

# Biodiversity Project

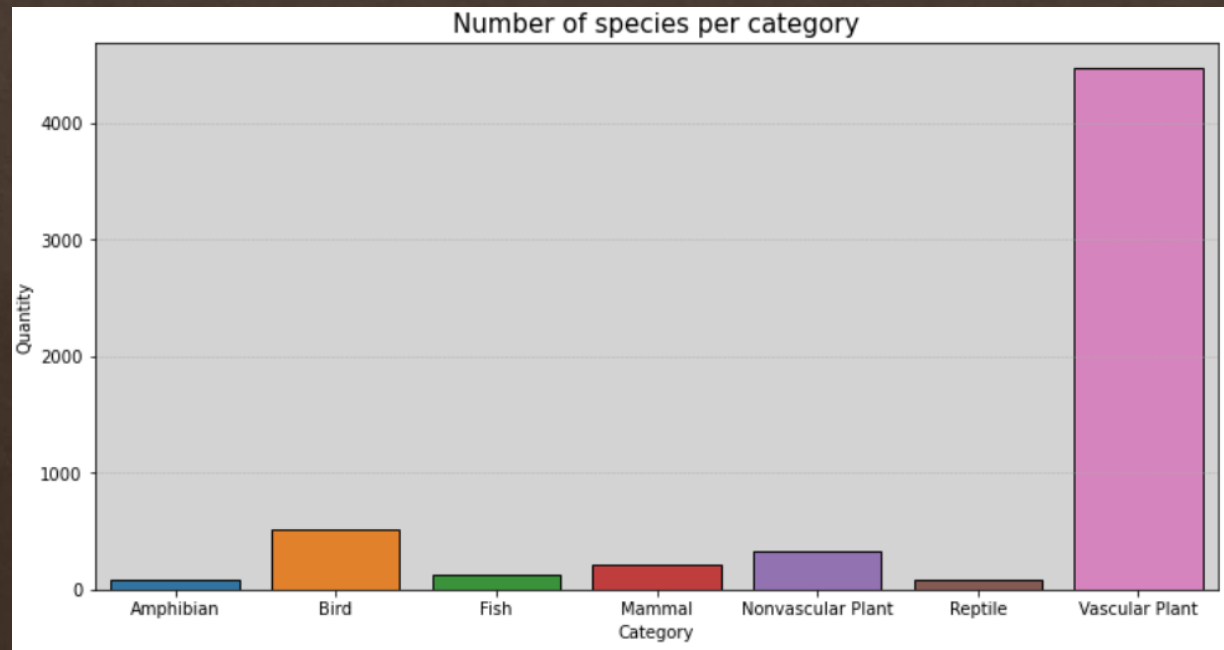
Codecademy

Data Science Career Path

2021

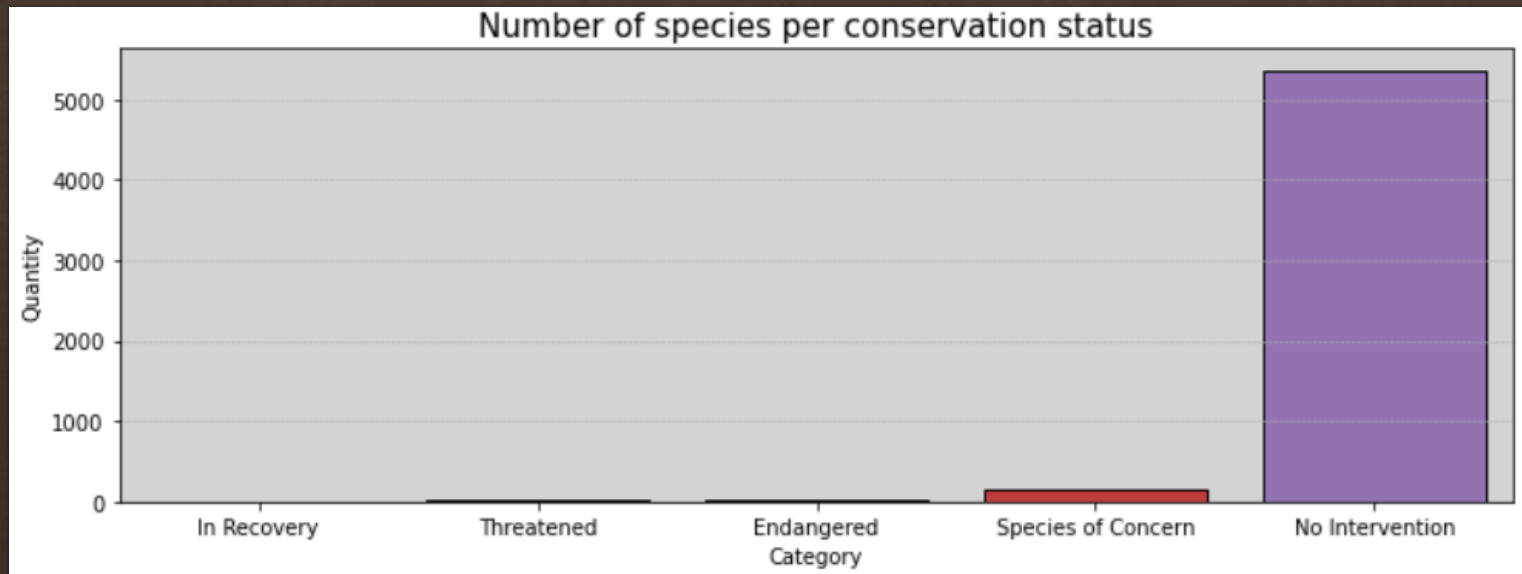
# Dataset N° 1: Species

- The dataset presents 5824 instances in 7 categories of different species, where the vascular plant is the category with the largest number of records.
- Contains 5633 null records in the attribute “conservation\_status”.



# CONSERVATION STATUS OF THE SPECIES

The number of animals without intervention in the national park is much greater than the other conservation states of the species.



The conservation status is:

- In Recovery: 4 species.
- Threatened: 10 species.
- Endangered: 15 species.
- Species of Concern: 151 species.

# HYPOTHESIS TEST RESULT

I proceeded to analyze what percentage of each species was within the "unprotected" category.

After that, I have proceeded to compare whether Species in category Mammal are more likely to be endangered than species in Bird

## CONCLUSION:

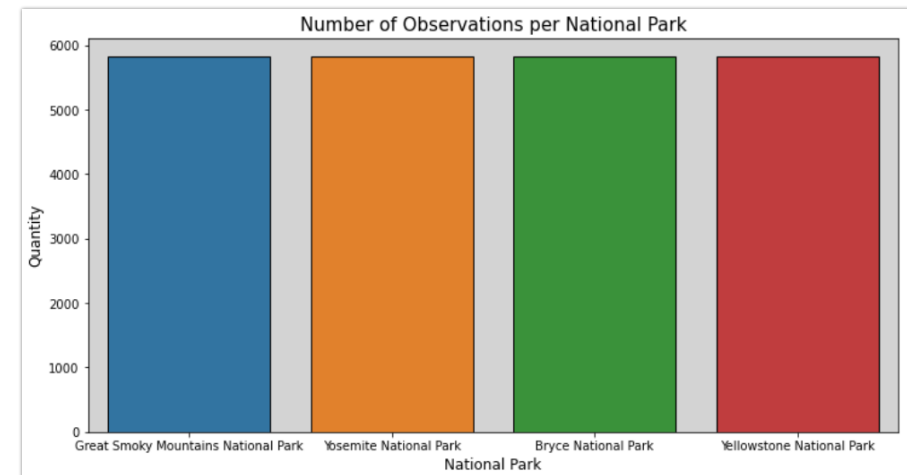
By performing a Chi-square test with a significance level of 0.05, it could be concluded that mammals have a higher probability to be endangered than the Birds



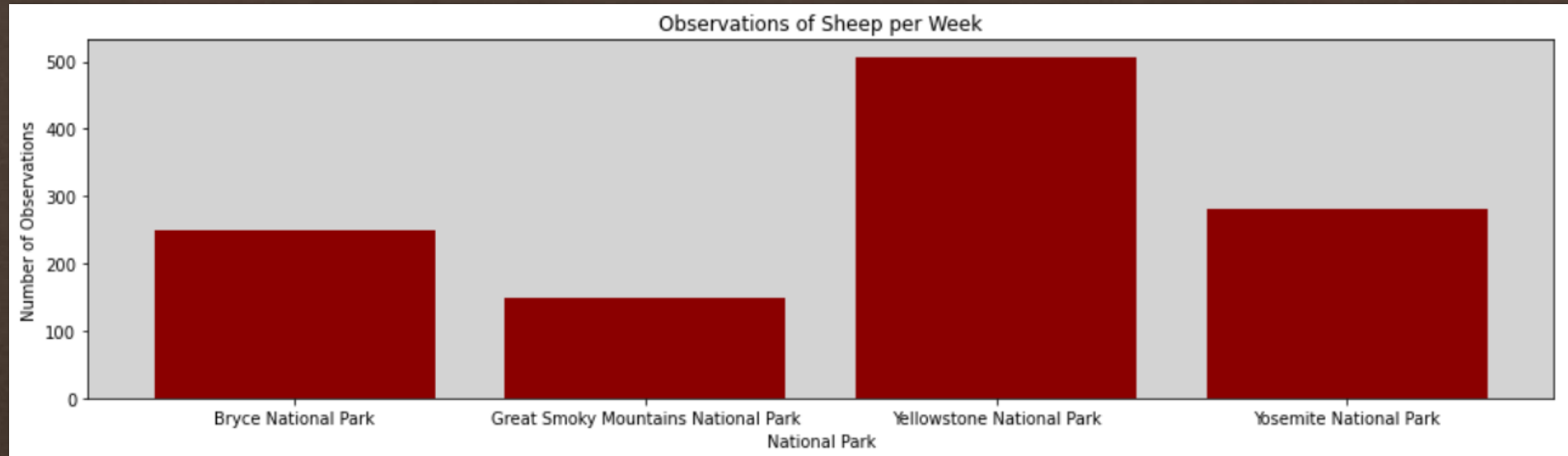
# Dataset N° 2: observations

The second data set:

- ◆ Presents 23.296 records of observations evenly distributed in 4 national parks.
- ◆ It does not have missing values.
- ◆ “*Streptopelia decaocto*” was the species with the highest number of sightings with 1785 views



¿HOW MANY TIMES THE DIFFERENT SPECIES OF SHEEP  
HAVE BEEN SIGHTED IN EACH NATIONAL PARK?



We can see that in Yellowstone National Park is possible to observe more sheeps than the others national parks.