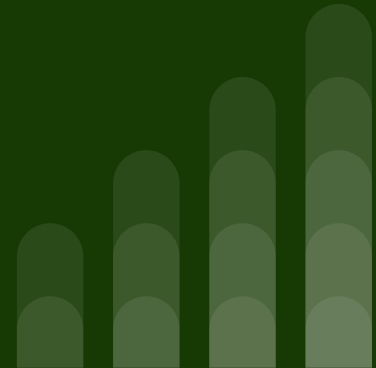# Biodiversity in National Parks

Data from the *National Parks Service* about endangered species in different parks

Datasets

species_info.csv

```
category
Amphibian            80
Bird                521
Fish                127
Mammal              214
Nonvascular Plant   333
Reptile              79
Vascular Plant     4470
dtype: int64
```

- **category:** represents the species' family. Can be:
  - Mammal
  - Bird
  - Reptile
  - Amphibian
  - Fish
  - Vascular plant
  - Nonvascular Plant
- **scientific_name:** species' name in latin

- **common_names:** common names in english language
- **conservation_status:** whether the species is at risk. Possible values:
  - Not at Risk
  - Species of Concern
  - Threatened
  - Endangered
  - In Recovery

Endangered	16
In Recovery	4
Not at risk	5633
Species of Concern	161
Threatened	10

observations.csv

- **scientific_name:** species' name in latin
- **park_name:** Name of the National Park where the species was found
- **observations:** Number of specimens observed in said National Park

park_name
Bryce National Park
Great Smoky Mountains National Park
Yellowstone National Park
Yosemite National Park

Species at risk

From a total of 5,824 species observed across four national parks...

191 are currently facing threats to their survival—falling into categories like Endangered, Threatened, In Recovery, or Species of Concern.
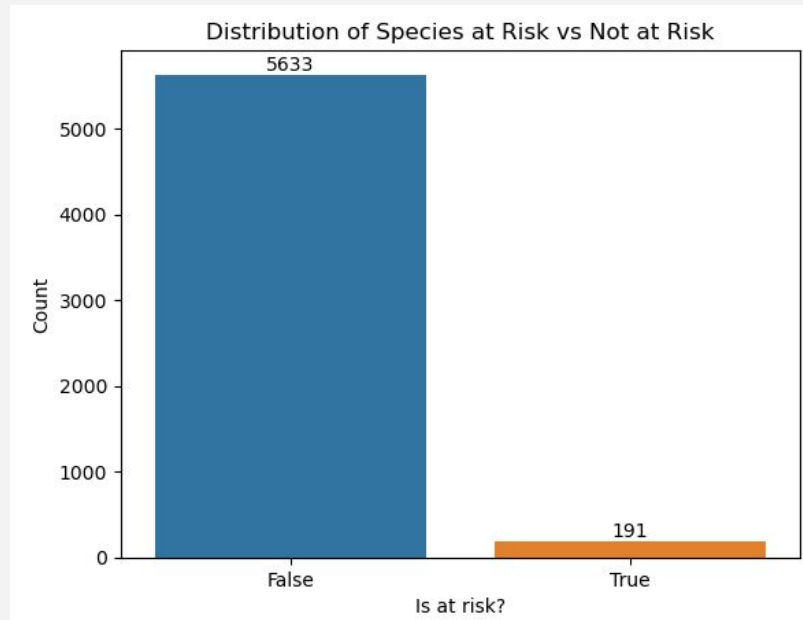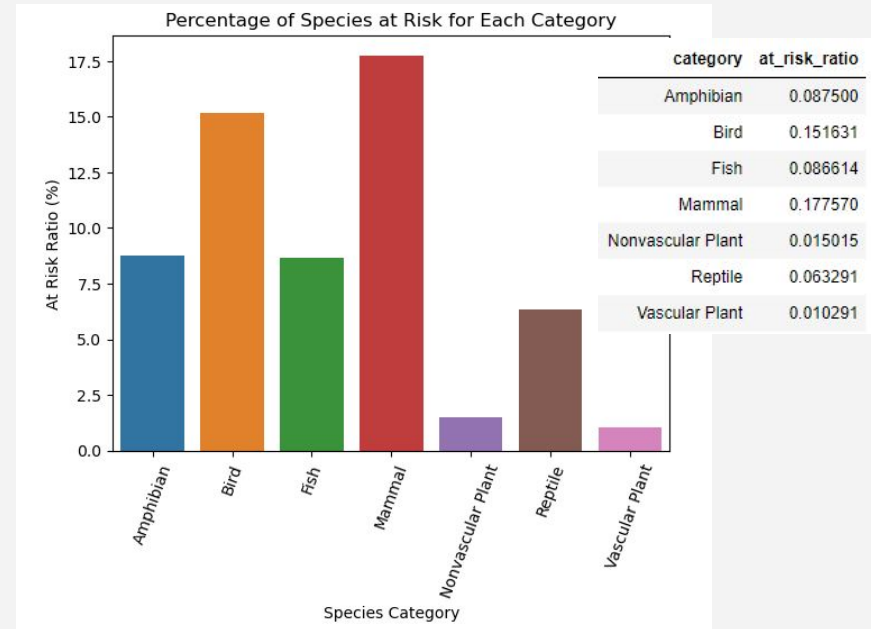
Risk ratio for individual categories

Comparing the Risk vs Not at Risk species on each category we find that...

Mammals (18%) and **Birds** (15%) are the categories facing the biggest threat.

The **plant categories face the least threat** — 1.5% and 1% for Nonvascular and Vascular respectively.

Percentage of Species at Risk for Each Category

| category	at_risk_ratio
Amphibian	0.087500
Bird	0.151631
Fish	0.086614
Mammal	0.177570
Nonvascular Plant	0.015015
Reptile	0.063291
Vascular Plant	0.010291

No. of species at risk per park

This being an **educational dataset**, it was cherry-picked and might have fake information for easier processing.

That caused the number of species to be exactly 179 for our 4 national parks.

In a real dataset, this would be almost impossible and we could choose to **allocate more resources to the parks with more species at risk**.
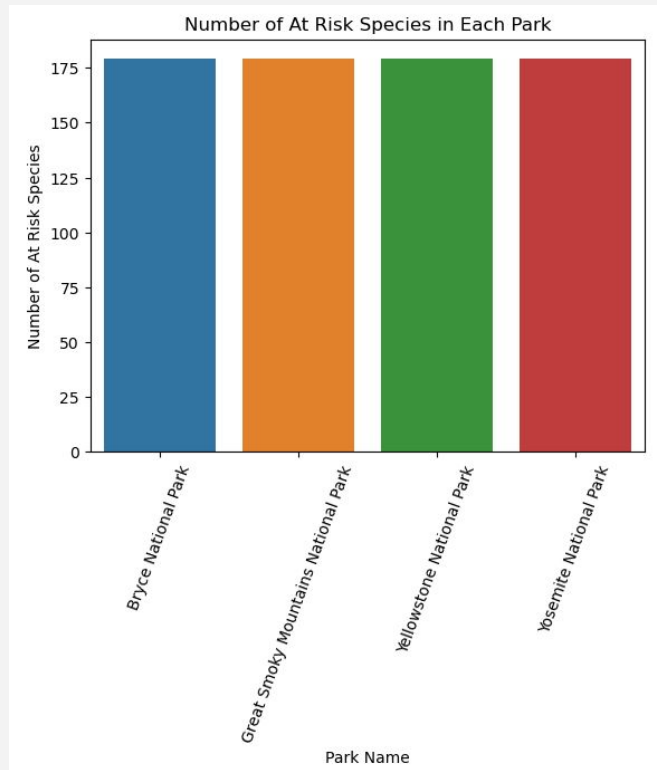
Recommendations

Based on our results, the most immediate course of action is to **focus on mammals and birds**.

Conservationists need to find out **what's causing these two categories' populations to drastically decrease**.

We need to get more info and **find out if there are specific species pushing the categories' numbers that high** or if it's a general thing.