Catherine Fernandez

Biodiversity for the National Parks

Capstone Project

species_info.csv

The CSV is a Data Base that shows the data from the National Park Service

	category	scientific_name	common_names	conservation_status
1	Mammel	Clethrionomys gapperi gapperi	Gapper's Red-Backed Vole	NaN
2	Mammel	Bos bison	American Bison, Bison	NaN
3	Mammel	Bos taurus	Aurochs, Aurochs, Domestic Cattle (Feral), Dom...	NaN
4	Mammel	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	NaN
...				

- The Code, `species = pd.read_csv('species_info.csv')` was used to upload the data
- The Code is then printed and examined through, `print species.head()`
- The chart above displays the Category, Scientific name, Common name, and Conservation Status of the animals

Endangered Status Between different Species

Many wonder if certain species are more likely to be endangered versus others. One way to find out would be to keep the code `species = pd.read_csv('species_info.csv')` to navigate through.

We now want to know if certain types of species are more likely to be endangered than others.

A new column in species called, `is_protected` which will be true if conservation status is will be equal to ‘no_intervention’ and false otherwise.

```
species['is_protected'] =

species.conservation_status != 'No Intervention'
```

We then grouped category and `is_protected` and saved results to `category_counts`.

Next, creating a pivot is vital to the chart we want to create to see if some species are more likely to be endangered.

Creating a Pivot will allow for the data to be viewed much easier as well as rearrange category_counts

Using a pivot will allow category_counts to be rearranged so the following happens:

- columns is is_protected
- index is category
- And values is scientific_name

Then the pivoted data was saved to category_pivot and .reset_index() at the end

This was the code;

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

category_pivot = category_counts.pivot(columns='is_protected',

        index='category',

        values='scientific_name')\

        .reset_index()
```

Below is the chart we currently have after we print

is_protected	category	False	True
0	Amphibian	72	7
1	Bird	413	75
2	Fish	115	11
3	Mammal	146	30
4	Nonvascular Plant	328	5
5	Reptile	73	5
6	Vascular Plant	4216	46

True and False doesn't tell us much. So we want to change false to not_protected and True to protected. We also want to find out the actual percent that is Protected so I will input these lines of code

```
category_pivot['percent_protected'] =

category_pivot.protected /

(category_pivot.protected + category_pivot.not_protected)

print(category_pivot)
```

Endangered Status Species Results!

	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

After print and analyze this graph we come to see the following:

- Out of all the species listed Birds are the most protected while nonvascular plants are least protected
- All of these species are far more non protected than protected
- Altogether combined there are 1,074 non protected species listed on this chart versus 128 protected species

Chi-Squared Test for Significance

Running a Chi-Squared test based on the previous chart we are testing to see if certain spaces are more endangered than others based n a Chi-Squared test chart. With this test we are looking to see if there is a significant difference in endangerment between Mammals and Birds and Mammals and Reptiles.

Mammals and Birds Test

```
contingency = [[30, 146], [75, 413]]

pval = chi2_contingency(contingency)[1]

print(pval)
```

p value = 0.688 (not significant)

Mammals and Reptiles

```
contingency_reptile_mammal = [[30, 146],  [5, 73]]

pval_reptile_mammal = chi2_contingency(contingency_reptile_mammal)[1]

print(pval_reptile_mammal)
```

p value = .038 (significant difference)

We can conclude there is a significant difference between mammals and reptiles.

Recommendations for conservationists

1. Advocate the importance of maintaining the population of our wildlife.
2. Get younger generations involved and create programs so that people get to understand the importance of these species by working with them.
3. Raise funding so that if we can get more species in captivity they are provided the best quality of life possible.
4. Based on the Chi-Squared test I recommend that all animals get more protection and supervision. However, noticing the difference between reptiles and mammals I'd say reptiles are more likely to be endangered and need more protection.

Foot and Mouth Disease

YellowStone national park service wants to know if Their Foot and Mouth Diseanse test actually works. Many sheep from various parks have been having foot and mouth disease so we are now testing this.

From being given a 15% base line from Bryant national park I concluded that 870 sheep in total would have to be observed to conclude that more than 5% of sheep have been rid of the disease.

The same size variant was found by using the sample size calculator. I input 15% for the baseline conversion rate with the info given, used a statistical significance of 90%, a minimum detectable effect of 33.3% which was found by using the code;

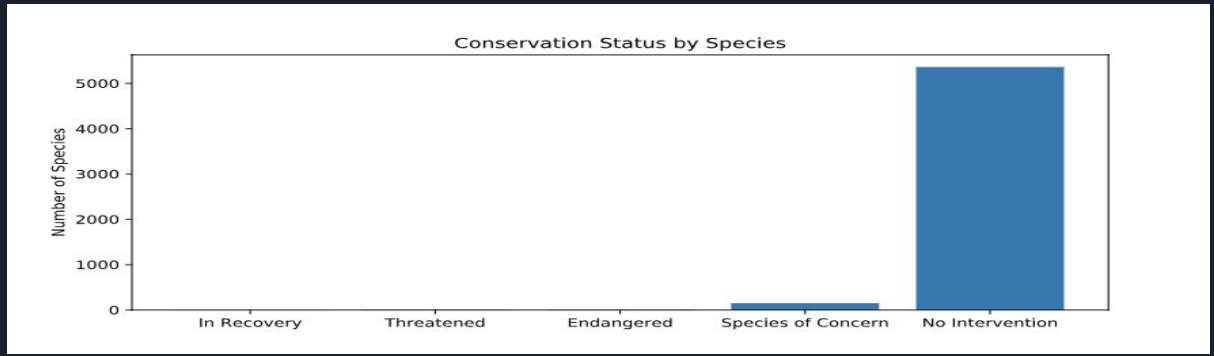
```
minimum_detectable_effect = 100*5./15
```

```
print(minimum_detectable_effect)
```

All together this gave me the total of 870

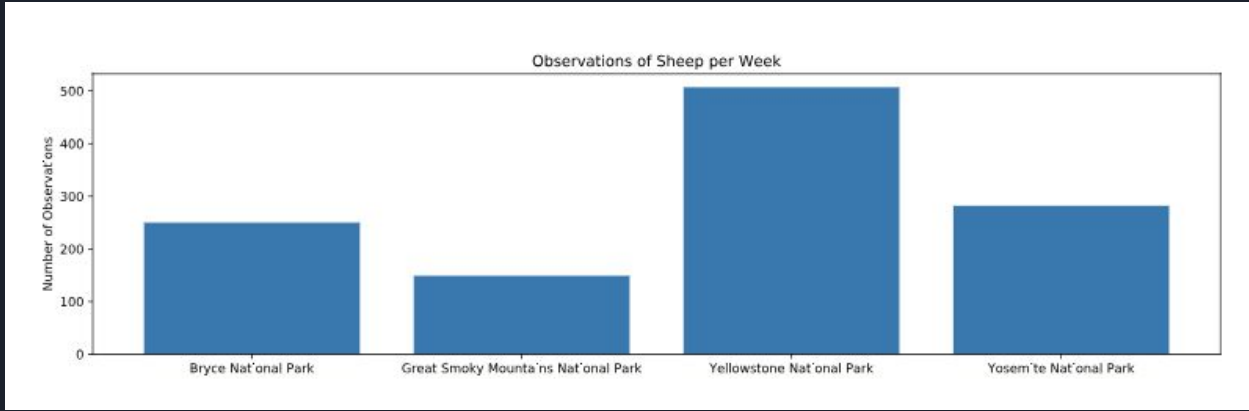
Conservation Status chart/graph

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



This bar graph i've created shows the number of species that are in recovery (4), threatened (10), endangered(15), species of concern (151), and have no intervention (5,363) based on the chart above.

Sheep Observation



This bar graph shows the number of sheep observed for each park tested for foot and mouth disease.

	scientific_name	park_name	observations	category	common_names	conservation_status	is_protected	is_sheep
0	Ovis canadensis	Yellowstone National Park	219	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
1	Ovis canadensis	Bryce National Park	109	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
2	Ovis canadensis	Yosemite National Park	117	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
3	Ovis canadensis	Great Smoky Mountains National Park	48	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
4	Ovis canadensis sierrae	Yellowstone National Park	67	Mammal	Sierra Nevada Bighorn Sheep	Endangered	True	True

Showing the conservation status of the amount of sheep in each park, showing scientific name, and showing they are protected.

Total Amounts of Sheep

	park_name	observations
0	Bryce National Park	250
1	Great Smoky Mountains National Park	149
2	Yellowstone National Park	507
3	Yosemite National Park	282

The chart above shows the total amount of sheep being observed in each park.

Thank You Code Academy!