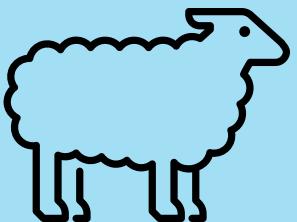
A large, abstract graphic on the left side of the slide consists of numerous 3D rectangular blocks of various sizes and colors, including red, orange, yellow, teal, light blue, and white. These blocks are stacked and arranged in a way that creates a sense of depth and perspective, resembling a city skyline or a complex architectural model.

# **BIODIVERSITY ANALYSIS & CONSERVATION PRIORITIES IN U.S. NATIONAL PARKS**

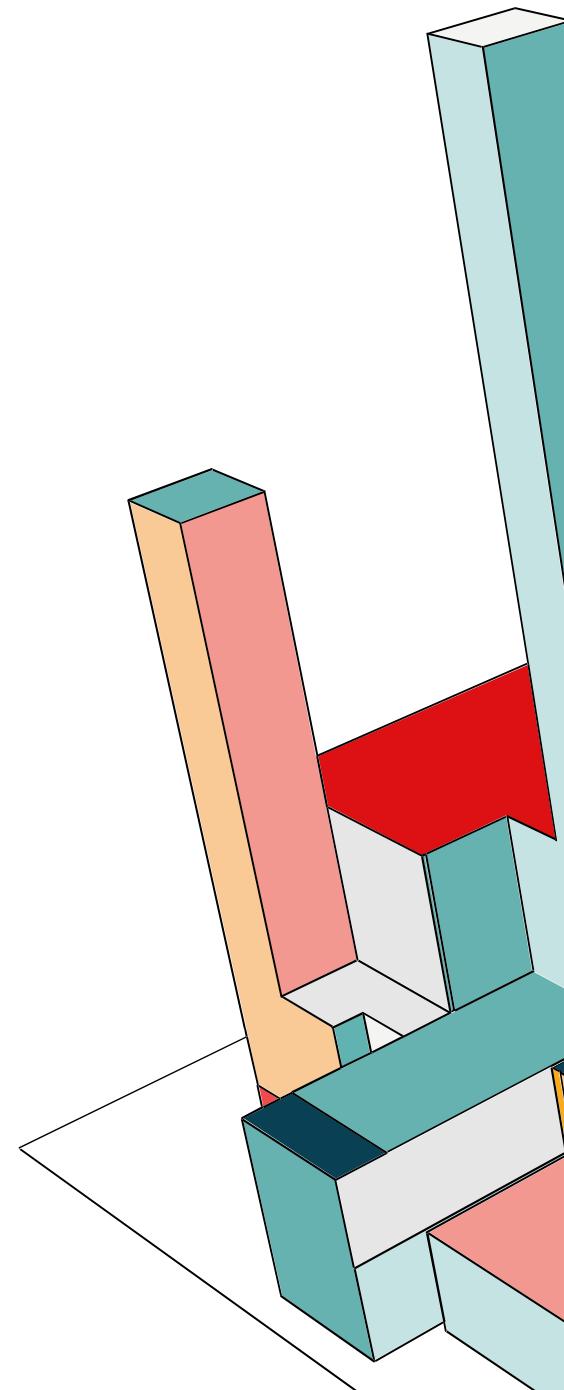
- Saurie Padayachee
- July 23, 2025
- MTN SA Data Science Programme



# PROJECT SCOPE

This project investigates biodiversity within national parks, with a focus on species conservation. The main goals are:

1. To understand the distribution of species by conservation status.
  2. To determine whether certain types of species are more likely to be endangered.
  3. To examine observations of sheep across different national parks.
  4. To provide data-driven recommendations for wildlife conservation, particularly for species at risk.
- The project involves scoping the research questions, preparing and exploring data, conducting statistical tests for significance, and visualizing key findings. This structured approach helps guide analysis and ensures the insights are actionable and well-supported.



## PRIMARY DATA SOURCES:

**species\_info.csv**: contains information about species found in national parks, including category (e.g., mammal, bird), scientific name, common names, and conservation status.

**observations.csv**: contains weekly observation counts of species recorded in four major u.s. national parks.





# KEY DATA INSIGHTS:

## Key Data Insights:

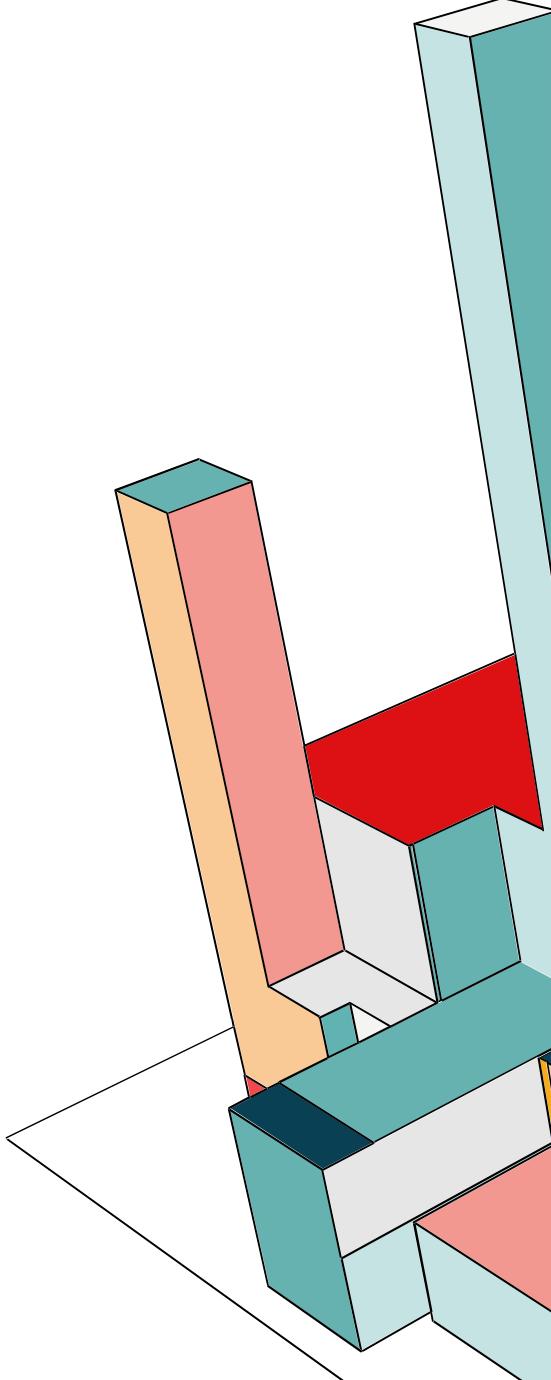
- **Species Count:** 5,541 unique species were recorded.
- **Categories:** Species are categorized into Mammals, Birds, Reptiles, Fish, Amphibians, Vascular Plants, and Nonvascular Plants.
- **Conservation Status:** The dataset includes labels like "Endangered," "Threatened," "Species of Concern," "In Recovery," and those with "No Intervention."

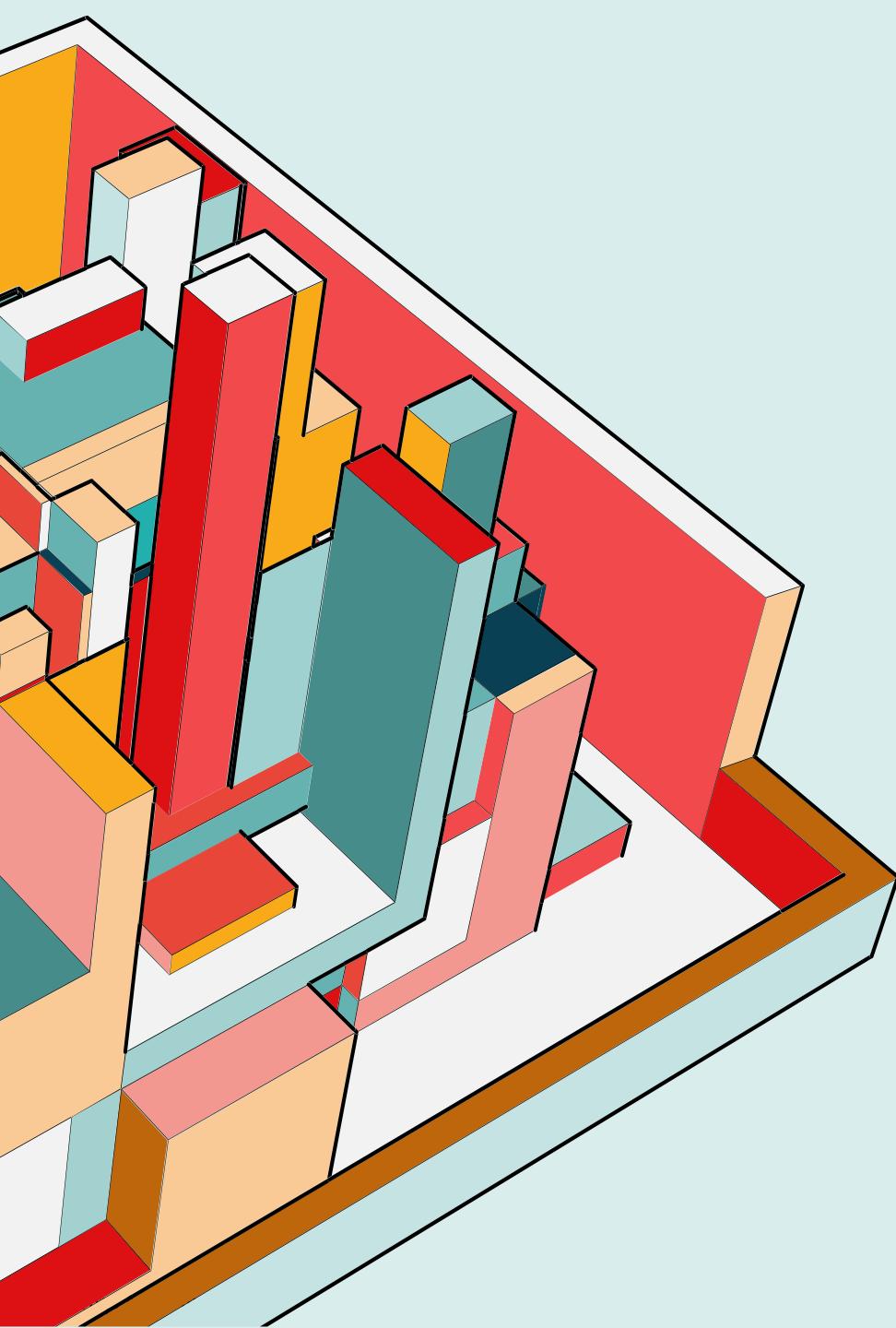
## Sheep-Specific Analysis:

- Using string matching on the common\_names column, species containing "Sheep" were identified.
- Only **mammal sheep species** were retained (e.g., *Ovis aries*, *Ovis canadensis*, *Ovis canadensis sierrae*).
- Merging `sheep_species` with the observations data resulted in a DataFrame (`sheep_observations`) focused on sheep sightings.

# **ANALYTIC STEPS FOR THE BIODIVERSITY PROJECT**

1. Clean and Prepare the Data
2. Visualize Conservation Status
3. Compare Protection Across Categories
4. Perform Significance Tests
5. Sample Size Calculation for Disease Study



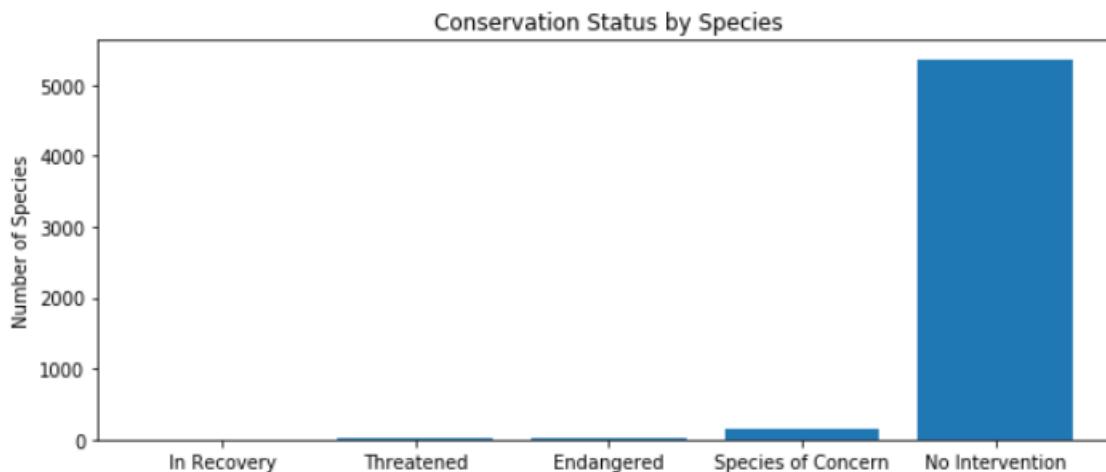


# EXPLORING THE DATA

# WHAT IS THE DISTRIBUTION OF CONSERVATION\_STATUS FOR ANIMALS?

A vast majority of species have no active conservation measures, while a small number fall under protection categories like "Endangered" or "Threatened." The category "Species of Concern" represents those being monitored but not formally protected.

	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?

Mammals and birds are more likely to be under protection, while most plants (especially vascular) have very low protection rates. This could reflect prioritization based on mobility, public awareness, or ecological impact.

[20]:

	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?

To test if differences in protection rates are **statistically significant**, we ran chi-squared tests on contingency tables:

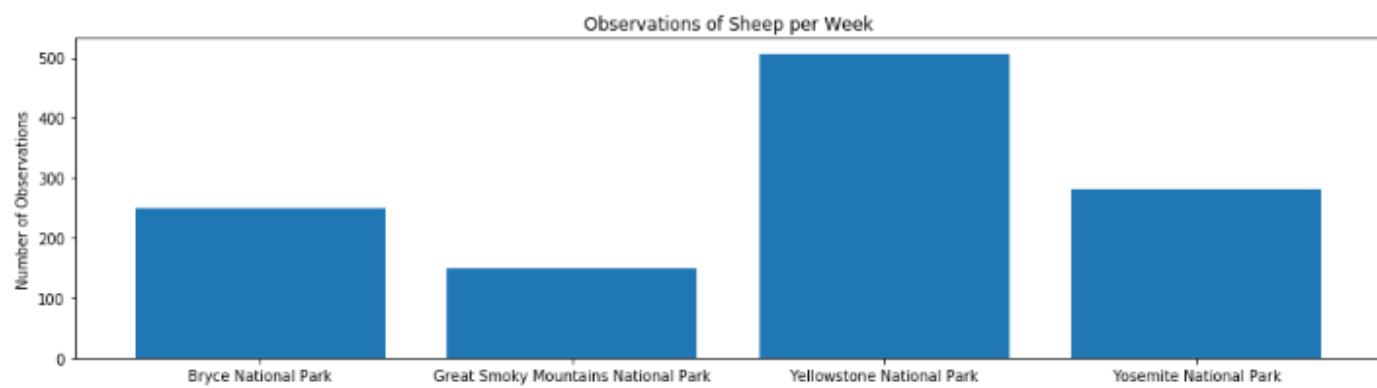
- **Mammals vs. Birds:**
  - p-value  $\approx 0.688 \rightarrow \text{Not significant}$
- **Mammals vs. Reptiles:**
  - p-value  $\approx 0.038 \rightarrow \checkmark \text{ Statistically significant}$

There is a **significant difference** in conservation rates between **mammals and reptiles**, with mammals being more likely to receive protection. However, the difference between mammals and birds was **not significant**.

# WHICH SPECIES WERE SPOTTED THE MOST AT EACH PARK?

Focusing on **sheep species**, we merged their records with the observations dataset and calculated weekly observations per park:

**Yellowstone** has the highest number of sheep observations, making it a strong candidate for any future field studies (like those related to disease or population health). **Great Smoky Mountains** had the fewest sightings.



# CONCLUSION

- This project highlighted several meaningful findings:
- **Protection Trends:** While a vast majority of species have no active conservation measures, Mammals have a higher proportion of protected species compared to plants and reptiles.
- **Statistical Insight:** A chi-squared test showed a significant difference in protection rates between mammals and reptiles ( $p \approx 0.038$ ), suggesting mammals are more often targeted for conservation.
- **Sheep Observations:** Yellowstone recorded the highest weekly sheep sightings, while Great Smoky Mountains had the fewest. This can help determine where conservation efforts might be most feasible or needed.



# THANK YOU

