

Extracting, Transforming, and Loading Earthquake, Population Size, and GDP Data: To Predict the Impact of Future Earthquakes



Team Members: Arjun, Christy, Ivy, Shifaa (Team 6)

Project Description:

We plan to extract earthquake data for the last 200 years using the USGS earthquake API and to find the nearest city for each earthquake. We will find the current population size of those cities and their country's current GDP. We will then transform and load the data into a SQL database.

Extract:

1. Extract earthquake data for the last 200 years using the USGS earthquake API:
<https://earthquake.usgs.gov/fdsnws/event/1/>
2. Identify the nearest city to each earthquake using citipy: <https://github.com/wingchen/citipy>
3. Extract current population size for those cities from a csv file downloaded from Kaggle:
<https://www.kaggle.com/max-mind/world-cities-database>
4. Extract current GDP data for all countries by scraping the UN Statistics website:
<https://unstats.un.org/unsd/snaama/Index>

Transform:

1. Clean and transform the data.

Load:

1. Create a new SQL database and SQL tables to hold the extracted and transformed data.
2. Load the data into the SQL database.

Future Analyses:

With the data we will extract, transform, and load into a SQL database, we will be able to calculate a risk predication number to determine the impact that an earthquake today would have on the cities closest to past earthquakes. The risk prediction number could factor in the number of people who would be affected (population size) and the resources countries have to rebuild and recover after an earthquake (GDP). As such, the database we create could be used to predict the impact of future earthquakes.