CODECADEMY

BIODIVERSITY CAPSTONE PROJECT

Doug McClymont

Jan 29, 2019 - Apr 23, 2019 Cohort

SPECIES_INFO DATA DESCRIPTION - 1

After loading the ***species_info.csv*** file some initial descriptive statistics of this DataFrame shows that there are:

	category	scientific_name	common_names	conservation_status
count	5824	5824	5824	191
unique	7	5541	5504	4
top	Vascular Plant	Procyon lotor	Brachythecium Moss	Species of Concern
freq	4470	3	7	161

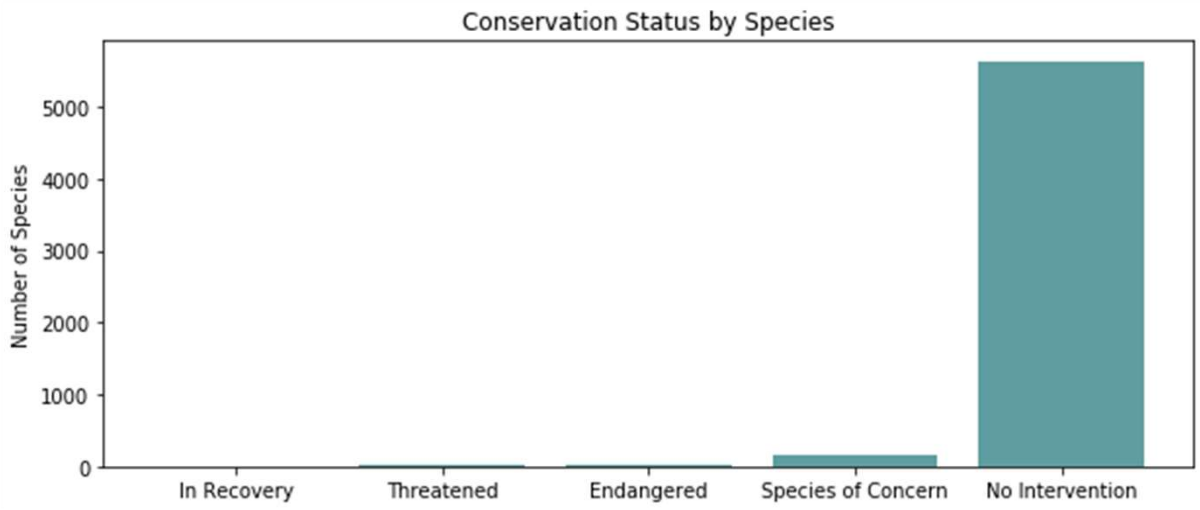
- 7 unique categories (Mammal, Bird, Reptile, Amphibian, Fish, Vascular Plant, Nonvascular Plant).
- 5541 unique Scientific Names.
- 4 different conservation status' (Species of Concern, Endangered' Threatened, In Recovery)

- The conservation status also contains a number of NaN values which are not captured in the descriptive statistics.
- These NaN values were replaced with *'No Intervention'* during this analysis.

SPECIES_INFO DATA DESCRIPTION - 2

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

After the NaN values were replaced in the DataFrame a new DataFrame was created which contained the sum of the different species grouped by conservation status. A bar plot of the conservation status by species is shown below.



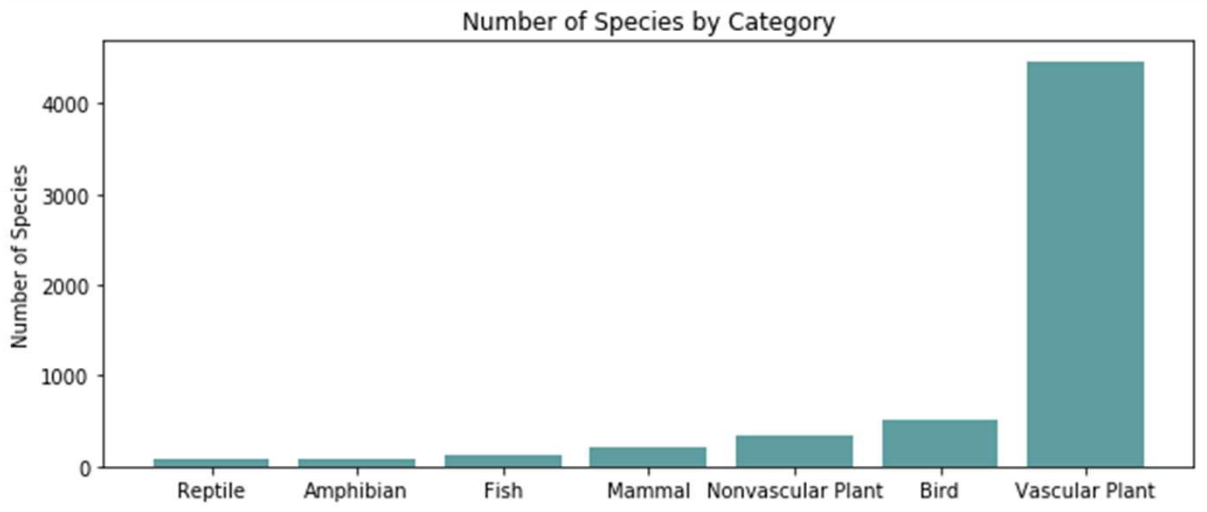
It can be seen that the vast majority of species in the National Park fall into the ***No Intervention*** conservation status .

SPECIES_INFO DATA DESCRIPTION - 3

	category	scientific_name
5	Reptile	79
0	Amphibian	80
2	Fish	127
3	Mammal	214
4	Nonvascular Plant	333
1	Bird	521
6	Vascular Plant	4470

The initial species_info DataFrame was subset a second time to analyse the number of species by category which is shown in the bar plot below.

This bar plot shows that the majority of species in the National Park are ***vascular plants***.

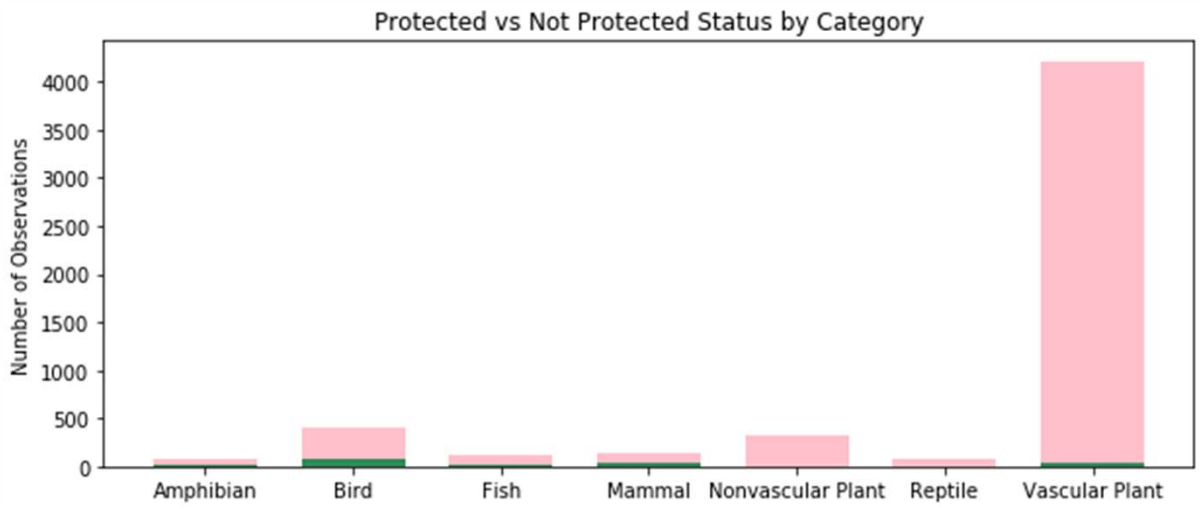


SPECIES_INFO DATA DESCRIPTION - 4

	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

Further analysis was undertaken to understand if certain species are more likely to be endangered.

The stacked bar plot below and data table to the left that mammals have the highest percentage of protected species.



It was proposed that it is more likely species in mammals are endangered than species in birds.

This was evaluated with a significance test

ENDANGERED STATUS SIGNIFICANCE CALCULATIONS

Based on the data table in the previous slide a Chi Squared Analysis was run to evaluate if species of Mammal were more likely to be endangered than species of Bird, and if species of Reptile were more likely to be endangered than species of Mammal.

Mammal vs Bird

	Protected	Not Protected
Mammal	30	146
Bird	75	413

Chi Squared Analysis of the above contingency table gives a p-value of **0.687** which indicates it **is not more likely** that species of mammal are endangered than birds.

Reptile vs Mammal

	Protected	Not Protected
Reptile	5	73
Mammal	30	146

Chi Squared Analysis of the above contingency table gives a p-value of **0.038** which indicates it **is more likely** that species of mammal are endangered than reptile.

RECOMMENDATION FOR CONSERVATIONISTS

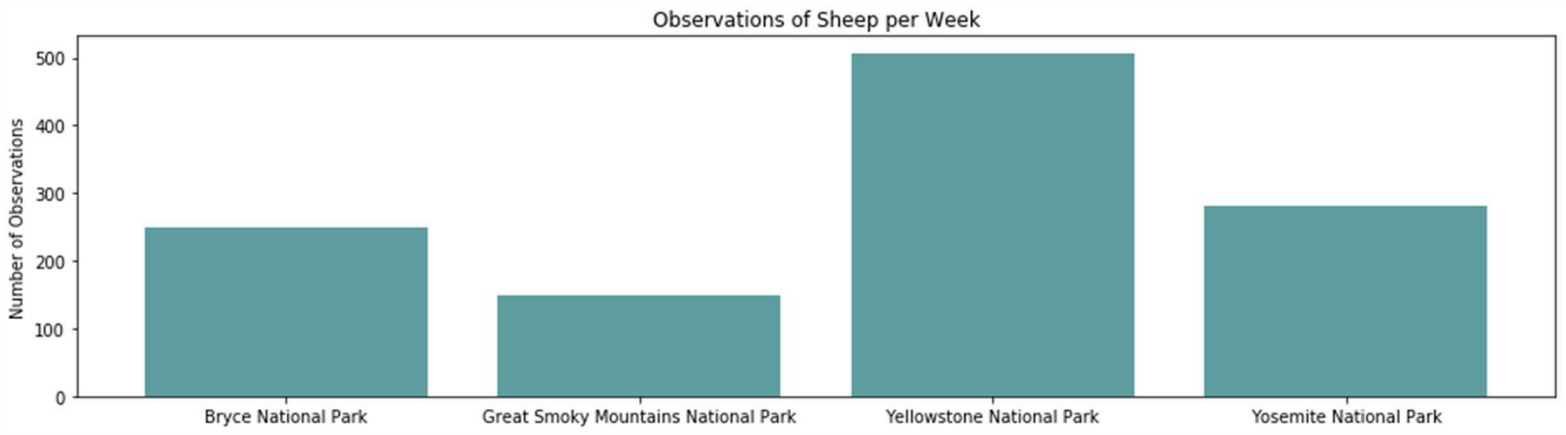
From the analysis of data collected about different species within the National Park presented in the above slides it appears that species within Mammals and within Birds are the most likely to be endangered.

Conservation efforts within the National Park should focus on both Mammal and Bird species.

OBSERVATIONS ON SHEEP

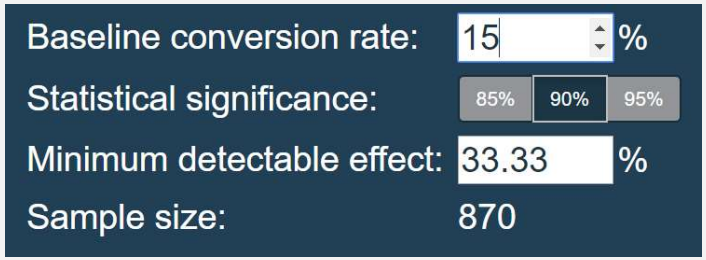
Data gathered from sightings of different species at different National Parks was analysed to understand the number of sheep sightings at different National Parks in the last 7 days. This is shown in the bar plot below.

It is known that 15% of sheep at Bryce National Park have Foot & Mouth Disease and a Sample Size Calculation has been run to understand how many weeks sheep need to be sighted to evaluate the success of a Foot & Mouth reduction programme at Yellowstone National Park.



SAMPLE SIZE CALCULATION

Using the information provided by the park rangers, the sample size calculation was parameterised with the inputs shown to the right.



This gave a required sample size of **870**.

Based on the number of sightings at each National Park in the last 7 days it would take approximately:

- 3½ weeks at Bryce National Park
- 1¾ weeks at Yellowstone National Park

to have enough sightings for a statistically valid result.