Optimal Birdwatching in Massachusetts: Species Frequency Forecasts & Birding Activity Analysis

Phase 2 - Final Presentation

Team 2: Michael Aldoroty, Neil Ghosh, Nhat Pham, Hajera Siddiqui

DS 5500

Objectives

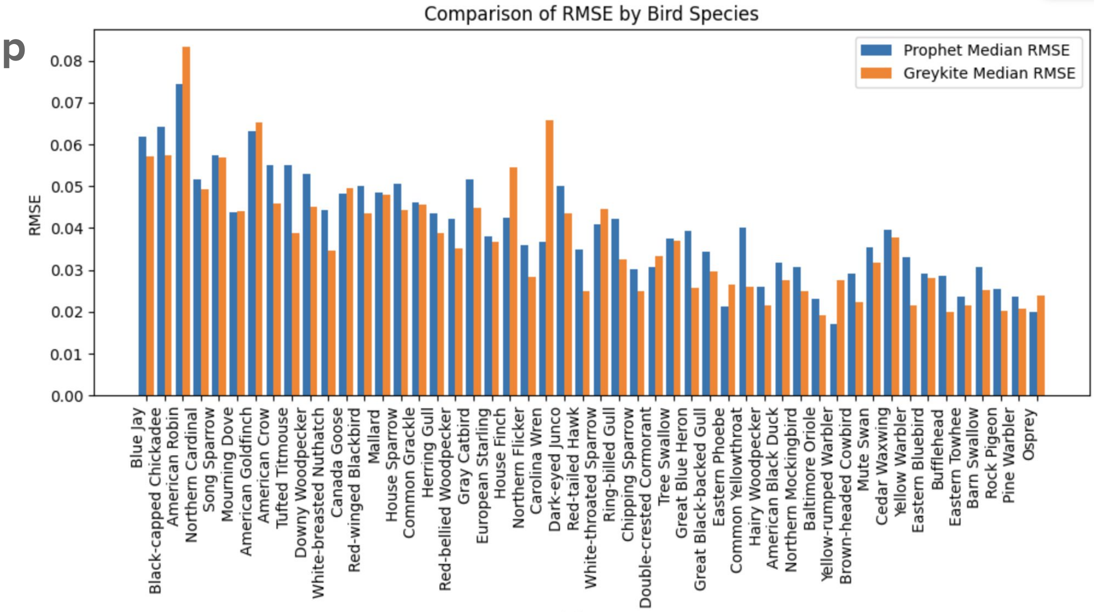
1. **Detection rate forecast**
    a. Improve performance and speed
    b. Scale model to all 465 birds in 14 counties
2. **Environmental regressors**
3. **Semi-automated pipeline**
4. **Birding activity analysis**

Forecast Models

Prophet vs Greykite

- **Testing on 50 species:** Fine-tuned hyperparameters (trend, seasonality etc.) for both models, employing grid search and CV for optimal performance, parallel processing to improve computational efficiency

- **Selected Prophet for final scale-up**
    - 3x faster than Greykite
    - Comparable accuracy



Final Prophet Model

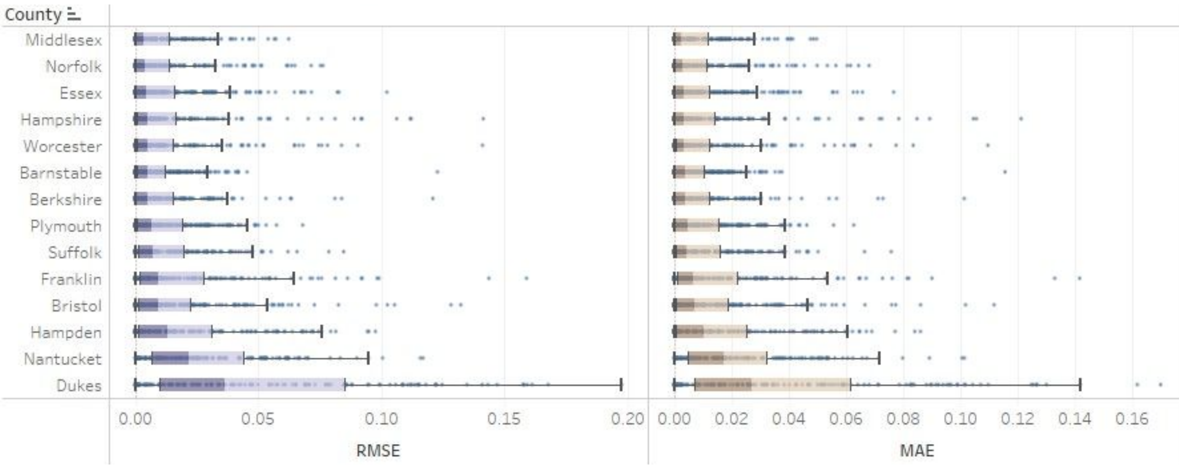
Data preparation	Streamlined dataset to Prophet's requirements Employed logistic regression: carrying capacity = 1, floor = 0 Applied exponential smoothing to reduce noise
Parameter tuning	Focused on changepoint & seasonality prior scale Grid search & CV for optimal parameter selection based on minimum RMSE Parallel testing with multiprocessing library, "Pool" object
Model execution	Ran each model with best changepoint parameter for 2-yr forecast Compiled forecast & actual data for Tableau dashboard Record RMSE and MAE values for model assessment

Results

- Total runtime (6510 models) is 1 hr 30 min
  - 11.6 seconds/ species vs. 27.6 seconds/ species in previous version
- Satisfactory accuracy

Metric	Min	Max	Median
RMSE	2.27e-9	0.197	0.007
MAE	1.36e-9	0.170	0.005

- Model performance varies across counties



Results

- High predictive accuracy for migratory birds because of pronounced seasonal patterns
- Lower predictive accuracy for resident birds due to variable detection patterns, underreporting, complex behaviors

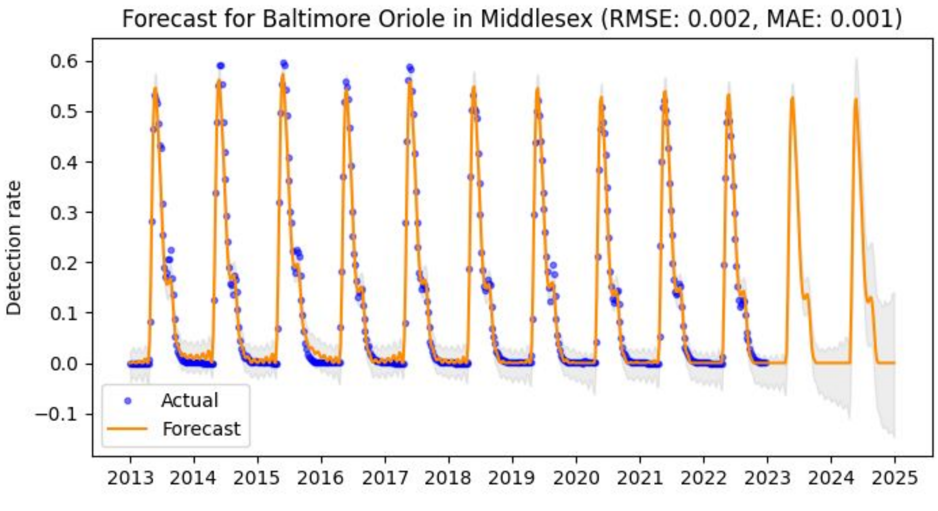


Fig 1. Forecast for a migratory bird

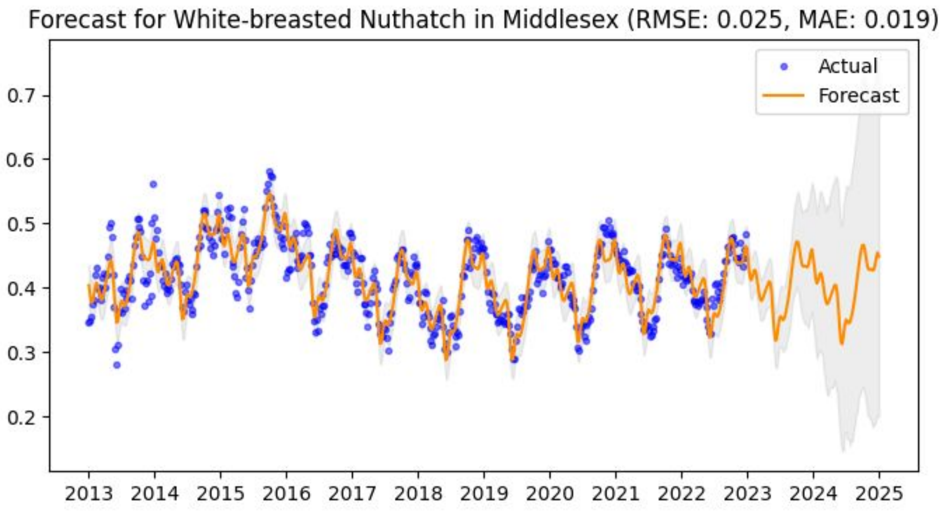


Fig 2. Forecast for a resident bird

Results

- Performs well with elusive, hard-to-detect species (falcons, owls)
- Captures seasonality and trends well, especially changes that occur gradually; struggles with abrupt fluctuations

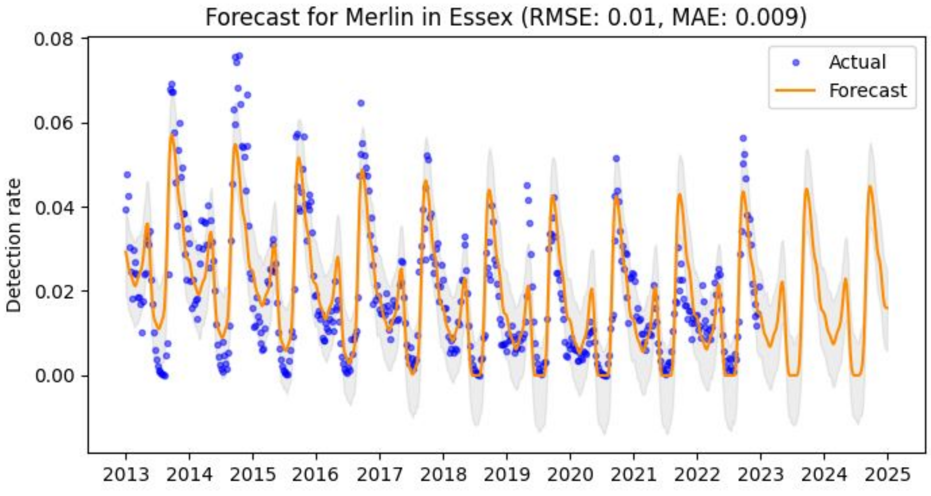


Fig 3. Forecast for an elusive bird

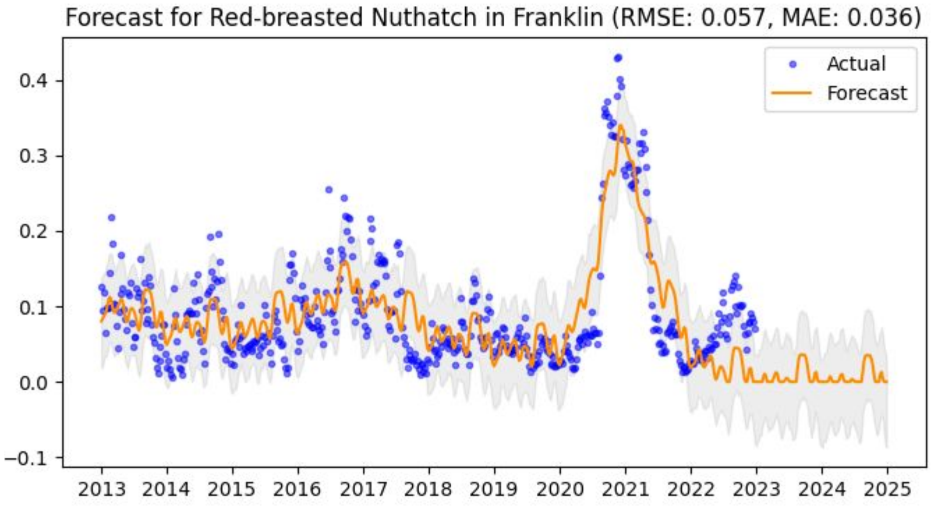


Fig 4. Forecast with changes in trend

Regressors

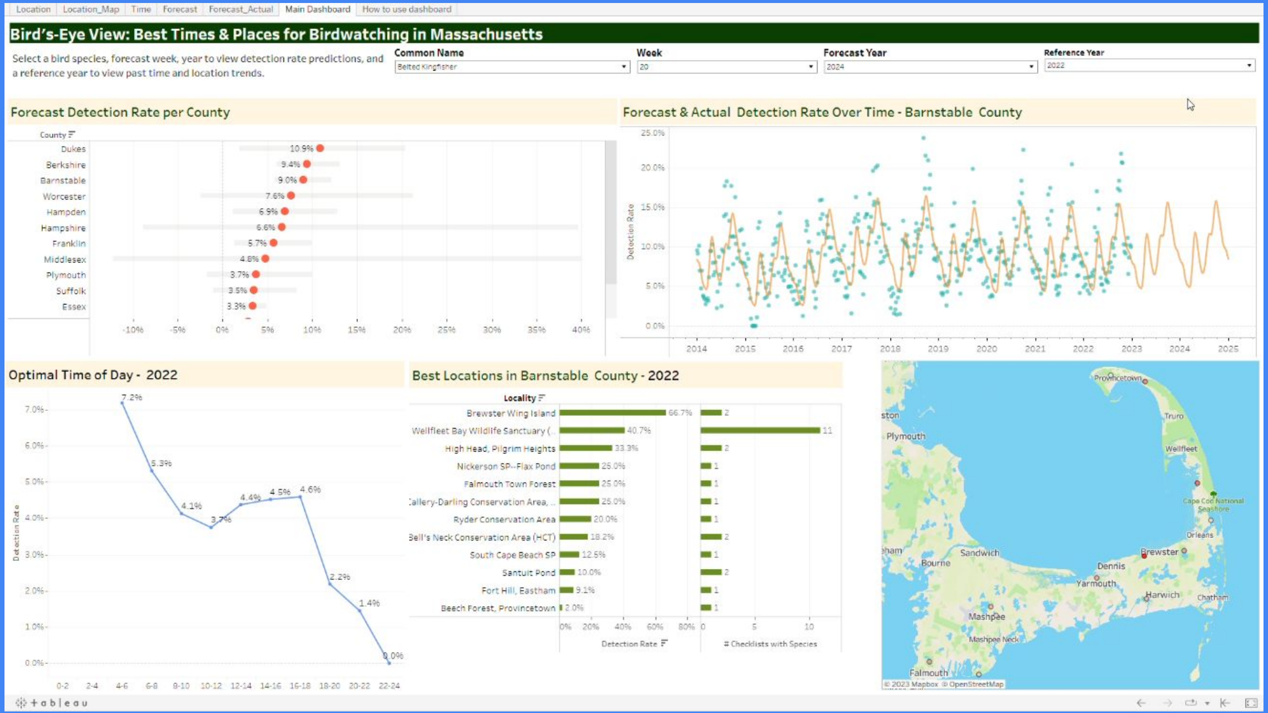
- We tested two regressors on the final Prophet model: average temperature and precipitation; data retrieved from the National Centers for Environmental Information
- Main challenge was sourcing a dataset that matched the weekly time scale of our eBird data. Data from NCEI was aggregated on a monthly basis, so four weekly eBird data points per month had the same temperature and precipitation value

County	Common Name	RMSE	
		No Regressor Model	Regressor Model
Barnstable	Northern Cardinal	0.0419	0.0422
Barnstable	Tufted Titmouse	0.0428	0.0443
Essex	Northern Cardinal	0.0435	0.0442
Essex	Tufted Titmouse	0.0419	0.0449

Automation Pipeline

- Download latest data (yearly)
- Data cleaning in R, Python
- Merge new entries into a rolling 10-year dataset for model training
- Retrain forecast models with new data
- Update analyses to reflect latest patterns
  Sync and display updated results on Tableau dashboard

https://tabsoft.co/3uFSm5S



Updated Dashboard

Birding Activity Analysis

Motivation

- User survey

Recommendation for “under birded” areas that could use more data collection

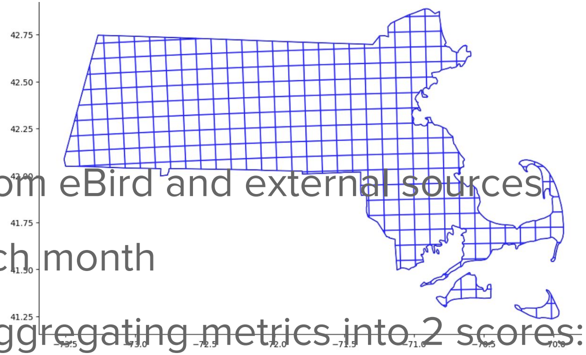
New locations or under birded locations

Perhaps under-birded areas near me

clearly bad at finding it. Also places/times that are underbirded and where my observation data would be useful.

- **Objective:** identify under-birded and heavily birded areas to guide birdwatchers and support conservation efforts
- **Data**: 2022 eBird records, American Birding Association, MA Avian Records Committee

Descriptive Analysis

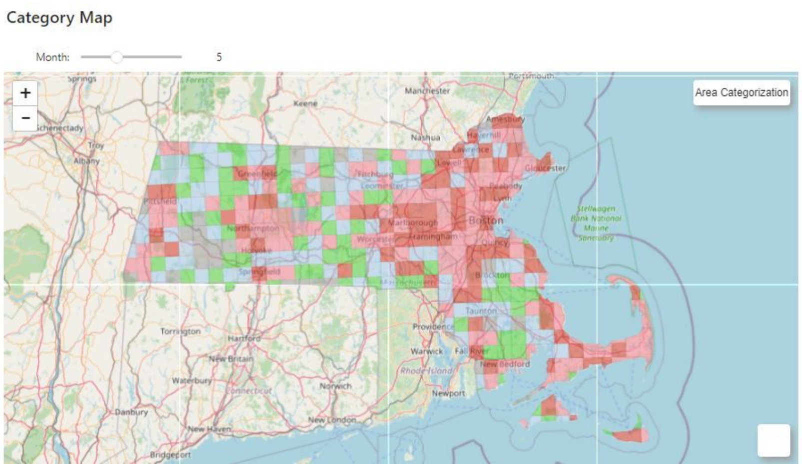


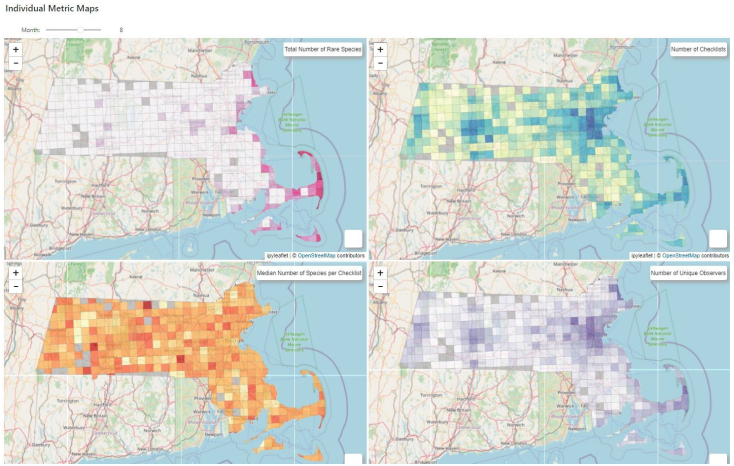
- Segmented MA into 8km x 8km grid cells
- Identified 4 key metrics, including rare species count, from eBird and external sources
- Assigned checklists to grids & computed metrics for each month
- Compare areas on species diversity & birding activity, aggregating metrics into 2 scores:
  - Diversity score = 0.4* median_num_species + 0.6* num_rarebirds
  - Popularity score = 0.4* num_checklists + 0.6* unique_observers
- Area categorization:

	Popularity	Diversity
Underbirded	low	high
Heavily birded	high	low
Popular	high	high
Low interest	low	low
Not enough info	n/a	n/a

Visualization

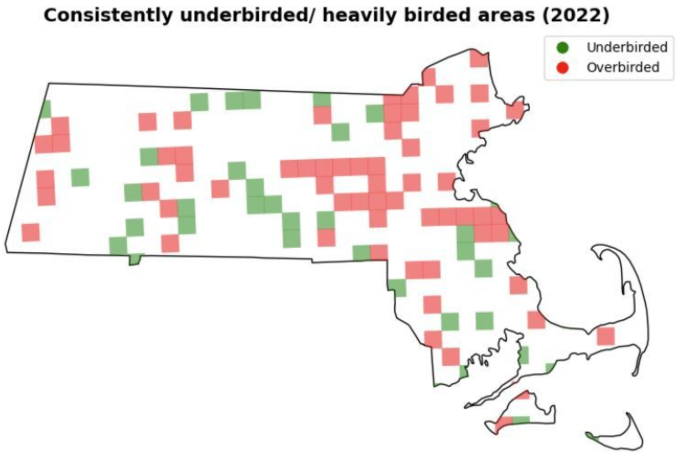
- Choropleth maps:
- Slider to capture month-specific birding trends
- View maps: https://bit.ly/3R7xhZH





Results

- Popular/heavily birded areas identified in coastal, urban, and valley regions, corresponding with populated and accessible regions
- Underbirded areas are more scattered geographically and temporally



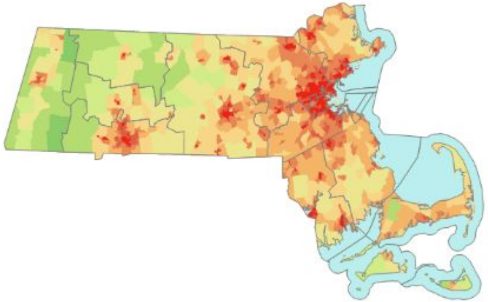
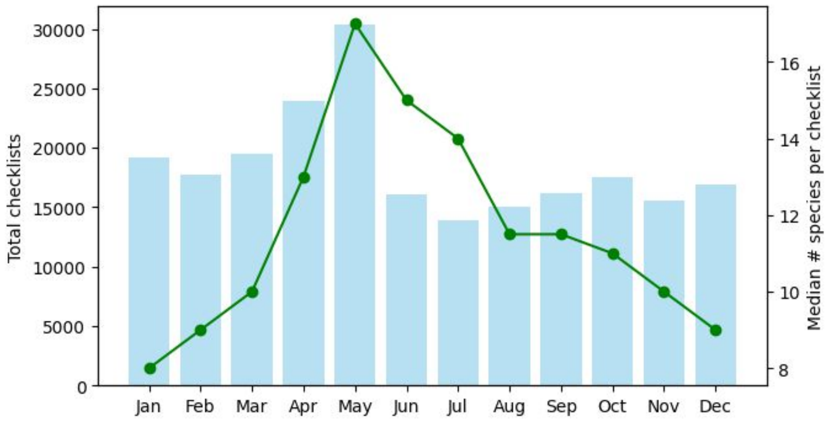


Fig. Consistently over/under-birded areas in 2022 (left), MA population density map 2010 (upper right), MA topology map (lower right)

Results

- There is a weak correlation between area popularity and number of species (coefficients 0.08 - 0.28) but stronger with rare species sightings (0.34 - 0.61)
- Birding activity peaks in May with spring migration
- Number of species observed fluctuates yet birding activity is relatively stable throughout the year



Discussion

- Developed a scalable forecast model with robust accuracy, a streamlined process for data updates and model retraining
- Dashboard guides birdwatchers in trip planning and assists habitat managers in monitoring and conservation efforts
- Choropleth maps identify overbirded and underbirded regions to balance environmental impact
- The tools facilitate comprehensive analysis of bird migration, species diversity, and urbanization effects

User Feedback

What do you find useful?

I appreciated seeing optimal times of the day for spotting bird. Interesting dashboard.

Location data.

I love seeing the guess about how likely I am to see a particular species at a particular place, based on a previous year!

I really like this tool! The information is clearly presented and easy to interpret.

First off, I'm thrilled to see grad students working on this! It could be a big contribution in the realm of the Patagonia picnic table effect

A particularly useful tab is Location_Map, since if a new birder is searching out a specific species, they can see possible locations at a quick glance.

How can we improve?

Narrow the list of species in the huge dropdown menu to those I might actually see. Show me results from birding areas within a 40 minute drive from a specific location that I can set on the map. Show the month and day in the dropdown menu, not the week number, and start with default of this week.

Some axes didn't have labels (weeks). I wondered about the accuracy of using 2-3 year-old data to forecast birding spots, especially rare birds. Seems like patterns are changing.

For bird species selection, it seemed to not have much/enough data for a number of species (particicularly for the location section). Some made sense since they wouldn't normally be seen this time of year, others less so

* It's not clear which filters are applied on each tab. For example, if you go to Location_Map, a user would have no idea that Week X is filtering the results.

* As for specific improvements, rewrite technical language. "Forecast Detection Rate" could simply be "Likelihood". "Week" needs to be a date range

Future Work

- Explore model scaling through species clustering and hierarchical modeling
- Search for high-quality habitat and weather data, integrate into both tools to improve forecast models and enhance understanding of birding activities
- Shift to a hotspot-focused approach and include accessibility information to give more practical insights to users
- Upgrade infrastructure with cloud solutions for better data hosting and develop a Flask app for centralized access to tools

References

- Cornell Lab of Ornithology. (n.d.). Trends maps. eBird Science. Retrieved November 19, 2023, from https://science.ebird.org/en/status-and-trends/trends-maps
- Cornell Lab of Ornithology. (2023). Use eBird data and tools. eBird Science. Retrieved November 3, 2023, from https://science.ebird.org/en/use-ebird-data
- Floyd, T., Minardi, R., & Retter, M. L. (2023). ABA Checklist. American Birding Association. Retrieved December 6, 2023, from https://www.aba.org/aba-checklist/
- Greykite - A flexible, intuitive and fast forecasting library. (n.d.). LinkedIn Open Source. Retrieved October 16, 2023, from https://linkedin.github.io/greykite/
- Khare, P. (n.d.). Understanding FB Prophet: A Time Series Forecasting Algorithm. Medium. Retrieved November 3, 2023, from https://medium.com/illumination/understanding-fb-prophet-a-time-series-forecasting-algorithm-c998bc52ca10
- NCEI. (2023). County Time Series | Climate at a Glance. National Centers for Environmental Information. Retrieved December 6, 2023, from https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/county/time-series
- Trimble, J., & Schain, R. (n.d.). Review List – Massachusetts Avian Records Committee. Massachusetts Avian Records Committee. Retrieved December 6, 2023, from https://maavianrecords.com/review-list/
- 12.2 Prophet model | Forecasting: Principles and Practice (3rd ed). (n.d.). OTexts. Retrieved October 16, 2023, from https://otexts.com/fpp3/prophet.html

Thank you!

Comments / Questions?