General Sir John Kotelawala Defence University

Faculty of Management, Social science and Humanities
Department of Languages
BSc. in Applied Data Science Communication

LB 2114: Advance SQL
Year: 2 Semester 4
Assignment 02

Group members

D/ADC/23/0008	**MAW Sammani**
D/ADC/23/0022	**HK Ahamat**
D/ADC/23/0026	**DMS Baddewithana**
D/ADC/23/0036	**MBSD Fernando**

CONTENTS

Task 01

CHILD WELL-BEING MONITOR USING MICROSOFT REPORT BUILDER

AND MICROSOFT SQL SERVER TO ANALYZE CHILD POVERTY IN LOWER INCOME COUNTRIES.

1.1. Introduction

The Young Lives study is the only available longitudinal survey on the dimensions of both poverty and inequality. It has been tracking 12,000 children in Ethiopia, India, Peru, and Vietnam since 2001; its leading team is based at the Department of International Development at the University of Oxford. The study attempts to follow the continuous lives of 12,000 children over 15 years of continuous change while data collection occurs once every three to four years. Young Lives has done Round 1 by interviewing two cohorts of children in each country at age one and five years. The same children were interviewed for round two aged between 5-12 years. At 7-8 years and 14–15 years old in Round 3 when they were 12 and 19 years old in Round 4 and when they were 15 and 22 years old in Round 5. It is run by the Young Lives team at the University of Oxford with funding from the Department for International Development. (A Guide to to young lives reserch, 2017)

The following report selects India and Ethiopia to describe and visualize, with information to show data on the Young Live Project for four key issues: education, health, social, and wellbeing. We have been assigned the task based on the requirement for knowledge in designing, implementation, and testing of a reporting system on this project working on SQL Server and Microsoft Report Builder. (Young lives, 2016)

1.2. Downloading the dataset

The dataset was downloaded from the link below.

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

We first signed up to the UK Data Service and requested a username. From this point forward, when we gained access to the UK Data Services, we downloaded the International Study of Child Poverty: Rounds 1-5 Constructed Files, 2002-2016 dataset; from this, we were able to upload the dataset into our account using the username and password created upon setup of our login within the UK Data service. Later, the abstract related to our work had to be submitted. The same has been done, and the downloading of the dataset is also complete.

1.3. Exploring the dataset

Archived with the unprocessed datasets, and to facilitate use of the Young Lives data by researchers, are 'constructed files'. This set is combined sub-sets of variables from Rounds 1 to 5 of the Young Lives household and child surveys which took place between 2002 and 2016. These data files represent an update of the previously archived Rounds 1 to 4 constructed files. All the variables are kept in the data, but in a few cases, some computations needed adjustment, and their variables were substituted to attain consistency with the definitions of previous rounds.

All these were converted to CSV format in a step-by-step process. Each of the tab files was opened in a text editor like notepad and then each was saved in text format. We used the facilities provided by Microsoft Excel since the files were now in text format. Each of these text files was imported in a different Excel worksheet and after importation each was saved as a CSV file.

This allowed us to transform each one of those sets into tabular data in CSV format, hence granting much easier access and usability: name, description, and number of rows in the csv file.

File name	No of rows	Content
all_countries_math_irt_scores	32632 rows	All Countries Mathematics Item Response Theory Scores Data Files

all_countries_ppvt_irt_scores	34010 rows	All countries Peabody picture vocabulary test item response theory score data file
all_countries_reading_irt_scores	17451 rows	All countries Reading Item response theory score data file
et_in_pe_sibling_ppvt_irt_scores	6164 rows	Ethiopia, India and Peru sibling Peabody Picture Vocabulary Test item response theory scores data file
ethiopia_constructed	14995 rows	Data file of Ethiopia
india_constructed	15097 rows	India created data file
india_sibling_math_irt_scores	3023 rows	India sibling math item response theory scores data file
peru_constructed	13830 rows	Peru data file created
vietnam_constructed	15000 rows	Vietnam Created data file

Other variables were also added to give the researchers a background insight into the Young Lives data and to represent the applicable questions asked of the household and children in the latter rounds (Azubuike & Briones, 2016).

These built files divide variables into four general categories: identification and location variables; panel information; child characteristics; and household characteristics.

- Identification and locating variables
- Panel Details
- Child characteristics

- housing characteristics

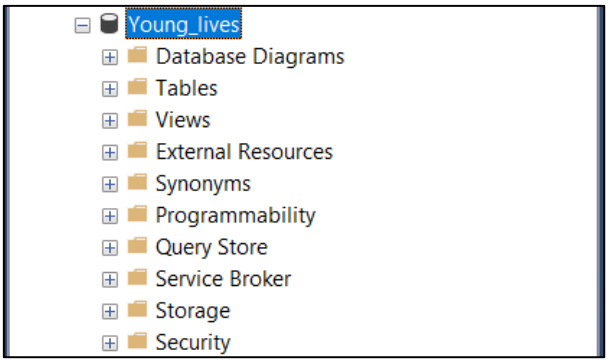
The tables selected for our case study in the report are india_constructed and Ethiopia_constructed.

1.4. Importing the dataset to SQL and designing the database

First, after downloading the dataset, I saved it as a csv file format. Then, in SQL server studio, a database named "Young_lives" was created.

```
CREATE DATABASE Young_lives
GO
```

Data Base Design



Further, we added a new column in order to include the name of each country into its respective row of data. First of all, add a new column into the whole table with 'ALTER TABLE' filled with 'NULL' values:

```
USE [youngLives];
GO

-- Preview the first 10 rows of the Ethiopia dataset
SELECT TOP 10 *
FROM [dbo].[ethiopia_constructed];

-- Preview the first 10 rows of the India dataset
SELECT TOP 10 *
FROM [dbo].[india_constructed];

-- Get the column names and data types of the Ethiopia table
SELECT COLUMN_NAME, DATA_TYPE
FROM INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_NAME = 'ethiopia_constructed';

-- Get the column names and data types of the India table
SELECT COLUMN_NAME, DATA_TYPE
FROM INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_NAME = 'india_constructed';

-- Add Country column to the India table
ALTER TABLE [dbo].[india_constructed]
ADD Country VARCHAR(50);

-- Update the Country column for India
UPDATE [dbo].[india_constructed]
SET Country = 'India'
WHERE Country IS NULL OR Country = '';

-- Add Country column to the Ethiopia table
ALTER TABLE [dbo].[ethiopia_constructed]
ADD Country VARCHAR(50);

-- Update the Country column for Ethiopia
UPDATE [dbo].[ethiopia_constructed]
SET Country = 'Ethiopia'
WHERE Country IS NULL OR Country = '';
```

So, dealing with this dataset a lot of columns were missing or having wrong detailed rows. We didn't do any step of data pre-processing with the Excel sheet; rather we did all the steps for data pre-processing using SQL server management studio. Thus, we decided to remove these inconsistent data columns by filtering them using SQL server.

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

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

Delete Null Values from tables.

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

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

The filtering done the next day yielded 6 new ideas related to the Young Life project. All the ideas were pursued by selecting the respective relevant columns.

1. Impact of Child's Age and Gender on Nutritional Status.
2. Child growth analysis in relation to householder food situation .
3. Child health and age distribution across region
4. Evaluating the performance of Household Food Situation vs Child Growth
5. Interaction between Parental Education and Child Alcohol Use
6. Comparing the child age gender vs nutrition

We first picked the columns based on the idea mentioned above, and then created the SQL views which could filter those columns according to the idea, removing the null values using 'NOT' Boolean Expression in SQL.

SQLQuery1.sql - D...4V8MKOQ\user (56))*

```sql
---Household_FoodSituation_vs_Child_Growth
CREATE VIEW Household_FoodSituation_vs_Child_Growth AS
SELECT
    country,
    foodsec, -- Household food security
    chweight, -- Child's weight
    chheight, -- Child's height
    bmi, -- Body Mass Index
    zwfa, -- Weight-for-age z-score
    zhfa -- Height-for-age z-score
FROM india_constructed
WHERE
    country != 'NA' AND
    foodsec != 'NA' AND
    chweight != 'NA' AND
    chheight != 'NA' AND
    bmi != 'NA' AND
    zwfa != 'NA' AND
    zhfa != 'NA'
UNION ALL
SELECT
    country,
    foodsec, -- Household food security
    chweight, -- Child's weight
    chheight, -- Child's height
    bmi, -- Body Mass Index
    zwfa, -- Weight-for-age z-score
    zhfa -- Height-for-age z-score
FROM ethiopia_constructed
WHERE
    country != 'NA' AND
    foodsec != 'NA' AND
    chweight != 'NA' AND
    chheight != 'NA' AND
    bmi != 'NA' AND
    zwfa != 'NA' AND
    zhfa != 'NA';
```

Like this we have created six views according to the 6 ideas mentioned above.

1.5. Create Views

1. Household_FoodSituation_vs_Child_Growth

```sql
---Household_FoodSituation_vs_Child_Growth
CREATE VIEW Household_FoodSituation_vs_Child_Growth AS
SELECT
    country,
    foodsec, -- Household food security
    chweight, -- Child's weight
    chheight, -- Child's height
    bmi, -- Body Mass Index
    zwfa, -- Weight-for-age z-score
    zhfa -- Height-for-age z-score
FROM india_constructed
WHERE
    country != 'NA' AND
    foodsec != 'NA' AND
    chweight != 'NA' AND
    chheight != 'NA' AND
    bmi != 'NA' AND
    zwfa != 'NA' AND
    zhfa != 'NA'
UNION ALL
SELECT
    country,
    foodsec, -- Household food security
    chweight, -- Child's weight
    chheight, -- Child's height
    bmi, -- Body Mass Index
    zwfa, -- Weight-for-age z-score
    zhfa -- Height-for-age z-score
FROM ethiopia_constructed
WHERE
    country != 'NA' AND
    foodsec != 'NA' AND
    chweight != 'NA' AND
    chheight != 'NA' AND
    bmi != 'NA' AND
    zwfa != 'NA' AND
    zhfa != 'NA';
```

	country	foodsec	chweight	chheight	bmi	zwfa	zhfa	chsex	weight_category
1	India	1	22.29999924	124.4000015	14.41000366	-1.230000019	-0.860000014	1	Underweight
2	India	1	21.5	117.5	15.57265759	-1.399999976	-1.929999948	1	Underweight
3	India	1	18.29999924	117.5	13.25486565	-2.440000057	-1.649999976	1	Underweight
4	India	1	19	119.3000031	13.34973621	-1.710000038	-1.100000024	2	Underweight
5	India	2	26.39999962	126	16.62887383	0.07	-0.449999988	2	Underweight
6	India	1	19.79999924	126.3000031	12.41247654	-2.180000067	-0.569999993	1	Underweight
7	India	2	19.39999962	120.3000031	13.40511227	-1.889999986	-1.309999943	2	Underweight
8	India	1	24.29999924	121.5	16.46090508	-0.469999999	-1.24000001	2	Underweight
9	India	3	18.5	119.5	12.9549551	-2.319999933	-1.549999952	2	Underweight
10	India	1	14	121.4000015	9.499279022	-4.349999905	-1.440000057	2	Underweight
11	India	2	22.29999924	123.0999985	14.71596527	-0.680000007	-0.519999981	2	Underweight
12	India	2	19	116.6999969	13.95121098	-1.889999986	-1.75999999	2	Underweight
13	India	2	19.79999924	115.1999969	14.91970539	-1.440000057	-1.830000043	2	Underweight
14	India	1	21.10000038	122.6999969	14.01500797	-1.350000024	-0.970000029	2	Underweight
15	India	1	26.5	116.5	19.52513313	0.370000005	-1.700000048	2	Normal
16	India	1	33	117	24.10694695	1.539999962	-2	1	Normal

By creating this view, users can analyse how **household food security** correlates with **child growth indicators** such as weight, height, and BMI across two countries (India and Ethiopia). The view allows researchers or analysts to perform comparative studies between the two countries using the same key variables, facilitating cross-national analysis of the impact of food security on child growth.

We create new column using this dataset:

```sql
-- New Column: Weight Category based on BMI
CASE
    WHEN CAST(bmi AS FLOAT) < 18.5 THEN 'Underweight'
    WHEN CAST(bmi AS FLOAT) >= 18.5 AND CAST(bmi AS FLOAT) < 24.9 THEN 'Normal'
    WHEN CAST(bmi AS FLOAT) >= 25 AND CAST(bmi AS FLOAT) < 29.9 THEN 'Overweight'
    ELSE 'Obese'
END AS weight_category  -- New column for child's weight category
```

In this step, we introduce a new column, **`weight_category`**, to categorize each child's BMI into distinct weight categories. This classification is based on the World Health Organization's BMI standards, which help identify potential health risks associated with different weight levels. The formula for this categorization is as follows:

- **Underweight**: BMI < 18.5
- **Normal**: BMI between 18.5 and 24.9
- **Overweight**: BMI between 25 and 29.9
- **Obese**: BMI ≥ 30

2. child_age_gender_vs_nutrition

```sql
----child_age_gender_vs_nutrition
CREATE VIEW child_age_gender_vs_nutrition AS
SELECT
    country,    -- Identify the country
    childid,    -- Child ID
    chsex,      -- Child's gender (1 = Male, 2 = Female)
    agemon,     -- Child's age in months
    chweight,   -- Child's weight (kg)
    chheight,   -- Child's height (cm)
    bmi,        -- Body Mass Index
    zwfa,       -- Weight-for-age z-score
    zhfa        -- Height-for-age z-score
FROM [dbo].[india_constructed]
WHERE
    country != 'NA' AND
    childid != 'NA' AND
    chsex != 'NA' AND
    agemon != 'NA' AND
    chweight != 'NA' AND
    chheight != 'NA' AND
    bmi != 'NA' AND
    zwfa != 'NA' AND
    zhfa != 'NA'

UNION ALL

SELECT
    country,  -- Identify the country
    childid,
    chsex,
    agemon,
    chweight,
    chheight,
    bmi,
```

	country	childid	chsex	agemon	chweight	chheight	bmi	zwfa	zhfa
1	India	IN131022	1	90	16.29999924	114.6999969	12.38969898	-3.099999905	-1.820000052
2	India	IN131023	1	99	17.29999924	116.3000031	12.79047394	-3.160000086	-2.150000095
3	India	IN131024	2	92	18.10000038	120.4000015	12.48606491	-1.980000019	-0.790000021
4	India	IN131025	1	99	20	123.0999985	13.19817543	-2	-1
5	India	IN131026	2	90	16.60000038	106.4000015	14.66306782	-2.49000001	-3.099999905
6	India	IN131027	1	98	19.29999924	114.8000031	14.64446449	-2.200000048	-2.339999914
7	India	IN131028	1	91	16.79999924	113.0999985	13.13360882	-2.940000057	-2.210000038
8	India	IN131029	1	91	17	112.0999985	13.52812767	-2.809999943	-2.349999905
9	India	IN131030	1	89	20	124.1999969	12.9654274	-1.360000014	0
10	India	IN131031	1	98	15.80000019	114.4000015	12.07271767	-3.779999971	-2.420000076
11	India	IN131032	1	90	18.20000076	119.1999969	12.80910873	-2.220000029	-1.039999962
12	India	IN131033	2	97	18	106	16.01993561	-2.369999886	-3.680000067
13	India	IN131034	1	99	20.10000038	115.1999969	15.14576244	-1.940000057	-2.349999905
14	India	IN131035	1	90	20.5	116.0999985	15.20860672	-1.25999999	-1.600000024
15	India	IN131036	2	89	17.20000076	110.5999985	14.06106567	-2.190000057	-2.319999933
16	India	IN131037	1	92	16	109.4000015	13.36858177	-3.359999895	-2.910000086

The **child_age_gender_vs_nutrition** view aggregates and cleans data related to children's demographic and nutritional information from the India_constructed and Ethiopian_constructed tables. It selects key attributes like country, child ID, gender, age (in months), weight, height, BMI , and z-scores for weight for age and height-for-age.

3. Household_Wealth_Index_and_Child_Health

```sql
---Household_Wealth_Index_and_Child_Health
CREATE VIEW Household_Wealth_Index_and_Child_Health AS
SELECT
        childid,
        region,
        commid,
        clustid,
        chillness,
        chinjury,
        country
FROM india_constructed
WHERE
    childid != 'NA' AND
    region != 'NA' AND
    commid != 'NA' AND
    country='NA' AND
    clustid != 'NA' AND
    chillness != 'NA' AND
    chinjury != 'NA'
    UNION ALL

SELECT
        childid,
```

	childid	region	commid	clustid	chillness	chinjury	country	age_group	health_status
1	IN131021	24	IN629	90	1	0	India	18+	Average
2	IN131022	22	IN045	13	NA	NA	India	6-12	Unknown
3	IN131022	22	IN045	13	1	0	India	6-12	Unknown
4	IN131022	22	IN045	13	NA	0	India	13-17	Good/Very Good
5	IN131022	22	IN045	13	0	0	India	18+	Good/Very Good
6	IN131022	22	IN045	13	0	0	India	18+	Good/Very Good
7	IN131023	22	IN045	13	NA	NA	India	6-12	Unknown
8	IN131023	22	IN045	13	0	0	India	6-12	Unknown
9	IN131023	22	IN045	13	NA	1	India	13-17	Good/Very Good
10	IN131023	22	IN045	13	0	1	India	18+	Good/Very Good
11	IN131023	22	IN045	13	0	0	India	18+	Average
12	IN131024	22	IN045	13	NA	NA	India	6-12	Unknown
13	IN131024	NA	IN252	90	0	0	India	6-12	Unknown

The **Household_Wealth_Index_and_Child_Health** view consolidates data related to household characteristics and child health outcomes from the India and Ethiopian constructed` table. It includes fields such as child ID, region, community ID, cluster ID, child illness, child injury, and country.

4. Maternal_Health_and_Child_Nutrition_Summary

```sql
----Maternal_Health_and_Child_Nutrition_Summary
CREATE VIEW Maternal_Health_and_Child_Nutrition_Summary AS
SELECT
    region,                                 -- Region
    caredu,                                 -- Caregiver's education level
    COUNT(childid) AS total_children,       -- Total number of children
    AVG(CAST(chweight AS FLOAT)) AS avg_weight,   -- Average child's weight (kg)
    AVG(CAST(chheight AS FLOAT)) AS avg_height,   -- Average child's height (cm)
    AVG(CAST(bmi AS FLOAT)) AS avg_bmi,           -- Average Body Mass Index (BMI)
    AVG(CAST(zwfa AS FLOAT)) AS avg_zwfa,         -- Average weight-for-age z-score
    AVG(CAST(zhfa AS FLOAT)) AS avg_zhfa          -- Average height-for-age z-score
FROM [dbo].[india_constructed]
WHERE
    childid != 'NA' AND
    region != 'NA' AND
    commid != 'NA' AND
    caredu != 'NA' AND
    hhsize != 'NA' AND
    chweight != 'NA' AND
    chheight != 'NA' AND
    bmi != 'NA' AND
    zwfa != 'NA' AND
    zhfa != 'NA'
GROUP BY region, caredu

UNION ALL

SELECT
    region,                                 -- Region
```

	region	caredu	total_children	avg_weight	avg_height	avg_bmi	avg_zwfa	avg_zhfa
1	23	11	6	14.8000001108333	105.299999241667	13.0865169116667	-2.12000000783333	-1.14166667416667
2	77	5	1	19.89999962	122.5	13.26114082	-1.799999952	-0.810000002
3	22	5	140	15.9699999674143	105.168571369	14.1321558158571	-1.71271429331429	-1.30721428381428
4	21	5	233	15.5108154689356	103.231545217511	14.223147374176	-1.73416309116309	-1.41038626848069
5	21	12	60	16.2874999285167	103.157499692333	14.8283802435	-1.03300000046667	-0.780666670266666
6	23	2	35	17.1100000922857	110.788571599143	13.8720301197143	-2.05514285665714	-1.74942856237143
7	23	14	37	16.5918920228919	104.160811144595	14.7697138389189	-1.20189190027027	-0.919999998459459
8	22	7	98	15.5173469029286	103.848979679184	14.1424163915306	-1.73357143037755	-1.2979591867449
9	22	4	54	15.0481480848333	103.674073962778	13.7234469388889	-2.10314813392593	-1.57296296709259
10	21	15	12	23.23333327	125.099998466667	14.7767457158333	-0.609999999333...	-0.264999995583333
11	77	15	1	34.09999847	133.1000061	19.248558404	1.580000043	0.610000014
12	21	28	2	21.125	115.79999925	15.621852395	-0.5899999885	-1.0449999865
13	22	3	58	16.5413792710345	107.951724481552	14.0341947337931	-1.78517241294828	-1.3272413762931

The view `**Maternal_Health_and_Child_Nutrition_Summary**` provides a summary of maternal health and child nutrition data based on regions and caregiver education levels. It includes key metrics such as the total number of children, average child weight, height, Body Mass Index (BMI), and z-scores for weight-for-age and height-for-age.

5. parental_education_vs_child_alcohol_summary

```sql
----Parental Education vs. Child's Alcohol Consumption
CREATE view  parental_education_vs_child_alcohol AS
SELECT
    country,  -- Identify the country
    childid,             -- Child ID
    caredu,              -- Caregiver's education level
    chalcohol            -- Child's alcohol consumption (1 = Yes, 0 = No)
FROM india_constructed
WHERE
    country != 'NA' AND
    childid != 'NA' AND
    caredu != 'NA' AND
    chalcohol != 'NA'
UNION ALL
SELECT
    country,  -- Identify the country
    childid,
    caredu,
    chalcohol
FROM ethiopia_constructed
WHERE
    country != 'NA' AND
    childid != 'NA' AND
    caredu != 'NA' AND
    chalcohol != 'NA';
```

	country	caredu	total_children	alcohol_consumers	non_consumers
1	India	15	26	0	26
2	India	28	62	1	61
3	India	12	43	1	42
4	India	3	81	1	80
5	India	4	68	0	68
6	India	10	235	2	233
7	India	5	244	2	242
8	India	7	150	0	150
9	India	13	6	1	5
10	India	1	20	0	20
11	India	0	1046	11	1035
12	India	6	61	0	61
13	India	8	38	1	37
14	India	9	97	0	97
15	India	14	56	0	56
16	India	29	2	0	2
17	India	2	87	0	87
18	India	11	4	0	4
19	Ethiopia	4	131	5	126

The **parental_education_vs_child_alcohol** view is designed to facilitate the analysis of the relationship between **caregiver's educational level** and **child's alcohol consumption**. By combining data from both India and Ethiopia, it allows for comparisons across different cultural and socio-economic contexts.

6. Parental_Education_vs_Child_Educational_Attainment_Summary

```sql
----Parental_Education_vs_Child_Educational_Attainment_Summary
CREATE VIEW Parental_Education_vs_Child_Educational_Attainment_Summary AS
SELECT
    country,
    caredu,
    AVG(CAST(hschool AS FLOAT)) AS avg_hschool,    -- Average hours child spent at school
    SUM(CAST(hschool AS FLOAT)) AS total_hschool,  -- Total hours child spent at school
    COUNT(hschool) AS count_hschool                -- Number of records
FROM india_constructed
WHERE
    country != 'NA' AND
    caredu != 'NA' AND
    hschool != 'NA'
GROUP BY country, caredu

UNION ALL

SELECT
    country,
    caredu,
    AVG(CAST(hschool AS FLOAT)) AS avg_hschool,    -- Average hours child spent at school
    SUM(CAST(hschool AS FLOAT)) AS total_hschool,  -- Total hours child spent at school
    COUNT(hschool) AS count_hschool                -- Number of records
FROM ethiopia_constructed
WHERE
```

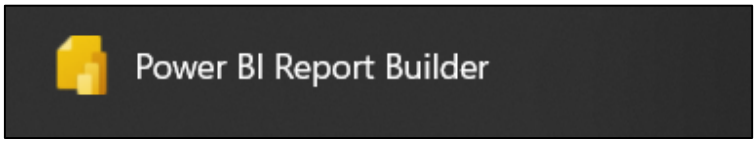
	country	caredu	avg_hschool	total_hschool	count_hschool
1	India	15	8.76811594202899	605	69
2	India	28	6.84848484848485	1582	231
3	India	12	7.99350649350649	1231	154
4	India	3	7.46223564954683	2470	331
5	India	4	7.23606557377049	2207	305
6	India	10	8.02884615384615	6680	832
7	India	5	7.50959232613909	6263	834
8	India	7	7.72106261859583	4069	527
9	India	13	8.23076923076923	214	26
10	India	1	7.17647058823529	610	85
11	India	0	6.59082813891362	29606	4492
12	India	6	7.29251700680272	2144	294
13	India	8	7.64285714285714	1177	154

This view is designed to enable analysis of the relationship between parental education and child's educational outcomes. The caregiver's education level is a crucial factor in determining a child's access to education, learning environment, and academic progress. By examining this relationship across two countries—India and Ethiopia—researchers can gain insights into whether and how parental education affects a child's time spent in school.

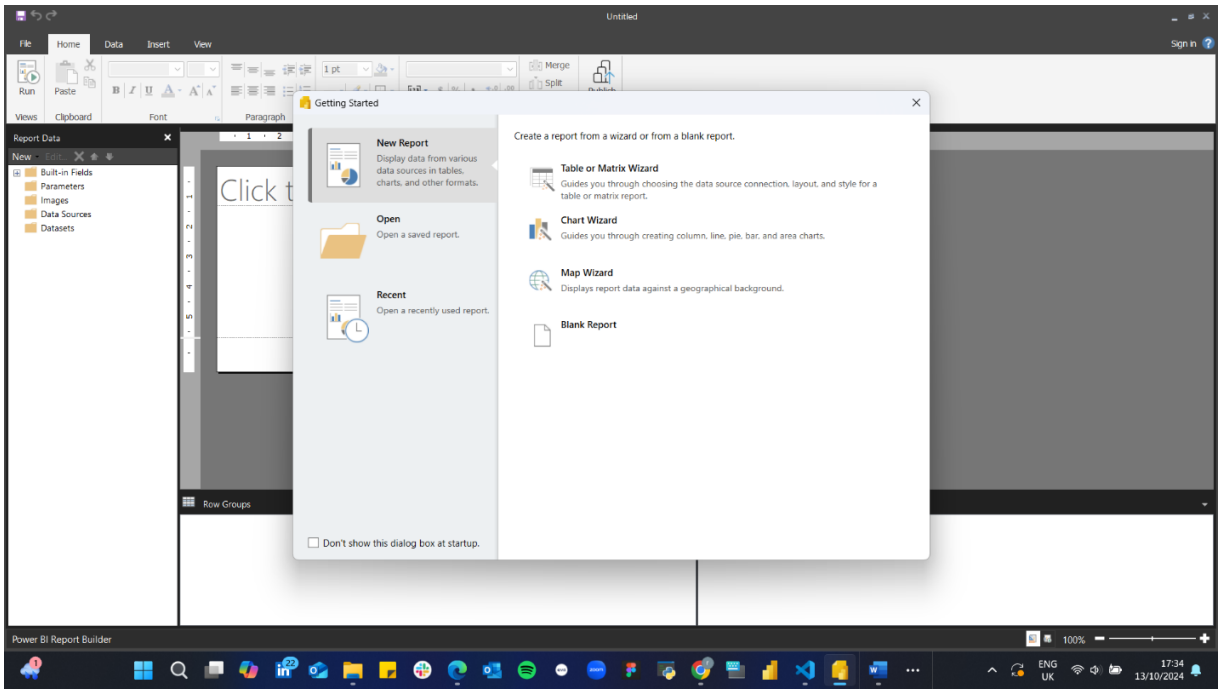
1.6. Creating reports using Microsoft Report Builder

Steps to Create the report using Microsoft report Builder.

01. After setting up the views in the SQL server, we now connect the report builder. Open Power BI report builder on the computer.

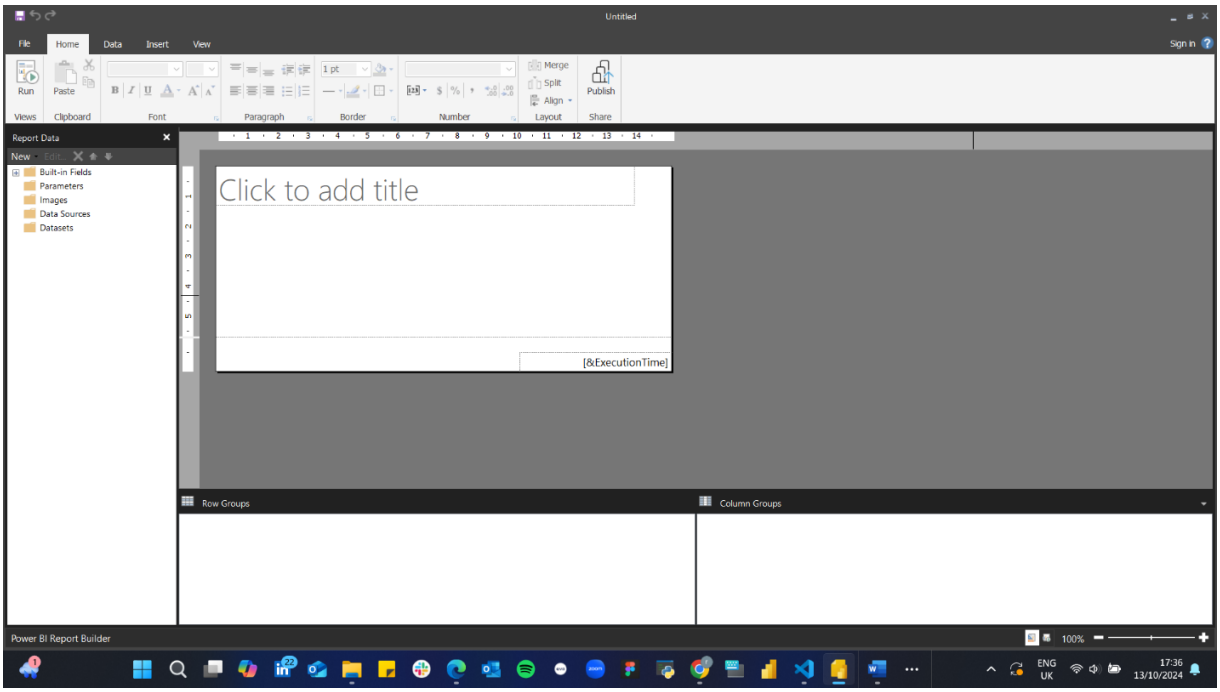


A New Report or Dataset dialog box pops up.



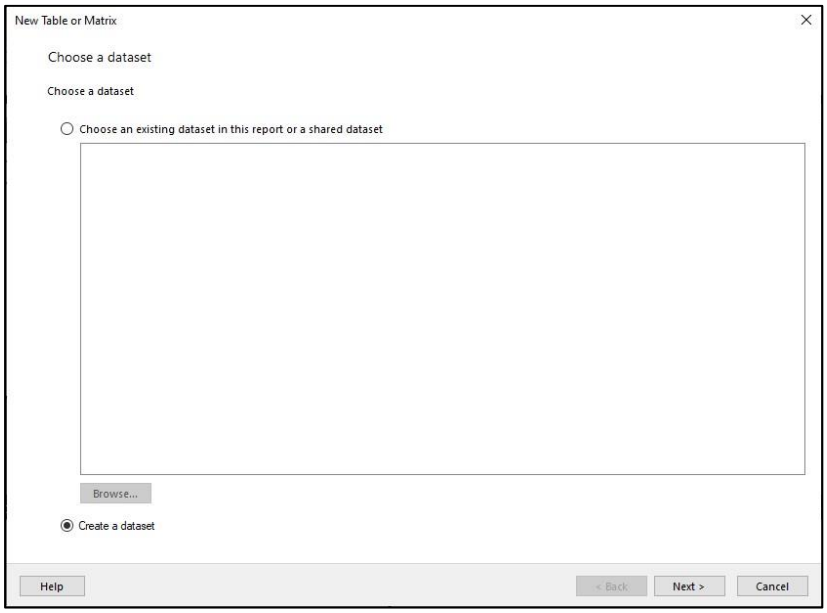
02. On the left side, click on New Report.

03. On the right-hand side, click on Table or Matrix Wizard. You will be presented with the following interface.

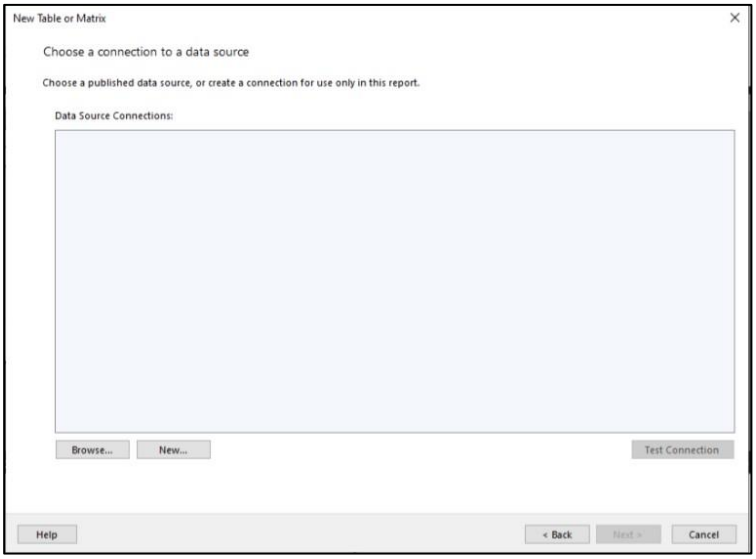


CREATE AN EMBEDDED DATA SOURCE

01. On the Choose a dataset page, select the option 'Create a dataset' and then click Next. The Choose a connection to a data source page opens.

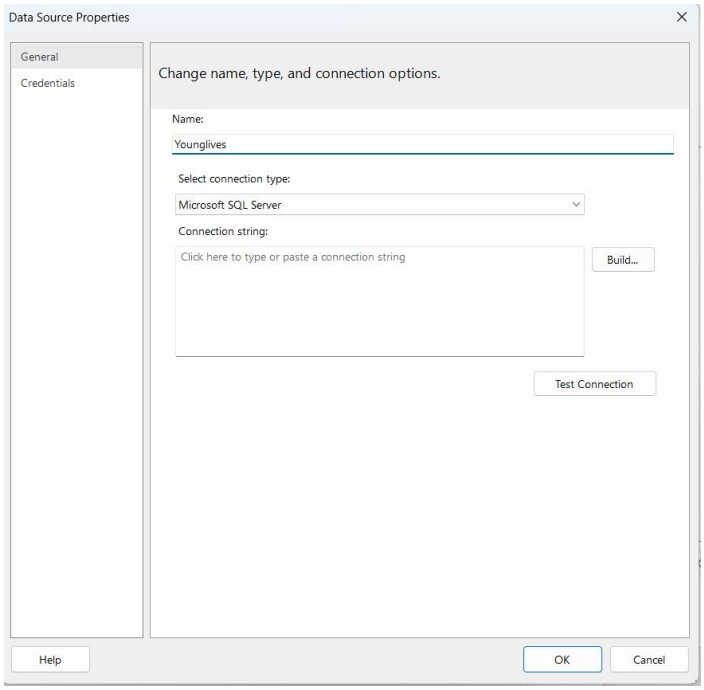


02. Click New. The Data Source Properties dialog box opens.



03. In the Name box, type "Younglives" as the new name of the data source.

04. In Select a connection type, make sure Microsoft SQL Server is selected.



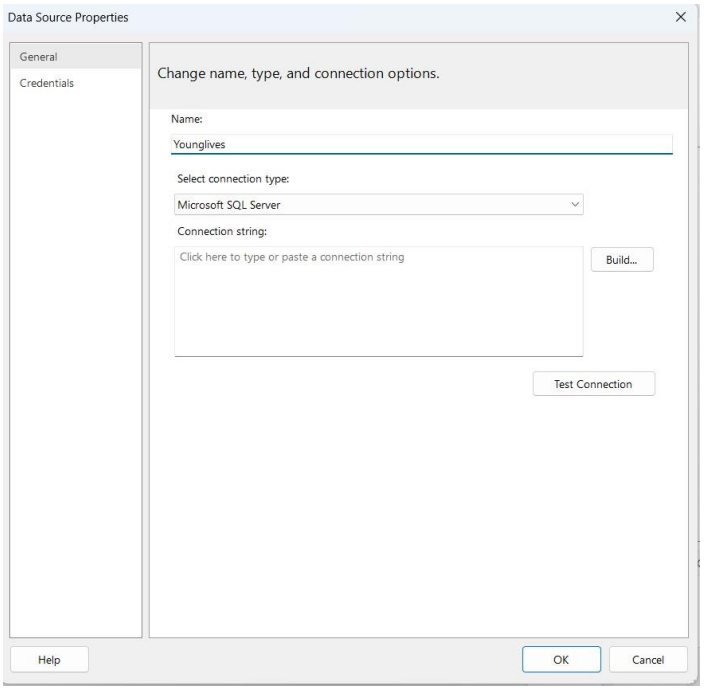
05. Click on Build in 'Connection string'.

06 Server Name box: Type the name of an instance of SQL Server or select it from the drop-down list.

07. Choose the needed database from dropdown, in this case, please choose "Younglifes" database.

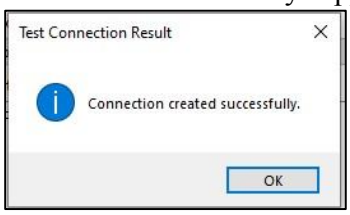
08. Click on Test Connection to test the Connection to the data source.

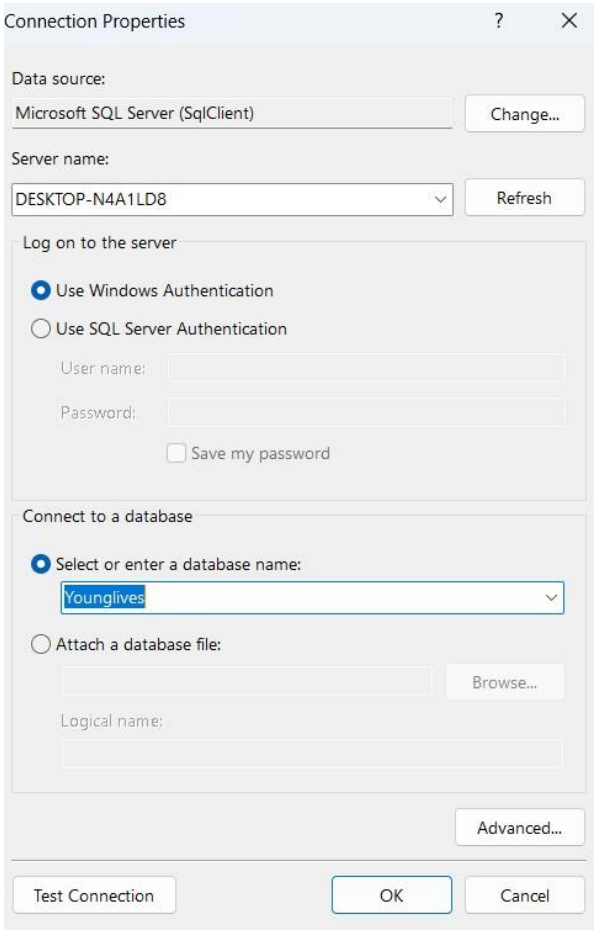
Thrown response "Connection successfully created".



09. Click Test Connection to test you can connect to the data source.

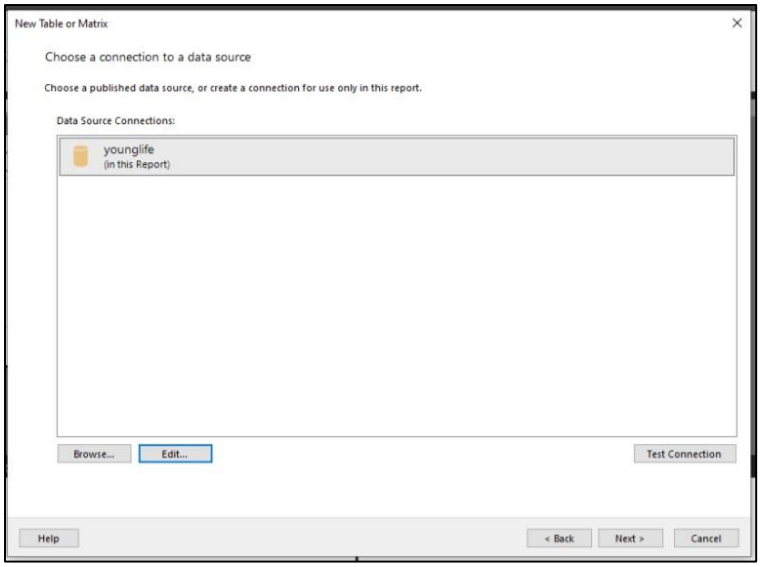
The message "Created connection successfully" appears.





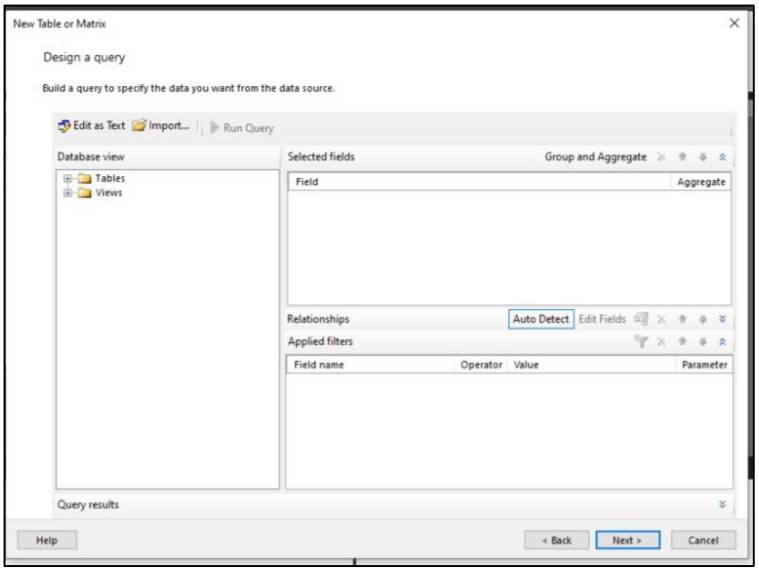
10. Click OK.

You will be taken back to the Choose a connection to a data source page. The new data source is now selected. Click the Next



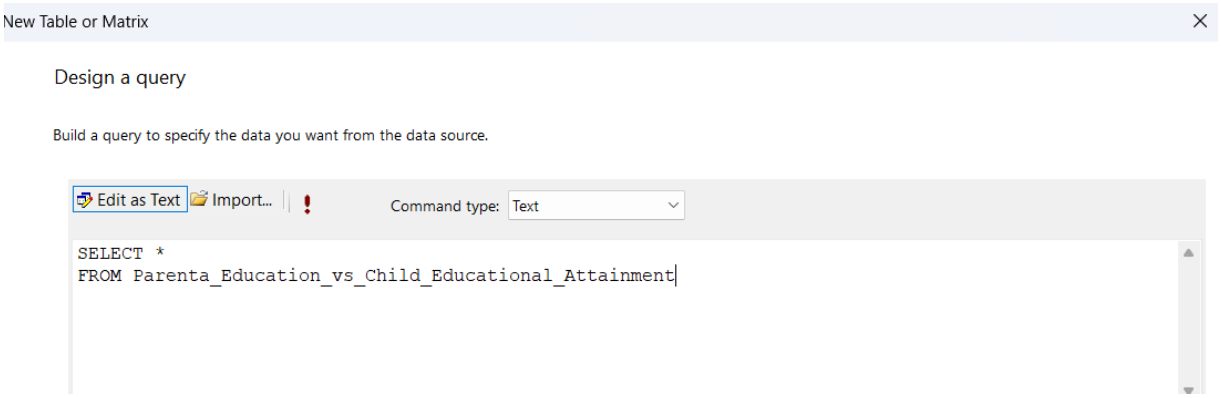
TO CREATE A QUERY

01. The relational query designer opens to the Design a query page. In this tutorial you use the text-based query designer.

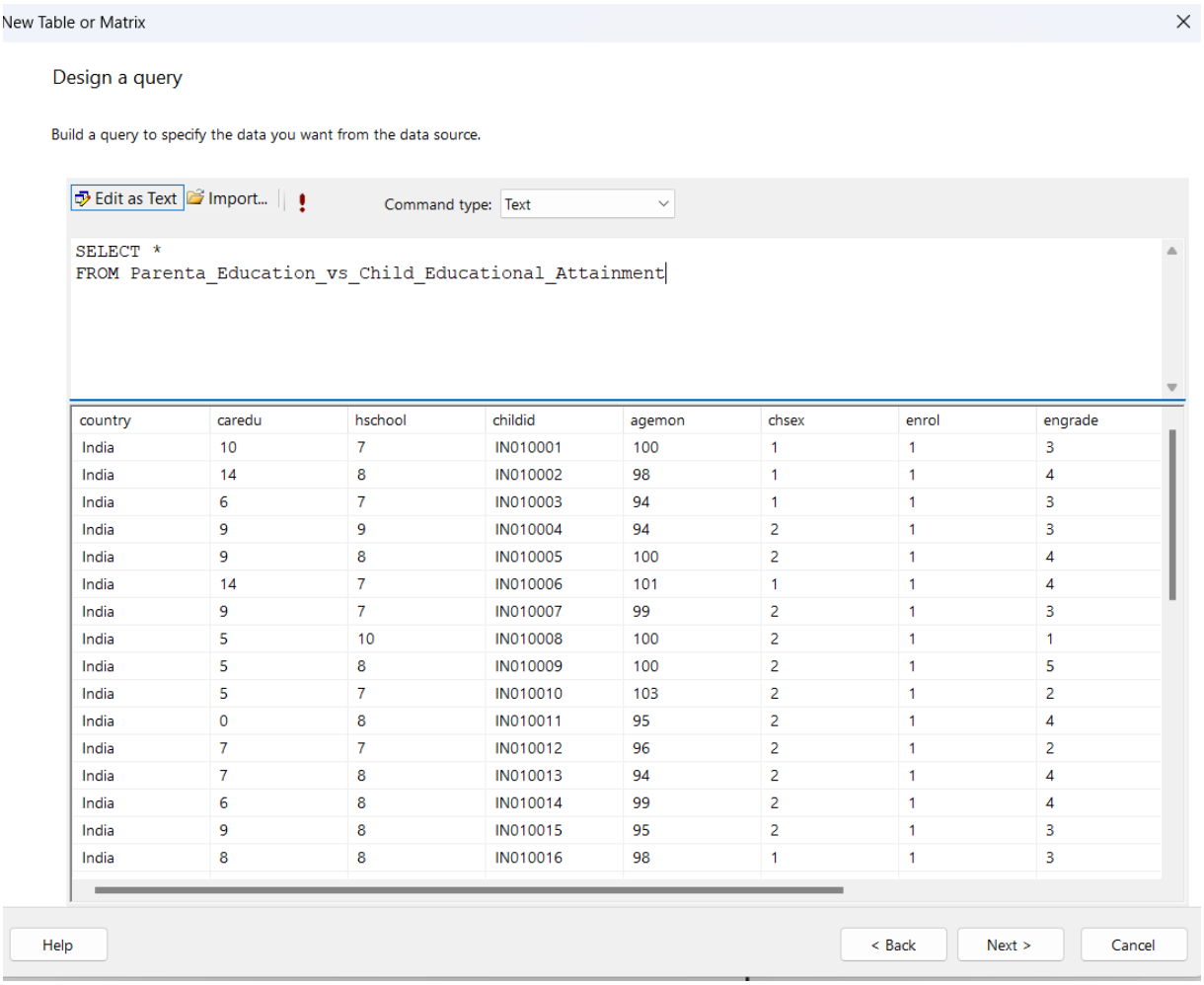


02. Click Edit as Text The text-based query designer displays a query pane and a results pane

03. Put the following Transact-SQL query in the empty top box.



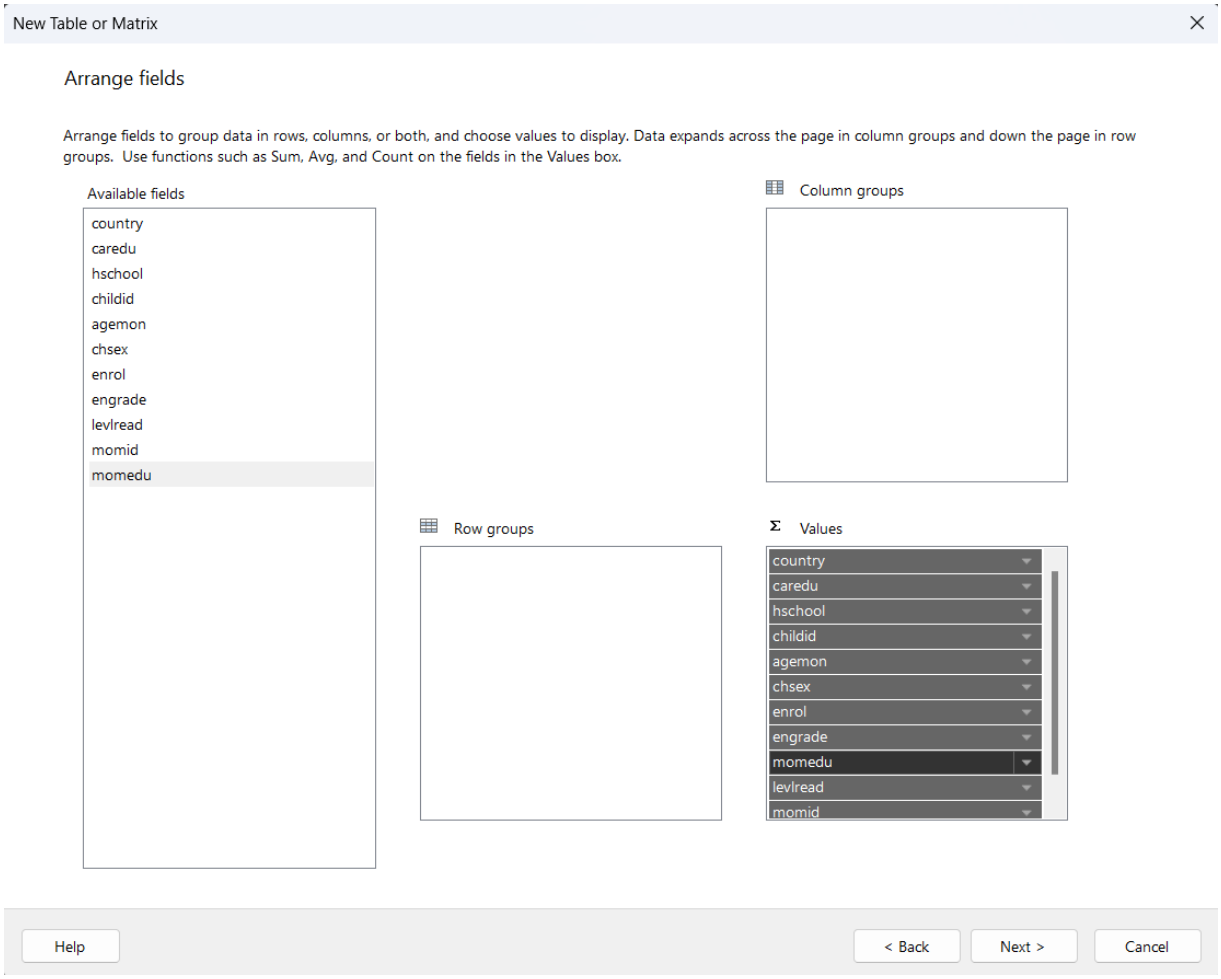
04. Click Run (!) on the query designer toolbar. After the query has completed, the result set for the fields is displayed.



country	caredu	hschool	childid	agemon	chsex	enrol	engrade
India	10	7	IN010001	100	1	1	3
India	14	8	IN010002	98	1	1	4
India	6	7	IN010003	94	1	1	3
India	9	9	IN010004	94	2	1	3
India	9	8	IN010005	100	2	1	4
India	14	7	IN010006	101	1	1	4
India	9	7	IN010007	99	2	1	3
India	5	10	IN010008	100	2	1	1
India	5	8	IN010009	100	2	1	5
India	5	7	IN010010	103	2	1	2
India	0	8	IN010011	95	2	1	4
India	7	7	IN010012	96	2	1	2
India	7	8	IN010013	94	2	1	4
India	6	8	IN010014	99	2	1	4
India	9	8	IN010015	95	2	1	3
India	8	8	IN010016	98	1	1	3

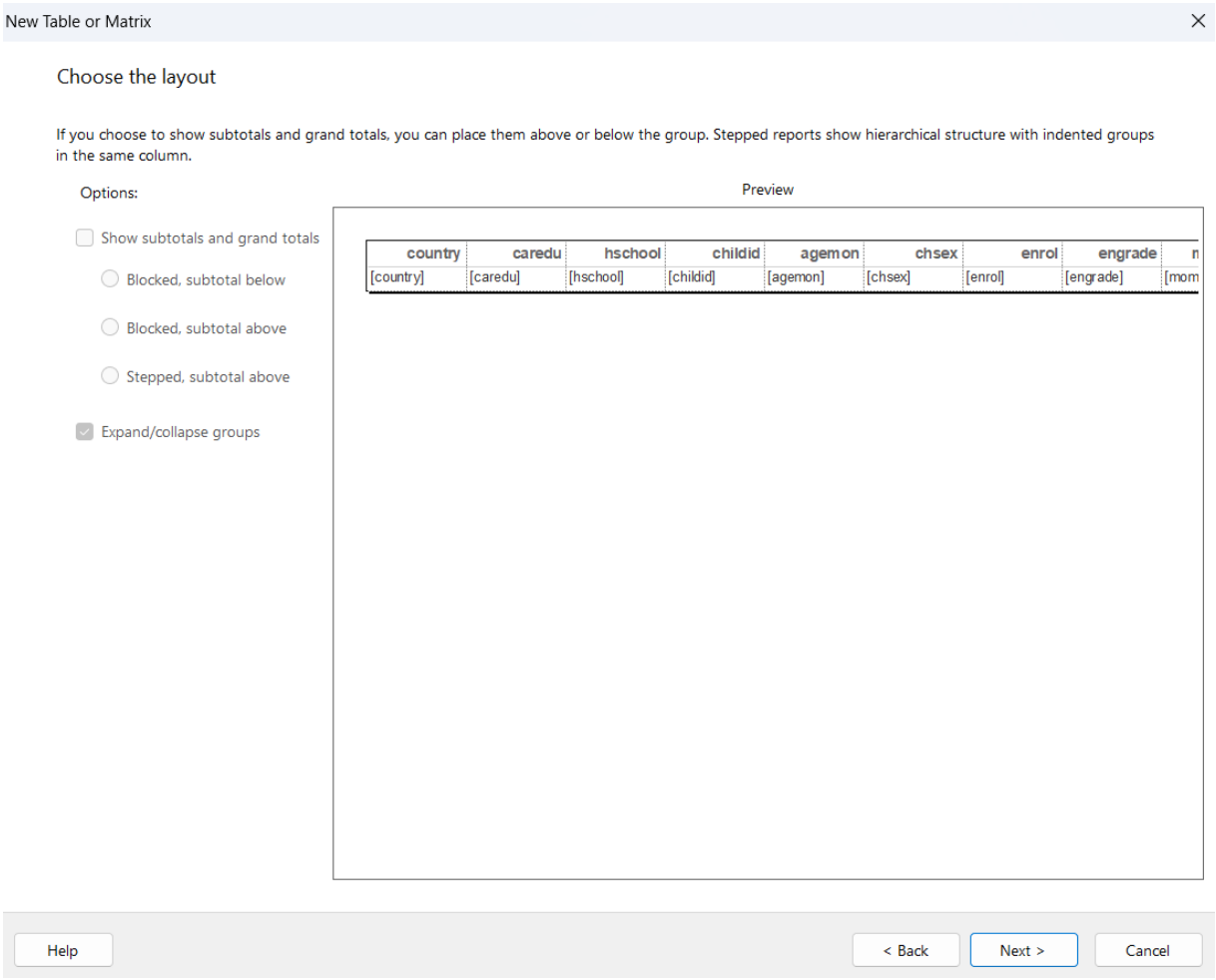
05.Select Next.
Next, the data needs to be grouped.

06. Move every column into the Values Section.



07.Select Next.

08. To display the table, select Next, and then click Finish.



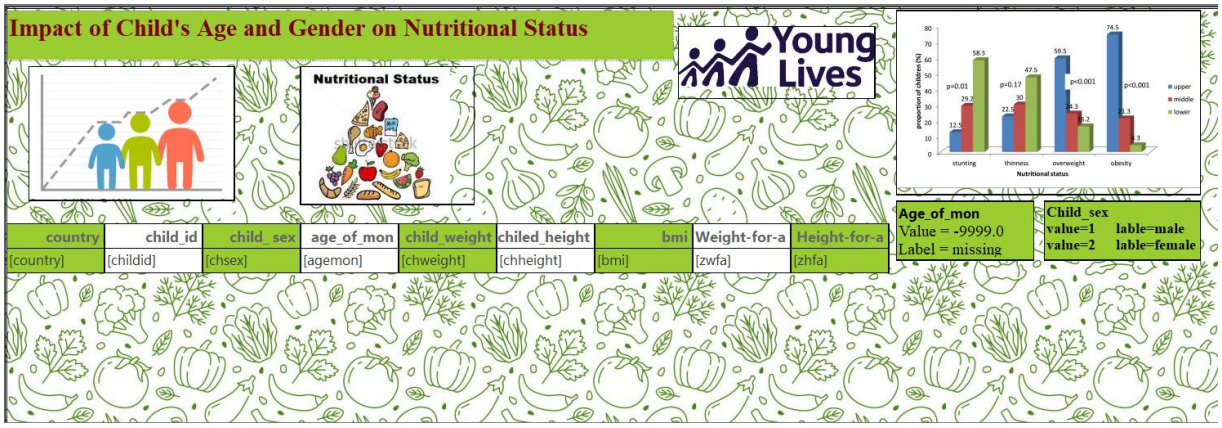
09. Now, Click Run to preview your report.

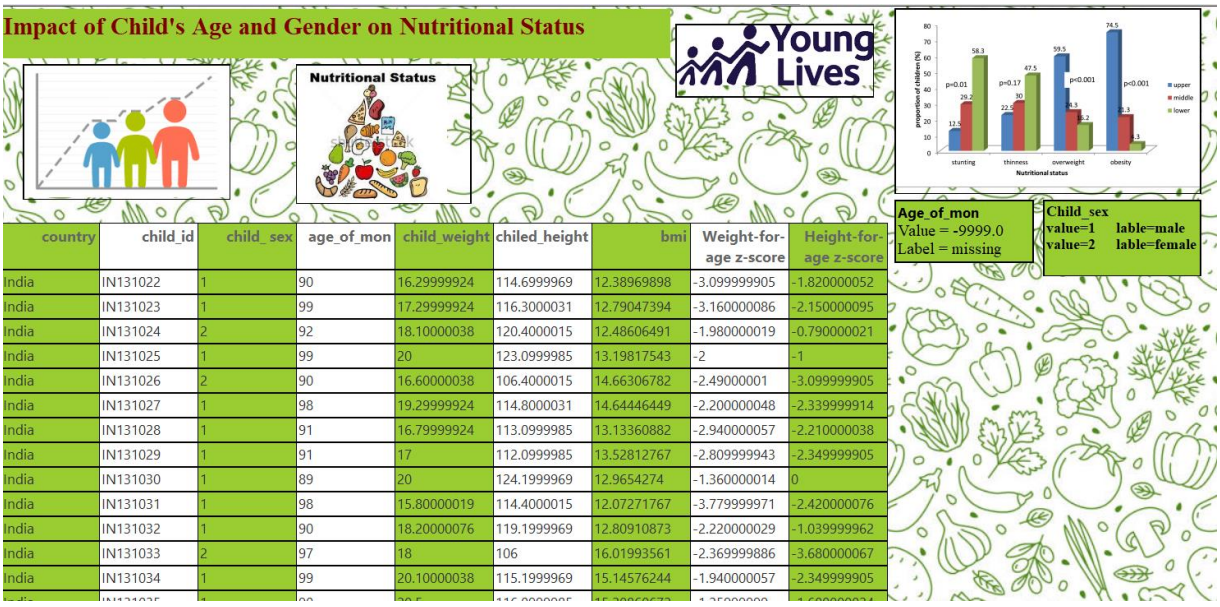
1.5. Creating reports using Microsoft Report Builder

We have created six reports under different topics, each addressing specific areas of analysis. These reports are designed to provide insightful data visualizations and detailed information based on our defined parameters. The topics for these reports are as follows:

➢ Report1 : Impact of Child's Age and Gender on Nutritional Status

The report shows about the impact of a child's age and gender on nutritional status and presents a nuanced examination of how age and gender influence a child's nutritional health, providing both statistical and graphical insights.





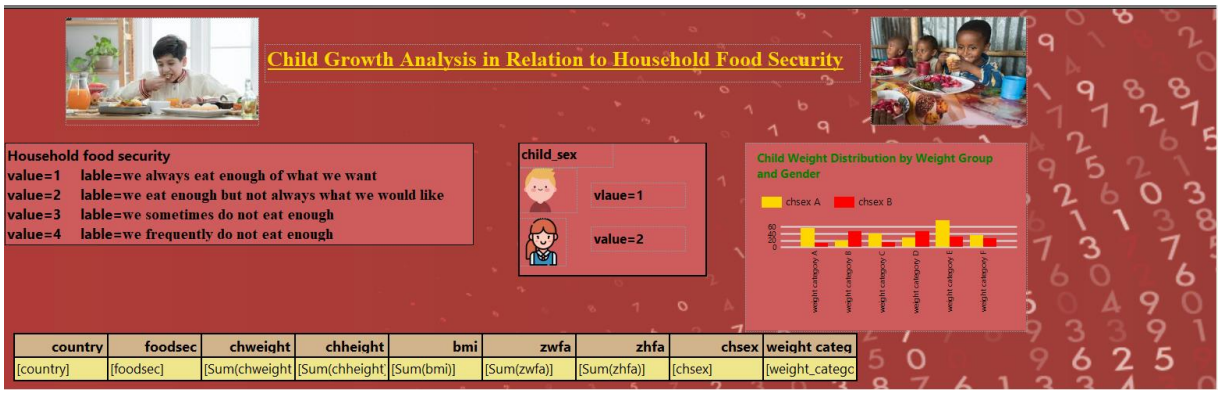
country	child_id	child_sex	age_of_mon	child_weight	chiled_height	bmi	Weight-for-age z-score	Height-for-age z-score
India	IN131022	1	90	16.29999924	114.6999969	12.38969898	-3.099999905	-1.820000052
India	IN131023	1	99	17.29999924	116.3000031	12.79047394	-3.160000086	-2.150000095
India	IN131024	2	92	18.10000038	120.4000015	12.48606491	-1.980000019	-0.790000021
India	IN131025	1	99	20	123.0999985	13.19817543	-2	-1
India	IN131026	2	90	16.60000038	106.4000015	14.66306782	-2.49000001	-3.099999905
India	IN131027	1	98	19.29999924	114.8000031	14.64446449	-2.200000048	-2.339999914
India	IN131028	1	91	16.79999924	113.0999985	13.13360882	-2.940000057	-2.210000038
India	IN131029	1	91	17	112.0999985	13.52812767	-2.809999943	-2.349999905
India	IN131030	1	89	20	124.1999969	12.9654274	-1.360000014	0
India	IN131031	1	98	15.80000019	114.4000015	12.07271767	-3.779999971	-2.420000076
India	IN131032	1	90	18.20000076	119.1999969	12.80910873	-2.220000029	-1.039999962
India	IN131033	2	97	18	106	16.01993561	-2.369999886	-3.680000067
India	IN131034	1	99	20.10000038	115.1999969	15.14576244	-1.940000057	-2.349999905

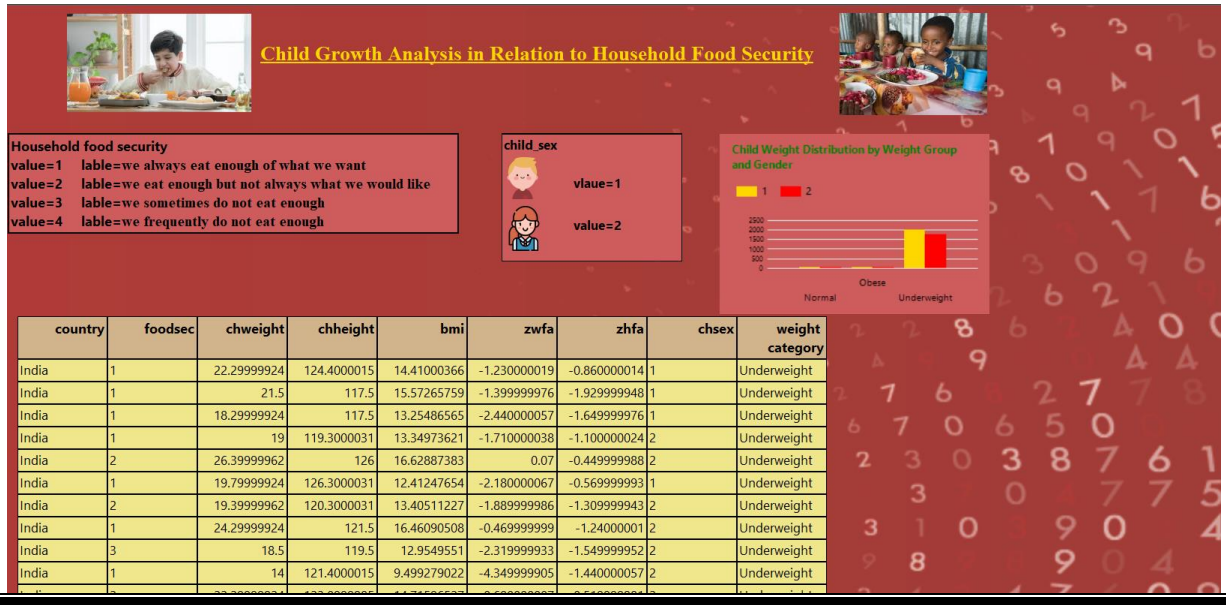
The report focuses on understanding how gender (male vs. female) and age (in months) correlate with various nutritional outcomes such as weight-for-age and height-for-age. The BMI and z-scores serve as indicators of whether children are underweight, normal weight, or overweight, with the corresponding socio-economic context illustrated in the chart. The data emphasizes the importance of addressing malnutrition both in terms of **underweight and stunting** in poorer populations and **overweight and obesity** in wealthier children.

This detailed report is ideal for assessing child nutrition and identifying trends related to age, gender, and socio-economic conditions.

➢ Report 2- Child Growth Analysis In Relation To Householder Food Situation

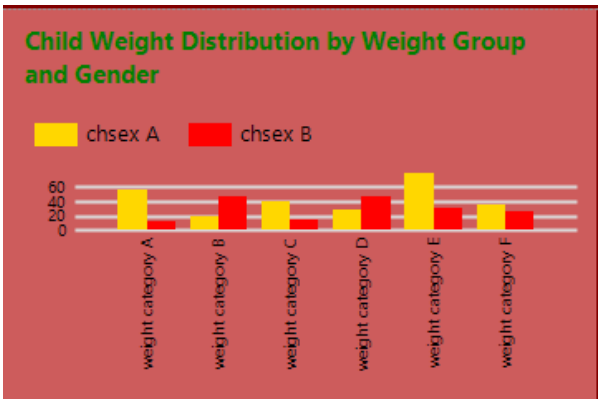
This report examines the relationship between household food security and child growth by analyzing key nutritional indicators. The data is displayed for children from India and includes anthropometric measures, food security status, gender, and weight classification.

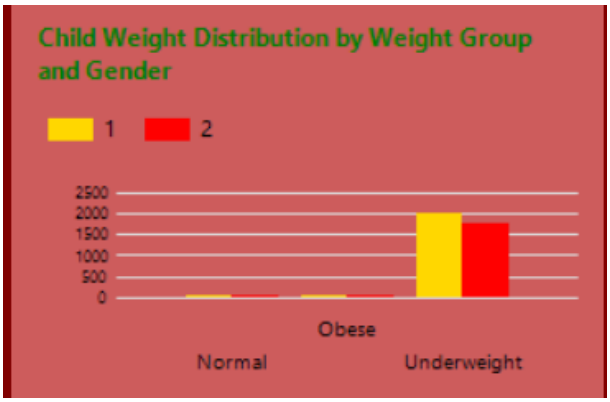




country	foodsec	chweight	chheight	bmi	zwfa	zhfa	chsex	weight category
India	1	22.29999924	124.4000015	14.41000366	-1.230000019	-0.860000014	1	Underweight
India	1	21.5	117.5	15.57265759	-1.399999976	-1.929999948	1	Underweight
India	1	18.29999924	117.5	13.25486565	-2.440000057	-1.649999976	1	Underweight
India	1	19	119.3000031	13.34973621	-1.710000038	-1.100000024	2	Underweight
India	2	26.39999962	126	16.62887383	0.07	-0.449999988	2	Underweight
India	1	19.79999924	126.3000031	12.41247654	-2.180000067	-0.569999993	1	Underweight
India	2	19.39999962	120.3000031	13.40511227	-1.889999986	-1.309999943	2	Underweight
India	1	24.29999924	121.5	16.46090508	-0.469999999	-1.24000001	2	Underweight
India	3	18.5	119.5	12.9549551	-2.319999933	-1.549999952	2	Underweight
India	1	14	121.4000015	9.499279022	-4.349999905	-1.440000057	2	Underweight

The report highlights a strong correlation between household food security and child growth, with a focus on children in India. The data shows that even in households where children have enough food to eat (food security level 2), the lack of dietary diversity leads to widespread underweight conditions, as indicated by low BMI and negative weight-for-age z-scores. Gender disparities are also suggested, with a higher prevalence of underweight girls compared to boys. Stunting, a sign of chronic malnutrition, is evident in several children, pointing to the need for improved food quality and targeted nutritional interventions to address both undernutrition and its long-term effects on child development.

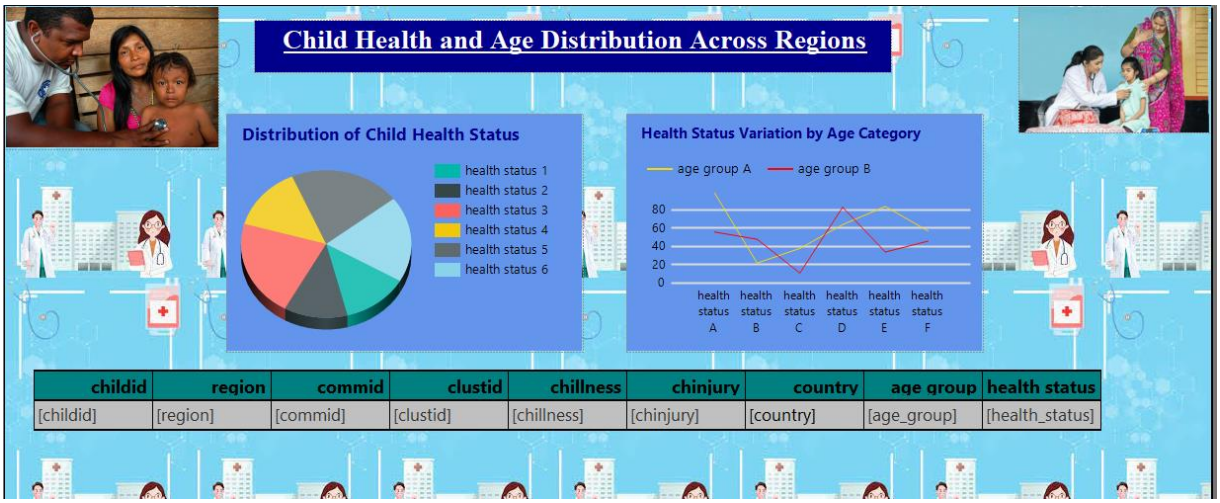




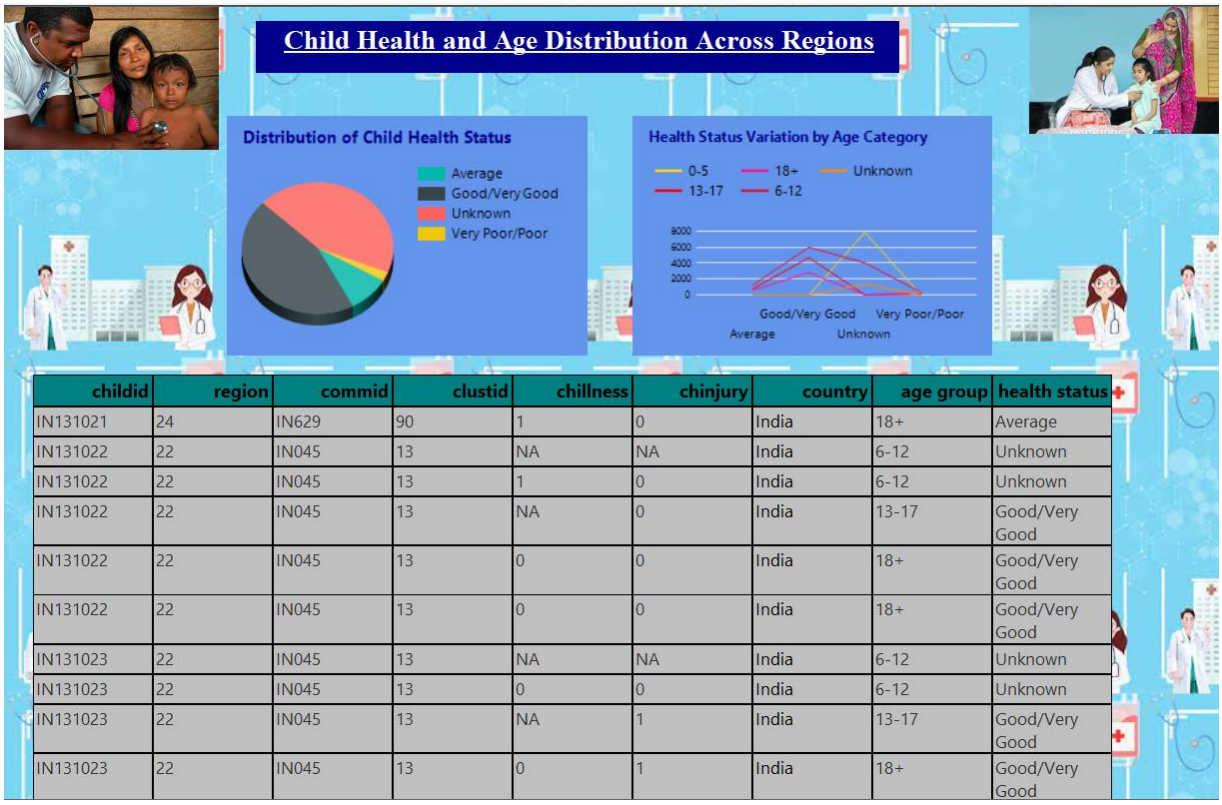
The bar chart provides a clear visual representation of the distribution of children by weight group and gender. Using a color-coded system, red bars denote males and yellow bars signify females, both of which predominantly occupy the underweight category, highlighting a significant nutritional concern affecting both genders. In contrast, the normal and obese categories show very few children, underscoring that the majority of the population is grappling with underweight issues rather than achieving healthy weight status. This stark visual reinforces the urgency of addressing malnutrition among children in the dataset.

➢ Report3-Child Health And Age Distribution Across Region

This report explores the connection between child health status and regional variations by analyzing critical health indicators. The data focuses on children from India and includes key factors such as age group, health status classifications, injury and symptom reports, and regional distribution. Through this analysis, the report aims to uncover trends and disparities in child well-being across different regions, offering insights into the health conditions affecting children and potential areas for targeted interventions.

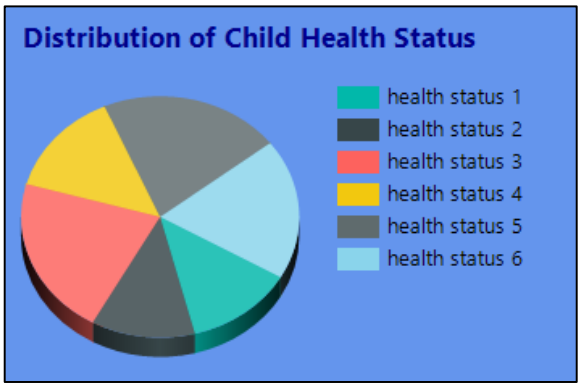


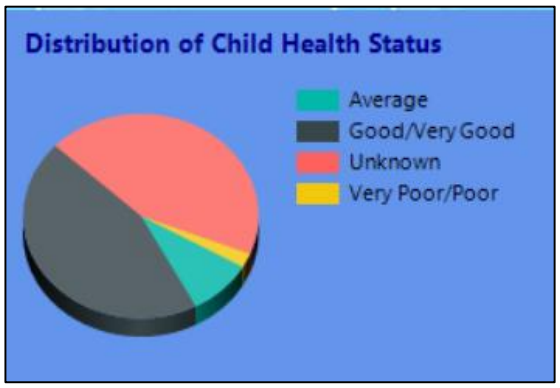
childid	region	commid	clustid	chillness	chinjury	country	age group	health status
[childid]	[region]	[commid]	[clustid]	[chillness]	[chinjury]	[country]	[age_group]	[health_status]



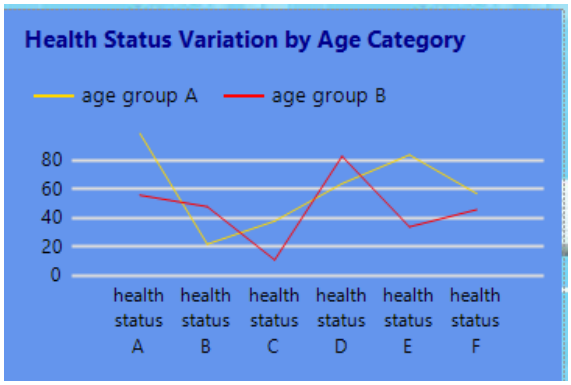
childid	region	commid	clustid	chillness	chinjury	country	age group	health status
IN131021	24	IN629	90	1	0	India	18+	Average
IN131022	22	IN045	13	NA	NA	India	6-12	Unknown
IN131022	22	IN045	13	1	0	India	6-12	Unknown
IN131022	22	IN045	13	NA	0	India	13-17	Good/Very Good
IN131022	22	IN045	13	0	0	India	18+	Good/Very Good
IN131022	22	IN045	13	0	0	India	18+	Good/Very Good
IN131023	22	IN045	13	NA	NA	India	6-12	Unknown
IN131023	22	IN045	13	0	0	India	6-12	Unknown
IN131023	22	IN045	13	NA	1	India	13-17	Good/Very Good
IN131023	22	IN045	13	0	1	India	18+	Good/Very Good

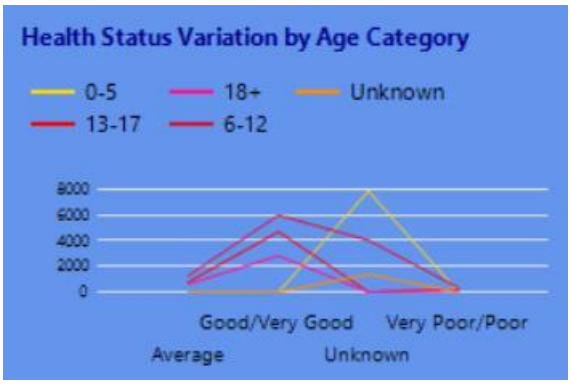
This report examines the regional disparities in child health across India and Ethiopia, focusing on key health indicators such as age, injury prevalence, and overall health status. The findings highlight the need for improved healthcare outreach, better data accuracy, and targeted interventions to address these disparities.





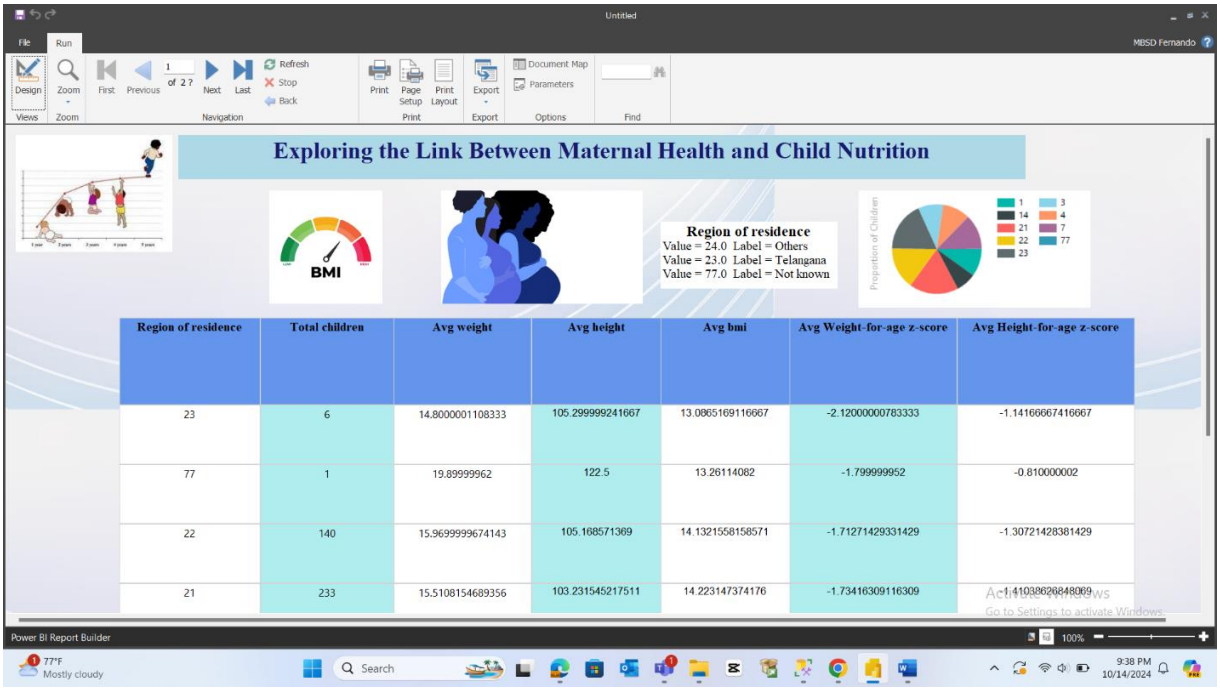
The pie chart reveals that the majority of children are classified as having good/very good health, suggesting that most children in these regions experience favorable health conditions. A smaller group of children is categorized as Very Poor/Poor, signaling possible health crises in specific areas that require urgent intervention. Additionally, the presence of an average health category suggests that while many children are in good health, moderate health issues persist and could benefit from improved healthcare measures.



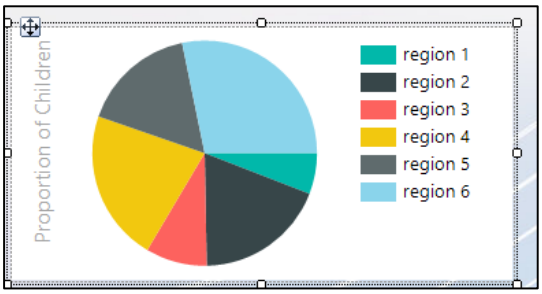


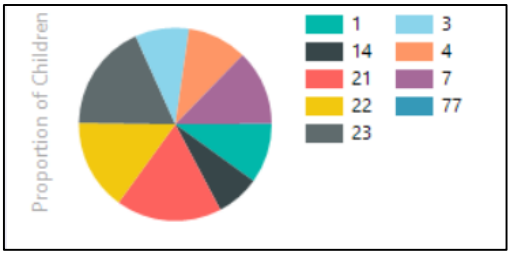
Children aged 6-12 years show the most variability in health, with some falling into the Very Poor/Poor category and others classified as Good/Very Good, indicating that this group may need more targeted health interventions. The 0-5 years and 13-17 years age groups generally have more children in the Good/Very Good category, but there are still significant cases of poor health in these groups. In the 18+ age group, most children are categorized as average or good/very good, with relatively few instances of poor health.

Report4: Exploring the link between menternal health and child nutrition



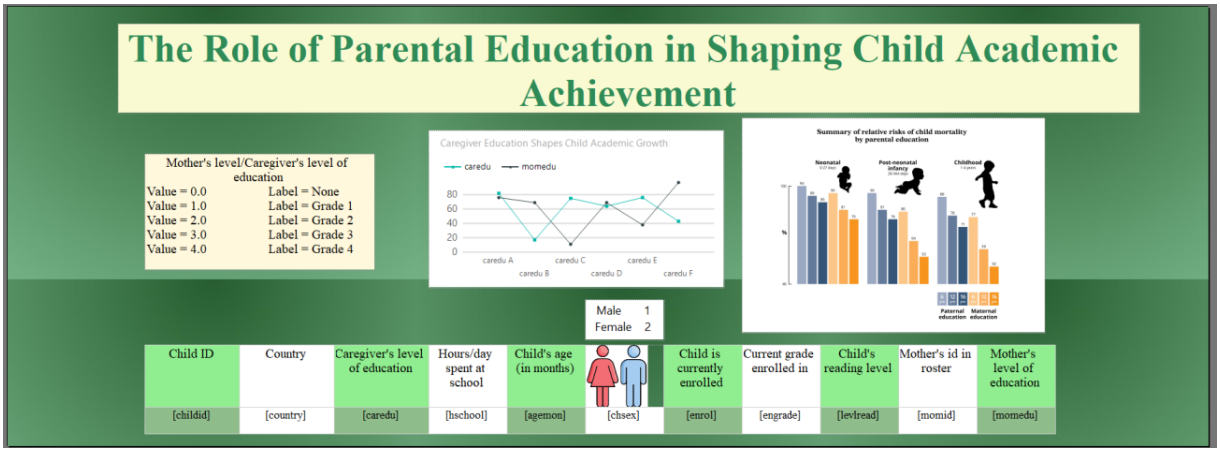
This report explores the link between maternal health and child nutrition, focusing on regional disparities. It presents data on children from various regions, including their average weight, height, body mass index (BMI), and z-scores for weight-for-age and height-for-age. Visual elements like a BMI gauge, a pie chart showing the proportion of children across regions, and illustrations of child growth emphasize key health indicators. The report highlights regional differences in nutritional metrics, which are critical for understanding child development in different areas.

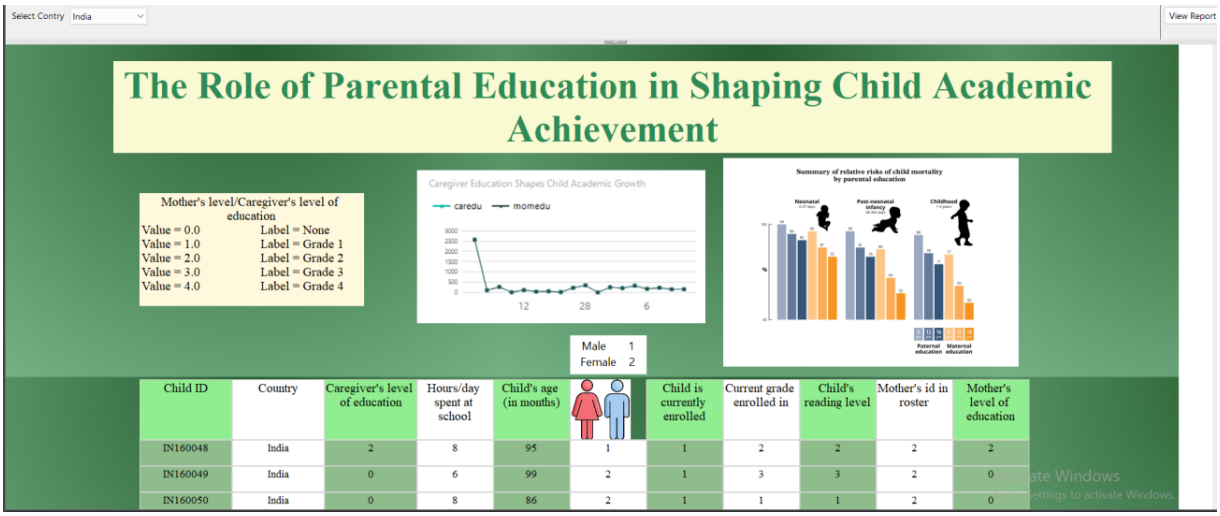




The pie chart in the report illustrates the proportion of children across different regions of residence. Each segment of the chart is color-coded to represent a specific region, labeled with numbers such as 1, 3, 14, 21, 22, 23, and 77. The size of each slice corresponds to the relative number of children from that region. The chart helps visualize the distribution of children across various regions, showing which areas have larger or smaller populations in the study. This visual is useful for comparing regional representation in terms of child nutrition data.

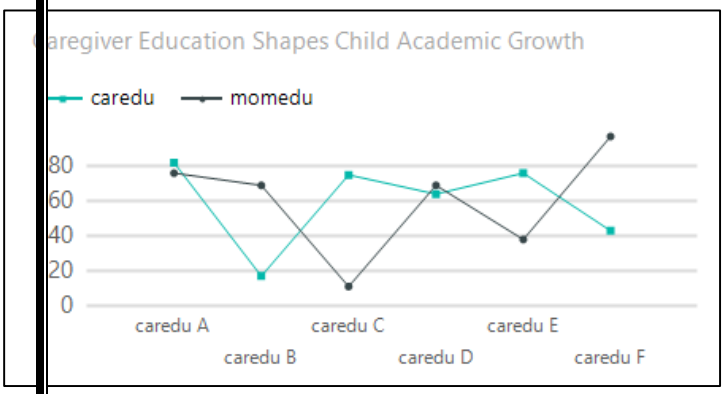
Report 5-The Role of Parental Education in Shaping Child Academic Achievement

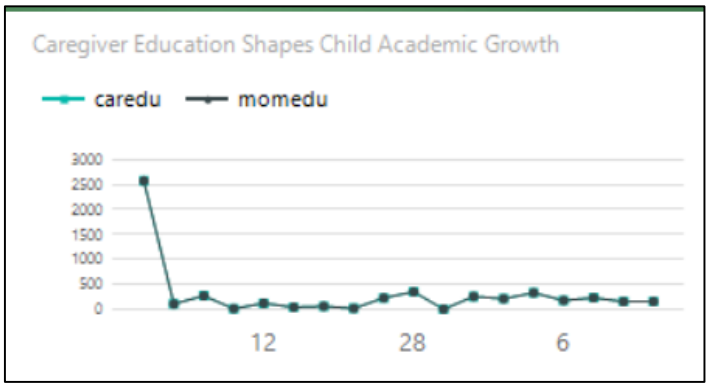




The report titled "The Role of Parental Education in Shaping Child Academic Achievement" explores the correlation between the educational level of parents, especially caregivers, and the academic growth of children. It presents data on how parental education affects various child development outcomes, such as school enrollment, time spent in school, reading level, and grade

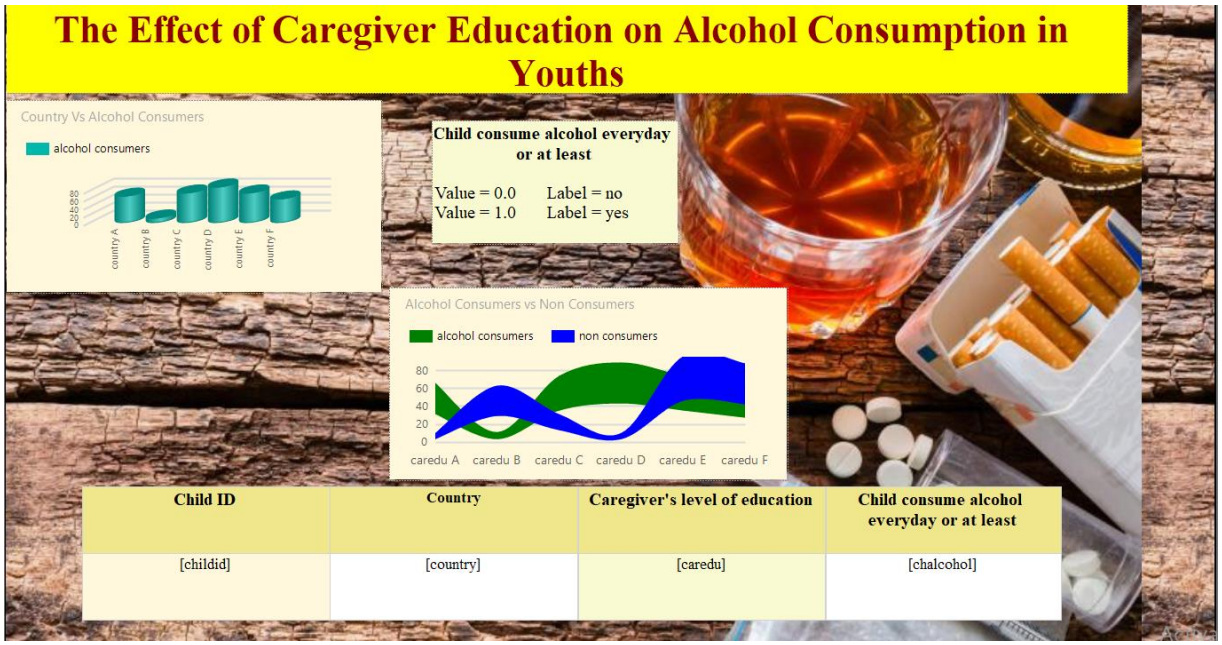
enrollment. The summary section compares relative risks of child mortality in different life stages—neonatal, post-neonatal, and childhood—based on parental education levels. The table at the bottom highlights specific child data, including the country (India and Ethiopia in this case), caregiver's education level, time spent at school, child's age, current enrollment status, reading level, and more.

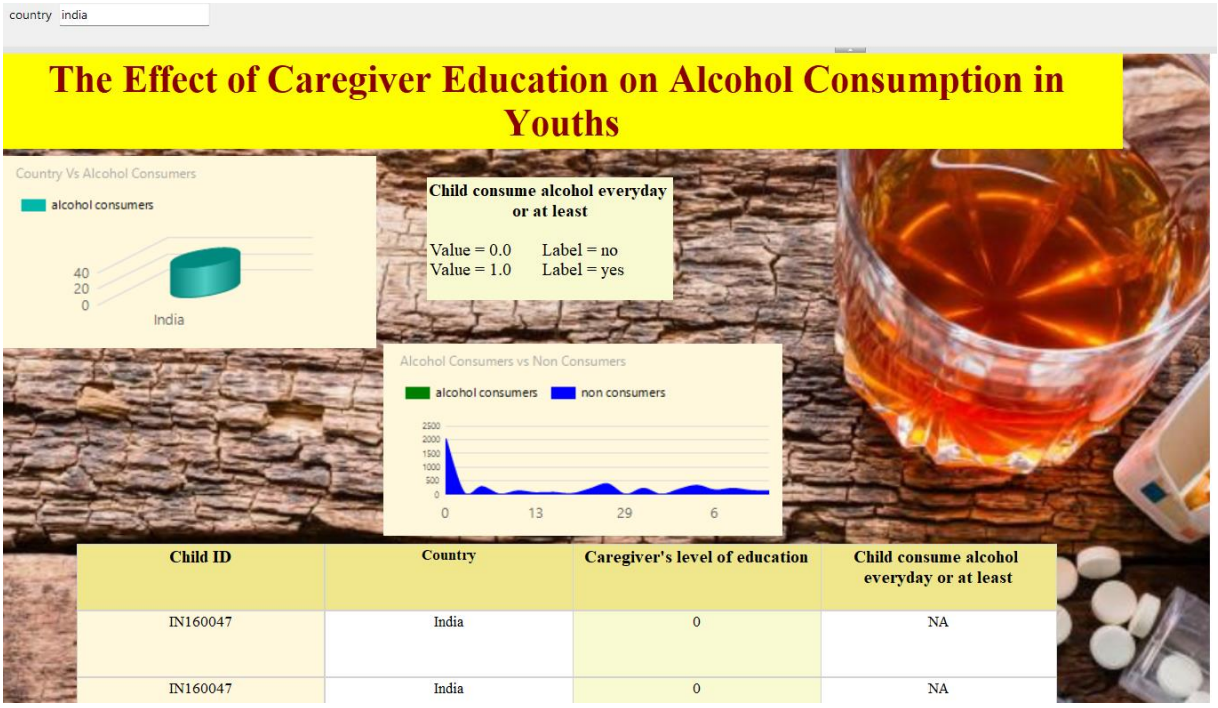




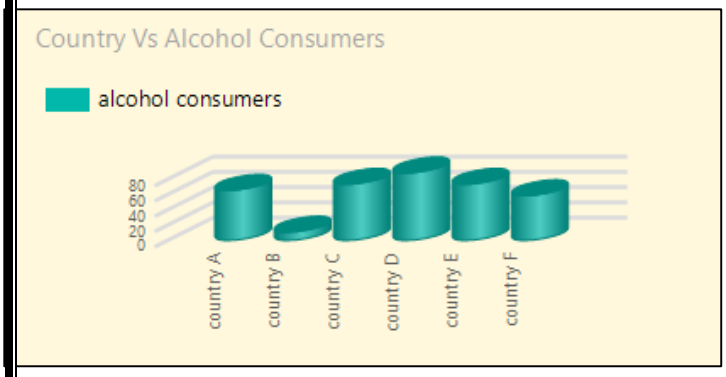
The line chart visualizes the relationship between caregiver education (caredue) and mother's education (momedu) over time in terms of child academic growth. The vertical axis represents the academic growth, while the horizontal axis tracks the progress across different months or years. Initially, there is a sharp increase, followed by a consistent decline and stabilization, indicating a critical early impact of parental education, particularly at the onset of the child's academic journey. The chart suggests that as the caregiver's education improves, there is a significant but short-term boost in the child's academic achievement, which gradually stabilizes over time.

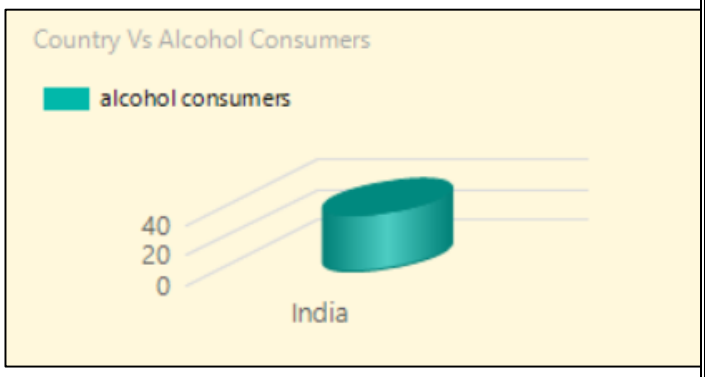
Report6- The Effect Of Caregiver Education On Alcohol Consumption In Youths



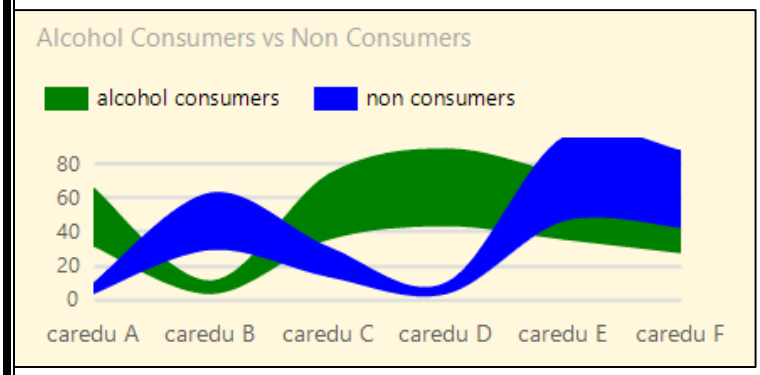


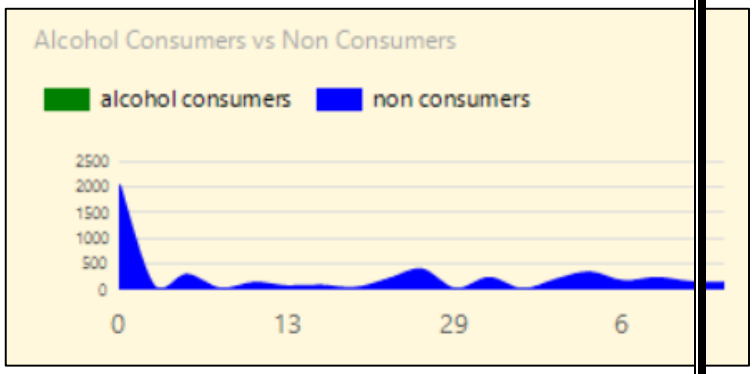
This report explores the effect of caregiver education on alcohol consumption in youths. It focuses on how the education level of caregivers influences youth behavior, specifically their likelihood of consuming alcohol regularly. The report includes a country parameter, allowing filtering by India, and presents visual data that compares the alcohol consumption habits of children in relation to their caregivers' education levels. A table at the bottom lists Child IDs, country, caregiver's level of education, and the status of child alcohol consumption (daily or occasional).





The chart shows the proportion of alcohol consumers in India. It indicates that a certain percentage of youths, relative to the total, are alcohol consumers. The chart is a simple visual representation of how many youths are consuming alcohol compared to the overall population within the selected country (India in this case).

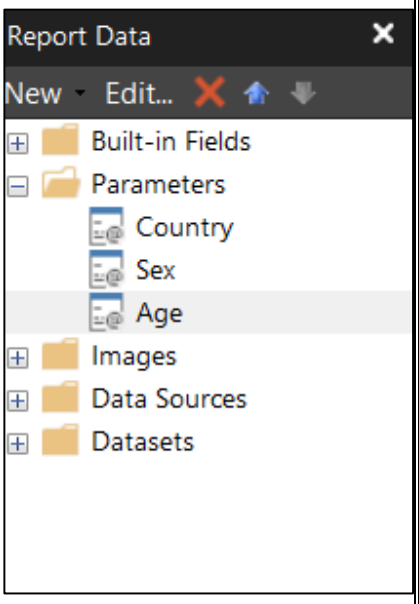




The line chart on the right compares alcohol consumers to non-consumers among youths. The "Alcohol Consumers" (green) and "Non-Consumers" (blue) trend lines show the distribution of these groups over a specified period or age range. This chart highlights that the number of non-consumers significantly outweighs the alcohol consumers, indicating that most children do not consume alcohol regularly.

ADD PARAMETER

1) Add a Country Parameter:
- Go to **Report Data pane → Parameters→ Right-click and choose Add Parameter**.
- Name: `Country`, Prompt: `Select Country`, Data Type: `Text`.



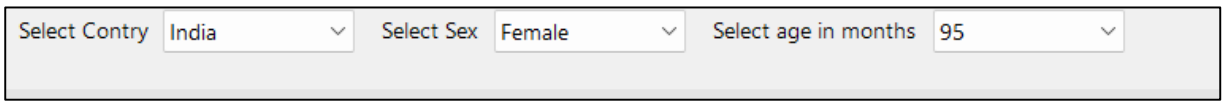
2) Set Available Values:
- Under Available Values tab, choose Specify Value.
- Add two values:
  - Label: 'India', Value: 'India'
  - Label: 'Ethiopia', Value: 'Ethiopia'

3) Filter Dataset by Country:
- **Right-click the dataset → Dataset Properties → Filters.**

- Add a new filter:
  - Expression: `Country`
  - Operator: `=` (equals)
  - Value: `=Parameters!Country.Value`

4) Preview the Report:

- Select **India** or **Ethiopia** from the drop-down and verify the report filters correctly by country.



Conclusion

This report provides an in-depth analysis of child poverty, health, and educational outcomes in lower-income countries, specifically focusing on India and Ethiopia. Utilizing data from the Young Lives project, this analysis offers valuable insights into the complex relationships between socio-economic factors such as household food security, parental education, and child well-being. By leveraging Microsoft SQL Server and Microsoft power bi Report Builder, the study organizes and visualizes key findings across various dimensions, including health, nutrition, education, and regional disparities.

One of the central findings is the significant impact of household food security on child growth and nutritional status. The report reveals that even in households with adequate food availability, malnutrition—especially underweight and stunting—remains prevalent, emphasizing the importance of dietary diversity. The data further highlights gender and age-related disparities, with girls and younger children being more susceptible to malnutrition. This suggests the need for targeted nutritional interventions that address both the quantity and quality of food.

Parental education, particularly maternal education, emerges as a critical factor influencing child health, academic performance, and even behavior. The study shows that children with better-educated caregivers tend to have better health outcomes and are more likely to succeed academically. Additionally, the report uncovers regional differences in child health, pointing to the need for localized strategies to address health inequalities. Overall, this report underscores the importance of comprehensive, targeted interventions in nutrition, education, and healthcare to improve the well-being of children in disadvantaged regions.

References

- A Guide to to young lives reserch. (2017).
- Young lives. (2016).
- https://epc.opendatacommunities.org/downloads/domestic#local-authority