codecademy

Biodiversity in National Parks

Cletus Norton
Oct. 14 2020

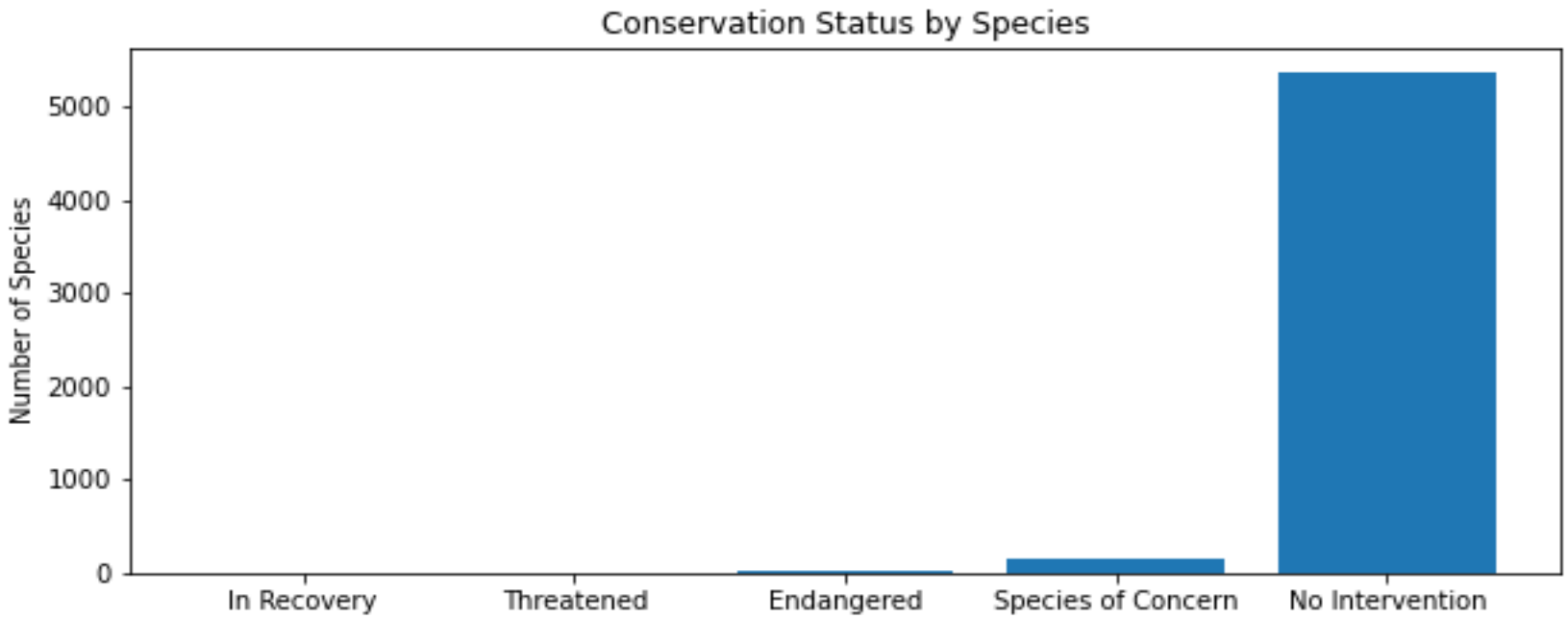
1.1 Information about the data

This is a capstone project for the data science path, the provided the data for this project is fictional. The idea was to act as a data analyst for the National Parks Services, with preset questions to be answered.

The data provided contains 5541 unique species ranging from plants to animals. Plants(4,803) were the most commonly counted species, with birds (521) and mammals(214) being the dominent species of animals. From this information I was able to check on conservation status, however this data didn't provide information about which national park the counts were made in.

1.2 Conservation Status

- 5363 Species with no intervention
- 151 species of concern
- 10 threatened
- 15 Endangered
- 4 in Recovery



The previous chart shows the majority of the species in the park do not need intervention. While it also points out that we have more concerns that we have recoveries.

1.3 Protection Status

- This provides a breakdown of the species type in protection.
- Birds and mammals have the highest percentage of species in protected status
- Note, that vascular plants have a higher quantity in protected status, but lower percentage because of the abondance of species counted.

category	not_protected	protected	percent_protected
Amphibian	72	7	8.860759
Bird	413	75	15.368852
Fish	115	11	8.730159
Mammal	146	30	17.045455
Nonvascular Plant	328	5	1.501502
Reptile	73	5	6.410256
Vascular Plant	4216	46	1.079305

On first glance, the data suggests that birds are more endangered than mammals (Pvalue of 0.68). When checked in a Chi squared test, the significance did not prove out, but it did when we compared reptiles to mammals (Pvalue = 0.03).

1.4 Sheep Observations in the National Parks

- The data below is based on observations made at the parks over a seven day period.
- We are able to this data and associate it with the park where the sheep were observed.
- We were then able to estimate observation rates, to assist in the research into foot and mouth disease.

scientific_name	park_name	observations	category	common_names	conservation_status	is_protected
Ovis canadensis	Yellowstone National Park	219	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True
Ovis canadensis	Bryce National Park	109	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True
Ovis canadensis	Yosemite National Park	117	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True
Ovis canadensis	Great Smoky Mountains National Park	48	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True
Ovis canadensis sierrae	Yellowstone National Park	67	Mammal	Sierra Nevada Bighorn Sheep	Endangered	True
Ovis canadensis sierrae	Yosemite National Park	39	Mammal	Sierra Nevada Bighorn Sheep	Endangered	True
Ovis canadensis sierrae	Bryce National Park	22	Mammal	Sierra Nevada Bighorn Sheep	Endangered	True
Ovis canadensis sierrae	Great Smoky Mountains National Park	25	Mammal	Sierra Nevada Bighorn Sheep	Endangered	True
Ovis aries	Yosemite National Park	126	Mammal	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False
Ovis aries	Great Smoky Mountains National Park	76	Mammal	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False
Ovis aries	Bryce National Park	119	Mammal	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False
Ovis aries	Yellowstone National Park	221	Mammal	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False

1.5 Sheep Observations in the National Parks Bar plot

- The chart below gives us a weekly observation rate in each park for the foot and mouth disease program.
- With 15% infection rate at Bryce and Yellowstone running a prevention program. We were able to figure out sample size.
- To be able to acheieve the minimum detechtable effect, we would need a sample size of 890.
- We will need about 3.5 weeks to reach this sample size at Bryce and 1.5 weeks at Yellowstone.