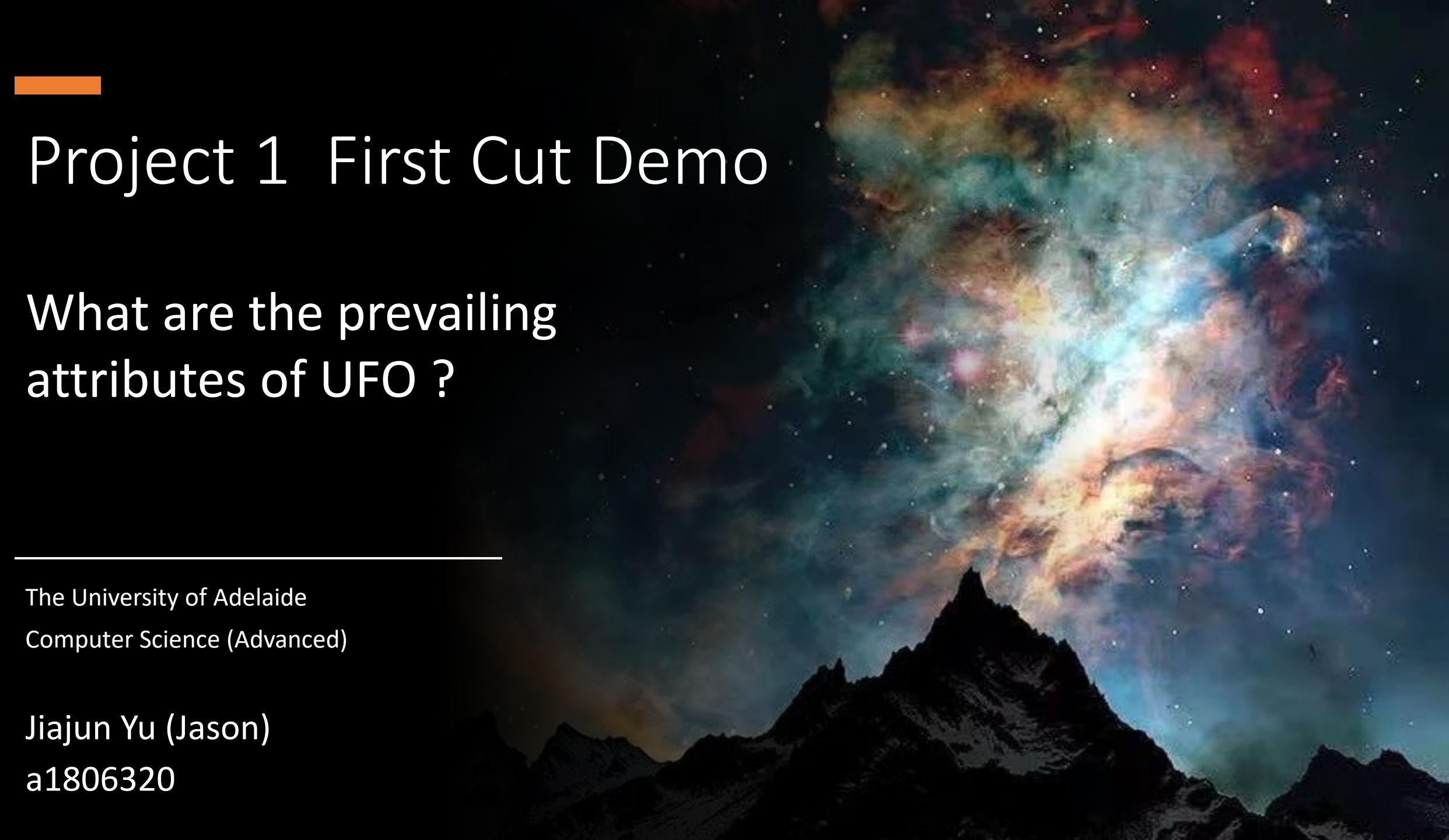


Project 1 First Cut Demo

What are the prevailing
attributes of UFO ?

The University of Adelaide
Computer Science (Advanced)

Jiajun Yu (Jason)
a1806320



Contents

01

Overview of the topic

02

Resources

03

Challenges

04

Analysis

05

Expected results

06

Initial visualizations



Overview 01

- The research of Peter Sturrock
- Research Form
- Benefits

Overview

In 1977, Peter Sturrock mailed 2,611 questionnaires about UFO sightings to members of the American Astronomical Society.

The research of Peter Sturrock



The outcome of the research

4.6% – reported witnessing or recording inexplicable aerial phenomena.

This rate is similar to the approximately 5 % of UFO sightings that are never explained.
(The Conversation 2018)

Analysis of UFO based on descriptions of eyewitnesses -> find some general patterns of UFO events.

Form of my research



Benefits

- To find the prevailing attributes of UFO
- It is a social obligation to help people who are concerned with UFO events to know more about UFO.
- Promote the development of philosophy



Resources 02

- Dataset from Kaggle --- Good points
- Resources that I need --- Need help ?

Resources

- From National UFO Reporting Center
- Updated at 2021-7-18
- Contains 134k + records
- Mainly focuses on USA UFO events

What I already have

	A	B	C	D	E	F	G	H
133818	6/15/50 03:00	Sugar House UT	Unknown	Unknown	I am '72 and this started when I was			
133819	6/12/1950 10:30	Budapest (H)	Unknown	Circle	1min	The 'object' made an extremely high		
133820	6/10/1950	Napa CA	Disk	2 minutes	Saw a circular craft approximately 3			
133821	6/08/1950 10:00	Kansas City MO	Disk	15minutes	The ufo came from the East, as I wa			
133822	6/07/1950 20:00	Kensington MD	Oval	10 minutes	Reflective oval seen hovering over V			
133823	6/01/1950 20:00	Pittsburgh PA	Oval	20/25 minutes	One larger white object & 2 smaller			
133824	6/01/1950 16:00	Philadelphia MS	Disk	ten minutes or less	Childhood sighting of disk witnessed			
133825	6/01/1950 12:00	Mexicali (Me Unknown)	Disk	1 minute	Large disk observed by townspeople			
133826	6/01/1950	Crowder MS	Unknown		Large circular area of cotton plants			
133827	5/15/50 13:00	New York Cit NY	Sphere	2 minutes	"Ball" shape object, size of a softba			
133828	5/12/1950 14:00	Fort Lewis WA	Triangle	5 min	the triangular shaped craft chased e			
133829	4/25/50 16:00	Vancouver BC	Cone	1 minute	Flying cone-shaped UFO seen by Va			
133830	4/15/50 14:00	Waynesboro VA	Disk	5 minutes	Fishburne Military School: 9 studen			
133831	4/15/50 08:00	Arkansas (ea AR)	Disk	1-3 minutes	Disk shaped unmoving over highway			
133832	3/22/50	Roswell NM	Unknown		Link to declassified, redacted, USAF			
133833	1/02/1950 0:00	New York Cit NY	Oval	15min	good and true info. 500 Lights On O			
133834	1/01/1950 22:00	Mount Hope WV	Unknown	unknown	One of first sightings in U.S. - 1950			
133835	12/23/49 07:30	Moody TX	Disk	8-10 minutes	UFO in central Texas, 3 spacemen c			
133836	10/15/49 20:00	Baltimore MD	Cigar	2-3 minutes	It was around 8:PM and I was walki			
133837	10/10/1949 21:00	Lackland AFE TX	Light	1-2 hrs	1949 Lackland AFB, TX. Lights racin			
133838	10/10/1949 20:30	San Marcos TX	Cylinder	45 minutes	This event took place in early fall an			
133839	9/15/49 21:00	Post TX	Disk	10 seconds	Disk object appeared very close to g			
133840	9/15/49 21:00	Mt Gilead OH	Circle	15 min.	Brilliant dime sized objects rotating			
133841	8/25/49 19:00	Steilacoom WA	Disk	15 min.	AS A GIRL OF ABOUT NINE AND MY			
133842	8/15/49 22:00	Tarrant City AL	Circle	ten+ minutes	Glowing large orange ball in Alabam			
133843	8/15/49 21:00	Mill Valley CA	Light	1 min. +	3 lights appeared, maneuvering rap			
133844	8/15/49 20:00	Wilmington DE	Cylinder	1 minute	I was ascending a external staircase			
133845	8/15/49 0:00	Excelsior MN	Circle	3-5 min.	large ball of brite light above pine t			
133846	7/15/49 20:30	Booneville MS	Unknown	hours	Returning home from a movie, we w			
133847	7/01/1949 16:00	Marion AL	Formation	Under 1 minute	GEORGIA UFO GROUP Report/Tom			

What resources do I need

- More detailed information about California
- Rough distribution of U.S. military bases
- Some general information of each state in the USA
- Python coding instructions for visualization

Need help ???



Challenges 03

- Multiple expression in the dataset
- Lacking information in the dataset
- Python beginner

Challenges



01

3 Min
5 Minutes
.5 hour
20 min
5 minutes or longer
10 minutes?
1-2 minutes
2 minutes
Hour
2 minutes
10 min
2 minutes
3 minutes
10-15 seconds
2 min
3 min
10 seconds
35 minutes
3 seconds
1.5 hours
10 min
~2
Aprox: 2 minutes
1 minute
10 seconds

'10 min',
'20 minutes',
'~5 mins',
'>3 Min',
'2 1/2 mins',
'four min',
'about a min',
'<10 Mins',
'1+ minute',
'2-3 minutes '



02

0:04	Alabama	AL	Diamond	60 seconds
1947-2008 00:00	Maple Valley	WA	Changing	4 years
06 00:00	Pittsburg	CA	Disk	3 minutes
5/20/20014 00:00	Atlanta	GA	Unknown	2 seconds
	New York City	NY	Circle	20 seconds I think
	Unknown	Unknown	Unknown	
	North Carolina	NC	Flash	
	Hollywood	Unknown	Unknown	
	Unknown	Unknown	Unknown	
	Jax	FL	Unknown	
	Hollywood	FL	Changing	
	Long Island (VNY)		Triangle	30 seconds
	Makawao	HI	Light	5 minutes
	Little Rock (5) AR		Unknown	25 minutes
	Unknown	Unknown	Unknown	
	Warfordsburg	PA	Changing	Minutes, maybe longer
	Unknown	Unknown	Unknown	
	Unknown	IN	Unknown	
	Poca	WV	Circle	All night all summer
	Brookhaven	MS	Disk	5 mins
	Dayton	OH	Disk	

Some of entries lack information

Multiple expression in the dataset

Lacking information in the dataset



03

```
1 class Bigfile:
2     def __init__(self, datadir, name):
3         self.dir = os.path.join(datadir, "t4.txt")
4         self.name = name
5         self.names2Index = {}
6         self.names = []
7         self.featurefile = os.path.join(datadir, "feature.txt")
8         print("Bigfile %s features, %d dimensions" % (len(self.names), len(self.names)))
9         print("binary: %s" % self.featurefile)
10        print("text: %s" % self.dir)
11
12    def read(self, requested, isname=True):
13        if isname:
14            index_name_array = [(self.names2Index[x], x) for x in requested]
15        else:
16            assert(len(requested) == 1)
17            assert(len(requested) == 1)
18            index_name_array = [(x, self.names[x]) for x in requested]
19            index_name_array.sort()
20
21        vecs = self.read(self.featurefile, self.names, 1000000, index_name_array)
22        return [x[1] for x in index_name_array], vecs
23
24    def shape(self):
25        return (len(self.names), len(self.names))
```

- Independent Learning
- Time management

Python beginner

A large, abstract network graph occupies the left two-thirds of the slide. It consists of numerous small, semi-transparent dots connected by thin, light-colored lines forming a complex web of triangles and quadrilaterals. The colors transition from orange at the top to red, then purple, and finally blue at the bottom, creating a vertical gradient.

Analysis 04

- Research approach
- Research tools

Analysis

Approach

- 1: **Pre-processing** dataset
- 2: **Count** word frequencies from descriptions
- 3: **Classfy** English words
- 4: **Visualize** data
- 5: **Explore** more ?



• Python algorithm

```
# Ranking
items=list(counts.items())
items.sort(key=lambda x:x[1])
print("\n");
for i in range(100):
    word,count=items[i]
    print("{0:<15}{1:>10}\n".format(word,count))
print("\n");

in          42446
the         37263
a           33294
and          28976
lights        26242
of            25035
light         24715
sky            22892
object        19825
bright        10525

The count is : 62467
```



• Excel Text Filter

The screenshot shows the 'Text Filter' dialog in Excel. It includes dropdown menus for 'Begins With', 'Ends With', 'Contains', 'Does Not Contain', and 'Custom Filter'. A checkbox for 'Advance' is also present. Below the dialog, a list of items is shown, with 'Triangle ufo' highlighted in yellow.

Filter Type	Value
Begins With	E
Ends With	J
Contains	U
Does Not Contain	b
Custom Filter	Triangle ufo
	Strange star
	Large Orange
	Saw cigar sh

Tools

• R algorithm

```
#import data
setwd("~/Desktop/Statistical Practice")
library(readxl)
ostro<-read_excel("bone.xlsx")
ostro

#scatterplot
plot(NonDom~Dom,data=ostro)

#regression line
ostro_lm<-lm(NonDom~Dom,data=ostro)

#know specific infro of ostro
summary(ostro_lm)
```



Expected Results 05

- Expected results list

Expected Results

- Advanced visualizations (Map)
 - Explore some further correlations
 - Inspirations for the future
-
- Time pattern of UFO
 - Duration pattern of UFO
 - Emergence year pattern of UFO
 - General position pattern of UFO
-
- Shape pattern of UFO
 - Color pattern of UFO
 - Audio pattern of UFO
 - State positions pattern of UFO

01

02

03

A large, abstract network visualization occupies the left two-thirds of the slide. It consists of numerous small, semi-transparent dots connected by thin, light-colored lines, forming a complex web-like structure.

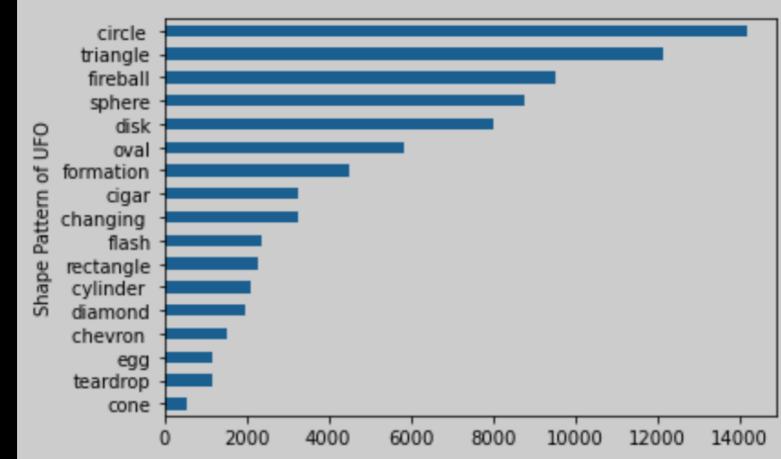
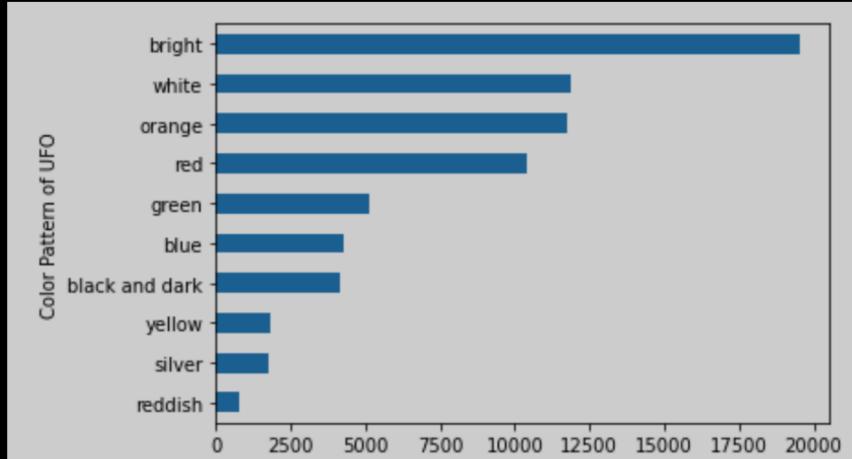
Initial Visualizations 06

- Shape pattern of UFO
- Color pattern of UFO
- Audio pattern of UFO
- State positions pattern of UFO

Initial visualizations



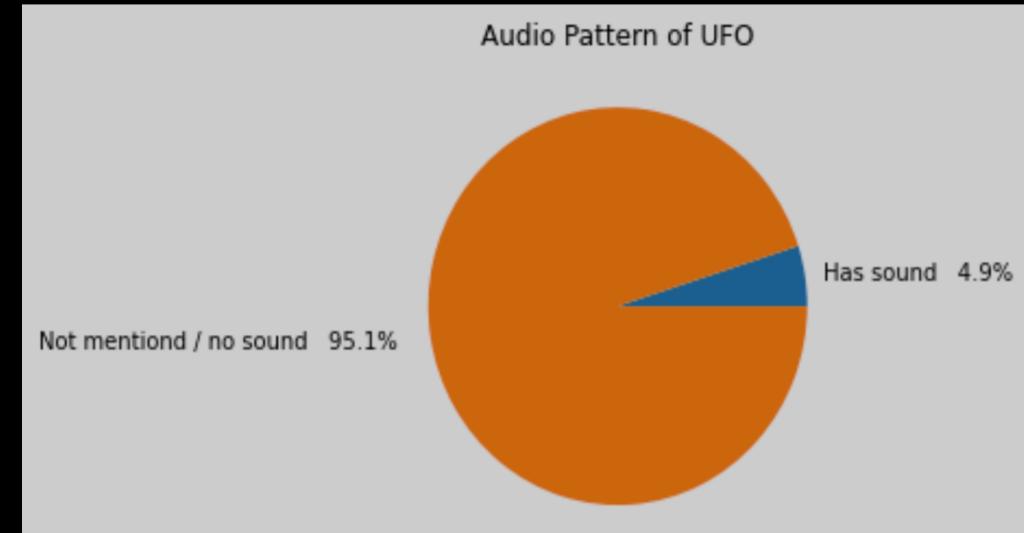
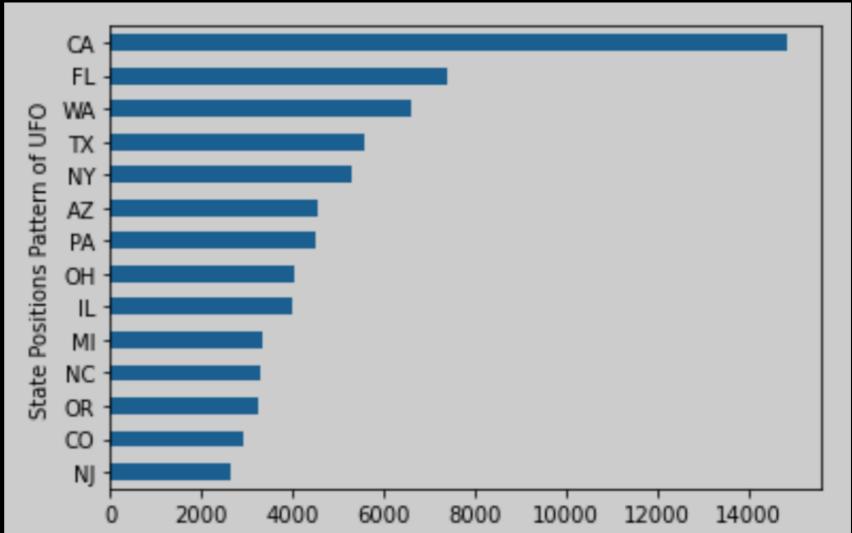
- 1: Bright
- 2: White
- 3: Orange



- 1: Circle
- 2: Triangle
- 3: Fireball



- 1: California
- 2: Florida
- 3: Washington DC



4.9 % sound



References

- https://www.google.com/search?q=ufo+background+&tbm=isch&ved=2ahUKEwjs7Kncma3yAhUxwYsBHe_8DwkQ2-cCegQIABAA&oq=ufo+background+&gs_lcp=CgNpbWcQAzIECAAQQzIFCAAQgAQyBQgAEIAEMgUIABCABDIFCAAQgAQyBQgAEIAEMgUIABCABDIFCAAQgAQyBQgAEIAEUM6HBVjOhwVgtokFaABwAHgAgAGUAYgBIAGSAQMwLjGYAQCgAQGqAQtnd3Mtd2I6LWltZ8ABAQ&sclient=img&ei=XfoVYazKM7GCr7wP7_m_SA&bih=1009&biz=1
- https://theconversation.com/are-we-alone-the-question-is-worthy-of-serious-scientific-study-98843792#imgrc=fmk_z2n8DrYn7M

Thanks for listening

