Biodiversity in National Parks(EDA)

Ali Pakzad

SPECIES AND CONSERVATION DATA

- The data contains 5541 unique species.
- Vascular Plant is the category with most species(4470).After that bird category has the most species in the national parks with 521 species.
- We have 5 conservation status:

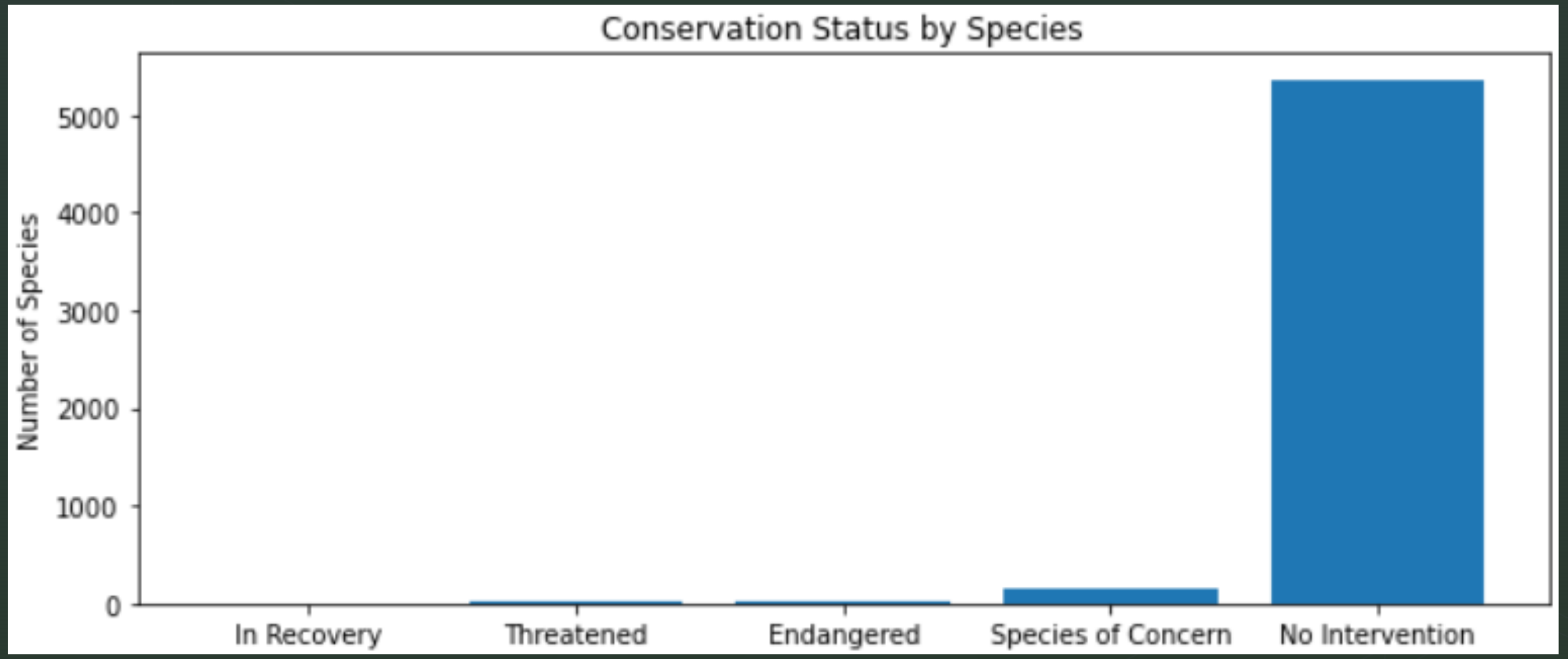
'Species of Concern', 'Endangered',
'Threatened', 'In Recovery'
and None (which I changed it to No Intervention)

```
print(species.category.value_counts())
```

```
Vascular Plant       4470
Bird                  521
Nonvascular Plant     333
Mammal                214
Fish                  127
Amphibian              80
Reptile                79
Name: category, dtype: int64
```

Conservation Status by Species

- The majority of species fall under the No Intervention status (5363 species).



Are certain types of species more likely to be endangered?

	category	not_protected	protected	percent_protected
0	Amphibian	72	7	8.860759
1	Bird	413	75	15.368852
2	Fish	115	11	8.730159
3	Mammal	146	30	17.045455
4	Nonvascular Plant	328	5	1.501502
5	Reptile	73	5	6.410256
6	Vascular Plant	4216	46	1.079305

- Birds, Vascular Plants, and Mammals have a higher number of species protected.
- Mammals and Birds have the highest percentage of being in protection.

Are the differences between species and their conservation status significant?

	protected	not protected
Mammal	30	146
Bird	75	413
Reptile	5	73

I run chi-2 test to see if there is significant difference in conservation status rate between two different species:

- Mammal to Bird chi square test yielded no statistical significance with a p-value of 0.687
- Reptile to Mammal chi square test yielded a statistical significance with a p-value of 0.038 which is less than 0.05

The null hypothesis is rejected and we can conclude that there is association between species category and their conservation status.

Sheep Observations

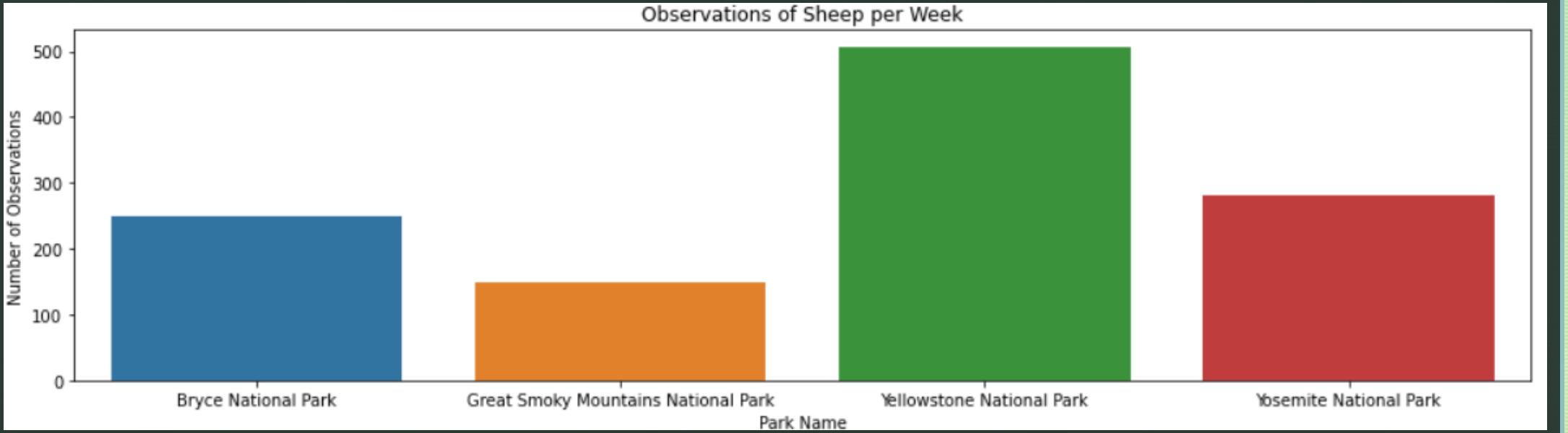
	scientific_name	park_name	observations	category	common_names	conservation_status	is_protected	is_bat	is_sheep
0	Ovis canadensis	Yellowstone National Park	219	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	False	True
1	Ovis canadensis	Bryce National Park	109	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	False	True
2	Ovis canadensis	Yosemite National Park	117	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	False	True
3	Ovis canadensis	Great Smoky Mountains National Park	48	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	False	True
4	Ovis canadensis sierrae	Yellowstone National Park	67	Mammal	Sierra Nevada Bighorn Sheep	Endangered	True	False	True
5	Ovis canadensis sierrae	Yosemite National Park	39	Mammal	Sierra Nevada Bighorn Sheep	Endangered	True	False	True
6	Ovis canadensis sierrae	Bryce National Park	22	Mammal	Sierra Nevada Bighorn Sheep	Endangered	True	False	True
7	Ovis canadensis sierrae	Great Smoky Mountains National Park	25	Mammal	Sierra Nevada Bighorn Sheep	Endangered	True	False	True
8	Ovis aries	Yosemite National Park	126	Mammal	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	False	True
9	Ovis aries	Great Smoky Mountains National Park	76	Mammal	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	False	True
10	Ovis aries	Bryce National Park	119	Mammal	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	False	True
11	Ovis aries	Yellowstone National Park	221	Mammal	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	False	True

I select the rows of species where is_sheep is True and category is Mammal. Save the results to the variable sheep_species.

Then merge sheep_species with `observations` to get a DataFrame with observations of sheep.

Sheep Observations

Here is the total number of sheep observations at each national park:



	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

Disease Reduction Program

- Our scientists know that 15% of sheep at Bryce National Park have foot and mouth disease.
- Scientists want to be able to detect reductions of at least 5 percentage point.

```
minimum_detectable_effect = (15 - 10) / 15 * 100

minimum_detectable_effect
```

```
33.33333333333333
```

```
baseline_conversion_rate = 15
statistical_significance = 90
```

- By using Codecademy sample size calculator, our sample size is 870.

• We need 3.5 weeks(870/250) in Bryce National Park and 1.7 weeks(870/507) in Yellowstone National Park to obtain enough number of samples.

Conclusions

- The most spotted species in Bryce National Park is `Valerianella radiata`.
- The most spotted species in Great Smoky Mountains National Park is `Sonchus asper ssp. asper`.
- The most spotted species in Yellowstone National Park is `Lycopodium tristachyum`.
- The most spotted species in Yosemite National Park is `Ivesia shockleyi var. shockleyi`.
- Yellowstone National Park has the highest and Great Smoky Mountains National Park has the lowest number of mammals.
- Yellowstone National Park has the highest number of endangered species(1558).We can see that all four parks have same number of different endangered species type(15)
- Yellowstone National Park has the most number of observed endagered Vascular plants(63).
- The most of species are not part of conservation(5633).There are 161 species of concern, 16 endangered, 10 threatened, and 4 in recovery.
- Mammals and Birds had the highest percentage of being in protection.
- While mammals and Birds did not have significant difference in conservation percentage, mammals and reptiles exhibited a statistically significant difference.
- The most prevalent animal in the data set is bat.
- Yellowstone National Park has the largest number of bats observations by 8362 and Great Smoky Mountains National Park has the lowest bats with 2411.