

---

# Waterfowl Travel: Team Iteration 5 Report

---



4/21/2025

Prepared by the students of CSC 4615 – Spring 2025

Jacob Sullivan

Tania Perdomo Flores

Revel Etheridge

Kenny Adams

Drew Burkhalter

Breanna Woosley

Under the direction of

Dr. William Eberle (Professor)

Stories Identified: .....	3
Plan For Next Iteration: .....	4
Burnup Chart: .....	5
Retrospective Summary:.....	5
Team Temperature: .....	7
Showcase:.....	8
Appendix: .....	8

## Stories Identified:

Authors: Revel Etheridge, Drew Burkhalter

Contributors: Revel Etheridge, Drew Burkhalter, Jacob Sullivan, Kenny Adams, Breanna Woosley, Tania Perdomo Flores

During this iteration, developers focused on model refinement and finalization with fine tuning of weather parameters. The GNN model was reformatted to utilize a dual edge weighting system for location clustering while the LSTM model finalized its k-d tree implementation. Weather data sourcing was transferred from NOAA to OpenWeather and an accuracy evaluation program was created to help decide which of the two models would be most preferable for the client. Developers also furthered their understanding and capabilities of/within the HPC to increase efficiency and scale of predictions; additionally, exportation of data structures was also standardized to support the automation process of pulling and modeling data from Movebank. Finally, UI improvements were made including migration visualizations, directional patterns, and heatmapting that were all in line with the client's vision.

- **Stories completed**

- As a developer, I want to implement the GNN model to the clustering and dual edge weighting to round off duck locations so that we give a more generic prediction (Revel Etheridge) [5pts]
- As a user, I want confidence intervals on predictions to gauge reliability and variability. (Kenny Adams) [5pts]
- As a developer, I want to convert from NOAA to OpenWeather so that we can have as many weather nodes possible to allow more accurate and faster readings. (Kenny Adams) [8pts]
- As a user, I want to finalize the GNN model and integrate a dual edge-weighting framework for efficient spatial querying and Open weather parameter verification, so that I can model duck migration as a dynamic graph and make accurate, environmentally informed predictions. (Kenny Adams, Revel Etheridge, Drew Burkhalter) [13pts]
- As a developer, I want to ensure the export data of both models follow a structure so that it will allow for easy integration into the front-end map. (Tania Perdomo Flores, Breanna Woosley, Revel Etheridge, Drew Burkhalter) [3pts]
- As a user, I want to move away from Wix and explore more robust hosting options, such as git pages, to better support our data needs and enable a complete redesign of our frontend and user interface. (Jacob Sullivan) [13 pts]
- As a developer, I need to automate the process of pulling Movebank data and the process of the HPC to enable scheduled runs for Dr. Renfro (Jacob Sullivan) [13pts]
- As a user, I want to see the probability/direction of ducks leaving specific areas, along with the covariance of environmental factors (e.g., weather, temperature) that influence these departures, to better understand migration triggers. (Jacob Sullivan) [5pts]
- As a client I want to make a name/logo change so that I can have my own DNS/Ownership of the name (Tania Perdomo Flores) [1pt]

- **Stories in progress**

- As a user, I want to finalize the LSTM model and integrate the k-d tree, data vectorization, and OpenWeather weather parameter verification so that I can generate accurate, location-aware predictions of duck migration patterns. (Tania Perdomo Flores, Breanna Woosley) [13pts]

## Plan For Next Iteration:

Author: Kenny Adams, Drew Burkhalter

Contributors: Revel Etheridge, Drew Burkhalter, Jacob Sullivan, Kenny Adams, Breanna Woosley, Tania Perdomo Flores

Iteration 6 is the final planned iteration and therefore must bring the product to a presentable state. We plan to make multiple changes to the website itself, including improving the UI and modeling ducks with a gradient cloud. Once the website is presentable to the public, we will also create a client package to allow our client to keep the website operational. It will also be necessary for the HPC and Movebank to be able to communicate for future duck data to be used. We will finalize our code to communicate between the data crunching websites, model prediction algorithm, and the public website.

Additionally, we want to improve the accuracy of predictions by fine tuning our models, which can be achieved by combining the best features of our current models or better filtering the weather data. The run time of our predictions is also something we hope to improve by further utilizing the HPC to run more efficiently. We will need to consolidate our efforts from the two models to one to ensure our final product has the most effective method of predicting data.

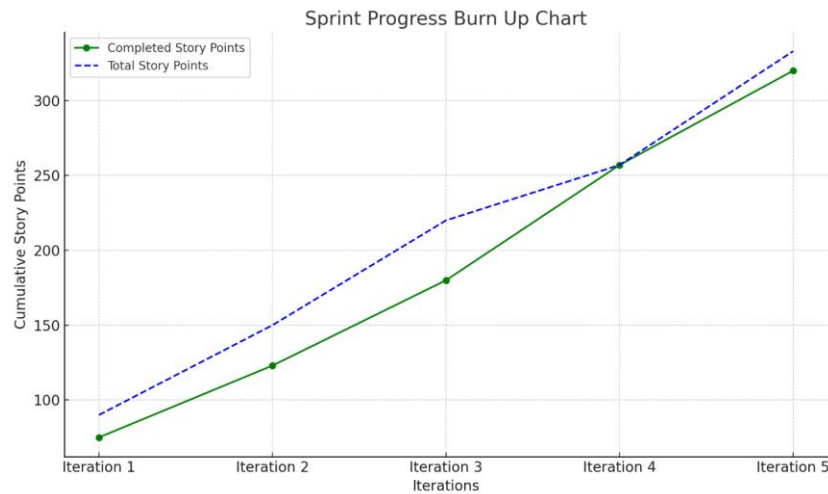
- **Stories Planned**

- As a developer, I want to make machine learning models that successfully predict the migration movements of mallards so that the client is given a fully functional app. (Tania Perdomo Flores, Drew Burkhalter, Kenny Adams, Revel Etheridge, Breanna Woosley) [13 pts]
- As a developer, I want a fleshed-out webpage so that it is completed for the client and for public usage. (Jacob Sullivan) [5 pts]
- As a client I want to change the map heat using a Geo JSON to reflect a gradient cloud and shapefile of the Mississippi Flyway so that it generalizes the duck's movement in a certain area. (Jacob Sullivan) [13 pts]
- SPIKE: As a developer, I want to finalize the model of choice so that we have the preferred model for this type of data handling. (Jacob Sullivan, Kenny Adams, Tania Perdomo Flores, Drew Burkhalter, Revel Etheridge, Breanna Woosley) [3 pts]
- As a developer, I want to create a client package so that the client will know the implementation and directions needed so that it can be handled after hand-off. (Tania Perdomo Flores, Drew Burkhalter, Kenny Adams, Revel Etheridge, Breanna Woosley) [5 pts]
- As a developer, I want to change the function of the Movebank pull script so that it can specifically pick the months or days non-chronologically. (Jacob Sullivan, Revel Etheridge) [5 pts]
- As a developer, I want to be able to run a large amount of data and automate it with the HPC to push to Git Pages so that we know it can handle the dataset without the client intervening. (Jacob Sullivan, Tania Perdomo Flores, Drew Burkhalter, Kenny Adams, Revel Etheridge, Breanna Woosley) [13 pts]
- As a developer, I want to convert the output of our model to a Geo JSON instead of a CSV so that it does not weigh down the front-end. (Tania Perdomo Flores, Drew Burkhalter, Kenny Adams, Revel Etheridge, Breanna Woosley) [8 pts]

## Burnup Chart:

Author: Jacob Sullivan

Contributors: Revel Etheridge, Drew Burkhalter, Jacob Sullivan, Kenny Adams, Breanna Woosley, Tania Perdomo Flores



We began Iteration 5 with a total of 76 points and finished with all but 13 completed, making this iteration successful overall. We made significant progress on our UI transition from WIX to GitHub Pages, which now allows us to host the site freely without the need for a dynamic backend, since all data is preprocessed. Thanks to the meetings with our client and other professors who guided us in the right direction, we were able to make key decisions on both the UI and the models we've been using. This has led to the successful prediction from our GNN model, which, using a month's worth of data, was able to closely predict duck movement a week in advance. As for the LSTM, we are still running into a few bugs, and it is currently under review to determine whether we want to continue development or end the creation of that model. Finally, with Kenny implementing our new weather API and comparison algorithm, we've taken a big step toward ensuring data consistency and accurately pulling only the weather data we need.

## Retrospective Summary:

Author: Breanna Woosley

Contributors: Revel Etheridge, Drew Burkhalter, Jacob Sullivan, Kenny Adams, Breanna Woosley, Tania Perdomo Flores

### What Has Gone Well?

- Retrospectives and Coordination**  
 Regular retrospectives have helped identify blockers early and improve coordination among team members. This rhythm has kept the project organized despite challenges.
- Clear Communication**  
 The team continues to communicate effectively across all channels, ensuring tasks are distributed clearly and dependencies are well-managed.

- **High Performance Computing (HPC) Utilization**

Despite initial delays, we've successfully resumed using HPC to train our models. Jobs are queued efficiently, and we're starting to make better use of the compute time.

- **Frontend Revamp**

The frontend component of our project was overhauled to improve usability and better visualize model outputs. This has helped validate results and share progress more effectively.

### What Could Be Improved?

- **Debugging and Code Testing**

Debugging has been a major pain point, especially given the long runtime on HPC. This delays iteration and makes it difficult to test small changes quickly.

- **Model Progress**

Progress on refining the models is slower than expected. While we've established a working pipeline, the accuracy is still off by over 300 miles — highlighting the need for better data preprocessing, feature selection, or model choice.

### What Questions Do You Have About the Project or Process?

- **Should we commit to either GNN or LSTM, or try a hybrid model?**

We need to determine if an ensemble or hybrid approach might help reduce our prediction error, or if simplifying to one architecture is more efficient.

- **How do we improve model accuracy?**

What additional data inputs, preprocessing methods, or tuning strategies can we apply to reduce the haversine error distance?

- **How much HPC time is reasonable to use per iteration?**

With runs taking 24+ hours, do we need a backup local model or smaller test subset to validate logic faster?

### What Action Items Need to Be Addressed?

- **Establish a Lightweight Debugging Workflow**

Design a local version of the model using a reduced dataset or fewer epochs to debug faster without needing full HPC jobs.

- **Set Evaluation Criteria for Model Selection**

Define metrics to help decide between GNN, LSTM, or a hybrid approach based on test performance and runtime efficiency.

- **Explore Hybrid Model Integration**

Investigate methods to blend GNN and LSTM predictions — e.g., weighted averaging, stacking, or temporal-spatial feature fusion.

- **Refine Preprocessing for Spatial Accuracy**

Revisit input features, particularly spatial interpolation and weather alignment, to reduce error in the prediction output.

## Team Temperature:

Author: Tania Perdomo Flores

Contributors: Revel Etheridge, Drew Burkhalter, Jacob Sullivan, Kenny Adams, Breanna Woosley, Tania Perdomo Flores

Spider chart along with a description and summary of our team's temperature.



### Summary:

After the fifth iteration of this project, our spider diagram reflects high results in *Team Communication*, *Stakeholder Engagement*, *Workload*, and *Collaboration* all clustered in the low-to-mid-4 range on a five-point scale. *Project Progress* is the outlier and the lowest of the five metrics. This dip reflects the fact that we have not yet processed the entire data set; running analyses on the full data set is a prerequisite for several downstream UI features, so those items remain on hold until the data work is complete.

High-performance-computing (HPC) runs had to be scheduled outside regular class hours, so much of the heavy lifting happened asynchronously and limited real-time collaboration. *Team communication* channels and *stakeholder engagement* touch-points are functioning well, and *workload* has been sustainable, but *project progress* will remain capped until full-dataset processing is finished. Priorities for the next phase are clear: (1) secure sufficient HPC slots to complete the end-to-end data run, and (2) carve out synchronous working blocks—virtual or in-person—to lift the *Collaboration* metric and accelerate UI development once the data bottleneck clears.

## Showcase:

Author: Tania Perdomo Flores

Contributors: Revel Etheridge, Drew Burkhalter, Jacob Sullivan, Kenny Adams, Breanna Woosley, Tania Perdomo Flores

At the end of this iteration, our team prepared a showcase of the effort completed. The showcase demonstrates the specific user stories completed during the iteration.

[https://youtu.be/\\_6Qy0Lehd1k](https://youtu.be/_6Qy0Lehd1k)

## Appendix:

Author: Jacob Sullivan

Contributors: Revel Etheridge, Drew Burkhalter, Jacob Sullivan, Kenny Adams, Breanna Woosley, Tania Perdomo Flores

- Kanban Board and GitHub
  - <https://github.com/users/JakeSul1023/projects/1/views/1> (Kanban Board GitHub)
  - [https://github.com/JakeSul1023/Team-7\\_Waterfowl](https://github.com/JakeSul1023/Team-7_Waterfowl) (Repository GitHub)
  - <https://github.com/JakeSul1023/duck-data-web> (New Web App Page)
  - <https://github.com/JakeSul1023/DuckData-WebApp> (Web-Page Repo) (Archived)