I. INTRODUCTION

For my Data Engineering Capstone project, I have selected the issue relating to the promotion of decent work and economic growth. Why did I choose this problem or objective? Decent work and economic growth are essential components of a sustainable development, propelling the Philippines toward becoming a livable and advanced nation. The pursuit for decent work and economic growth aims to make the Philippines, a country where every Filipinos can find good jobs, fair working conditions, and access to opportunities for career development. By addressing these challenges can help significantly lower poverty rates and reduce the necessity for Filipino to look for work abroad. It also intends to create a vibrant economy that can support a good life for all Filipinos. These two things, decent work and economic growth, are closely connected and really important for the welfare of the Filipino society and the country.

II. PROBLEM STATEMENT

1. What is the problem that you are trying to solve?

The main issue we're looking at is that in many parts of the Philippines, there aren't enough decent work opportunities or long-lasting economic growth. Lots of people can't find work, end up with jobs that don't pay enough, or are treated unfairly at work. Also, in some places, the money-making isn't done in a way that can keep going without damaging the environment and making the gap between rich and poor bigger. This problem makes life hard for every Filipino and stops us from reaching our national ambition to become an advanced country stated in Ambisyon Natin 2040 where no Filipino is left behind.

2. What are the objectives you want to achieve?

   a. To promote access to decent work for all Filipinos, regardless of their background, skills, or location.
   b. To encourage lasting economic growth that warranted a high standard of living for both present and future generations.
   c. To mitigate unemployment rate to help reduce income inequality, alleviate poverty, and enhance social inclusion through initiatives that prioritize decent work and economic growth.
   d. To support the global community in achieving Sustainable Development Goal by the United Nation specifically the Goal 8, which stated "*Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all*."

In this project, we will limit only for 2 reports about the economic and unemployment issues.

By addressing the challenges of decent work and economic growth, we aim to create a more equitable, sustainable, and prosperous future for all.

III. METHODOLOGY

1. Data Collection – Gather quantitative and qualitative data from diverse sources such as PSA, including national statistical agencies, surveys, and reports.

2. Database Modeling – Enabling organizations and policymakers to make informed decisions, improve labor conditions, and enhance overall productivity.

3. Data Analysis – Analyze data related to employment rates, wage levels, job quality indicators, GDP growth, industry sectors, and other relevant economic factors. Employ statistical tools to identify correlations and trends between decent work indicators and economic growth.

4. Data Visualization – Present the findings such as a dashboard through data visualization techniques such as graphs, charts, and maps to make the research results more accessible and comprehensible to a broader audience.

This multifaceted methodology aims to provide a comprehensive understanding of the relationship between decent work and economic growth, offering insights and practical recommendations.

IV. RESULT AND SOLUTION

We will collect data from psa.gov.ph to conduct an analysis and develop solutions for the challenges related to decent work and economic growth. The first step involves designing a relational database and creating a visualization, such as a dashboard, to gain insights into the issues. To initiate this process, we must create an Entity-Relationship Diagram (ERD). The initial step in this ERD creation is identifying the entities involved.

A. Selected Entities and Their Respective Attributes:

1. *Goal* – A broad desired outcome or result that our project aims to achieve. It represents the ultimate mission of our project.
   Attributes: *Goal_Num, Goal_Title, Goal_Desc*
2. *Target* – A specific, measurable level that is set to achieve within a specified time frame to reach a goal. Targets are more specific and quantifiable than goals.
   Attributes: *Target_Num, Target_Desc*
3. *Economics* – Economic variable or measure that provides evidence of progress toward achieving our target or goal. Indicators are used to assess economic performance or outcomes.
   Attributes: *Economics_ID, Economics_Num, Economics_Date, Economics_Desc, Economics_value*
4. *Unemployment* – a metric that measures the percentage of the workforce that is currently unemployed and actively seeking employment. Indicators are used to assess the extent of

joblessness within the given context, providing insights into the health of the labor market. Attributes: *Unemployment_ID, Unemployment_Num, Unemployment_Date, Unemployment_Desc, Unemployment_value*

B. ERD Process and Design

1. Assumptions:

a. Each Goal has a unique Goal_ID
b. Each Target has a unique Target_ID
c. Each Indicator has a unique Inidicator_ID

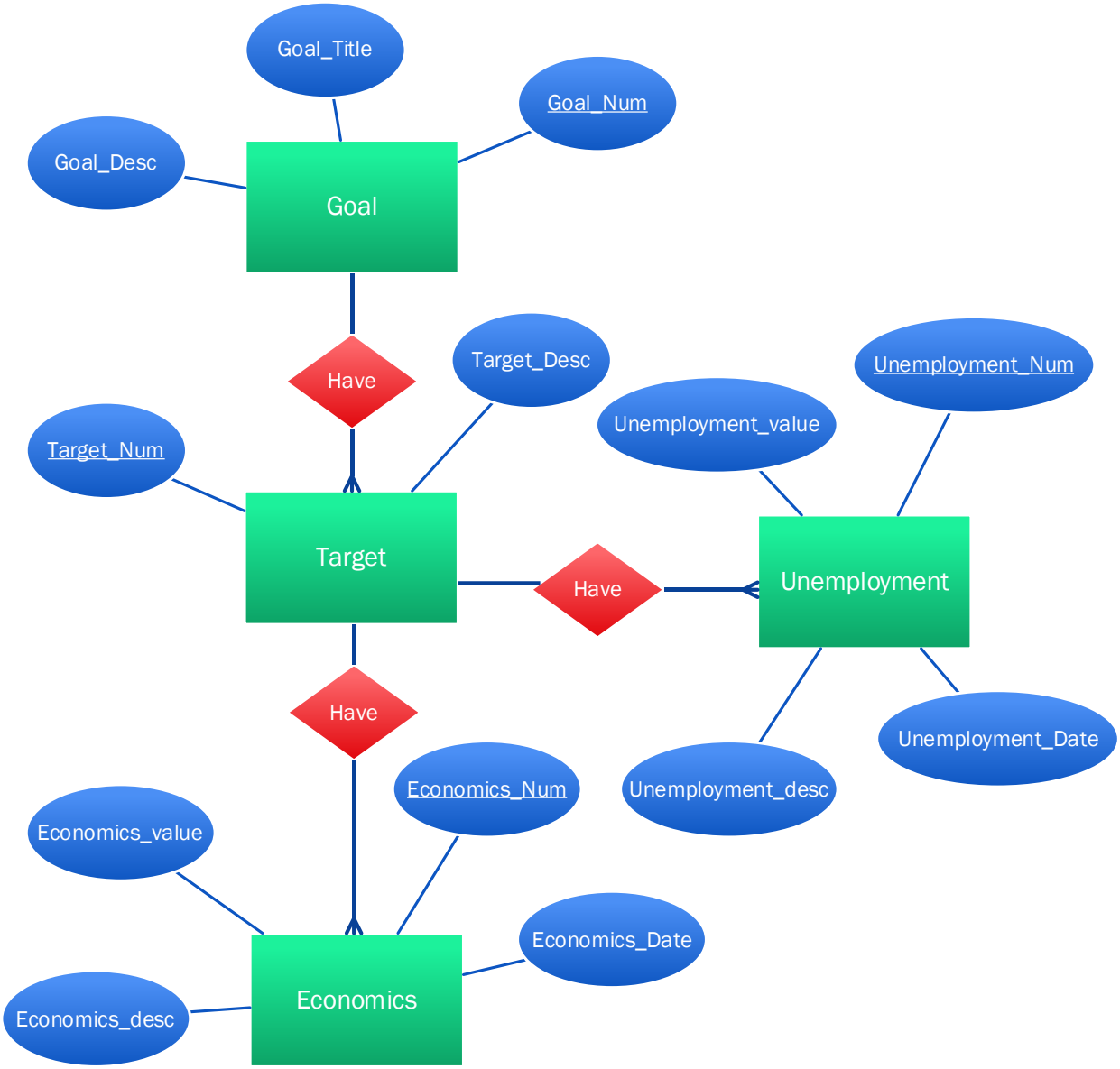
2. Cardinality:

a. Goal(One) – Target(Many): One Goal can have multiple Targets
b. Target(One) – Indicator(Many): One Target can have multiple Indicators. In our project, we have Economics and Unemployment indicators

3. Attributes Objectives:

a. Goal_Num – unique identification for each Goal
b. Goal_Title – textual representation of the title associated with the specific Goal
c. Goal_Desc – detailed description or explanation of the purpose of each Goal
d. Target_Num – unique identification for each Goal
e. Target_Desc – detailed description or explanation of the purpose of each Target
f. Economics_ID – unique identification for each Economics Indicator
g. Economics_Num – numerical identifier assigned to each Economics Indicator
h. Economics _Date – represents the date associated with the particular Economics Indicator
i. Economics _Desc – detailed description or explanation of the purpose of each Economics Indicator
j. Economics _value – represents the value associated with the particular Economics Indicator
k. Unemployment_ID – unique identification for each Unemployment Indicator
l. Unemployment_Num – numerical identifier assigned to each Unemployment Indicator
m. Unemployment_Date – represents the date associated with the particular Unemployment Indicator
n. Unemployment_Desc – detailed description or explanation of the purpose of each Unemployment Indicator
o. Unemployment_value – represents the value associated with the particular Unemployment Indicator

Figure 1. ENTITIY RELATIONSHIP DIAGRAM



Having identified the entities and their attributes along with their respective functions and relationships, we've successfully crafted the Entity Relationship Diagram, illustrated in Figure 1.

The next step is designing our normalized model, drawing guidance from the created ERD.

C. Normalized Model Process and Design

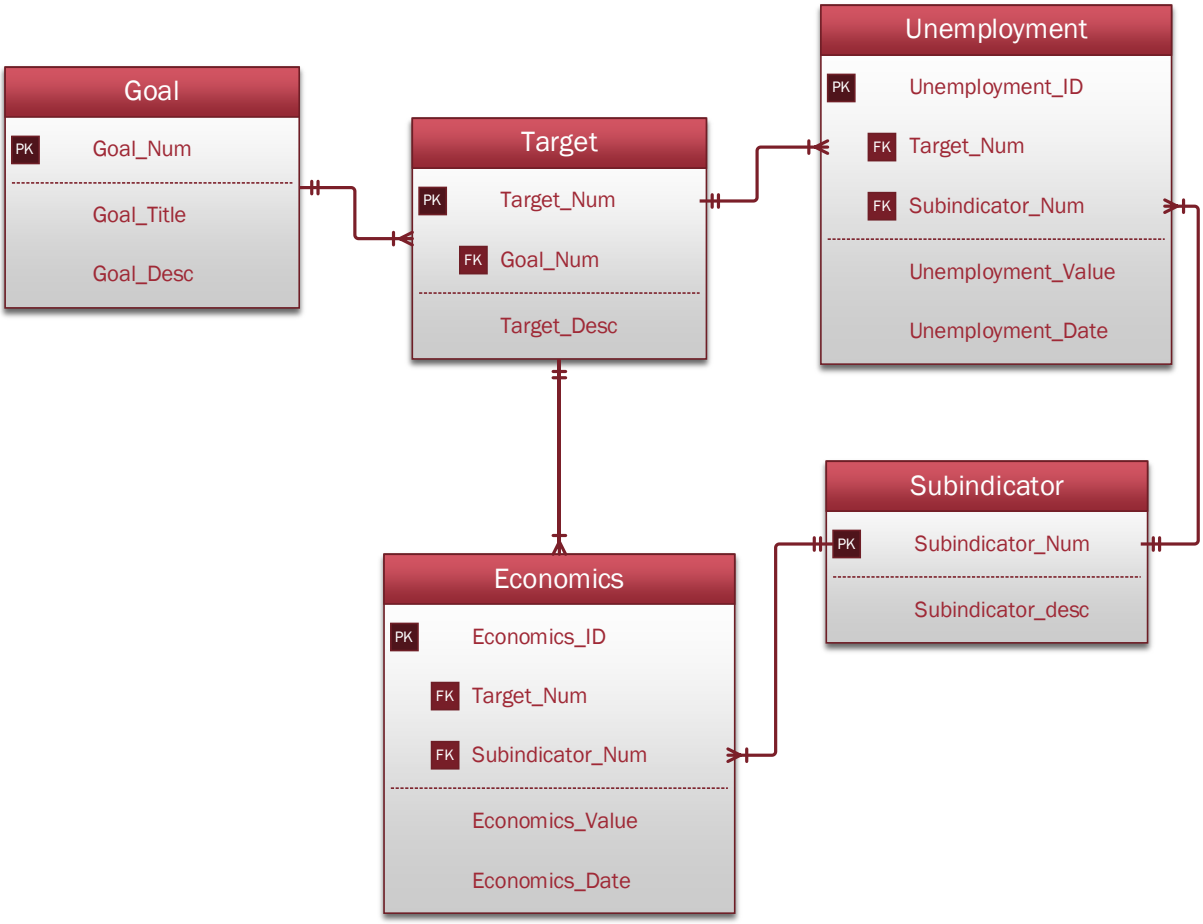
First Normal Form (1NF): Each column of all the tables contain atomic values and that each row is uniquely identifiable and is in First Normal Form or 1NF.

Second Normal Form (2NF): Remove partial dependencies. A table is in 2NF if it is in 1NF and all non-key attributes are fully functionally dependent on the entire primary key. In our design, economics_desc is too dependent on economics_num. So, we separate a table for it which is called subindicator and add a primary key subindicator_num to remove the dependency and to make sure that there is no duplication of rows in our table. We also change the name to subindicator_num and subindicator_desc, respectively. Similarly, apply the same approach to the table regarding unemployment. The tables are now already in 2NF because all attributes in each table are fully functionally dependent on the primary key.

Third Normal Form (3NF): Remove transitive dependencies. A table is in 3NF if it is in 2NF and it does not have transitive dependencies. In our design, it is already in 3NF.

After the normalized process done, the design from relational model haven't change to normalized model as shown in Figure 2.

Figure 2. NORMALIZED MODEL



From here, we can now design our dimensional model

D. Dimensional Model Process and Design

1. The 4-step Dimensional Modeling Process

Step 1: Select the business process

Business processes are the operational activities performed by the organization.

Economics – represented through the *Economic_Fact* table,
Unemployment – represented through the Unemploymet_Fact table.

These 2 business processes will be offering the essential metrics required to shape our insights. These insights guide our decision-making process in addressing the challenges related to decent work and economic growth.

Step 2: Declare the grain

Grain refers to the lowest level at which data is captured by a given business process.

a. Economics by goal by target by subindicator by year
b. Unemployment by goal by target by subindicator by year

Step 3: Identify the Dimensions

Dimensions provide the "who, what, where, when, why, and how" context surrounding a business process event.

a. Goal_Dim [who]
b. Target_Dim [what]
c. Date_Dim [when]
d. Subindicator_Dim[why]

Step 4: Identify the Facts

Facts are the measurements that result from a business process event and are almost always numeric.

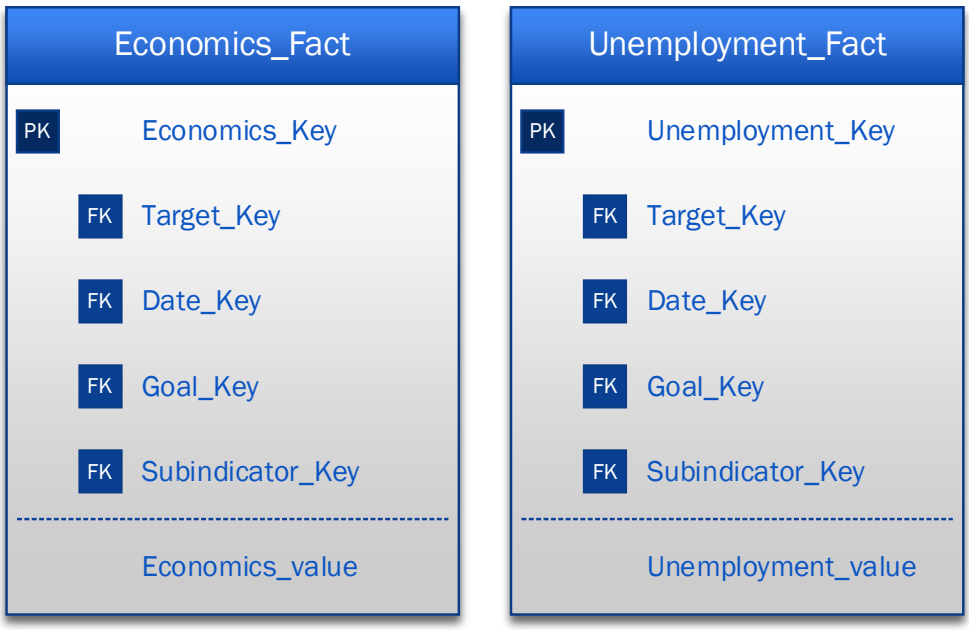
a. Economics_value – will be the fact for monitoring the Economic Indicator specifically the GDP per capita
b. Unemployment_value – will be the fact for monitoring the Unemployment Indicator specifically the Unemployment rate

2. Creation of Fact Table

A fact table contains the numeric measures produced by an operational measurement event in the real world. At the lowest grain, a fact table row corresponds to a measurement event and vice versa. Thus the fundamental design of a fact table is entirely based on a physical activity and is not influenced by the eventual reports that may be produced. In addition to numeric measures, a fact table always contains foreign keys for each of its associated dimensions, as well as optional degenerate dimension keys and date/time stamps. Fact tables are the primary target of computations and dynamic aggregations arising from queries. – Kimball Group

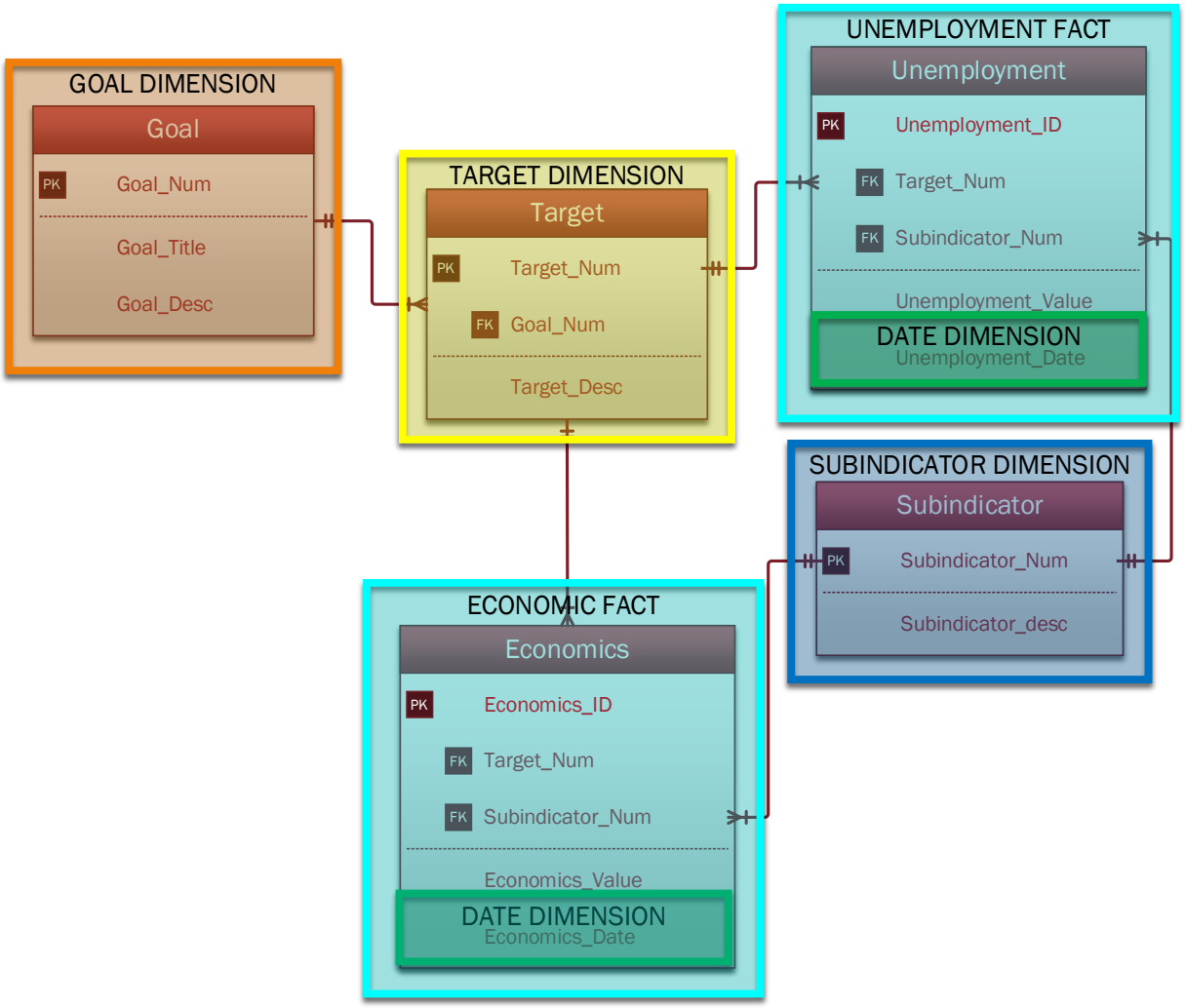
With the information above and the 4-step Dimensional Modeling, then we create the fact table as shown in Figure 3.

Figure 3. Indicator Fact Table



3. Source to Target Map

Figure 4. Source to Target Map



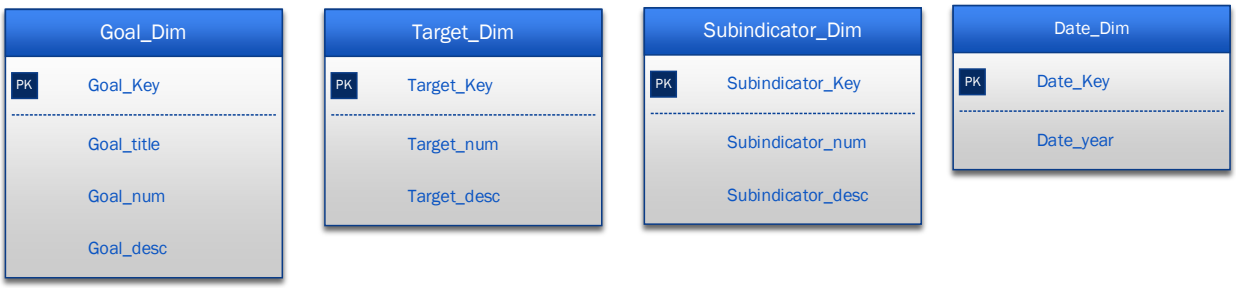
Source-to-target mapping refers to the process of identifying and defining how data is transferred and transformed from a source system to a target system. Figure 4 is the Source-to-target mapping for the Retail Sales Analysis System.

4. Designing Dimension Table Using Dimensional Normal Form

Step 1 (DNF1): Normalized Model (3NF)
We have previously completed the task, as illustrated in Figure 2.

Step 2 (DNF2): Denormalized into the Dimensions
First, we need to denormalized Goal_Dim, Target_Dim, Subindicator_Dim and Date_Dim as shown in Figure 5.

Figure 5. Denormalized Goal_Dim, Target_Dim, and Date_Dim



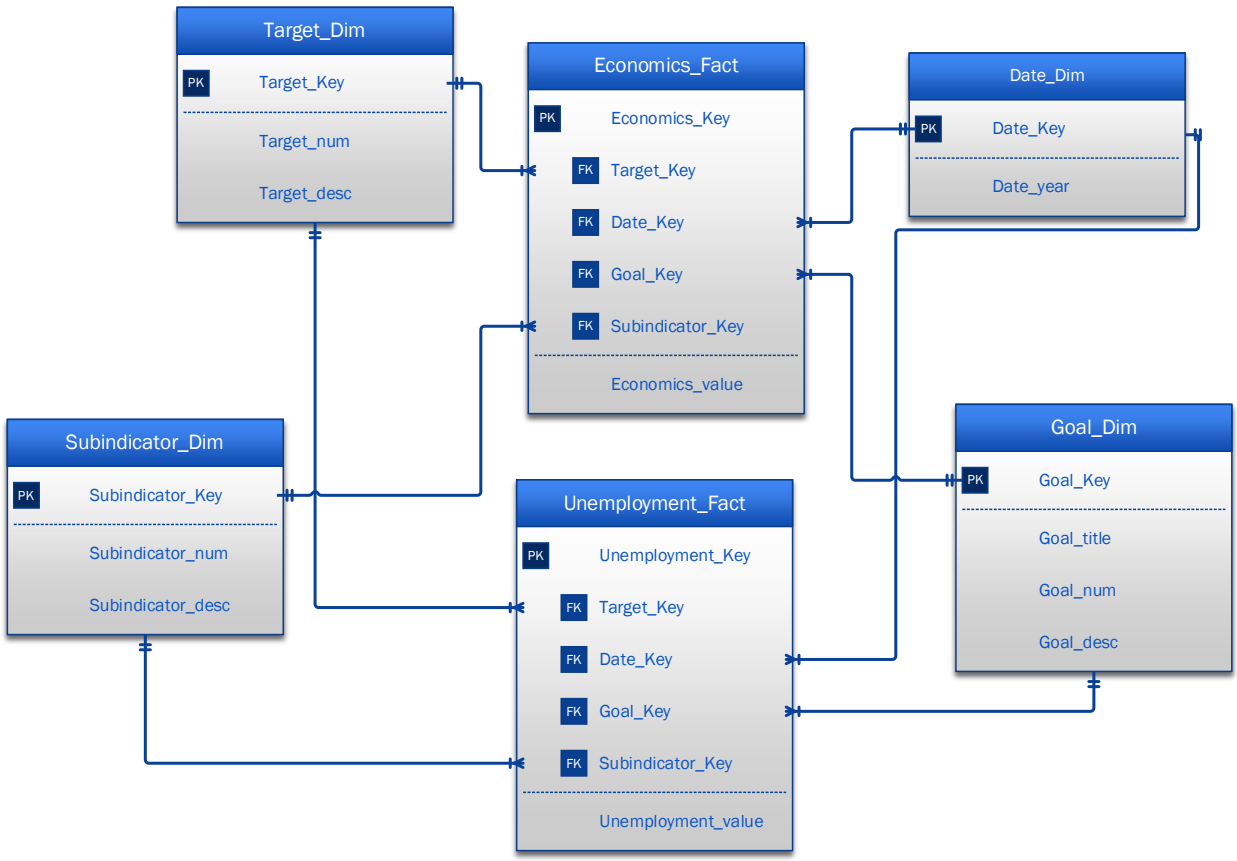
Process Summary of DNF2

- Goal Table → Goal_Dim
- Target Table → Target_Dim
- Subindicator Table → Subindicator_Dim
- Create Date_Dim

Step 3 (DNF3): Attach Dimensions to Fact Table

We have previously completed the task to create the Fact table, as illustrated in Figure 3. Now, we will attach it to our dimensional tale to complete as shown in Figure 6.

Figure 6. DIMENSIONAL MODEL



The ERD, Normalized Model, and Dimensional Model exhibit flexibility to accommodate additional Goals, Targets, and Indicators. This implies that the models are not exclusively tailored to the Decent Work and Economic Growth goal but can be adapted with minimal adjustments to incorporate other Goals as well in Sustainable Development Goals.

E. SQL SCRIPT

Figure 7. SQL Create Normalized Model

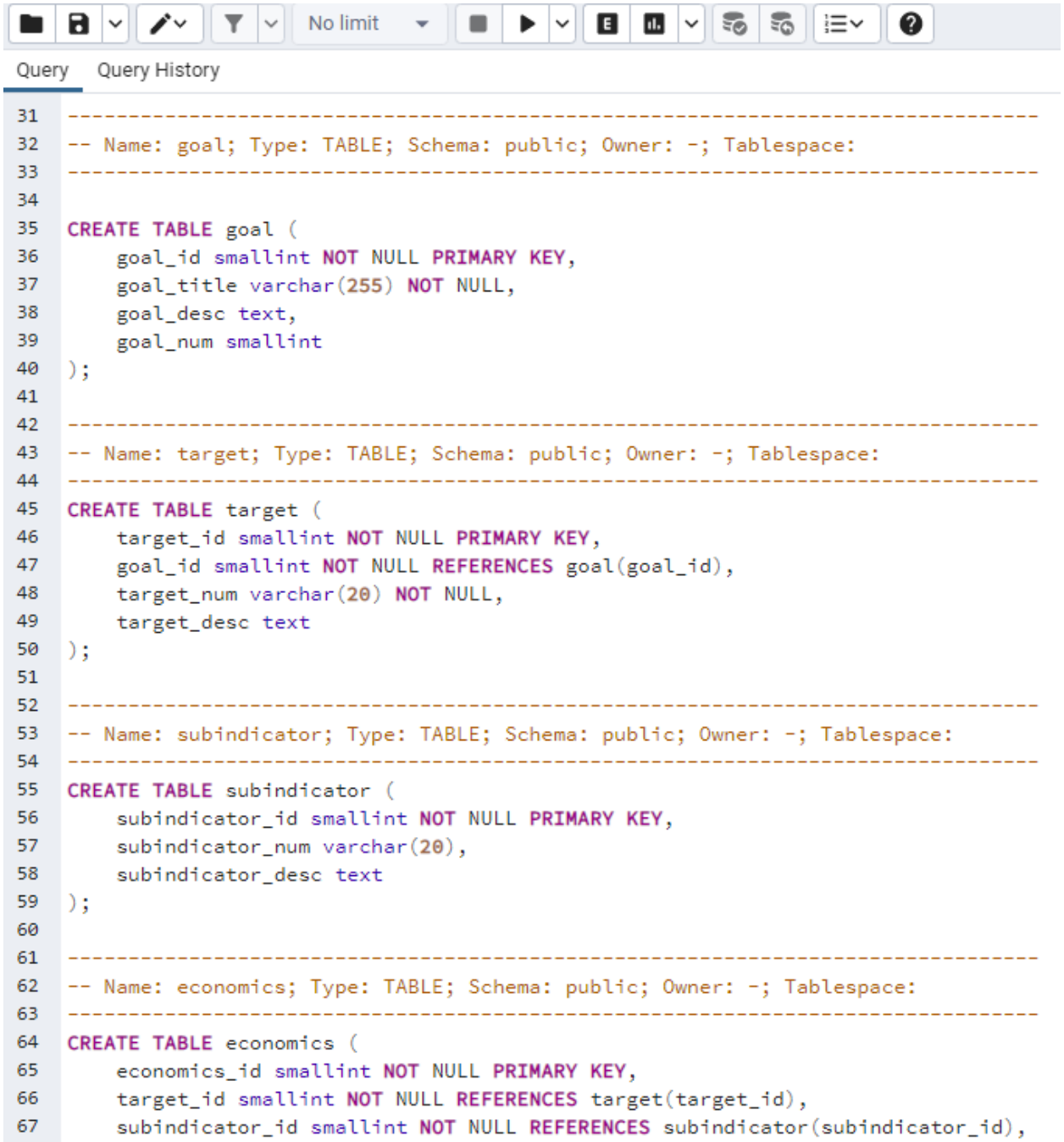
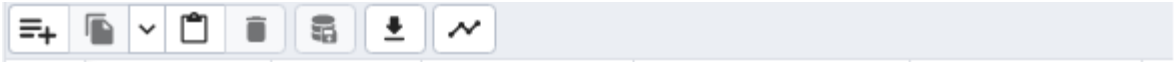


Figure 8. SQL Output for Normalized Model [Economics Table]



	economics_id [PK] smallint	target_id smallint	subindicator_id smallint	economics_date character varying (20)	economics_value numeric
1	1	1	1	2001	1.0
2	2	1	1	2002	1.7
3	3	1	1	2003	3.1
4	4	1	1	2004	4.5
5	5	1	1	2005	3.0
6	6	1	1	2006	3.4
7	7	1	1	2007	4.6
8	8	1	1	2008	2.5
9	9	1	1	2009	-0.4
10	10	1	1	2010	5.5
11	11	1	1	2011	2.1
12	12	1	1	2012	5.1
13	13	1	1	2013	4.9
14	14	1	1	2014	4.6
15	15	1	1	2015	4.6
16	16	1	1	2016	5.4
17	17	1	1	2017	5.2
18	18	1	1	2018	4.7
19	19	1	1	2019	4.6
20	20	1	1	2020	-10.8
21	21	1	1	2021	4.3

Figure 9. SQL Create Dimensional Tables

```
27
28  ------------------------------------------------------------------------------
29  --goal_dim
30  ------------------------------------------------------------------------------
31  CREATE TABLE goal_dim(
32      goal_key int PRIMARY KEY,
33      goal_id int,
34      goal_title varchar(255) NOT NULL,
35      goal_desc text,
36      goal_num smallint
37  );
38
39  ------------------------------------------------------------------------------
40  --target_dim
41  ------------------------------------------------------------------------------
42  CREATE TABLE target_dim(
43      target_key int PRIMARY KEY,
44      target_id int,
45      target_num varchar(20) NOT NULL,
46      target_desc text
47  );
48
49  ------------------------------------------------------------------------------
50  --subindicator_dim
51  ------------------------------------------------------------------------------
52  CREATE TABLE subindicator_dim(
53      subindicator_key int PRIMARY KEY,
54      subindicator_id int,
55      subindicator_num varchar(20),
56      subindicator_desc text
57  );
58
59  ------------------------------------------------------------------------------
60  --date_dim
61  ------------------------------------------------------------------------------
62  CREATE TABLE date_dim (
63      date_key int PRIMARY KEY,
64      date_year varchar(20)
65  );
66
```

Figure 10. SQL Create Fact Table

```
24  ---------------------------------------------------------------------------
25  --economics_fact
26  ---------------------------------------------------------------------------
27  CREATE TABLE economics_fact (
28      economics_key serial PRIMARY KEY,
29      goal_key int REFERENCES goal_dim (goal_key),
30      target_key int REFERENCES target_dim (target_key),
31      subindicator_key int REFERENCES subindicator_dim (subindicator_key),
32      date_key int REFERENCES date_dim (date_key),
33      economics_value numeric
34  );
35
36  ---------------------------------------------------------------------------
37  --unemployment_fact
38  ---------------------------------------------------------------------------
39  CREATE TABLE unemployment_fact (
40      unemployment_key serial PRIMARY KEY,
41      goal_key int REFERENCES goal_dim (goal_key),
42      target_key int REFERENCES target_dim (target_key),
43      subindicator_key int REFERENCES subindicator_dim (subindicator_key),
44      date_key int REFERENCES date_dim (date_key),
45      unemployment_value numeric
46  );
```

Figure 11. SQL Output for Fact Table [Unemployment_Fact Table]

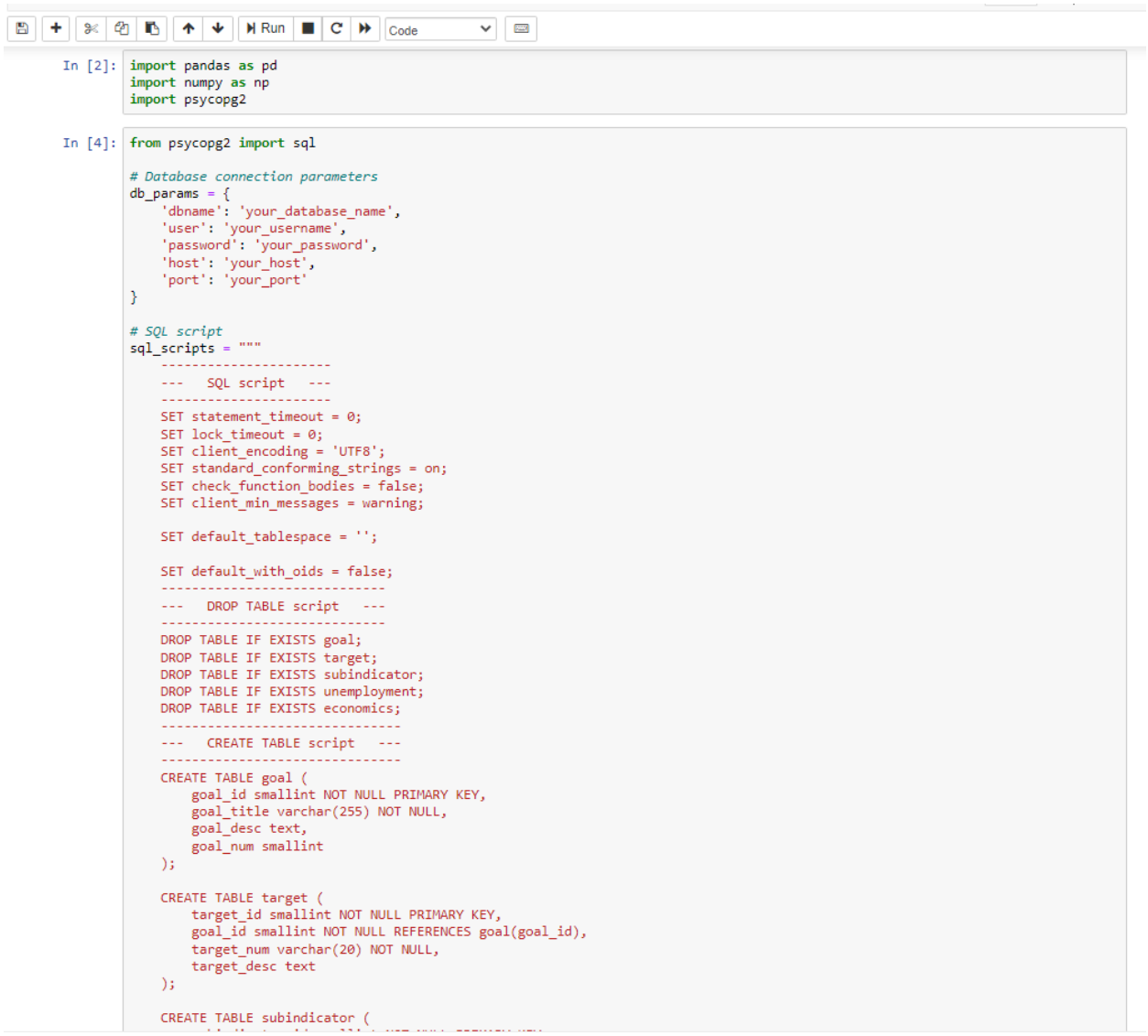
Data Output

	unemployment_key [PK] integer	goal_key integer	target_key integer	subindicator_key integer	date_key integer	unemployment_value numeric
1	1	1	2	2	6	7.8
2	2	1	2	2	7	8.0
3	3	1	2	2	8	7.3
4	4	1	2	2	9	7.4
5	5	1	2	2	10	7.5
6	6	1	2	2	11	7.4
7	7	1	2	2	12	7.0
8	8	1	2	2	13	7.0
9	9	1	2	2	14	7.1
10	10	1	2	2	15	6.6
11	11	1	2	2	16	6.3
12	12	1	2	2	17	5.4
13	13	1	2	2	18	5.7
14	14	1	2	2	19	5.3
15	15	1	2	2	20	5.1
16	16	1	2	2	21	10.3
17	17	1	2	2	22	7.8

Here, we've demonstrated the construction of a database using the SQL language in PostgreSQL, along with its corresponding output.

Please refer to the file inside the zipped file for the full SQL script.

Figure 12. Create Normalized Model using Python-SQL script in Jupyter Notebook



Run Code

In [2]:
```
import pandas as pd
import numpy as np
import psycopg2
```

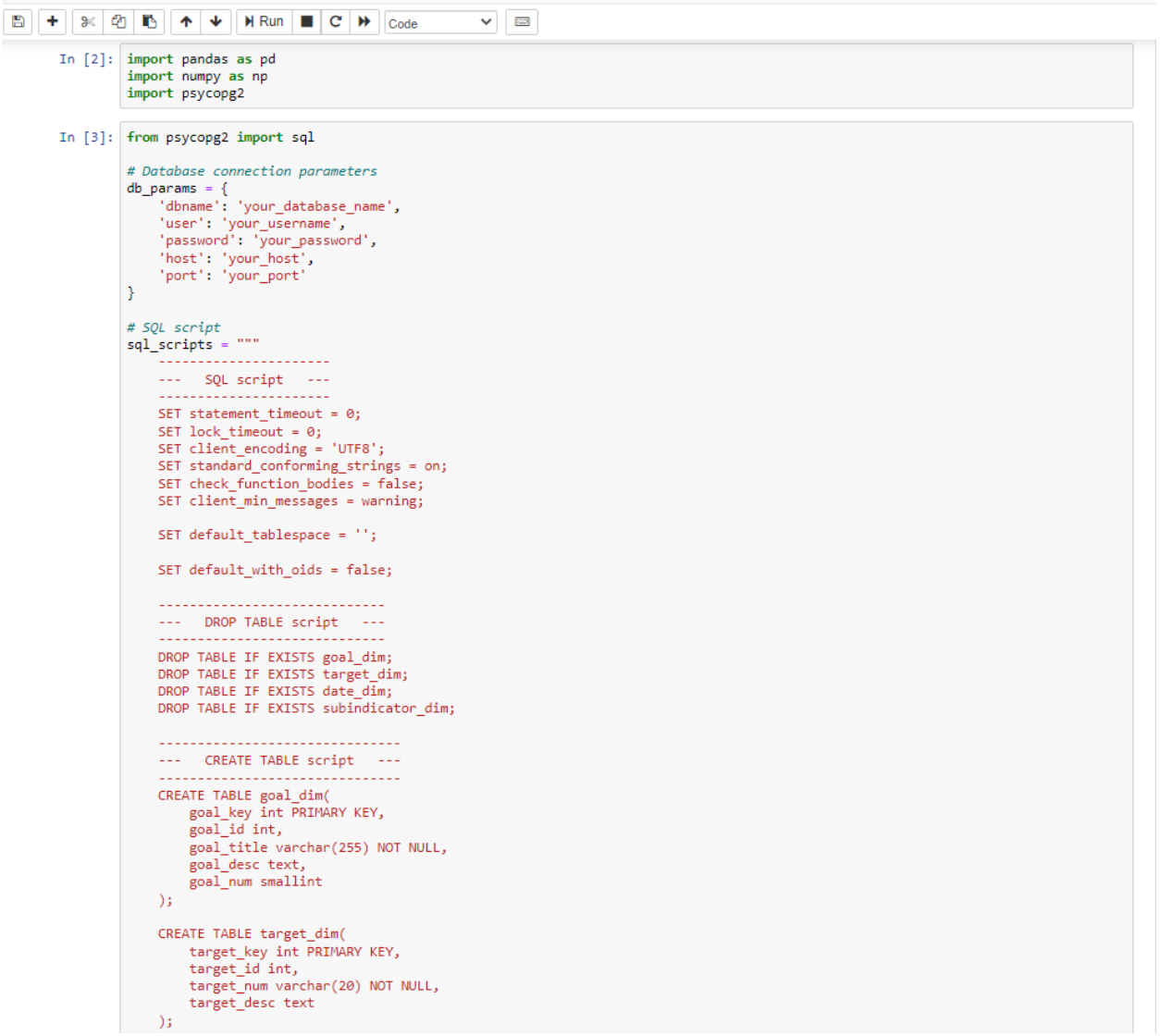
In [4]:
```
from psycopg2 import sql

# Database connection parameters
db_params = {
    'dbname': 'your_database_name',
    'user': 'your_username',
    'password': 'your_password',
    'host': 'your_host',
    'port': 'your_port'
}

# SQL script
sql_scripts = """
    ----------------------
    ---   SQL script   ---
    ----------------------
    SET statement_timeout = 0;
    SET lock_timeout = 0;
    SET client_encoding = 'UTF8';
    SET standard_conforming_strings = on;
    SET check_function_bodies = false;
    SET client_min_messages = warning;

    SET default_tablespace = '';

    SET default_with_oids = false;
    ----------------------------
    ---   DROP TABLE script   ---
    ----------------------------
    DROP TABLE IF EXISTS goal;
    DROP TABLE IF EXISTS target;
    DROP TABLE IF EXISTS subindicator;
    DROP TABLE IF EXISTS unemployment;
    DROP TABLE IF EXISTS economics;
    ------------------------------
    ---   CREATE TABLE script   ---
    ------------------------------
    CREATE TABLE goal (
        goal_id smallint NOT NULL PRIMARY KEY,
        goal_title varchar(255) NOT NULL,
        goal_desc text,
        goal_num smallint
    );

    CREATE TABLE target (
        target_id smallint NOT NULL PRIMARY KEY,
        goal_id smallint NOT NULL REFERENCES goal(goal_id),
        target_num varchar(20) NOT NULL,
        target_desc text
    );

    CREATE TABLE subindicator (
```

Figure 13. Normalized Model Output using Python-SQL script in Jupyter Notebook

```
# Display the DataFrame from the Database
print(goal.head())
print(target.head())
print(subindicator.head())
print(unemployment.head())
print(economics.head())
```

```
   goal_id                     goal_title  \
0        1  Decent Work and Economic Growth

                                           goal_desc goal_num
0  Promote sustained, inclusive and sustainable e...        8
   target_id  goal_id target_num  \
0          1        1        8.1
1          2        1        8.5

                                         target_desc
0  Sustain per capita economic growth in accordan...
1  By 2030, achieve full and productive employmen...
   subindicator_id subindicator_num  \
0                1            8.1.1
1                2            8.5.2

                                subindicator_desc
0         Annual growth rate of real GDP per capita
1  Unemployment rate by both sexes of 15 years an...
   unemployment_id  target_id  subindicator_id unemployment_date  \
0                1          2                2              2005
1                2          2                2              2006
2                3          2                2              2007
3                4          2                2              2008
4                5          2                2              2009

  unemployment_value
0                7.8
1                8.0
2                7.3
3                7.4
4                7.5
   economics_id  target_id  subindicator_id economics_date economics_value
0             1          1                1           2001             1.0
1             2          1                1           2002             1.7
2             3          1                1           2003             3.1
3             4          1                1           2004             4.5
4             5          1                1           2005             3.0
```

Figure 14. Create Dimensional Model using Python-SQL script in Jupyter Notebook



```
In [2]:
import pandas as pd
import numpy as np
import psycopg2

In [3]:
from psycopg2 import sql

# Database connection parameters
db_params = {
    'dbname': 'your_database_name',
    'user': 'your_username',
    'password': 'your_password',
    'host': 'your_host',
    'port': 'your_port'
}

# SQL script
sql_scripts = """
    ----------------------
    ---   SQL script   ---
    ----------------------
    SET statement_timeout = 0;
    SET lock_timeout = 0;
    SET client_encoding = 'UTF8';
    SET standard_conforming_strings = on;
    SET check_function_bodies = false;
    SET client_min_messages = warning;

    SET default_tablespace = '';

    SET default_with_oids = false;

    ----------------------------
    ---   DROP TABLE script   ---
    ----------------------------
    DROP TABLE IF EXISTS goal_dim;
    DROP TABLE IF EXISTS target_dim;
    DROP TABLE IF EXISTS date_dim;
    DROP TABLE IF EXISTS subindicator_dim;

    ------------------------------
    ---   CREATE TABLE script   ---
    ------------------------------
    CREATE TABLE goal_dim(
        goal_key int PRIMARY KEY,
        goal_id int,
        goal_title varchar(255) NOT NULL,
        goal_desc text,
        goal_num smallint
    );

    CREATE TABLE target_dim(
        target_key int PRIMARY KEY,
        target_id int,
        target_num varchar(20) NOT NULL,
        target_desc text
    );
```

Figure 15. Dimensional Model Output [Economics_Fact Table Only] using Python-SQL script in Jupyter Notebook

```
In [23]: economics_fact.head()
```

Out[23]:

	target_key	subindicator_key	date_key	economics_value	economics_key	goal_key
0	1	1	2	1.0	1	1
1	1	1	3	1.7	2	1
2	1	1	4	3.1	3	1
3	1	1	5	4.5	4	1
4	1	1	6	3.0	5	1

```
In [ ]:
```

In here, we also create the database designed from normalized and dimensional model using Python-SQL script. This is from raw data which is in csv file to the output database. With this we can automate any data migration.

Please refer to the attached file for the full Python-SQL script

F. DATA ANAYLYSIS AND VISUALIZATION

Figure 16. Annual Growth Rate of GDP per capita

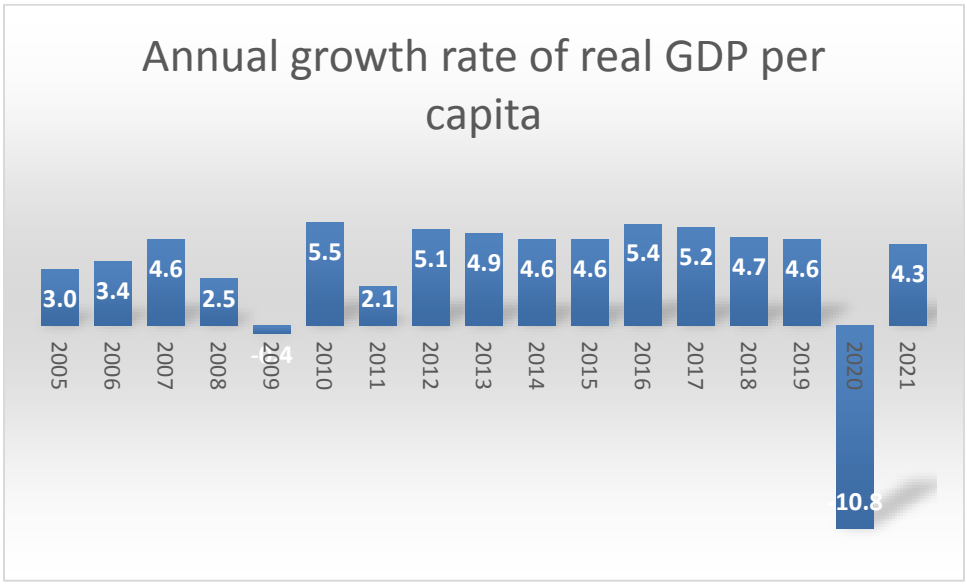


Figure 17. Unemployment Rate

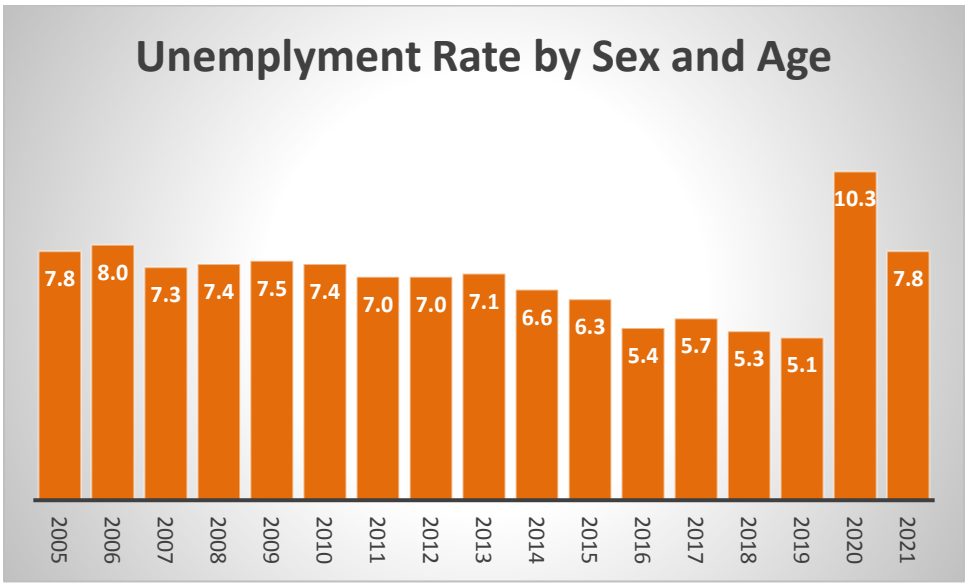


Figure 18. Correlation Matrix between Economics and Unemployment

	8.1.1 Annual growth rate of real GDP per capita	*8.5.2 Unemplyment Rate by Sex and Age*
8.1.1 Annual growth rate of real GDP per capita	1	
8.5.2 Unemplyment Rate by Sex and Age	-0.750407259	1

GDP per Capita Growth Rate:

The annual growth rate of real GDP per capita exhibits fluctuations over the years. Notable positive growth occurred in 2007, 2010, 2012, 2013, 2016, 2017, 2018, and 2021. However, a significant contraction is observed in 2009 and particularly in 2020, with a drastic decline of -10.8%, likely attributable to the global economic impact of the COVID-19 pandemic.

Unemployment Rate:

The unemployment rate by sex and age shows a general downward trend from 2009 to 2016, indicating an improvement in the labor market during this period. However, there's a slight increase in 2017, followed by a more noticeable decrease in 2018 and 2019. In 2020, there is a substantial spike in the unemployment rate, likely a direct consequence of the economic challenges posed by the pandemic. The rate decreases again in 2021, possibly reflecting economic recovery efforts.

Correlation:

The relationship between the annual growth rate of real GDP per capita and the unemployment rate seems complex. There is a strong negative correlation with -0.75 between GDP per capita growth rate and unemployment rate. This means economic growth leading to lower unemployment.

Having a broader range of indicators can enhance our understanding and facilitate informed decision-making, particularly in addressing challenges related to decent work and economic growth. We've illustrated that an improved and expansive database allows for real-time analysis, underscoring the value of comprehensive data in addressing complex issues.

V. RECOMMENDATION AND CONCLUSION

To derive meaningful insights and guide decision-making in addressing challenges related to decent work and economic growth in the Philippines, it is recommended to leverage the organized data in the relational database that this study created. Utilizing the relationships defined in the Entity Relationship Diagram (ERD) shown in Figure 1, our design for relational database management system (RDBMS) can facilitate the retrieval and analysis of key indicators. This database could serve as a foundation for constructing a dynamic and interactive dashboard.

The implementation of a user-friendly dashboard can further streamline the interpretation of key metrics or indicators. This visual representation enables stakeholders to instantly grasp trends, correlations and potential goal areas that required attention. For instance, the strong negative

correlation between economic indicator and unemployment suggests the need to prioritize and support all factors to achieve economic growth as it lowers the unemployment rate in the country. The dashboard, incorporating visualizations based on the analyzed data as we have shown on this study, would provide a real-time overview of economic and employment metrics.

In conclusion, the adoption of a relational database and a user-friendly dashboard can empower decision-makers with actionable insights and support to the Sustainable Development Goals. This technology-driven approach not only enhances data organization and accessibility but also facilitates a more informed and targeted strategy to promote decent work and foster economic growth. As the database and dashboard provide a dynamic platform for continuous monitoring and ongoing analysis, stakeholders can adapt strategies based on real-time trends and ensure a more responsive and effective approach to address the identified challenges.

VI. REFERENCE

The datasets generated during and/or analyzed during the current study are publicly available in the Philippine Statistics Authority at openstat.psa.gov.ph