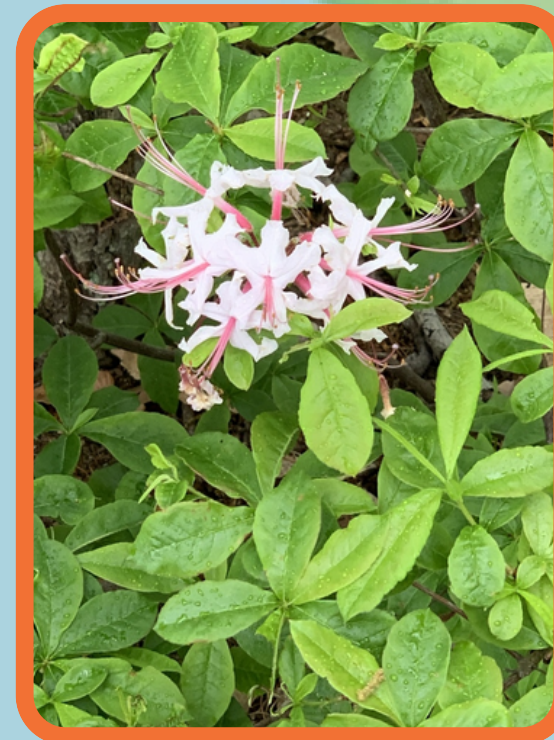


HIKING WITH INATURALIST

*Determining User Behavior
with Crowd-Sourced Data
for Environmental Science*

LISA MCCOOL-GRIME




iNATURALIST DATASET

SOCIAL NETWORK SHARING
BIODIVERSITY INFORMATION

OVER 100 MILLION NATURAL
OBSERVATIONS SINCE 2015

MY SUBSET: 1,050,151 U.S. OBSERVATIONS
FROM 300 USERS

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



Observed: July 2022 Submitted: July 2022

Map Satellite

Virginia, US (Obscured)

abelkinser and annegp faved 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 Compare Agree

catullus suggested an ID Jul '22

Crested Coralroot Orchid

Community Taxon

Crested Coralroot Orchid (*Bletia spicata*) VU

Cumulative IDs: 4 of 4

0 2/3rds 4

Agree Compare About

Annotations

Attribute	Value	Agree	Disagree
Plant Phenology	Select		
Sex	Select		

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

Encompassing Places

Standard:	Community Curated:
North America (Continent)	Rolling Coastal Plain (Region)
United States (Country)	Atlantic Southern Loam Hills, US (Region)
Virginia, US (State)	Mid-Atlantic States, US (Region)
King William County, US, VA (County)	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

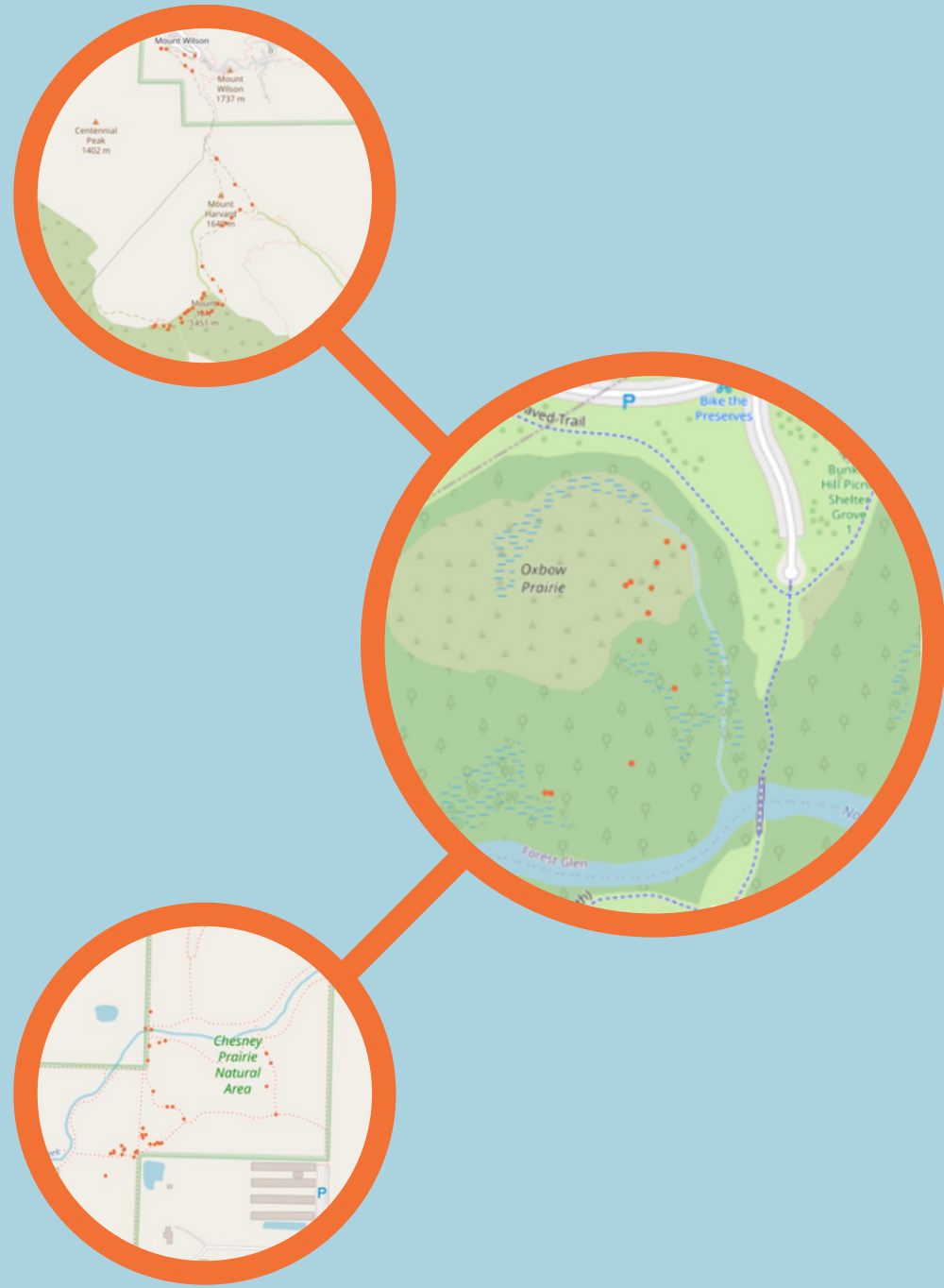
Geoprivacy is obscured: Observer has chosen to obscure the coordinates.

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.



iNATURALIST MISSION:
to connect people to nature and
advance biodiversity science and conservation

PROBLEM: HOW DOES USER BEHAVIOR OVERLAP WITH HIKING BEHAVIOR?



Can iNaturalist data be clustered over time and space to identify “hike-like behavior”?

What metrics can describe users who commonly use nature trails?
What metrics describe other ways of engaging with iNaturalist?

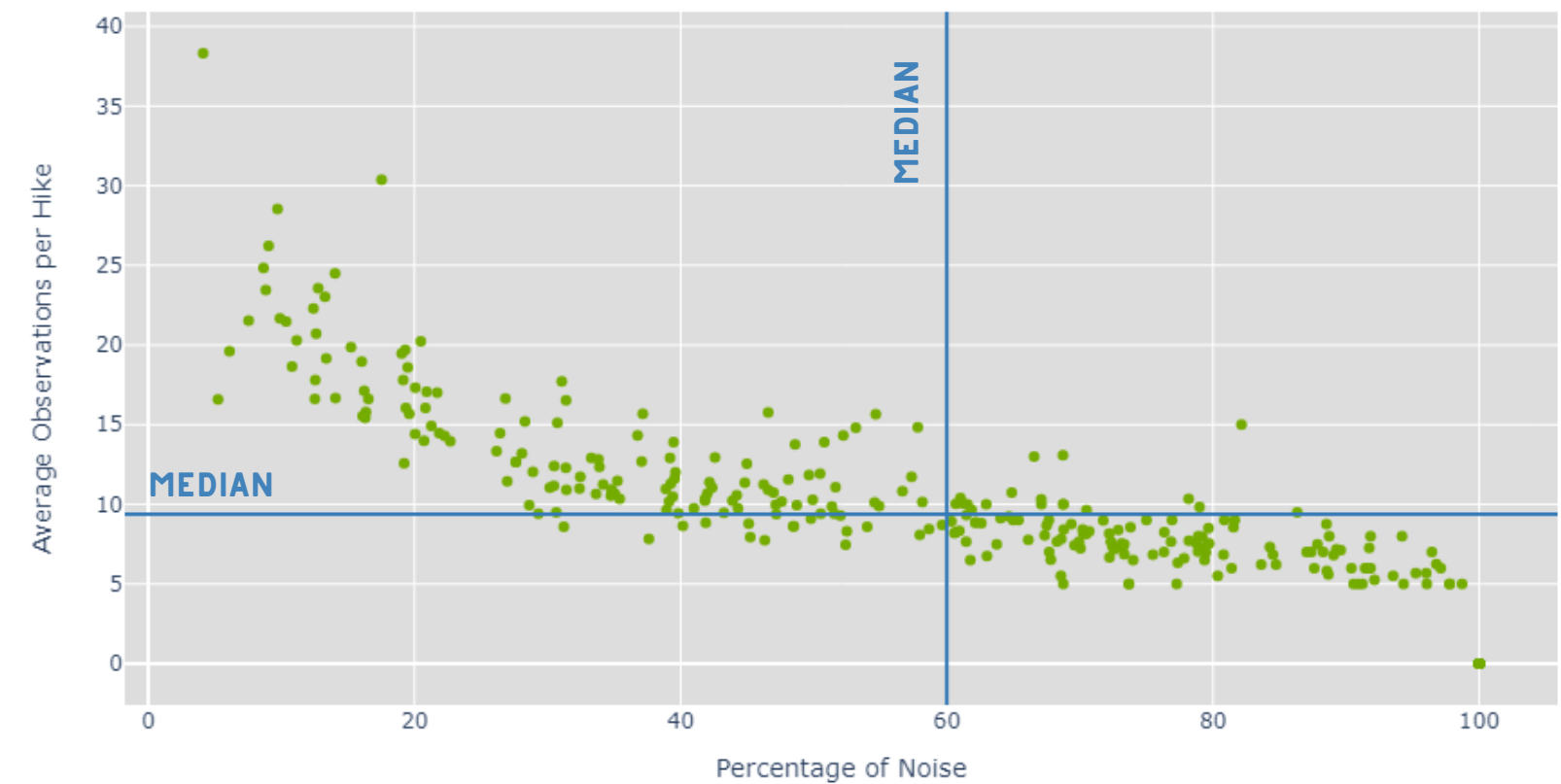
IMPACT

MORE DATA!! MORE SCIENCE!!

What percent of a user's data demonstrates "one-off observation" behavior?

If a user demonstrates "hike-like" behavior, what are the average number of observations they take per "hike"?

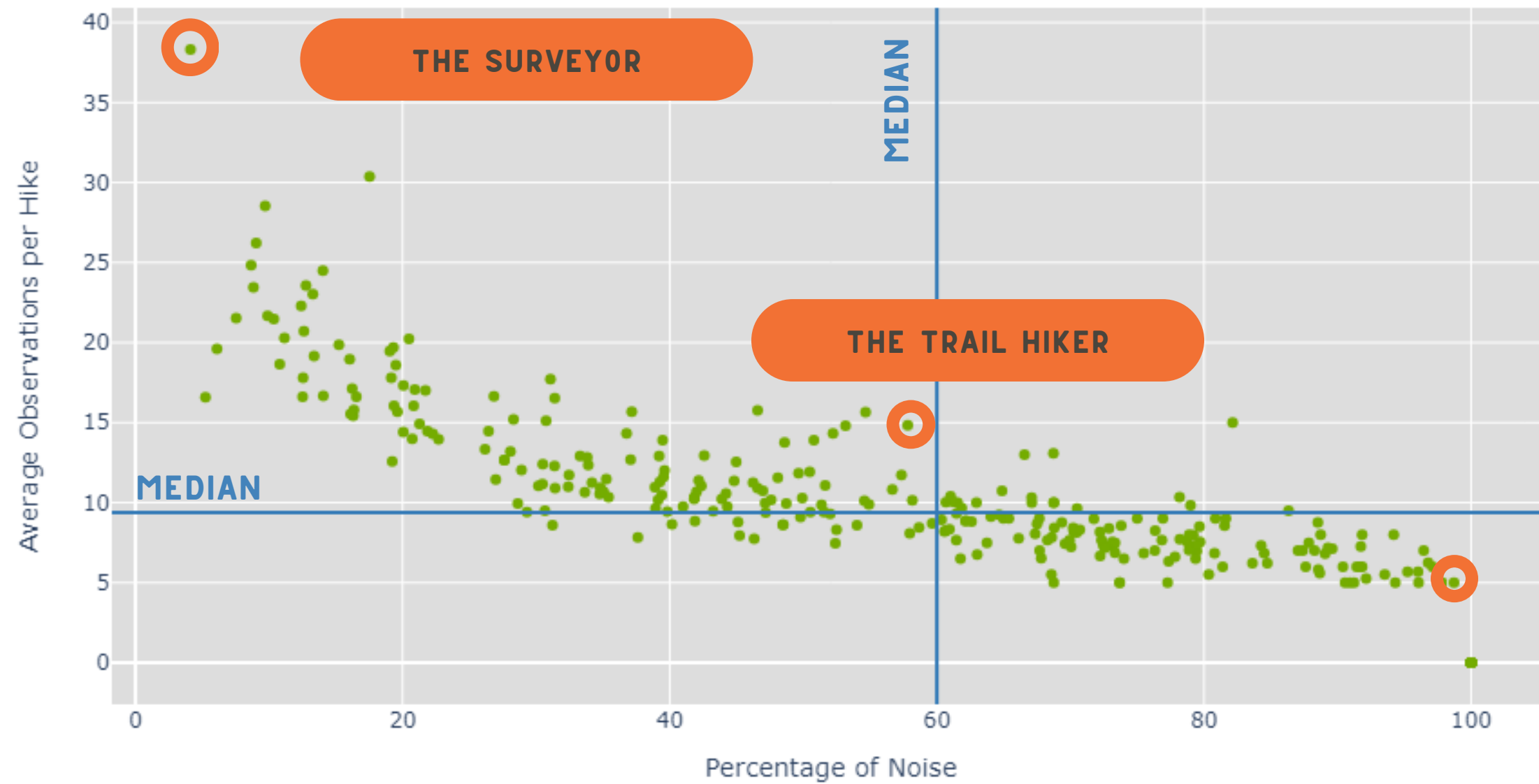
As "One-Off" Behavior Increases, Observations per Hike Decrease



IMPACT

TARGETED USER ENGAGEMENT

As "One-Off" Behavior Increases, Observations per Hike Decrease



THE NEIGHBORHOOD AMBLER

SOLUTION: CLUSTERING



HEURISTIC MODEL

RULES-BASED

- Groupby by date.
 - Any date with more than the specified minimum number of samples was grouped as a “hike-like” cluster.
-
- Silhouette scores performed very well.
 - But non-discerning between “hike-like” behavior and “one-off” happening in the same day.
 - Acted in place of “external validation” to compare against DBSCAN models.

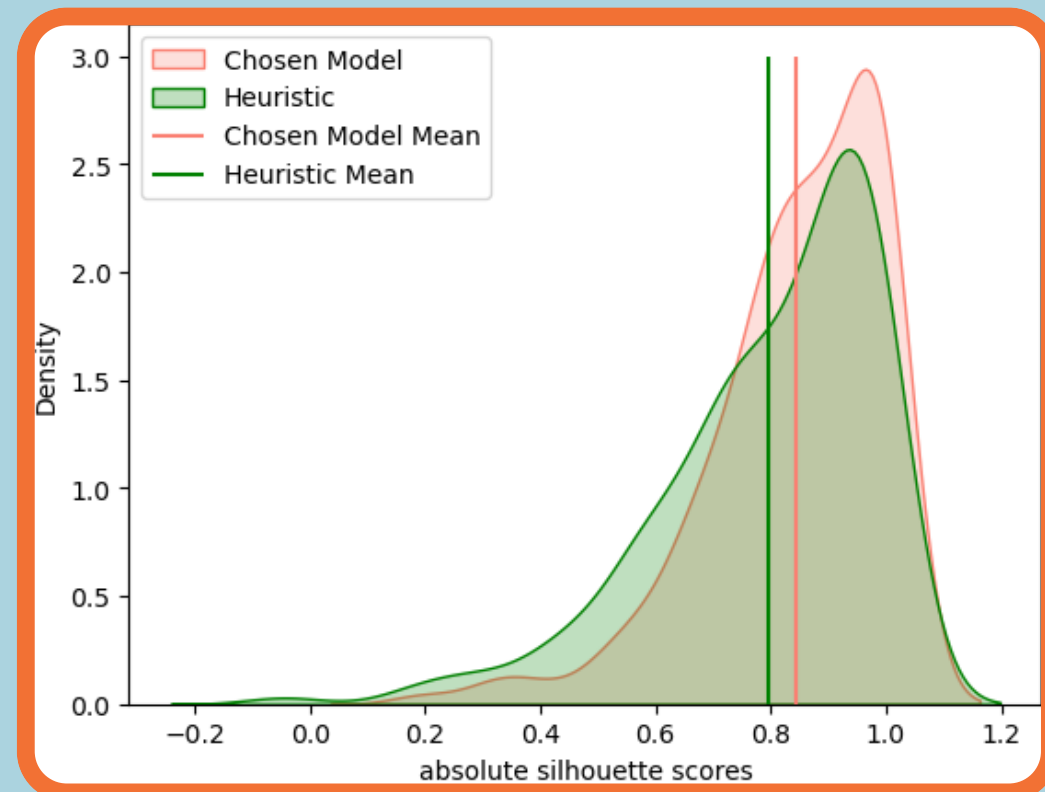
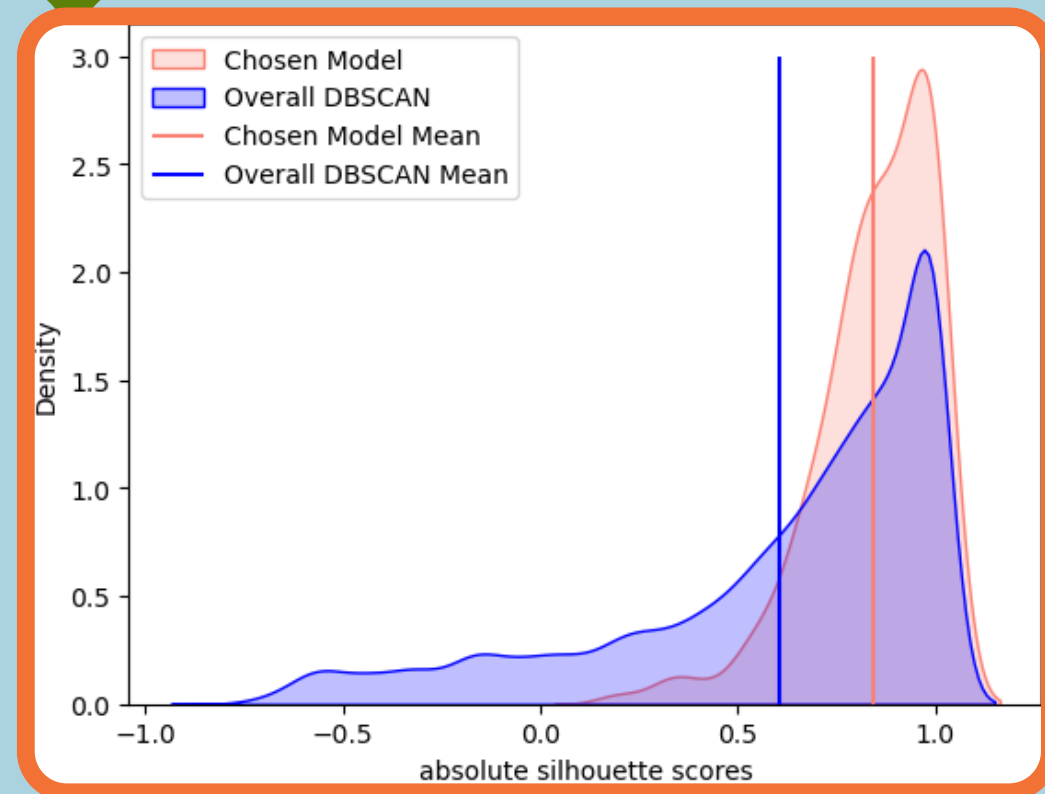


DBSCAN MODEL

CLUSTERING ALGORITHM

- Doesn't require knowing the number of clusters ahead of time
 - Identifies clusters (“hikes”) AND noise (“one-off observations”)
-
- Silhouette scores ranged from poor to good depending on feature selection and hyperparameters
 - Found a model that outperformed other DBSCAN models and Heuristic

SOLUTION: RESULTS



THE DBSCAN MODEL I CHOSE:

1) MEAN SILHOUETTE SCORE WAS
40% IMPROVEMENT OVER MEAN
FOR GENERAL RANGE OF DBSCAN
MODELS

2) MEAN SILHOUETTE SCORE WAS
6% IMPROVEMENT OVER HEURISTIC

THANK YOU

[HTTPS://GITHUB.COM/MCCOOLGRIME/](https://github.com/MCCoolGrime/)



[WWW.LINKEDIN.COM/IN/LISA-MCCOOL-GRIME](https://www.linkedin.com/in/lisa-mccool-grime)

