Capstone: Biodiversity for the National Parks

By- Daniel Day

Intro

The ‘species_info.csv’ contained the following information.

- The scientific name of each species
- The common name of the species
- The conservation status of the species
- There were 5541 unique species included in the data set.

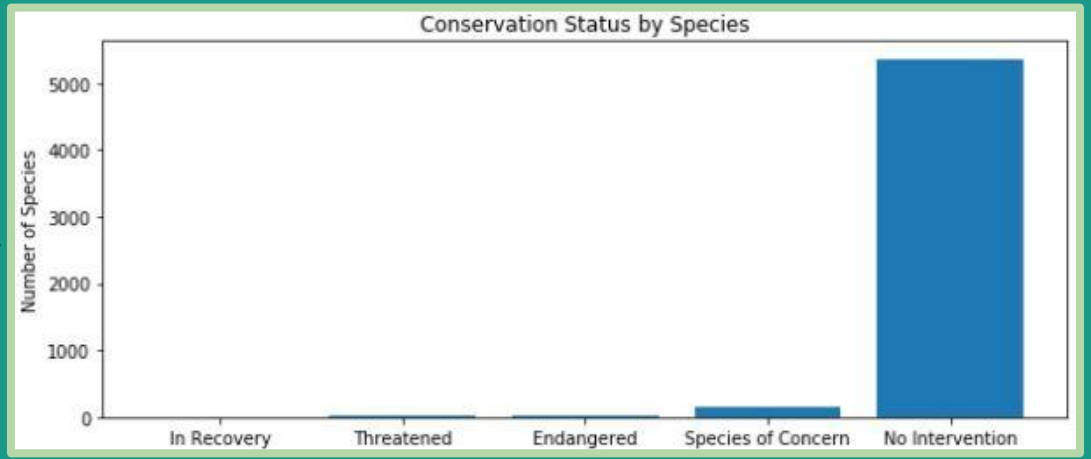
Conservation Status - definitions

1. Species of Concern : Declining or appear to be in need of conservation
2. Threatened: Vulnerable to endangerment in the near future
3. Endangered: Seriously at risk of extinction
4. In Recovery: Formerly Endangered, but currently neither in danger of extinction all or a significant portion of its range.

Conservation Cont’

- Out of all the species in the dataset 180 species were in one of the four conservation statuses. The rest currently require no intervention

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



Conservation Cont'

With this information we wanted to see if certain type of species were more likely to have a conservation status.

So we reorganized the data using pivot tables and made a chart to more accurately reflect the data.

	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

Recommendations

Based on the significance testing Mammals seem to be the most endangered type of species in the national parks.

While there is a significant difference between mammals and the other species, it is still recommended to continue efforts in protecting all wildlife/plant life within the national park.

Sheep Study

- In ‘observations.csv’ we received data detailing the amount of sheep observations at several national parks.
- The first thing we had to do was separate the sheep species, as well as record how many observations took place at each national park.

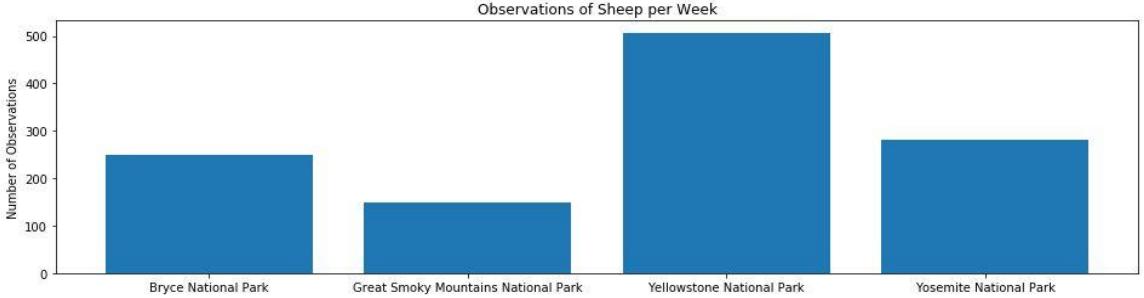
Organized Observation Chart

	scientific_name	park_name	observations	category	common_names	conservation_status	is_protected	is_sheep
0	Ovis canadensis	Yellowstone National Park	219	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
1	Ovis canadensis	Bryce National Park	109	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
2	Ovis canadensis	Yosemite National Park	117	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
3	Ovis canadensis	Great Smoky Mountains National Park	48	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
4	Ovis canadensis sierrae	Yellowstone National Park	67	Mammal	Sierra Nevada Bighorn Sheep	Endangered	True	True
5	Ovis canadensis sierrae	Yosemite National Park	39	Mammal	Sierra Nevada Bighorn Sheep	Endangered	True	True
6	Ovis canadensis sierrae	Bryce National Park	22	Mammal	Sierra Nevada Bighorn Sheep	Endangered	True	True
7	Ovis canadensis sierrae	Great Smoky Mountains National Park	25	Mammal	Sierra Nevada Bighorn Sheep	Endangered	True	True
8	Ovis aries	Yosemite National Park	126	Mammal	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True
9	Ovis aries	Great Smoky Mountains National Park	76	Mammal	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True
10	Ovis aries	Bryce National Park	119	Mammal	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True
11	Ovis aries	Yellowstone National Park	221	Mammal	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True

Combined Park Observations

	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

Observations Sheep Per Week



Sample size determination for sheep foot and mouth disease

- Since the scientist hypothesize that 15% of the sheep are infected with the disease, to test whether the scientist test were working , we used a sample size calculator to calculate the number of sheep that would have to be observed at each park.

- Using the default of 90% significance, a baseline of 15% and a minimum detectable effect of 33% . We Calculated that it would take 520 observations to have a viable sample size.