

Biodiversity for the National Parks


Author - Anrudh Kumar
Codecademy Data Analysis

Introduction

- This project is based on analyzing the data on the conservation statuses of the species present in Biodiversity and to investigate if there are any patterns or themes to the types of species that become endangered.



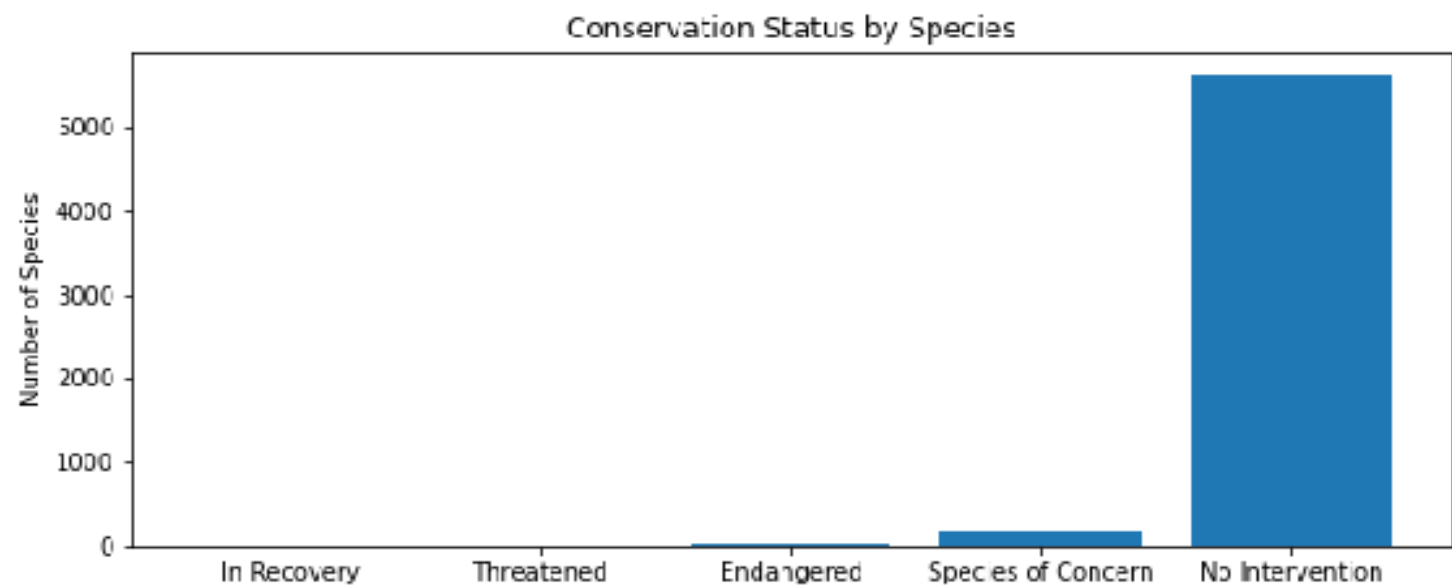
Species Information

- Our species_info.csv contains data on 5,541 unique species across various national parks.
 - Along with the scientific name of the species, we have included the common name, category, and conservation status
 - There are 7 categories of species: 'Mammal', 'Bird', 'Reptile', 'Amphibian', 'Fish', 'Vascular Plant', and 'Nonvascular Plant'
 - There are 5 conservation statuses: 'Species of Concern', 'Endangered', 'Threatened', 'In Recovery', and 'No Intervention' (status for species that had no conservation status)
- 


Conservation Status by Species

- Below, we've counted the number of species by their conservation status, and sorted them from least to greatest
- As you can see, most of our species require no protection (5,633), and 4 are in recovery
- We'll have to keep a careful eye on 151 species that may be in need of conservation (species of concern), and 25 species that are dangerously close to extinction (threatened or endangered)

	conservation_status	scientific_name	
1	In Recovery	4	
4	Threatened	10	
0	Endangered	15	
3	Species of Concern	151	
2	No Intervention	5363	




Exploring Category of Species

- Are there certain types of species more likely to be endangered?
 - To answer the above question, we created a pivot table (see next slide) and grouped species by their category and their protection status
 - 'not_protected' shows the count of unique species that do not require intervention (has a 'No Intervention' status)
 - 'protected' column displays count of unique species that have a conservation status not equal to 'No Intervention'
- 

Species by Category and Protected Status

	category	not_protected	protected	percent_protected
0	Amphibian	72	7	0.088608
1	Bird	413	75	0.153689
2	Fish	115	11	0.087302
3	Mammal	146	30	0.170455
4	Nonvascular Plant	328	5	0.015015
5	Reptile	73	5	0.064103
6	Vascular Plant	4216	46	0.010793

Comparing Species Categories

- How can we determine if there is a significant difference between 2 categories of species and their protection status data?
 - For example, ~17% of species in the Mammal category are protected, compared to the ~15% of Bird species protected
 - Is this difference between mammal and bird significant?
 - To compare differences in categorical data (which we have with our species categories data), we conducted two chi squared tests
- 


Comparing Mammal and Bird

- Null hypothesis: There is no significant difference between the mammal dataset and the bird dataset
 - To reject the null hypothesis, we will look for a p-value of less than 0.05
- After creating a contingency table and using the `chi2_contingency()` function from `scipy.stats`, a p-value of 0.686 was returned, greater than 0.05.
- This means we can't reject our null hypothesis and there isn't a significant difference between Mammal and Bird!


Comparing Mammal and Reptile

- Is the difference between Reptile (6% protected) and Mammal (17% protected) significant?
- Null hypothesis: There is no significant difference between Mammal dataset and Reptile dataset
 - Reject the null hypothesis if p-value is less than 0.05
- With a new contingency table and using the `chi2_contingency()` function, a p-value of 0.038 was returned, less than 0.05.
- We CAN reject our null hypothesis! There IS a significant difference between Mammal and Reptile!

Recommendations of Endangered Species

- While mammal and bird categories may not have a significant difference between each other, they are still the top two categories of species that are more likely to become endangered than the other categories.
 - This is determined by looking at percent_protected values: The higher this value, the more likely the species is to become endangered
 - Vascular and nonvascular plant species are least likely to become distinct.
- 

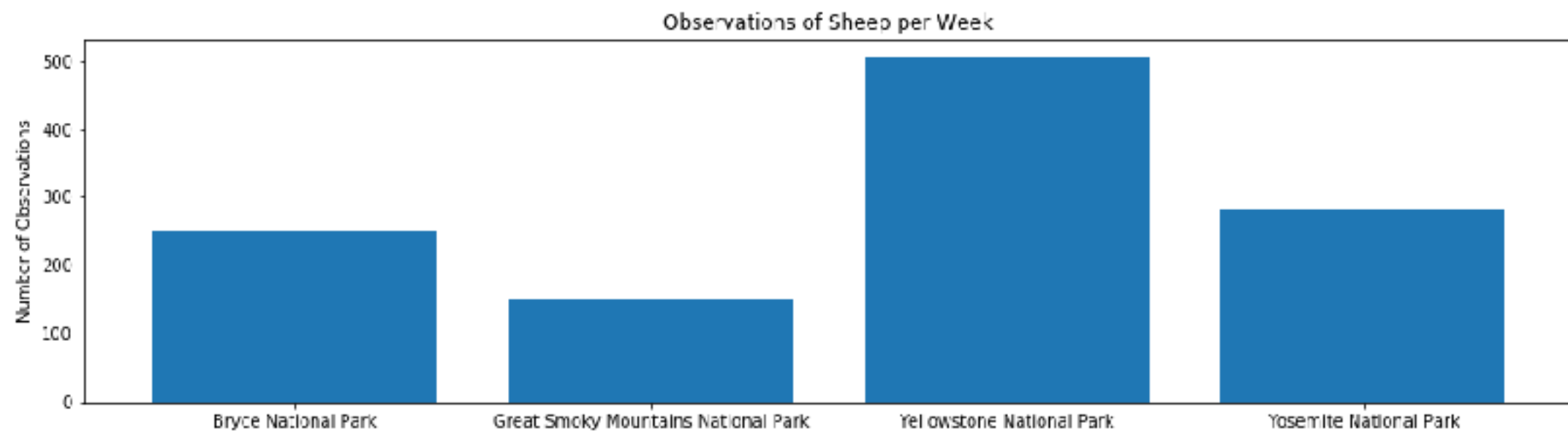
Sheep Observations

- Our observations.csv file contains recorded sightings of different species at several national parks over the past 7 days
 - We've added to our species data (discussed in the previous slides) an 'is_sheep' column and filtered the data to only include mammal sheep species
 - Our data contains 3 sheep species: Domestic sheep, Bighorn sheep, and Sierra Nevada Bighorn sheep
- 


- Our observations data was then merged with our species data.
- Below, we found the total sheep observations across the 3 species, grouped by the national park

		park_name	observations
0		Bryce National Park	250
1	Great Smoky Mountains National Park		149
2	Yellowstone National Park		507
3	Yosemite National Park		282


- Below is the data from the previous slide's dataframe as a
- bar chart.
 - Yellowstone National Park saw the most sheep observations (507 sightings) out of the four parks we have data.



Foot and Mouth Disease Study Amongst Parks' Sheep

- 15% of sheep at Bryce National Park have foot and mouth disease.
 - Park rangers at Yellowstone National Park have been running a program to reduce the rate of foot and mouth disease at that park.
 - We want to find out whether or not this program is working (A/B Test), and to detect reductions of at least 5 percentage points.
- 

Finding the Sample Size for our A/B Test

- We used Optimizely to find the sample size we needed for each variation, or the number of sheep observations that needed to be made at each park
 - For the calculator:
 - Baseline conversation rate: 15%
 - Minimum Detectable Effect: 33.33%
 - Statistical Difference: 90%
- 

Sheep Studies Conclusions

- Optimizely calculated a sample size of 510 per variation (510 sheep for both Bryce and Yellowstone)
- To observe enough sheep, we would need about 2 weeks at Bryce National Park ($510/250 = 2.04$ weeks) and about 1 week at Yellowstone National Park (1.0059 weeks)

	park_name	observations
0	Bryce National Park	250
1	Great Smoky Mountains National Park	149
2	Yellowstone National Park	507
3	Yosemite National Park	282

Thank You

