Project Scoping

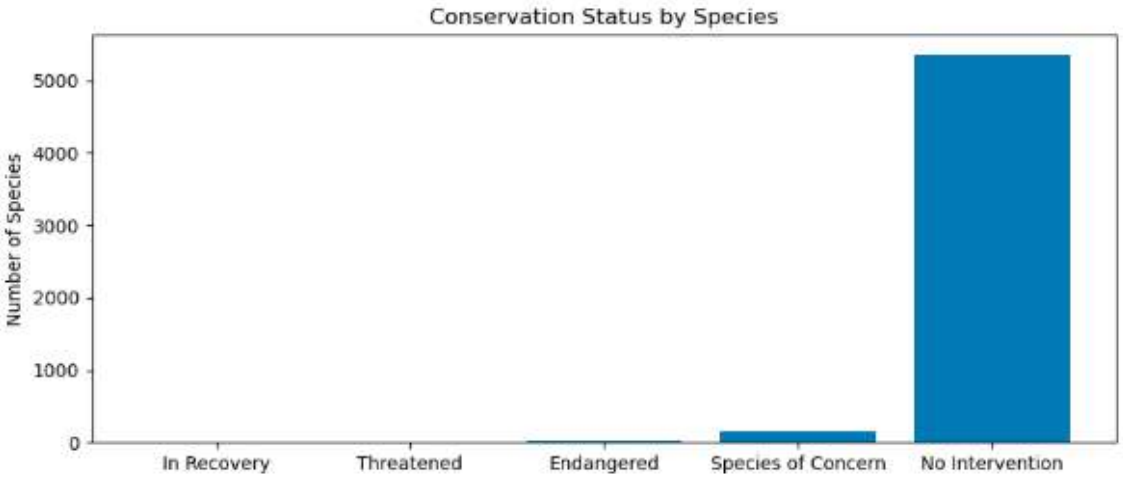
- **Goal:** To analyse species biodiversity and conservation status across many national parks to inform conservation efforts.
- **Data:**
    - species_info.csv
    - observations.csv
- **Analytical Steps:**
    - Data cleaning (fill missing values in conservation status)
    - Summary statistics on species by category and conservation status
    - Statistical tests to assess significance of protection status by category
    - Focus on specific species groups and their observation patterns
    - Calculate sample sizes for monitoring studies

2. Data Exploration and Explanation

Answer these questions:

- **What is the distribution of conservation_status for animals?**
    - Most species have "No Intervention" conservation status.

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



- **Are certain types of species more likely to be endangered?**

  Birds and mammals have a higher proportion of protected species while most plants (especially vascular) have very low protection rates.

	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

- **Are the differences between species and their conservation status significant?**
    - Chi-square test indicates significant difference in protection between mammals and reptiles but not between mammals and birds.

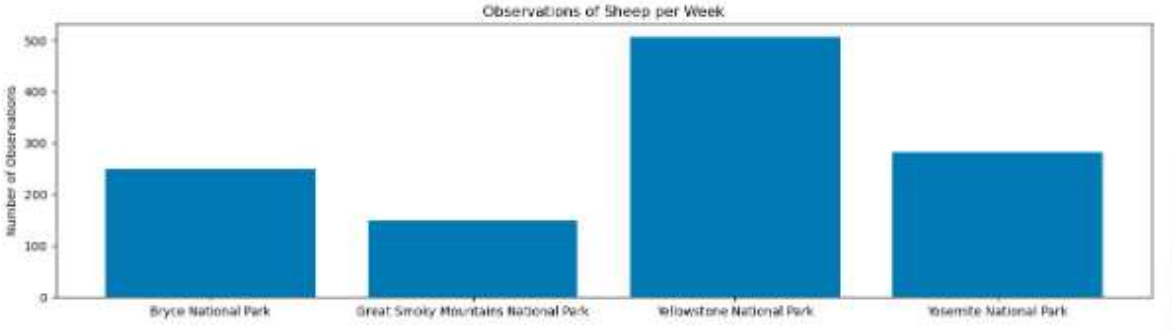
Mammals vs. Birds:

p-value ≈ 0.688 → Not significant

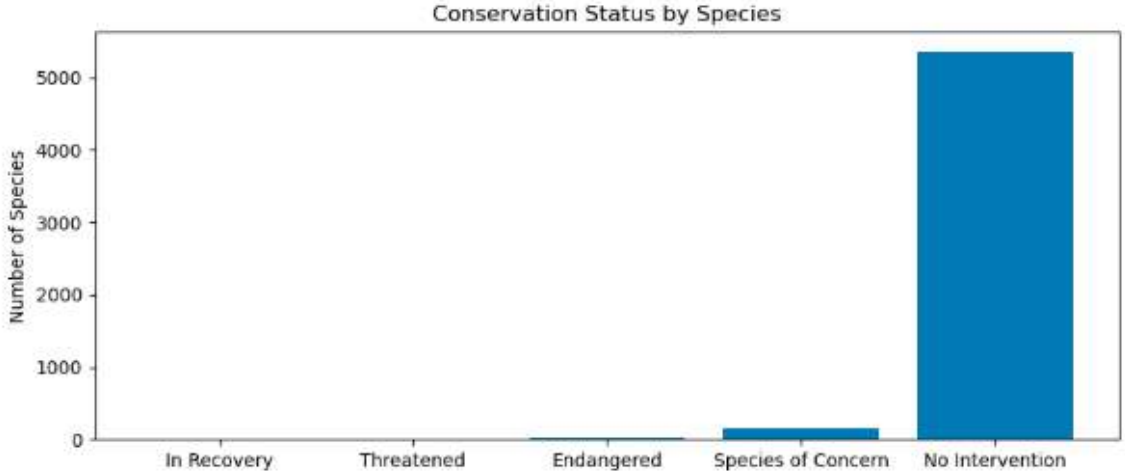
Mammals vs. Reptiles:

p-value ≈ 0.038 → Statistically significant

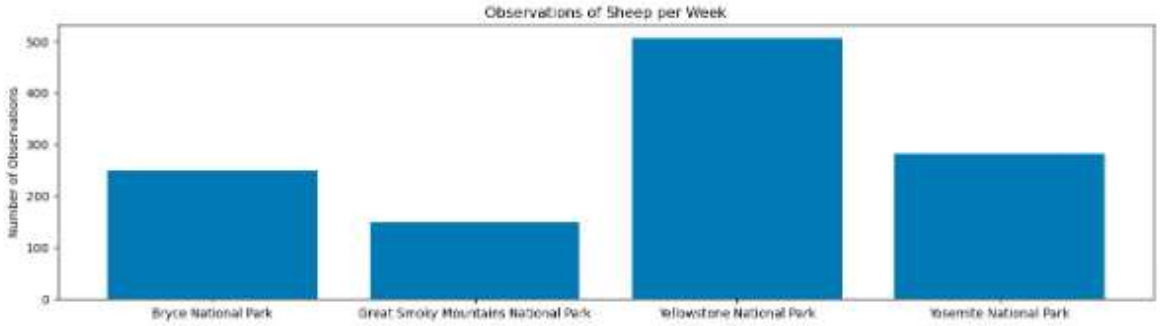
- **Which species were spotted the most at each park?**
    - Sheep species. Yellowstone National Park has the highest number of sheep observations and the Great Smoky Mountains National Park had the fewest sighting.



- **Visualizations:**
    - Bar plot of species counts by conservation status.



- Bar plot of sheep observations per park.



3. Conclusions

- **What did you learn?**
  - Majority of species are not currently protected.
  - Mammals have the highest protection percentage.
  - A chi-squared test showed a significant difference in protection rates between mammals and reptiles, suggesting mammals are more often targeted for conservation.
  - Certain species need focused conservation efforts.

- **Key findings:**
  - Significant difference in conservation status distribution between categories.
  - Yellowstone has the highest number of sheep observations.
  - Sample size calculations indicate monitoring will require many weeks of observations depending on the park.