Biodiversity in National Parks

Data Science Capstone Project

Oluwatosin Ishmeal

29.06.2020

Table of Contents

1. Project Introduction

Project Introduction

This project is based on data from the National Parks Service. They are interested in analysing some data about species at various national parks.

The National Parks Service would like you to perform some data analysis on the conservation statuses of these species and to investigate if there are any patterns or themes to the types of species that become endangered.

The data provided contains information such as the scientific name, common name and conservation status of each species.

Each species are separated into 7 different categories namely:

- Mammals
- Bird
- Reptile
- Amphibian
- Fish
- Vascular Plant
- Nonvascular Plant

The species are further classified into 5 different conservation status which are:

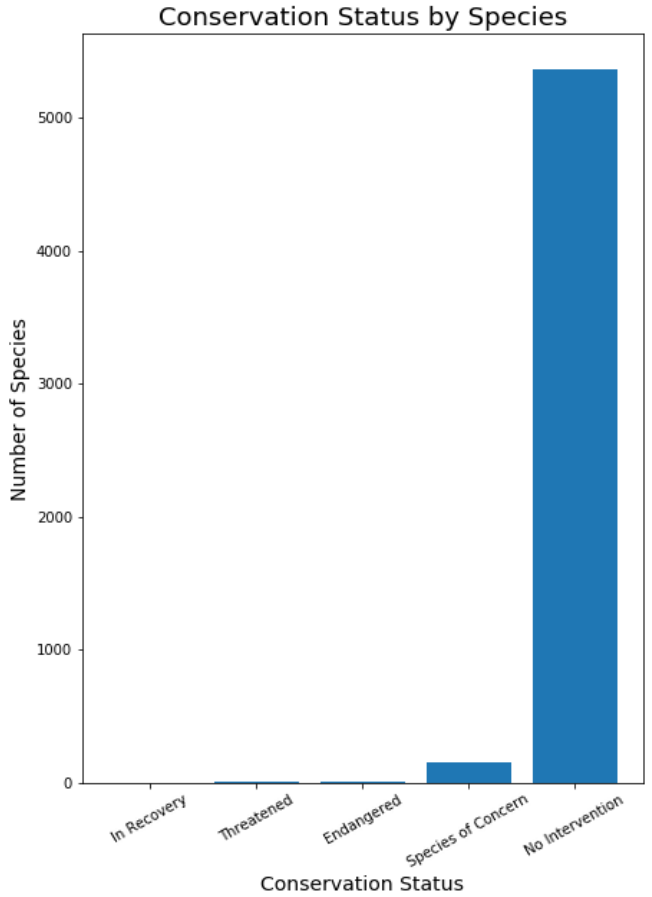
- Species of Concern
- Endangered
- Threatened
- In Recovery
- No Intervention

2. Analysis & Calculations

2.1 Number of Species in Conservation Status

Since we are interested in the conservation status of each species, we decide to see how many species are in each of the different conservation status. We have displayed this information both as a table and a graph.

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



2.2 Certain Species more endangered?

Next, we are interested in finding out if certain species are more likely to be endangered than others. To do this, we find the species that are not in the ‘No Intervention’ status and classify them based on their category

Category	Protected	Not Protected	Percentage Protected
Amphibian	7	72	0.09
Bird	75	413	0.15
Fish	11	115	0.09
Mammal	30	146	0.17
Nonvascular Plant	5	328	0.02
Reptile	5	73	0.06
Vascular Plant	46	4216	0.01

2.3 Significant Tests

After doing the calculations to see if some species are more susceptible to be endangered than others in the previous slides, we ran a binomial test on different species to check if this statement is true.

First, we ran the test on Mammals and Birds (two highest values in our table).

- pvalue > 0.05, thus the difference we noticed in the result is not significant.

Next, we ran the test on Mammals and Reptiles

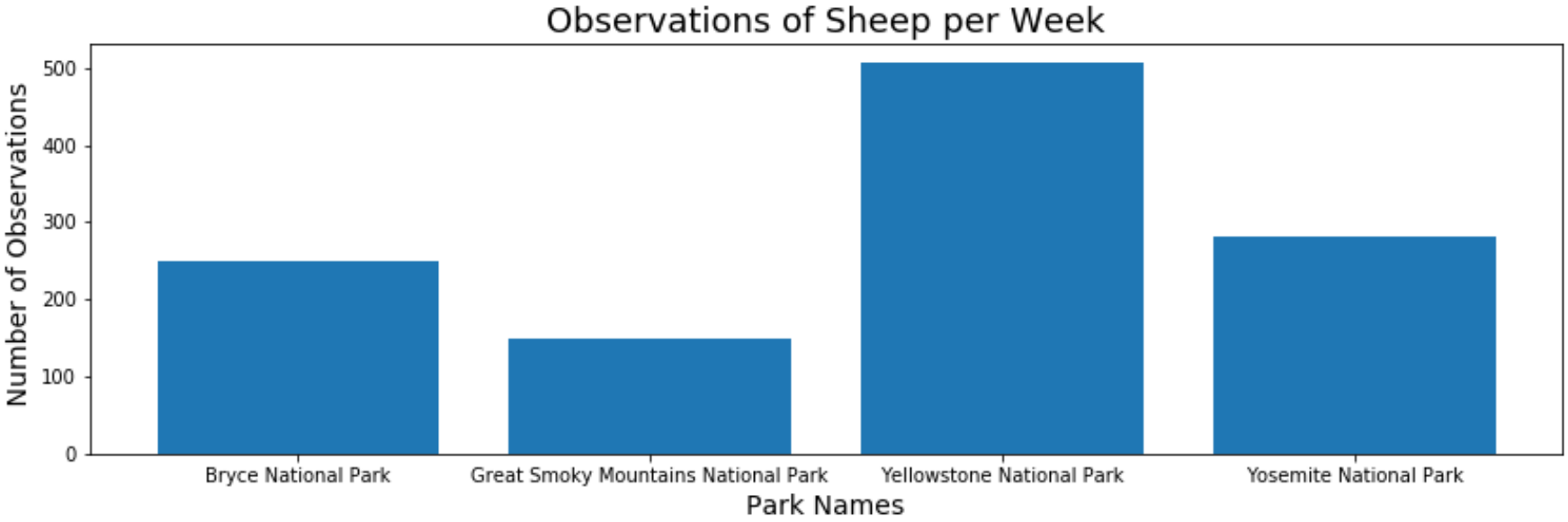
- pvalue < 0.05, this indicates that the different we noticed in the result is significant and Reptiles are more likely to be endangered than Mammals. Why?

Lastly, we ran the test on Vascular Plant and Non-Vascular Plant

- pvalue > 0.05, thus the difference noticed between both plant is not significant

2.4 Sheep Observations

Some scientists are studying the number of sheep sightings at different national parks. There are several different scientific names for different types of sheeps, and we have used this to filter our dataframe to include only sheeps. We have plotted a bar chart showing different number of observations per week at each park



2.5 Sample Size Calculations

Our scientists know that 15% of sheep at Bryce National Park have foot and mouth disease. Park rangers at Yellowstone National Park have been running a program to reduce the rate of foot and mouth disease at that park. The scientists want to test whether this program is working or not. They want to be able to detect reductions of at least 5 percentage points. Thus we need to calculate the sample size needed for this observation.

There are many online sample size calculator that we can use and for this project we decided to use the one provided by Codecademy. This needs some default inputs such as Confidence Level, Minimum Detectable Effect, and the Baseline Conversion Rate.

Confidence Level was provided by our scientists at 90%.
Minimum Detectable Effect is calculated as a percent of the baseline conversion rate, since the baseline conversion rate is 15% and we want a reduction of 5%, we can divide 5% by 15% which gives 0.333 and multiplying by 100% gives us 33.33%. Inputting all this information to the calculator return a sample size of 870 sheeps needed for the observation.

For our scientists at the Bryce National Park this means they will have to observe sheeps for 3.5 weeks and for our scientists at Yellowstone National Park for 2 weeks.

3. Recommendation

Recommendations

We are recommending the following steps to the National Park that we believe will help them to preserve the species:

- Focus on why reptiles are more likely to be endangered and to protect more species of reptiles.
- Try to get more Species to move from Endangered, Species of Concern and Threatened status to In Recovery.
- To campaign against animal trading as most endangered species might have been hunted due to their perceived medicinal or economic advantages.