

UFO SPOTTING

AUGUST 13, 2018

Joey Chen
Anuraag Mohile
Catherine McNabb
Preetika Srivastava



Agenda

- Description of Data
- Exploratory Analysis
- Insights
- Recommendations



DESCRIPTION OF DATA



Why UFO Data?

- UFOlogy – the study of UFO's
- Pop culture – Men in Black, ET, Star Wars, Independence Day
- Conspiracy Theories – Area 51, Stonehenge, Egyptian Pyramids
- “1996 Gallup Poll reported that 71 percent of the United States population believed that the U.S. government was covering up information regarding UFOs” (Wikipedia)



Data Background

In the data set:

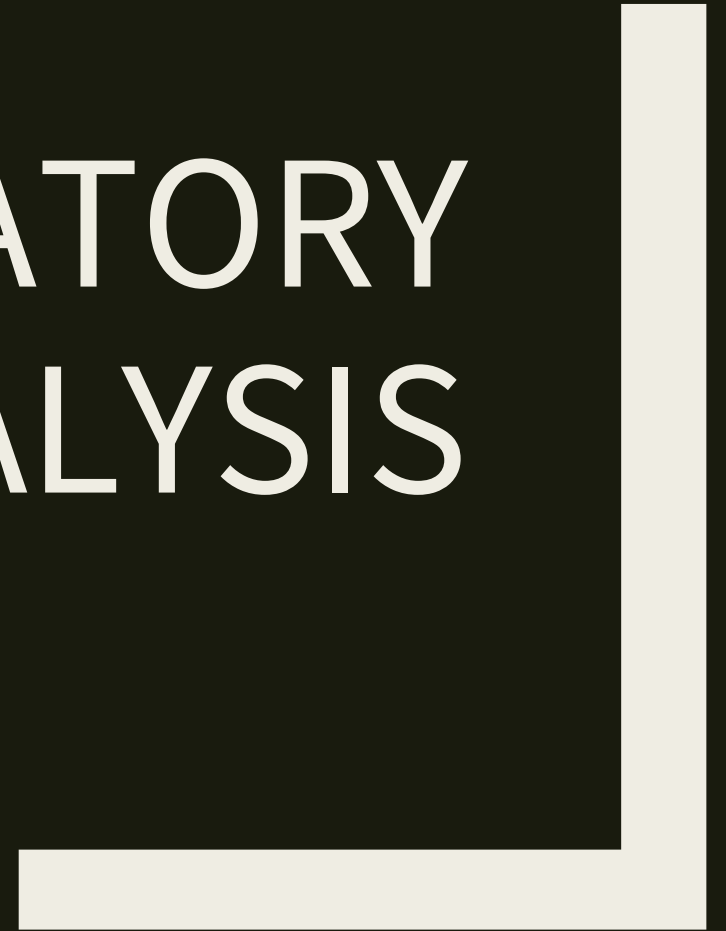
- Date and time
- Latitude and Longitude
- City, State, and Country
- Shape
- Duration
- Unexplainable v Explainable

Added:

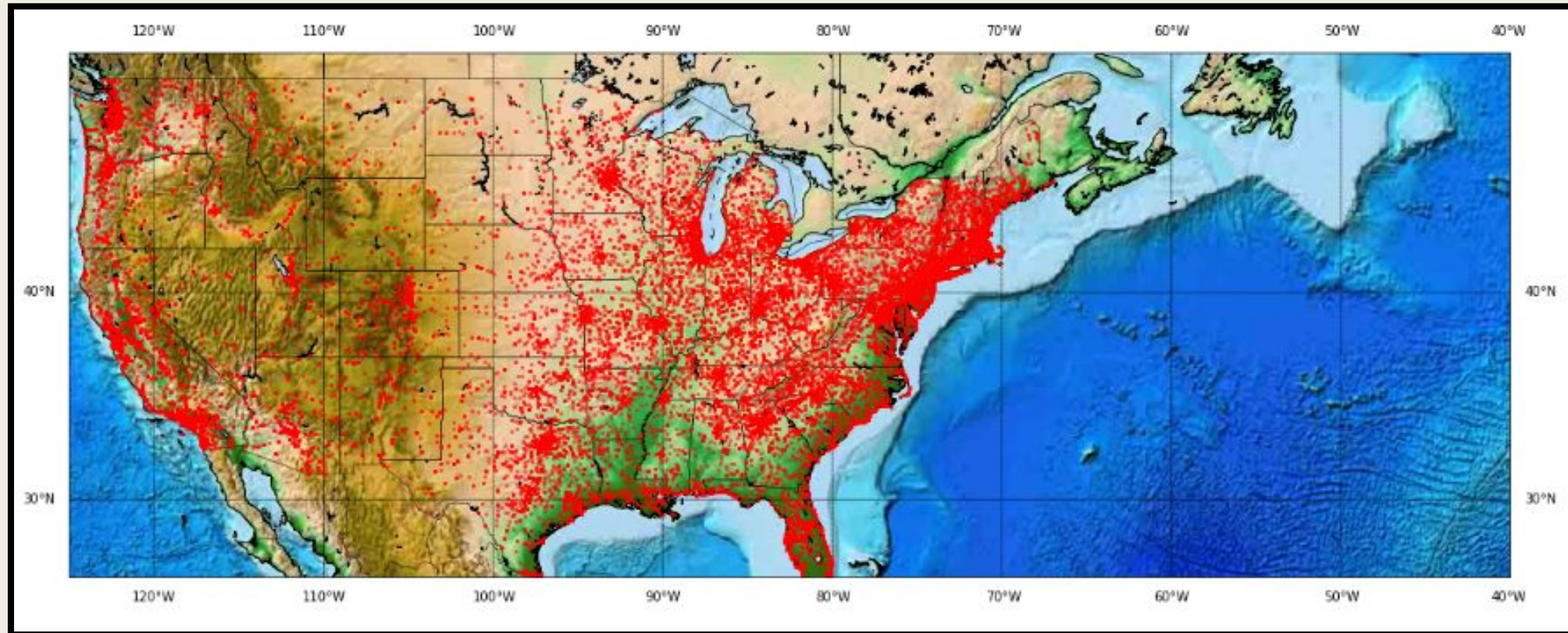
- Weather
- Day of the Week
- Moon Phase
- Proximity to US Airforce Base



EXPLORATORY ANALYSIS



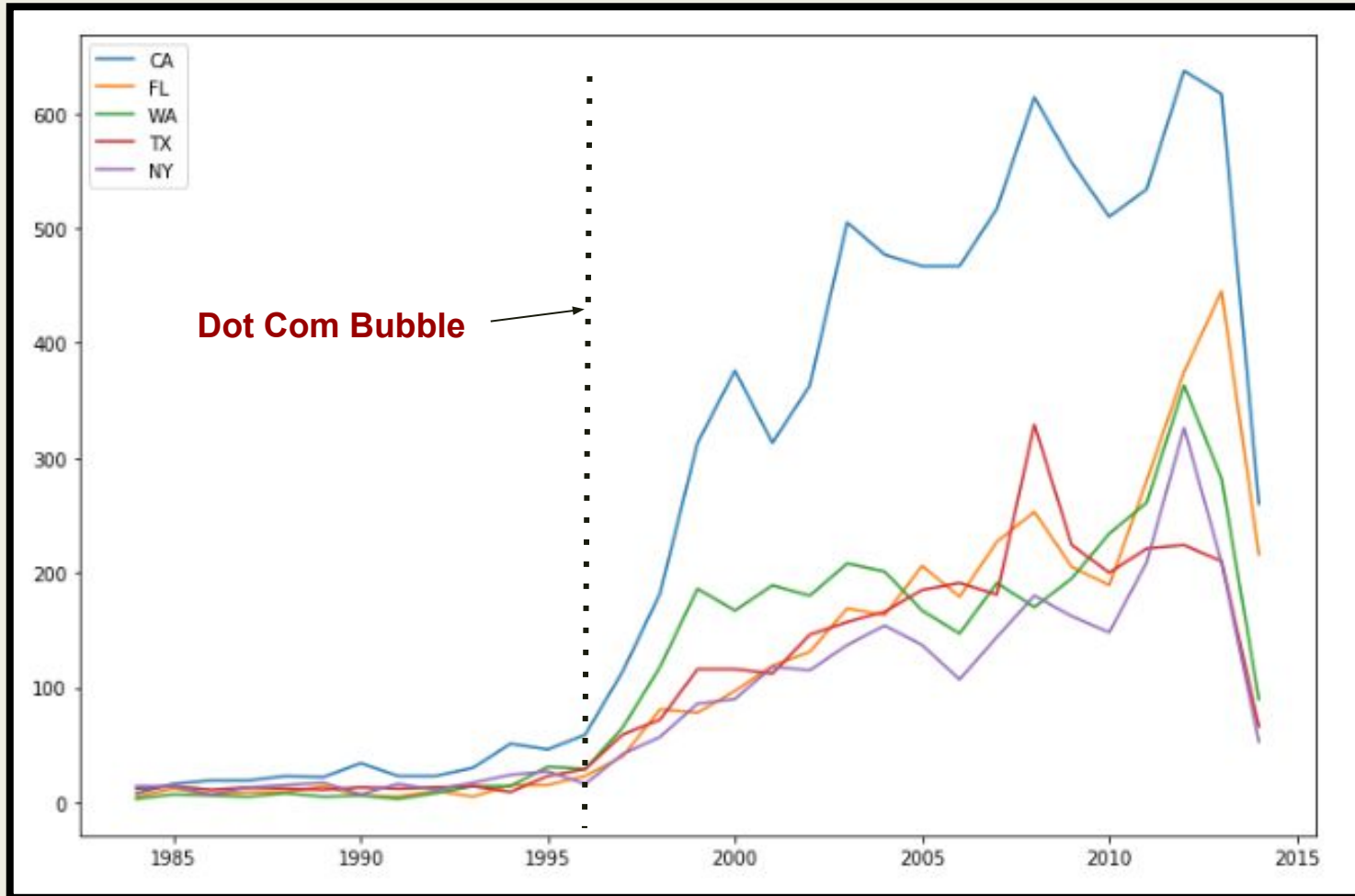
Sightings in the United States



created using `mpl_toolkits::basemap`



Sightings Over Time



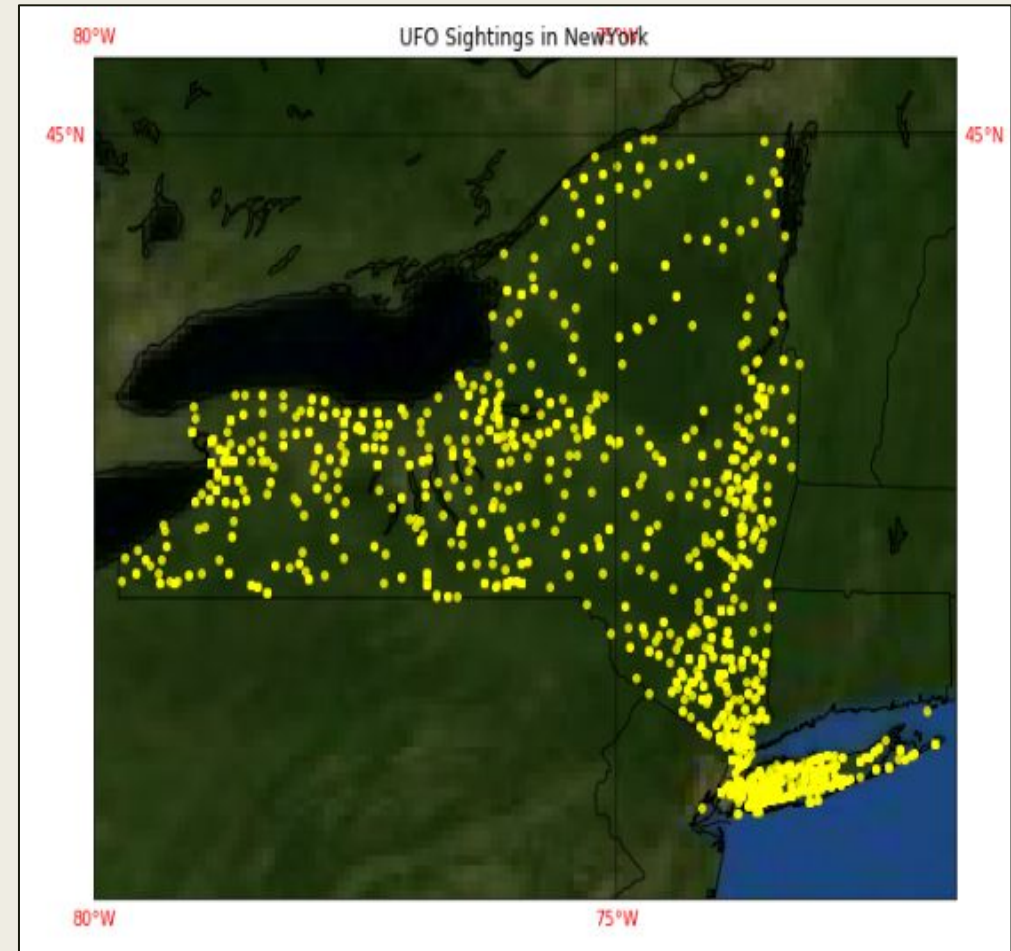
Do you see any patterns ?

FLORIDA

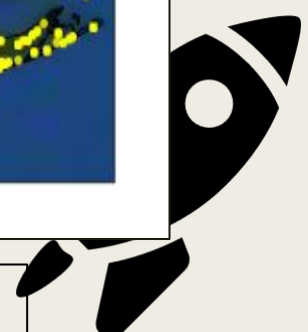


I - 95 Corridor

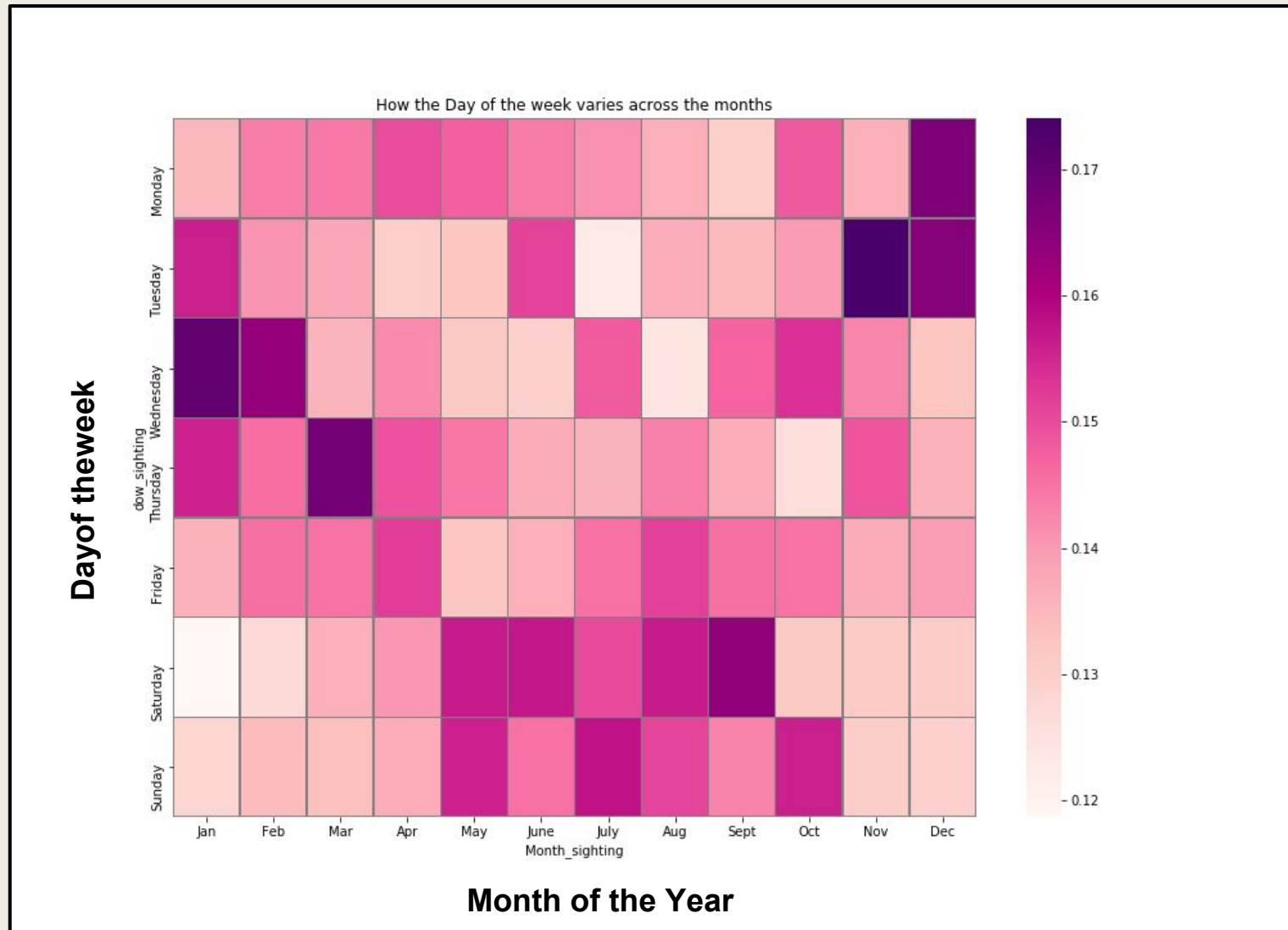
NEW YORK



I - 87 Corridor



When Will I See a UFO?

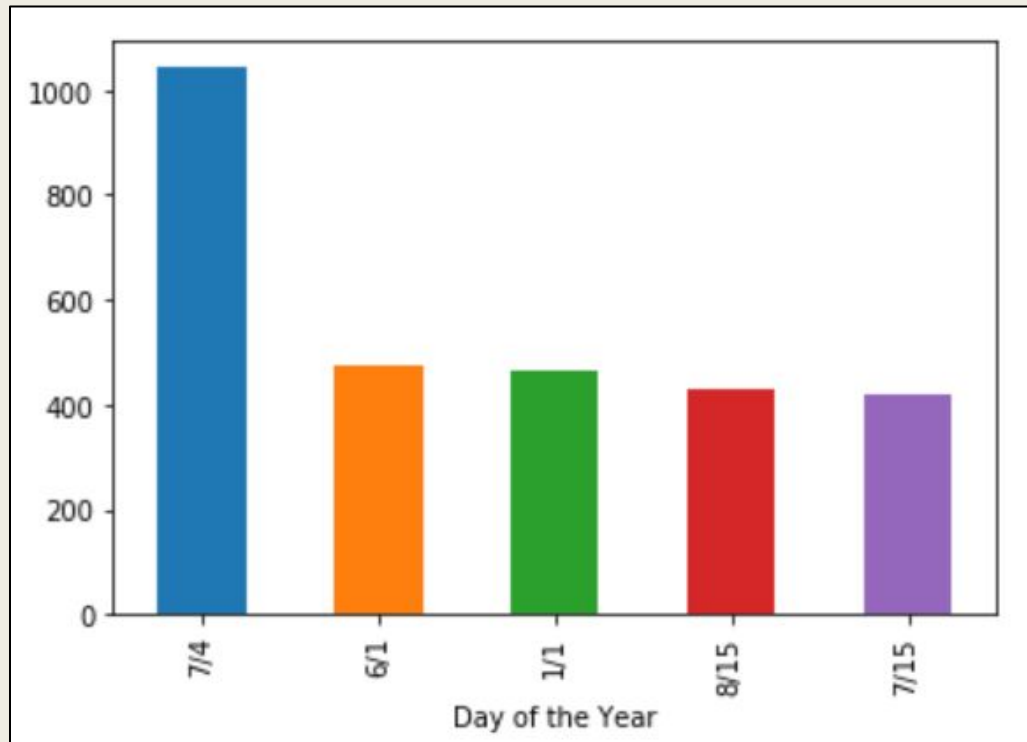


created using **seaborn**

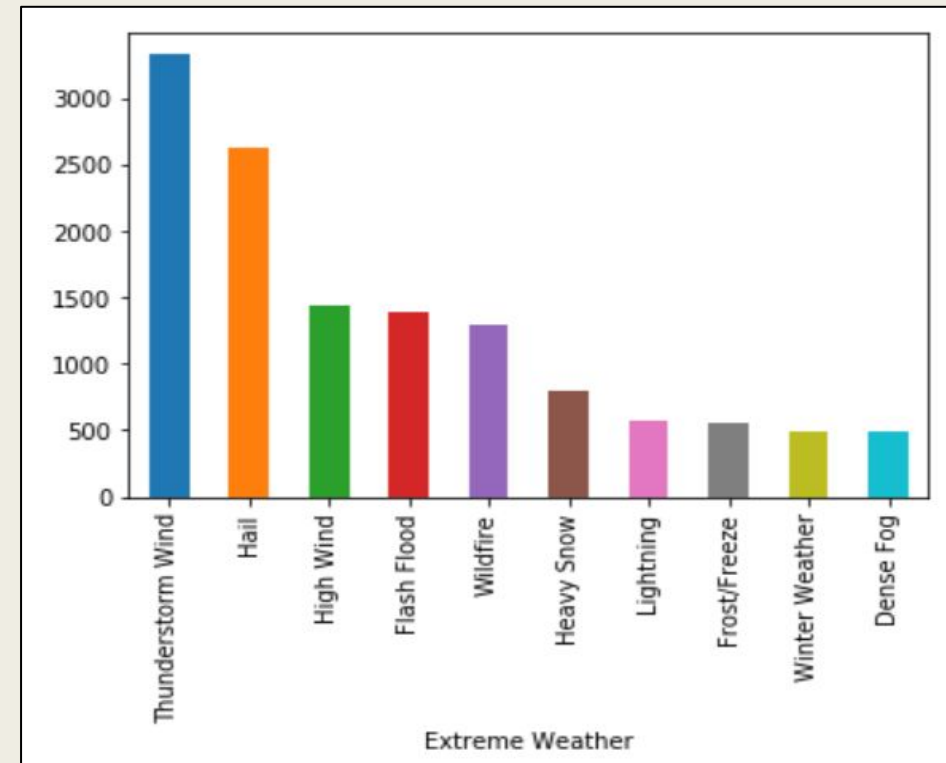


When Will I See a UFO?

Day of the Year

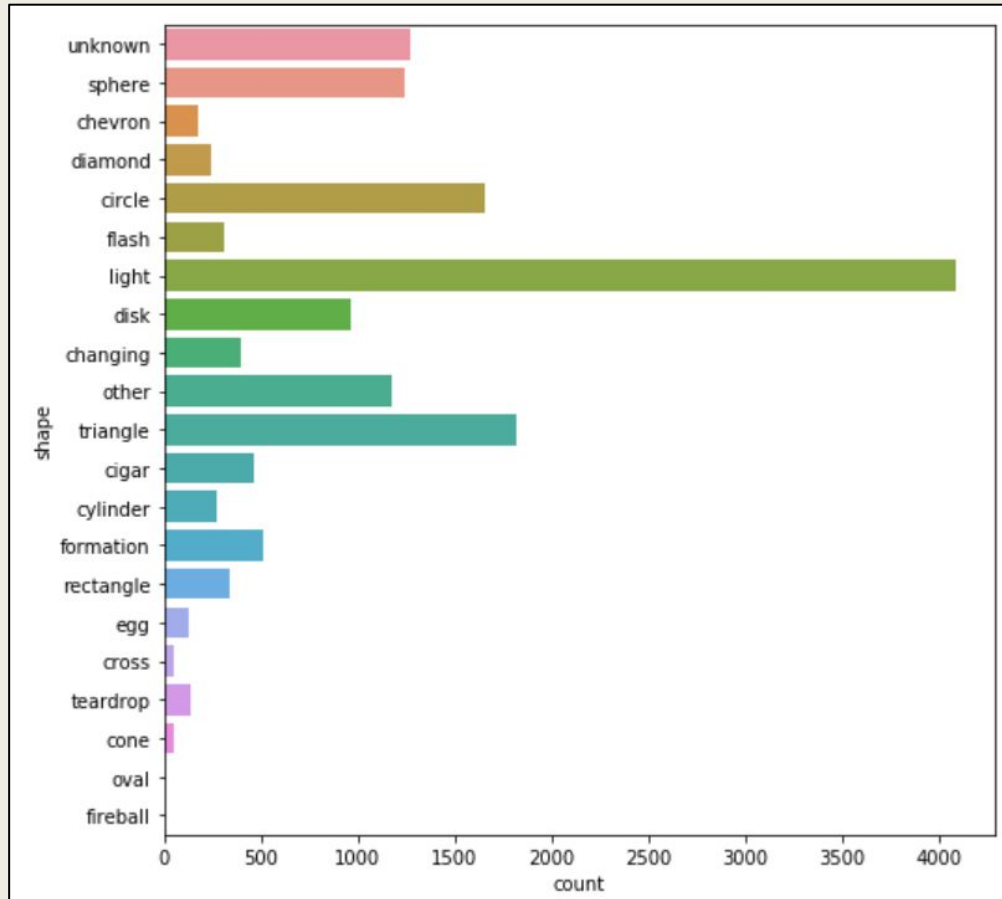


Severe Weather Day of Sighting

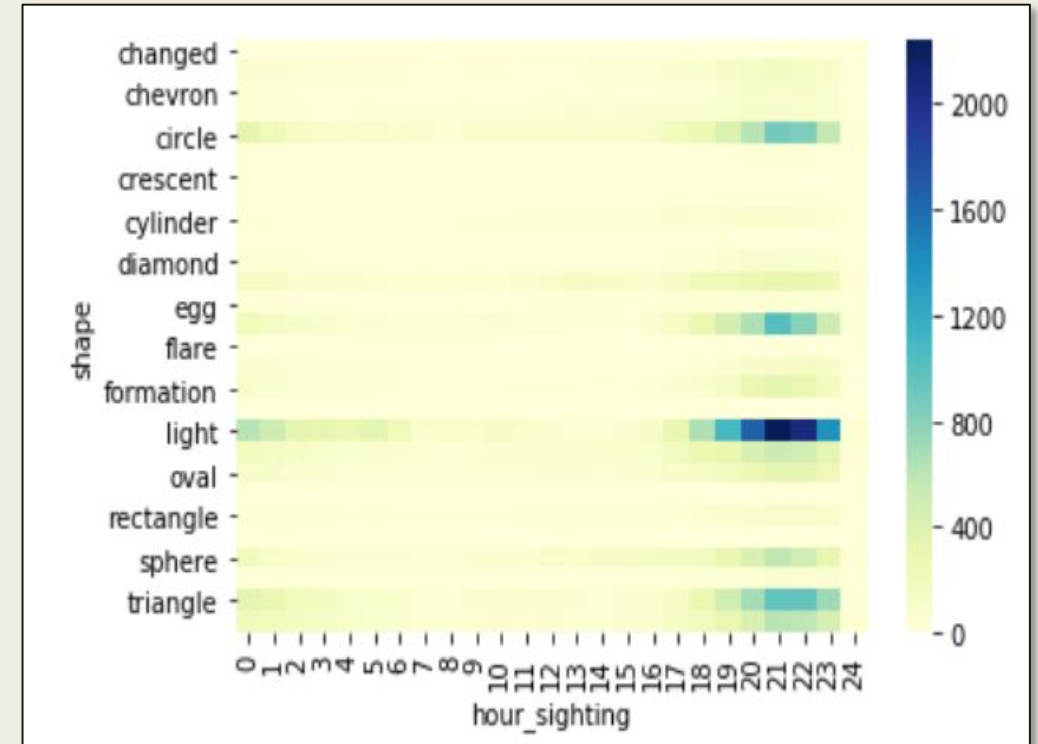


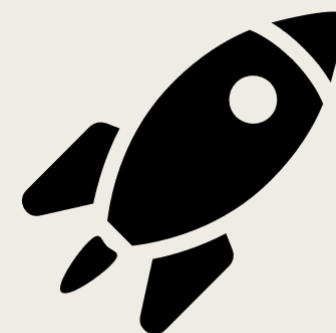
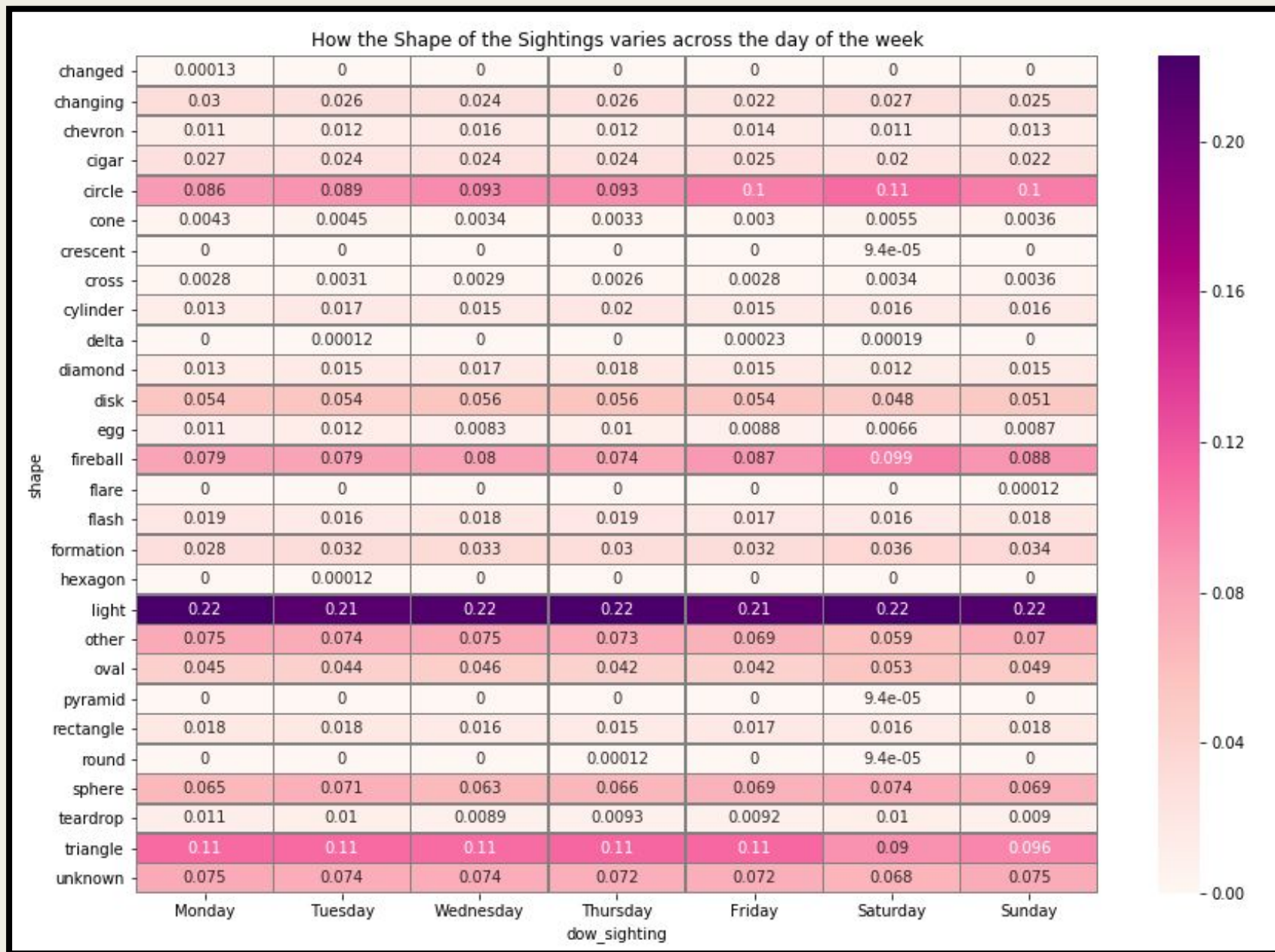
What Will I See?

Type of Shape



Shape vs. Hour





INSIGHTS



Logistic Regression

Y = Explainable Events

*created using **scikit-learn***

Test Data		Prediction		
Confusion Matrix		0	1	Total
Actual	0.0	28,742	44	28,786
	1.0	941	114	1,055
	Total	29,683	158	29,841

Precision-Recall Table			
	Precision	Recall	Support
0.0 (Unexplained)	0.97	1.00	28,786
1.0 (Explained)	0.72	0.11	1,055
Avg/Total	0.96	0.97	29,841

Most Unexplainable Shapes:

1. Cigar
2. Chevron
3. Triangle
4. Egg

Most Unexplainable States:

1. Arkansas
2. Idaho
3. Wisconsin



Random Forest

Training Data Confusion Matrix

Prediction Actual	0	1
0	34541	2
1	276	989

Accuracy : 0.9922

Test Data Confusion Matrix

Prediction Actual	0	1
0	22974	55
1	756	88

Accuracy : 0.9660

	precision	recall	f1-score	support
0.0	0.97	1.00	0.98	23029
1.0	0.62	0.10	0.18	844
avg / total	0.96	0.97	0.95	23873

created using **scikit-learn**



Random Forest: Threshold = 0.95

Training Data Confusion Matrix

Prediction Actual	0	1
0	31675	2868
1	0	1265

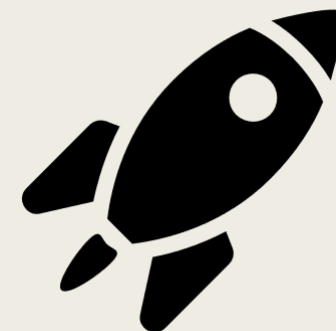
Accuracy : 0.9199

Test Data Confusion Matrix

Prediction Actual	0	1
0	18819	4210
1	472	372

Accuracy : 0.8039

	precision	recall	f1-score	support
0.0	0.98	0.82	0.89	23029
1.0	0.08	0.44	0.14	844
avg / total	0.94	0.80	0.86	23873



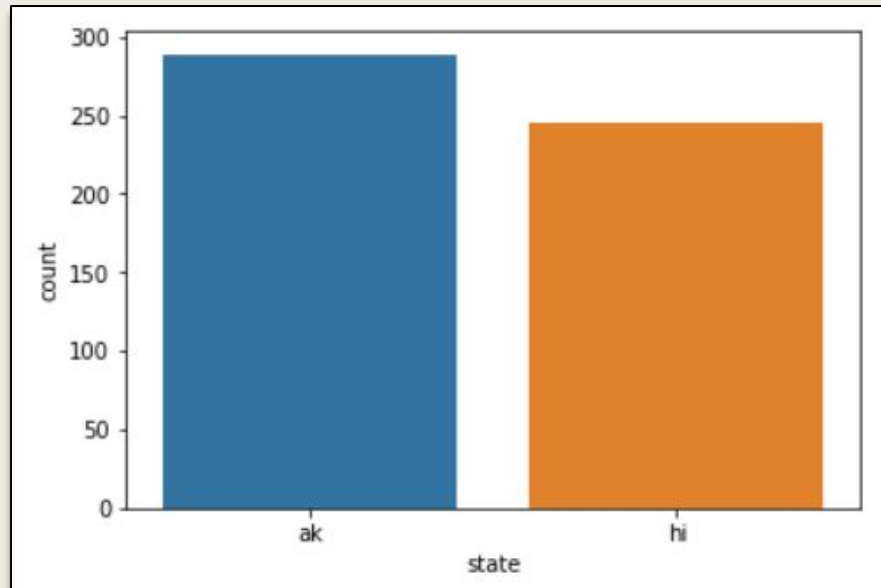
Random Forest: Feature Importance

Features	Importance
Time	0.010399
Proximity to USAFB	0.004348
Date	0.002605
Days Lapsed after Sighting	0.001957
Duration (seconds)	0.001336
Day of the Week	0.001215
Shape	0.001099
State	0.000900
Extreme Weather	0.000421
Moon Phase	0.000001

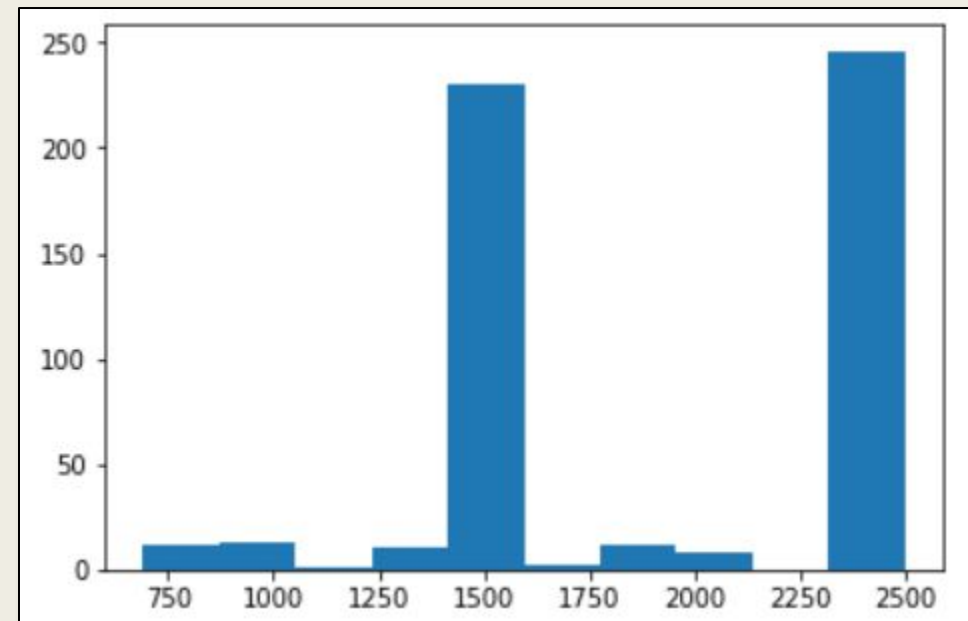


K Means Clustering

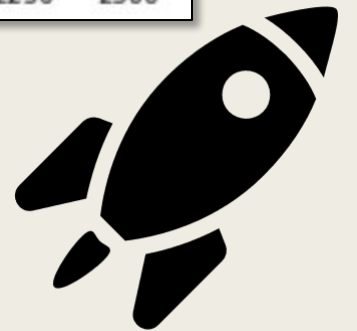
of Sightings by State



Proximity to Air Force Base

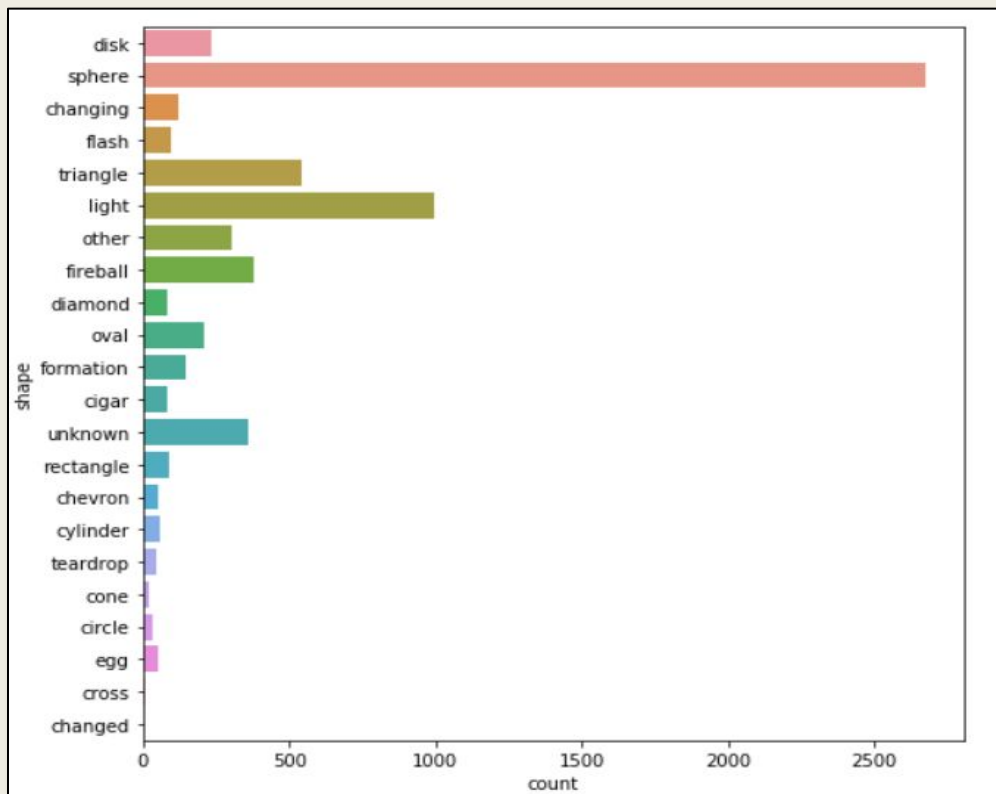


- Alaska and Hawaii
- ≥ 1500 miles from an airforce base

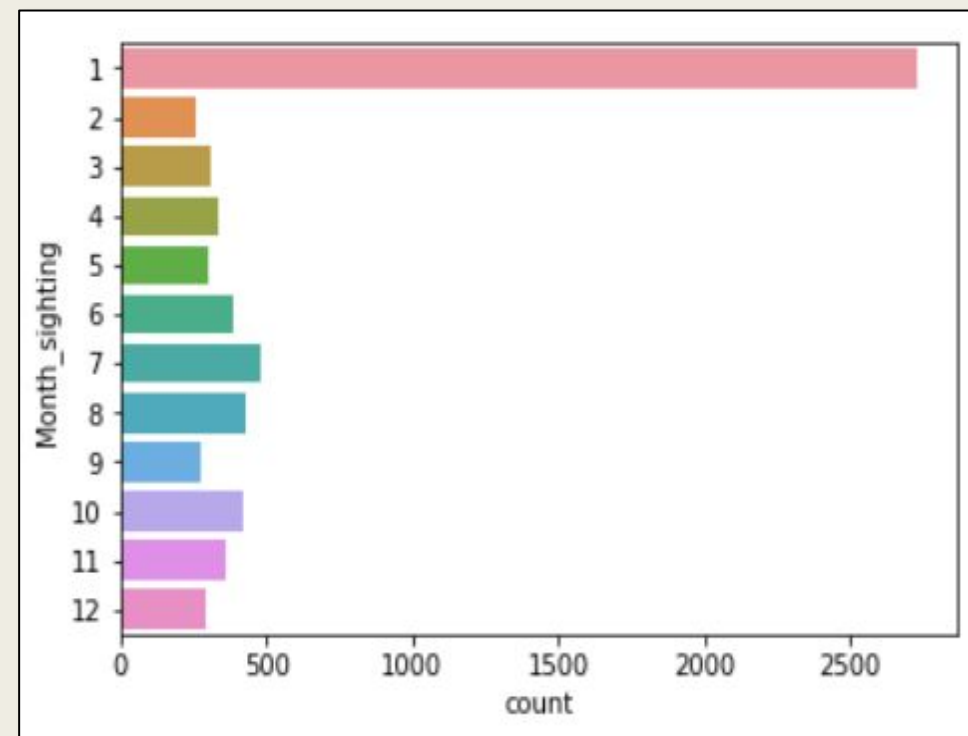


K Means Clustering

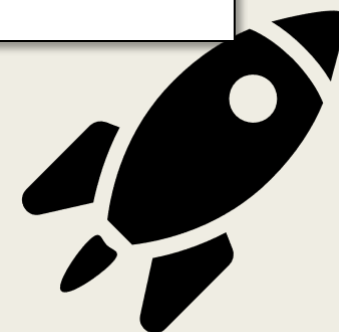
Shape of UFO



Month of Sighting



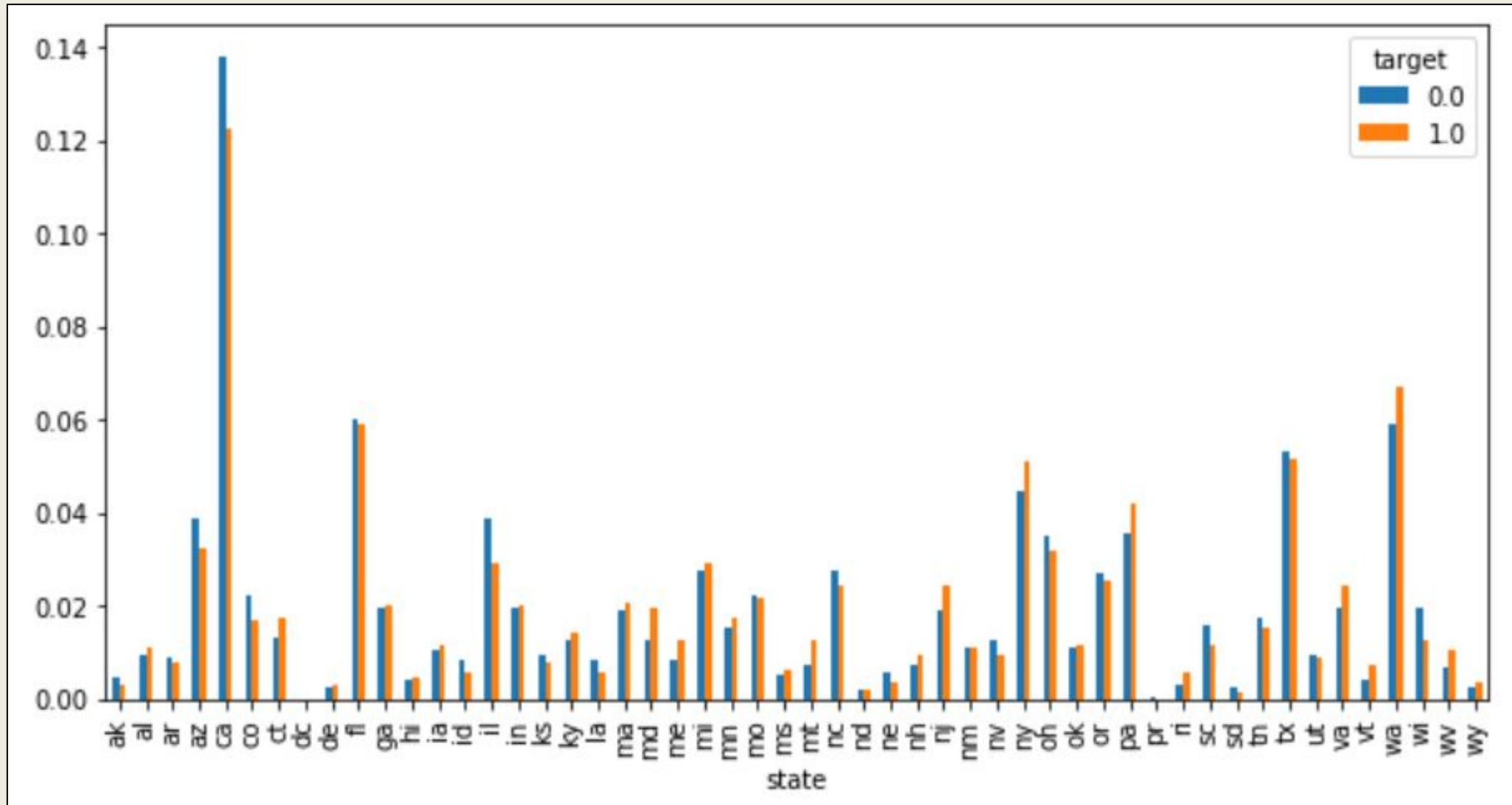
- Spherical UFOs
- January



RECOMMENDATIONS

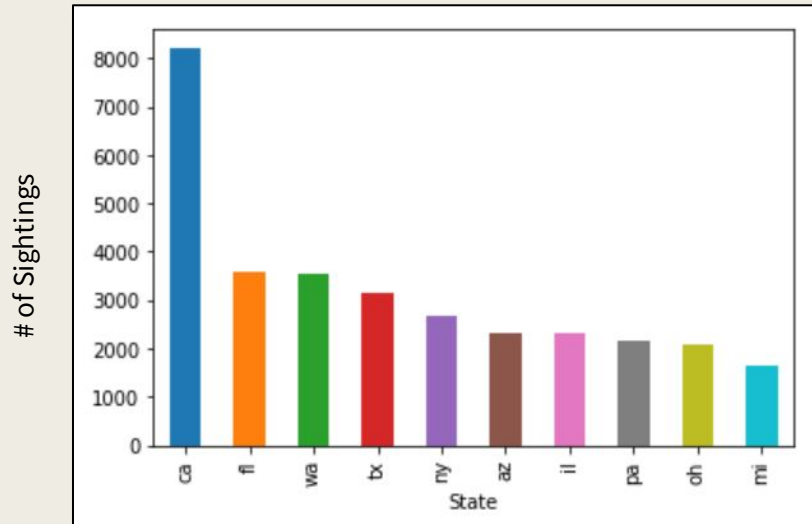


Number of Unexplainable and Explainable sightings by state

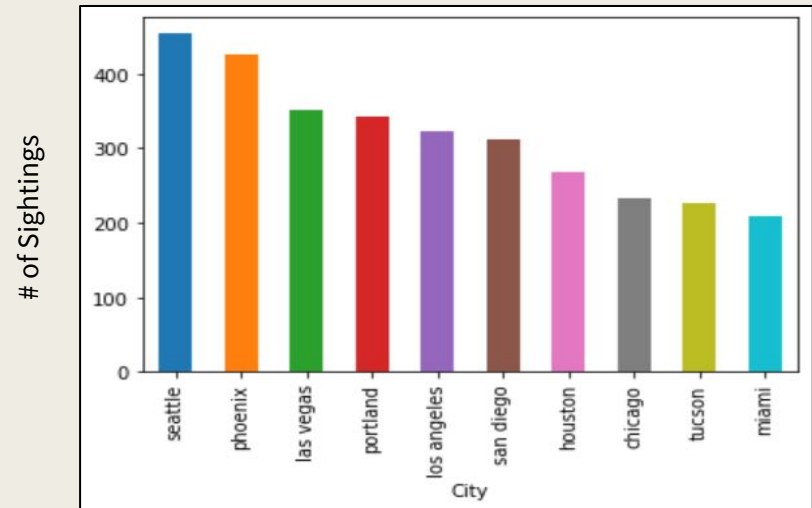


Where Will You See a UFO?

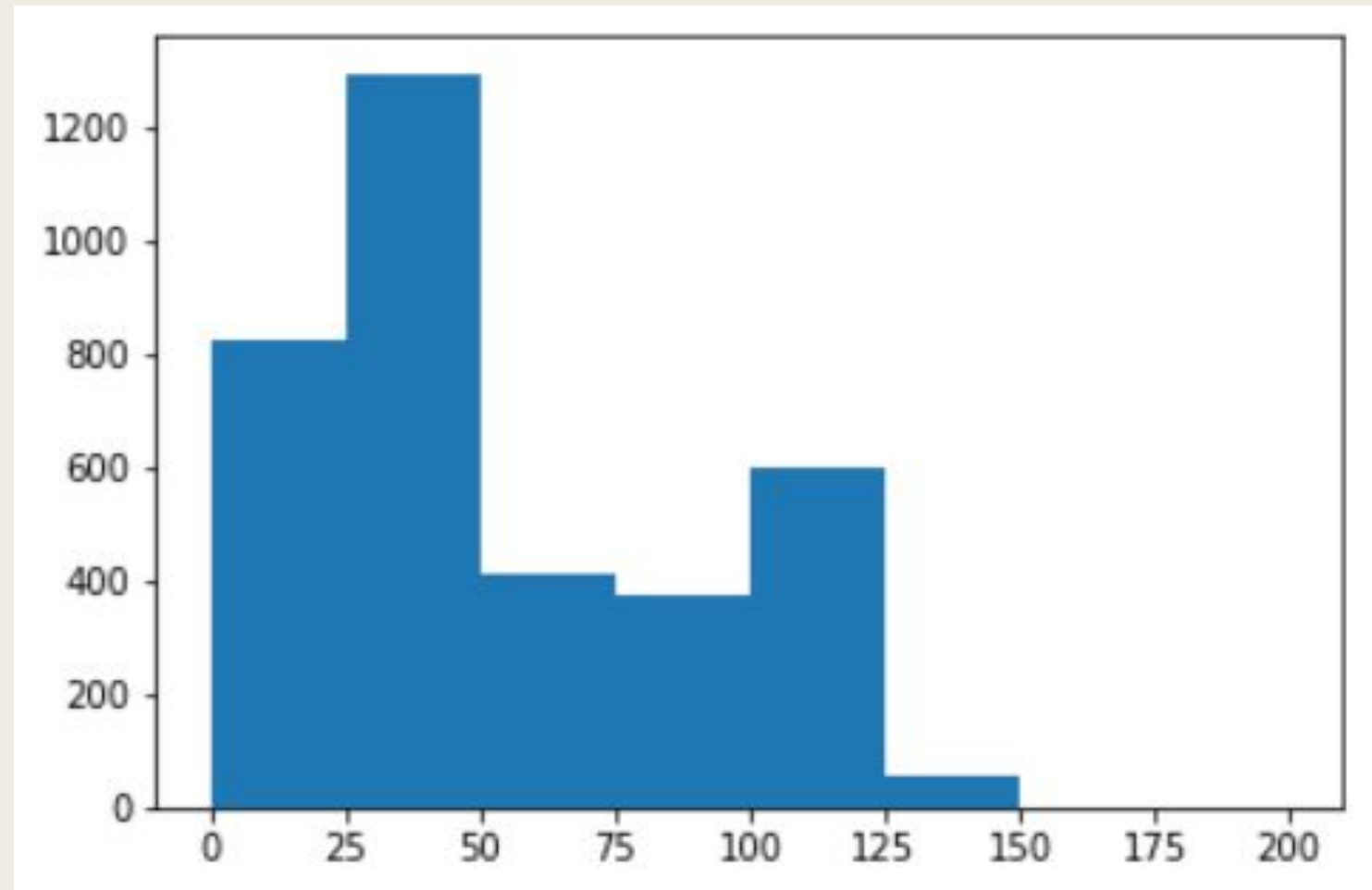
State



City



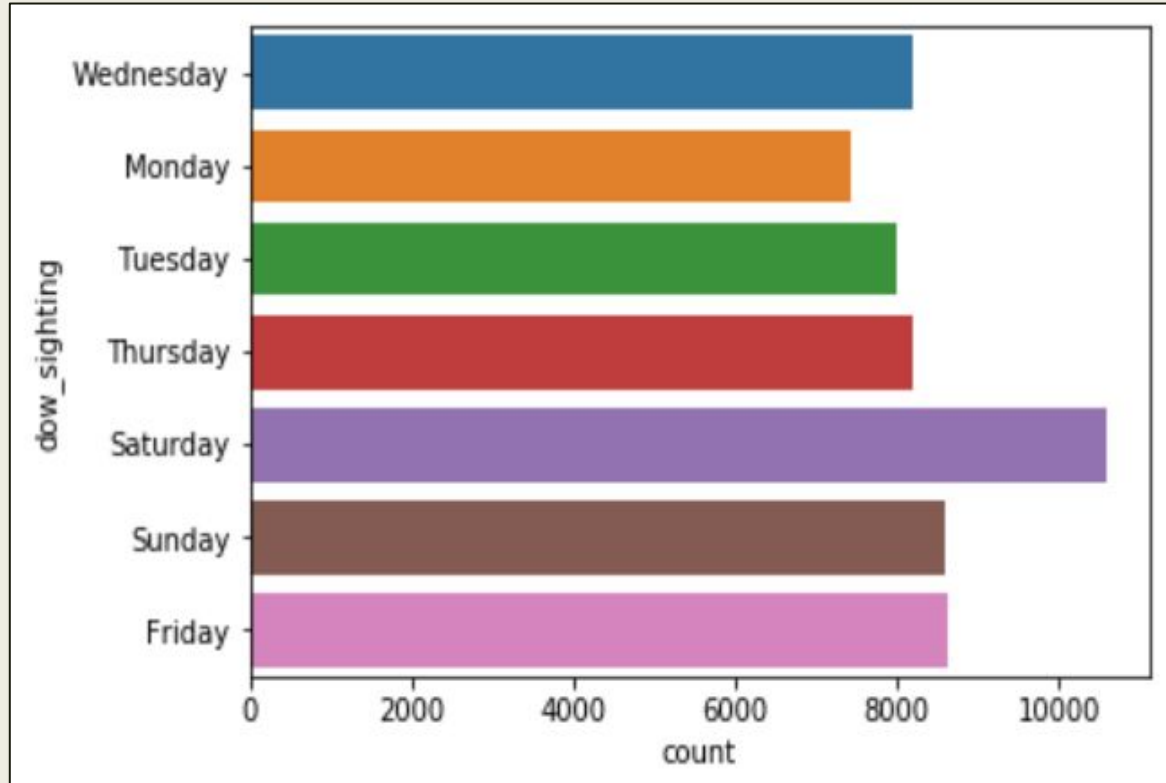
Proximity of Washington sightings to Air Force Bases



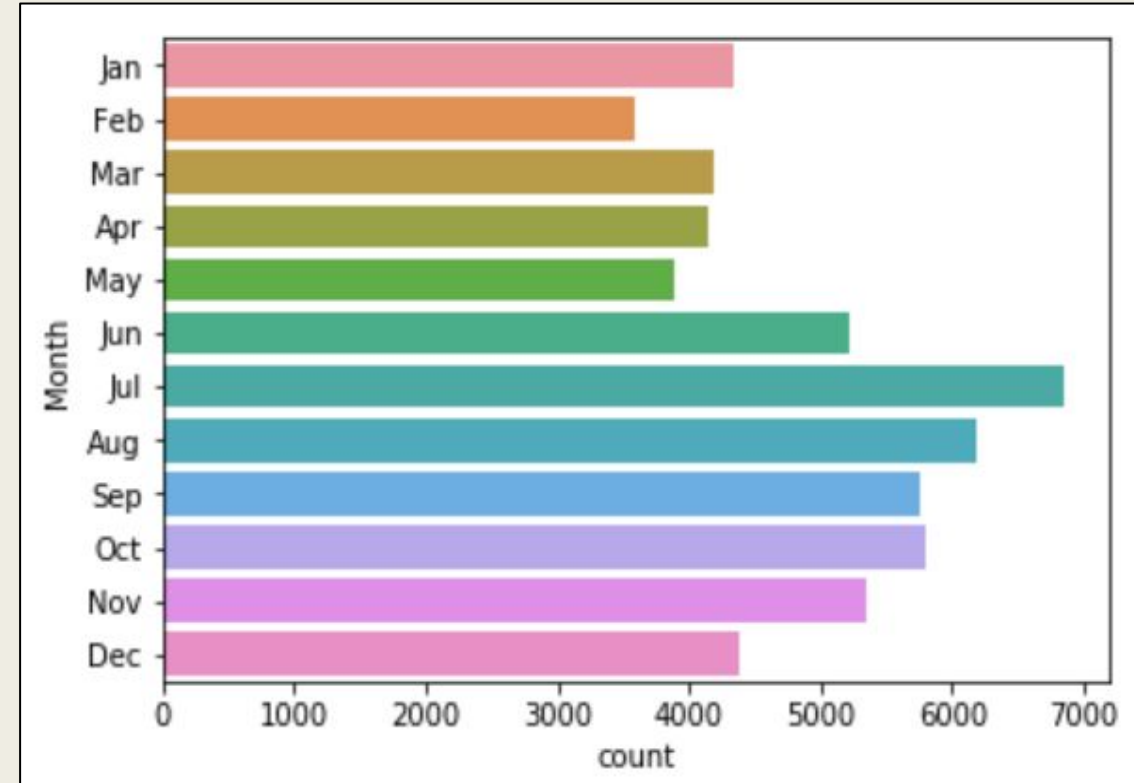
Miles

When Will You See a UFO?

Day of the Week



Month



How to See a “Real” UFO:

- Seattle, WA
- At night
- On a weekend
- July or August
- Triangle or cigar



Thank You!

QUESTIONS?

