Analytics of Electricity Usage in the United States

Dinghong Wang
New York University
dw2618@nyu.edu

Yitao Zhou
New York University
yz4171@nyu.edu

Qiyuan Huang
New York University
qh2086@nyu.edu

***Abstract*—In this project, an open-source dataset has been used to study the behavior of electricity usage and the potential of renewable energy across the U.S. The key results fit our hypothesis for electricity usage and also suggest the locations for utilizing a variety of renewable energy sources. Code for this project can be found here [1].**

I. Introduction

Electricity is a fundamental aspect of modern human society, powering everything from our homes and workplaces to our means of communication and transportation. It has revolutionized the way we live and work, providing us with a virtually limitless source of energy that we can use to power our daily activities and pursuits.

However, as the need for electricity grows every day and the need for carbon neutrality, we need to plan for better electricity usage and better utilizing emission-free energy. Existing technologies like Hadoop and Hive can be used to analyze the massive U.S. electricity usage dataset [2] provided by National Renewable Energy Laboratory, which is around 30 TB. In this project, we use around 300 GB of the dataset to get a better understanding of electricity usage at different granularity across different counties. We then analyze the potential of renewable energies across the U.S. mainly in terms of wind speed and direct normal irradiance.

II. Motivation

As more and more natural disasters occur and the macrogrid becoming more outdated, lots of the U.S. families are now considering microgrid, which is a smaller and more distributed version of a macro electrical system(macrogrid). A microgrid typically involves power supply/generation devices such as solar PV panels and wind turbines, a power storage device such as a battery and a power load device that provides services for human needs. The services can be provided for as electricity-consuming as data center, telecoms or can be provided just for personal electricity devices, such as electric vehicles.

The advantages of this kind of distributed system is actually quite similar to what we have learned in the lectures. This could lead to more resilience as well as potential income for example, through electricity arbitrage, for housing/building owners.

For example, U.S. residential utility customer uses an average of 886 kWh per month. On average, depending on location due to different sunlight received per unit area, a 10kW system will produce about 1255 kWhs of electricity. The extra electricity generated can be further sold back to the macrogrid or electrical companies during peak hours for profit. The solar panels can be costly as a 10 kWh cost \$20000 - \$25000. And this is the key motivation behind the project. By understanding the natural resources distribution as well as the electrical usage across the U.S., we will be able to know the locations and business opportunities for renewable energy generation devices such as solar panels and even a whole distributed electrical system(microgrid).

III. Datasets and Design Diagram

A. Counties Energy Consumption - 2018

The dataset reflects the sum of all profiles in a county by building type and end use with 15 minutes timeseries aggregation in 2018. This dataset has county Gisjoin as key, with columns such as building type, timestamp, total energy consumption, and detailed energy consumption by devices. It is extracted from End-Use Load Profiles for the U.S. Building Stock [3] (39.68 TB approx.). Sample rows are provided in Fig 1.

B. Weather - 2018

The dataset contains historical weather data of the U.S. in 2018 (2 GB approx.). This dataset has county Gisjoin as key, with columns such as temperature, humidity, wind speed, and direct normal irradiance. It is extracted from End-Use Load Profiles for the U.S. Building Stock [3] (39.68 TB approx.). Sample rows are provided in Fig 2.

C. Design Diagram

Our design diagram is shown in Fig 3. We first used MapReduce to extract columns we interested in and aggregate them from 15 mins granularity to 1 hour or even 1 day. We then built different customized writable for grouping the energy consumption with different key combinations. In addition to that, we applied combiner to accelerate the MapReduce jobs. We also used Hive to join weather data, Gisjoin identifiers(which are the keys in all of our datasets) with county name, and the energy consumption dataset to get one final table. And we finally utilized tableau to visualize our results.

county	in.building_type	timestamp	out.electricity.total.energy_consumption
G3600010	FullServiceRestaurant	2018-01-01 00:15:00	2019.12
G3600010	FullServiceRestaurant	2018-01-01 00:30:00	2063.83
G3600010	FullServiceRestaurant	2018-01-01 00:45:00	2111.76
G3600010	FullServiceRestaurant	2018-01-01 01:00:00	2233.22
G3600010	FullServiceRestaurant	2018-01-01 01:15:00	2091.89
G3600010	FullServiceRestaurant	2018-01-01 01:30:00	2084.37
G3600010	FullServiceRestaurant	2018-01-01 01:45:00	2096.55
G3600010	FullServiceRestaurant	2018-01-01 02:00:00	2105.31
G3600010	FullServiceRestaurant	2018-01-01 02:15:00	2093.84
G3600010	FullServiceRestaurant	2018-01-01 02:30:00	2133.0

Fig. 1. Counties Energy Consumption Dataset Snippet

date_time	Dry Bulb Temperature [°C]	Relative Humidity [%]	Wind Speed [m/s]	Wind Direction [Deg]	Global Horizontal Radiation [W/m2]	Direct Normal Radiation [W/m2]
2018-01-01 01:00:00	-19.4	54.51	2.1	350.0	0.0	0.0
2018-01-01 02:00:00	-20.0	56.33	1.05	350.0	0.0	0.0
2018-01-01 03:00:00	-22.2	63.93	0.0	299.17	0.0	0.0
2018-01-01 04:00:00	-22.2	63.93	3.1	320.0	0.0	0.0
2018-01-01 05:00:00	-21.4	61.04	2.85	330.0	0.0	0.0
2018-01-01 06:00:00	-21.7	60.89	3.6	320.0	0.0	0.0

Fig. 2. Weather 2018 Dataset Snippet

IV. Renewable Energy Calculation Methods

A. Wind

The physics behind wind energy calculation is quite simple. The amount of energy generated can be defined as:

$$\begin{aligned} W &= P \times t \\ &= \frac{1}{2} \times \rho \times A \times v^3 \end{aligned} \tag{1}$$

Here, W is the total work, P is power, t is time, ρ refers to the density of the air in kg/m^3, A is the cross-sectional area of the wind in m^2 and v is wind speed.

The wind value we defined later on is simply the average electricity price in each state, multiplies v^3 since energy is proportional to the cube of wind speed. The physical meaning behind this variable refers to the value generated by free wind energy in each county. This variable can be used to analyze the business potential of building wind turbines in different area.

B. Solar

The solar energy calculation can be quite complicated as factors like angles, longitude and altitude can influence the result. A good estimation is that solar energy is proportional to 75% of the direct normal irradiance, which is measured at the surface of the Earth at a given location with a surface element perpendicular to the Sun.

The solar value variable is also defined as the average electricity price in each state, multiplies direct normal irradiance. This variable can be used to analyze the business potential of building solar panels or solar power thermal plants.

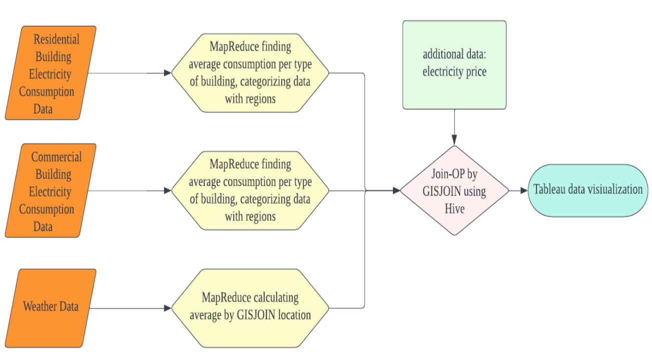

Fig. 3. Different Types of Residents' Hourly Electricity Usage

V. Results

A. Influence of Temperature on Electricity Usage

We first join U.S. res stock, U.S. com stock and weather data together to analyze the influence of temperature on electricity usage. Figure 4 shows how New York State resident's electricity usage is in proportional to temperature. The X-axis is a 2 Degree Celsius interval, and the Y-axis is the average electricity usage. Each bar is categorized by building types. From this graph, we learned that when temperature reaches peak in a year, the electricity usage also records the highest.

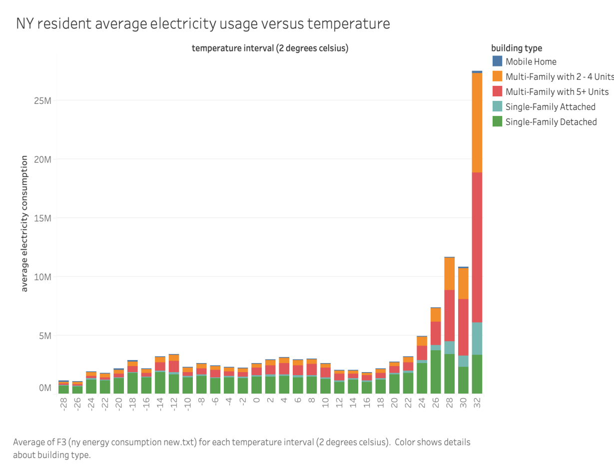

Fig. 4. NY Resident Average Electricity Usage Versus Temperature

Figure 5 illustrates how commercial electricity usage is relevant to temperature. As we can observe, comparing with residential usage, commercial usage increases gently when temperature rises, however, the peak usage is still recorded during the highest temperature season.

B. Electricity Usage Peak Time In a Day

We analyzed the peak time of electricity usage of U.S. residents in a day and we did a comparison between the resident's usage and commercial usage. From Fig 6, we observed that the usage peak for residents happens around 3 pm to 8 pm, around the time when people are getting back home from work.

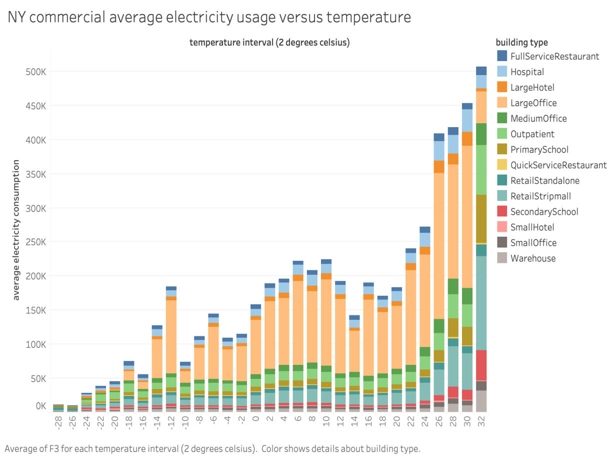

Fig. 5. NY Commercial Average Electricity Usage Versus Temperature

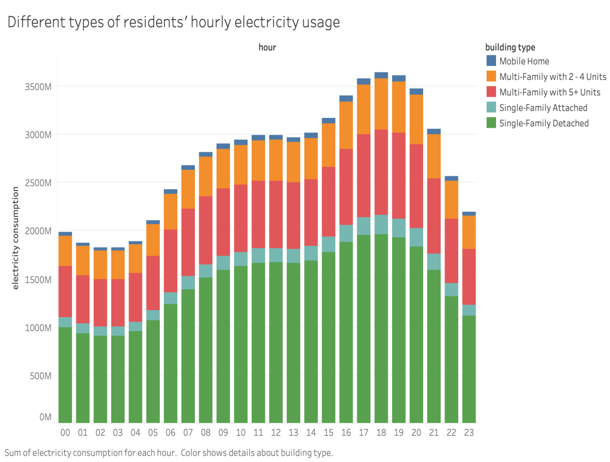

Fig. 6. Different Types of Residents' Hourly Electricity Usage

Contrarily, the commercial buildings have high electricity in day time, from 9 am to 5 pm, as demonstrated in Fig 7.

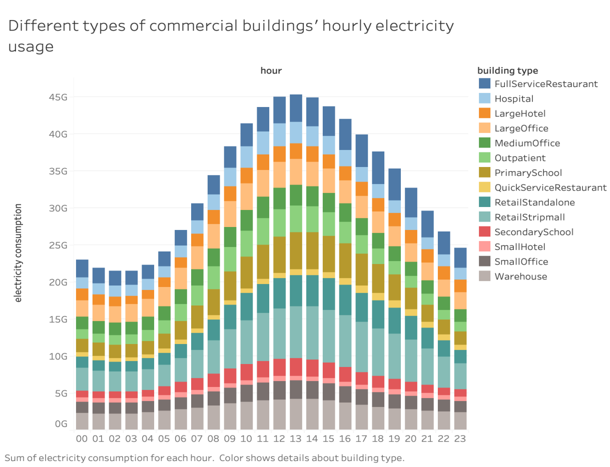

Fig. 7. Different Types of Commercial Buildings' Hourly Electricity Usage

From the analysis in section V-A and V-B, we can infer that the government agencies should be prepared for the supplies during peak times, especially during high temperature seasons and the daylight times in each day.

C. Clean Energy across the U.S.

We continued to assess the benefit that clean energy could bring to us, and we found that solar energy and wind energy can be best utilized at certain locations across the United States to create complimentary electricity supply.

Solar energy can be deducted from direct normal irradiance as described in IV-B. In Figure 8, we can observe that the irradiance is high in the southwest part of the United States. To make our analysis more precise, we have taken into account the influence that state electricity price could bring. With account of electricity price, as demonstrated in Figure 9, we recognized that it is best to build solar panels in California and northeast part of the United States.

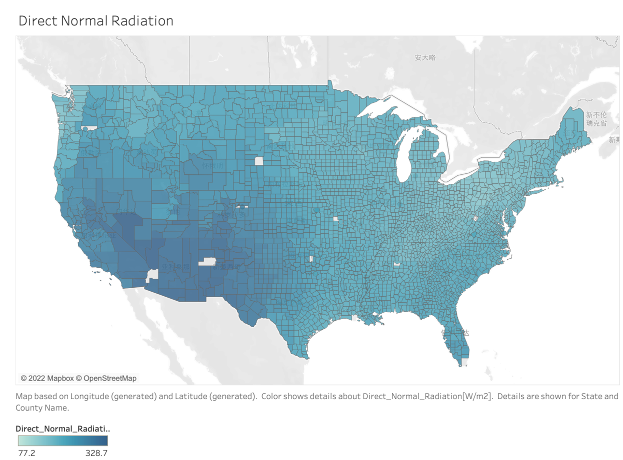

Fig. 8. Direct Normal Irradiance

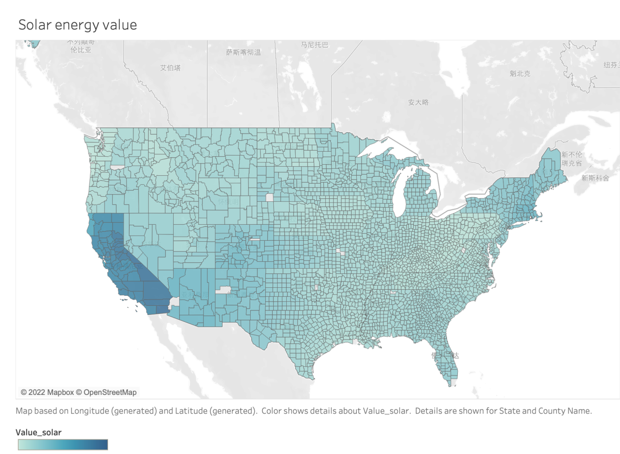

Fig. 9. Solar Energy Value

Part IV-A described the approach that we took to measure the wind energy. Figure 10 shows that wind speed in the central U.S. is of the highest compared to the two coasts. The conclusion is the same after we introduce the electricity price, as shown in Figure 11.

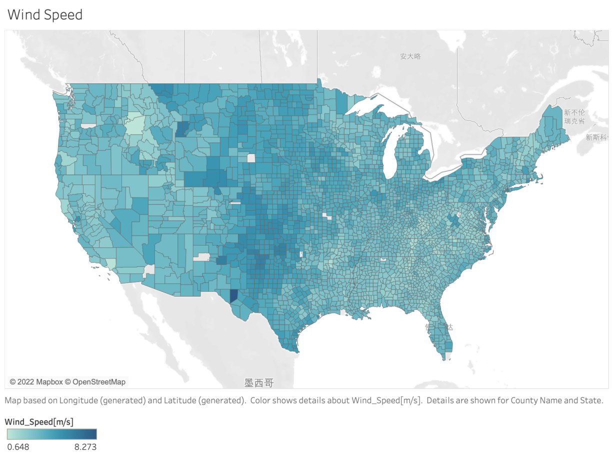

Fig. 10. Wind Speed

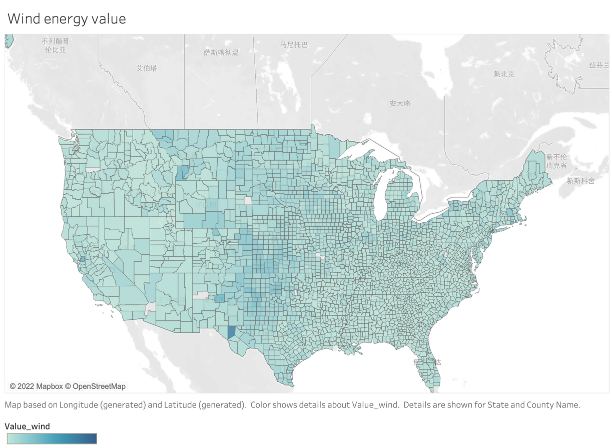

Fig. 11. Wind Energy Value

VI. Future Work

Our current results can be further combined with economics data, such as local average income, housing types and area etc., of local regions to further analyze the potential of solar panels and wind turbines.

VII. Acknowledgment

We would like to thank National Renewable Energy Laboratory (NREL) for providing the dataset and NYU Dataproc for providing the computational resources. We are also grateful to professor Yang Tang for his lectures on Realtime Big Data Analytics.

References

[1] https://github.com/dwang48/Realtime-Big-Data-Analytics.

[2] E. J. H. Wilson, A. Parker, A. Fontanini, E. Present, J. L. Reyna, R. Adhikari, C. Bianchi, C. CaraDonna, M. Dahlhausen, J. Kim, A. LeBar, L. Liu, M. Praprost, L. Zhang, P. DeWitt, N. Merket, A. Speake, T. Hong, H. Li, N. Mims Frick, Z. Wang, A. Blair, H. Horsey, D. Roberts, K. Trenbath, O. Adekanye, E. Bonnema, R. El Kontar, J. Gonzalez, S. Horowitz, D. Jones, R. T. Muehleisen, S. Platthotam, M. Reynolds, J. Robertson, K. Sayers, and Q. Li, "End-use load profiles for the u.s. building stock: Methodology and results of model calibration, validation, and uncertainty quantification." [Online]. Available: https://www.osti.gov/biblio/1854582

[3] P. A. F. A. P. E. R. J. A. R. B. C. C. C. D. M. K. J. L. A. L. L. P. M. W. P. Z. L. D. P. M. N. S. A. H. T. L. H. M. F. N. W. Z. B. A. H. H. R. D. T. K. A. O. B. E. E. K. R. G. J. H. S. J. D. M. R. P. S. R. M. R. J. S. K. Wilson, Eric and Q. Li, "End-use load profiles for the u.s. building stock."