This research report shows the methodology and conclusions of finding some prevailing attributes of UFO (undefined flying object) by analyzing the words frequencies from 134k descriptions of UFO eyewitnesses.

The University of Adelaide

Computer Science (Advanced)

Jiajun Yu

a1806320

What are the prevailing attributes of UFO

Project 1 Final Written Report

**Project 1 Final Written Report**

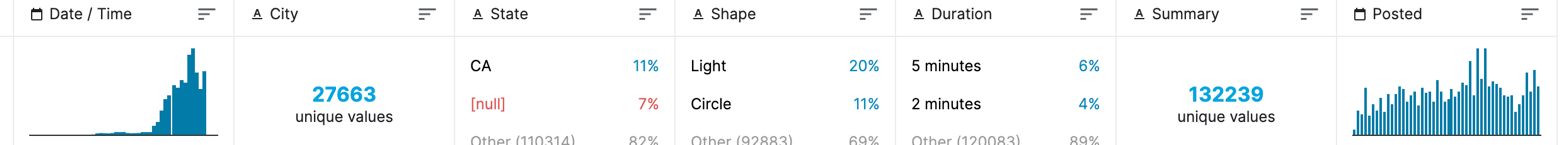
Topic Introduction

In 1977, Peter Sturrock, a professor of space science and astrophysics at Stanford University, did a research and found that 4.6% of members of the American Astronomical Society claimed that they have witnessed inexplicable aerial phenomena. This rate is close to the fact that approximately 5 percent of UFO (undefined flying object) sightings that are never explained (Knuth 2018). The topic of UFOs attracts people from all over the world. Investigations and studies for UFO have been conducted by various governments worldwide nowadays, along with private individuals and organizations (Daniela, G et al. 2021).

Motivation

Few months ago, the US government released a report about UAP (unidentified aerial phenomena). It is based on the review of 144 significant UAP reports involving observations made by military aviators between 2004 and 2021. However, the task force could only give a concrete explanation for one phenomenon, and the rest remain unexplained (Stieb & Danner 2021). That report raised the awareness of people who are concerned with UFO events. Hence, doing a research about identifying some prevailing attributes of UFO is necessary to help people know more about this unfamiliar field. This report will show the research methodology and conclusions in detail.

Dataset

The dataset utilized for this research is based on web scraped data from USA National UFO Reporting Center (Kaggle 2021). It contains 7 unique columns with 134,421 records of UFO sightings from 1946 to 2021 in the USA and was updated in July 2021. This dataset includes information such as time, locations, and descriptions from eyewitnesses about UFO sightings.

Methodology

This research mainly analyses the words frequencies of information contained in the dataset and draws conclusions by categorizing words to identify the potential patterns of UFOs. The research has explored the shape, color, USA state positions, audio, and general positions patterns of UFO. In addition, appearance year, month, time, and duration patterns of UFO have also been investigated. All visualizations were implemented by Python. Three main methods were adopted to explore the main characteristics of UFO, which were ‘All Words Frequencies Counting’ code (Code 1), ‘Target Words Frequencies Counting’ code (Code 2) and Microsoft Excel respectively. Firstly, ‘All Words Frequencies Counting’ code is able to count the frequency for every word in all descriptions. ‘Target Words Frequencies Counting’ code can count the word frequency for several specific words. These two scripts will not count repeated words from one piece of description. Lastly, Microsoft Excel has powerful searching functions which can identify some patterns that cannot be done with Python.

To find the shape pattern of UFO, ‘All Words Frequencies Counting’ code was used to obtain the word frequency for all words in a specific column of the dataset containing the shape information of UFO. After getting the frequencies for words describing shapes, the outcome was visualized (Figure 1). However, Figure 1 only shows 10 shapes of UFO. In order to show the overall shape pattern of UFO, word cloud visualization was utilized (Figure 2). It took a text file as its input which contained different numbers of shape words according to their frequencies and was generated by Code 15. This visualization is able to show the frequencies for all shape words by plotting them with various sizes. From the data, it can be concluded that the top 3 words describing the shapes of UFO are ‘circle’, ‘triangle’ and ‘fireball’, which totally account for 38.85% of all records. Taking into the consideration of people’s stereotype towards UFO, the ratio of circular shapes (circle, fireball, sphere, disk, oval, egg) should be calculated. It turned out to be 57.21% which indicates that people Text

Description automatically generatedChart, bar chart, funnel chart

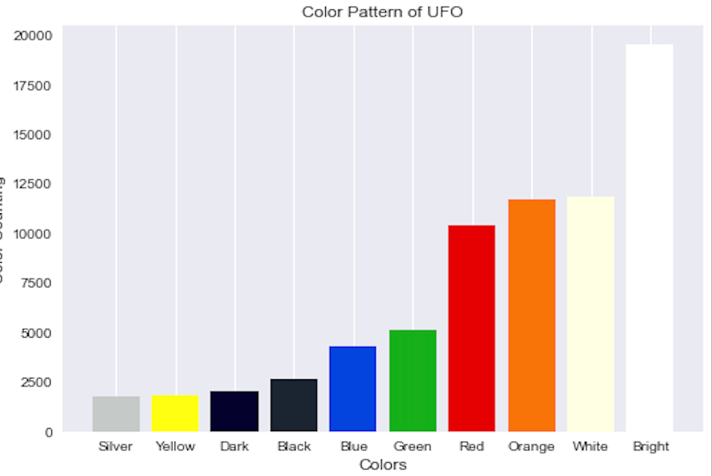
Description automatically generatedmight be affected by their stereotype towards UFO.

**Figure 2**

Generated by Code7

**Figure 1**

Generated by Code4

A picture containing shoji, building

Description automatically generatedTo find the color pattern of UFO, ‘All Words Frequencies Counting’ code was utilized to count the word frequency for all words from people’s descriptions. Then, color words as well as their words frequencies were recorded and visualized (Figure 3). To make Figure 3 more intuitive, waffle visualization was utilized (Figure 4). From the data, it can be concluded that the top 3 color words are ‘bright’, ‘white’ and ‘orange’, which totally account for 59.44% of all records. In addition, most colors are bright colors, which means that the vast majority of people might witness UFOs at night.

Generated by Code3

**Figure 3**

Generated by Code13

**Figure 4**

Map

Description automatically generatedChart, funnel chart

Description automatically generatedThe method of identifying the USA state positions of UFO is identical to finding the shape pattern of UFO. Using ‘All Words Frequencies Counting’ code to acquire the word frequency for state names in a specific column of the dataset could produce a list of state names with the corresponding numbers of UFO witnessed in those states. After that, the outcome was visualized (Figure 5). In order to have an overview of the distribution of UFO events, map visualization was adopted (Figure 6). It shows the overall number of UFO events in each state that witnessed from 1946 to 2021. It can be concluded from the data that people are more likely to witness UFOs in California, Florida, and Washington, and the ratio of UFOs to be witnessed in California is 12.44%. In addition, there are 76.61% of UFOs that were witnessed in coastal states.

Generated by Code12

**Figure 6**

Generated by Code5

**Figure 5**

Chart, pie chart

Description automatically generatedChart, funnel chart

Description automatically generatedFinding the general positions of UFO needs both ‘All Words Frequencies Counting’ code and ‘Target Words Frequencies Counting’ code. The first script was used to find all words related to general positions, and the second script was utilized to find the word frequency for a group of similar words. It can be seen from Figure 7 that it is easier for people to see UFOs from home, on the mountain, and on the street. In addition, the percentage of witnessing UFOs at home is about 41.39% higher than on the mountain, and it accounts for approximately 19.74% of all records.

Generated by Code6

**Figure 8**

Generated by Code10

**Figure 7**

Finding the audio pattern of UFO requires ‘Target Words Frequencies Counting’ code which can search and count the number of descriptions containing particular words. For example, setting the target words as ‘noise’, ‘noisy’, ‘hear’ etc. (30~50 words related to hearing sound) can be applied to investigate the audio pattern of UFO. From figure 8, it can be concluded that approximately 5% of eyewitnesses claimed that they heard the sound from UFOs.

Chart, line chart

Description automatically generatedA picture containing chart

Description automatically generatedOne advantage of the dataset is that all records have been sorted according to the time of UFO events. Hence, it is relatively easy to find the number of UFO events in a specific year or month by counting them in Microsoft Excel. It can be seen from Figure 9 that the number of UFO events has been on the rise since 1990 and reached its peak in 2014. After that, the number of sightings dropped sharply until 2018, and has been increasing thereafter. Figure 10 provides an intuitive overview of the number of UFO sightings in each month from 2009 to 2020. It can be concluded that people witnessed more UFOs than usual in July and August from 2012 to 2014 and March to April in 2020.

Generated by Code14

**Figure 10**

Generated by Code11

**Figure 9**

Finding the appearance time of UFO needs the ‘Text Filter’ from Microsoft Excel to avoid compatibility issues generated by Python. For example, to find the number of UFO sightings at 20:00 to 21:00, the content in the ‘Text Filter’ could be “contains ‘20:’ ”. It can be seen from Figure 11 that the number of UFO events increases sharply after 4 pm and it reaches a peak at 9 pm. After that, the number declines. This trend is consistent with previous speculation that the vast Chart, pie chart

Description automatically generatedChart, line chart

Description automatically generatedmajority of people might witness UFOs at night.

Generated by Code9

**Figure 12**

Generated by Code8

**Figure 11**

Considering the time constraint and actual situations, only data from 2020 were used to investigate the durations of UFO. Identifying this pattern requires Microsoft Excel combined with some manual work. Firstly, meaningless signs and English words were replaced by spaces. Then, empty records were deleted to make sure ‘Text Filter’ was able to work. Next, utilizing ‘Text Filter’ to filter various durations could generate outcomes. During the whole process, filtered data should be removed from the dataset after each round of searching to avoid repetition. It can be concluded from Figure 12 that there are 14% of UFOs will disappear within 10 seconds and 30% of UFOs will vanish within 1 minute. In addition, a vast majority of eyewitnesses (71%) can witness UFOs within 10 minutes.

Conclusion

In conclusion, this report provides people with an overview of the prevailing attributes of UFOs. Numbers of patterns have been identified and concluded from the descriptions of different eyewitnesses. However, the UFO phenomenon is still a mystery and needed to be studied continuously by human beings.

Summary of key findings

* Top 3 shape words describing UFOs are ‘Circle’, ‘Triangle’ and ‘Fireball’.
* Top 3 color words describing UFOs are ‘Bright’, ‘White’ and ‘Orange’.
* UFOs are most likely to be witnessed in ‘California’, ‘Florida’ and ‘Washington’.
* People are more likely to witness UFOs from home, on the mountain, and on the street.
* Approximately 5% of eyewitnesses heard the sound from UFOs.
* The number of UFO sightings has risen since 1990 and reached its peak in 2014.
* People witnessed more UFOs in July, August from 2012 to 2014 and March, April in 2020.
* The number of UFO sightings increases sharply after 4 pm and reaches a peak at 9 pm.
* A vast majority of eyewitnesses (71%) can witness UFOs within 10 minutes.

Reference:

Carvalho, T 2020, *Heatmap Basics with Seaborn*, Towards Data Science, viewed 1 September 2021, <https://towardsdatascience.com/heatmap-basics-with-pythons-seaborn-fb92ea280a6c>.

Chauhan, A 2020*, Find frequency of each word in a string in Python*, GeeksforGeeks, viewed 10 August 2021,<https://www.geeksforgeeks.org/find- frequency-of-each-word-in-a-string-in-python/>.

Daniela, G et al. 2021, *Unidentified flying object*, Wikipedia, viewed 5 September 2021, <https://en.wikipedia.org/wiki/Unidentified\_flying\_object>.

Gawande, J 2021, *Bar Plot in Matplotlib*, GeeksforGeeks, viewed 21 August 2021, <https://www.geeksforgeek s.org/bar-plot-in-matplotlib/>.

Gawande, J 2020, *Plot a pie chart in Python using Matplotlib*, GeeksforGeeks, viewed 23 August 2021, < https://www.geeksforgeeks.org/plot-a-pie-chart-in-python-using-matplotlib/>.

Kaggle 2021, *National UFO Reporting Data*, csv, viewed 7 August 2021, <https://www.kaggle.com /connerbrew2/national-ufo-reporting-data>.

Knuth, K 2018, *Are we alone? The question is worthy of serious scientific study*, The Conversation, viewed 4 September 2021, <https:// theconversation.com/are-we-alone-the-question-is-worthy-of-serious-scientific-study-98843>.

Li, G 2021, *Make Waffle Charts in Python*, Python Repo, viewed 25 August 2021, <https://pythonrepo.com/repo/gyli-PyWaffle-python-data-validation>.

*Line plot or line chart in python with legends* 2021, Data Science Made Simple, viewed 26 August 2021, < https://www.datasciencemade simple.com/line-plot-line-chart-in-python-legends/>.

Stieb, M & Danner C 2021, *What’s Inside the Pentagon’s Long-Awaited UFO Report*, Intelligencer, viewed 5 September 2021, <https://nymag.com/intelligencer/article/pentagon-ufo-report-what-we-know.html>.

Vu, D 2019, *Generating Word Clouds in Python*, Data camp, viewed 29 August 2021, <https://www.datacamp.com/ community/tutorials/wordcloud-python>.

Wu, L 2020, *13 lines of Python code to draw a map of the US epidemic*, CSDN, viewed 27 August 2021, < https://blog.csdn.net/wulishinian /article/details/105539212>.

Text

Description automatically generatedText

Description automatically generatedText

Description automatically generatedText

Description automatically generatedText

Description automatically generatedAppendix:

**Code 4**

Bar plot for Shape Pattern

**Code 3**

Bar plot for Color Pattern

**Code 1**

All Words Frequencies Counting Code

**Code 2**

Target Words Frequencies Counting Code

Text

Description automatically generated

**Code 5**

Bar plot for State Positions of UFO

**Code 6**

Pie chart for Audio Pattern

Code 11

Code 10

Text

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated

**Code 11**

Line chart for Appearance year of UFO

**Code 7**

Word Cloud Visualization for Shape Pattern

Graphical user interface, text

Description automatically generatedText

Description automatically generated

**Code 8**

Line chart for Appearance time of UFO

Text

Description automatically generatedText

Description automatically generatedText

Description automatically generated

**Code 12**

Map Visualization for State Positions of UFO

**Code 13**

Waffle Visualization for Color Pattern

**Code 10**

Bar plot for General Positions of UFO

**Code 9**

Pie chart for Durations of UFO

Code 9

Text, letter

Description automatically generatedText

Description automatically generated with low confidenceA picture containing graphical user interface

Description automatically generatedText

Description automatically generated

**Code 15**

Generating Input for Word Cloud Visualization

**Figure 13**

Text File Generated by Code 15

**Code 14**

Heat Map for Appearance Month of UFO

Table

Description automatically generatedTable

Description automatically generatedTable

Description automatically generatedTable

Description automatically generatedTable

Description automatically generatedGraphical user interface, table

Description automatically generatedExcel Worksheet

**Note:** Numbers in AT column represent the ratio of colors

**Note:** Numbers in AF column represent the relative numbers for shape words.

**Note:** Durations

of UFO (2020)

**Note:** Yellow entries represent coastal states in the USA

A picture containing table

Description automatically generatedTable

Description automatically generated with low confidenceTable

Description automatically generated with low confidenceTable

Description automatically generated with low confidenceTable

Description automatically generated with medium confidenceScatter chart

Description automatically generated with medium confidenceTable

Description automatically generated with medium confidenceScatter chart

Description automatically generated with low confidenceTable

Description automatically generated with medium confidenceTable

Description automatically generated with low confidenceA picture containing scatter chart

Description automatically generatedCalendar

Description automatically generated with medium confidenceCalendar

Description automatically generated

Data for appearance month of UFO