BIODIVERSITY FOR THE NATIONAL PARKS

Capstone Project by Kartik Adimulam

Species _info.csv

The species_info csv provided data about animals in national parks, including the scientific name of the species, the category, the common names used to describe the species, and the conservation status. While working with the data, I noticed that the vast majority of species had no intervention, or no data regarding the conservation status. This was surprising as most animals are safe and under close to no danger of extinction. I also noticed that out of the animals that did have information regarding the conservation status, most of them were species of concern. This is generally positive as this is the mildest of concerns. Overall, the species_info.csv data provided a lot of information on categories of animals and where they fell in the conservation status categories, providing a data analyst a lot of data to work with and analyze.

Calculations done on significance

I performed two calculations on the data provided in the csv. They were both chi squared tests focused mainly on the pvalue and testing the null hypothesis. The goal of these calculations was to look at each category of animals, such as mammals or birds, and group each category into two columns. The two columns were protected or not protected. From here i used a chi squared test to see whether there was a significant difference between categories, showing me if different categories of animals were prone to be endangered over other categories of animals. The results of both calculations were different and both provided different conclusions for a data analyst.

RECOMMENDATION

The first calculation was done between mammals and birds, trying to see whether either species was more prone to endangerment. The result was the null hypothesis being proven correct, the p value was very high, 0.68, so neither animal was more prone to going extinct. The second calculations, however, yielded interesting results. The second calculation was done between mammals and reptiles. The p value was 0.03, incredibly low, showing a significant difference. As the reptiles have the lower percent protected, I would suggest to a conservationist to focus more on reptiles than mammals on birds. Make sure to analyze every part of the reptiles habitat, as there is some factor that is causing them to die out quicker than other animals, it is not just a coincedence.

Sample size determination

Scientists from multiple national parks want to know how many sheep have foot and mouth disease. Last year, they recorded 15% percent of sheep with this disease, but they neeeded to know how many samples need to be taken for this percent to be significant. Three numbers were needed to be inputted into the calculator. First the baseline conversion rate was 15%, as that was the percent recorded last year, the base percent we are testing. Next the statistical significance is defaulted at 90%. Last, the minimum detectable effect is how much of a difference we are testing. Since scientists want a 5 percentage point difference, the difference will be in relation to the baseline rate. Since 5 percent of 15 percent is ⅓, our minimum detectable effect is 33.33333 percent, a repeating decimal. Our output is 870 samples needed to be confident in our findings.

Graphs Created