Biodiversity on America's National Parks

INTRODUCTION

Biodiversity conservation is a critical global concern, with approximately 1 million species at risk of extinction due to human activities. This project focuses on analyzing and investigating data related to the conservation statuses of species living in national parks of The United States.

Objective

The primary goal is to gain insights into the conservation statuses of species by examining their distribution across different categories and national parks.

Here are a few questions that this project has sought to answer:

- What is the distribution of conservation status for species?
- Are certain types of species more likely to be endangered?
- Are the differences between species and their conservation status significant?
- Which animal is most prevalent and what is their distribution amongst parks?

Methodology

1. **Data Import and Cleaning**:
    - Import datasets and assess their quality.
    - Address missing values (NaN) and duplicated records for accuracy and reliability.
    - The cleaned data will serve as a foundation for analysis.
2. **Investigation**:
    - Explore conservation statuses based on:
        - **Species Categories**: Analyze which groups (e.g., mammals, birds) are most impacted.
        - **National Parks**: Identify trends and patterns across parks.
3. **Visualization and Insights**:
    - Use charts and plots to communicate findings effectively.

Data Sources

The datasets used in this project were provided by Codecademy.com. They include:

- **`observations.csv`**: Records of species observations in various national parks.
- **`species_info.csv`**: Details on species, including their conservation statuses and categories.

Conservation Statuses and Categories

Conservation statuses provide critical information about the health of species populations. The main categories analyzed in this project are:

- **Least Concern**: Species with a stable population and low risk of extinction.
- **Species of Concern**: Species requiring monitoring but not at immediate risk.
- **Endangered**: Species facing a high risk of extinction in the near future.
- **Threatened**: Species likely to become endangered.
- **In Recovery**: Species whose populations are improving due to conservation efforts.

Analysis Goals

This project aims to:

1. Understand the distribution of species across these categories.
2. Identify which categories are most prevalent in national parks.
3. Visualize patterns across parks and species groups.

Visualization

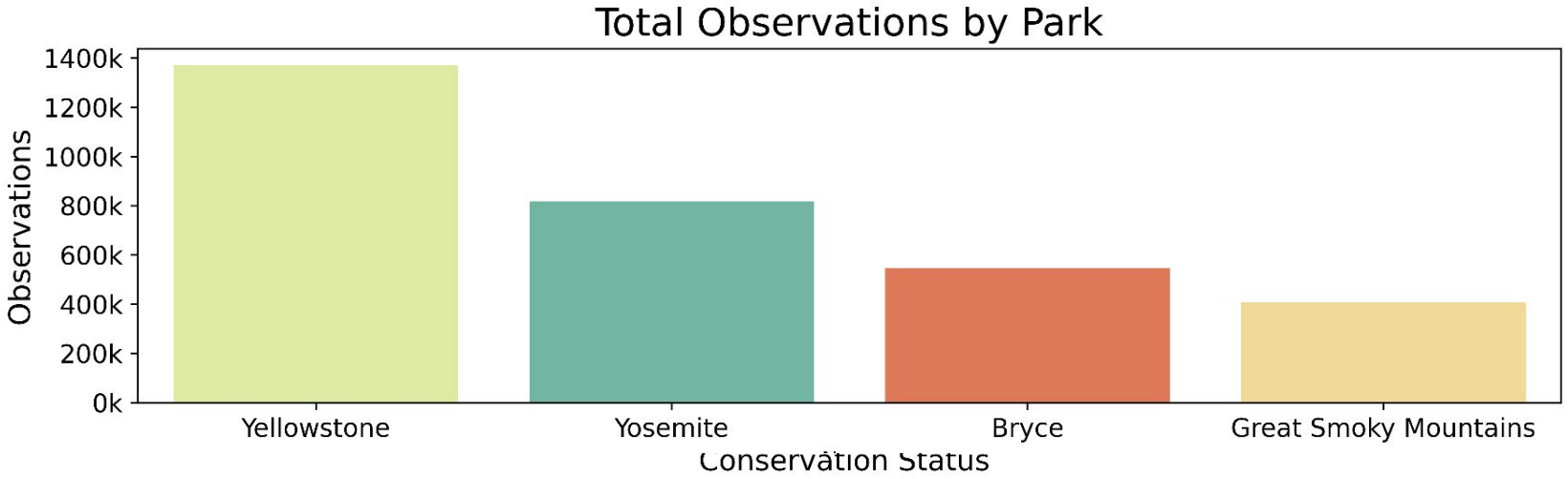
Visualizations are used to effectively communicate insights from the data. The following plots are included in the analysis:

- **Bar Charts**:
  - Illustrate the count of species in each conservation status category.
  - Compare the number of species across different categories within each conservation status.
- **Box Plots**:
  - Explore the distribution of species observations across parks for each status.
- **Histograms**:
  - Show the frequency distribution of species counts within categories to identify trends.

These visualizations provide an in-depth view of the species' conservation statuses and their distributions across parks.

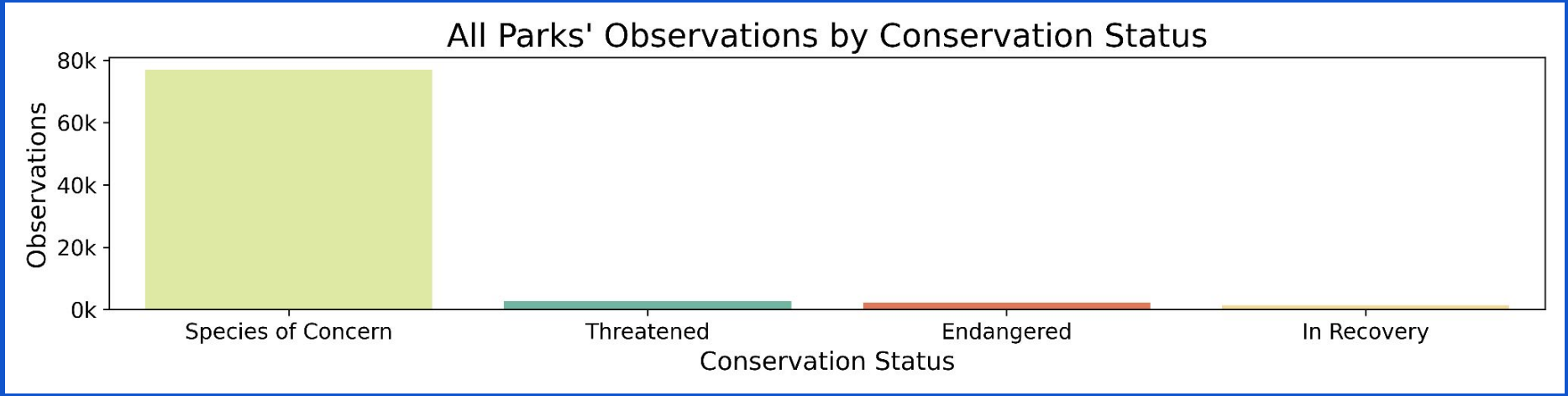
Total Observations by Park

Total count of species observations by park.



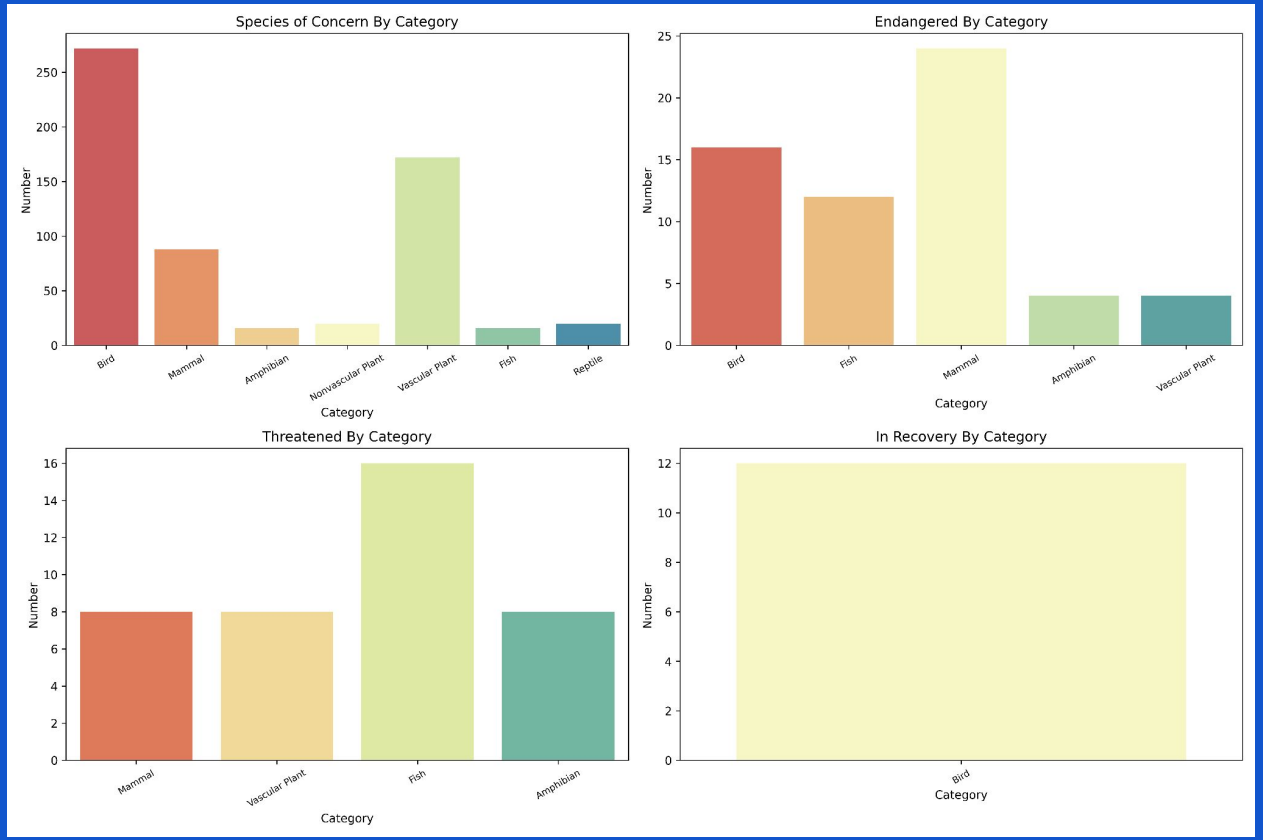
All Parks Observations by Conservation Status

Number of species observed by conservation status.



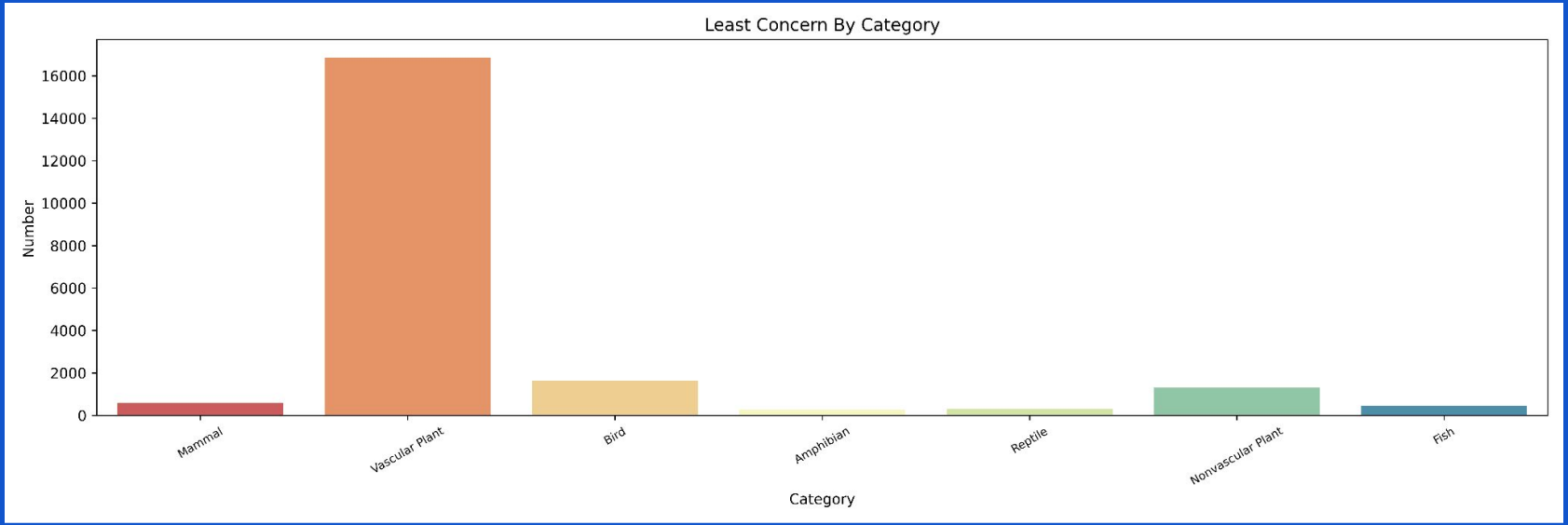
Species Categories By Conservation Status

Species with some degree of risk of extinction, by category and conservation status

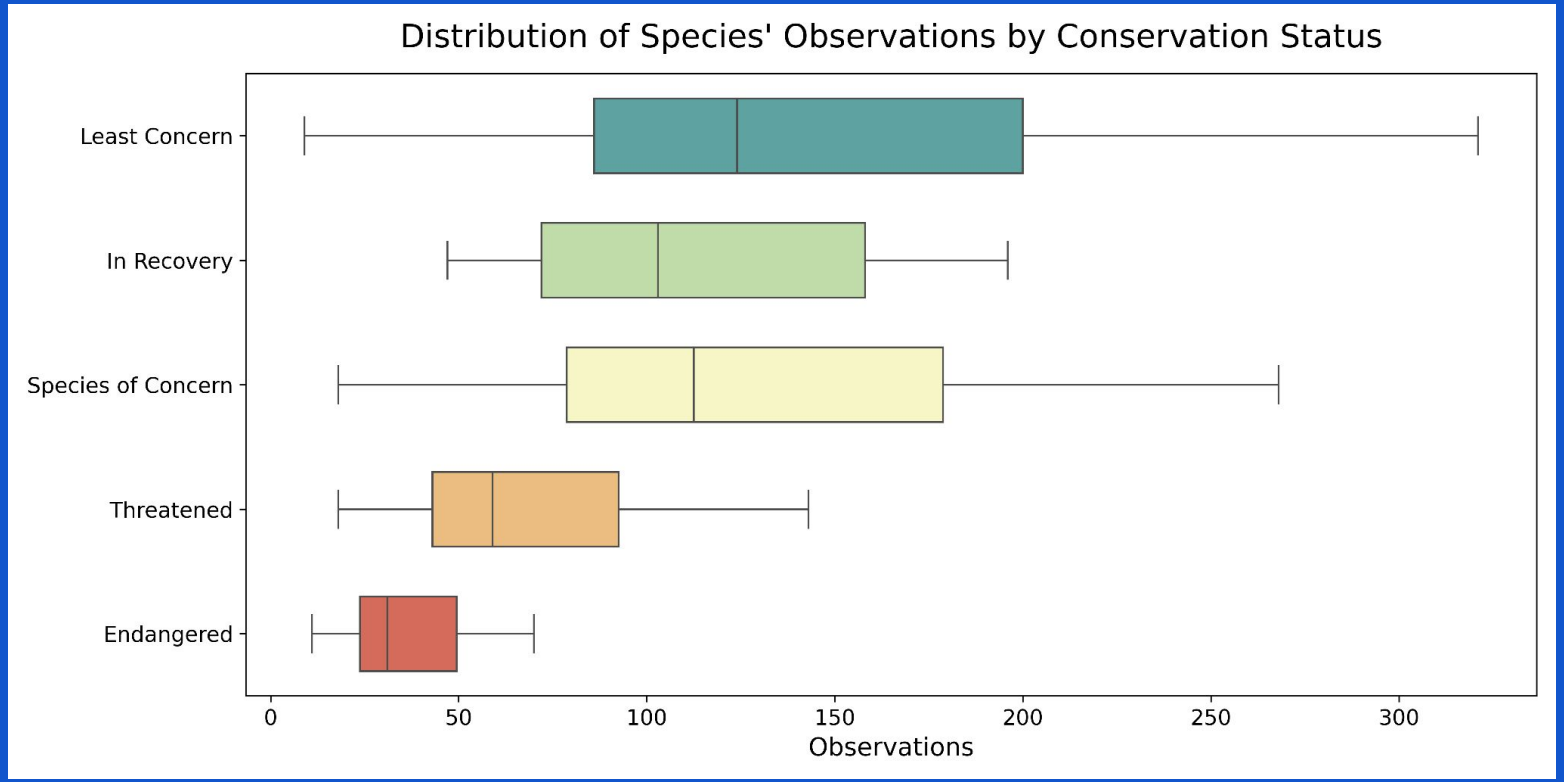


Species of Least Concern by Category

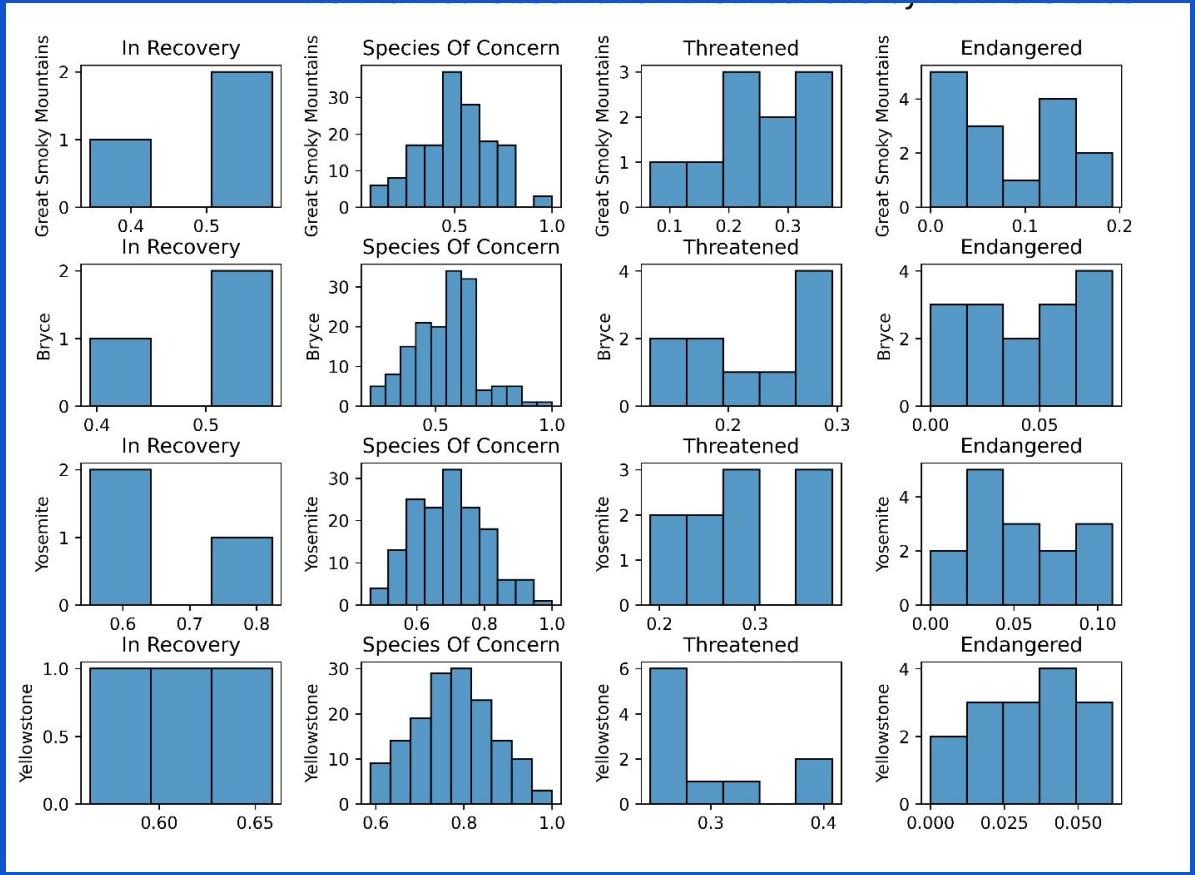
Species with a stable population and low risk of extinction.



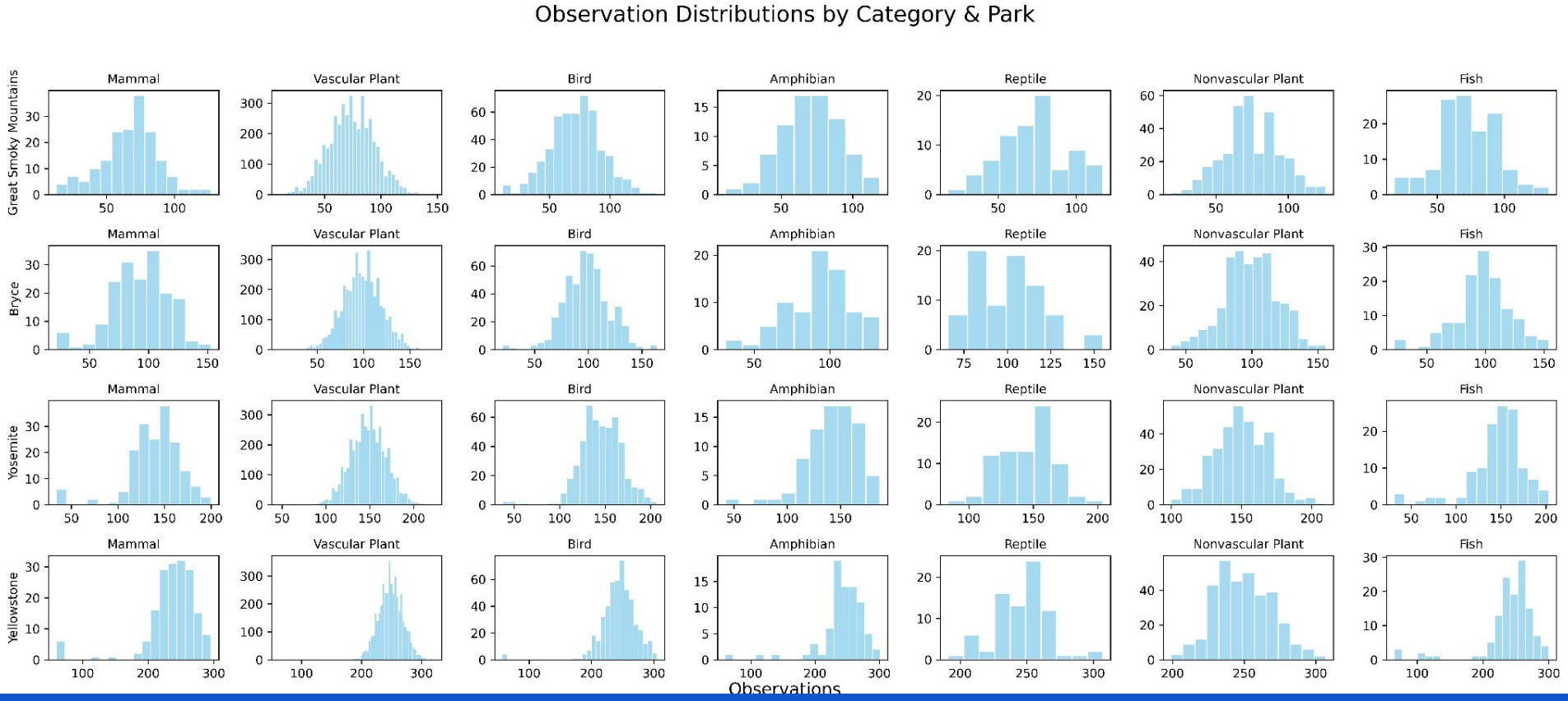
Distribution of Species Observations by Conservation Status



Observation Distribution by Category & Park



Observation Distribution by Category & Park



Conclusion

Key Learnings

This project provided valuable insights into biodiversity conservation and the use of data analysis for identifying patterns. Key skills developed include:

- Data cleaning and handling missing values and duplicates.
- Visualizing data with Python libraries like `pandas`, `seaborn`, and `matplotlib`.

Key Findings

Conservation Statuses

Most species are categorized as "Least Concern," but critical statuses like "Endangered" and "Threatened" highlight the need for focused conservation efforts.

Park Insights

Yellowstone and Great Smoky Mountains have higher observations of endangered species, making them priority areas for conservation.

Species Categories

Mammals and vascular plants are more frequently categorized under critical conservation statuses, indicating higher vulnerability.