Intelligent Energy Demand Forecasting in the US Using EIA and Weather Data

Aishanee Sinha Ketki Ramakant Maddiwar Leela Prasad Dammalapati Utkarsh Tripathi

Abstract

This project combines the electricity consumption data from the U.S. Energy Information Administration (EIA) and weather data from OpenMeteo API to study the trends observed in the conumption patterns and forecast the usage for next few days. Electricity consumption data is combined with weather features, which include temperature, precipitation, and wind speed, to deduce the impact of climate conditions on energy consumption. Predictive models are developed using Snowflake's ml capabilities to forecast daily electricity consumption. Results are then visualized in an interactive Power BI dashboards, showcasing regional electricity consumption trends, weather correlations, and seasonal demand variations. This comprehensive approach supports informed decision-making to optimize grid performance, reduce operational costs, and implement effective demand-response strategies during peak usage or extreme weather events.

I. Introduction

This project focuses on analyzing and forecasting electricity consumption trends in the United States by integrating historical energy consumption data from the U.S. Energy Information Administration (EIA) and weather data from Open Meteo. The aim is to provide actionable insights into energy demand patterns and their correlation with weather conditions.

The workflow begins with data ingestion through APIs, where historical and real-time datasets are collected. These datasets include detailed energy usage statistics and key weather variables such as temperature, precipitation, and wind speed. The collected data is stored in a Snowflake data warehouse, which serves as a centralized platform for managing and processing large datasets.

To prepare the data for analysis, ETL pipelines are implemented using Apache Airflow. These pipelines automate data extraction, cleaning, and loading into structured tables in Snowflake. The pipelines also validate data integrity, ensuring the reliability of the analysis.

Further data modeling and transformations are carried out using dbt (Data Build Tool). This includes the calculation of daily energy consumption trends, peak demands, and seasonal variations. Predictive analytics is performed using Snowflake's machine learning capabilities, where models are trained to forecast energy demand based on weather conditions and historical consumption data.

The project concludes with visualizations in Power BI. Interactive dashboards highlight key metrics such as regional energy consumption trends, weather correlations, and seasonal demand variations. These dashboards provide energy providers with valuable tools for optimizing grid performance, reducing operational costs, and implementing demand-response strategies during peak usage or extreme weather events.

By integrating reliable data pipelines, robust modeling techniques, and dynamic visualizations, this project offers a comprehensive approach to energy consumption analysis and demand forecasting. It enables energy providers to make informed decisions, optimize resource allocation, and enhance energy efficiency across the grid.

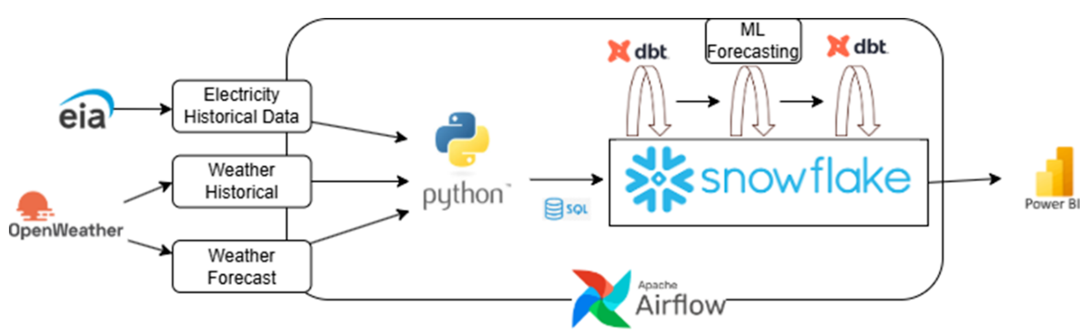


Fig. 1: System Architecture

II. Data Description

This section provides a comprehensive description of the datasets used in this analysis. The first dataset contains energy consumption data from various utility companies, and the second dataset provides weather forecast data. Both datasets are essential for analyzing energy usage patterns and correlating them with weather conditions.

A. Energy Consumption Dataset

The energy dataset contains daily energy consumption data from multiple utility companies in California, including Pacific Gas and Electric (PGAE), Southern California Edison (SCE), San Diego Gas and Electric (SDGE), and others. The data is used to assess energy usage trends across different regions.

1) Variable Descriptions for Energy Dataset:

1) **Period** (datetime):
   Description: The date for energy consumption data on which day it is recorded. It represents the specific day of energy consumption data.
2) **Subba** (String):
   Description: An abbreviation for the utility company, such as ”PGAE”, ”SCE”, or ”SDGE”.
3) **Subba Name** (String):
   Description: The full name of the utility company, e.g., ”Pacific Gas and Electric” or ”Southern California Edison”.
4) **Parent** (String):
   Description: The organization that is overseeing the utility company. Usually, CISO (California Independent System Operator).
5) **Parent Name** (String):
   Description: The full name of the parent organization, eg.. California Independent System Operator.
6) **Timezone** (String):
   Description: The timezone in which the data is recorded, eg..Pacific Standard Time.
7) **Value** (Float):
   Description: The energy consumption value for the utility company during the given period, measured in gigawatt-hours (GWh).

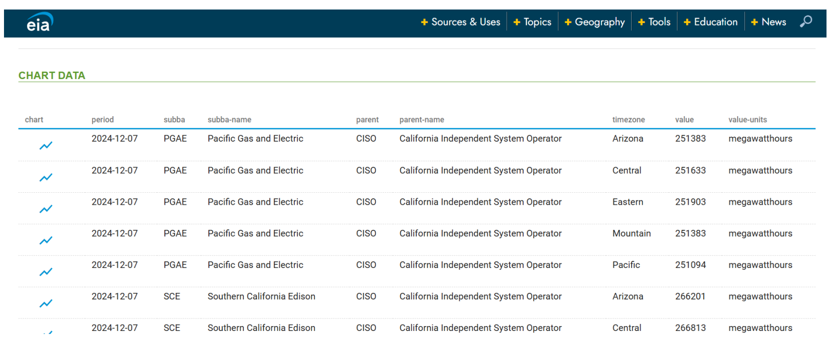


chart	period	subba	subba-name	parent	parent-name	timezone	value	value-units
	2024-12-07	PGAE	Pacific Gas and Electric	CISO	California Independent System Operator	Arizona	251383	megawatthours
	2024-12-07	PGAE	Pacific Gas and Electric	CISO	California Independent System Operator	Central	251633	megawatthours
	2024-12-07	PGAE	Pacific Gas and Electric	CISO	California Independent System Operator	Eastern	251903	megawatthours
	2024-12-07	PGAE	Pacific Gas and Electric	CISO	California Independent System Operator	Mountain	251383	megawatthours
	2024-12-07	PGAE	Pacific Gas and Electric	CISO	California Independent System Operator	Pacific	251094	megawatthours
	2024-12-07	SCE	Southern California Edison	CISO	California Independent System Operator	Arizona	266201	megawatthours
	2024-12-07	SCE	Southern California Edison	CISO	California Independent System Operator	Central	266813	megawatthours

Fig. 2: EIA API Overview

B. Weather Dataset

The weather dataset contains forecasted and historical weather data for specific locations and times, providing relevance and inferences to energy consumption trends.

1) Feature Descriptions for Weather Dataset:

1) **Date** (datetime):
   Description: The date for the weather measurement or forecast.
2) **Temperature (Max)** (Float):
   Description: The maximum temperature recorded or forecasted for that specific day, measured in degrees Celsius (°C).
3) **Temperature (Min)** (Float):
   Description: The minimum temperature recorded or forecasted for that specific day, measured in degrees Celsius (°C).
4) **Precipitation** (Float):
   Description: The total precipitation for the day, measured in millimeters (mm).
5) **Snowfall** (Float):
   Description: The total snowfall for the day, measured in millimeters (mm).

6) **Wind Speed** (Float):
Description: The maximum wind speed recorded or forecasted for the day, measured in meters per second (m/s).

7) **Average Temperature** (Float):
Description: The calculated average temperature for the day, derived from the maximum and minimum temperatures.

III. Functional Analysis

Our project consists of the following functional components:

A. Data Collection

- **Electricity Data**: The electricity consumption dataset is taken from an open source API, U.S. Energy Information Administration (EIA) API. This dataset contains daily electricity consumption values, time period where the energy value corresponds to, and regional information, such as the sub-balancing authority (SubBA), parent regions, and their respective time zones. The data is fetched with specific parameters to filter the data for relevant records, such as frequency (daily), time range, and location. The API response is extracted into into a Pandas DataFrame to further facilitate preprocessing and is subsequently loaded into Snowflake for further analysis.
- **Historical Weather Data**: Historical weather data is taken from Open Meteo API. This dataset contains weather related metrics, such as daily maximum and minimum temperatures, total precipitation, snowfall accumulation, and maximum wind speed. The data is retrieved for the State of California using longitude and latitude to ensure geographical relevance. Additional preprocessing steps are performed, such as calculating the average daily temperature, renaming the columns to maintain consistency and handling missing values. The processed data is then stored in Snowflake in a structured table format for further access and integration.
- **Forecast Weather Data**: Weather forecast data is sourced from the Open Meteo Forecast API. In comparison to historical weather data, the forecast includes daily maximum and minimum temperatures, precipitation levels, snowfall, and wind speeds. This data contains prediction of the aforementioned data inputs for 7 days period of time. This gives input for predictive models to estimate future electricity usage. The forecast data is processed and stored in Snowflake for consistency with historical weather datasets.
- **Data Integration and Storage**: The electricity and weather datasets are integrated into Snowflake tables. SQL queries ensure proper schema creation, data normalization, and primary key definitions. This integration enables seamless data access and supports downstream analytical tasks and model training processes.
- **Automation**: The data collection process is fully automated using Apache Airflow. The Airflow pipeline handles the tasks like data fetching , preprocessing, database insertion, and monitoring. This automation ensures scalability, reliability, and repeatability in the data collection workflow.

B. Data Modeling

- electricity_data_historical
- weather_data_historical
- weather_data_forecast
- electricity_weather_historical
- weather_forecast_processed
- electricity_data_forecast
- energy_historical_forecast
- energy_demand_final_data

C. Data Processing - Airflow (Data Pipeline using DAG)

For data pipeline we are using Apache Airflow

- The ETL pipeline in this project involves extracting electricity consumption and weather data from APIs using Airflow, setup on docker followed by transforming the data in Snowflake for further analysis.
- The raw data is cleaned, integrated, and processed, including the creation of a machine learning model for predicting future electricity demand based on weather conditions.
- The transformed data, along with the forecasted predictions, is stored in Snowflake for easy retrieval and analysis.

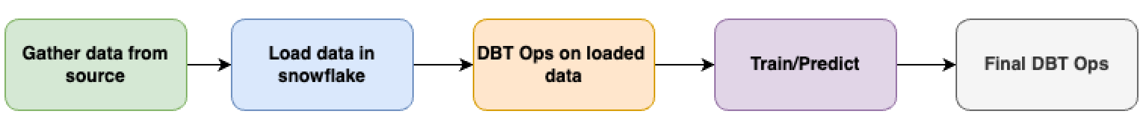


Fig. 4: Data Pipeline

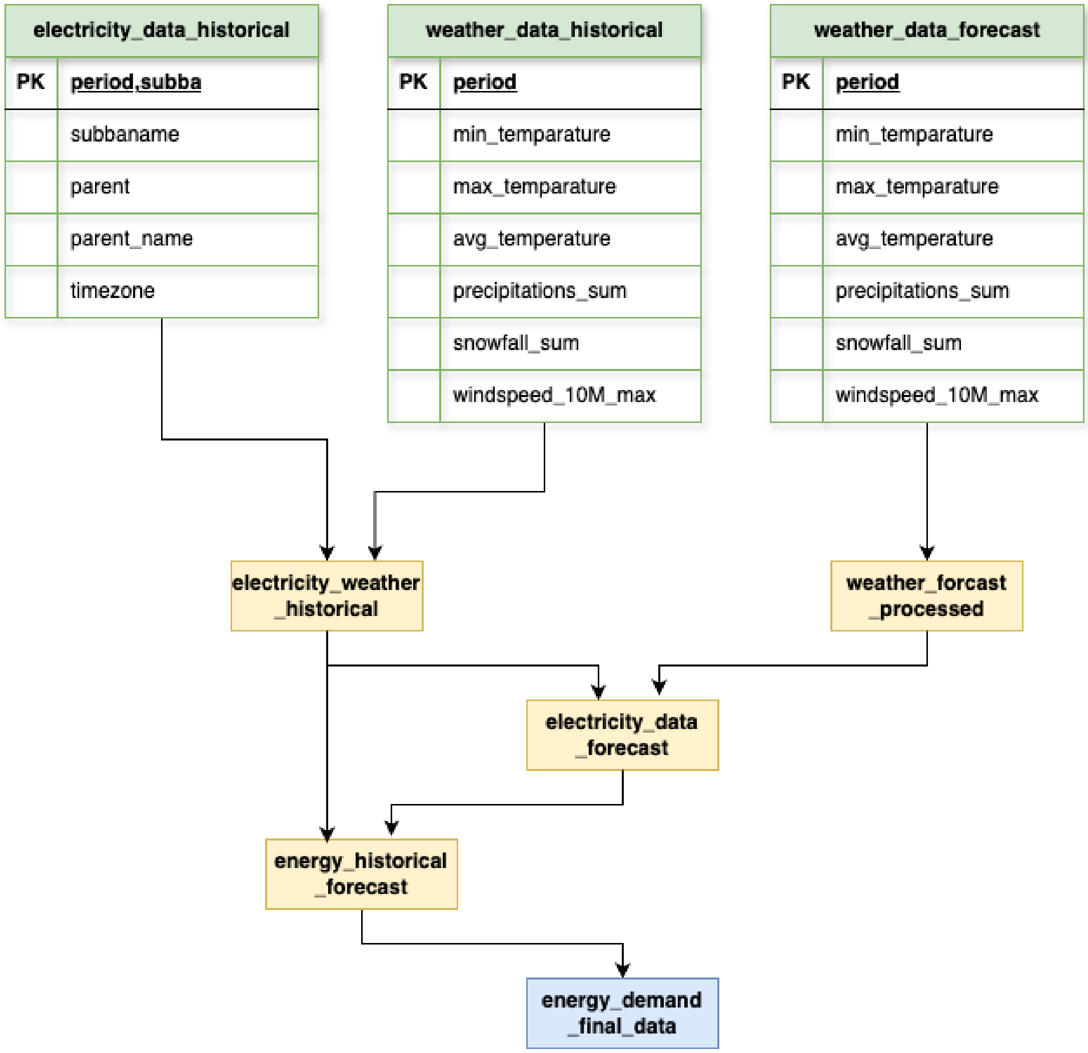


Fig. 3: Data Modeling

The DAG involved the following tasks:

- Gather data from electricity and weather API sources
- Load data into data warehouse/snowflake tables
- Perform data cleaning, merging and transformations using DBT
- Train a predictive model using historical data
- Evaluate weather's impact on energy generation and optimize predictions.

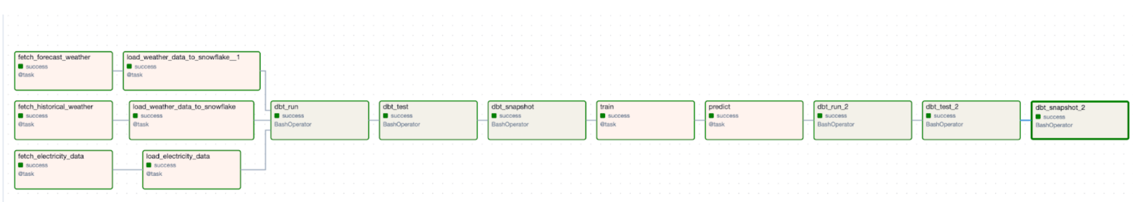


Fig. 5: DAG detailed tasks

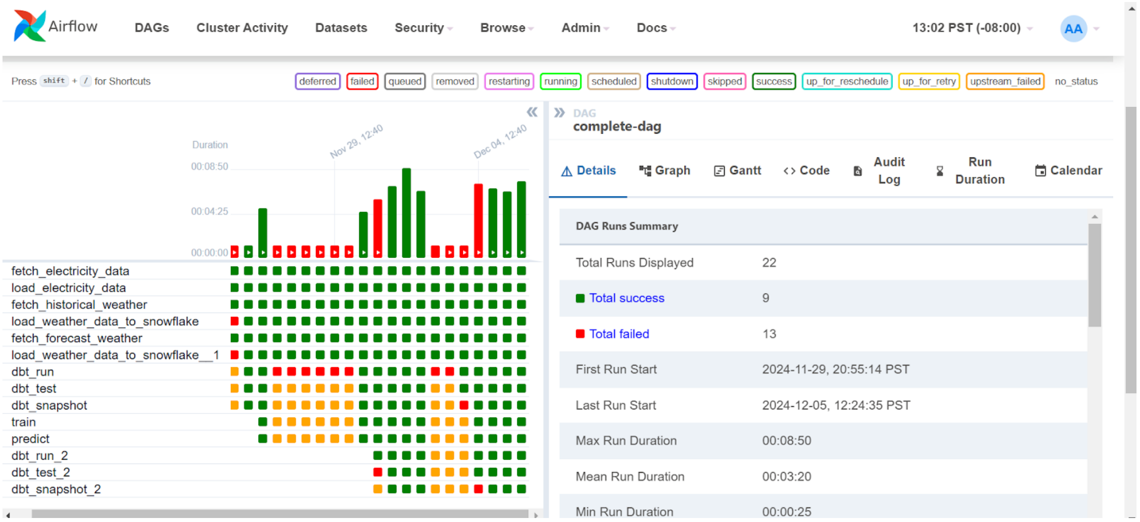


Fig. 6: DAG Executions

D. Data Transformation - DBT

- Merged the Electricity and Weather Historical Datasets on period using dbt.
- Transformed the Electricity consumption value from MWh to GWh
- Processed the weather forecast data. Performed a cross join to add subba to weather forecast data.
- Stored the final merged and processed data under Analytics schema in Snowflake dev database.
- Performed dbt test to check for null values in the merged dataset.
- Created snapshot on the merged table using check strategy checking on columns period and electricity consumption.

E. Machine Learning

We use a machine learning model to predict the Electricity Demand in GWh for the next 7 days.

- We retain only the necessary columns (PERIOD, ELECTRICITY_VALUE_GWH, SUBBA, MIN_TEMPERATURE, MAX_TEMPERATURE, AVG_TEMPERATURE, PRECIPITATION_SUM, SNOWFALL_SUM, WINDSPEED_10M_MAX) from electricity weather merged dataset as train input table.
- We use the Snowflake.ML.Forecast function to train our model. The target variable is ELECTRICITY_VALUE_GWh.
- We set the forecasting period as 7 days.
- The series column is subba and period is converted to timestamp for model training.
- All the other weather variables are treated as exogenous variables.
- We use the electricity_weather historical table for training the model and the weather_forecast_processed table for the next 7 days prediction.
- We setup two tasks train and predict in the Airflow DAG.
- The results are stored in the Snowflake table electricity_data_forecast

F. Data Transformations -dbt (2)

We have performed another set of dbt transformations after the train and predict tasks.

- Merged the Historical and Forecast Tables using dbt.
- Computed metrics temperature categories, wind speed categories, extreme weather indicator and predicted/actual flag in the final dataset.
- The final table is stored as energy_demand_final_data in Snowflake.
- Performed dbt test -not null on the final dataset to ensure there are no null values in ELECTRICITY_VALUE_GWH.
- Created snapshots by the ‘check’ strategy, checking on columns PERIOD AND ELECTRICITY_VALUE_GWH. The dbt snapshot table is stored as energy_demand_final_snapshot in Snowflake.

G. Dashboard Visualization

- Power BI Report
- Purpose of the dashboard

This report explores the trends in electricity consumption in California among major electricity providers. Also, it demonstrates the influence of other factors such as weather parameters (Average Temperature, Average Snowfall, Wind speed, and Precipitation). The dashboard effectively combines real-time and historical data to observe the changes in the pattern for the current year i.e. 2024 and last two years (2022, 2023) along with predicting next seven days of consumption.

- Tools Used for Visualization: Microsoft Power BI, Snowflake (for accessing data)
- Methodology:
- Extract:
  This report extracts data from snowflake target table named (ENERGY DEMAND FINAL DATA) using import method which requires manual refresh as of now to have latest data in report. In future advancements, Direct Query feature of Power BI can be leveraged to set up live connection with Snowflake to modify as per data change in Snowflake's table.

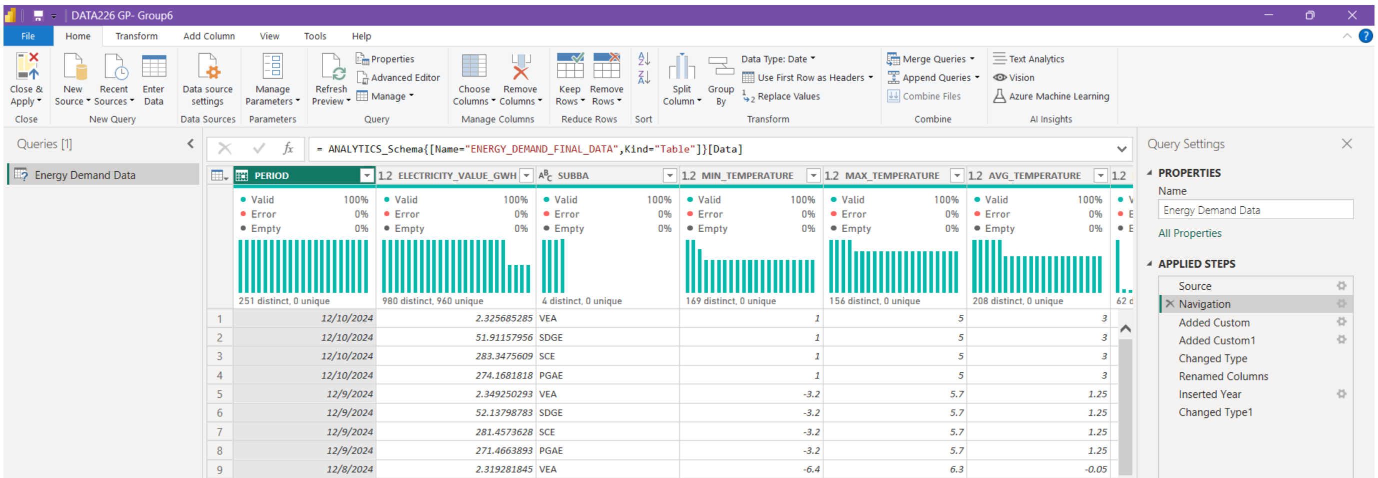


Fig. 7: System Diagram

- Transform:
  After extracting data, below transformation steps are being performed in Power Query Editor:
  - Creation of logic of categorizing data based on Peak and Off-Peak Period Types. As per current system design, Period Type has been decided to be assumed as Peak for Weekdays (i.e. Monday – Friday) and Weekends for Saturday and Sunday.
  - Converting AVG_TEMPERATURE field value to Fahrenheit using the formula { ((([AVG_TEMPERATURE]*9)/5)+32) }.
  - Renaming PERIOD column to Date for avoiding confusion with newly created Period Type column.
  - Extracting Year Column and converting to text for displaying it as an indicator on Home Page.

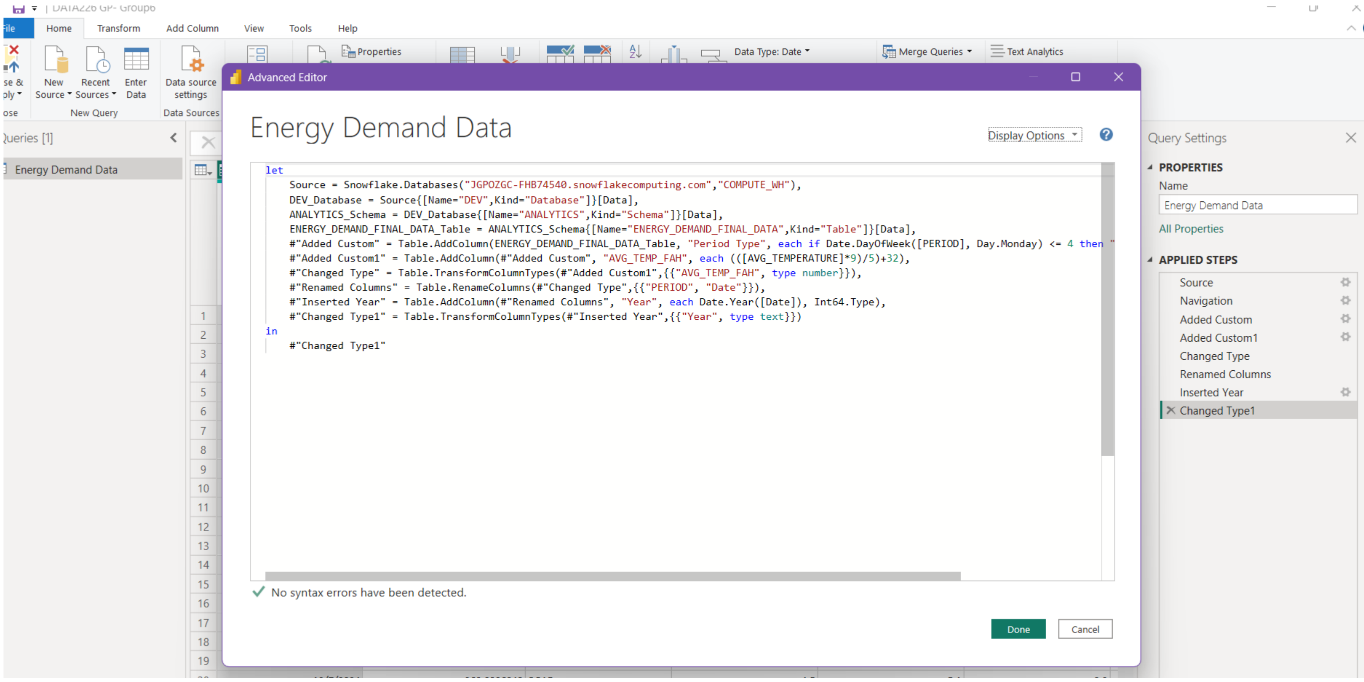


Fig. 8: System Diagram

– Load:
  In the report, we are using below mentioned indicators and displaying them in the form of visualizations.
  * S.No. Metric Name Definition Formula
  * Total Actual:
    For calculating cumulative sum of consumption for Actual Data

```
SUMX(
    FILTER(
        'Energy Demand Data',
        'Energy Demand Data'[ACTUAL_PREDICTED_FLAG] = "Actual"
    ),
    'Energy Demand Data'[ELECTRICITY_VALUE_GWH]
)
```

  * Total Predicted:
    For calculating cumulative sum of consumption for Predicted Data

```
SUMX(
    FILTER(
        'Energy Demand Data',
        'Energy Demand Data'[ACTUAL_PREDICTED_FLAG] = "Predicted"
    ),
    'Energy Demand Data'[ELECTRICITY_VALUE_GWH]
)
```

  * Min Temp:
    Minimum of the MIN_TEMPERATURE

```
((MIN('Energy Demand Data'[MIN_TEMPERATURE]))*9/5)+32
```

  * Max Temp:
    Maximum of the MAX_TEMPERATURE

```
((MAX('Energy Demand Data'[MAX_TEMPERATURE]))*9/5)+32
```

* Peak Demand Detail:
  To calculate Peak Demand Day and associated value

```
CONCATENATEX(
    TOPN(
        1,
        'Energy Demand Data',
        'Energy Demand Data'[ELECTRICITY_VALUE_GWH],
        DESC
    ),
    'Energy Demand Data'[Date] & ": " &
    FORMAT('Energy Demand Data'[ELECTRICITY_VALUE_GWH], "#,##0.00") & " GW
    ", "
)
```

- The report consists of following sheets:
  - Home Page
  - Energy Consumption Over Years
  - Other Factors Affecting Consumption
  - Electricity Consumption Breakdown by Providers
  - Temperature vs Electricity
  - Self-Analysis
- Home Page:
  This page displays the main dashboard of the report with several key performance indicators and slicer for selecting year. This page intends to display results for a selected year e.g. for year 2024 in below screenshot. This page also has a link to Consumption Trends dashboard which will direct the user on Energy Consumption over Years page. This link has been intended to set up as in after publication every other sheet will be hidden and first page to view is going to be home page.

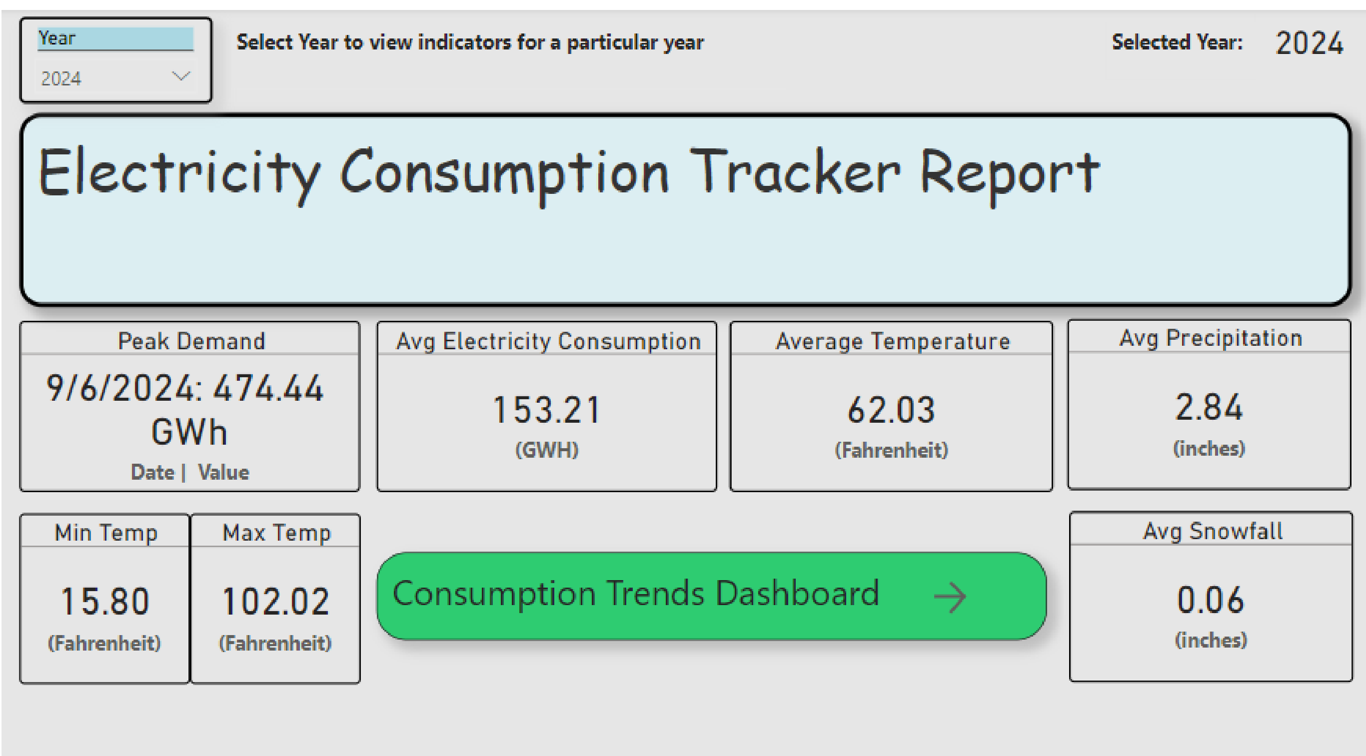


Fig. 9: System Diagram

- Key Metrics displayed:
  - Peak Demand displaying Date and respective value for the selected year like for year 2024 Peak Demand was observed on the day of 6th September 2024 with consumption accounting for 474.44 GWH.
  - Average Electricity Consumption (153.21 GWH) for the current year.
  - Also, it displays other supporting parameters such as Averages of Temperature, Precipitation, Snowfall along with Minimum and maximum temperature observed in that year.
- Electricity Consumption over Years:
  This page reveals trends in Electricity Consumption across time periods with first chart categorizing data based on Peak and Off-Peak Period Types. The second chart displays line chart for predicted data of electricity for next 7 days based on Actual and Predicted values and getting updated with each refresh.
- This page consists of page navigator and slicers (Data Type, Year and Month) for filtering data. The second chart will remain unaffected irrespective of the selection of the values of these filters.

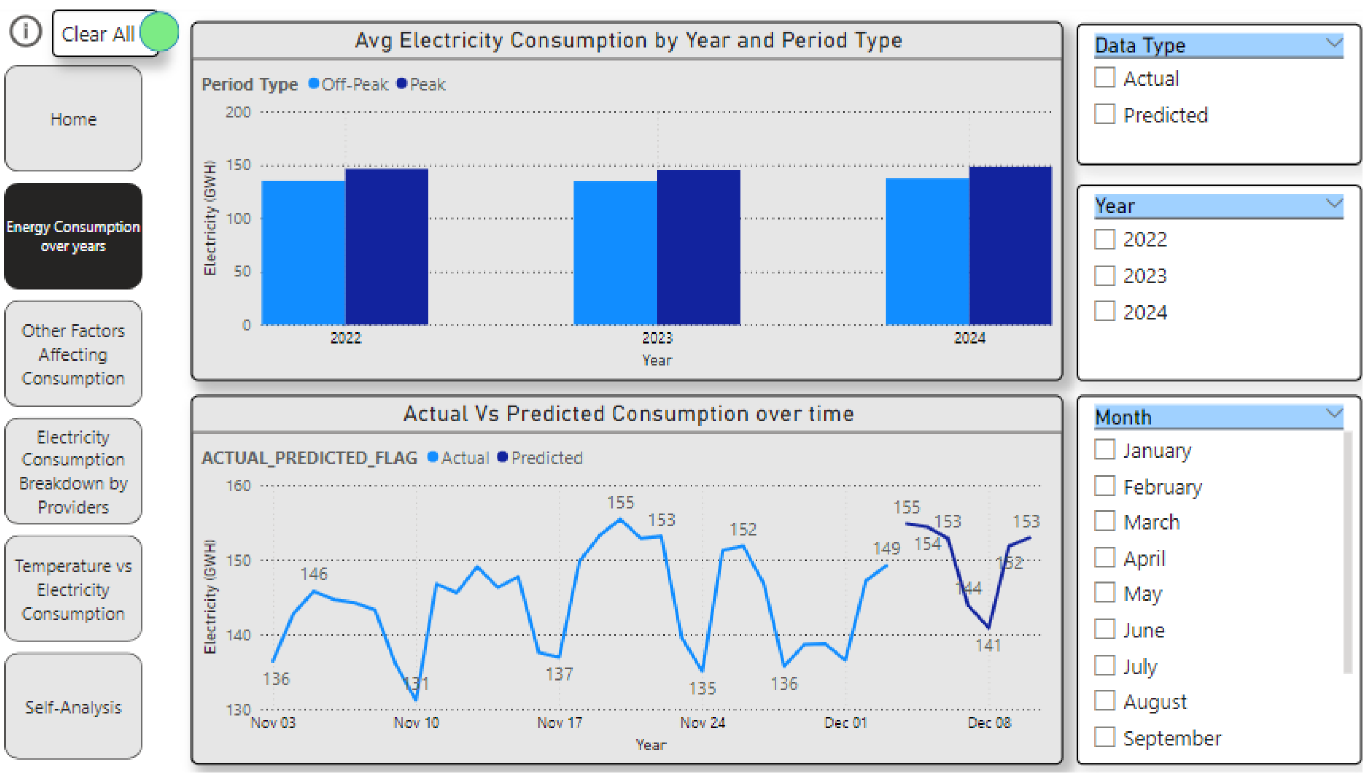


Fig. 10: System Diagram

- Other Factors Affecting Consumption:
  This page focuses on Electricity Consumption relation with weather dataset's parameters such as Average Temperature, Wind speed, Snowfall and Precipitation using scatter plot along with Page Navigator and Year filter.
  Key Observations: From these visuals, it can be clearly said that Temperature is the major factor impacting Consumption followed Wind speed and few impact by Snowfall and Precipitation. Overall, it seems that with the increasing temperature, consumption is also increasing and with wind speed consumption is almost same however, a significant drop after the speed exceeds 40 m/sec.

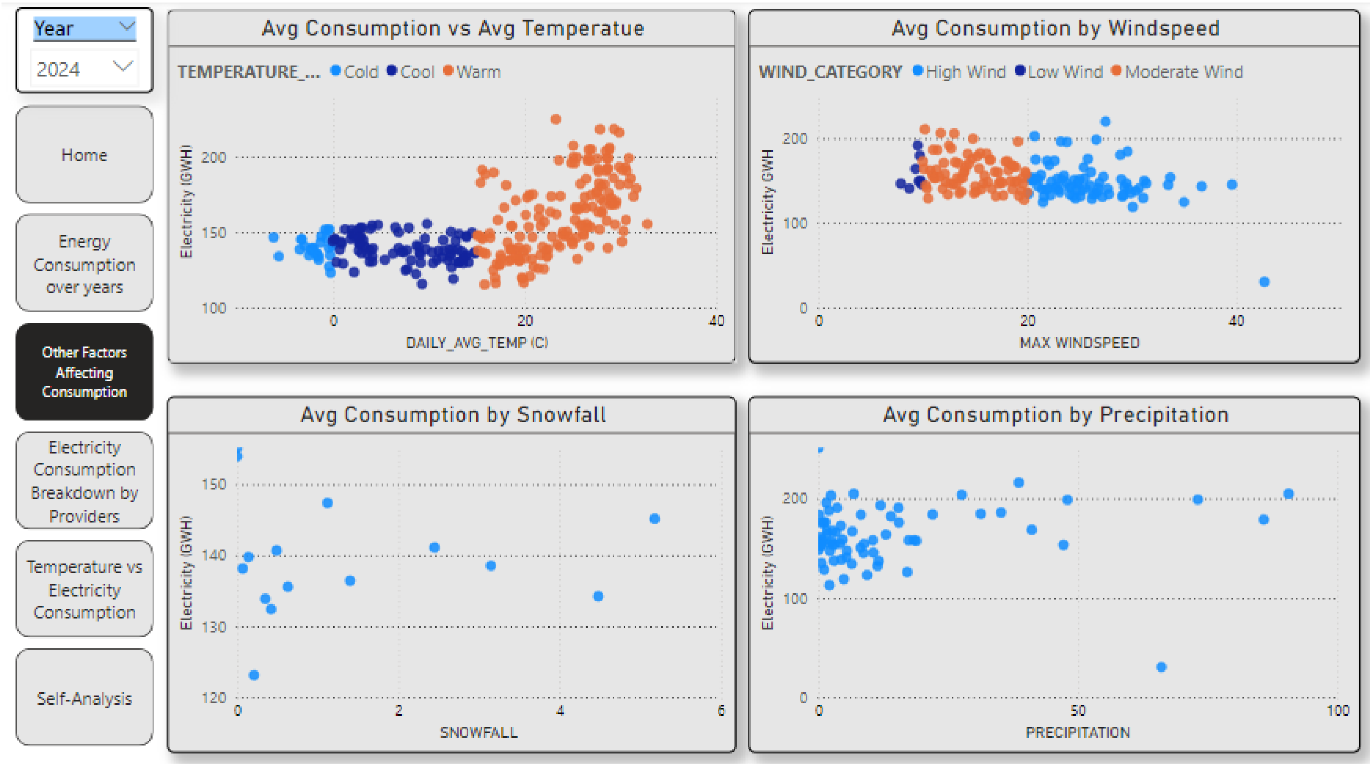


Fig. 11: System Diagram

- Electricity Consumption Breakdown by Providers:
  The major objective of this sheet is to display distribution of Electricity Distribution among major electricity distributors in California that are SCE (Southern California Edison), PGAE (Pacific Gas and Electric), SDGE (San Diego Gas and Electric) and VEA (Valley Electric Association).

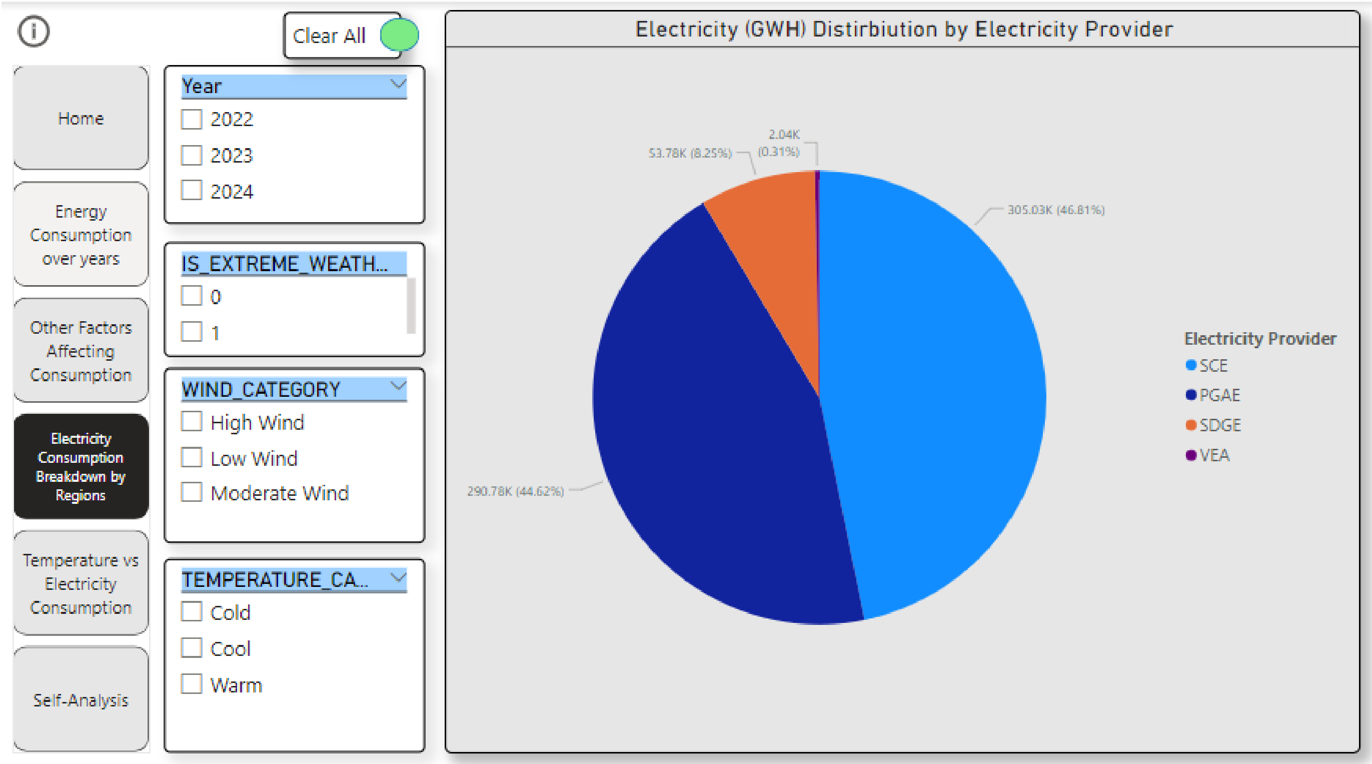


Fig. 12: System Diagram

The report consists of pie chart displaying distribution of electricity providers which reveals that the majority distribution is handled by two companies i.e. SCE and PGAE whereas VEA only serving rural areas.
Along with the visualization, there are several slicers available to filter the result based on the values of Year, IS_EXTREME_WEATHER (for filtering based on the classification of Extreme Weather day flag), Wind and Temperature Category. Also, in the left side, Page Navigator is available to display the title and link to other pages.

- Temperature vs Electricity Consumption:
  This page demonstrates the trends in electricity consumption in relation with average temperature and snowfall with slicer for year and page navigator.
  Key Observations: It seems in Quarter-3 i.e. from June to August electricity consumption is more as compared to other months due to hot summers and snowfall is playing a very less role here in California showing an increase in February and very less in March and to zero afterwards until December.

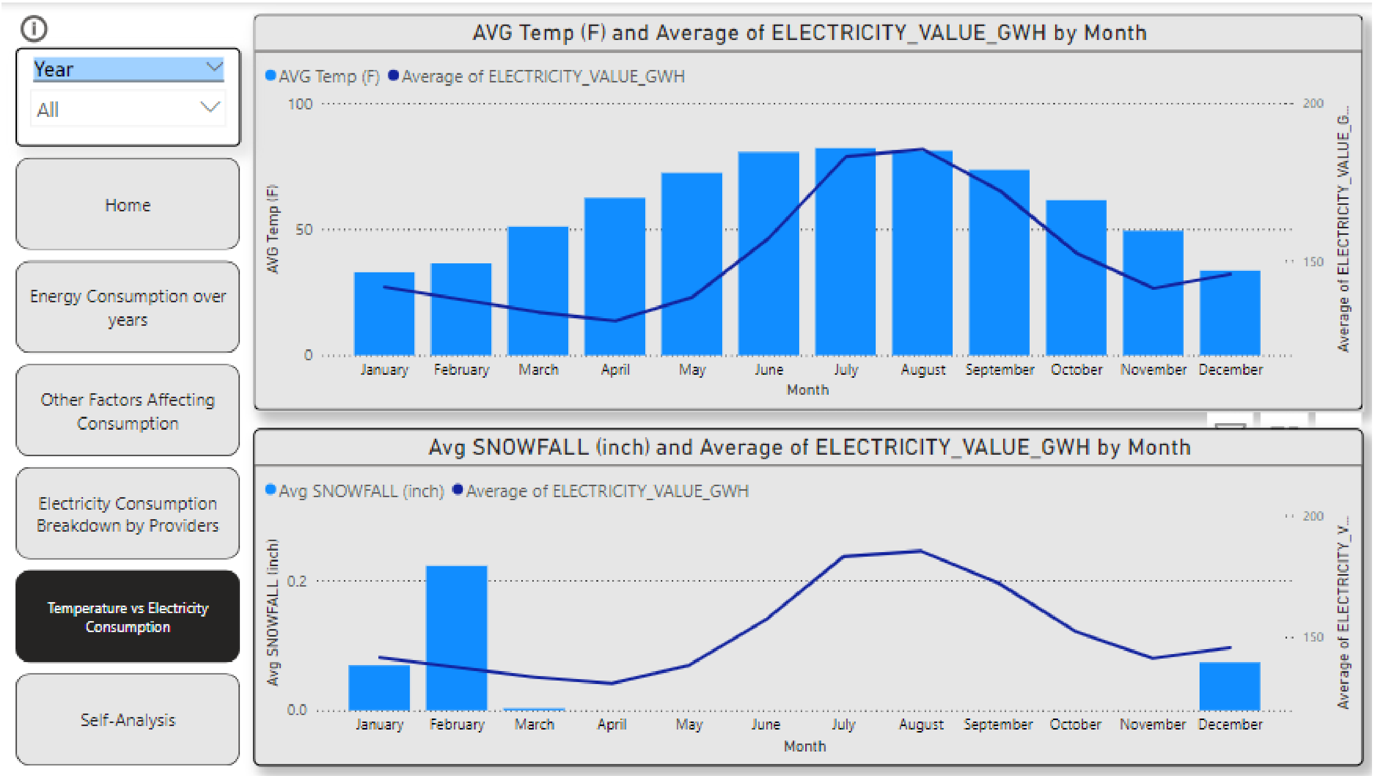


Fig. 13: System Diagram

- Self-Analysis:
  This page consists of self-service report leveraging Power BI's feature of analyzing trends in consumption data and accounts for several key influencers in the data.
  For example, analysis is being made for consumption to increase which explains electricity provider to be SCE with increase in 1.1k and PGAE 1.02k and Period Type is Peak.

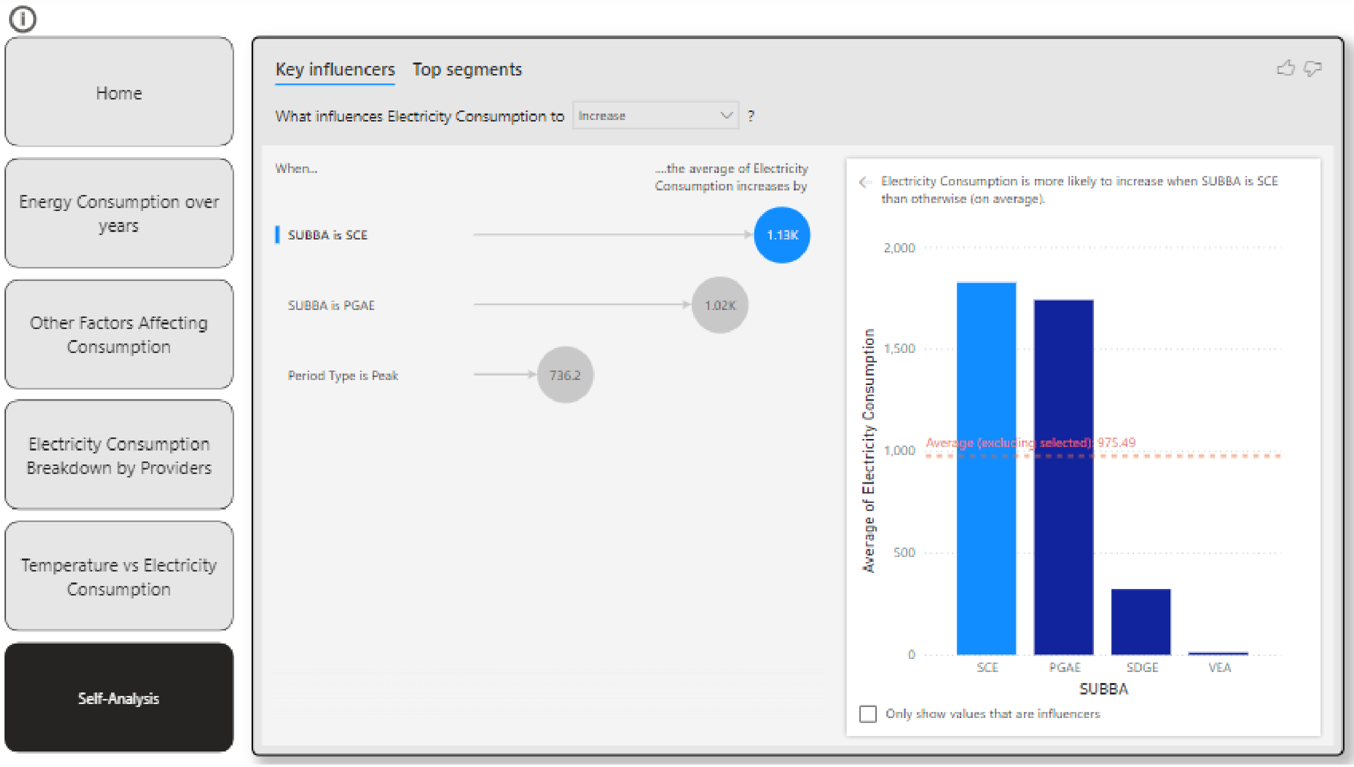


Fig. 14: System Diagram

Similarly, analysis can be made for consumption to look for data points to decrease.

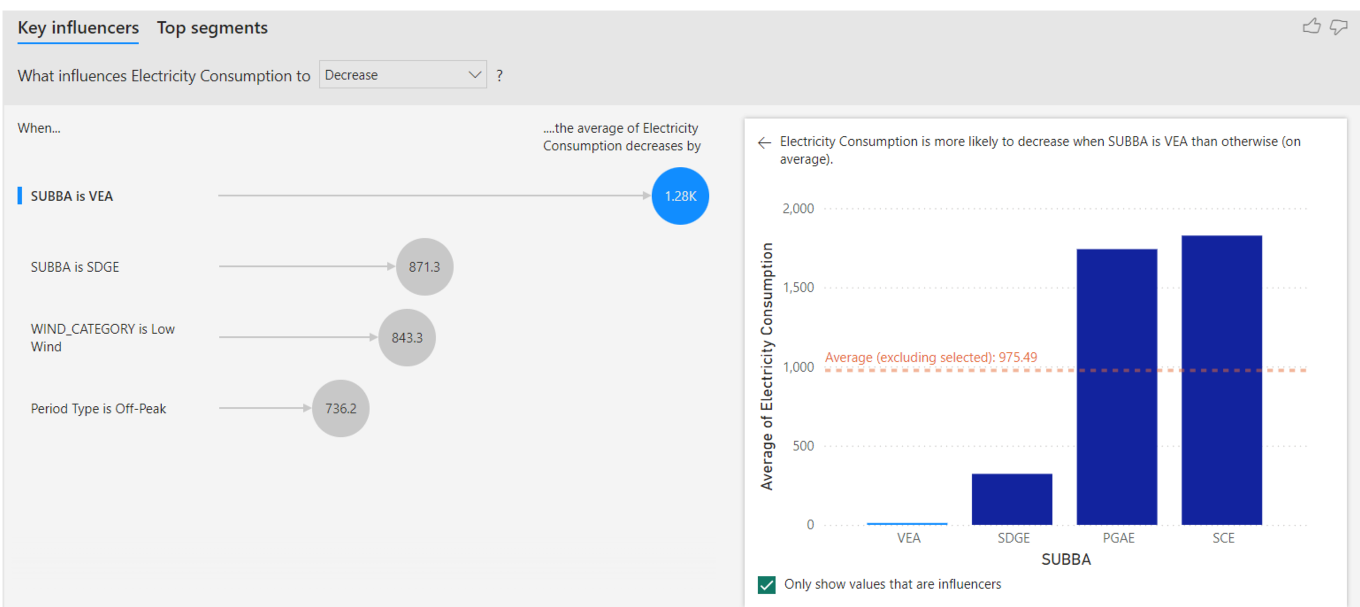


Fig. 15: System Diagram

Also, along with key influencers, it gives visibility on Top segments like for checking the segment when electricity consumption was 12.21, what were the value of data having that value.

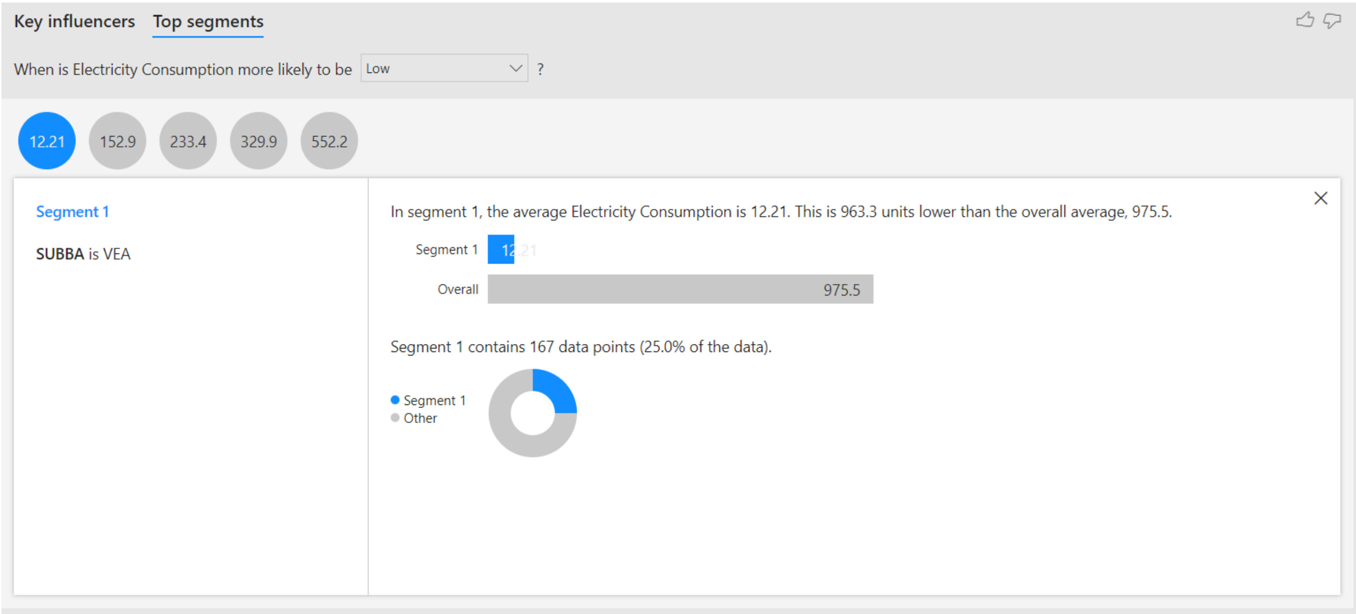


Fig. 16: System Diagram

Similarly, it can be used to see extreme values in the dataset like for the higher consumption value.

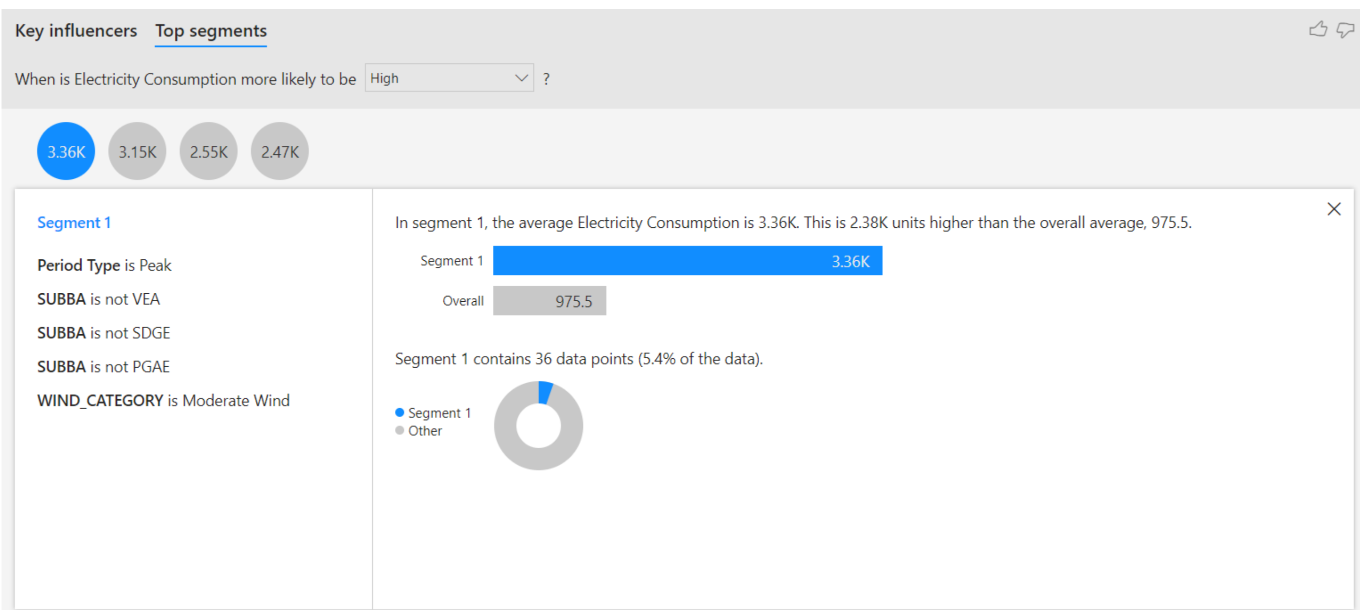
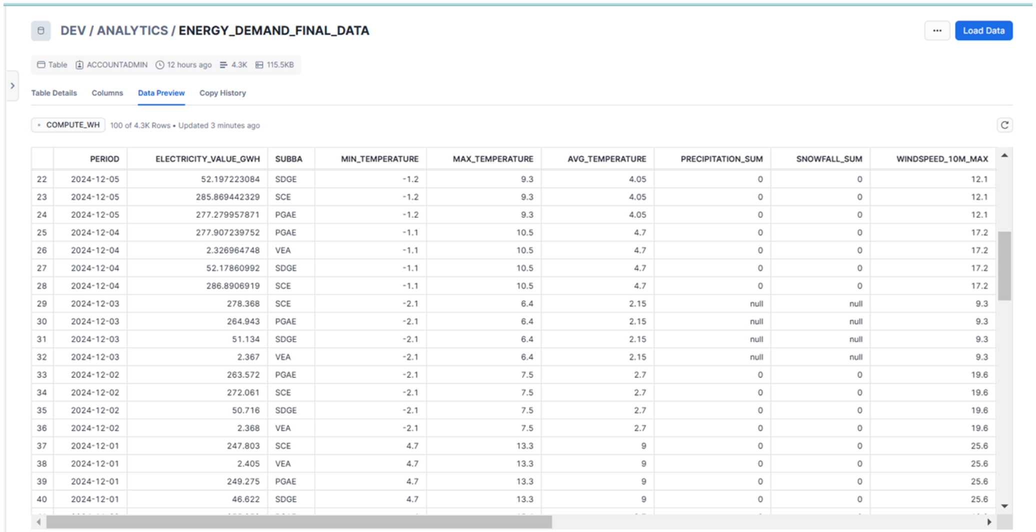


Fig. 17: System Diagram

IV. Output Tables



DEV / ANALYTICS / ENERGY_DEMAND_FINAL_DATA

Load Data

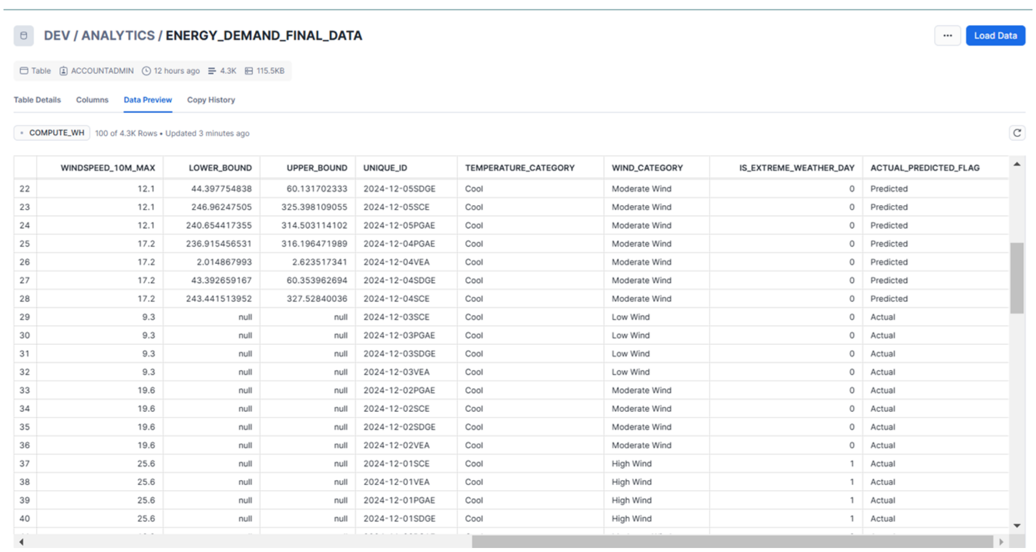
Table · ACCOUNTADMIN · 12 hours ago · 4.3K · 115.5KB

Table Details · Columns · Data Preview · Copy History

COMPUTE_WH · 100 of 4.3K Rows • Updated 3 minutes ago

	PERIOD	ELECTRICITY_VALUE_GWH	SUBBA	MIN_TEMPERATURE	MAX_TEMPERATURE	AVG_TEMPERATURE	PRECIPITATION_SUM	SNOWFALL_SUM	WINDSPEED_10M_MAX
22	2024-12-05	52.197223084	SDGE	-1.2	9.3	4.05	0	0	12.1
23	2024-12-05	285.869442329	SCE	-1.2	9.3	4.05	0	0	12.1
24	2024-12-05	277.279957871	PGAE	-1.2	9.3	4.05	0	0	12.1
25	2024-12-04	277.907239752	PGAE	-1.1	10.5	4.7	0	0	17.2
26	2024-12-04	2.326964748	VEA	-1.1	10.5	4.7	0	0	17.2
27	2024-12-04	52.17860992	SDGE	-1.1	10.5	4.7	0	0	17.2
28	2024-12-04	286.8906919	SCE	-1.1	10.5	4.7	0	0	17.2
29	2024-12-03	278.368	SCE	-2.1	6.4	2.15	null	null	9.3
30	2024-12-03	264.943	PGAE	-2.1	6.4	2.15	null	null	9.3
31	2024-12-03	51.134	SDGE	-2.1	6.4	2.15	null	null	9.3
32	2024-12-03	2.367	VEA	-2.1	6.4	2.15	null	null	9.3
33	2024-12-02	263.572	PGAE	-2.1	7.5	2.7	0	0	19.6
34	2024-12-02	272.061	SCE	-2.1	7.5	2.7	0	0	19.6
35	2024-12-02	50.716	SDGE	-2.1	7.5	2.7	0	0	19.6
36	2024-12-02	2.368	VEA	-2.1	7.5	2.7	0	0	19.6
37	2024-12-01	247.803	SCE	4.7	13.3	9	0	0	25.6
38	2024-12-01	2.405	VEA	4.7	13.3	9	0	0	25.6
39	2024-12-01	249.275	PGAE	4.7	13.3	9	0	0	25.6
40	2024-12-01	46.622	SDGE	4.7	13.3	9	0	0	25.6

Fig. 18: Final Energy Table: 1



DEV / ANALYTICS / ENERGY_DEMAND_FINAL_DATA

Load Data

Table · ACCOUNTADMIN · 12 hours ago · 4.3K · 115.5KB

Table Details · Columns · Data Preview · Copy History

COMPUTE_WH · 100 of 4.3K Rows • Updated 3 minutes ago

	WINDSPEED_10M_MAX	LOWER_BOUND	UPPER_BOUND	UNIQUE_ID	TEMPERATURE_CATEGORY	WIND_CATEGORY	IS_EXTREME_WEATHER_DAY	ACTUAL_PREDICTED_FLAG
22	12.1	44.397754838	60.131702333	2024-12-05SDGE	Cool	Moderate Wind	0	Predicted
23	12.1	246.96247505	325.398109055	2024-12-05SCE	Cool	Moderate Wind	0	Predicted
24	12.1	240.654417355	314.503114102	2024-12-05PGAE	Cool	Moderate Wind	0	Predicted
25	17.2	236.915456531	316.196471989	2024-12-04PGAE	Cool	Moderate Wind	0	Predicted
26	17.2	2.014867993	2.623517341	2024-12-04VEA	Cool	Moderate Wind	0	Predicted
27	17.2	43.392659167	60.353962694	2024-12-04SDGE	Cool	Moderate Wind	0	Predicted
28	17.2	243.441513952	327.52840036	2024-12-04SCE	Cool	Moderate Wind	0	Predicted
29	9.3	null	null	2024-12-03SCE	Cool	Low Wind	0	Actual
30	9.3	null	null	2024-12-03PGAE	Cool	Low Wind	0	Actual
31	9.3	null	null	2024-12-03SDGE	Cool	Low Wind	0	Actual
32	9.3	null	null	2024-12-03VEA	Cool	Low Wind	0	Actual
33	19.6	null	null	2024-12-02PGAE	Cool	Moderate Wind	0	Actual
34	19.6	null	null	2024-12-02SCE	Cool	Moderate Wind	0	Actual
35	19.6	null	null	2024-12-02SDGE	Cool	Moderate Wind	0	Actual
36	19.6	null	null	2024-12-02VEA	Cool	Moderate Wind	0	Actual
37	25.6	null	null	2024-12-01SCE	Cool	High Wind	1	Actual
38	25.6	null	null	2024-12-01VEA	Cool	High Wind	1	Actual
39	25.6	null	null	2024-12-01PGAE	Cool	High Wind	1	Actual
40	25.6	null	null	2024-12-01SDGE	Cool	High Wind	1	Actual

Fig. 19: Final Energy Table: 2

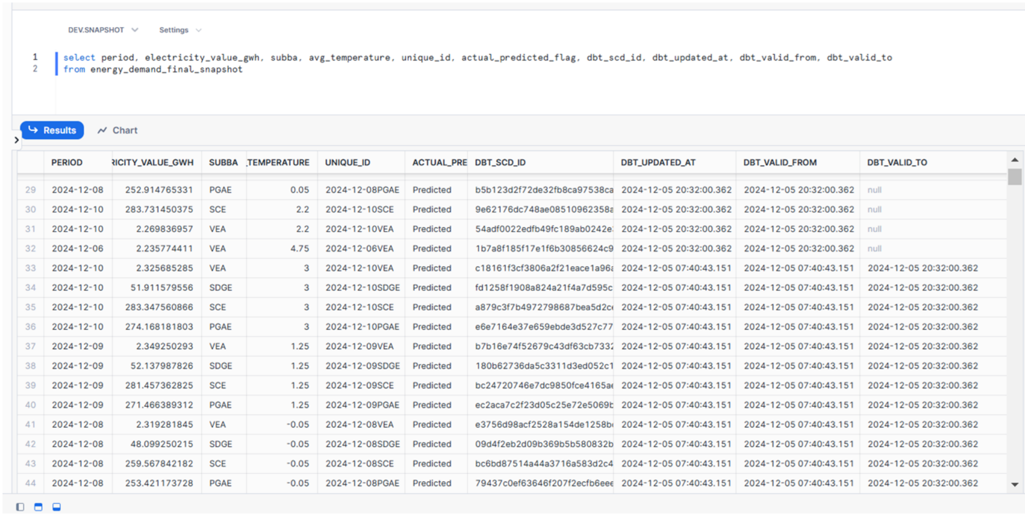
V. Code Repository

The entire code repository for this project can be accessed using the following link: https://github.com/aishanee-sinha/Energy_demand_fore

VI. Conclusion

This project integrates electricity consumption data from the U.S. Energy Information Administration (EIA) and weather data from Open Meteo to analyze and forecast energy demand trends effectively. By leveraging Snowflake for efficient data storage and management, advanced preprocessing techniques, and predictive modeling, the project delivers accurate forecasts of energy consumption. The inclusion of interactive Power BI dashboards provides actionable insights into energy trends, regional consumption patterns, seasonal variations, and the impact of weather on electricity usage.

These insights empower energy providers to optimize resource allocation, reduce operational costs, and implement demand-response strategies during peak periods or extreme weather events. The comprehensive approach ensures informed decision-making for improving grid reliability and efficiency.



DEV.SNAPSHOT Settings

```sql
select period, electricity_value_gwh, subba, avg_temperature, unique_id, actual_predicted_flag, dbt_scd_id, dbt_updated_at, dbt_valid_from, dbt_valid_to
from energy_demand_final_snapshot
```

Results Chart

	PERIOD	RICITY_VALUE_GWH	SUBBA	TEMPERATURE	UNIQUE_ID	ACTUAL_PRE	DBT_SCD_ID	DBT_UPDATED_AT	DBT_VALID_FROM	DBT_VALID_TO
29	2024-12-08	252.914765331	PGAE	0.05	2024-12-08PGAE	Predicted	b5b123d2f72de32fb8ca97538ca	2024-12-05 20:32:00.362	2024-12-05 20:32:00.362	null
30	2024-12-10	283.731450375	SCE	2.2	2024-12-10SCE	Predicted	9e62176dc748ae08510962358a	2024-12-05 20:32:00.362	2024-12-05 20:32:00.362	null
31	2024-12-10	2.269836957	VEA	2.2	2024-12-10VEA	Predicted	54adf0022edfb49fc189ab0242e	2024-12-05 20:32:00.362	2024-12-05 20:32:00.362	null
32	2024-12-06	2.235774411	VEA	4.75	2024-12-06VEA	Predicted	1b7a8f185f17e1f6b30856624c9	2024-12-05 20:32:00.362	2024-12-05 20:32:00.362	null
33	2024-12-10	2.325685285	VEA	3	2024-12-10VEA	Predicted	c18161f3cf3806a2f21eace1a96a	2024-12-05 07:40:43.151	2024-12-05 07:40:43.151	2024-12-05 20:32:00.362
34	2024-12-10	51.911579556	SDGE	3	2024-12-10SDGE	Predicted	fd1258f1908a824a21f4a7d595c	2024-12-05 07:40:43.151	2024-12-05 07:40:43.151	2024-12-05 20:32:00.362
35	2024-12-10	283.347560866	SCE	3	2024-12-10SCE	Predicted	a879c3f7b4972798687bea5d2c	2024-12-05 07:40:43.151	2024-12-05 07:40:43.151	2024-12-05 20:32:00.362
36	2024-12-10	274.168181803	PGAE	3	2024-12-10PGAE	Predicted	e6e7164e37e659ebde3d527c77	2024-12-05 07:40:43.151	2024-12-05 07:40:43.151	2024-12-05 20:32:00.362
37	2024-12-09	2.349250293	VEA	1.25	2024-12-09VEA	Predicted	b7b16e74f52679c43df63cb733	2024-12-05 07:40:43.151	2024-12-05 07:40:43.151	2024-12-05 20:32:00.362
38	2024-12-09	52.137987826	SDGE	1.25	2024-12-09SDGE	Predicted	180b62736da5c3311d3ed052c1	2024-12-05 07:40:43.151	2024-12-05 07:40:43.151	2024-12-05 20:32:00.362
39	2024-12-09	281.457362825	SCE	1.25	2024-12-09SCE	Predicted	bc24720746e7dc9850fce4165a	2024-12-05 07:40:43.151	2024-12-05 07:40:43.151	2024-12-05 20:32:00.362
40	2024-12-09	271.466389312	PGAE	1.25	2024-12-09PGAE	Predicted	ec2aca7c2f23d05c25e72e5069b	2024-12-05 07:40:43.151	2024-12-05 07:40:43.151	2024-12-05 20:32:00.362
41	2024-12-08	2.319281845	VEA	-0.05	2024-12-08VEA	Predicted	e3756d98acf2528a154de1258b	2024-12-05 07:40:43.151	2024-12-05 07:40:43.151	2024-12-05 20:32:00.362
42	2024-12-08	48.099250215	SDGE	-0.05	2024-12-08SDGE	Predicted	09d4f2eb2d09b369b5b580832b	2024-12-05 07:40:43.151	2024-12-05 07:40:43.151	2024-12-05 20:32:00.362
43	2024-12-08	259.567842182	SCE	-0.05	2024-12-08SCE	Predicted	bc6bd87514a44a3716a583d2c4	2024-12-05 07:40:43.151	2024-12-05 07:40:43.151	2024-12-05 20:32:00.362
44	2024-12-08	253.421173728	PGAE	-0.05	2024-12-08PGAE	Predicted	79437c0ef63646f207f2ecfb6ee	2024-12-05 07:40:43.151	2024-12-05 07:40:43.151	2024-12-05 20:32:00.362

Fig. 20: Final dbt Snapshot table

Future enhancements could include the integration of renewable energy data, such as solar and wind, to account for their growing influence on energy systems. Additionally, employing advanced machine learning techniques could further improve the accuracy of forecasts and enable more effective real-time energy management.

References

[1] DBT Integration with Airflow Available:https://docs.google.com/document/d/1NDo9FoG5JH-ssRB3G9xp6LDjEXxG9D2Nw4-T9oJdTAE/edit?tab=t.0.[Accessed: Nov. 07, 2024].

[2] K. Lee, "SJSU Data-226," GitHub. [Online]. Available: https://github.com/keeyong/sjsu-data226. [Accessed: Nov. 08, 2024].

[3] Apache Superset Documentation [Online]. Available: https://superset.apache.org/docs/intro/. [Accessed: Nov. 08, 2024].

[4] Apache Software Foundation, "Apache Airflow Documentation," Apache Airflow. [Online]. Available: https://airflow.apache.org/docs/apache-airflow/stable/index.html. [Accessed: Oct. 10, 2024].

[5] Snowflake Inc., "Snowflake Documentation," Snowflake. [Online]. Available: https://docs.snowflake.com/. [Accessed: Oct. 08, 2024].

[6] Dbt Documentation: https://docs.getdbt.com/