
Biodiversity Capstone Project

— Investigating Protected Species —

The dataset

File: species_info.csv

Dimensions: Category, Scientific_name, Common_names, Conservation_status

Category: contains 7 categories (Mammal, Bird, Reptile, Amphibian, Fish, Vascular Plant, Nonvascular Plant)

Scientific_name: contains 5,541 distinct species

Common_names: many different values

Conservation_status: contains 5 status (Endangered, In Recovery, No Intervention, Species of Concern, Threatened)

	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), Domesticated Cattle	nan
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	nan
4	Mammal	Cervus elaphus	Wapiti Or Elk	nan

Significant calculations

4 - Analyzing species conservation status:

```
conservation_counts =  
species.groupby('conservation_status').scientific_name.nunique().reset_index()
```

2 - Replacing in 'NaN' values with 'No Intervention':

```
species.fillna('No Intervention', inplace = True)
```

3 - Checking species conservation status again:

```
conservation_counts_fixed =  
species.groupby('conservation_status').scientific_name.nunique().reset_index()
```

Before replacement of 'NaN' values

conservation_status	scientific_name
Endangered	15
In Recovery	4
Species of Concern	151
Threatened	10

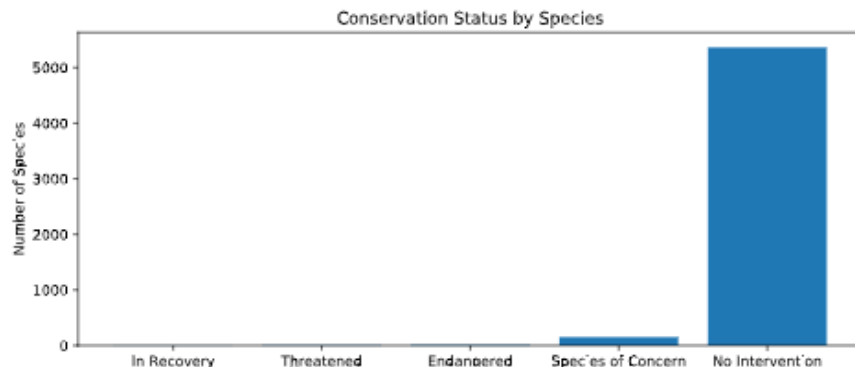
After replacement of 'NaN' values

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

Significant calculations

4 - Plotting Conservation Status by Species

```
plt.figure(figsize=(10, 4))
ax = plt.subplot()
plt.bar(range(len(protection_counts)),protection_counts.s
cientific_name.values)
ax.set_xticks(range(len(protection_counts)))
ax.set_xticklabels(protection_counts.conservation_status.
values)
plt.ylabel('Number of Species')
plt.title('Conservation Status by Species')
labels = [e.get_text() for e in ax.get_xticklabels()]
plt.show()
```



Significant calculations

5 - Creating a new column to check which species are protected

```
species['is_protected'] = species.conservations_status != 'No Intervention'
```

```
category_counts = species.groupby(['category', 'is_protected']).scientific_name.nunique().reset_index()
```

6 - Pivoting to better visualize the data

```
category_pivot = category_counts.pivot(columns='is_protected', index='category', values='scientific_name').reset_index()  
category_pivot.columns = ['category', 'not_protected', 'protected']
```

7 - Creating a new column to display the percentage protected

```
category_pivot['percent_protected'] = category_pivot.protected /  
(category_pivot.protected + category_pivot.not_protected)
```

	category	not_protected	protected	percent_protected
0	Amphibian	72	7	0.088608
1	Bird	413	75	0.153689
2	Fish	115	11	0.087302
3	Mammal	146	30	0.170455
4	Nonvascular Plant	328	5	0.015015
5	Reptile	73	5	0.064103
6	Vascular Plant	4216	46	0.010793

Significant calculations

8 - Chi-Squared Test for Significance - Checking whether Mammals are more likely to be endangered than Birds

```
contingency = [[30, 146], [75, 413]]  
pval = chi2_contingency(contingency)[1]  
print(pval)
```

No significant difference because $pval > 0.05$

9 - Chi-Squared Test for Significance - Checking whether Mammals are more likely to be endangered than Reptiles

```
contingency_reptile_mammal = [[30, 146], [5, 73]]  
pval_reptile_mammal = chi2_contingency(contingency_reptile_mammal)[1]  
print(pval_reptile_mammal)
```

Significant difference! $pval_reptile_mammal < 0.05$

10 - Conclusion:

Certain types of species are more likely to be endangered than others.

Recommendations for conservationists concerned about endangered species

- Conduct Chi-Squared tests to identify which species are more likely to be endangered than others
- Allocate greater amount of budget to species that are more likely to be endangered
- Promote initiatives to educate the public
- Create a program to facilitate reproduction of endangered species until the significance tests' results become no significant

Sample size determination

baseline = 15

minimum_detectable_effect = $100 \times 5 / 15$

sample_size_per_variant = 870

yellowstone_weeks_observing = $\text{sample_size_per_variant} / 507$.

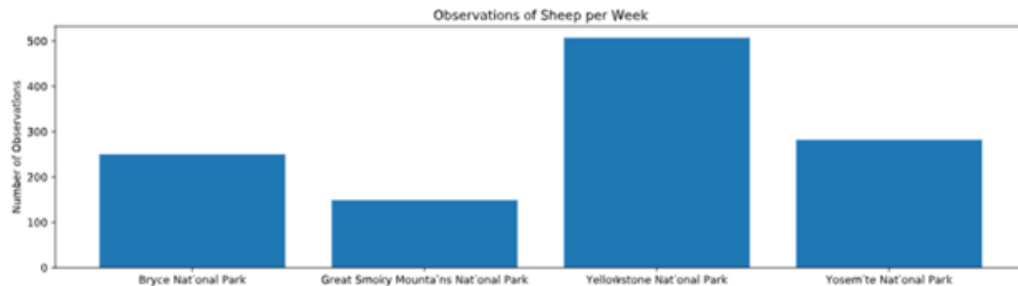
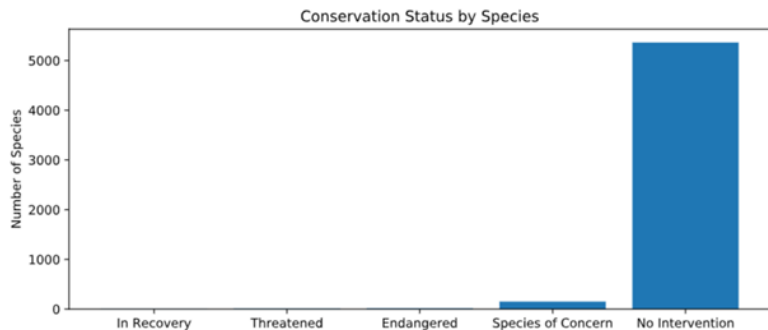
bryce_weeks_observing = $\text{sample_size_per_variant} / 250$.

Sample size calculator

Baseline conversion rate:	<input type="text" value="15"/>
	%
Statistical significance:	<input type="text" value="85%"/> <input checked="" type="text" value="90%"/>
	<input type="text" value="95%"/>
Minimum detectable effect:	<input type="text" value="33.3"/>
	%
Sample size:	<input type="text" value="870"/>

Appendix

Graphs created during analysis



Thank you!