UFO sightings – Meet ET in person

<https://www.kaggle.com/NUFORC/ufo-sightings>

80,000 records – around 68k when cleaned

We’ll use ML to predict where the next UFO sighting will take place. For example: Predict likelihood of which state, season, time of day, shape.

What areas of the country are most likely to have UFO sightings? California. Why? Aliens want to meet celebrities. Aliens like to go to warm vacation locations just like the rest of us. People use too much hallucination drugs thinking they are seeing ufos.

Are there any trends in UFO sightings over time? Do they tend to be clustered or seasonal?

-------------

Minimal data cleaning – remove rows with empty cells for state/city (use **.dropna()**)

Filter data to US only (also some dates are off)

Create bin for Months, Seasons, Morning/afternoon/night

Use **Tableau** to display UFO sightings by state (calculate seasons, duration, shape) US only

Use **Machine Learning** to predict next ufo sighting by state, season, shape, duration. Use 2 different models to predict and check score. Use “.get\_dummies()” to convert string columns into integers if necessary.

One long html page using AOS Javascript <https://michalsnik.github.io/aos/>

This page will have Tableau integrated and our machine learning findings, matplotlib charts.

Locally hosted html or if time allows use AWS to host the html page and data file

Also, if time allows we can import the data into a database in AWS.

**Technology used (project requires at least 2, we will use more than 2):**

Python Pandas – cleaning data

Machine Learning – scikit-learn

Matplotlib – machine learning charts

Tableau – map and additional charts

HTML, CSS, Bootstrap – present data

Javascript – AOS (dynamic elements)

**Optional if time allows:**

AWS – host the html page store csv

AWS Database – store and retrieve data

D3 or Plotly for ML charts