**Project Proposal**

**Title:** UFO Sightings

**Team Members:**

* Justin Nolan
* Darell Johnson
* Nathan Roller
* Ruby Bhatia

**Project Description/Outline:** This project aims to leverage machine learning to analyze patterns in UFO sightings and identify potential correlations with weather data. By exploring temporal and geographic trends, the project seeks to determine if certain weather conditions, seasons, or other variables correlate with an increase in reported sightings. We will use Scikit-learn for machine learning modeling and focus on building a front-end interface that can retrieve predictions and visualizations via an API. This project will showcase machine learning’s capability to uncover patterns in observational data, especially when exploring potentially obscure correlations.

**Research Questions:**

1. Which states and cities report the highest concentration of UFO sightings?
2. How have UFO sighting frequencies changed over the years, and are there noticeable monthly or seasonal patterns?
3. What are the average and most frequent durations of reported sightings, and do these vary by location?
4. How do descriptions of sightings vary across locations and time, and are there common characteristics reported in hotspot areas?
5. What are the most commonly reported shapes of UFOs, and how do these shapes correlate with specific locations or geographic regions?

**Datasets to be Used:**

* **Primary Dataset:** UFO sightings data (ufo\_sightings.csv)
* **Secondary Dataset:** Weather data (complete.csv)

**Technologies to be Utilized:**

* **Machine Learning Libraries:** Scikit-learn
* **Data Processing and Visualization:** Python (Pandas, Matplotlib), JavaScript (Plotly, Leaflet for map visualizations)
* **Database Options:** SQL or MongoDB for storing processed data
* **Web Interface and API:** HTML/CSS/Bootstrap for UI, connected to backend machine learning predictions via API

**Rough Breakdown of Tasks:**

1. **Data Collection and Database Setup:**
   * Verify and import UFO sightings and weather data.
   * Set up a SQL or MongoDB database to store and manage data efficiently.
2. **Data Cleaning and Preprocessing:**
   * Handle missing data and ensure compatibility across datasets for merged analysis.
   * Format data for input into Scikit-learn models.
3. **Machine Learning Analysis:**
   * Apply clustering algorithms (e.g., K-Means) to identify common factors in sightings.
   * Use classification models to predict sighting likelihood based on weather and location data.
   * Evaluate model performance and refine as needed.
4. **Visualization Development:**
   * Develop interactive geographic visualizations showing sighting hotspots and weather trends.
   * Implement a machine-learning-powered visualization in Plotly or Leaflet to update clusters based on new data.
5. **Front-End Interface Development:**
   * Create an interactive web interface allowing users to select geographic areas and weather conditions.
   * Implement API functionality for real-time model predictions.
6. **Final Integration and Testing:**
   * Conduct comprehensive testing to ensure usability and correct predictions.
   * Address ethical considerations and document results.