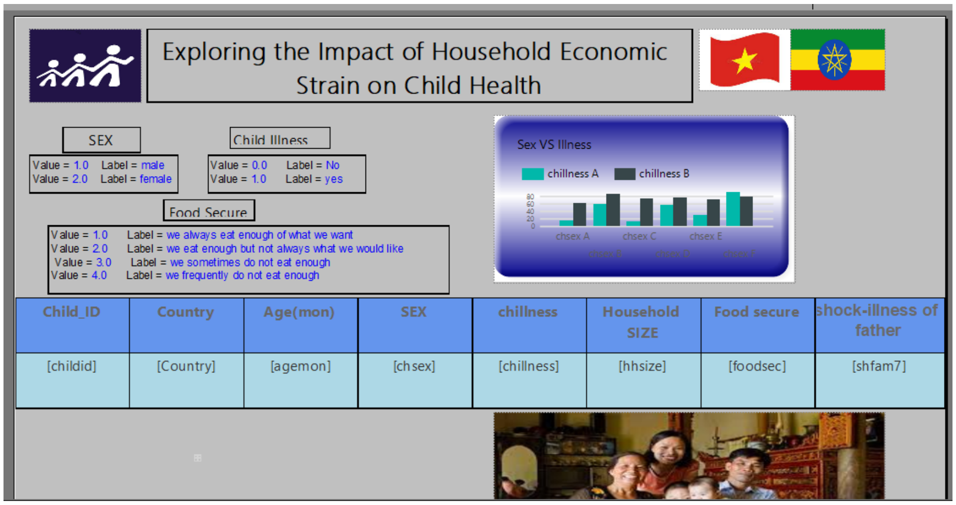
Value = 1.0 Label = male
Value = 2.0 Label = female

Child ID	Country	SEX	Age(mon)	Birth_wght	BMI
VN010006	Vietnam	1	12	3000	16.11364555
VN010017	Vietnam	1	12	2500	16.67333603
VN200003	Vietnam	1	12	3000	15.71713352
VN200020	Vietnam	1	12	3300	17.36765862
VN160041	Vietnam	1	12	2500	15.14813995
VN160046	Vietnam	1	12	2500	14.29837704
VN160052	Vietnam	1	12	3200	15.59147263
VN110083	Vietnam	1	12	3500	15.23931122
VN110084	Vietnam	1	12	3500	15.43209839
VN110104	Vietnam	1	12	4500	17.54307365

❖ Report 4: Exploring the impact of household economic strain on child health

In Ethiopia and Vietnam, household economic stress has a major impact on children's health. Children who live in families who are struggling financially are more likely to suffer from malnutrition and delayed growth. Children's immune systems can be weakened by inadequate nutrition, leaving them more vulnerable to illnesses and infections. Access to healthcare services is also limited by a lack of financial means. Families may postpone or forego medical treatment in rural areas, where financial difficulty is more common, which could impact health outcomes.

Furthermore, unstable living circumstances and a lack of support can have a detrimental impact on a child's mental health by producing stress and anxiety.



Exploring the Impact of Household Economic Strain on Child Health

SEX

Value = 1.0 Label = male
Value = 2.0 Label = female

Child Illness

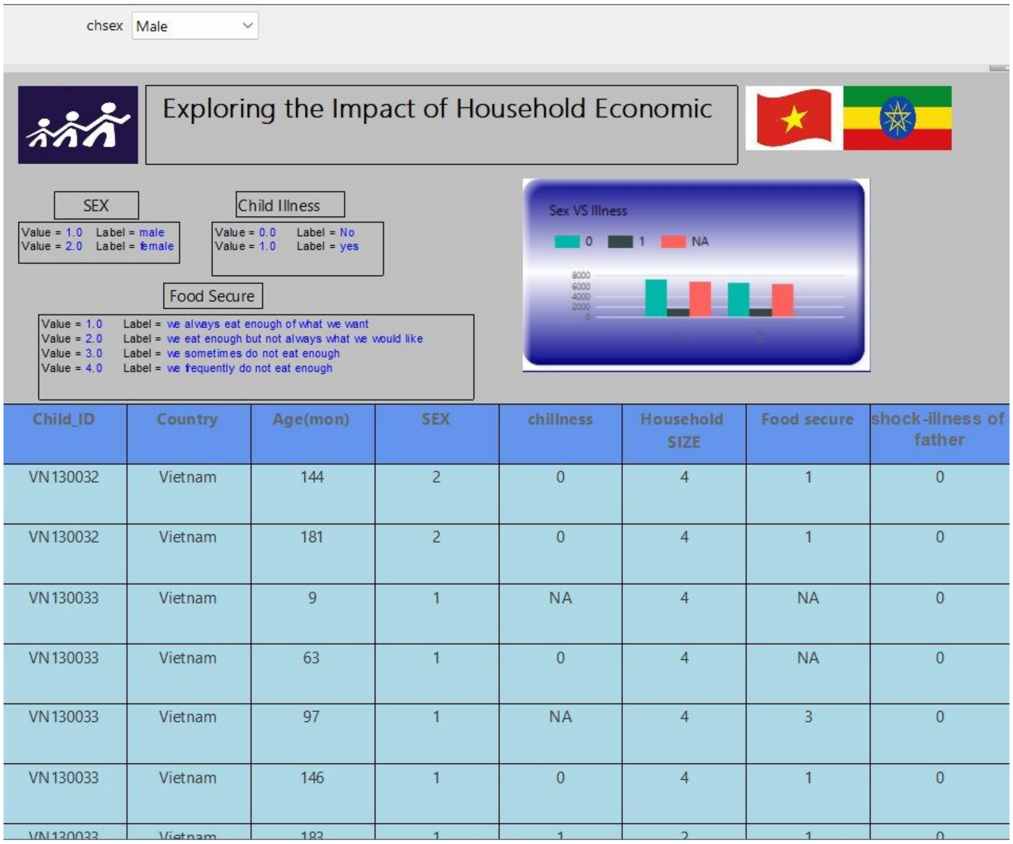
Value = 0.0 Label = No
Value = 1.0 Label = yes

Food Secure

Value = 1.0 Label = we always eat enough of what we want
Value = 2.0 Label = we eat enough but not always what we would like
Value = 3.0 Label = we sometimes do not eat enough
Value = 4.0 Label = we frequently do not eat enough

Child_ID	Country	Age(mon)	SEX	chillness	Household SIZE	Food secure	shock-illness of father
[childid]	[Country]	[agemon]	[chsex]	[chillness]	[hhsize]	[foodsec]	[shfam7]

Final Report



chsex Male

Exploring the Impact of Household Economic

SEX

Value = 1.0 Label = male
Value = 2.0 Label = female

Child Illness

Value = 0.0 Label = No
Value = 1.0 Label = yes

Food Secure

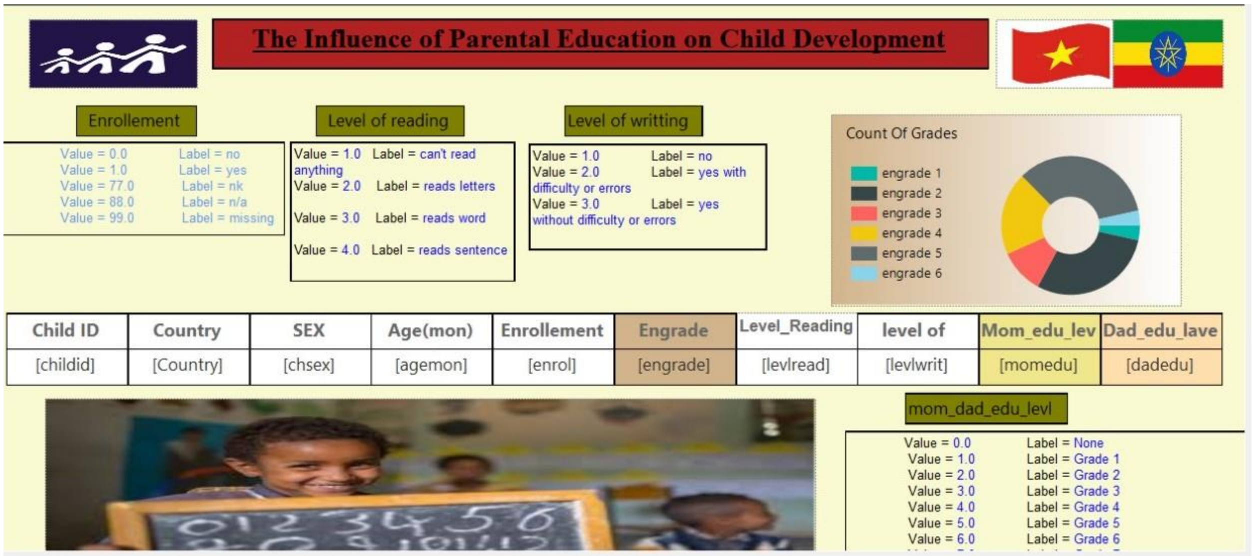
Value = 1.0 Label = we always eat enough of what we want
Value = 2.0 Label = we eat enough but not always what we would like
Value = 3.0 Label = we sometimes do not eat enough
Value = 4.0 Label = we fequently do not eat enough

Child_ID	Country	Age(mon)	SEX	chillness	Household SIZE	Food secure	shock-illness of father
VN130032	Vietnam	144	2	0	4	1	0
VN130032	Vietnam	181	2	0	4	1	0
VN130033	Vietnam	9	1	NA	4	NA	0
VN130033	Vietnam	63	1	0	4	NA	0
VN130033	Vietnam	97	1	NA	4	3	0
VN130033	Vietnam	146	1	0	4	1	0
VN130033	Vietnam	183	1	1	2	1	0

❖ Report 5: The Influence of Parental Education on Child Development

In Vietnam and Ethiopia, parental education has a big impact on how children develop. Higher educated parents in Ethiopia typically support their children's learning more effectively at home and through their knowledge of the educational system. Parents with higher levels of education are more likely to read to their kids or encourage school attendance, two activities that foster cognitive growth and improve academic performance.

Parental education has a similar impact on children's development in Vietnam. Formally educated parents frequently place a high value on their kids' education, promoting self-discipline and academic achievement. Additionally, their children's overall physical and mental development is aided by their increased access to resources and health and nutrition information. By encouraging healthier, more nurturing surroundings, raising parental education can greatly enhance children's prospects for the future in both nations.



The Influence of Parental Education on Child Development

Enrollement

Value = 0.0 Label = no
Value = 1.0 Label = yes
Value = 77.0 Label = nk
Value = 88.0 Label = n/a
Value = 99.0 Label = missing

Level of reading

Value = 1.0 Label = can't read anything
Value = 2.0 Label = reads letters
Value = 3.0 Label = reads word
Value = 4.0 Label = reads sentence

Level of writting

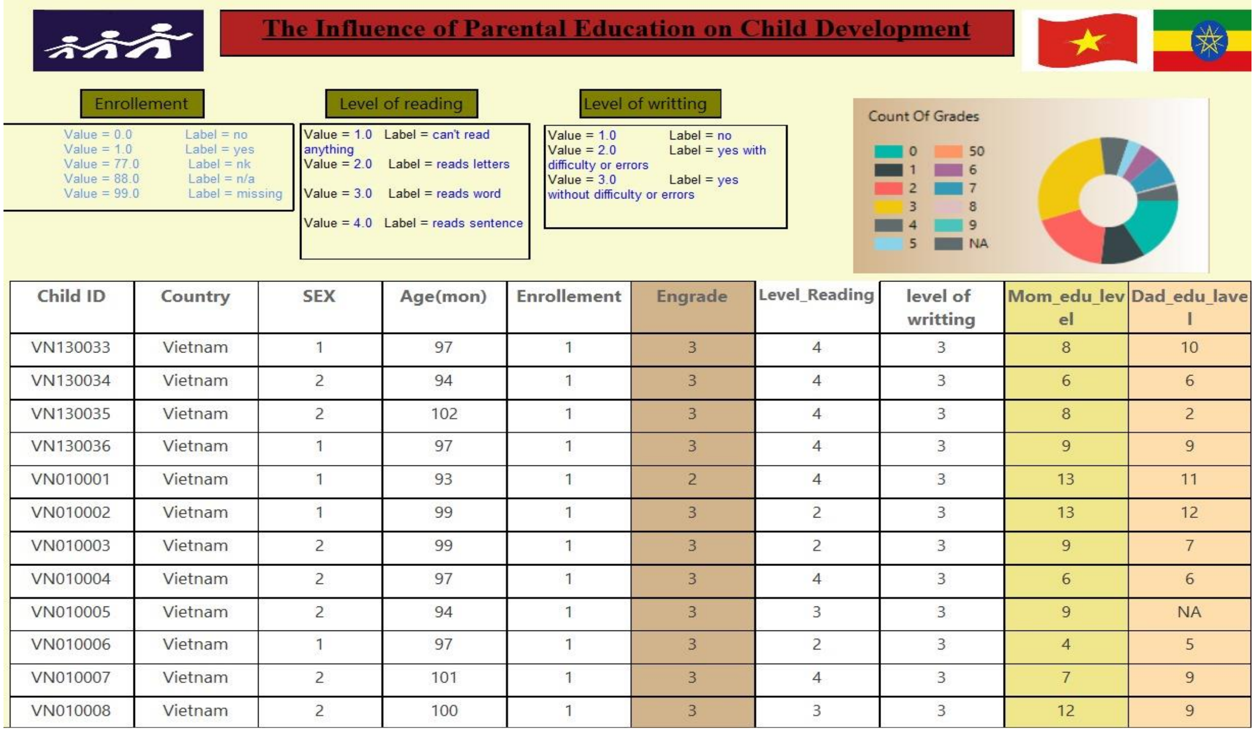
Value = 1.0 Label = no
Value = 2.0 Label = yes with difficulty or errors
Value = 3.0 Label = yes without difficulty or errors

Child ID	Country	SEX	Age(mon)	Enrollement	Engrade	Level_Reading	level of	Mom_edu_lev	Dad_edu_lave
[childid]	[Country]	[chsex]	[agemon]	[enrol]	[engrade]	[levlread]	[levlwrit]	[momedu]	[dadedu]

mom_dad_edu_levl

Value = 0.0 Label = None
Value = 1.0 Label = Grade 1
Value = 2.0 Label = Grade 2
Value = 3.0 Label = Grade 3
Value = 4.0 Label = Grade 4
Value = 5.0 Label = Grade 5
Value = 6.0 Label = Grade 6

Final Report



The Influence of Parental Education on Child Development

Enrollement

Value = 0.0 Label = no
Value = 1.0 Label = yes
Value = 77.0 Label = nk
Value = 88.0 Label = n/a
Value = 99.0 Label = missing

Level of reading

Value = 1.0 Label = can't read anything
Value = 2.0 Label = reads letters
Value = 3.0 Label = reads word
Value = 4.0 Label = reads sentence

Level of writting

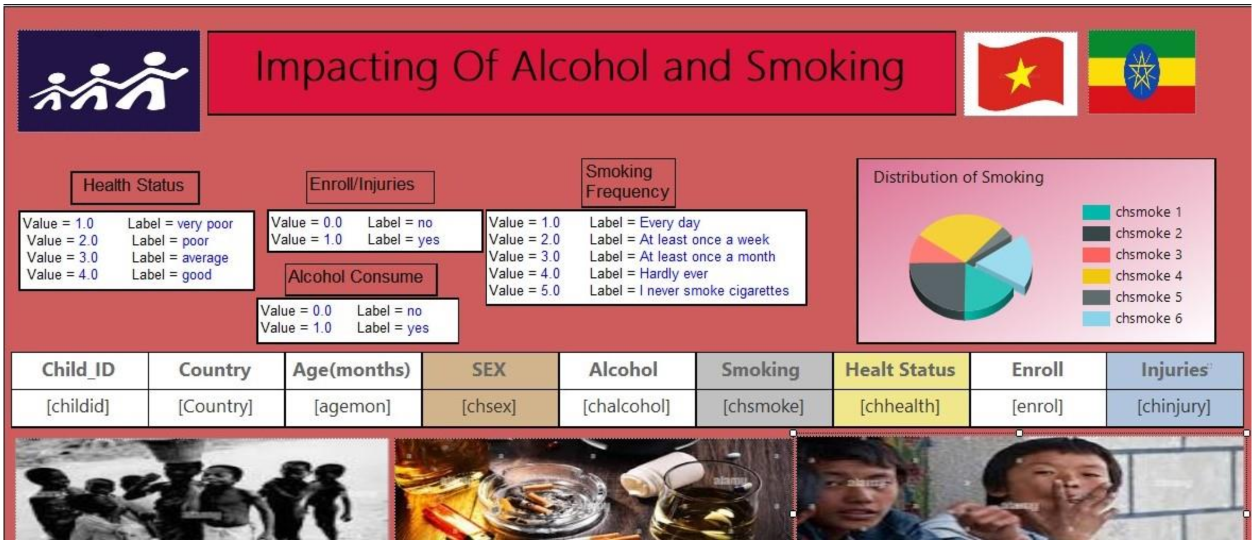
Value = 1.0 Label = no
Value = 2.0 Label = yes with difficulty or errors
Value = 3.0 Label = yes without difficulty or errors

Child ID	Country	SEX	Age(mon)	Enrollement	Engrade	Level_Reading	level of writting	Mom_edu_level	Dad_edu_lavel
VN130033	Vietnam	1	97	1	3	4	3	8	10
VN130034	Vietnam	2	94	1	3	4	3	6	6
VN130035	Vietnam	2	102	1	3	4	3	8	2
VN130036	Vietnam	1	97	1	3	4	3	9	9
VN010001	Vietnam	1	93	1	2	4	3	13	11
VN010002	Vietnam	1	99	1	3	2	3	13	12
VN010003	Vietnam	2	99	1	3	2	3	9	7
VN010004	Vietnam	2	97	1	3	4	3	6	6
VN010005	Vietnam	2	94	1	3	3	3	9	NA
VN010006	Vietnam	1	97	1	3	2	3	4	5
VN010007	Vietnam	2	101	1	3	4	3	7	9
VN010008	Vietnam	2	100	1	3	3	3	12	9

❖ Report 6: Impacting of Alcohol and Smoking

In Ethiopia and Vietnam, alcohol and tobacco use have a substantial negative influence on children's health, especially when poverty is present. In both nations, caregivers' smoking and alcohol use can have a detrimental impact on home income by taking funds away from necessities like healthcare, education, and nutrition. This can impede development and worsen poverty among children.

High alcohol use in Ethiopia is associated with socioeconomic stress, which deteriorates parental health and reduces child care. Similar to this, smoking is common in Vietnam, particularly among men. This has a direct effect on the welfare of children and their prospects for the future since it can result in health issues and decreased financial stability.



Impacting Of Alcohol and Smoking

Health Status

Value = 1.0	Label = very poor
Value = 2.0	Label = poor
Value = 3.0	Label = average
Value = 4.0	Label = good

Enroll/Injuries

Value = 0.0	Label = no
Value = 1.0	Label = yes

Alcohol Consume

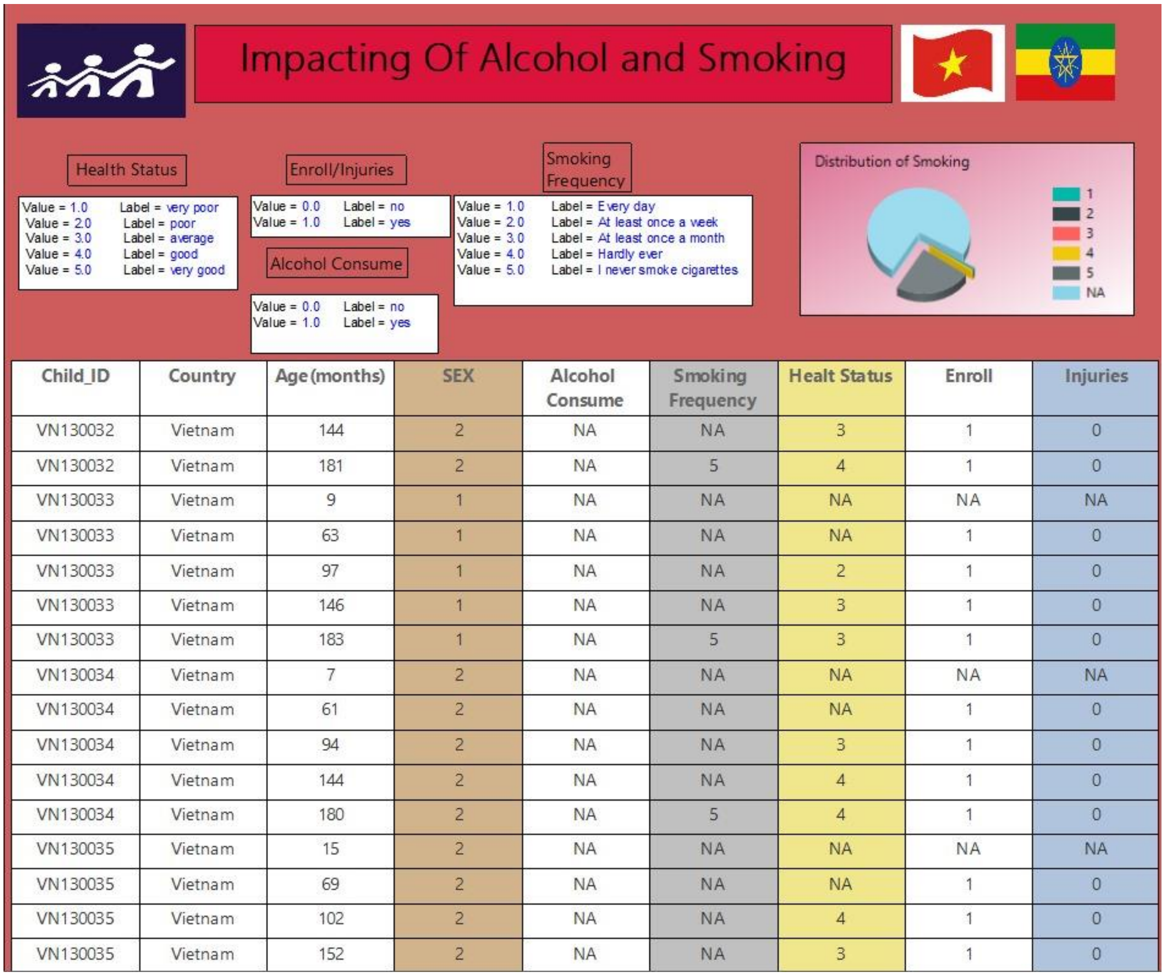
Value = 0.0	Label = no
Value = 1.0	Label = yes

Smoking Frequency

Value = 1.0	Label = Every day
Value = 2.0	Label = At least once a week
Value = 3.0	Label = At least once a month
Value = 4.0	Label = Hardly ever
Value = 5.0	Label = I never smoke cigarettes

Child_ID	Country	Age(months)	SEX	Alcohol	Smoking	Healt Status	Enroll	Injuries
[childid]	[Country]	[agemon]	[chsex]	[chalcohol]	[chsmoke]	[chhealth]	[enrol]	[chinjury]

Final report



Impacting Of Alcohol and Smoking

Health Status

Value = 1.0	Label = very poor
Value = 2.0	Label = poor
Value = 3.0	Label = average
Value = 4.0	Label = good
Value = 5.0	Label = very good

Enroll/Injuries

Value = 0.0	Label = no
Value = 1.0	Label = yes

Alcohol Consume

Value = 0.0	Label = no
Value = 1.0	Label = yes

Smoking Frequency

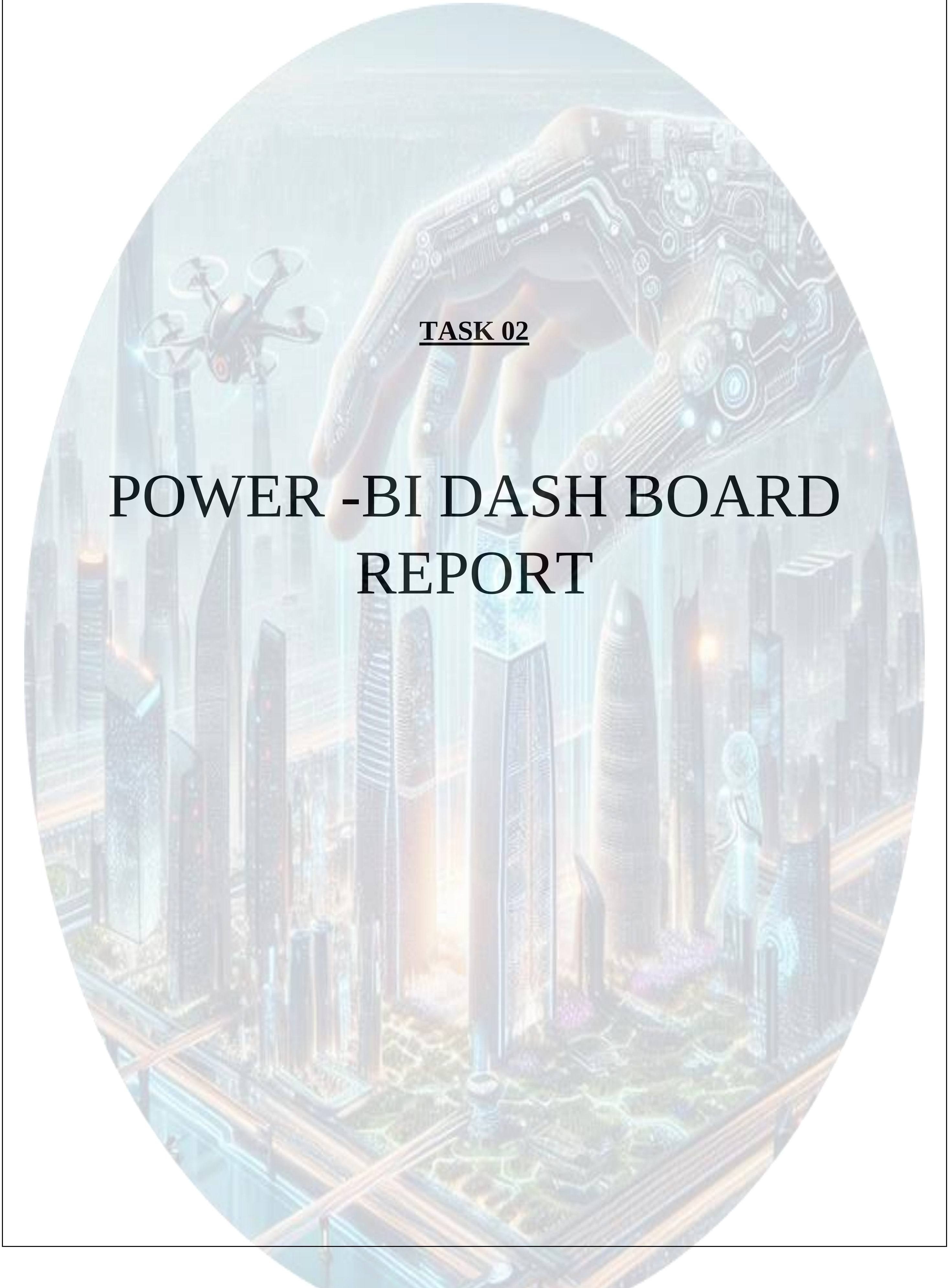
Value = 1.0	Label = Every day
Value = 2.0	Label = At least once a week
Value = 3.0	Label = At least once a month
Value = 4.0	Label = Hardly ever
Value = 5.0	Label = I never smoke cigarettes

Child_ID	Country	Age(months)	SEX	Alcohol Consume	Smoking Frequency	Healt Status	Enroll	Injuries
VN130032	Vietnam	144	2	NA	NA	3	1	0
VN130032	Vietnam	181	2	NA	5	4	1	0
VN130033	Vietnam	9	1	NA	NA	NA	NA	NA
VN130033	Vietnam	63	1	NA	NA	NA	1	0
VN130033	Vietnam	97	1	NA	NA	2	1	0
VN130033	Vietnam	146	1	NA	NA	3	1	0
VN130033	Vietnam	183	1	NA	5	3	1	0
VN130034	Vietnam	7	2	NA	NA	NA	NA	NA
VN130034	Vietnam	61	2	NA	NA	NA	1	0
VN130034	Vietnam	94	2	NA	NA	3	1	0
VN130034	Vietnam	144	2	NA	NA	4	1	0
VN130034	Vietnam	180	2	NA	5	4	1	0
VN130035	Vietnam	15	2	NA	NA	NA	NA	NA
VN130035	Vietnam	69	2	NA	NA	NA	1	0
VN130035	Vietnam	102	2	NA	NA	4	1	0
VN130035	Vietnam	152	2	NA	NA	3	1	0

1.6 Conclusion

To sum up, this project examined a number of important variables influencing children's health and development in diverse settings, with an emphasis on Ethiopia and Vietnam. The well-being of children is impacted by the inequalities in household sizes between urban and rural settings, which reflect discrepancies in resource allocation and access to essential services. Low birth weight frequently results in long-term health issues, making it an important predictor of a child's health. Household financial distress exacerbates these problems by restricting access to nutrition and medical treatment. Food security is important since a child's physical and mental development are directly impacted by poor nutrition. Since educated parents are better able to support their children's development and education, parental education has also emerged as a critical factor. Last but not least, the negative impacts of smoking and drinking, especially in low-income households, worsen health and financial issues and further impair children's general wellbeing.

Targeted strategies that take into account the social and economic determinants of health are necessary to address these interrelated problems. Significant progress can be made in improving child health outcomes and general development in these areas by concentrating on strengthening food security, healthcare access, and education while lowering harmful habits like smoking and alcohol use.



TASK 02

POWER -BI DASH BOARD REPORT

CONTENT

01.Introduction

02.Database Creation

03. Exploratory Data Analysis

04. Power BI Dashboard Design

05. Findings and Recommendations

06.Conclusion

TASK 02

02. ANALYSIS AND VISUALIZATION OF EPC DATA IN MANCHESTER (2013-2023) USING POWER BI

1 Introduction

The Greater Manchester Domestic Energy Performance Analyzer is a large-scale project that evaluates and visualizes the energy efficiency of buildings across all of Manchester's local authorities using Energy Performance Certificate (EPC) data from 2013 to 2023. This assignment highlights the significance of data-driven insights in understanding energy use patterns and improving local sustainability activities. Using Power BI, the goal is to transform unstructured data into intelligent, interactive visualizations that identify trends in energy efficiency, identify areas for improvement, and provide useful guidance to individuals in the energy sector. This study will serve as a roadmap for future energy efficiency projects, reflecting historical progress in energy performance and assisting lawmakers, real estate developers, and homeowners in making educated decisions. Manchester's energy sustainability will rise as a result of the interactive dashboard, which will make it easier for users to explore the data and discover trends in energy performance across various property types and neighborhoods.

1.2 Exploring the dataset

Downloading the dataset

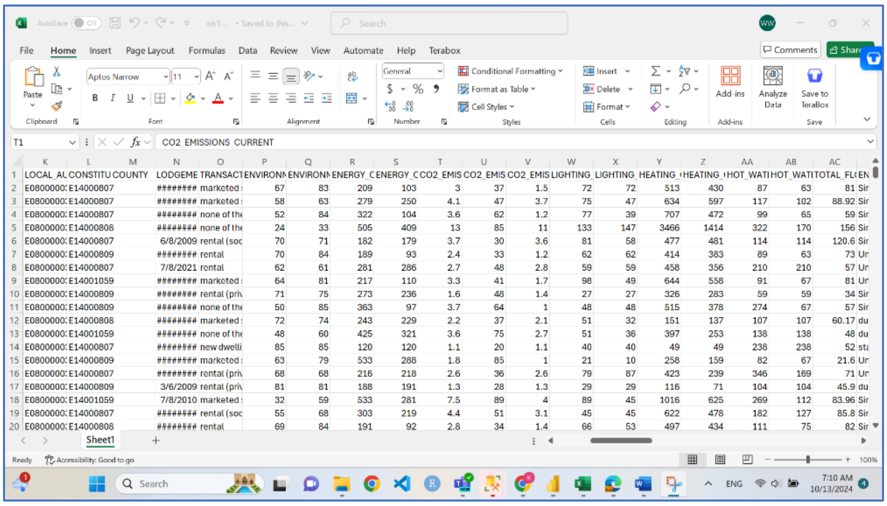
Using the provided URL, the dataset—which contains thorough information on residential Energy Performance Certificates (EPCs)was downloaded. It is important to review the dataset before beginning any analysis, making sure you understand the primary variables and have a good understanding of how it is organized. This exploratory study helps with trend identification, data quality assurance, and preparing the information for further research.

The dataset can be found at the following location:

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

Exploring the dataset

The EPC dataset for England and Wales provides comprehensive information on energy efficiency in residential and commercial premises. This information includes property type, insulation, heating type, emissions rating, and energy expenses. Lawmakers, tenants, homeowners, and landlords utilize this publicly available dataset to monitor their progress toward carbon reduction goals and to make well-informed decisions regarding energy efficiency. With 300,889 rows and 92 columns of property data, it provides an enormous collection that may be used to examine policy effects, energy performance trends, and opportunities for environmental improvements.



T1 — CO2_EMISSIONS_CURRENT

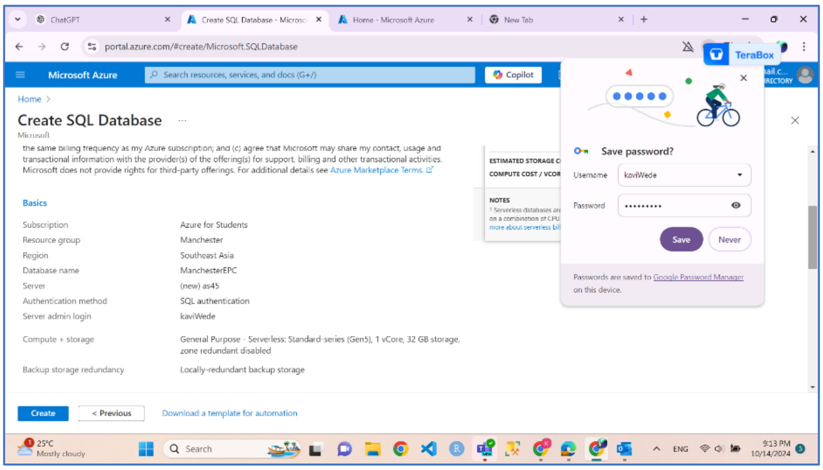
	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	
1	LOCAL_AU	CONSTITU	COUNTY	LODGEME	TRANSACT	ENVIRONM	ENVIRONM	ENERGY_C	ENERGY_C	CO2_EMIS	CO2_EMIS	CO2_EMIS	LIGHTING_	LIGHTING_	HEATING_	HEATING_	HOT_WATI	HOT_WATI	TOTAL_FL	EN
2	E0800000:	E14000807		########	marketed :	67	83	209	103	3	37	1.5	72	72	513	430	87	63	81	Sir
3	E0800000:	E14000807		########	marketed :	58	63	279	250	4.1	47	3.7	75	47	634	597	117	102	88.92	Sir
4	E0800000:	E14000807		########	none of the	52	84	322	104	3.6	62	1.2	77	39	707	472	99	65	59	Sir
5	E0800000:	E14000808		########	none of the	24	33	505	409	13	85	11	133	147	3466	1414	322	170	156	Sir
6	E0800000:	E14000807		6/8/2009	rental (soc	70	71	182	179	3.7	30	3.6	81	58	477	481	114	114	120.6	Sir
7	E0800000:	E14000809		########	rental	70	84	189	93	2.4	33	1.2	62	62	414	383	89	63	73	Ur
8	E0800000:	E14000807		7/8/2021	rental	62	61	281	286	2.7	48	2.8	59	59	458	356	210	210	57	Ur
9	E0800000:	E14001059		########	marketed :	64	81	217	110	3.3	41	1.7	98	49	644	558	91	67	81	Ur
10	E0800000:	E14000809		########	rental (priv	71	75	273	236	1.6	48	1.4	27	27	326	283	59	59	34	Sir
11	E0800000:	E14000809		########	none of the	50	85	363	97	3.7	64	1	48	48	515	378	274	67	57	Sir
12	E0800000:	E14000808		########	marketed :	72	74	243	229	2.2	37	2.1	51	32	151	137	107	107	60.17	du
13	E0800000:	E14001059		########	none of the	48	60	425	321	3.6	75	2.7	51	36	397	253	138	138	48	du
14	E0800000:	E14000807		########	new dwelli	85	85	120	120	1.1	20	1.1	40	40	49	49	238	238	52	sta
15	E0800000:	E14000809		########	marketed :	63	79	533	288	1.8	85	1	21	10	258	159	82	67	21.6	Ur
16	E0800000:	E14000807		########	rental (priv	68	68	216	218	2.6	36	2.6	79	87	423	239	346	169	71	Ur
17	E0800000:	E14000809		3/6/2009	rental (priv	81	81	188	191	1.3	28	1.3	29	29	116	71	104	104	45.9	du
18	E0800000:	E14001059		7/8/2010	marketed :	32	59	533	281	7.5	89	4	89	45	1016	625	269	112	83.96	Sir
19	E0800000:	E14000807		########	rental (soc	55	68	303	219	4.4	51	3.1	45	45	622	478	182	127	85.8	Sir
20	E0800000:	E14000808		########	rental	69	84	191	92	2.8	34	1.4	66	53	497	434	111	75	82	Sir

2. Database Creation

2.1 Azure cloud database

Azure MySQL Database Creation: We created a MySQL database instance in Azure to store the EPC data. This ensured reliable data storage with cloud-based scalability and security. The steps followed were:

- Azure Service Used: Azure Database for MySQL
- Database Name:ManchesterEPC
- Resource Group: Manchester
- MySQL Server Name: epc-mysql-server



portal.azure.com/#create/Microsoft.SQLDatabase

Create SQL Database

Microsoft

the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. For additional details see Azure Marketplace Terms.

Basics

Subscription	Azure for Students
Resource group	Manchester
Region	Southeast Asia
Database name	ManchesterEPC
Server	(new) as45
Authentication method	SQL authentication
Server admin login	kaviWede
Compute + storage	General Purpose - Serverless: Standard-series (Gen5), 1 vCore, 32 GB storage, zone redundant disabled
Backup storage redundancy	Locally-redundant backup storage

Save password?

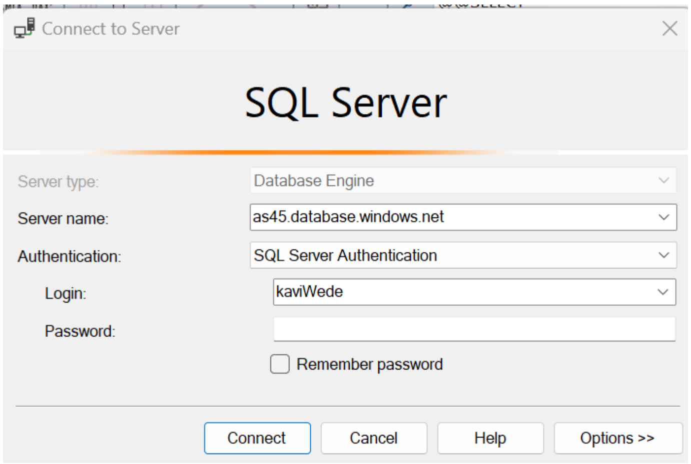
Username: kaviWede

Password: •••••••••

Passwords are saved to Google Password Manager on this device.

Connection to Microsoft SQL Server

- **SQL Database Connection**: The EPC data was imported into the Azure SQL Database using SQL Server Management Studio (SSMS) and Azure Data Factory. Once the database was set up, we established a secure connection to the SQL instance for querying and data management

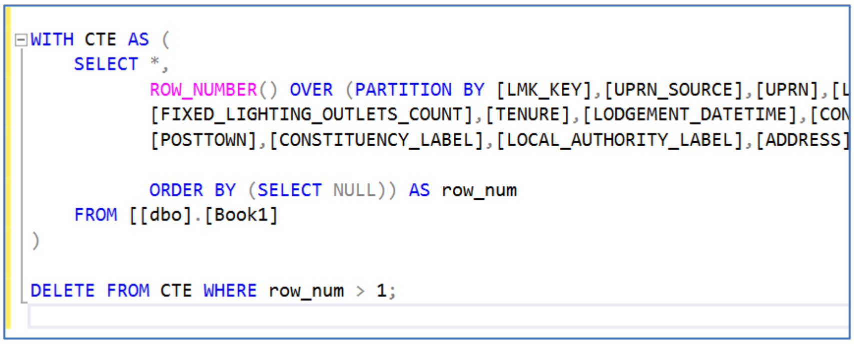


Server type: Database Engine
Server name: as45.database.windows.net
Authentication: SQL Server Authentication
Login: kaviWede
Password:
☐ Remember password

2.2Data Cleaning and Transformation Using SQL

2.2.1 Removing Duplicates

Duplicate records can result in skewed or incorrect analyses, especially when analyzing trends in energy performance. We used SQL's ROW_NUMBER() function in conjunction with a **Common Table Expression (CTE)** to identify and remove duplicates based on multiple columns such as LMK_KEY, UPRN, Energy Consumption, CO2 Emissions, and other related fields.



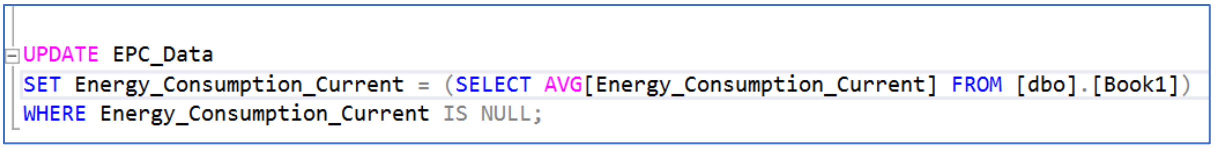
```
WITH CTE AS (
    SELECT *,
           ROW_NUMBER() OVER (PARTITION BY [LMK_KEY],[UPRN_SOURCE],[UPRN],[L
           [FIXED_LIGHTING_OUTLETS_COUNT],[TENURE],[LODGEMENT_DATETIME],[CON
           [POSTTOWN],[CONSTITUENCY_LABEL],[LOCAL_AUTHORITY_LABEL],[ADDRESS]

           ORDER BY (SELECT NULL)) AS row_num
    FROM [[dbo].[Book1]
)

DELETE FROM CTE WHERE row_num > 1;
```

2.2.2 Handling Missing Values

Missing values in key columns like Energy_Consumption_Current, CO2_Emissions_Current, and Current_Energy_Rating were handled by replacing them with calculated averages values.

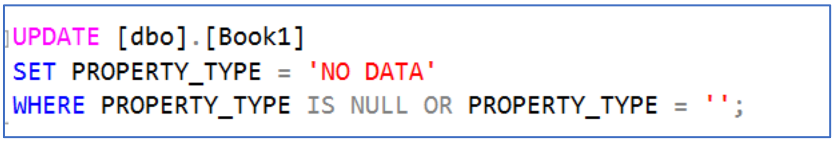
Replace Null Values:

we replaced the null values with the average values and blank property type fields were set to "No DATA" to avoid null values in our analysis::



```
UPDATE EPC_Data
SET Energy_Consumption_Current = (SELECT AVG[Energy_Consumption_Current] FROM [dbo].[Book1])
WHERE Energy_Consumption_Current IS NULL;
```

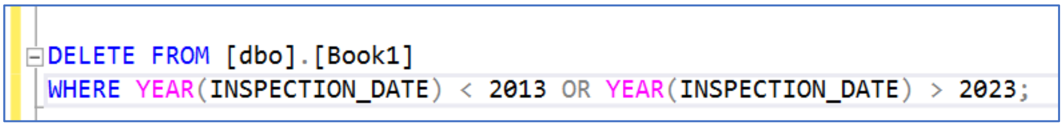
Replace blank column with No Data



```
UPDATE [dbo].[Book1]
SET PROPERTY_TYPE = 'NO DATA'
WHERE PROPERTY_TYPE IS NULL OR PROPERTY_TYPE = '';
```

2.2.3 Filtering Data Between 2013 and 2023

To focus on recent energy performance data, we filtered the dataset to include only records between the years 2013 and 2023:

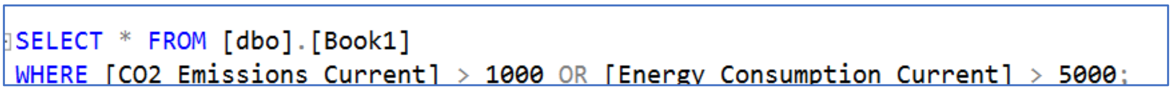


```
DELETE FROM [dbo].[Book1]
WHERE YEAR(INSPECTION_DATE) < 2013 OR YEAR(INSPECTION_DATE) > 2023;
```

2.2.4 Handling Outliers

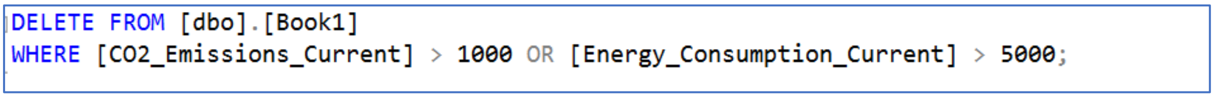
Outliers in CO2_Emissions_Current and Energy_Consumption_Current were identified using SQL aggregate functions and manually removed based on threshold values:

Identify Outliers:



```
SELECT * FROM [dbo].[Book1]
WHERE [CO2_Emissions_Current] > 1000 OR [Energy_Consumption_Current] > 5000;
```

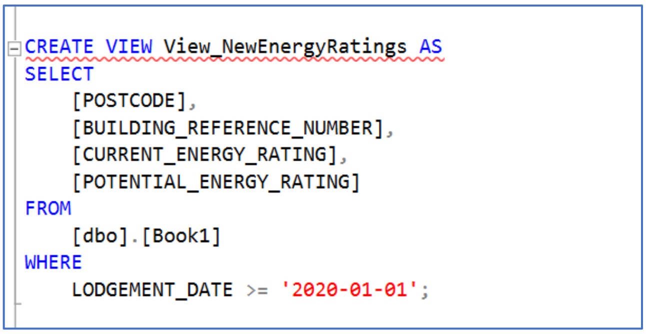
Remove Outliers:



```
DELETE FROM [dbo].[Book1]
WHERE [CO2_Emissions_Current] > 1000 OR [Energy_Consumption_Current] > 5000;
```

2.2.5 Creating Views

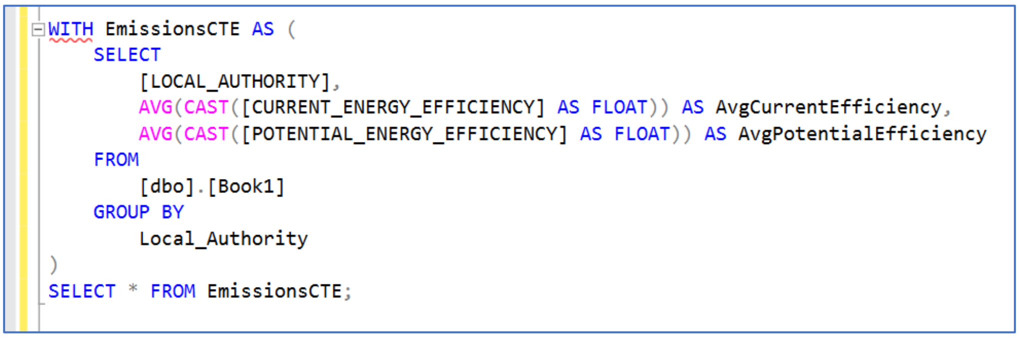
Create views for different aspects of the energy data, which could help simplify access to specific segments of the data, such as properties grouped by energy ratings or focusing on potential improvements.



```
CREATE VIEW View_NewEnergyRatings AS
SELECT
    [POSTCODE],
    [BUILDING_REFERENCE_NUMBER],
    [CURRENT_ENERGY_RATING],
    [POTENTIAL_ENERGY_RATING]
FROM
    [dbo].[Book1]
WHERE
    LODGEMENT_DATE >= '2020-01-01';
```

2.2.5. Using Common Table Expressions (CTEs)

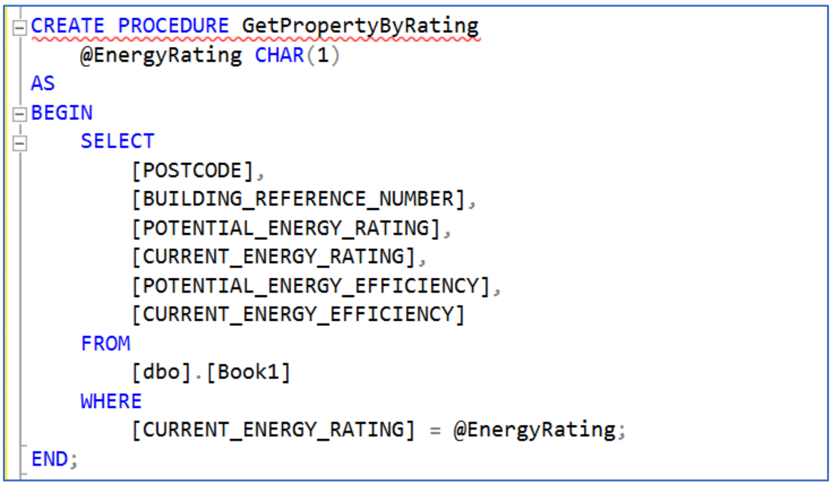
We use CTE for perform complex queries in a more readable manner, such as calculating average current and potential energy efficiencies.



```
WITH EmissionsCTE AS (
    SELECT
        [LOCAL_AUTHORITY],
        AVG(CAST([CURRENT_ENERGY_EFFICIENCY] AS FLOAT)) AS AvgCurrentEfficiency,
        AVG(CAST([POTENTIAL_ENERGY_EFFICIENCY] AS FLOAT)) AS AvgPotentialEfficiency
    FROM
        [dbo].[Book1]
    GROUP BY
        Local_Authority
)
SELECT * FROM EmissionsCTE;
```

2.2.6. Stored Procedures

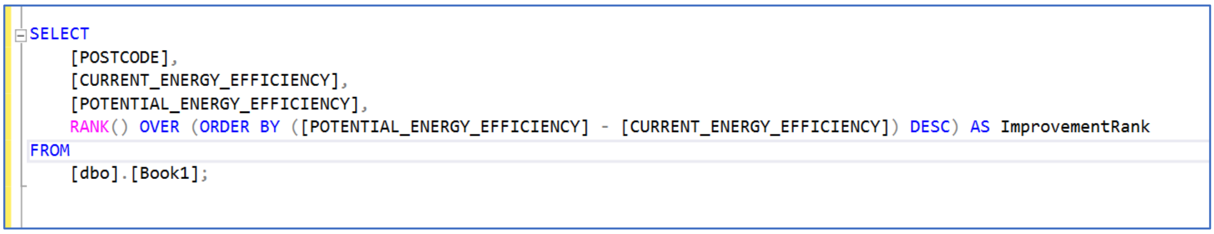
We are Creating a stored procedure to retrieve or manipulate data based on certain input parameters like energy rating or postcode.



```
CREATE PROCEDURE GetPropertyByRating
    @EnergyRating CHAR(1)
AS
BEGIN
    SELECT
        [POSTCODE],
        [BUILDING_REFERENCE_NUMBER],
        [POTENTIAL_ENERGY_RATING],
        [CURRENT_ENERGY_RATING],
        [POTENTIAL_ENERGY_EFFICIENCY],
        [CURRENT_ENERGY_EFFICIENCY]
    FROM
        [dbo].[Book1]
    WHERE
        [CURRENT_ENERGY_RATING] = @EnergyRating;
END;
```

2.2.7. Aggregate and Ranking Functions

We Utilize SQL's built-in functions to analyze the data, such as finding the top 10 properties with the highest potential for energy efficiency improvement.



```
SELECT
    [POSTCODE],
    [CURRENT_ENERGY_EFFICIENCY],
    [POTENTIAL_ENERGY_EFFICIENCY],
    RANK() OVER (ORDER BY ([POTENTIAL_ENERGY_EFFICIENCY] - [CURRENT_ENERGY_EFFICIENCY]) DESC) AS ImprovementRank
FROM
    [dbo].[Book1];
```

03. Exploratory Data Analysis (EDA)

Overview

Exploratory Data Analysis (EDA) is a crucial step in understanding the structure, patterns, and relationships within the data. For this assignment, we performed EDA on the **Energy Performance Certificate (EPC)** dataset for Manchester's local authorities, covering the period from 2013 to 2023. The primary goal was to explore the data, identify trends in energy performance, CO2 emissions, and energy consumption, and uncover potential areas for improvement in energy efficiency.

Objectives of EDA

- Understand the distribution of key metrics such as energy ratings, CO2 emissions, and energy consumption.
- Detect patterns and trends in energy performance over time.
- Investigate geographical differences in energy efficiency across different local authorities.
- Identify correlations between energy ratings, CO2 emissions, and property types.
- Highlight any potential anomalies, such as outliers, that might affect the analysis.

3.2.1 Energy Ratings Distribution

The **distribution of energy ratings** was one of the first aspects we explored. We used a **bar chart** to show the proportion of properties that fall into each energy efficiency category (A-G). This allowed us to quickly see whether most properties are energy efficient (A-C ratings) or whether the majority fall into the lower ratings (D-G), which indicates a need for improvement.

Key Insights:

- The majority of properties in Manchester fall within the D and E energy rating categories, indicating a significant opportunity for energy efficiency improvements.
- Only a small percentage of properties have an A or B rating, reflecting a potential area for energy upgrades.

3.2.2 CO2 Emissions Analysis

We performed an analysis of **current and potential CO2 emissions** across different local authorities. A **line chart** was used to track the trends in CO2 emissions over time, highlighting how emissions have fluctuated across Manchester between 2013 and 2023.

In addition, a **map visualization** was created to show **geographic variations** in CO2 emissions. This allowed us to identify specific areas where properties emit higher levels of CO2, which can guide policy interventions for reducing emissions.

Key Insights:

- Certain local authorities, particularly in densely populated areas, exhibited significantly higher CO2 emissions compared to suburban and rural areas.
- CO2 emissions have gradually decreased over the years, but there is still room for improvement, especially in older properties with inefficient heating systems.

3.2.3 Energy Consumption Trends

We analyzed trends in **current energy consumption** versus **potential energy consumption** to identify opportunities for energy savings. Using a **stacked column chart**, we compared current and potential energy consumption across different property types (e.g., flats, detached houses, terraced houses).

This analysis helped highlight which property types have the highest energy consumption and the most potential for improvement if energy-saving measures are applied.

Key Insights:

- Detached houses and older buildings exhibited higher energy consumption levels, particularly in heating costs.
- There is a significant opportunity for reducing energy consumption in terraced houses and flats by upgrading insulation and heating systems.

4.2.4 Geographic Differences in Energy Efficiency

To explore geographic differences in energy efficiency, we created a **heat map** that showed the average energy efficiency scores across local authorities in Manchester. This helped us identify regions that have consistently poor energy performance, as well as those that are performing well.

Key Insights:

- Central Manchester tends to have poorer energy efficiency ratings, likely due to older buildings with outdated insulation and heating systems.
- Suburban areas generally perform better in terms of energy efficiency, with more properties achieving higher ratings (A-C).

3.3 Outlier Detection

Outliers in the dataset, such as abnormally high CO2 emissions or extremely low energy consumption, were identified and analyzed.

For instance, properties with extremely high CO2 emissions (above a certain threshold) were flagged for further investigation. These outliers could indicate errors in the data or properties with significantly outdated infrastructure that require urgent intervention.

Steps Taken to Handle Outliers:

- **Filter Outliers**: We filtered out properties with extremely high CO2 emissions or energy consumption that were far outside the normal range.
- **Replace Missing Values**: For certain outliers with missing or inconsistent data, we replaced them with average values based on the property type and location.

04. Power BI Dashboard Design

The Power BI dashboard is designed to provide an intuitive, user-friendly interface for stakeholders to explore energy performance data across Manchester's local authorities. The dashboard is divided into four distinct categories to ensure that all key aspects of the data are covered and presented in a logical flow. Each category addresses a specific area of energy performance, allowing users to dive deep into relevant data and extract meaningful insights.

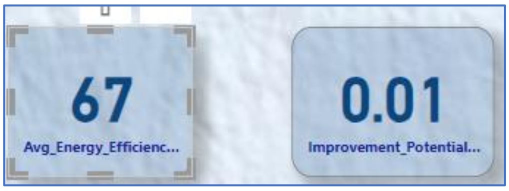
Dashboard Categories:

1. **Energy Efficiency Overview**
2. **Property Type & Energy Consumption Breakdown**
3. **CO2 Emission Analysis and Geography Insights**
4. **Risk & Opportunity Analysis**

4.1 Category 1: Energy Efficiency Overview

This section provides a high-level summary of the energy efficiency ratings of properties in Manchester. It is designed to give stakeholders an immediate understanding of the overall energy performance across local authorities, highlighting areas where energy efficiency improvements can be made.

1. **Key Performance Indicators (KPI Cards):**



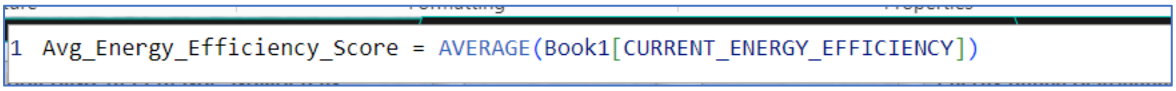
Display high-level metrics, such as the average energy efficiency score and the percentage of properties with potential for improvement..

Visual: KPI Cards

- Card 1: Average Energy Efficiency Score
- Card 2: Percentage of Properties with Potential for Improvement

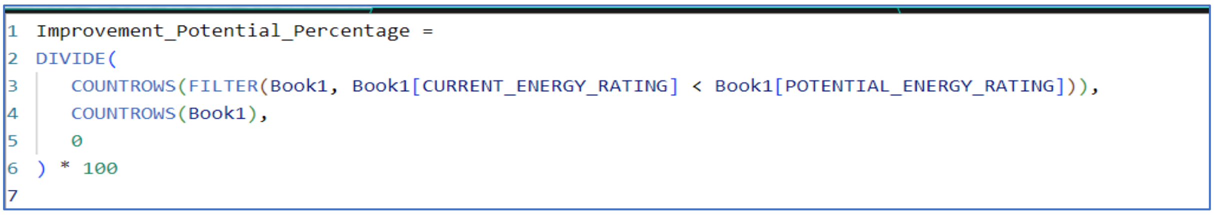
DAX Code for KPIs:

1. **Average Energy Efficiency Score**:



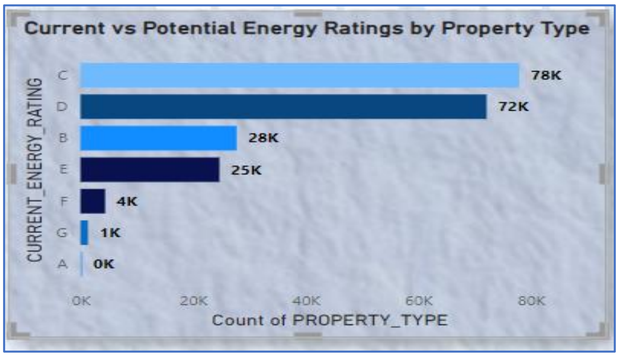
```
Avg_Energy_Efficiency_Score = AVERAGE(Book1[CURRENT_ENERGY_EFFICIENCY])
```

2. **Percentage of Properties with Potential for Improvement**



```
Improvement_Potential_Percentage =
DIVIDE(
    COUNTROWS(FILTER(Book1, Book1[CURRENT_ENERGY_RATING] < Book1[POTENTIAL_ENERGY_RATING])),
    COUNTROWS(Book1),
    0
) * 100
```

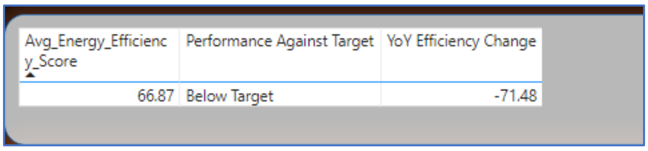
2. Bar Chart: Current vs Potential Energy Ratings by Property Type



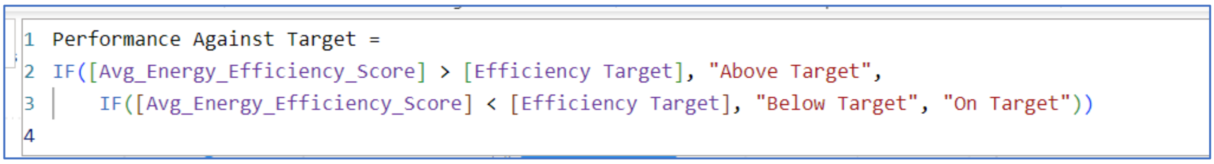
A **clustered bar chart** comparing **current energy ratings** to **potential energy ratings** across different property types (House, Flat, etc.) gives a clear indication of how properties could improve. This helps you quickly see which property types have the most room for improvement

3.Table - Detailed Energy Efficiency Data

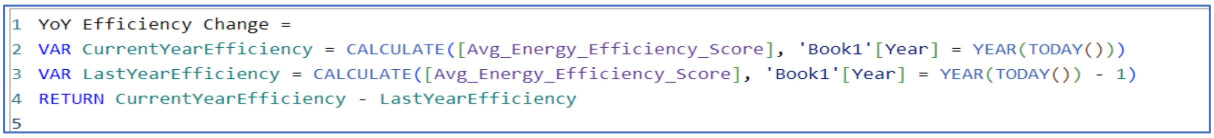
We Provides a detailed breakdown of energy efficiency data by property, including current ratings, targets, and performance against those targets.



Avg_Energy_Efficiency_Score	Performance Against Target	YoY Efficiency Change
66.87	Below Target	-71.48



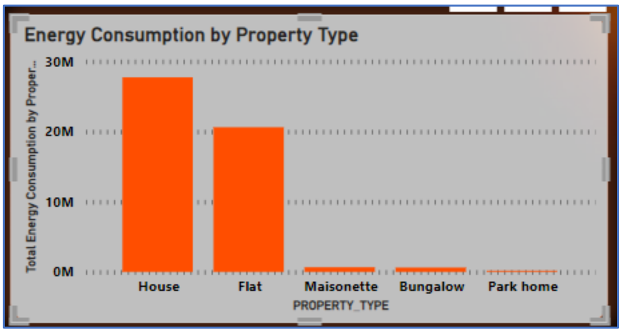
```
Performance Against Target =
IF([Avg_Energy_Efficiency_Score] > [Efficiency Target], "Above Target",
    IF([Avg_Energy_Efficiency_Score] < [Efficiency Target], "Below Target", "On Target"))
```



```
YoY Efficiency Change =
VAR CurrentYearEfficiency = CALCULATE([Avg_Energy_Efficiency_Score], 'Book1'[Year] = YEAR(TODAY()))
VAR LastYearEfficiency = CALCULATE([Avg_Energy_Efficiency_Score], 'Book1'[Year] = YEAR(TODAY()) - 1)
RETURN CurrentYearEfficiency - LastYearEfficiency
```

4.Energy Consumption by Property Type

The "Energy Consumption by Property Type" chart is a column chart that visually compares the total energy consumption across various property types. It specifically presents data for Houses, Flats, and Maisonettes, though only Houses and Flats show significant energy usage, nearly 30 million and just over 20 million units respectively. Maisonettes, Bungalows, and Park homes are also listed but do not display any visible energy consumption, indicating minimal or no usage in these categories. This chart effectively highlights which property types are the largest consumers of energy, aiding in targeted energy management and conservation efforts.

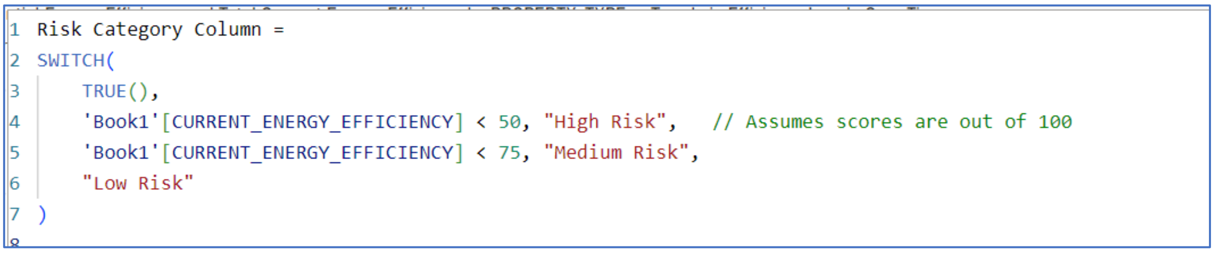


Category 2.Risk & Opportunity Analysis

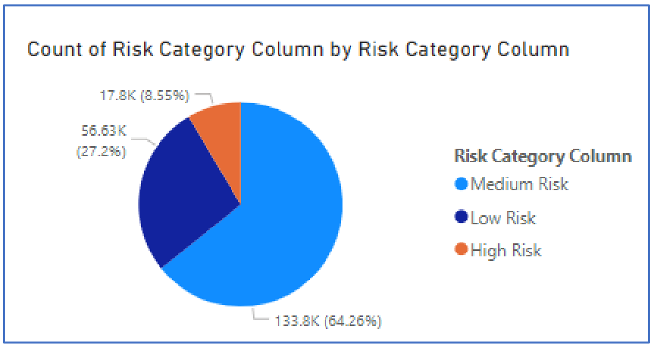
1.Pie Chart - Distribution of Risk Categories

We Visualize the proportional distribution of risk based on energy efficiency levels.it immediate understanding of which risk categories contain the most properties and ability to prioritize interventions based on the size of each risk category.

DAX for Risk Category Calculation:

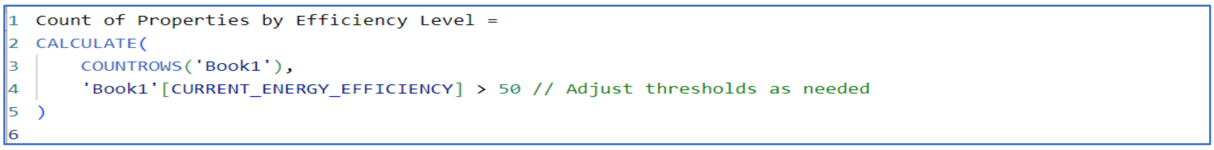


```
Risk Category Column =
SWITCH(
    TRUE(),
    'Book1'[CURRENT_ENERGY_EFFICIENCY] < 50, "High Risk",    // Assumes scores are out of 100
    'Book1'[CURRENT_ENERGY_EFFICIENCY] < 75, "Medium Risk",
    "Low Risk"
)
```

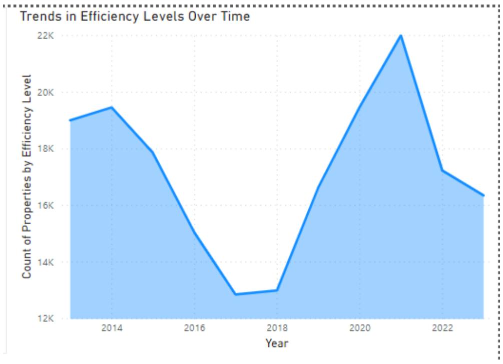


2.Stacked Area Chart - Efficiency Improvement Over Time

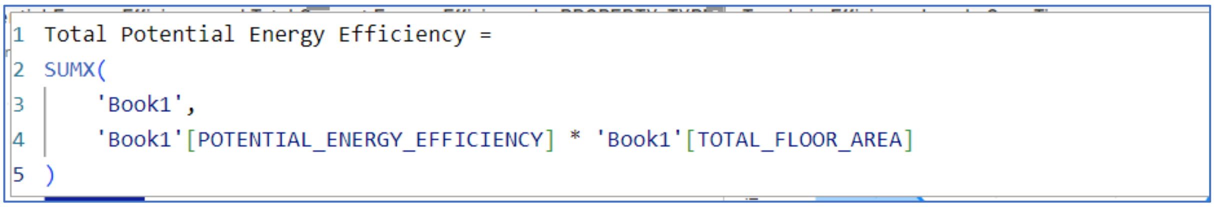
Monitor changes in energy efficiency levels over time across different property category.this trends in energy efficiency improvements or declines.



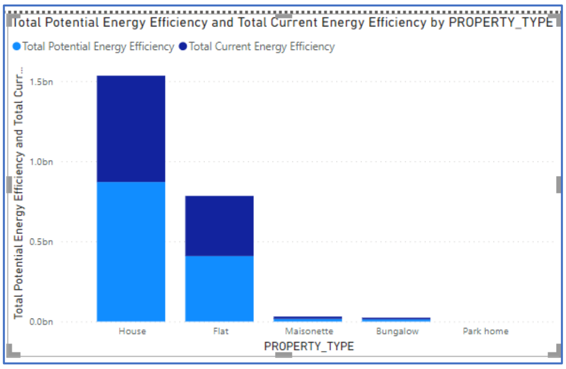
```
Count of Properties by Efficiency Level =
CALCULATE(
    COUNTROWS('Book1'),
    'Book1'[CURRENT_ENERGY_EFFICIENCY] > 50 // Adjust thresholds as needed
)
```



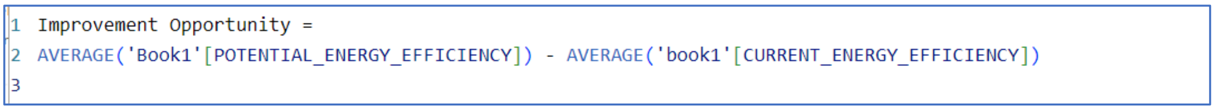
Total Potential Energy Efficiency and Total Current Energy Efficiency by PROPERTY_TYPE," is a stacked column chart that compares the total current energy efficiency against the total potential energy efficiency for various types of properties. The chart clearly shows that houses and flats are the primary focus, with houses displaying a significant gap between current and potential energy efficiency, indicating substantial room for improvement. Flats also show a notable difference but to a lesser extent, whereas Maisonettes, Bungalows, and Park homes have minimal or no data shown.



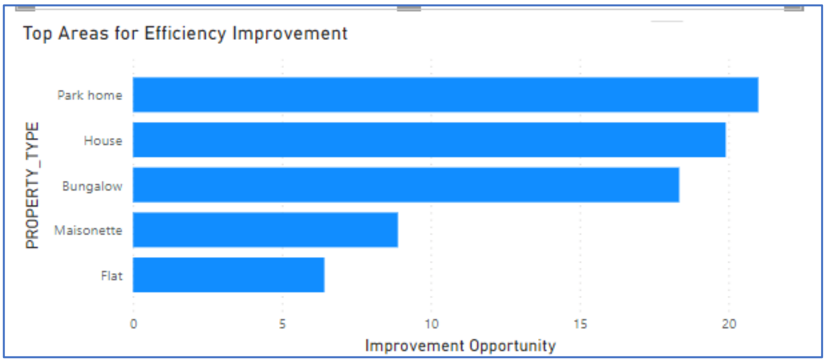
```
Total Potential Energy Efficiency =
SUMX(
    'Book1',
    'Book1'[POTENTIAL_ENERGY_EFFICIENCY] * 'Book1'[TOTAL_FLOOR_AREA]
)
```



Top Areas for Efficiency Improvement," is a horizontal bar chart that ranks property types by their potential for efficiency improvement. The chart lists 'Park home' as having the highest potential, followed by 'House,' 'Bungalow,' 'Maisonette,' and 'Flat,' indicating where efforts to improve energy efficiency could be prioritized. The use of a horizontal layout makes it easy to compare across types and immediately spot which property types require the most attention

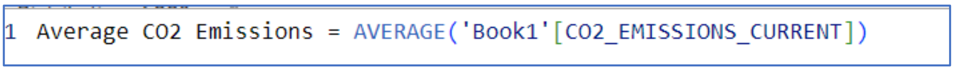


```
Improvement Opportunity =
AVERAGE('Book1'[POTENTIAL_ENERGY_EFFICIENCY]) - AVERAGE('book1'[CURRENT_ENERGY_EFFICIENCY])

```

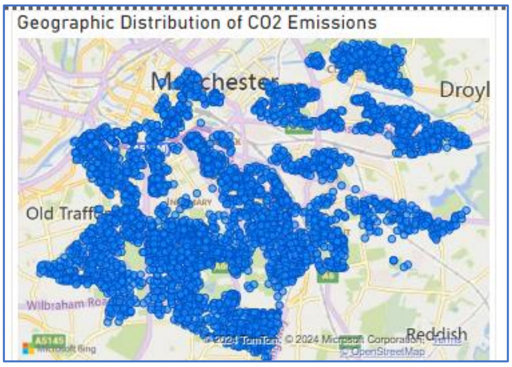


Category 3.CO2 Emission Analysis and Geography Insights

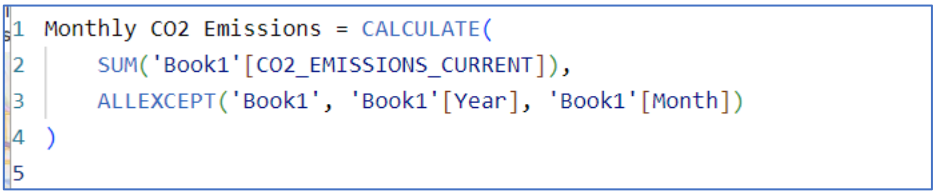
1. **Geographic Distribution of CO2 Emissions**: This map highlights the distribution of CO2 emissions across the Manchester area, using blue markers to indicate emission levels in different locations. This visualization helps identify specific areas with higher emissions, aiding targeted environmental interventions.



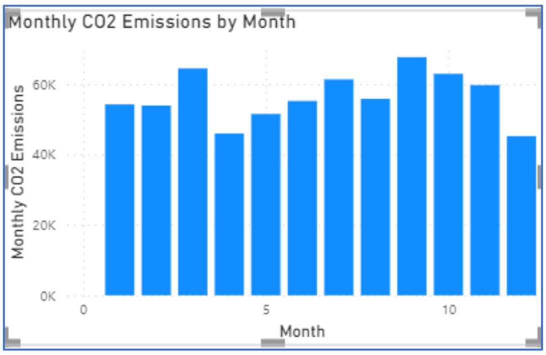
```
Average CO2 Emissions = AVERAGE('Book1'[CO2_EMISSIONS_CURRENT])
```


Geographic Distribution of CO2 Emissions
Manchester
Droyl
Old Traff
Wilbraham Roa
Reddish
A5145
Microsoft Bing
© 2024 TomTom, © 2024 Microsoft Corporation, Terms
© OpenStreetMap

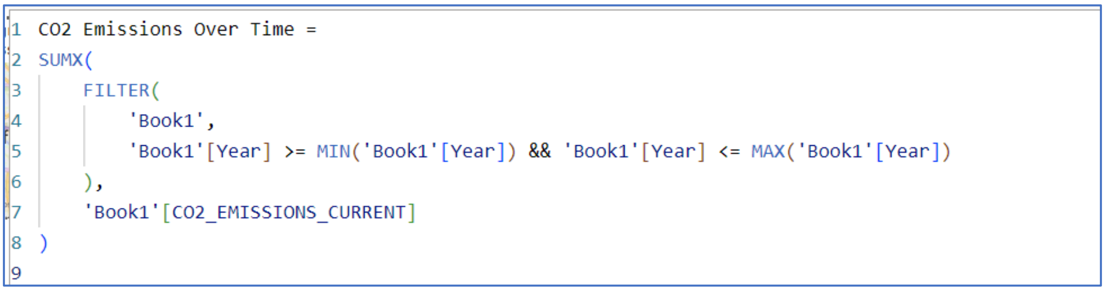
2.**Monthly CO2 Emissions by Month**: The column chart shows the fluctuations in CO2 emissions throughout the year, with each bar representing a month. This visualization reveals a trend where emissions peak around the middle of the year, possibly indicating seasonal variations in energy use or industrial activity.



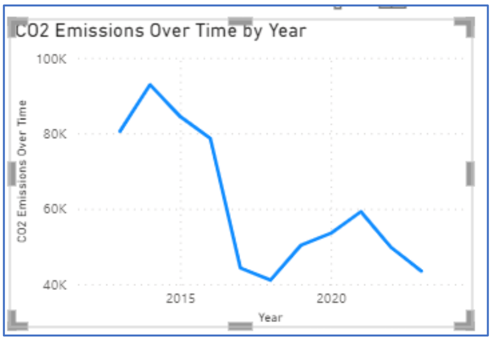
```
Monthly CO2 Emissions = CALCULATE(
    SUM('Book1'[CO2_EMISSIONS_CURRENT]),
    ALLEXCEPT('Book1', 'Book1'[Year], 'Book1'[Month])
)

```



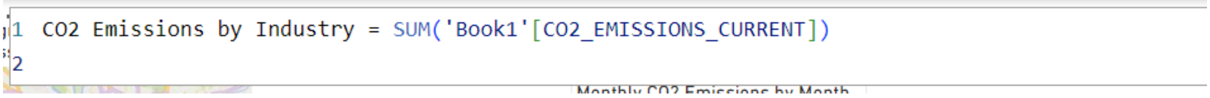
3.**CO2 Emissions Over Time by Year**: The line chart tracks the annual CO2 emissions from 2015 to beyond 2020, showing significant fluctuations over the years. The trend indicates a peak around 2015, followed by a sharp decrease and subsequent fluctuations, which could reflect changes in local industry practices, regulatory impacts, or economic factors.



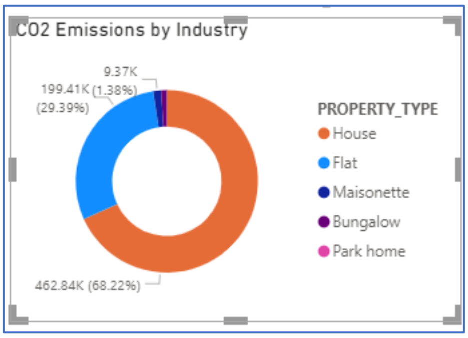
```
CO2 Emissions Over Time =
SUMX(
    FILTER(
        'Book1',
        'Book1'[Year] >= MIN('Book1'[Year]) && 'Book1'[Year] <= MAX('Book1'[Year])
    ),
    'Book1'[CO2_EMISSIONS_CURRENT]
)

```



4.**CO2 Emissions by Industry**: The donut chart breaks down CO2 emissions by property type, indicating a predominant contribution by 'Flat' and 'House' types, which together make up the majority of emissions. This chart helps in understanding which property types are the most significant contributors to emissions, informing policy and energy efficiency strategies.

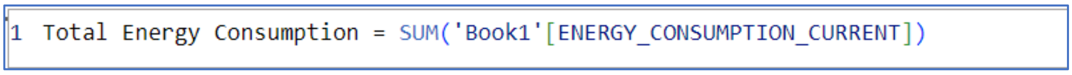


```
CO2 Emissions by Industry = SUM('Book1'[CO2_EMISSIONS_CURRENT])

```

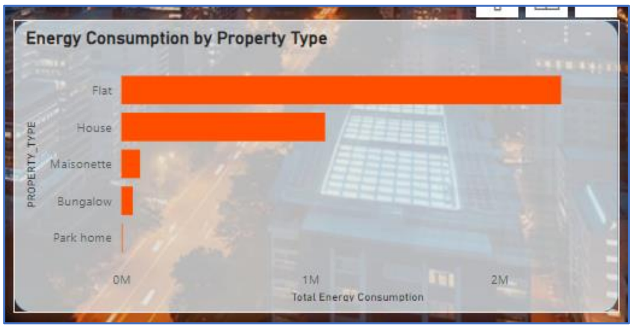


Category 4.Property Type & Energy Consumption Breakdown

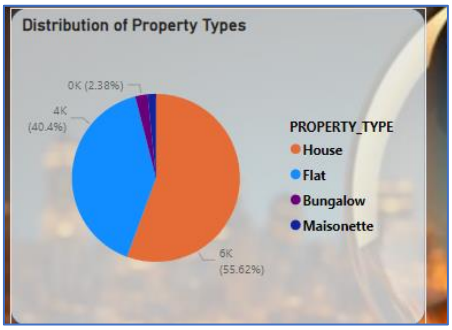
1. **Energy Consumption by Property Type**: This bar chart displays the total energy consumption segmented by property type. Flats consume the most energy, significantly more than houses, with maisonettes and bungalows showing relatively minimal consumption. The visual helps identify which property types are the highest energy consumers, potentially guiding energy efficiency initiatives.

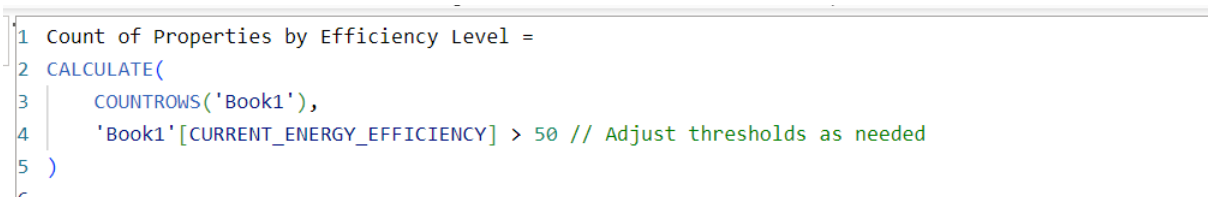


```
Total Energy Consumption = SUM('Book1'[ENERGY_CONSUMPTION_CURRENT])
```



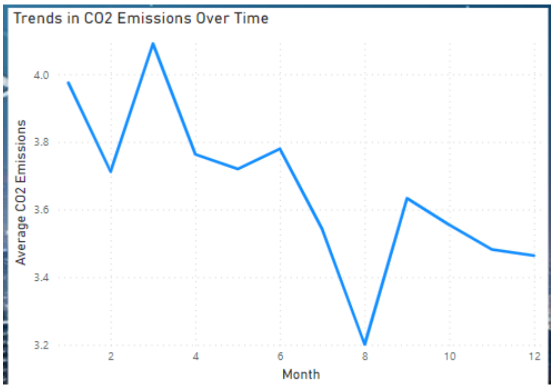
2. **Distribution of Property Types**: The pie chart shows the distribution of different property types. Houses dominate the mix, comprising 93.52% of the properties, followed by a small percentage of flats (4.45%) and an even smaller fraction of maisonettes (2.03%). This chart provides a quick view of the property landscape, useful for contextualizing energy consumption data.





```
Count of Properties by Efficiency Level =
CALCULATE(
    COUNTROWS('Book1'),
    'Book1'[CURRENT_ENERGY_EFFICIENCY] > 50 // Adjust thresholds as needed
)
```

3. **Trends in CO2 Emissions Over Time**: The line chart plots the average monthly CO2 emissions, revealing fluctuations over the course of a year. This visualization highlights the variability in emissions, which could correlate with seasonal changes, operational adjustments, or external factors affecting energy use.



4. **Map of Local Authority and Postcode**: The global map does not display specific data points but includes an indication of the local authority area (E08000003), suggesting a focus on regional analysis. This part of the dashboard could be utilized for geographically slicing the data to analyze energy and emissions data by specific areas.

5.Findings and Recommendations

Findings:

1. **Property Type Energy Efficiency Variability:**
   - Energy consumption and efficiency levels vary significantly across different property types, with flats showing the highest energy usage
2. **CO2 Emissions Insights:**
   - Analysis of CO2 emissions reveals a pronounced seasonal variability, suggesting higher emissions during colder months. There is also a noticeable reduction in emissions over the years, reflecting the impact of existing energy efficiency measures.
3. **Geographical Disparities:**
   - The geographic distribution of CO2 emissions and energy efficiency ratings showcases notable differences between various local authorities. Some areas, particularly urban centers, show higher emissions, which could be targeted for specific policy interventions.
4. **Opportunities for Improvement:**
   - There is a considerable potential for improvement in energy efficiency, especially in houses and park homes, as indicated by the gap between current and potential energy ratings.

Recommendations:

1. **Targeted Retrofit Programs:**
   - Implement targeted retrofit programs for flats and houses to enhance their energy efficiency. This could include upgrading insulation, installing energy-efficient windows, and modern heating systems.
2. **Seasonal Energy Efficiency Strategies:**
   - Develop seasonal energy efficiency strategies to address the peak in emissions during the colder months.
3. **Local Authority Collaboration:**
   - Collaborate with local authorities to implement region-specific energy efficiency initiatives.
4. **Enhanced Monitoring and Reporting:**
   - Strengthen monitoring and reporting mechanisms to track energy consumption and CO2 emissions more effectively.

   -

Conclusion

The analysis conducted using the Power BI dashboard provided profound insights into the patterns of energy consumption and CO2 emissions across Manchester. It highlighted significant opportunities for enhancing energy efficiency, particularly in residential properties. By focusing on the identified areas with the highest energy consumption and CO2 emissions, targeted interventions can be designed to improve energy efficiency, reduce emissions, and move towards greater sustainability. Continuing to leverage data-driven insights will be crucial in guiding effective policy-making and achieving long-term energy sustainability goals.