

# Using Historical Reports of UFO Sightings from NUFORC Dataset to Predict Future Sighting Trends

Shreyansh Misra, Lawrence Meng

2024-12-10

The NUFORC Databank is the largest independently collected set of unidentified flying object (UFO) sighting reports available on the internet. This analysis aims to uncover trends in UFO sightings by addressing three key questions:

1. **Where are UFO sightings most frequent?** Are they concentrated in specific countries or states?
2. **When are UFO sightings most likely to occur?** Are they influenced by seasonal patterns?
3. **What are the most commonly reported UFO descriptions?** What shapes stand out in accounts?

By investigating these questions, we aim to identify trends in UFO sighting reports. Furthermore, we aim to use predictive models like time series to forecast future sightings. This approach will not only highlight historical patterns but will also help us anticipate future occurrences of UFO sightings.

## 1 Data Preperation

### 1.1 Overview

The dataset contains 80,332 records of UFO sightings, each providing information on the location, timing, and characteristics of the sighting. While the dataset excludes records with insufficient information, some variables still contain missing values. These gaps are addressed on a case-by-case basis during our analysis. We also exclude outlier detection from this investigation as our goal is not to validate the accuracy of reports but to identify trends and patterns across the dataset.

### 1.2 Variables

- **Datetime:** date and time of each sighting.
- **City/State/Country:** location of the sighting, given by city, state, and country.
- **Shape:** shape of the UFO, categorized into multiple levels (explored in Analyzing Descriptions).
- **Duration (seconds):** length of time the UFO was observed.
- **Latitude/Longitude:** geographic coordinates of the sighting.
- **Comments:** free-text eyewitness accounts.

#### 1.2.1 Variables of Interest

- **Datetime**
- **City/State/Country:**
- **Shape**

These variables are essential to answering the core questions of this investigation.

## 2 Exploratory Analysis

### 2.1 Analyzing Location

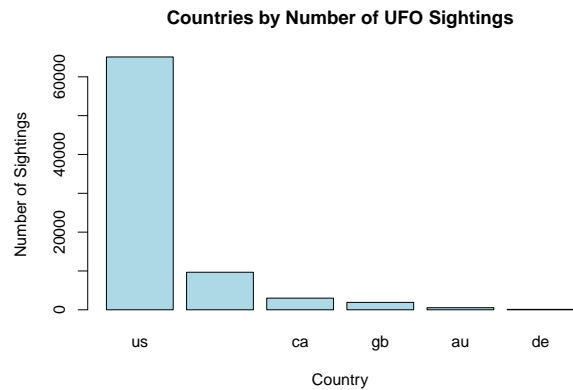


Figure 1: Countries by Number of UFO Sightings

The United States overwhelmingly has the highest number of UFO Sightings reported, followed by Canada, Great Britain, Australia, and Germany. There are also a large number of reports where the country is left blank. The NUFORC being based in the United States is likely why most of their reports are from the United States. Since the sightings from outside the US are too few to come to reasonable conclusions with, we will only use sightings based in the United States for this investigation.

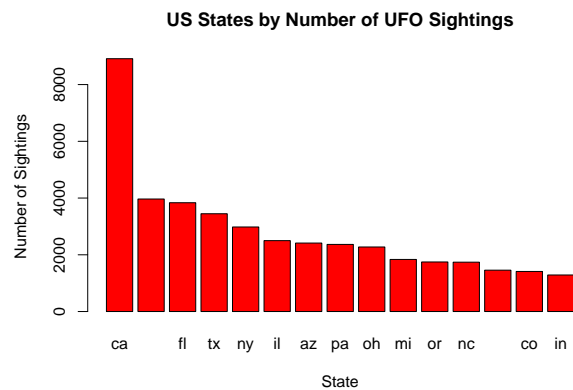


Figure 2: US States by Number of UFO Sightings

California had the most UFO sightings reported, followed by Washington, Florida, Texas, and New York. It is significant to note that California, Texas, Florida, and New York are the four most populous states in that order. Furthermore, all of those states have a coastline. California in particular has a very diverse landscape with many remote areas and a large amount of “deserted” space without light pollution that can make it easier to spot usual objects in the night sky. Those states also have a very significant military presense and drone testing, fighter plane tests, and other military activities can often be mistaken for UFOs.

## 2.2 Analyzing Descriptions

There were 29 different “shapes” that the NUFORC Databank classified each UFO sighting into. People reported what UFO looked like and it was grouped into a category based on their descriptions. While they are labelled as “shapes” in the dataset, a more accurate way to define the data column would be “descriptions” of UFOs. This is because some data points include “fireball”, “light”, “flash”, and “flare”, which are good descriptions of what the UFO would have looked like but are not explicitly shapes.

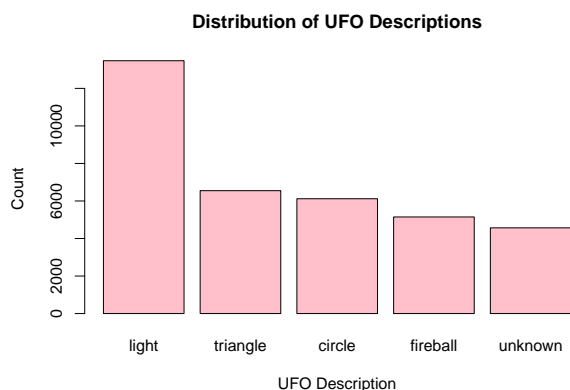


Figure 3: Distribution of UFO Descriptions, warning=FALSE

The top five “descriptions” of UFOs were light, triangle, circle, fireball, and unknown respectively. There is not much to analyze when it comes to UFO descriptions since there is a lot of variation in what people see, but it is significant to note that accounts of “light” and “fireballs” could be mistaken for shooting stars, comets, or airplane lights.

## 2.3 Analyzing Time

The dataset includes historical reports dating back to 1910. For the context of this investigation, we will be eliminating reports collected before 1990. This means that the oldest report was taken from “1990-01-03” and the most recent from “2014-05-01”.

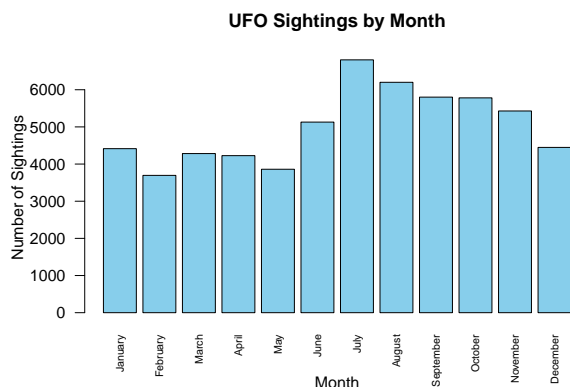


Figure 4: Distribution of UFO Sightings by Month

July, August, September, and October have the highest number of UFO sightings. These months likely coincide with increased outdoor activity due to favorable weather conditions in many regions. For example,

summer and early fall months are popular for outdoor gatherings, vacations, and celebrations, which may increase the likelihood of spotting UFOs or mistaking a flying object for a UFO.

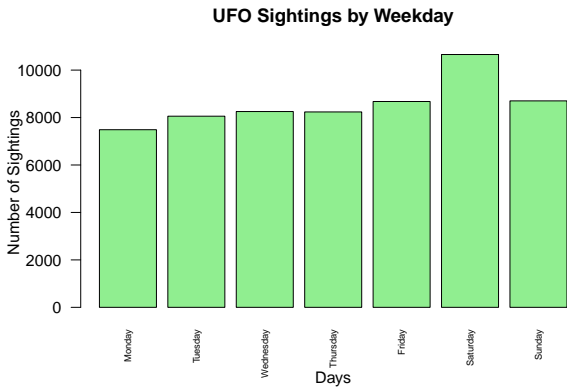


Figure 5: Distribution of UFO Sightings by Weekday

Saturdays have the highest number of sightings, with Fridays and Sundays following closely behind. This pattern aligns with increased leisure activities during weekends. People are more likely to be outdoors, socializing, or engaging in nighttime activities during these days, potentially explaining the surge in sightings.

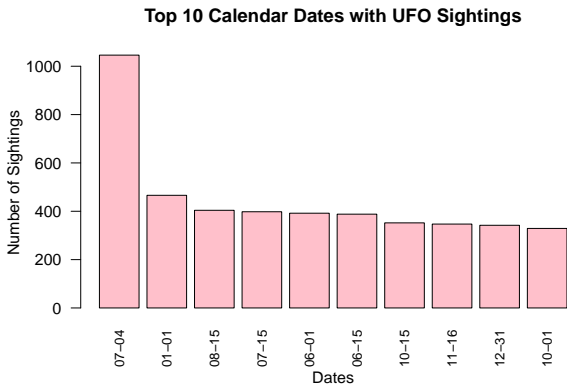


Figure 6: Top 10 Calendar Dates with UFO Sightings

The dates with the most UFO sightings across the 24 year period observed are:

- **4th of July:** U.S. Independence Day.
- **1st of January:** New Year’s Day.
- **15th of August:** A mid-summer date with unknown specific cultural or historical significance.

The 4th of July and New Year’s Day are associated with large public gatherings, fireworks, and increased alcohol consumption, all of which could contribute to more frequent sightings. We also notice that prevalence of sightings in July through October aligns with seasons where outdoor nighttime activities are common. Furthermore, sightings peaking on Saturdays and Fridays could reflect the combined effect of leisure time and increased alcohol consumption, which might lead to more reports.

Alcohol consumption in particular could lead to the misinterpretation of ordinary events such as fireworks, airplanes, skyscrapers, or shooting stars, leading people to believe that they are UFOs and reporting their sightings.

## 3 Predictive Analysis

### 3.1 Time Series

From the exploratory analysis, time emerged as the primary predictor of UFO sightings, with clear evidence of seasonal patterns. We selected a time series model as the best tool to forecast future sightings from the NUFORC dataset. The target variable is the monthly number of UFO sightings, aggregated by year and month. The dataset spans a 24 year period, from “1990-01-03” to “2014-05-01”, which is aggregated to a monthly frequency. The time series model created reflects all of these attributes, providing a baseline for our predictive analysis.

### 3.2 Exploration

We first processed the data to aggregate sightings by year and month, converting the datetime column into a monthly time series. Below is a plot of the time series, which shows trends, seasonality, and occasional spikes in sightings.

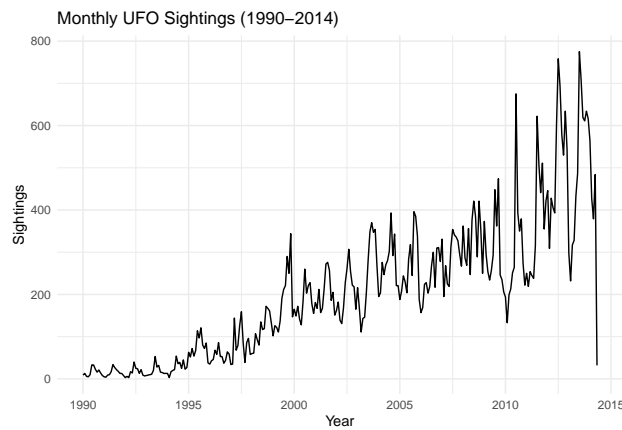


Figure 7: Monthly UFO Sightings (1990-2014)

This plot reveals:

- **Long-term trends:** Increasing overall trend in UFO sightings (multiplicative increase).
- **Seasonality:** Regular patterns of higher activity in certain periods or months.

### 3.3 Model Fitting

To forecast UFO sightings, we evaluated two models:

- **Benchmark Model:** Mean forecast.
- **ARIMA Model:** Captures trends, seasonality, and autocorrelations.

The dataset was divided into training and test sets:

- **Training Set:** Covers all data except the final 12 months.
- **Test Set:** The last 12 months, reserved for model evaluation.

### 3.3.1 Mean Forecast

The mean forecast assumes future sightings will equal the historical mean of past data. This provides a simple baseline for comparison.

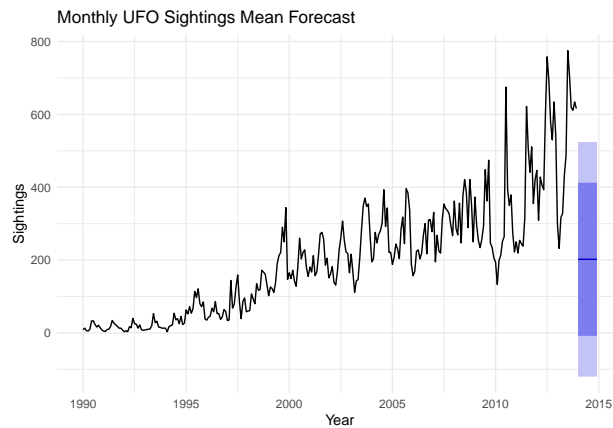


Figure 8: Monthly UFO Sightings Mean Forecast

### 3.3.2 ARIMA Model

The ARIMA (autoregressive integrated moving average) model captures additional patterns, automatically adjusting for trends and seasonality. Differencing was applied to ensure that the time series remained stationary, which we confirmed the need for after conducting the Augmented Dickey-Fuller (ADF) and KPSS tests.

- **Augmented Dickey-Fuller (ADF) Test:** -7.0054
- **KPSS Test:** 4.1877

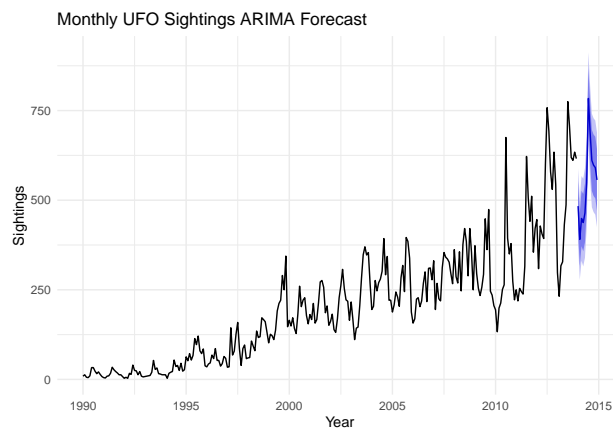


Figure 9: Monthly UFO Sightings ARIMA Forecast

Visually, we can already see a significant improvement in the ARIMA forecast as compared to the mean forecast.

### 3.4 Model Comparison

The models were compared using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), calculated on the test set.

Table 1: Performance Metrics for Mean Forecast and ARIMA

Model	MAE	RMSE
Mean Forecast	243.1875	253.5293
ARIMA	133.2323	200.9817

The ARIMA model significantly outperformed the benchmark model in both MAE and RMSE, demonstrating its ability to account for trends and seasonality. From our exploratory analysis, we were already aware that UFO sighting reports had clear trends and seasonality which made the ARIMA model a great choice for our predictive modelling.

### 3.5 Best Model Fit and Forecast

The ARIMA model, selected for its superior performance, was refitted to the entire dataset. Predictions were generated for a 36-month (3 year) forecast horizon.

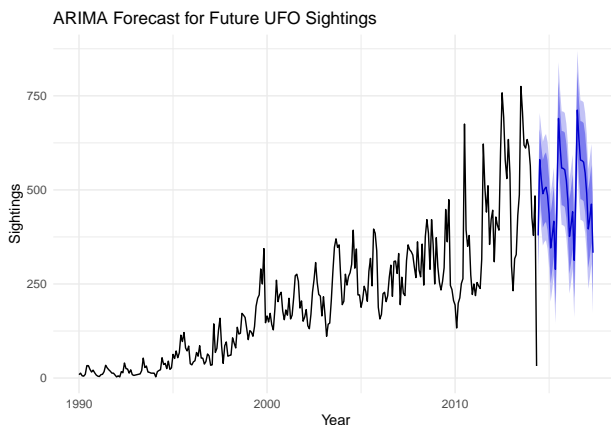


Figure 10: ARIMA Forecast for Future UFO Sightings

The forecast predicts that the upward trend in UFO sightings will continue, as seen in recent decades, although with much greater variability. The prediction interval, shaded in blue, highlights increasing uncertainty over time which is normal for long-term forecasting models. Seasonal fluctuations are still very evident in the forecast, which reflects the historical seasonality that we identified in the exploratory analysis, such as peaks during specific months.

External factors like changes in reporting behavior, technological advancements, or sociocultural events might influence future sightings, potentially impacting the accuracy of the forecasting model. While the ARIMA model very effectively captures historical trends and creates a forecast, it assumes the continuation of historical trends and cannot account for unprecedented shifts in reporting patterns.

## 4 Conclusion

### 4.1 Summary of Results

This investigation analyzed the NUFORC Databank to uncover trends in UFO sightings and predict future occurrences using predictive modeling techniques. By addressing key questions related to the timing, location, and descriptions of UFO sightings, we aimed to find patterns and generate forecasts for future observations.

From our exploratory analysis, we noticed that UFO sightings are most frequent during the summer months and on specific dates such as the 4th of July and New Year’s Day. It confirmed that time was a primary predictor for changes in UFO sightings and that reports were seasonal, leading us to move forwards with a time series forecasting model.

We compared a mean forecast model and an ARIMA model in our model selection process, and ARIMA model proved most effective in capturing historical trends and seasonality. It demonstrated a steady upward trajectory in UFO sightings over the past two decades and forecasted a continuation of this trend over the next three years, albeit with increasing uncertainty.

### 4.2 Comparison of Results and Main Questions

- Where are UFOs most likely to be sighted? UFO sightings are most concentrated in the United States, with regional patterns influenced by population and cultural factors.
- When are UFOs most likely to occur? The observed patterns reveal a seasonal and weekly cycle, with sightings peaking in the summer months and on weekends.
- What are the most common UFO descriptions? Observers frequently report simple, luminous shapes such as “light” or “circle”.

### 4.3 Final Conclusion

This investigation concludes that UFO sightings have geographic, seasonal, and descriptive trends. It also identified time as the primary predictor of UFO sightings, with the ARIMA model suggesting that sightings will continue to rise, reflecting the historical upward trend.

However, it is important to acknowledge the limitations of this analysis. The dataset relies on self-reported observations, which certainly include biases and inaccuracies. Moreover, the models assume that past trends will continue to take place, which does not account for shifts in reporting behavior or any external influences. Future investigations could explore the impacts of external variables, such as changes media coverage or technological advancements, and their effects on UFO sighting reports.

In summary, this study provides an analysis on UFO sighting trends and shares insights into the patterns that could predict future UFO sightings.

## 5 References

1. <https://nuforc.org/databank/>
2. <https://attheu.utah.edu/facultystaff/the-west-is-best-to-spot-ufos/>
3. <https://www.skyatnightmagazine.com/space-science/things-mistaken-for-ufos>
4. <https://a-little-book-of-r-for-time-series.readthedocs.io/en/latest/src/timeseries.html>