



HIKING with iNATURALIST

A BRAINSTATION CAPSTONE
BY LISA MCCOOL-GRIME

Crested Coralroot Orchid
© McCool-Grime

INTRO TO iNATURALIST

iNaturalist is an online social network of people sharing biodiversity information to help each other learn about nature.

The iNaturalist mission is to connect people to nature and advance biodiversity science and conservation.

From the iNaturalist export page, I was able to download and compile a dataset of 1050151 data points from 300 different iNaturalist users for analysis.

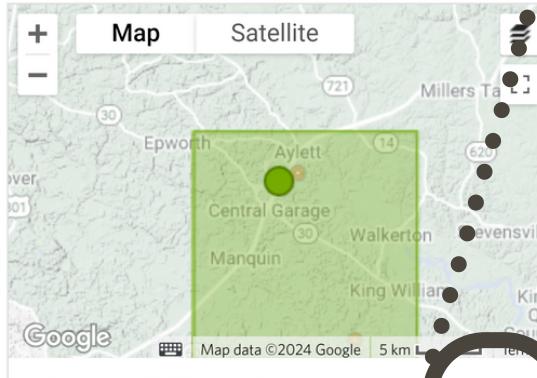
Crested Coralroot Orchid (*Bletia spicata*) VU Research Grade



annegp

Observed: July 2022 Submitted: July 2022

Map Satellite



Details ▾

abelkinser and annegp favor this observation

Notes

Group of 5 flower spikes (tallest is about 2 feet)
Flora of Virginia lists this as *Hexalectris spicata*
In King William County, VA

Activity

annegp suggested an ID Improving Jul '22

Crested Coralroot Orchid *Bletia spicata* Compare Agree

catullus suggested an ID Jul '22

Crested Coralroot Orchid *Bletia spicata* Select

Lat: 37.764612 Lon: -77.123083 Accuracy: 28.36km Geoprivacy: Obscured

Encompassing Places

Standard: North America (Continent), United States (Country), Virginia, US (State), King William County, US, VA (County)

Community Curated: Rolling Coastal Plain (Region), Atlantic Southern Loam Hills, US (Region), Mid-Atlantic States, US (Region), Chesapeake Bay watershed (Drainage), BIOL272 project area (Unknown), United States Atlantic Coastal Plain (Land feature), US Eastern States, US (Colloquial), Eastern United States (Region), Newcomb's Wildflower Guide Range (Region), Southeast US conifer savannas, US (Zone)

More

Why the Coordinates Are Obscured

Geoprivacy is obscured: Observer has chosen to obscure the coordinates.

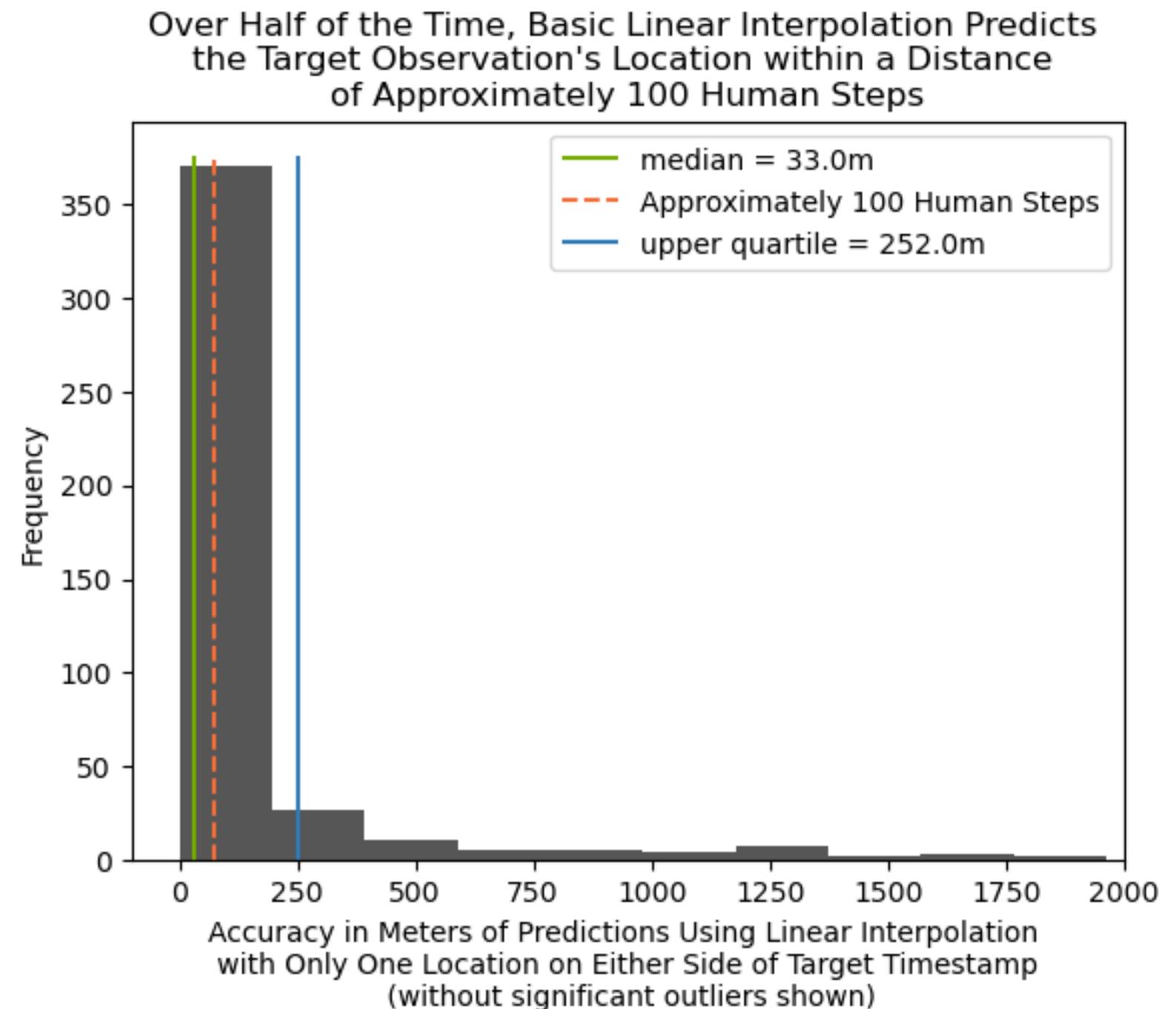
Taxon is threatened, coordinates obscured by default: One of the taxa suggested in the identifications, or one of the taxa that contain any of these taxa, is known to be rare and/or threatened, so the location of this observation has been obscured.

THE OLD PROBLEM

What can data analytics and data science tell us about best practices for better obscuring the location of vulnerable species?

Recommendations:

- 1) Obscure the “observed_at_time” online and in the dataset
- 2) Obscure the “created_at” time online and in the dataset
- 3) Research and implement a way to assign id numbers to observations that are not based on when the observation is created



THE NEW PROBLEM

Can a model learn to “see” an iNaturalist user’s hikes?

Potential Impact:

1) Individual User Engagement--

example: iNaturalist could include a slidedeck or interactive map of a user’s “longest hike” in the personal “Year in Review” materials

2) Regional User Engagement/Expansion--

example: iNaturalist could partner with a city, identify common walking trails from their data and create a “scavenger hunt” of observations found on the trails

3) Scientific --

example: identifying the number of “hikes” a user takes in relationship to the observations they contribute that aren’t clustered into hikes could be used as a potential measure for user behavior



Pale-Spiked Lobelia
© McCool-Grime

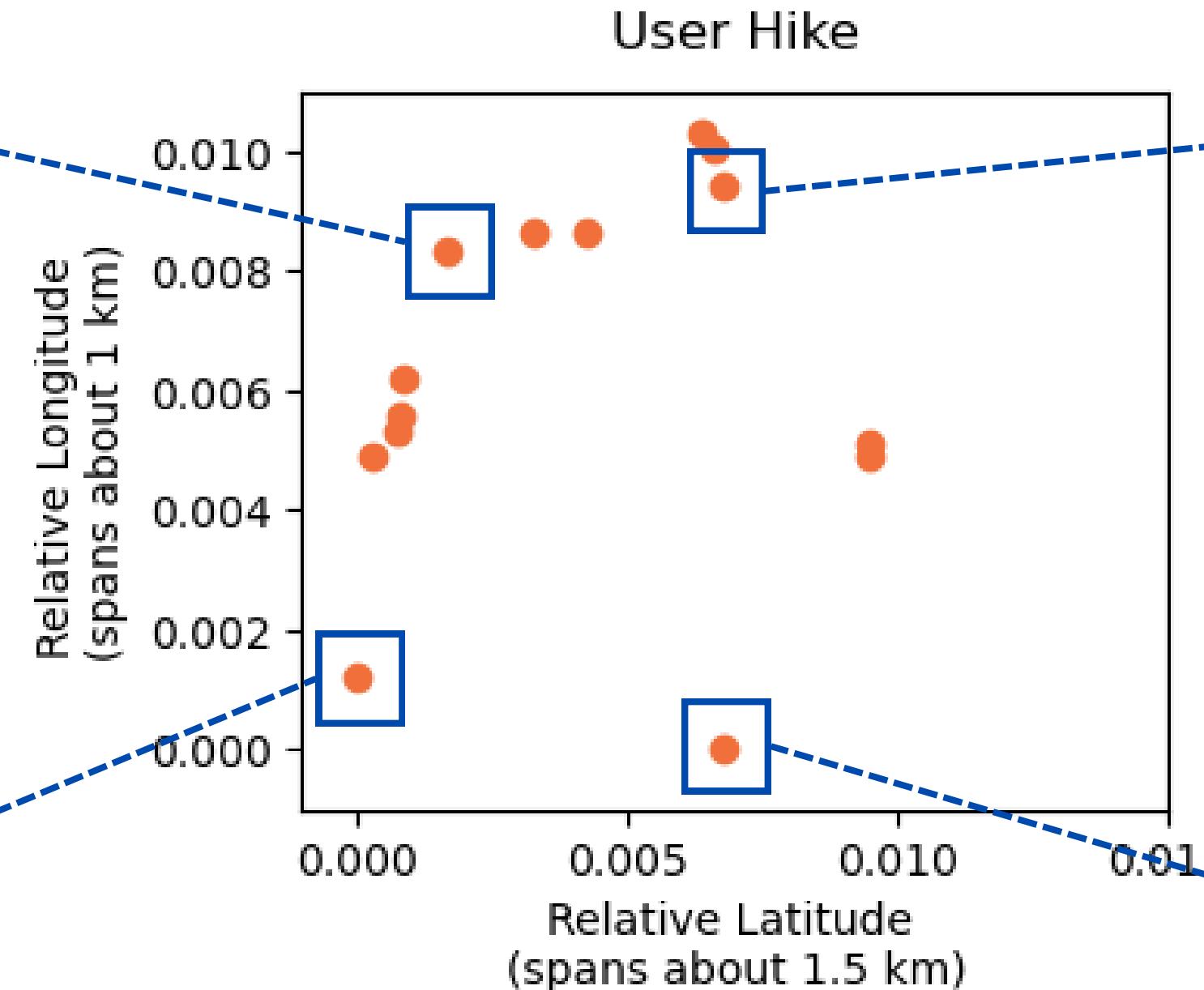
DEMONSTRATION



Eastern Poison Ivy



Pale Jewelweed



Monarda

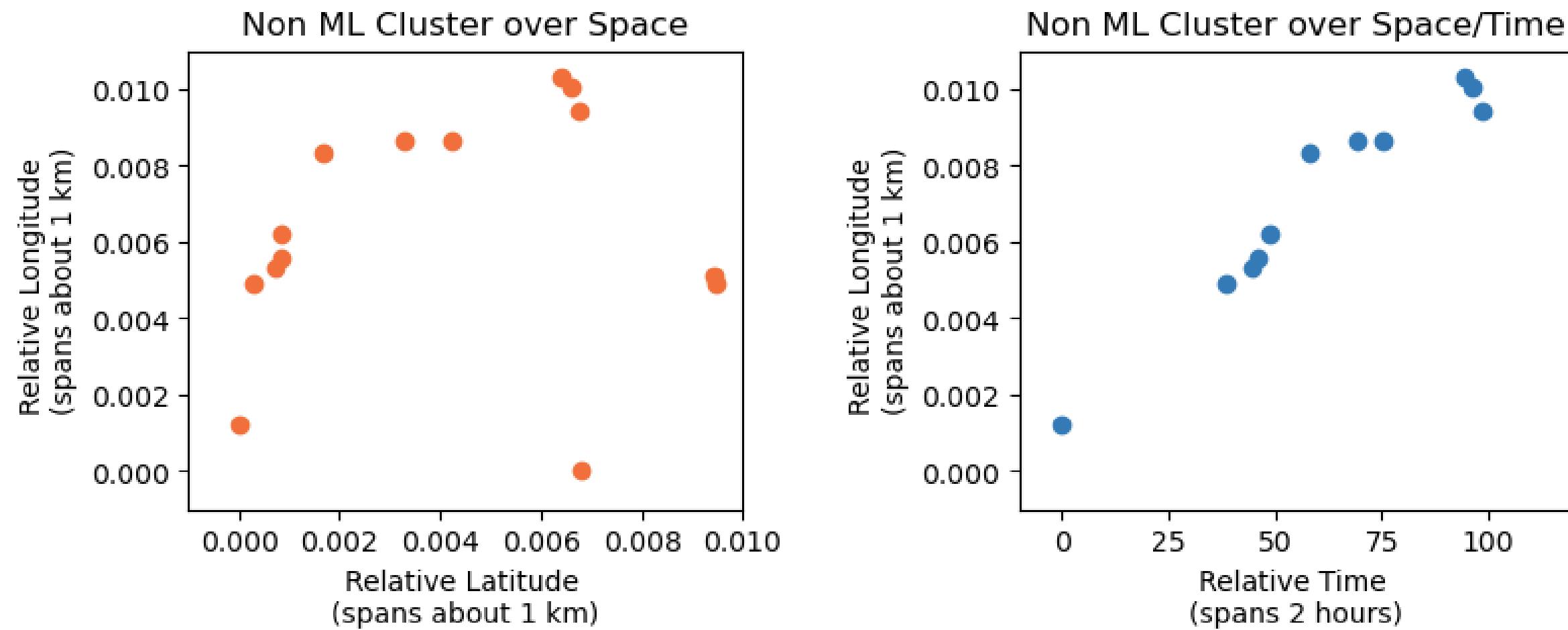


Jelly False Coral Fungus

NON-MACHINE LEARNING MODEL

Clustering By Dates

This straightforward method returns a good picture of likely hikes.

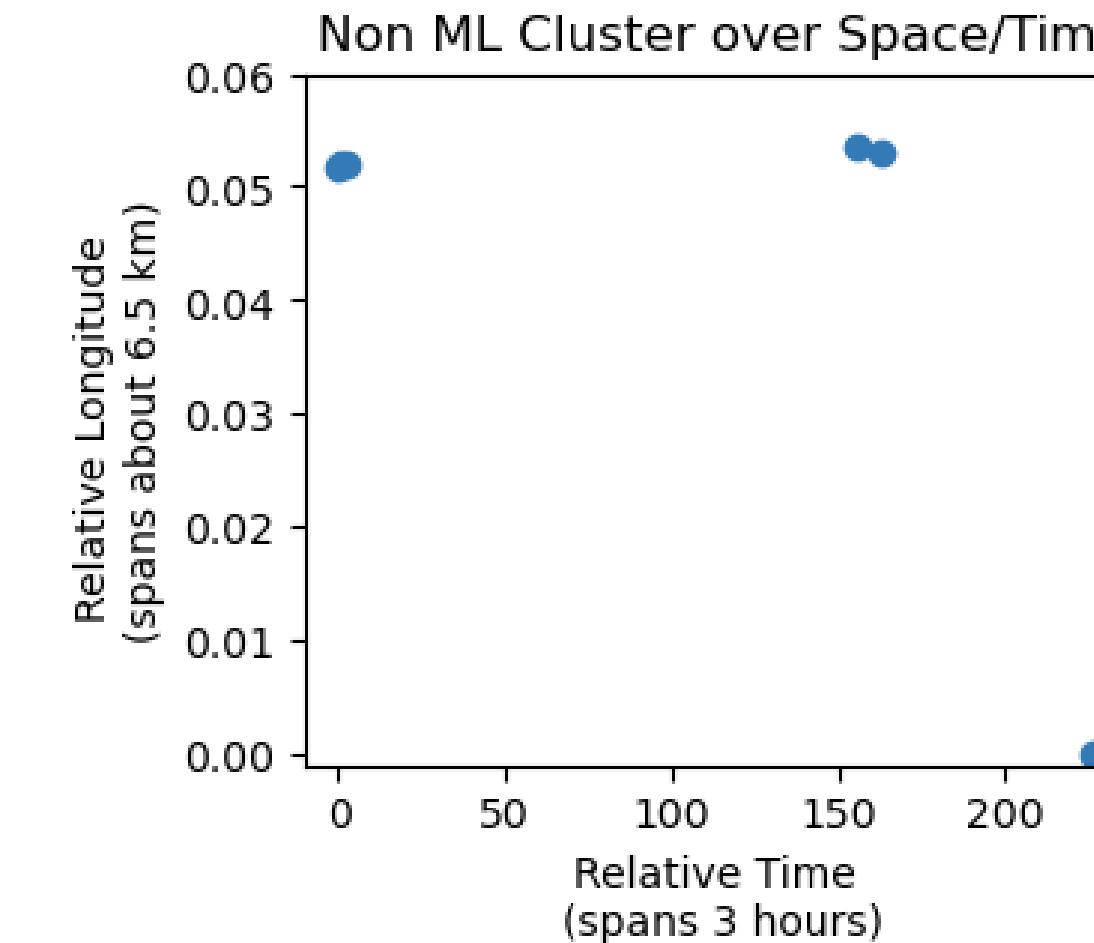
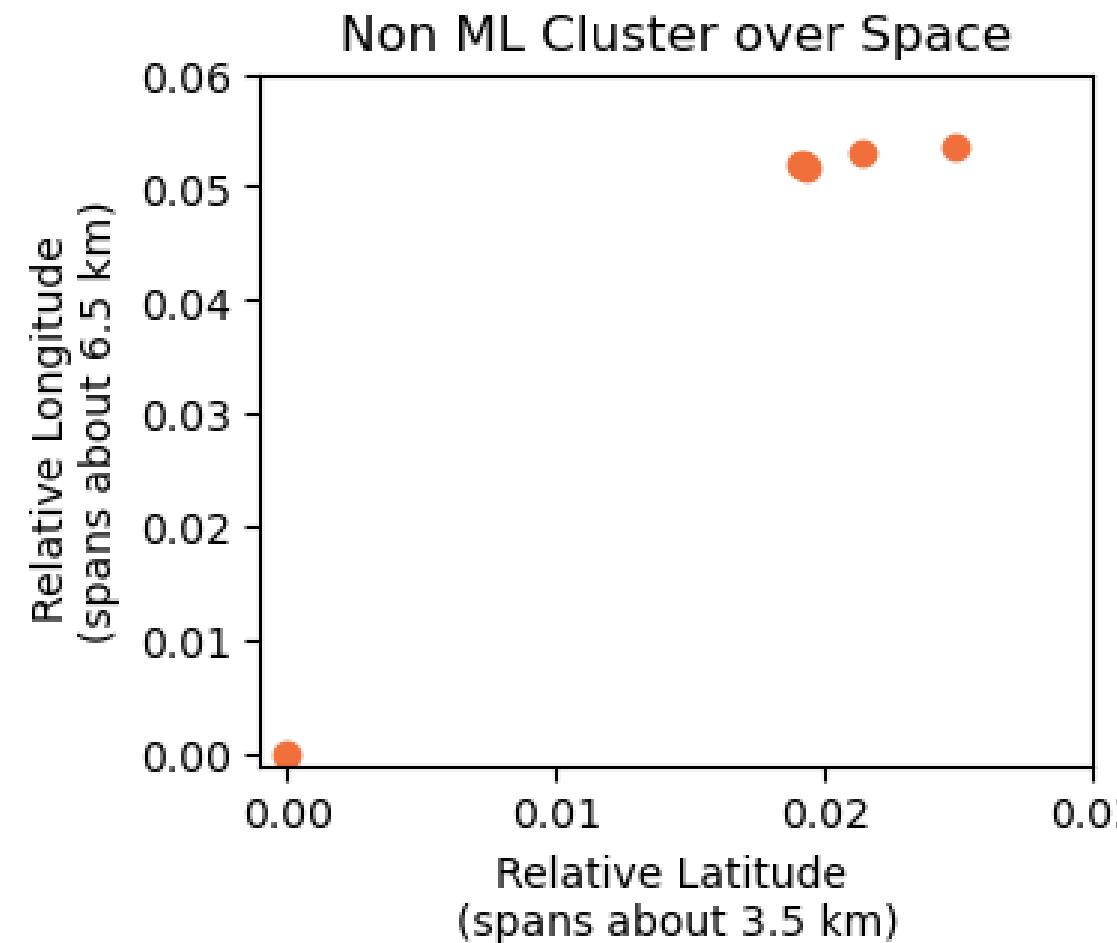


14 data points pictured for this cluster

NON-MACHINE LEARNING MODEL

Clustering By Dates

However, it is too liberal and will include “clusters” that are clearly not “hikes”.

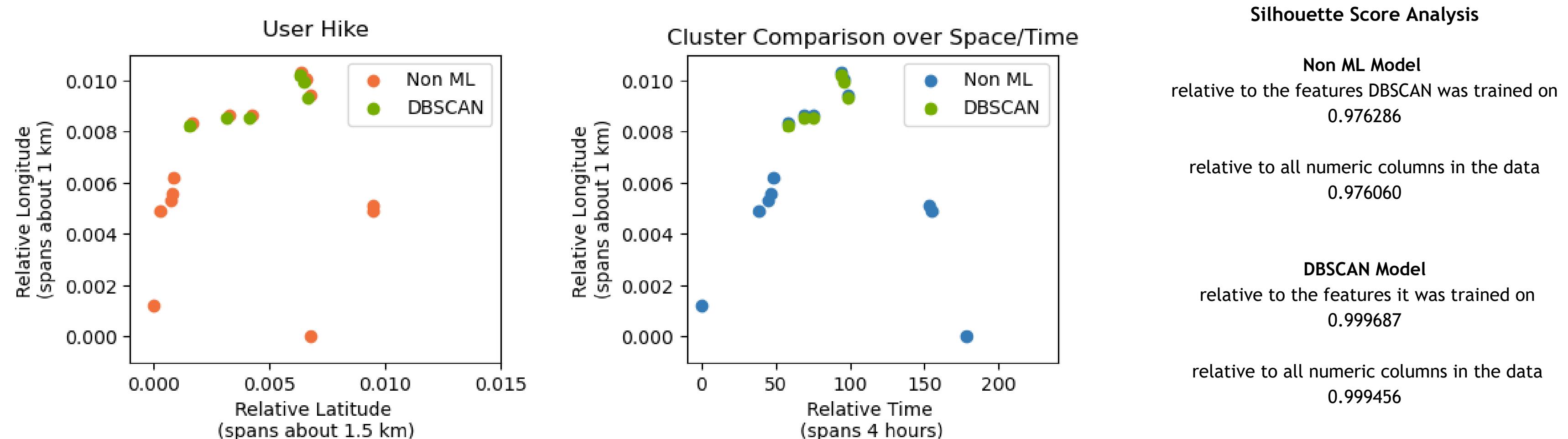


10 data points pictured for this cluster

MACHINE LEARNING MODEL

DBSCAN

However, the DBSCAN model with the “best” performance scores can be too conservative.



Of the 14 points identified as a cluster with the group-by-date method,
only 6 were chosen as a “hike” by the DBSCAN model

NEXT STEPS

- Goldilocks: What metrics are best for creating a model that is not too liberal and not too conservative when identifying a hike?
- Scaling up: What is the range of performances for these models on other users?



Yellow Lady's Slipper
© McCool-Grime