UNIVERSITI MALAYSIA PAHANG
AL-SULTAN ABDULLAH

BSD2343 DATA WAREHOUSING

GROUP PROJECT

TITLE:

Exploring Global Inequalities Indicators By Analyzing Socioeconomic Disparities Among Nations

(SDG 10: REDUCE INEQUALITIES)

PREPARED FOR:

DR AZUANA BINTI RAMLI

NO	MATRIC ID	NAME	SECTION
1.	SD22011	NOR MIMI AZURA BINTI HUZAIMI	01G
2.	SD22019	NUR SABIHAH BINTI ANUAR	
3.	SD22027	HONG ZIEYIE	
4.	SD22048	CHONG JIN JYE	
5.	SD22060	FARAH HANAN BINTI TAJUDIN	02G

TABLE OF CONTENT

1.0 BACKGROUND

1.1 Description of Project

Currently, the world is dealing with a lot of unfairness in politics and society. This has made life really tough for people in certain areas. Some examples of these unfair situations include the growing gap between the rich and the poor. Some groups of people are not able to live a better life simply because of the economic inequality. Besides, there are also gender gaps that undervalue women's work and limit their chances for high-paying jobs. With increasing awareness about these issues, the United Nations has set up the 2030 Sustainable Development Goals (SDGs). One of these goals, SDG 10 focused on how to reduce inequalities within and between countries.

First and foremost, economic inequality is obviously one of the most common forms of inequality. Everyday, we witness the reality that some people possess great wealth but others struggle to make a living and survive. Many factors such as education and politics have contributed to this sad reality. For example, the rich in Brazil owns most of the wealth while the poor are unable to meet basic needs such as food and healthcare. Many professionals have started to question the position of billionaires in the economy and investigate what should be done to distribute this wealth. This inequality has severely affected the living standards of the people. The next social inequality is the people's status and power in society which is affected by stereotypes of gender and race. The world cannot live in harmony if such ignorant bias continues to spread, and eventually leads to mindless conflict among the people.

Furthermore, the surge in health inequality poses substantial health risks to the people struggling with poverty. The most common issue would be rural populations struggling with unhygienic living environments along with inadequate medical resources. Certain countries in Africa face poor health status because they lack the proper technology and medical resources. Diseases such as AIDS, malaria, and many more have a serious impact on health in these countries. Next, educational inequality also affects the future of the people, with some people unable to obtain a good education, limiting their career development and quality of life. For example, prolonged war and conflict have led to the collapse of the education system in Afghanistan, with many children unable to attend school. In addition, regional inequality is manifested due to the differences in the development rate and resource distribution between regions. In Pakistan, for example, the issue of inequality relates to differences between urban and rural areas. Urban areas are usually more prosperous while rural areas have to face biased and unequal treatment.

Essentially, the goal of SDG 10 is to reduce economic, social, health, education, and regional inequalities to secure equal opportunities and rights for all. Achieving this goal requires governments, international organizations, and all sectors of society to work together to ensure that human development is comprehensive and sustainable. We need to build a pluralistic society with mutual respect, where respect and tolerance are demonstrated between different ethnic groups so that they can co-exist peacefully and without conflict with each other. Therefore, we need to identify and understand the cause of these inequalities to create a more just and inclusive social environment.

1.2 Problem to Be Solved

Inequalities in all aspects have persisted as a liability to everyone in the world, leading to biased and unfair treatment of the unfortunate. Many unfortunate groups of people have been burdened by these socioeconomic inequalities, which includes aspects such as finance, quality of life and education. If the basic needs and rights of humans are unprotected, the people will have to suffer unjust treatment and low living standards. The impact of this matter affects beyond just individuals, as it also poses a risk to many organizations and governments. There are countries that are struggling with financial inequalities, which has led to a downfall of education and technology levels. Subsequently, the people are the ones who have to suffer from the consequences.

A thorough investigation is required to identify the various forms of inequalities that have struck people from all around the world. As this issue remains unresolved, we can never expect a peaceful and prosper future on a global scale. The obligation falls on the shoulders of every party involved, including governments, organizations and businesses to make an effort in addressing this matter. The world must accept the reality that a unified effort is mandatory to resolve all the underlying intricacies of inequalities. By identifying the underlying causes, these socioeconomic inequalities can gradually be addressed to build a better and brighter future for all.

Currently, there has been an effort to gather data regarding all these inequalities as accomplished by the United Nations (UN). The UN has made a global statement called the Sustainable Development Goals (SDGs) to ensure that every nation works together to address these issues including inequalities. The problem remains to transform these data into value by discovering actionable insights that can determine what kind of inequalities currently exist and the cause of their existence.

1.3 Objective

1. To identify which demographic group is most susceptible to socioeconomic inequalities by analyzing multidimensional datasets using OLAP techniques.
2. To deliver meaningful insights using suitable data visualizations to uncover the patterns within the data.
3. To provide reasonable explanation and recommendations regarding the various inequality issues.

1.4 Data schema

A database schema usually refers to the structure and organization of data within a database or a dataset. Database schema also defines how data will be organized within a relational database. There are various ways that could be used to arrange data schema objects in order to build a designed data warehouse for our database.

The following are the definitions of various scientific abbreviations:

- country: Country or Area
- year: Year
- mean_years_of_schooling: Mean years of schooling completed by population
- expected_years_of_schooling: Expected years of schooling completed by population
- human_development_index: Human Development Index, a summary measure of average achievement in key dimensions of human development (a long and healthy life, being knowledgeable and having a decent standard of living)
- life_expectancy_years: Life Expectancy at Birth (years)
- gross_national_income_per_capita: Gross National Income per Capita (2017 PPP$)
- hdicode: Human Development Index Categories
- region: UNDP Developing Regions
- poverty_percent: Poverty headcount ratio at $6.85 a day (2017 PPP) (% of population)
- gross_expenditure_rd: Gross domestic expenditure on research and development (GERD) in '000 current PPP$

- population_using_improved_water_facilites: Population using improved drinking-water sources (%)
- gdp: Gross Domestic Product, PPP (current international $)

Figure below shows all tables under the public schema of the "warehouse" database.

	table_catalog name	table_schema name	table_name name
1	Warehouse	public	world_region
2	Warehouse	public	world_gdp
3	Warehouse	public	world_indicator
4	Warehouse	public	world_poverty
5	Warehouse	public	world_rd_expenditure
6	Warehouse	public	world_water_faci

Figure 1.4.1: Table of public schema

Figure 1.4.1 shows the database consists of 6 tables such as world_region, world_gdp, world_indicator, world_poverty, world_rd_expenditure, and world_water_faci.

TABLE 1

No	Table Name	Description	Column Name	Data Type	Key
1.	world_region	Used to store information about countries, including years of the data, code represented human development and the region the country belongs to.	country	text	Primary key
			year	integer	Primary key
			hdicode	text	
			region	text	

Table 1.4.1: World Region

```
1  SELECT * FROM public."world_region";
```

Data Output | Messages | Notifications

	country [PK] character varying	year [PK] integer	hdicode character varying	region character varying
1	Afghanistan	2022	Low	Sub-Saharan Africa
2	Albania	2022	High	Europe and Central Asia
3	Algeria	2022	High	Arab States
4	Andorra	2022	Very High	Others
5	Angola	2022	Medium	Sub-Saharan Africa
6	Antigua and Barbuda	2022	Very High	Latin America and the Caribbean
7	Argentina	2022	Very High	Latin America and the Caribbean
8	Armenia	2022	High	Europe and Central Asia
9	Australia	2022	Very High	Others
10	Austria	2022	Very High	Others
11	Azerbaijan	2022	High	Europe and Central Asia
12	Bahamas	2022	Very High	Latin America and the Caribbean
13	Bahrain	2022	Very High	Arab States
14	Bangladesh	2022	Medium	Sub-Saharan Africa
15	Barbados	2022	Very High	Latin America and the Caribbean
16	Belarus	2022	Very High	Europe and Central Asia
17	Belgium	2022	Very High	Others

Total rows: 193 of 193 | Query complete 00:00:00.099

Figure 1.4.2: World Human Development Index Categories by Region

Figure 1.4.2 shows the world Human Development Index (HDI) categories by region from the dataset. Based on the output, this table contains 4 columns and 193 rows which are country, year, Human Development Index categories and region.

TABLE 2

No	Table Name	Description	Column Name	Data Type	Key
2.	world_gdp	Holds information about the gross domestic product of various countries for a specific year.	country	text	Primary key
			year	integer	Primary key
			gdp	numeric	

Figure 1.4.2: World Gross Domestic Product (GDP)

```
1  SELECT * FROM public."world_gdp";
```

Data Output | Messages | Notifications

	country [PK] character varying	year [PK] integer	gdp numeric
1	Afghanistan	2021	67125058170.0
2	Afghanistan	2020	81007485783.0
3	Afghanistan	2019	81889326057.0
4	Afghanistan	2018	77417896456.0
5	Afghanistan	2017	74711922906.0
6	Afghanistan	2016	70097956089.0
7	Afghanistan	2015	71831696728.0
8	Afghanistan	2014	69058343420.0
9	Afghanistan	2013	65039839449.0
10	Afghanistan	2012	59667003515.0
11	Afghanistan	2011	51184181904.0
12	Afghanistan	2010	49929490245.0
13	Afghanistan	2009	43140533618.0
14	Afghanistan	2008	35312296534.0
15	Afghanistan	2007	33339216567.0
16	Afghanistan	2006	28518821098.0
17	Afghanistan	2005	26258419344.0

Total rows: 1000 of 7728 | Query complete 00:00:00.108

Figure 1.4.3: World Gross Domestic Product (GDP)

Figure 1.4.3 shows the world gross domestic product (GDP) dataset. Based on the output, it shows that the table contains 3 columns and 7728 rows which are country, year, and gross domestic product (GDP).

TABLE 3

No	Table Name	Description	Column Name	Data Type	Key
3.	world_indicator	Stores information on development indicators, including educational dimensions, human development index, life expectancy, and income per capita for each company by year.	country	text	Primary key
			year	integer	Primary key
			mean_years_of_schooling	numeric	
			expected_years_of_schooling	numeric	
			human_development_index	numeric	
			life_expectancy_years	numeric	
			gross_national_income_per_capita	numeric	

Table 1.4.3: World Indicator

```
1  SELECT * FROM public."world_indicator";
```

Data Output | Messages | Notifications

	country [PK] character varying	year [PK] integer	mean_years_of_schooling numeric	expected_years_of_schooling numeric	human_development_index numeric	life_expectancy_years numeric	gro nur
1	Afghanistan	2000	1.264052241	5.856421507	0.34	55.298	
2	Afghanistan	2001	1.315550666	6.148417656	0.344	55.798	
3	Afghanistan	2002	1.367049091	6.440413805	0.368	56.454	
4	Afghanistan	2003	1.418547515	6.732409954	0.379	57.344	
5	Afghanistan	2004	1.47004594	7.600430012	0.395	57.944	
6	Afghanistan	2005	1.521544365	7.858029938	0.402	58.361	
7	Afghanistan	2006	1.595280745	8.115629864	0.41	58.684	
8	Afghanistan	2007	1.669017126	8.37322979	0.426	59.111	
9	Afghanistan	2008	1.742753507	8.630829716	0.431	59.852	
10	Afghanistan	2009	1.816489888	8.888429642	0.441	60.364	
11	Afghanistan	2010	1.890226268	9.180809975	0.449	60.851	
12	Afghanistan	2011	1.937043018	9.473190308	0.457	61.419	
13	Afghanistan	2012	1.983859768	9.803686778	0.467	61.923	
14	Afghanistan	2013	2.030676517	10.13418325	0.475	62.417	
15	Afghanistan	2014	2.077493267	10.46467972	0.48	62.545	
16	Afghanistan	2015	2.124310017	10.48297477	0.479	62.659	

Total rows: 1000 of 4581 | Query complete 00:00:00.296 | Ln 1, Col 38

Figure 1.4.4: World Indicator

Figure 1.4.4 shows the world indicator dataset. Based on the output, it shows that the table contains 7 columns and 4581 rows which are country, year, and mean years of schooling, expected year of schooling, Human Development Index, life expectancy years, gross national income per capita.

TABLE 4

No	Table Name	Description	Column Name	Data Type	Key
4.	world_poverty	Contains data on poverty levels in multiple countries for specific years.	country	text	Primary key
			year	integer	Primary key
			poverty_percent	numeric	

Table 1.4.4: World Poverty

```
1  SELECT * FROM public."world_poverty";
```

Data Output | Messages | Notifications

	country [PK] character varying	year [PK] integer	poverty_percent numeric
1	Albania	2020	13.7
2	Albania	2019	14.9
3	Albania	2018	18.2
4	Albania	2017	25.5
5	Albania	2016	25.4
6	Albania	2015	26.2
7	Albania	2014	38.7
8	Albania	2012	36.7
9	Albania	2008	33.4
10	Albania	2005	40.4
11	Albania	2002	52.1
12	Albania	1996	47.9
13	Algeria	2011	36.6
14	Algeria	1995	61.1
15	Algeria	1988	63.6
16	Angola	2018	78.0
17	Angola	2008	69.1

Total rows: 1000 of 2498 Query complete 00:00:00.209

Figure 1.4.5: World Poverty

Figure 1.4.5 shows the world poverty dataset. Based on the output, it shows that the table contains 3 columns and 2498 rows which are country, year, and poverty percent.

TABLE 5

No	Table Name	Description	Column Name	Data Type	Key
5.	world_rd_expenditure	To store information about the gross expenditure on research and development activities for each country.	country	text	Primary key
			year	integer	Primary key
			gross_expenditure	numeric	

Table 1.4.5: World Research and Development Expenditure

```
1  SELECT * FROM public."world_rd_expenditure";
```

Data Output | Messages | Notifications

Graph Visualiser

	country [PK] character varying	year [PK] integer	gross_expenditure_rd numeric
1	Albania	2008	39831.94
2	Albania	2007	19875.92
3	Algeria	2005	242321.25
4	Algeria	2004	549074.75
5	Algeria	2003	614641.59
6	Algeria	2002	1049650.27
7	Algeria	2001	615279.99
8	Argentina	2011	4655464.79
9	Armenia	2014	58412.2
10	Armenia	2013	51322.61
11	Armenia	2012	52560.9
12	Armenia	2011	54549.21
13	Armenia	2010	45612.71
14	Armenia	2009	53142.37
15	Armenia	2008	47381.85
16	Armenia	2007	40891.63
17	Armenia	2006	39911.74

Total rows: 1000 of 1494 Query complete 00:00:00.114

Figure 1.4.6: World Research and Development Expenditure

Figure 1.4.6 shows the world research and development expenditure dataset. Based on the output, it shows that the table contains 3 columns and 1494 rows which are country, year, and gross expenditure for research and development.

TABLE 6

No	Table Name	Description	Column Name	Data Type	Key
6.	world_water_faci	Used to store information about the percentage of the population using improved water facilities in each country for each specific year.	country	text	Primary key
			year	integer	Primary key
			population_using_improved_water_facilities	integer	

Table 1.4.6: World Water Facilities

```
SELECT * FROM public."world_water_faci";
```

Data Output | Messages | Notifications

	country [PK] character varying	year [PK] integer	population_using_improved_water_facilities integer
1	Afghanistan	2012	64
2	Afghanistan	2010	57
3	Afghanistan	2005	40
4	Afghanistan	2000	22
5	Afghanistan	1995	5
6	Albania	2012	96
7	Albania	2010	96
8	Albania	2005	96
9	Albania	2000	96
10	Albania	1995	96
11	Algeria	2012	84
12	Algeria	2010	84
13	Algeria	2005	86
14	Algeria	2000	89
15	Algeria	1995	93
16	Algeria	1990	94
17	Andorra	2012	100

Total rows: 1000 of 1096 Query complete 00:00:00.106

Figure 1.4.7: World Water Facilities

Figure 1.4.7 shows the world water facilities dataset. Based on the output, it shows that the table contains 3 columns and 1096 rows which are country, year, and population using improved water facilities.

2.0 ARCHITECTURE AND ETL PIPELINE

2.1 Pipeline Structure

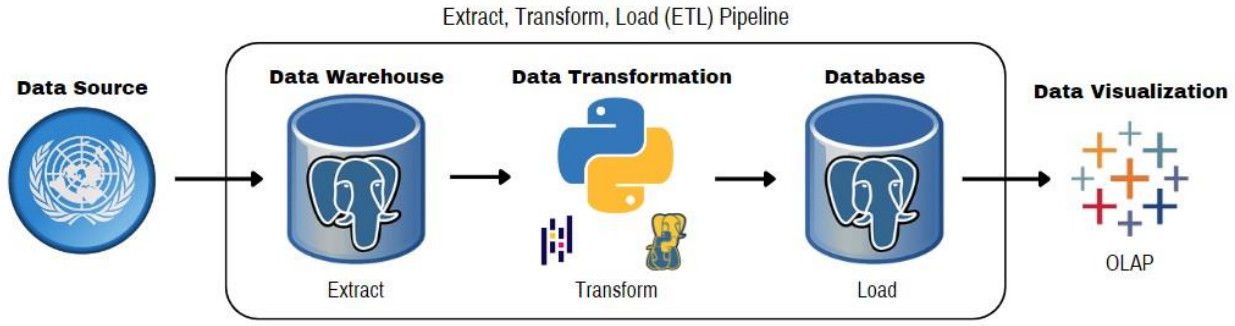


Figure 2.1 Structure of ETL

The data warehouse for the project integrates the bottom-up technique by Kimball. The pipeline structure is designed with ease of use and simplicity in mind to provide an optimal setting for data analysis. Kimball's approach is preferred for its smart warehouse structure for the delivery of meaningful and rapid insights using data marts. Therefore, this technique is suitable for a small-scale analytics team in this study to focus on efficient and cost-effective operations.

The primary data source for the pipeline is collected from the United Nations (UN) as a third-party data source for the investigation. The UN is an international diplomatic body whose main duty involves introducing the Sustainable Development Goals (SDGs) for the sake of world peace. The study decides to obtain datasets from the official website of UN because of its trustworthiness and reliability where the data is collected through a joint effort of multiple countries for various data and indicators that are compatible with the SDGs. This is why it is very important to establish a reliable data source to ensure that the results obtained are accurate.

First, the data is loaded into the data warehouse acting as a single storage system that stores raw data from the data source. We want to be able to store the raw structured data as tables in the database. Therefore, the pipeline will rely on a relational database management system (DBMS) known as PostgreSQL. Not only for the feature of open source, it also has main functions that allow users to handle the data that comes from data warehouse effectively and perform queries on it. Next, data transformation is done by using IPython in Jupyter Notebook along with various open-source libraries like pandas and psycopg2. The datasets are then imported straight from the data warehouse to Jupyter Notebook for a comprehensive transformation operation such as handling the missing values, dropping the unwanted columns and checking on the data types.

After the data is transformed, it is then loaded into a new database in PostgreSQL which stores clean datasets that are ready for online analytical processing (OLAP). OLAP operations are performed in PostgreSQL to perform querying on multi-dimensional data. After that, data visualization tools like Power BI will be employed to analyze the queried data and examine the patterns and insights that can be used in further decision-making. Power BI is a popular business intelligence tool that is widely used and helpful when it comes to reporting purposes. In conclusion, this complete pipeline structure allows for the extraction of raw data from third-party sources, followed by the subsequent transformation and loading of data into databases for visualization and analysis.

2.2 ETL Pipeline

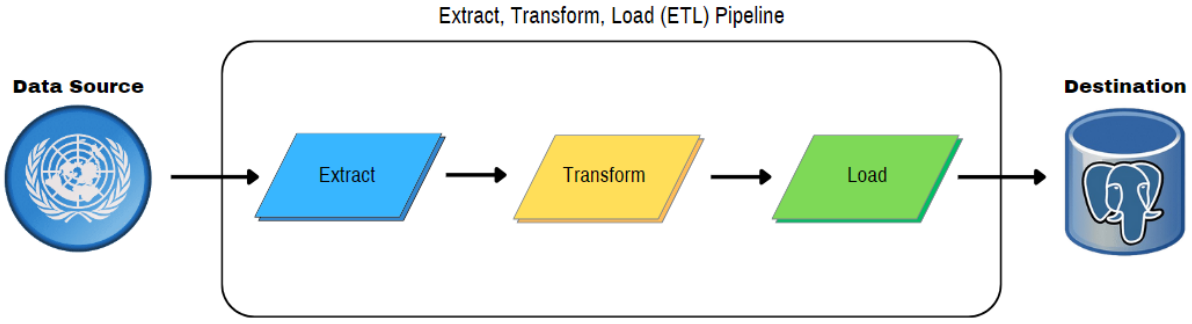


Figure 2.2 ETL Pipeline

Extract, Transform, Load (ETL) is a well-known method applied for various data operations in the industry. This study depicts a clear and comprehensive ETL pipeline that is used to process the data before it enters its designated destinations. The main objective of the ETL pipeline is to extract raw data from the data source to transform it into clean data before loading it into the database. A total of 6 datasets are selected from the official website of the United Nations (UN), thus proper data cleaning and integration is required. Initially, the raw data is stored in the data warehouse using PostgreSQL before extracted into Jupyter Notebook for transformation. The data transformation process is present in order to ensure dirty data is thoroughly cleaned so that the data can be converted into valuable and meaningful insights. Finally, the clean data will be imported into the database and it is ready for analysis. One thing worth noting is that data integration was not an issue because the file format was kept as a csv file throughout the pipeline. Any queries to join and filter the datasets can be made to meet the needs of the target audience for meaningful data visualizations.

2.3 Extracting Raw Data into PostgreSQL

1. A database named “WarehouseProject” is created in PostgreSQL as the data warehouse to store raw data.

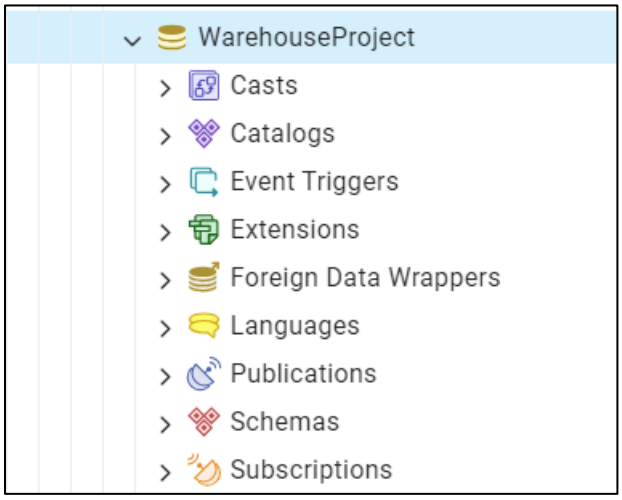


Figure 2.3.1

2. A table named “world_indicator” is created to store raw data on human development indicators.

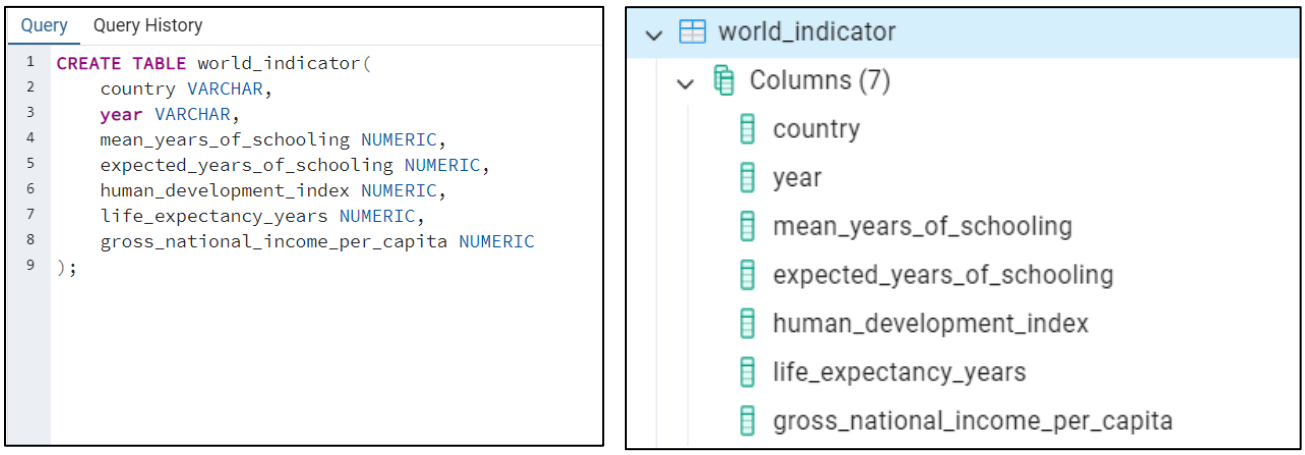


Figure 2.3.2

3. The local file is imported into PostgreSQL.

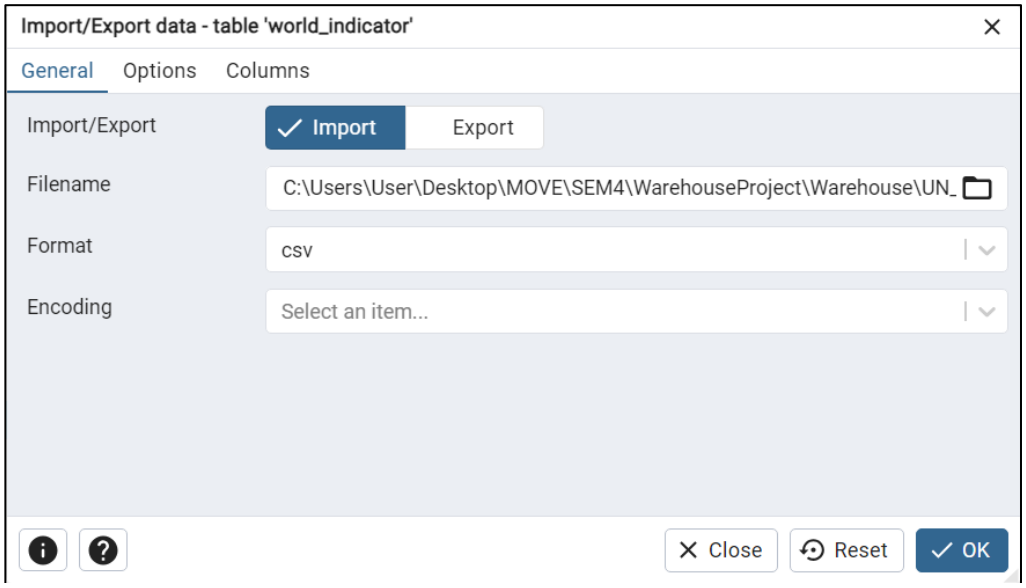
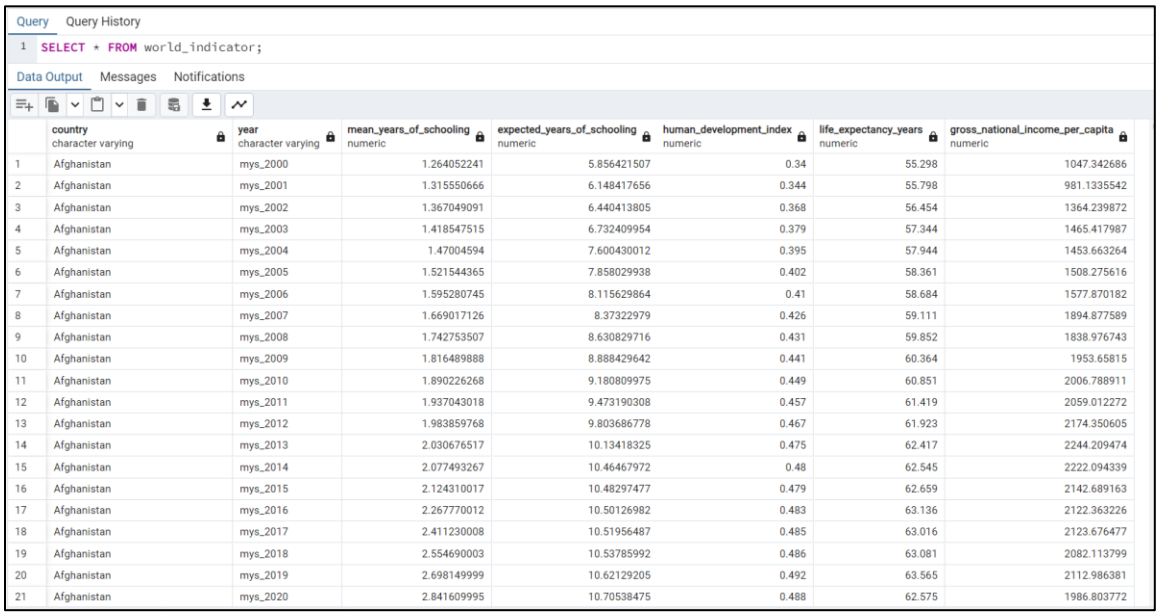


Figure 2.3.3

4. The "SELECT * FROM world_indicator" query is used to confirm the successful import of dataset into the data warehouse.



	country character varying	year character varying	mean_years_of_schooling numeric	expected_years_of_schooling numeric	human_development_index numeric	life_expectancy_years numeric	gross_national_income_per_capita numeric
1	Afghanistan	mys_2000	1.264052241	5.856421507	0.34	55.298	1047.342686
2	Afghanistan	mys_2001	1.315550666	6.148417656	0.344	55.798	981.1335542
3	Afghanistan	mys_2002	1.367049091	6.440413805	0.368	56.454	1364.239872
4	Afghanistan	mys_2003	1.418547515	6.732409954	0.379	57.344	1465.417987
5	Afghanistan	mys_2004	1.47004594	7.600430012	0.395	57.944	1453.663264
6	Afghanistan	mys_2005	1.521544365	7.858029938	0.402	58.361	1508.275616
7	Afghanistan	mys_2006	1.595280745	8.115629864	0.41	58.684	1577.870182
8	Afghanistan	mys_2007	1.669017126	8.37322979	0.426	59.111	1894.877589
9	Afghanistan	mys_2008	1.742753507	8.630829716	0.431	59.852	1838.976743
10	Afghanistan	mys_2009	1.816489888	8.888429642	0.441	60.364	1953.65815
11	Afghanistan	mys_2010	1.890226268	9.180809975	0.449	60.851	2006.788911
12	Afghanistan	mys_2011	1.937043018	9.473190308	0.457	61.419	2059.012272
13	Afghanistan	mys_2012	1.983859768	9.803686778	0.467	61.923	2174.350605
14	Afghanistan	mys_2013	2.030676517	10.13418325	0.475	62.417	2244.209474
15	Afghanistan	mys_2014	2.077493267	10.46467972	0.48	62.545	2222.094339
16	Afghanistan	mys_2015	2.124310017	10.48297477	0.479	62.659	2142.689163
17	Afghanistan	mys_2016	2.267770012	10.50126982	0.483	63.136	2122.363226
18	Afghanistan	mys_2017	2.411230008	10.51956487	0.485	63.016	2123.676477
19	Afghanistan	mys_2018	2.554690003	10.53785992	0.486	63.081	2082.113799
20	Afghanistan	mys_2019	2.698149999	10.62129205	0.492	63.565	2112.986381
21	Afghanistan	mys_2020	2.841609995	10.70538475	0.488	62.575	1986.803772

Figure 2.3.1

The SQL queries and csv of raw data for other tables can be accessed here:

https://drive.google.com/drive/folders/1beXvNyF0is9Fth1Ypy6vct63v0cDewky?usp=sharing

2.4 Transforming Data Using Jupyter Notebook

1. Before proceeding to data transformation, the required Python packages should be installed and imported first. One of the required packages is psycopg2 which enables the connection of databases in PostgreSQL to Jupyter Notebook. This essentially allows datasets from PostgreSQL to be imported into Jupyter Notebook to carry out data transformation.

```
[46]
# import necessary packages
import pandas as pd
import psycopg2 as ps
import pandas.io.sql as sqlio
import warnings
warnings.filterwarnings("ignore", message="pandas only supports SQLAlchemy connectable")

[47]
# the psycopg2 package is used to establish a connection to the datawarehouse in PostgreSQL
conn=ps.connect(dbname="WarehouseProject",user="postgres",password= "123123",host="localhost",port="5432")
```

Figure 2.4.1

2. The "world_indicator" table from PostgreSQL is imported into a data frame variable named "df1". The info() method is used to briefly explore the data and identify the data types.

Table 1 - World Human Development Indicators

```
[48]
# sql code to be executed
sql="""SELECT * FROM "world_indicator" """

# read the table from PostgreSQL to create a dataframe in Python
df1=sqlio.read_sql_query(sql,conn)

[49]
# explore the data
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4738 entries, 0 to 4737
Data columns (total 7 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   country                         4738 non-null   object
 1   year                            4738 non-null   object
 2   mean_years_of_schooling         4600 non-null   float64
 3   expected_years_of_schooling     4669 non-null   float64
 4   human_development_index         4581 non-null   float64
 5   life_expectancy_years           4738 non-null   float64
 6   gross_national_income_per_capita 4679 non-null  float64
dtypes: float64(5), object(2)
memory usage: 259.2+ KB
```

Figure 2.4.2

3. The first 10 rows of the data are printed for further exploration. In this case, it is observed that the "year" column is not properly formatted.

```
[50]
# review first 10 rows of data
df1.head(10)
```

	country	year	mean_years_of_schooling	expected_years_of_schooling	human_development_index	life_expectancy_years	gross_national_income_per_capita
0	Afghanistan	mys_2000	1.264052	5.856422	0.340	55.298	1047.342686
1	Afghanistan	mys_2001	1.315551	6.148418	0.344	55.798	981.133554
2	Afghanistan	mys_2002	1.367049	6.440414	0.368	56.454	1364.239872
3	Afghanistan	mys_2003	1.418548	6.732410	0.379	57.344	1465.417987
4	Afghanistan	mys_2004	1.470046	7.600430	0.395	57.944	1453.663264
5	Afghanistan	mys_2005	1.521544	7.858030	0.402	58.361	1508.275616
6	Afghanistan	mys_2006	1.595281	8.115630	0.410	58.684	1577.870182
7	Afghanistan	mys_2007	1.669017	8.373230	0.426	59.111	1894.877589
8	Afghanistan	mys_2008	1.742754	8.630830	0.431	59.852	1838.976743
9	Afghanistan	mys_2009	1.816490	8.888430	0.441	60.364	1953.658150

Figure 2.4.3

4. A loop is applied to correct the "year" column by removing the unnecessary letters. In addition, the column is converted to integer type. To check that the column is clean and properly formatted, the unique() method is used to confirm each distinct value in the column. It can be observed that the years are now represented as integers.

```
[51]
# clean and transform 'year' column into integer type
for x in range(2000, 2023):
    df1['year'] = df1['year'].str.replace(f'mys_{x}', f'{x}')

df1['year'] = df1['year'].astype(int)

# check values of 'year' column
df1['year'].unique()
```

```
array([2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010,
       2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021,
       2022])
```

Figure 2.4.4

5. The data type for columns is checked again. Then, the isnull().sum() method is applied to check for missing values. As observed, there are indeed missing values that may interfere with the OLAP operations.

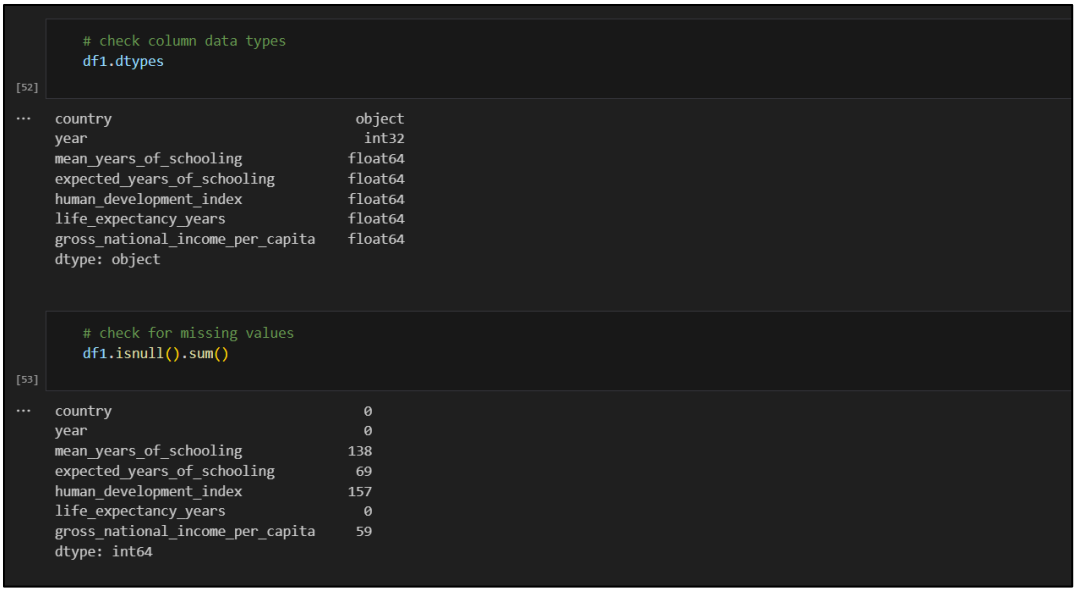


Figure 2.4.5

6. Therefore, the dropna() method is used to remove rows containing missing values. After checking for the final shape and data types of the dataset, it is confirmed that the data is clean and prepared for analysis.

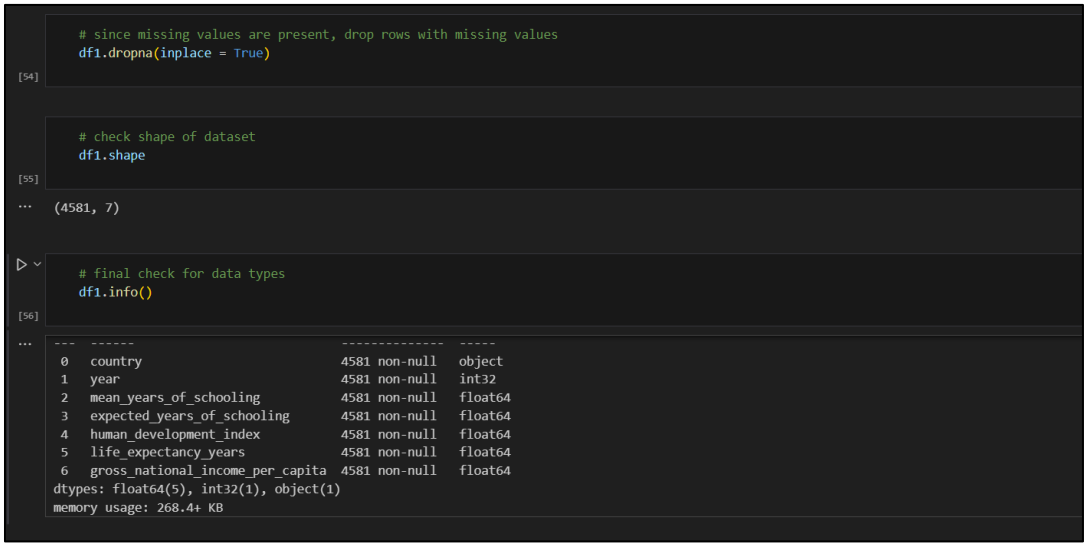


Figure 2.4.6

The complete ipynb file containing data transformation code for every table can be accessed here: https://drive.google.com/drive/folders/1AZx2Xu8vG6OFWlnZyUWFypUyf1UVddEc?usp=sharing

For a full display of output from the data transformation, please refer to the Appendix.

2.5 Loading Clean Data into PostgreSQL

1. A database named “WarehouseProjectDatabase” is created in PostgreSQL to load and store the clean data.

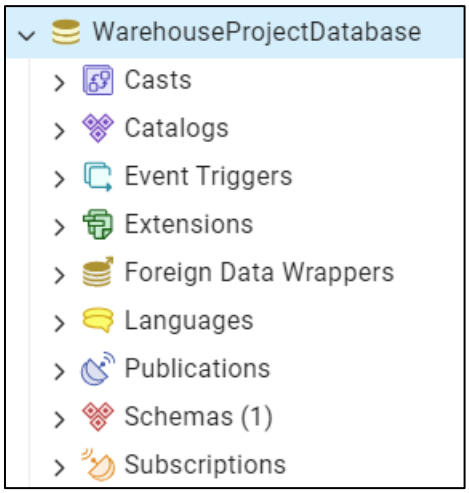
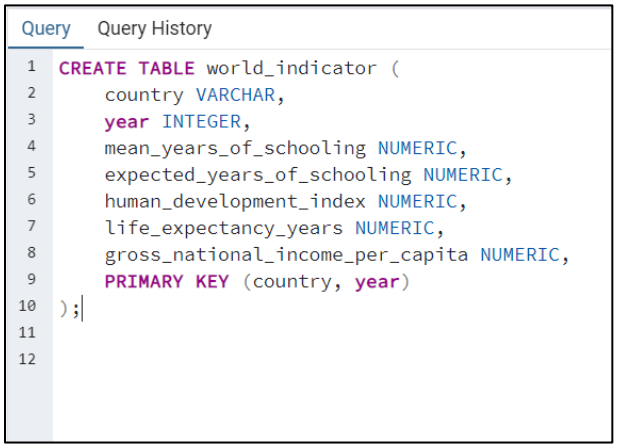


Figure 2.5.1

2. A table named “world_indicator” is created to store clean data on human development indicators. For the database, a composite primary key is utilized. The combination of “country” and “year” attributes will form a unique key that is able to uniquely identify each record in the table.



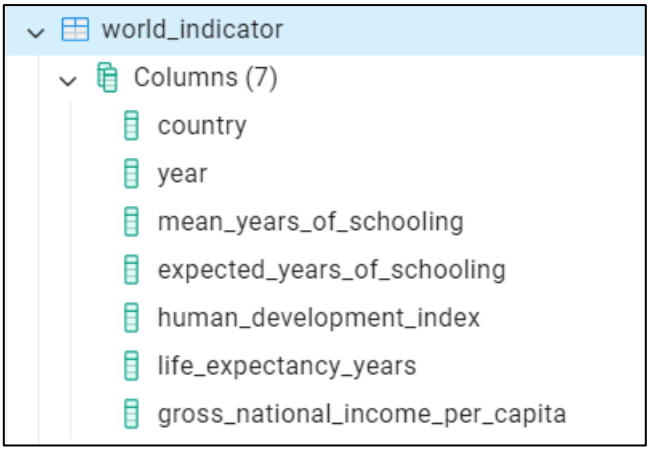


Figure 2.5.2

3. The values of the clean data is imported into PostgreSQL.

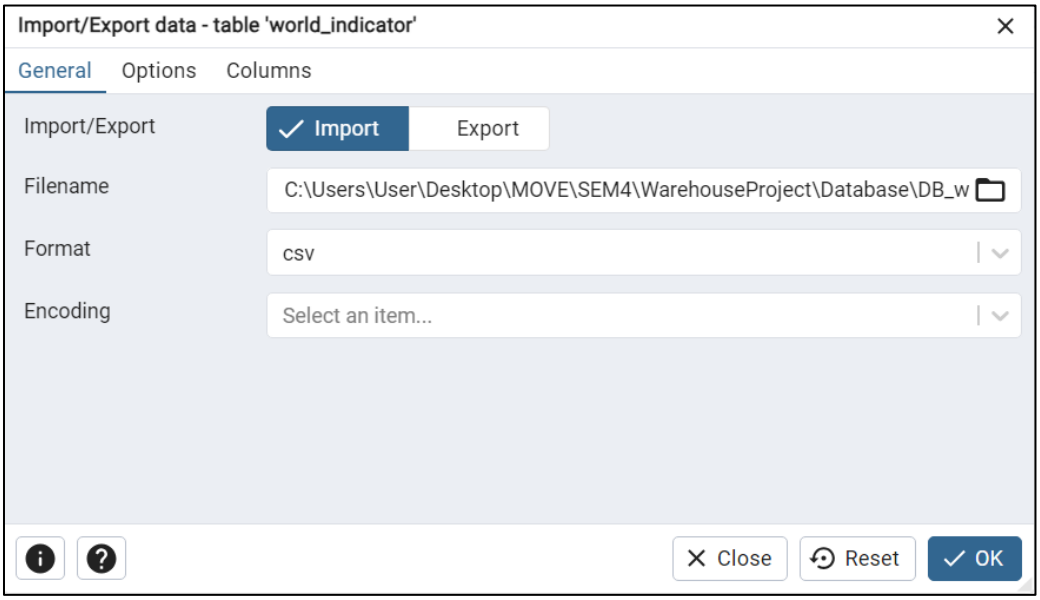
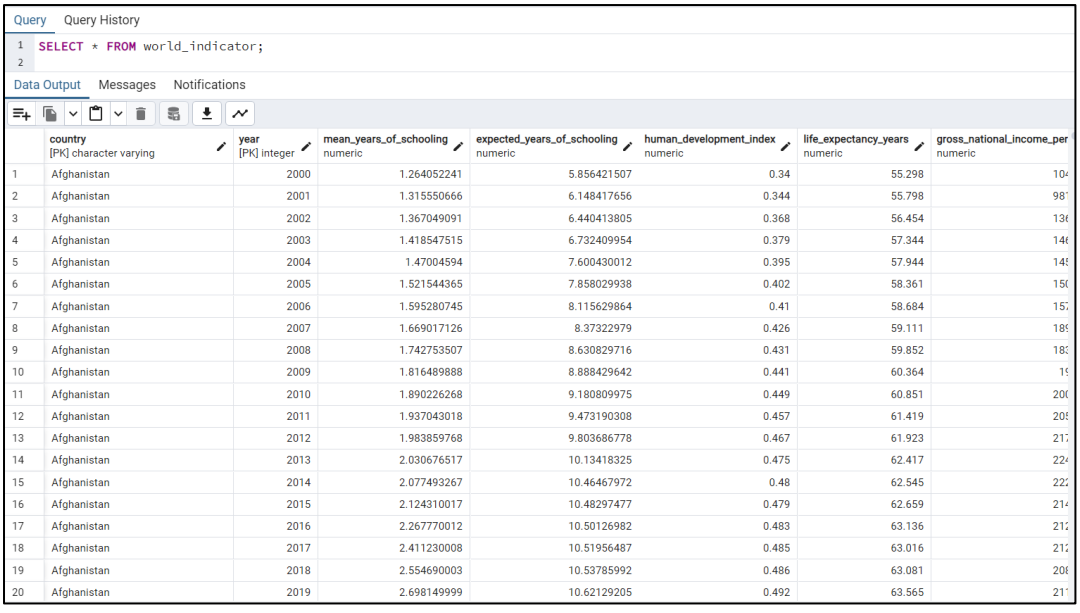


Figure 2.5.3

4. The "SELECT * FROM world_indicator" query is used to confirm the successful import of dataset into the database.



	country [PK] character varying	year [PK] integer	mean_years_of_schooling numeric	expected_years_of_schooling numeric	human_development_index numeric	life_expectancy_years numeric	gross_national_income_per numeric
1	Afghanistan	2000	1.264052241	5.856421507	0.34	55.298	10
2	Afghanistan	2001	1.315550666	6.148417656	0.344	55.798	98
3	Afghanistan	2002	1.367049091	6.440413805	0.368	56.454	13
4	Afghanistan	2003	1.418547515	6.732409954	0.379	57.344	14
5	Afghanistan	2004	1.47004594	7.600430012	0.395	57.944	14
6	Afghanistan	2005	1.521544365	7.858029938	0.402	58.361	15
7	Afghanistan	2006	1.595280745	8.115629864	0.41	58.684	15
8	Afghanistan	2007	1.669017126	8.37322979	0.426	59.111	18
9	Afghanistan	2008	1.742753507	8.630829716	0.431	59.852	18
10	Afghanistan	2009	1.816489888	8.888429642	0.441	60.364	19
11	Afghanistan	2010	1.890226268	9.180809975	0.449	60.851	20
12	Afghanistan	2011	1.937043018	9.473190308	0.457	61.419	20
13	Afghanistan	2012	1.983859768	9.803686778	0.467	61.923	21
14	Afghanistan	2013	2.030676517	10.13418325	0.475	62.417	22
15	Afghanistan	2014	2.077493267	10.46467972	0.48	62.545	22
16	Afghanistan	2015	2.124310017	10.48297477	0.479	62.659	21
17	Afghanistan	2016	2.267770012	10.50126982	0.483	63.136	21
18	Afghanistan	2017	2.411230008	10.51956487	0.485	63.016	21
19	Afghanistan	2018	2.554690003	10.53785992	0.486	63.081	20
20	Afghanistan	2019	2.698149999	10.62129205	0.492	63.565	21

Figure 2.5.4

The SQL queries and csv of clean data for other tables can be accessed here:
https://drive.google.com/drive/folders/15glvZr57zEjV-pXeB91zpTjB_nkuYdqV?usp=sharing

3.0 DATABASE

3.1 Relational Model (ERD)

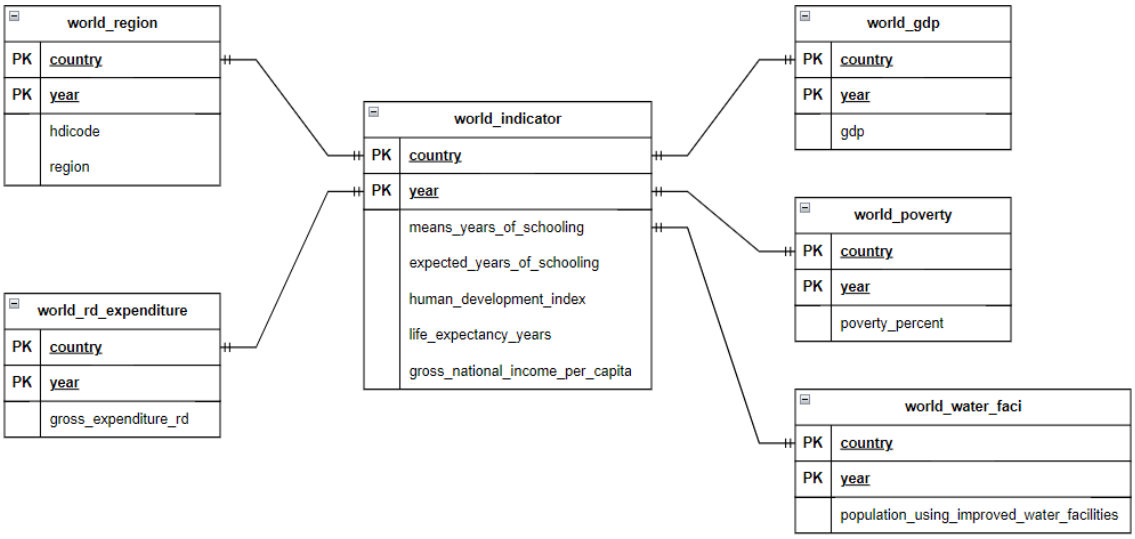


Figure 3.0 ERD

3.2 Identification of Data Warehouse Schema

Figure 3.0 shows that the relationship of this data is star schema. It can be observed that the star schema contains 6 tables which are world_ indicators as a fact table in the center. The 5 tables surrounding the world_indicator are known as the dimension table which are world_region, world_rd_expenditure, world_gdp, world_poverty and world_water_faci. These 5 tables are connected to the fact table with a one-to-one relationship using a composite primary key consisting of the country and year attribute. The use of composite primary key is possible as each combination of country and year attribute will produce a unique key which uniquely identifies each record in the dataset. The star schema design is chosen for this project because it allows for more efficient data retrieval by simplifying join operations between tables. Therefore, it is possible to study and analyze the global inequality indicators across multiple dimensions resulting in a full understanding of the factors contributing to these inequalities.

4.0 RESULTS AND DATA ANALYSIS

4.1 OLAP Coding

1) CUBE

```
SELECT g.country,  r.region, SUM(gdp) AS total_gdp
FROM world_gdp g, world_region r
WHERE g.country = r.country AND g.year = r.year
AND g.country like 'D%'
GROUP BY CUBE (g.country, r.region);
```

	country character varying	region character varying	total_gdp numeric
1	[null]	[null]	700592627696.5
2	Dominican Republic	Latin America and the Caribbean	256000000000.0
3	Denmark	Others	437000000000.0
4	Djibouti	Arab States	6605370361.0
5	Dominica	Latin America and the Caribbean	987257335.5
6	Dominican Republic	[null]	256000000000.0
7	Dominica	[null]	987257335.5
8	Djibouti	[null]	6605370361.0
9	Denmark	[null]	437000000000.0
10	[null]	Latin America and the Caribbean	256987257335.5
11	[null]	Others	437000000000.0
12	[null]	Arab States	6605370361.0

Figure 4.1.1

Figure 4.1.1 shows the data cube operation applied on the tables “world_gdp” and “world_region” connected using the composite primary key of country and year attribute. This connection allows data aggregation using the GROUP BY clause with CUBE function. The CUBE function is useful to provide a multi-dimensional view for the gross domestic product. The output showed total gross domestic product across different countries and regions. This means the dimension is created by a combination of countries and regions. At the top of the hierarchy is the grand total, represented by NULL value. This row indicates the overall total gross domestic

product across all countries and regions which is $ 700592627696.5$. This provides a high-level overview of the total gross domestic product around the world.

According to the figure, “Denmark” in “Others” regions has the higher total gross domestic product which is $ 437000000000.00 compared to other countries. These results show that Denmark has high productivity, technological advancements, strong social welfare, political stability that can contribute to economic growth and lead to higher total gross domestic product levels in Denmark and other regions. While “Djibouti” in “Arab States” we found that they have the lowest total gross domestic product which is $ 6605370361.00 compared to Denmark and Dominican Republic. This is due to Djibouti's status as one of the smallest countries in both Africa and the Arab States. The size of its economy restricts its capacity to diversify output and increases its reliance on foreign markets, making it more sensitive to market downturns and limiting its access to external financing.

At row 6, it shows the total gross domestic product across all regions in the country. The NULL in the “region” column indicates that the data is aggregated over all regions. The results are the same as previous explanations because in the table in Figure 4.1.1, each country has one region which is in “Dominican Republic” and “Dominica” has “Latin America and the Caribbean”, “Denmark” has “Others” and “Djibouti” has “Arab States”. But for the region, we can see the difference results at row 10 which is “Latin America and the Caribbean”. The NULL in the “country” column indicates that the data is aggregated across all countries. So, the total gross domestic product in the region “Latin America and the Caribbean” across countries is $ 256987257335.50. This is the sum up of total gross domestic product in Dominican Republic and Dominican.

To summarize, the GROUP BY CUBE operation is useful in providing a more complete and multi-dimensional view of the data, allowing us to analyze insight data at different levels of granularity and show subtotals for various combinations of grouped columns like country, regions and overall. This can be very useful for financial reporting, data analysis, and producing pivot-style summaries.

2) ROLL UP

```
SELECT r.region,  r.hdicode,  AVG(life_expectancy_years) AS avg_lifeExpectancy
FROM world_region r, world_indicator i
WHERE r.country = i.country AND r.year = i.year
GROUP BY ROLLUP ( r.region,  r.hdicode);
```

	region character varying	hdicode character varying	avg_lifeexpectancy numeric
1	[null]	[null]	71.8421
2	Arab States	High	73.6824
3	Arab States	Low	62.0660
4	Europe and Central Asia	Medium	71.2880
5	Arab States	Medium	72.8697
6	East Asia and the Pacific	Very High	78.6560
7	Latin America and the Caribbean	Medium	69.4428
8	East Asia and the Pacific	Medium	68.1685
9	Others	High	71.5280
10	Sub-Saharan Africa	High	72.2288
11	Latin America and the Caribbean	High	72.8105
12	Europe and Central Asia	Very High	75.3263
13	East Asia and the Pacific	High	70.8959
14	Sub-Saharan Africa	Very High	71.7380
15	Sub-Saharan Africa	Medium	64.7066
16	Europe and Central Asia	High	72.1627
17	Arab States	Very High	78.6842
18	Sub-Saharan Africa	Low	61.1118
19	Others	Very High	81.1067
20	Latin America and the Caribbean	Very High	76.5764
21	Latin America and the Caribbean	[null]	73.1352
22	East Asia and the Pacific	[null]	70.9374
23	Others	[null]	80.8786
24	Europe and Central Asia	[null]	73.3444

Total rows: 26 of 26 Query complete 00:00:00.125

Figure 4.1.2

Figure 4.1.2 shows the roll up operation on the tables "world_region" and "world_indicator" that connect both using the common identifiers "country" and "year". This connection allows data aggregation using the GROUP BY clause with ROLLUP function. The ROLLUP function is useful for generating subtotal and grand total by aggregating data from various levels of hierarchy within a dimension. The output showed average life expectancy across different regions and HDI categories. This means the dimension is created by combination of region and HDI categories. At the top of the hierarchy is the grand total, represented by NULL

value. This NULL value indicates the overall average life expectancy across all regions and HDI categories which is 71.8421. This provides a high-level overview of the average life expectancy globally.

It can be observed that in East Asia and the Pacific with a very high Human Development Index (HDI), the average life expectancy is clearly high. This possibly reflects the presence of good healthcare systems, financial stability, and social development. It is safe to deduce that countries with higher levels of human development have longer life expectancies than regions with lower levels. Meanwhile, in Sub-Saharan Africa with a low HDI, it is found that the average life expectancy is lower. This points to serious issues in healthcare access, economic stability, and social growth. Factors such as a high rate of infectious diseases, poor healthcare infrastructure, restricted access to clean water and sanitation, and socioeconomic gaps all contribute to this region's low life expectancy.

In conclusion, the GROUP BY ROLLUP operation is crucial for data analysis as it is a hierarchical aggregation from the grand total to the most detailed levels. This allows a better understanding of the data at different levels of detail so that precise insights can be discovered.

3) SLICING

```
SELECT *
FROM world_water_faci
WHERE population_using_improved_water_facilities >50
ORDER BY population_using_improved_water_facilities DESC;
```

	country [PK] character varying	year [PK] integer	population_using_improved_water_facilities integer
1	Lebanon	1995	100
2	Lebanon	1990	100
3	Monaco	1995	100
4	Monaco	2000	100
5	Monaco	2005	100
6	Monaco	2010	100
7	Luxembourg	2012	100
8	Luxembourg	2010	100
9	Luxembourg	2005	100
10	Luxembourg	2000	100
11	Luxembourg	1995	100
12	Luxembourg	1990	100

Figure 4.1.3

	country [PK] character varying	year [PK] integer	population_using_improved_water_facilities integer
998	Kiribati	1995	53
999	Swaziland	2000	52
1000	Ethiopia	2012	52
1001	Guinea	1990	52
1002	Kenya	2000	52
1003	Guinea-Bissau	2000	52
1004	Niger	2012	52
1005	Equatorial Guinea	1995	51
1006	Chad	2012	51
1007	Togo	1995	51
1008	Niger	2010	51
1009	Cameroon	1990	51
1010	Zambia	1995	51
1011	Equatorial Guinea	2005	51
1012	Equatorial Guinea	2000	51
1013	Burkina Faso	1995	51

Total rows: 1013 of 1013 Query complete 00:00:00.077

Figure 4.1.4

Figure 4.1.3 shows the slicing operations that retrieve from the table “world_water_faci”. The first condition is that the percentage of the population using improved water facilities must be greater than 90. It then will execute the results in descending order based on the percentage population using improved water facilities. The results show the table with three columns which is country, year and the population using improved water facilities in percentage. As we can see here, Lebanon had 100% of its population using improved water facilities in 1995 and 1990. Monaco also reached 100% in the years 2000, 2005, and 1990. Luxembourg achieved 100% in several years from 2012, 2010, 2005, 2000, 1995, and 1990. Followed by other countries. It shows that these countries have provided access to safe water and improved water sources for all its citizens.

The figure shows that the lowest percentage of population using improved water facilities around the world can reach as low as 60%. Countries with below 60% of the population using improved water facilities are likely facing major challenges in providing clean and safe water to all their citizens.

In conclusion, the data highlights both successes and challenges in providing access to safe water around the world. Countries with high coverage percentage demonstrate progress and commitment to water infrastructures, while those with lower coverage percentages underscore the ongoing need for the targeted interventions and investments to improve water access and quality for all citizens in their country.

4) DICING

```
SELECT *
FROM world_rd_expenditure
WHERE country like 'K%'
AND year between 2010 AND 2018;
```

	country [PK] character varying	year [PK] integer	gross_expenditure_rd numeric
1	Kazakhstan	2013	691386.13
2	Kazakhstan	2012	620829.77
3	Kazakhstan	2011	540741.2
4	Kazakhstan	2010	480856.86
5	Kenya	2010	788176.89
6	Kuwait	2013	830461.54
7	Kuwait	2012	260791.86
8	Kuwait	2011	244664.79
9	Kuwait	2010	223992.11
10	Kyrgyzstan	2014	24507.49
11	Kyrgyzstan	2013	27500.24
12	Kyrgyzstan	2012	27208.44
13	Kyrgyzstan	2011	25179.19
14	Kyrgyzstan	2010	23153.73

Total rows: 14 of 14 Query complete 00:00:00.173

Figure 4.1.5

Figure 4.1.4 shows the dicing operation on the tables “world_rd_expenditure”. Output displays the gross domestic expenditure on research and development by different countries and year. It filtered data with countries starting with initial “K” and year between 2010 and 2018. The result shows 3 columns which are country, year and gross expenditure. In Kazakhstan from 2010 to 2013, there was a significant increase for gross expenditure. The significantly increasing gross expenditure means economic growth in Kazakhstan. For example, Kazakhstan may have introduced principles and increased funds to encourage gross expenditure efforts across multiple

sectors. This consists of investments in research infrastructure, funding for research programs, and subsidies to encourage private sector participation in research and development.

This trend also applies with other countries such as Kenya, Kuwait and Kyrgyzstan which yearly increase gross expenditure. It seems that Kazakhstan grew rapidly compared to other countries followed by Kenya, Kuwait and Kyrgyzstan. This may be because of the different economic conditions for these countries. For example, Kazakhstan with actively developing technology is able to invest in further research and development that are more advanced than other countries. In addition, Kazakhstan with its strategic location between Europe and Asia make it an advantage. In conclusion, dicing operations help to extract specific subset of data for targeted analyzing.

4.2 Data Visualization

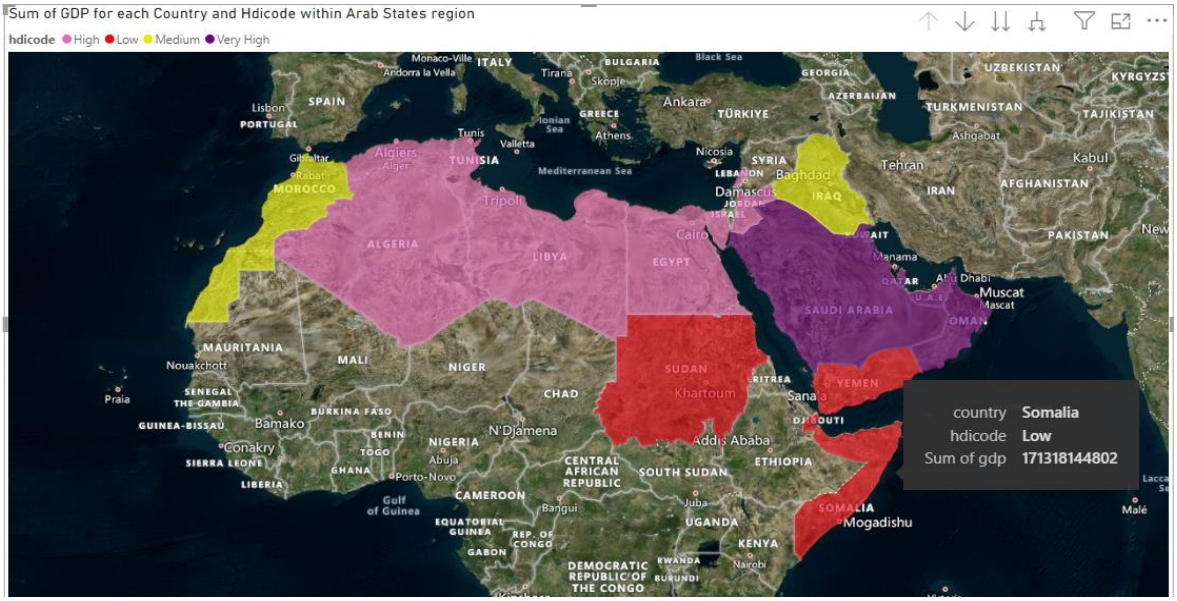


Figure 4.2.1 Sum of GDP for Each Country and HDI Category Within Arab States Region

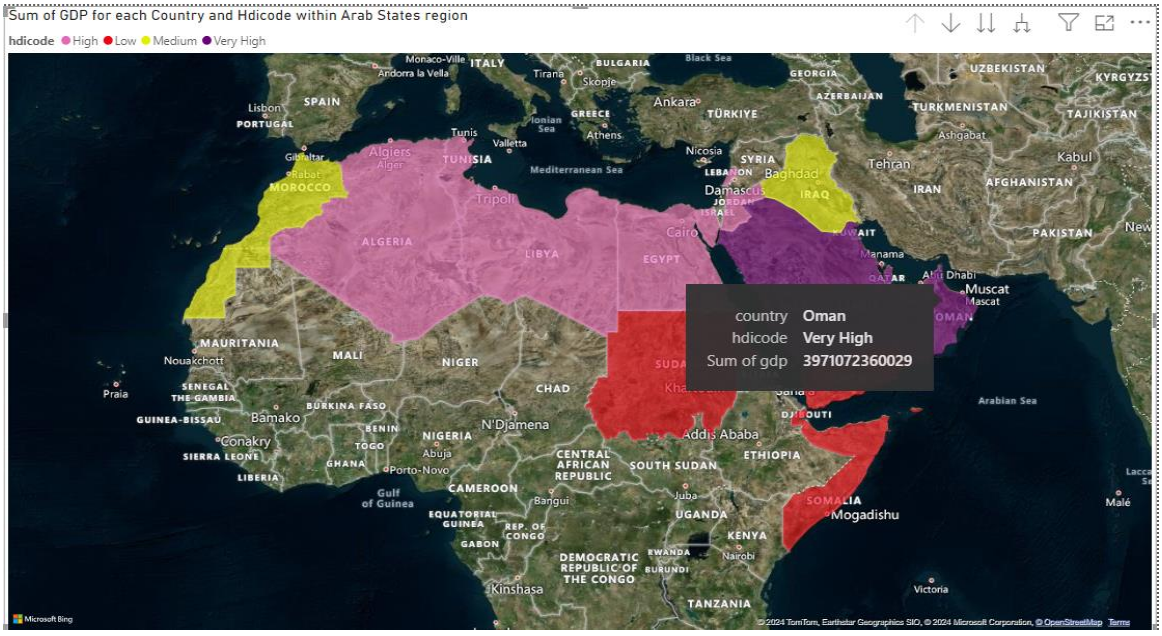


Figure 4.2.2 Sum of GDP for Each Country and HDI Category Within Arab States Region

Figure 4.2.1 and figure 4.2.2 show a graph filled map for the sum of gross domestic product (GDP) of each country and Human Development Index (HDI) within the Arab States region. Each different colour in the map represents a certain HDI category. Countries in the “Very High” HDI category are shaded purple, those in the “High” HDI category are shaded pink, those with a “Medium” HDI category are shaded yellow, and those with a “Low” HDI category are shaded red.

In Somalia with low HDI, we found that the sum of GDP is 171318144802 while for Oman has high HDI with the sum of GDP is 3971072360029. This indicates that Somalia has low economic growth compared to Oman. Somalia may find it difficult to stabilize their socioeconomic conditions in terms of generating income and economic possibilities for the people. Other than that, Somalia may face limited infrastructure for human development, such as access to education, healthcare, and basic needs. On the other hand, Oman has strong economic development compared to Somalia. This indicates that the citizens of Oman have a higher living quality with proper infrastructure for human development. For example, Oman has a quality healthcare system that is easy to access for their citizens. The citizens can get treatment when they are sick without facing difficulties going to the hospital. In conclusion, by visualizing the sum of GDP for different countries and HDI categories using a map graph, the economic and human development conditions within the Arab States region can be thoroughly comprehended.

Figure 4.2.3 Average Poverty Rate by Country

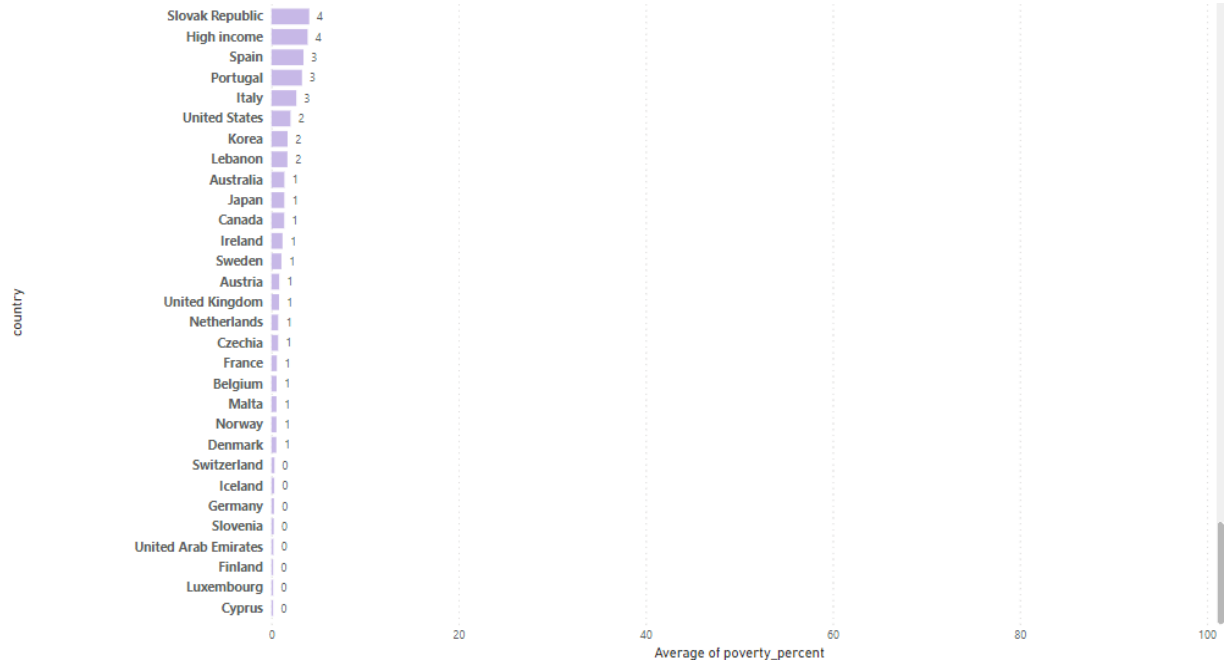


Figure: 4.2.4 Average Poverty Rate by Country

From Figure 4.2.3, show the horizontal bar chart with the x-axis representing the average of poverty in percentage and the y-axis representing the country. As we can see here, the top 3 highest average of poverty in percentage is Democratic Republic of Congo, Uzbekistan and Burundi. Democratic Republic of Congo has the highest average of poverty in percentage which is 99%. Political corruption limited economic opportunities and ongoing conflict could be potential causes of the persistent poverty in the country. Uzbekistan has the second highest average of poverty in percentage which is 98%. This is because Uzbekistan's economy is heavily dependent on natural resources and agriculture, which leaves it vulnerable to fluctuations in commodity prices. Burundi has the same average of poverty in percentage as Uzbekistan. However, Burundi's struggles with poverty are deeply rooted in a complex history, caused by political instability, a rapidly increasing population and food insecurity, leaving its citizens in need of long-term solutions. Poverty can lead to poor access to healthcare, clean water and poor living conditions which is happening in this country.

From the figure, it is observed that the lowest average of poverty in percentage is 0% which is in Switzerland, Iceland, Germany, Slovenia, United Arab Emirates, Finland, Luxembourg and Cyprus. This country has the lowest average poverty percentage because it has a decent economy, comprehensive social welfare systems, investment in education and healthcare. Furthermore, equitable governance structures, and a commitment to sustainable development and social inclusion can help reduce poverty.

In conclusion, addressing poverty requires unified efforts that consider economic development, social welfare, education, healthcare and sustainable practices within the country. Learning from the successes of countries with low poverty rates can provide valuable insight and strategies to ensure that the people can always be free from poverty.

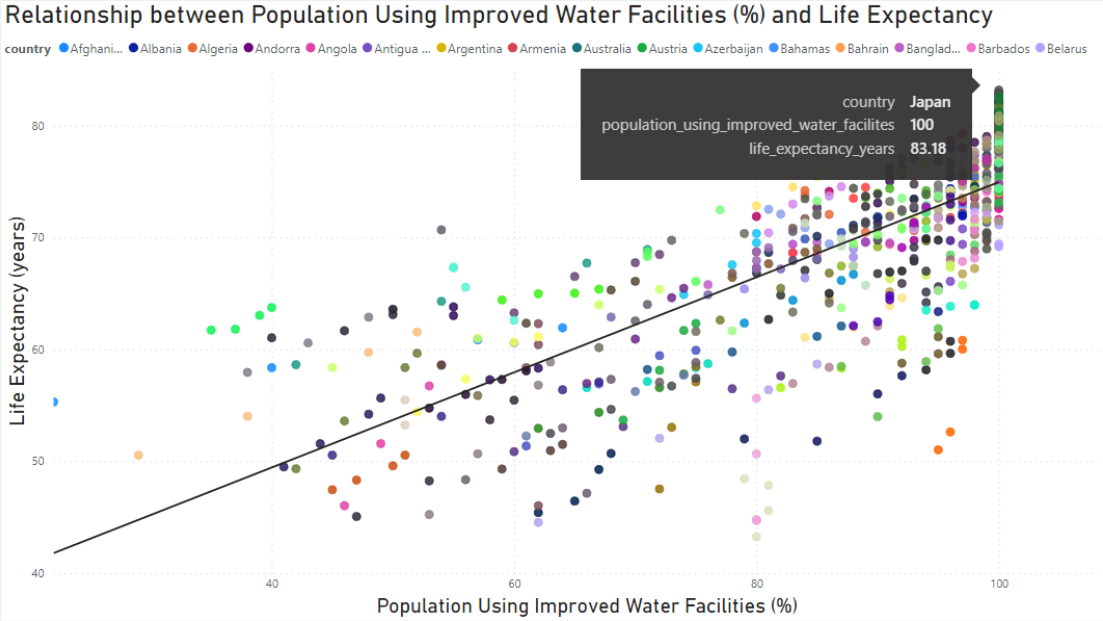


Figure 4.2.5 Relationship Between Population Using Improved Water Facilities and Life Expectancy

The figure above displays the relationship between the percentage of population using improved water facilities and life expectancy. Each point in the data represents a country. It can be observed that the trend line reflects a positive correlation between the percentage of the population using improved water facilities and their life expectancy. A deduction can be made that as the percentage of population using improved water facilities increases, the life expectancy of the population also increases. Japan is observed to have the highest life expectancy of 83.18 years with 100% of its population having access to improved and clean water facilities.

However, the inequality in living standards is obvious. The Ethiopian population appears to have a life expectancy of about 50 years because a majority of the people do not have access to clean water. The population using improved water facilities in Ethiopia is only 29%. The low quality of life in Ethiopia is alarming, as its lack of basic infrastructure has potentially led to the declining health of its people. Various economical and historical issues have burdened Ethiopia causing its underdeveloped civilization. This stark inequality of living standards between nations is certainly alarming.

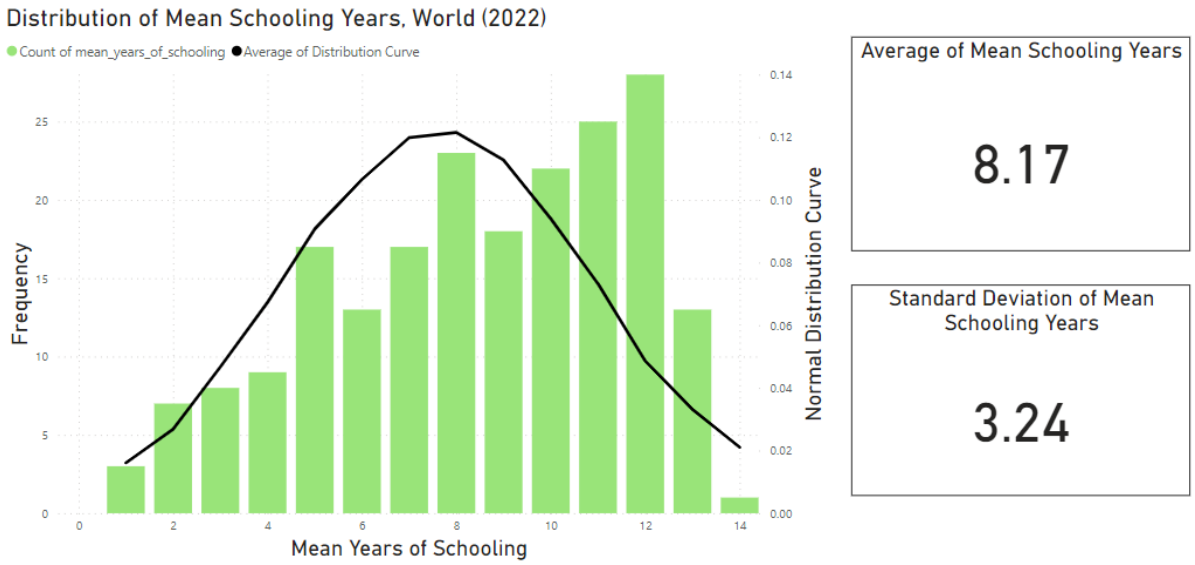


Figure 4.2.6 Distribution of Mean Schooling Years, World 2022

The histogram displays the distribution of the global mean schooling years, accompanied by info cards of the average and standard deviation of mean schooling years in 2022. From the graph, it can be observed that the distribution is slightly skewed. The distribution curve line represents the normal distribution of the mean schooling years and is used to compare with the actual distribution of the histogram. Using the info cards, the average of global mean schooling years is calculated at 8.17. This indicates that education has not been prioritized for many unfortunate children at a global scale. At the same time, we can assume that a country with mean schooling years of less than 11 years indicates that a majority of the population did not receive tertiary education. Tertiary education lengthens the schooling year of an individual but is crucial to improve the overall intelligence and civilization of the country. The count of mean schooling years with the highest frequency is 12 years. It can be deduced that most first-world countries who focus on nurturing their young talents will fall in this category, thus contributing to its high frequency.

The graph brings several implications regarding the inequality of education opportunities across all countries. Even though more countries have a mean schooling year higher than the overall average of 8.20, there are still countries that fall behind in ensuring their population is well-educated. It can be observed from the histogram that there are still about 20 countries that have their population receive a mean schooling year of less than or equal to 4 years. This inequality in education will bring serious issues in the development of the nations. A population with low education levels limits the ability of the nation to develop in terms of economy and technology. Thus, this inequality in education must be addressed.

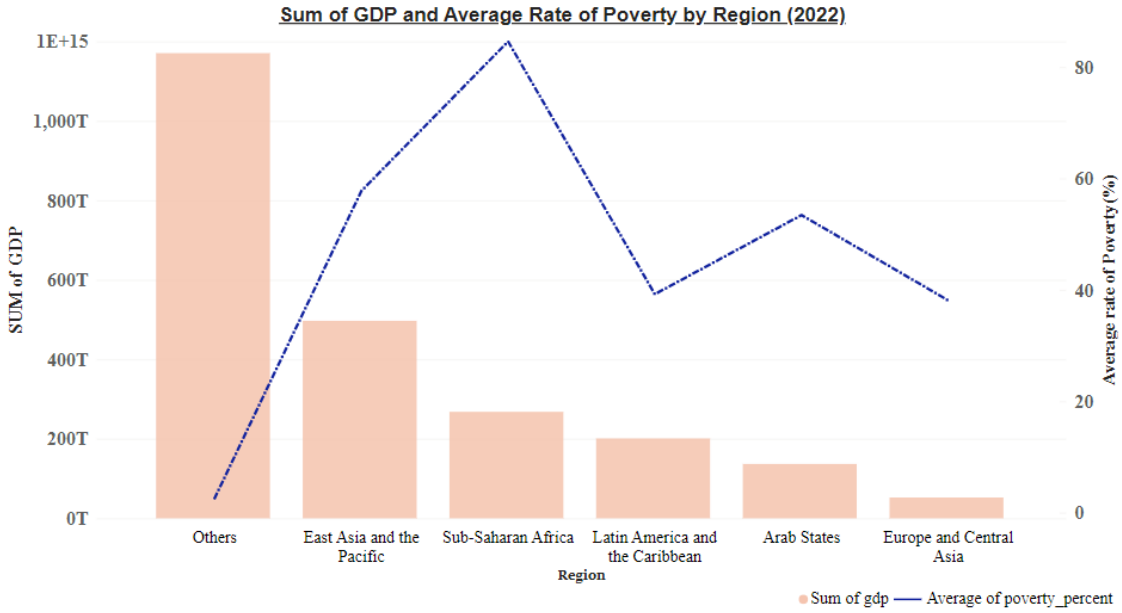


Figure 4.2.7 Sum of GDP and Average Rate of Poverty by Region, 2022

From the combined graph above, the sum of GDP and the average poverty rate across different regions in 2022 has been displayed. The x-axis represents the different regions and the y-axis displays two different scales where the left scale represents the sum of GDP in trillions while the right scale shows the average poverty rate percentage. The bars represent the sum of GDP and the line graph depicts the average poverty rate. From the bar chart, the "Others" region has the highest GDP sum at around $1,200T, followed by East Asia and the Pacific at around 500T.

On the other hand, the region which recorded the lowest average poverty rate is from "Others" region also at around 2%. Besides, Europe and Central Asia regions have the second lowest average poverty rate at around 38%. The region with the highest average poverty rate is Sub-Saharan Africa peaking at around 84%. The combined graph highlights the significant economic disparities between regions. Some regions have higher GDP sums than others. However, poverty rates vary considerably. Some regions have a high GDP but also a high average poverty rate, while others have a low GDP and a low average poverty rate. These circumstances indicate differences in socioeconomic development and living standards across regions. In short, the combined graph shows a significant economic disparities and variations in poverty rates across different regions in 2022.

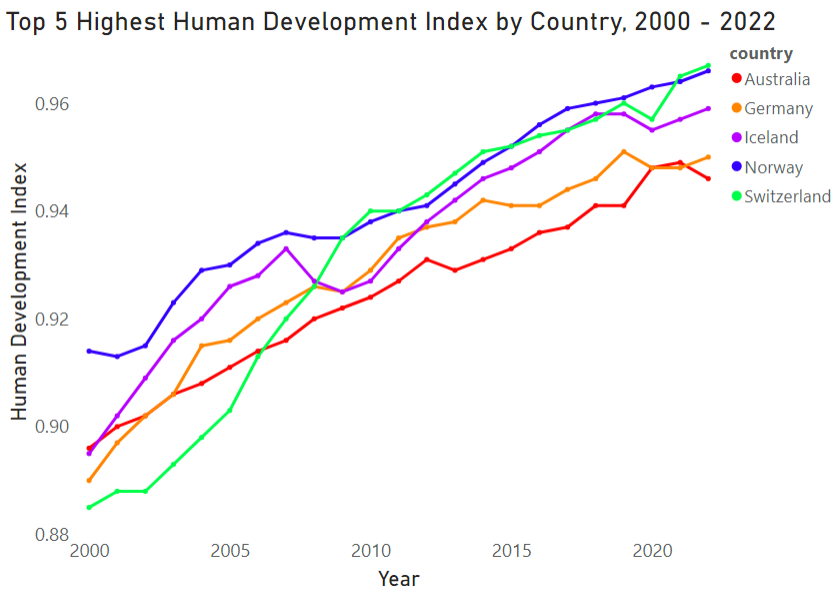


Figure 4.2.8 Top 5 Highest Human Development Index by Country, 2000 – 2022

Figure 4.2.8 is a line graph of the top 5 countries that have the highest human development index (HDI) from 2000 to 2022. For context, HDI is a metric set by the UN to measure the average achievement in key dimensions of human development such as health, education and living standard. A high HDI indicates that the nation is well-developed and favorable for living. As observed, the top 5 countries are first-world countries with a majority of them located in Europe. Moreover, it shows that their HDI has consistently increased in the last two decades, with Switzerland having a spike in its HDI from 2005 to 2010.

There is a lot to be learned from this simple line graph. The many European countries with higher than average HDI should be studied thoroughly to become a guideline for other countries struggling with inequality. Europe has focused on providing strong welfare systems such as healthcare, education and pensions to its people. Along with a strong economic presence, they are politically stable and have low levels of corruption. Their advanced infrastructure and sustainable development have become a core part of their culture. Therefore, many countries should learn and focus on these key traits to increase their HDI to improve the overall quality of life for their population.

5.0 CONCLUSION

In summary, the study has successfully conducted a study which aligns with the Sustainable Development Goal 10 of addressing various inequalities all around the world. A thorough investigation has been conducted on various aspects of socioeconomic inequalities that have burdened many nations and individuals. Moreover, a clear demonstration using data warehouse management skills combined with online analytical processing (OLAP) techniques has contributed to the extraction of valuable insights from various datasets. The study has provided reasonable deductions regarding the cause of these inequalities and their further consequences if these issues remain unresolved. The project concludes that there exist inequalities in terms of financial stability of nations, human development standards, poverty rates and access to basic human needs. It was also found that individuals from all around the world are still suffering from these issues in low HDI nations.

The study has implemented the concept of data warehouse to store raw data, acting as a single source of truth for the entire operation. The data is ensured to be trustworthy and reliable as it is gathered from the prominent United Nations. The following process involves a complex operation of constructing a reusable and efficient ETL pipeline to carry out data transformation and load into a database for visualization purposes. By utilizing advanced OLAP operations and visualizations, valuable insights can be discovered from the data. As actionable insights are extracted from the data, the study offers helpful deductions and analysis that can aid in the efforts of resolving these issues of inequalities. The act of addressing inequalities remains one of the most urgent challenges in the modern world as it constantly impacts the overall quality of life and opportunities for many individuals. Therefore, governments must collaborate to address this urgent matter and achieve the SDG 10 within the specified timeframe.

6.0 REFERENCES

30 years on, South Africa still dismantling racism and apartheid's. (2024, April 19). Africa Renewal. https://www.un.org/africarenewal/magazine/april-2024/30-years-south-africa-still-dismantling-racism-and-apartheid%E2%80%99s-legacy

Half of Afghan children out of school, due to conflict, poverty, discrimination: UNICEF. (2018, June 5). UN News. https://news.un.org/en/story/2018/06/1011211

Lifestyle diseases pose new burden for Africa. (2017, February 23). Africa Renewal. https://www.un.org/africarenewal/magazine/december-2016-march-2017/lifestyle-diseases-pose-new-burden-africa

Maqbool, A., & Bashir, M. K. (2009). Rural development in Pakistan: issues and future strategies. *ResearchGate*. https://www.researchgate.net/publication/216413804_Rural_Development_in_Pakistan_Issues_and_Future_Strategies

Munir, F., Ahmad, S., Ullah, S., & Wang, Y. P. (2021). Understanding housing inequalities in urban Pakistan: An intersectionality perspective of ethnicity, income and education. *Journal of Race, Ethnicity and the City*, *3*(1), 1–22. https://doi.org/10.1080/26884674.2021.1986442

Sustainable Development Goal 10: Reduced Inequalities | The United Nations in China. (n.d.). Sustainable Development Goal 10: Reduced Inequalities | the United Nations in China. https://china.un.org/en/sdgs/10

7.0 DATASETS

Latest Human Development composite indices tables:
https://hdr.undp.org/data-center/documentation-and-downloads

GDP, PPP (current international $) dataset:
https://data.un.org/Data.aspx?d=WDI&f=Indicator_Code%3aNY.GDP.MKTP.PP.CD

Population using improved drinking-water sources (%) dataset:
https://data.un.org/Data.aspx?d=WHO&f=MEASURE_CODE%3aWHS5_122

Gross domestic expenditure on research and development (GERD) in '000 current PPP$ dataset:
https://data.un.org/Data.aspx?d=UNESCO&f=series%3aST_SCGERDPPP

Poverty headcount ratio at $6.85 a day (2017 PPP) (% of population):
https://data.un.org/Data.aspx?d=WDI&f=Indicator_Code%3aSI.POV.UMIC