Taylor Woodington

IUCN Red List most Endangered Species Patterns: White Paper

1. Business Problem

This project will analyze patterns of endangerment through the classification established globally from the IUCN Red List. Many animals have been on this list and for a long time struggling to have a healthy and stable population. The patterns being explored will be across different taxonomic groups to see what groups are at the most risk compared to which are at the least risk. The questions being posed are which groups are high in count across all the endangered categories and which groups are the highest count at the most severe category. This will help states and countries know what groups to focus conservation efforts towards. In addition, this could help support campaigns and organizations advertised to the public and provide reasons for the public to support them.

1. Background/History

The dataset being used for analysis is from Kaggle and includes information about different species classified under the IUCN Red List as endangered and more. The data within the file includes taxonomic groups as well as different conservation status categories for extinct species, extinct in the wild, critically endangered, endangered, and vulnerable, going from the most extreme to least extreme status. This allows for some kind of biodiversity risks that may be present amongst different species. The dataset will allow for different visuals focusing on specific species as well as categories.

1. Data Preparation

In order to analyze this data, I want to make sure the dataset is organized and clean for use. First, I check the data to make sure it is loaded properly and from there I remove the missing values in the data. The reason for missing data removal is to minimize the bias that the missing values would cause in the data analysis. In addition to this I decided to remove the “DD” column in the dataset, which is a column for “data deficiency”, meaning that there were counts that are not categorized under any of the columns. This could sway the data to be more biased away from the categories (such as endangered and extinct) that we are focusing on because it includes so many different species that may not be the ones in need. Following these removals, I converted numeric columns to numeric values for easier manipulation of the data for analysis.

1. Methods and Analysis

Once the data was prepped, I decided to visualize it through a bar graph and tree map before modeling. The first bar graph as shown in Appendix A, shows that the highest count for taxonomic groups in conservation categories is “LC or LR/lc” which is the column for species that are Least Concern. This is good that the majority of the species fall into this category with a count above 2,000. However, the following two categories with counts above 500 were “VU” (vulnerable) and “EN” (endangered). While this may seem low it is a high number of taxonomic groups, and some groups might be at more risk than others. To analyze this, I made a tree map which focuses on the top ten taxonomic groups with the highest endangered species as shown in figure A.2. In this tree map we can see that “Polypodiopsida” and “Anthozoa” are the taxonomic groups with the highest risk. These groups cover different ferns and sea corals which can be essential to the world for taking carbon dioxide and producing oxygen.

With this important information at hand, I wanted to make a model that would assist in identifying if a species was “high risk” meaning endangered or critically endangered to give states points of focus in their wildlife management projects. After training this dataset and running it, I found that the accuracy of the model was 92%. This means that the model is highly accurate and can predict whether or not species are at high risk.

1. Limitations and Challenges

While this data is useful it can pose some limitations. The possibility of missing data could create bias or inconsistencies in the analysis. As well as potential human error or classification errors in the data itself. One of the largest challenges in place is that the status of different species changes consistently which makes it more difficult to keep up with and understand the dynamics of conservation efforts in place. This can be combatted by clear communication of when this data is from and stay up to date by checking it often. Taxonomic groups are an overarching umbrella for species names which change consistently in scientific literature. Biologists and data scientists need to make sure the data is accurate and up to date with changes in this or statuses.

1. Future Uses and Recommendations

In the future, I think it would be important to break down those taxonomic groups that are labeled the class name into different rows for genus names to see which specific species may need more wildlife management and attention. This can allow for more detailed management plans to help increase a species population but make sure it is a healthy stable ratio and not just an inflow of the species that were of least concern but in the high-risk taxonomic group.

1. Implementation Plan

Knowing that a model with high confidence could detect species of high risk, each states wildlife commission should identify species in their areas that are of high risk. Upon identifying those species, biologists and habitat managers can look at the habitats which those species are living in the areas under their care and test water, soil, air quality, and more to find parts of the habitat that need to be improved. Once those parts are understood, there can be better management over the area to increase the species population within that state and provide relief to the status of the species overall.

1. Ethical Assessment

When analyzing this data, it is incredibly important to communicate clearly about what the different categories are, and taxonomic groups involved. If not done clearly it can create a misconception on how the IUCN Red List works or imply improper management of the species groups that are involved in the data. While some groups may have many species threatened, it does not mean that there are no conservation efforts in place to care for them. It is important to consider and note that the data involved in this analysis does not label the care or conservation that is currently happening but provides a group to focus on. Providing this focus can allow for alterations or additional efforts to be made in the future but does not state them now. Clear communication on this will limit the ethical challenges.

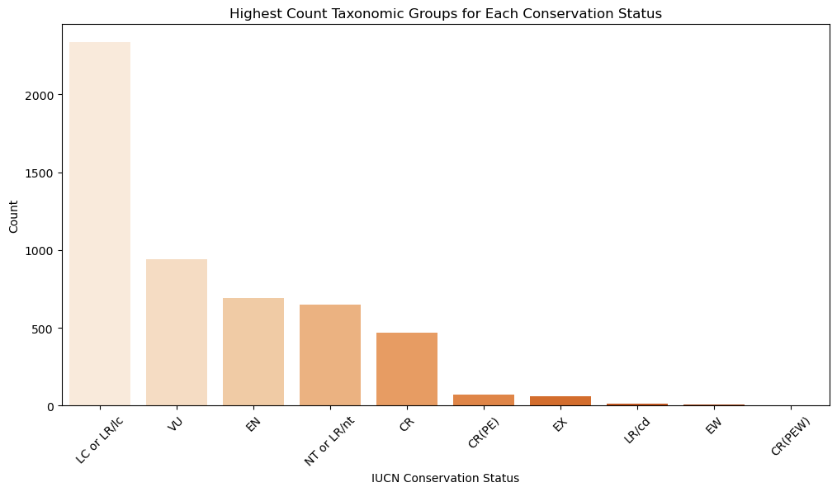
1. Audience Questions
2. What is the lowest count of taxonomic groups for conservation status?
   1. The lowest count of conservation status for taxonomic groups is Endangered in the Wild.
3. What does total threatened groups look like versus least concern?
   1. Total Threatened (Critically Endangered, Endangered, and Vulnerable) is around 2,100 whereas the total of least concern is around 2,330. Showing that the total amount of species threatened is very close to those of least concern and that these species do need attention.
4. What are some of the other taxonomic groups with high risk?
   1. Other taxonomic groups with high risk are Arachnida (spiders, scorpions, etc.) and Bivalvia (clams, oysters, etc.) as well as decomposers from the Lecanoromycetes group (meaning fungi) and Diplopoda (Millipede)
5. Which different species are in the class under the most risk from this dataset?
   1. Some different species of fern under this group are Angiopteris evecta (King Fern), Platycerium bifurcatum (Elkhorn Fern), and Asplenium australasicum (Birds Nest Fern).
6. What is the count of endangered species for the class at the bottom of the top ten compared to the class at the top of the top ten taxonomic groups of high risk?
   1. The count of endangered species for the class at the bottom of the top ten is 55 species whereas the class at the top ten has 238 listed.
7. What is the importance of the decomposers at risk? What could happen if they were not protected?
   1. Decomposers are extremely important when it comes to the environment. These are the organisms which break down all kinds of unwanted natural material to ensure a clean ecosystem, halting the spread of disease and continuing the flow for the environment. If they were not protected, an increase of disease could exist in the habitat from spreading of disease or an increase of infection and bacteria from the debris pile up.
8. What are actions that could directly help manage the habitat of at-risk species?
   1. Actions could be planting more species of plants that will be beneficial interacting or supporting the species that need the most help.
9. How many total critically endangered and endangered species are there all together?
   1. There are 1,163 total of critically endangered and endangered species.
10. Should we focus at all on species of least concern?
    1. There should be attention or focus on some species with least concern to see how they interact with species of high concern and if they have a carrying affect on them.
11. What number quantifies a species as endangered?
    1. What will determine a species listed status is a combination of their population size, rate of decline, geographic distribution and the degree of population and distribution fragmentation.
12. References

Harshith, J. (2021, November 29). *Number of species in each IUCN Red List Category*. Kaggle. <https://www.kaggle.com/datasets/johnharshith/number-of-species-in-each-iucn-red-list-category?resource=download>

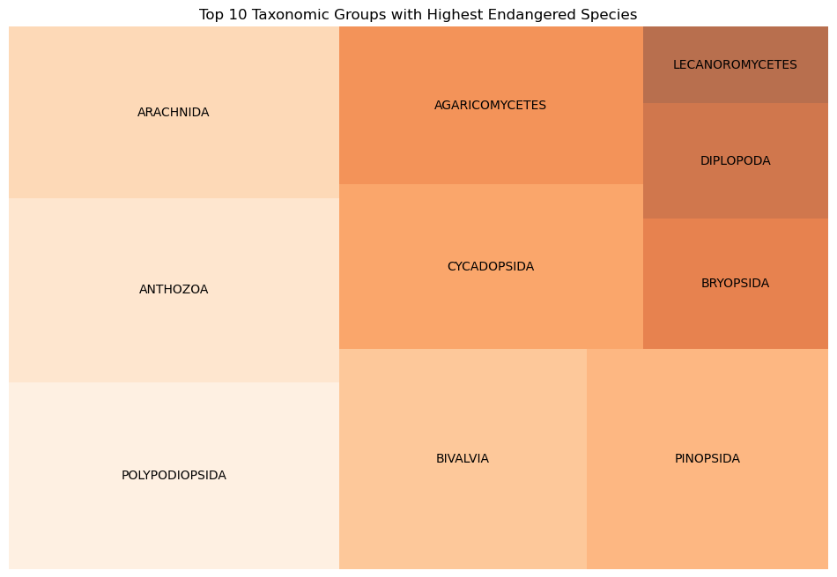
*The IUCN Red List of Threatened Species*. IUCN Red List of Threatened Species. (n.d.). <https://www.iucnredlist.org/>

1. Appendix A

A.1 Count of Taxonomic Groups for Each Status



A.2 Top 10 Taxonomic Groups with Highest Count of Endangered Species



A.3 Species Classification High-Risk Confusion Matrix

