**PROJECT 4 - UFO Sightings**

**What is the purpose and scope of the project?**

For Project 4, participants Kevin Lindstam, Duy Nghiem, Justin Trout and Yvonne Duncan will work as a group to create an presentations for users to view UFO Sightings. The aim of our project is to discover where UFO sightings occur, determine if UFO sightings occur in more locations than others, and to use ML to determine predictive analysis with this UFO sightings dataset.

**What are the sources and formats of the data files?**

– Kaggle - [UFO Sightings (kaggle.com)](https://www.kaggle.com/datasets/willianoliveiragibin/ufo-sightings)

– ufo-sightings-transformed.csv - row heading information:

* Date\_time
* Date\_documented
* Year
* Month
* Hour
* Season
* Country\_Code
* Country
* Region
* Locale
* Latitude
* Longitude
* UFO\_shape
* Length\_of\_encounter\_seconds
* Encounter\_Duration
* Description

**How are the data files organized and named?**

Project4

ufo-sighting-transformed.csv

UFO\_sightings\_cleaned.ipynb

Cleaned\_ufo\_data.csv

UFO\_sightings\_initial\_analysis

**List out processes**

* Project ideation: possible ideas: 1) Flight Price Prediction, 2) UFO Sightings
* Data fetching/API integration: pulled data from kaggle
* Data analysis:
  + Bar plotted number of UFO sightings by year
    - This plot didn’t display as nicely as it would if we focused on the last 30 years, replotted from 1997-2014
  + Discovered the first year of a recorded UFO sighting was 1910 and the total number of sightings from 1910-1996 was 8,661
  + Discovered the total number of sightings from 1997–2014 was 69,176
  + Pie plotted the distribution of UFO Shapes (initial)
  + Re-plotted the distribution of UFO Shapes combining like shapes - output: Light = 30.8%, Circle/Circle-like = 30.6%, Other/Unknown = 23.6%, Triangle = 12%, Formation = 3.1%
  + Plotted the geographical distribution of UFO Sightings
  + The country with the most UFO sightings is: United States
  + Pie plotted countries with more than 500 UFO sightings: U.S. = 91.5%, Canada = 4.7%, United Kingdom = 3%, Australia = .8%
  + Bar graphed UFO Sightings by “State” in the United States, top 5: California, Florida, Washington, Texas and New York
* Building the ML model
  + Possible analyses and predictions:
    - 1. Predicting UFO Shapes: predict the UFO shape based on other features in the dataset. This could be treated as a classification problem.
    - 2. Temporal Patterns: predict the month or season when UFO sightings are more likely to occur.
    - 3. Geographical Analysis: identify regions or countries with a higher likelihood of UFO sightings.
    - 4. Encounter Duration Prediction: predict the length of the UFO encounter based on other features. This could be treated as a regression problem.
    - 5. Text Analysis of Descriptions: use natural language processing (NLP) techniques to analyze the text in the 'Description' column, could identify common themes, keywords, or sentiment associated with UFO sightings.
    - 6. Anomaly Detection: Treat UFO sightings as anomalies and use machine learning to detect unusual patterns or outliers in the dataset.
    - 7. Cluster Analysis: Apply clustering algorithms to group similar UFO sightings based on features such as location, date, or duration.
    - 8. Time Series Analysis: Treat the dataset as a time series and use machine learning to forecast future UFO sighting trends.

**What are the programs used for data processing, analysis, and visualization?**

csv  
jupyter notebook  
pandas  
Matplotlib.pyplot as plt  
geopandas as gpd, from shapely.geometry import Point

**Slide deck found HERE:**