Using Data Analysis to Examine Electricity Demand and Renewable Energy in Southern Africa

Tiana Curry

Environmental Studies, University of California, Santa Barbara

Abstract

Access to electricity is essential for development and economic growth. This study explores the potential for renewable energy to provide affordable energy and to alleviate energy poverty in the southern African region. Data analysis and visualization can help us understand the trends in, and characteristics of, energy demand and generation and draw insights into energy inequity across southern Africa. Patterns of electricity demand dictate the scale of energy infrastructure investments required to meet that demand reliably. However, electricity demand forecasts, dictated by the country's paying capacity and international financial support, highlight the continuation of energy inequities across the southern African region well into the future. This electricity demand analysis project is part of a larger project examining cost-effective planning and implementation of renewable energy generation to provide reliable electricity to countries in southern Africa.

Introduction

As the world continues to grow in population and countries continue to become more industrialized, the conversation about electricity accessibility has become part of a broader conversation about equality, health, and environmental impacts. Electricity is still not easily accessible in many different parts of the world. Thirteen percent of the world's population, about 940 million people, do not have access to electricity (Ritchie, 2019). Many developing countries cannot meet modern standard needs for electricity; in other words, these countries are considered to be energy-poor regions. This paper is about our electricity demand analysis project and is part of a larger study that will explore how renewable energy can provide sustainable energy and alleviate energy poverty in southern Africa regions. In this project, we analyzed energy demand data to find statistical correlations and to determine the electricity generation capacity needed to provide reliable electricity in southern Africa regions. We also examined the relationship between energy access and human development to explore the impacts of energy poverty for countries in southern Africa including Eswatini, Lesotho, Namibia, Malawi, Mozambique, South Africa, Tanzania, Zambia, and Zimbabwe.

Methods

We collected energy demand data from various electricity companies in the southern Africa region, including Eswatini, Lesotho, Namibia, Malawi, Mozambique, South Africa, Tanzania, Zambia, and Zimbabwe. For the purpose of this paper, we will define energy demand as the energy consumption by human activity, that drives the whole energy system by influencing the total amount of energy used (Creds, 2020). The data we collected were unstructured energy demand data in the format of business Excel spreadsheets, which tracked energy usage measured in megawatts (MW) and energy supply measured in megawatt hours (MWh) over the years for each country. We also collected hydro energy data from Malawi, which showed the amount of hydro energy used from various hydropower dams in Malawi. The energy demand data were from multiple years (2016-2019). We decided to focus on the year 2018 and use the 2018 data for the electricity planning model in GridPath (GridPath, 2021). GridPath is an advanced open-source software electricity system modeling platform developed by several researchers, including my advisor Dr. Ranjit Deshmukh. The energy demand data sets are considered unstructured or "raw" because they are not in a format that can be used in a computer algorithm for data analysis and data visualizations. Each country's data collector used a different method to format their business spreadsheets, and as a result, each country's raw data were processed differently in order to create a uniformed system. More explicitly, the task was to structure and clean each data set using Python so they can be used for further analysis.

The process of cleaning and structuring data for further use is referred to as data wrangling (Trifacta, 2020). For our data wrangling process, we used a computer programming language called Python and used popular libraries in Python, such as Pandas and NumPy, to structure the data. The best structuring method for this project was to create data frames that could be used as input for an algorithm we created in Python. A data frame is defined as a table or as a two-dimensional array-like structure in which each column contains values of one variable, and each row contains one set of values from each column. Mathematically, we can think of data frames as matrices, where the columns are considered the vectors of a matrix, and the rows are considered the linear equations of these vectors. We can also think of data frames as a table, where each row is a separate entry, and the columns are the variables. For example, if we wanted to collect the age and class level of each student at UC Santa Barbara, each student would be listed as an entry (row), and their age and class level would be the variables (columns). It is important to understand how data frames work because they are structured based on the type of algorithm for which they will be used. Algorithms are instructions given to the computer to conduct a particular task, in this case, to conduct statistical analyses. We can think of algorithms as recipes for a particular dish or as a set of directions to a particular location. To optimize the process of data wrangling, we created algorithms, also called functions, for each country's data set, to create data frames in the same format.

We created eleven algorithms, one for each of the countries a part of our research, that outputs a data frame with the date, hour, and energy load for each hour in a year. We built functions for each country's 2018 energy demand and for the Malawi hydro demand data. Each country has a different method to organize its data, so as a result, we created a different function for each country to make our data frames. Some of the challenges we faced were dealing with missing data points in the raw data sets or incomplete spreadsheets. To overcome these challenges and make sure our functions were working properly, we filled missing values using the average with a process called mean substitution (Roy, 2020). In every function, we indicate any changes or substitutions made to account for the impact the substitutions may have on our final results. Once the errors were fixed, our functions output the energy demand data frames that consisted of columns for each hour in the year, the date, and the total energy load for each hour (energy load is the energy demand). By structuring the data frames in this way, we are able to calculate a statistical summary, create time-series, data visualizations, and analyze the data to make decisions.

Summary statistics and data visualizations are useful because they provide detailed information about our data without the need to look at complicated tables or data frames. Once the data frames were created for each country, we then decided to create data visualizations and calculate basic statistics to summarize the information we have for each country. The libraries in Python that were most useful to create the data visualizations were Plotly and Pandas (Wijaya, 2020). We created time series data visualizations for each country to show the amount of energy demand in a given hour for the year 2018. We created the peak demand curve data visualization to see the range of energy demand for each country. Knowing the energy peak demand is particularly important to establish the required generation for our renewable energy sources. By having the energy demand for a given year, we can calculate the total energy demand for the year and calculate the energy per capita value. We define energy per capita as the total energy used per person in a given country. To calculate the energy per capita, we added up the total energy demand in a year in kWh and divided it by the total population in the given country. We calculated the energy per capita for all nine countries in southern Africa, and we calculated the energy per capita for the United States (US) and California for comparison. We used the energy per capita values to create a bar plot to clearly show the energy access for countries in the southern Africa region. The US and California energy per capita were calculated to reference how much energy we have access to in the US compared to developing countries in southern Africa. The energy per capita visualization will be used to examine energy poverty across the southern Africa region.

Results

We used the energy demand data frames to create time series to show the amount of energy used for each hour in the year 2018. We can define time series as a series of data points ordered by time. The time-series starts from hour one corresponding to January 01, 2018, at 12 am, to hour 8760 corresponding to December 31, 2018, at 11 pm. If we look at Figure 1 called the Energy Demand Time Series plot for Lesotho in 2018, we see that it ranges from an energy demand of 5 MW to an energy demand of about 175 MW. In Figure 2 called the Energy Demand Time Series plot for South Africa in 2018, it ranges from an energy demand of 15,000 MW to about 36,000 MW. Notice the lower values are a result of lower energy demand, and there is a large difference between energy demand in Lesotho compared to the energy demand in South Africa. At a glance the time series plots are harder for untrained eyes to interpret, but they serve a great purpose for researchers, because we can see the range and consistency of energy demand over time

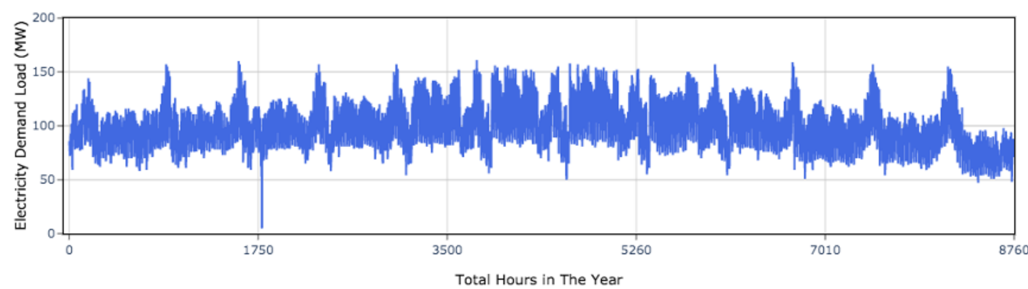

Figure 1. Energy Demand Time Series plot for Lesotho in 2018

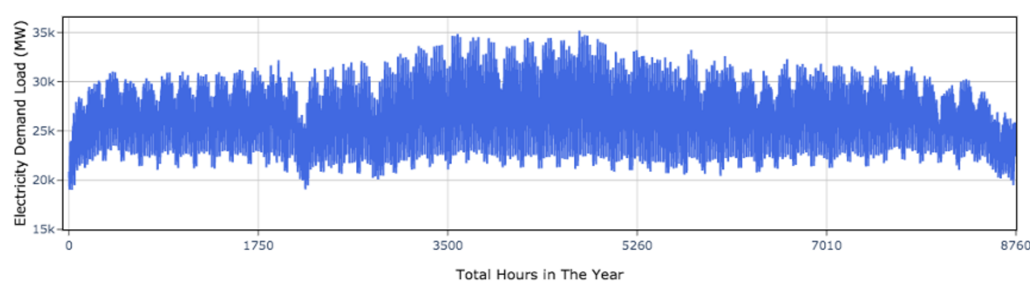

Figure 2. Energy Demand Time Series plot for South Africa in 2018

In Figure 3 called the Cumulative Energy Demand Curve, we have the energy demand for 2018 (in hours) for the nine countries located in sub-Saharan Africa. We notice there is a large variety of usage between the nine countries. By looking at the leftmost point in Figure 3, we can see the highest amount of energy used by each country in 2018. It is important to note that this graph is not a time series. In other words, the data is not ordered by time but rather it is ordered by highest energy demand to lowest energy demand. Looking at the rightmost point of Figure 3, we can see the lowest energy demand value for each country, where some countries have approached zero. Energy demand in the amount of zero could mean a country-wide electricity outage, or it could mean there are missing data. The leftmost point in Figure 3 is significant because this value indicates approximately the minimum amount of electricity generation capacity required to meet the electricity demand for each country. This is important because we want the amount of energy generated to be stable for every hour throughout the year. South Africa's cumulative energy demand curve is shown in Figure 4 because its energy demand is significantly greater than the other countries in the southern Africa region. If we were to represent South Africa in Figure 3 the other countries' energy demand curve would appear to be close to zero. By separating South Africa in Figure 4, we are able to better represent the energy demand for each country. Most of the countries in our study have similar energy demand results except for South Africa, because South Africa is economically more developed and has a significantly greater energy demand. These data visualizations are useful because they give a clear summary of energy demand for multiple countries. By using data visualizations, we can derive insights from our data without referring to the data frames. These figures are useful to determine how much energy is needed to meet the demand for each country throughout the year.

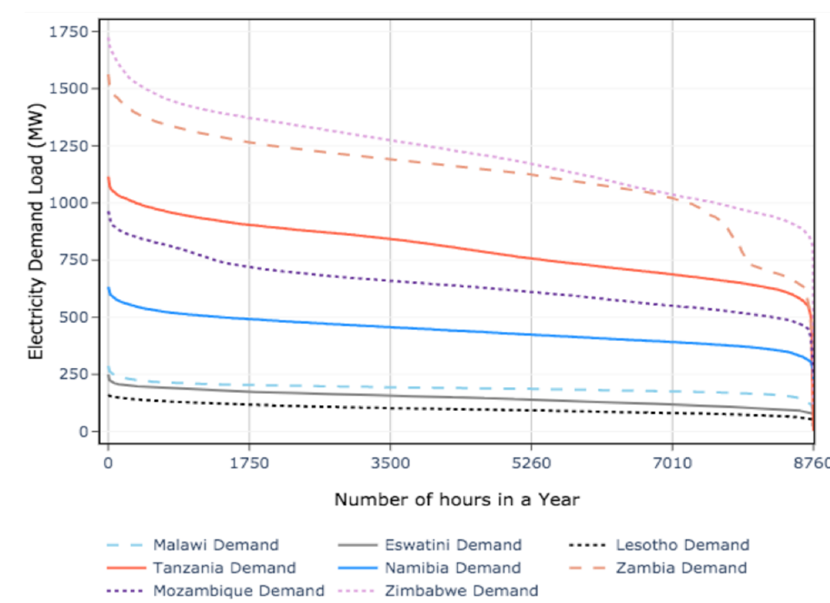

Figure 3. Cumulative Energy Demand curve for countries in Southern Africa in 2018

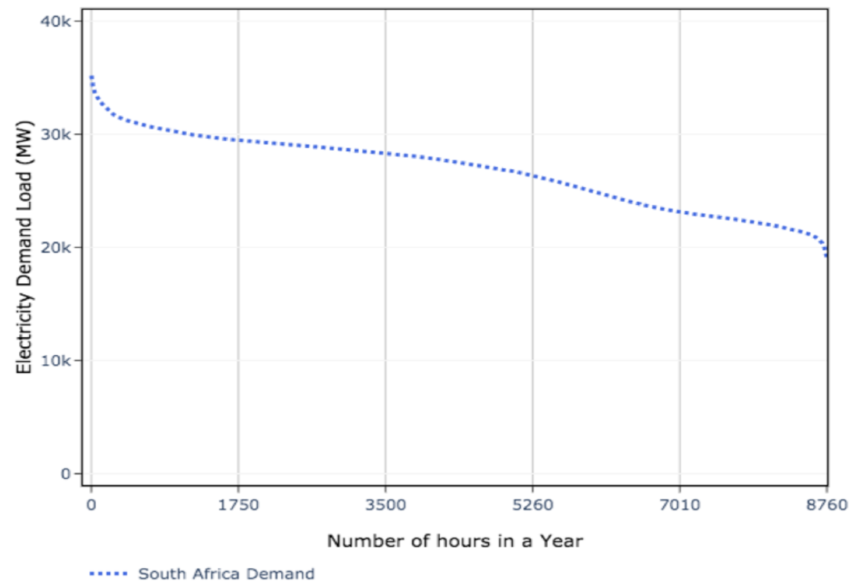

Figure 4. Cumulative Energy Demand curve for South Africa in 2018.

In Figure 5, called the Energy per Capita in Southern Africa for the year 2020 and 2040, we calculated the results using the data in, "Annual Reports," from the Southern African Power Pool database. These data were used to calculate the energy per capita for the nine countries in Southern Africa. Also, we used data from the EIA to calculate the energy per capita for the United States and California represented in the same plot. Figure 5 shows the energy per capita value in the year 2020 and 2040 for the nine countries in Southern Africa. We can see a large disparity between South Africa's per capita electricity consumption and that of other countries. We also see a significant difference in the expected growth for Malawi and Tanzania from the year 2020 to the year 2040.

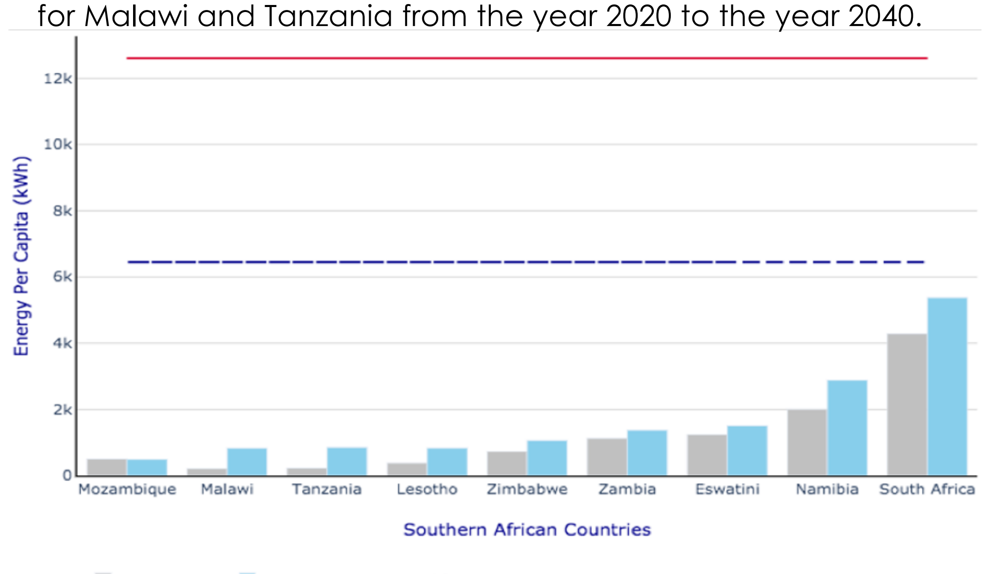

Figure 5. Energy Per Capita in Southern Africa for the year 2020 and 2040.

Discussion

Our results for energy per capita provide key insights into energy poverty and inequity in sub-Saharan Africa. To get a better understanding, in 2018, the energy per capita for the United States was 12,600 kWh, and in California it was 6,445 kWh (U.S. Energy Information Administration [EIA], 2020). This means that the state of California in the US currently has a higher electricity per capita than every country in the southern Africa region. In fact, most of the countries in our study are projected to reach less than ten percent of the United States electricity per capita by 2040. Our results highlight the inequity in electricity access in sub-Saharan Africa. Our next step is to run a simple linear regression (SLR) model on electricity per capita vs. the Human Development Index (HDI) for each country in southern Africa. We will apply a SLR model because we can determine if there is a correlation between electricity per capita and the HDI. This means we will use the SLR model to explore the relationship between the amount of energy a country uses and how much that country will develop.

This electricity demand analysis project is part of a larger project examining the cost-effective planning and implementation of renewable energy generation to provide reliable electricity to southern African countries. Our results for energy demand are essential to the larger project that focuses on the development of an electricity system planning model. We will continue with this research to develop the statistical summary for each time series and to create more informative data visualizations. With the time series we have created, we will develop heat maps to indicate energy demand per month and provide monthly averages for the given year. The remaining figures will be used to develop a statistical report that will be available to the public on the GridPath website. GridPath is advanced software that provides analytical power-system modeling for renewable energy generation development. The southern Africa data we have collected will serve as a model for other researchers to develop renewable energy generation systems across the world. We will also analyze the correlation between energy demand and wind and solar power generation, to determine the need for other generation resources. Our analysis will help us provide reliable electricity supply for sub-Saharan Africa. Our analysis will lead to a better understanding of the costs and benefits of developing renewable energy. This will help us to ensure a cleaner, sustainable, and affordable electricity system for the countries in sub-Saharan Africa.

About the Author

Tiana Curry is a graduating senior completing her degree in mathematics. Currently, she works as a research assistant through the Faculty Assistant Research Program, working with Dr. Ranjit Deshmukh in the environmental studies department at UC Santa Barbara. Tiana is also a transfer student and is a mentor in the Transfer Student Mentorship Program and was a student speaker at the Transfer Student Class 2020 Graduation. In 2019, Tiana participated in a summer internship at the Smithsonian Institution Data Science Lab and was featured in a Smithsonian article about her contributions to the American Women's History Initiative. After participating in this internship, she was inspired to continue her research in data science. She plans to participate in a summer internship with UC Santa Barbara and further her research experience with three research projects in the fall of 2020. In the fall of 2020, she plans to participate in three research projects related to renewable energy in southern Africa, environmental justice and COVID-19, and laser simulation at UC Santa Barbara. Her goal is to pursue a graduate degree with a focus in data science to become a data scientist.

Works Cited

"Annual Reports." (1970) Southern African Power Pool, www.sapp.co.zw/annual-reports

Creds. (2020). "What is Energy Demand?". www.creds.ac.uk/what-is-energy-demand/#:~:text=Energy%20demand%20is%20the%20term,use%20technologies%20that%20consume%20energy

"Population Division World Urbanization Prospects 2018". (2018) United Nations Department of Economic and Social Affairs, Population Dynamics [Data Query], population.un.org/wup/DataQuery/

Ritchie, Hannah. Roser, Max. (2019). "Access to Energy," Our World In Data, ourworldindata.org/energy-access#access-to-electricity

Roy, Baijayanta. (2020) "All About Missing Data Handling." Medium, towardsdatascience.com/all-about-missing-data-handling-b94b8b5d2184

Trifacta. (2020). "What is Data Wrangling?". www.trifacta.com/data-wrangling/

U.S. Energy Information Administration (EIA). (2020). "United States, Primary Energy." www.eia.gov/international/overview/country/USA