**MIS 6060,**

**Energy Mix Optimization:**

**Enhancing Sustainability**

**Deepa Palariya, Soundarya Bachu**

Final Project Report

## ***Organization description, problem statement and approach used.***

Our project aims to optimize energy usage across various sectors by conducting a comprehensive analysis of energy mixes, efficiency levels, and the utilization of renewable and non-renewable energy sources. By examining these factors, we intend to develop strategies that promote sustainability and reduce environmental impact within the energy sector.

The primary objective of this project is to minimize the environmental footprint associated with energy consumption while enhancing sustainability through the implementation of sustainable effective strategies and policy recommendations. We will achieve this goal by leveraging data from the [U.S. Energy Information Administration (EIA)](https://www.eia.gov/opendata/) to inform our analysis and decision-making process.

To address the complexities of energy data analysis, we have adopted a systematic approach:

1. Data Cleaning and Filtering: We have meticulously cleaned and filtered a comprehensive dataset spanning from 1949 to 2023. Our focus is on the most recent 10 years (2013-2023) and 24 key variables relevant to energy consumption.
2. Analysis of Energy Consumption Patterns: Our analysis will involve comparing the contributions of renewable and non-renewable energy sources across different sectors of the economy. This comparative analysis aims to identify trends and opportunities for optimizing energy usage.
3. Identification of Optimization Opportunities: By examining energy consumption patterns and sector-specific trends, we aim to identify areas where efficiency improvements and policy interventions can lead to sustainable energy practices.

Overall, our approach is data-driven and focused on generating insights that can drive meaningful change in energy management practices. Through this project, we seek to contribute to the advancement of sustainability goals and environmental stewardship within the energy sector.

***Overview of Decision-Making Challenges and the Role of DW/BI Applications:***

1. **Analyzing Energy Consumption Patterns:** The DW/BI application will facilitate the analysis of energy consumption from renewable and non-renewable sources across industrial, commercial, and residential sectors. This analysis will help identify the predominant energy sources used in each sector and uncover trends and patterns in energy usage over the past 10 years.
2. **Identifying Most Used Energy Sources:** By leveraging DW/BI tools, stakeholders will determine which energy source (renewable or non-renewable) is predominantly used across different sectors. This identification will provide insights into areas where non-renewable energy usage can be minimized to reduce carbon emissions and promote sustainability.
3. **Studying 10-Year Trends:** Our project will examine the 10-year trend of the most used energy source identified in each sector. This analysis will reveal shifts in energy consumption patterns over time, allowing stakeholders to understand historical usage trends and forecast future energy demands.
4. **Enabling Data-Driven Decision-Making:** Through real-time data monitoring and analysis, the DW/BI application will empower stakeholders to make informed decisions regarding energy optimization strategies. This includes prioritizing efforts to transition towards renewable energy sources and implementing policies to reduce reliance on non-renewable sources.

In summary, the DW/BI application plays a critical role in our project by providing actionable insights into energy consumption patterns and trends across sectors. By leveraging these insights, stakeholders can prioritize initiatives to minimize environmental impact, promote sustainable energy practices, and drive effective decision-making towards a greener future.

## ***Raw data and cleaning using python code:***

We retrieved the dataset from the Energy Information Administration (EIA) in a non-relational format. To make it compatible with a relational database, we employed Python to transform the data.

Figure 2: Raw Energy Data

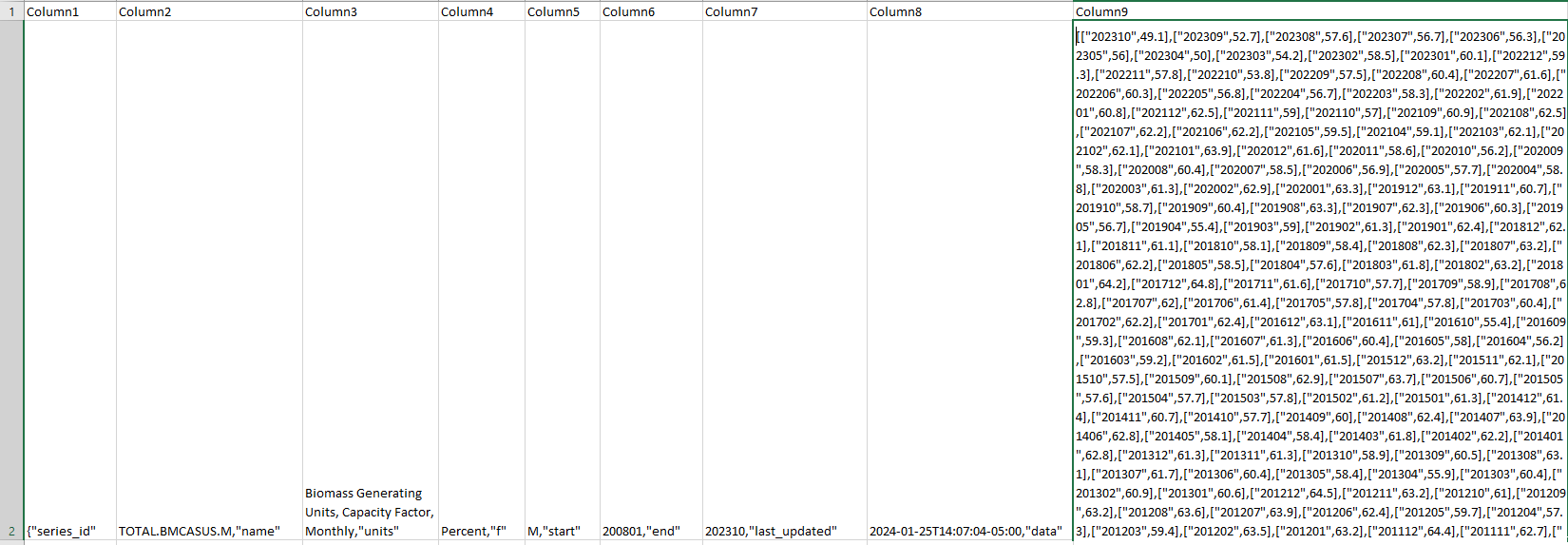
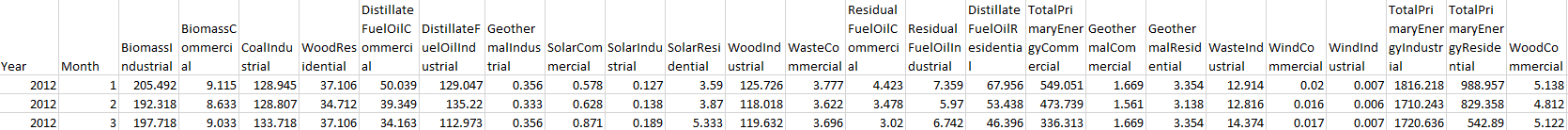


Figure 3: Cleaned Energy Data



Key steps in this transformation process include:

* **Loading the Data:** Initially, the raw data was loaded from a CSV file into a Pandas Data Frame named energy.
* **Filtering Columns:** We filtered the dataset to include only columns relevant to specific energy-related variables such as total primary energy consumption.
* **Cleaning Column Names:** Column names were cleaned to enhance readability and consistency by removing unnecessary text and formatting.
* **Splitting Data in Column 9:** The data in Column 9 was split into separate rows for each year-month-value combination, resulting in a new Data Frame called new.
* **Iterating Over Data:** For each variable, we iterated over the data in Column 9, splitting the year-month-value data and creating a new Data Frame.
* **Concatenating Data Frames:** All individual Data Frames for each variable were concatenated into a single Data Frame named new, which now contains all the cleaned and transformed data.
* **Saving the Data:** The final Data Frame new was saved as a new CSV file named MonthlyData.csv for further analysis.
* **Checking the Results:** To ensure the success of the transformation, we checked the results by loading and displaying the final data from the saved CSV file.

This systematic approach ensured that the data from EIA, initially in an incompatible format, was successfully converted into a structured format suitable for storage and analysis in a relational database environment.

## ***Sample Master Table and Transaction Table Structure:***

Figure 1 showcases the stage tables employed in the initial data loading phase, while Figure 2 depicts our data warehouse structure, embracing a Snowflake schema and comprising 3 master and 3 transactional tables. Our comprehensive analysis confirms the Snowflake schema as the optimal choice for our needs. Each master table acts as a dimension table, while every transaction table serves as a Fact table, embodying the essence of a snowflake schema. These tables collectively play a crucial role in deriving essential Key Performance Indicators (KPIs) vital for formulating strategic solutions aimed at energy optimization.

Figure 1:

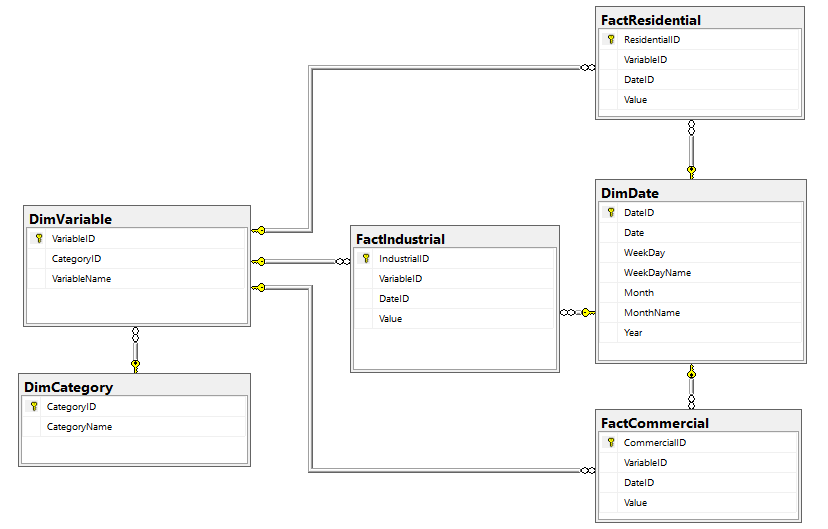
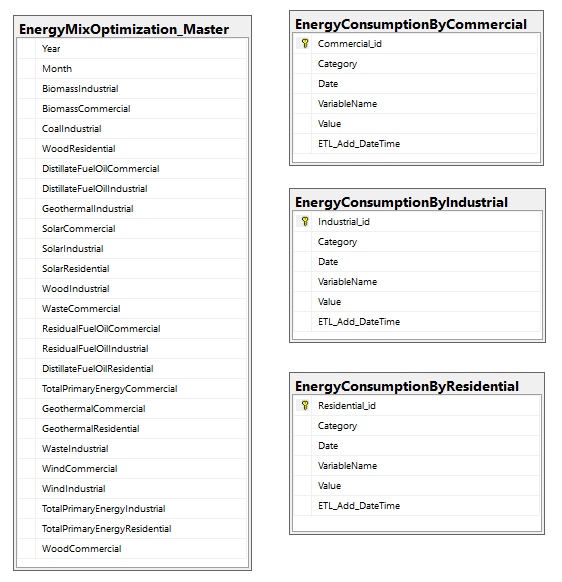


Figure 2:



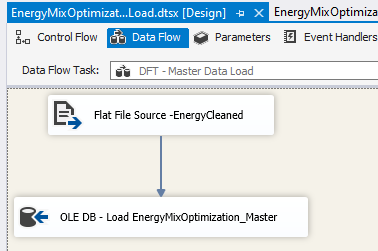
## ***The specifications for the DW/BI solution and Schema and design of the DW/BI application.***

In our approach for the Energy Mix Optimization project, we employed a comprehensive toolset including Excel, Python, MS SQL, SSIS, and SSAS. This combination enabled us to effectively manage and analyze data, covering various aspects of data processing from initial extraction and transformation to storage and analysis.

When designing the database schema for the Energy Mix Optimization project, our first step was to create the EnergyMixOptimization\_Master table. This table serves as the foundation for the project, storing energy consumption data for different sources and sectors on a monthly basis. This comprehensive view of the energy mix data is essential for analysis and decision-making.

To populate the EnergyMixOptimization\_Master table, we utilized the EnergyMixOptimization\_MasterDataLoad SSIS package for data dump, as shown in Figure 3. This package facilitated the loading of data into the table, ensuring its accuracy and consistency for further analysis and reporting.

Figure 3: EnergyMixOptimization\_Master Data Load



After creating the EnergyMixOptimization\_Master table, we utilized pivot functionality in MS SQL to create three stage tables: Commercial, Industrial, and Residential. These stage tables were instrumental in reducing redundancy to some extent and facilitating further data processing.

Subsequently, we loaded the data from these stage tables to the corresponding Energy Consumption by Commercial, Industrial, and Residential tables. During this process, we ensured to handle transformations such as data conversion and no duplicate insertion. This was achieved using the EnergyMixOptimization\_DataLoad SSIS package, which streamlined the loading process and ensured the integrity of the data being inserted into the tables.

Figure 4: EnergyMixOptimization\_DataLoad

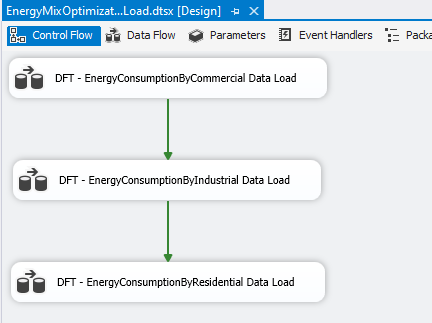
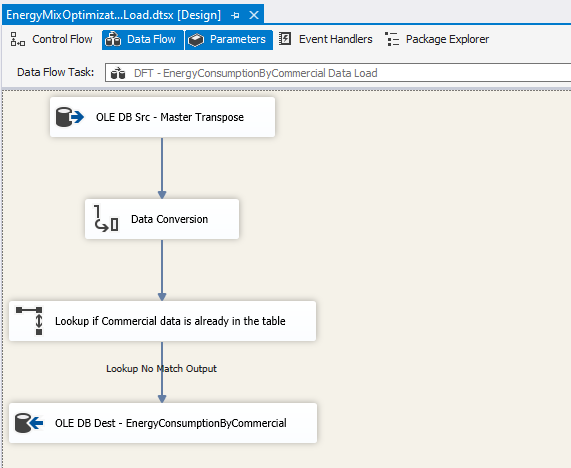


Figure 5: EnergyMixOptimization\_DataLoad Transformation



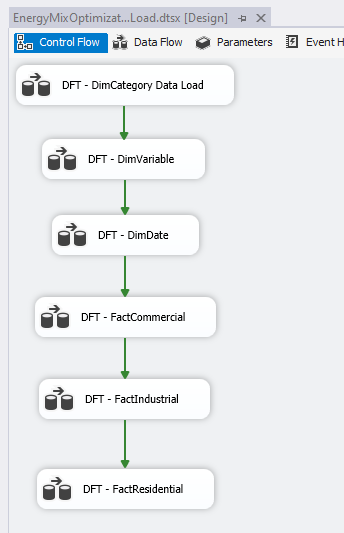
In our efforts to further eliminate redundancy and enhance the analytical capabilities of the Energy Mix Optimization project, we implemented dimensional modeling. This involved creating separate dimension tables (DimCategory, DimVariable, DimDate) and fact tables (FactCommercial, FactIndustrial, FactResidential) in SQL Server.

Dimensional modeling allows us to organize the data into manageable parts, making it easier to navigate and query for specific information. For example, DimCategory stores distinct categories related to energy consumption, while DimVariable stores variables related to each category. DimDate stores date-related information, enabling time-based analysis.

To maintain data integrity and relationships between dimension and fact tables, we established primary and foreign key relationships, as depicted in Figure 1. These relationships ensure that data is correctly linked and can be queried efficiently.

To load these tables, we utilized the EnergyMixOptimization\_DWDataLoad SSIS package. This package streamlined the process of populating the tables, ensuring that data was loaded accurately and efficiently.

Normalization was a key aspect of our approach, aimed at eliminating redundancy and improving data integrity. Instead of duplicating data in the fact tables, I used foreign keys to reference entries in the dimension tables. This ensures that the data is consistent and reduces the risk of errors or inconsistencies.

Figure 6: EnergyMixOptimization\_DWDataLoad

We are able to see the diced and sliced data using SSAS Cubes in data. Below figures are showing few samples.

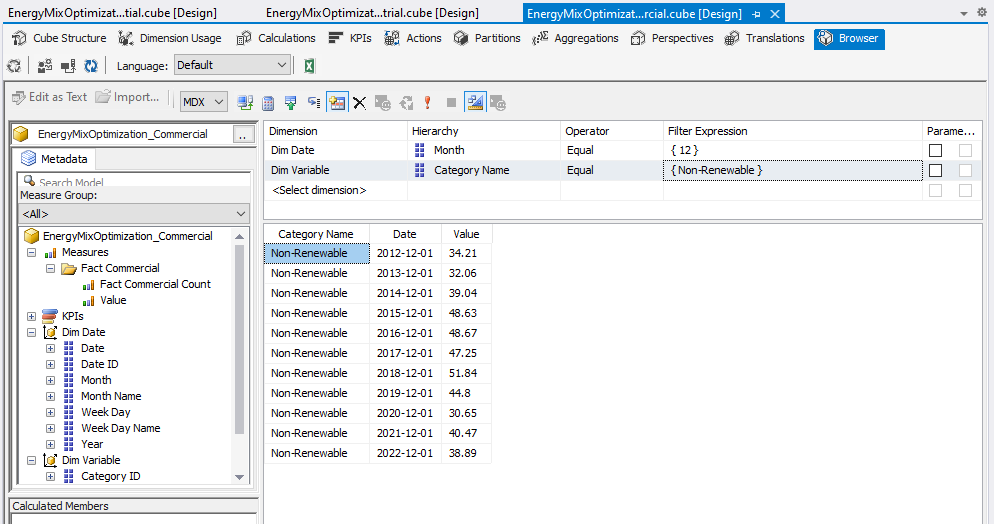
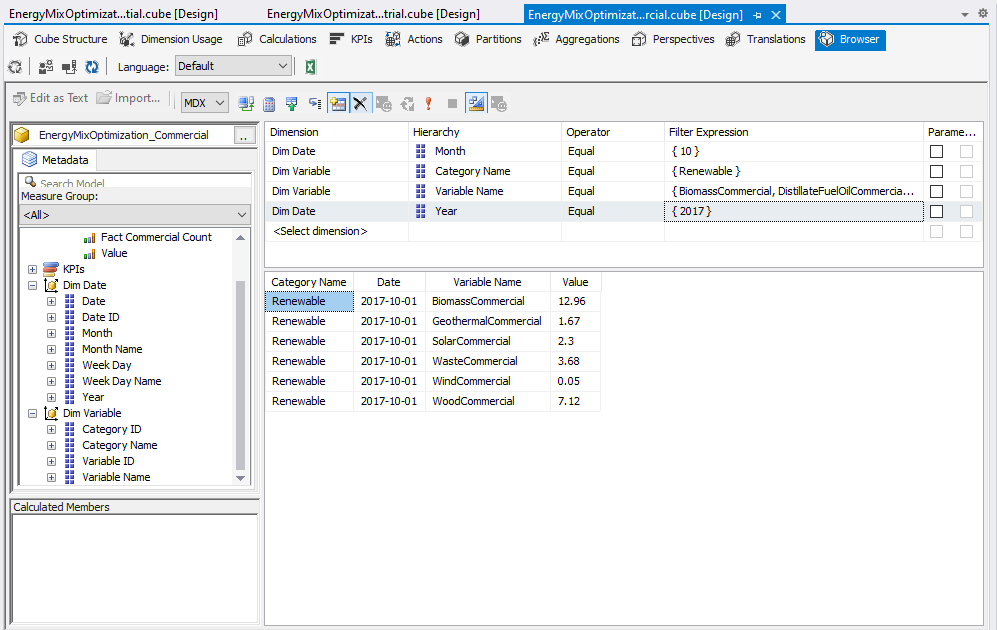
Figure 7:

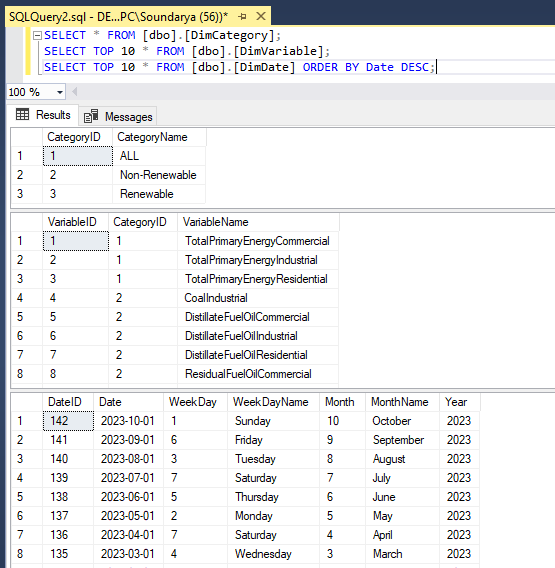
Figure 8:



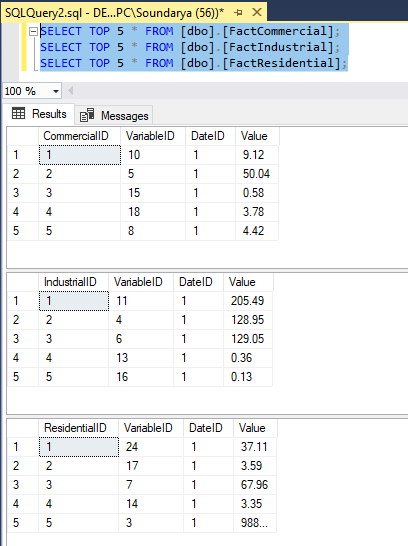
## ***Helping views and queries created to account for the KPIs demonstrating how my application is used to address the problems/KPIs identified.***

With the data warehouse (DW) set up and populated, our next task was to create views for calculating Key Performance Indicators (KPIs). We deliberated between using MS SQL for queries or Power BI for DAX functions, but we chose MS SQL for both views and KPIs. Here are the tables and examples of the views we created, forming the foundation for querying and building KPIs for visualization.

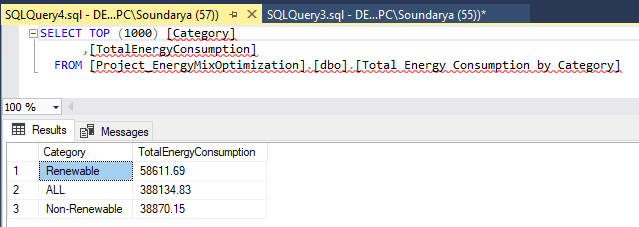
1. Dimension tables:



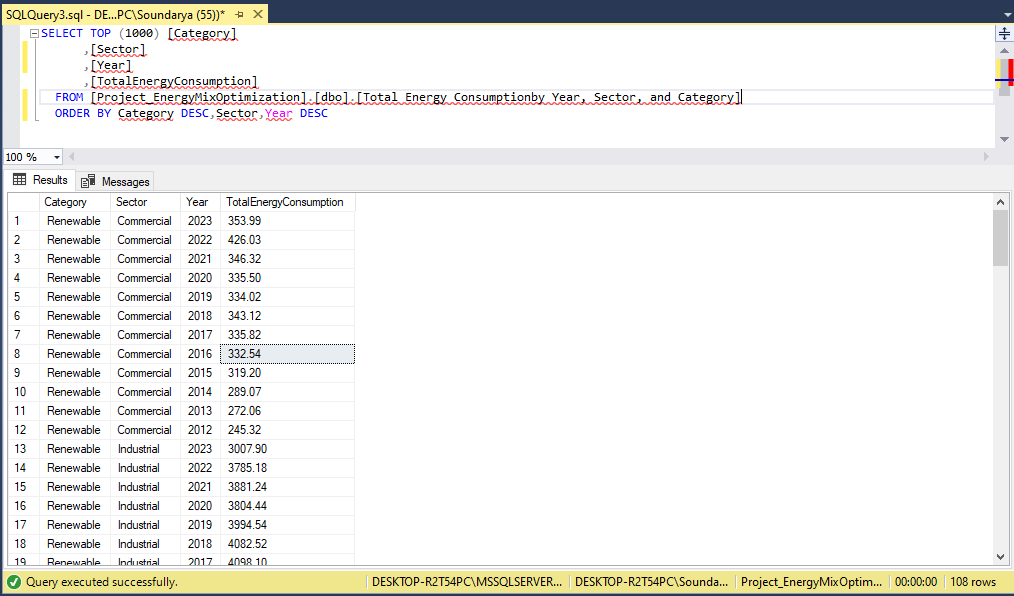
1. Fact Tables:



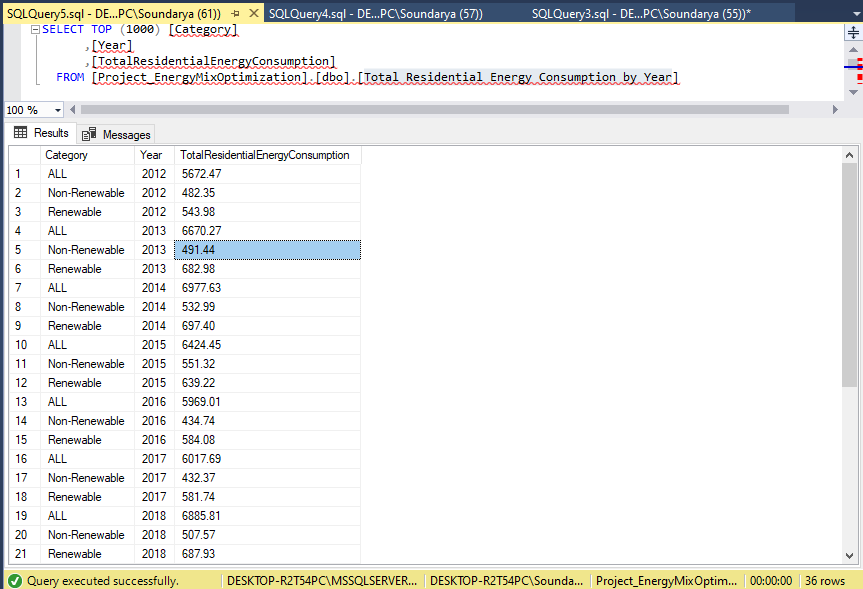
1. KPI - Total Energy Consumption by Category



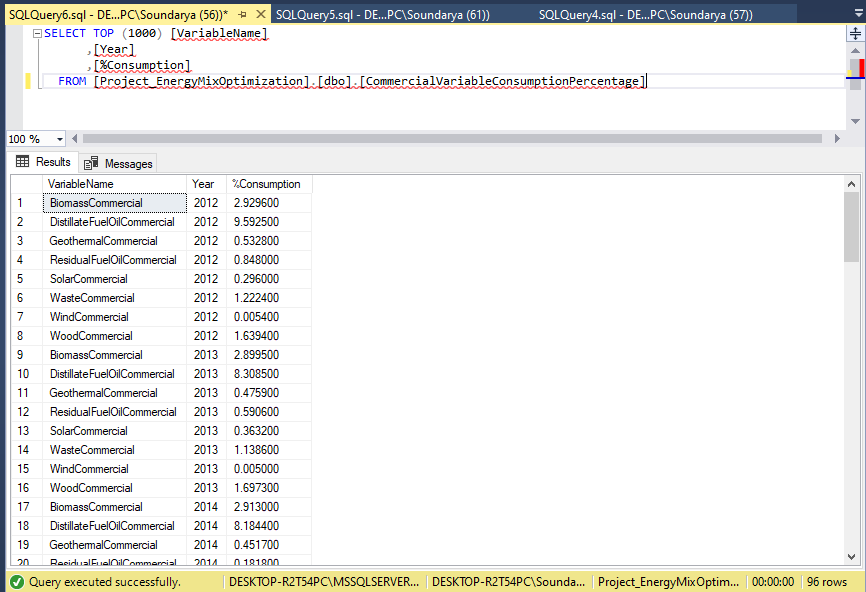
1. KPI – Total Energy Consumption by Year, Sector, and Category



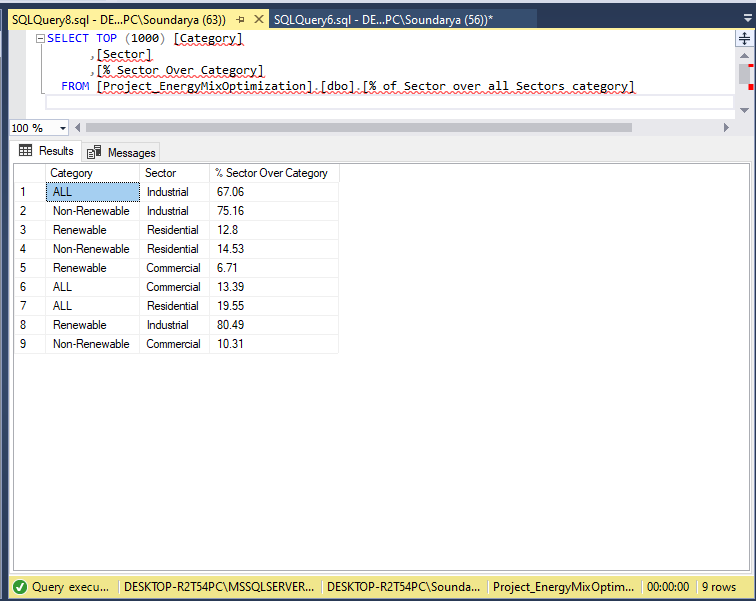
1. Total Energy Consumption by Category



1. Commercial Variable Consumption Percentage



1. KPI - % of Sector over all Sectors category

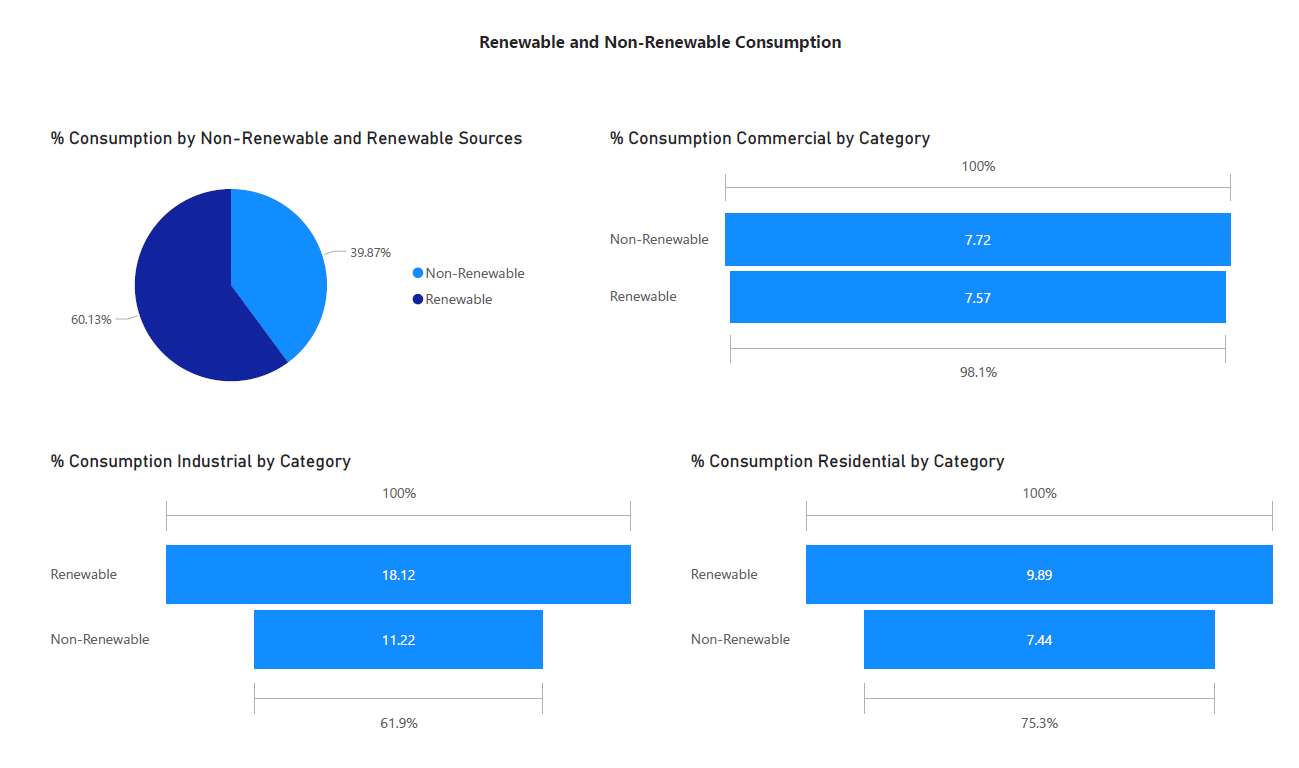
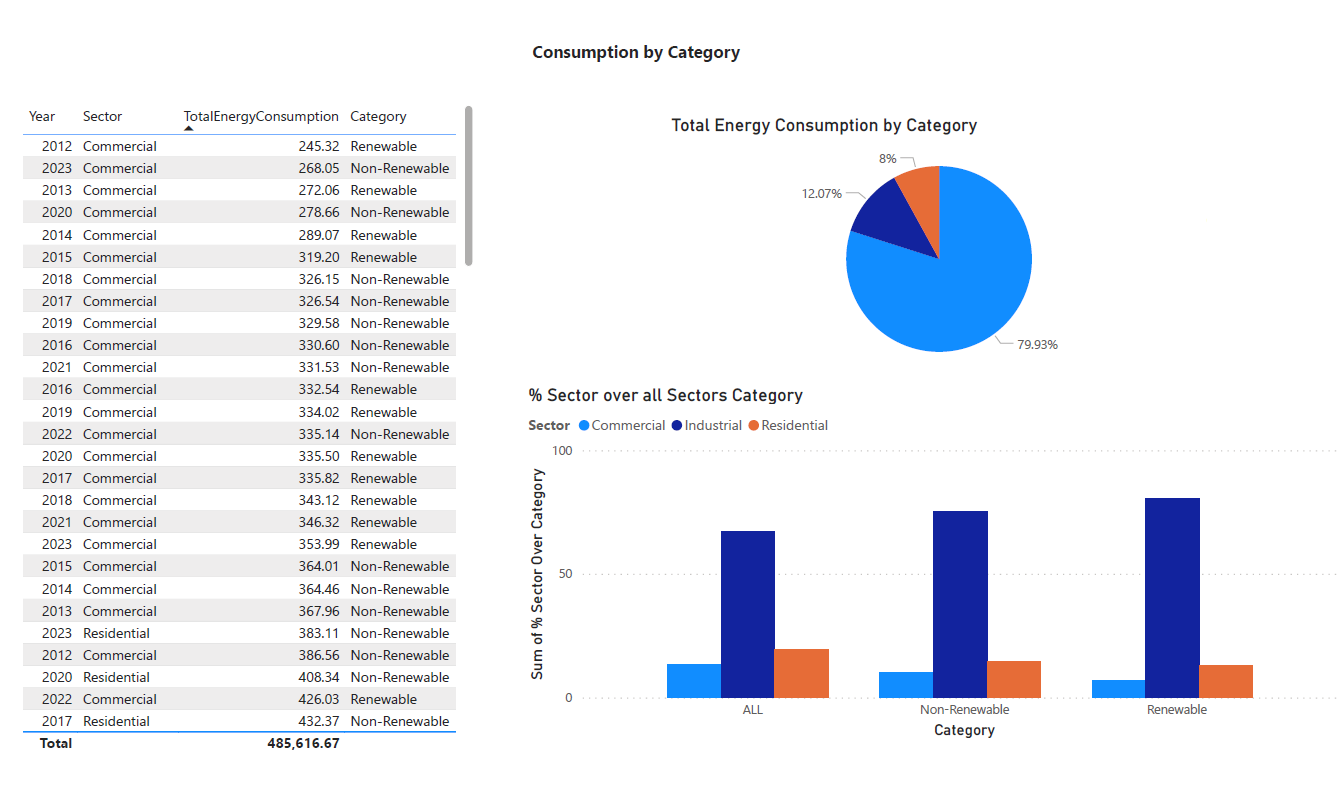
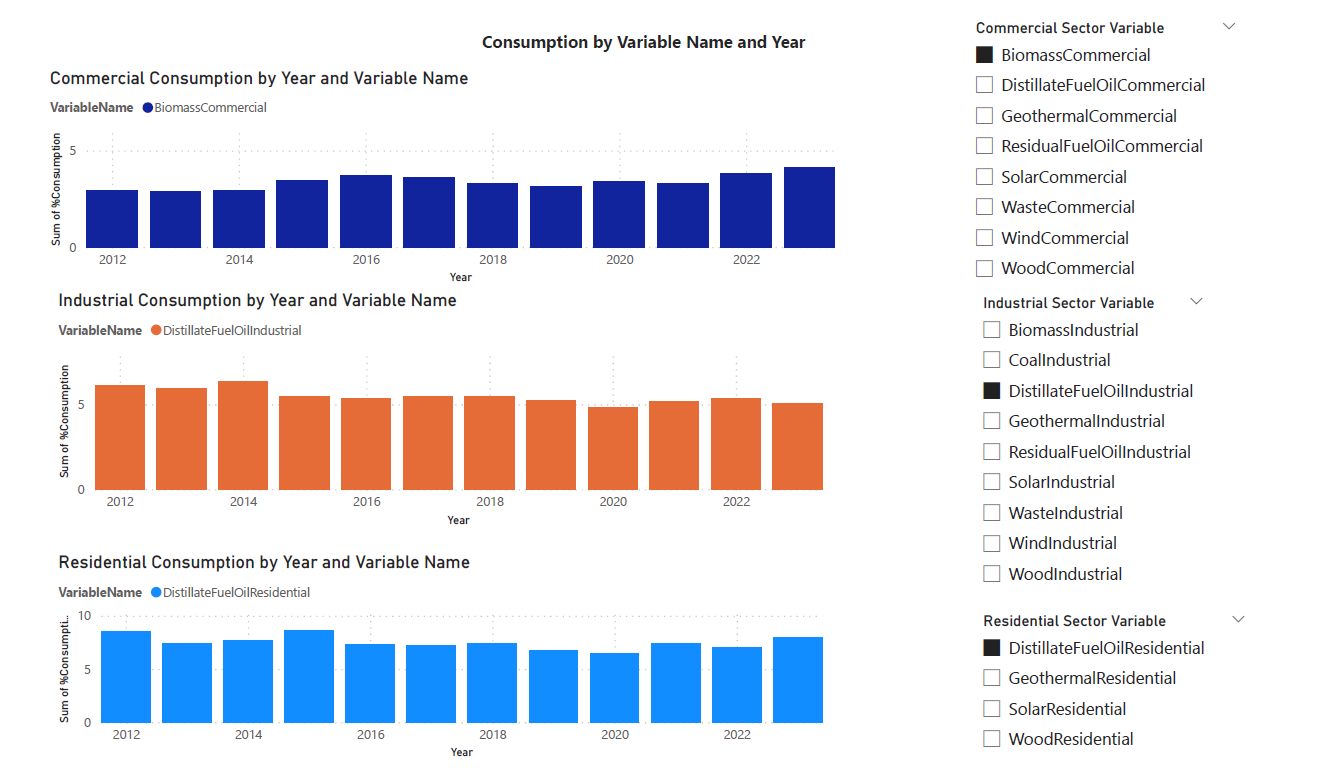
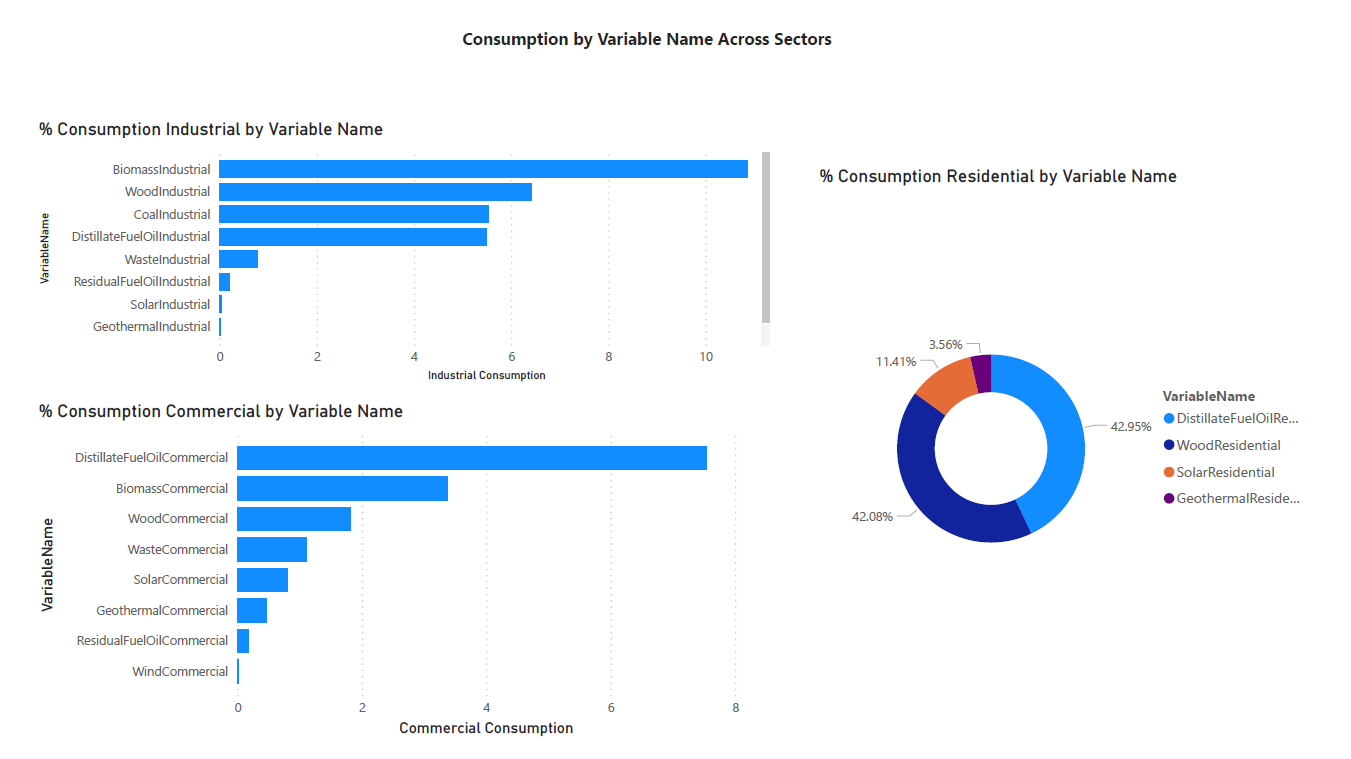


**Key queries for informed decision - making with the decision support systems Dashboard**

* Total Energy Consumption across all sectors by Year and Category
* Total Energy Consumption across all sectors by Source (renewable/non-renewable)
* Trend and patterns for both renewable and non - reneweable sources across sectors over the last 10 years
* Percentage of Total Energy Consumption by category
* Percentage of total renewable and non-renewable consumption
* Percentage of total renewable and non-renewable consumption across all sectors
* Percentage of total renewable and non-renewable consumption across all sectors by Category

A graph of different colored bars

Description automatically generated with medium confidence**The below screenshots helped answer all the above-listed questions.**



## ***f) Final Analysis/Findings:***

Based on findings in Power BI, below are the key observations:

**Commercial Sector Energy Consumption:**

1. Total energy consumption in the commercial sector is higher from renewable sources compared to non-renewable.
2. Maximum consumption for both renewable and non-renewable sources was observed in 2022, with a decrease in 2023.
3. Despite fluctuations, renewable energy consumption remains higher than non-renewable.

**Residential Sector Energy Consumption:**

* + 1. Renewable energy consumption in the residential sector peaked in 2019 but has since fluctuated, decreasing in 2020 and then increasing till 2022, followed by a decrease in 2023.
    2. Similar trends were observed for non-renewable sources, but renewable energy consumption remains higher overall.

**Industrial Sector Energy Consumption:**

1. Renewable energy consumption in the industrial sector is nearly double that of non-renewable.
2. Maximum renewable consumption was observed in 2014, followed by fluctuations in subsequent years.
3. Non-renewable consumption peaked in 2014 and has steadily decreased since then.

**Top Contributors in Industrial Sector:**

1. Biomass is the top contributor (10%) to energy consumption in the industrial sector, followed by wood (6.42%) and coal (5.53%).
2. Biomass consumption peaked in 2015 and has since declined.

**Top Contributors in Commercial Sector:**

1. Distillate fuel oil is the top contributor (7.54%) in the commercial sector, followed by biomass (3.38%) and wood (1.81%).
2. Distillate fuel oil consumption decreased from 2012 to 2020, with spikes in 2021 and 2023.

**Top Contributors in Residential Sector:**

1. Distillate fuel oil is the top contributor (42.95%) in the residential sector, followed by wood (42.08%) and solar (11.41%).
2. Distillate fuel oil consumption peaked in 2015, with fluctuations in subsequent years.

**Total Energy Consumption and Sources:**

1. Renewable sources account for 12% of total energy consumption across all sectors, while non-renewable sources account for 8%.
2. Industrial sector has the highest renewable consumption (80%), followed by residential (12.8%) and commercial (6.7%).

**Non-Renewable Consumption Distribution:**

1. The industrial sector has the highest non-renewable consumption (75.16%), followed by residential (14.53%) and commercial (10.31%).

**Distribution of Renewable and Non-Renewable Sources:**

1. Overall, renewable sources account for 60% of total energy consumption, while non-renewable sources account for 40%.
2. Distribution varies across sectors: commercial (7.57% renewable, 7.72% non-renewable), residential (9.89% renewable, 7.44% non-renewable), and industrial (18% renewable, 11.2% non-renewable).

Below are the solutions and recommendations proposed to optimize energy usage to minimize environmental impact and enhance sustainability:

## **Promote Renewable Energy Adoption:**

* Encourage increased adoption of renewable energy sources, especially in sectors where renewable energy consumption is already higher than non-renewable (e.g., commercial and residential sectors).
* Offer incentives and subsidies to businesses and households to invest in renewable energy technologies like solar panels, wind turbines, and geothermal systems.

## **Implement Energy Efficiency Measures:**

* Focus on improving energy efficiency across all sectors by implementing energy-saving technologies and practices.
* Conduct energy audits and provide recommendations for energy-efficient upgrades in industrial, commercial, and residential buildings.

## **Develop Sector-Specific Energy Plans:**

* Develop tailored energy plans for each sector based on their energy consumption patterns and trends identified in the analysis.
* Set sector-specific goals for reducing reliance on non-renewable energy sources and increasing the share of renewable energy in the overall energy mix.

## **Invest in Biomass and Wood Energy Solutions:**

* Given the significant contribution of biomass and wood to energy consumption in the industrial sector, invest in sustainable biomass and wood energy solutions.
* Promote responsible forestry practices and support the development of biomass energy technologies.

## **Explore Alternative Fuel Options:**

* Encourage the use of alternative fuels, such as biofuels and hydrogen, to replace conventional fossil fuels like distillate fuel oil in the commercial and residential sectors.
* Support research and development efforts to advance alternative fuel technologies.

1. **Policy and Regulatory Interventions:**

* Advocate for policies and regulations that prioritize renewable energy integration and penalize excessive reliance on non-renewable energy sources.
* Implement carbon pricing mechanisms or emissions trading schemes to incentivize the transition to cleaner energy sources.

## **Educational Campaigns and Awareness Programs:**

* Launch educational campaigns to raise awareness about the benefits of renewable energy and energy efficiency.
* Engage with stakeholders, including businesses, policymakers, and the public, to promote sustainable energy practices and behaviors.

## **Monitor and Evaluate Progress:**

* Establish monitoring and evaluation mechanisms to track the progress of energy optimization initiatives.
* Regularly assess the effectiveness of implemented solutions and adjust strategies based on performance metrics and feedback.

By implementing these solutions and recommendations, our project can contribute to achieving its goal of optimizing energy usage, reducing environmental impact, and enhancing sustainability across various sectors. Collaboration with stakeholders and alignment with government policies will be essential for successful implementation and long-term impact.

## ***g) Summary/Lessons learned/Limitations:***

Through this project focused on optimizing energy usage across sectors by analyzing a dataset from 2013 to 2023, we gleaned insights into consumption patterns, emphasizing the dominance of renewable energy in certain sectors and the importance of promoting renewable adoption for sustainability.

Key lessons include the significance of thorough data cleaning and tailored sector-specific energy solutions, facilitated by tools like Power BI for effective data visualization. Also, implementation of SSAS and SSIS with real world data was one of the learnings.

Limitations encompassed data scope and generalizability due to external factors, highlighting the complexity of energy systems and the need for interdisciplinary approaches to address sustainability challenges. Moving forward, refining data strategies and acknowledging these insights, we recognize the need for additional data such as emission data and a wider range of sources for both renewable and non-renewable energy to enable a more granular analysis and informed decision-making.

## ***h) Instructions to run the entire DW/BI application:***

The essential components required to run the entire Data Warehouse (DW) and Business Intelligence (BI) application include the Database Management System (DBMS) Microsoft SQL Server (MSSQL), SSIS, SSAS and the BI tool Power BI. Attempting to execute scripts in alternative DBMS platforms like PostgreSQL or Oracle may encounter issues due to differences in syntax, as the tables are created using MS SQL Server queries.

To initiate the process, attach the provided .mdf file named **Project\_EnergyMixOptimization.mdf** to your MSSQL server, ensuring seamless operation of the DW. Subsequently, import the data directly into Power BI via SQL Database Once the data is loaded into Power BI, the pre-configured dashboards will be readily accessible.

**List of files inside the folder -** **MIS 6060\_EnergyMixOptimization\_FinalProjectFiles :**

1. **EnergyOriginal.csv:** This file contains all the Excel data sourced from the US Energy Administration website for 10 years (2013 - 2023).
2. **DataCleaning\_EnergyConsumption.ipynb:** This is a Jupyter notebook where data cleaning is done and final file is derived for further analysis.
3. **EnergyCleaned.csv:** This file is the final cleaned version used for our analysis.
4. **EnergyMixOptimization\_Master Table Creation.sql:** This is a script to create EnergyMixOptimization\_Master Table.
5. **Master\_Transpose View Creation.sql:** This view will help us to sagregate data according to sectors.
6. **EnergyConsumptionBy Tables Creation.sql:** This script creates stage tables for all the sectors.
7. **Dim and Fact Tables Creation.sql:** This script will create datawarehouse, dimension and fact tables.
8. **EnergyMixOptimization\_SSIS.zip:** This folder contains SSIS package which loads all the MSSQL server tables.
9. **EnergyMixOptimization\_SSAS.zip:** This folder contains SSIS package which helps us see diced and sliced data.
10. **All Queries with views for Power BI.sql:** This script has views creation for all the KPIs used in our analysis.
11. **EnergyMixOptimization.pbix:** This is the Power BI desktop file where visuals are created based on the KPIs.
12. **EnergyMixOptimization.pdf:** This is the pdf version of the Power BI desktop file.
13. **Project\_EnergyMixOptimization.mdf:** This is the final Database which can be attached as is in MSSQL Server and views can be used to run the Power BI report. This file can be used standalone without importing the files listed from 1-9.

A white screen with black text

Description automatically generatedThese resources collectively facilitate the setup and execution of the DW/BI application, providing comprehensive insights into Energy consumption from Renewable and Non-Renewable sources across sectors.