codecademy

BioDiversity Project
Capstone

Data Analysis using Python, MatpplotLib, Hypothesis Testing

Kaustubh Joshi

April 2020

Biodiversity Project - Table of Contents

Note :

All analysis and visualisations have been done by Kaustubh Joshi, data visualization developer.

Species Data - What data do we have?

- To start this project, we have a well documented data set covering **5541 unique species** across **7 categories.** Listed below :
    - 'Mammal'
    - 'Bird'
    - 'Reptile'
    - 'Amphibian'
    - 'Fish'
    - 'Vascular Plant'
    - 'Nonvascular Plant'
- While species have been identified y scientific names, the data also provides details of all the common names by which the species is known for easier analysis and interpretations from real world observations.
- Species have also been classified as follows :
    - **Species of Concern:** declining or appear to be in need of conservation
    - **Threatened:** vulnerable to endangerment in the near future
    - **Endangered:** seriously at risk of extinction
    - **In Recovery:** formerly Endangered, but currently neither in danger of extinction throughout all or a significant portion of its range
    - The remaining species require ‘No Intervention’

Endangered Species - Observations

	conservation_status	scientific_name
1	In Recovery	4
4	Threatened	10
0	Endangered	15
3	Species of Concern	151
2	No Intervention	5363

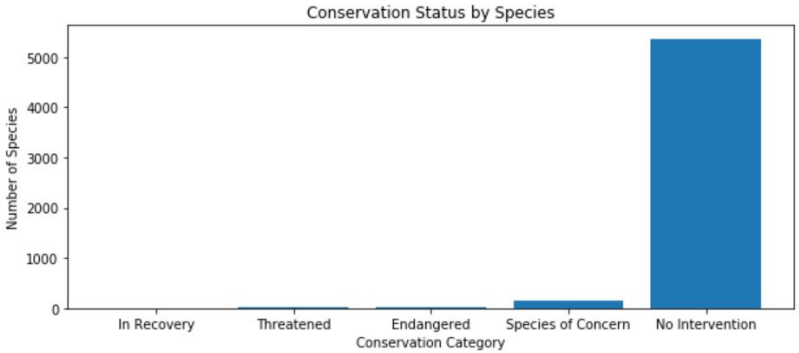

Table alongside shows the count of species by endangered status.
As we can see, 151 species are species of concern, and need early attention.

Endangered Species - Recommendations

While analyzing % of a category endangered, the data for Birds and Mammals is particularly alarming.
Running a Chi Square Test confirms that the differences observed are statistically significant and need special attention

We recommend, additional resources and conservation efforts be deployed towards **Mammals and Bird categories.**

is_protected	category	not_protected	protected	percent_protected
0	Amphibian	72	7	0.097222
1	Bird	413	75	0.181598
2	Fish	115	11	0.095652
3	Mammal	146	30	0.205479
4	Nonvascular Plant	328	5	0.015244
5	Reptile	73	5	0.068493
6	Vascular Plant	4216	46	0.010911

Foot & Mouth Sample Sizing

For evaluating conservation efforts done to prevent foot & mouth disease, we recommend a sample size of 870 with the following assumptions :

#Baseline conversion rate = 15%
Minimum detectable effect = 33.33% (i.e : 5%/15%)
Confidence = 90%

These observations will require 4 weeks at Bryce and 2 weeks # at Yellowstone National Parks based on the below weekly sheep observation estimates.

Baseline conversion rate:	15	%
Statistical significance:	85% 90% 95%	
Minimum detectable effect:	33.3333	%
Sample size:	870