

Project: Analysis of World Bank Data for Energy Consumption

Saif Ahmed
ahmedss@iu.edu
chameleon: ahmedss
futuresystems: ahmedss

Snehil Vishwakarma
snehvish@iu.edu
chameleon: snehvish
futuresystems: snehvish

Sruthi Mallina
smallina@indiana.edu
chameleon: smallina
futuresystems: smallina

ABSTRACT

This paper provides a detailed overview of the proposal for the project on analysis of world energy consumption. This will be analysed using the World Development Indicators data-set provided by World Bank. The world is amidst an energy crisis due to our over dependence on fossil fuels as the primary source of energy. Through this paper we are interested in analysing the different aspects of energy consumption in a country and its surrounding region. The primary aim of the project is to provide a statistical solution by incorporating a comprehensive analysis of the aforementioned data-set. This data-set in particular consists of thousands of annual indicators of economic development from hundreds of countries around the world. In order to get the relevant statistics and provide its visual presentation, technologies like Python, R and d3.js will be used. Technologies and frameworks are discussed briefly towards the later part of the paper. In addition to that, deliverables of the project are outlined so as to provide a quick estimation of the project scope.

Keywords

Fossil Fuels; Nuclear Fission; Carbon Footprint; Statistical Operations; Data Visualisation;

1. PROPOSAL

1.1 Nature of the project and its context

This paper uses the world development indicators data-set. Specifically the area of study would be the energy consumption of a country and how it stacks up against other countries in the same region. Efficient production and use of energy is a crucial question that needs to be tackled by each country. While addressing this, an important aspect to consider is, what percentage of the population has access to electricity. Under this another point is how was the access to electricity distributed over time. Over the last century or so, the dependence on fossil fuel as the primary form of energy generation has led to serious depletion of fossil reserves. Even though the percentage of energy generated using renewable resources and nuclear fission has considerably increased over the few decades but still it is not comparable to that of fossil fuels. And, this is a problem that every country is facing and needs to be addressed going forward. Through this project, analysis would be done on the energy

mix of the country and the percentage of energy generated using different sources. Another aspect to consider while doing analysis on energy would be to see how a country is consuming energy. Further, this project aims to analyse how the energy has been consumed over a specific period of time. Energy consumption using fossil fuels has vast applications along with implications of its use. With the rise of electric consumption there has been a substantial increase in our carbon footprint over the last few years. This has devastating results in contribution towards global warming and climate change in general. The comprehensive analysis and visualisation will provide information about the unprecedented effects of use of fossil fuels towards the consumption footprint.

1.2 Technologies used

The data analytics frameworks used in our project are the Python 2.7, R 3.3.1, D3.js. These technologies provide a lot of libraries which will be used for any scientific computations and data analysis. Python being used as a general purpose programming language to do a significant deep data analytics with concepts like Clustering, Regression, Structured Prediction and others, with strong libraries like NumPy, SciPy, Matplotlib, Pandas, Scikit Learn, Blaze. [2] Whereas R will be used for statistical and field-centric operations and also for data visualization because of its extensive library for statistical analysis with quantmod, gmodels and for data visualization with quantmod, ggplot2, metric-sgraphics, RColorBrewer. [3] This project uses D3.js as a primary tool to design dynamic and interactive data visualizations for the generated results.

1.3 Data Set

The data-set being used here in the project has been taken from kaggle which is an open source repository. [1]

1.4 Proprietary issues

There are no proprietary issues with the concerned project.

1.5 Aim of the Project

As discussed briefly in the abstract, the application aims at providing the statistics of energy consumption all over the world. The world bank data has a myriad of data based on mortality, gender, internet users over the years, etc. However, the project is focused on exploring the data related to energy as the world is amidst energy crisis. Using the world bank data-set with energy as world development indicator, this project intends to analyze and visualize the following:

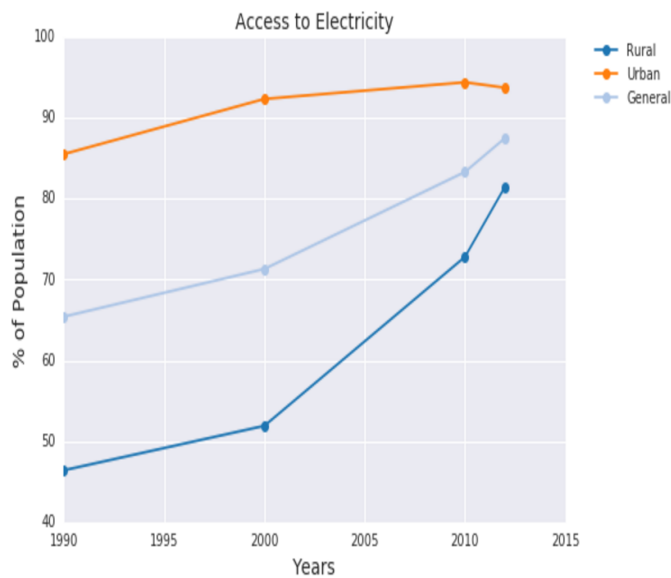


Figure 1: Energy Consumption
[4]

3. REFERENCES

- [1] Dataset. Webpage.
- [2] K. Jain. Tutorial for datascience with python. Webpage Article, Jan. 2016.
- [3] S. Machlis. R packages. Webpage Article, July 2016.
- [4] L. Miranda. Energy consumption statistics. Webpage.

- Resources from which energy is generated
- Sectors in which energy is being consumed
- Comparison between the countries based on the world energy consumption
- Percentage of population with access to electricity
- Rate of change of energy consumption over time
- Consumption Footprint

1.6 Intended Deliverables

With this application, we intend to produce deliverables such as visualization charts, screenshots of the application etc., besides the source code of the project.

2. ARTIFACTS

All of the artifacts for the project "Analysis of World Bank Data for Energy Consumption" will be uploaded to the project repository on GitLab. All the by-products of each phase will be included in it. The artifacts produced during the development of this project are:

- Project Proposal - Project plan and execution abstract. It also specifies the deliverables
- Project Status Report - Reporting periodic progress of the project
- Source Code - Executable files. Periodic update of source code based on project milestones
- Screenshots - Graphics showing Analytics Results and Visualized Data
- Project Paper - Concluding the technological approach, development and results of the project