**ETL Project: Bigfoot and UFO Sightings**

Clay Bazzle, Peter Drozdzewicz, Rodercick Gee, Valerie Pippenger, Donald Yakam Ngandeu

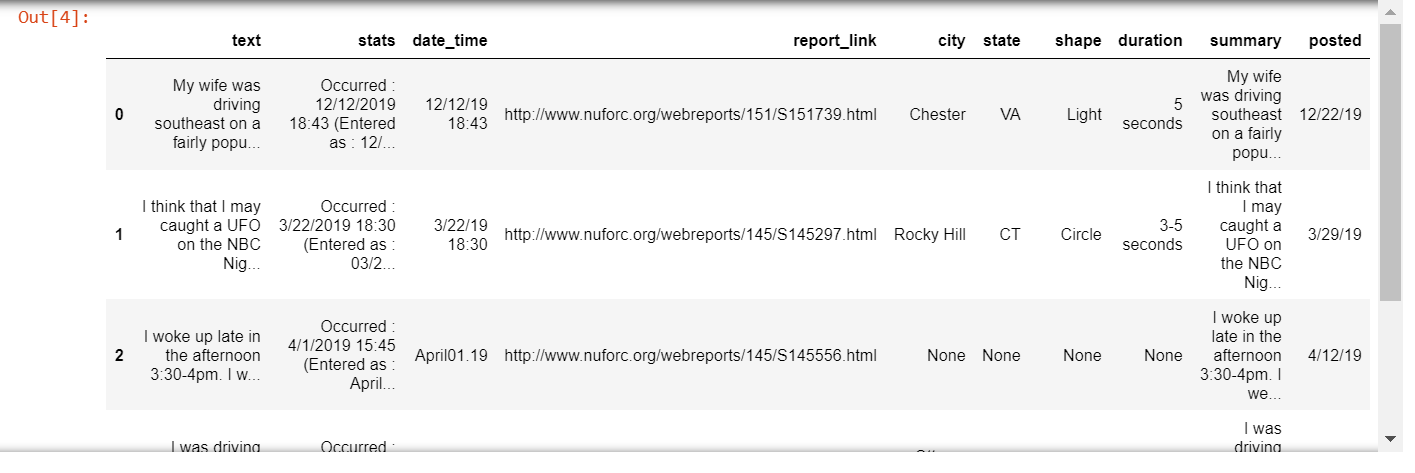
**Extraction**

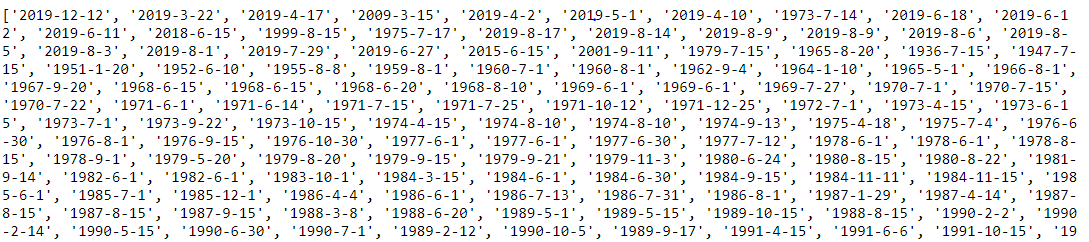
The group used two datasets from data.world. One documented Bigfoot sightings and the other documented UFO sightings. Both datasets were in .csv and JSON formats and the group extracted the .csv version of Bigfoot sightings and the JSON version of UFO sightings.

**Datasets used:**

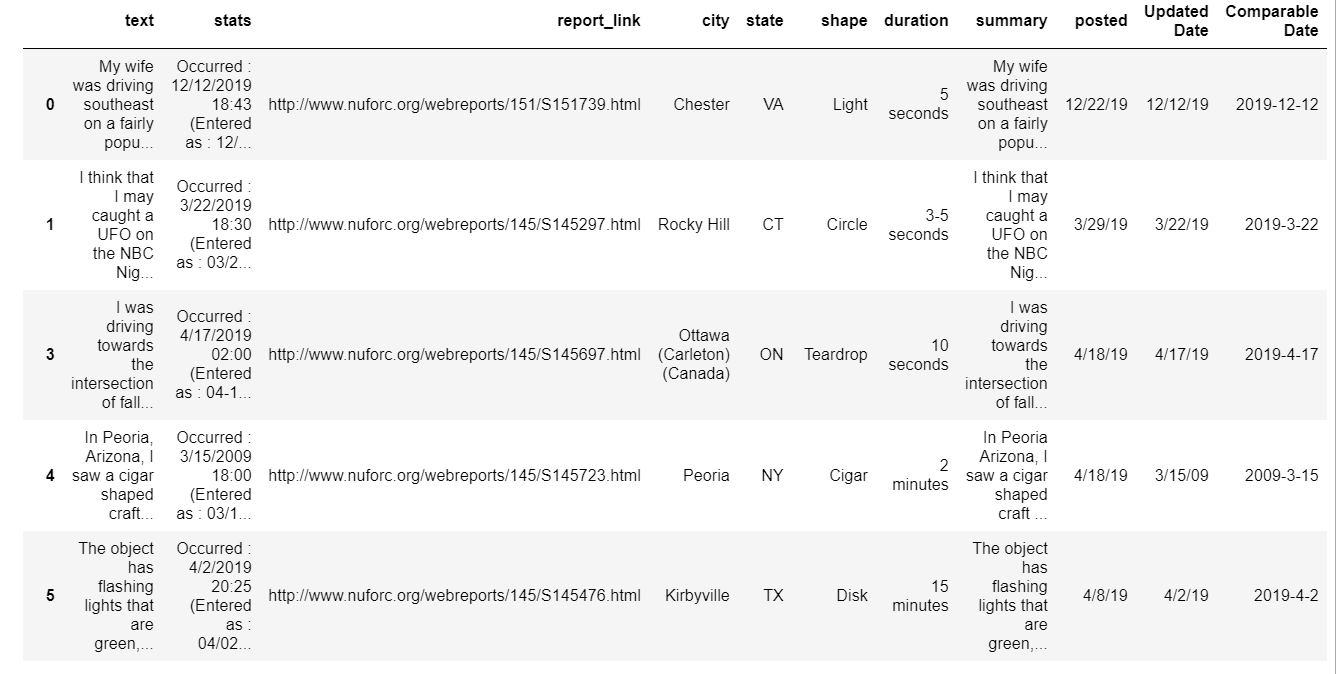
* Bigfoot sightings: <https://data.world/timothyrenner/bfro-sightings-data/workspace/file?filename=bfro_reports_geocoded.csv>
* UFO sightings: <https://data.world/timothyrenner/ufo-sightings/workspace/file?filename=nuforc_reports.json>

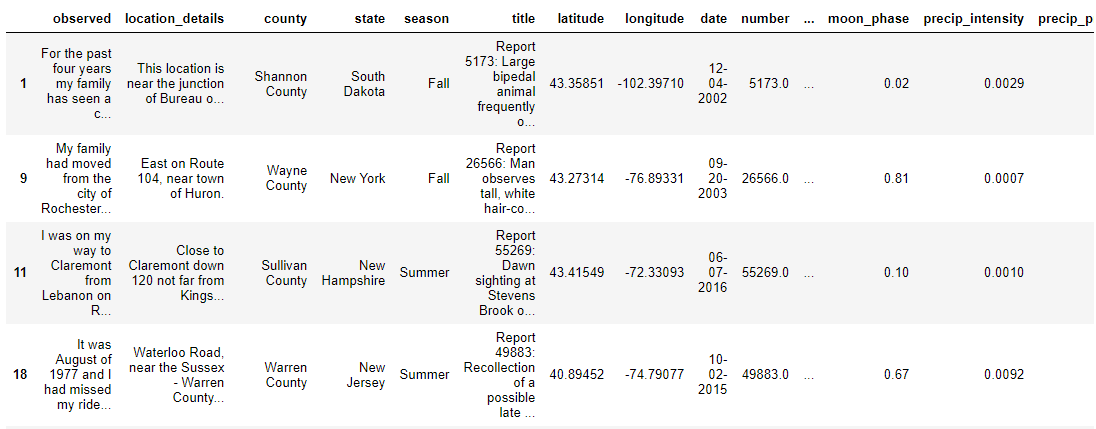
**Transformation**

The group used pandas in Jupyter Notebook to load the .csv and JSON files and turn them into DataFrames. Both datasets included rows missing data, and those were filtered out of each DataFrame once they were loaded into pandas. The dates in the Bigfoot data included time and a different format from the UFO data, which did not include the time. The time was deleted and the date was changed to match the format in the UFO dataset (yyyy-mm-dd).

*UFO Original DataFrame: New date column with the time removed. Rows with missing data have been removed.*

*UFO sightings: Partial list of dates with format changed to match dates of Bigfoot sightings.*

*UFO sightings: Updated Dataframe comparable to Bigfoot Dates*



*Bigfoot DataFrame: Rows with missing data have been removed.*

**Loading**

After the data was cleaned, it was loaded into MongoDB. The group chose MongoDB because there are disparate types of data in the two datasets. In case we wanted to use them in the future we wanted to have the datasets in a non-relational database.

**Data Use**

The purpose of this process was to ultimately be able to see if there was any correlation between UFO and Bigfoot sightings by date and/or location.

