Endangered Species Analysis for National Parks

Xing Lu

Available Data

The following table is an example of the available data:

category	scientific_name	common_names	conservation_status
Mammal	Clethrionomys gapperi gapperi	Gapper's Red-Backed Vole	NaN
Mammal	Bos bison	American Bison, Bison	NaN
Mammal	Bos taurus	Aurochs, Aurochs, Domestic Cattle (Feral), Dom...	NaN
Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	NaN
Mammal	Cervus elaphus	Wapiti Or Elk	NaN

We have a table that gives us the category, scientific name, common names, and conservation status of different species.

Looking into Data

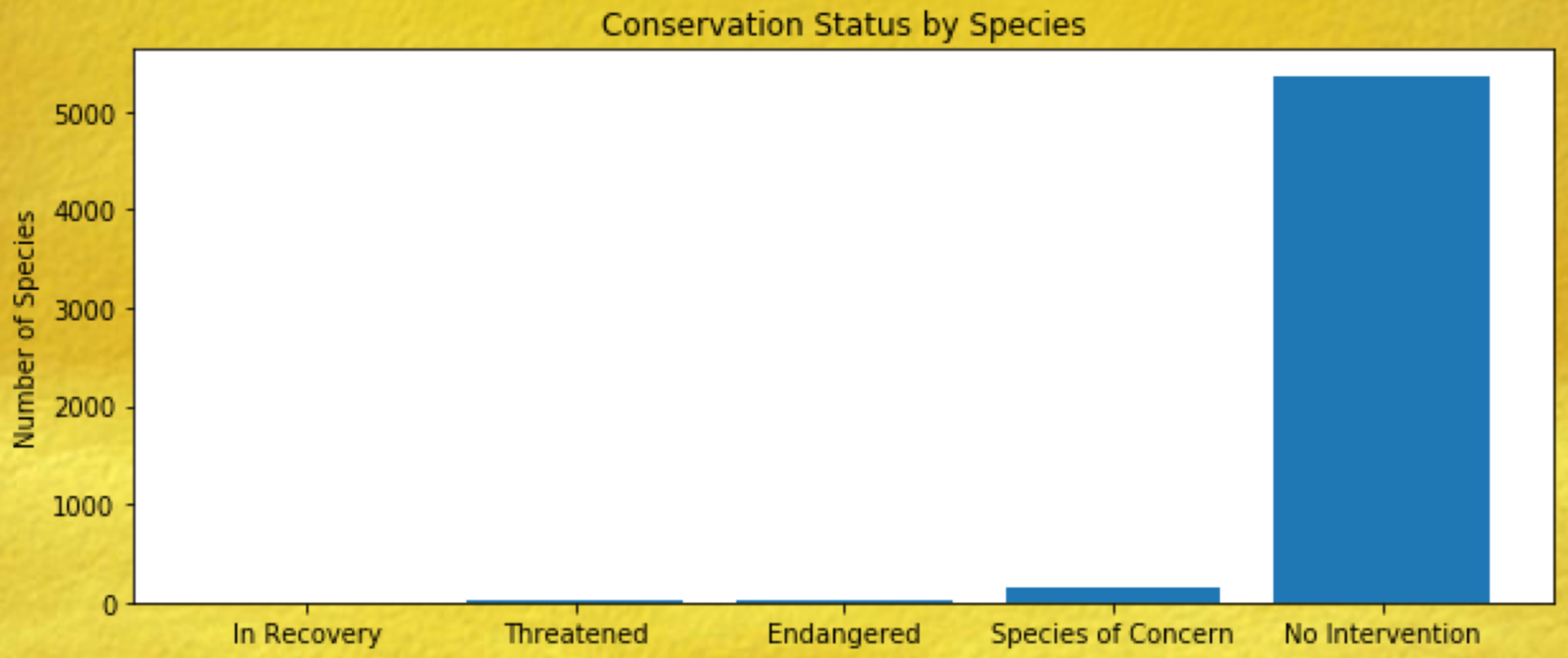
- Number of different species: 5541
- List of different categories:
  - Mammal
  - Bird
  - Reptile
  - Amphibian
  - Fish
  - Vascular Plant
  - Nonvascular Plant

Conservation Status

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

This table shows us the conservation status of different species. No intervention means the species do not need protection.

Conservation Status



As the plot shows, most of the species do not need protection. All other labels are considered as “endangered” in the following analysis.

Endangered Species

The table shows the number and the percent of species that need protection in each categories:

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

Bird and mammal are the two categories that need the most attention.

Significance Test and Recommendation

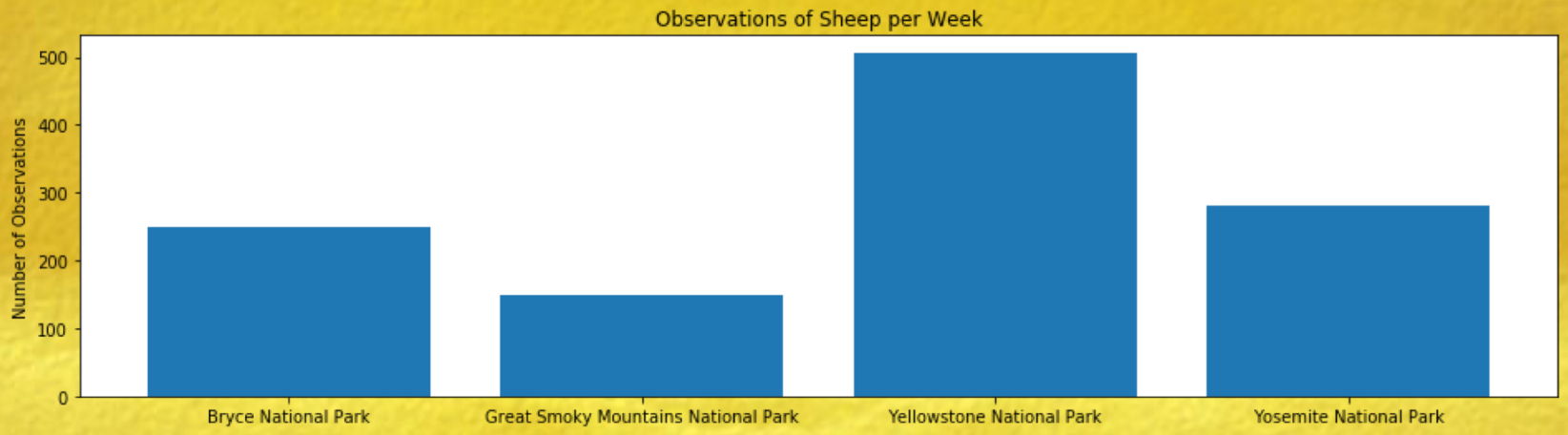
We performed two chi square tests:

1. To see if species in category mammal are more likely to be endangered than species in bird.
    p-value = 0.69
    The answer is no.

2. To see if species in category mammal are more likely to be endangered than species in reptile.
    p-value = 0.04
    The answer is yes!

Based on the significance test, we should put more attention on mammal and bird.

Observations of Sheep



The chart shows number of sheep observed in different parks per week.

Sample Size Determination

Our scientists know that 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. The scientists want to test whether or not this program is working. They want to be able to detect reductions of at least 5 percentage point. For instance, if 10% of sheep in Yellowstone have foot and mouth disease, they'd like to be able to know this, with confidence.

Use the sample size calculator at Optimizely to calculate the number of sheep that they would need to observe from each park. Use the default level of significance (90%).

Baseline Conversion Rate: 15%

Minimum Detectable Effect: 5%/15% = 33.33%

Sample Size: 510