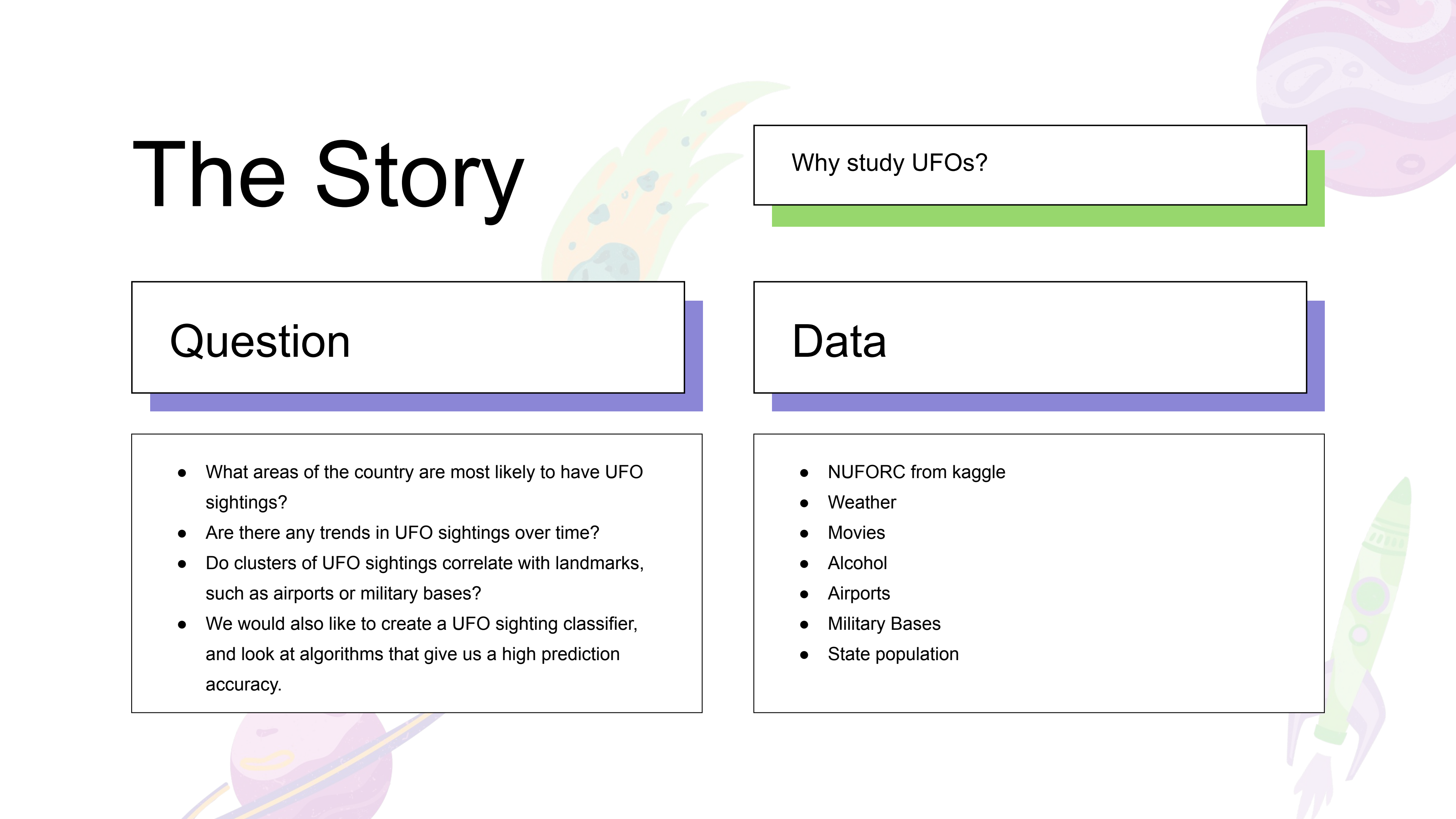


The background is a dark blue space scene. In the top right, a pink planet with a blue ring and orange ring gaps is visible. In the bottom left, a pink planet with white wavy patterns is shown. A green comet with orange and yellow streaks is on the left. A green rocket with purple fins and a purple flame is on the right. The text is centered in the upper half.

IST 718 Big Data Analytics Close Encounters

By Laura Larregui & Tiffannie Mac Donald

The Story



Why study UFOs?

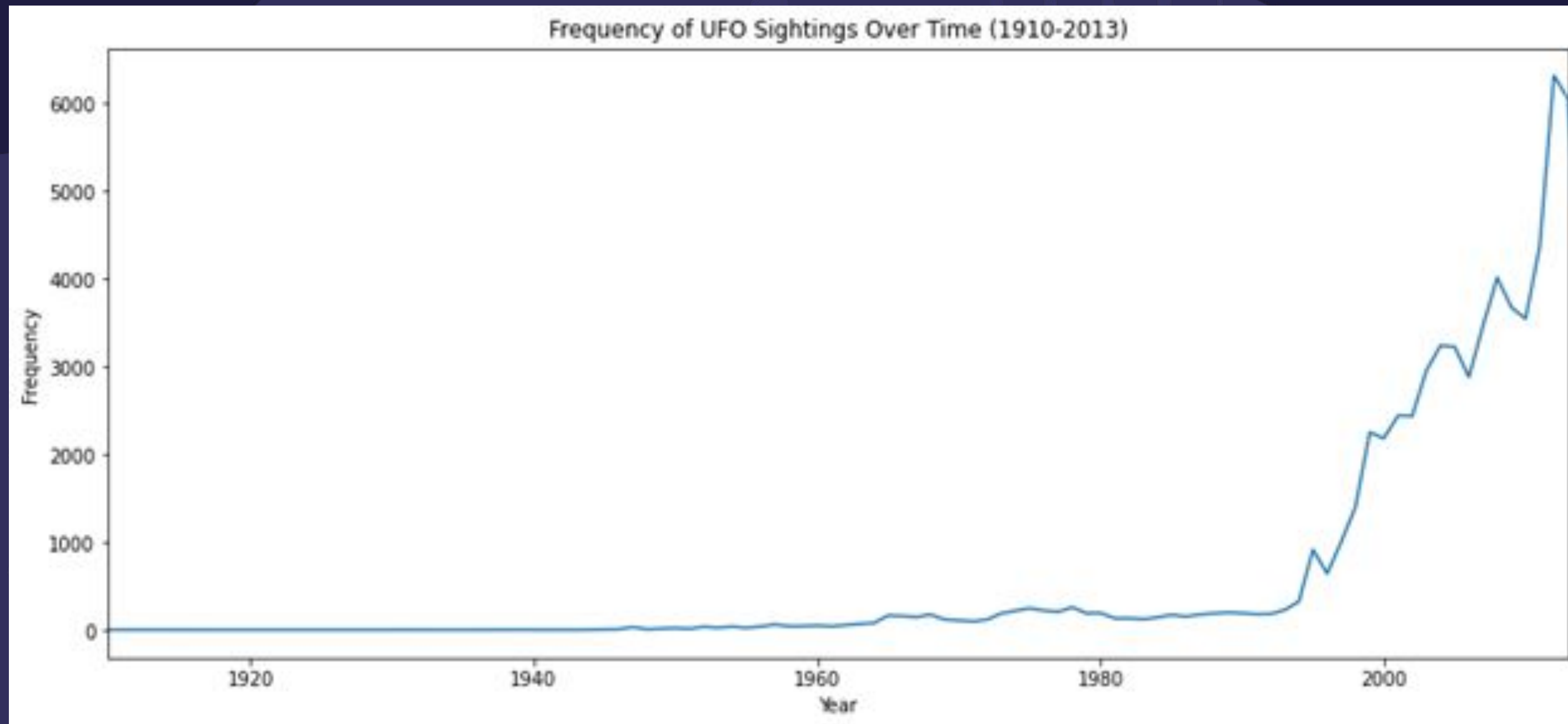
Question

- What areas of the country are most likely to have UFO sightings?
- Are there any trends in UFO sightings over time?
- Do clusters of UFO sightings correlate with landmarks, such as airports or military bases?
- We would also like to create a UFO sighting classifier, and look at algorithms that give us a high prediction accuracy.

Data

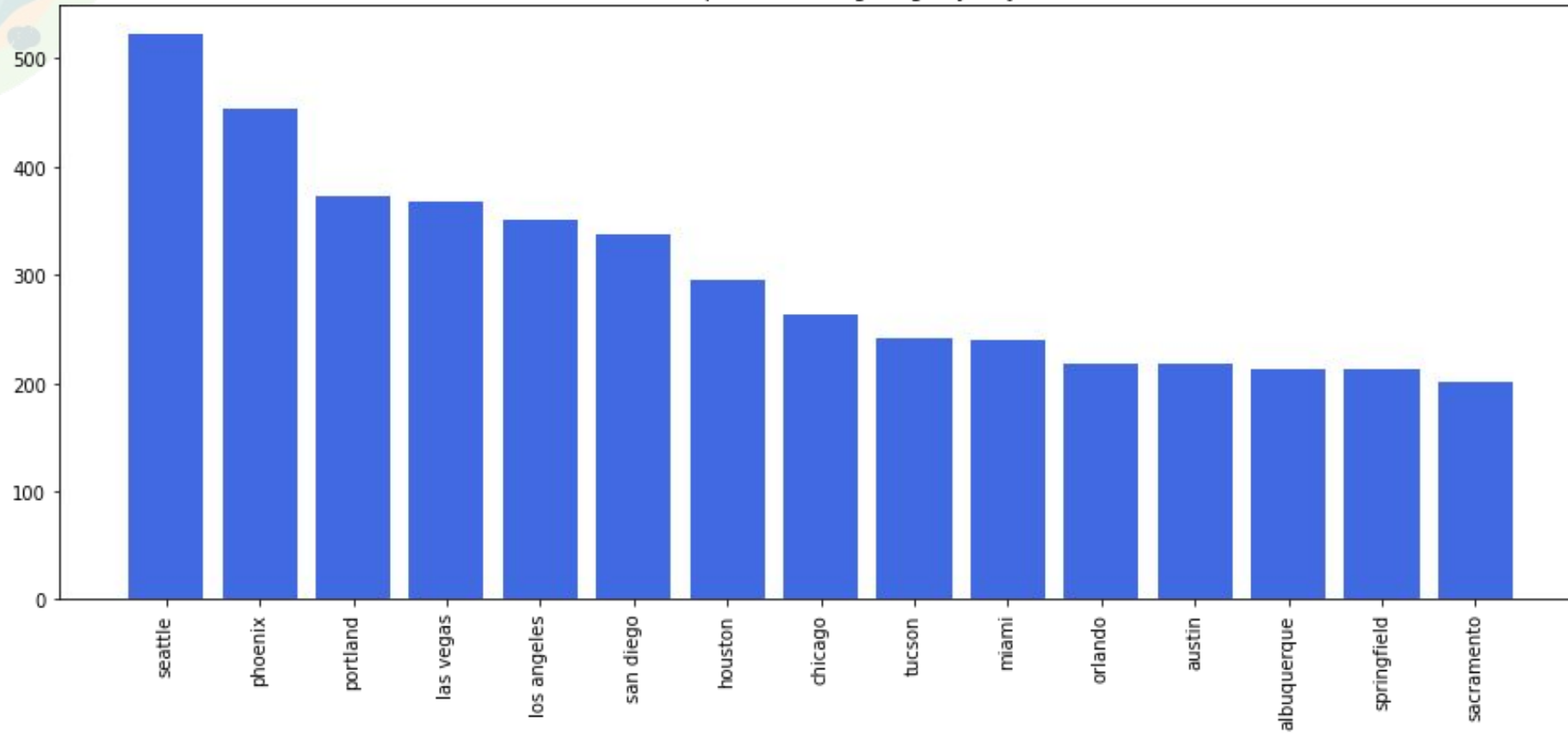
- NUFORC from kaggle
- Weather
- Movies
- Alcohol
- Airports
- Military Bases
- State population

We are under attack!

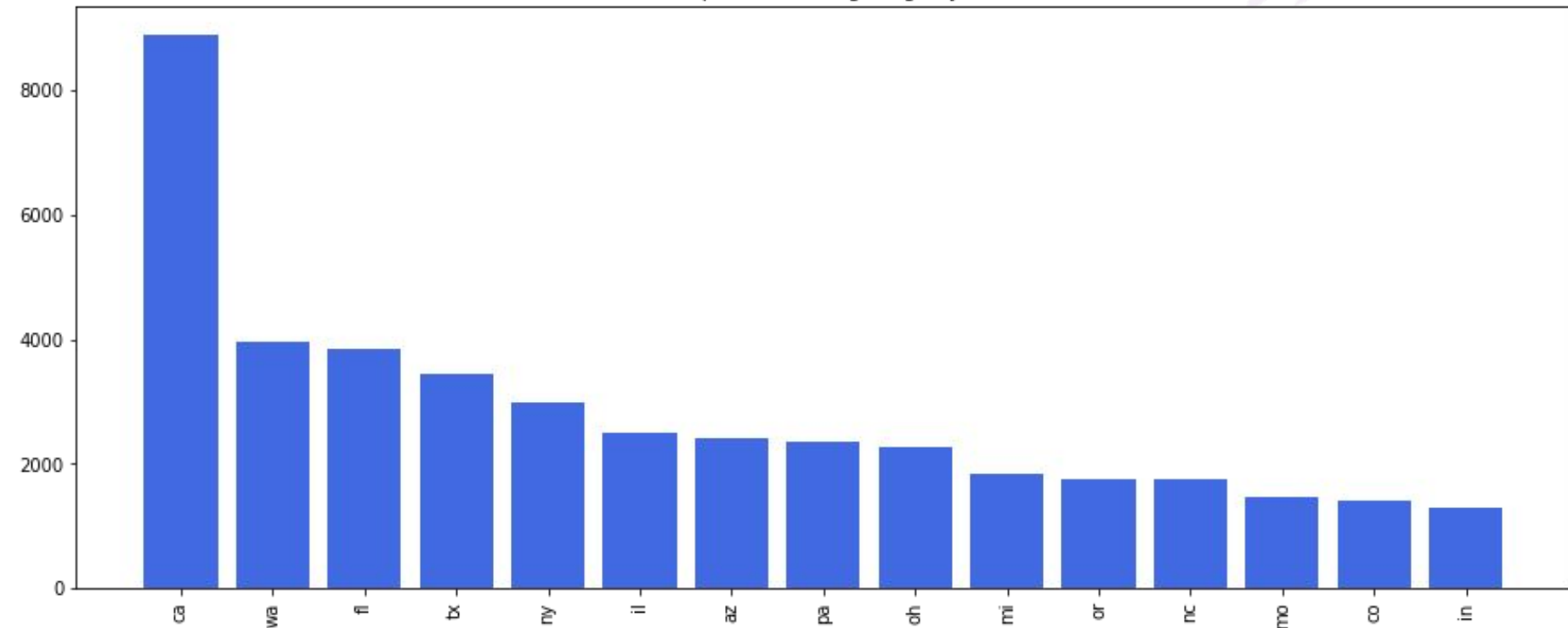


Where do UFOs hang out?

Most reported UFO sightings by City

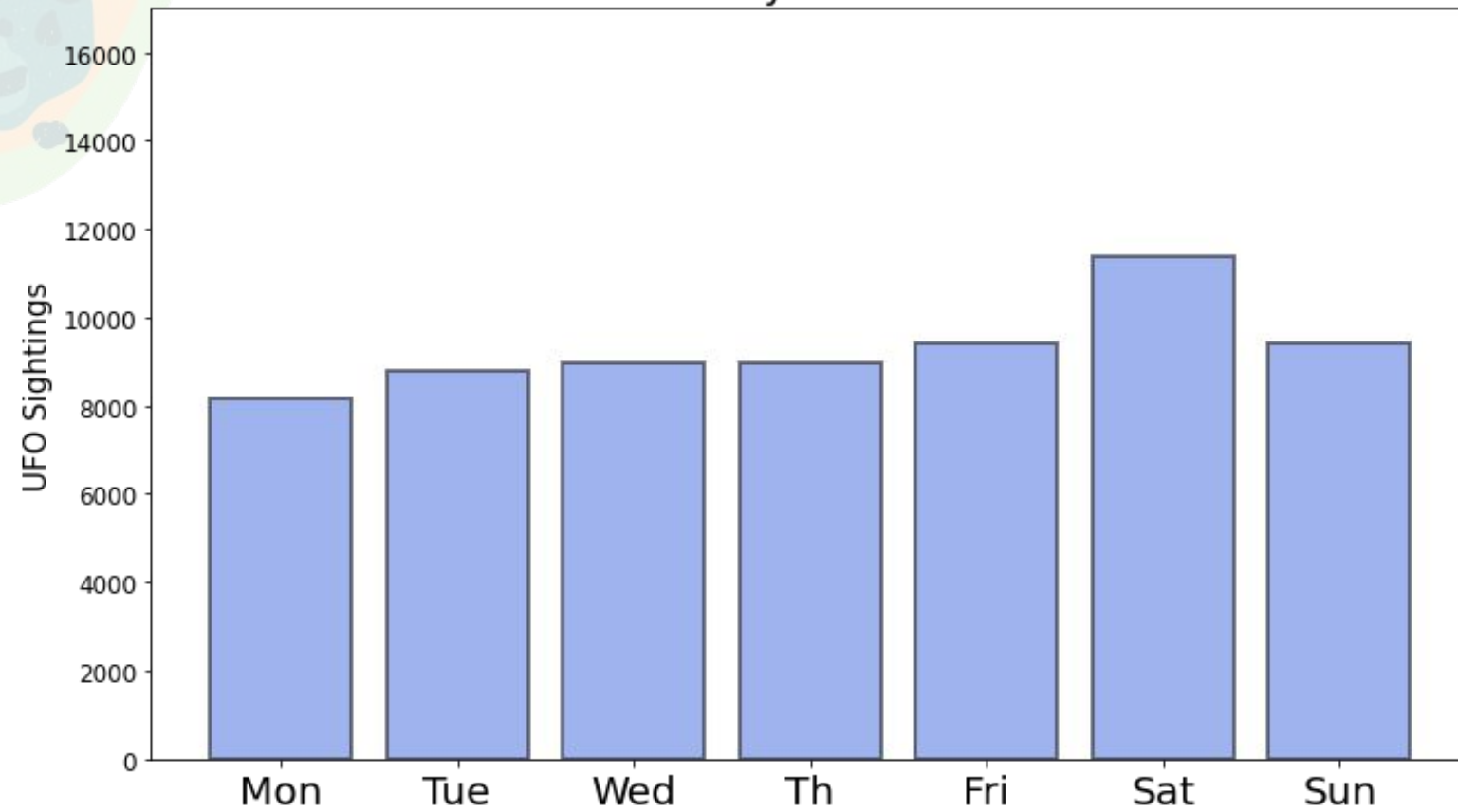


Most reported UFO sightings by State

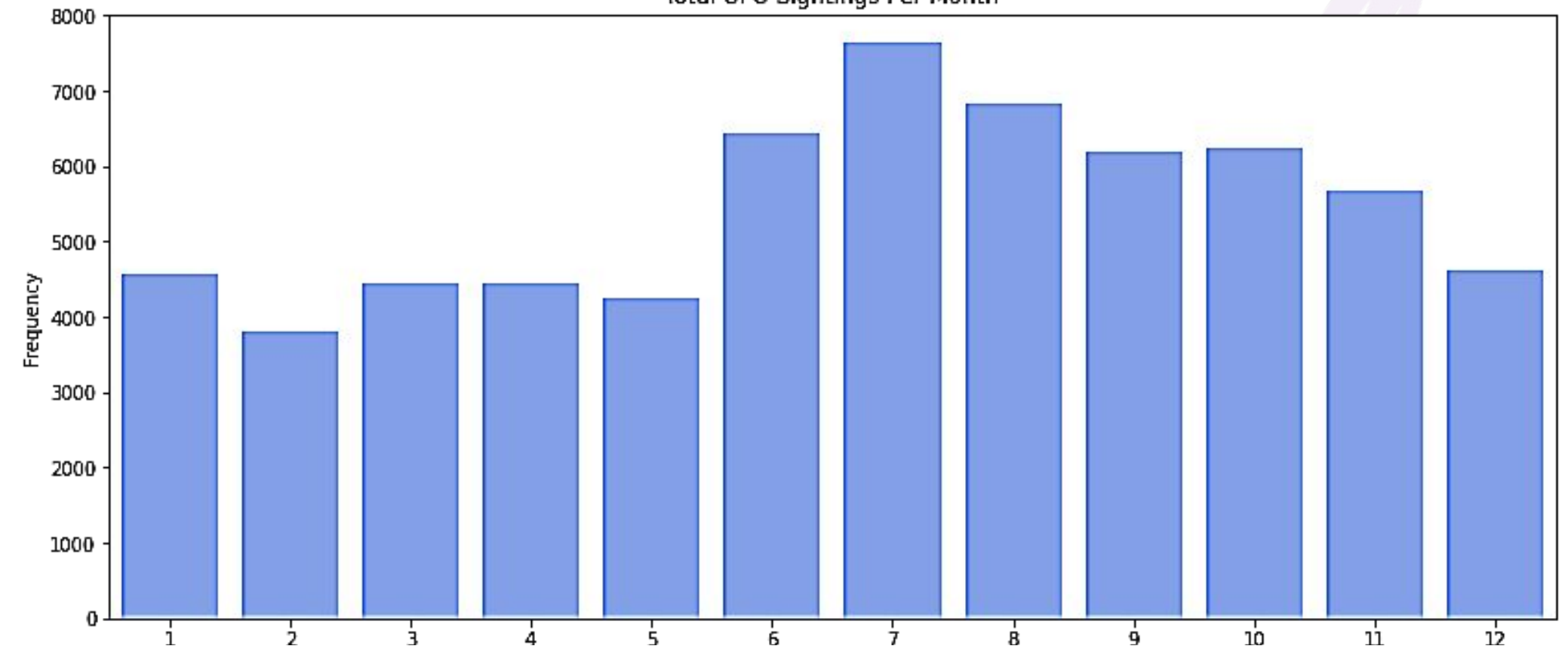


When do UFOs hang out?

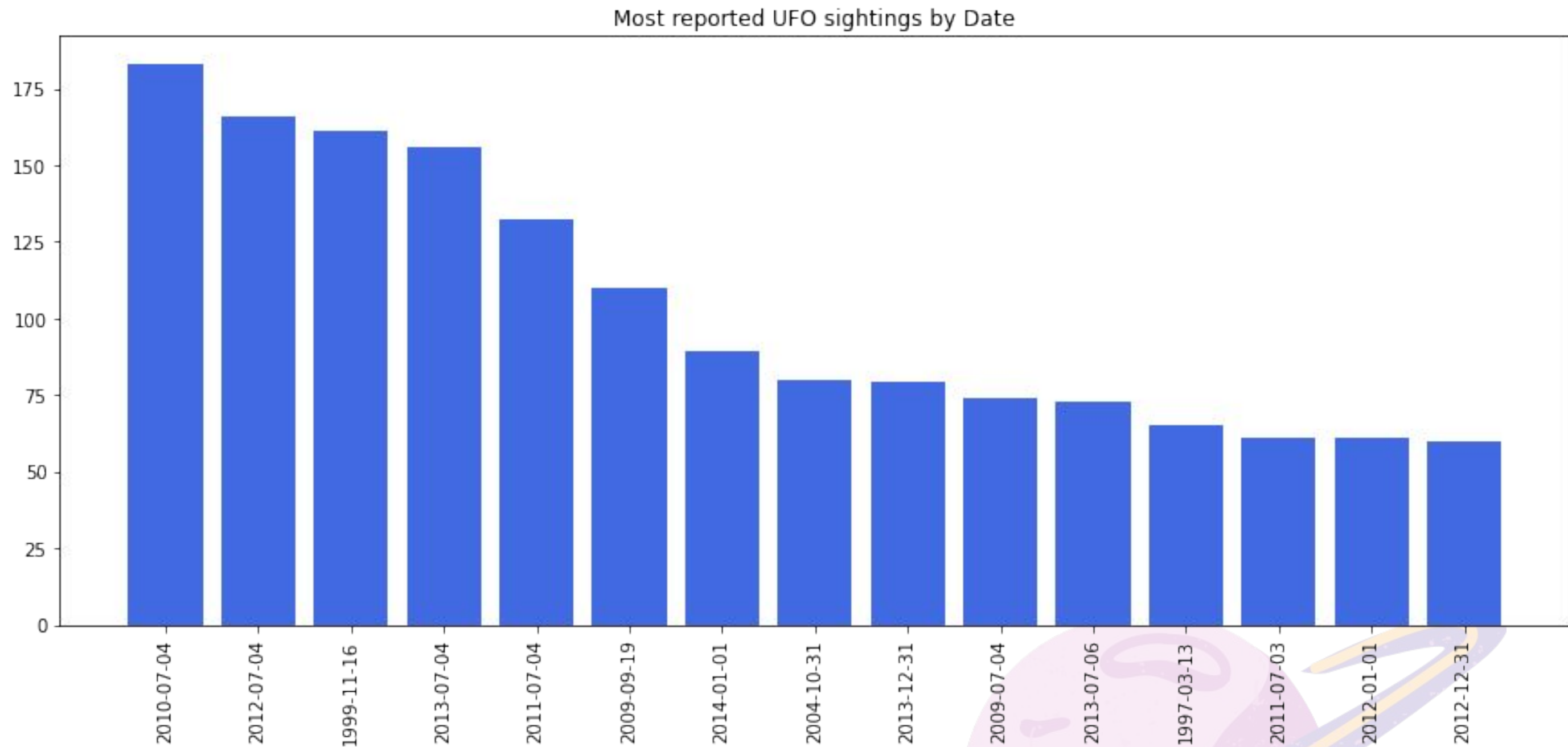
UFOs& Days of the Week



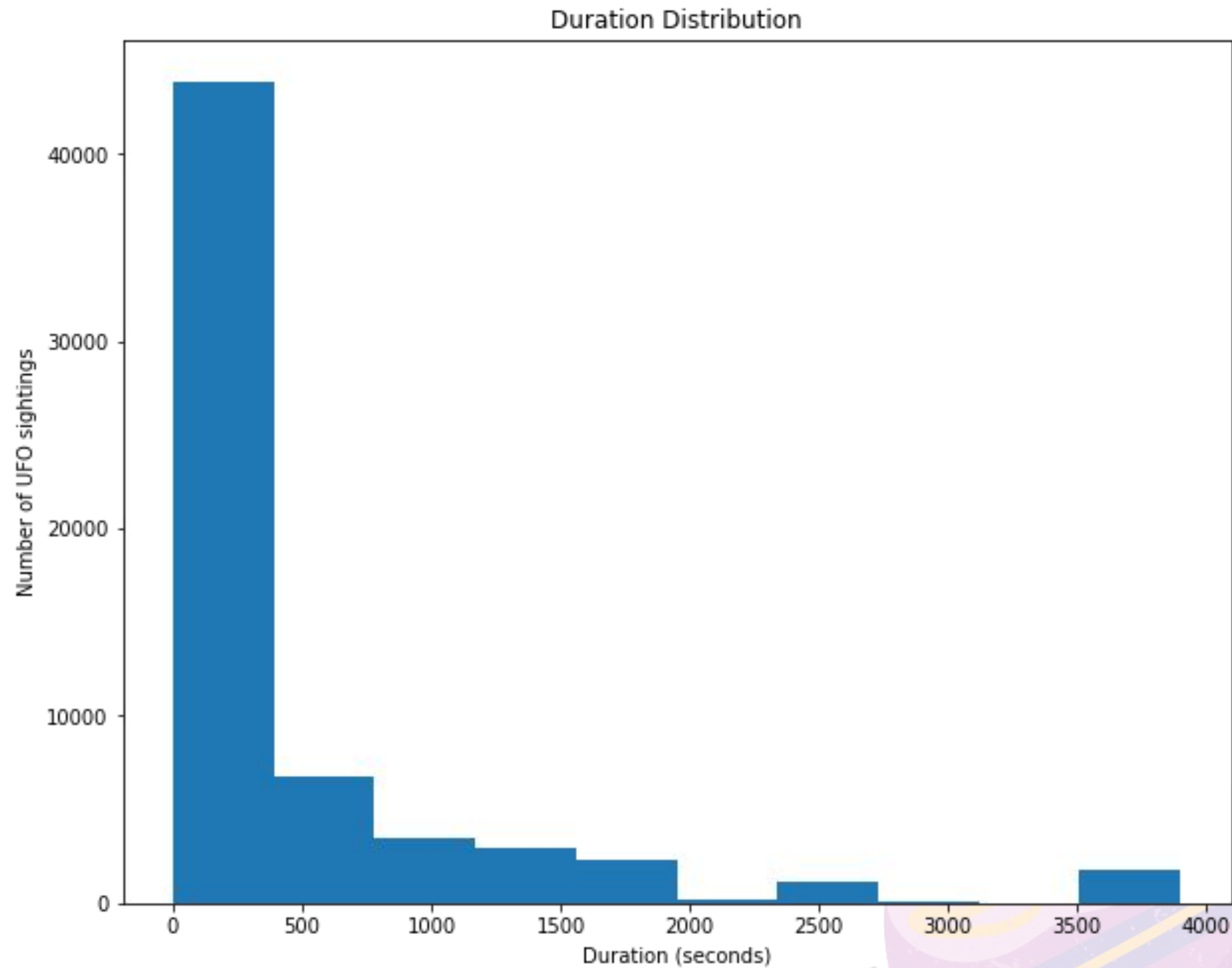
Total UFO Sightings Per Month



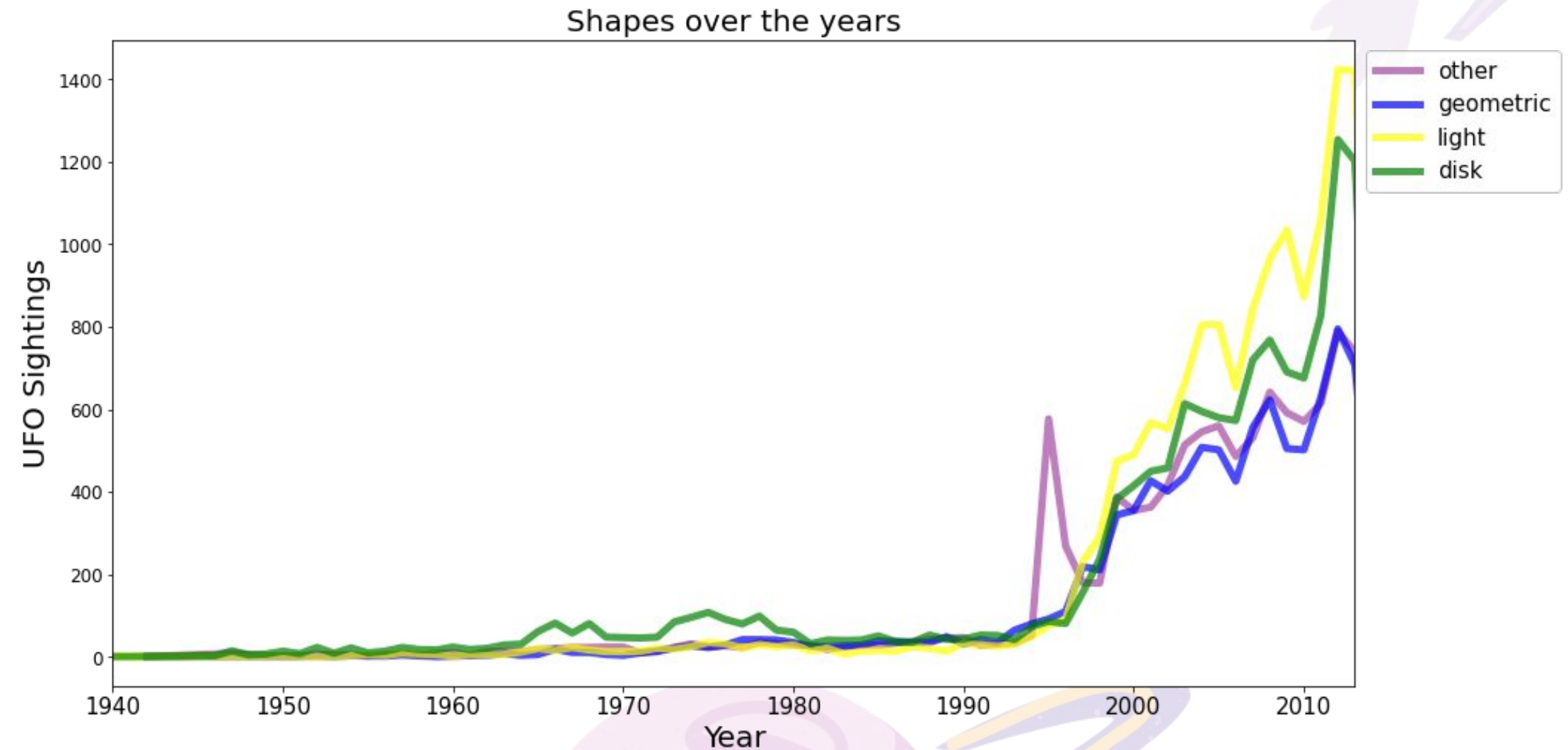
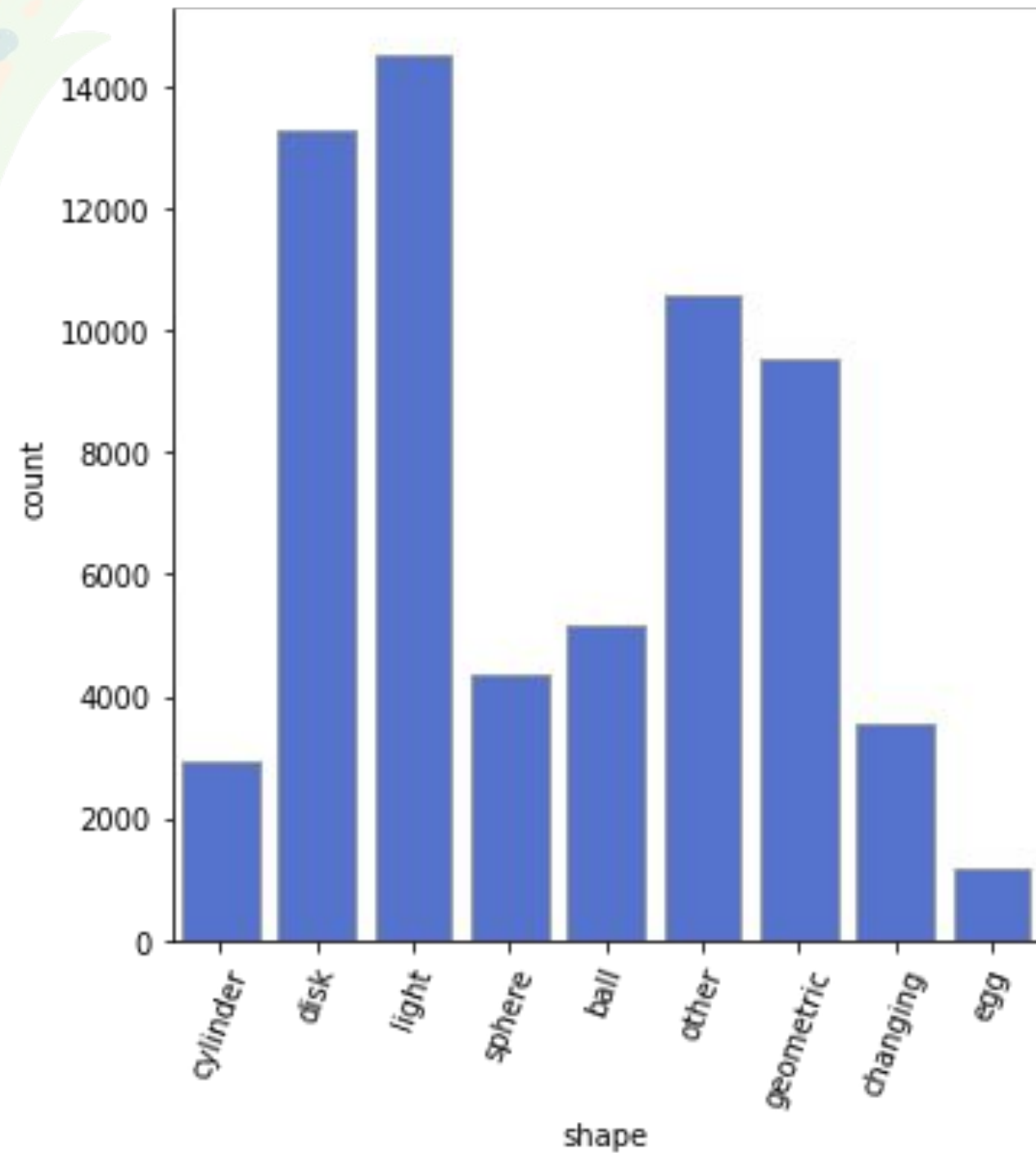
When do UFOs hang out?



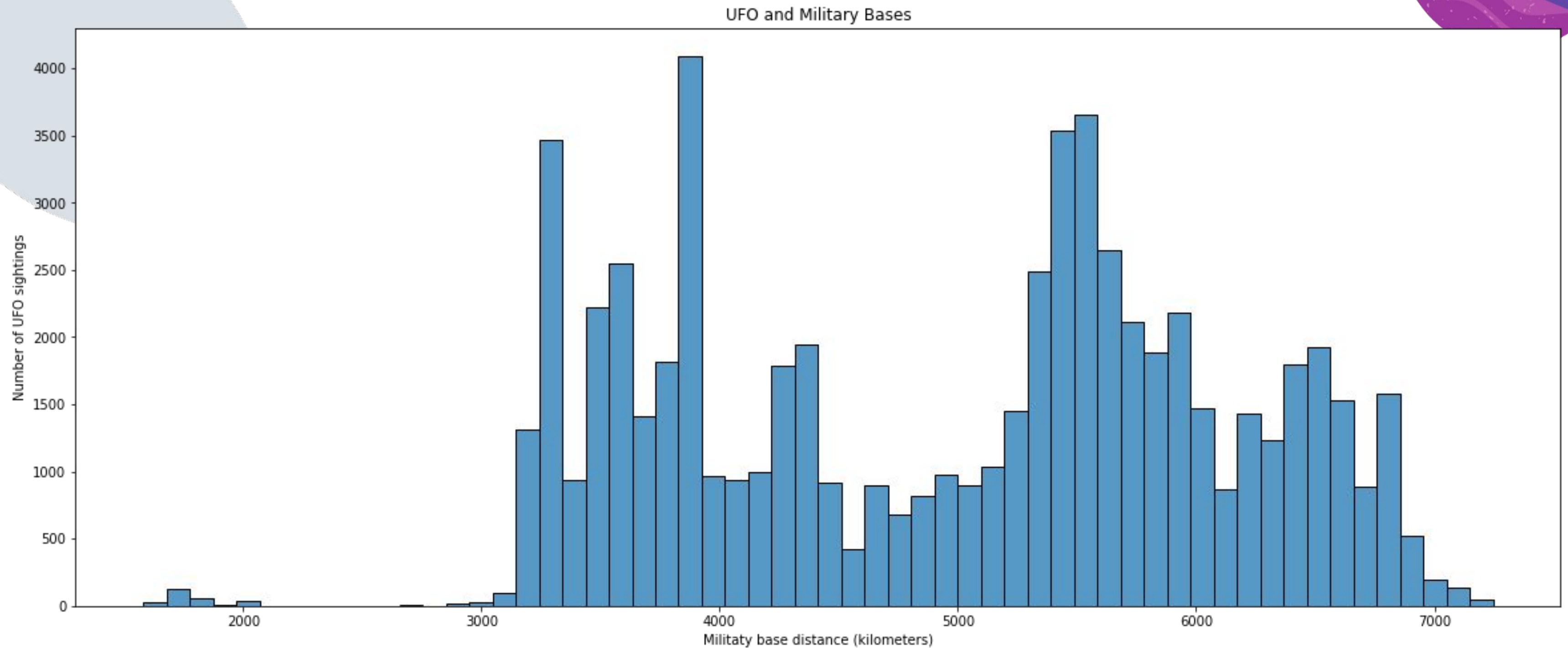
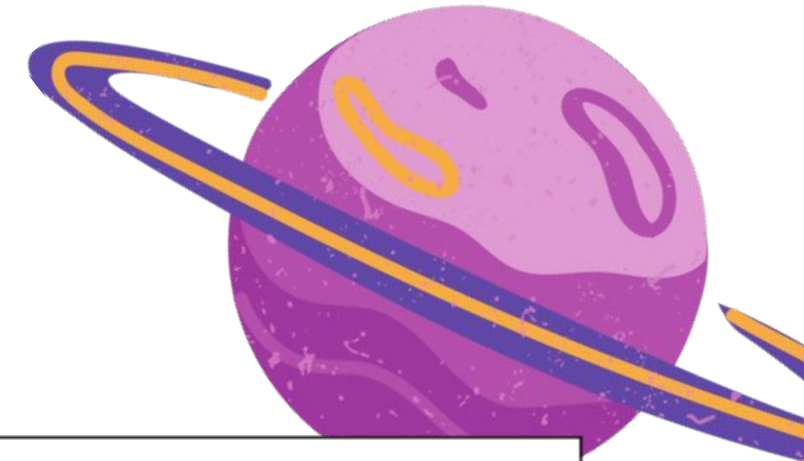
Duration of UFO sightings



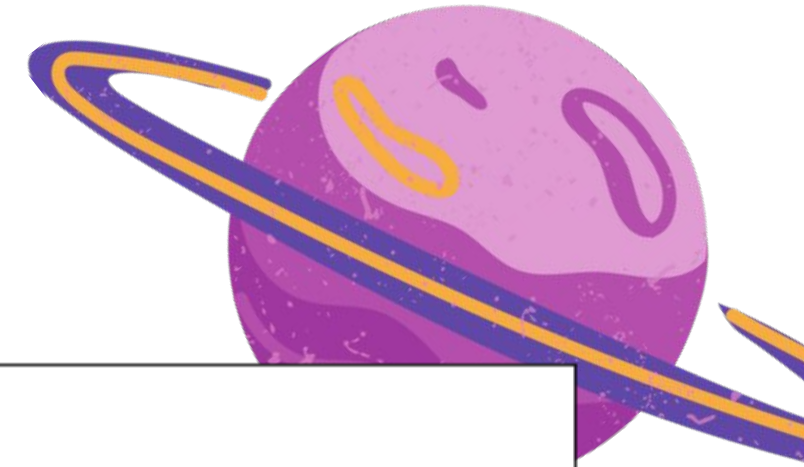
Shape of UFO sightings



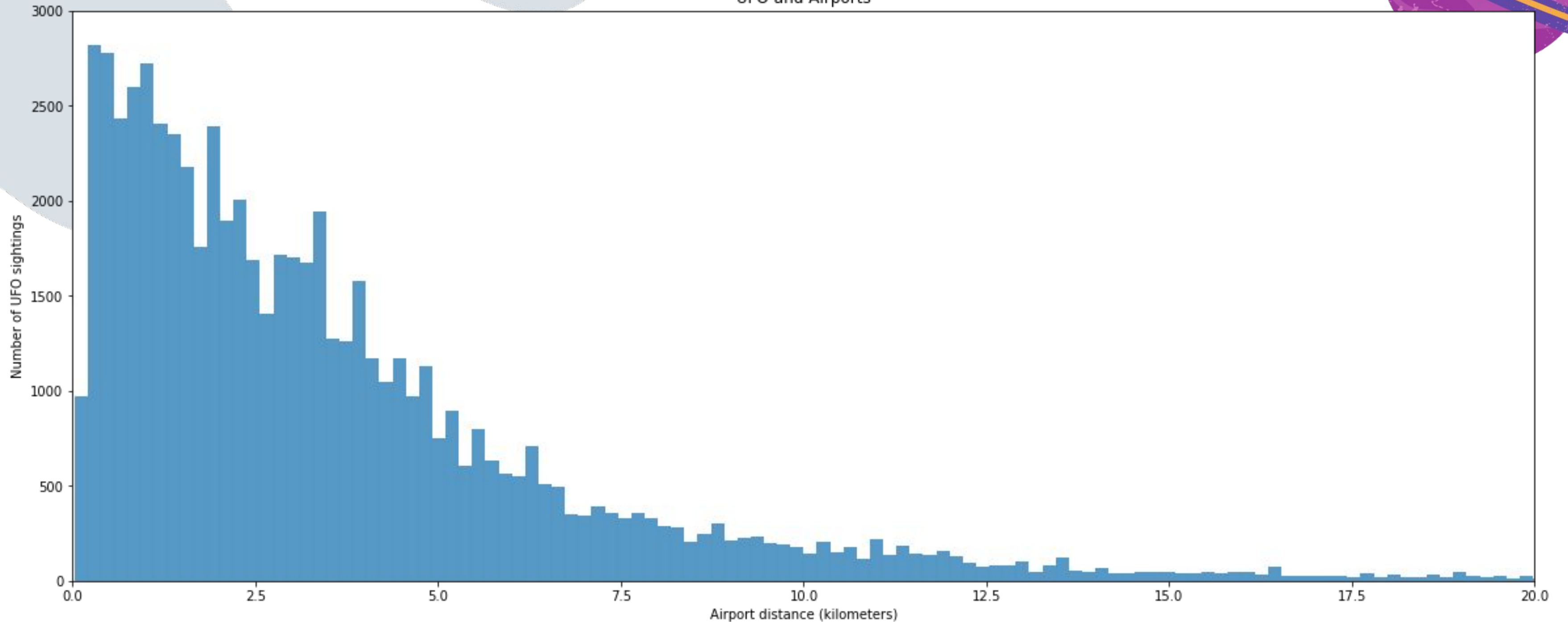
Is there correlation?

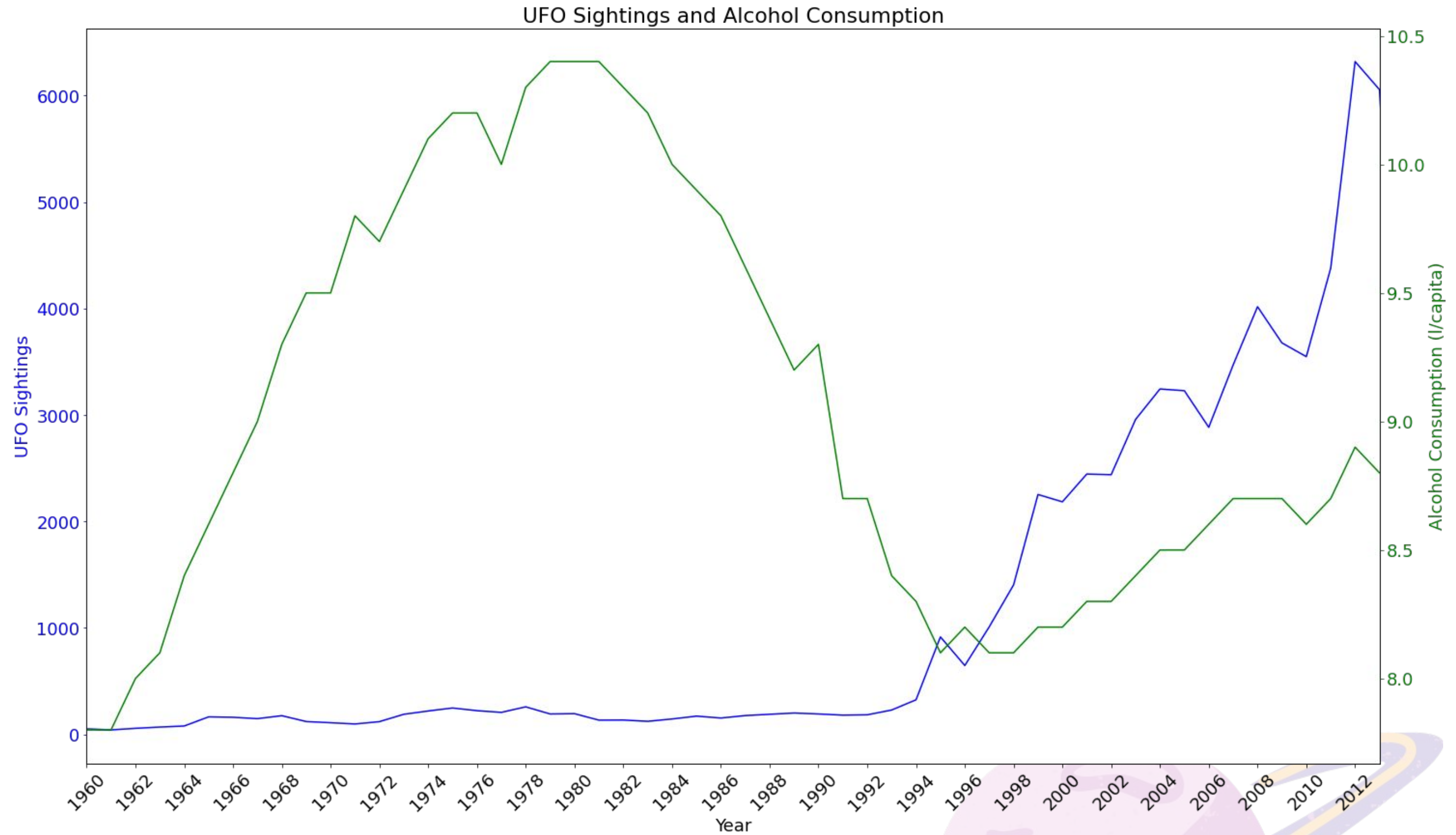


Is there correlation?



UFO and Airports





"I was driving home from Thanksgiving dinner at my friends house about five minutes away. It was a beautiful day and I had one heineken"

API Data

The TMDB logo, featuring the letters "TMDB" in a light blue, sans-serif font, followed by a solid blue rounded rectangle. The logo is set against a dark blue rectangular background.

- The API service is for anyone interested in using the database's movie, TV show or actor images and/or data in their application.
- The API is a system provided for anyone to programmatically fetch and use our data and/or images.
- The focus in this analysis was
 - Title
 - Release Date
 - Overview

CERTIFICATIONS

CHANGES

COLLECTIONS

COMPANIES

CONFIGURATION

CREDITS

DISCOVER

FIND

GENRES

GUEST SESSIONS

KEYWORDS

LISTS

MOVIES

NETWORKS

TRENDING

PEOPLE

REVIEWS

SEARCH

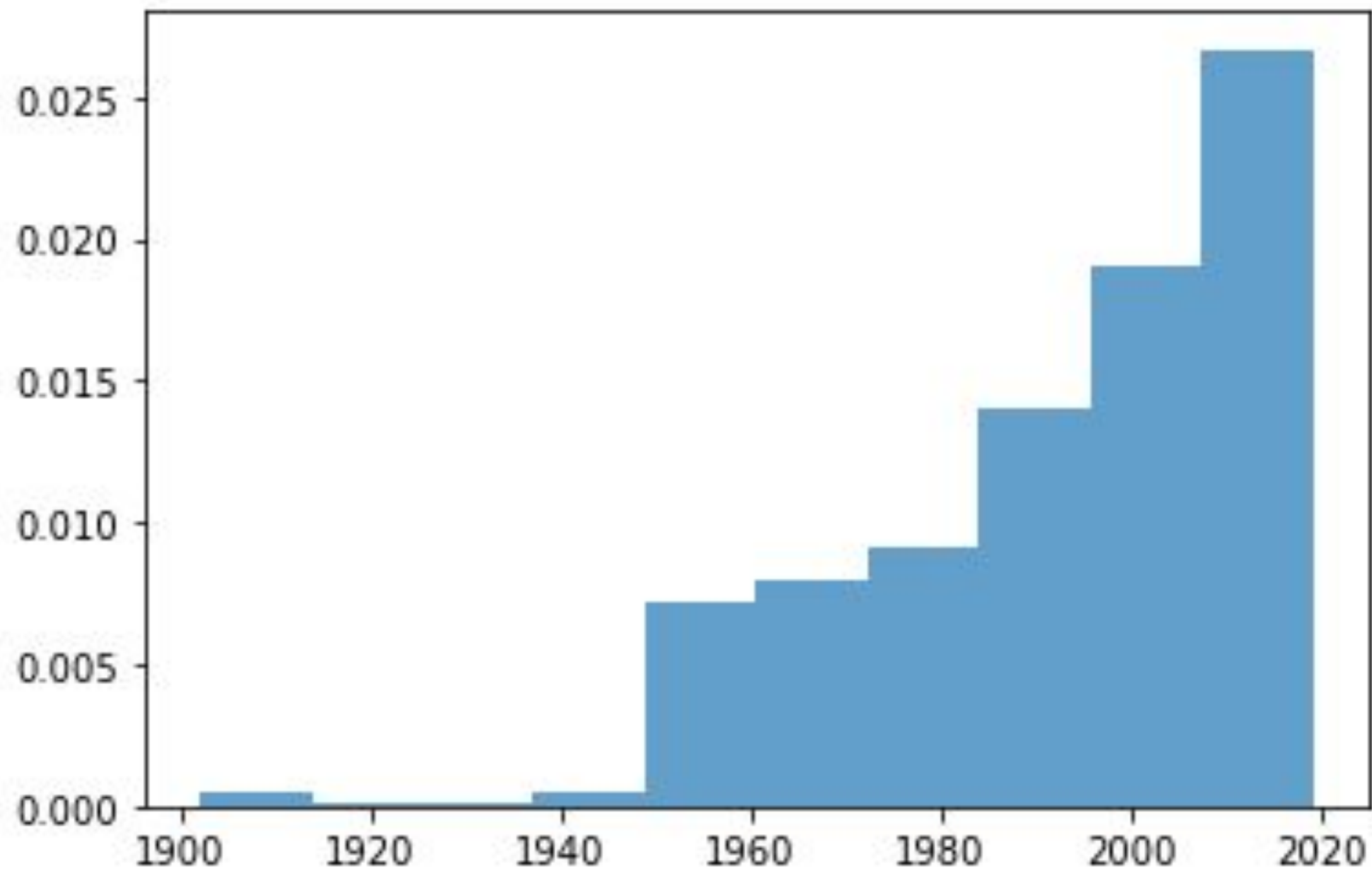
TV

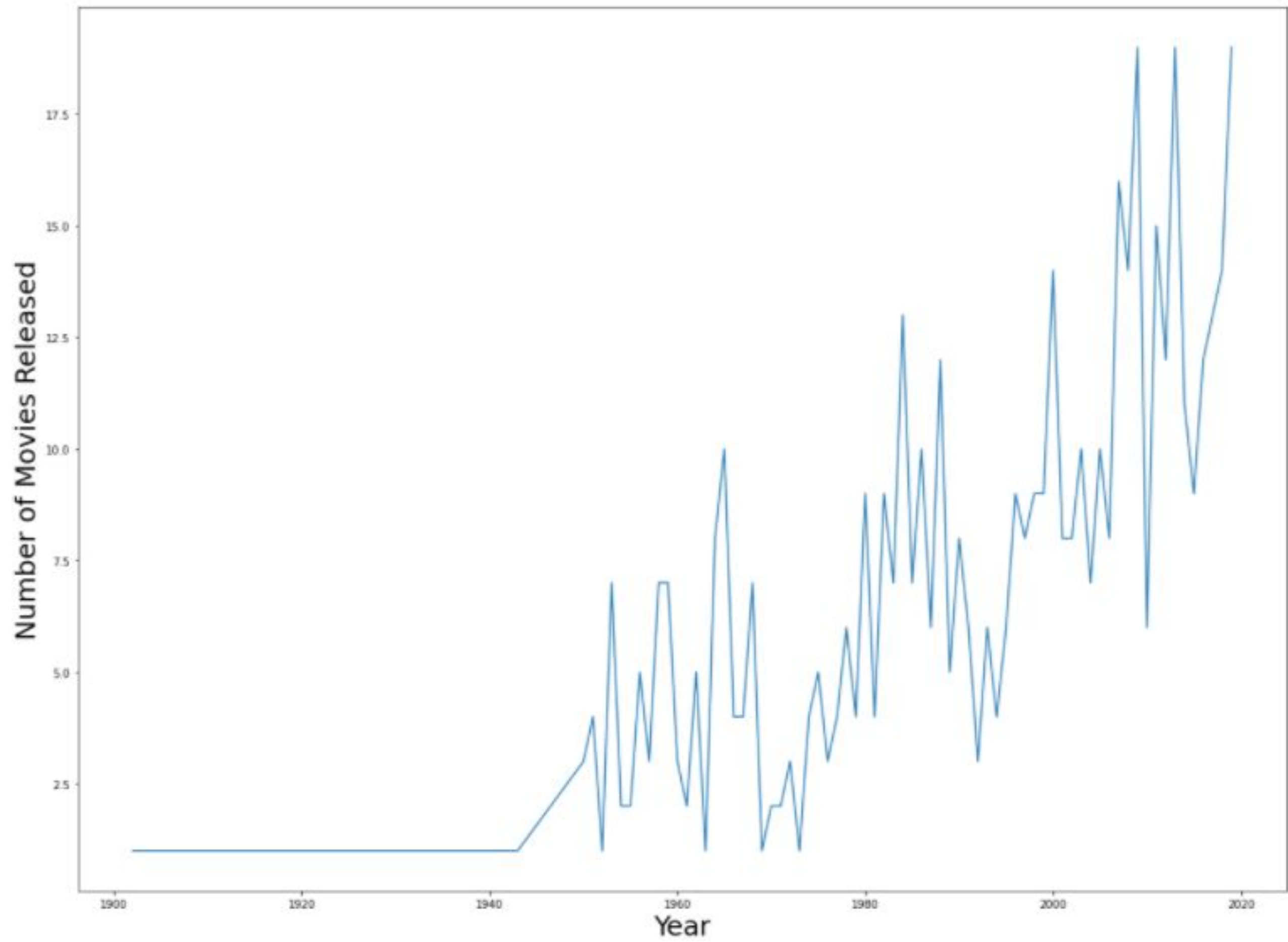
TV SEASONS

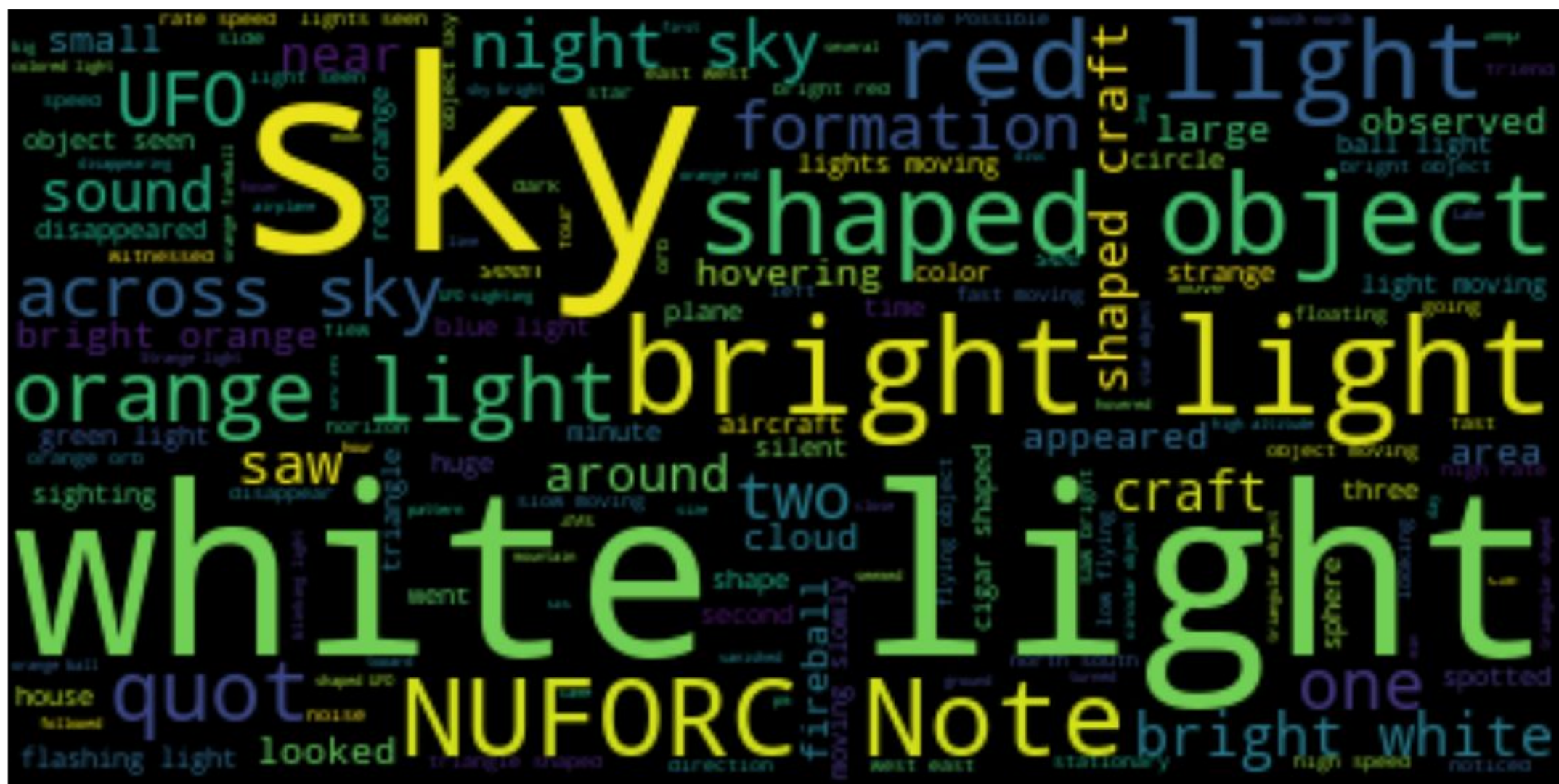
TV EPISODES

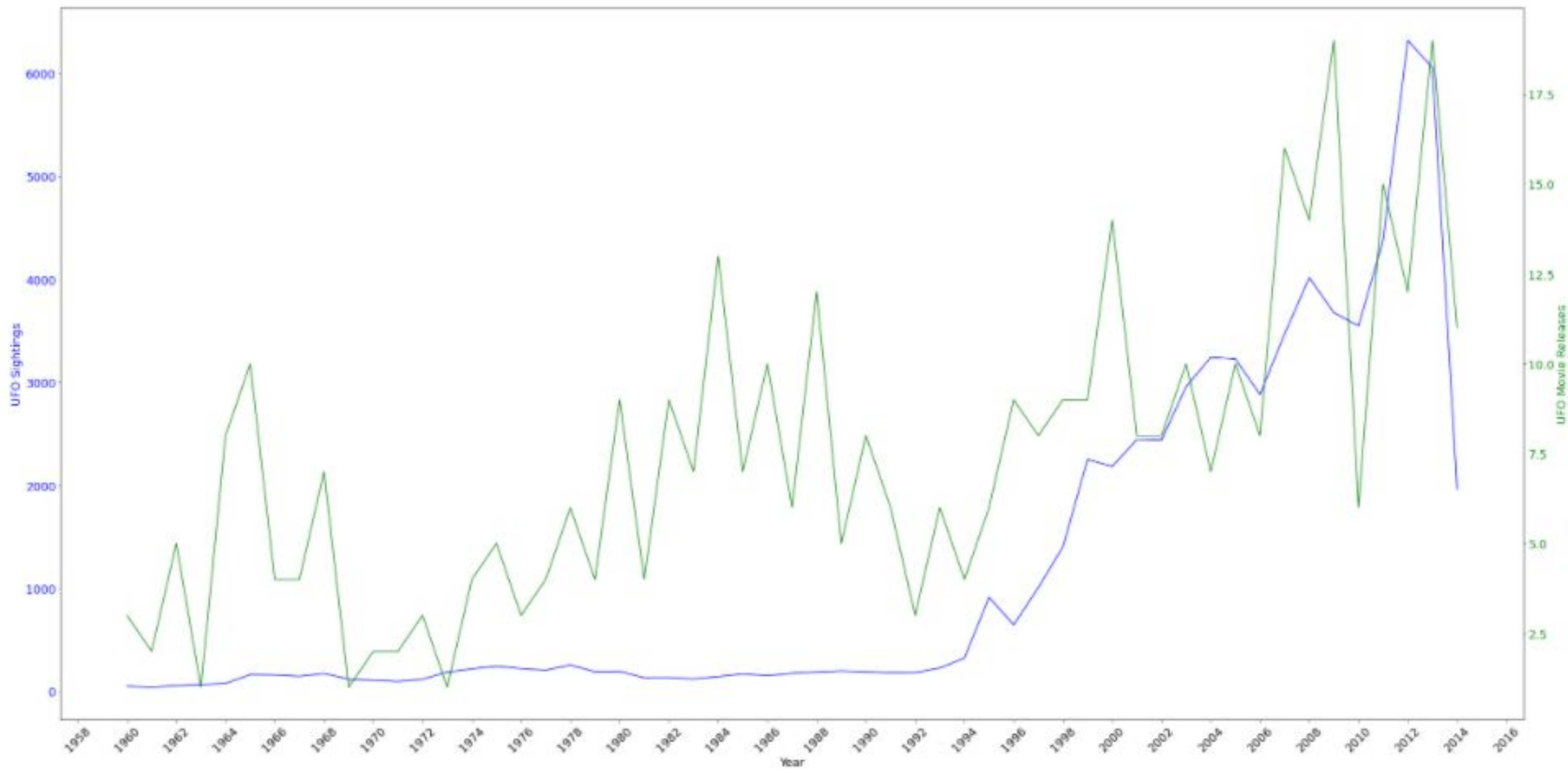
TV EPISODE GROUPS

Movie Release Distribution

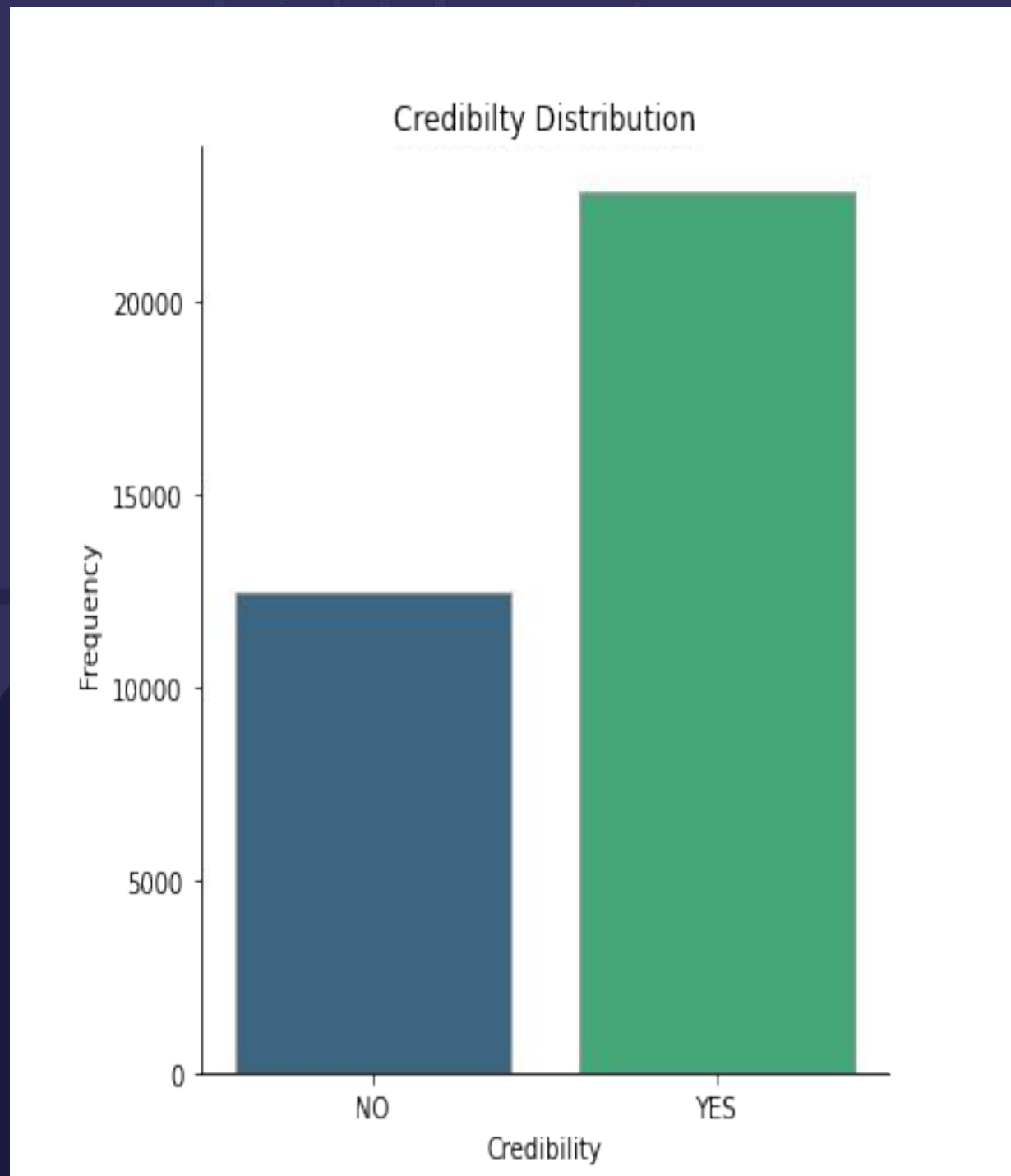








Classifying Reports



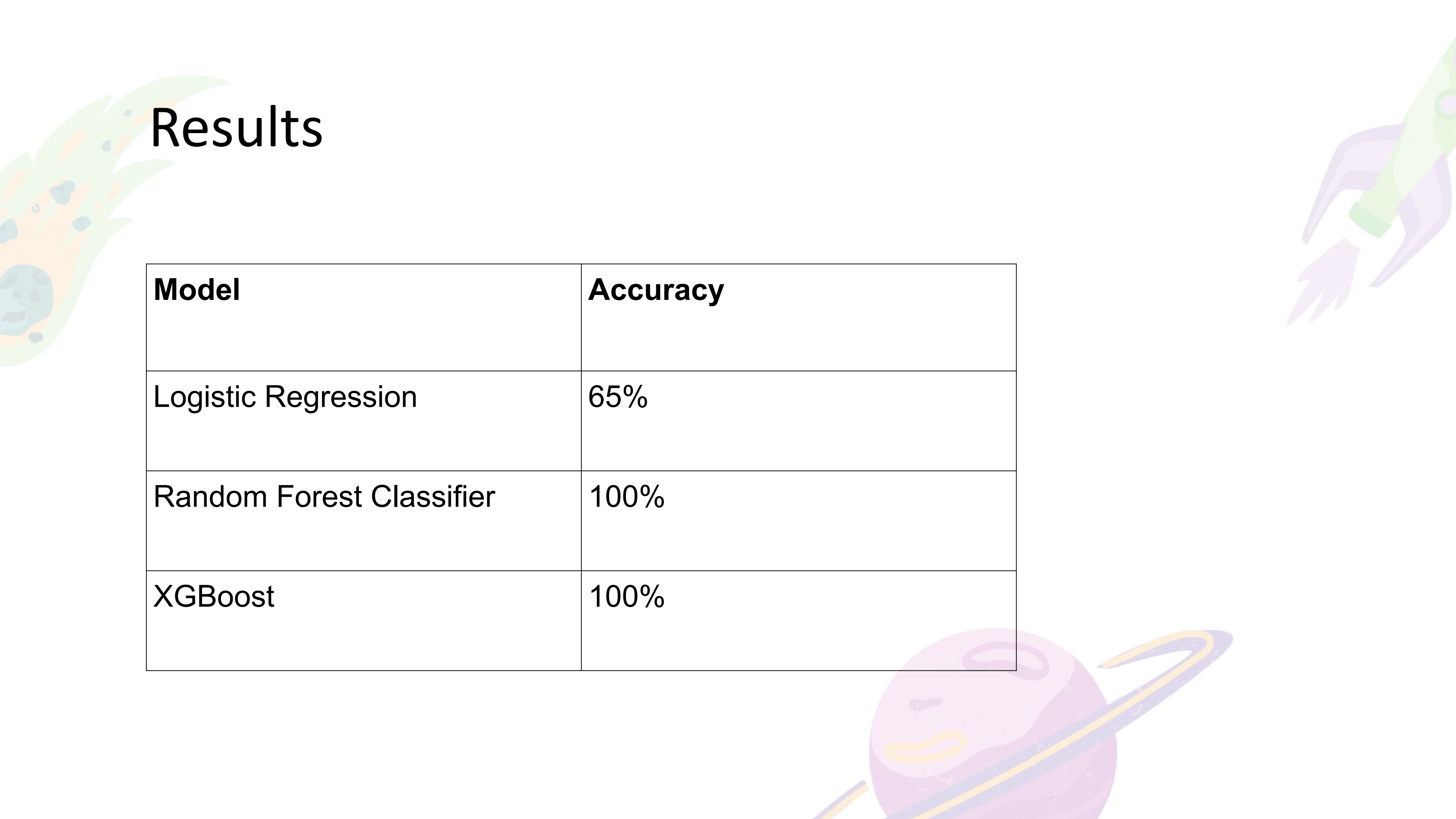
Reports were classified as not credible if:

- New Years Day
- Fourth of July
- rain
- hail
- tornado
- fog
- snow
- thunder
- < 1.6 kilometers from airport/base



Models

- Logistic Regression
- Random Forest Classifier
- XGBoost
- Tensorflow Sequential Model (Neural Network)
- K-Nearest-Neighbors



Results

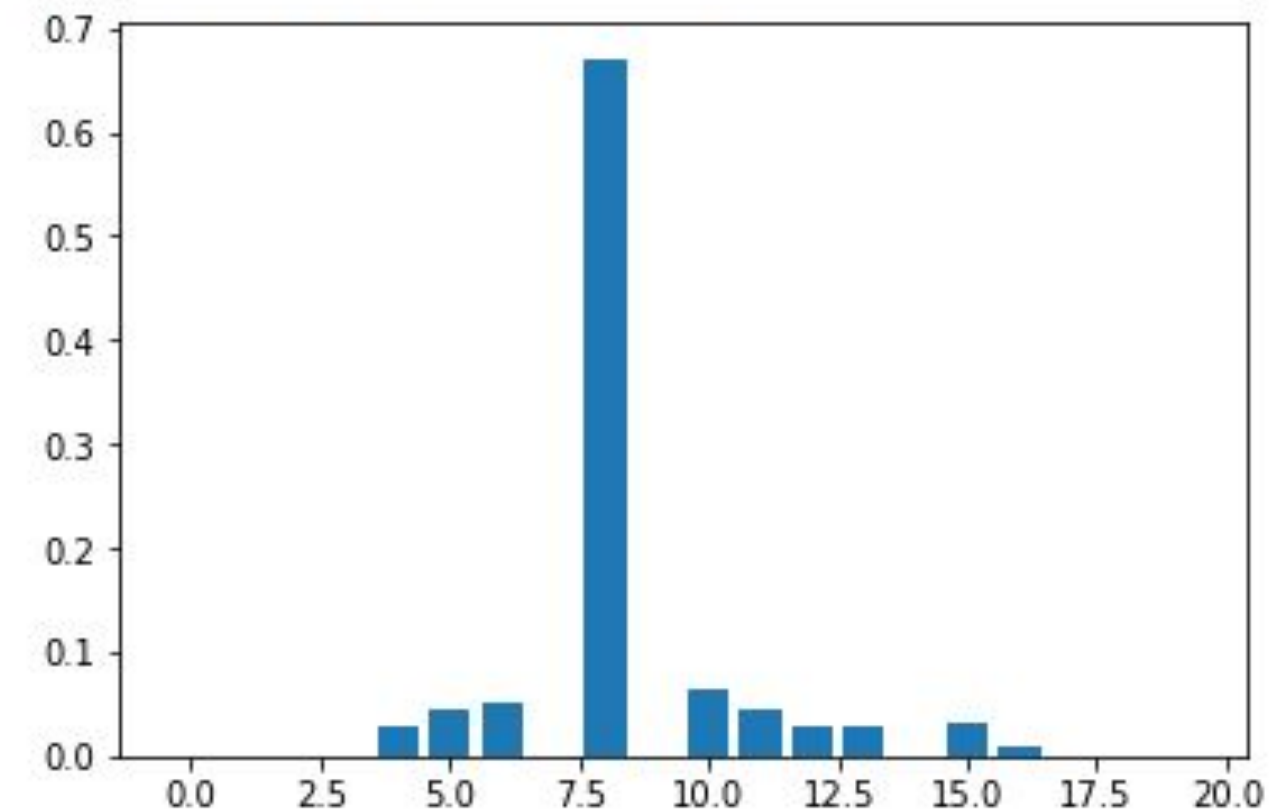
Model	Accuracy
Logistic Regression	65%
Random Forest Classifier	100%
XGBoost	100%

Results

Feature Importance per model:

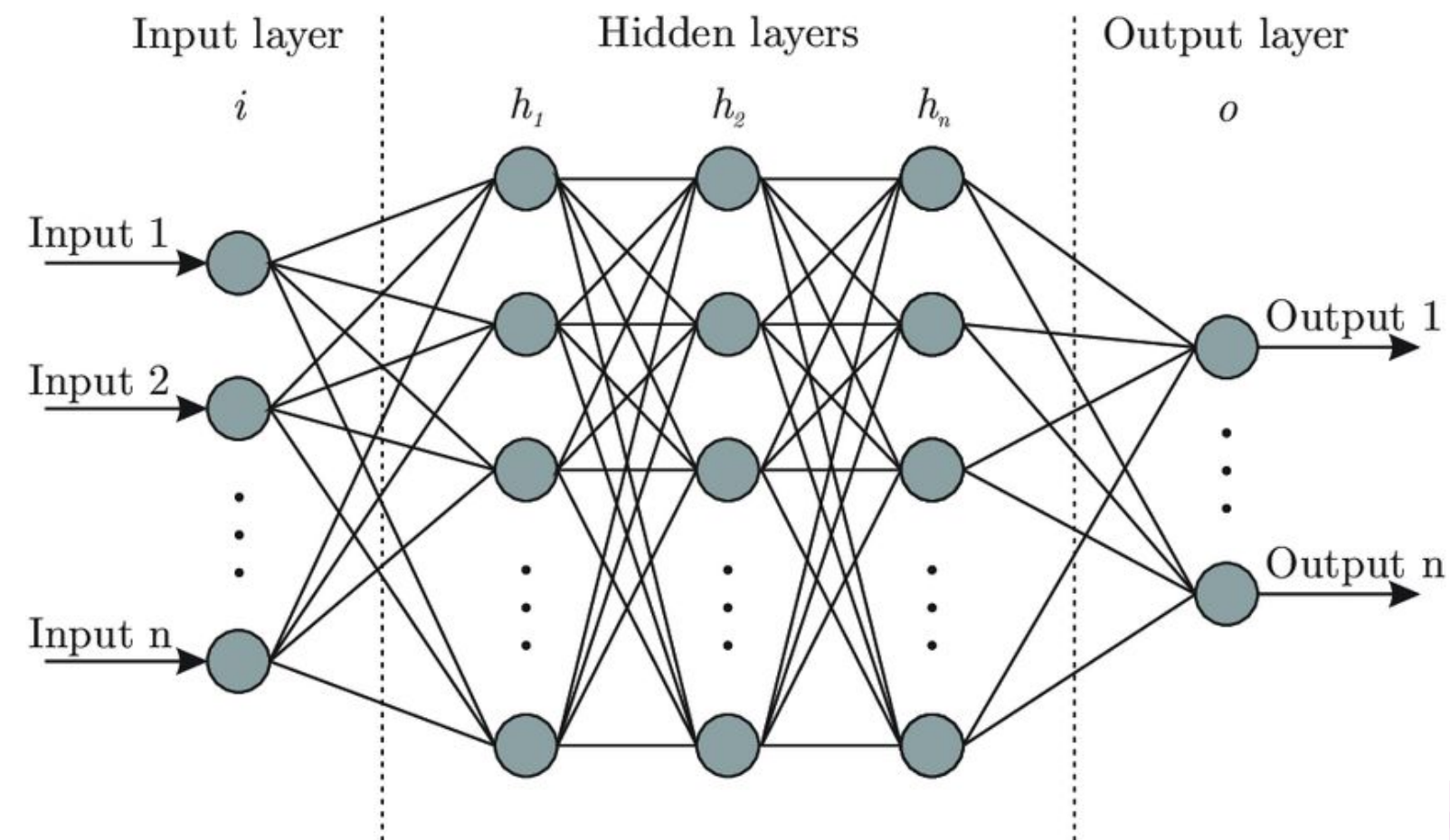
- Logistic Regression:
 - military base distance
- Random Forest Classifier:
 - airport distance
 - city
 - month
 - day
 - rain
- XGBoost:
 - airport distance
 - month
 - day
 - rain
 - thunder
 - snow

Feature: 0,	Score: 0.00000
Feature: 1,	Score: 0.00000
Feature: 2,	Score: 0.00000
Feature: 3,	Score: 0.00000
Feature: 4,	Score: 0.02819
Feature: 5,	Score: 0.04358
Feature: 6,	Score: 0.04986
Feature: 7,	Score: 0.00000
Feature: 8,	Score: 0.67089
Feature: 9,	Score: 0.00000
Feature: 10,	Score: 0.06411
Feature: 11,	Score: 0.04363
Feature: 12,	Score: 0.02993
Feature: 13,	Score: 0.02879
Feature: 14,	Score: 0.00000
Feature: 15,	Score: 0.03127
Feature: 16,	Score: 0.00975
Feature: 17,	Score: 0.00000
Feature: 18,	Score: 0.00000
Feature: 19,	Score: 0.00000



Results

Neural Network - yielded accuracy of 99%



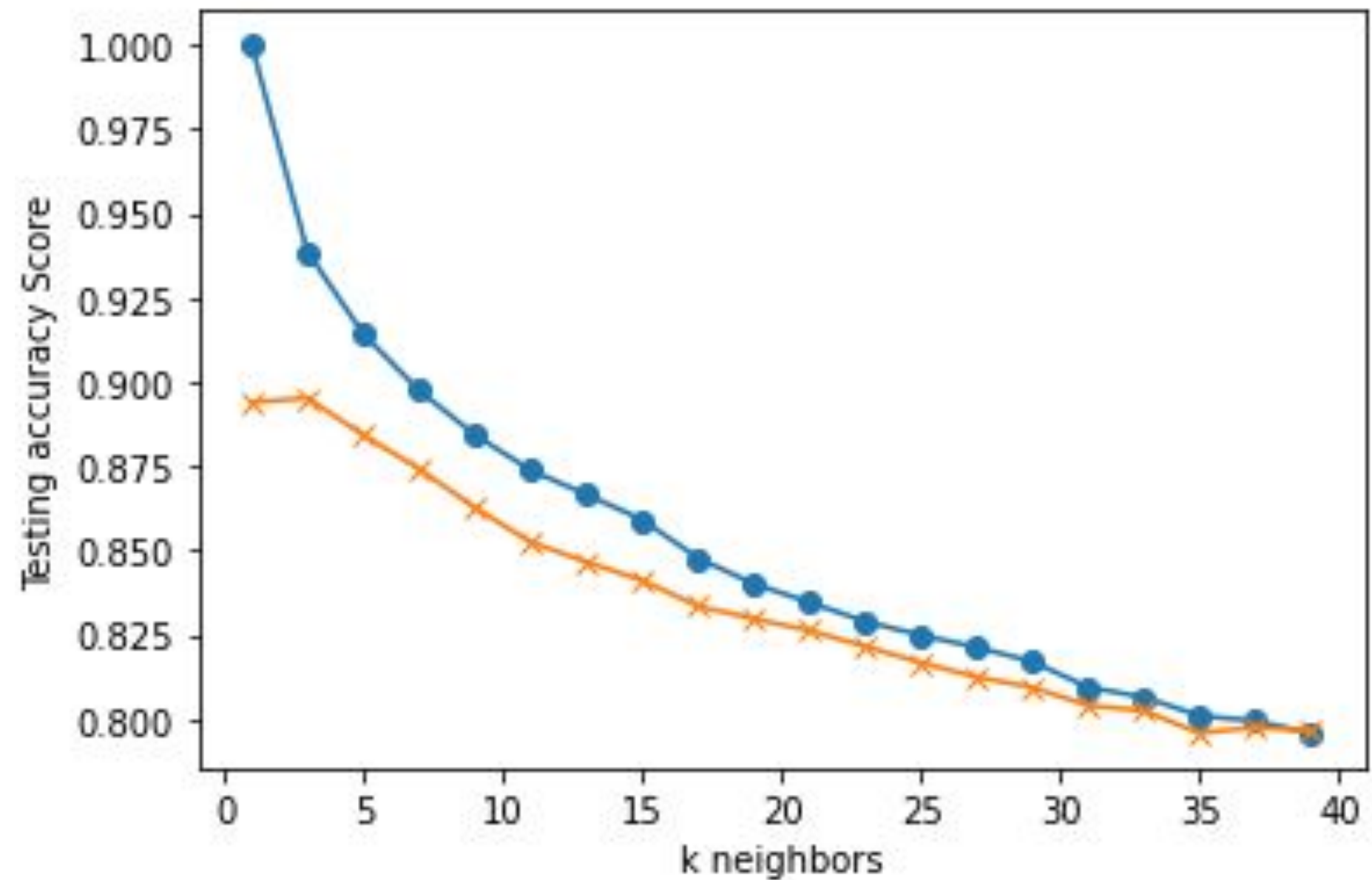
275/275 - 0s - loss: 0.0131 - accuracy: 0.9954

Normal Neural Network - Loss: 0.01306922361254692, Accuracy: 0.9954452514648438

Results

KNN - Accuracy for predicting:

- Credibility: 85%
- City: 35%





Recommendations

1

If you want to see an UFO, we recommend spending time outside during the Summer in a big city (somewhere in California). UFOs favorite day to come out is the Fourth of July.

2

If you want to avoid UFOs, stay indoors, don't hang around airports, and keep sober.



Thank You!

Questions or Comments?

References

Airport Dataset: <https://data.humdata.org/dataset/ourairports-usa>.

Alcohol Dataset <https://www.cdc.gov/alcohol/data-table-text.htm#intensity>

Drug Use: <https://www.cdc.gov/drugoverdose/maps/rxstate2010.html>

Military Base Dataset:

https://www.google.com/maps/d/viewer?mid=1hvB-oq9gE0H8gEwKJ4XHFOKaY5k&hl=en_US&ll=32.131165347452765%2C-104.03214912642196&z=5

Movie Dataset: <https://www.themoviedb.org/>

Population Dataset https://simple.wikipedia.org/wiki/List_of_U.S._states_by_population

UFO Dataset: <https://www.kaggle.com/NUFORC/ufo-sightings>

Weather Dataset: Consolidated UFO and Weather Data | Kaggle

Map Code: <https://www.kaggle.com/daveianhickey/how-to-folium-for-maps-heatmaps-time-data/>