Natural Language Processing and Sentiment Analysis of the Interstellar Travel Industry

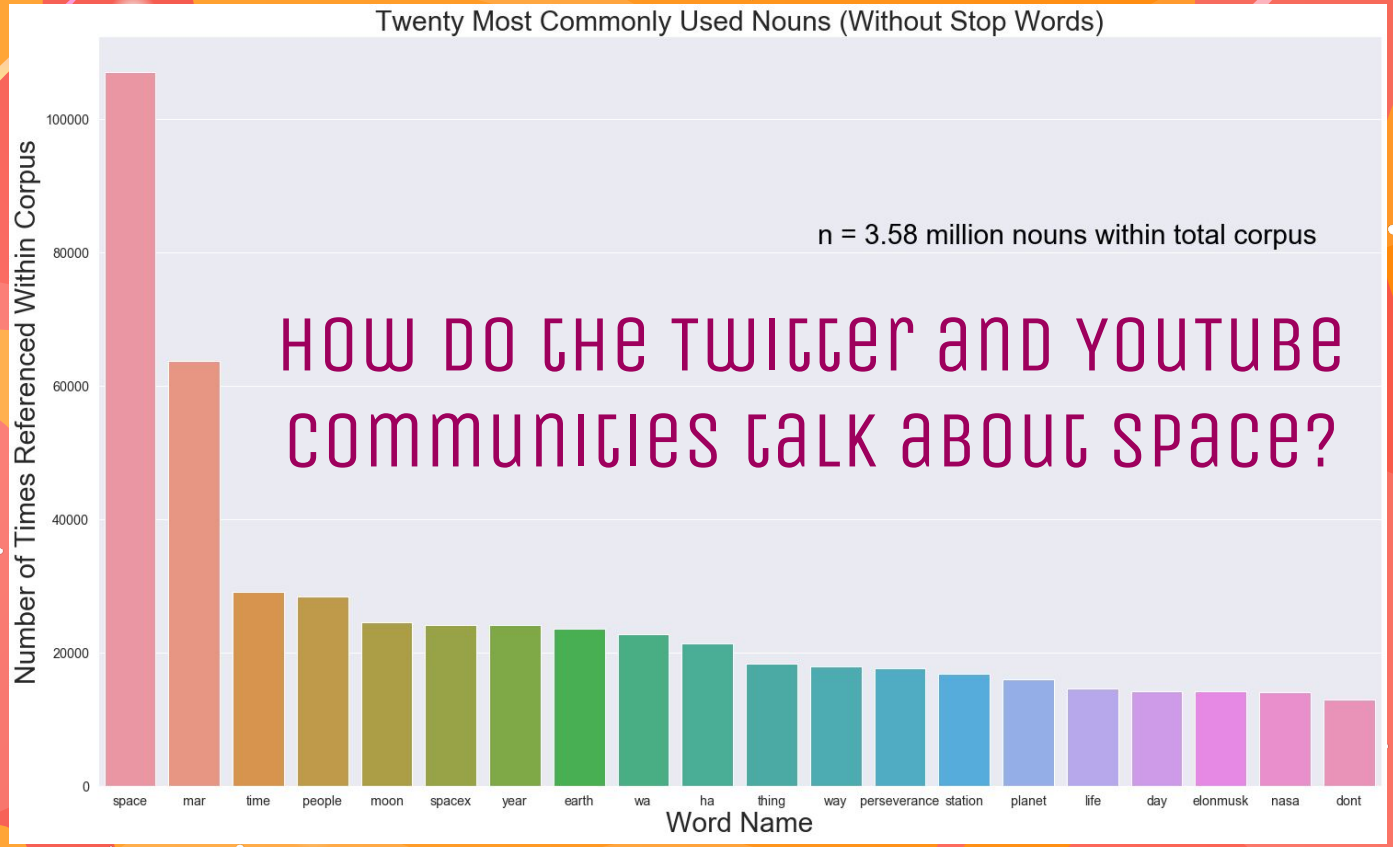
Fennec Nightingale + Matthew Lipman

April 23, 2021

BUSINESS PROBLEM

"It's a fixer upper of a planet, but we can make it work."

—ELON MUSK



How do the Twitter and YouTube communities talk about space?

METHODOLOGY

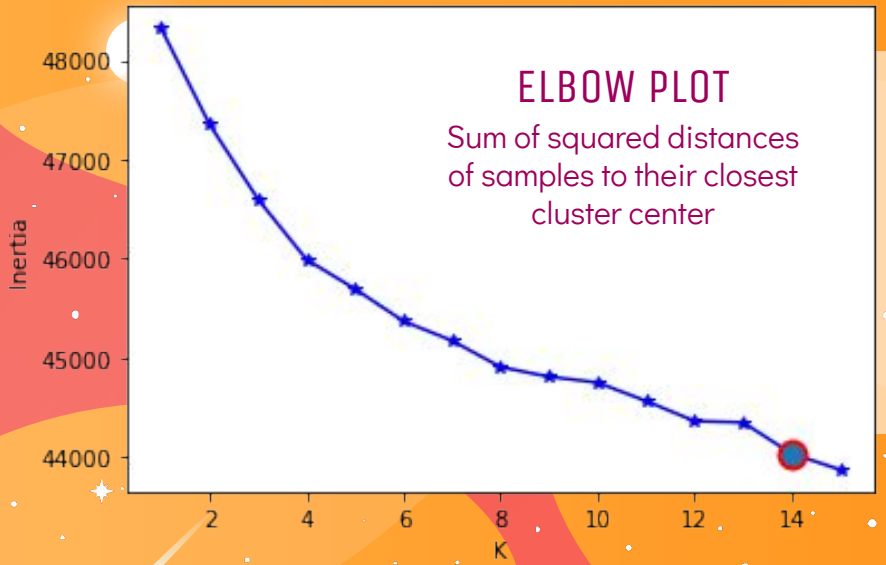
01 ___ Data collection

02 ___ Preprocessing

03 ___ Clusterization

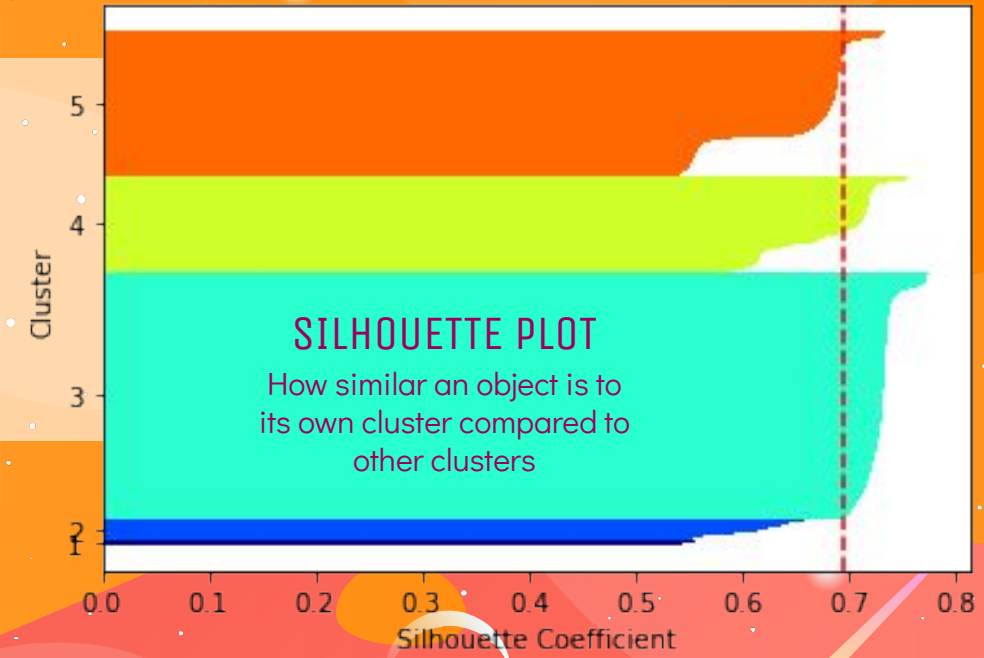
04 ___ Modeling

The Results: Finding the right clusters



ELBOW PLOT

Sum of squared distances of samples to their closest cluster center

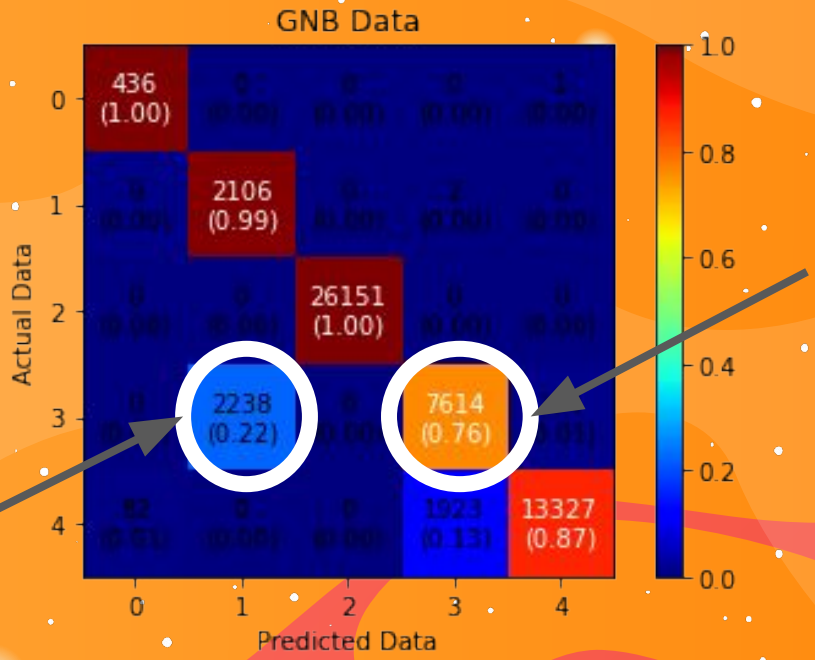


SILHOUETTE PLOT

How similar an object is to its own cluster compared to other clusters

The Results: Modeling

CLUSTERS 0,1,2 ACCURATE!

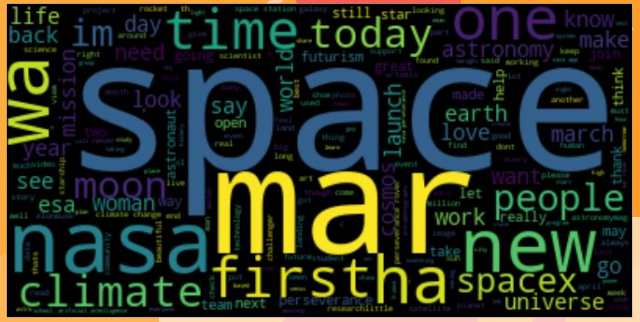


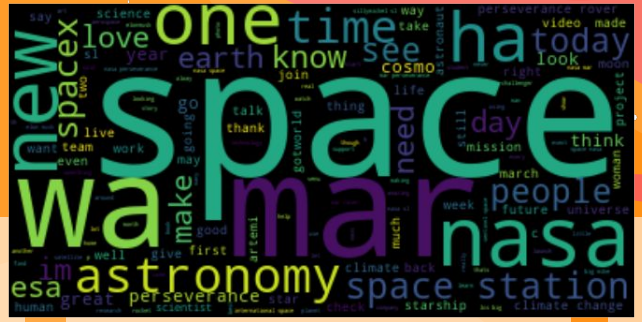
76% CORRECTLY CLASSIFIED

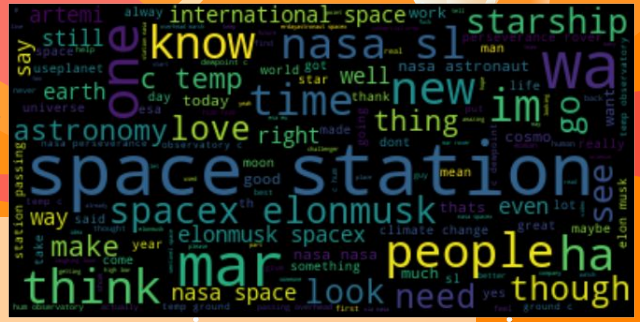
FALSE NEGATIVES / FALSE POSITIVES

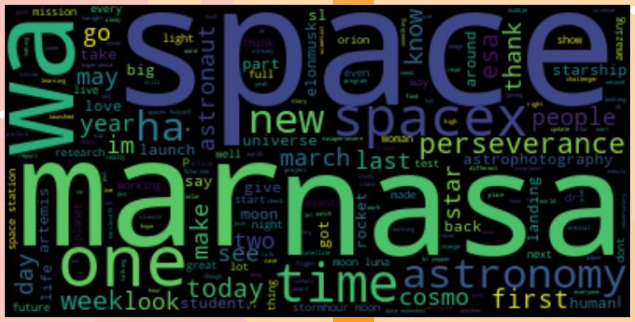
Accuracy = (False Positives + False Negatives)/Total*100% = 92%

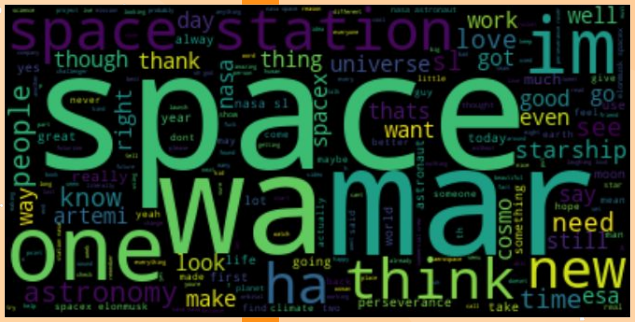
our reccomendations











LEVERAGE SPACE

Space is commonly found within the discourse, but in congruence with other words will add significance

PARTNER WITH THE BEST

Institutions like NASA and SpaeX not only have scientific headway, but they have built reputations.

EVOKE EMOTION

The power building an emotional response adds tremendous value to your marketing capabilities

Future work

Incorporate more platforms

- Reddit
- Facebook
- Instagram

Time series analysis

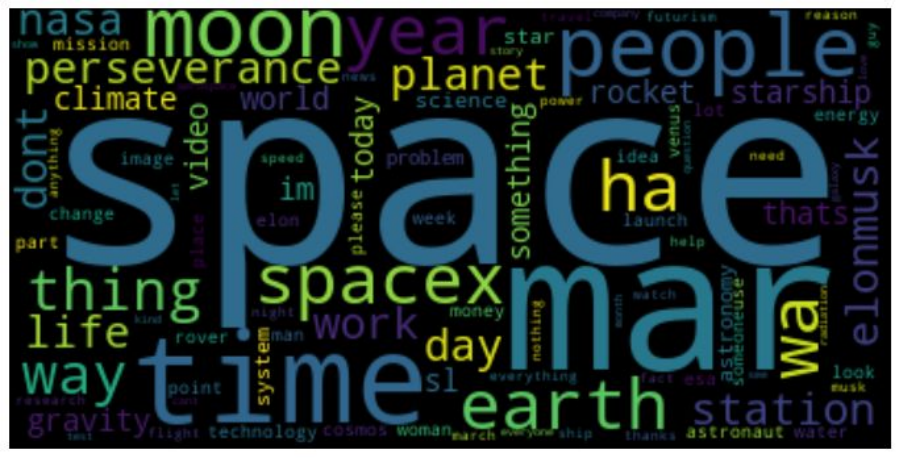
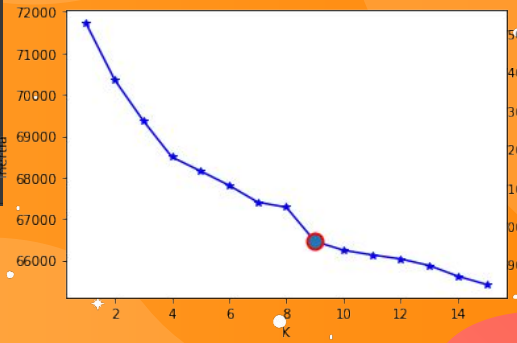
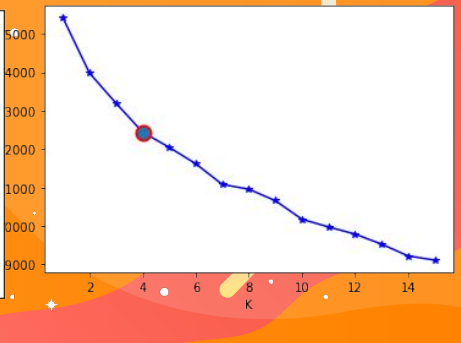
- Looking at change of space sentiment after major events
- Long term study of how specific words have changed in sentiment

Thank you!

Please do not hesitate to reach out to request any additional information.

Fen: fenneccharcles@gmail.com
Matt: matthewrlipman@gmail.com

www.github.com/Fennecnightingale/space-hospitality

APPENDIX

```
Training Accuracy Score : 92.32000000000001
Training Precision Score : [0.99390345 0.80843589 1.         0.85376908 0.98553626]
Training Recall Score : [0.99843064 0.99823926 1.         0.75230938 0.8667815 ]
Training F1 Score : [0.9961619  0.89336751 1.         0.79983451 0.92235211]

Training residual count:
0    361764
2     18617
1         1

Testing Accuracy Score : 91.91
Testing Precision Score : [0.82732448 0.48480663 1.         0.79819688 0.99144473]
Testing Recall Score : [0.99771167 0.99480397 1.         0.76399759 0.86922776]
Testing F1 Score : [0.90456432 0.65191147 1.         0.78072289 0.92632237]

Testing residual count:
0    49634
2     2240
1

Testing R2 Score: 0.7479017913768848
Testing R2 Score: 0.8799230883065321
Cross validated model accuracy:
Testing:
0.96 with a standard deviation of 0.00
Training:
0.92 with a standard deviation of 0.00

Testing R2 Score: 0.7479017913768848
Testing R2 Score: 0.8799230883065321
```