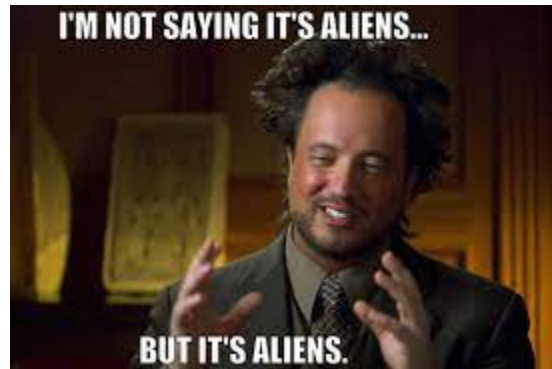


UFO Technical Report

Project 2: ETL Challenge

This project was completed by Rosaicela, Bitty, and Kellen.



Extracting

Data Collection:

With recent news coverage of the upcoming Congressional UAP report due later this month (June 2021), and increased media coverage of the UFO phenomenon, our database consists of UFO sightings within the United States for the year 2016 and documented astrological fireball sightings within the United States for the same period. We then compared the sighting locations of UFO and fireballs to known U.S. airports/airfields within 50-miles of the sightings in addition to a keyword search of U.S. military installations/locations within that same 50-mile radius. Our final step was a merge of the data to see if any of the known fireballs (obtained from NASA via Kaggle) could be associated with any of the UFO sightings.

Using the Google Maps API, we used geo-location data (Latitude and Longitude) to roughly pinpoint sighting locations. Our initial plan was to use a .csv of U.S. airports/airfields (including military airbases), but we decided that it was simpler and more efficient to use the Google Maps API instead.

The following sources were used for data:

- <https://data.humdata.org/dataset/ourairports-usa> -- The Humanitarian Data Exchange -- (us-airports.csv)*
- Fireballs (Sightings): <https://www.kaggle.com/nasa/fireballs> - Kaggle (via NASA) -- (cneos_fireball_data.csv)
- <https://www.kaggle.com/salmanfaroz/ufo-sightings-in-the-us> -- UFO Sightings in U.S. -- (ufo_location_shape.csv)
- <https://public.opendatasoft.com/explore/dataset/military-bases/table/> -- Open Data Soft -- (military-bases.json)*

***note: we ultimately ended up not using these datasets**

Transformation

Data Cleansing:

We used Python/Jupyter Notebook to import the data we collected from their original .csv and .json formatting into dataframes. Each of us took a section of the data to cleanse. Due to the time constraints around this project, we dropped dates that were outside of our year of interest, 2016, and converted the dates to datetime format.

Fireball Data Cleansing:

We needed to make the latitude and longitude columns into a format that was consistent with the other dataframes. The latitude and longitude information in this dataset was initially formatted with compass point letters rather than positive and negative numbers.

UFO Data Cleansing:

We filtered the UFO data to only show UFO sightings in the U.S. We used Google Maps APIs to grab the latitude and longitude information from the approximate locations of the UFO sightings. We then used Google Maps APIs to pull the closest airport locations into our UFO database, using the latitude/longitude information we pulled from the previous API query.

Loading

Python and SQLAlchemy:

Using the dataframes we created in Pandas/Python, we used the `to_csv()` function to export and read the data frames and used the `to_sql` function to push the dataframes and create the corresponding tables in SQLAlchemy which was then put into Postgres.

The tables are being joined by date - so as to determine if any dates of known fireballs coincide with reported UFO sightings. We set the primary keys using the Serial Key function in Postgres.

The tables in our project were: `ufo_airports` and `ufo_fireballs`.

- The `ufo_fireballs` table contains the date in 2016 of a fireball, and the latitude and longitude of the sighting.
- The `ufo_airports` table contains names of U.S. civil and military airports and physical addresses. However, due to the nature of our 'military' keyword search using the Google Maps API many of the 'military' locations are any U.S. military-related location (including air bases, recruiting stations, national guard facilities,

etc). This table additionally contains the 2016 date of UFO sightings, country of the sighting (U.S. only in our case), shape of the UFO, location (non-approximate), U.S. state of the sighting, and the latitude and longitude of the sighting.

Future considerations:

If we were to have more time with this project, we would create a distance matrix request to measure the distance between our UFO sighting locations, the U.S. military bases, fireball sightings, and airports. This would allow us to have a column that confirms or denies the possibility that some of these UFO sightings may have been airplanes or fireball sightings - and not aliens!

As it stands, we have fireball sightings that include more than just locations within the U.S. We would require further research to find out what the sighting range would be for the fireballs. This way we could determine if a fireball sighting in one country could be viewed in another country.

Image: obtained from Google (presumably a screen grab meme of a History Channel special)