Identifying Investment Opportunities for Thermal Energy Storage

Chandler Angell

27 June 2020

[Article on Medium](https://medium.com/@chandlerangell/identifying-investment-opportunities-for-thermal-energy-storage-da10f7530c9d)

Table of Contents

[Project Overview 3](#_Toc44120294)

[Problem Statement 3](#_Toc44120295)

[Background 3](#_Toc44120296)

[Data Sources 5](#_Toc44120297)

[Methodology 7](#_Toc44120298)

[Scoring Model 7](#_Toc44120299)

[Metrics 9](#_Toc44120300)

[Data Preparation 9](#_Toc44120301)

[Extract 10](#_Toc44120302)

[Transform 10](#_Toc44120303)

[Load 13](#_Toc44120304)

[Data Exploration 13](#_Toc44120305)

[Outlier Analysis 13](#_Toc44120306)

[Data Understanding 14](#_Toc44120307)

[Weight Selection 22](#_Toc44120308)

[Model Preparation 24](#_Toc44120309)

[Normalization 24](#_Toc44120310)

[Model Verification & Validation 25](#_Toc44120311)

[Single Indicator Verification 25](#_Toc44120312)

[Dual Inverted Indicator Verification 26](#_Toc44120313)

[Results 28](#_Toc44120314)

[Conclusion 31](#_Toc44120315)

[Appendix A – Indicators and Abbreviations 33](#_Toc44120316)

# Project Overview

This project seeks to help businesses and investors identify country-level markets that could be pursued for thermal energy storage (TES) projects. Traditionally, market analyses like these are conducted by evaluating several key indicators for each market and then making a subjective assessment on which markets to invest in. This project reduces subjectivity in market analyses by using a weighted scoring system. Once investors and businesses agree on the most important market indicators, this scoring system will identify countries that should be further assessed for TES investment. The key attribute of this scoring system is that it is repeatable and deterministic- as market statistics change over time, the system can be rerun using the preselected market indicators.

# Problem Statement

The World Bank’s International Finance Corporation (IFC) is looking for commercial opportunities in developing countries to support deployments of renewable energy and, specifically, thermal energy storage.[[1]](#footnote-1) Institutions like the IFC along with businesses and investors around the world should have a platform that simplifies the process of evaluating and identifying potentially promising markets.

# Background

In 2018, 26% of global electricity generation was from renewable sources. By 2040, this is expected to grow to 45%. Comprising the bulk of this renewable regeneration was solar, wind and hydropower. However, the all-too-familiar limitation of solar and wind is that they are dependent on weather and the availability of sunlight or wind.

Energy storage addresses some of the limitations of renewable sources and has many more benefits. It can provide continuity of power when wind or sunlight is not available. It can provide electrical continuity during power outages. It can also provide load balancing to meet peak power demands and reduce the need for expensive backup power plants and equipment. By 2022, the market revenue for thermal energy storage is expected to be nearly 5x that of any other storage technology and TES is expected to maintain a respectable growth rate of 17%.

Thermal energy storage operates by storing heat or the absence of heat in a medium such as water, rocks, metal or molten salt. By storing that heat in an insulated enclosure, the heat can be recovered at a later time to produce electricity. There are three major types of TES: sensible, latent, and thermo-chemical. Each of these differs on the thermodynamic level in how heat is stored and recovered.

Sensible and latent TES designs are the most mature technologies. Latent technologies are more expensive because they contain more complex components to handle phase changes in the storage medium. Latent storage is more energy dense and both, latent and sensible storage solutions, have similar efficiencies; however, there is a high range of efficiency on energy output. About 30% efficiency is obtained when extracting only power. When extracting both heat and power, about 95% efficiency can be achieved. Storage periods for these technologies can be upwards of 24 to 48 hours.

TES solutions operate by accepting one of three different types of input. First, leftover heat from industrial processes can be stored. Alternatively, electricity from the grid, photovoltaics, or wind turbines can be converted into heat. Also, heat can be generated from concentrated solar power. At a later time, this heat can be recovered and used for heating and cooling, converted to steam for industrial systems, or converted to electricity through systems like steam generators or sterling engines. Batteries, on the other hand, maintain a direct connection between an electrical source and the electrical output and they forgo all of the mechanical systems in between.

Digging into the specifics, thermal energy storage excels with much higher power ratings, storage durations, and component lifetimes in comparison to batteries. The largest lithium-ion battery installation is about 150 MW in Australia. These batteries have discharge durations of about 8 hours and have component lifetimes of about 15 years. TES can provide power outputs of up to 400 MW for up to 48 hours. TES components are also typically rated to last between 30 and 40 years.

Energy storage technologies can be viewed in terms of discharge duration and power output. Lithium ion batteries will fall in the lower left on a plot of these two parameters and TES would be plotted adjacent to the right of batteries. TES has a fair amount of overlap with batteries but, as noted previously, it maintains much longer discharge durations.

In reviewing the maturity level of energy storage systems and their associated capital requirements and risk, TES is at the forefront in comparison to all other storage technologies. TES solutions are predominantly moving from the demonstration and deployment phase to commercialization.

# Data Sources

There is a plethora of data that can support an analyst’s review of markets for TES investment. Before an analyst searches for data to support a market survey, he or she would typically pose a set of questions that can provide information on particular markets. Once these questions are defined, the search for data to support answering these questions begins.

An example of a few questions an analyst could ask include:

1. What is the market size of the energy sector in a given country?
2. How developed is a country’s electrical infrastructure?
3. How often do electrical outages occur?
4. What is the price of electricity in a given country?
5. What is the risk of doing business in a country?
6. What is the wealth of a country and its citizens?
7. How do countries compare in terms of pollution?
8. How do countries compare in terms of renewable energy generation?
9. How committed is a country to renewable energy expansion and investment?

As new data is discovered in the future, the scoring system is designed such that this data can be integrated into this project’s consolidated database of information and into future market rankings. This project uses the data sets listed below to show the viability of the scoring system. The information from all resources below comprises nearly 200,000 data elements.

1. BMI Research Dataset
   1. **Description:** BMI provides in-depth market research reports on various industries and global markets, emphasizing emerging markets. It also provides extensive economic and political risk ratings and analysis, macroeconomic analysis and forecasts, and financial analysis of debt and equity. It includes profiles for many multinational companies and their subsidiaries, and intra-daily alerts on economic, industrial, and political developments, business deals, multinational joint ventures, and regulatory changes. The data pull from BMI includes 210 unique regions and 63 different indicators.
   2. **Total Items:** 152,172
   3. **Source:** <https://www.fitchsolutions.com/bmi-research>
2. Infrastructure Enterprise Survey
   1. **Description:** This dataset summarizes findings from a survey of 152 regions and countries over 12 questions related to the countries’ electrical infrastructure. In each of these regions, businesses were investigated to provide data on indicators such as: the percent of firms experiencing electrical outages, number of outages in a typical month, the percent of annual sales that were lost due to outages, and the percent of firms owning a backup generator.
   2. **Total Items:** 1,949
   3. **Source:** <https://tcdata360.worldbank.org/>
3. Regulatory Indicators for Sustainable Energy (RISE)
   1. **Description:** This dataset consists of scores that provide a snapshot of a country’s policies and regulations in the energy sector, organized into the following categories: Energy Access, Energy Efficiency, and Renewable Energy.
   2. **Total Items:** 5,488
   3. **Source:** <https://rise.worldbank.org/scores>
4. TES Market Potential
   1. **Description:** This dataset is from a report on the global thermal energy storage market and provides total market potential for TES in US dollars for over 200 countries.
   2. **Total Items:** 395
   3. **Source:** The 2020-2025 World Outlook for Thermal Energy Storage, <https://www.mordorintelligence.com/industry-reports/global-thermal-energy-storage-market-industry>
5. Access to Electricity Indicator
   1. **Description:** This dataset is from the World Bank Sustainable Energy for All database and indicates access to electricity as a percentage of each country’s population. The dataset provides breakdowns of electricity access overall, in urban areas, and in rural areas from 1990 to 2019.
   2. **Total Items:** 23,760
   3. **Source:** <https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS>
6. CO2 Emissions Report
   1. **Description:** This data was collected by the Oak Ridge National Laboratory Carbon Dioxide Information Analysis Center and published by the World Bank. It provides information on CO2 emissions in metric tons per capita for 264 regions and countries from 1960 to 2014.
   2. **Total Items:** 13,373
   3. **Source:** <https://data.worldbank.org/indicator/EN.ATM.CO2E.PC>

# Methodology

## Scoring Model

Weighted scoring models are common in project management and finance for evaluating projects against predefined criteria. Criteria such as net present value, market share, and project risk are popular metrics seen in the business sector. Weights are assigned to each criterion and indicate the relative importance of that criterion to the firm. The criteria values for each project are multiplied by their corresponding weights and then summed together to result in the total weighted score for each project. An example of this can be found here:

<https://baelearn.uncg.edu/wordpress/ism678/unit-2/unit-2-2-selection-models/unit-2-2-2-numeric-selection-models/2225-weighted-scoring-model/>

This project will be using the following scoring model:

**Step 1 – Identify Criteria**

After reviewing all indicators available for the dataset, a subset will be selected for use in the scoring model. The indicators used should have a hypothetical correlation with the project’s success. For example, it can be hypothesized that the percentage of people with access to electricity is inversely related to TES project success: countries with low access to electricity may exhibit a greater need for the reliability and energy supply provided by TES and thus result in greater demand for TES.

**Step 2 – Associate Weights with Criteria**

Once all criteria is selected, the weight table is populated. Weights indicate how important the associated criteria is to the project. The maximum and minimum values of the weights are arbitrary but should be consistent across all weights. For this project, only weights between 1 and 10 are selected, with 1 being least important and 10 being the most important. A column titled ‘Inverted’ is used to indicate that the criteria values should be reversed- high values should correspond to low scores and low values should correspond to high scores.

|  |  |  |
| --- | --- | --- |
| Indicator | Weight | Inverted |
| Access to Electricity (% of population) | 5 | True |
| Market Size ($ Millions) | 8 | False |
| Price of electricity | 3 | False |

Table 1 Example Weight Table

**Step 3 – Normalize Data**

With all criteria and associated weights, the data table is created, checked for missing values, and then normalized. Table 2 is an example of a data table, with the first column listing the countries and the remaining columns listing the values of each indicator for the countries.

|  |  |  |  |
| --- | --- | --- | --- |
| Country | Access to Electricity | Market Size | Price of Electricity |
| Mozambique | 78% | $ 20 M | $ 0.12 |
| Paraguay | 95% | $ 10 M | $ 0.08 |
| Pakistan | 98% | $ 15 M | NaN |
| Myanmar | 97% | $ 30 M | $0.12 |

Table 2 Example Data Table

Missing values are imputed with the median of each column and then each column is normalized to range between 0 and 1. For indicators that are inverted, like Access to Electricity in Table 1, the normalized value is subtracted from 1.

The equation for normalization is as follows, using the maximum and minimum values of each column:

Table 3 is an example of a normalized data table.

|  |  |  |  |
| --- | --- | --- | --- |
| Country | Access to Electricity | Market Size | Price of Electricity |
| Mozambique | 1.00 | 0.50 | 1.00 |
| Paraguay | 0.23 | 0.00 | 0.00 |
| Pakistan | 0.00 | 0.25 | 0.75 |
| Myanmar | 0.05 | 1.00 | 1.00 |

Table 3 Example of Normalized Data Table

Step 4 – Calculate Score

The country score is calculated by multiplying the indicator weights by the country’s normalized data. In equation form, the following calculation is performed:

The score matrix in Table 4 shows what the individual and total scores would look like.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Country | Access to Electricity | Market Size | Price of Electricity | Total  Score |
| Mozambique | 5 | 4 | 3 | 12 |
| Paraguay | 0.75 | 0 | 0 | 0.75 |
| Pakistan | 0 | 2 | 2.25 | 4.25 |
| Myanmar | .25 | 8 | 3 | 11.25 |

Table 4 Example Score Matrix

**Step 5 – Evaluate Results**

With all scores calculated, the analyst should then review the rankings of the countries and determine if any changes to the indicator weights are necessary. Once the analyst is satisfied with the results, he or she can use them to determine which countries should be considered for further analysis and a final evaluation for investment.

The score results are not the final recommendation of which countries to invest in. This scoring method is meant to help highlight potential countries of interest and save time in the analysis process. The corporation looking to make the investment decision should complete its standard investment evaluation for the top countries that are suggested by this scoring model. From Table 4, we can see that Mozambique and Myanmar had the highest scores and Paraguay and Pakistan had the lowest. The markets of Mozambique and Myanmar should thus be further investigated for TES project investment.

## Metrics

Successful implementation of the scoring model is based on being able to create a justifiable business case for each of the top countries using the criteria selected. Using the results from the example above, it makes sense that Mozambique and Myanmar are the highest scored countries. They both have large market sizes, high prices of electricity, and relatively low percentages for access to electricity- all of these factors are important based on the weighting table that was created.

Another way to check for proper implementation of the scoring model is to use only one criterion for weighting the data. After running through the scoring model’s calculations using any weight value, the final scores should result in an ordering of countries that matches the ordering of countries in the original dataset based on that sole indicator. This kind of test could also be run with the inverted value set to True for a reversed ordering of countries.

# Data Preparation

Data preparation is accomplished following an extract, transform, and load (ETL) pipeline. Data is first pulled from all sources, transformed into a useful form for this project, and then saved into a database for use by the scoring model.

## Extract

The first step along the data preparation pipeline is extraction. For each of the sources listed in the Data Sources section, the data was manually downloaded from the respective website or source document. Fortunately for this project, most of the data was available in the form of multiple Excel spreadsheets. Only the TES market potential data was manually captured into a spreadsheet. It is noted that the World Bank sources have application programming interfaces (APIs) that permit downloading information on demand in the form of JSON files; this was not used for this project. All data was saved as Excel spreadsheets into a single repository as shown in Figure 1. Pandas was used to open the spreadsheets.

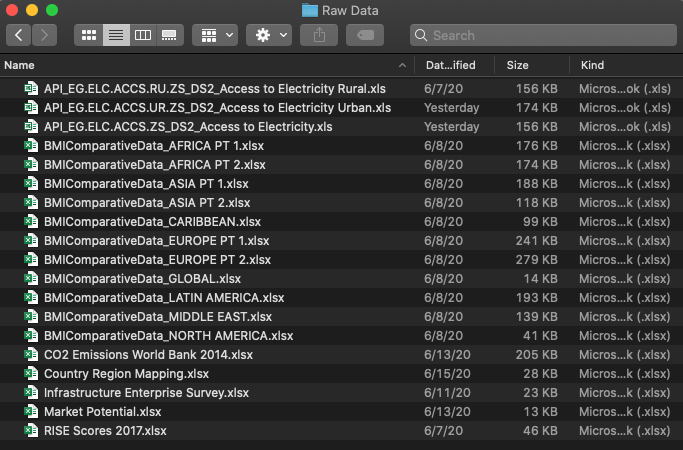


Figure 1 Consolidated Data Folder

## Transform

The next step is to transform all extracted data into a form usable by the scoring model. In Table 2, the scoring model requires a dataframe with all indicators along along the columns and all countries along the rows. As shown in Figure 2, the source data does not comply to this format and contains columns with country information, the indicator name, and then the years that the data was taken over.



Figure 2 Screenshot of a source dataframe

**Indicator Abbreviations**

Spanning all datasets, a total of 112 indicators were extracted. These indicators are fully spelled out and require abbreviation to improve readability in the final dataframe. For example, in Figure 2, the first row has an indicator name of “ELECTRICITY GENERATION: Generation, Total, TWh.” This indicator should be simplified as “GEN\_TOTAL\_TWH.” Appendix A – Indicators and Abbreviations lists all indicators and their associated abbreviations. Figure 3 provides a screenshot of the dataframe with abbreviated indicator names.

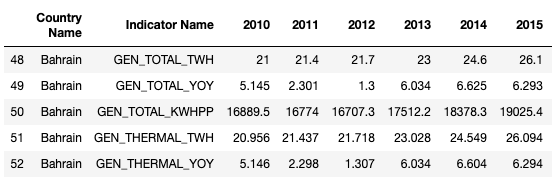


Figure 3 Dataframe with abbreviated indicators

**Country Code Mapping**

Country spelling and inclusion of regional statistics differed depending on the dataset. The country names need to be consistently formatted and used as the row names in the final dataframe. For countries with the same name but different spellings (e.g., Laos being spelled Lao PDR or Laos), using Pandas to join rows based on similar indices would not work because of differing index values. Pycountry was used to convert all full-length country names to their alpha-3 representation (e.g., Bahrain becomes BHR). In some instances, Pycountry’s fuzzy search method returned the wrong alpha-3 country code. To remedy this, a dictionary was created with the correct country and country code associations and then applied to the dataframe. All regional statistics like ‘North America’ or ‘South Africa’ were dropped from the dataframe since the focus of this model is on individual countries. Figure 4 provides an example of the country code associations in the dataframe. The resulting dataframe consists of statistics for 266 unique countries.

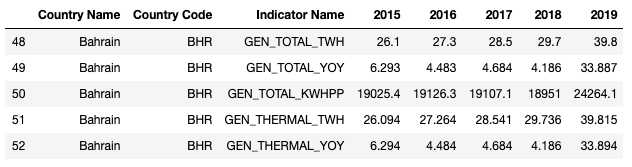


Figure 4 Country code additions

**Data Types**

With abbreviations of the indicator names and countries, a review of the datatypes in the dataframe was conducted. Many columns were treated as object datatypes due to the existence of commas and hyphens to represent missing data. Multiple Pandas replace method calls were used to remove the commas and properly represent missing values as NaNs.

**Pivot**

The desired dataframe, similar to what is represented in Table 2, has rows that are indexed with country names and columns of the different indicators. To achieve this, a Pandas pivot operation is performed using only 2017 data. Data for this year was selected based on the year’s recency and that this year consisted of the least amount of missing data. After pivoting, a dataframe consisting of 266 countries (rows) and 112 indicators (columns) was produced. A screenshot of this dataframe is provided in Figure 5.

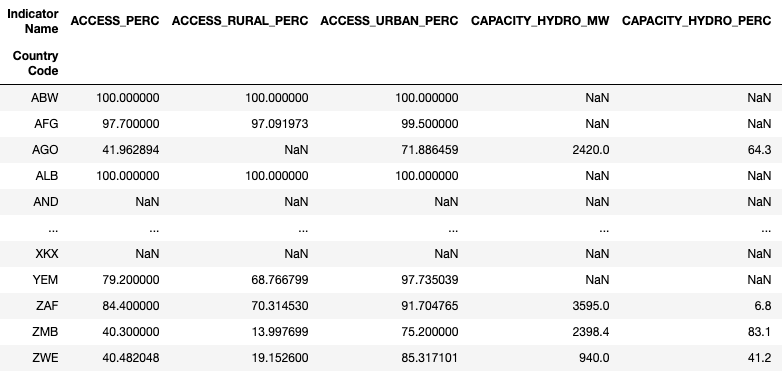


Figure 5 Dataframe after pivot operation

**Missing Values & Imputation**

The dataframe is now in the format needed for the scoring model. The last transform step is to establish a strategy for handling missing values. Using Pandas to create statistics on the number of missing values yielded a total of 166 countries that exceeded 25% missing data. The cutoff of 25% was selected to ensure countries are properly represented in this scoring model. In future versions of this scoring model, sensitivity studies should be performed to identify the optimal cutoff level that drops the least amount of countries but is still able to maintain a proper representation of them. The 166 countries were dropped and, with the remaining countries, missing data was imputed using the median of each column (indicator). Median imputation was selected to minimize sensitivity to outliers in each column in comparison to mean imputation.

## Load

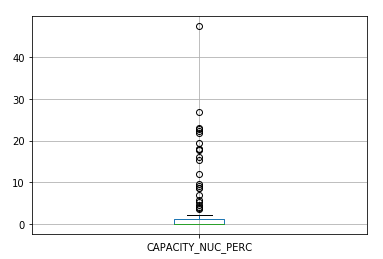
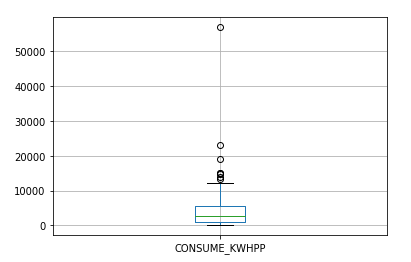
The final ETL step is to load or save the dataframe for further processing. Multiple options exist for preserving the data, such as using an SQL database, CSV file, pickle file, or an Excel spreadsheet. For this project, an Excel spreadsheet was selected to make it easier to open and view the data using standard office productivity software. The resulting dataset contains 99 countries (rows) and 112 indicators (columns) for a total of 11,088 data values.

# Data Exploration

## Outlier Analysis

Prior to exploring the data to better understand what it may be representing, a review of the outliers in the dataset is warranted. Outlier analysis can help identify values that are in error and that could result in erroneous results in the scoring model. The standard method of evaluating outliers using the interquartile range was used: values that are less than Q1 – 1.5\*IQR or greater than Q3+1.5\*IQR are labeled as outliers.

Only 36 out of 112 indicators (columns) in the dataset do not contain outliers. After further evaluation, this high number of columns that contain outliers is expected. For instance, there is extremely wide variability in country GDP values, electricity usage, and electricity production values. Looking through the rest of the indicators in Appendix A – Indicators and Abbreviations, outliers here are not implausible. Figure 6 shows a few boxplots of indicators in the dataset. By looking at the percentage of total power generation capacity from nuclear sources (CAPACITY\_NUC\_PERC) in the top right of Figure 6, it can be seen that many countries use little to no nuclear power; however, countries like the United States and China produce large amounts of power from nuclear sources. The data here is not alarming and does not warrant additional scrutiny. A similar story holds for the other box plots in Figure 6, average per capita electricity consumption (CONSUME\_KWHPP), total renewable power generation (GEN\_RENEW\_KWHPP), and total market potential (MKT\_POT\_USDMN).

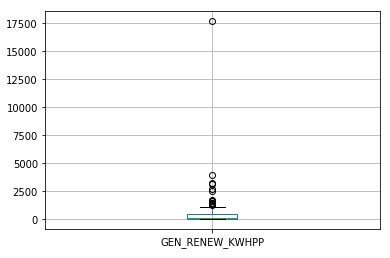
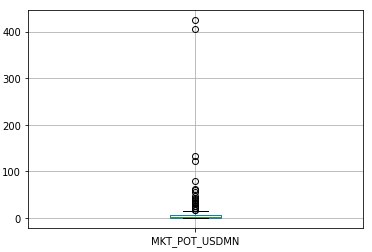
 

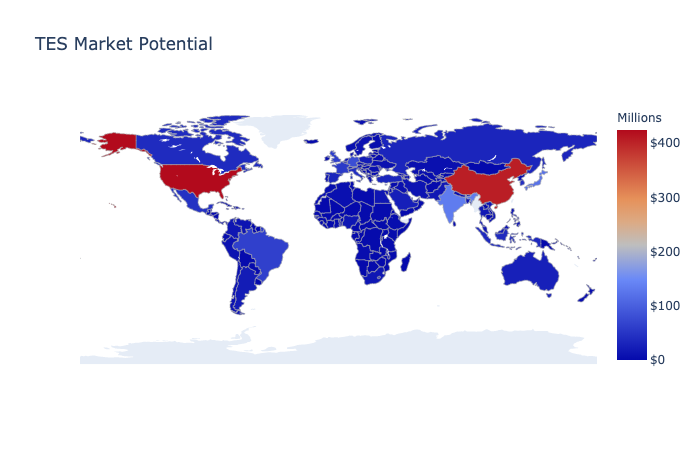
Figure 6 Boxplots showing outliers

## Data Understanding

The questions posed in the Data Sources section above are used to better understand the data and identify how it could best be used in the scoring model. For each question, a choropleth is used to provide a geographic visual of the data differences across countries. The Jupyter notebook contains more detailed statistics on each of these questions.

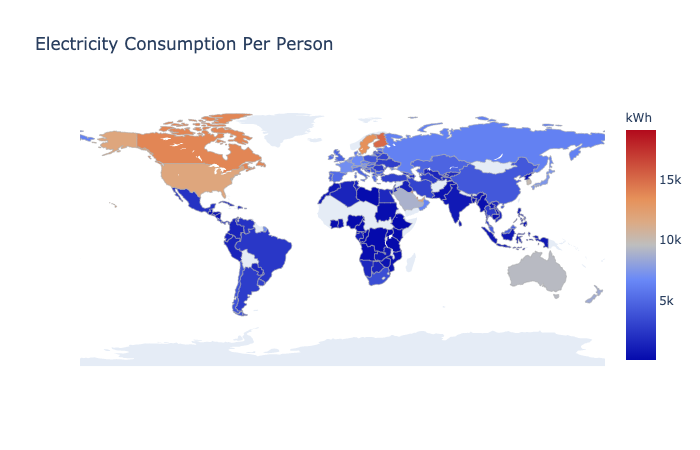
**What is the market size of the energy sector in a given country?**

The market size of the energy sector is determined with the MKT\_POT\_USDMN indicator which is the market potential for TES in millions of US dollars. As can be seen in the choropleth, the US and China are the largest markets for TES. India, Japan, Germany, and Brazil are the next largest markets.



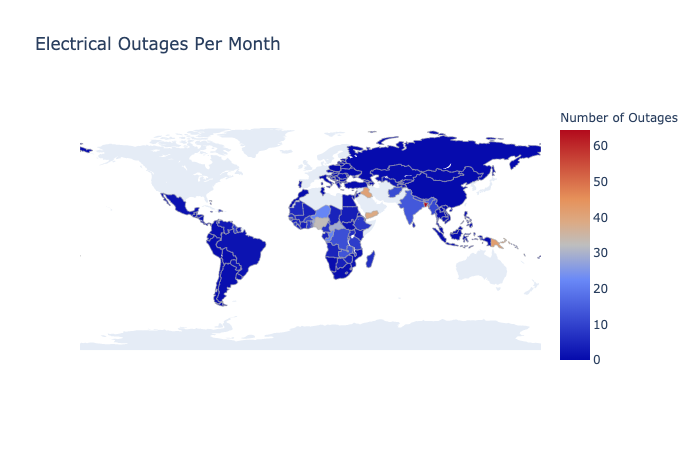
**How developed is a country’s electrical infrastructure?**

One way to gauge the health of a country’s infrastructure is through the amount of electricity consumed per person. In the dataset, this indicator would be CONSUME\_KWHPP which is the kilowatt-hours of electricity used per person in a country. People in developed countries such as USA, Sweden, and Finland are the greatest consumers of electricity. This may correspond to the health of the developmental state of each country’s electrical infrastructure.

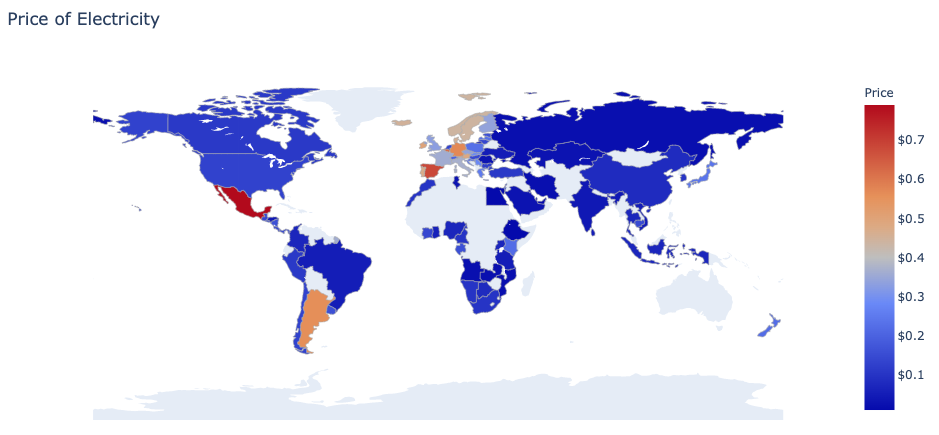


**How often do electrical outages occur?**

The indicator SURV\_OUTAGE\_QTY shows the results of a survey within each country that asks companies how many electrical outages they experience in a given month. From this data, Bangladesh, Iraq, Yemen, and Papua New Guinea have the highest number of outages.

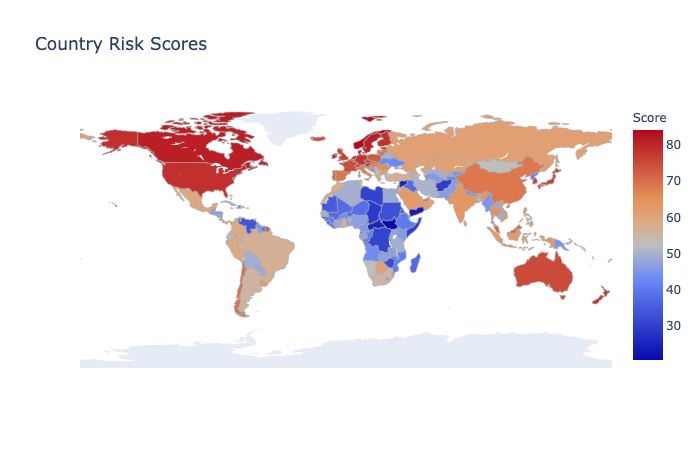
**What is the price of electricity in a given country?**

High prices of electricity can indicate problems that could be solved through TES. For instance, high electricity prices from inefficient grids or expensive energy production methods can be reduced with TES. The indicator PRICE\_RES\_USDKWH provides the prices of electricity per kilowatt-hour. Here, Mexico, Argentina, Spain, and Germany have the highest costs.



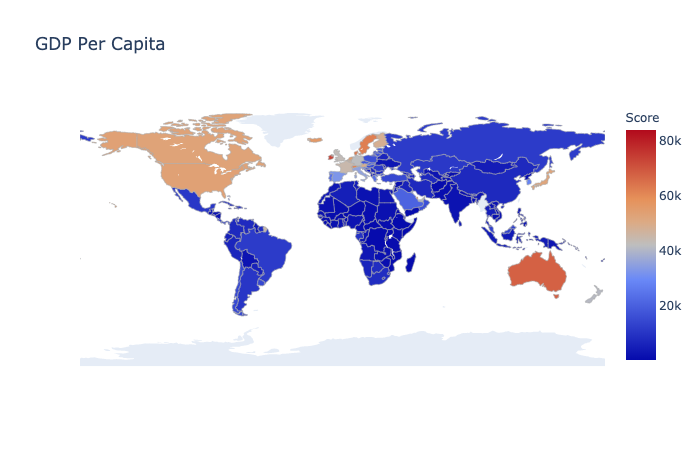
**What is the risk of doing business in a country?**

Investors and companies looking to do business in a country are interested in the level of risk associated with each country. Business risks could include investment risk, political risk, and logistical risk. The BMI dataset includes these risk indicators which are summarized with the RI\_COUNTRY indicator. This risk indicator is inverted whereby high risk scores indicate countries with low risk. As expected, developed countries rank the highest and have the least amount of risk associated with doing business. Africa is notable for a large portion of the continent having high levels of risk; companies looking to do business here should evaluate mitigations against this risk.

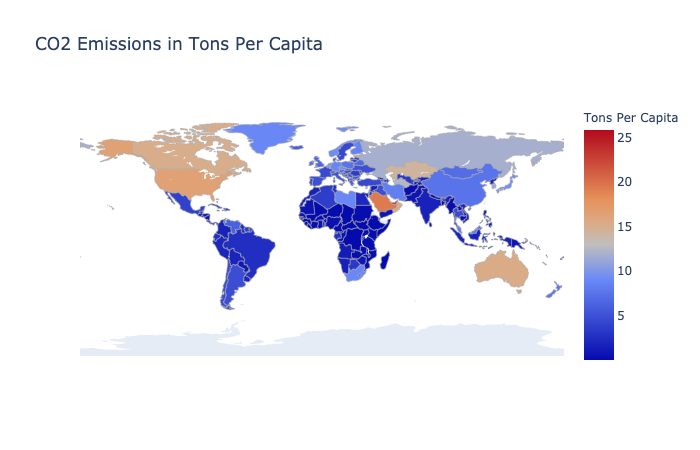


**What is the wealth of a country and its citizens?**

Countries with greater wealth could be markets that have a higher willingness to pay for TES which would result in greater profits to investors or companies. Real gross domestic product per capita provides an idea of the wealth of individuals in each country and this is captured by the indicator RGDP\_PP. As expected, developed countries such as Liechtenstein, Luxembourg, and Norway top the list for real GDP.

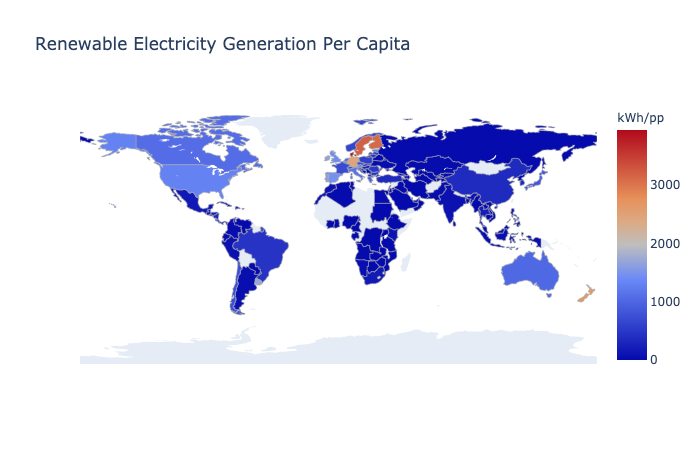


**How do countries compare in terms of pollution?**

High pollution countries are likely interested the reduction effects offered by renewable energy and TES. These high polluting countries could be good markets for TES as TES integrates well with renewable power plants. The top polluting countries on a per capita basis, indicated by CO2\_EMISS\_TONSPP, include Qatar, Curacao, Trinidad and Tobago. Larger countries with high pollution amounts include USA, Saudi Arabia, and Australia. This data is not surprising as these countries are heavily reliant on fossil fuels for electricity generation.****

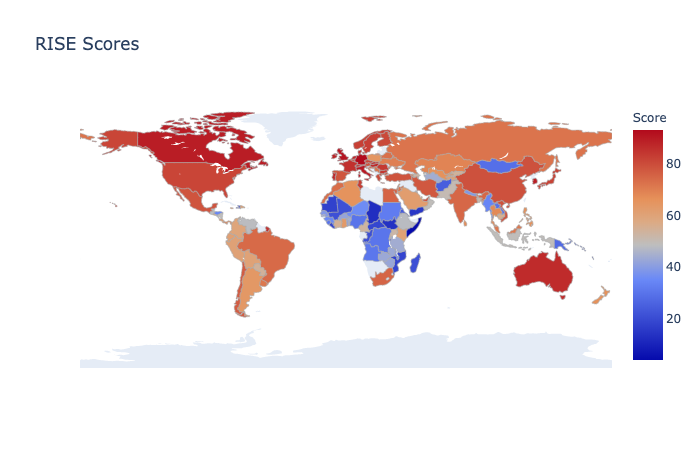
**How do countries compare in terms of renewable energy generation?**

Countries that are already producing large amounts of renewable energy may be interested in TES to provide load balancing and stability to their renewable sources. The total amount of renewable energy produced per country and normalized on a per-person level is provided by the indicator GEN\_RENEW\_KWHPP. Countries topping this statistic are all in Europe and include Iceland, Denmark, Sweden, Finland, New Zealand, and Germany.



**How committed is a country to renewable energy expansion and investment?**

The World Bank Regulatory Indicators for Sustainable Investment scores capture the strength of a country’s strategy with regard to renewable energy. The RISE scores evaluate countries across three categories, access, efficiency, and renewables. Access refers to polices and incentive that are in place to increase electrical network connections and reliability of the grid. Efficiency refers to efforts the country is taking to reduce waste and pursue efficient generation methods. Renewables, as the name implies, refers to policies to support the country’s adoption of renewable energy generation methods. The RISE\_OVERALL indicator is the average of individual scores for access, efficiency, and renewables. European countries are in the lead with Germany, Gibraltar, and Italy with the highest scores. South America and Southeast Asia are notable for their high scores in comparison to Sub-Saharan Africa. South America and Southeast Asia both contain several developing countries and if these developing countries have a national strategy towards renewable energy (as indicated by the RISE score), these countries may be a good investment for TES companies.



## Weight Selection

Using the preceding analysis along with a look at other parameters in the dataset, 11 indicators were selected for use in the final scoring. These weights also accounted for the World Bank’s IFC mission of focusing on emerging, developing markets and looking for companies that may be interested in delivering solutions to developing countries but may not have the financial or business facilities to do so (i.e., experience in international project finance/structuring and handling complex energy transactions). All weights were selected to range between 1 and 10 for consistency. The rational for these parameters is as follows:

* Mission of IFC:
* <https://www.ifc.org/wps/wcm/connect/corp_ext_content/ifc_external_corporate_site/about+ifc_new>
* <https://www.ifc.org/wps/wcm/connect/Industry_EXT_Content/IFC_External_Corporate_Site/Infrastructure/Priorities/Power/>

ACCESS\_PERC

This indicator provides the percentage of the population that has access to electricity. The World Bank is looking for countries that are in need of electrical infrastructure and this indicator is expected to align with that goal. A weight of 5 (medium importance) is given and this indicator is inverted such that countries with low access percentages have higher scores.

CAPACITY\_RENEW\_PERC

This indicator provides the percentage of total electricity generation that is from renewable sources. High relative amounts of renewable generation indicates possible opportunities for TES installations. A weight of 2.5 (low importance) is given since countries that have future plans for renewables will not be captured by this statistic.

CONSUME\_KWHPP

This indicator provides electricity consumption on a per capita basis. Low per capita consumption could indicate expensive, inefficient, or inaccessible electricity which could potentially be resolved by TES systems. A weight of 2.5 (low importance) is given and this value is inverted such that countries with low consumption countries have higher scores.

CONSUME\_YOY

This indicator provides the year over year percent change in electricity consumption for a country. Countries with high positive year over year growth values indicate potentially growing opportunities for renewables and TES. A weight of 2.5 (low importance) is given.

LOSS\_PERC

This indicator provides the amount of electrical losses due to transmission and distribution. High losses are usually due to inefficient and old grids, and in some cases, poor governance of the electrical infrastructure. Losses in the electrical grid directly translate to losses in revenue and could be partially solved with TES. A weight of 4 (moderate-low importance) is given.

MKT\_POT\_USDMN

This indicator provides the market potential in millions of US dollars for TES. This indicator is important because investors are profit seeking and there must be a market potential for TES in a given country before business commences. A weight of 5 (moderate) is given. A higher score is not given because developed markets like USA and China will dominate this metric.

PRICE\_RES\_USDKWH

This indicator provides the price of residential electricity in US dollars. Countries with high electricity costs indicate potentially inefficient markets and TES may be able to improve that efficiency. A weight of 3 (low importance) is given since electricity prices are known to fluctuate widely in developing markets.

RISE\_OVERALL

This indicator provides a picture of the regulatory incentives surrounding renewable energy in a country. Countries that offer incentives and have policies in place to support renewables will likely welcome TES investments. A weight of 8 (high importance) is given since government actions in a sector can widely influence business success.

RI\_COUNTRY

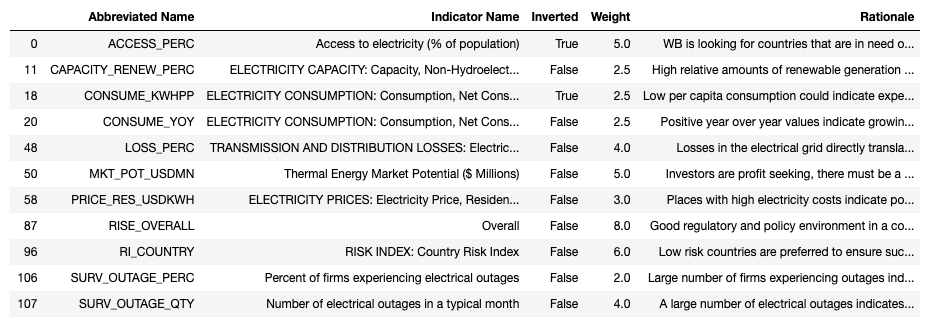
This indicator provides a perspective on the amount of risk is involved in doing business in a country. Low risk countries are preferred to ensure TES investments will be successful. A score of 6 (moderate-high importance) is given.

SURV\_OUTAGE\_PERC

This indicator presents survey results that shows the percent of firms in a country that experience electrical outages. A large number of firms experiencing outages indicates a need for grid stability which could be offered by TES. A score of 2 (low importance) is given since this is survey data and may be subject to bias.

SURV\_OUTAGE\_QTY

This indicator also presents survey results regarding the number of electrical outages that firms typically see in a given month. A large number of electrical outages indicate serious problems with the grid, again warranting a look at TES investments. A score of 4 (moderate-low importance) is given.



# Model Preparation

## Normalization

The final step before being able to run the model is data normalization. As noted previously, normalization takes all columns in the dataset and normalizes them from zero to one. This normalization removes the effects of extreme values and allows the selected weights to better capture the relative strengths of each country. Normalization applied to this dataset results a dataframe similar to that shown in Figure 7.

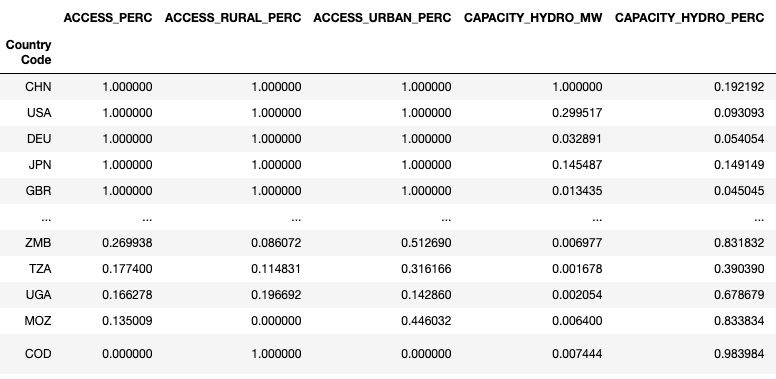


Figure 7 Normalized dataset

## Model Verification & Validation

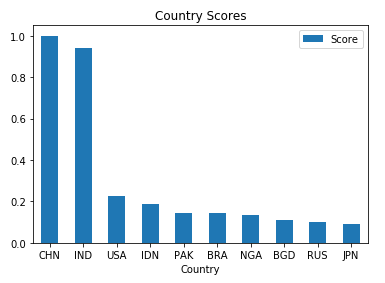
Model verification consists of evaluating and checking the model to ensure it is meeting the requirements it was designed with. Model validation is related and ensures that the model is meeting the original design intent since design intent is not always adequately captured in system requirements.

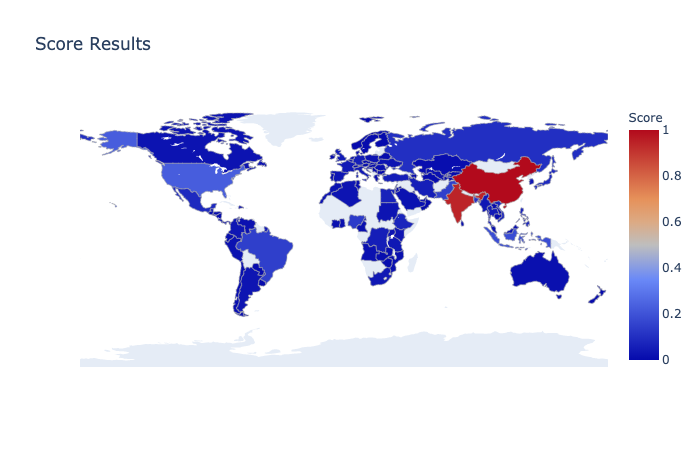
Single Indicator Verification

This model is verified by using a single indicator with arbitrary weight. With only one indicator being used, the resulting model will merely sort the countries based on this indicator. This sorting can be checked against an independent sort of the data.

The single indicator used is population. When applied to the scoring model, China, India, USA, Indonesia, and Pakistan rank at the top with the highest populations. Checking these results against InternetWorldStats.com, these results are accurate.



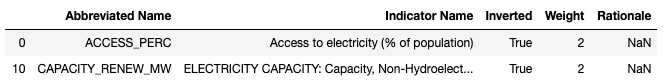


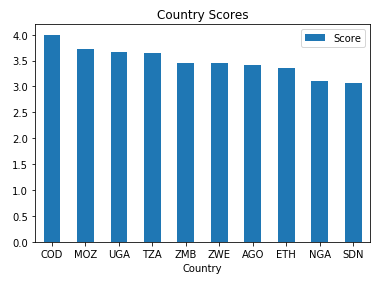


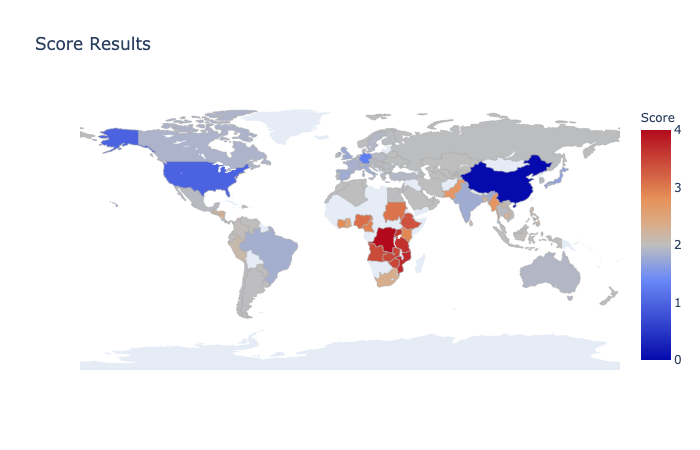
### Dual Inverted Indicator Verification

A second verification can be performed by using two indicators with equal weights but with their inverted values set to True. This run of the model will verify that the scoring calculation is operating and combining indicator weights correctly.

The two indicators used for this verification is ACCESS\_PERC which is the percentage of the population that has access to electricity and CAPACITY\_RENEW\_MW which is the total renewable energy generation capacity in megawatts. Note that both of these indicators are inverted in the table below; this should result in a reversed ordering of countries. The countries with the highest scores in this model are Congo, Mozambique, Uganda, Tanzania, Zambia, and Zimbabwe. These results are plausible as Africa is known for its poor electrical infrastructure. The scoring system is concluded to be functioning correctly: country rankings are in the expected order and the maximum equals the sum of the weights.

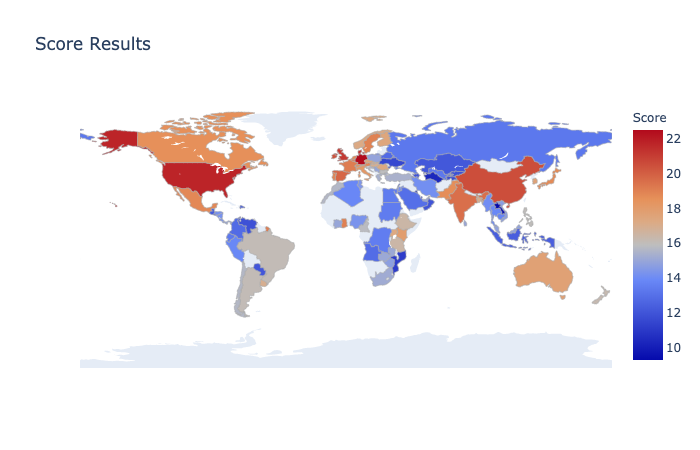






# Results

Using the weights described above, the scoring model was able to highlight a geographically diverse set of markets. As expected, developed countries like USA, Spain, Germany, Australia, and China all had the highest scores. Since the World Bank is focused on emerging markets, these developing countries were removed from the analysis.

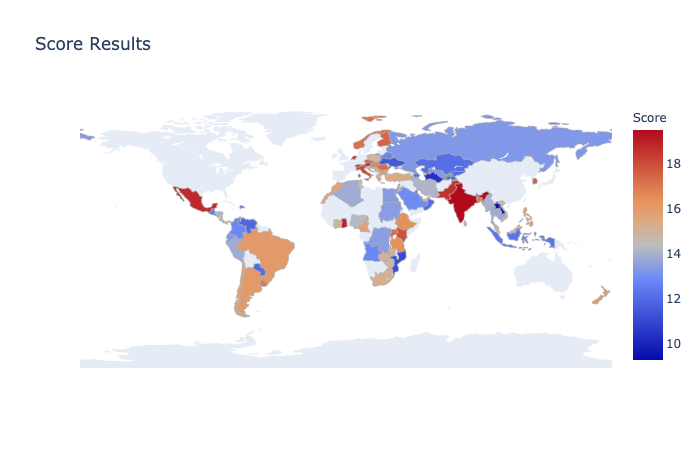


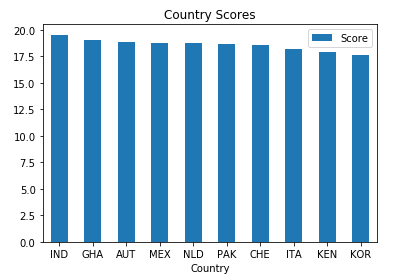
The choropleth was rerun to better highlight market opportunities in less developed regions. Topping the scores this time were India, Ghana, Brazil, Mexico, Argentina, Pakistan, Kenya, and Uganda. There was low disparity between the scores for these top countries indicating that they all are nearly equally recommended by the scoring system. The results of this ranking are consistent with current economic conditions in those countries. For example, recent news reports show that India is pursuing a massive $6 billion solar energy investment. Also, for Pakistan, power outages have been plaguing the country and the scoring model picked up on these extreme outage statistics. In Latin America, Brazil, Argentina, and Mexico are also seemingly good investment opportunities. These countries have relatively good government support for renewable investments and have lower amounts of country risk compared to Africa. In Southeast Asia, the Philippines is in dire need of better electrical infrastructure and has been opening up to more foreign investment in this sector.

<https://www.cnbc.com/2020/06/10/india-targets-lofty-climate-goals-with-6-billion-deal-to-boost-solar.html>

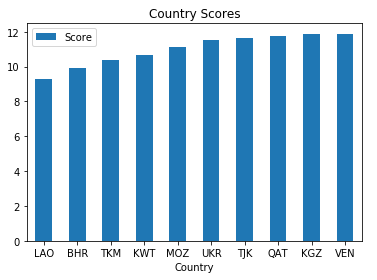
<https://www.reuters.com/article/pakistan-energy/pakistan-sees-end-to-routine-power-outages-by-years-end-idINKBN15S1L2>

<https://asiafoundation.org/2015/03/18/energy-crisis-in-the-philippines-an-electricity-or-presidential-power-shortage/>





Looking at the lowest scores, the results are not surprising. Laos is at the bottom mostly due to its poor legal and administrative framework for supporting renewable energy along with the difficulty of doing business in this country (i.e., country risk). Bahrain and Turkmenistan are also have very low scores. This is likely due to these countries’ high reliance on fossil fuels and slow adoption of renewable energy. Both of these countries are situated near large oil reserves which reduces their immediate need for renewable energy.



# Conclusion

The scoring model created in this project adequately assessed global markets for their viability of TES investments. Factors such as the regulatory environment, market potential for TES, and the state of the country’s electric grid were the salient indicators used in the scoring model. The results from the scoring model were logical and interpretable using the weighting criteria that was established. This indicates that it is possible to use this scoring model as an automated and more objective filtering mechanism for identifying countries to invest in. With this scoring model, several tens of hours of analysts’ time can be saved and faster investment decisions can be reached by management.

In going forward, there are many improvements that can be made to the scoring model created herein:

1. Unsupervised learning can be explored to attempt binning countries together based on similar indicator values. Given the over 100 indicators used in this set, finding the indicators that are most representative of the dataset could be useful for analysis and creating the scoring weights.
2. Many online data sources have APIs that permit automated downloading of indicator data. If this scoring model is to be used over prolonged periods of time, this scoring system should be revised to automatically pull data through these APIs. This will be useful for indicators that are changing quarterly and even yearly.
3. Lastly, ways to mathematically combine data across indicators should be investigated. For example, electrical losses in terawatt-hours is available; however, those losses translated into US dollars was not able to be found. However, it is possible to calculate these losses by multiplying the cost per kilowatt hour of electricity in a country by the total amount of electrical losses it experiences. These calculated columns may be useful to analysts and be able to paint a better financial picture to investors.

# Appendix A – Indicators and Abbreviations

|  |  |
| --- | --- |
| **Abbreviated Name** | **Indicator Name** |
| ACCESS\_PERC | Access to electricity (% of population) |
| ACCESS\_RURAL\_PERC | Access to electricity, rural (% of rural population) |
| ACCESS\_URBAN\_PERC | Access to electricity, urban (% of urban population) |
| CAPACITY\_HYDRO\_MW | ELECTRICITY CAPACITY: Capacity, Hydropower, MW |
| CAPACITY\_HYDRO\_PERC | ELECTRICITY CAPACITY: Capacity, Hydropower, % of total capacity |
| CAPACITY\_HYDRO\_YOY | ELECTRICITY CAPACITY: Capacity, Hydropower, % y-o-y |
| CAPACITY\_MW | ELECTRICITY CAPACITY: Capacity, Net, MW |
| CAPACITY\_NUC\_MW | ELECTRICITY CAPACITY: Capacity, Nuclear, MW |
| CAPACITY\_NUC\_PERC | ELECTRICITY CAPACITY: Capacity, Nuclear, % of total capacity |
| CAPACITY\_NUC\_YOY | ELECTRICITY CAPACITY: Capacity, Nuclear, % y-o-y |
| CAPACITY\_RENEW\_MW | ELECTRICITY CAPACITY: Capacity, Non-Hydroelectric Renewables, MW |
| CAPACITY\_RENEW\_PERC | ELECTRICITY CAPACITY: Capacity, Non-Hydroelectric Renewables, % of total capacity |
| CAPACITY\_RENEW\_YOY | ELECTRICITY CAPACITY: Capacity, Non-Hydroelectric Renewables, % y-o-y |
| CAPACITY\_THERMAL\_MW | ELECTRICITY CAPACITY: Capacity, Conventional Thermal, MW |
| CAPACITY\_THERMAL\_PERC | ELECTRICITY CAPACITY: Capacity, Conventional Thermal, % of total capacity |
| CAPACITY\_THERMAL\_YOY | ELECTRICITY CAPACITY: Capacity, Conventional Thermal, % y-o-y |
| CAPACITY\_YOY | ELECTRICITY CAPACITY: Capacity, Net, % y-o-y |
| CO2\_EMISS\_TONSPP | CO2 emissions (metric tons per capita) |
| CONSUME\_KWHPP | ELECTRICITY CONSUMPTION: Consumption, Net Consumption, KWh per capita |
| CONSUME\_TWH | ELECTRICITY CONSUMPTION: Consumption, Net Consumption, TWh |
| CONSUME\_YOY | ELECTRICITY CONSUMPTION: Consumption, Net Consumption, % y-o-y |
| GEN\_COAL\_KWHPP | ELECTRICITY GENERATION: Generation, Coal, KWh per capita |
| GEN\_COAL\_TWH | ELECTRICITY GENERATION: Generation, Coal, TWh |
| GEN\_COAL\_YOY | ELECTRICITY GENERATION: Generation, Coal, % y-o-y |
| GEN\_GAS\_KWHPP | ELECTRICITY GENERATION: Generation, Natural Gas, KWh per capita |
| GEN\_GAS\_TWH | ELECTRICITY GENERATION: Generation, Natural Gas, TWh |
| GEN\_GAS\_YOY | ELECTRICITY GENERATION: Generation, Natural Gas, % change y-o-y |
| GEN\_HYDRO\_KWHPP | ELECTRICITY GENERATION: Generation, Hydropower, KWh per capita |
| GEN\_HYDRO\_TWH | ELECTRICITY GENERATION: Generation, Hydropower, TWh |
| GEN\_HYDRO\_YOY | ELECTRICITY GENERATION: Generation, Hydropower, % change y-o-y |
| GEN\_NUC\_KWHPP | ELECTRICITY GENERATION: Generation, Nuclear, KWh per capita |
| GEN\_NUC\_TWH | ELECTRICITY GENERATION: Generation, Nuclear, TWh |
| GEN\_NUC\_YOY | ELECTRICITY GENERATION: Generation, Nuclear, % y-o-y |
| GEN\_OIL\_KWHPP | ELECTRICITY GENERATION: Generation, Oil, KWh per capita |
| GEN\_OIL\_TWH | ELECTRICITY GENERATION: Generation, Oil, TWh |
| GEN\_OIL\_YOY | ELECTRICITY GENERATION: Generation, Oil, % change y-o-y |
| GEN\_PUMPED\_KWHPP | ELECTRICITY GENERATION: Hydro-Electric Pumped Storage, KWh per capita |
| GEN\_PUMPED\_TWH | ELECTRICITY GENERATION: Hydro-Electric Pumped Storage, TWh |
| GEN\_PUMPED\_YOY | ELECTRICITY GENERATION: Hydro-Electric Pumped Storage, % y-o-y |
| GEN\_RENEW\_KWHPP | ELECTRICITY GENERATION: Generation, Non-Hydropower Renewables, KWh per capita |
| GEN\_RENEW\_TWH | ELECTRICITY GENERATION: Generation, Non-Hydropower Renewables, TWh |
| GEN\_RENEW\_YOY | ELECTRICITY GENERATION: Generation, Non-Hydropower Renewables, % change y-o-y |
| GEN\_THERMAL\_KWHPP | ELECTRICITY GENERATION: Generation, Thermal, KWh per capita |
| GEN\_THERMAL\_TWH | ELECTRICITY GENERATION: Generation, Thermal, TWh |
| GEN\_THERMAL\_YOY | ELECTRICITY GENERATION: Generation, Thermal, % y-o-y |
| GEN\_TOTAL\_KWHPP | ELECTRICITY GENERATION: Generation, Total, KWh per capita |
| GEN\_TOTAL\_TWH | ELECTRICITY GENERATION: Generation, Total, TWh |
| GEN\_TOTAL\_YOY | ELECTRICITY GENERATION: Generation, Total, % y-o-y |
| LOSS\_PERC | TRANSMISSION AND DISTRIBUTION LOSSES: Electric power distribution losses, % of output |
| LOSS\_TWH | TRANSMISSION AND DISTRIBUTION LOSSES: Electric power distribution losses, TWh |
| MKT\_POT\_USDMN | Thermal Energy Market Potential ($ Millions) |
| POP | POPULATION: Population |
| POP\_PERC | POPULATION: Population, % world population |
| POP\_RURAL\_PERC | POPULATION: Rural population, % of total |
| POP\_URBAN\_PERC | POPULATION: Urban population, % of total |
| POP\_YOY | POPULATION: Population, % y-o-y |
| PRICE\_IND\_USDKWH | ELECTRICITY PRICES: Electricity Price, Industrial Users, USD/KWh |
| PRICE\_IND\_YOY | ELECTRICITY PRICES: Electricity Price, Industrial Users, USD/KWh, % y-o-y |
| PRICE\_RES\_USDKWH | ELECTRICITY PRICES: Electricity Price, Residential Users, USD/KWh |
| PRICE\_RES\_YOY | ELECTRICITY PRICES: Electricity Price, Residential Users, USD/KWh, % y-o-y |
| RGDP\_BN | GDP: Real GDP, USDbn 2010 prices & exchange rate |
| RGDP\_PP | GDP: Real GDP per capita, USD 2010 prices & exchange rate |
| RGDP\_PPYOY | GDP: Real GDP per capita, USD 2010 prices & exchange rate, % y-o-y |
| RGDP\_YOY | GDP: Real GDP, USDbn 2010 prices & exchange rate, % y-o-y |
| RISE\_ACCESS | Electricity Access |
| RISE\_EA1 | EA Indicator 1: Existence and monitoring of officially approved electrification plan |
| RISE\_EA2 | EA Indicator 2: Scope of officially approved electrification plan |
| RISE\_EA3 | EA Indicator 3: Framework for grid electrification |
| RISE\_EA4 | EA Indicator 4: Framework for minigrids |
| RISE\_EA5 | EA Indicator 5: Framework for stand-alone systems |
| RISE\_EA6 | EA Indicator 6: Consumer affordability of electricity |
| RISE\_EA7 | EA Indicator 7: Utility Transparency and Monitoring |
| RISE\_EA8 | EA Indicator 8: Utility Creditworthiness |
| RISE\_EE1 | EE Indicator 1: National energy efficiency planning |
| RISE\_EE10 | EE Indicator 10: Energy labeling systems |
| RISE\_EE11 | EE Indicator 11: Building energy codes |
| RISE\_EE12 | EE Indicator 12: Transport |
| RISE\_EE13 | EE Indicator 13: Carbon Pricing and Monitoring |
| RISE\_EE2 | EE Indicator 2: Energy efficiency entities |
| RISE\_EE3 | EE Indicator 3: Information provided to consumers about electricity usage |
| RISE\_EE4 | EE Indicator 4: EE incentives from electricity rate structures |
| RISE\_EE5 | EE Indicator 5: Incentives & mandates: Industrial and Commercial End users |
| RISE\_EE6 | EE Indicator 6: Incentives & mandates: public sector |
| RISE\_EE7 | EE Indicator 7: Incentives & mandates: utilities |
| RISE\_EE8 | EE Indicator 8: Financing mechanisms for energy efficiency |
| RISE\_EE9 | EE Indicator 9: Minimum energy efficiency performance standards |
| RISE\_EFFICIENCY | Energy Efficiency |
| RISE\_OVERALL | Overall |
| RISE\_RE1 | RE Indicator 1: Legal framework for renewable energy |
| RISE\_RE2 | RE Indicator 2: Planning for renewable energy expansion |
| RISE\_RE3 | RE Indicator 3: Incentives and regulatory support for renewable energy |
| RISE\_RE4 | RE Indicator 4: Attributes of financial and regulatory incentives |
| RISE\_RE5 | RE Indicator 5: Network connection and use |
| RISE\_RE6 | RE Indicator 6: Counterparty risk |
| RISE\_RE7 | RE Indicator 7: Carbon Pricing and Monitoring |
| RISE\_RENEWABLES | Renewable Energy |
| RI\_COUNTRY | RISK INDEX: Country Risk Index |
| RI\_LOGISTIC | RISK INDEX: Logistics Risk Index |
| RI\_OPERATIONAL | RISK INDEX: Operational Risk Index |
| RI\_TRADE | RISK INDEX: Trade And Investment Risk Index |
| SURV\_ELEC\_CONNECT | Days to obtain an electrical connection (upon application) |
| SURV\_ELEC\_CONSTRAINT | Percent of firms identifying electricity as a major constraint |
| SURV\_GEN | Percent of firms owning or sharing a generator |
| SURV\_GEN\_USE | If a generator is used, average proportion of electricity from a generator (%) |
| SURV\_OUTAGE\_DUR | If there were outages, average duration of a typical electrical outage (hours) |
| SURV\_OUTAGE\_LOSS\_PERC | If there were outages, average losses due to electrical outages (% of annual sales) |
| SURV\_OUTAGE\_PERC | Percent of firms experiencing electrical outages |
| SURV\_OUTAGE\_QTY | Number of electrical outages in a typical month |
| SURV\_SHIP | Proportion of products lost to breakage or spoilage during shipping to domestic markets (%)\* |
| SURV\_TRANS | Percent of firms identifying transportation as a major constraint |
| SURV\_WATER\_PERC | Percent of firms experiencing water insufficiencies |
| SURV\_WATER\_QTY | Number of water insufficiencies in a typical month\* |

1. <https://www.ifc.org/wps/wcm/connect/Industry_EXT_Content/IFC_External_Corporate_Site/Infrastructure/Priorities/Power/> [↑](#footnote-ref-1)