

# Predicting Great Salt Lake Waterfowl Populations

*Zach Eliason*

Level of interest: 7/10

## 1. Description of the project:

The goal of this project would be to predict total waterfowl populations at the Great Salt Lake using environmental data from the lake. The bird population numbers will come from researchers at Westminster University who conduct twice-a-month surveys. The environmental data will come from the USGS website<sup>1</sup>, which provides up to hourly information on the Great Salt Lake's depth, (imputed) area and volume, salinity, temperature, chemical composition, and other features. A machine learning model will then be developed to predict overall waterfowl populations based on correlation with a predetermined subset of these environmental factors. The ability to model waterfowl population is important because these birds rely exclusively on the Great Salt Lake for their migration. This information could be helpful for knowing at what point lake levels endanger the species' overall survival.

## 2. Data Features

I don't currently know which features would most affect waterfowl populations. Determining which to use would necessitate reading about these birds in ecology research papers and/or trying different feature combinations with the model. Potential features encompass a range of useful biogeochemical features, as well as waterfowl population data. Training data would take the following form (or something similar to it):

Date	Depth (m)	Area (km <sup>2</sup> )	Volume (km <sup>3</sup> )	Conductance (uS/cm)	Temp (C)	Salinity (ppt)	Waterfowl Population
10/4/23	40	10000	40000	10	20	100	3500

## 3. How the Data Set Will be Gathered:

The environmental data for the lake will be gathered by downloading TSVs where possible (it's unfortunately rarely possible) and by scraping the USGS website. I am in contact with researchers at Westminster University and have already collected their waterfowl data, although it exists in a very messy format across 5 or 6 separate sheets. In addition to obtaining the USGS data and cleaning the waterfowl dataset, our main challenge will be in reconciling inconsistencies in terms of time frames, spatial regions, and measurement units. Careful and evidence-backed data cleaning will be required to integrate the data into a consistent and usable format for the model to be able to pick up on meaningful signal.

---

<sup>1</sup> USGS water data for the nation. <https://waterdata.usgs.gov/nwis>