Codecademy Project

Biodiversity for the National Parks

Capstone Project Overview

The National Park Service is an entity that monitors and observes the conservation statuses of wildlife species at select National Parks.

This report contains two sections.

Section 1: Contains analysis on various protected and nonprotected species from different national parks and determines if certain species are more susceptible to being endangered than others.

Section 2: Contains a determination of the sample size needed to help show if the reduction efforts of sheep with foot and mouth disease is working.

*Section 1:
Are certain types of species more likely to be endangered?*

Exploring the Data in species_info.csv

The information contained in the species_info.csv file includes the following columns about animals and plants in the parks:

- Category of animal/plant species
- Scientific names of the species
- Common names of the species
- Conservation statuses.

	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), Domesticated Cattle	nan
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	nan
4	Mammal	Cervus elaphus	Wapiti Or Elk	nan

Exploring the Data in species_info.csv

When calculating the different number of species at the various national parks, there are 5541 different species in the DataFrame.

The various categories and conservation statuses are as follows:

Category: Mammal, Bird, Reptile, Amphibian, Fish, Vascular Plant, Nonvascular Plant
Conservation Status: NaN (NULL), In Recovery, Species of Concern, Threatened, Endangered

|	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), Domesticated Cattle	nan
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	nan
4	Mammal	Cervus elaphus	Wapiti Or Elk	nan

Number of species in each conservation status

Upon further inspection of the Dataframe, there appears to be a small number of species that have a declining population or appear to be in need of conservation, compared to the total number of species records. A small number of 4 were once considered endangered, but are now in recovery. In total, only 180 species appear to be in need of protection.

A total of 5,363 species have no conservation status and appear to not require a protected status.

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

Number of species in each conservation status
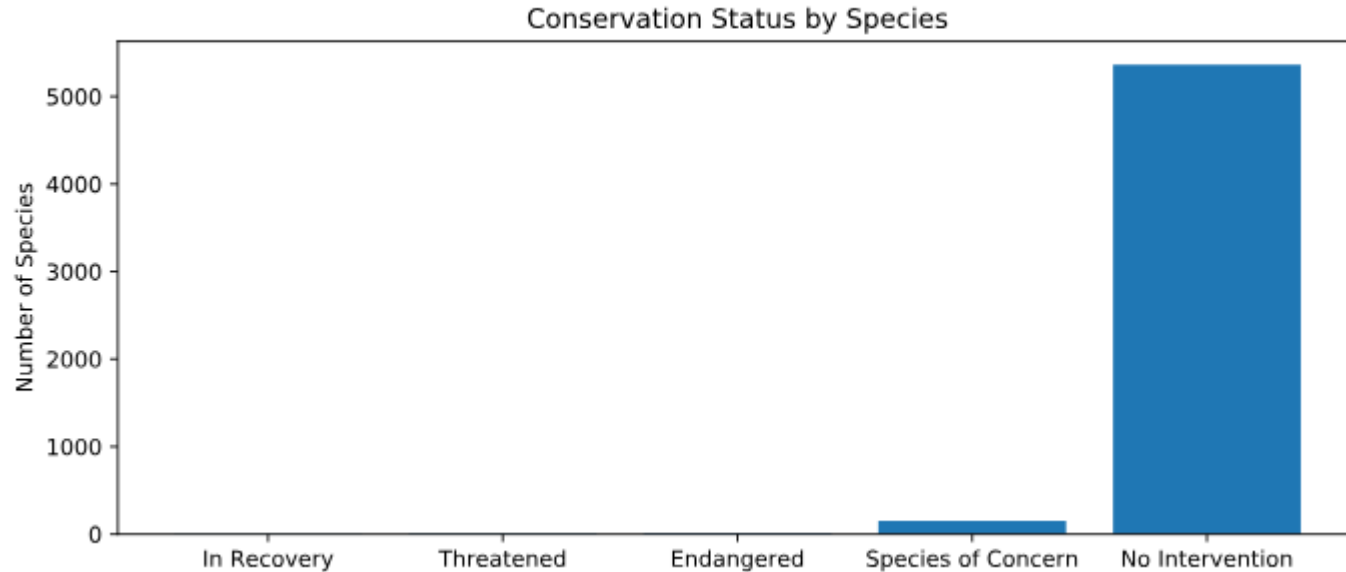
The bar graph above was generated using Matplotlib and shows the number of species in each conservation status. This further indicates that a large number of species do not require protection intervention at this time.

Category of species vs conservation status

Digging deeper into the Dataframe reveals that vascular and nonvascular plants are a species with the least amount of needed protection.

Mammals and birds are the species with the highest percentage of protection needed. With a higher number of birds needing protection, but a slightly lower percentage compared to mammals.

Is it true that mammals have a higher chance of needing protection than birds? What about other species? We'll examine these questions further to determine statistical significance.

	Not protected	Protected	% Protected
Amphibian	72	7	8.86%
Bird	413	75	15.37%
Fish	115	11	8.73%
Mammal	146	30	17.05%
Nonvascular Plant	328	5	1.50%
Vascular Plant	4216	46	1.08%
Reptile	73	5	6.41%

Are certain types of species more likely to be endangered?

In order to answer this question, the Chi-Squared Test was used to determine significance.

While looking at the percentages of protected vs non-protected, it appears that mammals are more likely to be endangered than birds. However, after running the test, we cannot reject the null hypothesis that the difference is because of chance, since the p-value is more than .05.

However, after running the test against mammals and reptiles, we can reject the null hypothesis, which reveals that some species are more susceptible to endangerment than others.

Test	P-value	Significance
Mammal /Bird	0.687594809666	No
Mammal/ Reptile	0.0383555902297	Yes
Fish/Reptile	0.740652461591	No
Amphibian/Fish	0.824794298152	No
Bird/Vascular Plant	4.61226803169e-79	Yes

Recommendation for conservationists

To conclude, there is no real significance between mammals and birds in terms of which are more likely to be endangered. However, when it comes to reptiles vs mammals or vascular plants vs birds, there is a significance in terms of which are more likely to be endangered.

A recommendation for conservationists is to understand that certain species are more likely become endangered than others.

The data points to mammals having the largest number of protected status (17%), with birds following closely behind (15%). These two are the species more likely to need future protection status/intervention than fish, reptiles and plants.

Test	P-value	Significance
Mammal /Bird	0.687594809666	No
Mammal/ Reptile	0.0383555902297	Yes
Fish/Reptile	0.740652461591	No
Amphibian/Fish	0.824794298152	No
Bird/Vascular Plant	4.61226803169e-79	Yes

Section 2: What sample size of sheep is needed to determine reduction of foot and mouth disease effort?

Exploring the Data in observations.csv

The next dataset to analyze involves sightings recorded of different species at select national parks over the last week.

The information contained in the observations.csv file includes the following columns about sheep sightings at the parks:
- Scientific names of the species
- Park Name
- Observations

	scientific_name	park_name	observations
0	Vicia benghalensis	Great Smoky Mountains National Park	68
1	Neovison vison	Great Smoky Mountains National Park	77
2	Prunus subcordata	Yosemite National Park	138
3	Abutilon theophrasti	Bryce National Park	84
4	Githopsis specularioides	Great Smoky Mountains National Park	85

Sheep Observations

Breaking down the data further shows the number of sheep observations 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

Further exploration of the file reveals the number of sheep observations per week at each national park under investigation.
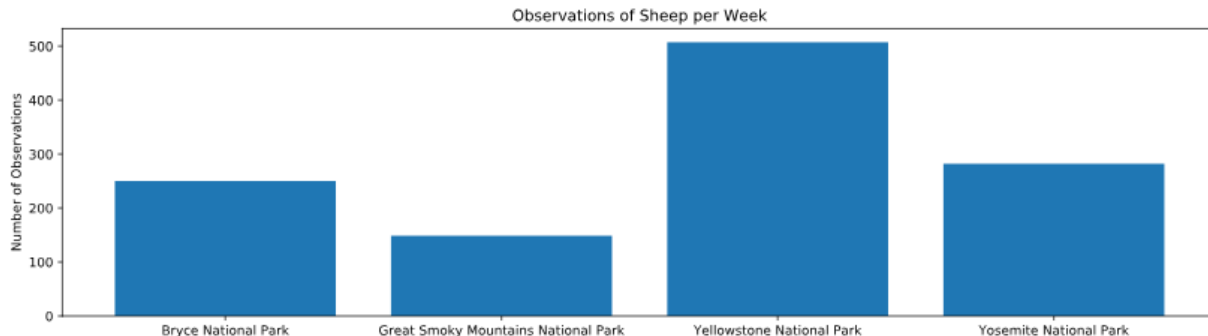
Foot and Mouth Reduction Effort
Sample Size Determination

Park Rangers are trying to reduce the rate of foot and mouth disease at. But in order to determine if the program efforts are paying off, we need to calculate the number of sheep that would need to be observed from each park. This helps ensure confidence that the foot and mouth disease percentages are significant.

The only information we have is that last year, 15% of sheep at Bryce National Park had foot and mouth disease.

To run this test we need to provide the A/B Sample Size Calculator with the following numbers.

- 15% Baseline conversion rate (number of sheep with foot and mouth disease at Bryce National Park)
- 33.3% minimum detectable effect ((5)/15)*100)
- 90% statistical significance

These inputs indicate that the sample size of sheep that needs to be observed is 870. Further calculations also reveal that scientists would need to spend 1.7 weeks at Yellowstone National Park to observe enough sheep. To observe enough sheep at Bryce National Park, scientists would need to spend nearly 3.5 weeks.