Biodiversity Capstone Project – Investigating Protected Species

Mike Truscott

species_info.csv

The species_info.csv file contains details about the different species in National Parkes Service National Parks.

The file provides details of both the scientific and common names of each species, the category that these fall into and the conservation status of each species. A sample of the data in the file is provided below:

	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

Biodiversity Capstone Project

species_info.csv

A total of 5,541 unique species are included in the dataset

The Species Type Category is divided into 7 groups:
- Mammal
- Bird
- Reptile
- Amphibian
- Fish
- Vascular Plant
- Nonvascular Plant

The Conservation Status is divided into 5 categories – with the details of the number of species in each of these categories provided below:

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

Biodiversity Capstone Project

species_info.csv

The chart below shows the Conservation Status by Species form the species_info.csv file.

These results were generated from doing a groupby function on the 'conservation_status' column in the file, having ensured that all blank cells were given 'No Intervention' as a status:

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

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


Conservation Status by Species
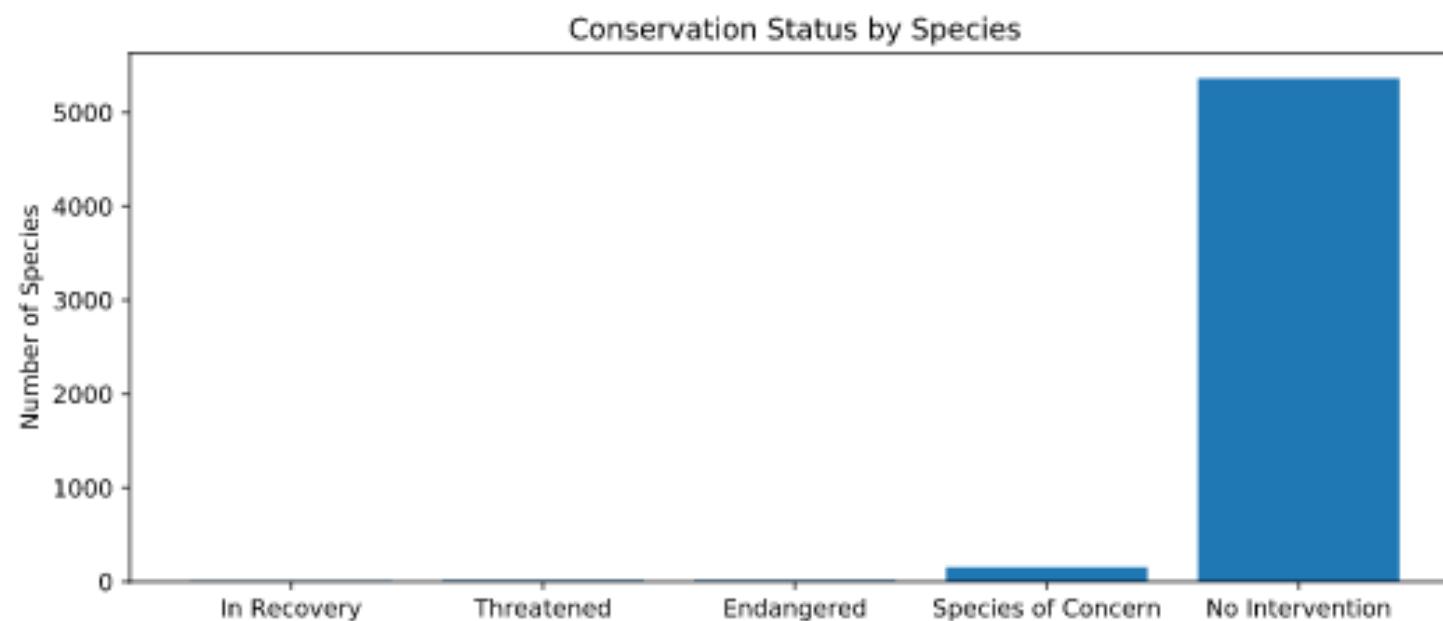
Biodiversity Capstone Project

species_info.csv

Detailed below is an overview of protected and not protected species in the file, grouped by category.

The table shows that Mammals and Birds seem to be the most likely to be endangered.

	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

Biodiversity Capstone Project

species_info.csv

It looks like Mammals are more likely to be endangered than birds, but is it a significant difference?

To determine this we do a Chi-Square test for Significance. First we create a contingency table for Mammals and Birds:

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

Then, after importing the chi-squared test from scipy we run the test:

```
pval = chi2_contingency(contingency)[1]
print(pval)
```

This gives a pval of 0.687594809666 – so there is no significant difference as this is above 0.05.

Biodiversity Capstone Project

species_info.csv

We also have run a significance test between Reptiles and Mammals

We have created a contingency table for Reptile and Mammals:

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

Then, after importing the chi-squared test from scipy we run the test:

```
pval_reptile_mammal = chi2_contingency(contingency_reptile_mammal)[1]
print(pval_reptile_mammal)
```

This gives a pval of 0.0383555902297 – so there is significant difference as this is below 0.05.

Biodiversity Capstone Project

species_info.csv

From these results we can conclude that the difference between birds and mammals was as a result of chance and not significant.

Similarly, from the tests we have performed we calculated that the difference between reptiles and mammals is significant.

We can therefore conclude that certain types of species are more likely to be endangered than others. This is something that should be considered and monitored by conservationists.

Biodiversity Capstone Project

Observations DataFrame

Looking at a file of species sightings at a range of National Parks. A sample of the information in the DataFrame is provided below:

	scientific_name	park_name	observations
0	Vicia benghalensis	Great Smoky Mountains National Park	68
1	Neovison vison	Great Smoky Mountains National Park	77
2	Prunus subcordata	Yosemite National Park	138
3	Abutilon theophrasti	Bryce National Park	84
4	Githopsis specularioides	Great Smoky Mountains National Park	85

Biodiversity Capstone Project

Observations DataFrame

Conservationists are particularly concerned about the movement of sheep in these parks. To determine which of the species in the DataFrame are sheep, we use the following Lambda function and then save to a variable called species_is_sheep:

```python
species['is_sheep'] = species.common_names.apply(lambda x: 'Sheep' in x)

species_is_sheep = species[species.is_sheep]
```

The totals for each of the sheep species are provided in the table below – for this we have slected all those where 'is_sheep' is True and 'category' is mammal :

	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

Biodiversity Capstone Project

Observations DataFrame

We have then performed a groupby to get the total number of sheep observations made at each National Park. The results of this are shown below:

```
obs_by_park =
sheep_observations.groupby('park_name').observations.sum().reset_index()
```

	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

Observations of Sheep per Week

Biodiversity Capstone Project
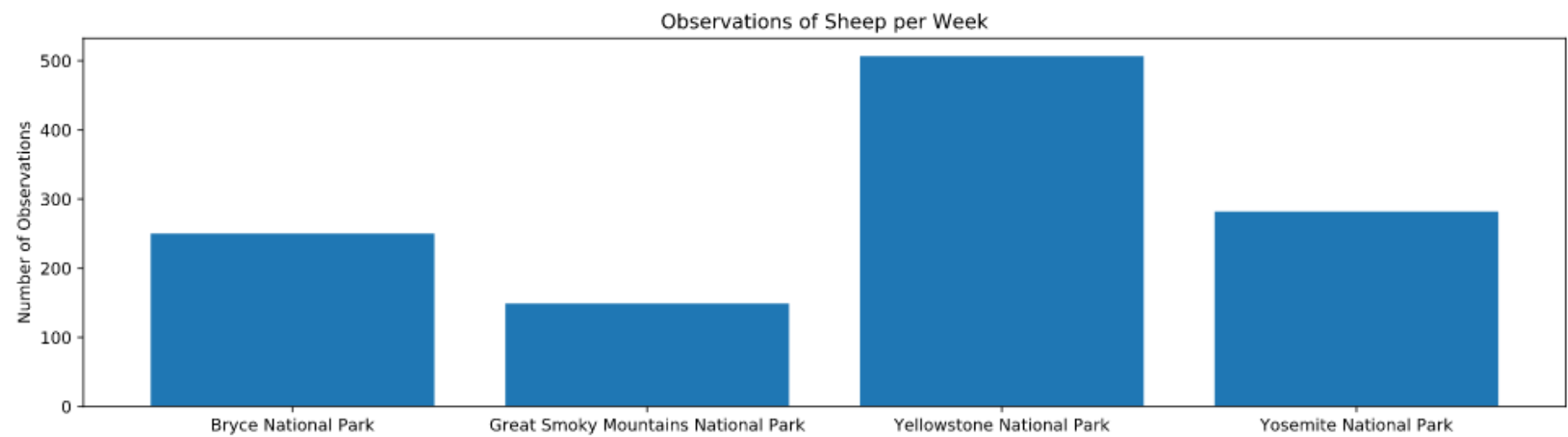
Observations DataFrame

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 or not this program is working. They want to be able to detect reductions of at least 5 percentage point.

The only information that the scientists currently have is that last year it was recorded that 15% of sheep at Bryce National Park have foot and mouth disease.

We have used a sample size calculator to determine the number of sheep from each park that they would need to observe from each park to make sure their foot and mouth percentages are significant.

```
baseline = 15

minimum_detectable_effect = 100 * 0.05 / 0.15
```

The result of this is that the Park Rangers would have to observe at least 510 sheep to ensure that a >5% drop in observed cases of foot and mouth disease in the sheep at Yellowstone was significant.

Using the observation data you analysed earlier, we found that this would take approximately one week of observing in Yellowstone to see that many sheep, or approximately two weeks in Bryce to see that many sheep.

Biodiversity Capstone Project