Biodiversity Capstone project

Codecademy course

Input data: species_info.csv

File species_info.csv contains details about analyses species, such as:

- category
- scientific_name
- common_names
- conservation_status

There are 5541 unique species recorded in that file.

Input data: conservation statuses

Species have following conservation statuses linked to each of them:

- Species of Concern: declining population or appears to be in need of conservation.
- Threatened: vulnerable to endangerment in the near future.
- Endangered: seriously at risk of extinction.
- In Recovery: formerly Endangered, but currently not in danger of extinction throughout all or a significant portion of its inhabitable range.
- No Intervention: when none of above mentioned statuses are applied

Input data: species that are endangered

Species have following categorization:

- Mammal
- Bird
- Reptile
- Amphibian
- Fish
- Vascular Plant
- Nonvascular Plant

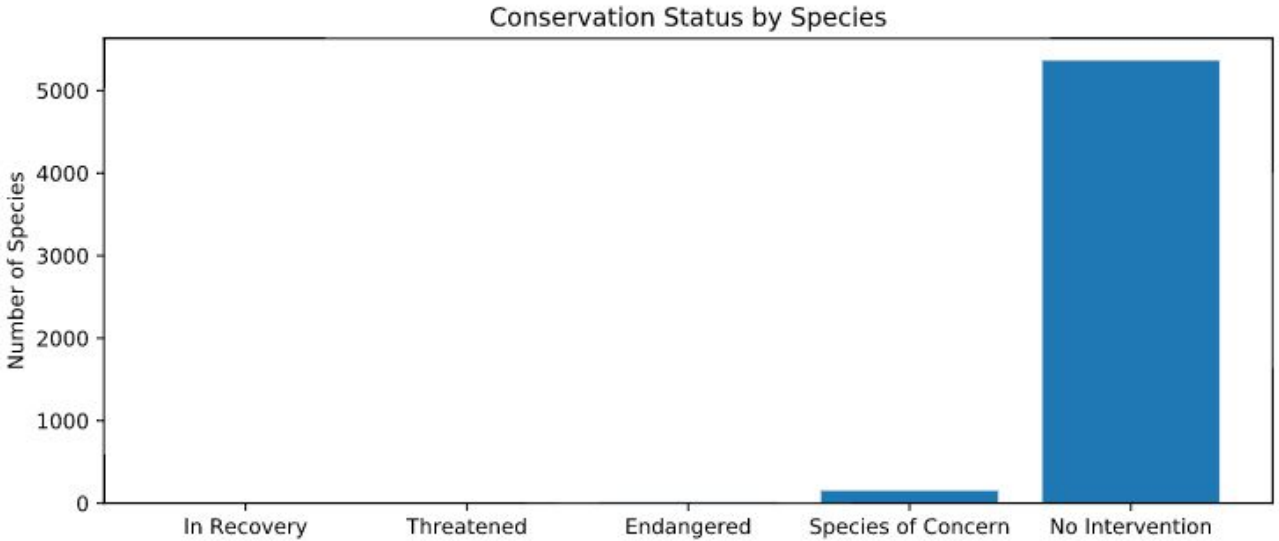
Analysis: conservation status

In numbers:

	Conservation Status	Count of species
0	Endangered	15
1	In Recovery	4
2	No Intervention	5363
3	Species of Concern	151
4	Threatened	10

Analysis: conservation status

In graph:



As we can see from this graph less endangered or species where no intervention is needed are a significant majority.

Analysis: species that are endangered

From input data we can see that some species are endangered more than others, however to confirm this, we have to conduct additional checks.

	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

Analysis: species that are endangered

In order to determine of data is consistent I took two sets of data for chi-squared test.

Compared Mammal vs Birds

	protected	not-protected
Mammal	30	146
Bird	75	413

Compared Mammal vs Reptile

	protected	not-protected
Mammal	30	146
Reptile	5	73

Analysis: species that are endangered

When we ran our chi-squared test, we found a p-value of ~0.688, so we can conclude that the difference between the percentages of protected birds and mammals **is not** significant and is a result of chance.

But, when we compared the percentages of protected reptiles and mammals and ran the same chi-squared test, we calculated a p-value of ~0.038, which **is** significant.

Therefore, we can conclude that certain types of species **are** more likely to be endangered than others.

Input data: observations.csv

File observations.csv contains data of sightings of different species at several national parks for the past 7 days.

Contains following data:

- Species name
- Park in which it was observed
- Amount of sightings

Observations.csv was **merged** with species.csv in order to have full stack of data for analysis.

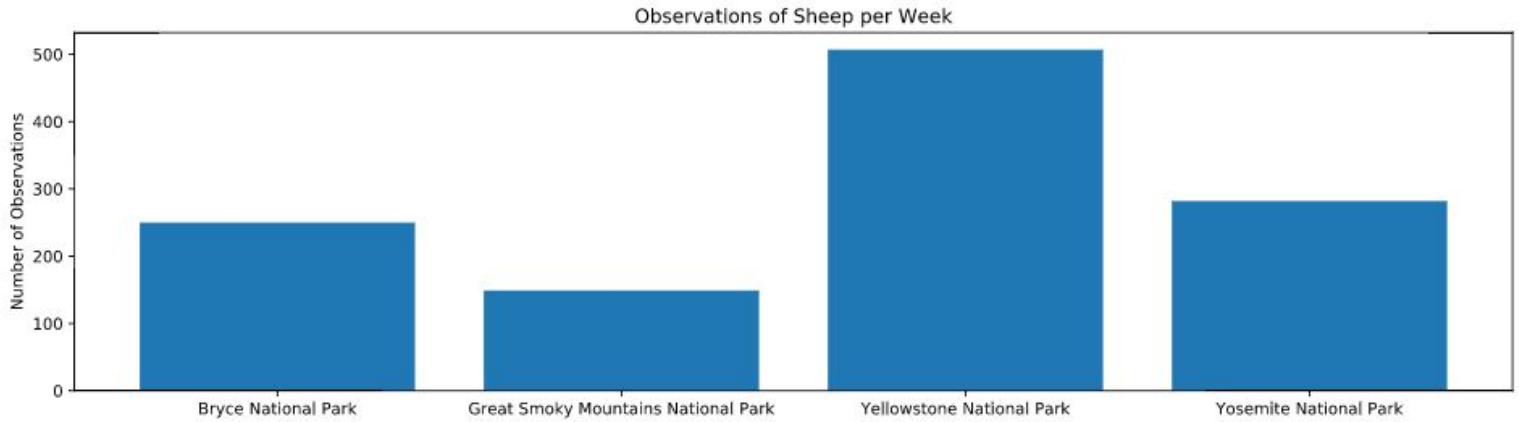
Analysis: “Sheep” sightings in parks

Since “Sheep” is a common name for Mammal in our data sheet, we used it to analyse sightings per each 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

Analysis: “Sheep” sightings in parks

Graphical display of sheep sightings in parks:



“Sheeps” are primarily seen in Yellowstone National Park, therefore we can conclude that Mammals are also located in Yellowstone.

Foot and Mouth Reduction Effort

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.

Requirement was to calculate the number of sheep that they would need to observe from each park to make sure their foot and mouth percentages are significant.

Input: Foot and Mouth Reduction Effort

- 15% of sheep at Bryce National Park have foot and mouth disease
- Requirement is to drop this to **10%**
- Minimum detectable effect is 33%

- Default level of significance was used (90%)

Analysis: Foot and Mouth Reduction Effort

Using Sample Size Survey Calculator we were able to determine needed sample size.

According to this outcome weekly rate of sightings:

- Bryce National Park will collect needed data in ~4 weeks
- Yellowstone National Park will collect needed data in ~2 weeks

Thank you