National UFO Reporting Center Data:
An Exploratory Analysis of UFO Sightings and Descriptions

Team Members: Blake Bormes, Steve Hewitt, Tymon Silva

Executive Summary

The following are the key findings from our analysis:

- The top three UFO shapes reported are “light”, “circle”, and “triangle” at 25%, 13%, and 11%, respectively.
- California has the most reported UFO sightings than any other state, representing over 12% of all reporting sightings across the country.
- The city with the most reported sightings is New York City at about 0.7% (781 UFO sightings). The city of Seattle and Phoenix were close runners up, each accounting for 0.6% of all UFO sightings.
- The most UFO sightings reported in one day occurred in 2004 on Halloween in Tinley Park, Illinois. That evening, 57 people reported seeing a UFO.
- The month of July is when UFO sightings occur most frequently. Furthermore, Saturday evenings between 9 and 10 PM are the most frequent day of the week and time of day reports of UFO sightings.
- Apparent spikes in UFO sightings occurred within the 90 days after box office release of some films such as 2001: A Space Odyssey, Alien, Avatar, and Avengers: End Game.

Overview and Background

Throughout recent years, discussions of aliens, UFO sightings, and paranormal activity have increased, including reports of top-secret UFO sightings documented by the US Government and increasing speculation of whether aliens are kept in Area 51 in Nevada. On July 8th, 1947, the first major US alien-related event was documented. Many conspiracy theories surround the Roswell Incident, many extraterrestrial believers allege that an alien spaceship crashed into the desert and both the spaceship, and its alien occupants were hidden at a top-secret location called Area 51 for further experimentation.

An exploratory data analysis of US UFO sightings was conducted to use historical data to understand where and when UFO sightings have occurred and what was seen during the sightings. The purpose of this analysis was to gain an understanding of the details of previous sightings to answer the questions:

- Where should I go to have the best chance to see a UFO?
- When should I go to have the best chance to see a UFO?
- What should I expect to see when viewing a UFO?

This analysis uses data from the National UFO Reporting Center on 134,057 documented UFO sightings with data including city and state, date and time, shape of UFO, and duration of

sighting. Further details on the datasets, data cleaning methodology, analysis, and conclusion are presented below.

Methodology

Datasets Used

UFO Data: Our primary data source was extracted from the NATIONAL UFO REPORTING CENTER website (http://www.nuforc.org/). The raw extract of this data is saved in ufo.csv for reference. The dataset contained 134,057 entries with the following information:

- Date/Time of UFO sighting (date and time)
- City (text)
- State (text)
- Shape of UFO (text)
- Duration of sighting (text)
- Summary of event (text)
- Date posted (date)

GIS Data: Longitude and Latitude data was gathered at the city level. This data was later joined to our primary data source to develop key visualizations. (http://www.geonames.org/)

Movie Data: List of 10 popular movies featuring aliens and their original theatrical release data. Selection of movies was determined by researcher discretion.

Data Cleaning and Sanity Checks

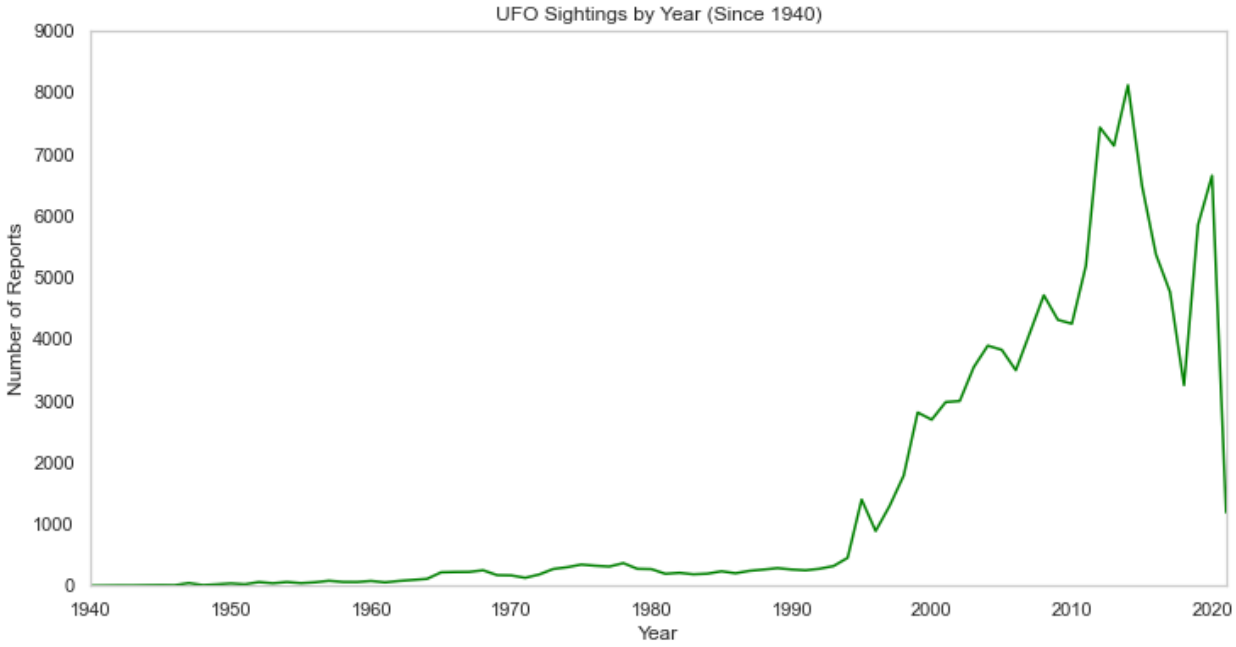
UFO Data: The UFO dataset was not available for direct download. The data needed to be extracted from several webpages. There is a directory page (http://www.nuforc.org/webreports/ndxevent.html) and this page links to 955 other pages that contain the relevant data. The HTML structure of these pages is relatively simple, with all the data on each page contained in a single table. We wrote a python script that uses Beautiful Soup to scrape the data from each of these pages, then we converted it into one Pandas dataframe and saved the result as a CSV file.

The quality of data collection in the original dataset is somewhat questionable, with many seemingly invalid data points being included.

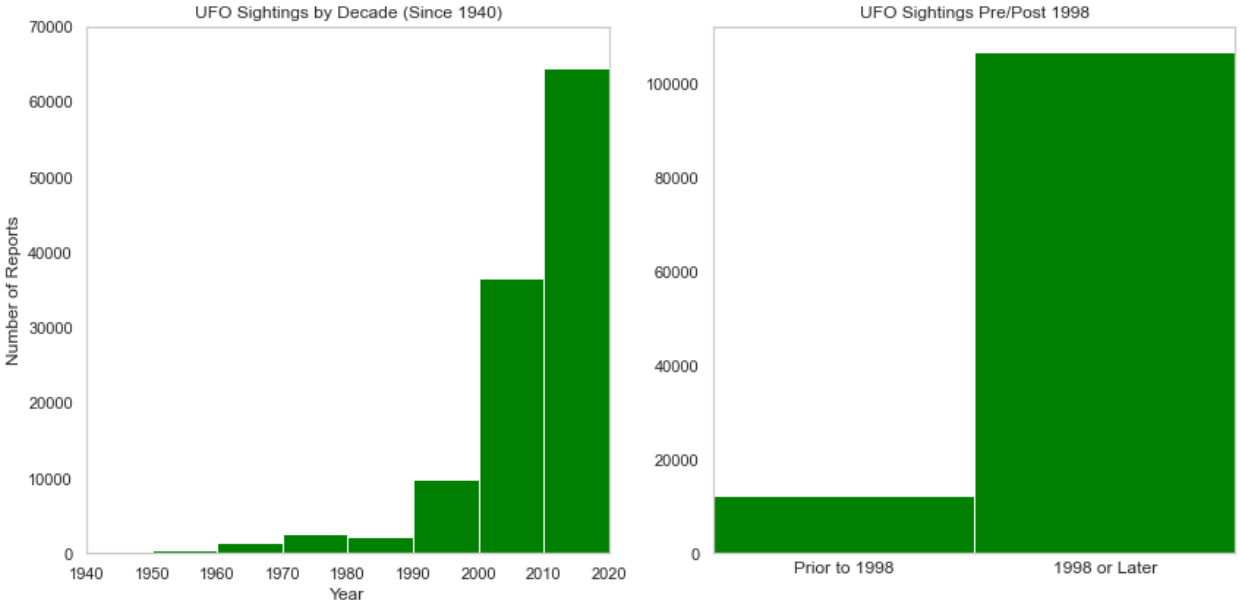
- City (included blanks and values that are not city names, some values were city/country pairs of locations outside the United States)
- State (included values that are not valid states)
- Summary (free text values, unstructured, sometimes laced with profanity)
- Duration (long-form free text values, not suitable for comparison/analysis)

The posted date indicates that this dataset was initially created in 1998, so there is a concern that as events get further back in time from this point the reliability of the data will decrease. The overall distribution of dates of events skews heavily towards events that occurred in 1998 or later. The earliest reported incident was alleged to have occurred all the way back in 1600.

Based on these facts, it was determined that the dataset should be truncated based on a minimum date. Given the significance of the Roswell Incident and our desire to group data by decades, we decided to discard events from years prior to 1940; this filtering removed less than 0.1% of all data. The chart below shows the distribution of reports by year after all data cleaning.



These additional charts below illustrate the proportion of the reports for each decade (below left) and before/after the year of the first posting, 1998 (below right). These further confirmed that our data skews heavily towards the time period after the launch of the NUFORC website in 1988.



The next important decision in the cleaning process was how to handle the location data. The NUFORC is a US-based organization that tracks national UFO sightings. Therefore, the data was primarily USA-centric. This fact, combined with the inconsistencies in the use of the city and state fields, led to our decision to focus our efforts on data related to the United States. We removed all entries without a valid US state in the state field, and those entries where the city name was not in a text format. This set of rules was the heaviest filter applied to our dataset, removing roughly 11% of the rows, but it put us in a position where we could be relatively confident that the geographic area of each incident was located within the United States, which would allow us to create interesting and insightful mapped visualizations. The city field was further modified to remove any text enclosed in brackets (e.g. “New York City (Brooklyn)” was converted to “New York City”) because this field was sometimes used to store other notes or descriptions in brackets beyond the name of the city.

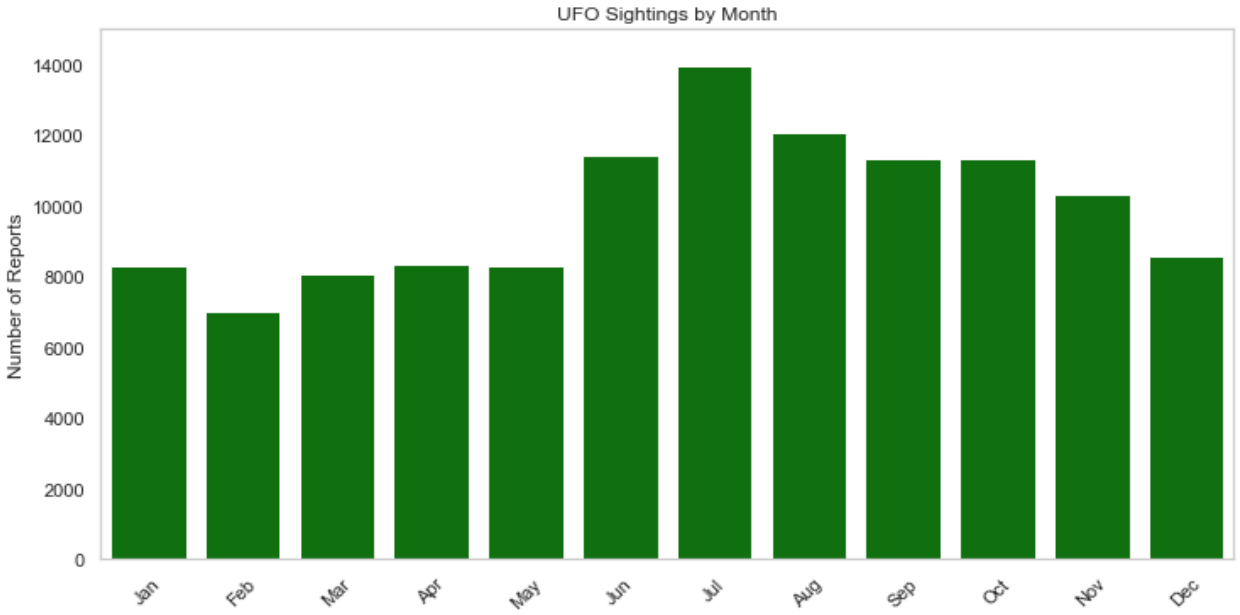
Aside from filtering, we also created some logic to clean and categorize the free text values:

- Duration was parsed using an algorithm that looked for numbers and units in the string and attempted to convert them to a standardized unit: number of minutes, and then stored in a new column. This method had some limitations such as being unable to handle all-text values like “five minutes”, and converting ranges of time into flat numbers based on the first figure in the range (e.g. “1-2 hours” becomes 60 minutes). With these limitations, the blanks in the original duration data, and values removed by a sanity check (a handful of durations longer than 72 hours); we were able to convert 92.8% of the duration data into a usable number of minutes.
- Shape date was converted into a standard case and categorized down into the top 15 most commonly reported shapes. This helped remove some of the noise from the raw text data.

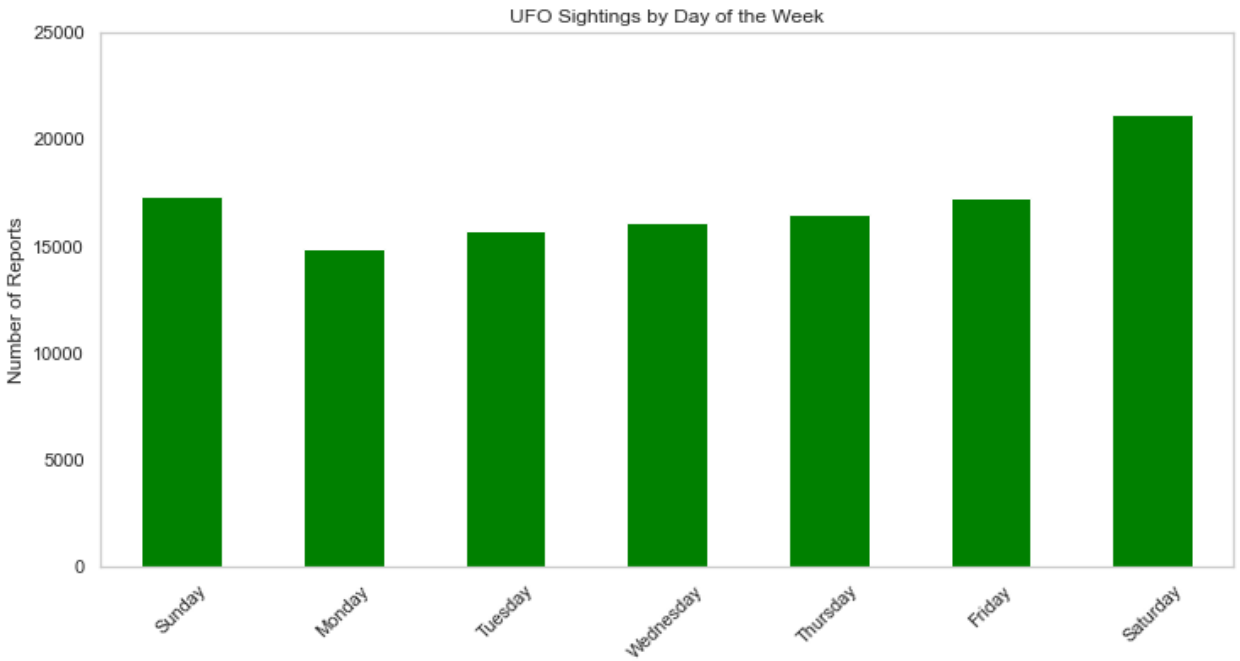
Analysis Summary

When Do UFO Sightings Occur?

To add to the data about the year of sightings, we also looked to categorize the time of sightings in a variety of other ways. The chart below shows the distribution of reports by month of the year. There is a clear increase during the warmer months, with July being the peak. In the Northern Hemisphere these correspond to the months with the most daylight hours. While it is initially counterintuitive to have more sightings when there are fewer nighttime hours, the increase in people spending time outdoors due to the warmer weather seems to be a stronger factor influencing sighting frequency. It should be noted that this chart is skewed by the fact that not all months have the same number of days.

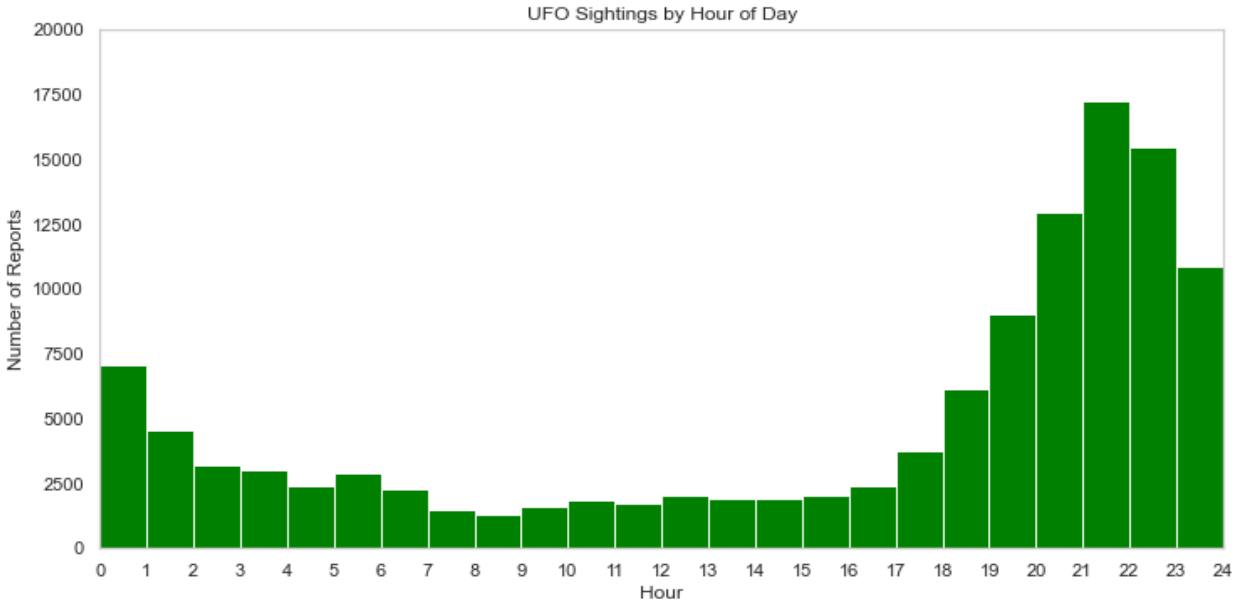


The chart below shows a breakdown by day of the week. The day of the week has a similar effect on the frequency of sightings that we saw with the months. Sightings are highest on Saturdays and lowest on Mondays. People are less likely to be working on the weekend so it seems reasonable that they would be spending more time outside and therefore have more opportunities to notice lights in the sky that they would perceive as UFOs.



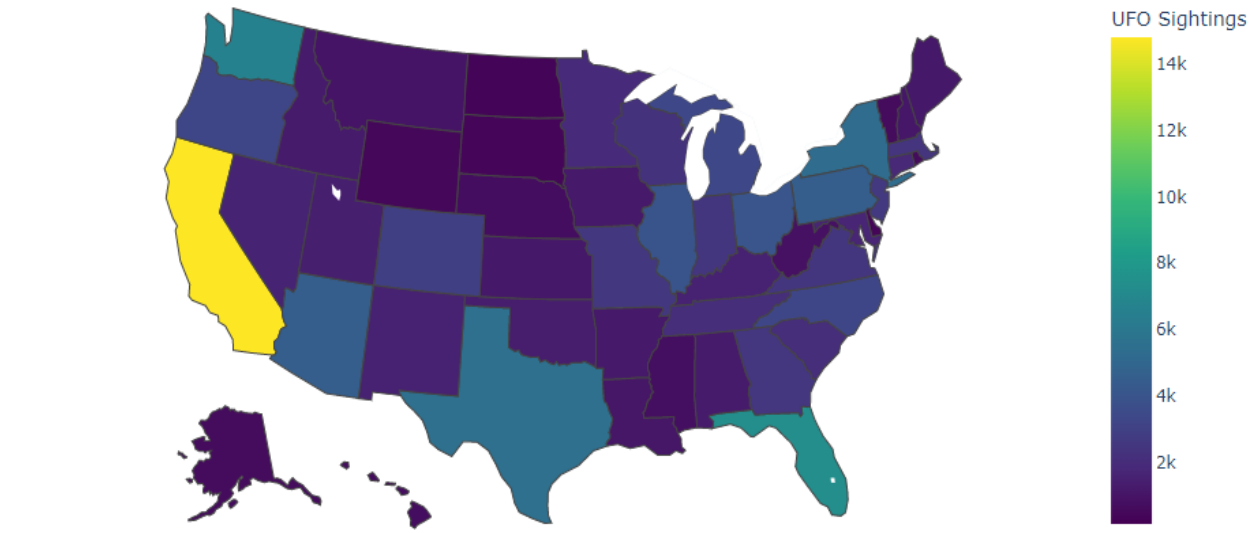
The next chart below shows a breakdown by hour of the day. This distribution aligns with our expectation that sightings primarily occur during the night. Light pollution during daytime hours makes it more difficult to see lights in the sky. The contrast of a dark background causes objects

with the same luminosity to appear more clearly than they would against a bright background. This is the same reason that the stars are only visible at night despite putting out a constant amount of light. The hour that saw the most sightings was 9-10 PM, with a very prominent grouping of the hours directly before and after. Sightings drop off precipitously after 11 PM despite still being dark outside and this can likely be explained by the fact that less people are awake during the middle of the night to potentially witness UFOs.



Where Do UFO Sightings Occur?

The State of California, by far, has the most reported UFO sightings (12.4%) when compared to other US States. The heat map below illustrates the distribution of sightings across the country.



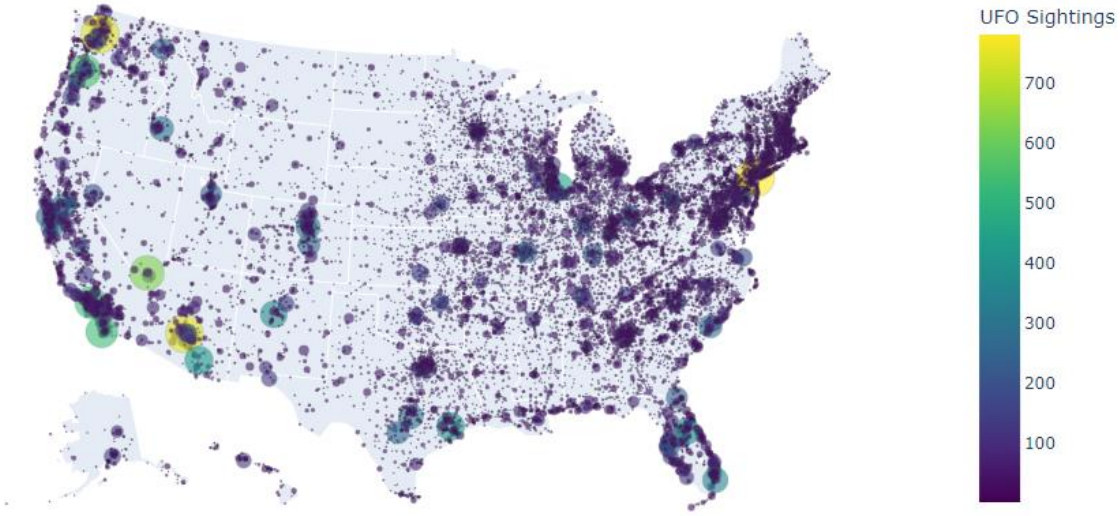
The table to the right shows the five states with the highest frequency of reported sightings. You can see that California more than doubles any other state when it comes to the number of UFO sightings. Altogether, this list represents roughly one-third of all reported sightings for the US.

State	UFO Sightings	Percentage
CA	14816	12.4%
FL	7371	6.2%
WA	6599	5.5%
TX	5531	4.6%
NY	5263	4.4%

Looking at the city level, New York City, Seattle, and Phoenix hold the most sightings at 781, 755, and 753, respectively. The table at right shows the top five cities with the highest frequency of sightings. The following bubble map helps illustrate which cities have the highest amount of UFO sightings, indicated by large yellow and green bubbles.

City	State	UFO Sightings	Percentage
new york city	NY	781	0.7%
seattle	WA	755	0.6%
phoenix	AZ	753	0.6%
las vegas	NV	628	0.5%
portland	OR	559	0.5%

Bubble Map of UFO Sightings

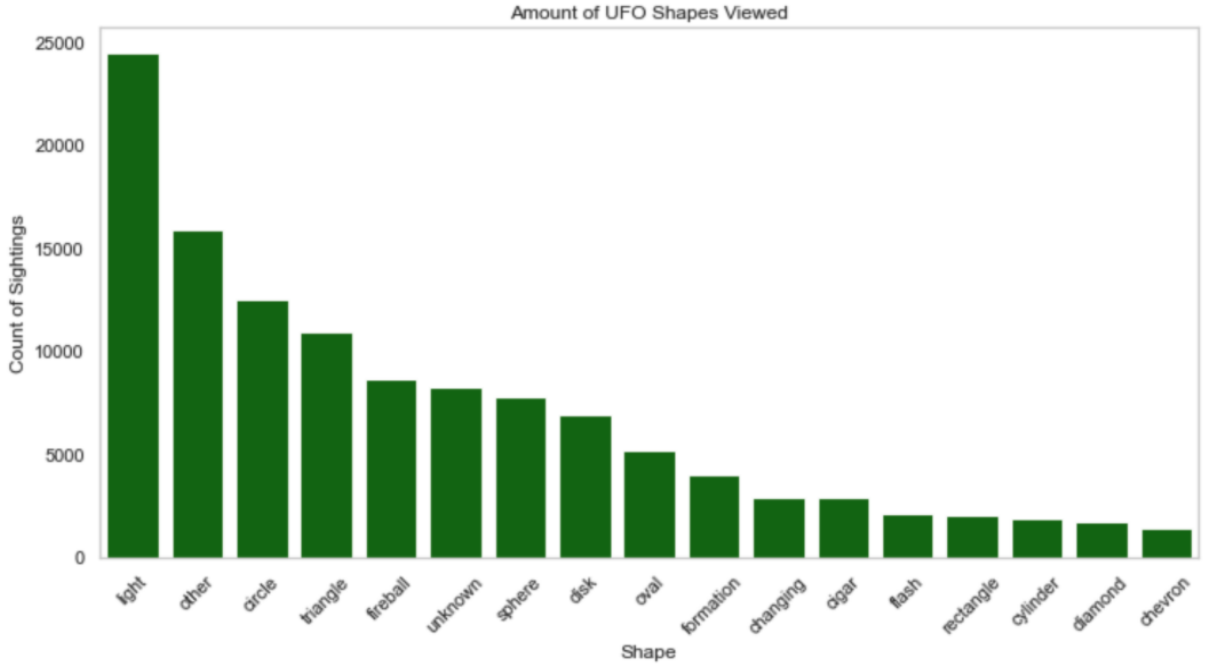


You may be asking yourself, “where and when was the most likely time to see a UFO?” To answer that question, let’s go back to October 31st, 2004. If you were out celebrating Halloween in Tinley Park, Illinois that day, that was your best chance in spotting a UFO. That date represents the most UFO sightings on a given day, which totaled 57 sightings. The table below describes other places and dates where high amounts of UFO sightings were reported.

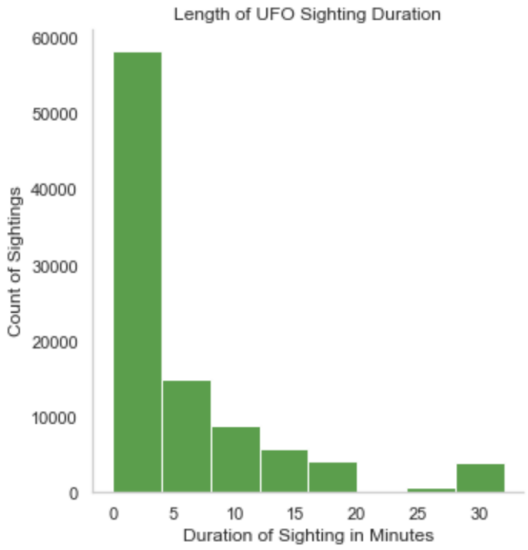
Year	Month	Day_Of_Month	City	State	UFO Sightings
2004	10	31	tinley park	IL	57
1997	3	13	phoenix	AZ	33
2004	8	21	tinley park	IL	31
2001	1	11	rockford	IL	23
2001	7	15	carteret	NJ	14

What Do UFO Sightings Look Like?

To determine what to expect when viewing a UFO, data analysis on UFO shape and duration of UFO sightings occurred. The initial analysis of UFO shape determined the number of sightings per shape given the top fifteen most viewed shapes. The most frequently viewed shapes were “light” and “other.” During data cleaning, “other” was calculated by taking the original amount of "other" shape values, which was 9,188 and changing the names of any data values that were not part of the most frequently used top 15 shapes. This resulted in a total of 15,855 occurrences of “other” shapes, because about 6,700 data entries’ shape values were changed to “other” for data consistency. Furthermore, it is interesting that light is the highest recorded shape at almost 25,000 sightings of a UFO light. This could be attributed to UFO's being so distant that it is difficult for viewers to determine a true shape aside from the bright light it produces in the sky. The next most frequently viewed shapes were circle, triangle, and fireball however, a fireball is similar to a light and a circle shape combined. Given this data, one can determine that when searching for a UFO in the sky, a bright light either in the shape of a circle or triangle is most likely to be viewed if a UFO is present.



The second part of the analysis to determine what a UFO sighting would look like included an analysis of the duration data. To gain a better understanding of the distribution of duration data values, ten bins were created across all duration data, which determined that the upper 10% of duration data was for sightings over 30 minutes to a maximum length of 3 days. Given this information, the following chart only includes data to show the distribution of sightings where the duration was 30 minutes or less.

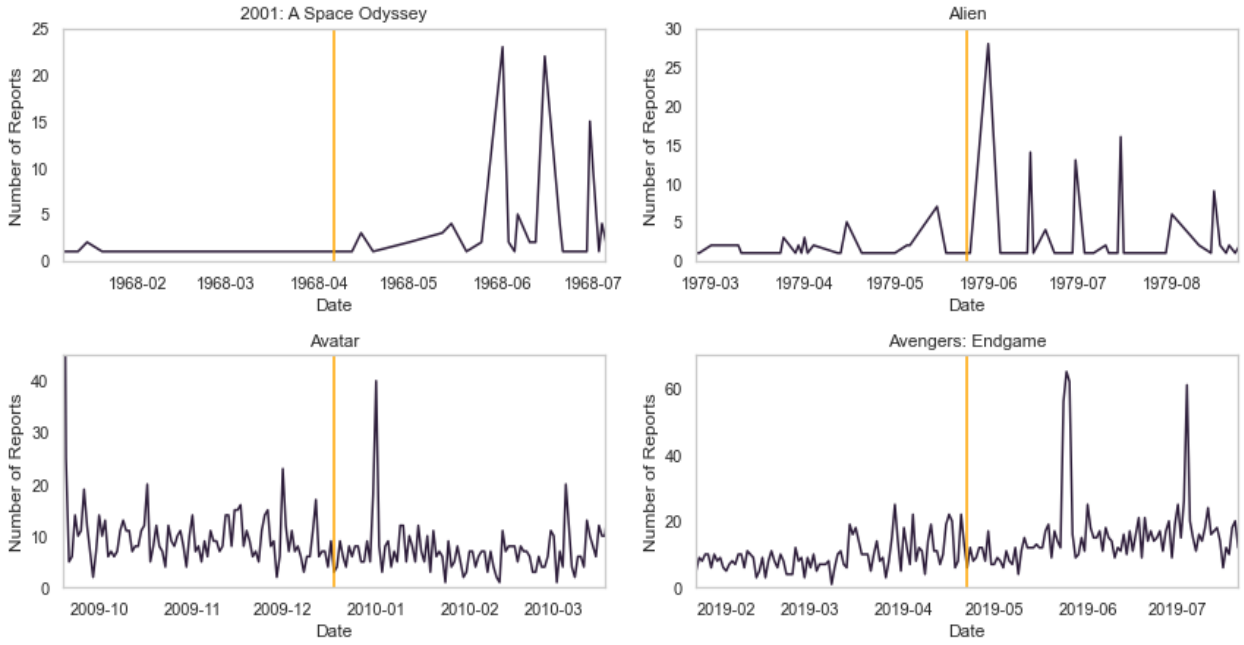


Count	105806.000000
Mean	15.163968
Std	71.106655
Min	0.000333
25%	0.500000
50%	3.000000
75%	10.000000
Max	4320.000000

The analysis of the duration of UFO sightings determined that even though the average duration of sightings is 15 minutes, 75% of sightings last 10 minutes or less. The graph further illustrates how most sightings lasted five or less minutes, which means that most UFO sightings recorded do not hover in one spot for a long time, they last only for a few minutes. Additionally, the top 10% of sightings recorded last anywhere from over 30 minutes to 3 days. This means that if someone were to try to see a UFO, they would need to ensure they are keeping a close eye on the sky because the UFO will not stay for a long period of time.

Movie Analysis

UFOs have captured the collective imagination of millions of people. The idea of “little green men” coming down from space to visit the Earth is a mainstay in popular culture. If UFO sightings are largely a byproduct of misinterpretations and imagination, then it stands to reason that rates of UFO sightings would increase following events in popular culture that stimulate the imagination. We looked for patterns in the frequency of sightings reported for the 90 days before and after the release of 10 major motion pictures that feature aliens. Some of the movies showed a noticeable spike in reports immediately following their release. The charts below show the result for a few of these movies, with the orange bar representing each film’s release date.



Conclusion & Recommendations

Our analysis has determined that summer months, especially July are when most UFO sightings occur. Additionally, Saturdays are the day of the week when most UFOs are viewed. The most likely time one can see a UFO is from 9 - 10PM at night and sightings drop off after 11 PM because less people are awake at that time to view a UFO. The times most UFOs are viewed can be attributed to people spending more time outside during warmer months, on weekends, and before going to bed. When analyzing where UFO sightings have occurred, we determined that the top five states to view a UFO are CA, FL, WA, TX, and NY, making up one third of UFO sightings. Additionally, the single most viewed UFO sighting occurred in Tinley Park, IL on Halloween 2004 with 57 sightings. When viewing UFOs, a bright light tends to be widely viewed. There is less certainty surrounding the exact shape of UFOs seen, but often a triangular or circular spaceship is recorded. Additionally, 75% of UFOs were seen over a duration of 10 minutes or less, meaning they don't usually hover in one area for a long period of time. Lastly, UFO sightings seemingly increased after the release of major blockbuster alien movies such as 2001: A Space Odyssey, Alien, Avatar, and Avengers: Endgame. This shows that the validity of each UFO sighting may be impacted by viewers' confirmation bias and are not necessarily indicative of an actual UFO. In conclusion, if you want to view a UFO, we recommend that you go to California and search for a bright circular or triangular light from 9 - 10 PM on a Saturday in July.