TEAM - IC24006

EY CHALLENGE: BIODIVERSITY STUDY

Thilak Mohan, Sadaf Davre, Ayan Hassen

Our Challenge: Developing a Species Distribution Model (SDM)

- Our mission is to develop a tool that can forecast where Litoria Fallax frogs might live. By analyzing weather patterns and geographic data, we aim to construct a model that accurately predicts frog habitat.

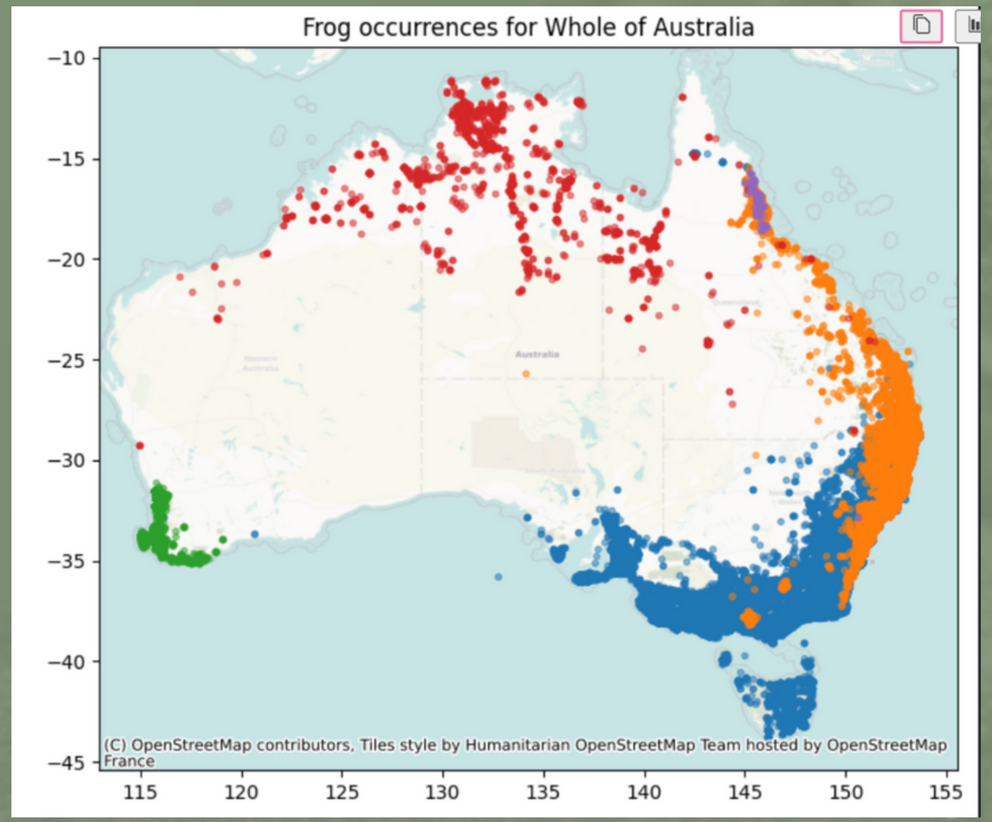
- We also aim to provide valuable insights into the factors influencing frog distribution, and eventually generate actionable insights to make a meaningful positive impact.

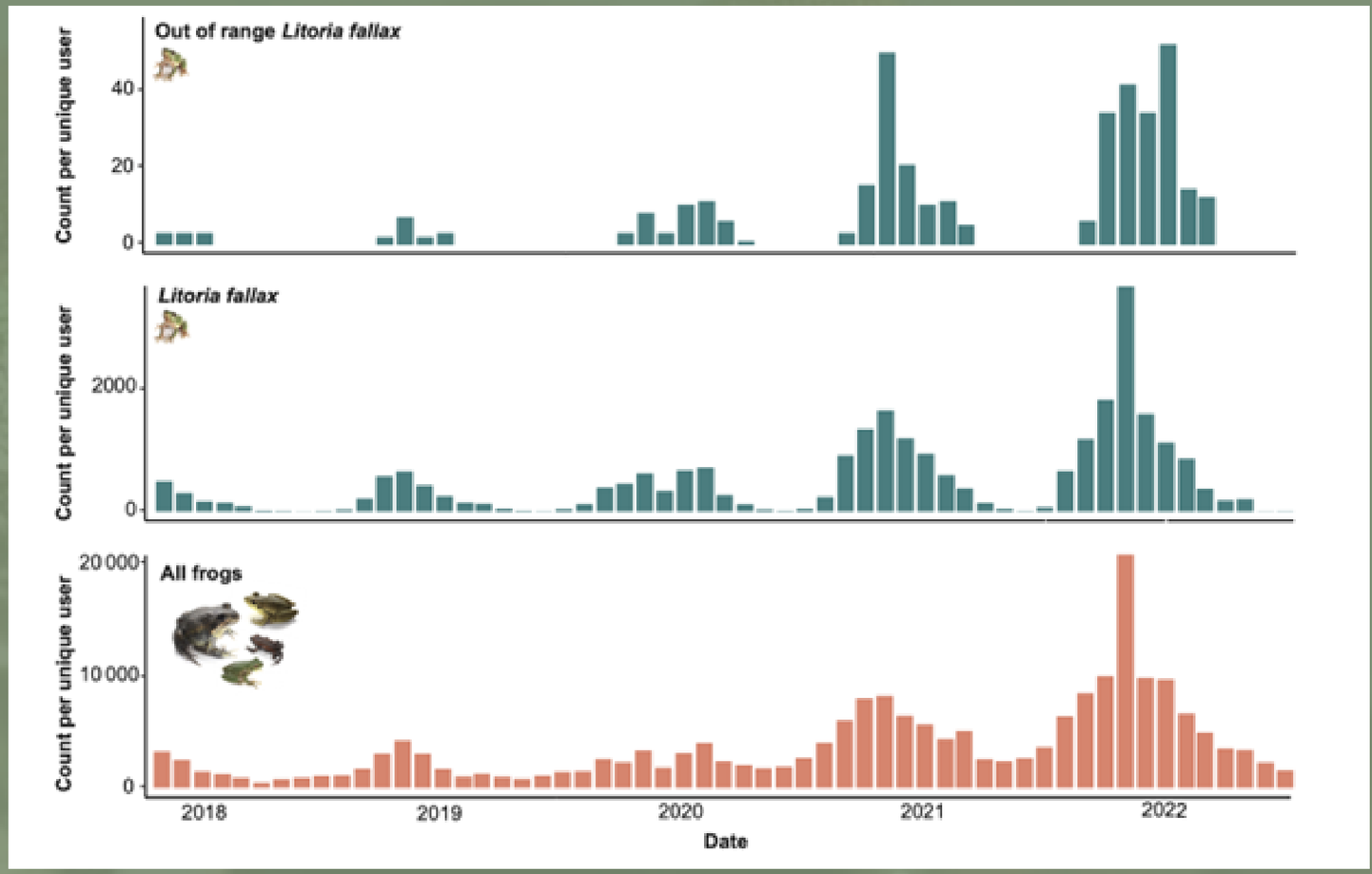
Why frogs matter

- Frogs serve as indicator species in ecosystems.
- Protecting frog populations is essential for ecosystem health and sustainable living.
- They provide insights into environmental health and biodiversity.

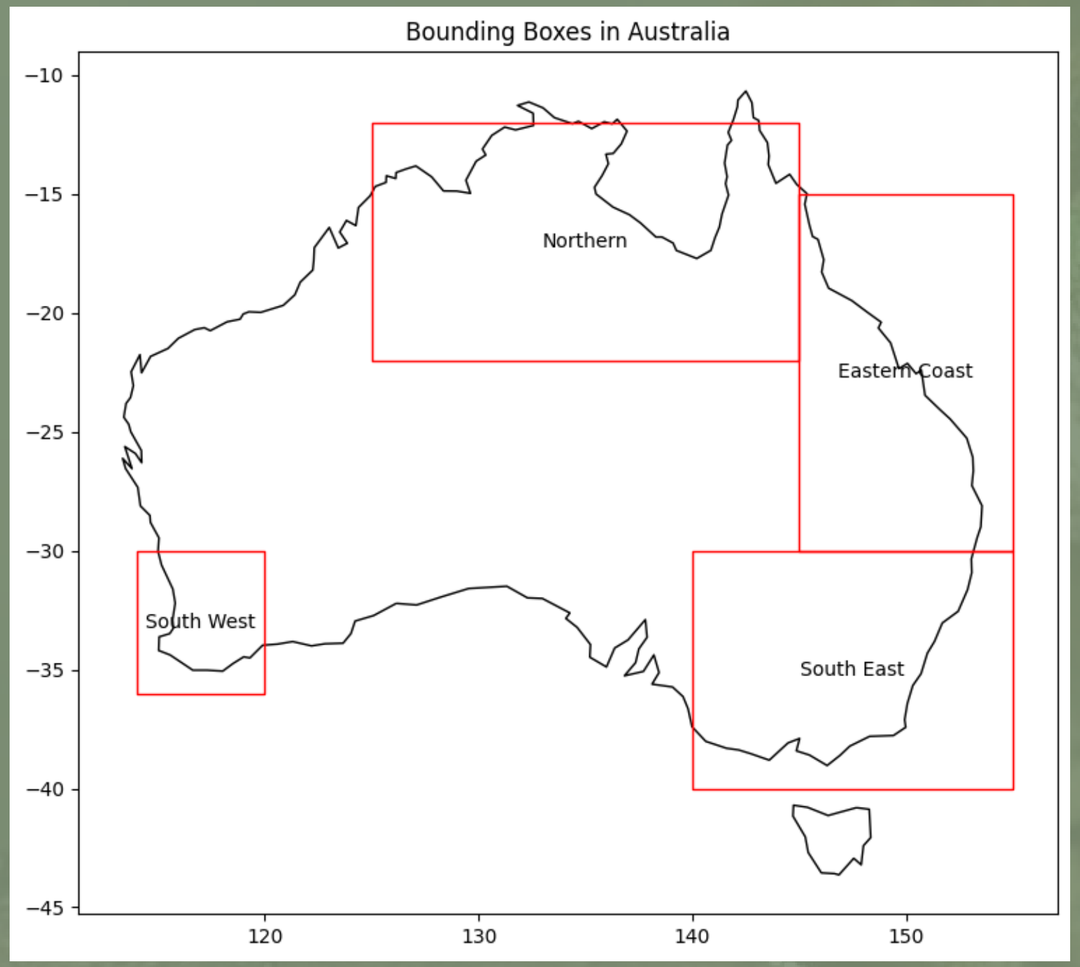
Analysis

- Seasonal Behavior
- Geographical Distribution
- Breeding Season Peak





Spatial Sampling

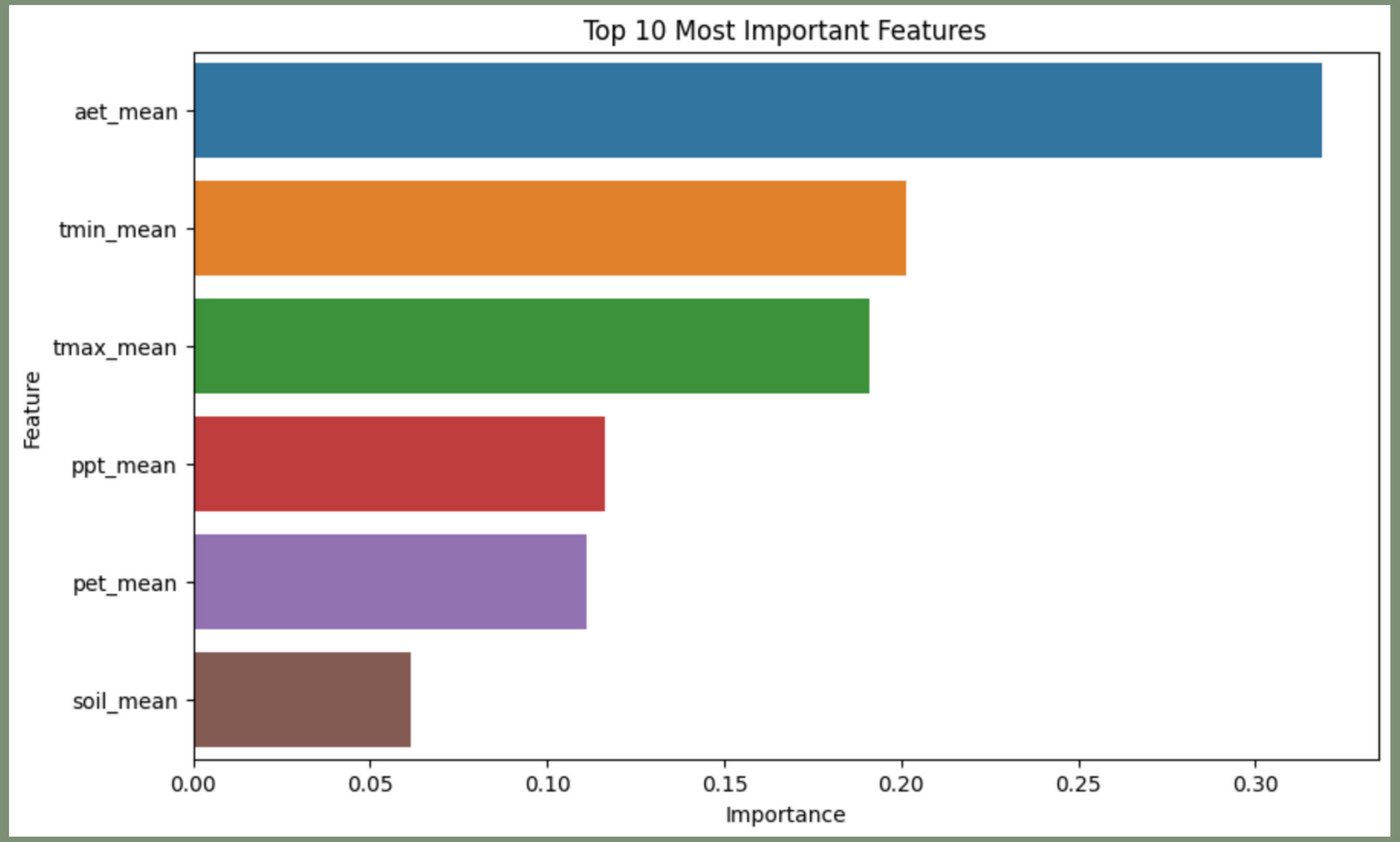


Sampling Bias

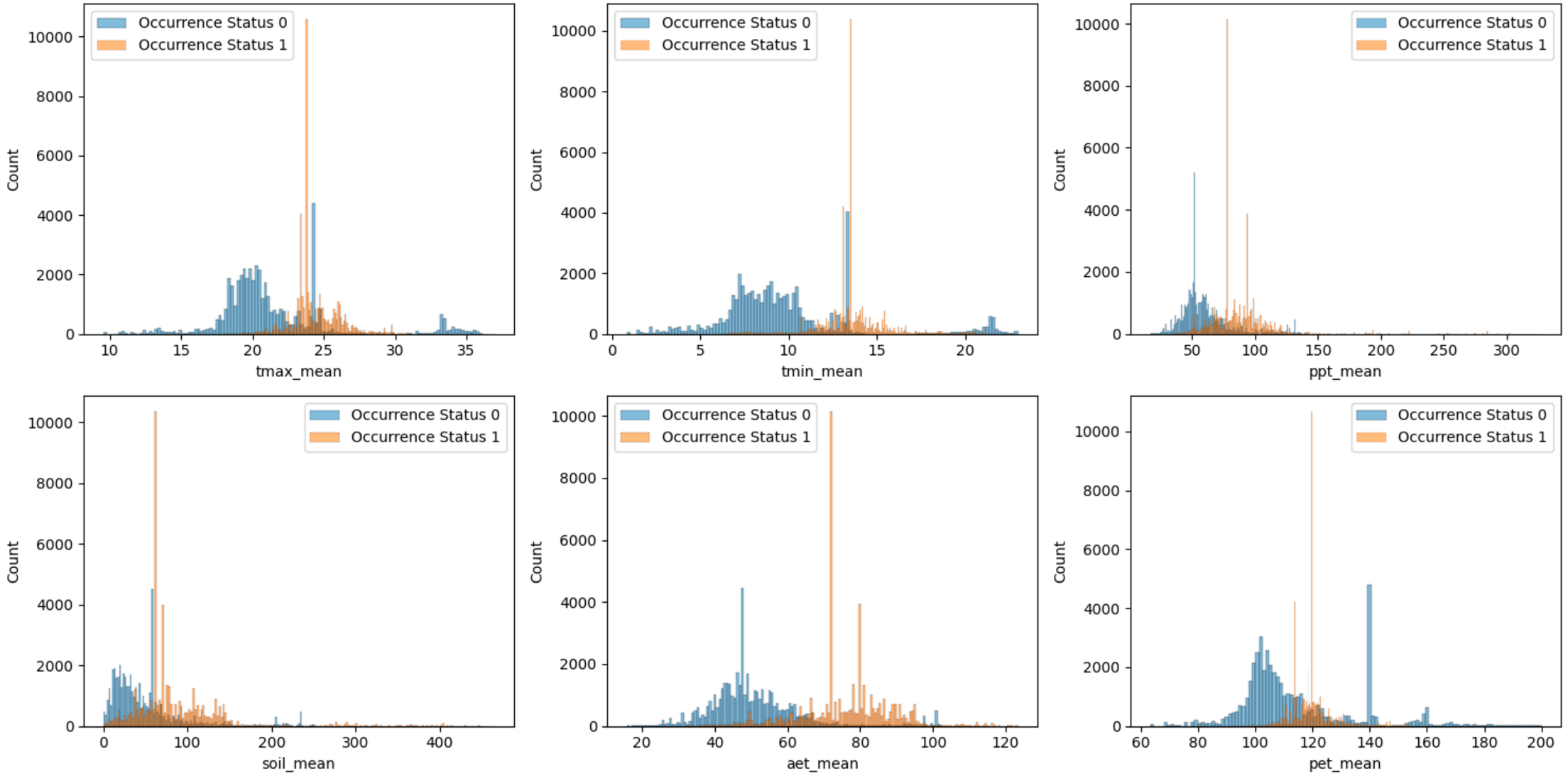
- Removed occurrences of species other than Litoria Fallax if they were close to the target species geographically
- KD-trees enabled efficient computation for finding the nearest neighbors

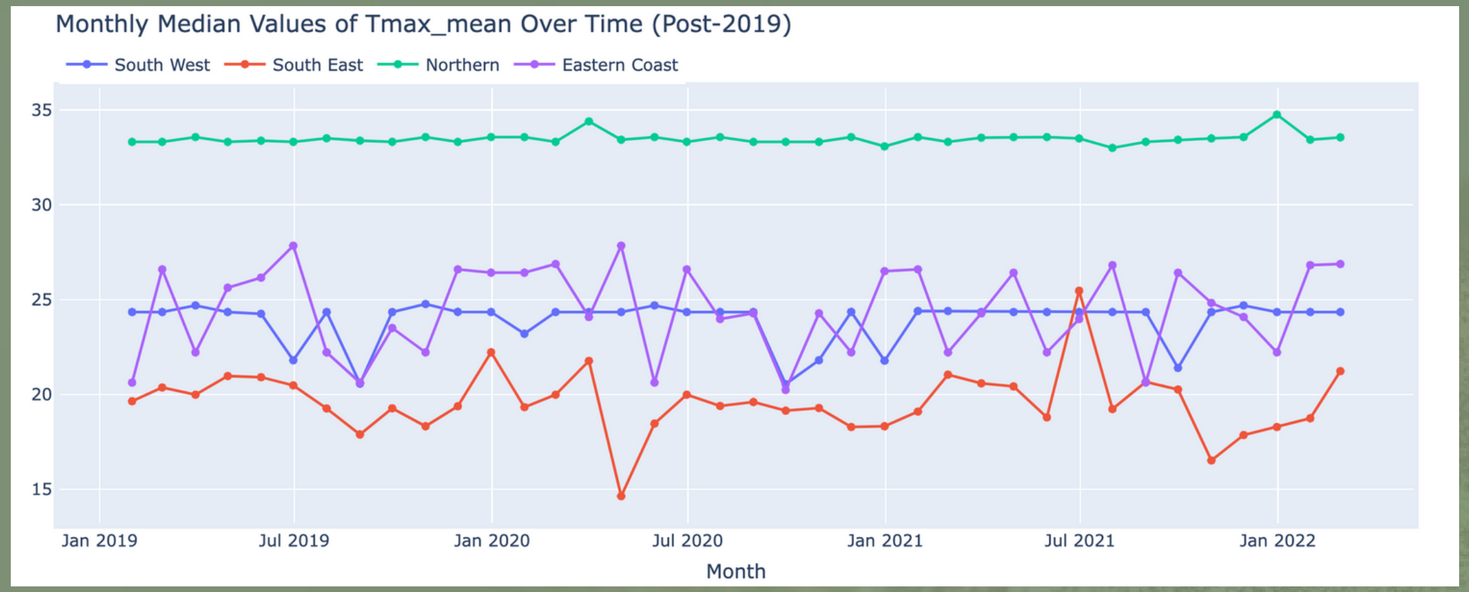
Feature Engineering

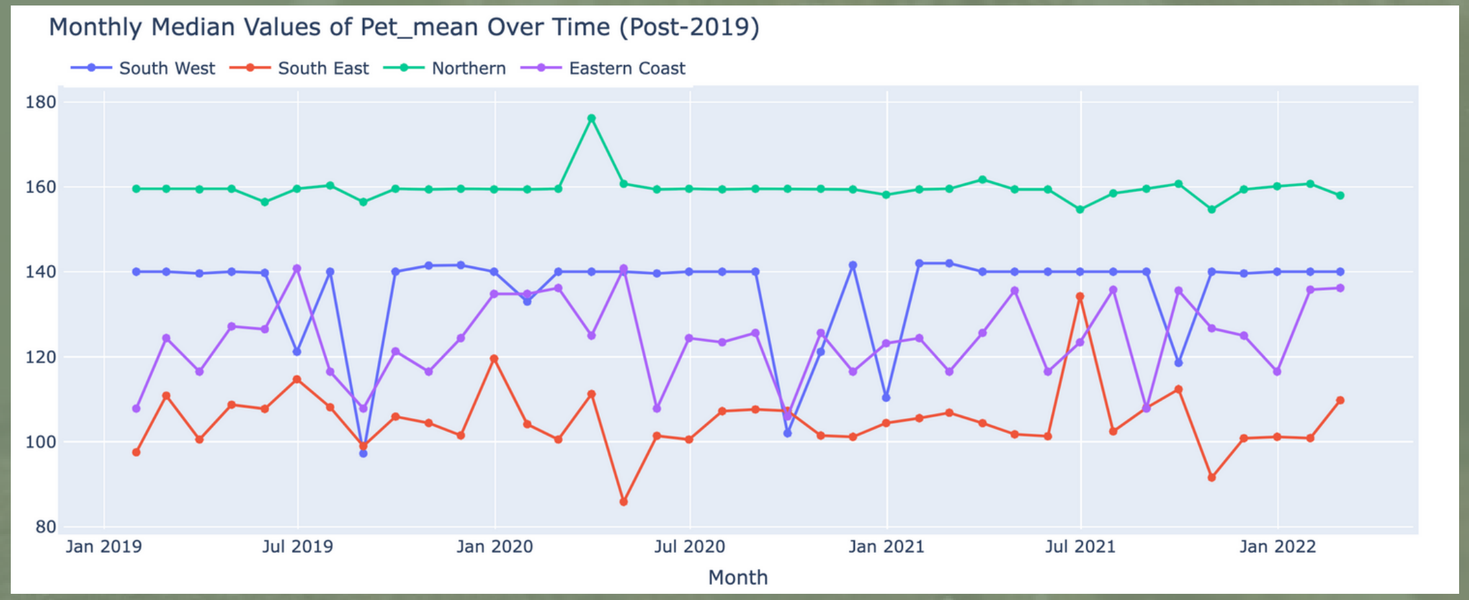
- **Extracted Additional Features: Actual Evapotranspiration (AET) and Potential Evapotranspiration (PET).**
- **Feature Standardization**
- **Handling Missing Values and outlier data**
- **Creation of Binned Variables: Binned variables were generated based on histograms analysis, facilitating a more granular understanding of feature distributions.**
- **Exploration of Feature Correlation and Importance: The correlation between features was examined, and feature importance was studied to identify key predictors influencing the target variable.**

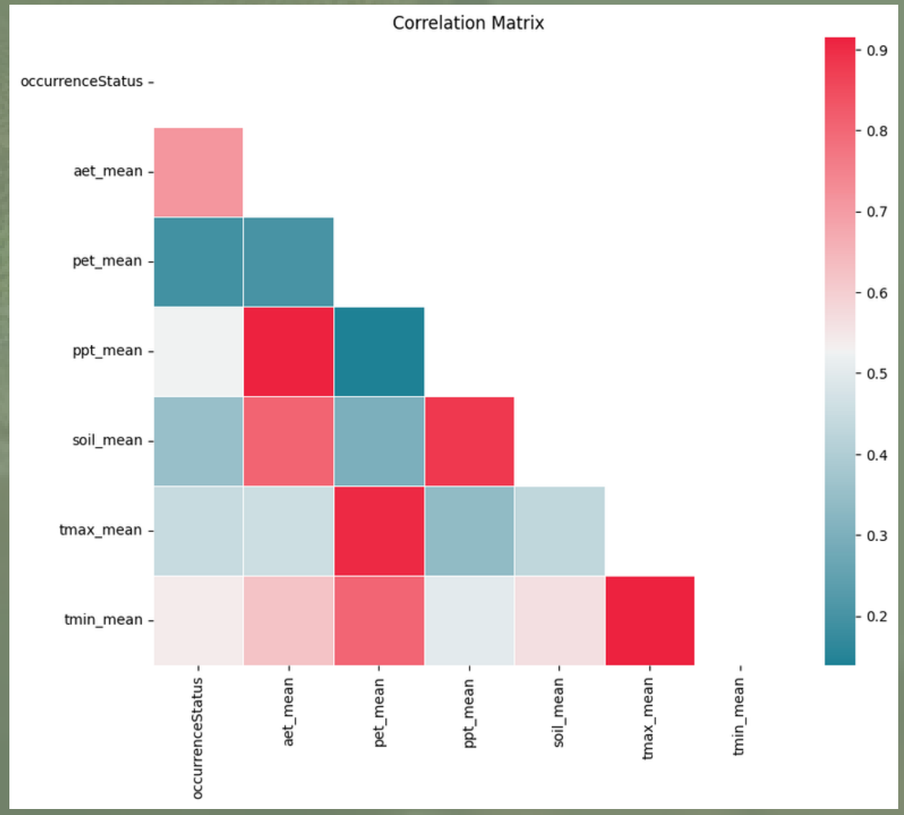

Top 10 Most Important Features
aet_mean
tmin_mean
tmax_mean
ppt_mean
pet_mean
soil_mean
Feature
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Importance

Frequency distribution of TerraClimate variables by Occurrence Status



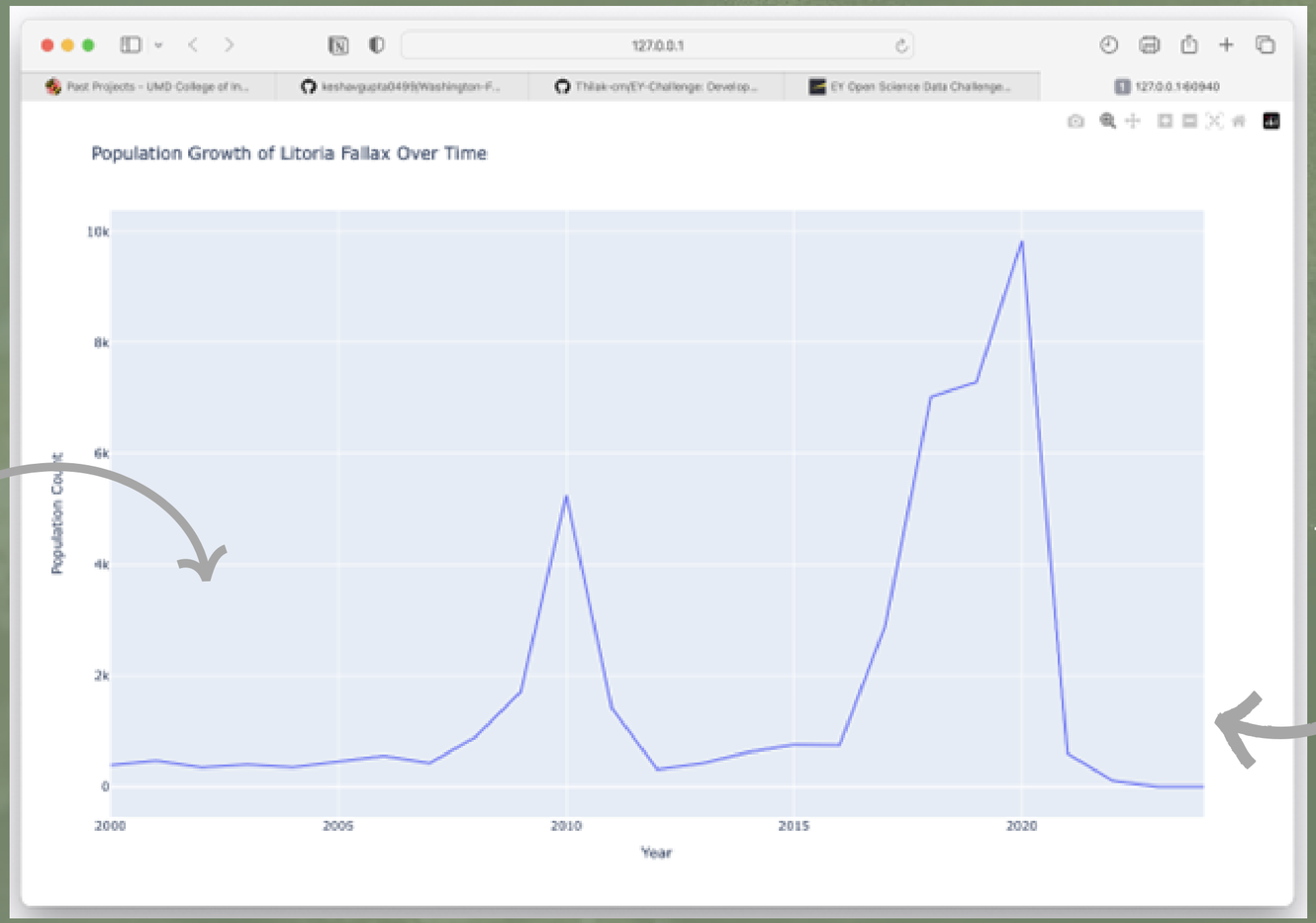

Monthly Median Values of Tmax_mean Over Time (Post-2019)
South West
South East
Northern
Eastern Coast
35
30
25
20
15
Jan 2019
Jul 2019
Jan 2020
Jul 2020
Jan 2021
Jul 2021
Jan 2022
Month


Monthly Median Values of Pet_mean Over Time (Post-2019)
South West
South East
Northern
Eastern Coast
180
160
140
120
100
80
Jan 2019
Jul 2019
Jan 2020
Jul 2020
Jan 2021
Jul 2021
Jan 2022
Month

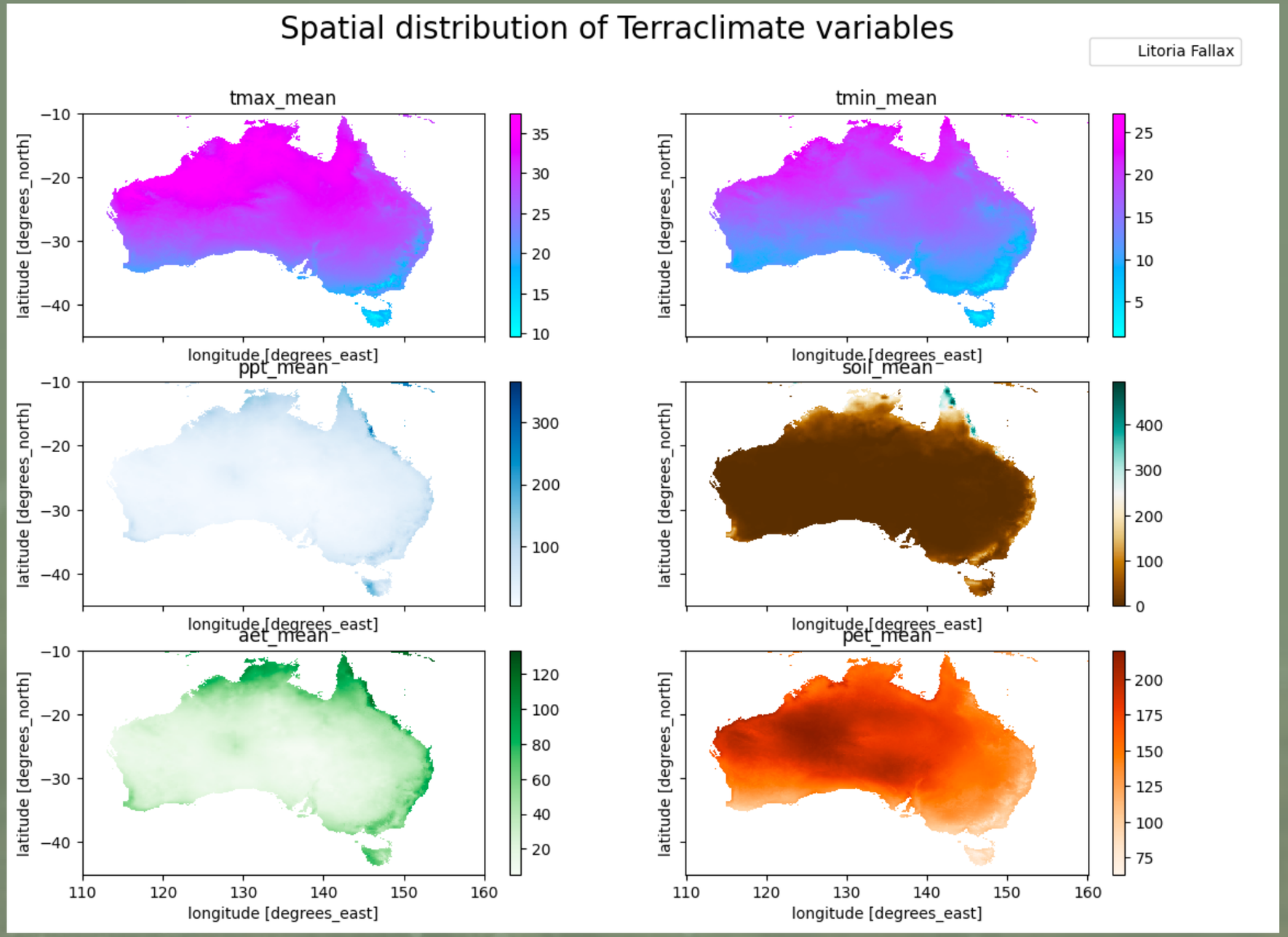

Correlation Matrix
occurrenceStatus
aet_mean
pet_mean
ppt_mean
soil_mean
tmax_mean
tmin_mean
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2

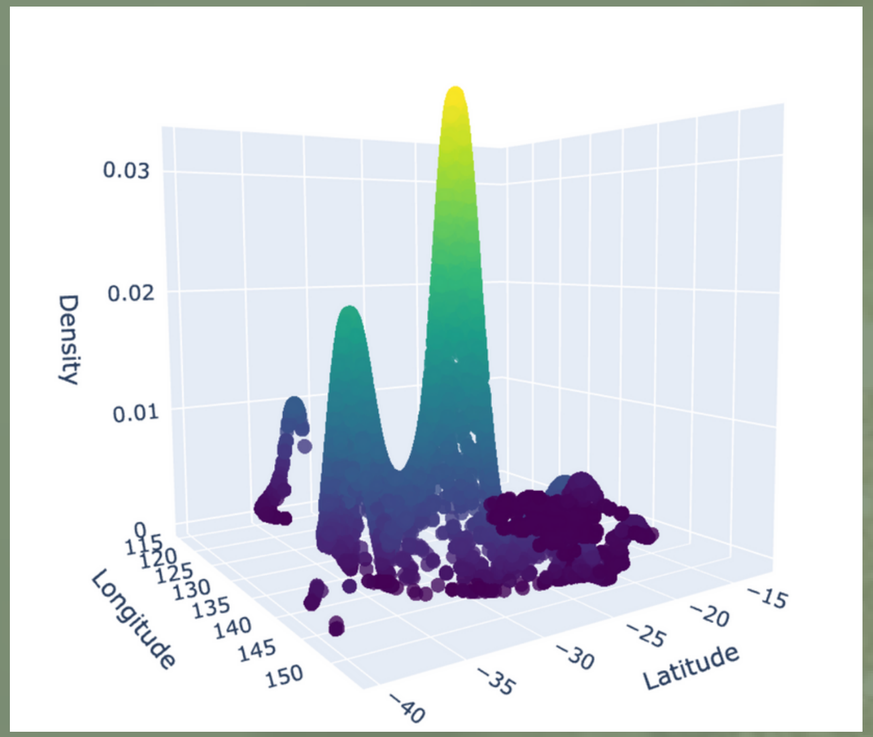
Why the dips in population?

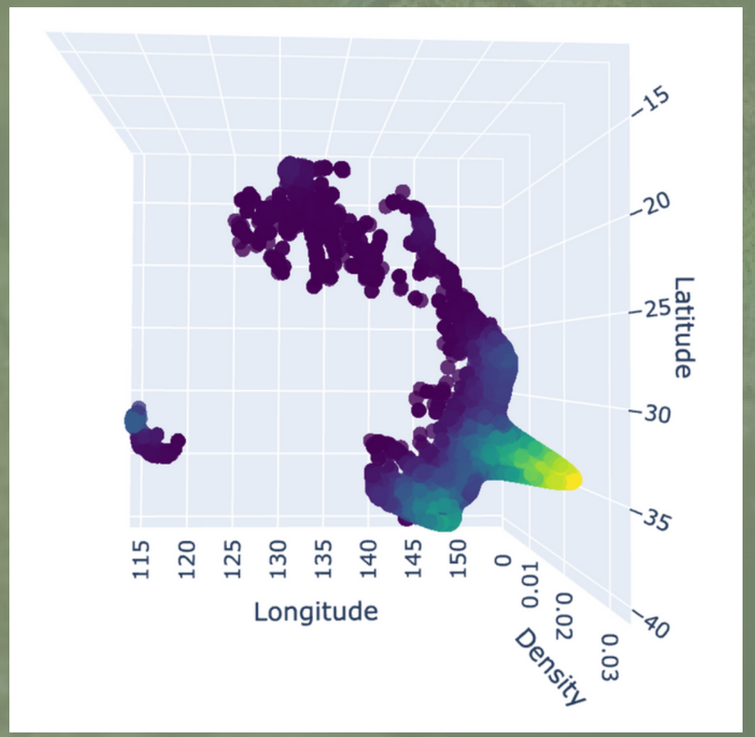
Climate change and habitat loss

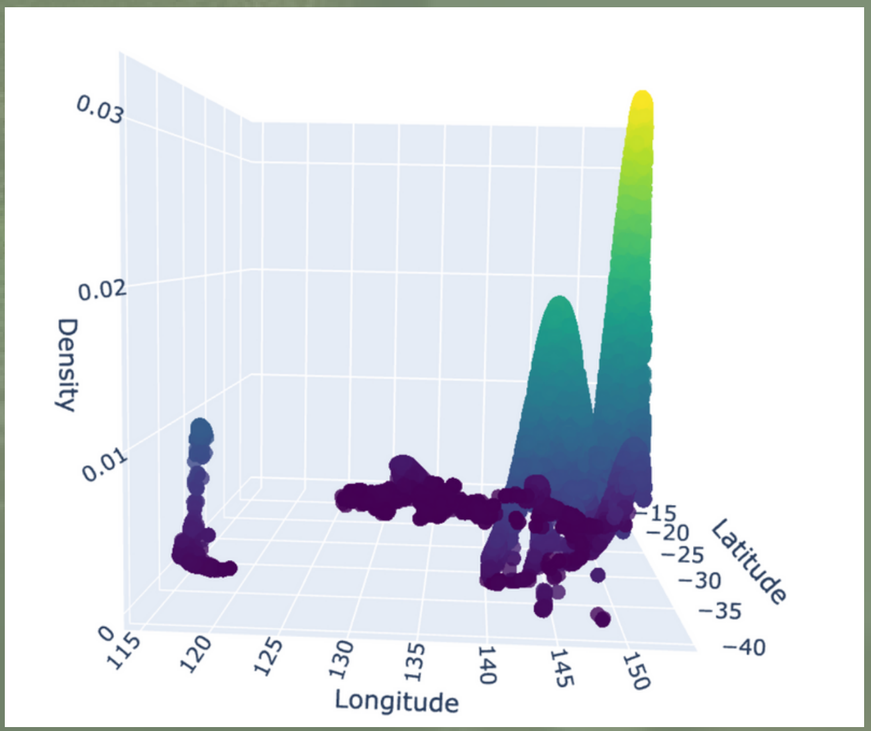


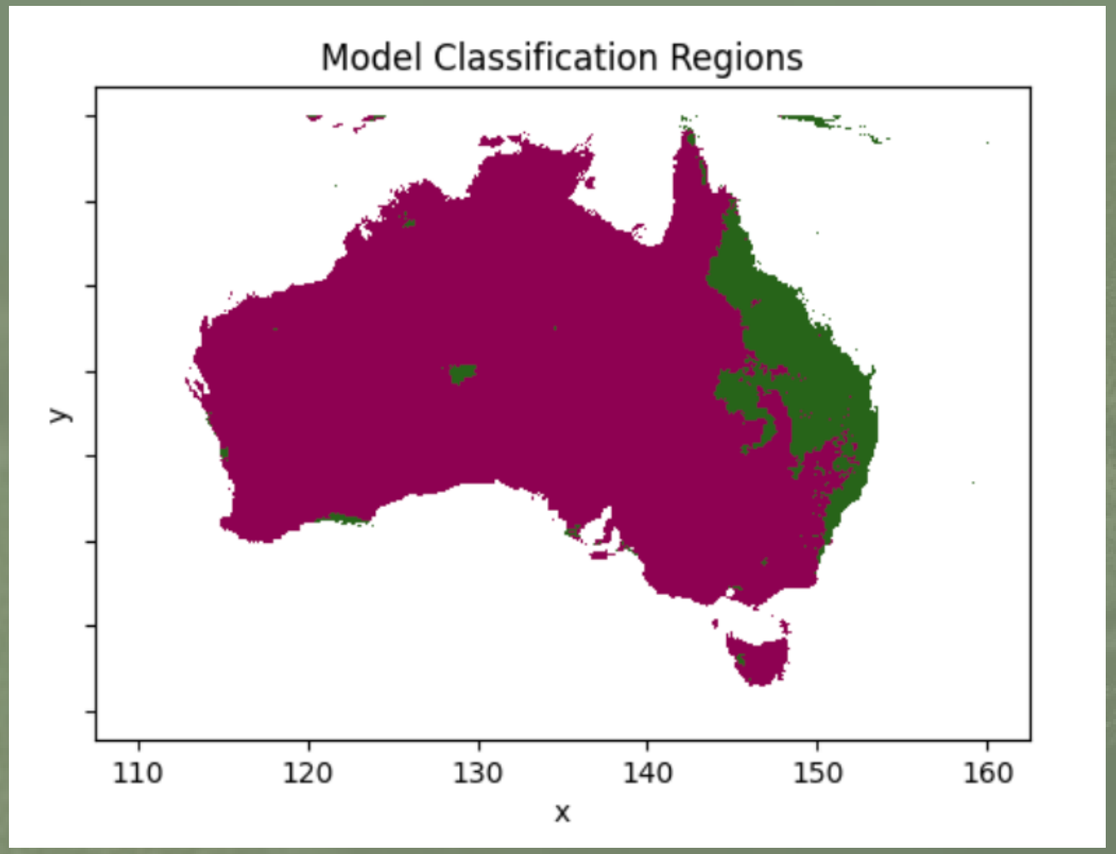
Amphibian chytrid fungus outbreak


Spatial distribution of Terraclimate variables
Litoria Fallax
tmax_mean
tmin_mean
ppt_mean
soil_mean
aet_mean
pet_mean
latitude [degrees_north]
longitude [degrees_east]
−10
−20
−30
−40
110
120
130
140
150
160
35
30
25
20
15
10
5
300
200
100
400
0
120
100
80
60
40
200
175
150
125
75


Density
0.03
0.02
0.01
0
Longitude
115
120
125
130
135
140
145
150
Latitude
−15
−20
−25
−30
−35
−40


Longitude
115
120
125
130
135
140
145
150
Density
0
0.01
0.02
0.03
Latitude
−15
−20
−25
−30
−35
−40


Density
0.03
0.02
0.01
0
Longitude
115
120
125
130
135
140
145
150
Latitude
−15
−20
−25
−30
−35
−40


Model Classification Regions
y
110
120
130
140
150
160
x


0.0
0.2
0.4
0.6
0.8
1.0

Findings

- Presence of separated populace (Melbourne)
- Eastern dwarf frogs are named so for a reason
- We now have a classification model that should aid in understanding frog population in non-urban areas which is a big area of focus

Model evaluation

- K-Fold cross-validation
- Random forest classifier with tuned hyperparameters
- F1 score and accuracy (submission score: 0.72)

Thank you!