Suspense at the Wildlife Preserve: “Cheep” Shots

Sammantha Firestone

COMP4449: Capstone | Midterm Project

Dr. Claudio Delrieux

Summer Quarter 2023

Contents

Suspense at the Wildlife Preserve: “Cheep” Shots

Scenario Description

At the Boonsong Lekagul Wildlife Preserve in March of 2018, investigations behind the decreasing population of the Rose-Crested Blue Pipit bird species have been underway by Mitch Vogel, post-doc student at Mistford College, and his professors. They are implicating Kasios Office Furniture Manufacturing firm in Mistford City to be responsible. They believe Kasios has been using banned substances in their manufacturing process and have been dumping it in the Southeast area in the preserve leading to the endangerment of the Blue Pipit and other bird species. Kasios claims they have done nothing wrong, and that Mitch Vogel and his professors are media-seekers and have flawed and biased research. Kasios backs up their claims by doing their own investigation and provide fifteen recordings of the Blue Pipit which the community of Mistford accepts.

Overall Goal

There are three main goals behind this investigation. The first goal is to utilize bird call collection metadata and the included map of the Wildlife Preserve to characterize the patterns of all the bird species in the Preserve over time and detect any trends or anomalies in the patterns. The second goal is to determine if the bird calls provided by Kasios support their claim that the Blue Pipits are flourishing by utilizing machine learning techniques. The last goal of this investigation is to formulate hypotheses concerning the state of the Rose-Crested Blue Pipit based on evidence collected throughout the analysis, as well as proposing next steps to be taken in the investigation to support or refute Kasios’ claim that the Blue Pipits are thriving across the Boonsong Lekagul Wildlife Preserve.

Data Exploration and Preparation

Data Description

The dataset consists of 2,081 audio recordings of bird calls and songs, compiled by a vetted Ornithology Group within the Wildlife Preserve. These recordings are in MP3 format and vary in duration. Accompanying the audio files is a metadata CSV file containing eight essential columns: File ID, English_name, Vocalization_type, Quality, Time, Date, X, Y. The File ID field serves as the index for the audio recordings, while the English_name field represents the common English name of the specific bird captured in each audio file. The Vocalization_type column specifies the type of bird sound present in the recording. The Quality field provides a score reflecting the qualitative measure of the bird sound, encompassing aspects such as purity and absence of background noise. The Time and Date columns denote the capture time of each recording. Lastly, the X and Y columns contain coordinates on the enclosed map, indicating the locations where the bird sounds were recorded.

Additionally, the dataset includes 15 bird recordings claimed by Kasios to be Blue Pipits within the Preserve. These recordings were obtained over the past few months. The metadata for these Kasios recordings specifies the respective locations on the map where the bird sounds were recorded.

Moreover, a 200 x 200 pixel map of the Lekagul Preserve is provided, offering a general representation of the site's layout, including indicative markings of roadways. Notably, the alleged dumping site for Kasios' waste products is centered around coordinates (148,40) on the map.

Metadata Cleaning & Audio Preparation

The metadata files for both the Ornithology recordings and Kasios recordings underwent a cleaning process before analysis. To enhance clarity, certain column names were revised: File ID became file_id, English_name was renamed bird_species, Vocalization_type was changed to vocalization_type, Time was updated to time, and Date was modified to date.

A visual representation of the distribution of vocalization types was generated, revealing that the majority of recordings belonged to call and song vocalization types. As outliers, unknown values, drumming, scold, and bill-snapping columns were removed from the dataset.

Next, the distribution of quality scores was plotted, demonstrating that a significant proportion of recordings received A and B quality ratings. To ensure only the highest quality audio data was utilized in the analysis, recordings with no quality score or scores below B were eliminated, thereby optimizing the performance of the audio classifier.

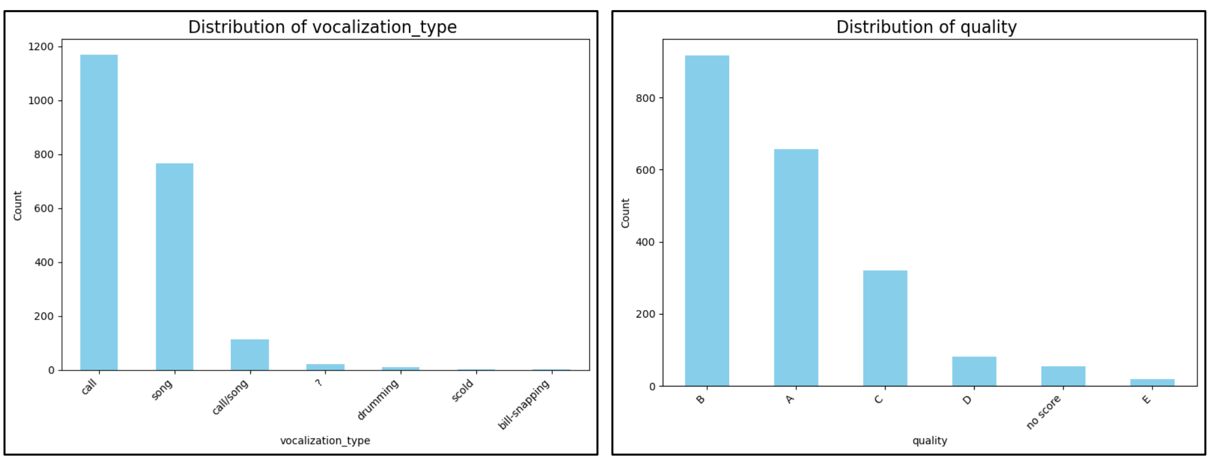

Figure 1: Distribution of vocalization type and quality

As the focus of the analysis centered around broader trends at a macro level of month and year, the time column was deemed unnecessary and subsequently removed from the metadata.

The date column underwent a cleaning process to ensure a standardized format of month, day, year. Furthermore, any rows lacking complete year or month information were excluded from the dataset, as this information is crucial to achieving the investigation's overall goals.

Finally, any rows with missing X and Y coordinate values were removed from the metadata since these coordinates play an essential role in the analysis.

The audio preparation process was primarily integrated into the pipeline alongside model building, and further elaboration on this preparation and cleaning is provided in the relevant section. Before this however, all audio data was converted from MP3 format to WAV, making use of the Python library Librosa easier. To improve data processing efficiency, the audio was transformed into mono.

Goal 1

Patterns of Bird Species in the Preserve

The initial part of this analysis focused on understanding the distribution of bird species over time within the Preserve. An animation was created to observe the variation in unique bird species counts across all recorded years. Additionally, time series plots were generated to visualize the count of each bird species over the entire timeframe, with specific attention given to the years 2005 to 2014 and 2011 and 2017.

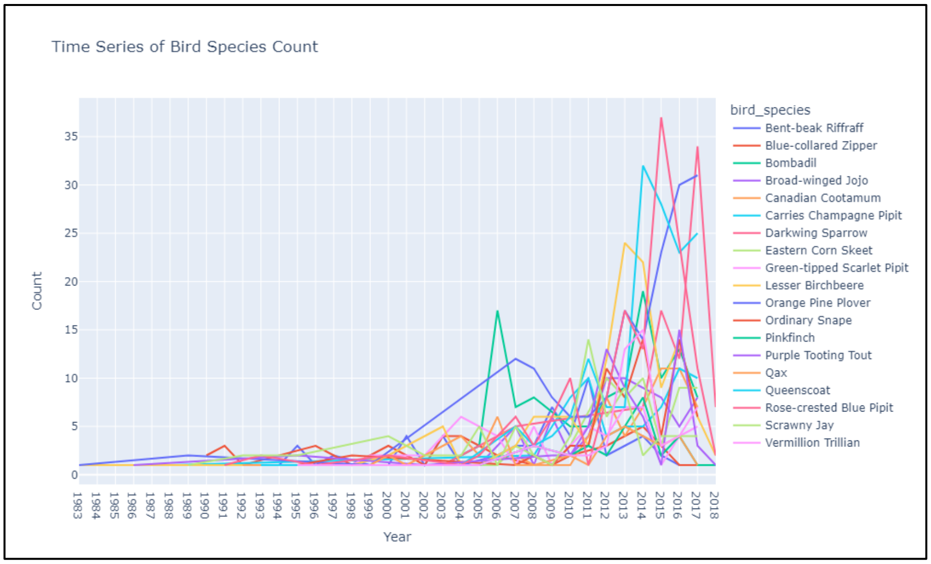

Figure 2: Time series of bird species count over time

The findings revealed a significant decline in the Blue Pipit population between 2015 and 2017. In 2015, there were 37 recorded instances of the Blue Pipit, but this number drastically reduced to only 11 in 2017, representing a substantial 70.27% decrease. Similar declining trends were observed in most bird species, except for the Darkwing Sparrow, which experienced an increase in recordings after 2016.

To better illustrate the count of Blue Pipit recordings per year, a bar plot was constructed. This visual representation emphasized that the number of recordings presented by Kasios in a few short months (fifteen recordings) is remarkably higher than the total counts for Blue Pipit calls in most years. Such an abnormal disparity raises doubts about the authenticity of the provided recordings and warrants a thorough examination of Kasios' claims.

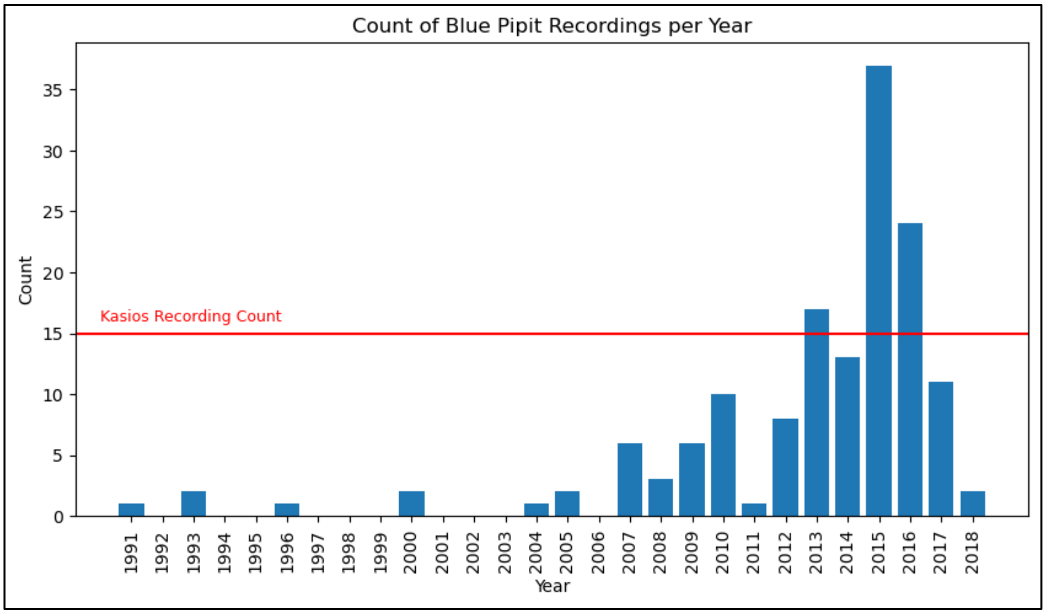

Figure 3: Count of Blue Pipit recordings per year

The next phase of this analysis delves into the congregation and migration patterns of birds within the preserve across time. The Lekagul map was utilized to plot the recording locations of different bird species, with the alleged dumping site marked as a red X, and the locations provided by Kasios represented as black points. Figure 4 presents the bird species recording locations spanning all years. While the plot contains numerous points, it reveals that the Rose-crested Blue Pipit and other bird species tend to remain in the same area. Specifically, Blue Pipit recordings are primarily concentrated in the Northeast corner of the Preserve. A few additional locations can be found on the westside (both North and South), but the density of recordings is lower there. Only one recording location was identified in the Southeast corner, where the alleged dumping site is situated.

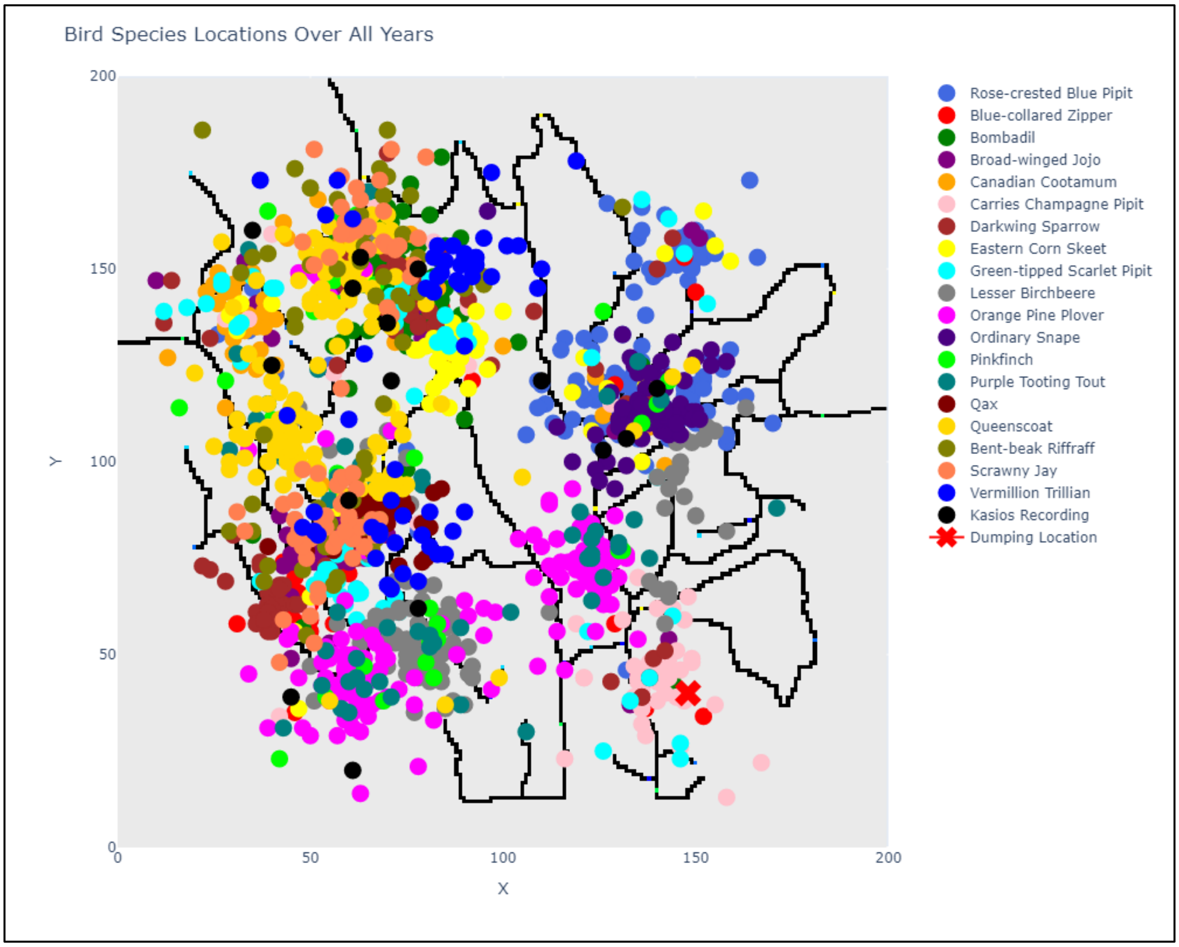

Figure 4: Bird species locations over all years

To gain deeper insights into congregation patterns over time, the same plot was created for each lustrum. Additionally, the plot was generated for the more recent three years (2016, 2017, and 2018), further highlighting the overall decline in bird species recordings during this period. Notably, certain bird species exhibited interesting behaviors. For instance, the Blue-collared Zipper bird predominantly resided in the Southwest corner of the map with a few outliers near the alleged dumping site between 2013 and 2016, and their presence ceased after 2017, except for one sighting in the Southeast corner. The Carries Champagne Pipit tended to congregate around the alleged dumping site, particularly evident in recent years (2017-2018), although there was a slight decrease in their population. The Eastern Corn Skeet population was concentrated in the western portion of the map, but no recordings of these birds were made after 2017, despite their significant presence in previous years.

In summary, for the year 2018, only two confirmed recordings of the Blue Pipit were found, while Kasios claimed to have recorded fifteen. Moreover, many of Kasios' reported recordings do not align with the Blue Pipits' congregation patterns; particularly their recordings observed in the deep Southwestern region of the Preserve, where Blue Pipits were not known to frequent. These disparities raise concerns and cast doubt on the credibility of Kasios' claims and the accuracy of their provided recordings.

Goal 2

Bird Call Audio Analysis

To gain a general perspective and make comparisons across bird species, box and whisker plots were employed to depict the duration, max amplitude, and max frequency of each audio recording. The duration plot showed that most audio files exceeded 30 seconds in length. The max amplitude plot indicated that the Rose-Crested Blue Pipit exhibited slightly higher maximum amplitude compared to other birds. As for Kasios' recordings, the max amplitude seemed to align, but the limited sample size hinders definitive conclusions.

To make further comparisons, a Blue Pipit song (file ID 130) and call (file ID 45) were loaded and played, contrasting them with Kasios' recordings. The song displayed distinct characteristics, featuring high-pitched sounds with a repeating pattern of notes. The call also shared similarities with the song, being high-pitched. However, Kasios' first provided sample appeared long, noisy, and with multiple bird calls, making it difficult to identify a Blue Pipit. The sixth recording from Kasios seemed less noisy, with only one or two distinguishable bird calls, but it repeated the same 30-second audio segment throughout the 2 minutes and 31 seconds, diverging from Lekagul's patterns.

To delve deeper into the audio data, various visualizations were implemented, including waveform plots, frequency-domain plots, spectrograms, and Mel-spectrograms. The waveform plot illustrated amplitude variations over time, while the frequency-domain plot showed the signal's amplitude in different frequency bands. The spectrogram depicted the signal's loudness over time at various frequencies, achieved by segmenting the time-domain signal and applying the fast Fourier transform (FFT) [1]. Finally, the Mel-spectrogram converted frequency bands to the mel scale, relating perceived frequency to actual measured frequency [2].

This suite of visualizations were plotted for fifteen samples of the Lekagul's Blue Pipit recordings, the fifteen samples provided by Kasios, and a single sample of all other bird species.

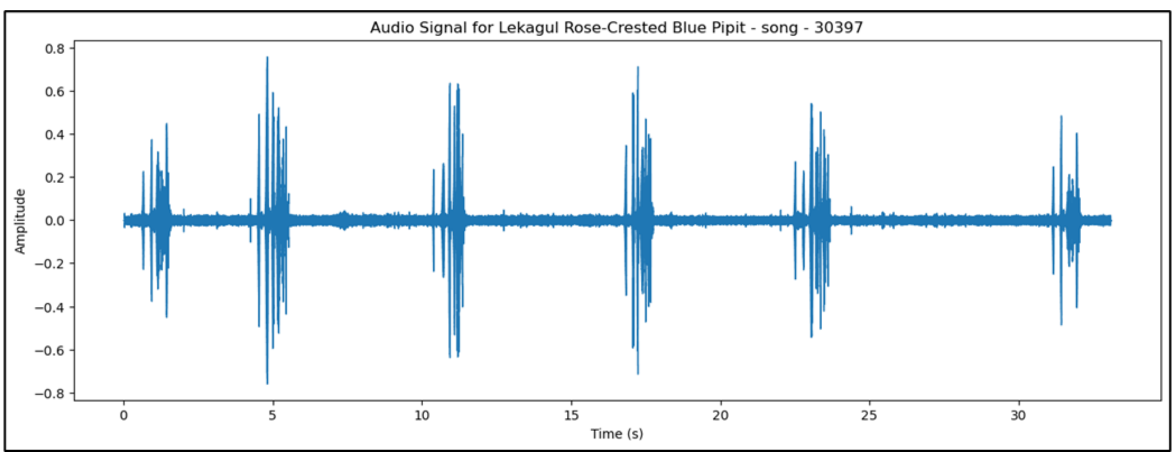

Figure 5: Waveform of Lekagul Blue Pipit Song

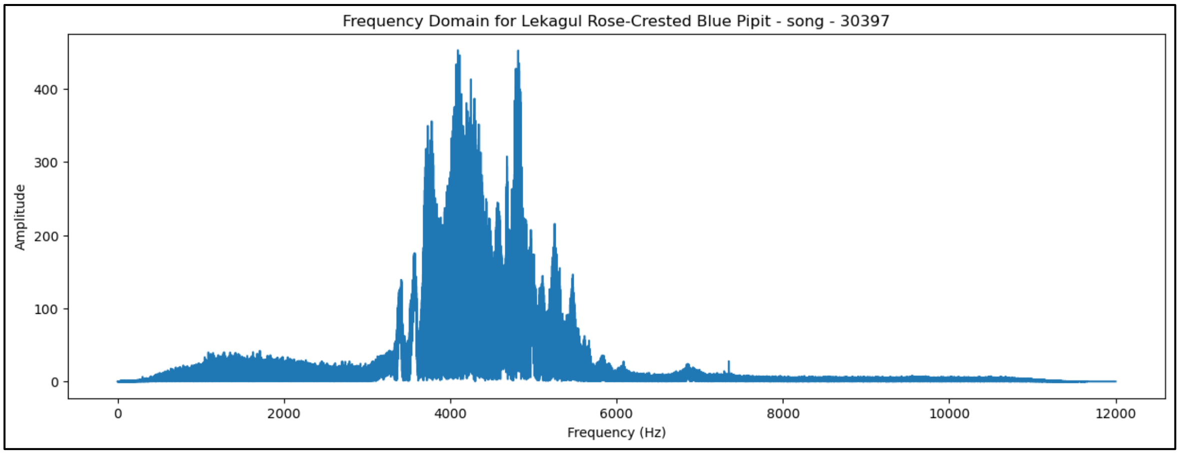

Figure 6: Frequency Domain plot of Lekagul Blue Pipit Song

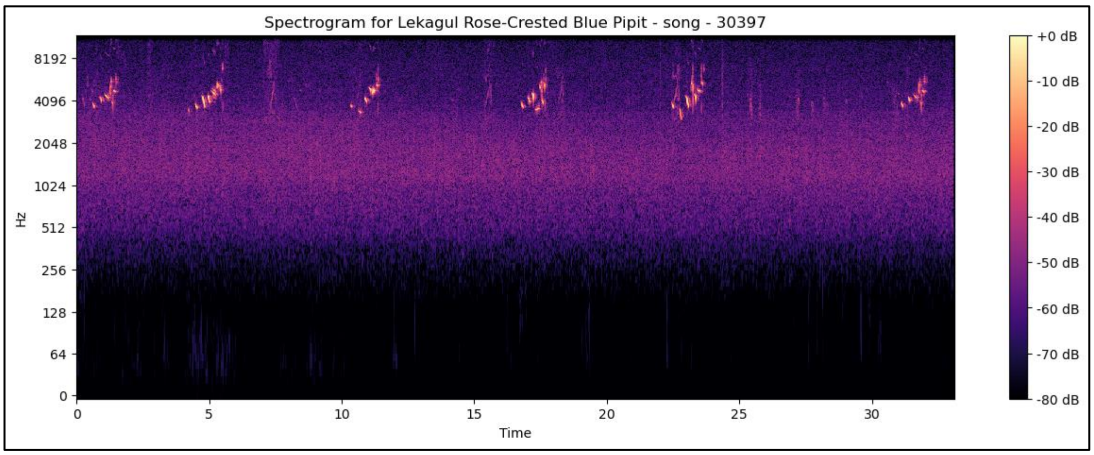

Figure 7: Spectrogram of Lekagul Blue Pipit Song

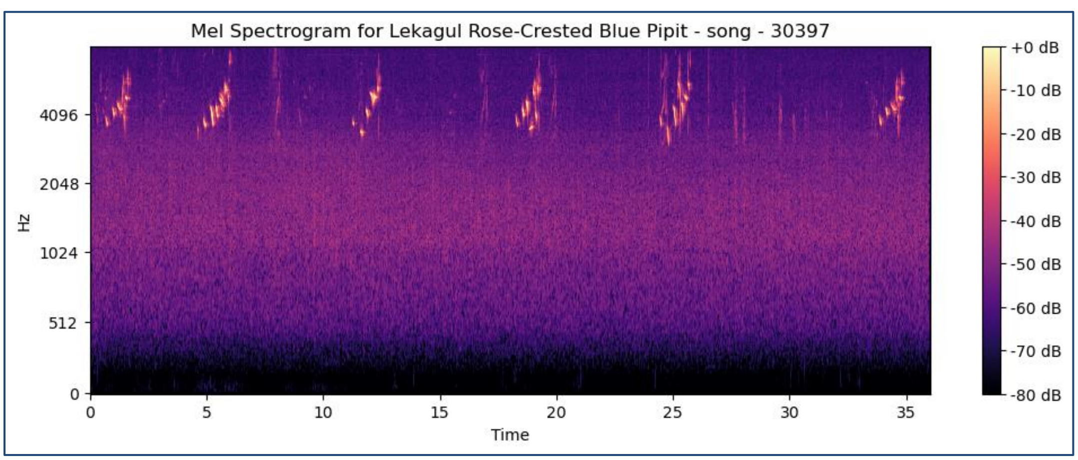

Figure 8: Mel Spectrogram of Lekagul Blue Pipit Song

The visualization suite shows the same characteristics as the audio file; the Blue Pipit typically has a high-frequency call and song around 4096 Hz, that repeats in short bursts.

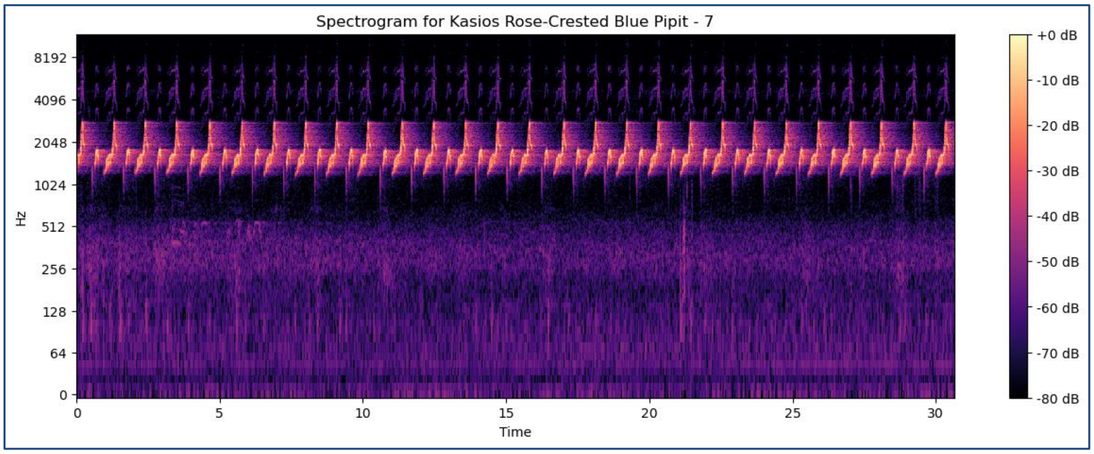

Figure 9: Spectrogram for Kasios' sample 7

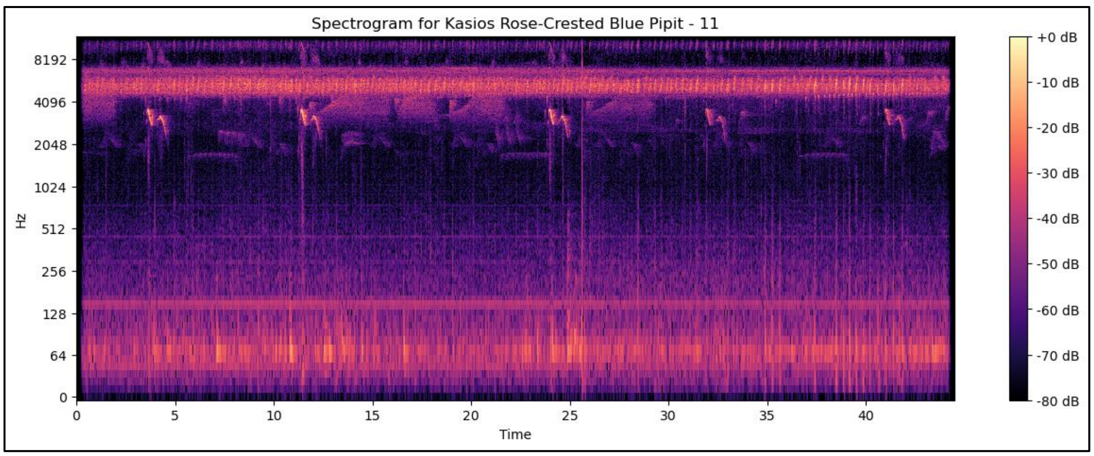

Figure 10: Spectrogram for Kasios' sample 11

In contrast, the visualizations for Kasios' recordings showcased inconsistencies and at times, noisy chirp patterns (Figure 10). Additionally, in eight of the fifteen recordings, the highest amplitude was observed around 2048 Hz (Figure 9), significantly lower than the amplitudes observed in Lekagul's recordings. Further notes and plot are available via the Jupyter Notebook.

CNN Binary Audio Classifier

To classify each audio clip provided by Kasios as a Rose-Crested Blue Pipit or not, a binary Convolutional Neural Network (CNN) model was developed. By converting the audio data into Mel spectrograms, the CNN could process the array of frequency, time, and decibels to identify specific patterns associated with a Blue Pipit call [3].

Prior to creating the Mel spectrograms, the audio data was segmented into fixed-size frames of 30 seconds. This duration was chosen because it offers sufficient information about the bird recording while minimizing the machine's memory load. For audio files shorter than 30 seconds, zero-padding was applied at the beginning and end to center the content. Conversely, for files longer than 30 seconds, the center of the audio was identified, and the 15 seconds before and after that center were considered.

During the Mel spectrogram computation, 128 Mel bands (Mel-frequency bins) were used to strike a balance between resolution and computation time. The audio signal was divided into short overlapping frames to capture its temporal dynamics, with 5 frames chosen to capture rapid changes in the audio signal while maintaining reasonable computation time [3]. This selection ensures efficient feature extraction for the CNN, enhancing its capacity to recognize distinct patterns indicative of a Blue Pipit call.

The target variable, "Is Blue Pipit," is significantly underrepresented, comprising only 150 audio files, while the non-target class, "Not Blue Pipit," contains 1,439 files. To address this class imbalance, random oversampling was applied to the target class, and random undersampling was performed on the non-target class. Following this preprocessing, both classes were balanced, each containing 500 files. Subsequently, Min-Max scaling was applied to all audio data, a crucial step for CNN models, as it aids in faster convergence and improved performance.

The CNN model architecture (Figure 11) consists of two convolutional layers, two max pooling layers, and two dense layers, with a dropout layer in between. The final layer employs a sigmoid activation function, which is appropriate for binary output [4].

	CNN Binary Classifier
Layer 1	**Conv2D (32 filters, kernel size 3x3, stride 1x1, 'same' padding, ReLU activation)**
Layer 2	**MaxPooling2D (pool size 2x2)**
Layer 3	**Conv2D (64 filters, kernel size 3x3, 'same' padding, ReLU activation)**
Layer 4	**MaxPooling2D (pool size 2x2)**
Layer 5	**Flatten**
Layer 6	**Dense (64 units, ReLU activation)**
Layer 7	**Dropout(0.5)**
Layer 8	**Dense (1 unit, sigmoid activation)**

Figure 11: CNN Binary Classifer Architecture

The Lekagul audio data was split into training and testing datasets, with 80% for training and 20% for testing. During training, the model underwent 15 epochs with a batch size of 32. Impressively, the model achieved an accuracy of approximately 98% by the 13th epoch, with a total loss of 0.0552 over the 15 epochs. Given that the model is used for binary classification, precision, recall, and F1 scores are particularly relevant. When predicting on the testing dataset, the model performed exceptionally well (Figure 12).

Model	*Precision*	*Recall*	*F1 Score*	*Accuracy*
Binary CNN Classifier	93.38%	96.95%	95.13%	94.58%

Figure 12: Test results for the Binary CNN Classifier

The model's performance on the testing data is remarkable, with a precision of 93.38%, indicating a high proportion of correct positive predictions among all positive predictions. Moreover, the recall score of 96.95% indicates that the model effectively captured a significant portion of actual positive instances, minimizing the number of false negatives (misclassifying Blue Pipits as non-target). This result is crucial for Kasios, as it maximizes the chances of identifying the Blue Pipit in the audio recordings they provided. In summary, the binary CNN classifier achieved an impressive overall F1 score of 95.13% and an accuracy of 94.58%, showcasing its efficacy in accurately distinguishing Blue Pipit calls from non-target calls.

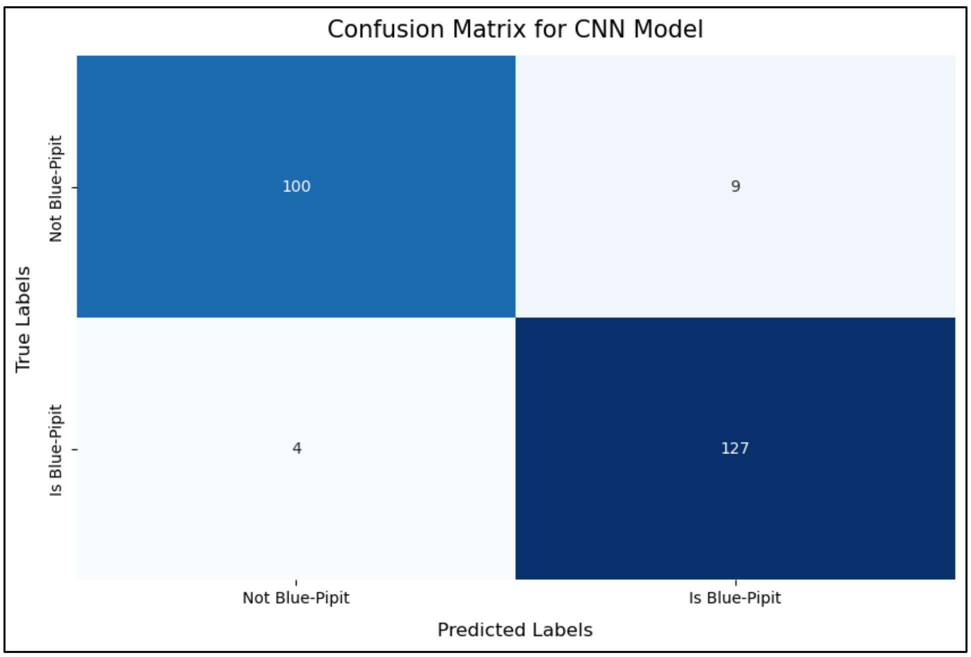

Figure 13: Confusion Matrix for testing data

When the model was applied to Kasios' audio data, it revealed that none of the provided recordings matched the Blue Pipit category. This outcome is consistent with the suspicions raised earlier during the audio visualization analysis, which showed inconsistencies and divergent patterns in Kasios' recordings. To ensure thorough and rigorous examination, additional modeling techniques were employed.

Additional Binary Classifier Models

To ensure thorough due diligence, common binary classifiers, including Logistic Regression, Random Forest, and Support Vector Machines, were implemented. As these classifiers do not handle image data like the CNN model, additional preprocessing steps were necessary. To extract features from the audio data, Mel-frequency cepstral coefficients (MFCCs) were chosen. MFCCs are widely used in speech recognition tasks and represent the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale frequency [5].

In this analysis, 16 Mel-spectral cepstral coefficients (MFCCs) were extracted, which is slightly higher than the typical 12 or 13 used in speech recognition [5]. This decision was made in the hope of potentially improving accuracy. This feature space reduction required addressing the imbalanced dataset. To tackle this issue, only random oversampling of the minority class "Is Blue Pipit" was applied, as our feature space had significantly decreased from the feature space used in the CNN model.

By applying these alternative classifiers and feature extraction techniques, the goal was to gain comprehensive insights and validation of the classification results, corroborating the CNN model's findings and providing further evidence for the lack of credibility of the audio data provided by Kasios.

Each of the models, Logistic Regression, Random Forest, and Support Vector Machine, underwent training using grid search to determine the best parameters for optimal performance. Comparing the Logistic Regression model to the CNN model, the model exhibited slightly lower performance. Despite this, it achieved an overall accuracy score of 82% on the testing data, indicating a substantial portion of audio samples were correctly classified. The precision-recall tradeoff favored recall, with 87.90% of Blue Pipit instances correctly identified as positive cases. The Support Vector Machine model outperformed the Logistic Regression model, achieving an overall accuracy score of 84.49% on the testing data, correctly classifying a majority of audio files. Once again, the precision-recall tradeoff favored recall, with a value of 95.02%.

Model		*Precision*	*Recall*	*F1 Score*	*Accuracy*
Logistic Regression	*Training*	76.03%	86.27%	80.83%	79.40%
	Testing	78.41%	87.90%	82.29%	82.29%
Random Forest	*Training*	99.83%	100.00%	99.91%	99.99%
	Testing	100.00%	100.00%	100.00%	100.00%
Support Vector Machine	*Training*	80.16%	92.49%	85.89%	84.71%
	Testing	84.49%	95.02%	89.45%	89.06%

Figure 14: Summary Table of Model Metrics

Among all models implemented, the Random Forest model demonstrated the highest metrics. It achieved 100% precision, recall, F1 score, and accuracy on the test dataset. When using these three models to predict whether the Kasios recordings were of the Blue Pipit, all models predicted that some amount of the recordings were not of the Blue Pipit. The Logistic Regression model predicted 8 out of the 15 recordings to be of the Blue Pipit, but this prediction should be taken with caution considering the lower precision and accuracy of the model. The Support Vector Machine model predicted 9 out of the 15 recordings to be of the Blue Pipit, though many of these recordings that were deemed to be the Blue Pipit are quite noisy, and the higher recall may have influenced this outcome in Kasios' favor. Finally, the Random Forest model predicted only 1 out of the 15 recordings (sample 5) to be of the Blue Pipit, which aligns well with the results obtained from the CNN model. This prediction would also be more reasonable if Kasios had provided only one recording of the Blue Pipit, as it matches the number of Blue Pipit calls recorded at the Lekagul Preserve.

Overall, the implementation of various models and techniques provided additional validation, reinforcing the initial findings of the CNN model and raising further doubts about the authenticity of Kasios' provided audio data.

Goal 3

Hypothesis and Next-Steps

Following a comprehensive analysis and investigation, significant findings have come to light. The overall bird population, including the Blue Pipit, has experienced a noticeable decline in the Northeast corner of the preserve over the last three years. Additionally, all bird species exhibit a tendency to congregate in specific locations within the preserve, and this is no exception for the Blue Pipit. However, the recording locations provided by Kasios appear irregular, particularly in the Southwest corner of the preserve. All binary classification models utilized in this analysis consistently indicate that the majority, if not all, of the audio recordings provided by Kasios do not match the Blue Pipit.

Given the analysis from both goals, it is clear that Kasios' is falsifying information. These findings raise the possibility that the Blue Pipit may either be an endangered species within the preserve or have migrated beyond its boundaries.

To advance the investigation, further audio cleaning techniques are recommended. By removing background noise, and lulled periods, the accuracy of the binary classification models can improve. Additionally, a Multiclass Classification model capable of identifying various bird species in the recordings provided by Kasios should be developed for a comprehensive examination. Lastly, Mitch and his professors should consider installing additional microphones on the Lekagul preserve's outskirts to monitor bird activity as it could shed light on the Blue Pipit's current status.

Incorporating these proposed steps will lead to a more comprehensive understanding of the Blue Pipit's situation and the overall bird population in the Boonsong Lekagul Wildlife Preserve.

Appendix

For further information and step-by-step analysis and code, please refer to the following repo:

https://github.com/sjmfirestone/comp-4449-final-cheepshots/tree/main

References

1. Vibration Research. (n.d.). What is a Spectrogram? Retrieved from https://vibrationresearch.com/blog/what-is-a-spectrogram/
2. Verlardo, V. Mel Spectrograms explained easily. Retrieved from https://www.youtube.com/watch?v=9GHCiiDLHQ4&t=1031s
3. Verlardo, V. Mel Spectrograms explained easily. Retrieved from https://www.youtube.com/watch?v=TdnVE5m3o_0&t=4s
4. Machine Learning Mastery. (n.d.). How to Choose an Activation Function for Deep Learning. Retrieved from https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning/
5. Verlardo, V. Extracting Mel-Frequency Cepstral Coefficents with Python. Retrieved from https://www.youtube.com/watch?v=WJI-17MNpdE&list=PL-wATfeyAMNqIee7cH3q1bh4QJFAaeNv0&index=20&pp=iAQB