

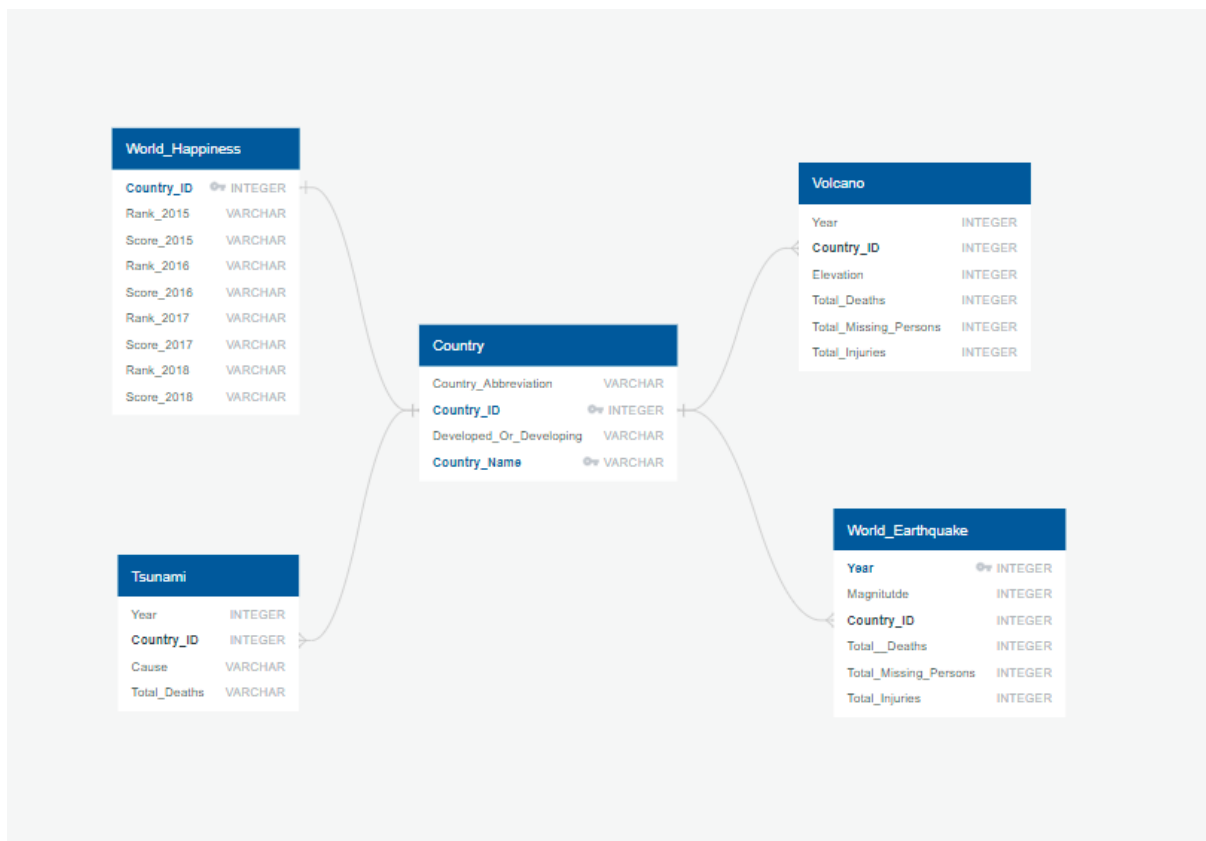
Technical Report:

Group members: Dylan, Nga, Nasrin, Franco

We are the natural disaster support team and are helping clients who have recently been affected by natural disasters to relocate. We are providing data-driven advice to clients about the countries with the least natural disasters and their respective world happiness score and index.

The ETL process was conducted using python in jupyter notebooks for the extraction and transformation, and PostgreSQL to load into the relational database via sqlalchemy in the notebook itself.

1. The relational database schema was defined quickdatabasediagrams.com as follows:



2. The export function was used to predefine the database in PostgreSQL (see [schema.sql](#))
3. Once the database is ready, the ETL process can run.

EXTRACTION PROCESS:

The following 5 (8 csv files) data sources are loaded. Csv files can be found in the Resources folder.

1. Volcano_data_2010.csv
 - a. Source: Volcanic Eruption Dataset (Kaggle)
<https://www.kaggle.com/texasdave/volcano-eruptions/version/2>

2. Worldwide-Earthquake-database.csv
 - a. Source: Earthquake Dataset (Kaggle) <https://www.kaggle.com/mohitkr05/global-significant-earthquake-database-from-2150bc>
3. tsunami_dataset.csv
 - a. Source: (Kaggle) <https://www.kaggle.com/andrewmvd/tsunami-dataset>
4. WH_2015.csv
 - a. Source: (Kaggle) <https://www.kaggle.com/unsdsn/world-happiness>
5. WH_2016.csv
 - a. Source: (Kaggle) <https://www.kaggle.com/unsdsn/world-happiness>
6. WH_2017.csv
 - a. Source: (Kaggle) <https://www.kaggle.com/unsdsn/world-happiness>
7. WH_2018.csv
 - a. Source: (Kaggle) <https://www.kaggle.com/unsdsn/world-happiness>
8. UNSD - Methodology.csv
 - a. Source: Country Data (United Nation Statistics Division) <https://unstats.un.org/unsd/methodology/m49/overview/>

TRANSFORMATION PROCESS:

1. World Happiness Data (WH_data_cleaning.ipynb)
 - The data was imported from five CSV files each represent a year
 - Import all csv files into Pandas dataframe and drop unwanted columns
 - Re-named Country column to prepare for merging
 - Re-named "Happiness Rank" and Happiness Score" according to their given year
 - Merge each data frame on "Country" to make one dataframe (combined.df)
 - Obtain the "Country" column from combined_df and make a new dataframe called country_df (later not used due to incomplete country data)
2. Country Data (WH_data_cleaning.ipynb)
 - As we noticed inconsistencies within Country names in all tables, Country Data was obtained from the UN Statistics Division to provide a standardised Country Name (country_df)
 - After obtaining relevant columns, merged with other data from main_cleaning.ipynb to detect country name inconsistencies
 - Re-launched main_cleaning.ipynb to change country names that were detected in the merging process.
3. Volcano, Earthquake, Tsunami (main_cleaning.ipynb)
 - Import each csv file into Pandas dataframe and drop unwanted columns
 - Re-named Country column to prepare for merging
 - Re-named each columns to ensure the names of columns are in the same format
 - Merge each data frame on "Country_Name" to make one dataframe (combined.df)
 - Filtered the dataset to match with the country happiness data which is from 2015 to 2018

LOADING PROCESS:

1. While going through the loading process, a mis-match data type error occurred because the Year columns in the data frames were of 'date' datatype. Re-casted into integer datatype.
2. Each column name for all data frames were re-casted into lowercase letters to match the column names in the database.
3. The index was set as the Country_ID for each table to avoid errors when loading into the database.
4. The final database would give an idea of the number of each natural disasters happened each year (2015 - 2018) and the country happiness index throughout each country
5. This topic is chosen because we want to help people who fear natural disasters choose the most appropriate country based on the 3 most common types of natural disasters and the country's happiness index.

FINDINGS:

- Some countries that had occurrences of natural disasters were not listed in the World Happiness Index
- Some islands were listed as the island name rather than the country itself (The Kermadec Islands near New Zealand) but is uninhabited.
- Their may be links between the occurrences of tsunamis and earthquakes, which can be further explored.