# CAPSTONE PROJECT

## Predicting Monthly Energy Needs for Trenton Falls, NY


### Abstract

Using a machine learning models and ETL pipeline built with the assistance of AI, the team analyzes historical weather patterns to predict energy needs.



Jake Halsey, David Haddad, Michelle Armstrong, Appal Swamy Badi, Michael Davis

**Executive Summary:**

**Requirements:**

The intention is to combine at least one (in this case, 2) archived (herein referred to as static) data sets with one live API to analyze energy consumption patterns**.**

**Business case:**

The electrical grid in the United States does not have significant energy storage capacity, and there is a cost to transporting and storing fuel for consumption. This creates a cost burden for energy companies that produce too little or too much energy in each time window.

**Proposed Solution:**

To facilitate the building of a future optimization model, we have designed and built an ETL pipeline and proof-of-concept prediction model for energy consumption requirements in Trenton Falls, NY. By Analyzing historical weather and energy consumption data, we attempt to use live weather data for zip codes in Trenton Falls, NY, to estimate the likely energy needs in the current month. Using a mature prediction model to create an optimization model, energy producers may be able to reduce storage and transportation costs.

**Key Findings:**

We find a strong, negative correlation (-0.83) between temperature and energy consumption, suggesting that colder months require significantly more energy than warmer months.

**Model Strengths:**

While temperature is not the only factor in determining usage, many other prediction models, such as predictions based on cell phone usage, cannot be adjusted by seasonal factors, affecting their reliability. We believe our model could be used to augment such models in the future, addressing the limitations of both models.

**Model Limitations:**

Few states track energy data below the state level, and only a limited number of weather stations provide local temperature records. Our model can predict energy needs for areas with both data points but not for those without. Correlations will likely vary by location, requiring the model to adjust accordingly. For example, while the current model suits New York's climate, a different pattern might emerge in warmer places, like Florida. Further research is needed to refine the model.

**Recommendations:**

To enhance consumption prediction, local weather and energy data should be closely monitored and integrated to expand the model to more locations. Where data exists, the model should augment current systems for greater accuracy. The final model should then be used to create an optimization model for utility production.

**Appendix:**

**Project Resources**

- **Static Data:**
    - Historical Energy Data: https://statics.teams.cdn.office.net/evergreen-assets/safelinks/1/atp-safelinks.html](https://catalog.data.gov/dataset/utility-energy-registry-monthly-zip-code-energy-use-beginning-2016
    - Historical Weather Data: https://www.climate.gov/maps-data/dataset/past-weather-zip-code-data-table
- **Live Data**
    - Weather API: https://api.weatherapi.com/v1/current.json?key=7c8218f18550417496b43123242902&q={area}
- **Competing Models (Cell phone data)**
    - https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-016-0075-3#:~:text=An%20accurate%20prediction%20of%20energy,allowing%20an%20efficient%20energy%20storage.

- **GitHub Repository Link:** https://github.com/jhalsey87/BigData_Capstone824/tree/main

- **Looker Studio Dashboard:** https://lookerstudio.google.com/u/0/reporting/00b216a9-b9e3-4ba4-a8a1-739f389aefa4/page/KAd8D

- **Presentation (Youtube):** https://www.youtube.com/watch?v=X1AMTUo97A0

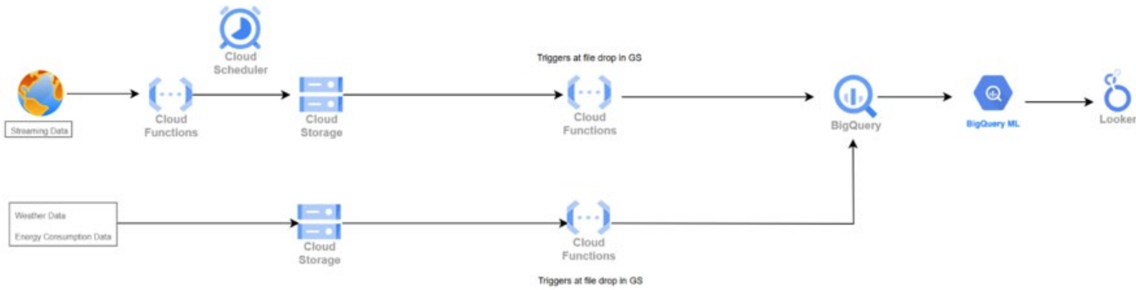
**Methodologies**

- Data Integration: Combined historical weather and energy consumption data with real-time weather information via a live API.

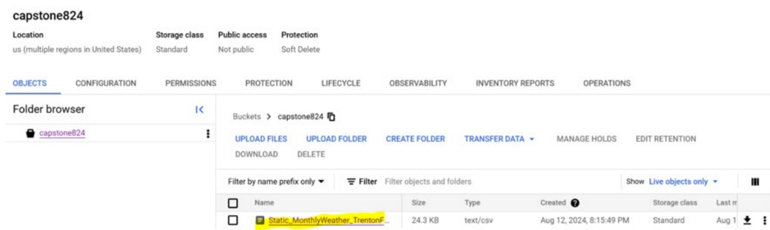
- Correlation Analysis: The relationship between temperature and energy consumption was assessed to identify actionable patterns.

- Predictive Modeling: Developed a proof-of-concept model to project energy consumption based on live weather inputs, serving as a foundation for enhancing future prediction accuracy.

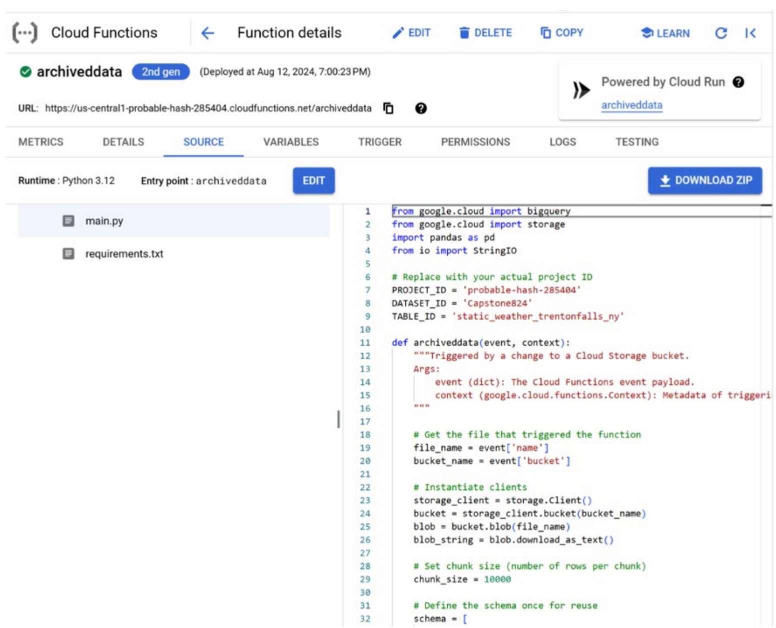
**Pipeline:**



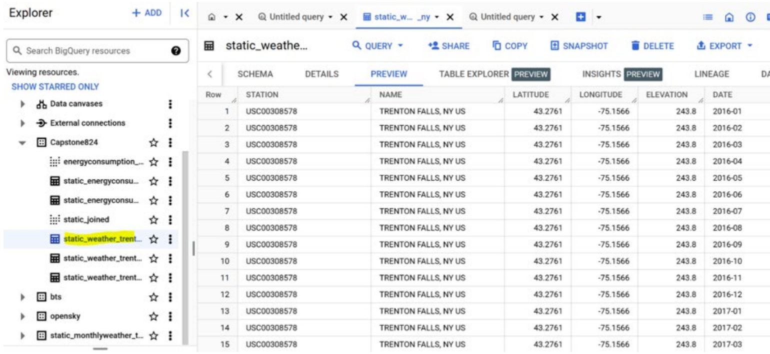

**Static Monthly Weather Archival Data Loaded Into Cloud Storage:**



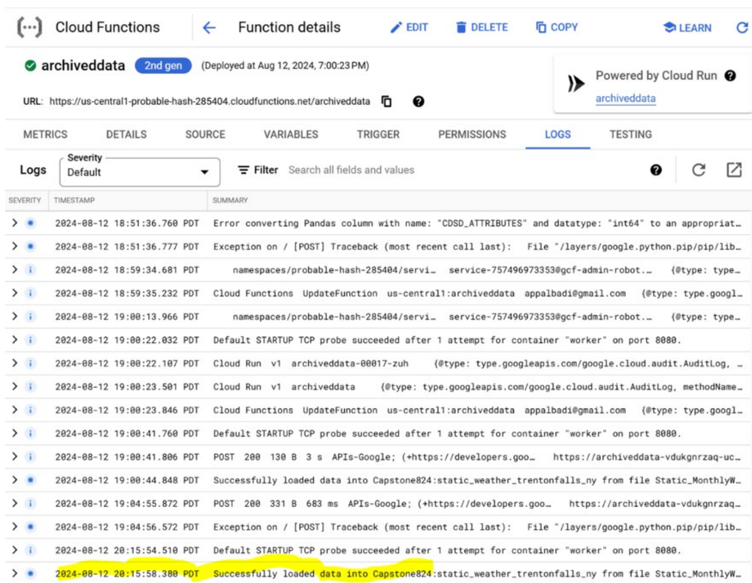

**Cloud Function for a One-Time Load of Data from Cloud Storage to Big Query Table:**



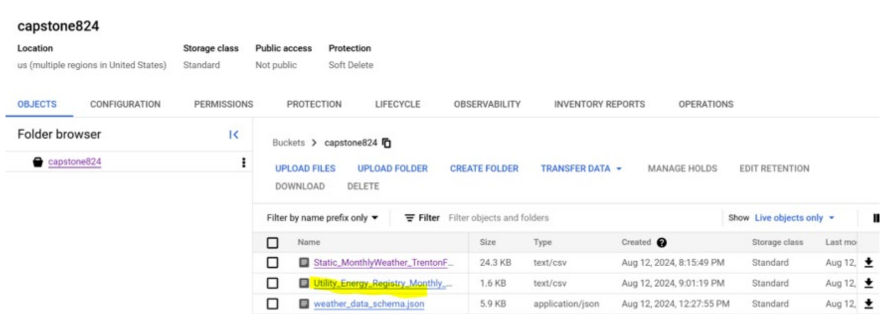
## Data Loaded Into Big Query Table:



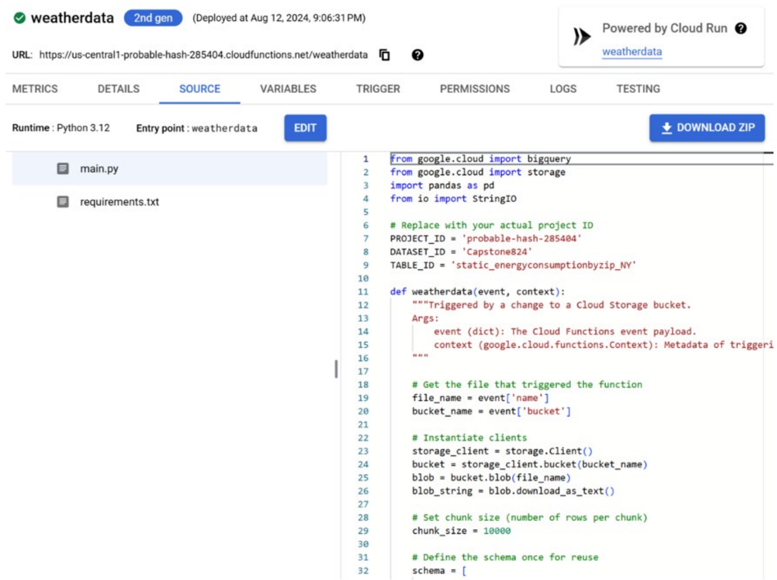

## Log Extract of Cloud Function Trigger from Cloud Storage File Upload:



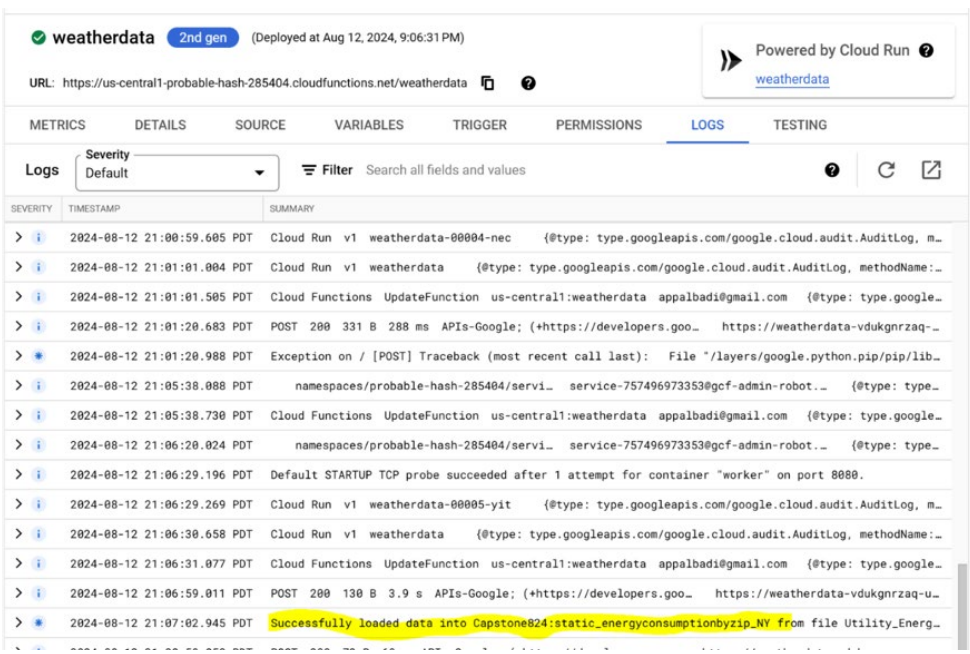

## Energy Archival Data Loaded Into Cloud Storage:



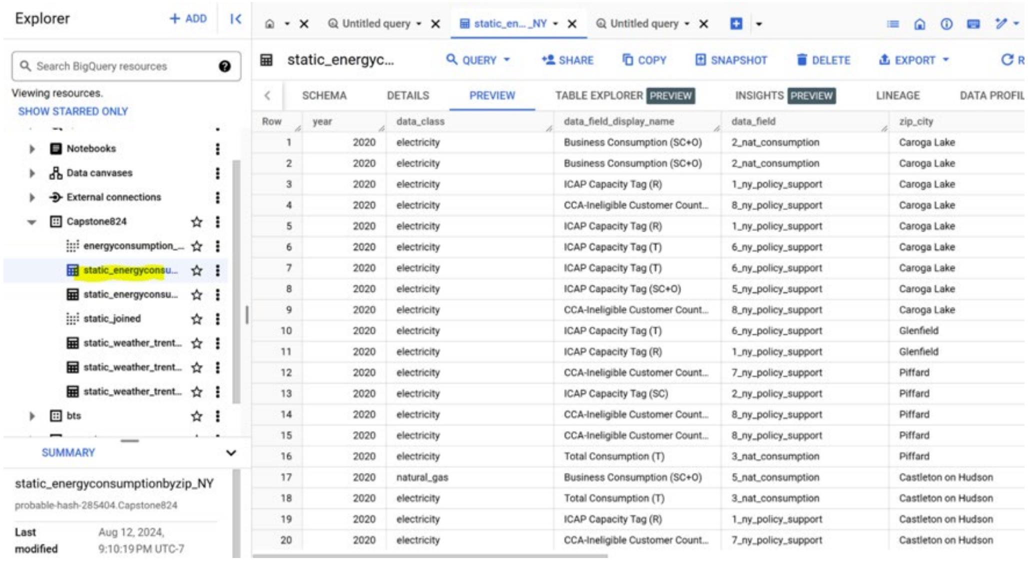
## Cloud Function For A One-Time Load Of Data From Cloud Storage To Big Query Table:



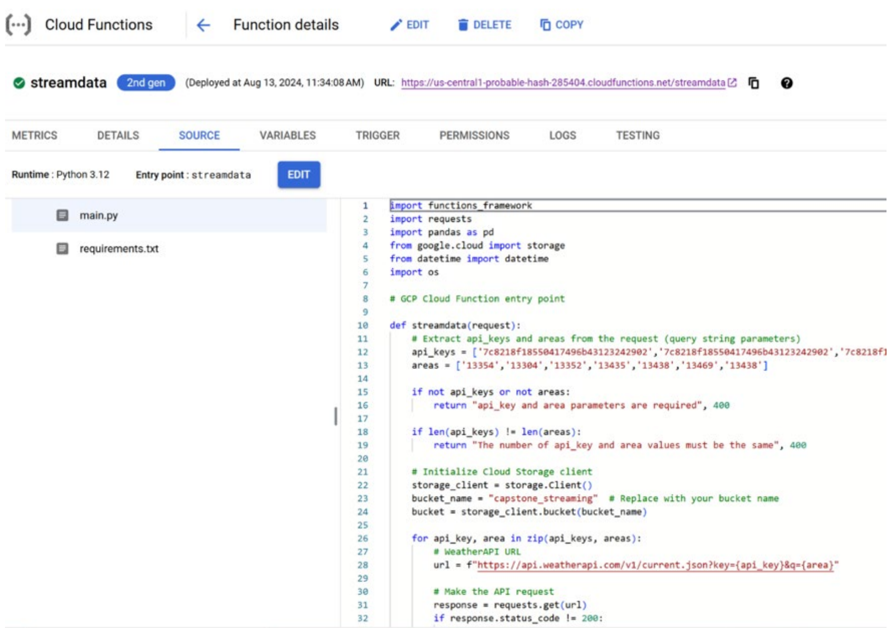

## Cloud Function Trigger Log:



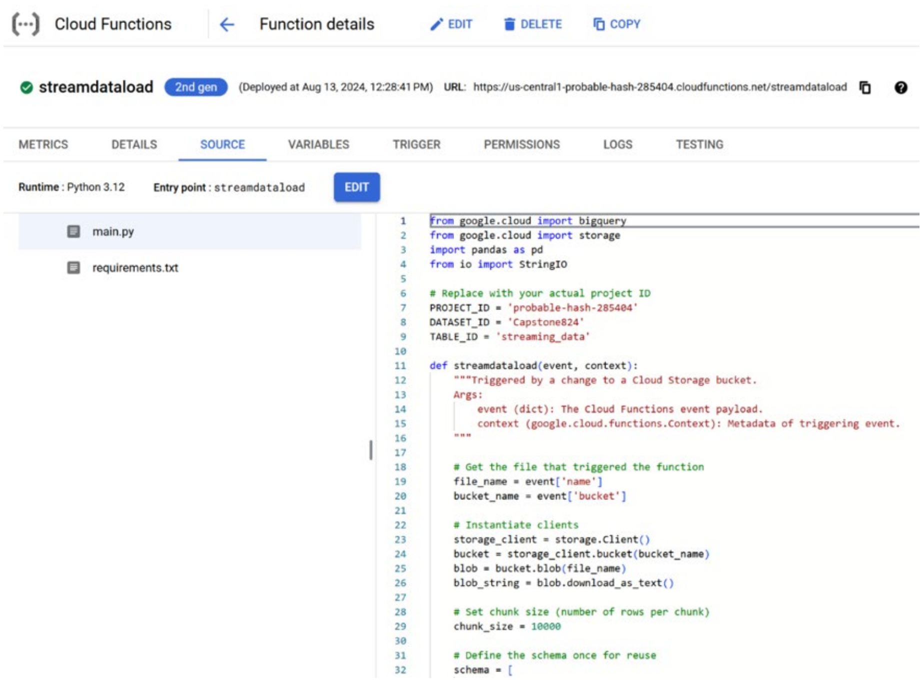
## Big Query Table Loaded Through Cloud Function Trigger:



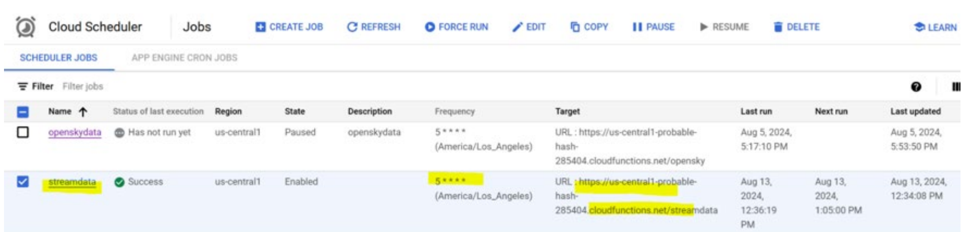

## Streaming Data Load Cloud Function to Call Apis and Load The Data Into Cloud Storage:



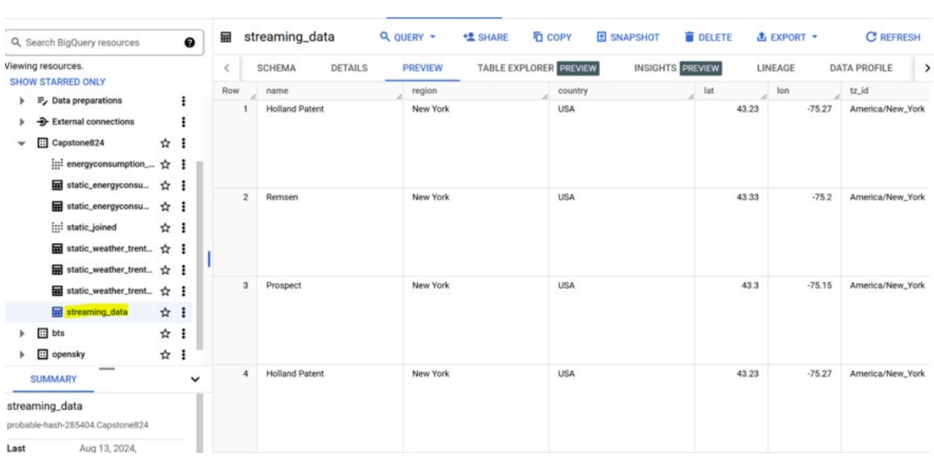
## Cloud Function to Load Data from Cloud Storage to Bigquery On The Cloud Storage Trigger:




## Scheduler to Run the API Call Function Every 5 Minutes:




## Big Query Table Loading Streaming Data



# Looker Dashboard (2 Pages)