Unidentified Flying Object Sightings Analysis

Sathvik Maridasana Nagaraj

**Introduction:**

UFO sightings have intrigued and mystified people for decades. In this project, we delve into the National UFO Reporting Center (NUFORC) database, focusing on sightings in 2016. The aim is to uncover trends, patterns, and insights into UFO sightings. By analyzing the dataset, we seek to understand the distribution of sightings, popular shapes, and any temporal patterns that might exist.

**Dataset:**

The dataset comprises UFO sightings reported in 2016 and is sourced from the National UFO Reporting Center. The dataset is a snapshot of the UFO phenomena recorded during the year 2016 in the region of USA & Canada. The dataset was sourced from Kaggle (https://www.kaggle.com/datasets/utkarshx27/ufo-sights-2016-us-and-canada/data)

Variables:

Date/Time: Data & time of the sighting

Country: Country the sighting was reported in

City: City the sighting was reported in

State: State the sighting was reported in

UFO shape: Shape reported by the reporter

Summary: summary of the explanation by the reporter

Latitude and longitude of the sighting

**Methods:**

1. **Data Preprocessing:**

The analysis was commenced by the phase of data wrangling, laying a robust foundation for subsequent explorations. This initial step involved a thorough examination for missing values, duplicates, and other data quality considerations. The objective was to ensure the integrity of the dataset before delving into the analyses.

1. **Geographical Analysis :**

Following the data refinement, the analysis progressed to geographical plotting, where insights into the distribution of UFO sightings across both countries were sought using seaborn library. Count plot was generated utilizing the 'coolwarm' color palette for improved visibility and an interactive map was generated using the Plotly Express library, showcasing the latitude, longitude, and reported UFO shapes. The map utilized a natural earth projection and featured markers with hover information displaying sighting summaries.

Subsequent granular analyses:

* **State wise distribution**: To identify the state which has more sighting reports in both country.
* **City wise distribution:** To identify the city which has more sighting reports in both country.
* **Popular Shapes per State:** To identify the most commonly reported UFO shape in each state. The data was grouped by state and shape, and the shape with the highest count in each state was determined.

Cross-comparisons between different dimensions enriched the analysis, unveiling intricate correlations and dependencies. The application of clustering techniques further added a spatial layer to the exploration, identifying clusters of heightened UFO activity.

1. **Temporal Analysis**

Temporal analysis provided a through monthly, daily, and hourly analyses, shedding light on time-dependent trends in UFO sightings. Concurrently, shape-wise analyses provided insights into the diverse forms reported in sightings.

* **Month and Weekly Analysis:**

Analysis based on month and day of the week was explored by converting the 'Date/Time' column to datetime format. Subsequently, the sightings were grouped by month first, and the count of UFO sightings for each month was visualized using an interactive line graph. Then a bar chart was generated using seaborn, illustrating the distribution of sightings across different days.

* **Hourly Analysis of Distinct shape**

Further analysis based on the hourly distribution of sightings for each distinct shape was conducted. The dataset was further refined to include only the 'Date/Time' and 'Shape' columns. The hour with the highest count for each shape was identified, revealing peak sighting hours. This information was visualized through an interactive bar plot using Plotly Express, providing a comprehensive overview of when each UFO shape was most frequently observed.

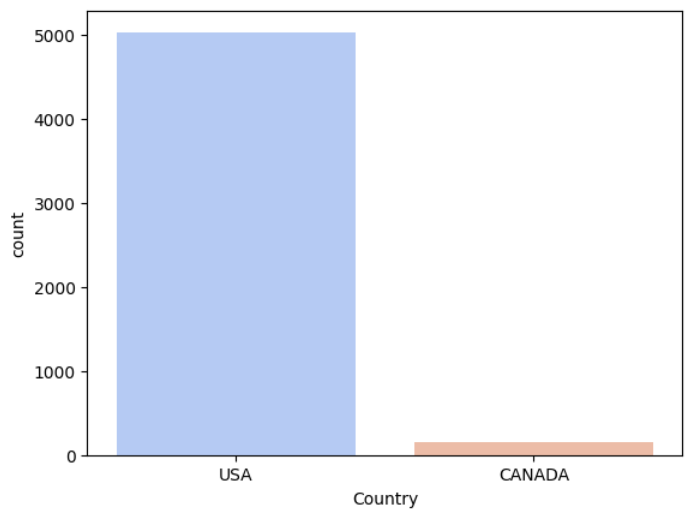
1. **Text analysis**

For a deeper understanding of the textual content in UFO sighting summaries, a text analysis was conducted. All summary texts were combined, and the frequency of individual words was determined. A word cloud was generated using the WordCloud library, visually representing the most frequently occurring words in the dataset. To add more details to the text analysis, another analysis was carried out just to check the shapes mentioned in the summary column using the bar plotting. This helped to draw the relationship between the overall summary and the shapes reported.

**Results:**

1. **Geographical Analysis**

* **Country-wise Distribution:**

****

The plot Figure (1) offered the dynamic perspective on the spatial distribution of UFO sightings in the United States and Canada. Notably, that the dataset contains limited information for Canada, prompting a shift in focus towards broader trends and analyses rather than emphasizing individual countries.

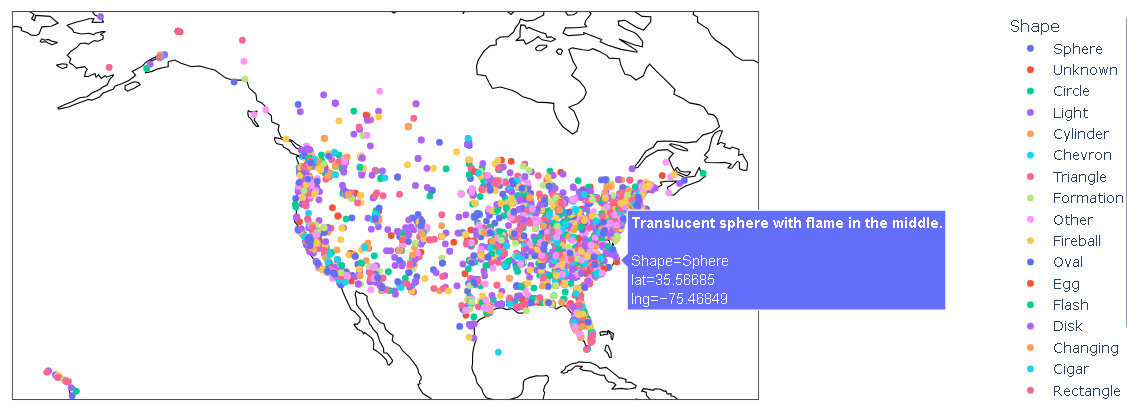
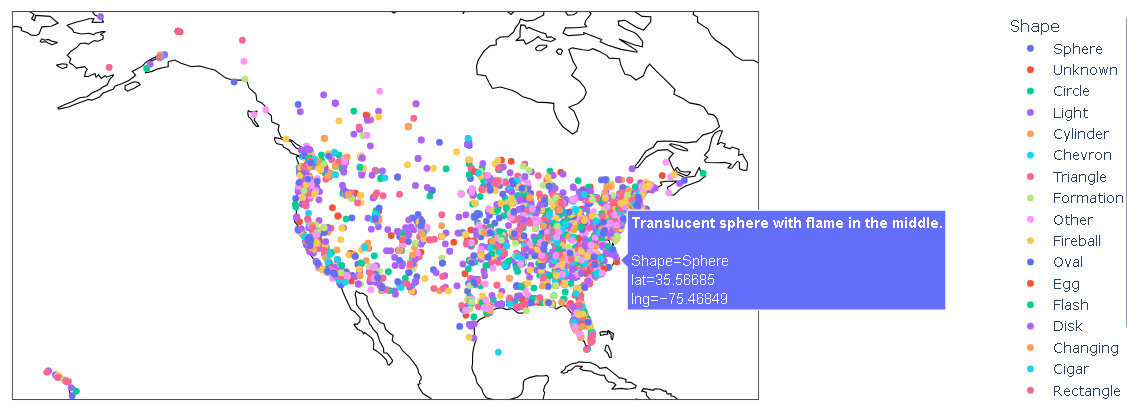
****

Figure 1

****If we look at the Figure (2) we can notice that the sighting density is more in the Eastern region compared to the Western region of the analyzed region.

We will go further and granulate the analysis to get to know more about the sightings reporting with geographical perspective.

Figure 2

* **State-wise Distribution:**

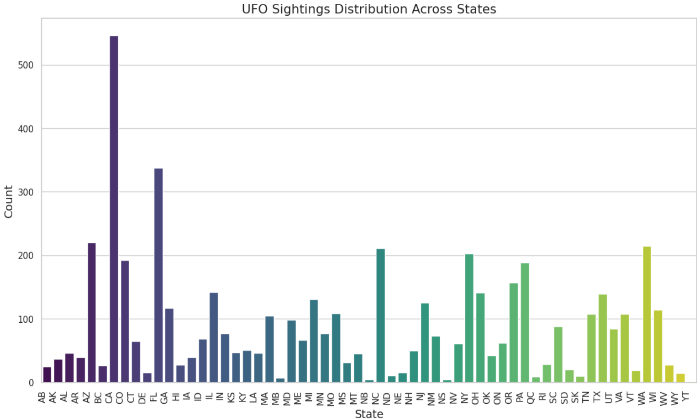
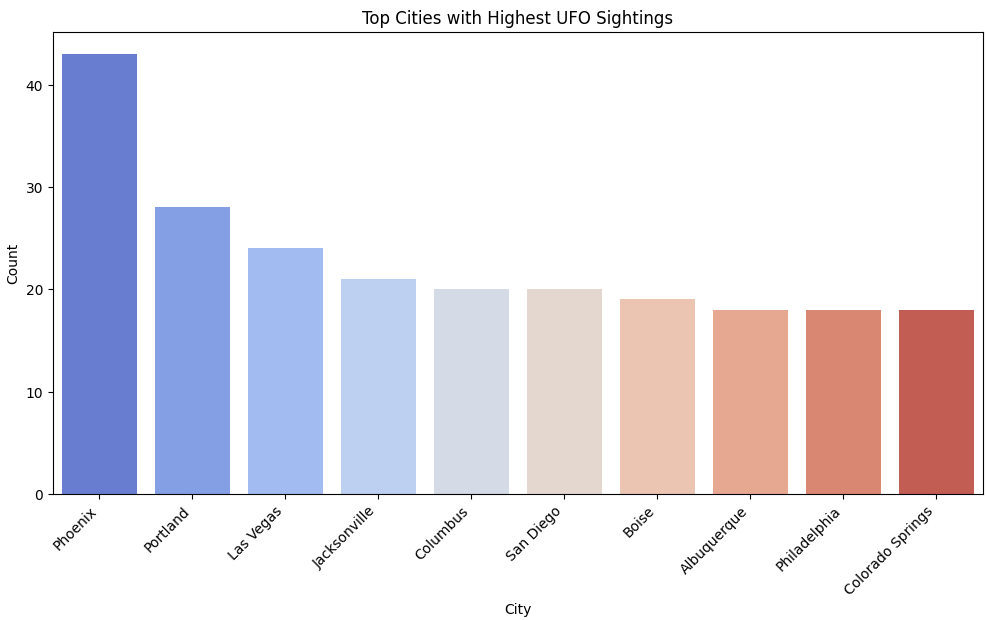
For a more granular analysis, when the distribution of reported UFOs within each state was explored, California (USA) emerged with the highest sightings, while Yukon (Canada) had the least, as evident in Figure (3). Figure (4) further emphasizes California's prominence, revealing it as the state with the highest occurrences of the reported shape "Light."

Figure (3)

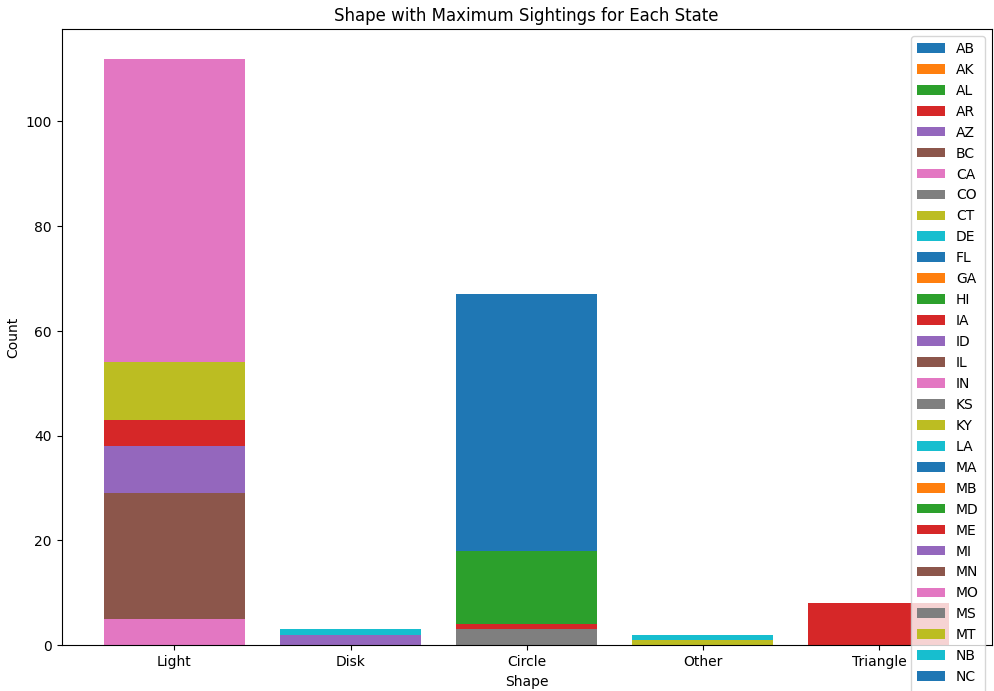
* **City-wise Distribution:**



The analysis was also narrowed down to counts per city. As we can see in Figure (4), the city Phoenix had the highest reported sighting in whole USA and Canada together. Phoenix is a city which is the Arizona city and it stands 3rd in the high reported sightings state list if we see Figure (3)

Figure 4

* **Popular Shapes per State:**

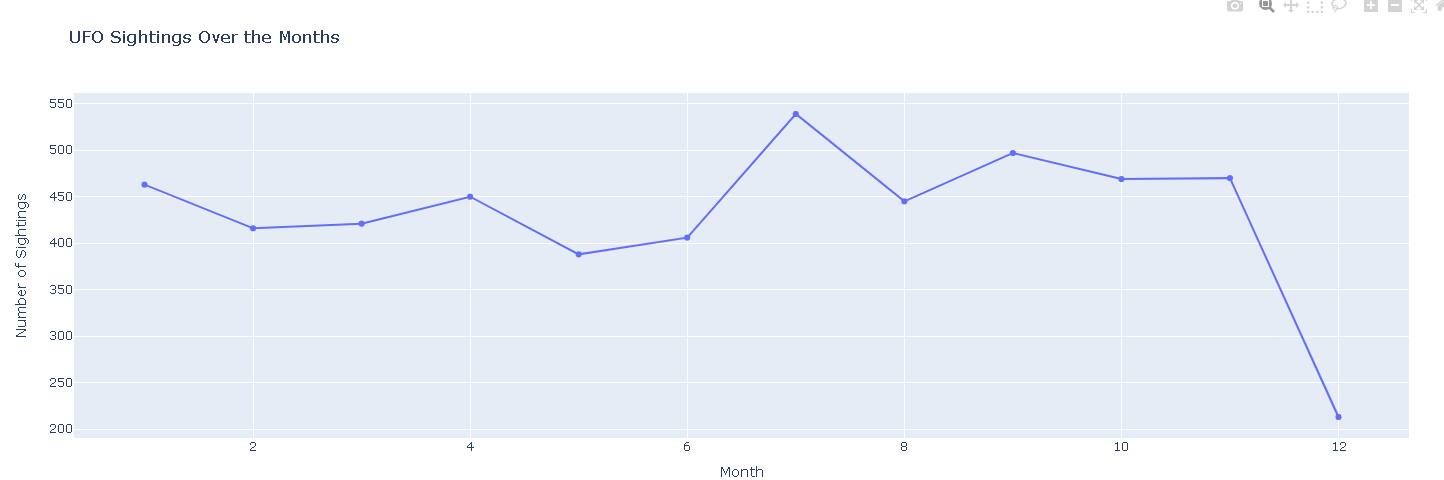


The Figure (5) shows that the overall highest shape reported was “Light” and the city that stood out in the analysis was CA. The second highest sighted and reported shape was circle shape and the city having most Circle shaped sightings is FL.

Figure 5

1. **Temporal Analysis:**

* **Month and Weekly analysis:**



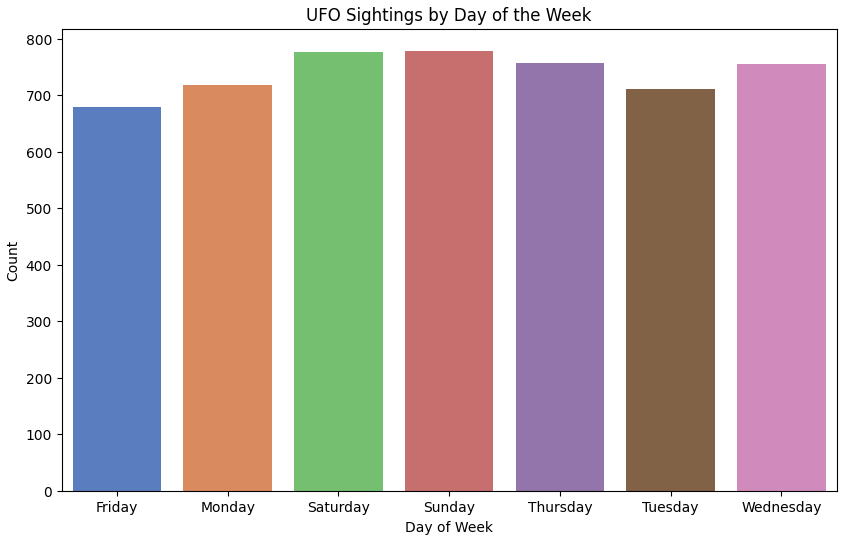
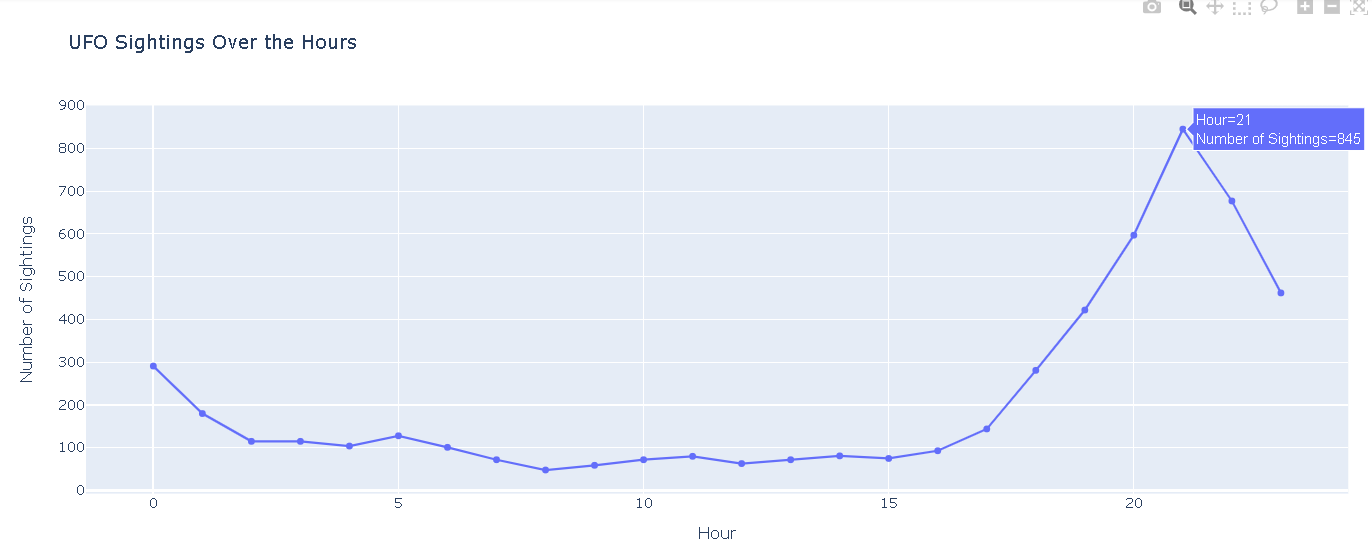
By looking at the Figure (6) we can notice that months from July to November had the highest sigthings in average, especially July had the peak reportings. If we look at the week wise sighting plot Figure (7), Saturday and Sunday have highest sightings.

Figure 6

Referencing to the Figure (6) and Figure (7) we can draw a conclusion that “More the people go outside or stay outdoors, the more the sightings reported”. Because, the months having highest sigthings are mostly summer end and fall months, where we can expect people going out more compared to other seasons. And if we look at the weeks, weekeds are when people are more outdoor.

Figure 7

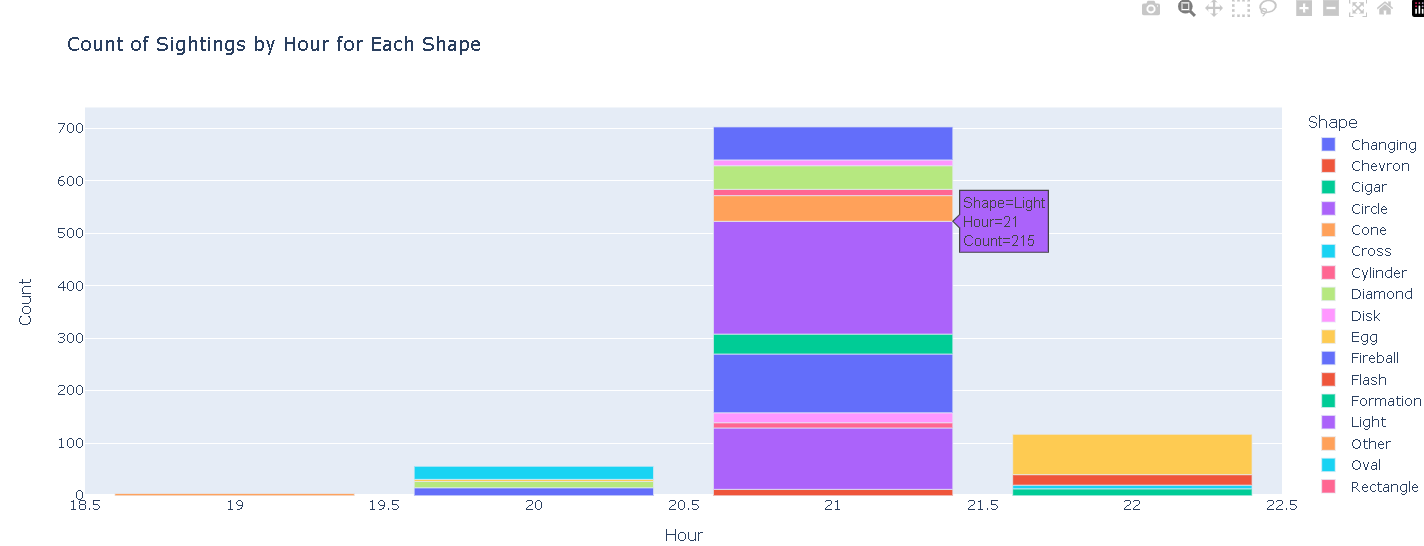
* **Hourly analysis of distinct shape:**



If we look at the hourly vs sightings reports, most of the sightings have happened in between 18:00-22:30 and the highest is at 21:00. This can be seen in Figure (8).

If we compare it with the shapes having their highest sightings in hourly axis, there are also we can see most of them are recorded during 21:00 Figure (9). Also notice the light shape which is seen more often has the highest sighting also at 21:00.

Figure 8



Observing both Figure (8) and Figure (9), it is evident that the bulk of UFO sightings occurred in the evening as the sun started setting.

Upon comprehensive comparison of Figures (6), (7), (8), and (9), a discernible pattern emerges, suggesting that the autumnal months experienced the highest frequency of UFO sightings, particularly after sunset. By connecting the insights from figure (9) with the monthly and hourly analyses, it becomes evident that the surge in sightings during these times may be attributed to the heightened visibility of lights against the backdrop of the darkened sky. Notably, in the fall, the sun sets after 18:00, providing optimal conditions for the observation of luminous phenomena, thereby contributing to the increased reporting of UFO sightings during this period.

Figure 9

1. **Text Analysis:**

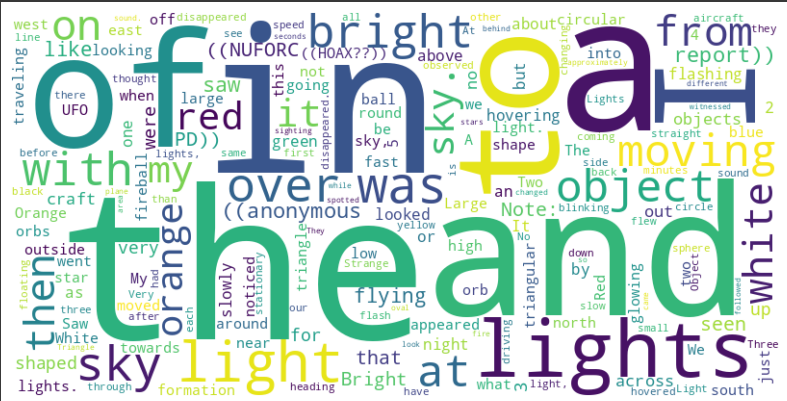
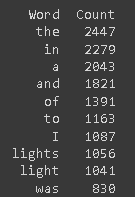
The text analysis delved into the summaries provided by viewers and reporters, unveiling the top 10 frequently used words, as depicted in Figure (11). This analytics provided valuable insights into prevalent themes and keywords encapsulated within UFO sighting reports. The corresponding word cloud plot, vividly illustrated in Figure (10)

fig 11

Figure 10

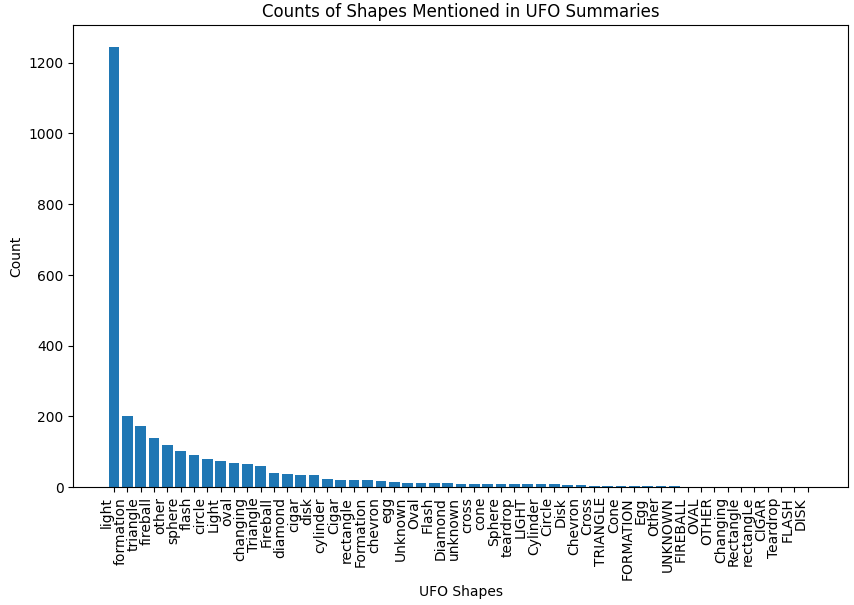
Furthermore, upon narrowing the focus to words specifically related to shapes within the summary explanations provided by reporters, a noteworthy finding emerged: "light" stood out as the most frequently reported shape, as evidenced by Figure (12). Intriguingly, this observation aligns with the top 10 words identified in the broader text analysis, showcased in WordCloud figures (10) and (11). The consistent prominence of "light" in both analyses underscores its significance, emphasizing its recurrent association with reported UFO shapes and reinforcing its prevalence within the dataset.

Figure 12

**Conclusion:**

The analysis of this UFO sighting project unmistakably points to the United States as the epicenter of reported sightings, with California emerging as the state with the highest frequency of occurrences. The majority of these reports clustered during the warmer months and were most prevalent during the evening, coinciding with the darker hours of the day. Notably, the findings suggest a correlation between the surge in sightings and periods when people are more likely to be outdoors, such as during the summer months when the evenings are conducive to sky observation. Furthermore, the text analysis underscores a recurring theme – the predominant mention of 'light' in sighting descriptions, reinforcing the notion that many reported UFOs are perceived as luminous phenomena. This lends credence to the reliability of the dataset and suggests that the reported sightings align with observable patterns, strengthening the veracity of the UFO phenomenon in the public eye.

**Future work:**

* Utilizing advanced natural language processing techniques for text analysis could uncover subtler themes within sighting descriptions, potentially unveiling hidden patterns. Exploring external factors such as weather conditions, celestial events, or human activities might contribute to a more holistic interpretation of UFO occurrences.
* Comparative analysis of UFO reports with existing databases on celestial events or military activities could shed light on potential misidentifications or correlations with known phenomena.

**Appendix:**

All code can be found at the following linked Google Colab notebook:

<https://colab.research.google.com/drive/1oj7EV8LJm6kXf_voMqBsTai0QsfETBZh?usp=sharing>