

Codecademy Capstone Project - Biodiversity for the National Parks

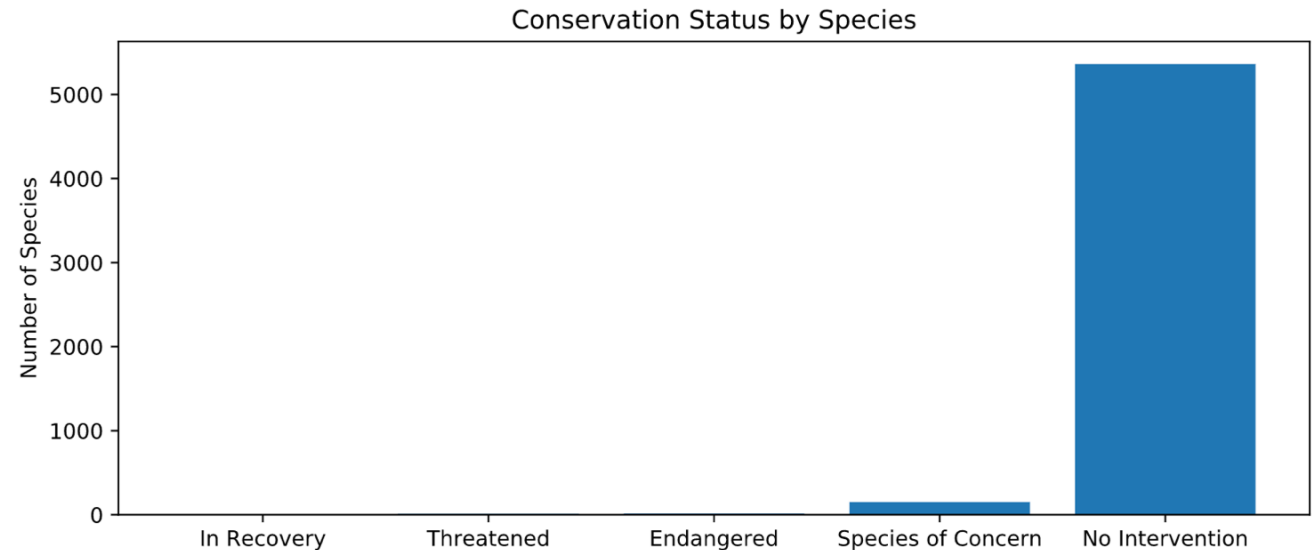
By Ashish Babaria

Digging into the data – *species_info.csv*

- Category
- Scientific_name
- Common_name
- Conservation_status

Conservation Status	Scientific Name
Endangered	15
In Recovery	4
Species of Concern	151
Threatened	10

After cleaning the data, we replaced the null values to “No Intervention” leaving us with the following results



Are certain types of species more likely to be endangered?

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

We ran a chi-squared test with a null hypothesis that the difference in percent protected is due to chance.

When we ran the chi-squared test, we found a p-value of ~ 0.68759 , so we can conclude the percent protected of mammals and birds is not significant and is due to chance.

We ran another chi-squared test and we found a p-value of ~ 0.0383556 , which is significant so we can conclude the percent protected of mammals and reptiles are not due to chance.

We can conclude that certain types of species are more likely to be endangered.

Sample Size – Foot & Mouth Disease Study

We began with a 15% baseline for the occurrence of foot and mouth disease in the sheep at Bryce National Park. When given a 90% confidence interval and the desire to notice a 5+% drop in observations at Yellowstone National Park, the sample size was at least 510.

From this we can deduce that we would need about one week at Yellowstone and about two weeks at Bryce to achieve our desired sample size.

park_name	observations
Bryce National Park	250
Great Smoky Mountains National Park	149
Yellowstone National Park	507
Yosemite National Park	282

