**Dataset Analysis:**

**Positives about the dataset: nuforc\_reports.csv**

* Freely available for public use
* Easy way for witnesses to give an account of a UAP sighting without peer judgment, interference, or retaliation
* Can be combined with other datasets for cross-referencing and for further analysis
* Contains geocoded data for many US cities in city latitude and city longitude columns

**Problems**

* Editorial judgment passed on in some reports in the summary and/or text columns: certain
* reports labeled as ((HOAX))
* Hoax filter relies on editorial judgment about the sincerity of reports, but the filter doesn't appear to
* affect the ranking of the data when it comes to grouping them by count (e.g. highest counts
* by city, highest counts by year)
* Collection bias: English-only language
* Numerous misspellings, formatting anomalies, and a few out-of-bound errors in city, state, country, date\_time, duration, and shape columns
* I wrote a fair amount of code to "fix" the city, state, country, date\_time, and duration columns so I may have introduced some errors into the dataset that may not have been there.

**Model Analysis:**

<b>**Metric Table for Prophet time series model**:</b>

A white background with black text

Description automatically generated

1. Prophet: results were fair.

2. Naive:

**Metric Table for LinearRegression, Lasso Regression, DecisionTreeRegressor, RandomForestRegressor, and SGD Regressor models**

A black text on a white background

Description automatically generated

3. LinearRegression: results were unremarkable.

4. Lasso Regression: results were unremarkable.

5. DecisionTreeRegressor: results were fair.

6. RandomForestRegressor: results were good.

7. Stochastic Gradient Descent Regressor models: results were unremarkable to poor

</pre>

<b>**Apparent Trends**:</b>

A screenshot of a graph

Description automatically generated- See the breakdown of trends by the Prophet model's forecast charts:

Weekly trend

Seasonal trend

Daily times

A graph with a red line

Description automatically generated- The trend line has a positive slope, which means that UFO sightings will generally increase.

- The probability of witnessing a UAP sighting is rare: Percent probability ranges from a mere 0.0001% to 0.0003% for the vast majority of the observations collected (around 105,000 observations out of 139,674 observations that frequencies could be computed). See percent\_prob\_hist.png.

- Certain cities like New York, Phoenix, Seattle, and Las Vegas, have greater frequencies of sightings, as well as certain states such as CA, Fl, WA, and TX.

- When it comes to probabilities, the top cities were Tinley Park (2004:0.014%), Portland (2014:0.008%), Phoenix (1997:0.008%), New York City (2014:0.008%), and Seattle (2012:0.007%).

**Supplemental Questions Analysis:**

<ul>

<b>**General Observations**:</b><br><li>

The US is the leading country with the most UAP sightings, with Canada and the United Kingdom as the countries having the next highest sightings. Being that the NUFORC is an American organization, the NUFORC would naturally collect more US reports than in other countries. This collection bias is probably due in part to people from other countries not being aware of the NUFORC's existence. (I didn't know of it until a few months ago.) At this time, the NUFORC appears to accept only English reports. This fact excludes non-English-speaking people. Now we have AI language-translating apps that may make it easier for reporting through nuforc.com. People from countries where computer resources are limited and web access is intensely censored are also less able to make reports.</li>

<b>**Apparent Trends**:</b><br>

<li>The most popular months for UFO sightings are July, August, and June, in that order. See graph XX.

<br>[A little speculation as to reasons behind this include: higher visibility in summer months for North America (where most reports are located), more summer activities involving being in the outdoors or stargazing, and more people taking vacations during the summer and being more attentive to nature.]</li>

<li>According to the year by counts line graph (see graph XX), the trend line for counts of UFO sightings shows a general postive increase. The counts peaked at an all-time high in 2014 with 8474 sightings. However, there are recent fluctuations after 2014, where the counts dipped in 2018 and 2012. </li>

<li>If we get more actual data from the US government, especially from military sources, we can gain a more complete picture of what the sightings data is telling us and gain more confidence in the integrity of the data.</li>

</ul>