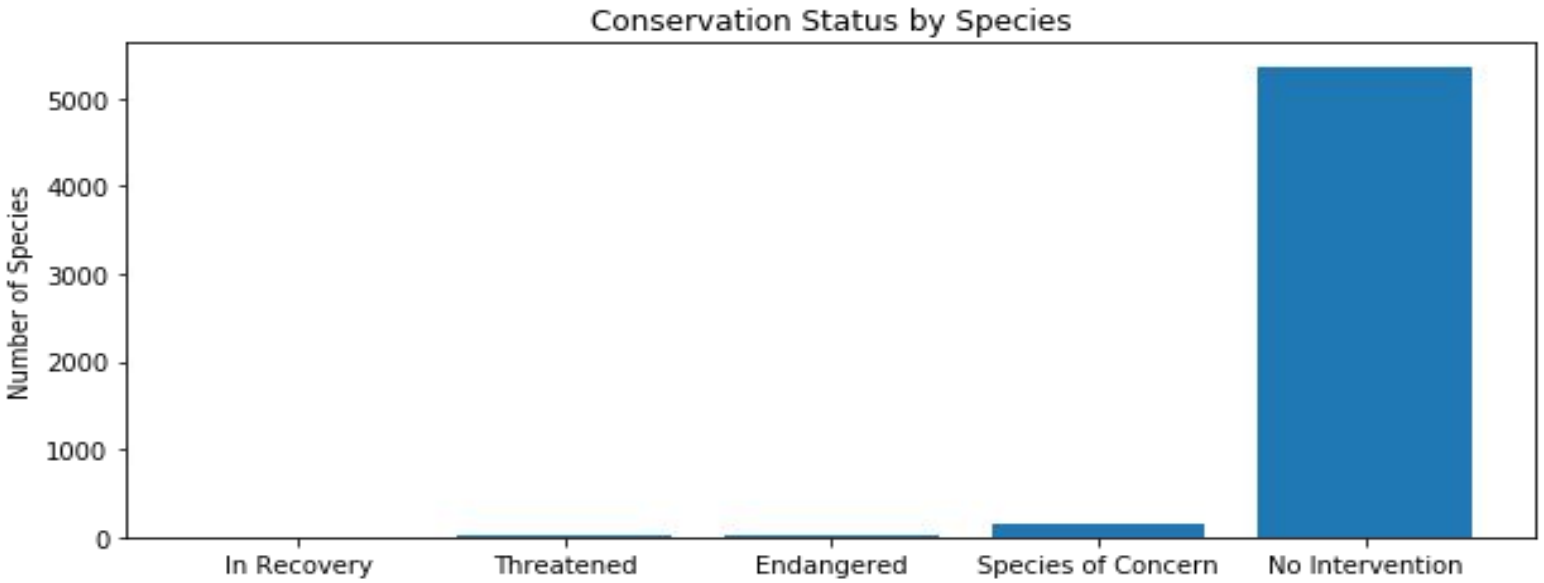
Codecademy Capstone Project - Biodiversity

By Hang Dao

Information about species.csv

1. It is a CSV - Comma Separated Values, so we need to use Pandas to read it
2. It contains information about category, scientific names, common name, and conservation status
3. There are 5541 species including in this CSV. They are belonged to 7 categories: “Mammal”, "Bird", "Reptile", "Amphibian", "Fish", "Vascular Plant", "Nonvascular Plant"
4. There are 4 conservation status: “ species of concern”, “endangered”, “threatened”, and ” in recovery.” The species not having intervention are marked as nan (none). 5363 species do not need intervention, which is the majority (Figure 1)

Figure 1 - Conservation Status by Species



Endangered Status between Different Categories

1. **Birds vs. Mammal**

- Alternative Hypothesis: Species in Mammal categories are more likely to be endangered than those in birds
- Null Hypothesis: Species in Mammal categories are NOT more likely to be endangered than those in birds
- Since we compared between 2 categories (Mammal and Bird) with 2 sets of data (protected, not protected), we need to use Chi-Squared Test
- Pval = 0.44 is greater than 0.05 => insignificant => we failed to reject the null hypothesis
- Conclusion: Species in Mammal categories are NOT more likely to be endangered than those in birds

Endangered Status between Different Categories

2. Reptile vs. Mammal

- Alternative Hypothesis: Species in Reptile categories are more likely to be endangered than those in Mammal.
- Null Hypothesis: Species in Reptile categories are NOT more likely to be endangered than those in Mammal.
- Since we compared between 2 categories (Reptile and Mammal) with 2 sets of data (protected, not protected), we need to use Chi-Squared Test
- Pval = 0.02 is less than 0.05 => significant => we reject the null hypothesis
- Conclusion: Species in Reptile categories are more likely to be endangered than those in Mammal.

Endangered Status between Different Categories

3. Recommendation:

- Between mammal and reptile, conservationists need to pay more attention to reptiles, because they are more likely to be endangered.

Foot and Mouth Disease Case Study

1. **Bryce National Park**

- Baseline Conversion Rate = 15%
- Minimum Detectable Effect = 5*100/15 = 33.33%
- Significance = 90%
- After calculating, sample size is 870
- Observation is 250 subjects per week (Figure 2)
- Time period to observe this sample size is 870/250 = 3.5 weeks

Foot and Mouth Disease Case Study

2. Yellowstones National Park

- Baseline Conversion Rate = 10%
- Minimum Detectable Effect = 5*100/10 = 50%
- Significance = 90%
- After calculating, sample size is 610
- Observation is 507 subjects per week (Figure 2)
- Time period to observe this sample size is 610/507 = 1.2 weeks

Figure 2 - Observations of Sheep per Week