

ETL project

Group 4

Glenda Decapia

Christopher Sadlo

Input sources:

- <https://www.kaggle.com/noaa/hurricane-database>
 - This data is originally source from HURDAT2, NOAA's hurricane database
- CitiPy
- country-converter from <https://pypi.org/project/country-converter/>

Transformation:

- Using both Atlantic and Pacific csv files data from Kaggle but limiting to:
 - (1) Year \geq 2000
 - (2) Maximum Wind \geq 100
 - (3) Latitude and Longitude value conversion from two different conventions of decimal degree format e.g. (15.1N, 103.1W) to (15.1, -103.1)
 - (4) Use other columns data: ID, Name, Date, Time, Status, Latitude, Longitude, Maximum Wind, Minimum Pressure*Note: filters #1 & #2 limits the data to less than 1000 records*
- using CitiPy
 - (1) Use Kaggle latitude and longitude data to determine the city name and country codes
 - (2) New column cityname will be created
- Using OpenWeatherMap API
 - (1) Use country to get the countryname
 - (2) Countryname will be a new column
 - (3) Use other column data: temp_max, humidity, clouds, wind speed
- Using country-converter module
 - (1) Retrieve the country names using the country code from CitiPy*Note: country code not found will be dropped*
- Other data manipulations
 - (1) Convert city names to title case
 - (2) Rename column names that uses spaces
 - (3) Convert column names to lower case

Database:

- SQL
- Database name: hurricanes

Output table:

- Table name: hurricanes

- Google maps plot of hurricane locations

Number of records to write:

- Less than 1000 records