UFO SPOTTING

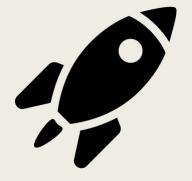
AUGUST 13, 2018

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

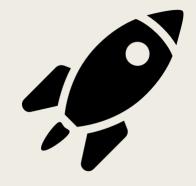
In the data set:

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

Added:

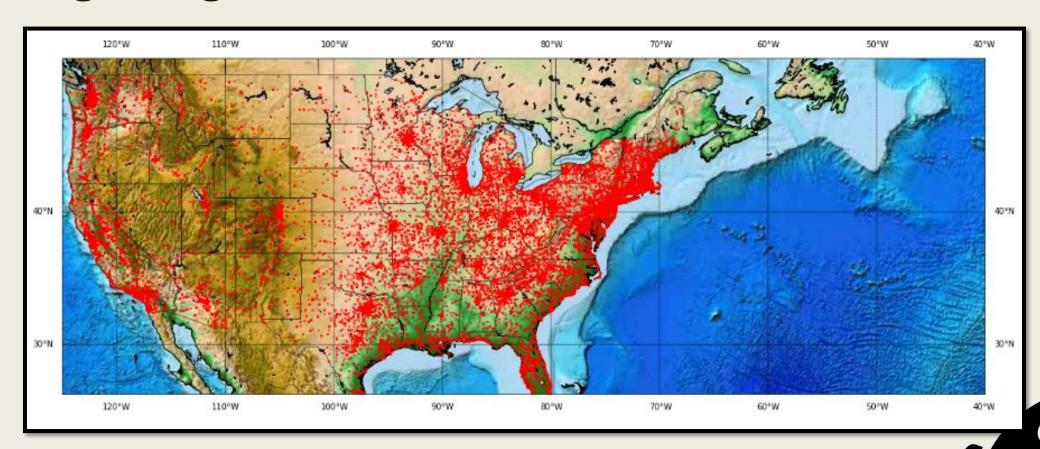
- Weather
- Day of the Week
- Moon Phase
- Proximity to US Airforce Base



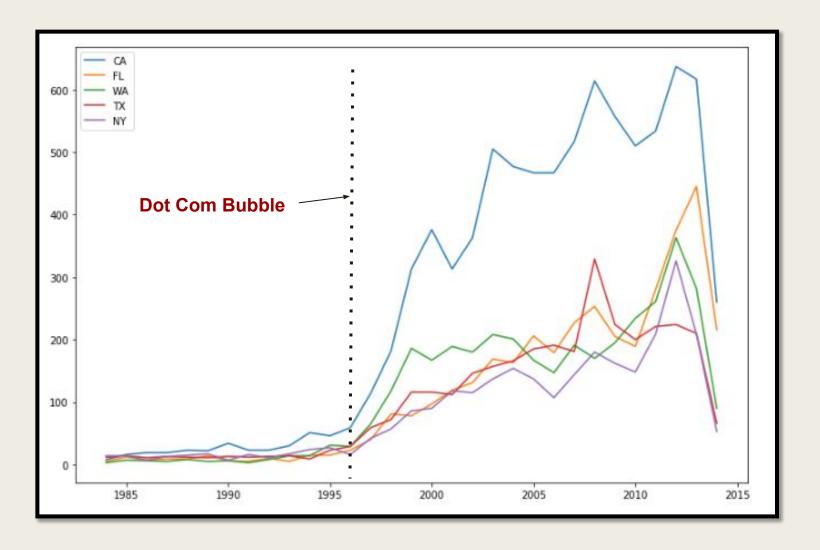


EXPLORATORY ANALYSIS

Sightings in the United States

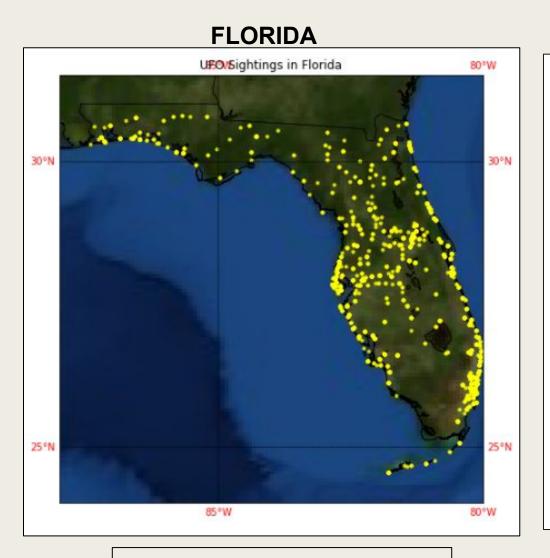


Sightings Over Time





Do you see any patterns?

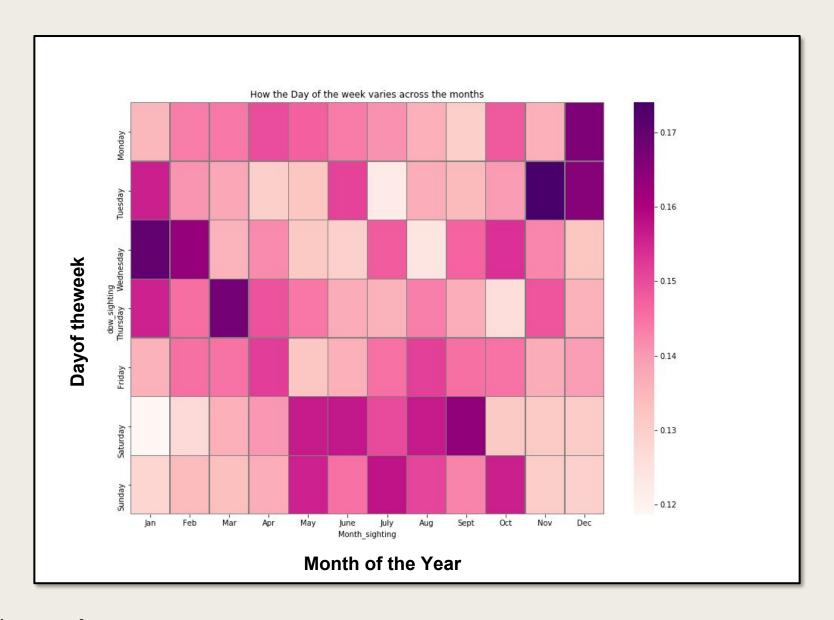


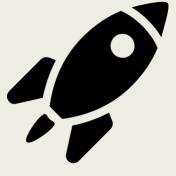
NEW YORK



I - 87 Corridor

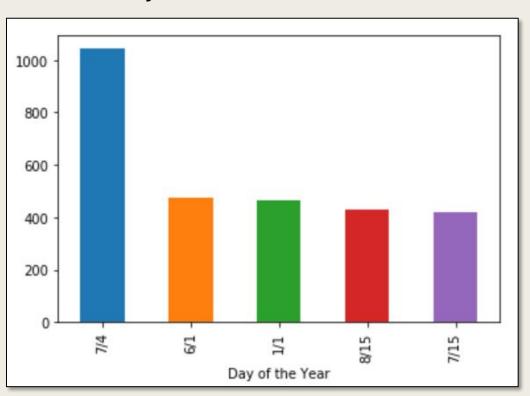
When Will I See a UFO?



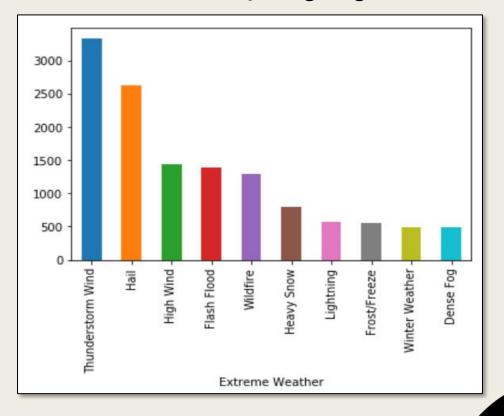


When Will I See a UFO?

Day of the Year

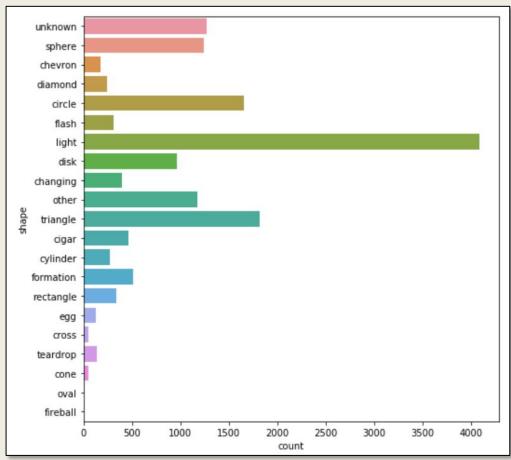


Severe Weather Day of Sighting

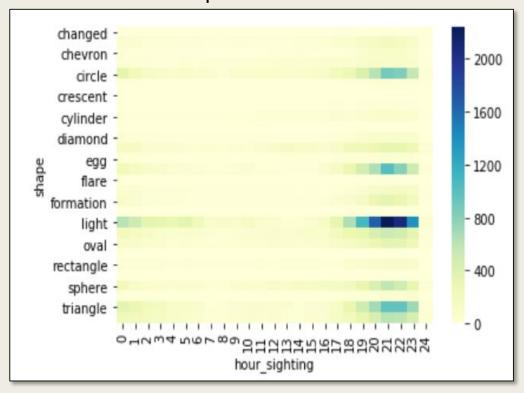


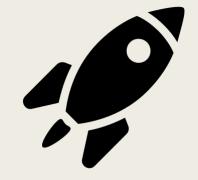
What Will I See?

Type of Shape

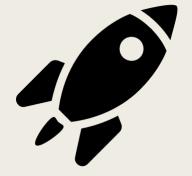


Shape vs. Hour





change	d - 0.00013	0	0	0	0	0	0	
changin	g - 0.03	0.026	0.024	0.026	0.022	0.027	0.025	
chevro	n - 0.011	0.012	0.016	0.012	0.014	0.011	0.013	- 0.20
ciga	ar - 0.027	0.024	0.024	0.024	0.025	0.02	0.022	0.20
circ	e - 0.086	0.089	0.093	0.093	0.1	0.11	0.1	
cor	ne - 0.0043	0.0045	0.0034	0.0033	0.003	0.0055	0.0036	
cresce	nt - 0	0	0	0	0	9.4e-05	0	
cros	s - 0.0028	0.0031	0.0029	0.0026	0.0028	0.0034	0.0036	- 0.16
cylinde	er - 0.013	0.017	0.015	0.02	0.015	0.016	0.016	0.10
delt	ta - 0	0.00012	0	0	0.00023	0.00019	0	
diamon	d - 0.013	0.015	0.017	0.018	0.015	0.012	0.015	
dis	ik - 0.054	0.054	0.056	0.056	0.054	0.048	0.051	
eg	ıg - 0.011	0.012	0.0083	0.01	0.0088	0.0066	0.0087	0.12
fireba	0.079	0.079	0.08	0.074	0.087	0.099	0.088	- 0.12
fireba flar	re - 0	0	0	0	0	0	0.00012	
flas	h - 0.019	0.016	0.018	0.019	0.017	0.016	0.018	
formatio	n - 0.028	0.032	0.033	0.03	0.032	0.036	0.034	
hexago	on - 0	0.00012	0	0	0	0	0	0.00
lig	nt - 0.22	0.21	0.22	0.22	0.21	0.22	0.22	- 0.08
othe	er - 0.075	0.074	0.075	0.073	0.069	0.059	0.07	
ov	al - 0.045	0.044	0.046	0.042	0.042	0.053	0.049	
pyram	id - 0	0	0	0	0	9.4e-05	0	
rectang	le - 0.018	0.018	0.016	0.015	0.017	0.016	0.018	0.04
rour	nd - 0	0	0	0.00012	0	9.4e-05	0	- 0.04
spher	e - 0.065	0.071	0.063	0.066	0.069	0.074	0.069	
teardro	p - 0.011	0.01	0.0089	0.0093	0.0092	0.01	0.009	
triangl	e - 0.11	0.11	0.11	0.11	0.11	0.09	0.096	
unknow	n - 0.075	0.074	0.074	0.072	0.072	0.068	0.075	0.00
	Monday	Tuesday	Wednesday	Thursday dow_sighting	Friday	Saturday	Sunday	- 0.00



INSIGHTS

Logistic Regression

Y = Explainable Events

created using scikit-learn

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

Precision-Reca

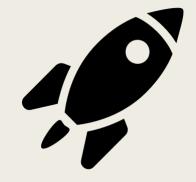
ll Table	Precision	Recall	Support
0.0 (Unexplained)	0.97	1.00	28,786
1.0 (Explained)	0.72	0.11	1,055
Avg/Total	0.96	0.97	29,841

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

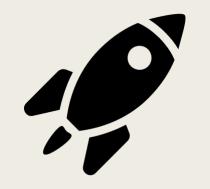
Test Data Confusion Matrix

Prediction Actual	0	1
0	22974	55
1	756	88

Accuracy: 0.9922

Accuracy: 0.9660

	precision	recall	f1-score	support
0.0	0.97 0.62	0.10	0.98 0.18	23029 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

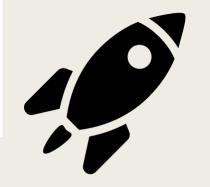
Test Data Confusion Matrix

Prediction Actual	0	1
0	18819	4210
1	472	372

Accuracy: 0.9199

Accuracy: 0.8039

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

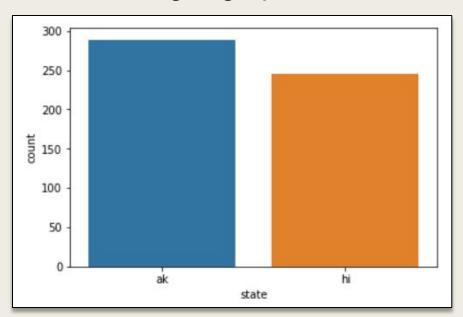


Random Forest: Feature Importance

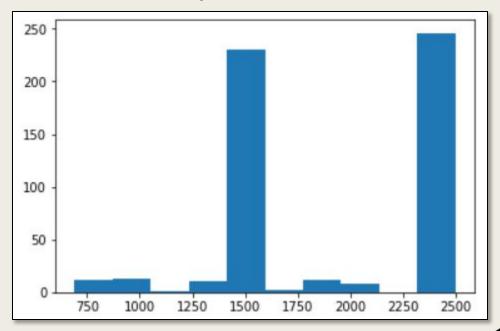
Features	Importance
Time	0.010399
Proximity to USAFB	0.004348
Date	0.002605
Days Lapsed after Sighting	0.001957
Duration (seconds)	0.001336
Day of the Week	0.001215
Shape	0.001099
State	0.000900
Extreme Weather	0.000421
Moon Phase	0.000001

K Means Clustering

of Sightings by State



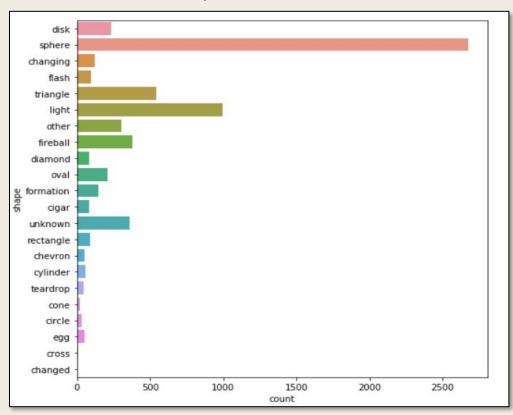
Proximity to Air Force Base



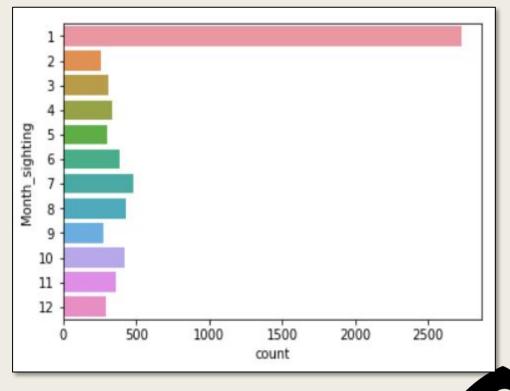
- Alaska and Hawaii
- >=1500 miles from an airforce base

K Means Clustering

Shape of UFO



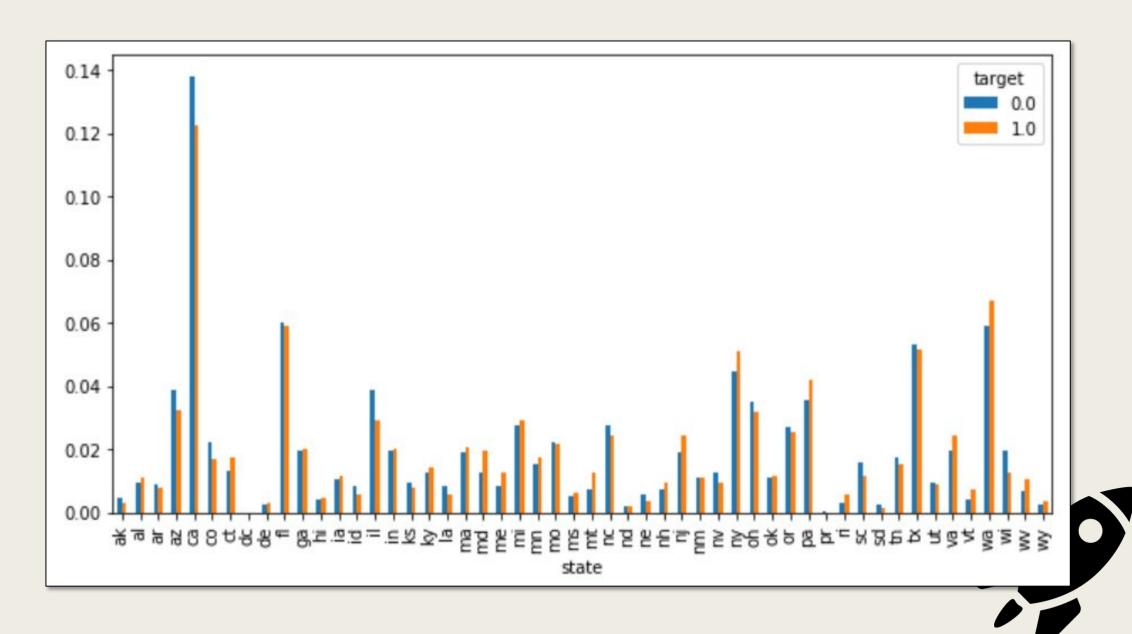
Month of Sighting



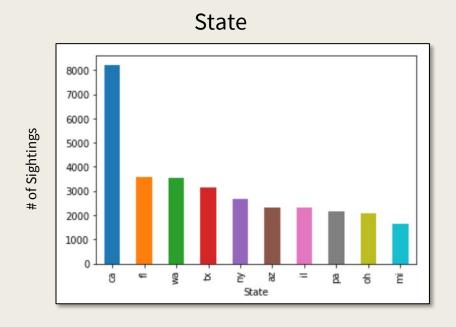
- Spherical UFOs
- January

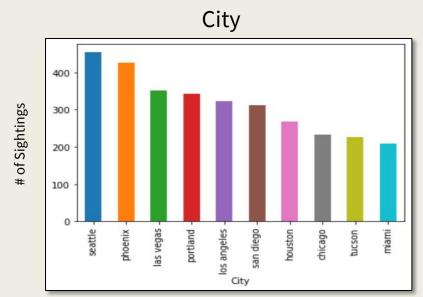
RECOMMENDATIONS

Number of Unexplainable and Explainable sightings by state



Where Will You See a UFO?

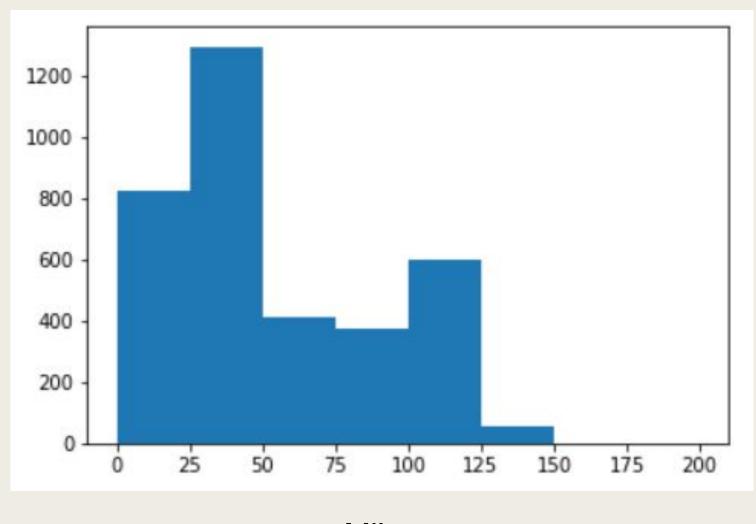






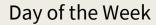


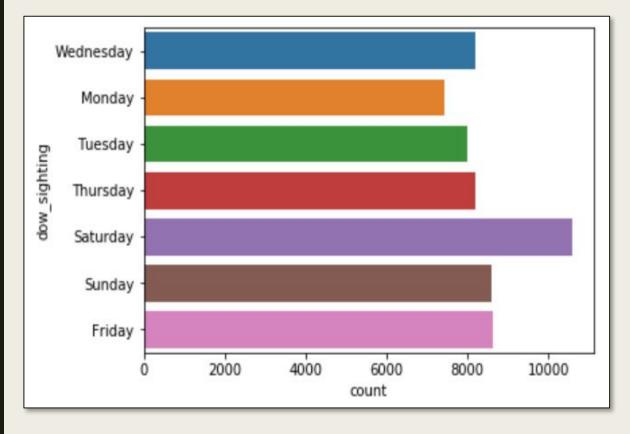
Proximity of Washington sightings to Air Force Bases



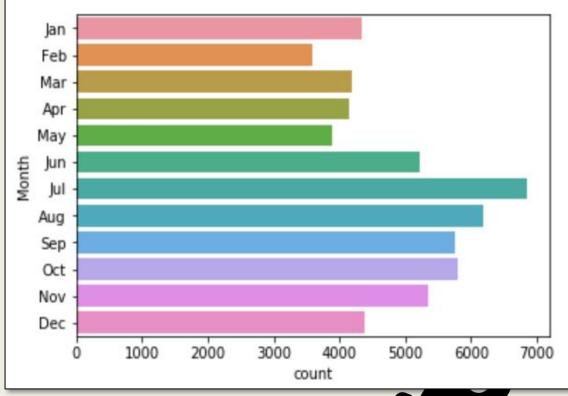
Miles

When Will You See a UFO?





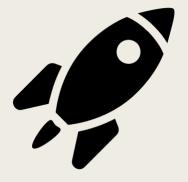
Month





How to See a "Real" UFO:

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



Thank You!

QUESTIONS?