|  |  |  |
| --- | --- | --- |
| World Happiness vs. CO2 Emission Ratings Database | |  |
| Data Analytics | April 14, 2019 | Technical Documentation | | |
| ETL Technical Report This report was built to walk through the creation of the Happiness Database. The information contained will display the reference sources for all the data, the extraction method, the transformation of the data, and the final upload into the database.  This report will also explain the HTML process used to display our data. |  | Report Index  PG. 2  Data Source Introduction and Explanation  PG. 3  Outline of (E) xtract, (T) ransform, and (L) oad process.  PG. 4  HTML Process Outline |

|  |  |  |
| --- | --- | --- |
| What Does the Data Mean? The first data set pulled is the “World Happiness Report” (<https://www.kaggle.com/unsdsn/world-happiness>) from Kaggle.com. This intent of this survey was to display global happiness and was released by the United Nations to rank countries around the world. The report also rates other indicators via survey other indicators such as economics, health, and public policies.  The scores in this data set are pulled from the Gallup World Poll 2017 data. The rating scale is a simple 0-10 scale from a nationally represented sample size. The other factors (columns) from this data set are ratings given by scholars in their respective fields which are independent from the overall happiness score.  The countries are organized from highest ranking (Norway) to lowest ranking (Central African Republic). |  | Inspiration….  The inspiration of this project came from exploring different datasets and finding a ranking of happiness as that is a relevant thought among people; everyone wants to be happy!  But, what could potentially affect, or decrease one’s happiness……  This is where we land on the CO2 data set |
| “Happiness, true happiness, is an inner quality.” |  | CO2 Emissions 🡪 What’s the Affect? CO2 Emission is carbon dioxide discharge from fossil fuels. As humankind has increased its fossil fuel use it has increased the CO2 emissions into the atmosphere. CO2 is measured by metric tons per year. The CO2 emissions can affect people’s health and plays a part in climate change. As a column in the original data set we thought this may be an interesting contributing factor.  This data came from the European Commission, a central location for policy and data relational to European Union. <http://edgar.jrc.ec.europa.eu/overview.php?v=CO2andGHG1970-2016&dst=CO2pc> Let the database building and correlation discovery begin! |

|  |  |  |
| --- | --- | --- |
|  |  | The Process… The two data sets were stored in CSV files. Jupyter Notebook was used to first upload the datasets (read csv). A new data frame was created to capture the relevant columns needed from both files. From the happiness csv, the country and ranking; from the CO2 file the corresponding emission rates was merged.  A connection to the database was created (create engine). Pandas was used to load the csv converted data frame into the database. After that data has been submitted confirmation is done by querying the database. |
| Extract – Transform - Load  * **Extract**is the process of *reading data* from a database. In this stage, the data is collected, often from multiple and different types of sources. * **Transform** isthe process of *converting the extracted data* from its previous form into the form it needs to be in so that it can be placed into another database. Transformation occurs by using rules or lookup tables or by combining the data with other data. * **Load** is the process of *writing the data* into the target database. |
|  | Challenges…  The issues incurred during the ETL process came with displaying and standardizing the datasets. The first merge did not produce the correct data combination, so the join type was changed. Also the headers were changed to make simple and normalize the table  The next issue was the country names. Gaps were produced in the result because of naming schema; this was corrected to give more outputs. However there were still some items without matching outcomes. A dropna.() function was used to remove NaN results to clean up the data. |
|  |

|  |  |
| --- | --- |
| HTML –CSS -Javascript  * **HTML –** is the language used to create web pages and web applications. * **CSS -** is a [style sheet language](https://en.wikipedia.org/wiki/Style_sheet_language) used for describing the [presentation](https://en.wikipedia.org/wiki/Presentation_semantics) of a document written in. * **Javascript –** a programming language which enhancesthe interactive experience on HTML pages. |  |
|  |

## The Process…

All the information from the data set (including this technical document) is uploaded onto HTML pages to give enhanced presentation capabilities. The styling of the website was done with some CSS and Javascript language loaded onto a style sheet. Enhancements including a sliding photo bar to show the team, a navigation bar to give access to multiple pages, and use of color coding to make user friendly.

The first page gives explanation and introduction to the overall project and business plan. Page two references the technical document for guidance. The third page gives insight to the actual database information.

Challenges…

The biggest challenges were. …

creating and formatting the proper margins. Proper alignment within the in individual web pages was a tedious task to get the look just right. In addition, image sizing and parameters were a bit of trial and error before it fit in the right array.

