

Biodiversity for the National Parks

Wei Wu

Describe the data I

	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), Dom...	NaN
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	NaN
4	Mammal	Cervus elaphus	Wapiti Or Elk	NaN

- The original **species_info.csv** has four columns:
 - category, scientific_name, common_names, conservation_status
- Category contains ['Mammal', 'Bird', 'Reptile', 'Amphibian', 'Fish', 'Vascular Plant', 'Nonvascular Plant']
- Conservation_status contains [nan, 'Species of Concern', 'Endangered', 'Threatened', 'In Recovery']

Describe the data II

	conservation_status	scientific_name
0	Endangered	16
1	In Recovery	4
2	No Intervention	5633
3	Species of Concern	161
4	Threatened	10

- Data group by conservation_status and count of scientific_name;
- Majority of conservation_status is No Intervention;

Significance calculations for chi squared test

	category	not_protected	protected	percent_protected
0	Amphibian	72	7	0.911392
1	Bird	413	75	0.846311
2	Fish	115	11	0.912698
3	Mammal	146	30	0.829545
4	Nonvascular Plant	328	5	0.984985
5	Reptile	73	5	0.935897
6	Vascular Plant	4216	46	0.989207

- Two significance calculations have been conducted for this project:

- category Mammal v.s. Bird.

	protected	not protected
Mammal	30	146
Bird	75	413

- p Value = 0.687594809666
- No significant difference between Mammal and Bird

- category Mammal v.s.Reptile.

	protected	not protected
Mammal	30	146
Reptile	5	73

- p Value = 0.0383555902297
- There is significant difference between Mammal and Reptile

Recommendation for conservationists

- From the chi squared test result, we found species in category Mammal are more likely to be endangered than species in Reptile.
- Since Mammal and Birds are not significantly different, this means Birds are also more likely to be endangered than species in Reptile.
- More actions need to be taken toward these two categories.

Sample Size Calculation

Baseline Conversion Rate

15 %

Your control group's expected conversion rate. [\[?\]](#)

Minimum Detectable Effect

33.3 %

The minimum relative change in conversion rate you would like to be able to detect. [\[?\]](#)

Statistical Significance

90%

95% is an accepted standard for statistical significance, although Optimizely allows you to set your own threshold for significance based on your risk tolerance. [\[?\]](#)

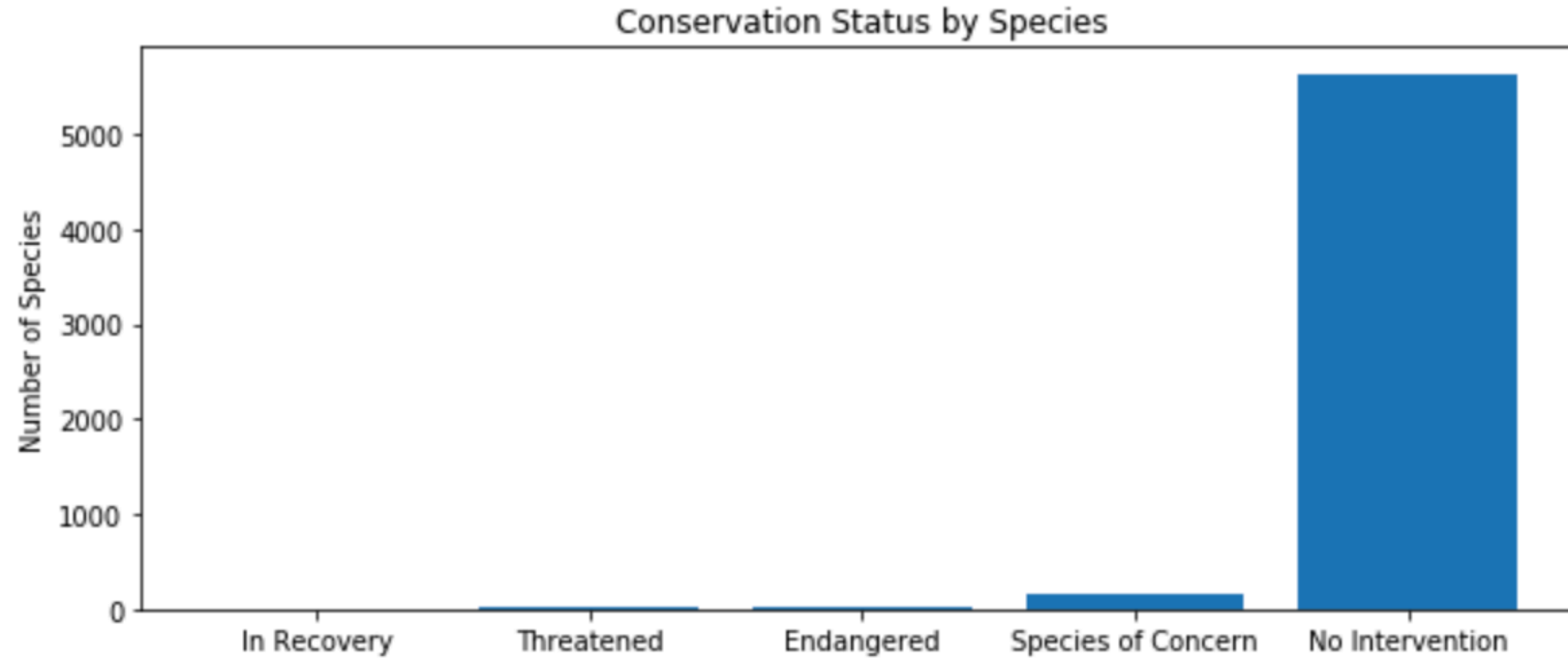
[EDIT](#)

Sample Size per Variation

510

- Baseline = 15%
- Since they want to be able to detect reductions of at least 5 percentage point.
- Minimum Detectable Effect = $5\%/15\%*100 = 33.33\%$
- Statistical Significance = 90%
- Using the tool: Optimizely to find the sample size per variation = 510;

Graph I Conservation Status by Species



Graph II Observations of Sheep per Week

