SYNOPSIS

Problem Statement:

Biodiversity is declining at an alarming rate due to habitat destruction, climate change, and human activities. Many species are either endangered, threatened, or at risk of extinction, but the lack of proper data analysis and visualization makes it difficult to track and take action. National parks play a crucial role in conservation, yet we often fail to analyze trends in species populations and understand which species need immediate protection.

Proposed Solution:

To address the challenges of biodiversity loss and species conservation (WildStat), this project proposes an analytical approach using Python and data visualization techniques to study species distribution in U.S. National Parks. By processing and analyzing biodiversity datasets, the project will uncover trends in species populations, conservation status, and park-wise diversity. Using libraries like Pandas, Matplotlib, and Seaborn, the data will be transformed into meaningful visualizations such as bar charts, pie charts, and heatmaps, helping researchers and conservationists identify species at risk. This solution not only provides data-driven insights for better conservation planning but also lays the groundwork for future AI/ML-based predictions of species population trends and extinction risks.

Objectives:

- Analyze species distribution across U.S. National Parks.
- Identify conservation statuses of different species (Endangered, Threatened, Safe, etc).
- Visualize biodiversity trends using Python, Pandas, Matplotlib, and Seaborn.
- Provide insights that can assist in better conservation planning.

Literature Review:

Biodiversity conservation has been widely studied, with research emphasizing the impact of climate change, habitat destruction, and human activities on species extinction. Studies from organizations like the IUCN and GBIF highlight the need for data-driven approaches to track species populations and their conservation status. Traditional methods rely on field surveys and ecological modeling, but recent advancements in data science have enabled better analysis through visualization and predictive analytics.

Existing biodiversity projects use GIS mapping, machine learning, and statistical analysis to study species distribution. However, there is a gap in easily accessible, interactive data visualization that helps conservationists make real-time decisions. This project builds on past research by using Python, Pandas, and visualization libraries to analyze biodiversity data from U.S. National Parks, providing insights that can support conservation efforts and future AI-based predictions.

Methodology:

The Prototype Methodology is being used in this project, which enables ongoing improvement through user feedback and iterative development. The process includes:

1. Data Collection: Import biodiversity datasets (species information + observations in national parks).
2. Data Cleaning & Processing: Handle missing values, filter important data.
3. Exploratory Data Analysis (EDA): Find patterns, conservation status distribution, and park-wise species count.
4. Visualization: Use bar charts, pie charts, and heatmaps to represent findings.
5. Insights & Conclusions: Identify key patterns and suggest conservation efforts.

Tools:

- Programming Language: Python
- Development Environment: Jupyter Notebook
- Data Processing & Analysis: Pandas, NumPy
- Data Visualization: Matplotlib, Seaborn
- Database : SQLite / CSV-based datasets
- Dataset Source: National Park Service Biodiversity Dataset
- Flask for development

Research Gap:

There is a lack of data-driven, interactive approaches that provide real-time insights into species distribution and conservation status. Existing projects primarily rely on static datasets and general visualizations, which may not fully capture the dynamic nature of biodiversity changes across different parks. Furthermore, there is limited use of machine learning to predict trends or provide actionable recommendations based on historical data.

This project aims to fill these gaps by combining real-time data analysis with interactive visualizations, and by exploring the potential of AI/ML for predicting species population trends in U.S. National Parks. The goal is to create a more dynamic, user-friendly tool that can assist conservationists with data-driven decision-making.

References:

- Global Biodiversity Information Facility (GBIF): Official Website. https://www.gbif.org
- IUCN Red List: Official Website. https://www.iucnredlist.org
- Matplotlib: Official Documentation. https://matplotlib.org/
- Seaborn: Official Documentation. https://seaborn.pydata.org/
- Enhancing biodiversity conservation and monitoring in protected areas through efficient data management"

- Authors: Ferdinando Urbano, Ramona Viterbi, Luca Pedrotti, Enrico Vettorazzo, Cristina Movalli, Luca Corlatt Published in: Environmental Monitoring and Assessment, December 2023
  https://link.springer.com/article/10.1007/s10661-023-11851-0
- A guide to analyzing biodiversity experiments" Authors: Cajo J. F. Ter Braak, Jarrett E. K. Byrne Published in: Journal of Plant
  https://academic.oup.com/jpe/article/10/1/91/2966834