**📌 Scope of the Project – Biodiversity in National Parks**

**Objective:**

The goal of this project is to explore and analyze biodiversity data from various U.S. national parks to understand patterns of species observation in relation to their biological category and conservation status. Through this analysis, we aim to identify which parks have the highest presence of endangered species, explore observation trends across categories, and test assumptions about the visibility of endangered species versus those with unknown conservation status.

**Data Sources:**

1. **species\_info.csv** – Contains taxonomic and conservation information of species:
   * category: Type of organism (e.g., Mammal, Bird, Vascular Plant).
   * scientific\_name: Scientific name of the species.
   * common\_names: Common names.
   * conservation\_status: Status such as Endangered, Threatened, or NaN (no known conservation concern).
2. **observations.csv** – Contains species sightings across parks:
   * scientific\_name: Matches species from the species\_info file.
   * park\_name: National park where the species was observed.
   * observations: Number of observations reported.

**Key Questions to Explore:**

* How are species observations distributed by biological category?
* Which Park has the highest number of endangered species?
* Which Park reports the most observations per category?
* How does species observation vary between parks?
* How does the number of observations vary by conservation status?

**Statistical Hypothesis to Test:**

* **Null Hypothesis (H₀):** Species without a listed conservation status (NaN) have on average, the same or **higher** number of observations than those categorized as endangered or threatened.
* **Alternative Hypothesis (H₁):** Species with a listed conservation status have a **different (typically lower)** number of observations than those with NaN status.

**Out of Scope:**

* No predictive modeling or machine learning is performed.
* Geographic mapping or external ecological data sources are not included.
* The analysis does not evaluate observation bias or data collection methodology.

**Tools:**

* Python (Pandas, Seaborn, Matplotlib)
* Jupyter Notebook
* Git & GitHub (for version control and sharing)