

Introduction to Data Analysis Codecademy

Biodiversity Project

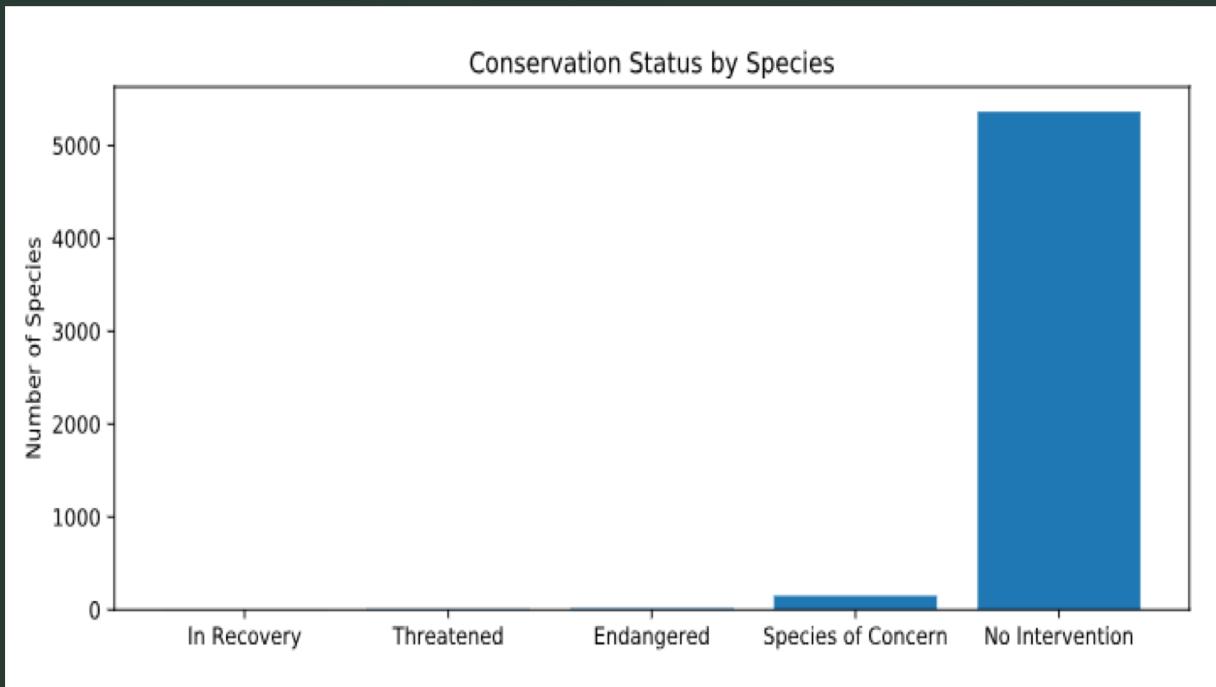
By Emma Briones

Part 1 Inspection the Different Species in U.S. National Parks

- The Species Dataset looks into the classifications, conservation status, scientific and common names of the living species in National Parks.
- The species in the dataset are ‘Mammal’, ‘Bird’, ‘Reptile’, ‘Amphibian’, ‘Fish’, ‘Vascular Plant’, and ‘Nonvascular Plant’
- Out of the 5824 animals and plants in this dataset, we can see that 15 species are seen to be endangered, 4 species in recovery, 10 threatened, and 151 species of concern.

Analyze Species Conservation Status

```
Vascular Plant      4470
Bird               521
Nonvascular Plant  333
Mammal              214
Fish                127
Amphibian            80
Reptile              79
Name: category, dtype: int64
```



The majority of the species in National Parks are Vascular Plants. The majority of the species in National Parks are considered safe ('No Intervention'); leaving 180 species to be of concern.

Breakdown of the Endangered Species

- Since the data handles various species, we need to know if there are certain species more likely to be endangered? There may be a relationship between the species' conservation status and whether or not they are protected in the parks.
- It looks like amphibians and fish are more likely to be protected than plants. Vascular plans are less likely to be protected.

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

Significant Calculations for Endangered Species – Animals are more endangered, but there are more plants in the National Parks

- Reptile and Mammal (0.038) : Mammals more likely to be endangered than Reptiles
- Fish and Nonvascular Plant (1.49 e -12) : Fish are more endangered than Nonvascular Plants
- Mammal and Vascular Plant (1.44 e -55): Mammals are more endangered than Vascular Plants
- Reptile and Vascular Plant (0.00015) : Reptiles are more endangered than Vascular Plants
- Mammal and Nonvascular Plant (0.00049) : Mammals are more endangered than Vascular Plants
- Bird and Nonvascular Plant (1.05 e -10): Birds are more endangered than Nonvascular Plants
- Nonvascular Plant and Amphibian (0.0018): Amphibians are more endangered than Nonvascular Plants
- Vascular Plant and Amphibian (1.044 e -8) : Amphibians are more endangered than vascular plants.

Recommendations and Final Thoughts

Part 1

- The majority of the species in National Parks are plants. Even though there are more plants than animals endangered or of concern in the National Parks, animals are more likely to be endangered than plants.
- When performing significance tests against the animals in the dataset, mammals more likely to be endangered than reptiles.
- For the National Parks, watch over the endangered animals and be sure to put more focus on mammals.

Part 2 Sample Size Determination Foot and Mouth Disease

- Across the different national parks lie various movements and species of sheep that can cultivate the parks. There has been an increase in Foot and Mouth Disease, so the following analysis will be done to figure out how the National Parks can decrease the disease.
- Yellowstone, Bryce, and Great Smoky Mountains are a few National Parks that have sheep.
- In the main dataset, the types of Sheep are ‘Ovis Canadensis’, ‘Ovis Canadensis Sierrae’, and ‘Ovis Aries’

Observation of Sheep in National Parks

When comparing the number of sheep observations in National Parks, it looks like Yellow National Park has the highest sightings of sheep.



Sample Size Determination

- Yellowstone National Park has been running a program to reduce the rate of foot and mouth disease from sheep. Their goal is to detect reductions of at least 5%.
- Using the sample size calculator, we are able to calculate to observe from each park to make sure if the mouth and food percentages are significant by 90%.
 - baseline = 0.15
 - minimum_detectable_effect = $100*5./15$
 - sample_size_per_variant = 870

Sample Size Per 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

With the observations above, we are able to determine how many weeks would the scientists need to spend at the park to observe enough sheep.

- Bryce National Park: about 3.5 weeks
- Great Smoky Mountains : 5 weeks
- Yellowstone National Park: about 2 weeks
- Yosemite National Park : 3 weeks