ETL Report: Wildfire economic impact

Team:

Yukie Kajita

John Falcone

Katy Luquire

**Sources:**

Beginning with the two given databases, Data.World and Kaggle, our group decided initially that we would be using Kaggle for this project. Using both websites, we unanimously agreed it was easier to use and find data sets in Kaggle, making it a clear choice for a project we only have a week to work on. Then we began thinking about what type of data we would be wanting to use in this project, Katy was visiting family in northern California at the beginning of the project and mentioned the wildfire situation there right now. Which brought the group to the topic of natural disasters, focused on wildfires and their economic impact, seeing how much economic damage they have been attributed in each year since 1911, with FEMA data beginning in 1953.

<https://www.kaggle.com/fema/federal-disasters>

<https://www.kaggle.com/dataenergy/natural-disaster-data?select=number-of-natural-disaster-events.csv>

**Data Transformation:**

The team began with importing the data from the three CSV files into Jupyter notebook and defining their data paths for use in Pandas.

* Economic data: Renamed one of the column names to a shorter path, located only the Wildfires, and finally dropped the unused columns (Code, Entity).
* Numerical data: The same cleaning types that were used on the economic data to give us two coherent databases that can be merged.
* Annual data: The resulting database of the combination of the cleaned economic data and numerical data.
* FEMA data: Similar to the economic and numerical data cleaning. Data was .loc for only fires, numerous unneeded columns were dropped, the date is extracted using a split and made into a column in the database.

Finally, both the cleaned and combined Annual data, along with the newly cleaned FEMA data are exported back to CSV files and saved in a new folder.

**Production database:**

Using PgAdmin we import the CSV files and create tables based on the data using the year as a key to connect the tables, creating the variables, and showing the tables for each database after they have been created.

* Trying to find out the wildfire data in California compared to North Carolina for the entire dataset, this brings up over 200 plus individual events recorded for California and 5 for North Carolina.
* 2012 was one of the later years in the dataset, the team wanted to find the total amount of damage measured in dollars done by wildfires in that year. With 38 results in the United States total damage done was estimated at 1 billion USD. With most of the wildfires happening in western states, with dryer longer higher heat days and low water levels due to droughts.

**Conclusions:**

The databases created can be utilized for wildfire research or exported and transformed again with more data added. Captain Planet would be proud of the work we’ve done with data, and the result is available for the world to see. Hopefully in the future gender reveal parties will be left to inside activities that don’t include burning a forest down.