Primary and Secondary Energy Consumption

Assessment of energy consumption and electricity mix in light of economic decision making

Students:

Leonard Brenk (158287)

Rasmus Krebs (135550)

Felix Wltschek (158368)

Class:

Visual Analytics

Programme: MSc Business Administration and Data Science

Date of submission: 06/01/2023

Number of pages/characters: 25/52160 (incl space)

Access dashboard here

Table of content

1. Introduction and Approach **1**

2. Target Audience **2**

3. Data **5**

3.1. Data source and context 5

3.2. Datasets 5

3.3. Extract, Transform, Load of Data on Energy 8

3.4. Tech stack 12

4. Dashboard **13**

4.1. Description and Quality of Dashboard 14

4.2. Relevance and Importance of KPIs shown 15

4.3. Analysis and evaluation 16

4.4. Justification of Design Choices and Accessibility 17

4.4.1. Dashboard Objects 18

4.4.2. Choice of Colors 19

4.5. Accessibility of the dashboard 20

5. Findings **21**

5.1. Discussion 21

6. Conclusion **24**

6.1. Limitations 24

6.2. Future Work 25

7. References **26**

8. Appendix **28**

1. Introduction and Approach

During the last couple of years companies have not only been experiencing a higher pressure from the society in regards to acting more environmentally conscious, but more and more have consequences of climate change affected their way of doing business through natural disasters, rising sea level or changing temperatures - undeniable signes that the the environment needs to be protected better. The economy accounts for a substantial portion of air pollution and revenue-driven destruction of natural ecosystems and has thus significant leverage to exert a positive impact on the environment. In order to enhance this impact without significantly limiting economic viability, we will look at energy consumption and related greener alternatives in this project. Fossil energy generation is currently still much more efficient, reliable and cheaper, leading companies to prioritize it over sustainable sources - even though sun, wind and water are available limitless and free of charge without having a negative impact on the environment. With this dashboard the global and country-wise energy mix and electricity profiles are assessed from different perspectives, providing companies a tool to find suitable sustainable locations for their energy procurement. With the help of charts that compare companies and put numbers into perspective, companies might now be able to grasp how large their negative impact of picking the wrong energy source can actually be and can hence result in a high impact with a low effort of switching energy suppliers. Although companies need to choose whether to take a greener route in the end, the dashboard can help to lead them into a more sustainable direction for the future. The dashboard is published under this link:
https://public.tableau.com/views/PrimaryandSecondaryEnergyConsumptionDashboard-VisualAnalyticsExam2223/LandingPage?:language=en-US&:display_count=n&:origin=viz_share_link

2. Target Audience

Climate change has gained awareness over the last couple of years and decades. Through this development, companies had to learn that they need to consider a new factor when conducting key strategic decisions: the way climate change could affect their operations in the future. Where in the past the sheer existence and price of resources was a major factor for selecting which operations should happen where on the planet, this might change in the future. Even though the price for resources will still be an important figure when conducting resource-based analysis, its computation might change severely. When comparing sustainable alternatives for existing fossil-based procedures[1], these alternatives often get assigned a so-called "green premium". A green premium is an additional cost created by a sustainable solution due to the fact that fossil resources are usually much cheaper (Gates, 2021). This holds only, as the existing prices for fossil energy resources do not include the costs caused by them through pollution and fostering climate change (Barbir et al, 2003). Already, some measures exist, that aim towards pricing energy sources with their actual price[2] and therefore also making climate-harming sources more expensive. These measures are called carbon pricing. Carbon pricing makes emitting greenhouse gases more expensive. As this is especially connected to the use of fossil fuels, their usage increases in price. Within carbon pricing, there are multiple different ways to execute it and receive the wanted effect of a reduction of emissions. Options for installing an efficient carbon pricing system can for example be based on trading emissions or on taxing their pollution directly (UNFCCC). With the increasing importance of climate change to the global public and politics, carbon pricing is getting more and more established. Today, already 30% of global emissions are covered by some form of carbon pricing (Black et. al, 2022). This will only have to increase if the target of the Paris Agreement of limiting global warming to below 2 degrees celsius should be fulfilled (United Nations). Consequently, it is highly likely that carbon pricing will become even more established and thereby relying on carbon-heavy energy would become more expensive for companies in the future. Therefore, when deciding on certain locations for their operations[3], companies will need to analyze the energy mix as one factor to find the strategically most suitable location. A country's energy mix is the combination of the energy sources the country uses to cover its total energy consumption (Planete energies, 2015). Under the assumption of a globally existing carbon pricing system, that makes the usage of fossil energy resources much more expensive, consuming energy in a country that mostly relies on such types of energy sources, the prices for energy could be expected to be much higher than for those which rely on more sustainable solutions.

Especially for energy-intense industries such as mechanical production or the chemicals industry, this decision could be crucial. For a CEO or COO of such a company who wants to decide about a new location or where to conduct an operation, especially two things would

[1] e.g., heating or energy production
[2] including their negative climate effect
[3] e.g., building new production sights or carrying out certain activities

become important. First, the person in charge would want to find a country with a primary energy mix that is as independent from fossil fuels as possible. The primary energy mix of a country covers all the energy consumed in a country unrelated to its actual use[4]. The executive would therefore get an intuition of whether consuming energy in a certain country would be in general rather expensive or not.

Secondly, due to the overall trend of electrification in the global economy, electricity becomes especially important within the secondary energy mix. The secondary energy mix includes the transportable form of primary energy[5]. When transforming primary into secondary energy[6], some portion of the energy is lost. Because of that and different efficiencies with the different sources of energy, primary and secondary energy are not identical. The term electrification speaks to the increasing reliance on electricity[7] to power processes (Deloitte, 2020). This is only sustainable if the electricity used is based on sustainable resources. Therefore, by looking at electricity as a secondary energy resource, a potential executive could find out whether using electricity in a certain country with existing carbon pricing would be expensive.

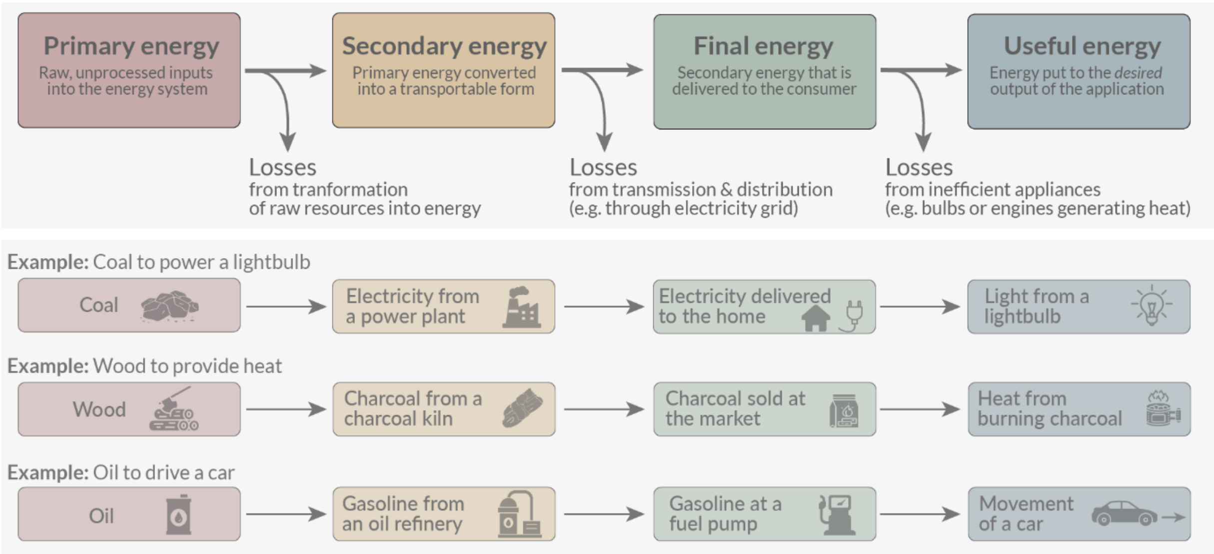


Figure 1 - Level of energy mix (Ritchie, 2022)

The relevant executive would be supported by experts or general staff. As the time of a person in a C-level position is really limited, the team's job would it be to further investigate the selected locations and back up the decision made by the executive. Therefore, more detailed information would be required.

---

[4] e.g., heating, electricity or transportation
[5] e.g., electricity or gasoline
[6] e.g. coal to electricity
[7] instead of for example burning fossil fuels

3. Data

3.1. Data source and context

The core dataset used for our project is called “Data on Energy” and was obtained through Our World in Data’s large catalog of datasets. Our World in Data is a publication platform focused on providing research and data for some of the greatest threats to the world, e.g., poverty, disease, hunger, and climate change (Roser, n.d.). Most of the research and data on the platform are contributions from various external sources, however, to ensure sufficient quality, all content is moderated by Oxford researchers and the non-profit organization Global Change Data Lab (Our World In Data, n.d.). The platform is primarily funded through grants from numerous large institutions, including the Bill and Melinda Gates Foundation, Musk Foundation, and World Health Organization. Besides grants, more than 4000 individuals and organizations have also supported Our World in Data’s efforts through numerous donations and contributions (Global Change Data Lab, 2022). Considering the grants by renowned institutions, the thousands of donations as well as content moderation being performed by researchers from one of the world’s most prestigious universities, we do not see any reason to question the trustworthiness of the data.

3.2. Datasets

In its original form, the dataset “Data on Energy” by Our World in Data contains 129 columns with 21.861 rows and is mostly a numerical dataset, with a few exceptions. Each row represents a country and a year, along with various energy-related information for the country in that given year, with the earliest year dating back to 1900 and the latest rows being from 2021. The energy-related columns include those that contains characteristics of the primary energy mix[8] and the computation of the secondary source of electricity[9].
Additionally, there are some general information about the respective country: country name, year, ISO country code, population and GDPR. The ISO country code column contains the ISO country code in alpha-3 format for the relevant country. The population and GDPR columns, contains the total population and GDPR for the given country in that given year. Because the dataset is concerned with the development over the period between 1900-2021, each entity is

[8] those raw unpressed materials inputted into the energy system
[9] the primary energy that have been converted in a storable form

represented many times and therefore the dataset does not contain any primary key. One could argue that the ISO code country column, with the codes being unique for each country and internationally recognized (ISO, n.d.), could serve as a unique identifier for each entity. While this could be the case, this is seemingly not intentional by the author, as the dataset includes nulls in multiple instances where they have included geographical aggregations, which will be expanded on later. However, for the purpose of joining datasets, this column will be the most sufficient as country names can vary depending on various factors, such as language, abbreviations and more. For that reason this will also be the column used for joining the “Data on Energy” dataset with our secondary dataset, “ISO 3166 Countries with Regional Codes”, which will be expanded on below. Due to the high dimensionality of the dataset we have deliberately not included descriptions nor general exploratory statistics for the whole dataset in this report. We will be expanding on the columns used for our dashboard below, however, we have limited to those. To compensate and allow curious readers to explore our raw dataset in more detail, we have collected all of the descriptions along with general statistics, such as unique count, null count and more, within appendix 1.

The energy-related columns represent detailed information about the country's energy consumption and are either related to the primary energy mix or to electricity, a secondary energy source. Additionally, the columns either include some aggregate information[10] or more specific information regarding certain energy sources. These sources are biofuel, coal, gas, hydropower, nuclear, oil, wind, solar and other

Overview of used columns

	Name	**Sublevel**	**Explanation**	**Primary Energy**	**Secondary Energy**
General information	**country**	-	Geographic location	-	-
	year	-	Year of observation	-	-
	iso_code	-	ISO 3166-1 alpha-3 three-letter country codes	-	-
Aggregated information	**primary_energy**	_consumption	Consumption of primary energy sources, measured in terawatt-hours	x	

[10] e.g., on the energy created by fossil fuels, the overall energy consumption or electricity demand of a country

	energy	_cons_change_pct	1)	x	
	low_carbon	_consumption	Combined consumption of energy coming from biofuel, hydropower, nuclear power, solar, wind and other renewable sources	x	
		_cons_change_pct	1)	x	
		_share_energy	2)	x	
		_share_elec	3)		x
		_electricity	4)		x
	fossil	_fuel_consumption	Fossil fuel consumption, measured in terawatt-hours, sum of primary energy from coal, oil and gas	x	
		_cons_change_pct	1)	x	
		_share_energy	2)	x	
		_share_elec	3)		x
		_electricity	4)		
	electricity	_demand	4)		x
		_generation	5)		x
		_net_elec_imports	6)		x
	renewables	_consumption	Same as low carbon but without nuclear	x	
Specific source	**biofuel, coal, gas, hydro, nuclear, oil, other_renewables, solar, wind**	_consumption	Primary energy consumption from respective energy source, measured in terawatt-hours	x	
		_cons_change_pct	1)	x	
		_share_energy	2)	x	

Table 1

Additional explanations

Number	Explanation
1)	Annual percentage change in primary energy consumption (of given source)

2)	Share of primary energy consumption that comes from given source - shares of all specific sources add up to 100% - low carbon (= nuclear + renewable energy) + fossil fuels add up to 100%
3)	Share of electricity generation that comes from given source
4)	Electricity demand, measured in terawatt-hours
5)	Electricity generation, measured in terawatt-hours
6)	Net electricity imports, measured in terawatt-hours

Table 2

renewables. To be able to fulfill the need of our target audience as precisely as possible, we decided to include not all of the existing columns, but only the especially relevant ones. The included columns and their related explanations can be seen in tables 1 and 2.

The data that is not included in the dashboard but was in the original dashboard covered some more granular information[11]. Additionally, the data set included information such as the consumption per capita or GDP for each individual source. We decided that these columns do not add enough information to be included and therefore rather minimized the included data to make the overall dashboard more intuitive.

Besides our core dataset "Data on Energy", we also used a dataset for mapping countries, more specifically ISO country codes, to their respective regions. The dataset we used was obtained through the Github repository "ISO 3166 Countries with Regional Codes" (Duncalfe, 2011), which contains all countries in the world, their ISO 3166 alpha-2 and alpha-3 code and the regions they belong to. By using this dataset and mapping on the ISO country code column instead of a general country to region mapping, we overcome common issues like different ways of spelling countries, e.g. United States of America versus United States. Seeing this dataset would not need any transformation or cleaning, and be a one-time upload, we argued that it would be less efficient passing it through an extensive extract, transformation and load (ELT) flow and instead would upload it manually to our data storages.

3.3. Extract, Transform, Load of Data on Energy

To download the data set, Our World in Data offers a GitHub repository[12] with prebuild Python scripts, that one can easily clone and run locally, which will extract and save the data in a CSV, XLSX and JSON format on one's local machine. While this does give access to their data,

[11] e.g., computation of the overall electricity consumption by the individual energy sources
[12] Find repository here: https://github.com/owid/energy-data

having to first run these scripts to download the data and then having to create a separate script for transforming and loading the data is very insufficient. To combat this insufficiency, limit friction and streamline this process for our case, we have decided to build our own Python script for extracting, transforming, and loading the data to suit our needs.

The first part of our ETL script is to extract data, where we utilized an existing python library by Our World in Data, which contains the necessary credentials and functions to enable easy access through their API. Importing data using this library is as easy as calling a *find* function, specifying which table you want access to and applying a *load* function on the results. Conveniently, the library returns the dataset in a Pandas DataFrame format meaning there is no need to perform any additional operations to format it into one.

After having extracted the data, we apply three different functions, all tasked with different areas of transformation. The first function is *restructuring,* which as the name suggest is focused with restructuring the dataset. Extracting data through Our World in Data's API returns a multi-dimensional index, that is unordered, which can complicate transformation and loading of the data. Therefore, the very first step of *restructuring* is to transform the DataFrame into a single-dimensional index, which is achieved using Pandas *reset_index* function. Following this, we sort the rows and columns conveniently by country and year so it is intuitive for the end user. Lastly, we reorganize the column order to ensure country, years and iso code are the first three columns, as these will be the first the end user sees and thereby make it easier to navigate.

Following *restructuring* we apply the *filtering* function, which will apply various filters on the DataFrame to remove unwanted rows and columns. The first part of the function removes all records before year 2000, as data prior this year is mostly uncollected and thereby nulls in the dataset and because we argued that having data prior this was redundant for our purpose. After removing all records prior 2000, we next remove all columns from the dataset that will not be used in our dashboard. It can be argued that keeping data that exceeds what is currently being processed or analyzed can be good, as it might come in handy in the future. However, we found that for the purpose of this assignment, that having all of those extra dimensions would add extra computation as well as an extra layer of granularity that is unwanted. Lastly, we remove all records without a value in the ISO country code column. We found that within the dataset, many rows were high level geographical aggregates of other rows on various dimensions[13] which we did not wish to include in our dataset. All these geographical aggregations were built based on some sort of logic by the author that is unbeknown to us and seeing that these aggregations are nested within the existing datasets, their values might cause confusion for the end-user. Because all of these aggregations are not a country, none of them includes an ISO country code and therefore we can easily remove all records using the *notna* Pandas function.

The last function in the transformation process is *divide_share*, which divides all columns containing *share* in their name by 100, as these are formatted by the authors to be percentages in whole numbers e.g. 25% instead of 0.25. The reason for converting the percentage into decimals, is when formatting the numbers in Tableau, a whole number like 25, would result in a percentage of 2500 instead of 25%, which is undesired.

After all the transformations have been applied, the dataset is reduced to a size of 4787 rows and 34 columns. Following the transformation the last part of our ETL process is loading the data. Typically, one would use a cloud database as the destination for this part of the process, however, after having searched for database providers that we could use, none proved sufficient and within a reasonable price. With no plausible offerings, we argued that we would instead use an alternative, however, one that would also include the ability to easily connect to it when using Tableau. We found that within Tableau, one of the many data sources that one can pick from was Google Drive, which seemed like a sufficient alternative. Therefore, for the loading of our data, we decided to use Google Drive, as the destination for our data.

As mentioned in the beginning, we decided to build our own ETL flow to streamline and reduce friction and therefore it was also important for us that uploading onto Google Drive would be achieved directly within our Python script. Google has already created a Google Drive API library for this exact purpose, however, a more convenient option is PyDrive, a wrapper library that simplifies many of the common tasks of this library (PyDrive, n.d.). There are a few steps involved with setting up the API, first of which is to get authentication for Google Service API, which is done by creating a client id in Google's Developer Consoler. The credentials for this client id are then downloaded and stored in the working directory, which the GoogleAuth class in the PyDrive library can then access for the API client. With the client id

[13] Geographical aggregates include EU, High-income Countries, North America etc.

created, the next step is to save the ID for the folder that the final file should be uploaded to, which can easily be found within Google Drive's user interface. Finally, one must get the file ID for the dataset in Google Drive. Unfortunately, this ID can only be accessed when the dataset has been uploaded, therefore, one would have to first run this script to generate the file ID and then update the script afterward with the file ID. It is possible to run the script without the file ID, however, without it it, the Google API client will not overwrite any already existing files and instead duplicate the file in Google Drive, resulting in multiple of the same files. Moreover, Google Drive has version control built into it, thereby, when overwriting an earlier version of the dataset, a user would be able to revert any changes to an earlier version. It is therefore easy to see how including the file ID can be beneficial and for that reason, we have also included it. After having downloaded the credentials and saved the relevant IDs, the next step is creating the connection in Python, create a file, setting content of the file, and uploading it. PyDrive includes functions for all these steps, with the first being *CreateFile*, which creates a Google Drive file and assigns the relevant metadata in Python. It is within this function that the folder ID and file ID are passed as arguments, along with the title and media type (mime type) of the final file. After having created the file, the next step is to set the content of the file created, which can only be done by referencing a local file, which implies that the final DataFrame must first be saved locally on one's machine. Luckily this can be achieved easily through Pandas *to_csv* function, however, saving the DataFrame locally we felt was redundant and went against the purpose of building our own ETL. To get around this, we implemented the *remove function* from the OS library in Python, which enable us to remove the file from the local computer when the loading process had finished. With the DataFrame saved locally as a CSV, the file is then set as content for the Google Drive file using the *SetContentFile* function. Lastly, the file is uploaded to Google Drive by calling the *Upload* function and subsequently deleted from the local machine.

While Google Drive did prove sufficient for our use case, not having the dataset in a traditional database with a corresponding database management system does include some inadequacy, like not being able to easily run SQL queries against it. Because of this inadequacy we decided to create a separate SQLite database file and load the final data into that using the sqlite3 library in Python. By doing so, interested parties can easily open up the file in SQLiteStudio to explore, create views and much more using SQL. Creating a database file is as easy as running sqlite3 followed by database name and the .db extension in the command line. This file is then stored appropriately and referenced using the connect function within the sqlite3 python library. With the connection created in Python, we used Pandas and its *to_sql* function to easily pass the final DataFrame into the database.

All of the code for the ETL flow can be found in the va_exam_externals.zip, which have been attached as an external file. The SQLite database can also be found within this file..

3.4. Tech stack

As described in the above section on our core dataset, we used *Our World in Data*'s large catalog as the source for our data. We extracted the dataset using their API python library, “owid-catalog”, which provided us with all the necessary credentials and functionalities for easy access and extraction. Besides extracting the data, Python was also our choice for transformation, more specifically, we used Jupyter Notebook for the transformation and loading during the exploration phases. When the ETL process was finalized, we converted the notebook into a separate Python script, allowing anyone to use either files for running ETL-flow. For loading our data we used “PyDrive”, a wrapper library of Google's API client, to load it onto Google Drive and the “SQLite3” library to load it onto our SQLite database, both of which would store our data. For the front-end portion of our tech stack we used “Tableau Desktop”, which allowed us to create beautiful and interactive dashboards in their straightforward, intuitive and minimal graphical user interface. Tableau includes an abundance of functionalities and capabilities, assisting us in performing various analysis and diving deep into our data. All of these features made Tableau a solid choice for communicating our narrative in a graphical and meaningful way. Besides Tableau, we also enabled the ability to explore and perform various operations on our the finalized dataset by opening the SQlite database in SQLiteStudio. For the ISO 3166 Countries with Regional Codes dataset that was downloaded from GitHub, we did as mentioned not perform any operations on it and uploaded it manually to onto our Google Drive and SQLite database. The final tech stack therefore looks as follow:

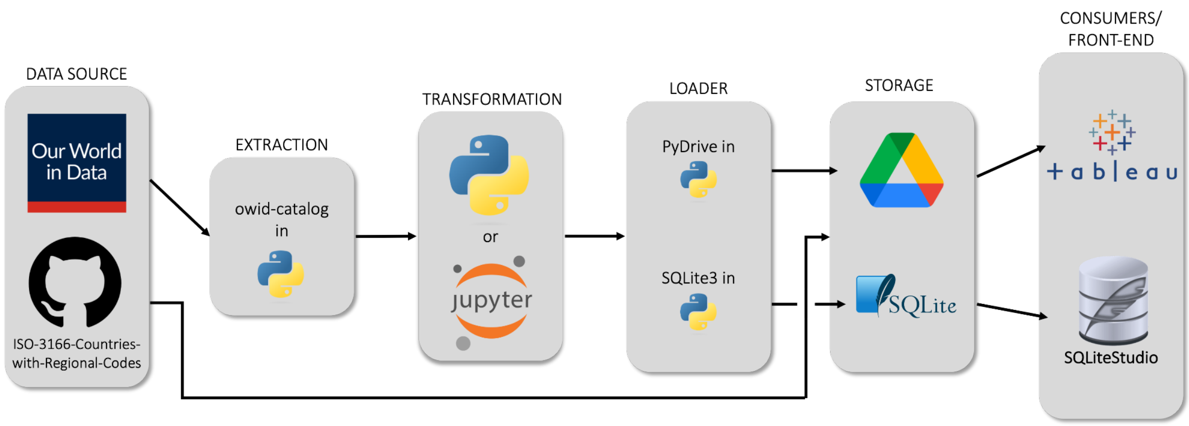


Figure 2 - Tech stack

4. Dashboard

The main underlying purpose of the dashboard is to illustrate and summarize the worldwide comparison of energy consumption and therefore its availability in countries. With an economic focus, it is mainly targeted at companies looking for cheap energy. As described in section 2, under the influence of globally existing carbon pricing, the target audience would want to identify a location that relies on fossil fuels[14] as little as possible. This would be the case for both, the overall energy mix as well as the secondary energy source electricity. In order to support the target audience in this matter, we have created a dashboard, that provides information regarding the overall energy mix of a country[15] and also deep dives into one part of the secondary energy mix, electricity. By using the dashboard, a company will be able to assess the overall (primary) energy mix of a country, its historical development and also understand how much a country's electricity production relies on fossil energy sources. As we assume that carbon pricing will make fossil fuels much more expensive in the future, our dashboard will enable companies to make the best strategic decisions.

This chapter provides an overview of the process behind creating the dashboard. Generally, we tried to make the workflow as simple as possible for the user, limiting the number of actions required by him/her, as well as minimizing the included data to deliver a clear message. Next to that good readability and accessibility were important to us. Additionally, we focused on the cognitive load for the user. We have tried to limit all objects and other necessary elements in both our visualizations and also on our dashboard, to limit what users need to process to

[14] the carbon-intensive ones
[15] regarding its primary energy source

understand the message that our dashboards create. Within all of that, we kept our target audience in mind and the specific use cases where our dashboard could serve handy, thereby understanding how to better frame the narrative for those people and situations.

The design of the dashboard was an iterative process which will be described later. Moreover, building a second dashboard was more or less to limit too much information on one dashboard, which would just cause cognitive overload.

4.1. Description and Quality of Dashboard

In order to ensure smooth navigation onto the Dashboard we added a landing page with the most relevant meta information and one big button that navigates to the dashboard. That page gives the user the first little impression on what they can expect on the next page and introduces the color palette.

The dashboard consists of two pages. The first page consists of a column with different selection options on the left and two visualizations on the right. In the selection section, a selection can be made between different continents (one or many can be selected). Below, three filters can be selected. All of them work as ranges. The first one ranges from the *year* 2000 until 2021. The second one is about the *average low carbon share* regarding energy and the third one regards electricity. If a range of years (e.g., from 2000 to 2021) is selected, then the other two filters are concerned with the average values over this period of time. In case just one year is selected, the two filters also cover just this year. The table below adjusts these filters according to them. It lists the names of the country fulfilling the previously defined criteria as well as the two values addressed in the other filters.

While designing the dashboard we realized that for some countries no values for at least one of the two KPIs exist. This applied especially to countries in Africa, which unfortunately limits the usability of the dashboard in this region.

Next to the selection/filtering column are two visualizations. Both follow the filters set on the left. They also include the *average low carbon share* for energy and electricity. The visualization on the top is a map. According to the selected time frame, the filling of the countries adjusts according to a color scale depending on a high or low share of low carbon energy. Regarding electricity, the circles placed inside the countries change their color, depending on a different color scale. When the users hover over a country, the exact information is displayed.

The other visualization is a two-by-two matrix. It includes the same KPI's, one on each axis. Each axis is divided by a line, highlighting the middle of the possible values on this axis. The countries are displayed within the matrix according to the two KPI's and differentiated by different colors depending on their continent.

If the user has selected the relevant criteria (year and KPIs) and the table, map and matrix have adjusted accordingly, it is also possible to select individual countries. This can be done within the table and both visualizations. If the user clicks on a certain country in the table, only this one is highlighted in the map and matrix. The selected country gets displayed on top of the filter column. By clicking on the button on the lower right-hand corner, the user can switch to the second page.

The second page provides a more detailed view on the country selected on the first page. It is split between the fields of concern for the two main KPIs from the first page (primary energy consumption and electricity consumption) into left and right. The left side is about the consumption of primary energy and consists of three major elements. A stacked bar-chart that shows the overall composition of the primary energy mix by energy source. Next to the bar chart is the overall consumption of primary energy presented as a number as well as a pie chart and a line chart. The pie chart provides an overview of how much of the energy mix is based on fossil- or on low carbon-based energy sources. The line chart gives an overview over the development the shares of each energy source had over the given time frame. Next to the individual sources, the aggregated values *fossil fuel, renewables* and *low carbon* are also included.

The right side, the one on electricity, gives an overview of the amount of electricity required, generated and imported by the selected country. Below that is an overview of how the consumed electricity can be split up in fossil- and low carbon-based electricity.

4.2. Relevance and Importance of KPIs shown

As described in the section above, there is a clear differentiation between both pages in the level of detail they display. When it comes to KPIs, the first page only includes the *Low Carbon Share* for *Energy* and *Electricity*.

The second page includes the *Share* that *Biofuel, Coal, Gas, Hydro energy, Nuclear energy, Oil, Solar, Wind and other renewables* have in the composition of the overall primary energy mix. Similar KPIs also exist for the absolute *consumption* of each source of energy. Next to that, the second page also includes the aggregated *Share of Low Carbon* and *Fossil* energy sources in the primary energy mix as well as for the *electricity* consumption. Lastly, it includes KPIs on the *Electricity Demand, Generation* and *Import.*

The two different levels of detail in combination with KPIs used for them, come from the different user groups that are supposed to work with the dashboard. The first page is supposed to give a quick high-level overview and implications to an executive who cannot focus on every detail of a decision. Therefore, the countries are presented based on the two most critical KPIs. The second page is thought to be used by the executive's supporting staff in order to back up the decision or raise potential issues.

An example could be that the executive wants to have an overview of the countries in Asia ordered by *Low Carbon Energy Share.* As the production activities of the respective company are quite electricity-intensive, the executive wants the selected country to also have at least 25% of its electricity based on low-carbon energy sources. This would lead to a list of seven countries. If the executive would then eliminate most of them due to other reasons (e.g. the political situation or location as an Eastern Asian country is required), only China, Japan and South Korea would remain. All of them seem to be quite equal in their "greenness" for their energy and electricity mix, with a slide advantage for China. Therefore the executive's decision would probably fall on China. The staff could then check the decision and would identify that in China the use of fossil energy (oil, gas and coal) has been increasing, whereas especially in Japan, it has been shrinking over the last ten years and the use of renewable energy has been strongly increasing. If this development continues, Japan will become a much more attractive location quickly. Such insides would be especially useful if certain energy resources would provide technical benefits in the future. Another inside that could potentially exclude a country from a list of candidates, would be whether it imports a lot of its consumed electricity and depends on other countries. A more detailed case study is described in the discussion part.

4.3. Analysis and evaluation

The first dashboard approach started with one page displaying the consumption and production next to the developments and certain relevant numbers and calculated scores out of those KPIs. The first change we realized needed to be made was to focus on consumption only instead of the production. When selecting a country for a future factory for instance it doesn't matter how much energy the country is producing if only a part is consumed in the country itself. The second change was to discard the score, as companies might have different measures to determine their best pick while a score would have specific weights and characteristics defined by us. In order to let the companies build their own score through a number of filters. Another change was the distinction between primary and secondary energy and putting a new focus on electricity. The biggest change however might have been the decision to add a second dashboard. We realized that the amount of relevant information and the nature of use in regards to two different groups of users[16] would not work as well on one page only. That is why we divided the dashboard into two pages that are smoothly connected by a navigating button. Until the final dashboard was finished, ten different dashboards were built over time and lead us to the final one. Thereby we made use of Tableau's advantage compared to PowerBI of separating charts from dashboards. As soon as the charts are ready one can build many different dashboards in a short time with drag-and-dropping the charts in the dashboards whereas in PowerBi one would have to build the charts from scratch when creating a new Dashboard. The final dashboard was then optimized many times and in the end we had 8 different visions saved until we had nothing more to improve.

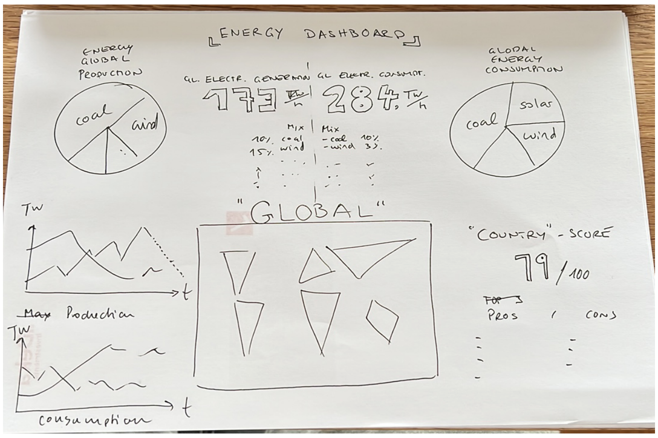


Figure 3: First dashboard draft

4.4. Justification of Design Choices and Accessibility

There were certain design choices made throughout this project. Next to theoretical input on how to design a dashboard, the core functionalities of the dashboard were key for its design. Under the assumption that relying as little as possible on non-fossil energy sources, these functionalities are

1. Giving a rough overview of global energy and electricity consumption

[16] C-level for rough overview and assistant for detailed assessment

2. Finding possible targets through filters for certain characteristics and creating a shortlist of possible countries by satisfying them
3. Providing detailed information on a specific country regarding that country's energy mix

As described in the previous section, is the first page for a rather high-level overview (serving functionalities 1 & 2) and the second page for providing detailed information (serving functionality 3). The selection of used visualizations was made accordingly. The connectivity of the two pages was extremely important to provide a good workflow for the user.

4.4.1. Dashboard Objects

As described previously, the dashboard consists of two pages with their own dashboard objects. An important concept that was used for designing the dashboard and implementing its objects, was to limit the activities required by the user (e.g., selecting filters). Therefore, the majority of the activities happen on the first page. In order to establish a good reading/working flow (left-to-right) in the dashboard, all the filter options for the user were set to the left side. Next to the filtering options (filters and table), the first page consists of a global map and a matrix. Both elements add some extra value to the dashboard, even though they simply present the same values that can also be seen in the table.

The map provides a geographical mapping that enables the user to identify potential relevant countries that lie right next to each other. Thereby, the effectiveness of the dashboard usage is not limited to the user's geographical knowledge.

The matrix with its differentiation into four clusters also provides some extra benefits. On one hand, it enables the user to see whether a selected country is more or less advanced in regards to carbon pricing depending on which field it is positioned in. It also allows the user to focus on specific needs. If the user is especially interested in countries with a low carbon share in the primary but not interested in the secondary energy mix, they should have a look at the two quarters at the top. If especially electricity is relevant, the most relevant target countries will probably be positioned in quarters on the right of the matrix. Due to the continent-based color coding, the matrix also allows for some peer-to-peer comparison.

In order to further limit the actions required by the user and to increase the interconnectivity of the two pages, it was ensured during the design phase that any selection on the first page would also transfer to the second one.

The second page includes, next to some facts that are displayed as numbers, a stacked bar chart, a pie chart, a line chart and a bar chart including two bars. As already described, the left side is about energy (primary energy) and the right side is about electricity (secondary energy). The stacked bar chart is used to show the exact composition of the primary energy mix. The pie chart simply provides a good overview of how big the share of the overall energy mix is, which can be assigned to low-carbon energy. For a similar reason, the two bars on the right side are used for electricity.

The line chart gives a good overview of the development in the primary energy mix and on how the share of each source is changing over time. The legend can also be used to highlight specific energy sources.

4.4.2. Choice of Colors

Color is one of the most powerful preattentive attributes and the use, and disuse of it can play a paramount role in the ability to communicate a narrative in a compelling and captivating way (Knaflic, 2015). Using color can help draw an audience's attention to the story, thereby, reducing the cognitive load of users. In her renowned book "Storytelling with Data: A Data Visualization Guide for Business Professionals" (2015) Cole Nussbaumer Knaflic highlights various good rules for picking colors. One of the rules she mentions is "Be thoughtful of tone that color conveys", which refers to the tone and emotion that a color can reinforce. In our case, one undeniable fact is the implicitly predefined color palette that is so oftenly used for representing sustainability or similarly in charts. A quick Google search on sustainability charts and one's Google feed will be filled with charts in shades of red, green and blue. When it comes to the tone of these colors, particularly in relation to visualizations regarding environmental performance of countries, then most people associated these with either good and bad, in fact the word "green" is widely used as an adjective for describing how environmental friendly an entity is. For this reason, we have also decided to draw on this color palette, as the preexisting tone from these colors can help highlight our story more sufficiently.

Another rule highlighted by Knaflic (2015) is "Design with colorblind in mind". There is no arguing that green and red are not very colorblind friendly, especially when used in the same visual. In fact, in here book, Knaflic (2015) writes "*In general, you should avoid using shades of red and shades of green together*" (p. 121). For this reason, we have limited the use of these colors within the same charts, and instead created and highlighted a narrative in which the need for these two colors have not been necessary. Thereby, throughout our dashboard shades of green have been used to represent generally sustainable energy sources. Besides green, we also draw upon shades of blue representing electricity shares and grey for generally carbon-intensive sources. Together all of these colors share a lot of similarities and neither of them contains enough contrast to draw and fight for the viewer's attention, letting the audience's attention be focused on whichever visual they are exploring. Collectively, these colors will be the primary colors used throughout our dashboards. Some secondary colors are used for each region in the two-by-two matrix on the first dashboard page. These are only used to distinguish between the regions in that one region throughout the chart and have been assigned randomly using Tableau’s Tableau 10 color palette. The final color palette used throughout thereby looks as follows:

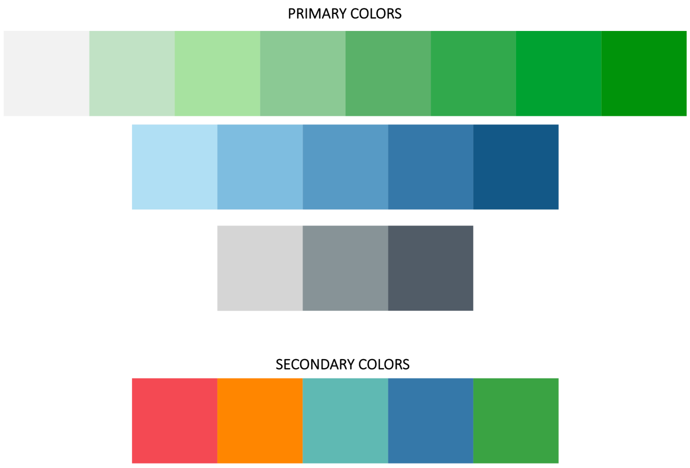


Figure 4

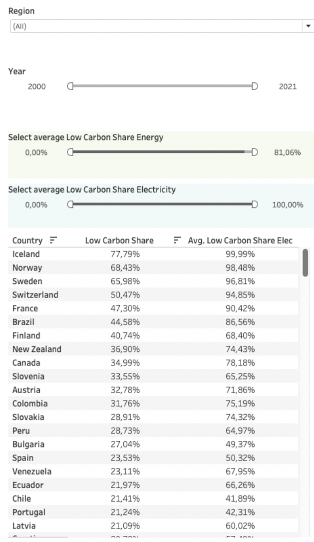
4.5. Accessibility of the dashboard

Seeing that the target group of the dashboard is mainly employed in larger companies working a lot with large screens and presentation tools like powerpoint, the dashboard size was oriented at the standard powerpoint slide size: 1600x900 pixels. Thereby users are not only used to the size but can comfortably open it on their device and navigate back an forth easily. As a decision like the one the dashboard is intended to assist with can have major consequences, the dashboard is not meant to be used on an Ipad or Iphone, but rather on a big screen in an environment that is appropriate for making fact-based longterm decisions.

5. Findings

5.1. Discussion

In order to illustrate how a company could benefit from this dashboard for internal decision making, we will go through the process like intended. When accessing the dashboard, the user is directly navigated to the landing page which presents the large button to the main dashboard and a few of relevant facts about it. When having clicked the button, the user then accesses the first site of the dashboard and is presented with two default, neutral charts not showing any relevant information as no region has been selected yet. That's where the user comes in and can select filters from top to bottom in order to find the country that matches their requirements the best. The filters (image right) let the user first pick one or more continents, then select years that are relevant for the users assessment, followed by the share of green energy/electricity. The resulting countries will remain in the table below while all countries not fulfilling those requirements will have disappeared both in the table as well as in the two charts on the right. Lets imagine a company that is located in Europe and is considering expanding to the Asian region. The first filter would therefore select Asia and Europe as continents. The company knows that the asian economy has developed especially fast in the last decade and would therefore like to only consider the last ten years, filtering for years 2011-2021. In regards to sustainability goals the company is aiming for at least 15% renewable energy sources in energy and electricity. Having adjusted the filters accordingly will show the following result.



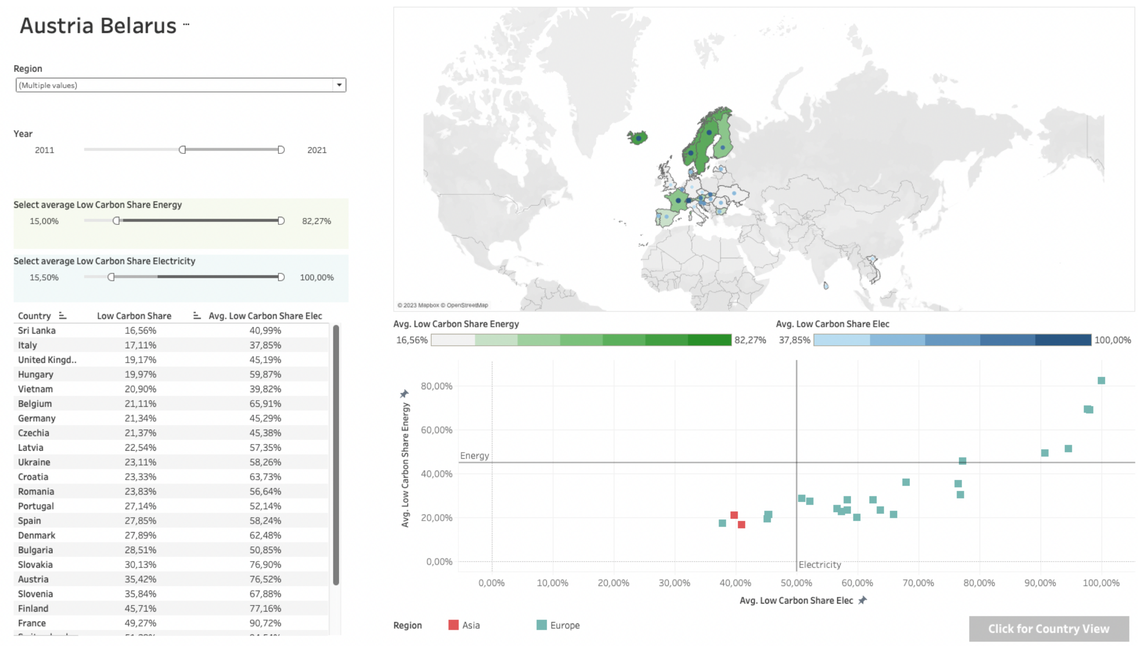


Figure 6

As described in Knaflic (2015) dashboard users tend to discover the dashboard in a “zick-zack” pattern, meaning the average user will start at the upper dashboard showing the world map highlighting those countries that still remained in the table in different tones of green. The chart clearly indicates that the european region satisfies the users requirements much better than the asian area. In the lower chart the green energy and green electricity share is shown in percentage, blue squares showing european countries and red squares countries in the Asian room. There are two asian countries within the users expectations although as visible in the lower chart the asian countries have a much smaller percentage of green energy and electricity. Let’s assume the company still wants to assess an asian country as labour and material costs are much lower than in the european region. In that case the user clicks on Vietnam highlighting Vietnam in all charts and tables. In order to get a more detailed view on Vietnam’s energy profile, the user can click the button in the lower right corner and is instantly navigated to the second page containing more specific charts and numbers about the country.

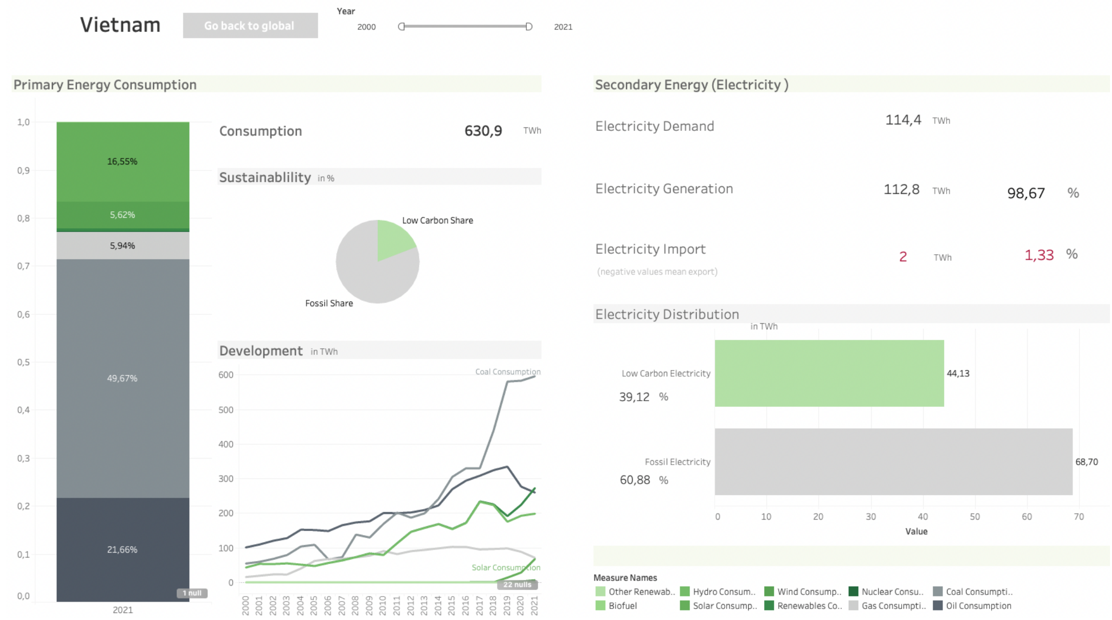


Figure 7: the second page

This page gives the user two important parts of the country's energy profile: the energy mix and development on the left side, and the electricity mix and numbers on the right. Taking a closer look to the contribution and sustainability charts on the left side, the user can identify that a large part - 77% to be exact - of the column is marked gray, meaning fossil. In the development chart the coal graph has taken a steep curve up in the last 5 years while green charts are located in the bottom of the chart. Thereby the user can anticipate that non-sustainable energy sources are more popular and might lead to a share of green energy under 15% in a couple of years which would interfere with the company's sustainability goals. The right side paints a quite similar picture as 61% of electricity consumed is based on fossil sources. In order to quickly find out how the development is in regards to the composition of electricity the user can pull the years range closer to 2016 - 2021 changing the fossil share from 61% to almost 63%, thereby underlining the conclusion drawn from the left side. With the numbers on both sides the company could compare countries in terms of the amount of energy generated, consumed and imported or exported in absolute numbers and percentages. Especially the percentage of electricity that needs to be imported is directly connected to the reliability in regards to other countries which could be influenced by political or environmental events. That percentage should thereby be as low as possible. Even though Vietnam does not import much electricity from other countries the company might in this case still decide to focus more on the european region as the propable development of the country's energy sources with a stronger focus on fossil origins does not work well with the company's sustainability objectives. In conclusion, the dashboard successfully let the user create a shortlist of relevant countries and showed the most important aggregated data - in this case the development chart - that assisted the user to make the best possible longterm, data-driven and result-oriented decision.

6. Conclusion

This report has concerned itself with the design process of the primary and secondary energy consumption dashboard. Before diving into the dashboard design process, this report highlighted the process of extracting, transforming and loading the core dataset, “Data on Energy”, from the publication platform Our World in Data. Throughout the description of this process the various tools and softwares used for this ETL flow were described in details, finalizing in the consumer/front-end section that in conjunction makes up the tech-stack used for this project. This transformed data was then accessed in Tablaeu to construct the dashboard, presented through this paper. The dashboard itself has undergone an extensive process. First there was only one dashboard displaying both energy consumption and production but no electricity data, while the end dashboard has two sites focusing on the consumption and has incorporated details on electricity aswell. As the narrative changed so has the dashboard; both have benefited significantly from the intensive engagement with the subject matter and dashboard theory during the time period of this project.

6.1. Limitations

One pressing limitation when working with energy data from around the world is the lack of tracking, particularly in the developing countries. Many of the cells in the dataset are unrecorded (null), which unfortunately complicates analysis and visualization, which can lead to misleading charts which is why we excluded affected countries as described earlier. Unfortunately, this is the reality of working with such kinds of data and relying on foreign data infrastructures, which is not prioritized equally around the world. Moreover, as highlighted, the source for our core dataset is seemingly very trustworthy, with large and renowned institutions backing it, which could lead one to believe that such discrepancies had already been addressed by them, implying that the inconsistencies in tracking are the nature of working with such data. While this can be the case, there is no arguing that this can be a significant weakness, one which should be given more thought and examined more closely had the time and resources been available.

Another weakness of the given project is the assumption that there would be one general carbon pricing mechanism, due to which emitting carbon in one country would have a similar price effect on the used energy as in another country. However, this is highly unlikely. In general, countries establish carbon pricing individually with their own conditions or as part of conglomerates (e.g., EU). Therefore, it oblies to them how effective the pricing is designed and therefore how expensive emitting carbon gets. Due to this, some countries might not install an effective carbon price to not limit any potential location advantage (Chateau et al., 2022). If in the future every country would actually have installed a way of carbon pricing, the effects the different mechanism would have on the actual price would have to be included in the dashboard. Without this the dashboard can just give an indication.

6.2. Future Work

In light of limited time and resources not all possibly insightful features could be implemented. For future improvement there are a couple of aspects that could potentially increase the impact of the dashboard and improve the transmission of relevant information. Even though the primary idea of the dashboard is to let the user get a first impression of the global energy consumption distributions, implementing energy prices in different countries might definitely be relevant to many dashboard users. Such could either be implemented on the first, more general, site with providing a rough overview of where energy is most or least expensive (a heat map could be a useful chart here) or in a more specific view on the second side. For that the data source would have to be merged with another dataset proving the energy prices, if possible in real time through the use of e.g., an API . Furthermore, due to the large size of the dataset including many years and every country, not all records are complete and should be further filled in order to ensure a realistic and complete view.

7. References

Barbir, F, Veziroğlu,T.N., PlassJr, H.J. (2003) *Environmental damage due to fossil fuels use, International Journal of Hydrogen Energy*. Pergamon. Available at: https://www.sciencedirect.com/science/article/pii/036031999090005J (Accessed: January 3, 2023).

Black, S., Parry, I. and Zhunussova, K. (2022) *More countries are pricing carbon, but emissions are still too cheap*, *IMF*. Available at: https://www.imf.org/en/Blogs/Articles/2022/07/21/blog-more-countries-are-pricing-carbon-but-emissions-are-still-too-cheap (Accessed: January 3, 2023).

Chateau, J., Jaumotte, F. and Schwerhoff, G. (2022) *Why countries must cooperate on Carbon Prices*, *IMF*. Available at: https://www.imf.org/en/Blogs/Articles/2022/05/19/blog-why-countries-must-cooperate-on-carbon-prices#:~:text=While%20carbon%20pricing%20is%20among,such%20as%20steel%20or%20chemicals. (Accessed: January 4, 2023).

Deloitte (2020) *Electrification in industrials*, *Deloitte Insights*. Available at: https://www2.deloitte.com/us/en/insights/industry/power-and-utilities/electrification-in-industrials.html (Accessed: January 5, 2023).

Duncalfe, L. (2011, May 18). *ISO 3166 Countries with Regional Codes*. GitHub.

Gates, B. (2021) *Bill Gates: How 'green premiums' can help us solve climate change*, *Fortune*. Fortune. Available at: https://fortune.com/2021/02/16/bill-gates-climate-change-research-green-premiums/ (Accessed: January 3, 2023).

GLOBAL CHANGE DATA LAB. (2022). *GLOBAL CHANGE DATA LAB: A company limited by guarantee - TRUSTEES' ANNUAL REPORT AND FINANCIAL STATEMENTS* .

ISO 3166 — Country Codes. (n.d.). ISO. https://www.iso.org/iso-3166-country-codes.html
Knaflic, C. N. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Wiley.

Our World In Data. (n.d.). *How we're funded*. Our World in Data. Retrieved January 5, 2023, from https://ourworldindata.org/funding

Planete energies (2015) *What is the energy mix?*, *Planète Énergies*. Available at: https://www.planete-energies.com/en/medias/close/what-energy-mix (Accessed: January 3, 2023).

PyDrive. (n.d.). *Welcome to PyDrive's documentation!* https://pythonhosted.org/PyDrive/. Retrieved January 5, 2023, from https://pythonhosted.org/PyDrive/

Retrieved January 5, 2023, from https://github.com/lukes/ISO-3166-Countries-with-Regional-Codes

Ritchie, H. (2022) *Primary, secondary, final, and useful energy: Why are there different ways of measuring energy?*, *Our World in Data*. Available at: https://ourworldindata.org/energy-definitions (Accessed: January 5, 2023).

Roser, M. (n.d.). *About*. Our World in Data. Retrieved January 5, 2023, from https://ourworldindata.org/about

UNFCCC *About Carbon Pricing*, *Unfccc.int*. Available at: https://unfccc.int/about-us/regional-collaboration-centres/the-ciaca/about-carbon-pricing (Accessed: January 3, 2023).

United Nations *The Paris Agreement*, *United Nations*. United Nations. Available at: https://www.un.org/en/climatechange/paris-agreement (Accessed: January 3, 2023).

8. Appendix

Appendix 1: Data on Energy by Our World in Data: Column descriptions and basic statistics

Link to sheet:

https://docs.google.com/spreadsheets/d/1zyaIsdXZeMne4q9hCdjGK8DBKiusCtCrCa6jki6upL0/edit?usp=sharing

	column	Count of Null	Unique Count	Mean	Median	Min	Max	description
General	country	0	319	Null	Null	Null	Null	Geographic location
	year	0	122	1974	1984	1900	2021	Year of observation
	iso_code	6162	219	Null	Null	Null	Null	ISO 3166-1 alpha-3 three-letter country codes
	population	4461	18040	1.04739E+16	7057871	1833	7909295104	Population
	gdp	11442	11113	358511083520	41674108928	164206000	113630171365376	Total real gross domestic product, inflation-adjusted
Biofuel	biofuel_cons_change_pct	21549	922	3.63597E+15	7.75316E+16	-100	5659328125	Annual percentage change in biofuel consumption
	biofuel_cons_change_twh	21010	950	3.03022E+16	4.60589E+15	-5458447265625	1.36261E+16	Annual change in biofuel consumption, measured in terawatt-hours
	biofuel_cons_per_capita	21603	614	159620361328125	1.8697E+15	0	1.74747E+16	Per capita primary energy consumption from biofuels, measured in
	biofuel_consumption	20971	955	4.1526E+15	2.48539E+16	0	1.13992E+16	Primary energy consumption from biofuels, measured in terawatt-h
	biofuel_elec_per_capita	17362	2620	627093620300293	1.11496E+16	0	2.40694E+15	Per capita electricity generation from biofuels, measured in kilowat
	biofuel_electricity	17142	1037	1.09988E+16	1E+16	0	6.6634E+15	Electricity generation from biofuels, measured in terawatt-hours
	biofuel_share_elec	17177	2701	1.99055E+16	638243556022644	0	7.14286E+15	Share of electricity generation that comes from biofuels
	biofuel_share_energy	20971	1046	5.46897E+15	7.45982E+15	0	86996431350708	Share of primary energy consumption that comes from biofuels
Carbon intensity	carbon_intensity_elec	17505	4016	4.48707E+15	4.87369E+15	0	4258064453125	Carbon intensity of electricity production, measured in grams of ca
Coal	coal_cons_change_pct	16883	5458	2.42013E+16	1.23966E+16	-100	55400	Annual percentage change in coal consumption
	coal_cons_change_twh	16535	5442	2.77466E+16	4.00559E+15	-173592578125	2646751953125	Annual change in coal consumption, measured in terawatt-hours
	coal_cons_per_capita	17712	4458	597837060546875	2.66227E+15	0	961699375	Per capita primary energy consumption from coal, measured in kil
	coal_consumption	16421	5482	162500146484375	4.68959E+15	0	4516120703125	Primary energy consumption from coal, measured in terawatt-hour
	coal_elec_per_capita	16805	2697	6.98804E+15	0	0	91038984375	Per capita electricity generation from coal, measured in kilowatt-ho
	coal_electricity	15919	3008	3.07291E+16	625	0	1011098046875	Electricity generation from coal, measured in terawatt-hours
	coal_prod_change_pct	12622	8904	1.85619E+16	2.20877E+16	-100	44965.75	Annual percentage change in coal production
	coal_prod_change_twh	5975	8712	1.5555E+16	0	-235952734375	305705859375	Annual change in coal production, measured in terawatt-hours
	coal_prod_per_capita	7945	8084	3.41833E+15	2.90456E+16	0	15166228125	Per capita coal production, measured in kilowatt-hours
	coal_production	5711	8648	9.35961E+15	1.51423E+15	0	4655060546875	Coal production, measured in terawatt-hours
	coal_share_elec	15954	3504	1.97861E+15	2.56952E+15	0	523074951171875	Share of electricity generation that comes from coal
	coal_share_energy	16421	5660	1.81059E+16	1.24897E+16	0	8.91387E+15	Share of primary energy consumption that comes from coal
Electricity	electricity_demand	17014	3341	5852919921875	1.059E+15	0	277827890625	Electricity demand, measured in terawatt-hours
	electricity_generation	15147	5125	7.20988E+15	2.3506E+16	0	277827890625	Electricity generation, measured in terawatt-hours
	electricity_share_energy	18389	4155	1.37507E+16	1.38135E+16	1.03271E+15	3.36054E+15	Electricity generation as a share of primary energy
Energy	energy_cons_change_pct	10294	11831	4.12526E+16	2.46042E+16	-9.50051E+15	155310498046875	Annual percentage change in primary energy consumption
	energy_cons_change_twh	10142	11933	8.43151E+15	1.11438E+16	-1287189453125	89174375	Annual change in primary energy consumption, measured in teraw
	energy_per_capita	12328	10112	26082701171875	138201796875	0	6575391875	Primary energy consumption per capita, measured in kilowatt-hour
	energy_per_gdp	15344	7211	1.82898E+15	1.32109E+16	783647820353508	2.52525E+16	Energy consumption per unit of GDP. This is measured in kilowatt-

<u>What do the dashboard users ask themselves?:</u>

- Where is the **cheapest energy** (cheapest in the sense of greenest → carbon pricing)?
- What is the **energy mix** at country X?
- How much energy is **available** where?
- How **reliable** is the energy mix of a country (reliability of sources, % imported energy)
- How did the energy mix **develop**?

Type	**Content**
Number	Total energy production globally

Number	Total energy consumption globally
Bar chart	(Solar, etc.) energy production per country
Pie chart	energy mix per country
Map	country heat map regarding consumption or production
Graph	Development over the last x years + prediction