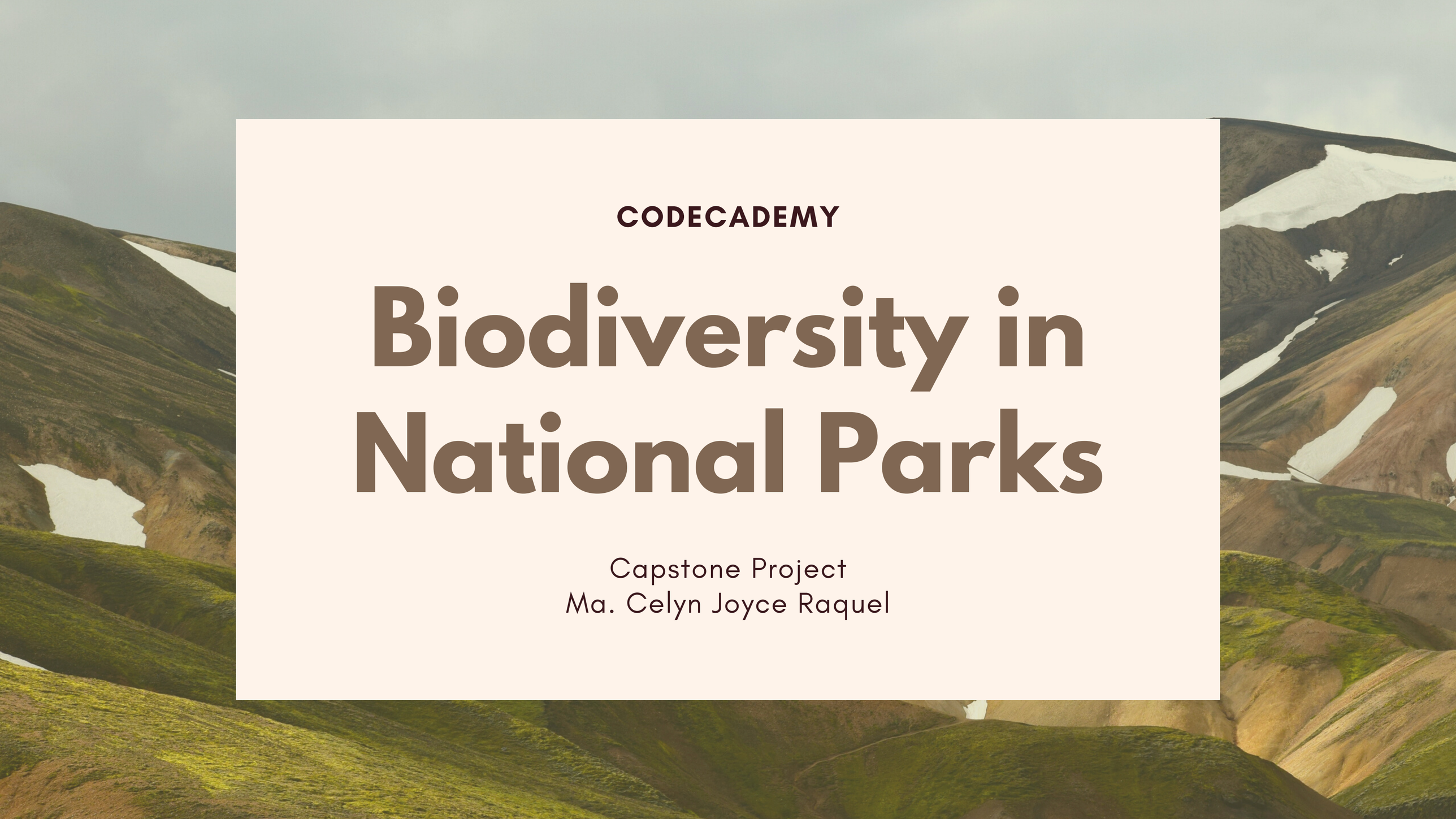


CODECADEMY

Biodiversity in National Parks

Capstone Project

Ma. Celyn Joyce Raquel

CONTENTS

CSV SOURCE

NaN values are present

	category	scientific_name	common_names	conservation_status
0	Mammal	Clethrionomys gapperi gapperi	Gapper's Red-Backed Vole	NaN
1	Mammal	Bos bison	American Bison, Bison	NaN
2	Mammal	Bos taurus	Aurochs, Aurochs, Domestic Cattle (Feral), Dom...	NaN
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	NaN
4	Mammal	Cervus elaphus	Wapiti Or Elk	NaN

SPECIES_INFO.CSV

DATA CLEANING

In order to analyze the data properly, the data was cleaned using Pandas. The NaN values previously shown in the 'conservation_status' column were filled with the string 'No Intervention'.

DATA ANALYSIS

1

CONSERVATION STATUS BY SPECIES

How many species are endangered, threatened, species of concern, in recovery and protected?

2

PROTECTION PER CATEGORY

Are certain types of species more likely to be endangered?

3

SIGNIFICANCE TESTS

Are mammals more endangered than birds?
Are reptiles more endangered than mammals?

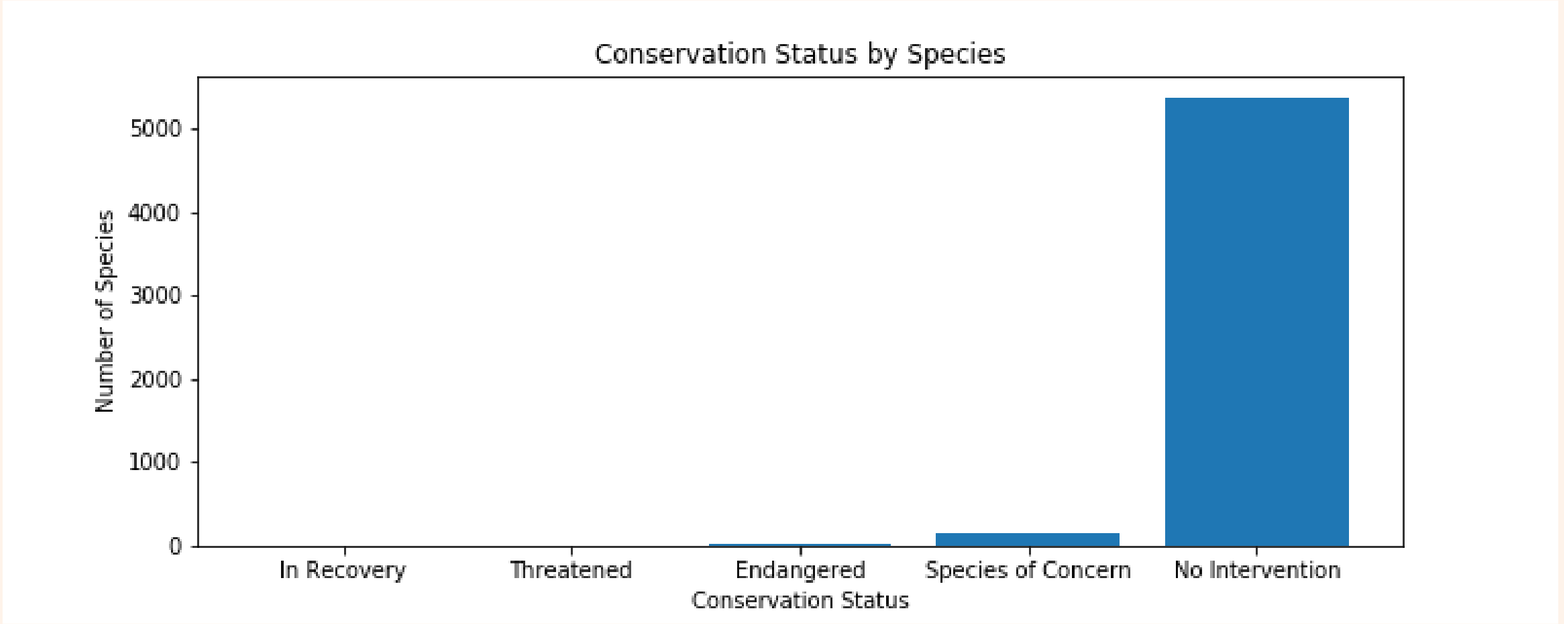
4

SHEEP PER PARK

How many sheep were sighted per park?

CONSERVATION STATUS BY SPECIES

The conservation statuses are arranged ascendingly by number of species. Majority of the species are protected and under no intervention.



PROTECTION PER CATEGORY

The most endangered category in the data is the mammal, while the least endangered is the vascular plant.

	category	not_protected	protected	percent_pivoted
0	Amphibian	7	72	0.911392
1	Bird	75	413	0.846311
2	Fish	11	115	0.912698
3	Mammal	30	146	0.829545
4	Nonvascular Plant	5	328	0.984985
5	Reptile	5	73	0.935897
6	Vascular Plant	46	4216	0.989207

SIGNIFICANCE TESTS

ARE MAMMALS MORE ENDANGERED THAN BIRDS?
ARE REPTILES MORE ENDANGERED THAN MAMMALS?

SIGNIFICANT DIFFERENCE?

NO

MAMMALS VS BIRDS

Using the chi square test from the SciPy library, and with a resulting p-value of 0.6875948096661336, the difference is not significant on a 95% confidence level.

SIGNIFICANT DIFFERENCE?

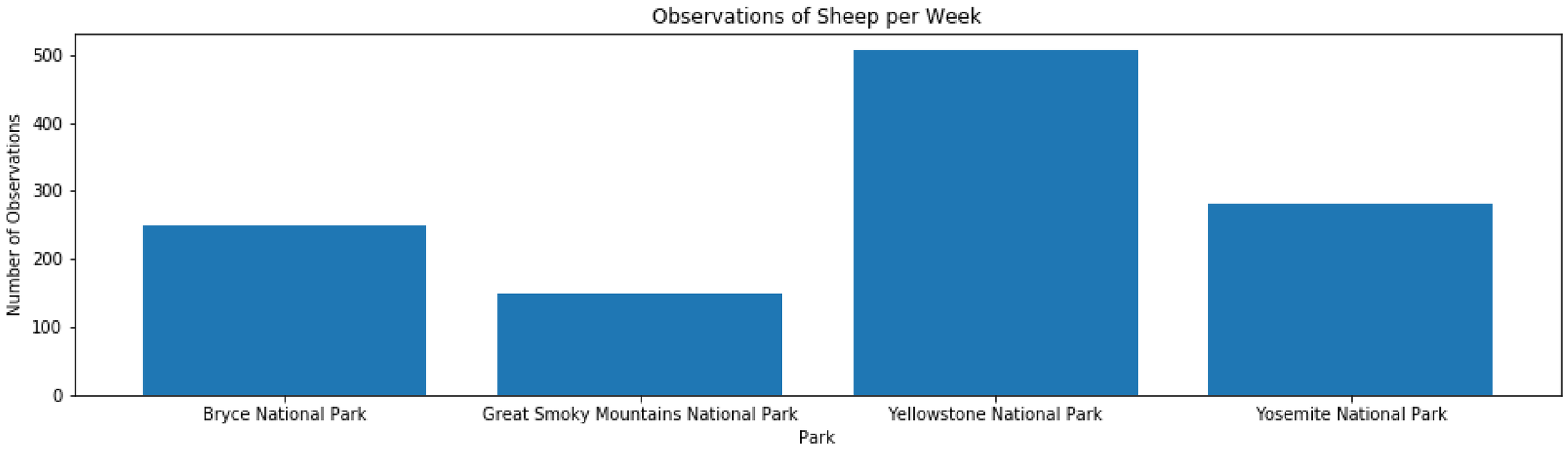
YES

REPTILES VS MAMMALS

Using the chi square test from the SciPy library, and with a resulting p-value of 0.038355590229698, the difference is significant on a 95% confidence level.

SHEEP PER PARK

Yellowstone National Park has the most sheep, while Great Smoky Mountains National Park has the least.



Foot and Mouth Disease in Sheep

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

Sample Size Calculation

Using the Codecademy sample size calculator, the following values were used:

Baseline Conversion Rate: 15%

Minimum Detectable Effect: 33%

Statistical Significance: 95%

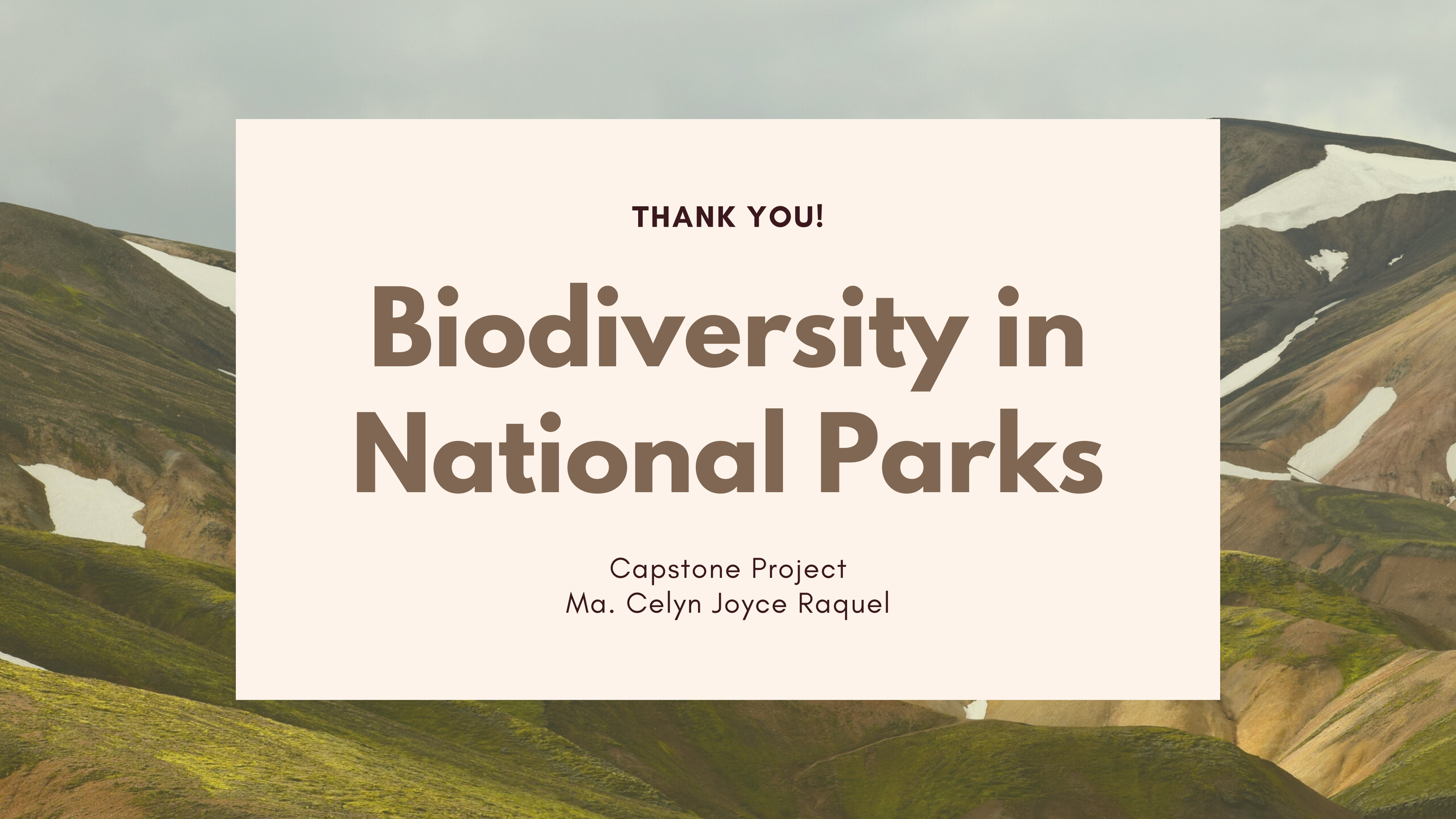
This calculation resulted to a **sample size of 890**.

No. of Weeks Calculation

With a sample size of 890 for each park, sheep populations of 250 and 507 in Bryce National Park and Yellowstone National Parks, respectively, the following number of weeks are needed for each park to gather enough data for the experiment:

Bryce National Park: almost 4 weeks

Yellowstone National Park: almost 2 weeks



THANK YOU!

Biodiversity in National Parks

Capstone Project
Ma. Celyn Joyce Raquel