



GENERAL SIR JOHN KOTELAWALA DEFENCE UNIVERSITY

Faculty of Management, Social Sciences and Humanities

Department of Languages

BSc in Applied Data Science Communication

INTAKE 40 – 2024

Assessment Title: Assignment 2 – Task 01

Child Well-Being Monitor & Greater Manchester Domestic Energy
Performance Analyzer

Module Name : Advanced SQL and Cloud Databases

Module Code : LB 2224

Submission Date : 14/10/2024

Student Name	Reg. No
D.L.S. Nadavi	D/ADC/23/0028
Y.M.S.S.B. Senevirathna	D/ADC/23/0030

Content

Task 01 - Child Well-Being Monitor (Power BI Reports)

- 1) Introduction
- 2) Explanation and Preparation of Data Set
 - a. Explanation of the data set
 - b. Preparation of the data set
- 3) Report Analysis - No. of Children in Peru & Vietnam
- 4) Report Analysis - Children's Average Weight against their Average Height by Age in Peru & Vietnam
- 5) Report Analysis - Summary of BMI value with Age in Peru & Vietnam
- 6) Report Analysis - Child's Health Level in Peru & Vietnam
- 7) Report Analysis- Household head's relationship to YL child In Peru & Vietnam
- 8) Report Analysis - Food Security Level in Peru & Vietnam
- 9) Conclusion
- 10) References for Task 01

Task 02 - Greater Manchester Domestic Energy Performance Analyzer

(Power BI Dashboard)

- 1) Introduction
- 2) Explanation and Preparation of the data set
 - a. Explanation of the data set
 - b. Preparation of the data
- 3) Dashboard Design & Implementation
- 4) Conclusion
- 5) References for Task 02

Task 01 –

Child Well-Being Monitor

(Power BI Reports)

1) Introduction

This report describes the processes of creating a report of Child Well-Being Monitor, depicts child poverty in the low- and middle-income countries. (Peru and Vietnam). This tool harnesses Power BI Report Builder and SQL to transform data into charts and tables.. In use of SQL Server as a data storage technology and T-SQL as a data manipulation language, the project is assured data management and excellent reporting techniques. This report and the design of the Power BI reports will be more user friendly for the decision makers.

2) Explanation and Preparation of Data Set

a. Explanation of the data set

In here the data set of Young Lives: an International Study of Childhood Poverty: Rounds 1-5 Constructed Files, 2002-2016 is used. (Boyden, J. (2022). Young Lives: An International Study of Childhood Poverty: Rounds 1-5 Constructed Files, 2002-2016. [data collection]. 5th Edition. UK Data Service. SN: 7483, [DOI: http://doi.org/10.5255/UKDA-SN-7483-5](http://doi.org/10.5255/UKDA-SN-7483-5)) Moreover “peru_constructed” and “vietnam_constructed” datasets are used to analyze through Power BI Report Builder.

The dataset of “peru_constructed” consists of data of Peru. It has 215 columns and 13,830 rows.

And the dataset of “vietnam_constructed” consists of the data of Vietnam. It has 213 columns and 15,000 rows.

b. Preparation of the data set

For the data preparation, T-SQL was used. The following codes & results show how the data set was cleaned by SQL.

1 – Preview the data

```
-- Preview the first 10 rows of the dataset
SELECT TOP 10 *
FROM [dbo].[peru_constructed]
```

```
SELECT TOP 10 *
FROM [dbo].[vietnam_constructed]
```

2. Get the column names and data types.

```
SELECT COLUMN_NAME, DATA_TYPE
FROM INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_NAME = 'peru_constructed';
```

```

SELECT COLUMN_NAME, DATA_TYPE
FROM INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_NAME = 'vietnam_constructed';

```

COLUMN_NAME	DATA_TYPE
childid	int
yc	varchar
inout	varchar
panel2345	varchar
deceased	varchar
def	varchar
commed	varchar
dusid	varchar
hypode	varchar
lgon	varchar
chid	varchar
cixer	varchar
chang	varchar
chdhac	varchar
agemon	varchar
marmocab	varchar
marmocab_age	varchar

Query executed successfully.

3. Check Missing Values or NULL Values.

```

SELECT
CASE WHEN [childid] IS NULL OR [childid] = '' THEN 1 ELSE 0 END) AS Missing_ChildID,
CASE WHEN [agemon] IS NULL OR [agemon] = '' THEN 1 ELSE 0 END) AS Missing_AgeMon,
CASE WHEN [foodsec] IS NULL OR [foodsec] = '' THEN 1 ELSE 0 END) AS Missing_Foodsc,
CASE WHEN [chheight] IS NULL OR [chheight] = '' THEN 1 ELSE 0 END) AS Missing_Chheight,
CASE WHEN [chweight] IS NULL OR [chweight] = '' THEN 1 ELSE 0 END) AS Missing_Chwheight,
CASE WHEN [bmi] IS NULL OR [bmi] = '' THEN 1 ELSE 0 END) AS Missing_BMI,
CASE WHEN [chhealth] IS NULL OR [chhealth] = '' THEN 1 ELSE 0 END) AS Missing_chhealth,
CASE WHEN [headrel] IS NULL OR [headrel] = '' THEN 1 ELSE 0 END) AS Missing_headrel,
CASE WHEN [dadedu] IS NULL OR [dadedu] = '' THEN 1 ELSE 0 END) AS Missing_dadedu,
CASE WHEN [momedu] IS NULL OR [momedu] = '' THEN 1 ELSE 0 END) AS Missing_momedu
FROM peru_constructed;

```

Query executed successfully.

```

SELECT
CASE WHEN [childid] IS NULL OR [childid] = '' THEN 1 ELSE 0 END) AS Missing_ChildID,
CASE WHEN [agemon] IS NULL OR [agemon] = '' THEN 1 ELSE 0 END) AS Missing_AgeMon,
CASE WHEN [foodsec] IS NULL OR [foodsec] = '' THEN 1 ELSE 0 END) AS Missing_Foodsc,
CASE WHEN [chheight] IS NULL OR [chheight] = '' THEN 1 ELSE 0 END) AS Missing_Chheight,
CASE WHEN [chweight] IS NULL OR [chweight] = '' THEN 1 ELSE 0 END) AS Missing_Chwheight,
CASE WHEN [bmi] IS NULL OR [bmi] = '' THEN 1 ELSE 0 END) AS Missing_BMI,
CASE WHEN [chhealth] IS NULL OR [chhealth] = '' THEN 1 ELSE 0 END) AS Missing_chhealth,
CASE WHEN [headrel] IS NULL OR [headrel] = '' THEN 1 ELSE 0 END) AS Missing_headrel,
CASE WHEN [dadedu] IS NULL OR [dadedu] = '' THEN 1 ELSE 0 END) AS Missing_dadedu,
CASE WHEN [momedu] IS NULL OR [momedu] = '' THEN 1 ELSE 0 END) AS Missing_momedu
FROM vietnam_constructed;

```

Query executed successfully.

4. Add a new column for Country name.

```

alter table [dbo].[peru_constructed]
add Country_Name Varchar(50);

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

```

5. Create a view for building a relationship.

```

CREATE VIEW Child_Poverty_View AS
SELECT
    'Peru' AS country,
    foodsec,
    agemon,
    CAST(chweight AS FLOAT) AS chweight, -- Convert chweight to numeric type
    CAST(chheight AS FLOAT) AS chheight, -- Convert chheight to numeric type
    CAST(bmi AS FLOAT) AS bmi, -- Convert bmi to numeric type
    CAST(chhealth AS FLOAT) AS chhealth,
    CAST(headrel AS FLOAT) AS headrel,
    CAST(dadedu AS FLOAT) AS dadedu,
    CAST(momedu AS FLOAT) AS momedu
FROM peru_constructed
WHERE
    foodsec IS NOT NULL
    AND agemon IS NOT NULL
    AND ISNUMERIC(chweight) = 1 -- Ensure chweight contains numeric values
    AND ISNUMERIC(chheight) = 1 -- Ensure chheight contains numeric values
    AND ISNUMERIC(bmi) = 1 -- Ensure bmi contains numeric values
    AND CAST(chweight AS FLOAT) > 0
    AND CAST(chheight AS FLOAT) > 0
    AND CAST(bmi AS FLOAT) > 0
    AND CAST(chhealth AS FLOAT) > 0
    AND CAST(headrel AS FLOAT) > 0
    AND CAST(dadedu AS FLOAT) > 0
    AND CAST(momedu AS FLOAT) > 0
    AND CAST(chweight AS FLOAT) BETWEEN 2 AND 60
    AND CAST(chheight AS FLOAT) BETWEEN 40 AND 200
UNION ALL

SELECT
    'Vietnam' AS country,
    foodsec,
    agemon,
    CAST(chweight AS FLOAT) AS chweight, -- Convert chweight to numeric type
    CAST(chheight AS FLOAT) AS chheight, -- Convert chheight to numeric type
    CAST(bmi AS FLOAT) AS bmi, -- Convert bmi to numeric type
    CAST(chhealth AS FLOAT) AS chhealth,
    CAST(headrel AS FLOAT) AS headrel,
    CAST(dadedu AS FLOAT) AS dadedu,
    CAST(momedu AS FLOAT) AS momedu
FROM vietnam_constructed
WHERE
    foodsec IS NOT NULL
    AND agemon IS NOT NULL
    AND ISNUMERIC(chweight) = 1 -- Ensure chweight contains numeric values
    AND ISNUMERIC(chheight) = 1 -- Ensure chheight contains numeric values
    AND ISNUMERIC(bmi) = 1 -- Ensure bmi contains numeric values
    AND CAST(chweight AS FLOAT) > 0
    AND CAST(chheight AS FLOAT) > 0
    AND CAST(bmi AS FLOAT) > 0
    AND CAST(chhealth AS FLOAT) > 0
    AND CAST(headrel AS FLOAT) > 0
    AND CAST(dadedu AS FLOAT) > 0
    AND CAST(momedu AS FLOAT) > 0
    AND CAST(chweight AS FLOAT) BETWEEN 2 AND 60
    AND CAST(chheight AS FLOAT) BETWEEN 40 AND 200;

```

6. Create a new view by ordering the column order and adding a new column with age (age category).

```

CREATE VIEW Child_Poverty AS
SELECT
    [country],
    CASE
        WHEN FLOOR([agemon] / 12) < 1 THEN '<1'
        WHEN FLOOR([agemon] / 12) BETWEEN 1 AND 5 THEN '1-5'
        WHEN FLOOR([agemon] / 12) BETWEEN 6 AND 10 THEN '6-10'
        WHEN FLOOR([agemon] / 12) BETWEEN 11 AND 15 THEN '11-15'
        WHEN FLOOR([agemon] / 12) BETWEEN 16 AND 20 THEN '16+'
    END AS age_category,-- New column for age category
    FLOOR ([agemon] / 12) AS age,
    chweight= ROUND([chweight], 2),
    chheight= ROUND([chheight], 2),
    bmi= ROUND([bmi], 2),
    [chhealth],
    [headrel],
    [momedu],
    [dadedu],
    [foodsec]
FROM [dbo].[Child_Poverty_View];

```

7. Preview the new data of the view (Child_Poverty).

	country	age_category	age	chweight	chheight	bmi	chhealth	headrel	momedu	dadedu	foodsec
1	Peru	6-10	7	22.2	112.3	17.6	4	1	11	2	1
2	Peru	11-15	11	34.1	133.2	19.	4	1	11	2	1
3	Peru	11-15	14	46.2	143.8	22.	4	1	11	2	2
4	Peru	6-10	7	19.5	116	14.	4	1	11	3	2
5	Peru	11-15	11	31.5	139.2	16.	3	1	11	3	1
6	Peru	11-15	14	39.2	149.1	17.	4	1	13	3	1
7	Peru	6-10	8	34	125.5	21.	5	1	11	11	2
8	Peru	11-15	12	51.6	142	25.	4	1	11	11	2
9	Peru	6-10	8	32.8	134.5	18.	3	1	14	14	2
10	Peru	11-15	12	56.9	163.6	21.	4	1	15	14	2
11	Peru	6-10	7	16.2	111.4	13.	3	1	4	4	3
12	Peru	11-15	11	28.3	132.7	16.	4	1	4	4	3
13	Peru	11-15	14	37.7	150.6	16.	4	1	4	4	2
14	Peru	6-10	8	35.2	124.9	22.	4	1	16	15	2
15	Peru	6-10	8	22	119.2	15.	4	1	13	13	2
16	Peru	11-15	12	38.8	138.4	20.	4	1	13	13	1
17	Peru	11-15	15	48	160.2	18.7	4	1	13	13	2
18	Peru	6-10	7	26.1	119.7	18.	4	1	11	11	1
19	Peru	11-15	11	49.7	148	22.	3	1	11	11	1
20	Peru	11-15	12	50	161	19.	4	3	9	11	2
21	Peru	6-10	7	17.05	107.7	14.7	4	1	8	6	2

Query executed successfully. (local) (16.0 RTM) DESKTOP-FNNU06P\ASUS (72) | The Young Lives Repository 00:00:00 9,144 rows

The created “Child_Poverty” data view consists of 11 columns and 9,144 rows.

Column names are ‘country’, ‘age_category’, ‘age’, ‘chheight’, ‘chweight’, ‘bmi’, ‘chhealth’, ‘headrel’, ‘momedu’ , ‘dadedu’ and ‘foodsec’.

- country :- The name of the country (Peru / Vietnam)
- age_category :- Age groups (1 -5 / 6 - 10 / 11 – 15/ 16 +)
- age :- Age of children in years (1 - 16)

- chheight :- Child's height (cm)
- chweight :- Child's weight (kg)
- bmi :- Calculated BMI

- chhealth :- Health level of children

1 - Very Poor 2 - Poor 3 - Average 4 - Good 5 - Very Good

- headrel :- Household head's relationship to the child

0 - YL Child 1 - Biological Parent 2 - Non-Biological Parent
 3 – Grandparent 4 - Uncle/Aunt 5 – Sibling 6 - Other-Relative
 7 - Other non-relative 8 - Partner/Spouse of child
 9 - Father-in-law/Mother-in-law

- momedu :- The education level of the mother

0 - None 1 to 11 - Grade 1 to Grade 11
 13 - Technical, pedagogical, CETPRO (incomplete)
 14 - Technical, pedagogical, CETPRO (complete)
 15 - University (incomplete) 16 - University (Complete)
 28 - Adult literacy 30 – Other

- dadedu :- The education level of the father

0 - None 1 to 11 - Grade 1 to Grade 11
 13 - Technical, pedagogical, CETPRO (incomplete)
 14 - Technical, pedagogical, CETPRO (complete)
 15 - University (incomplete) 16 - University (Complete)
 28 - Adult literacy 30 - Other

- foodsec :- Household's food situation in the last 12 months

1 - Always eat enough of what want
 2 - Eat enough but not always what like
 3 - Sometimes not eat enough
 4 - Frequently not eat enough

***** Note :-**

Although the assignment says to clean the dataset using Power BI, we did that through SQL Server because we couldn't import directly the published data set in my workspace in Power BI to the Power BI Report Builder.

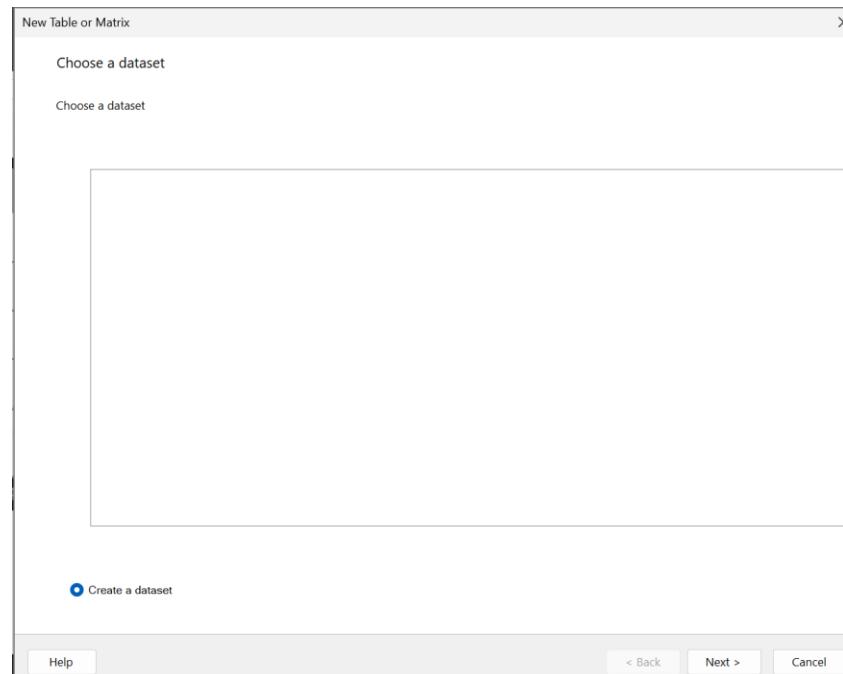
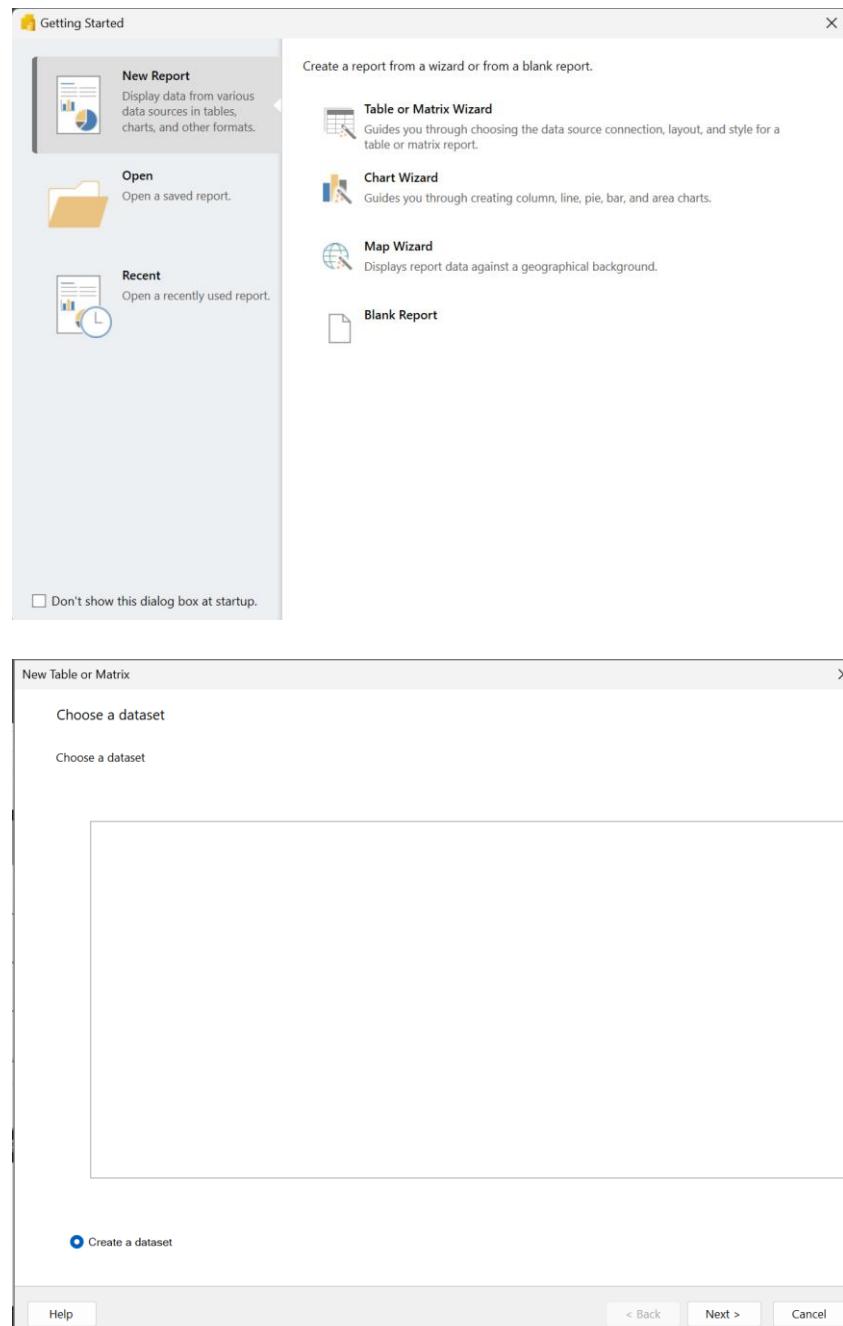
We tried to do that using Microsoft fabric. But we didn't have the premium version of Power BI because our free trial was finished before.

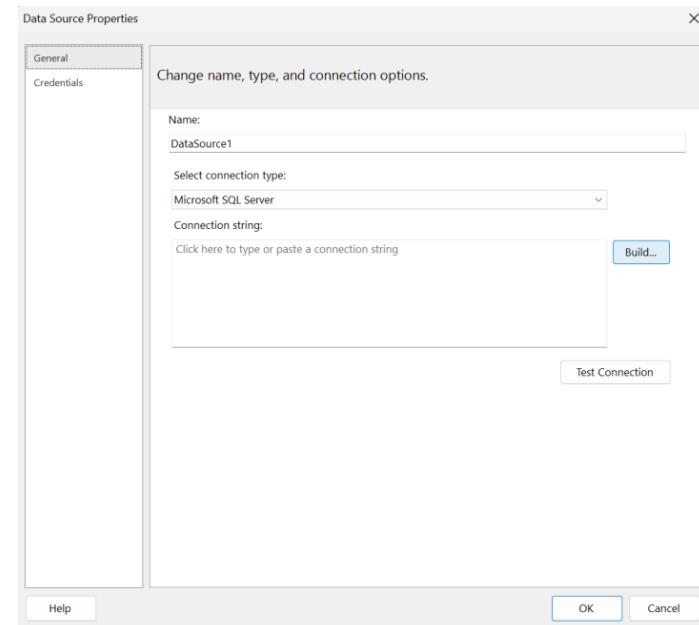
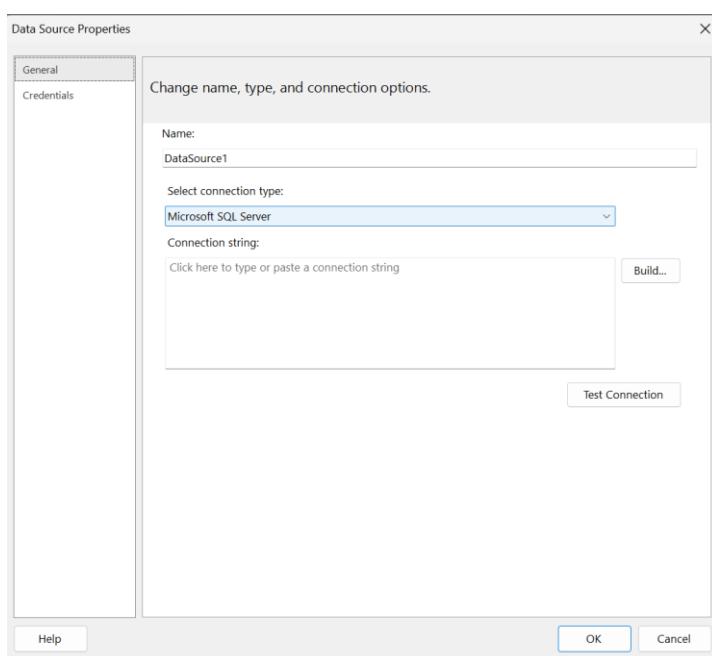
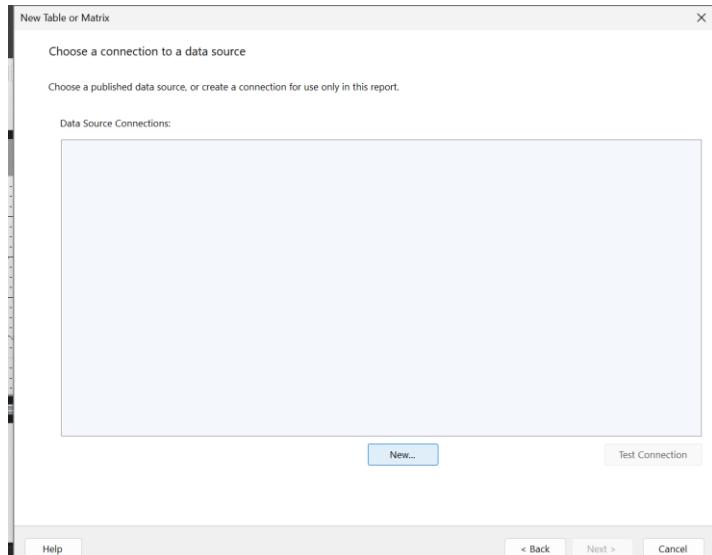
Therefore, please consider that we cleaned the data set using SQL Server.

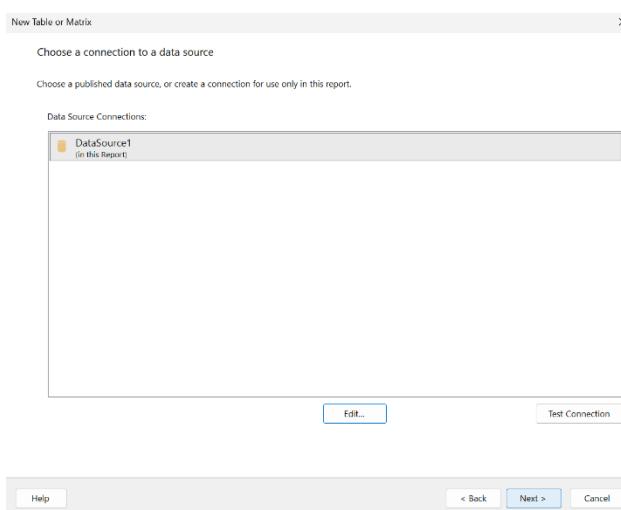
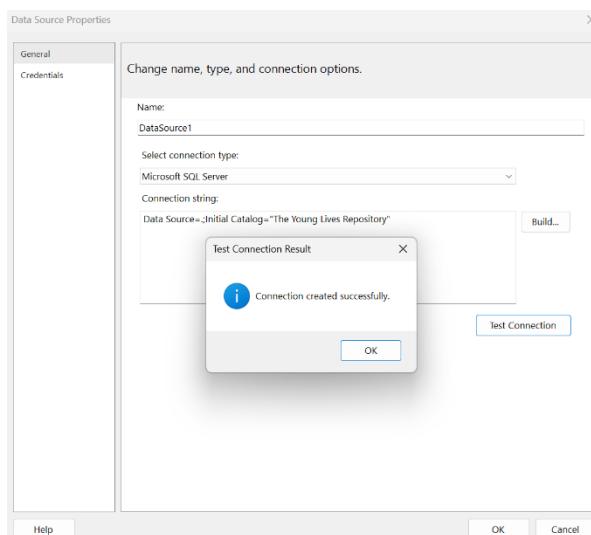
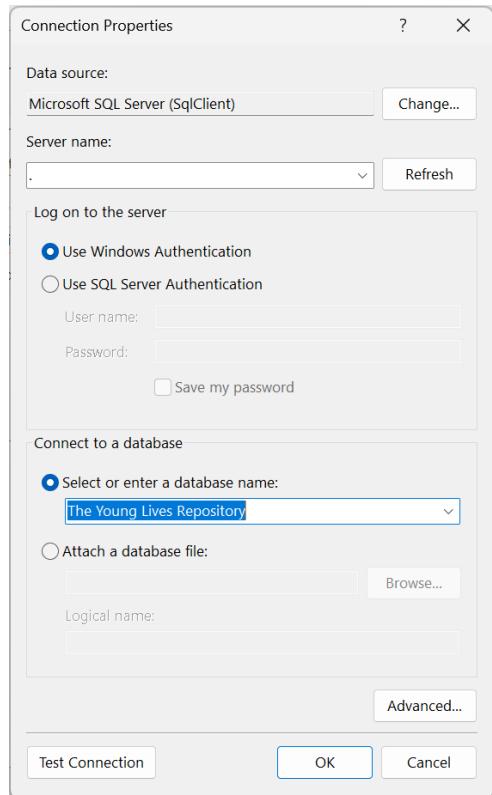
3) **Report Analysis – No. of Children in Peru & Vietnam**

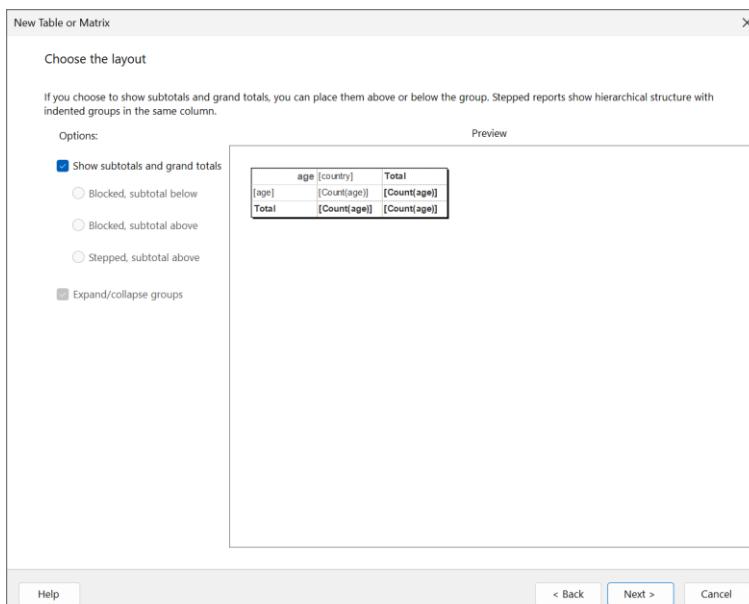
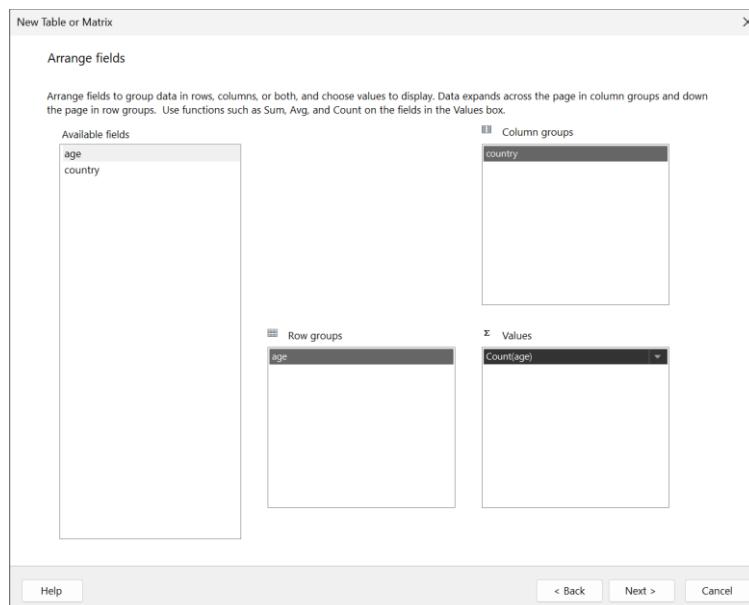
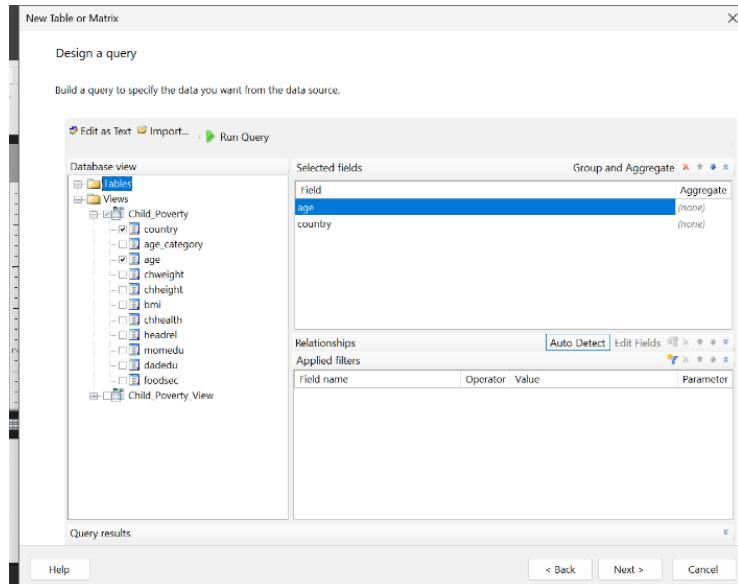
For create this, we used Power BI Report Builder. The following results show how the report was created.

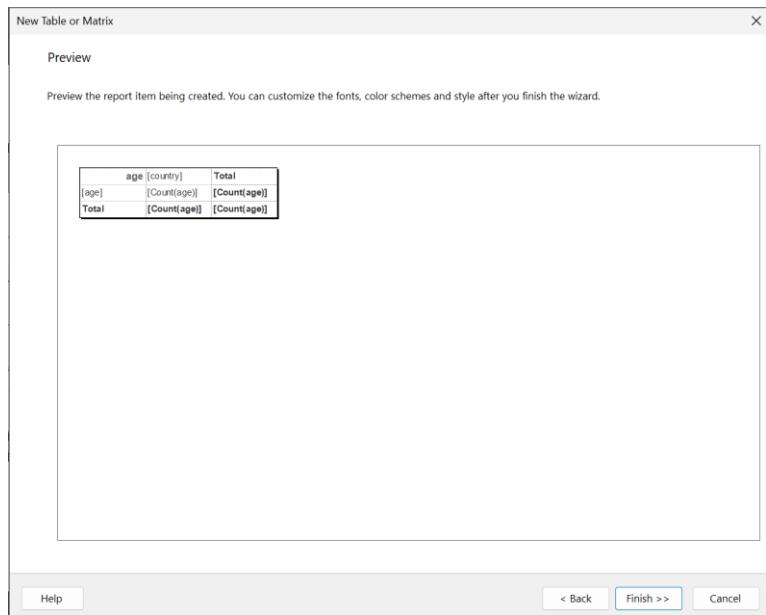
1. Open the Power BI Report Builder and follows.



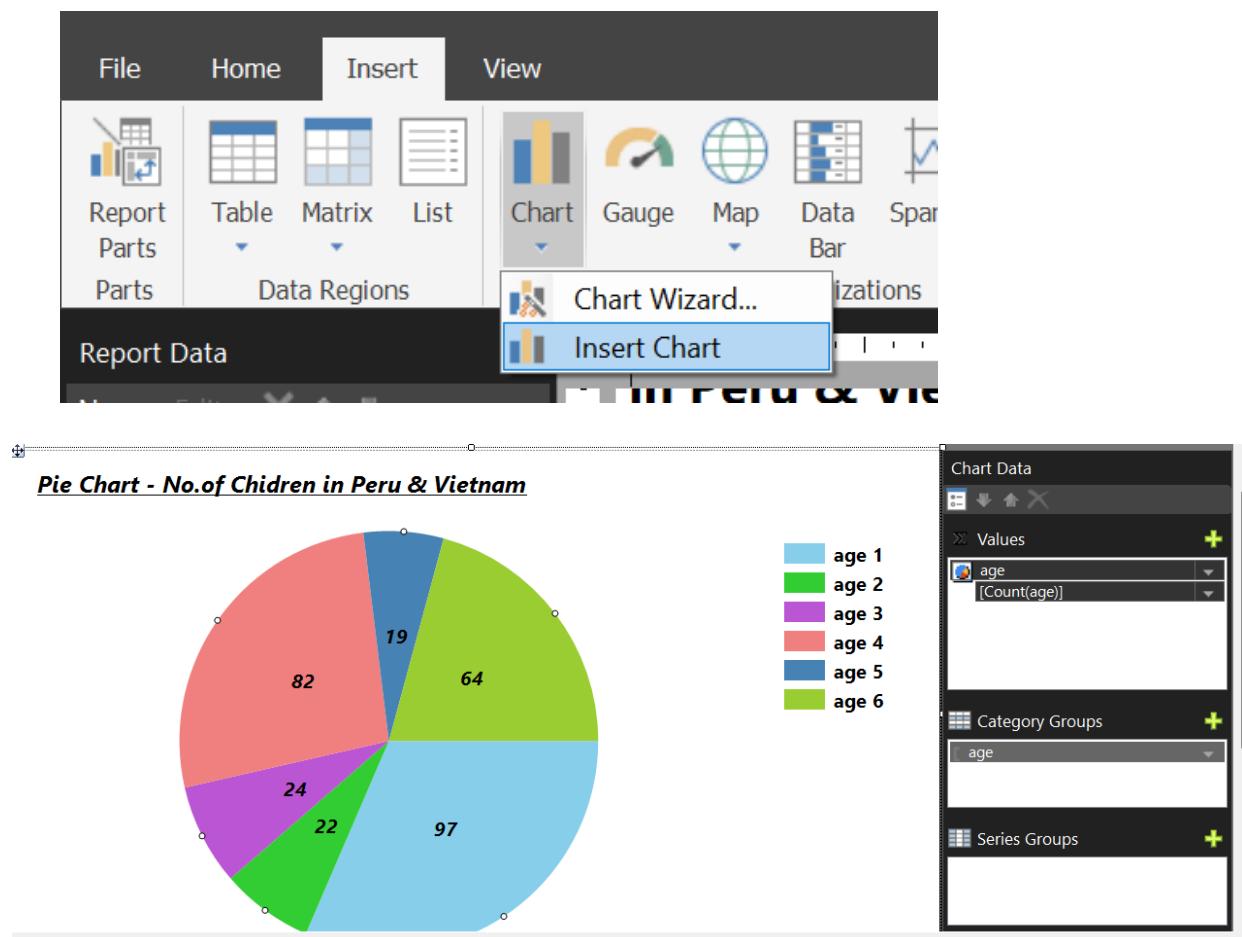




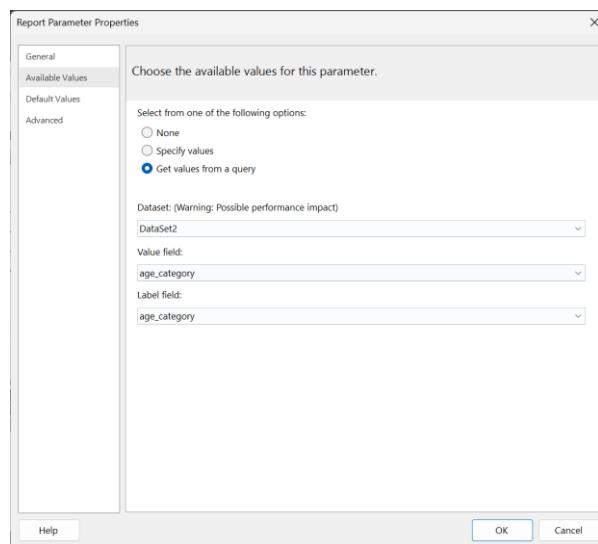
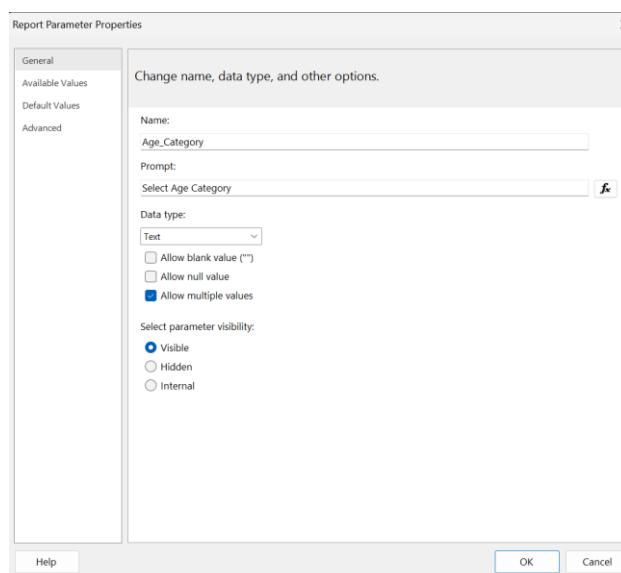
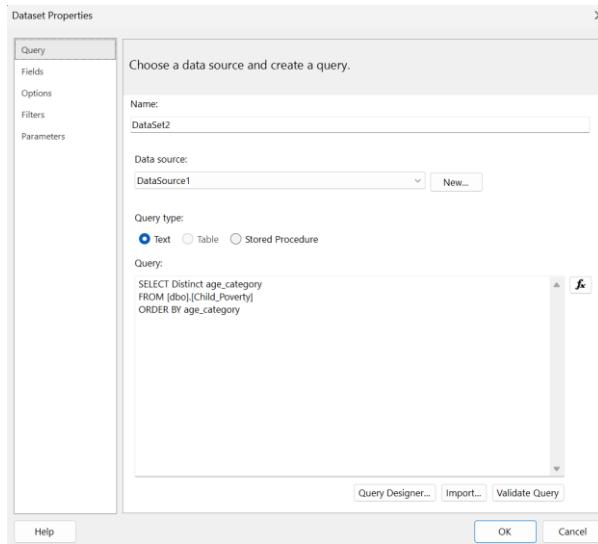


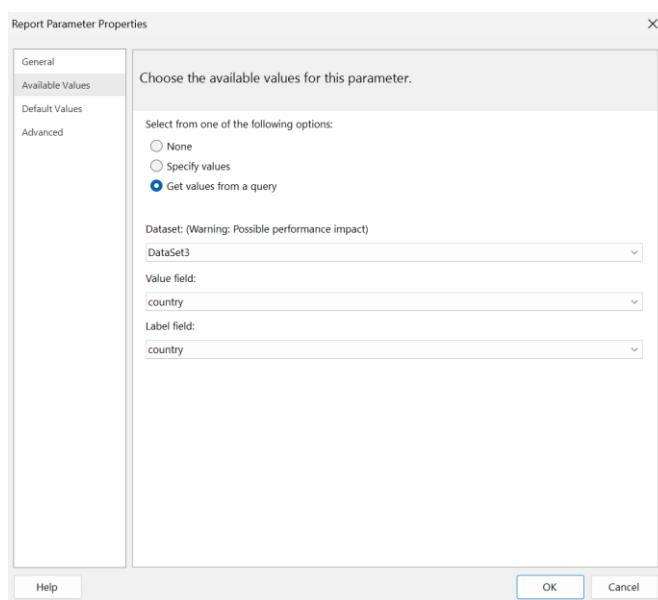
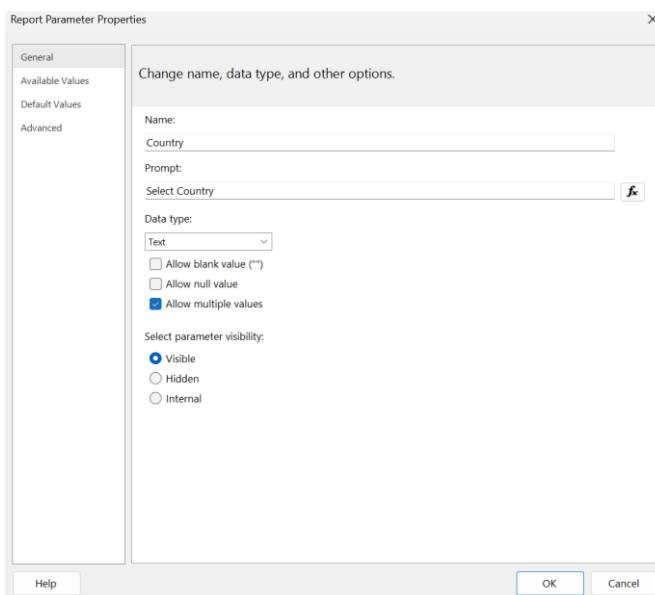
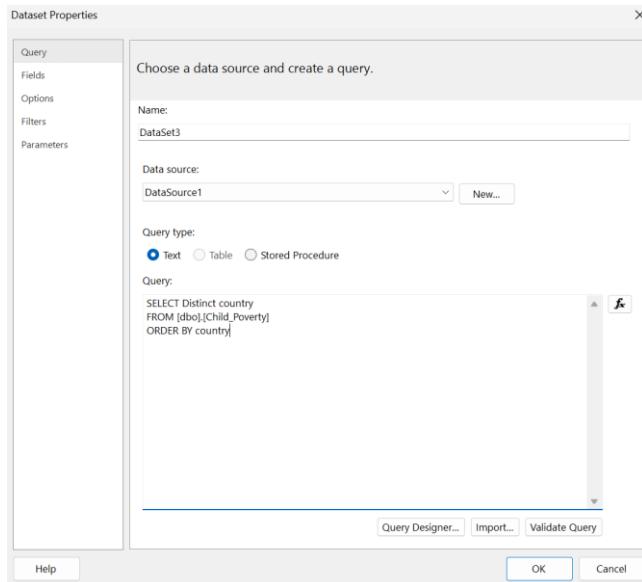


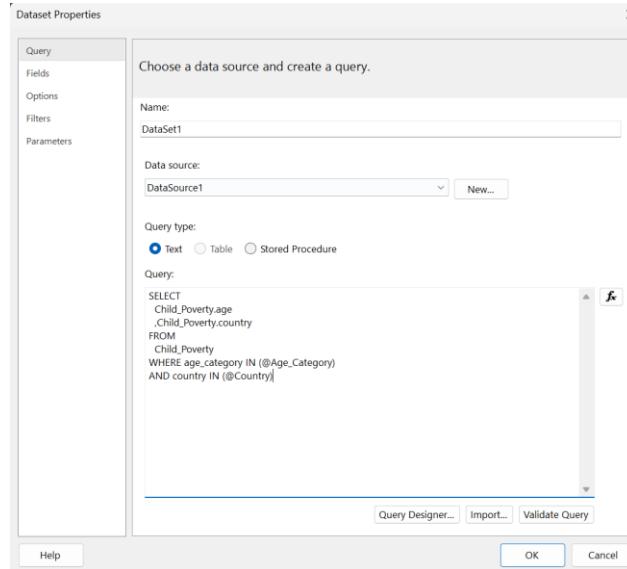
2. Insert Chart



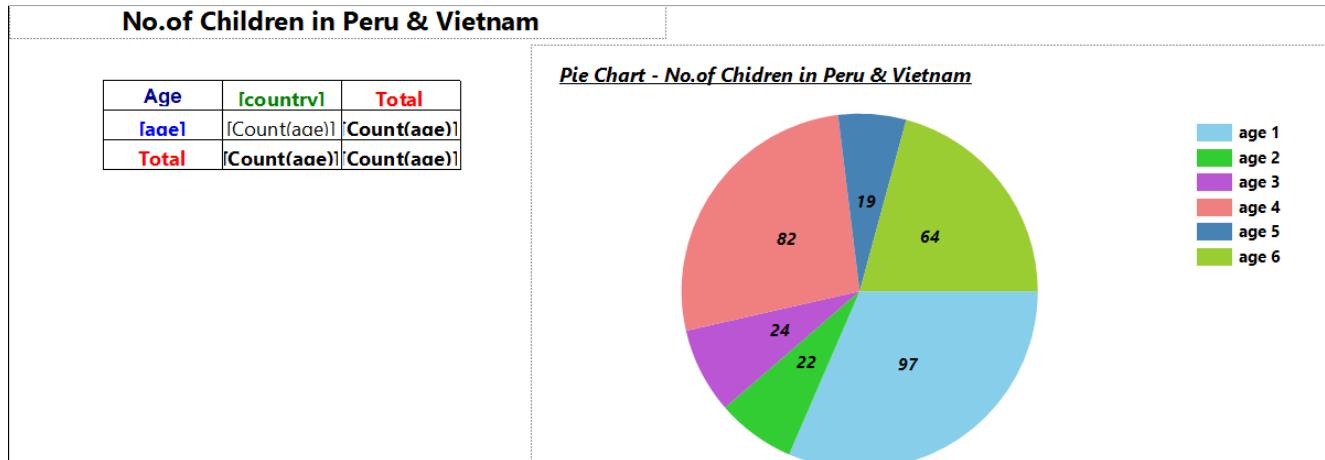
3. Create Parameters. (Age Category & Country)







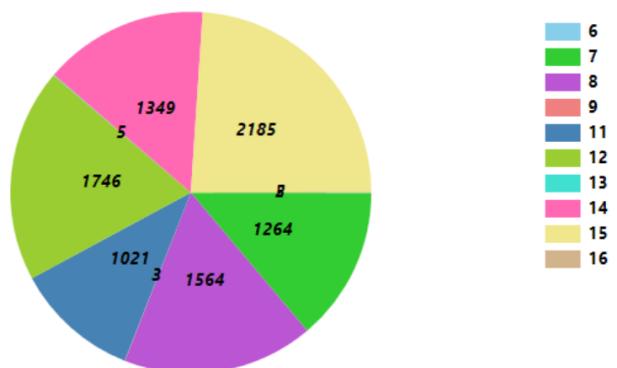
4. Final Output



No.of Children in Peru & Vietnam

Age	Peru	Vietnam	Total
6	0	2	2
7	673	591	1264
8	584	980	1564
9	0	3	3
11	678	343	1021
12	566	1180	1746
13	1	4	5
14	744	605	1349
15	655	1530	2185
16	0	5	5
Total	3901	5243	9144

Pie Chart - No.of Children in Peru & Vietnam



5. Analysis - No. of Children in Peru & Vietnam

The report, “No. of Children in Peru & Vietnam” depicts the number of children for ages 6 to 16 years has been stated. It shows that Vietnam has more children in all age groups than Peru. The total number of children in Peru is 3,901 and in Vietnam it is 5,243.

The pie chart shows the breakdown of total number of children in Peru & Vietnam for each age group. By this visual analysis, it is easier and quick to note the demographic distribution and contrast child population between the two countries.

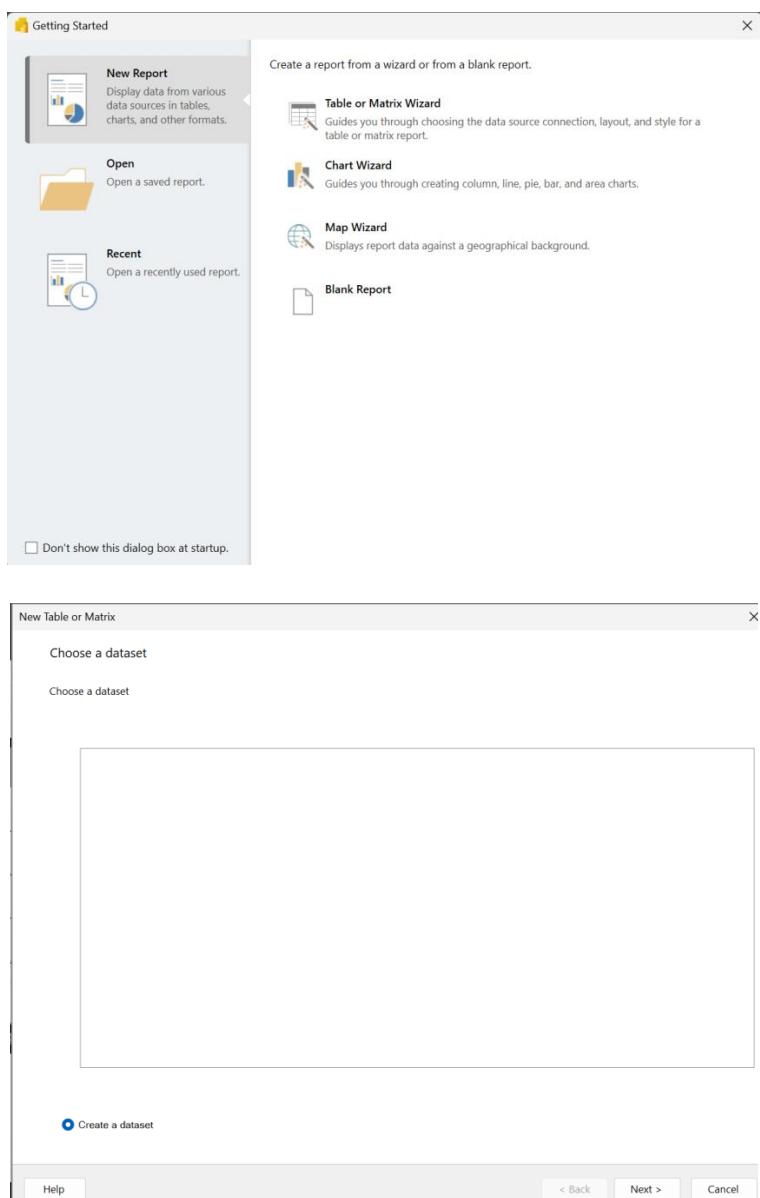
All these details can be helpful in different ways, like organizing educational materials, healthcare options and other programs relating to children. The age distribution also points out the certain ages that seem to need more focus or funds in each of the states.

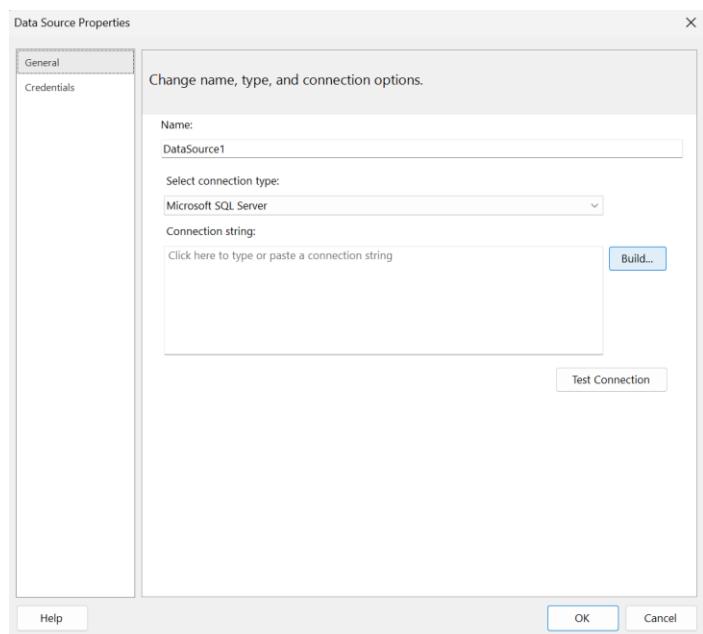
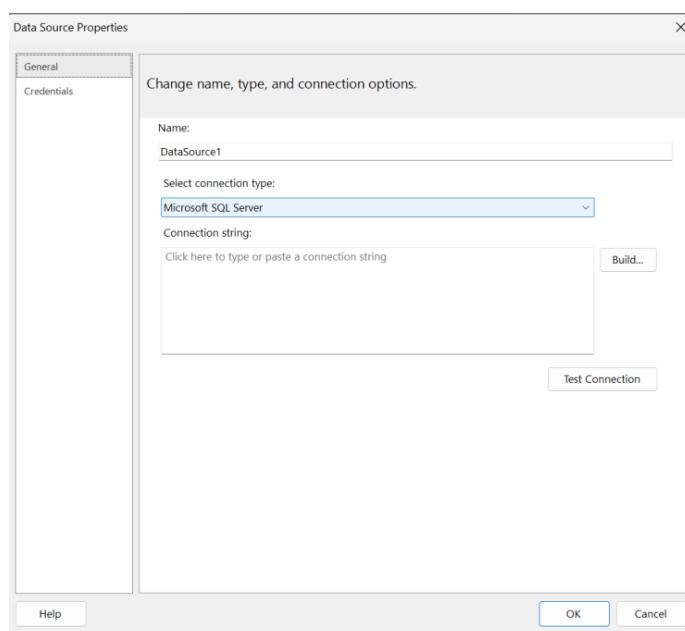
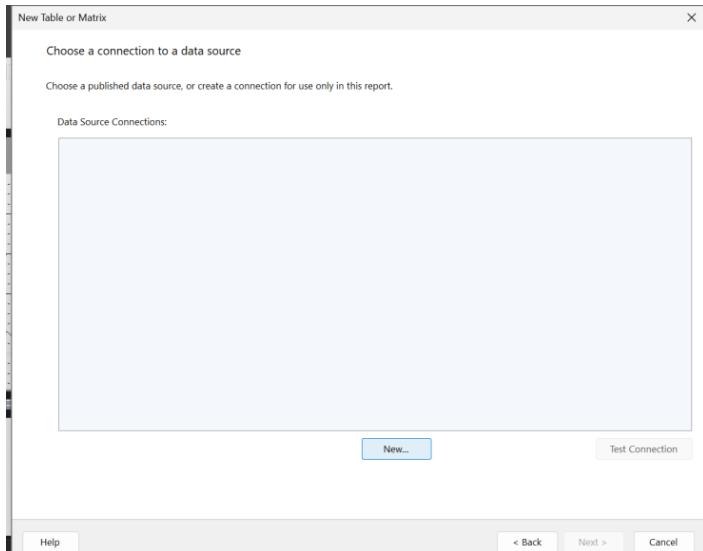
In conclusion, this presents a lot of light in the demographic pattern regarding children in Peru and Vietnam towards arriving at relevant decisions and resource utilization.

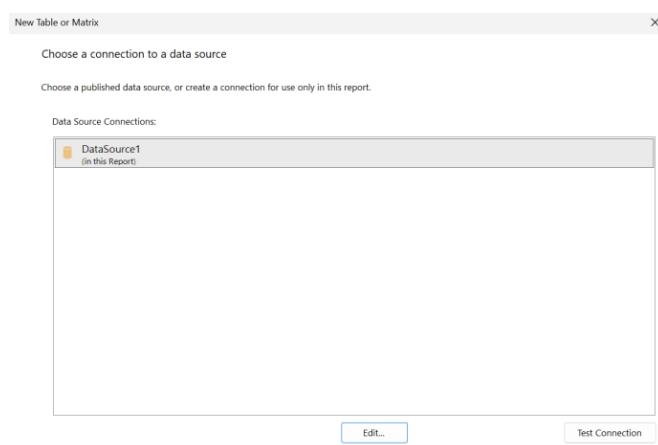
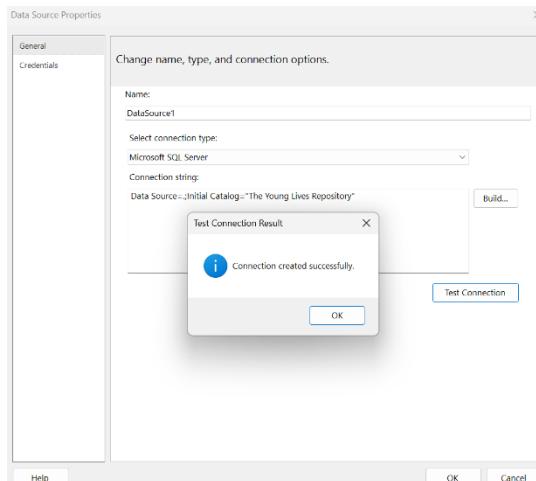
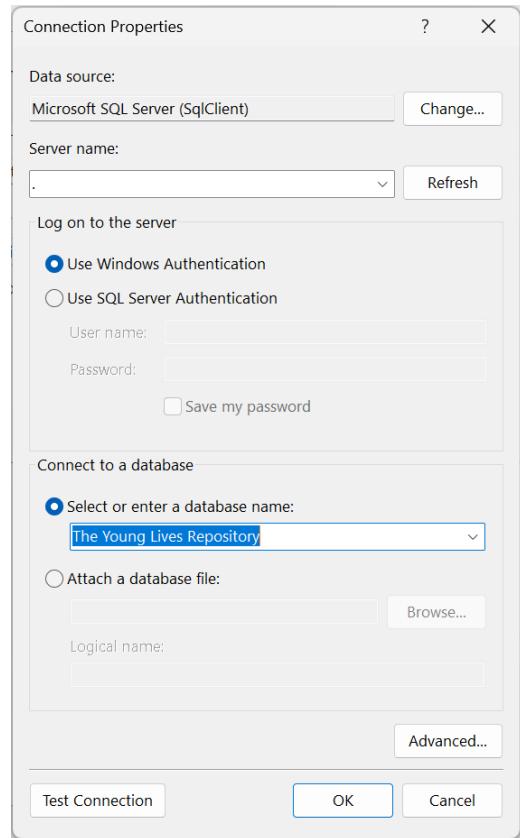
4) Report Analysis - Children's Average Weight against their Average Height by Age in Peru & Vietnam

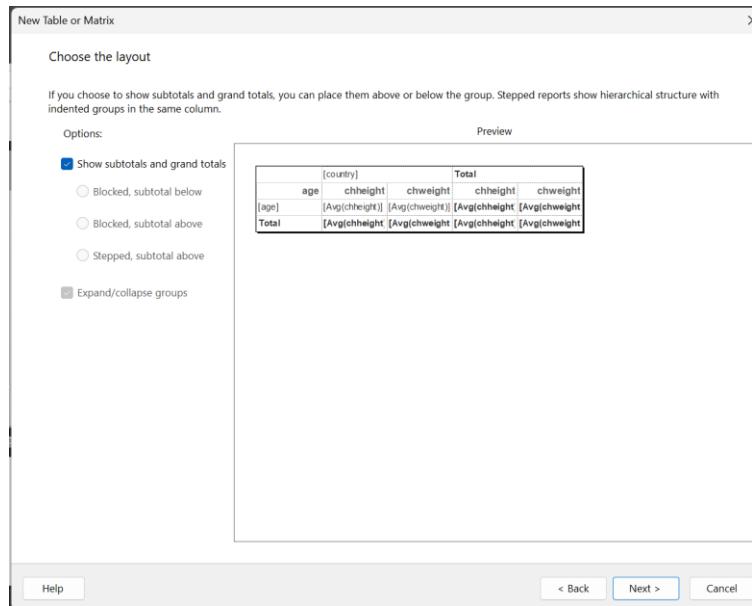
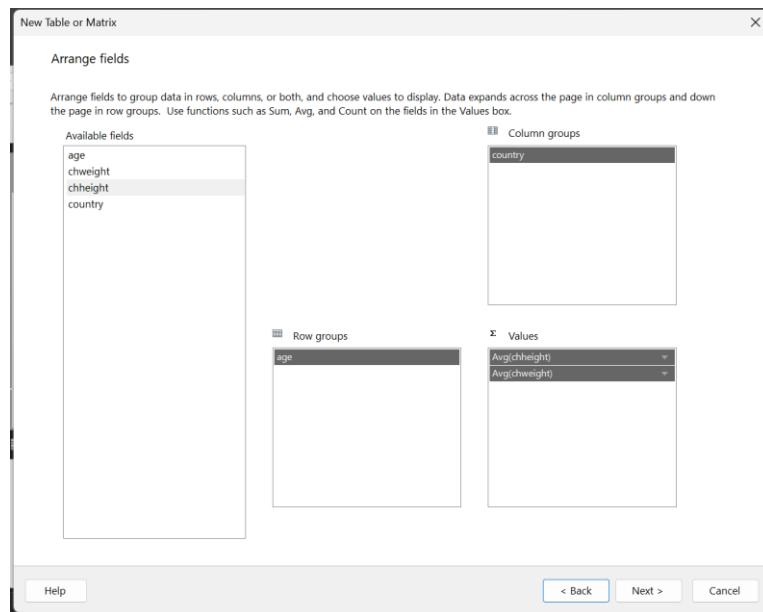
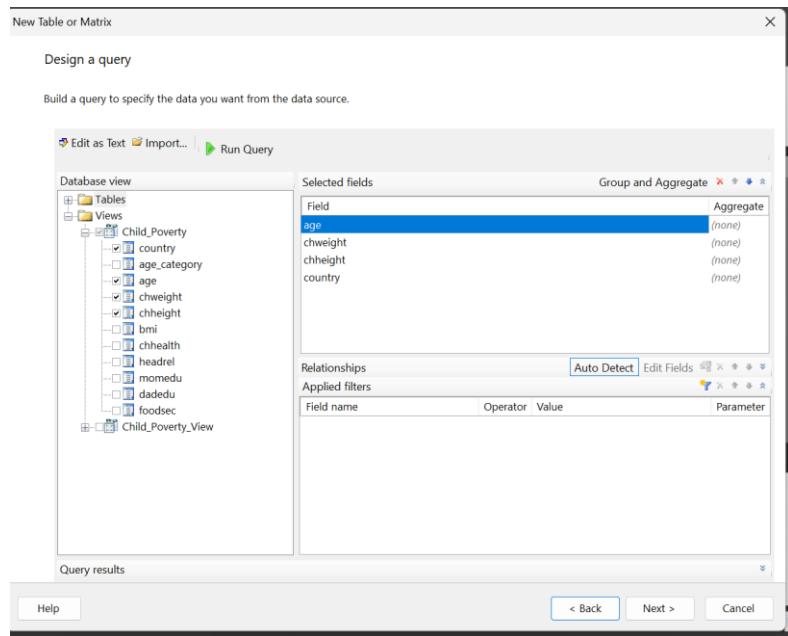
For create this, we used Power BI Report Builder. The following results show how the report was created.

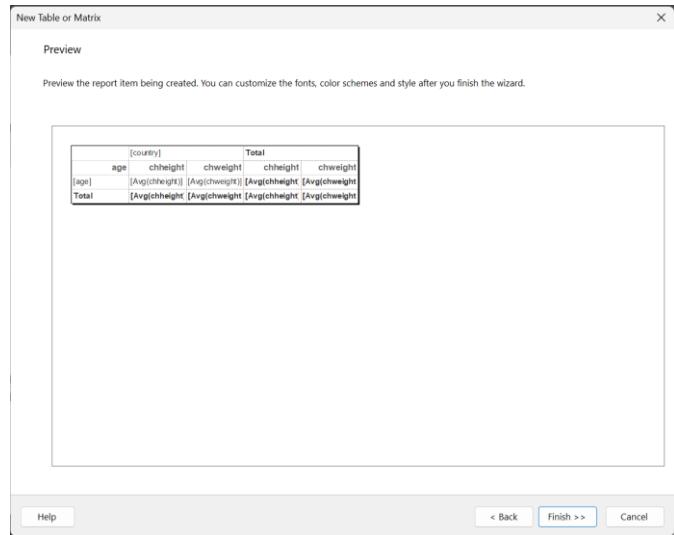
1. Open the Power BI Report Builder and follows.



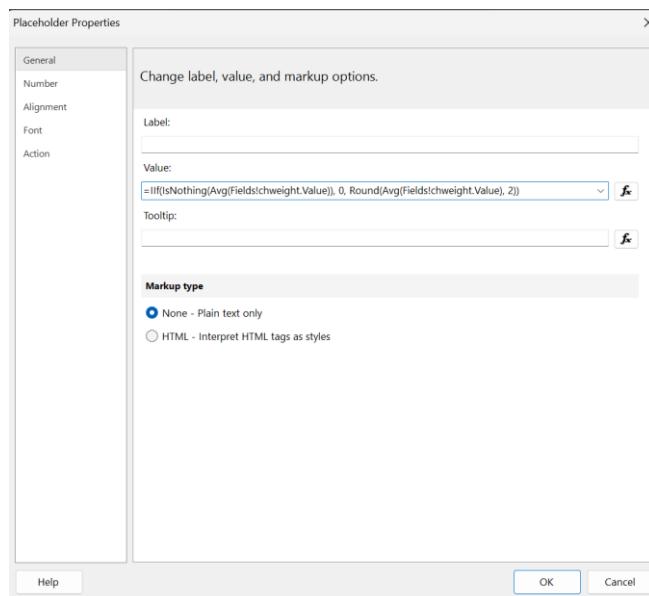
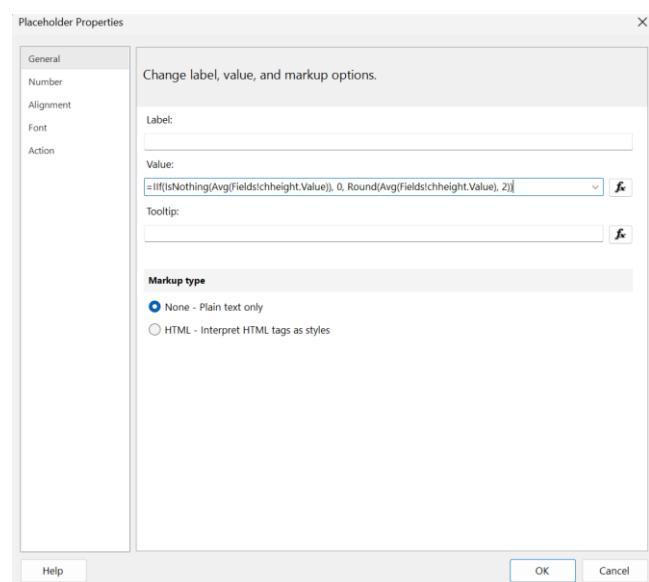




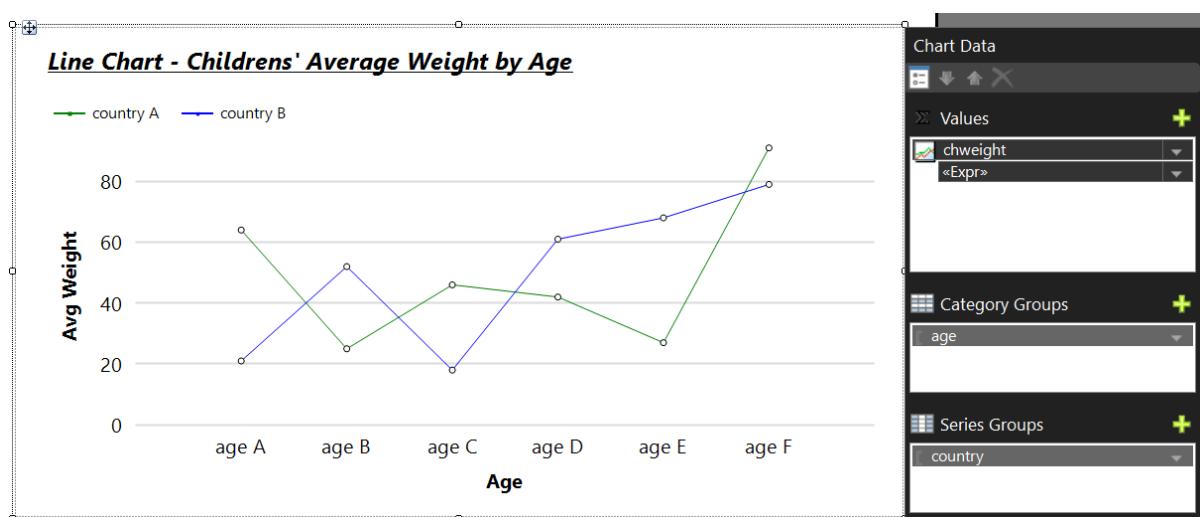
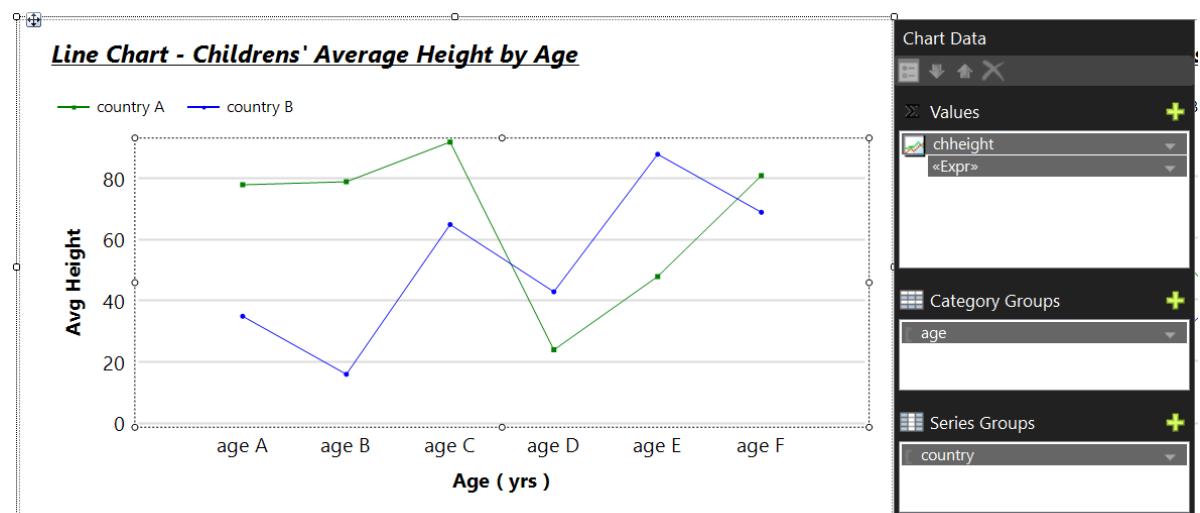
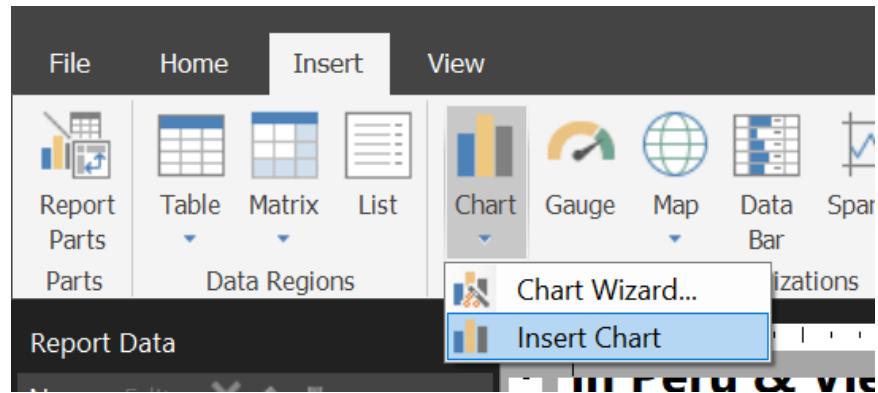




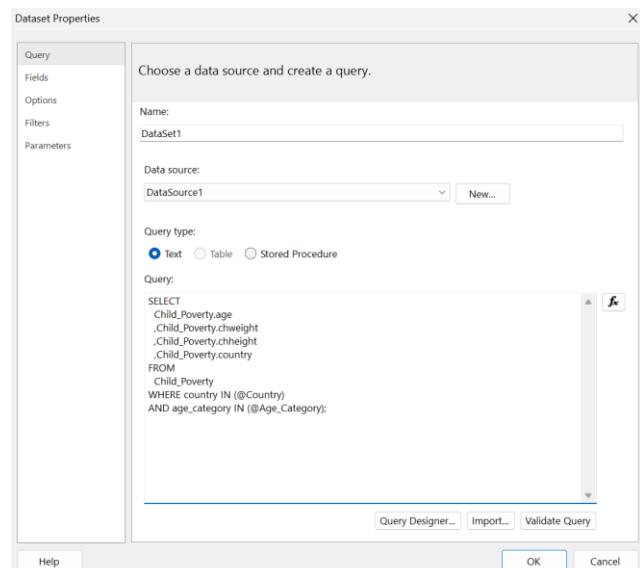
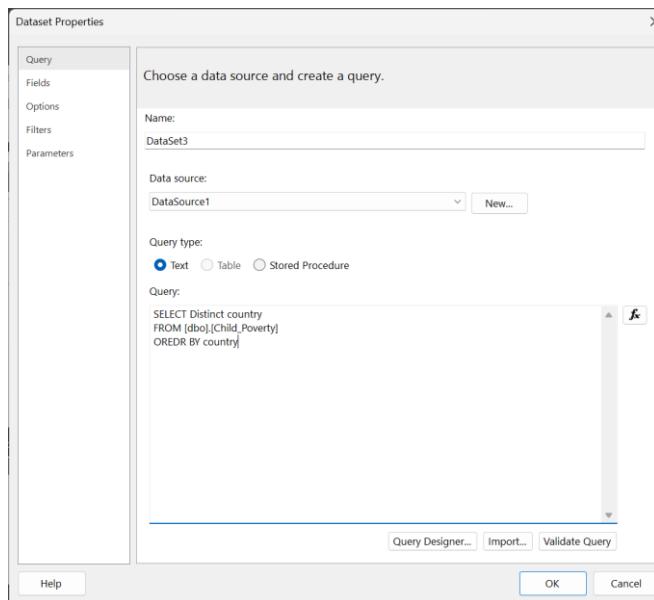
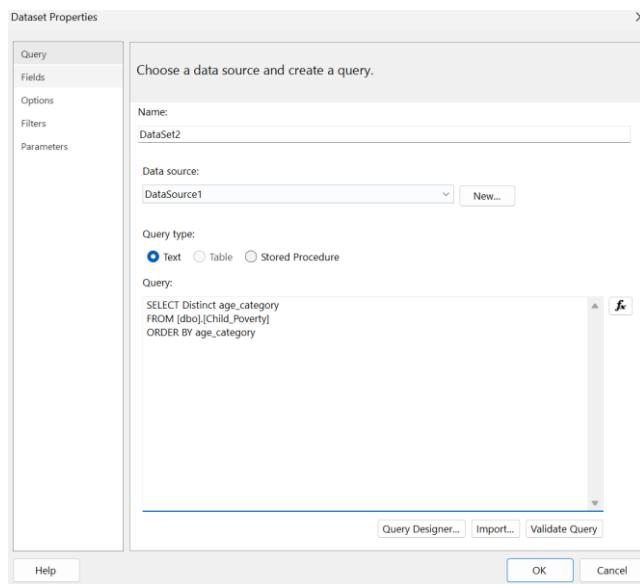
2. Create Expressions to get rounded average values.

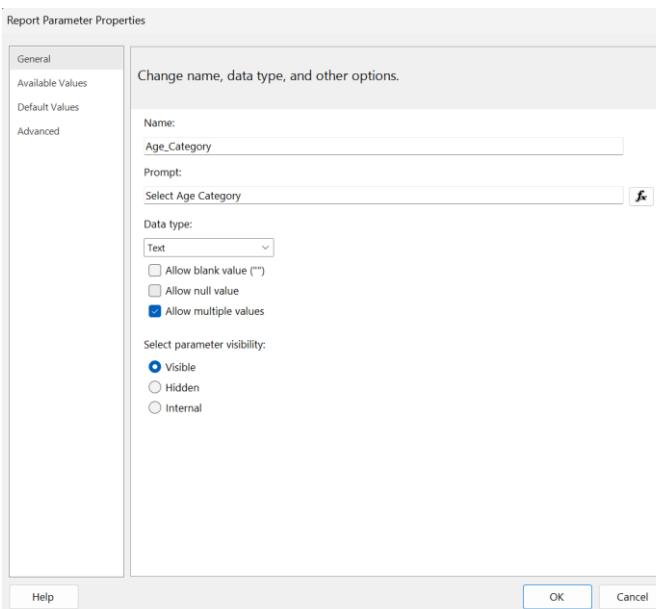
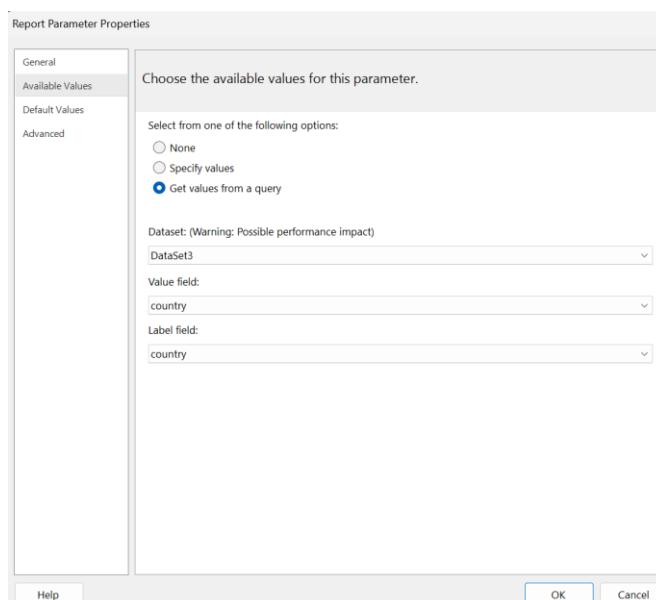
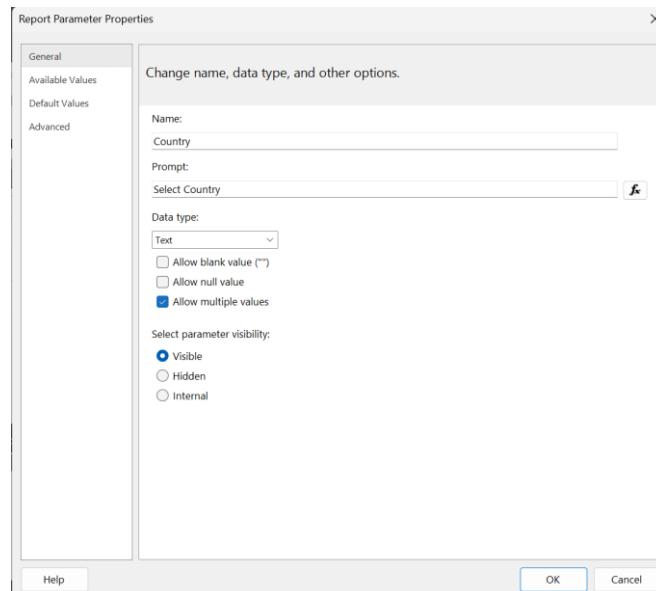


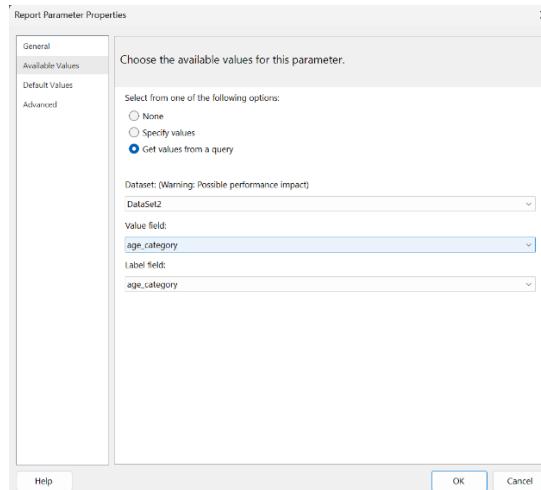
3. Insert Chart



4. Create Parameters (Age Category & Country)







5. Final Output

Childrens' Average Weight against their Average Height by Age in Peru & Vietnam

	[country]		
Age	Avg Height	Avg Weight	
[age]	«Expr»	«Expr»	

Line Chart - Childrens' Average Height by Age

Age	Country A (Avg Height)	Country B (Avg Height)
age A	~80	~35
age B	~80	~20
age C	~85	~60
age D	~25	~45
age E	~55	~85
age F	~80	~65

Line Chart - Childrens' Average Weight by Age

Age	Country A (Avg Weight)	Country B (Avg Weight)
age A	~65	~25
age B	~25	~55
age C	~45	~20
age D	~45	~60
age E	~25	~65
age F	~85	~75

Select Age Category
11-15, 16+, 6-10
Select Country
Peru, Vietnam

Childrens' Average Weight against their Average Height by Age in Peru & Vietnam

Age	Peru		Vietnam	
	Avg Height	Avg Weight	Avg Height	Avg Weight
6	0	0	118.35	23.45
7	119.14	24.05	120.2	21.95
8	122.3	25.73	122.86	22.92
9	0	0	121.57	24.3
11	141.54	39.15	142.96	35.33
12	144.28	40.71	145.47	36.42
13	152.6	44.9	154.25	46.72
14	154.86	48.98	156.08	44.53
15	155.88	49.98	157.62	45.89
16	0	0	165.4	53.88

Line Chart - Childrens' Average Height by Age

Age	Peru (Avg Height)	Vietnam (Avg Height)
7	~120	~120
9	~20	~120
11	~140	~140
12	~145	~145
14	~150	~150
16	~20	~160

Line Chart - Childrens' Average Weight by Age

Age	Peru (Avg Weight)	Vietnam (Avg Weight)
7	~25	~25
9	~20	~25
11	~30	~25
12	~35	~35
14	~40	~40
16	~5	~55

6. Analysis - Children's Average Weight against their Average Height by Age in Peru & Vietnam

The first two tables present the average height and weight of students in respective age ranges children within 6-16 years in Peru as well as Vietnam. The detailed data is rather helpful for tracing the tendencies of children's growth in those countries.

The first line chart on the left has the average height of children from the ages of 6-16 for both countries. The chart on the right shows the average weight for those age brackets. These charts facilitate early examination of trends and difference in the growth patterns among children in Peru and Vietnam.

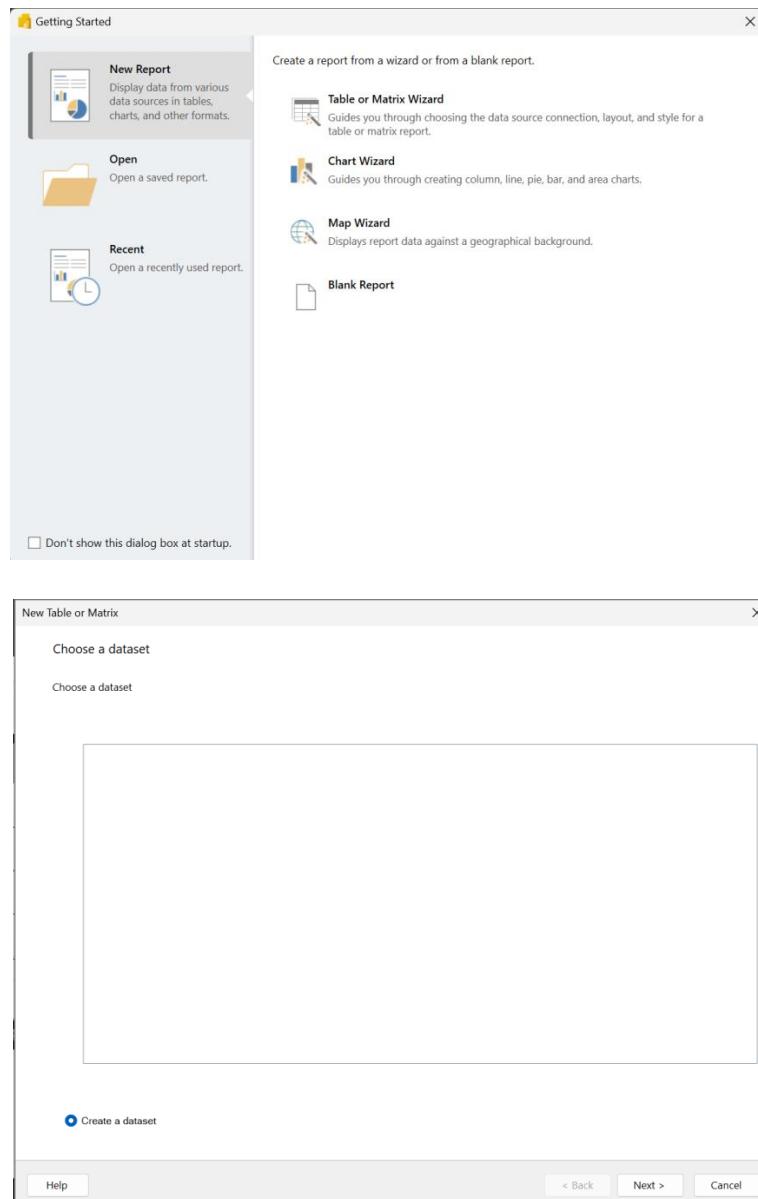
It has been learned from the data that there is variation in height and weight mean of children that exist between Peru and Vietnam. These differences could also be due to genetic, nutritional and broad health factors that exist in each country. Knowledge of these patterns is particularly helpful in designing effective health and nutritional support interventions for children in both countries.

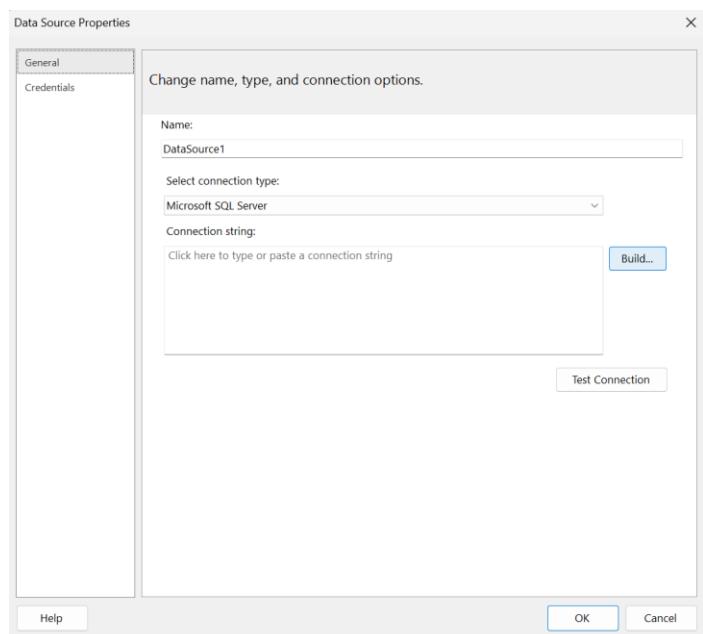
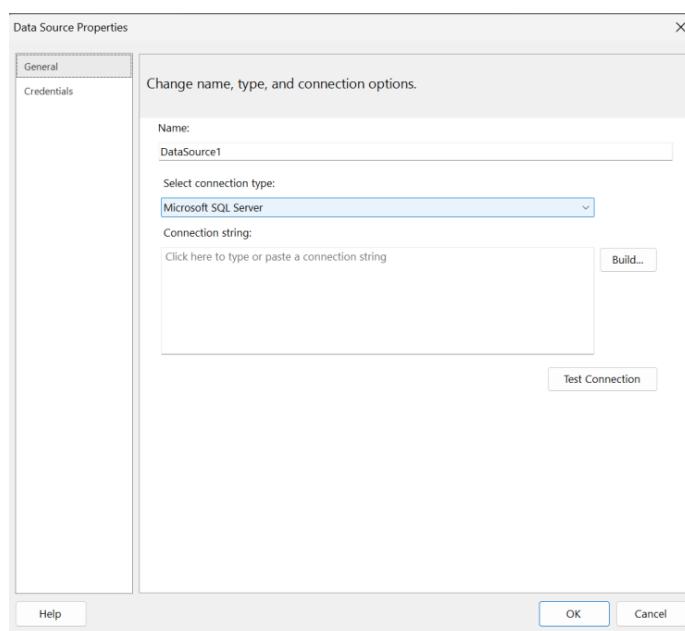
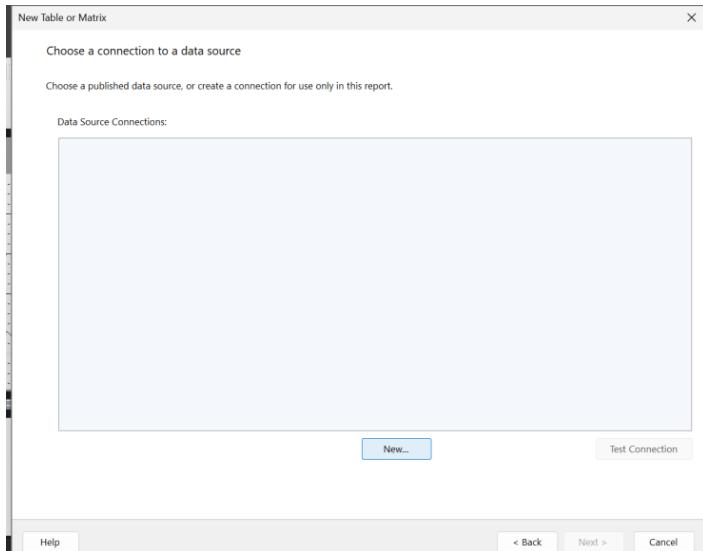
In conclusion, the present work gives vital data about the increasing trends in children of Peru and Vietnam useful for proper decision making for their development concerning their health and nutrition, quality education, etc.

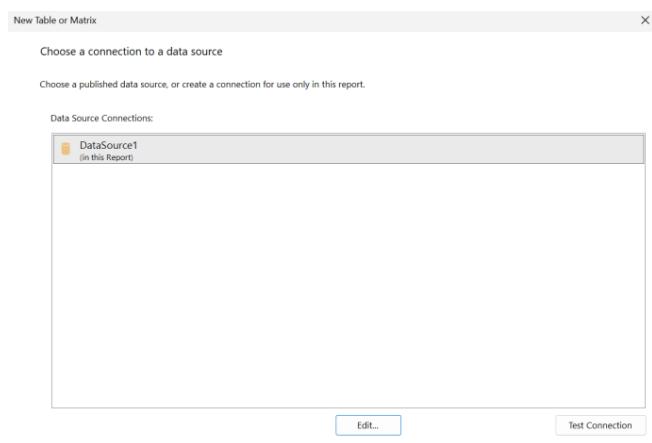
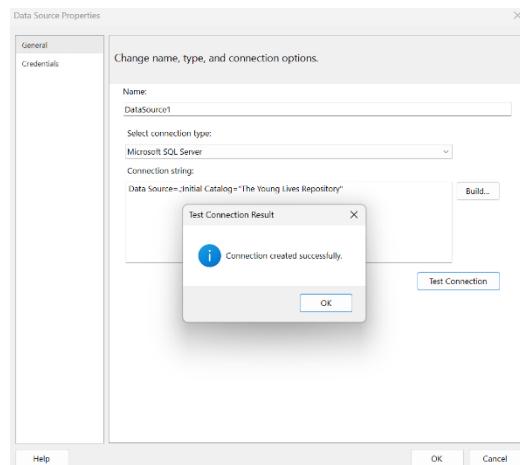
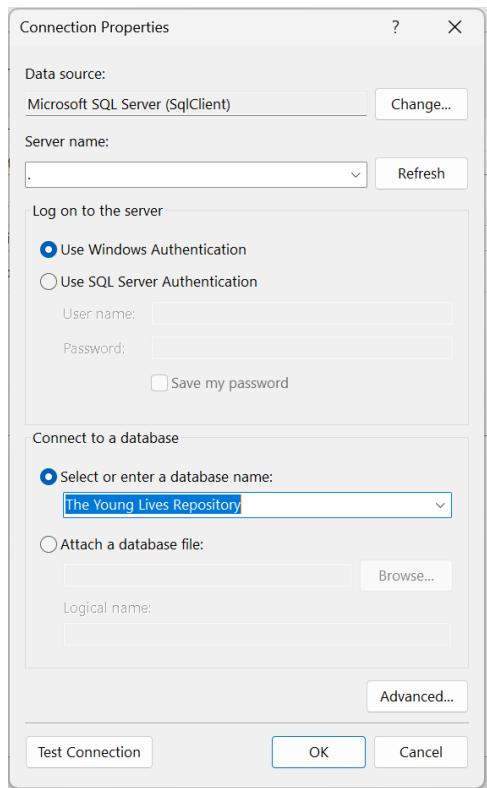
5) Report Analysis - Summary of BMI value with Age in Peru & Vietnam

For create this, we used Power BI Report Builder. The following results show how the report was created.

1. Open the Power BI Report Builder and follows.







New Table or Matrix

Design a query

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

Selected fields

Field	Group and Aggregate	Aggregate
age	(none)	(none)
bmi	(none)	(none)
country	(none)	(none)

Relationships

Applied filters

Query results

Help < Back Next > Cancel

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

- country
- age
- bmi

Column groups

- country

Row groups

- age

Values

- Max(bmi)
- Min(bmi)
- Avg(bmi)

Help < Back Next > Cancel

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

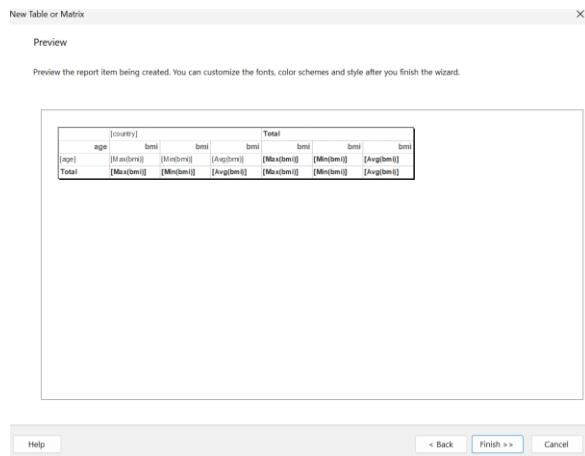
Stepped, subtotal above

Expand/collapse groups

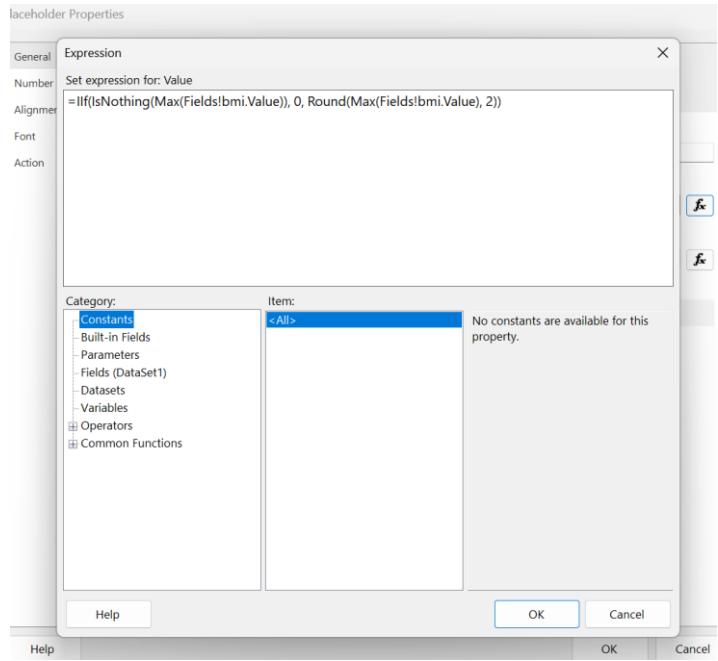
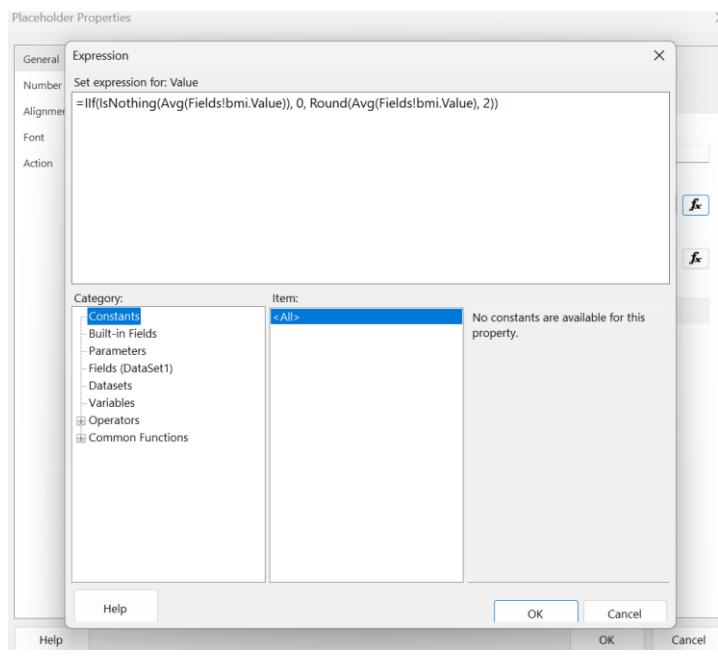
Preview

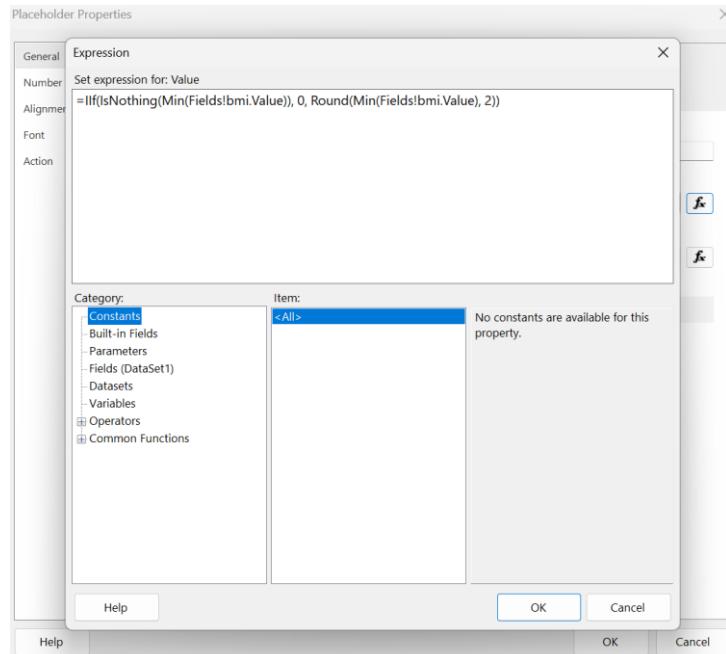
	[country]			Total		
age	bmi	bmi	bmi	bmi	bmi	bmi
[age]	[Max(bmi)]	[Min(bmi)]	[Avg(bmi)]	[Max(bmi)]	[Min(bmi)]	[Avg(bmi)]
Total	[Max(bmi)]	[Min(bmi)]	[Avg(bmi)]	[Max(bmi)]	[Min(bmi)]	[Avg(bmi)]

Help < Back Next > Cancel

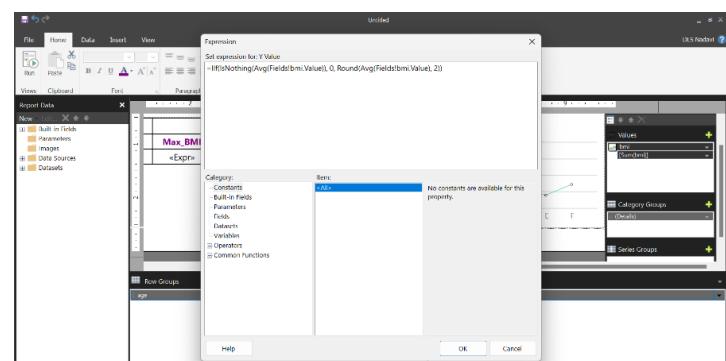
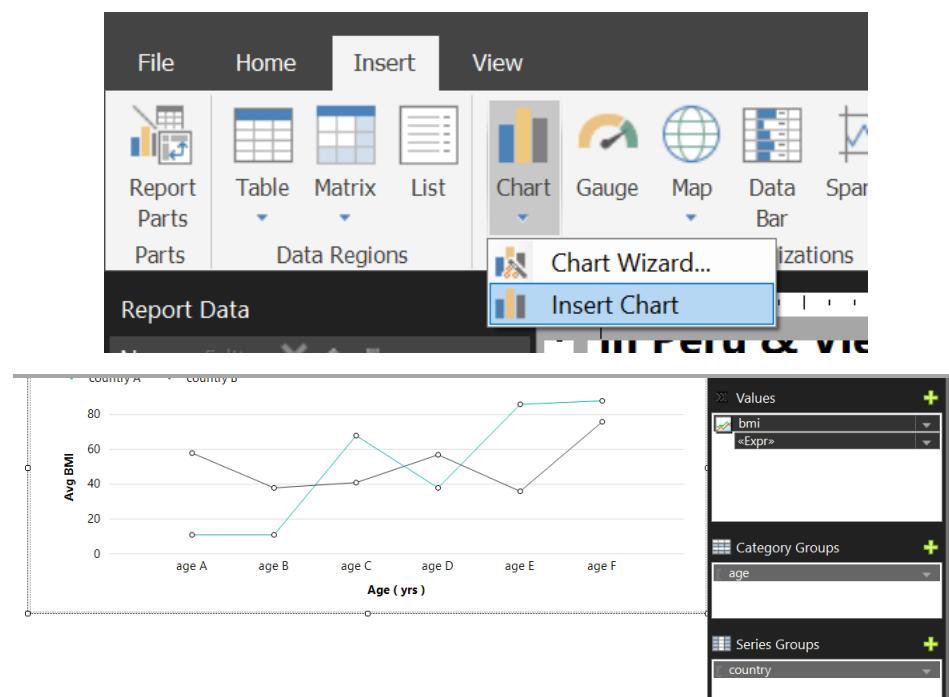


2. Create Expressions to get rounded max, min and average values.

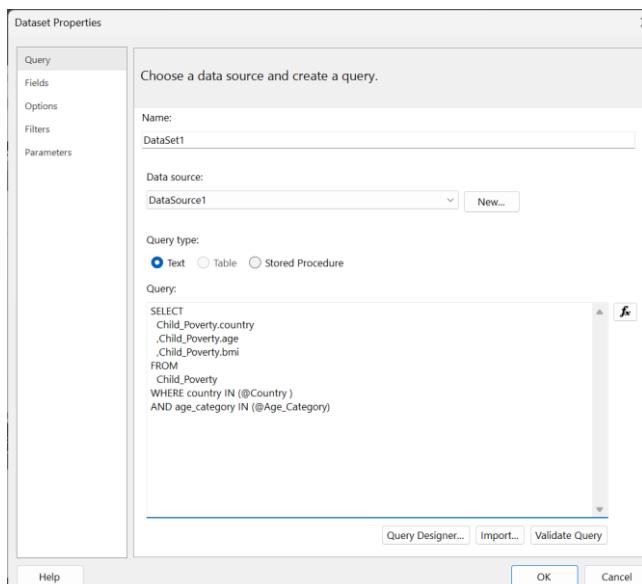
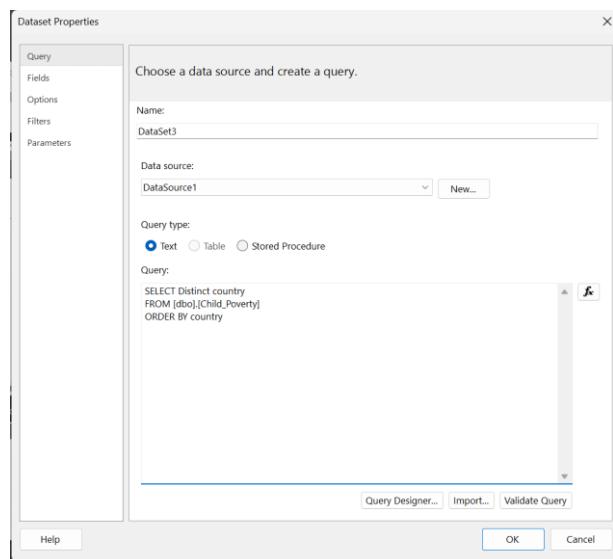
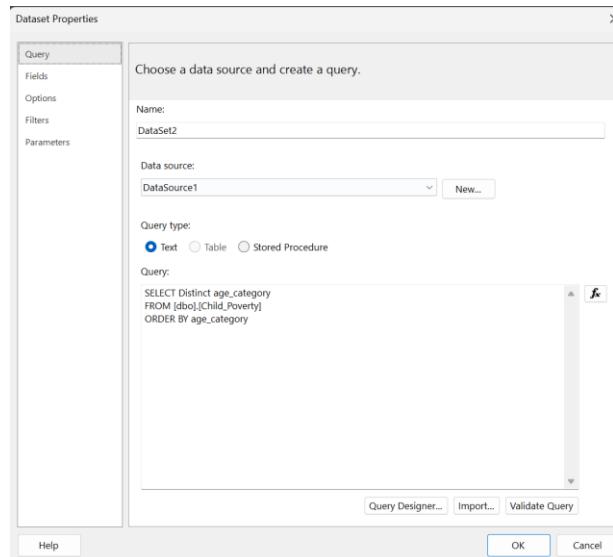


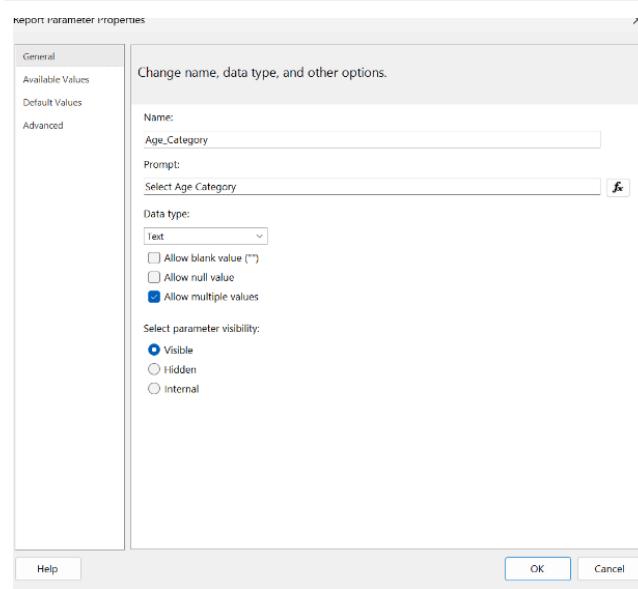
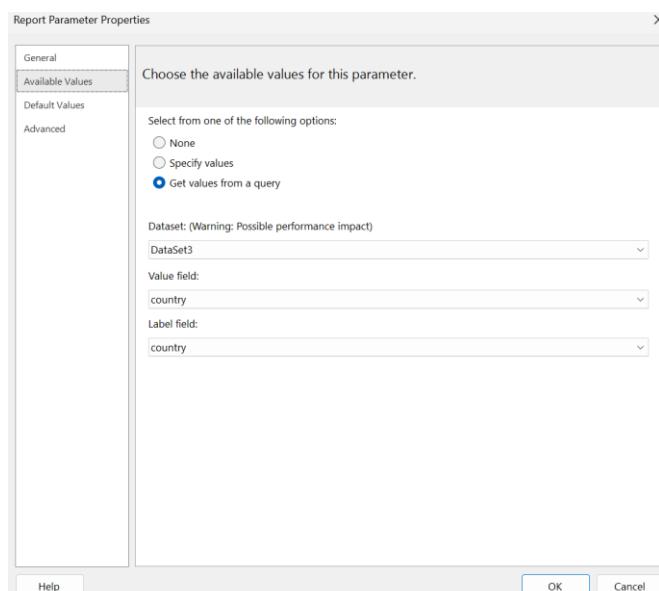
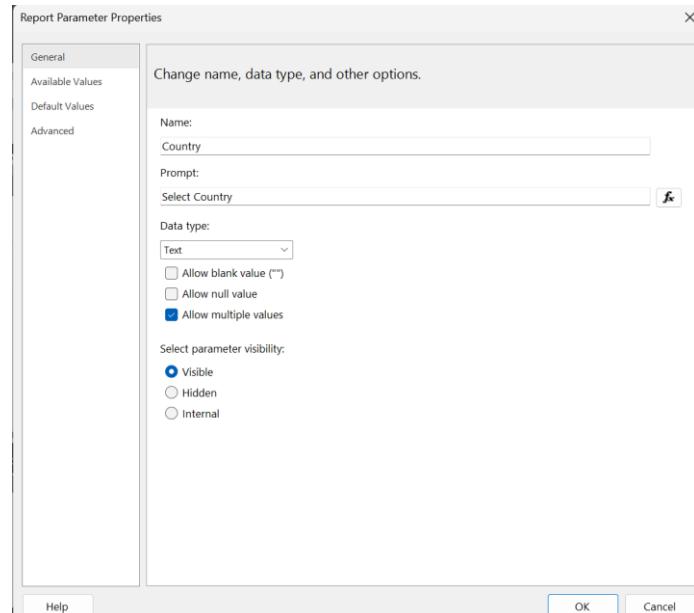


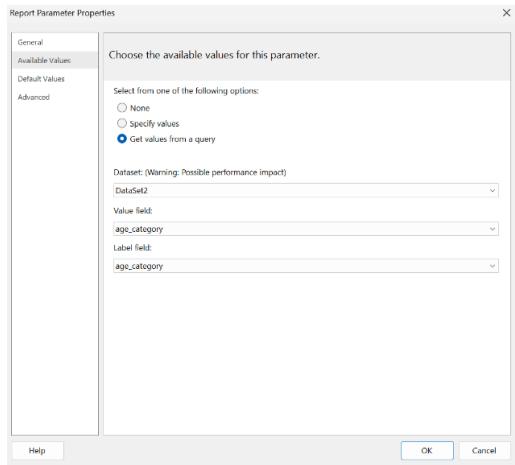
7. Insert Chart



3. Create Parameters. (Age Category & Country)







4. Final Output

Summary of BMI value with Age in Peru & Vietnam

[country]			
Age	Max_BMI	Min_BMI	Avg_BMI
[age]	«Expr»	«Expr»	«Expr»

Line Chart - Average BMI value with Age

Select Country: Peru, Vietnam Select Age Category: 11-15, 16+, 6-10

Summary of BMI value with Age in Peru & Vietnam

Age	Peru			Vietnam		
	Max_BMI	Min_BMI	Avg_BMI	Max_BMI	Min_BMI	Avg_BMI
6	0	0	0	18.63	14.69	16.66
7	29.11	10.2	16.84	25.24	10.91	15.09
8	33.3	11.31	17.07	26.51	8.11	15.09
9	0	0	0	18.09	14.35	16.04
11	29.54	12.84	19.4	27.09	11.64	17.13
12	28.84	13.88	19.44	29.59	12.07	17.09
13	19.28	19.28	19.28	21.91	16.17	19.44
14	26.76	12.98	20.42	29.45	12.56	18.25
15	27.74	14.79	20.58	96.9	13.27	18.5
16	0	0	0	22.68	17.21	19.85

Line Chart - Average BMI value with Age

5. Analysis - Summary of BMI value with Age in Peru & Vietnam

The table indicates the BMI among the children of 6 – 16 years in both Peru and Vietnam. For each age group, it classifies the maximum , minimum and average BMI values. This makes non-parametric comparisons of BMI between the two countries regarding children's growth and nutritional status.

The line chart presents the average BMI values of children of age group of 6 to 16 of Peru and Vietnam. On the x-coordinate, age is measured, and on the y-coordinate, the mean BMI values are indicated.

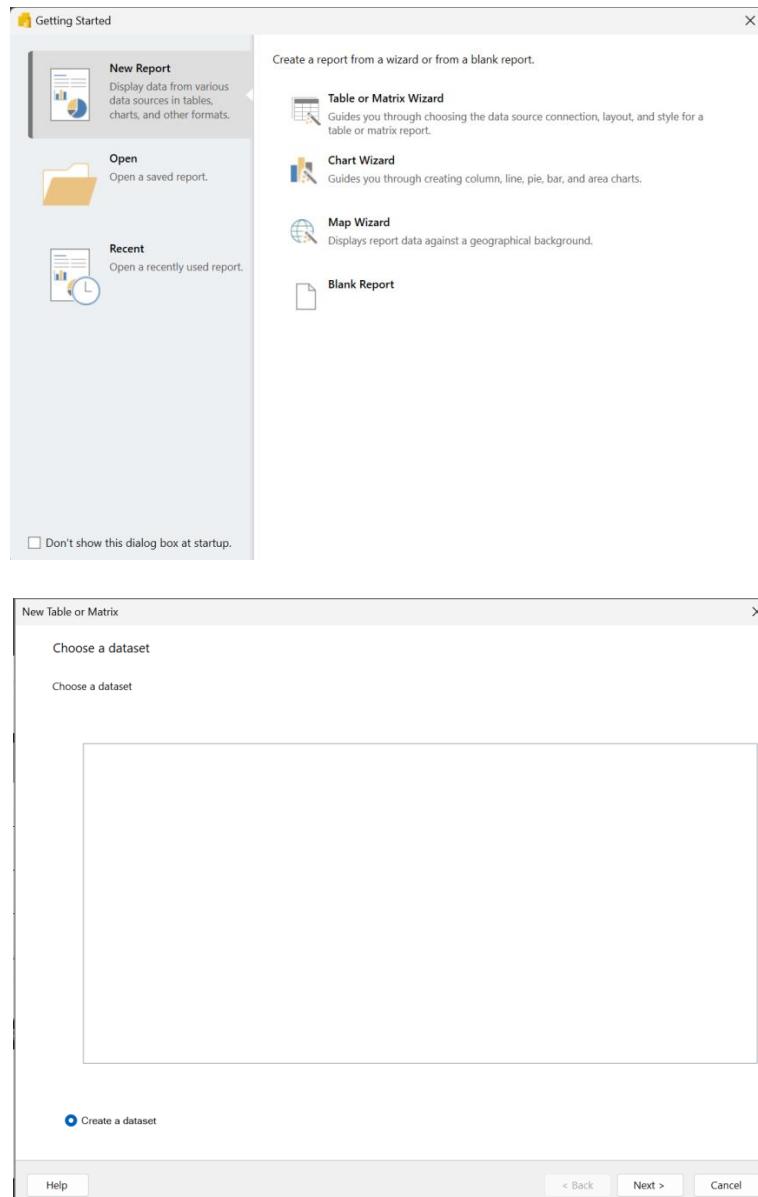
The differences in here could be due to differences in genetics, diet, social class and health care facilities. It is important to understand these patterns to properly tailor healthcare and nutrition-based programs for children of these countries. For example, if the calculated average BMI values in one country are smaller, respectively, it speaks about the imperative need for better nutritional approaches.

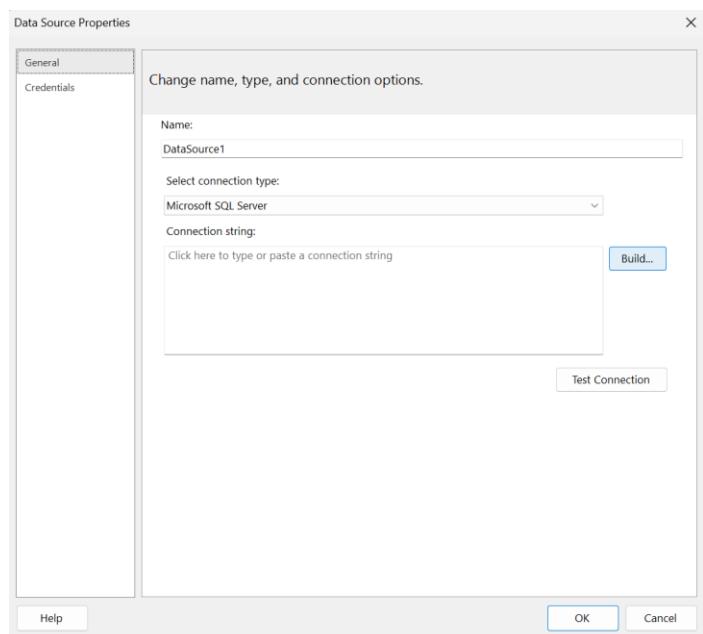
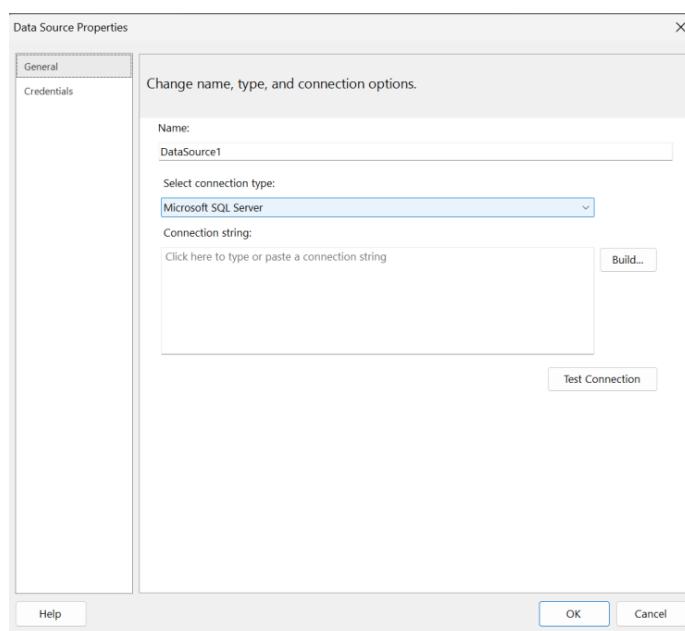
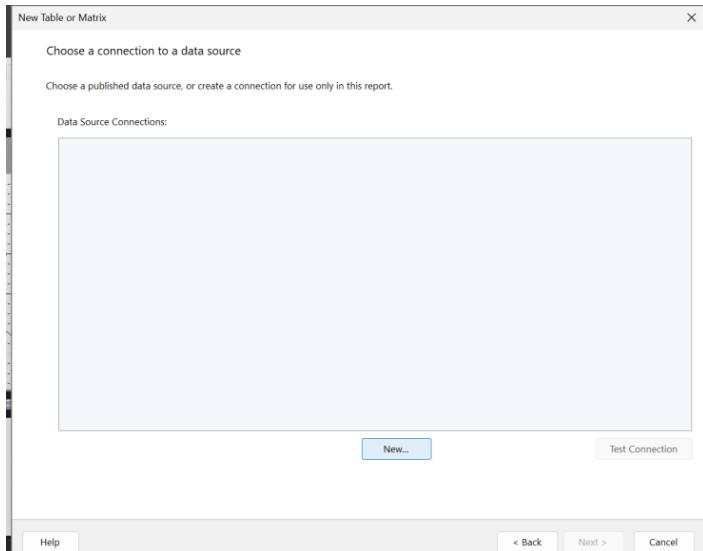
In general, this is useful for understanding about the BMI of children in two countries: Peru and Vietnam, thus, to support some decision-making process in future such as health policy, nutrition for children, and educational programs.

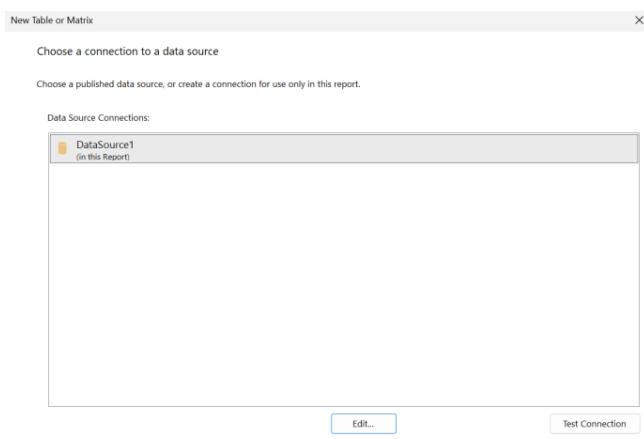
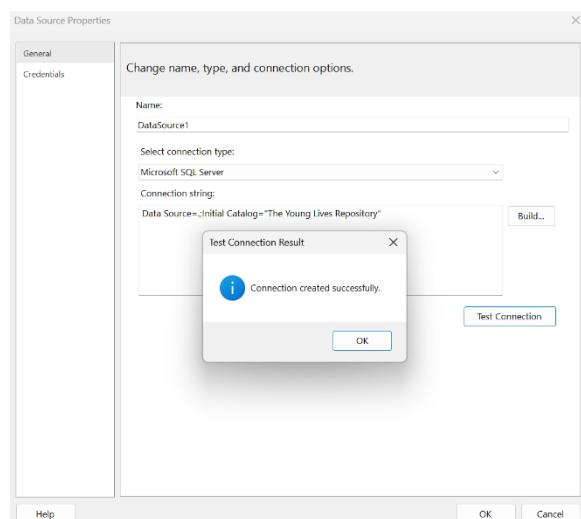
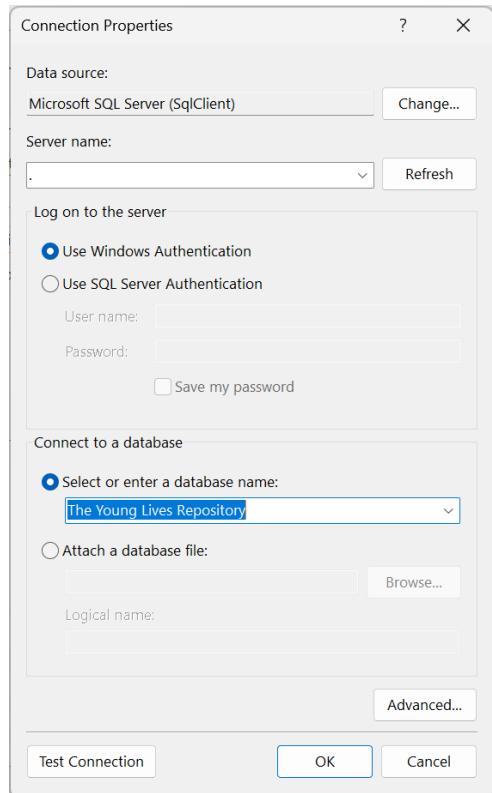
6. Report Analysis - Child's Health Level in Peru & Vietnam

For create this, we used Power BI Report Builder. The following results show how the report was created.

1. Open the Power BI Report Builder and follows.







New Table or Matrix

Design a query

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

Edit as Text Import... Run Query

Database view	Selected fields	Group and Aggregate
Tables		
Views		
Child_Poverty	Field country chhealth age	Aggregate (none) (none) (none)
Child_Poverty_View		

Relationships Auto Detect Edit Fields

Applied filters

Field name	Operator	Value	Parameter

Query results

< Back Next > Cancel

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	Column groups
country chhealth age	country

Row groups	Values
chhealth	Count(age)

Help < Back Next > Cancel

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

chhealth	[country]	Total
[chhealth]	[Count(age)]	[Count(age)]
Total	[Count(age)]	[Count(age)]

Help < Back Next > Cancel

New Table or Matrix

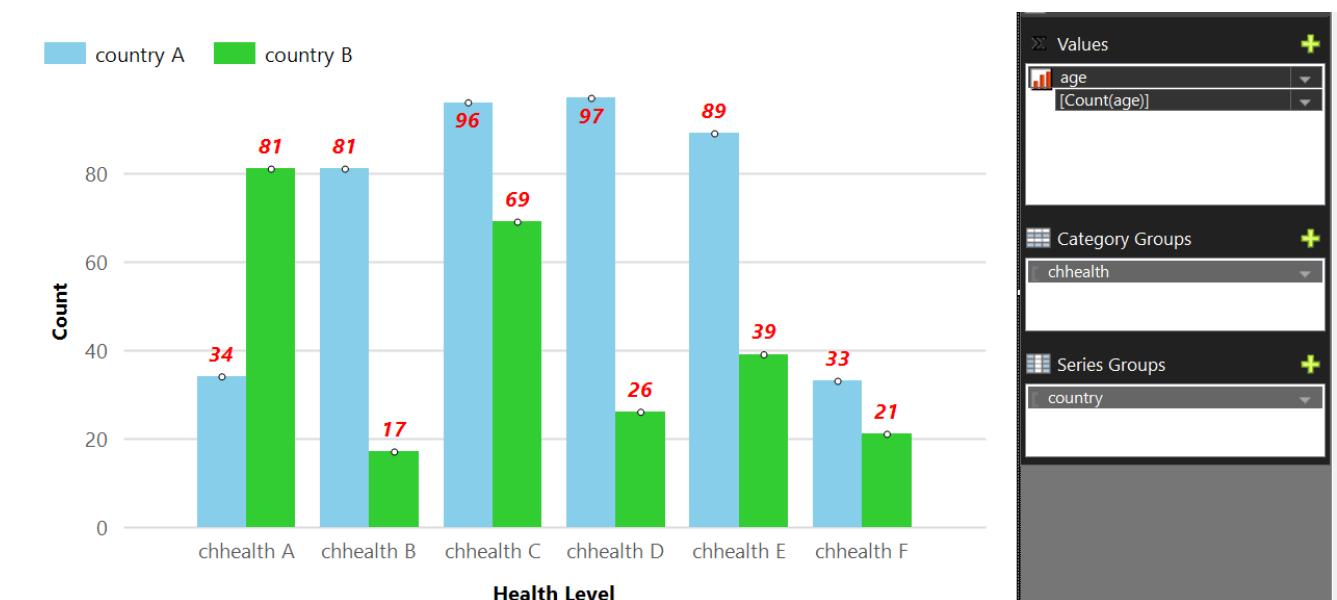
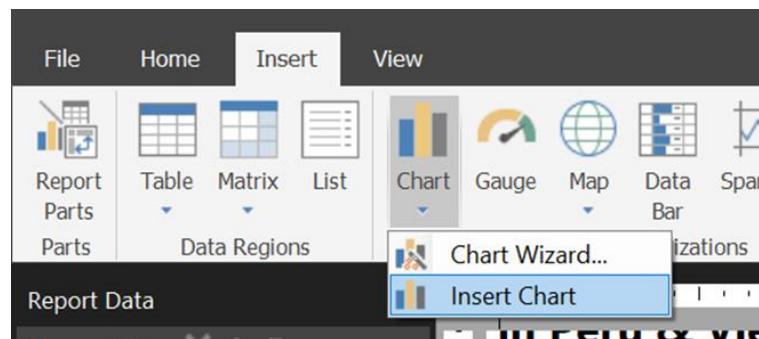
Preview

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

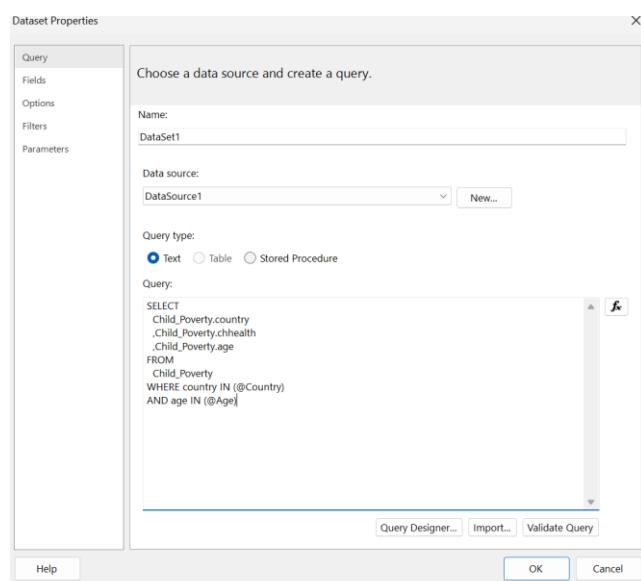
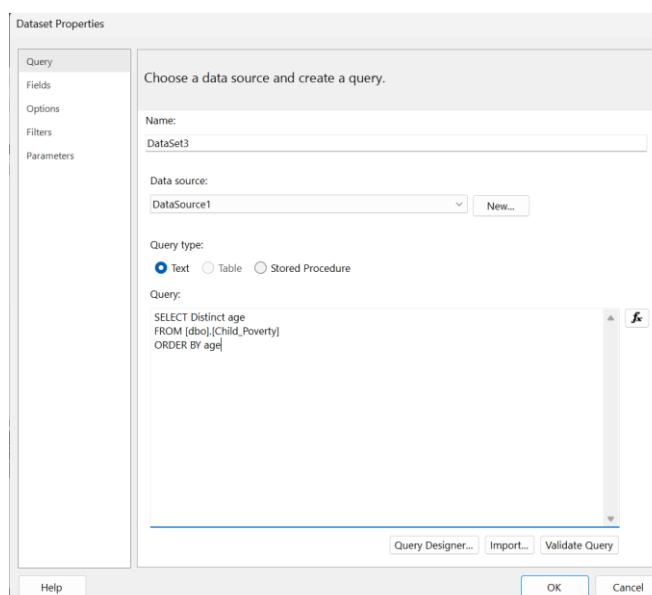
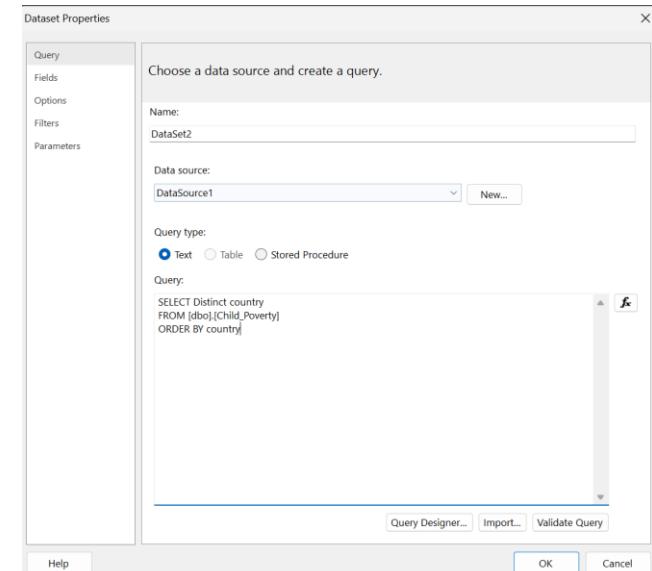
chhealth	[country]	Total
[chhealth]	[Count(age)]	[Count(age)]
Total	[Count(age)]	[Count(age)]

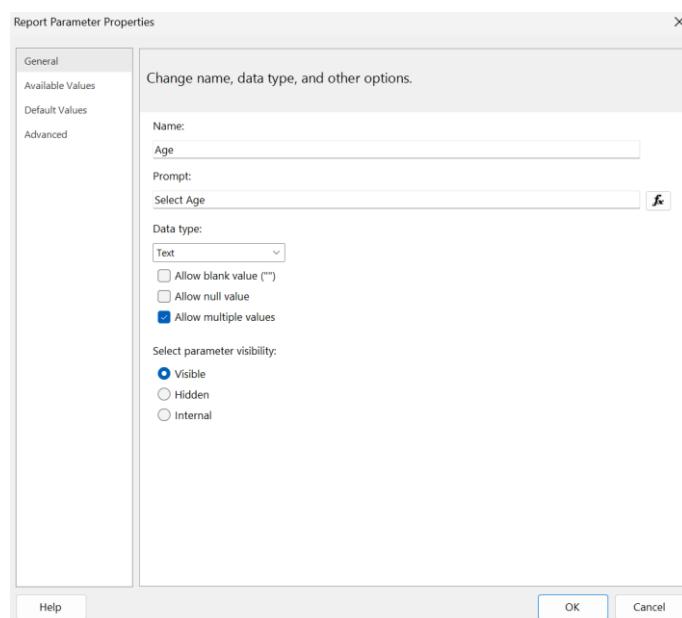
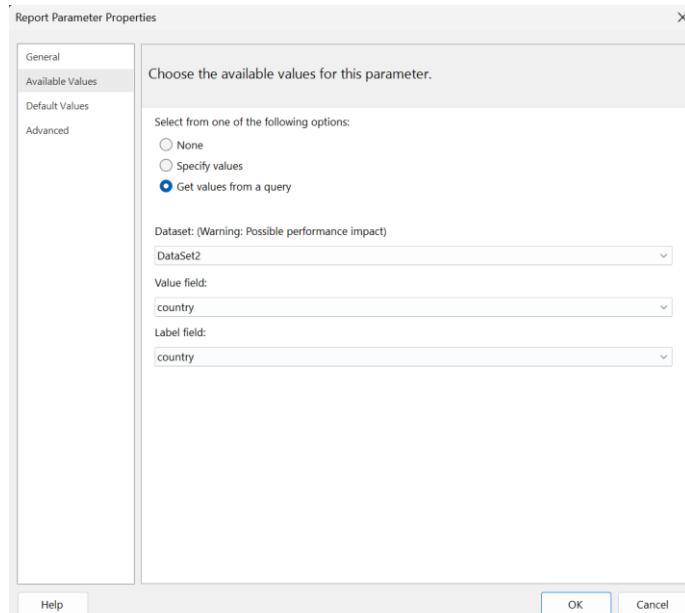
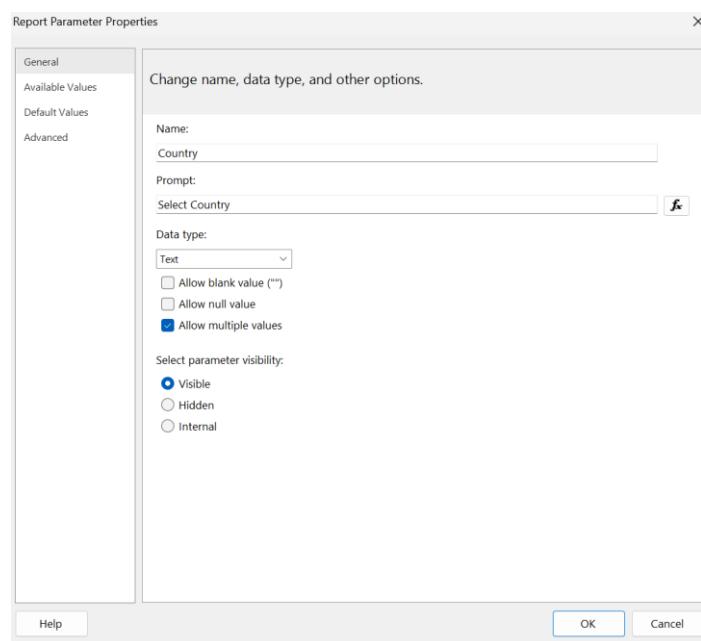
Help < Back Finish >> Cancel

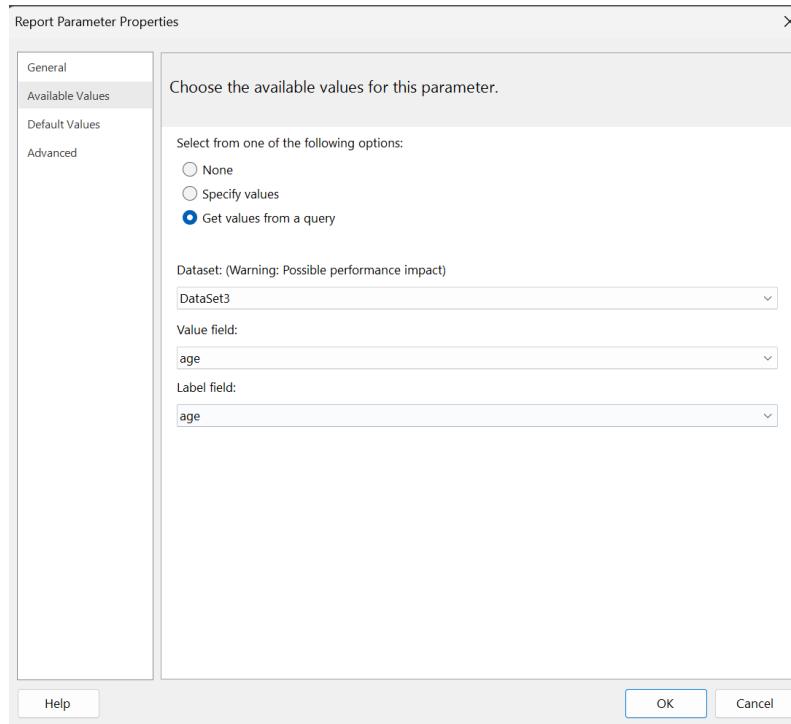
2. Insert chart.



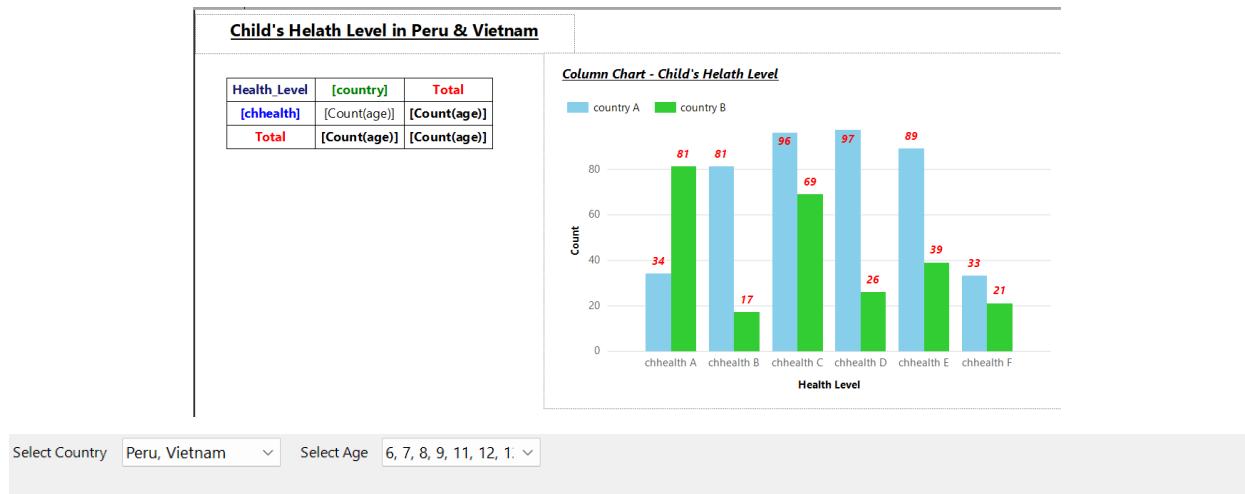
3. Create parameters. (Age & Country)







4. Final Output



Child's Health Level in Peru & Vietnam

Health_Level	Peru	Vietnam	Total
1	2	20	22
2	40	449	489
3	1244	2935	4179
4	2299	1658	3957
5	316	181	497
Total	3901	5243	9144

Column Chart - Child's Health Level



5. Analysis - Child's Health Level in Peru & Vietnam

This is about the child health levels in Peru and Vietnam. The table and column chart, respectively, shows the distribution of health level where the level of health regards as from level 1 up to level 5 in both countries.

This point to the fact that many children in the population fall in the mid to higher health standards in the population in Peru.

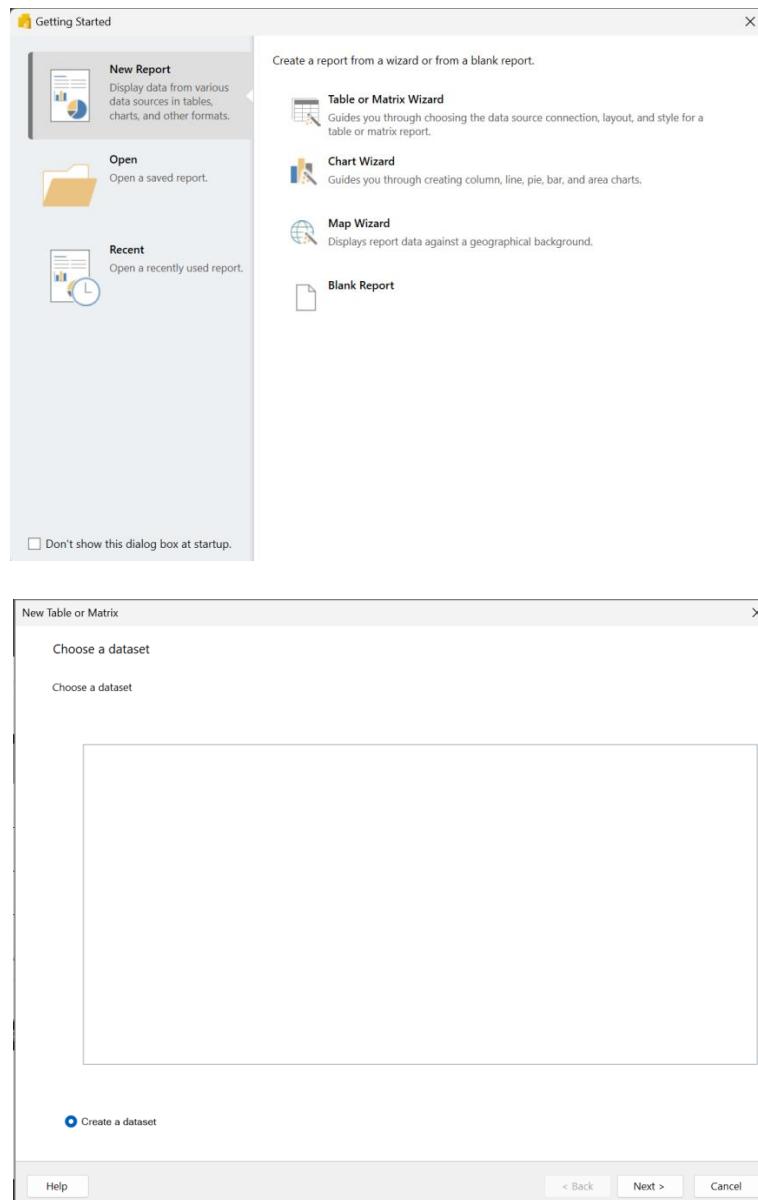
The total column therefore reflects the individual totals for both the countries and depicts the child health distribution. Such combined data can be also informative for policymakers and health professionals to find out where more effort and funds are needed.

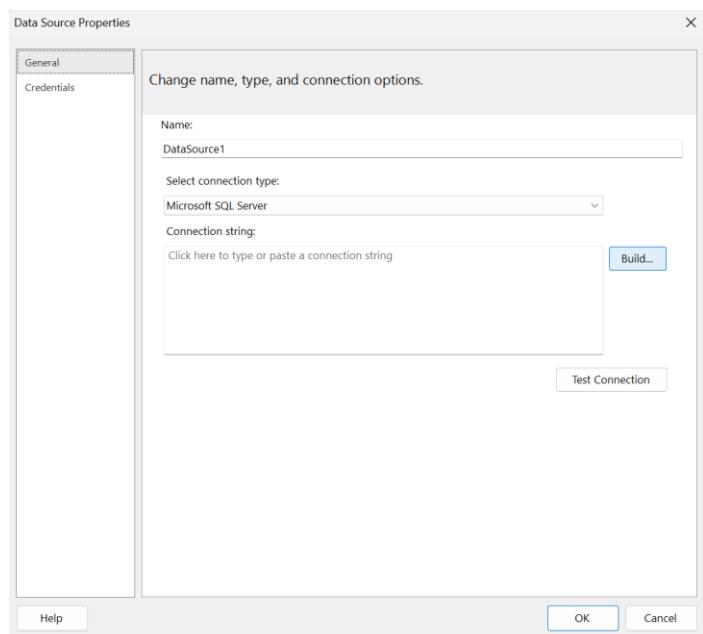
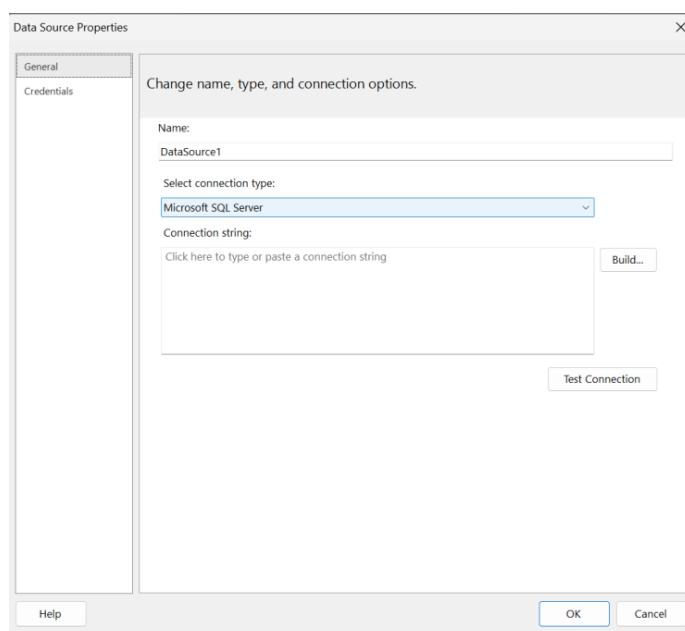
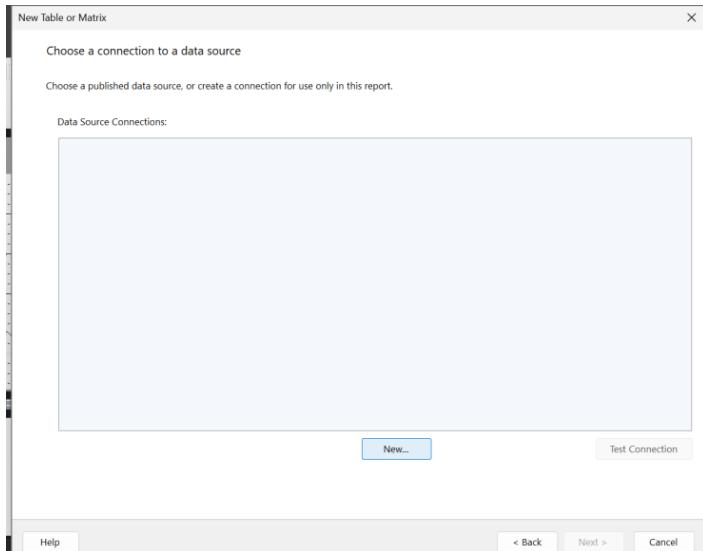
Altogether, this evidence has helped for the health systems demands in both Peru and Vietnam. It would also be beneficial to put emphasis on the health levels where the children are many with the intention of rectifying their health disorders.

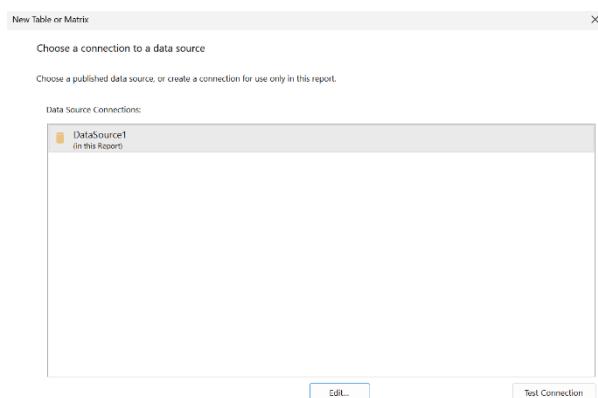
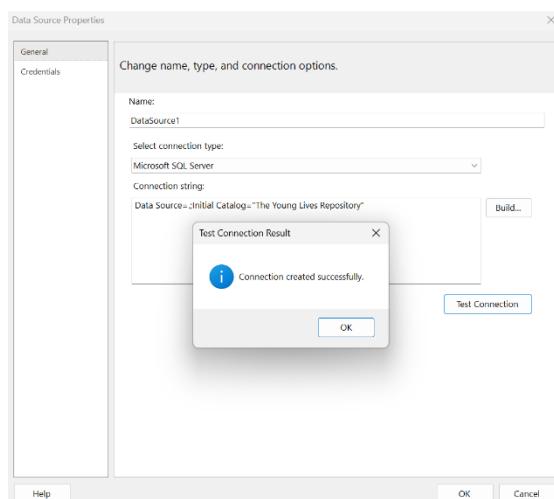
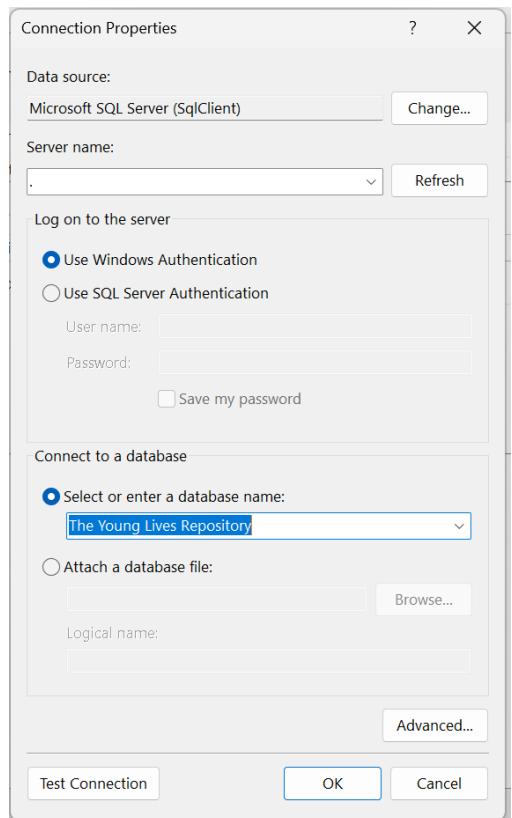
6. Report Analysis- Household head's relationship to YL child in Peru & Vietnam

For create this, we used Power BI Report Builder. The following results show how the report was created.

1. Open the Power BI Report Builder and follows.







New Table or Matrix

Design a query

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

Selected fields

Field	Aggregate
headrel	(none)
age	(none)
country	(none)

Relationships

Applied filters

Query results

Help < Back Next > Cancel

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

- headrel
- age
- country

Column groups

- country

Row groups

- headrel

Values

Count(age)

Help < Back Next > Cancel

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.

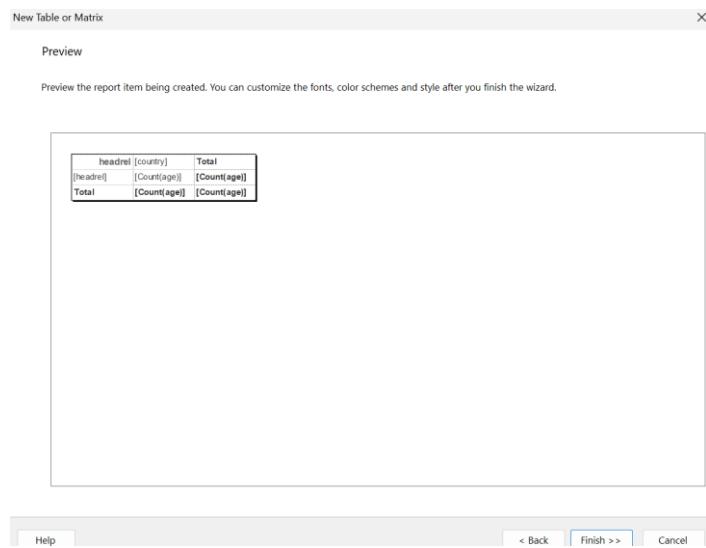
Options:

- Show subtotals and grand totals
- Blocked, subtotal below
- Blocked, subtotal above
- Stepped, subtotal above
- Expand/collapse groups

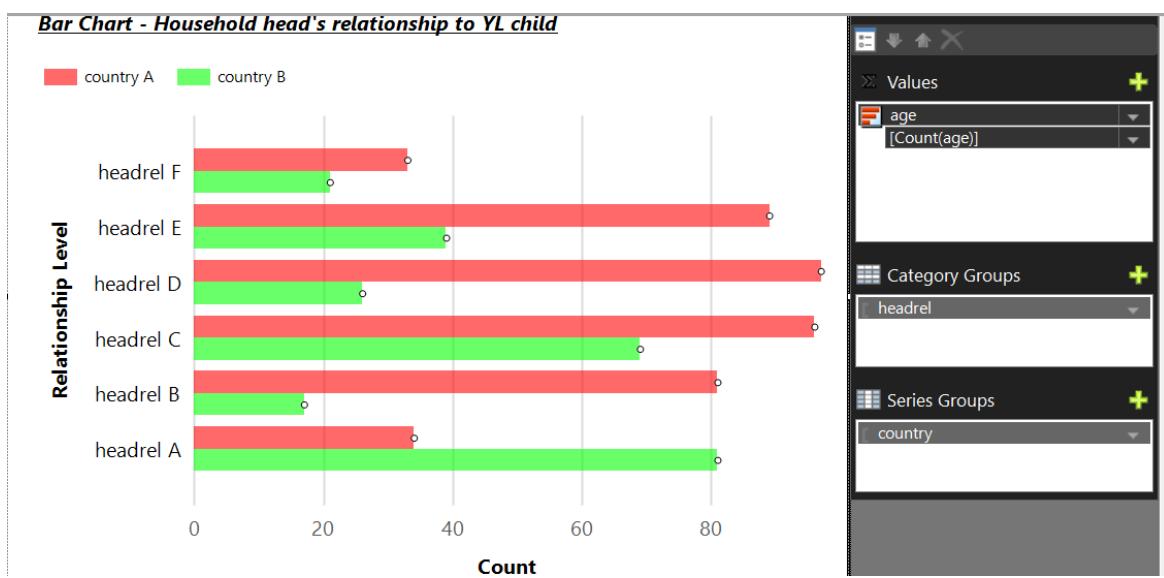
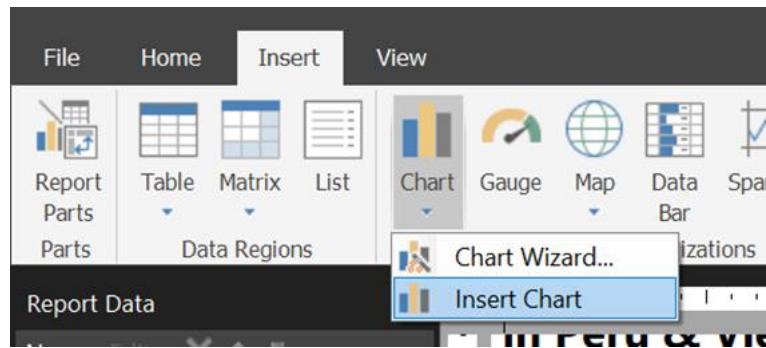
Preview

headrel	[country]	Total
[headrel]	[Count(age)]	[Count(age)]
Total	[Count(age)]	[Count(age)]

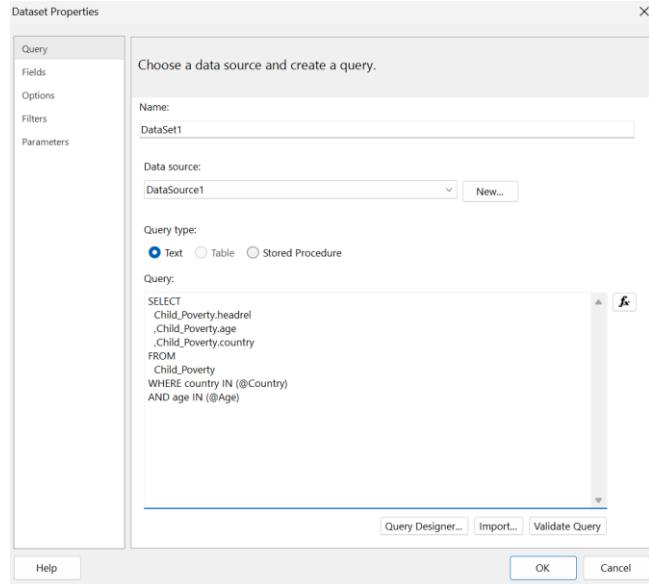
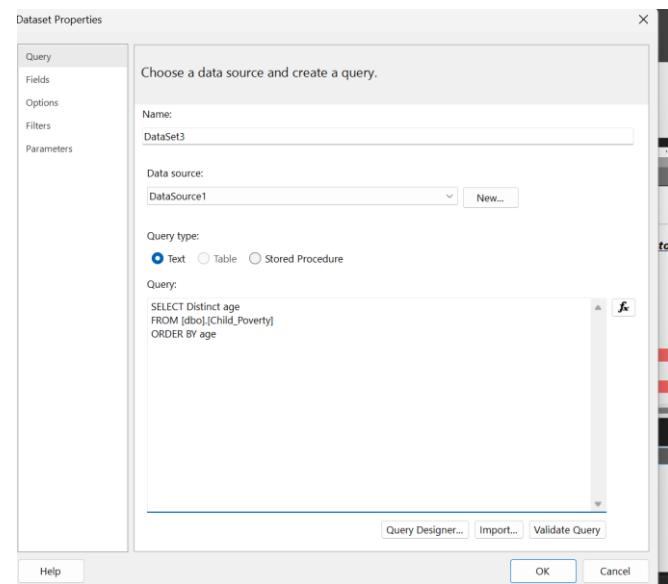
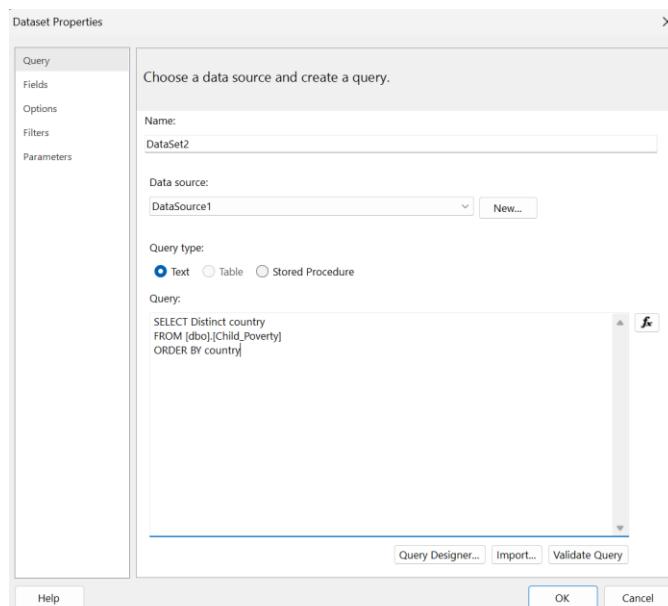
Help < Back Next > Cancel

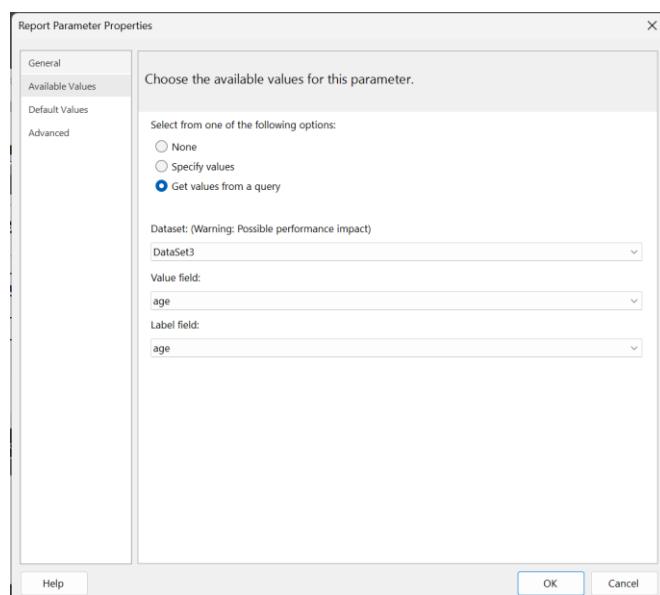
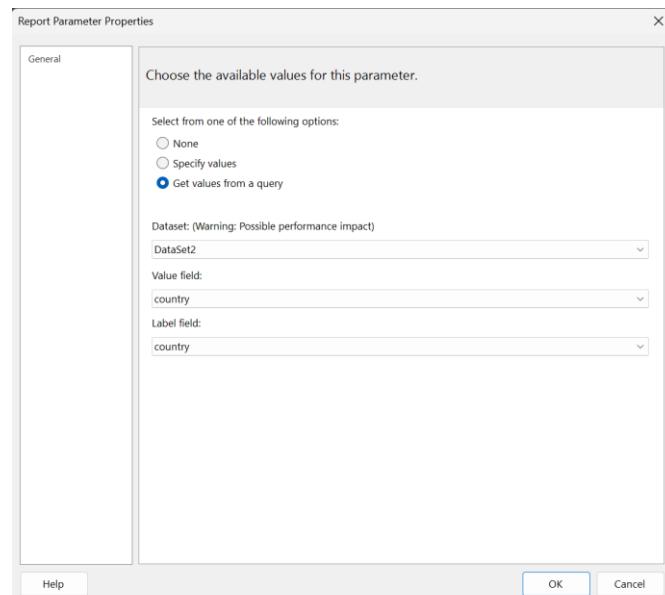
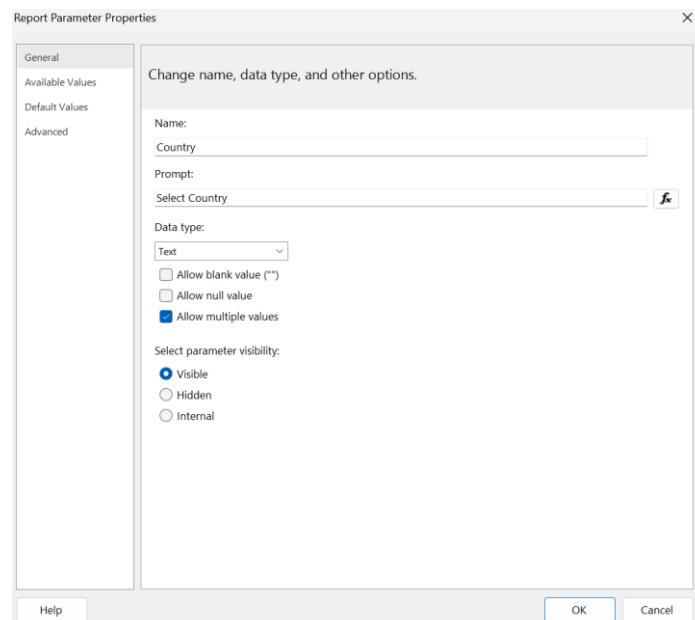


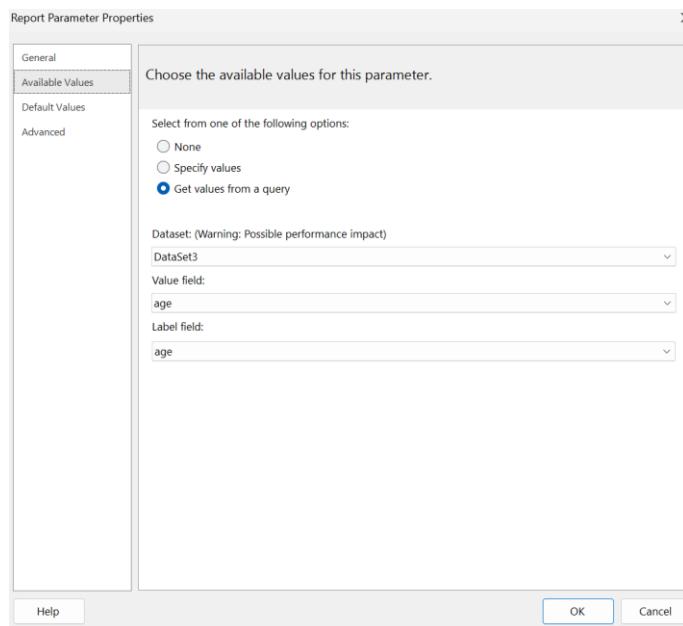
2. Insert Chart



3. Create parameters. (Age & Country)







4. Final Output

Household head's relationship to YL child

Relationship Level	[country]	Total
[headrel]	[Count(age)]	[Count(age)]
Total	[Count(age)]	[Count(age)]

Select Country: Peru, Vietnam Select Age: 6, 7, 8, 9, 11, 12, 1.

Bar Chart - Household head's relationship to YL child

Relationship Level	country A (Count)	country B (Count)
headrel F	~35	~20
headrel E	~85	~40
headrel D	~85	~45
headrel C	~85	~70
headrel B	~80	~15
headrel A	~35	~80

Household head's relationship to YL child

Relationship_Level	Peru	Vietnam	Total
1	3723	4737	8460
2	34	2	36
3	112	482	594
4	16	10	26
5	6	1	7
6	4	7	11
7	5	2	7
8	0	1	1
9	1	1	2
Total	3901	5243	9144

Bar Chart - Household head's relationship to YL child

Relationship Level	Peru (Count)	Vietnam (Count)
1	~3800	~4500
3	~100	~500

5. Analysis - Household head's relationship to YL child in Peru & Vietnam

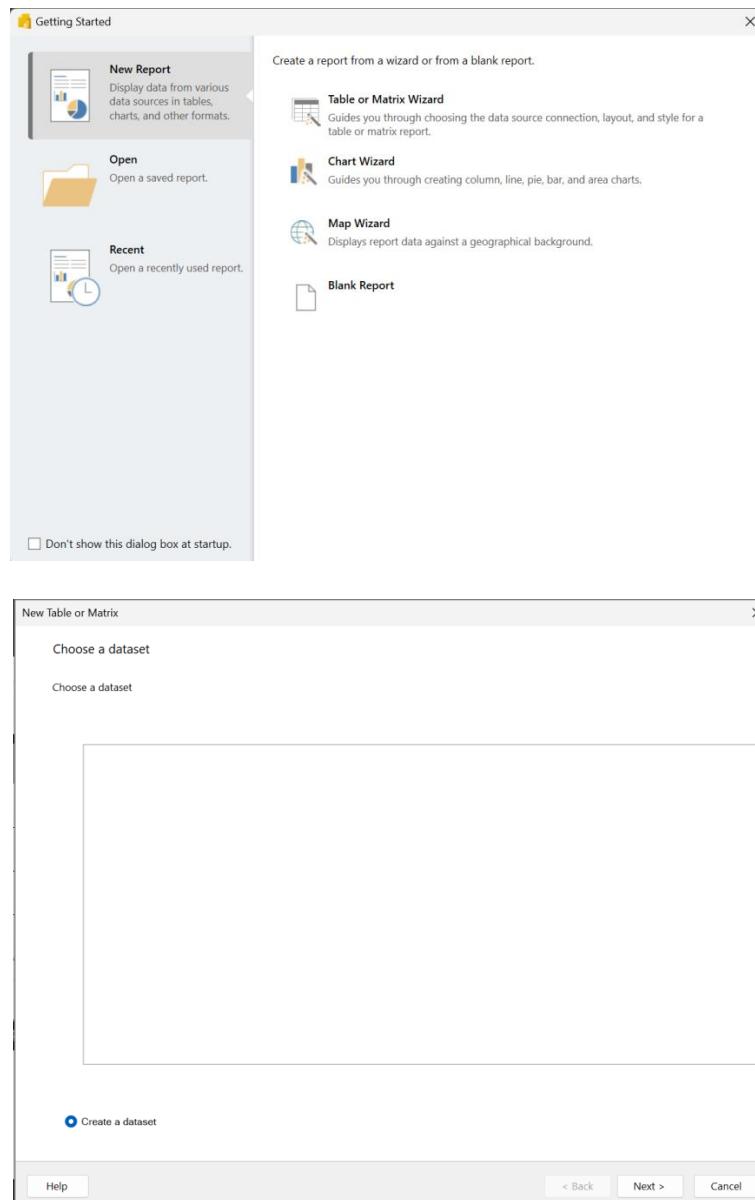
The table gives corresponding count for Peru, Vietnam. The bar chart provides the trend from the table. On the y-axis, there are different types of relationships, and on the x-axis, there are households' numbers. Since a greater proportion of children in both countries are with biological parents, the chart also makes it easier to perform a comparison of other relations in childcare.

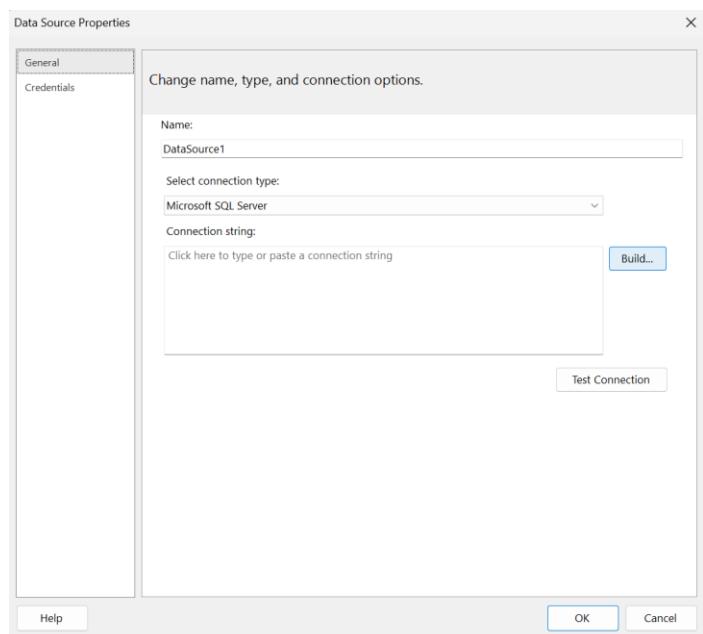
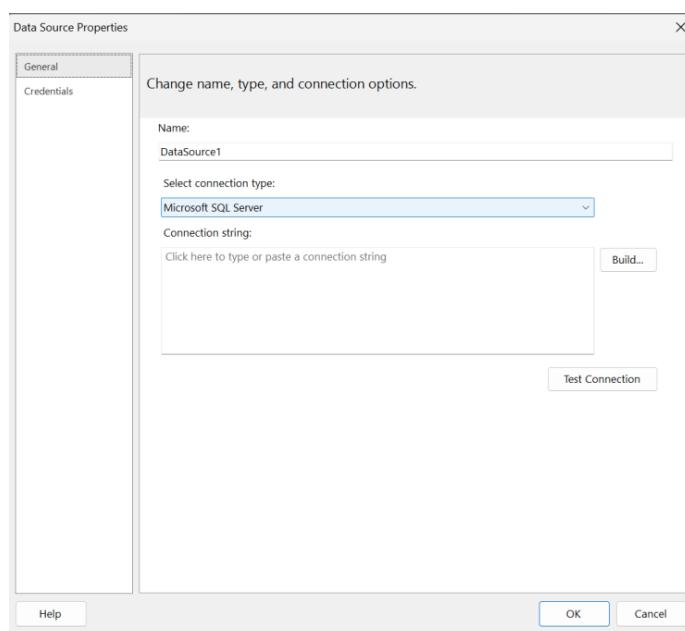
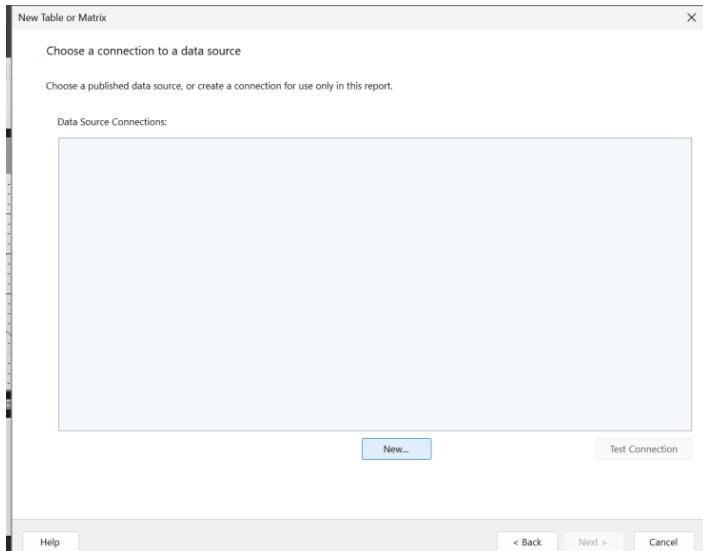
These statistics show that the biological parents are the main heads of the households in the two countries- Peru, and Vietnam. The counts for other relationships like those of stepparent or other relations are comparatively quite low. This indicates that there was a presence of a high-parent-led traditional family in the two countries. It is important for people interested in cross cultural analysis and comparison of family systems and social status of the heads of households.

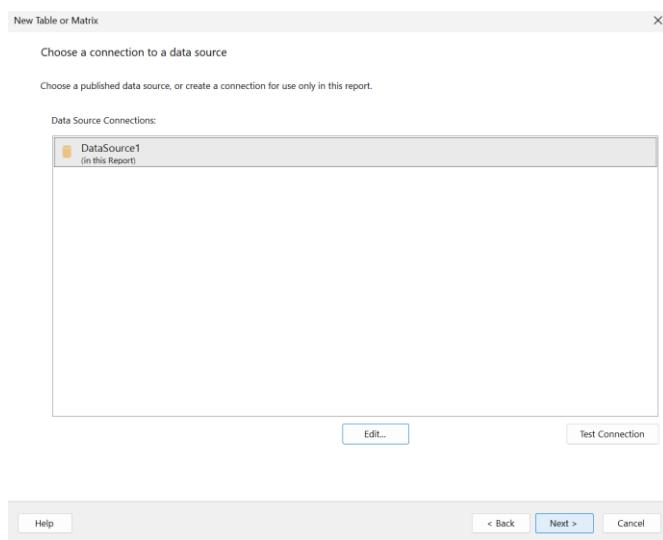
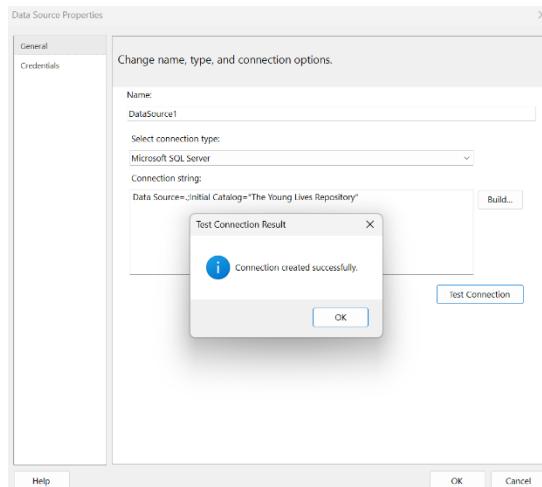
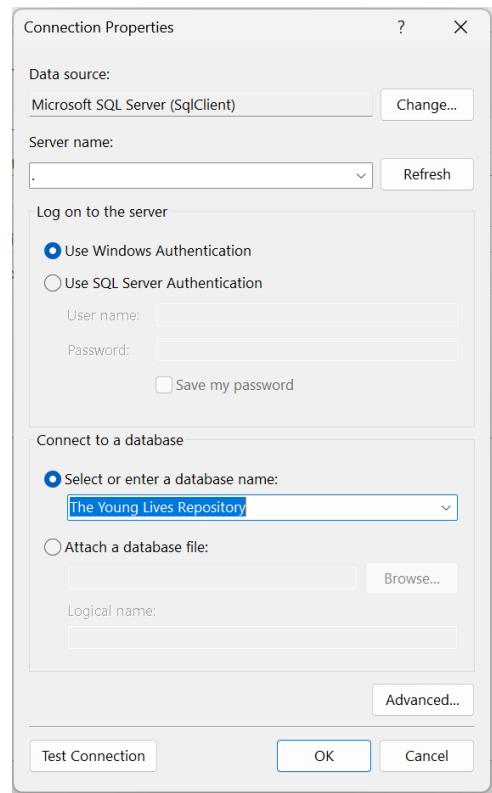
6. Report Analysis - Food Security Level in Peru & Vietnam

For create this, we used Power BI Report Builder. The following results show how the report was created.

1. Open the Power BI Report Builder and follows.







New Table or Matrix

Design a query

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

Selected fields

Field	Group and Aggregate
country	(none)
foodsec	(none)

Relationships

Applied filters

Query results

Help < Back Next > Cancel

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

- country
- foodsec

Column groups

- country

Row groups

- foodsec

Values

- Count(foodsec)

Help < Back Next > Cancel

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.

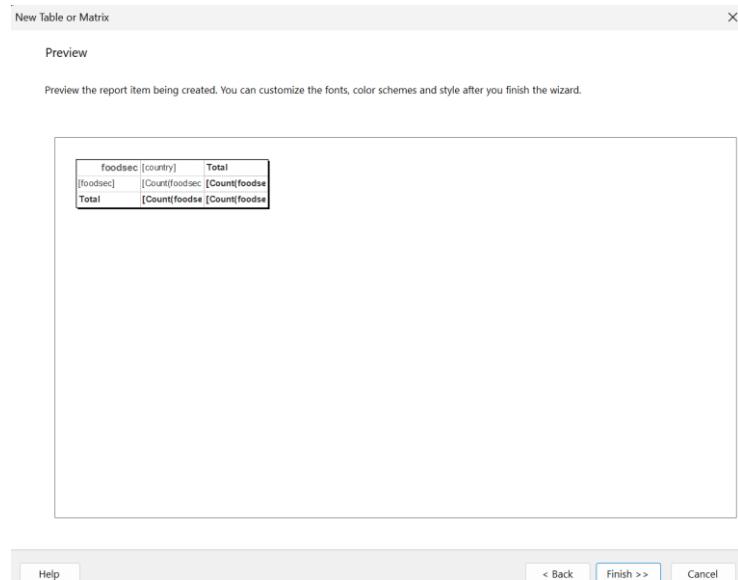
Options:

- Show subtotals and grand totals
 - Blocked, subtotal below
 - Blocked, subtotal above
 - Stepped, subtotal above
- Expand/collapse groups

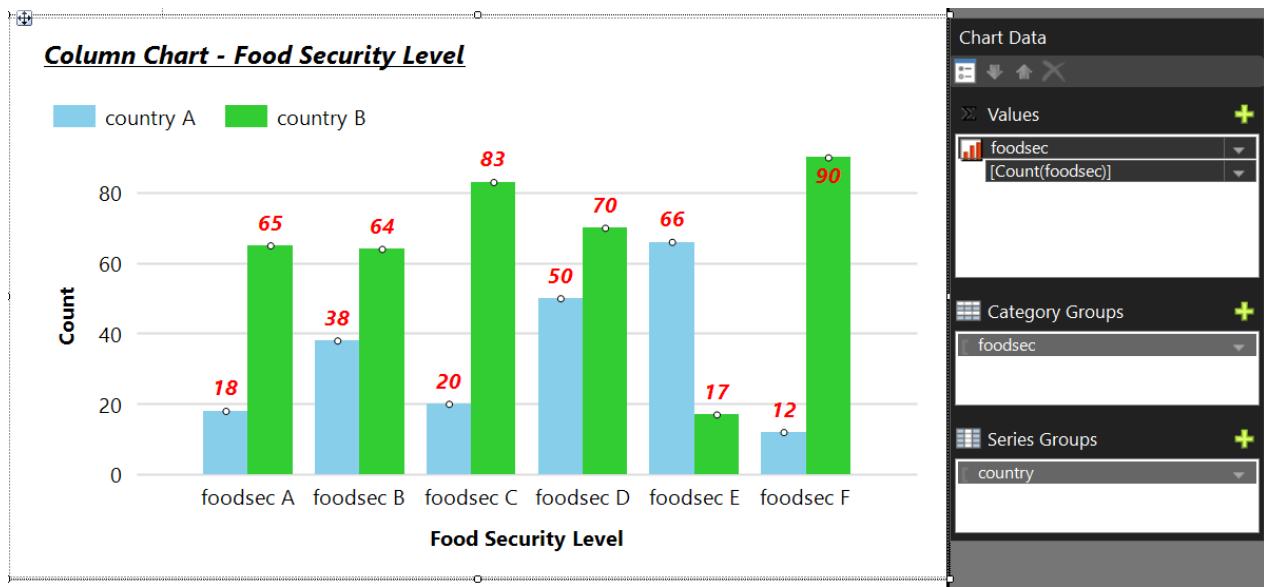
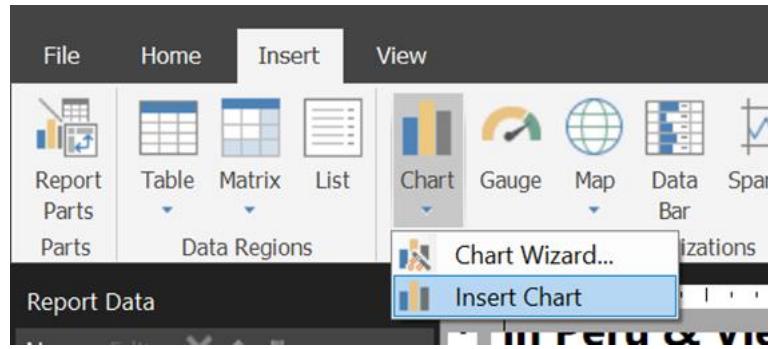
Preview

foodsec [country]	Total
[foodsec]	[Count(foodsec)]
Total	[Count(foodsec)]

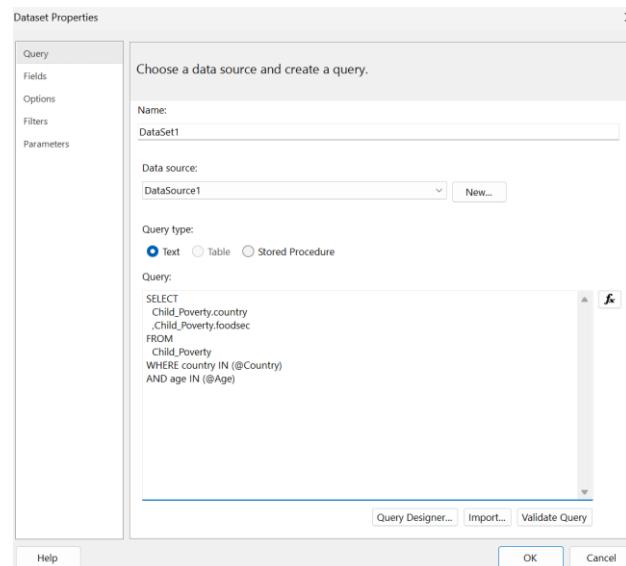
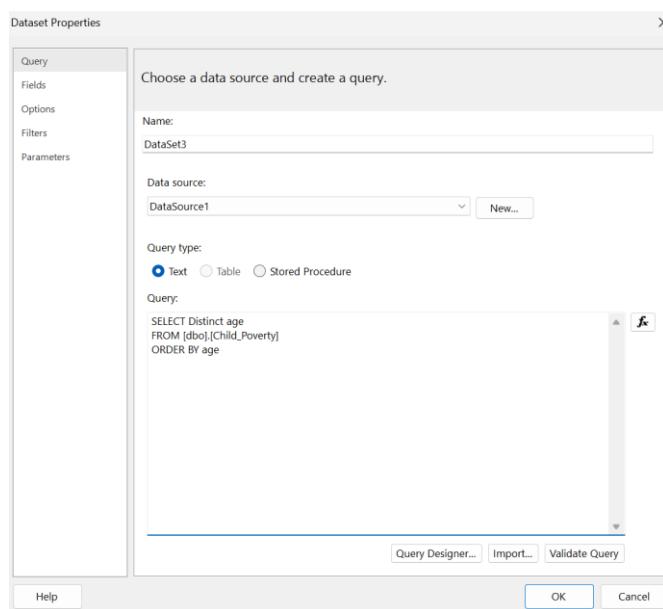
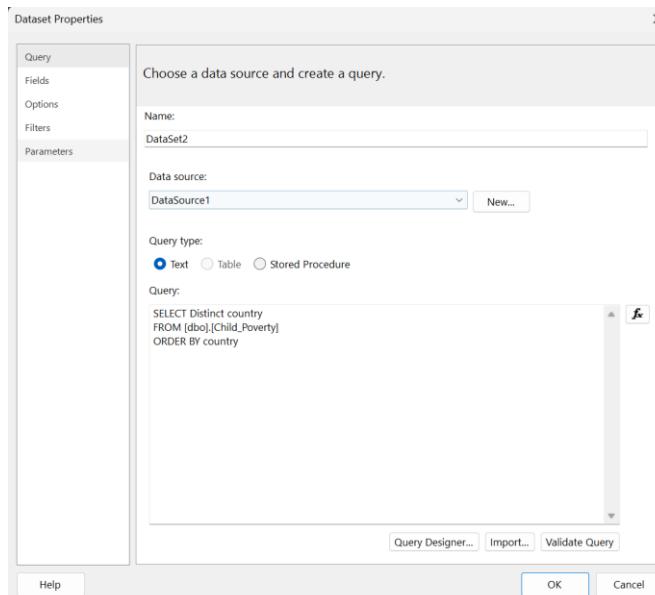
Help < Back Next > Cancel

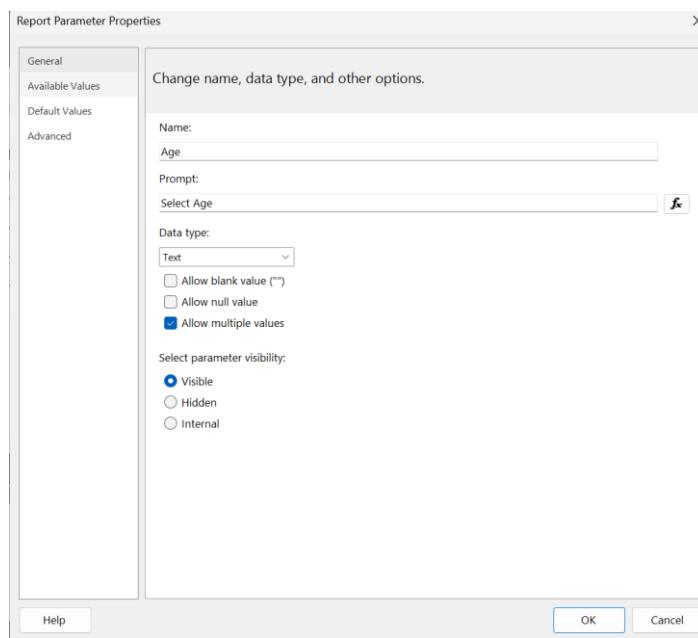
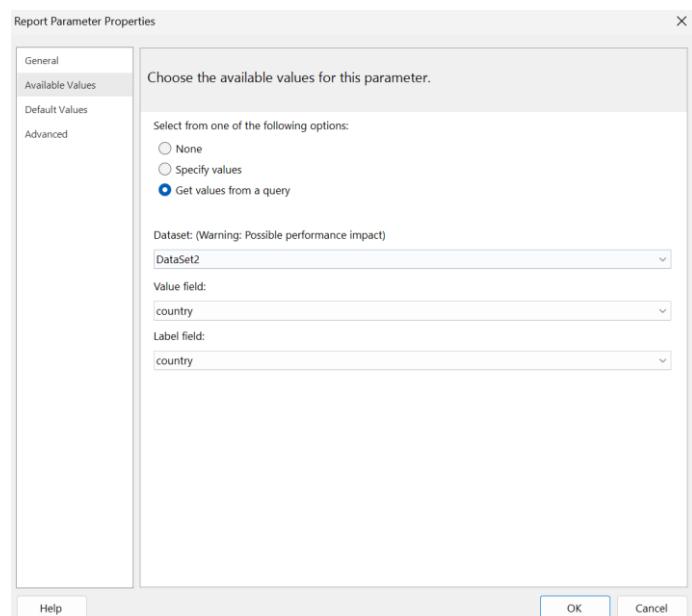
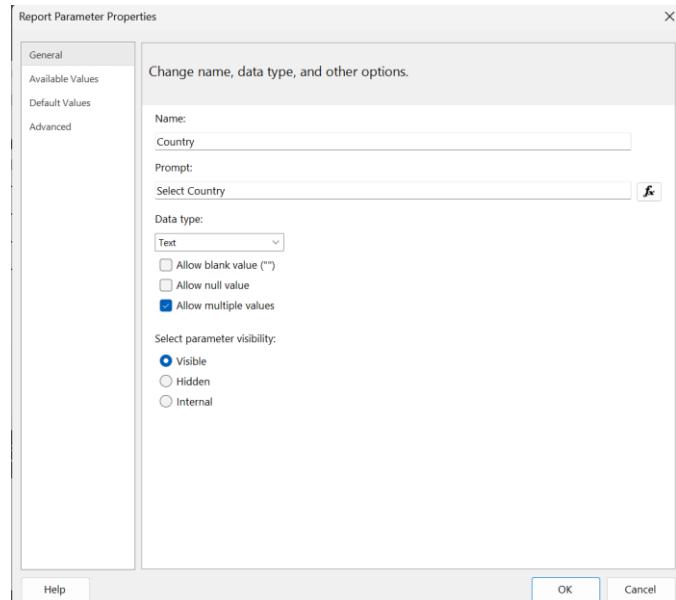


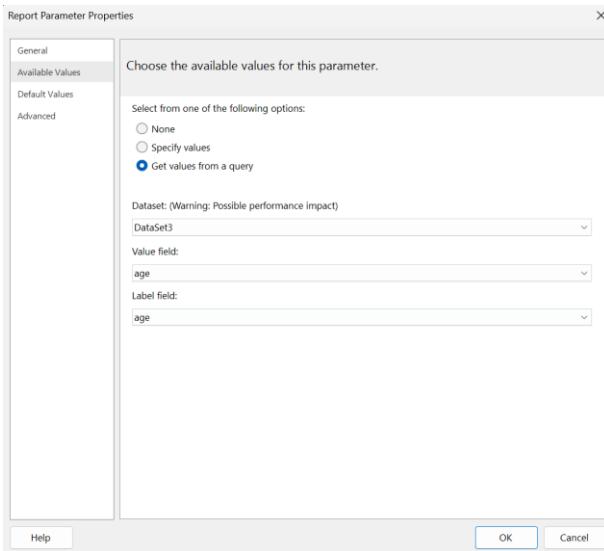
2. Insert Chart



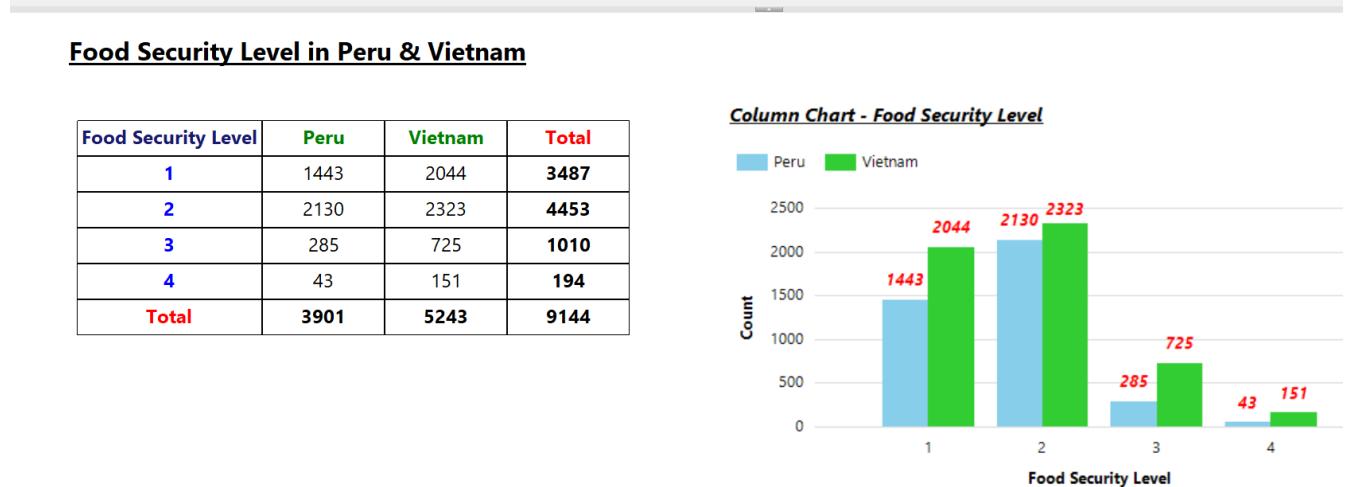
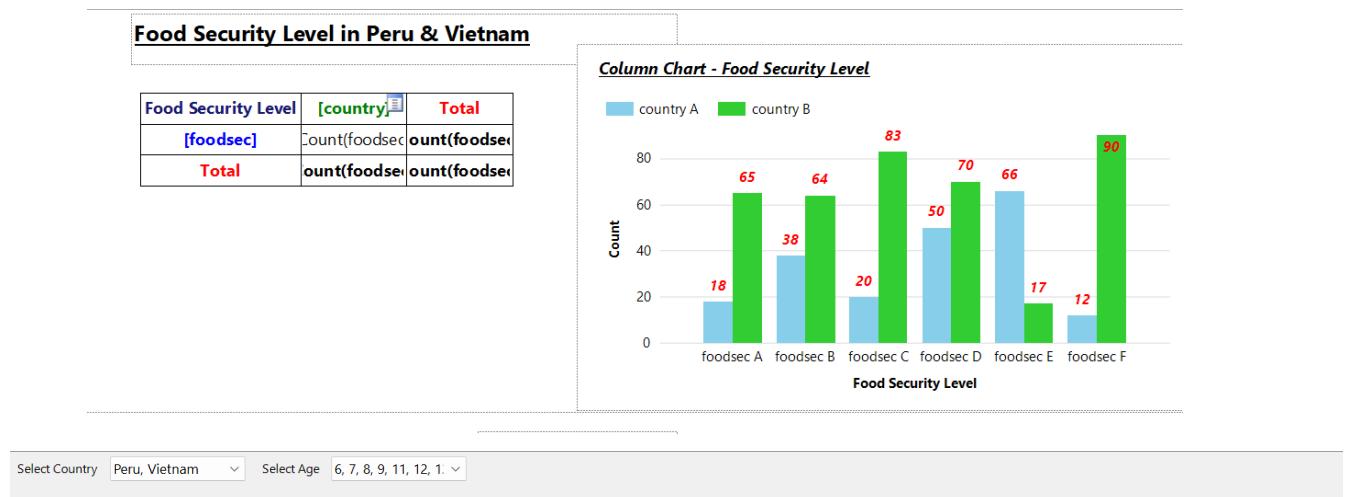
3. Create Parameters. (Age & Country)







4. Final Output



5. Analysis - Food Security Level in Peru & Vietnam

The table gives information on the food security level in both countries classified into four levels. From this it can be recommended that either Vietnam has a better sample size or indeed a better food security status than that of Peru.

The column chart shows the data from the table given below. It is easier to understand that the cross tabulation shows food security levels along the x-axis with levels ranging from 1 to 4 and the count values along the y-axis.

These insights can be used to improve the understanding of the nature of FS context in both countries and guide formulation of policies as well as distribution of resources.

6. Conclusion

The findings regarding demographics of children in need as well as child and health in Peru and Vietnam contain important implications for policymaking and practice. The distribution of child population for planning and development in prioritization of resources such as educational needs, health facilities, and child projects. Using data on heights and weights and BMI, distinctions involving genetic, nutritional and health care status are made for designing interventions for childhood development and health. Biological parents' participation across the households is also important for cross national family studies while FSEC data provide relevant directions in combating food insecurity. In general, the findings presented in this paper provide the basis for evidenced-based decision-making aimed at promoting children's wellbeing in Peru and Vietnam.

7. References for Task 01

- Young Lives: An International Study of Childhood Poverty: Rounds 1-5 Constructed Files, 2002-2016
(Boyden, J. (2022). Young Lives: An International Study of Childhood Poverty: Rounds 1-5 Constructed Files, 2002-2016. [data collection]. 5th Edition. UK Data Service. SN: 7483, [DOI: http://doi.org/10.5255/UKDA-SN-7483-5](http://doi.org/10.5255/UKDA-SN-7483-5))
- SQL Server / T-SQL Tutorial – YouTube Video
(<https://youtube.com/playlist?list=PLWf6TEjiuIDvJ4P5l7Bzrmpzv8hW9CAO&si=pQZpSsPMkclObU5d>)
- Paginated Report in Power BI with Parameters – YouTube Video
(<https://youtu.be/Ga7ntWwv5hU?si=sDKAmMwwcOliLicY>)

Task 02 - Greater Manchester

Domestic Energy Performance

Analyzer

(Power BI Dashboard)

1) Introduction

This assignment aims at data analysis and mapping of Energy Performance Certificate (EPC) data for the local authorities in Manchester for the year 2013-2023. Energy Performance Certificates are informative documents about the existing energy efficiency of buildings, which indicates trends contributing to decision-making for stakeholders in the energy industry. The purpose of implementing this project is to design an effective Power BI dashboard to investigate these trends and discover.

The data cleaning, transformation, and visualization is done through the Power BI. A big bonus for this solution will be the dashboard, which will help users navigate through the data and make the best decision. This project targets to provide a clear and efficient tool for the stakeholders in the energy sector with interactive tools in Power BI.

2) Explanation and Preparation of the data set

- Explanation of the data set

The EPC database (<https://epc.opendatacommunities.org/downloads/domestic#local-authority>) is open to England and Wales, that contains information on energy efficiency of buildings for residential and commercial purposes. In here, we used Energy Performance data in Manchester. The dataset contains certificate information recorded between 2006 and 2024. The goal of this process was to clean and prepare the data for further analysis, ensuring accuracy, removing duplicates, and handling missing or irrelevant values. Also there are 92 columns and 300888 rows in the data set.

- Preparation of the data set

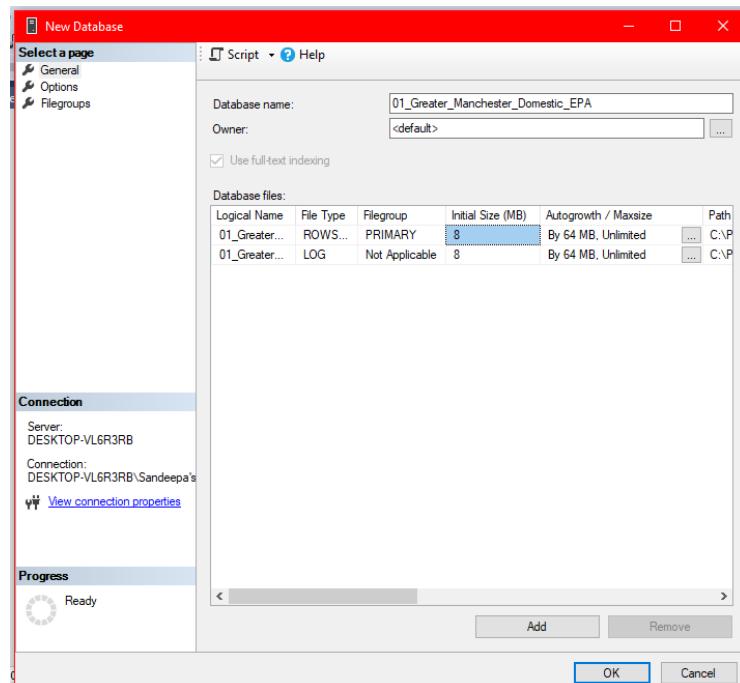
The dataset originally contained several columns representing details such as:

- Address fields (Address1, Address2, Address3, County)
- Energy efficiency ratings
- Costs related to energy consumption
- CO2 emissions
- Heating, lighting, and water costs, among others.

The data cleaning process aimed to remove unnecessary columns, address duplicates, handle missing values, and ensure data consistency.

1. Import Data to SQL Server Management Studio(SSMS)

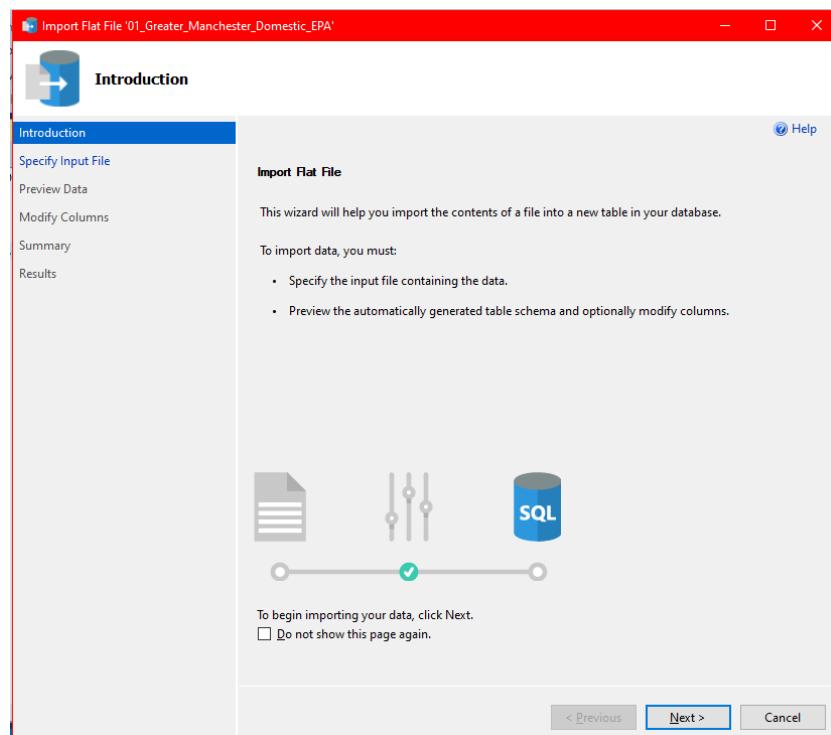
First of all we created a Data base called **01_Greater_Manchester_Domestic_EPA**.



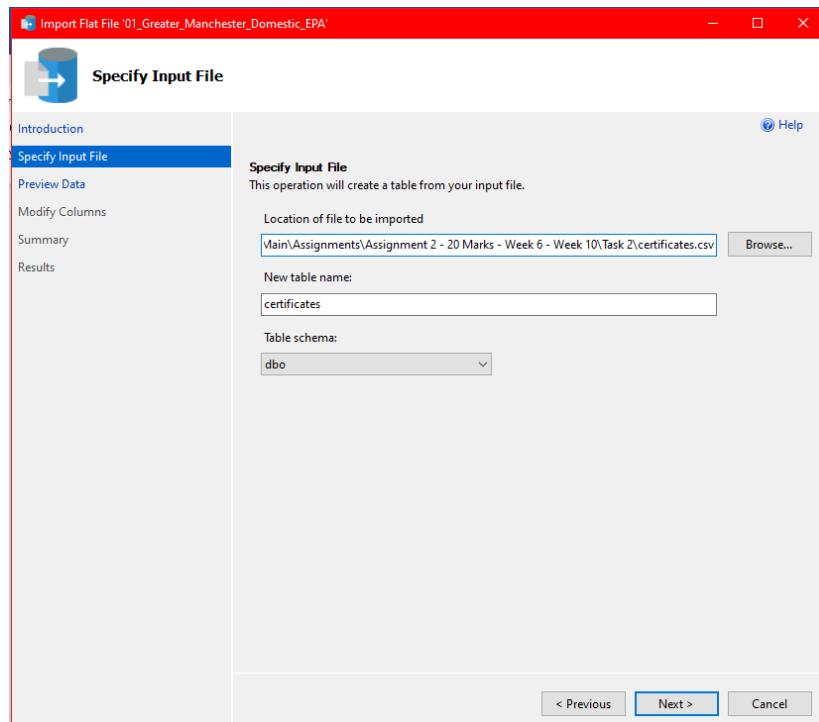
Step 01

Then, by clicking the following path we have imported the data set.

Right click on the data base → Tasks → Import Flat File



Step 02



Step 03

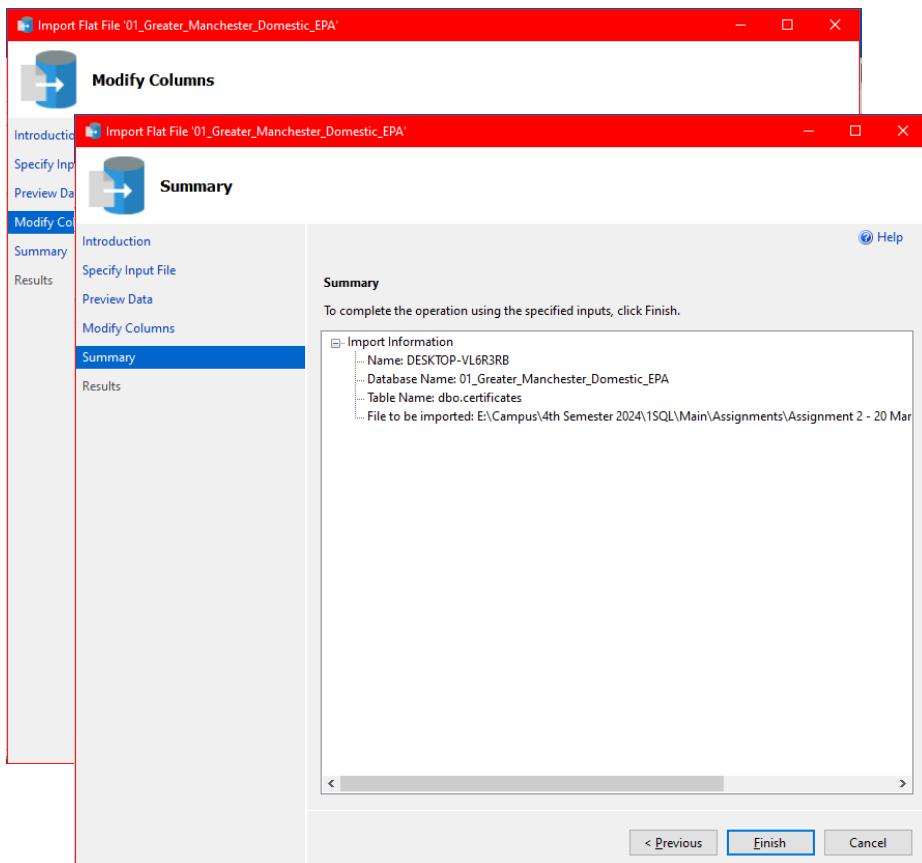
Here is the preview of the data set,

The screenshot shows the 'Preview Data' step of the 'Import Flat File' wizard. The left navigation pane includes 'Introduction', 'Specify Input File', 'Preview Data' (selected), 'Modify Columns', 'Summary', and 'Results'. The main panel title is 'Preview Data' with a subtitle 'This operation analyzed the input file structure to generate the preview below for up to the first 50 rows.' A large table displays 50 rows of data with columns: LMK_KEY, ADDRESS1, ADDRESS2, ADDRESS3, POSTCODE, and BUILDIN. Below the table is a checkbox for 'Use Rich Data Type Detection' with the note 'may provide a closer type fit. However, cells with anomalous values may be dropped.' At the bottom are buttons for '< Previous', 'Next >', and 'Cancel'.

LMK_KEY	ADDRESS1	ADDRESS2	ADDRESS3	POSTCODE	BUILDIN
20e4f9426b3cec...	428 St. Marys R...			M40 0DE	1000364
5453838010320...	20, Parkleigh Dr...			M40 3RY	0895420
2272285617520...	22, Silverwell St...			M40 1PA	5991187
1405676299212...	Apartment 8	1, Pennington S...		M12 4QJ	3746991
2884251108420...	10, Morna Walk			M12 6WP	7874712
25cb44ed6c67d...	42 GRINDLEY A...			M21 7NF	1000150
25d00f64a40b...	APTAMENT 20	LOCKES YARD	4 GREAT MARL...	M1 5AL	1000159
1072598235512...	14, Montpellior ...			M22 0DZ	6331548
1640172669842...	167B Fog Lane			M20 6FJ	9866068
1261844233352...	24, Carrgreen C...			M19 1LT	0518381
1836873209420...	14, Chervil Close			M14 7DP	6935624
1216165678852...	27, Riverstone ...			M23 9QW	7247858
1484708669962...	Flat 88	1 Cambridge St...		M1 5GB	2551957
1467796018320...	Flat A/1 Elm Co...	Barlow Moor R...		M20 2QQ	0645271
2195283899620...	Apartmen 6.4 ...	108, High Street		M4 1HT	4951807
2290393402220...	80, The Bouleva...			M20 2EU	3885428
2290393402220...	80, The Bouleva...			M20 2EU	3885428

Step 04

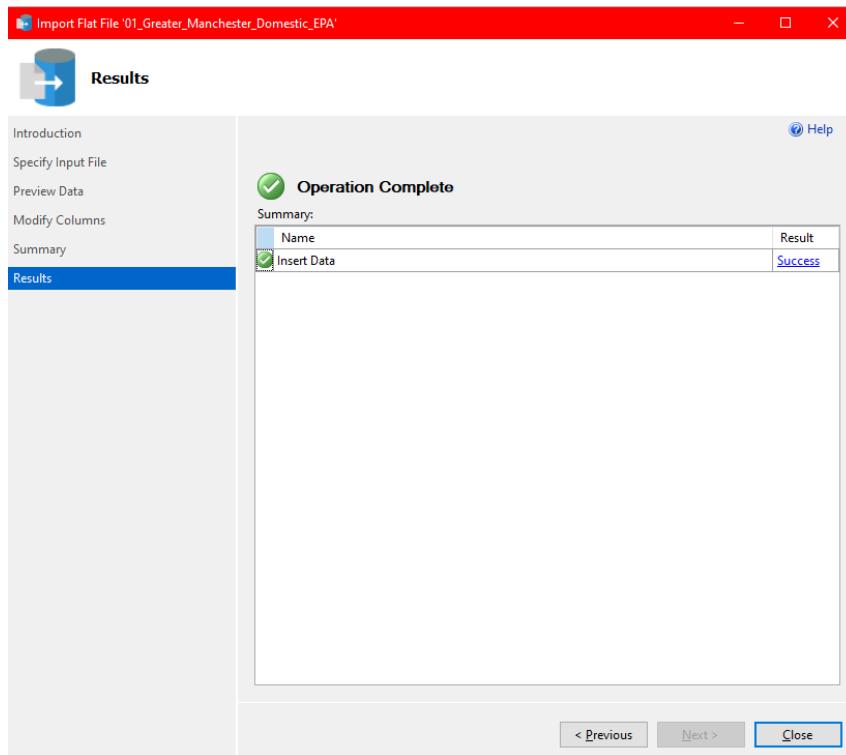
Here, we have changed the data type accordingly and allowed NULLs for all the columns.



Step 05

Step 06

After importing is successfully done, the window looks like below.

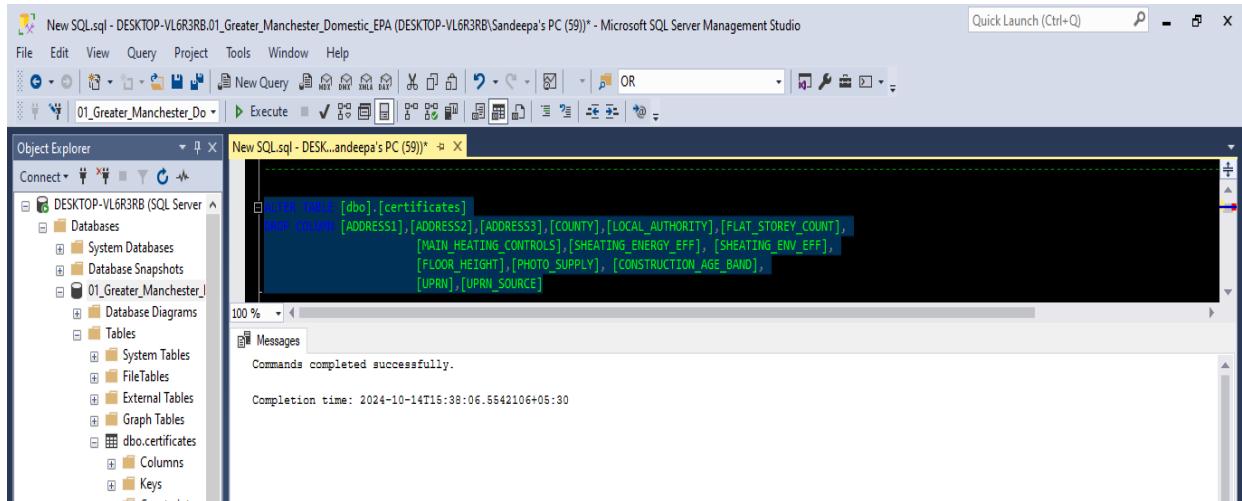


2. Cleaning Data

1) Preview the data set

	ADDRESS1	ADDRESS2	ADDRESS3	POSTCODE	BUILDING_REFERENCE_NUMBER	CURRENT_ENERGY_RATING	POTENTIAL
1	21810960213200902030531078058306	Fat 2	7 Keppel Road	NULL	M21 0AT	0678337562	C
2	10403719973201312813091989278896	170, Claremont Road	NULL	NULL	M14 4TU	9679816178	D
3	31798112042050701595716042816	Fat 28 Alley House	1, Andover Drive	NULL	M23 2HP	1247324668	D
4	41237602950201002061124578200656	Fat 1	257-259, Barlow Moor Road	NULL	M21 7GJ	357422768	E
5	9159565498020130915222030379818	180, Station Street	NULL	NULL	M11 4RT	3852447078	D
6	8626525295201212031304469592697	90, Neil Lane	NULL	NULL	M21 7DA	6789953078	C
7	133384589402014110612052494940968	2, Chilworth Street	NULL	NULL	M14 7PL	3413389468	D
8	19627164161212015952153599595640	5, Cockcroft Street	NULL	NULL	M9 4DE	5082658378	B
9	1380664318120151030128889759444	Fat 2	1, Dean Lane	NULL	M40 3AZ	4339120478	B
10	1128242996201404191314136268244	17, Randolph Street	NULL	NULL	M19 3AU	4639932278	D
11	119748546705201408291055195244020	5 Peepdale Road	NULL	NULL	M40 0AN	407572278	B
12	119748543855201408291056889240327	2a, Farnie Drive	NULL	NULL	M20 5WS	377827278	D
13	13641305793201509521001605858835	30, Neville Drive	NULL	NULL	M26 4TL	3081709378	C
14	9826556230201602121203109858828	30, Longlevens Ro...	NULL	NULL	M22 1AX	7330628078	D
15	104923189910201312515325912672058	6, Glency Avenue	NULL	NULL	M22 5AY	214628178	E
16	36078503952014101059124149038374	5, Rumbold Street	NULL	NULL	M18 0SF	7441595688	D
17	1485231819062019110722565057168876	12, Beaminch Close	NULL	NULL	M13 9RW	751617478	C
18	96215698970201809361840714062488	45, Ames Street	NULL	NULL	M9 4WL	67452669178	C

2) Dropping Irrelevant Columns



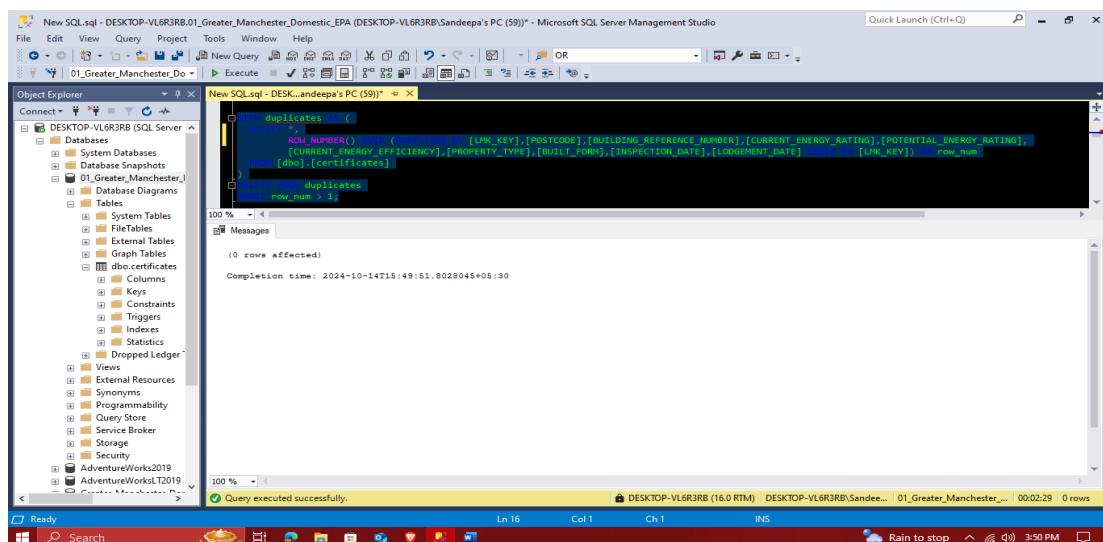
```
File Edit View Query Project Tools Window Help
Quick Launch (Ctrl+Q) P X
Object Explorer
Connect Databases Tables
System Tables FileTables External Tables Graph Tables
Columns Keys
ALTER TABLE [dbo].[certificates]
DROP COLUMN [ADDRESS1], [ADDRESS2], [ADDRESS3], [COUNTY], [LOCAL_AUTHORITY], [FLAT_STOREY_COUNT],
[MAIN_HEATING_CONTROLS], [SHEATING_ENERGY_EFF], [SHEATING_ENV_EFF],
[FLOOR_HEIGHT], [PHOTO_SUPPLY], [CONSTRUCTION_AGE_BAND],
[UPRN], [UPRN_SOURCE]
```

Messages

Commands completed successfully.

Completion time: 2024-10-14T15:38:06.5542106+05:30

3) Removing Duplicate



```
File Edit View Query Project Tools Window Help
Quick Launch (Ctrl+Q) P X
Object Explorer
Connect Databases Tables
System Tables FileTables External Tables Graph Tables
Columns Keys Constraints Triggers Indexes Statistics
Dropped Ledger
Views External Resources Synonyms Programmability Resource Store Service Broker Storage Security
AdventureWorks2019 AdventureWorksLT2019
ALTER TABLE [dbo].[certificates]
SELECT *, ROW_NUMBER() OVER (PARTITION BY [LPC_KEY], [POSTCODE], [BUILDING_REFERENCE_NUMBER], [CURRENT_ENERGY_RATING], [POTENTIAL_ENERGY_RATING], [CURRENT_ENERGY EFFICIENCY], [PROPERTY_TYPE], [BUILT_FORM], [INSPECTION_DATE], [LODGEMENT_DATE]) AS row_num
INTO #temp_duplicates
DELETE FROM #temp_duplicates
WHERE row_num > 1;
```

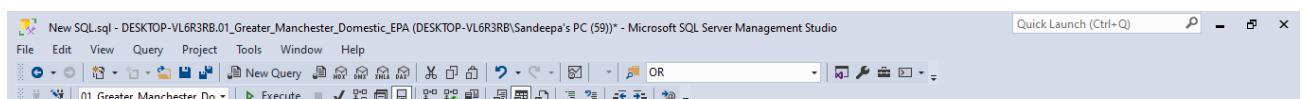
Messages

(0 rows affected)

Completion time: 2024-10-14T15:49:51.8028046+05:30

Query executed successfully:

4) Handling Missing Values



New SQL.sql - DESKTOP-VL6R3RB.01_Greater_Manchester_Domestic_EPA (DESKTOP-VL6R3RB\Sandeepa's PC (59)) - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

New Query Execute OR

Object Explorer

Connect Databases System Databases Database Snapshots 01_Greater_Manchester_ Database Diagrams Tables System Tables FileTables External Tables Graph Tables dbo.certificates Columns Keys Constraints Triggers Indexes Statistics Dropped Ledger Views External Resources Synonyms Programmability Query Store Service Broker Storage Security AdventureWorks2019 AdventureWorksLT2019

New SQL.sql - DESKTOP-VL6R3RB.01_Greater_Manchester_Domestic_EPA (DESKTOP-VL6R3RB\Sandeepa's PC (59))

Results Messages

	CURRENT_ENERGY EFFICIENCY
1	1
2	10
3	100

	POTENTIAL_ENERGY EFFICIENCY
1	1
2	10
3	100

	ENVIRONMENT_IMPACT_CURRENT
1	1
2	10

	ENVIRONMENT_IMPACT_POTENTIAL
1	1
2	10
3	100

	ENERGY_CONSUMPTION_CURRENT
1	1004
2	1008

	ENERGY_CONSUMPTION_POTENTIAL
4	1033
5	1007
6	103

	CO2_EMISSIONS_CURRENT
1	0
2	0.1
3	0.1

Query executed successfully.

DESKTOP-VL6R3RB (16.0 RTM) DESKTOP-VL6R3RB\Sandeepa's PC (59) 01_Greater_Manchester_Domestic_EPA 00:01:05 14,569 rows

Ready Search 80°F Light rain 3:58 PM

5) Data Standardization

New SQL.sql - DESKTOP-VL6R3RB.01_Greater_Manchester_Domestic_EPA (DESKTOP-VL6R3RB\Sandeepa's PC (59)) - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

New Query Execute OR

Object Explorer

Connect Databases System Databases Database Snapshots 01_Greater_Manchester_ Database Diagrams Tables System Tables FileTables External Tables Graph Tables dbo.certificates Columns Keys Constraints Triggers Indexes Statistics Dropped Ledger Views External Resources Synonyms Programmability Query Store Service Broker Storage Security AdventureWorks2019 AdventureWorksLT2019

New SQL.sql - DESKTOP-VL6R3RB.01_Greater_Manchester_Domestic_EPA (DESKTOP-VL6R3RB\Sandeepa's PC (59))

```
UPDATE [dbo].[certificates]
SET [FLOOR_LEVEL] =
CASE
    WHEN [FLOOR_LEVEL] IN ('06', '0') THEN '6th'
    WHEN [FLOOR_LEVEL] IN ('02', '2') THEN '2nd'
    WHEN [FLOOR_LEVEL] IN ('1', '-1', '01') THEN '1st'
    WHEN [FLOOR_LEVEL] IN ('3', '03') THEN '3rd'
    WHEN [FLOOR_LEVEL] IN ('4', '04') THEN '4th'
    WHEN [FLOOR_LEVEL] IN ('5', '05') THEN '5th'
    WHEN [FLOOR_LEVEL] IN ('6', '06') THEN '6th'
    WHEN [FLOOR_LEVEL] IN ('7', '07') THEN '7th'
    WHEN [FLOOR_LEVEL] IN ('8', '08') THEN '8th'
    WHEN [FLOOR_LEVEL] IN ('9', '09') THEN '9th'
    WHEN [FLOOR_LEVEL] IN ('10') THEN '10th'
    WHEN [FLOOR_LEVEL] IN ('11') THEN '11th'
    WHEN [FLOOR_LEVEL] IN ('12') THEN '12th'
    WHEN [FLOOR_LEVEL] IN ('13') THEN '13th'
    WHEN [FLOOR_LEVEL] IN ('14') THEN '14th'
    WHEN [FLOOR_LEVEL] IN ('15') THEN '15th'
    WHEN [FLOOR_LEVEL] IN ('16') THEN '16th'
    WHEN [FLOOR_LEVEL] IN ('17') THEN '17th'
    WHEN [FLOOR_LEVEL] IN ('18') THEN '18th'
    WHEN [FLOOR_LEVEL] IN ('19') THEN '19th'
    WHEN [FLOOR_LEVEL] IN ('20') THEN '20th'
    WHEN [FLOOR_LEVEL] IN ('20+') THEN '21st or above'
END;

SELECT DISTINCT [FLOOR_LEVEL] FROM [dbo].[certificates];
SELECT COUNT(*) FROM [dbo].[certificates] WHERE [FLOOR_LEVEL] IS NULL;
```

Results Messages

Query executed successfully.

DESKTOP-VL6R3RB (16.0 RTM) DESKTOP-VL6R3RB\Sandeepa's PC (59) 01_Greater_Manchester_Domestic_EPA 00:01:05 14,569 rows

Ready Search 80°F Light rain 4:01 PM

```

END;

select [FLOOR_LEVEL] from [dbo].[certificates];
SELECT count(*) as NULLs from [dbo].[certificates] where [FLOOR_LEVEL] is null;

UPDATE [dbo].[certificates]
SET [FLOOR_LEVEL] = 'NO DATA'
WHERE [FLOOR_LEVEL] is NULL;

----- 

alter table [dbo].[certificates] alter column [INSPECTION_DATE] DATE;
alter table [dbo].[certificates] alter column [LODGEMENT_DATE] DATE;

delete from [dbo].[certificates] where [INSPECTION_DATE] <'2013-01-01' or [INSPECTION_DATE] >='2024-01-01';

```

Messages

(261608 rows affected)

Completion time: 2024-10-14T17:08:33.9470951+05:30

Query executed successfully.

6) Date Consistency

```

alter table [dbo].[certificates] alter column [INSPECTION_DATE] DATE;
alter table [dbo].[certificates] alter column [LODGEMENT_DATE] DATE;

delete from [dbo].[certificates] where [INSPECTION_DATE] <'2013-01-01' or [INSPECTION_DATE] >='2024-01-01';

```

Messages

(92382 rows affected)

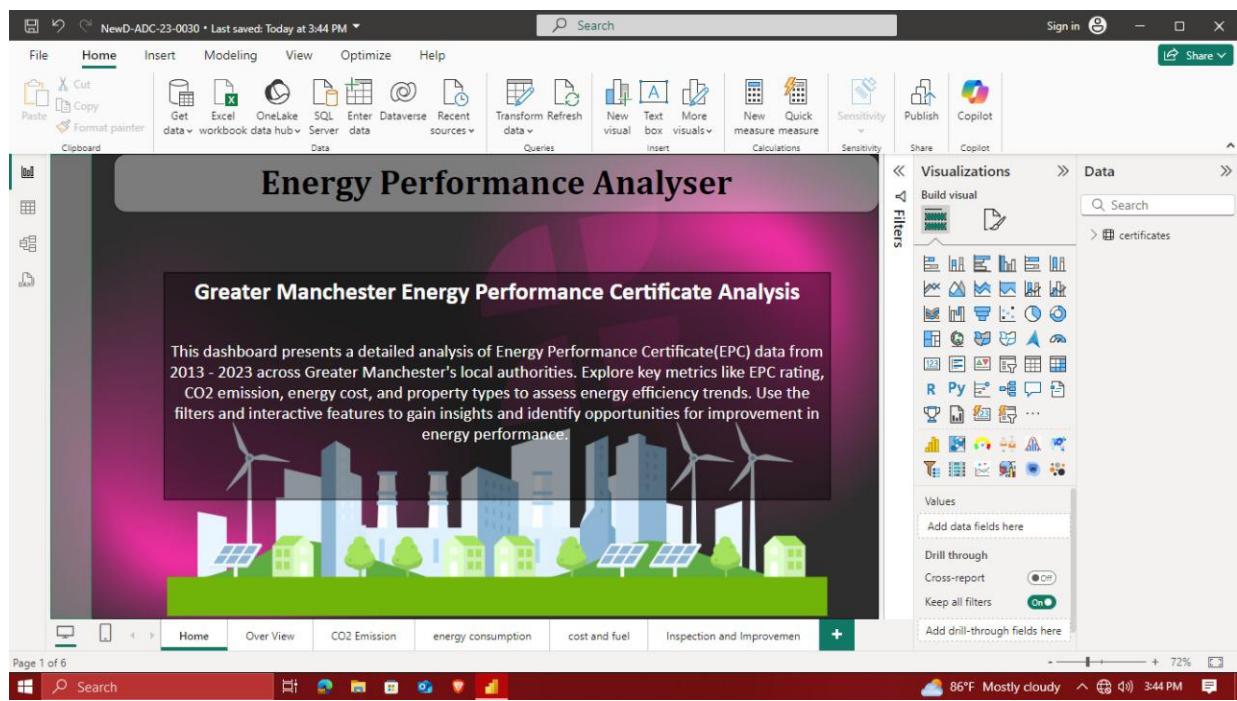
Completion time: 2024-10-14T17:11:22.3777436+05:30

Query executed successfully.

The remaining dataset is consistent, relevant, and prepared for more Power BI analysis after the data cleaning procedure. The dataset has been cleaned up, missing values have been handled, and formats have been standardized to provide useful insights into patterns in energy efficiency throughout Manchester's local authorities.

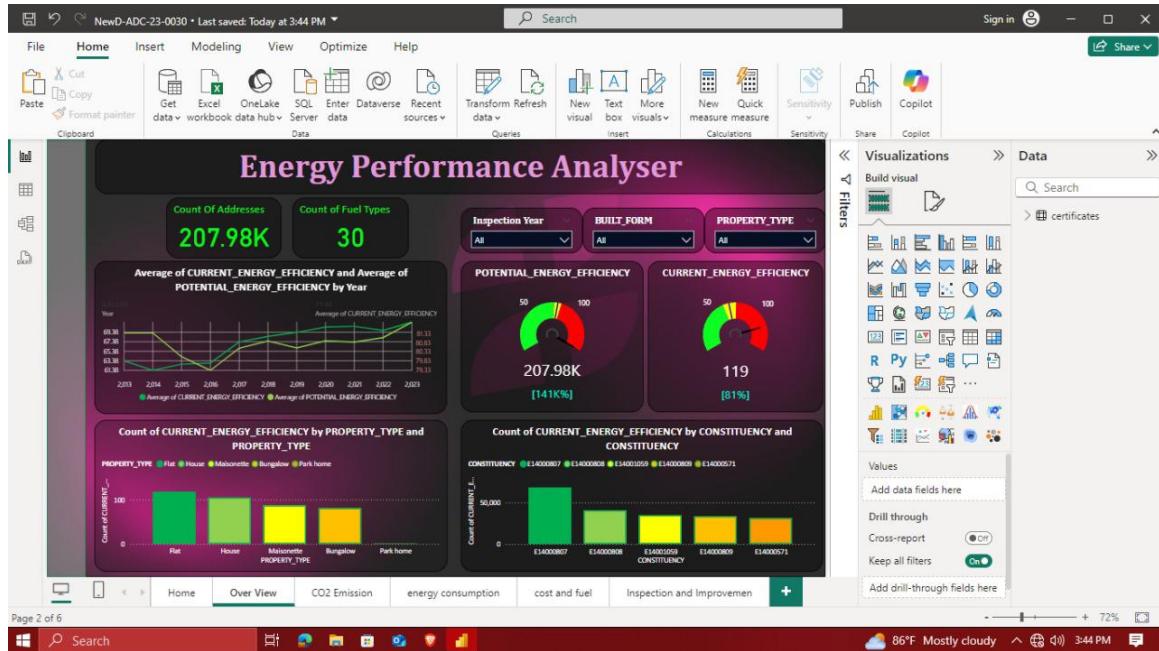
3) Dashboard Design & Implementation

● 01st Page of Dashboard

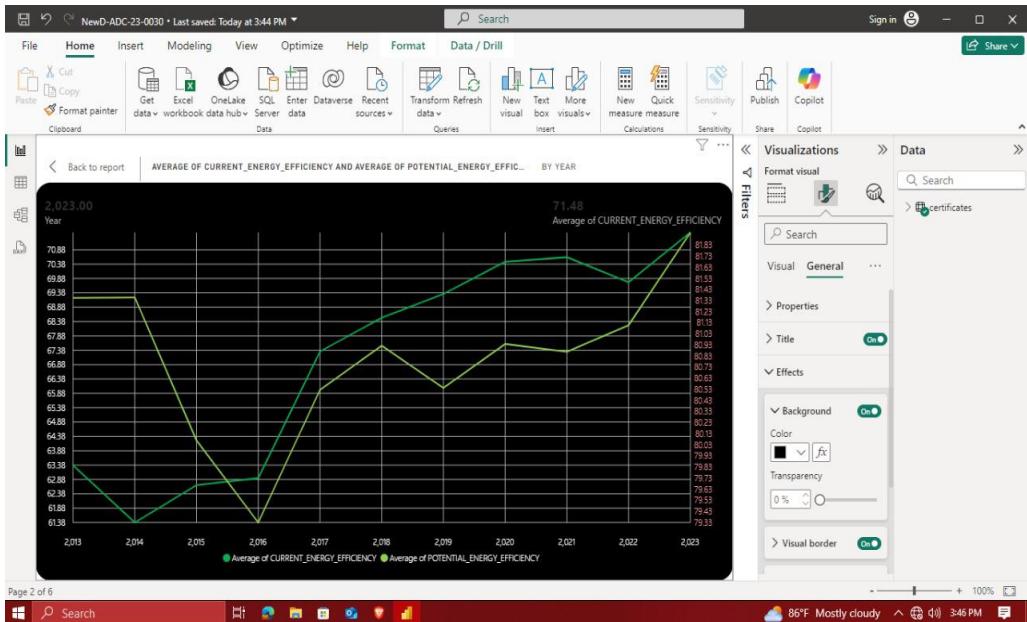


This first page of the dashboard depicts the Greater Manchester energy performance data for 2013 to 2023. The columns studied include EPC rating, CO2 emission, energy cost, and the property types. Consumers can then search and query the data through relevant radar and other forms, and see the current energy efficient status and prospect.

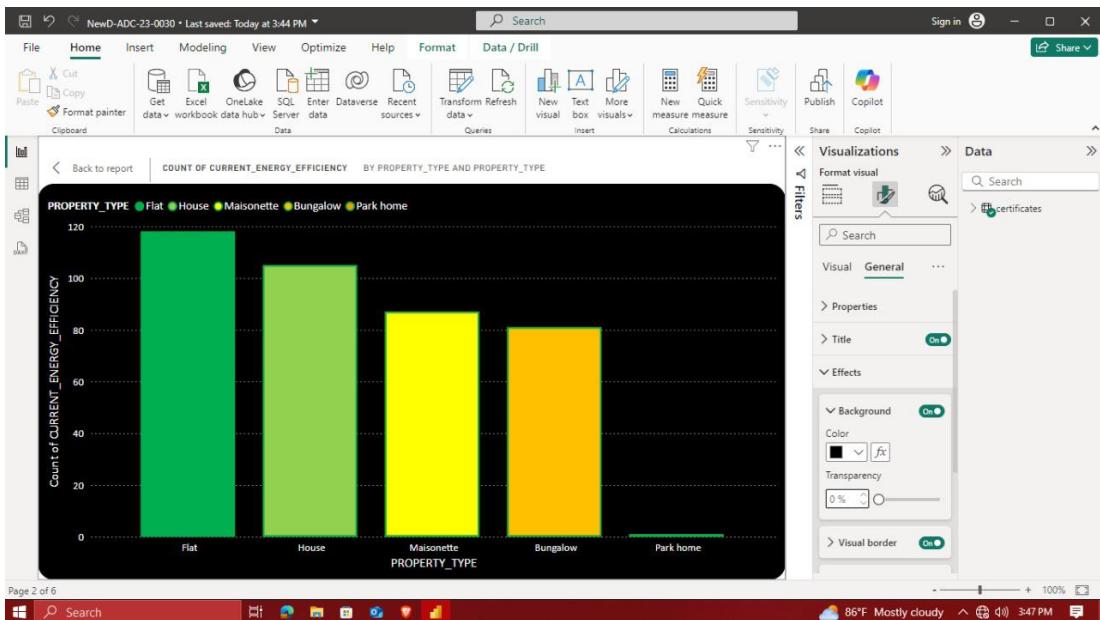
● 02nd Page of Dashboard



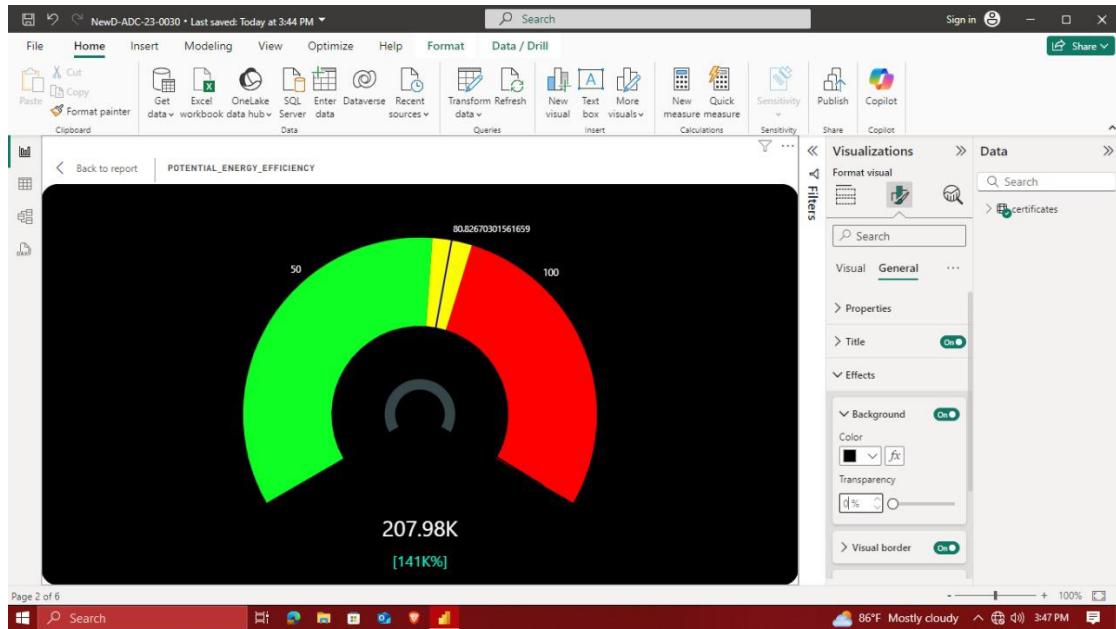
This page two focuses on the energy performance analysis. Specifically a bar graph which represents the Count of Current energy efficiency by property type is available along with a line graph which represents both the Average Current and Potential energy efficiency over some time interval. There is also a filter bar to select the distinct property type or a group of constituents to compare. In any case, this dashboard provides important information on various properties regarding their energy performance and improvement opportunities.



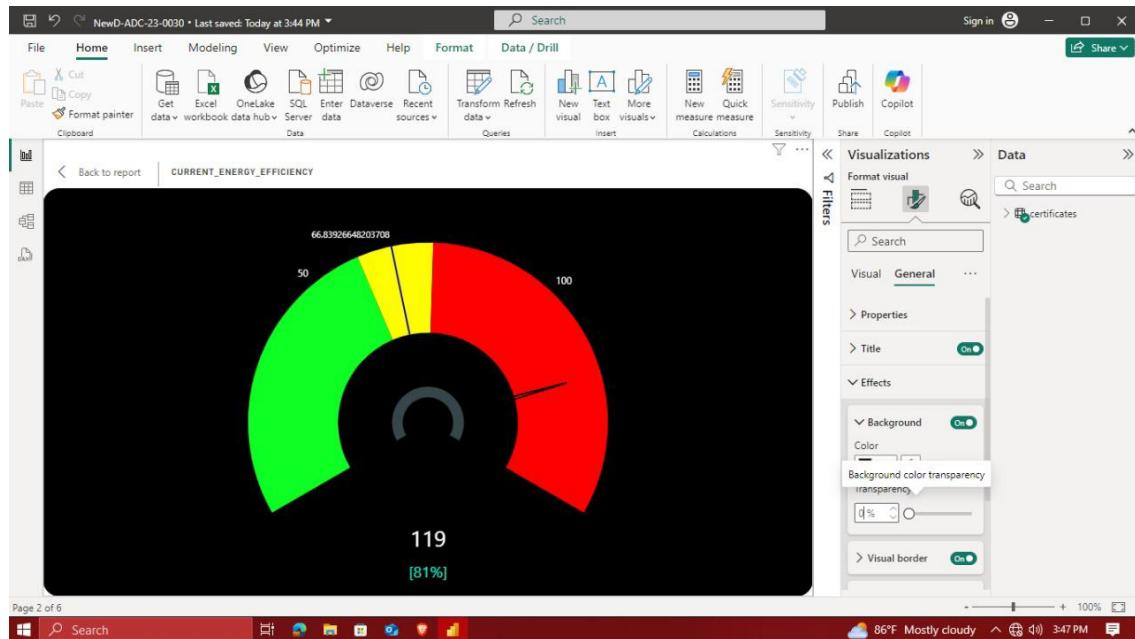
The line chart represents the average current energy efficiency and the average potential energy efficiency. On the horizontal axis, there is a temporal scale beginning at 2013 and ending at 2023, while the vertical axis shows average energy efficiency. It is clearly seen from the chart that the average current energy efficiency has been constantly rising over the years except in the years 2014 and 2015. The similar trend is observed for the average potential energy efficiency, which is however higher in this case than the current average energy efficiency. As a result this shows that there is an opportunity for energy savings for many of the properties.



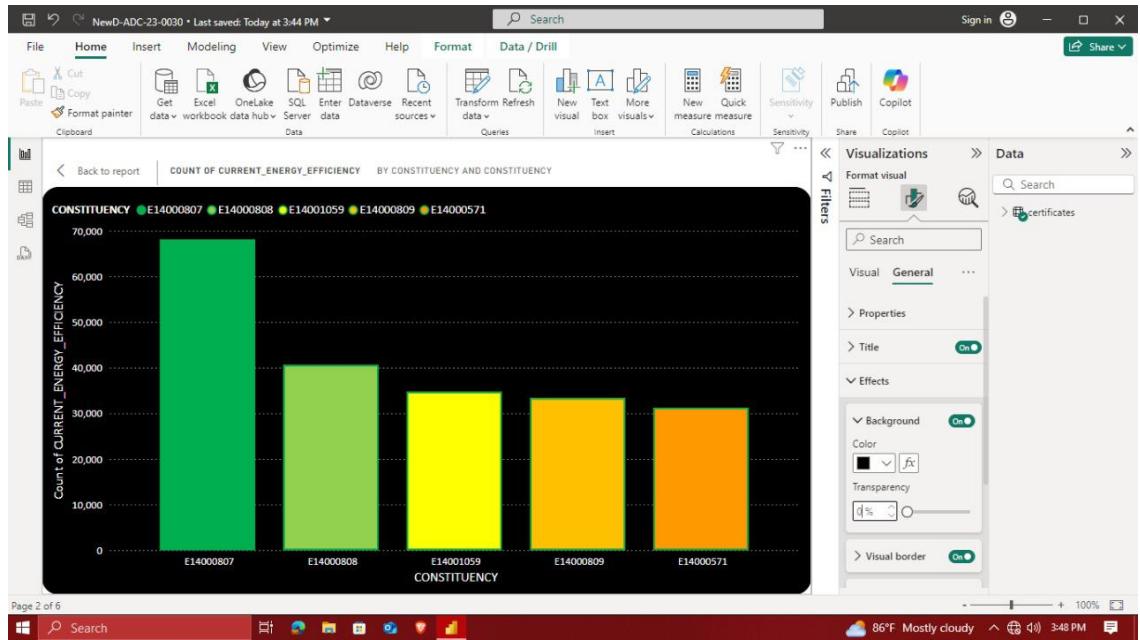
This bar chart represents the number of the current energy efficiency. The property types are Flat, House, Maisonette, Bungalow, Park home. The dependent variable is the count of current energy efficiency. According to this analysis, Flats as a property type seem to be relatively energy efficient than other types of properties in the dataset.



This gauge chart represents the potential energy efficiency with a scale of the range of 0-100; the green zone is from 81 to 100, the yellow zone is from 51 to 80 and the red zone is less than 50. This current value, 207.98 K is depicted within the chart and a percentage 141% is also indicated. This shows that probably the potential energy efficiency is very high, registering even more than the maximum potential indicated on the gauge chart. This implies that there is still much improvement that can be made regarding energy efficiency for the kinds of properties captured in the data.

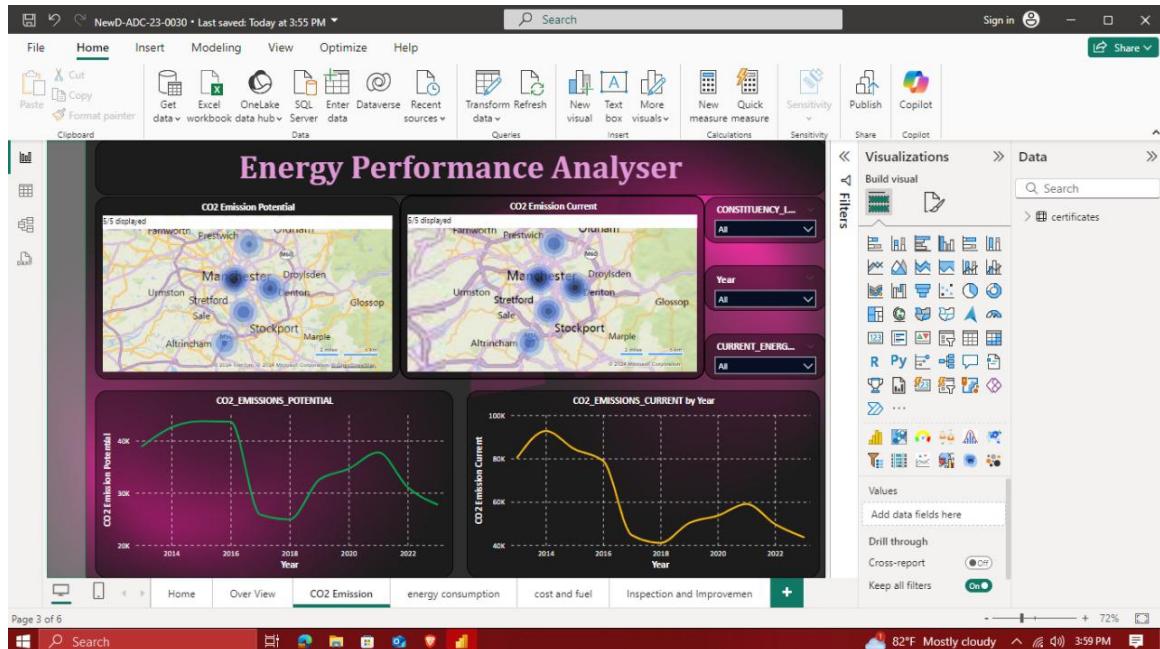


This gauge chart represents the current energy efficiency with a scale of 0-100, green representing 81-100, yellow 51-80 and red below 50. The current value is also indicated as 119 within the chart while the percentage is at 81%. This means that the current efficiency is high, within the green segment meaning that the energy performance is optimal or close to the optimum.

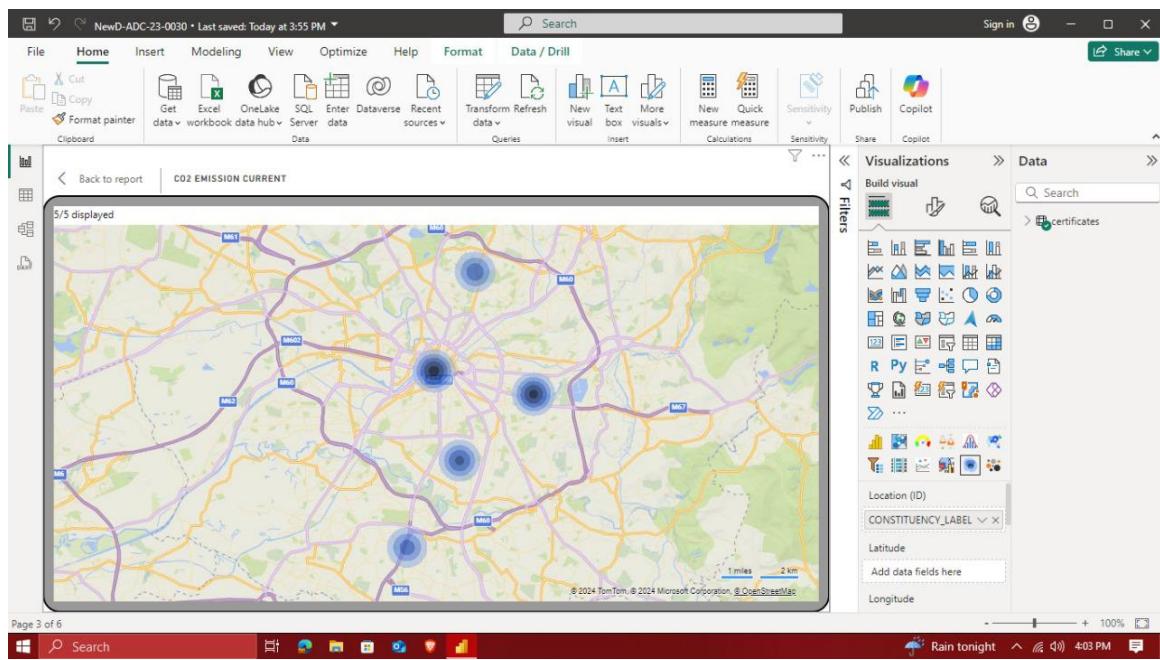


This bar chart visualises the count of current energy efficiency by constituency. The x-axis represents different constituencies and the y-axis represents the count of current energy efficiency. The chart indicates that constituency E14000807 has the highest count.

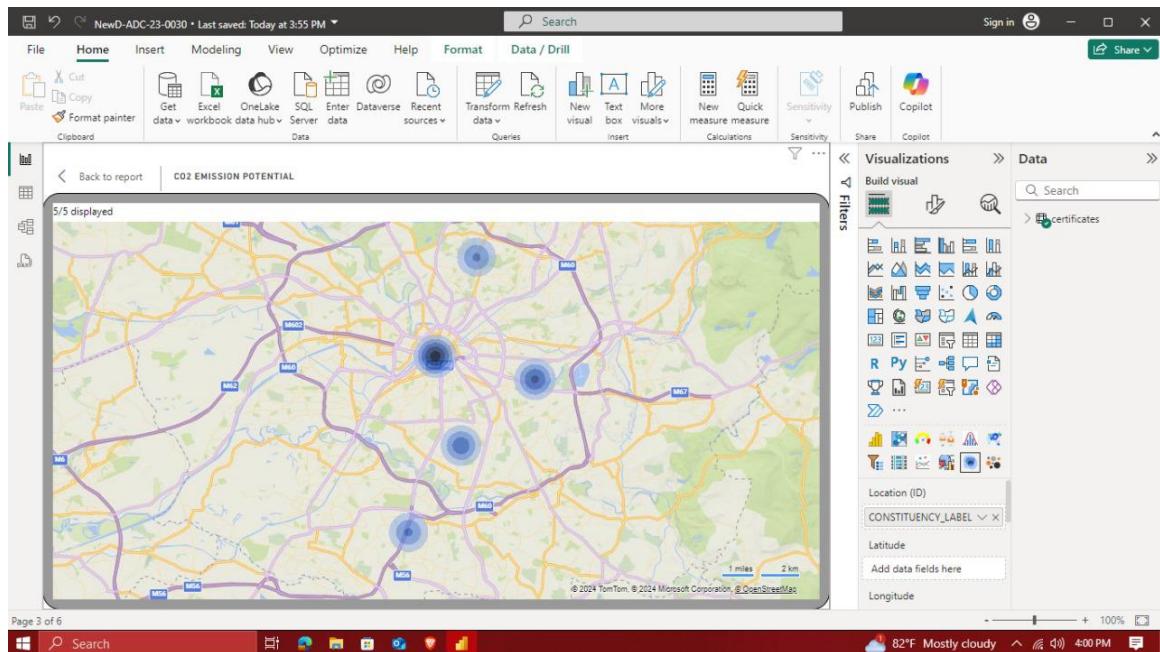
03rd Page of Dashboard



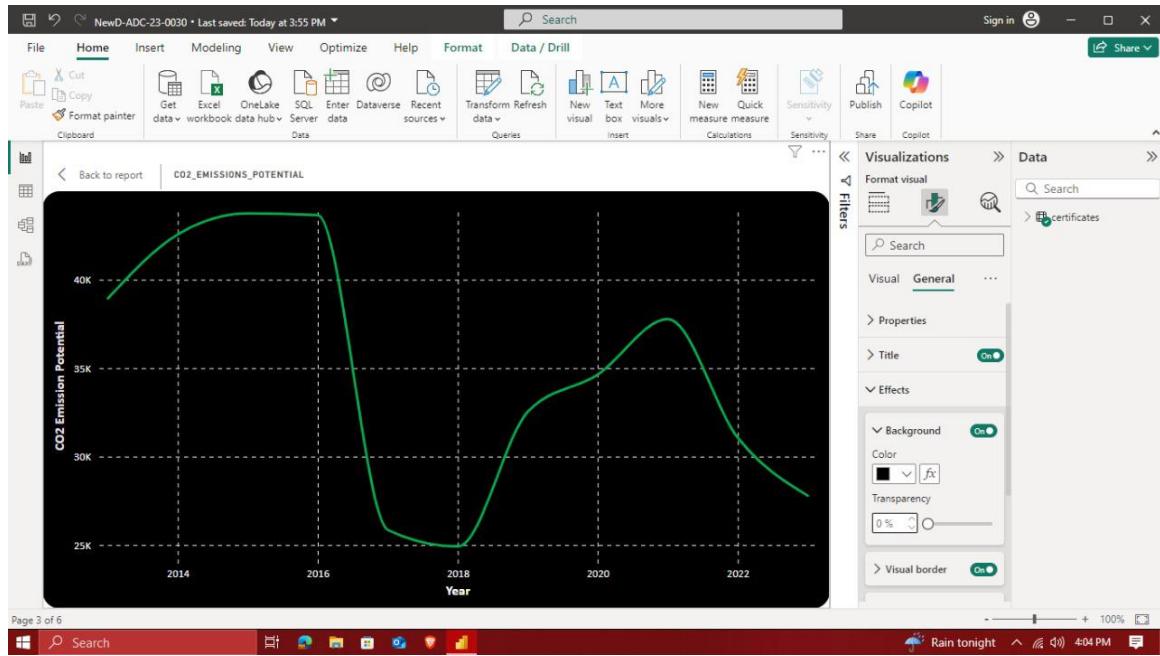
This gives a visualization of CO2 emission data. The dashboard has more features including maps : illustrating the CO2 emission potential and current by constituencies, line charts showing the cumulated and probability of having high CO2 emission potential and current by years. The dashboard also come with a filter bar where one can filter for any specific composition or period for comparison.



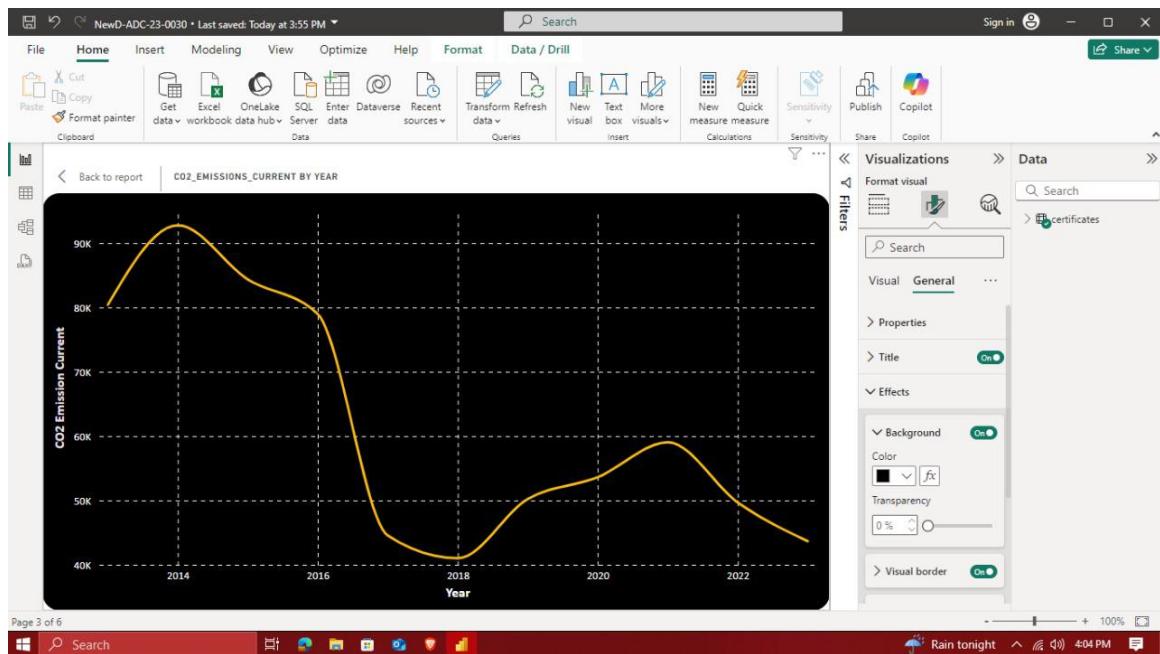
This map displays the CO₂ emission current by constituency where different places are marked as constituencies. The size of bubbles in the map gives the information about the current CO₂ emission for each of the constituency present at the particular location. It is clear from the chart that some constituencies produce higher levels of CO₂ as compared to the others, it may mean that these constituencies need some extra effort making them environmentally friendly.



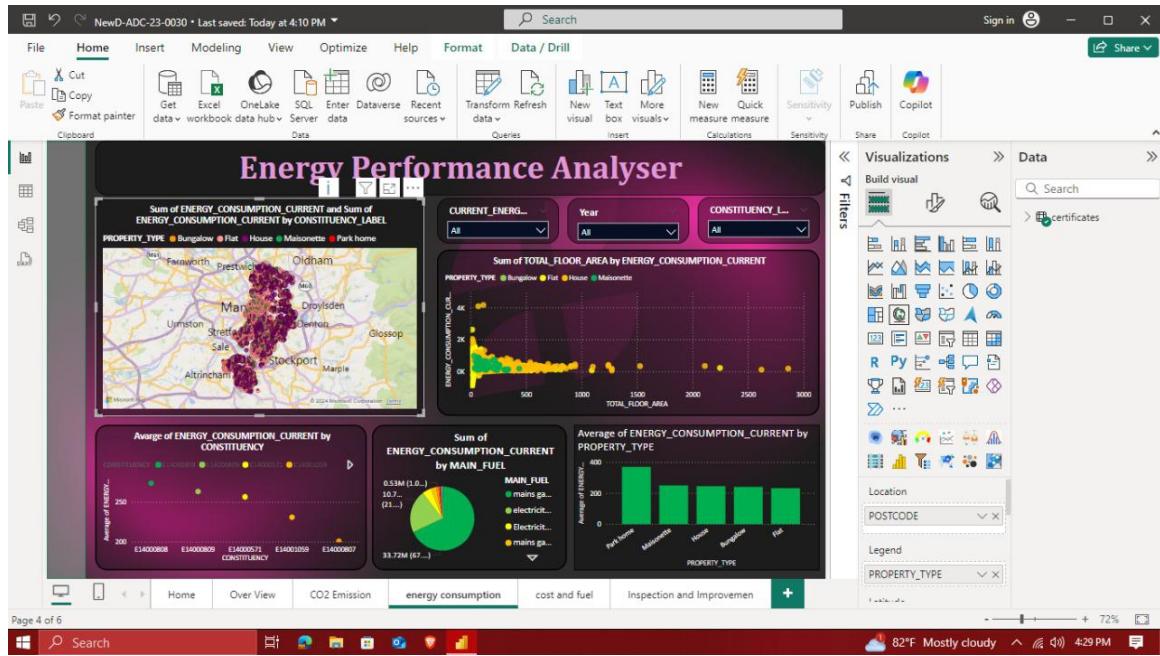
The is also a map, which may indicate information concerning CO₂ emission potential. The map is superimposed upon another map which can be referred to as the background map because it gives context such as roads, buildings and water bodies.



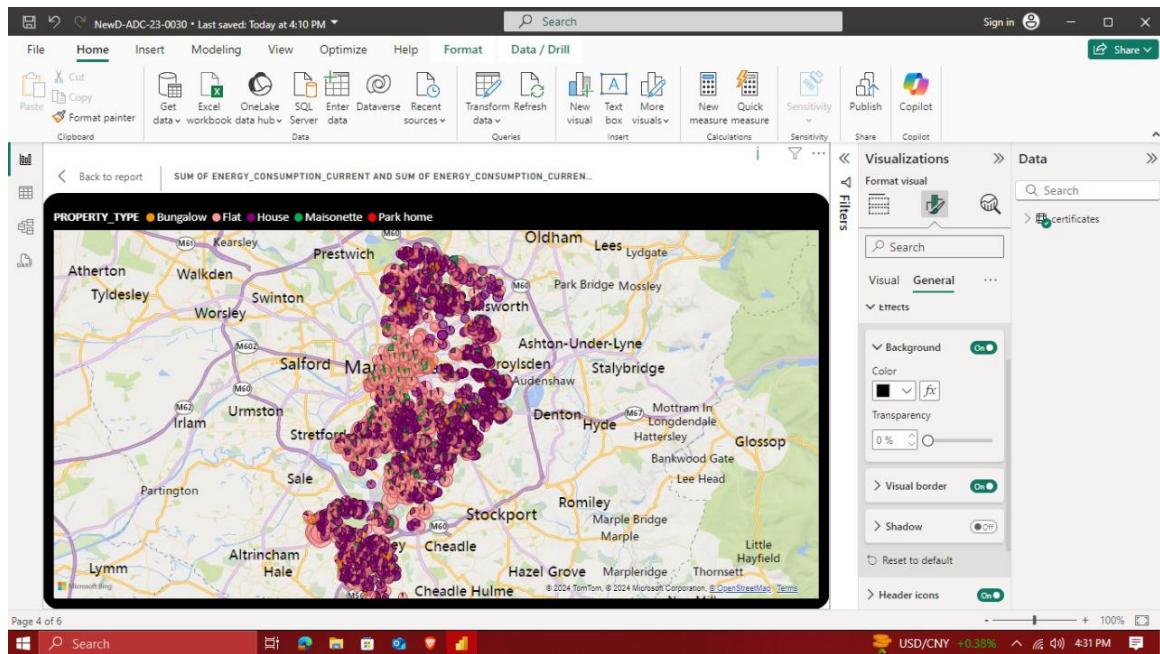
This looks a line chart , which represents the CO2 emission potential in different time points. The horizontal axis is the years from 2014 to 2021, while the vertical axis is the CO2 emission potential. The chart trends of emissions suggests a relatively stable level from 2014 to 2016, a slight drop in emissions from 2016 to 2018, and then again a rise to emissions in 2021.



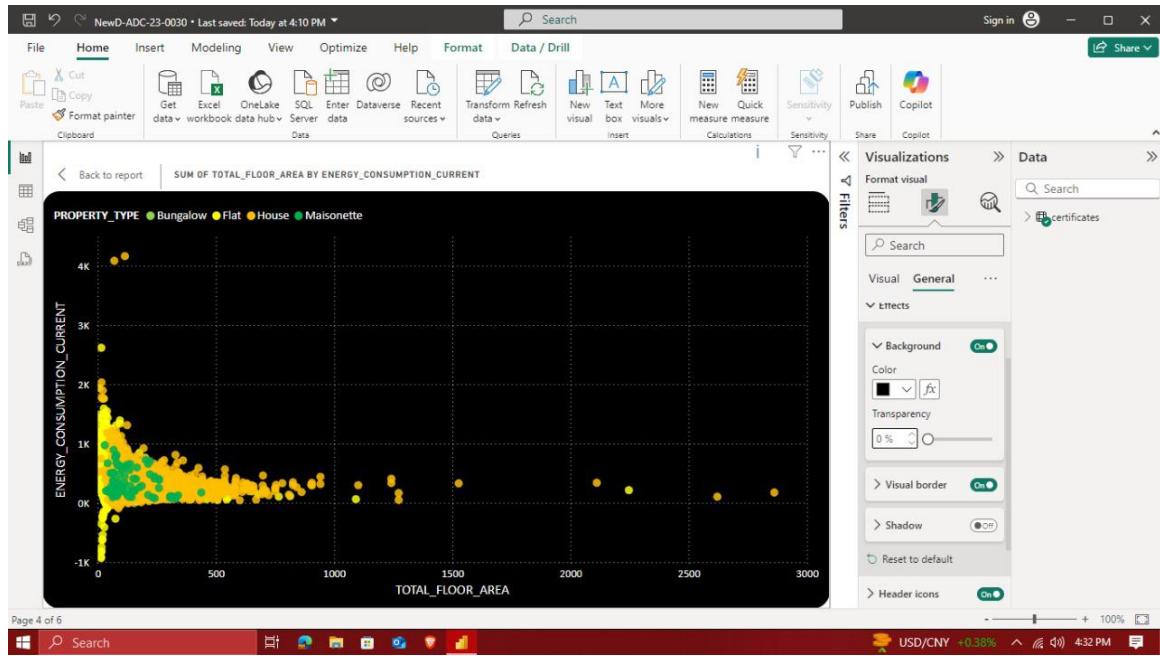
This line chart gives an insight into the emissions of CO2 in the range of 2014-2022. This shows that CO2 emissions over the various years with a general increase in the emission up to the year 2018, a decrease in the year 2020 and then the slight increase again in the year 2022.



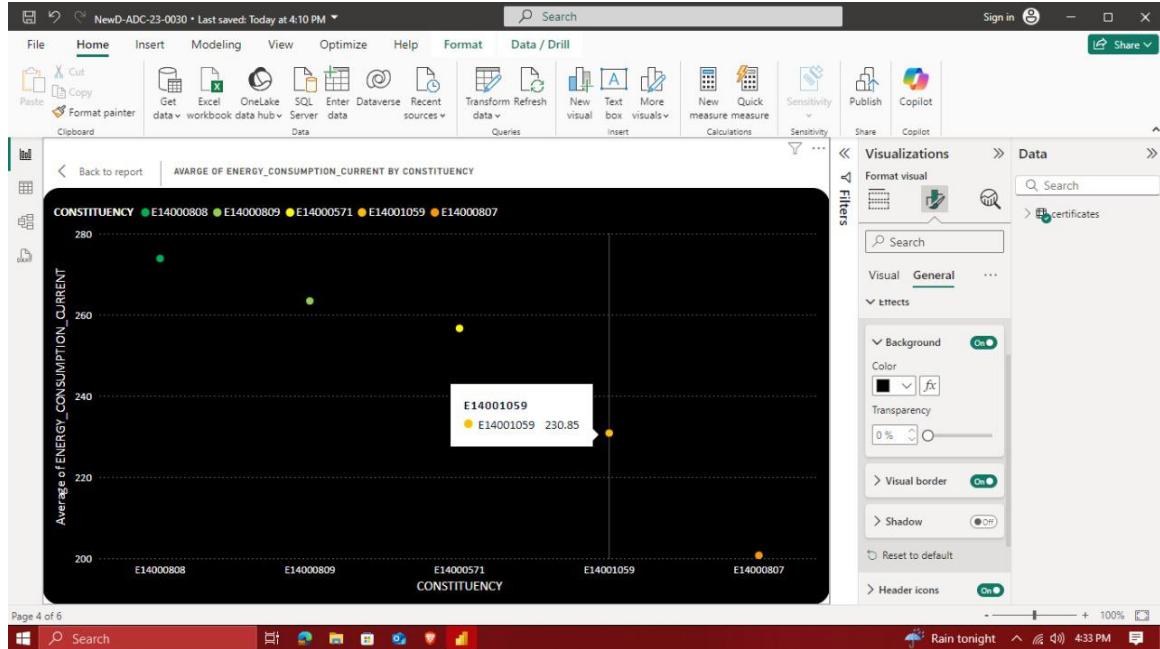
This dashboard page includes: Energy consumption map , bar chart, pie chart, scatter plots and filter bars for choosing property types or constituent categories of interest. In general, it has useful information in relation to energy consumption of some properties.



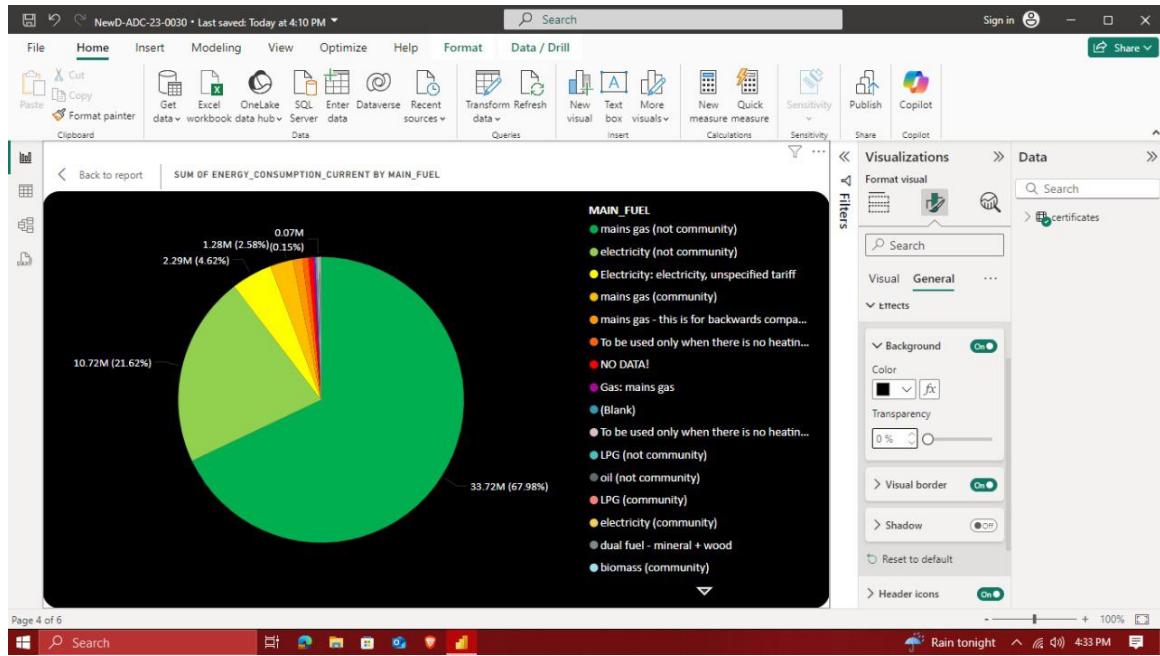
This map provides the geographic environment, including roads, buildings and water bodies. In general, the structure of the analysed dashboard is coherent and allows receiving essential and comprehensible insights from the data, that makes the use of the dashboard efficient for the analysis of the productivity and decision-making.



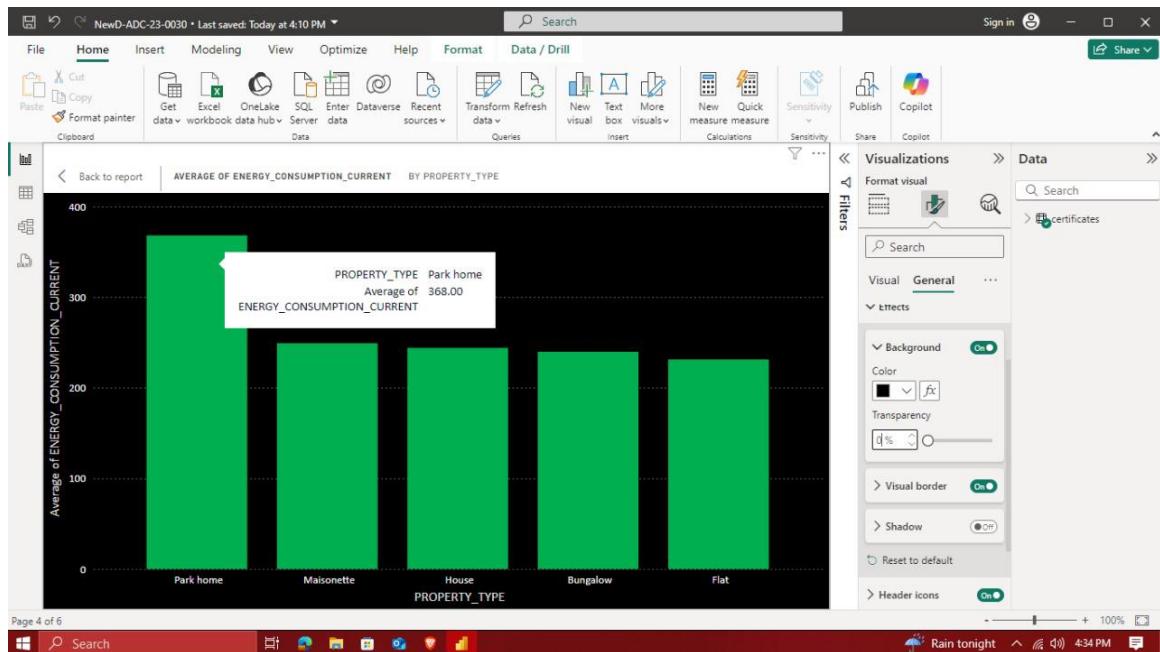
This scatter plot shows the correlation of total floor area zone with the current na energy consumption. The x-axis set is the total building floor area and the y-axis set is the total building energy use. The dots' colours stand for the kinds of properties under offer.



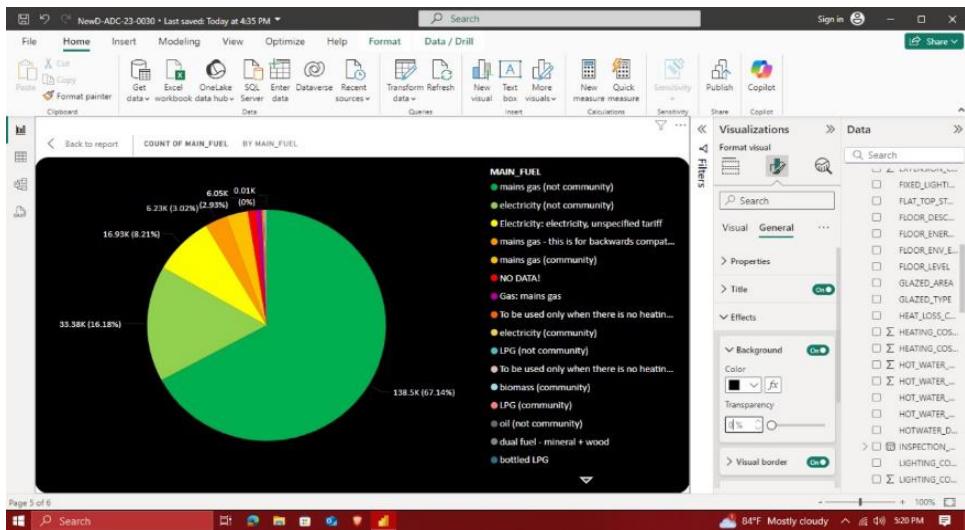
This bubble chart shows the average current energy consumption by constituency. The constituencies are located on the x axis while the y axis indicates the average energy consumption.



This is a pie chart that gives the breakdown of energy consumption by major fuel type. It is not one chart of multiple segments but divided based on categories of fuel types. The largest segment is for oil (not community) and is followed by electricity (not community) and mains gas (not community).

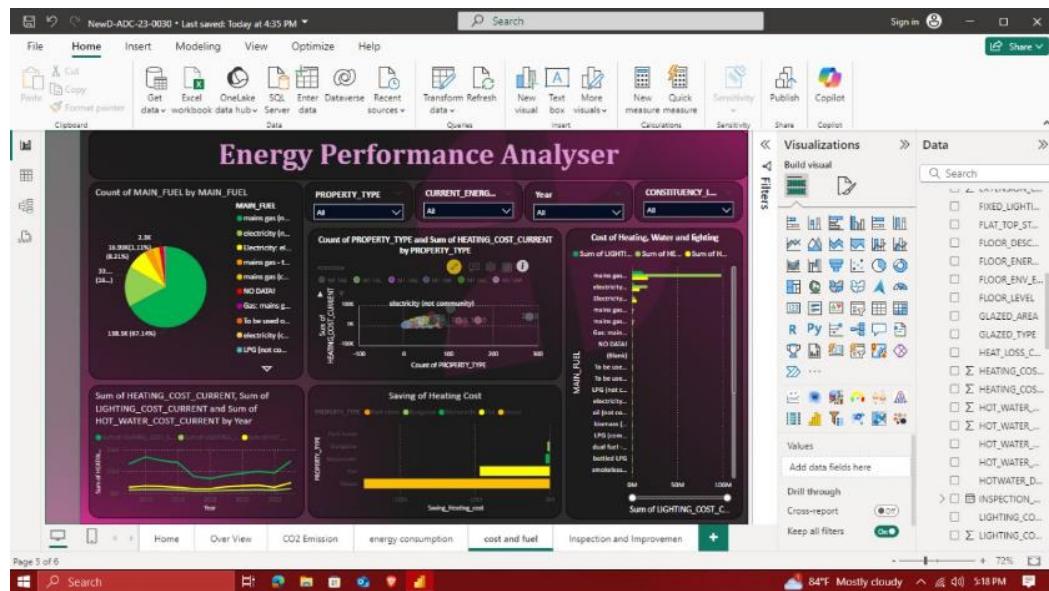


This column chart represents the average of current energy consumption by the property type. The horizontal axis is the property types which include Park home, Maisonette, House, Bungalow, and Flat, while the vertical axis is average energy consumption. This shows that the modern Park homes take the lead in energy consumption than others.

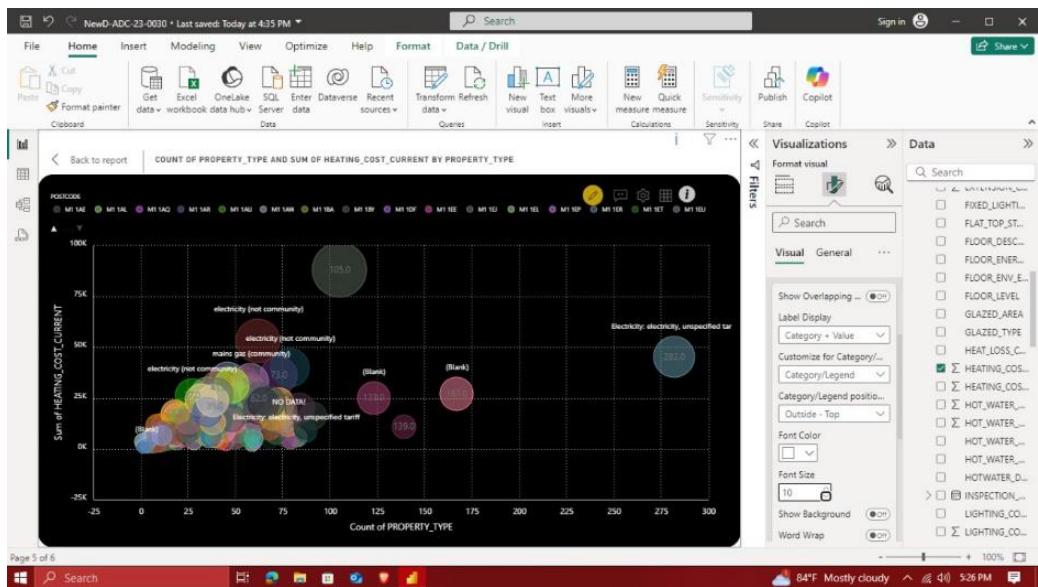


The chart shows the distribution of main fuel types used in a particular area. The most common fuel type is mains gas (not community), accounting for 67.14% of the total. Comparing the fuel types, mains gas is the most widely used fuel, suggesting its availability and affordability in the area. Electricity is the second most common fuel, highlighting the importance of electricity infrastructure in the area. The high proportion of LPG (not community) suggests a dependence on bottled gas for those who are not connected to the mains gas network. The relatively low proportion of biomass (community) and other alternative fuels suggests that these options may not be as widely adopted in the area.

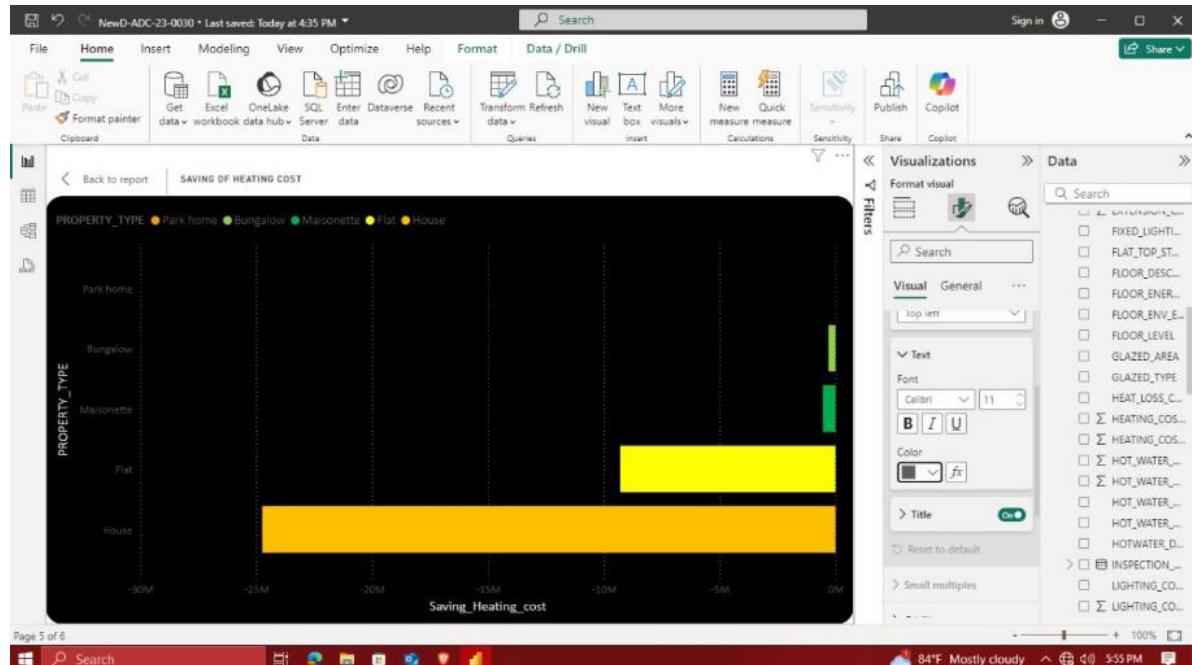
05th Page of Dashboard



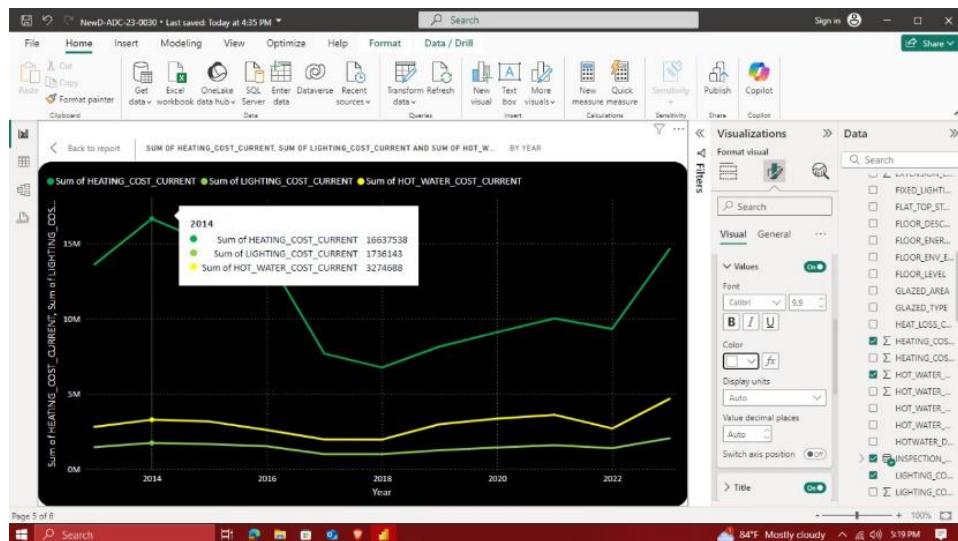
This page contains various charts and tables that are focused on different aspects of energy consumption, including main fuel types, property types, heating costs, lighting costs, hot water costs, and saving of heating costs. It also includes a visualization builder and filters for analysing data..The dashboard can be used to identify areas where energy consumption can be improved, such as by optimizing the use of specific fuel types or by implementing more energy-efficient heating systems.



The chart shows the count of property types and the sum of heating cost current for each property type. This scatter plot consists from the count of property type on the x-axis and the sum of heating cost current on the y-axis. The chart shows that the property type with the highest count is "Electricity: electricity, unspecified tariff", with a count of around 280. The property type with the highest sum of heating cost current is "M1 1EP", with a sum of around 105. The chart also shows that there is a positive correlation between the count of property type and the sum of heating cost current. This means that as the count of a property type increases, the sum of heating cost current also tends to increase. This can be used to identify property types that are more expensive to heat, and to compare the heating costs of different property types. Further analysis can be done to understand the reasons for these differences.

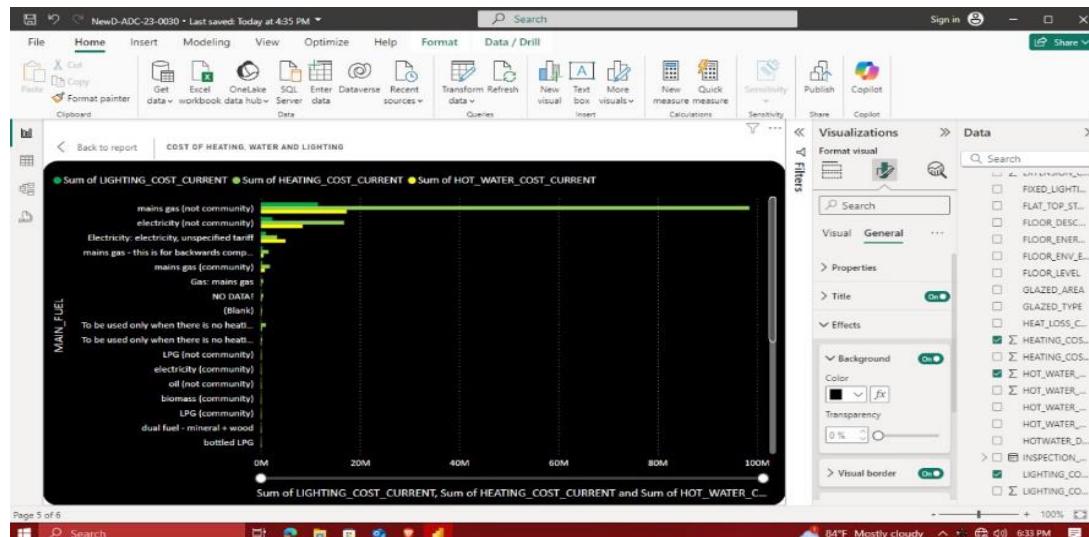


The chart shows the saving of heating cost for different property types. The property types are: Park Home, Bungalow, Maisonette, Flat, and House. The chart shows that the largest savings were achieved for houses, followed by flats. The smallest savings were achieved for park homes. This data suggests that houses are the most energy-efficient property type, followed by flats. Park homes are the least energy-efficient property type. It is important to note that this is just a snapshot of the data and that there are many other factors that could affect the energy efficiency of a property.



This line chart shows the sum of heating costs, lighting costs, and hot water costs over the years 2014-2022. The lines for each cost category show a general upward trend, although there are some dips and rises in the lines. The total cost of heating has the steepest slope, indicating the most growth in this cost category over the years.

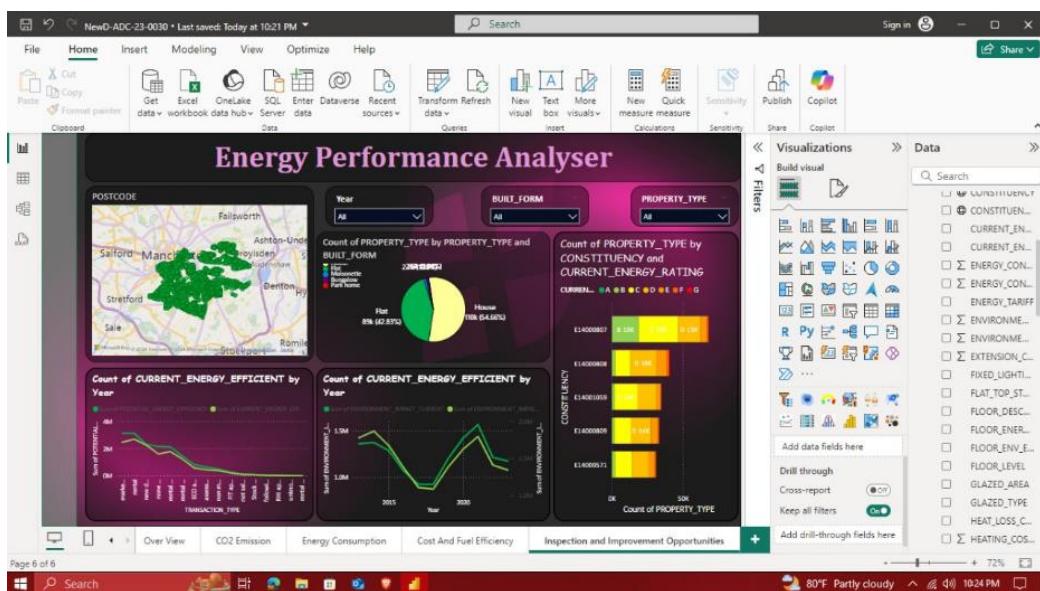
- Heating cost: The heating cost started at around 13 million in 2014, peaked around 17 million in 2014, then dropped to 14 million in 2016, and again dropped to 7 million in 2018. The heating cost steadily increased to around 14 million in 2022
- Lighting cost: The lighting cost started at around 3 million in 2014, increased slightly and plateaued around 4 million in 2016, and then continued to increase reaching 4 million in 2022.
- Hot water cost: The hot water cost started at around 2 million in 2014, decreased to 2 million in 2016, and again decreased to 1.5 million in 2018. The hot water cost started increasing again and reached 4 million in 2022.



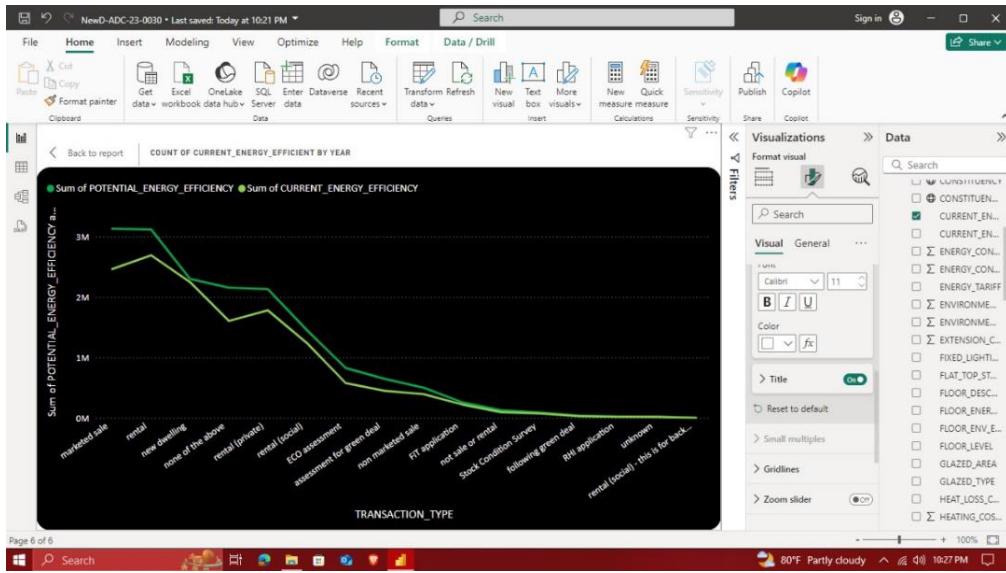
This bar chart displays the cost of heating, water, and lighting by fuel source. The x-axis shows the cost in millions of dollars, and the y-axis shows the fuel source. The bar chart is segmented into three parts, each representing a different cost category: lighting cost, heating cost, and hot water cost. The largest cost is from hot water, followed by heating. The lighting cost is the smallest. The bar chart shows that the most expensive fuel source for heating is mains gas followed by electricity. The least expensive fuel source for heating is bottled LPG. Overall, the chart shows that the cost of heating, water, and lighting is highly variable depending on the fuel source used.

- Hot water is the most expensive type of energy use, followed by heating and then lighting.
- The cost of hot water is significantly higher than the cost of heating and lighting.
- The cost of heating and lighting is relatively similar.
- The data panel shows that the user can drill down into the data to get more information about the different types of fuel and their costs. For example, they could drill down into the "Hot water" category to see which types of fuel are used for hot water, and how much each type of fuel costs.

06th Page of Dashboard

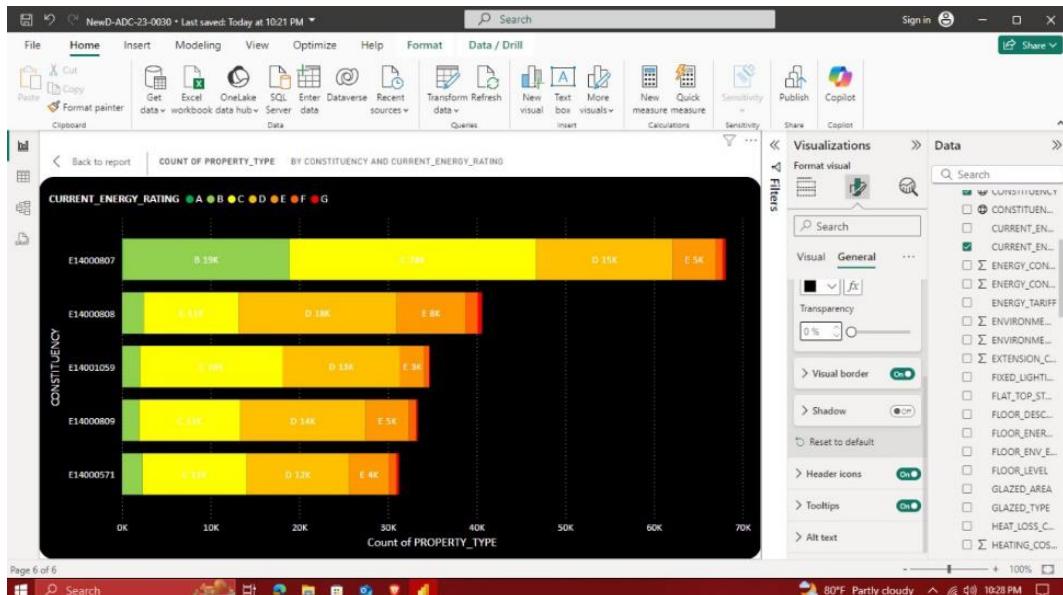


The image shows a dashboard of a building energy performance analysis. It includes a map of buildings, data on energy consumption, and charts showing the count of different property types, energy ratings, and built forms. The user is able to interact with the data by filtering, drilling through data points, and creating custom visualizations. The dashboard provides insights into the energy performance of buildings, allowing users to identify areas for improvement and track progress over time.



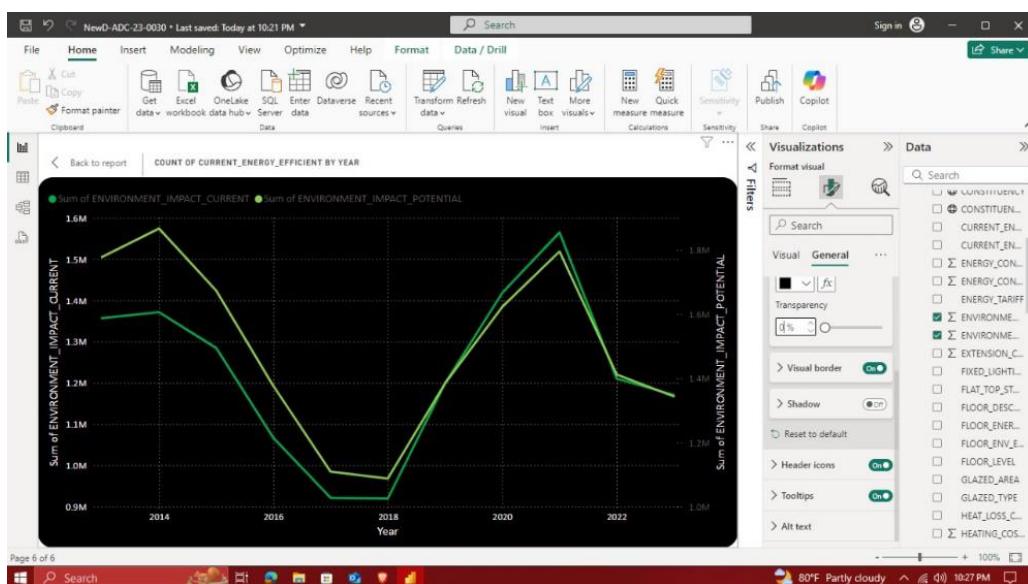
This line chart shows the sum of potential energy efficiency and the sum of current energy efficiency by transaction type. The chart shows that the sum of potential energy efficiency is higher than the sum of current energy efficiency for all transaction types. The report also includes a filter pane, which allows the user to filter the data by different criteria.

- Trend: The chart shows a general trend of decreasing energy efficiency, with the highest values for "marketed sale" and the lowest for "unknown".
- Comparison: It's clear that the "potential energy efficiency" is much higher than the "current energy efficiency" for all transactions. This suggests a gap between potential and actual energy saving, highlighting the need for improvement.
- Categorization: The data is categorized by transaction type, which helps understand energy efficiency across different property markets or deal types.
- Filters: The filter pane allows for deeper analysis by selecting specific transaction types or other criteria, providing a more granular view of the data.



This bar chart displays the count of property types, broken down by constituency and energy rating. The chart uses a colour scheme to represent the energy rating, with each colour representing a different rating, from A to G.

- Most Common Rating: The most common energy rating across all constituencies is "C," indicating that a significant number of properties fall into this category.
- Constituency Differences: There are variations in the distribution of energy ratings across constituencies. For instance, E14000807 has a higher concentration of properties with rating "B" compared to other constituencies.
- Trends: The chart suggests that the number of properties with higher energy ratings (A-B) is generally lower compared to those with lower ratings (D-G). This could imply a higher proportion of older or less energy-efficient properties in these areas.
- This visualization provides valuable insights into the energy efficiency of properties within different constituencies. The data can be used for various purposes, including:
- Energy Policy: Identifying areas with a higher proportion of less efficient properties can help prioritize energy retrofitting programs and initiatives.

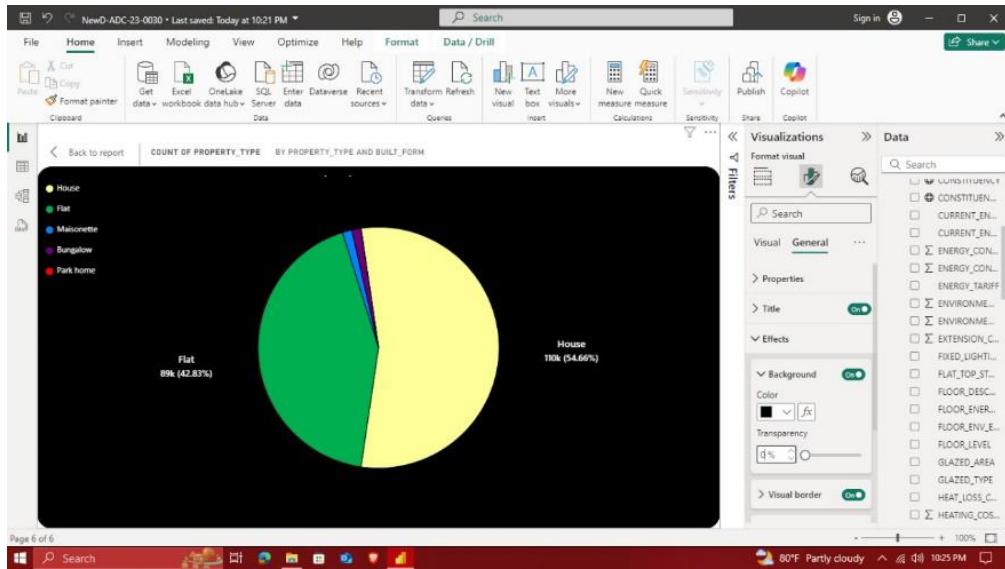


This chart shows the sum of Environment Impact - Current and Environment Impact - Potential from 2014 to 2022.

The Environment Impact - Current had a peak in 2014 and declined from then on. It bottomed out in 2018, but has been steadily increasing since.

The Environment Impact - Potential also peaked in 2014. However, it declined sharply until 2018 and then bounced back. It peaked in 2021, but then declined again.

While Environment Impact - Potential is higher than Environment Impact - Current for most of the chart, the gap between the two metrics has shrunk.



There are six property types represented in the pie chart: Flat, House, Maisonette, Bungalow, Park home, and Enclosed Mid-Terrace. The pie chart shows that the most common property type is House, followed by Flat. The least common property type is Enclosed Mid-Terrace.

4) Conclusion

dashboard provides a comprehensive overview of energy performance in Greater Manchester. By analyzing data on property types, energy ratings, and other factors, it can help identify areas where improvements can be made. Based on the visualizations, we can see that most properties in the area are flats. We can also see that the energy consumption of the properties has been increasing over time.

5)References for Task 02

<https://community.fabric.microsoft.com/t5/Themes-Gallery/Energy-Consumption-Report/m-p/4004452>

https://youtu.be/QdKOqlD_3jw?si=qWbku55LYKiyCyBF

<https://youtu.be/xCSYLrcLW00?si=VPBy1SVXsvXjluz1>