Biodiversity Capstone Project

Henry Kim
trust0219@gmail.com

Take a glimpse into species_info.csv

As observed first 5 rows of dataframe for species_info.csv, we can some information about animals. Columns were category, scientific_name, common_names, and conservation status.

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

Some facts about our data for species

- Species contain data of 5541 different species.
- Species consist of 7 different types (categories) of animals. The types are Mammal, Bird, Reptile, Amphibian, Fish, Vascular Plant, and Nonvascular Plant.
- Also species can be categorized with respect to 5 conservation status. The status are unknown, Species of Concern, Endangered, Threatened and In Recovery.

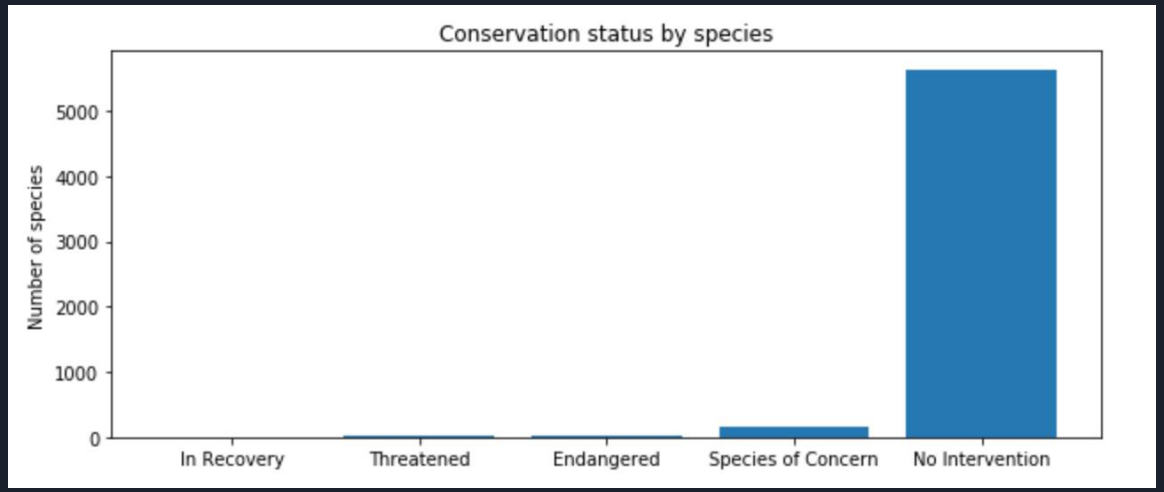
In terms of conservation status...

We can see there are 15 species endangered, 4 species in recovery, 151 species of concern, and 10 species threatened, and 5363 of N/A.

	conservation_status	scientific_name
1	In Recovery	4
4	Threatened	10
0	Endangered	16
3	Species of Concern	161
2	No Intervention	5633

Conservation Status by Species

As we can see from the bar graph, majority of species belong to “No Intervention”.



Current protection status of each category of species

Generally, all the species for each category are currently protected less than 20%. In other words, they all need more protection.

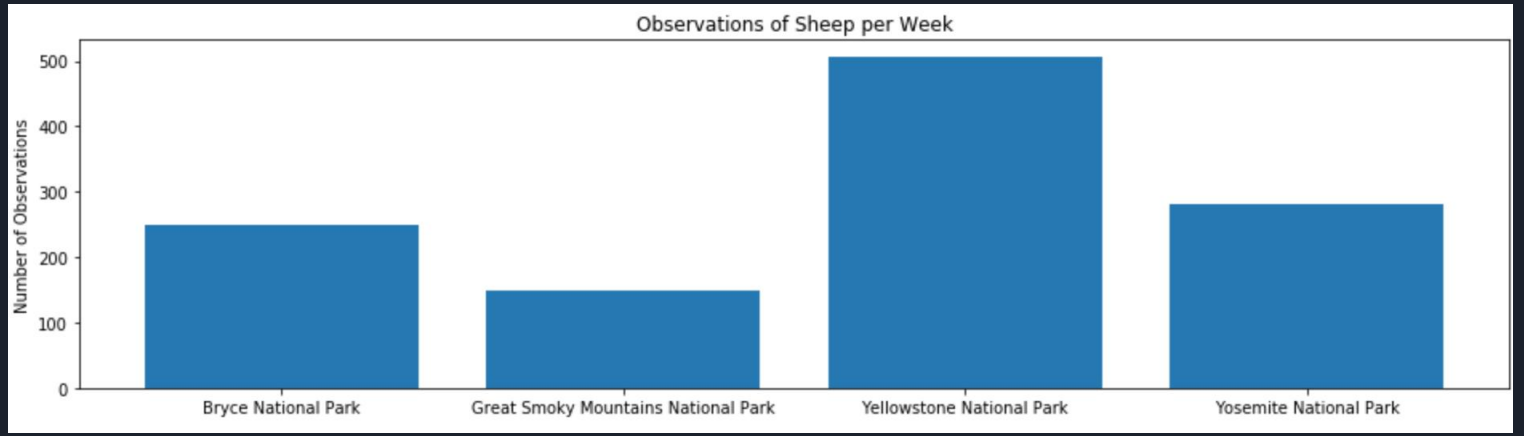
	category	not_protected	protected	percent_protected
0	Amphibian	72	7	0.088608
1	Bird	413	75	0.153689
2	Fish	115	11	0.087302
3	Mammal	146	30	0.170455
4	Nonvascular Plant	328	5	0.015015
5	Reptile	73	5	0.064103
6	Vascular Plant	4216	46	0.010793

Which species is more likely to endangered?

Based on our chi squared test between, Mammal and Bird, we got approximate p-value of 0.69, which is greater than 0.05. Hence, the data is insignificant. While between Mammal and Reptile, we got approximate p-value of 0.04 which is smaller than 0.05, Hence there is a significant difference between Mammal and Reptile in terms of protection. In other words, Reptile is much more prone than Mammal.

Observation of Sheeps at different parks

As the graph tells, sheeps are most likely to be found at Yellowstone National Park.



Weeks to take to see if the new program is effective

Currently, 15% of sheeps at Bryce National Park have foot/mouth disease while it is 10% for Yellowstone National Park. In order to detect at least of 5% of deduction with 90% of statistical significance, it takes 3.48 weeks for Bryce National Park, and 1.2 weeks for Yellowstone National Park.

In detail, sample size of sheeps to be examined from Bryce National Park is 870 while the sample size of that is 610 from Yellowstone National Park. The required sample size for Yellowstone National Park is smaller than one for Bryce National Park is due to its lower initial baseline.