

Project Proposal: Project 1 Group #11

Kerry Oostdyk, Shahla Shahnawaz, William Berry, Revano Harahap

- **Our Dataset:**
 - [Biodiversity in National Park](#)(Prof Rec)
 - [United States National Parks](#) from [Kaggle.com](#)
 - The data set provides an overview of the United States National Park.
 - We selected this because our group has expertise in this area, and it is a topic that personally interests us.
- **Guiding Questions:**
 - #1 What are the oldest parks in the country?
 - How does park age relate to visitor count? (Limitation: we can only answer that in relation to 2021)
 - #2 How does the size of the park affect its number of visitors?
 - #3 How are national parks distributed across the U.S.?
 - Which states have the highest density of national parks?
 - Are visitor counts and park location related? (Heatmap)
 - Are the more visited parks closer to other parks?
- **Inspiration** - other code/articles similar to your dataset
 - <https://www.kaggle.com/code/zveroz/top-9-national-parks>
 - 9 parks were picked based on top visitor counts. Ideal route maps were created based on geographical coordinates.
- **Example Visualization:**
 - Visualization for Guiding Question 1:
 - Data Viz: Line graph of total number of parks over time
 - Follow up chart: line chart of total visitors over time
 - Data Viz: Scatter plot of date establishment
 - Visualization for Guiding Question 2:
 - Data Viz: Scatter plot of size vs visitor count
 - Visualization for Guiding Question 3:
 - Data Viz: Map of the country with dots bigger for average visitor count per day
 - Data Viz: Average visitor counts vs average distance from other parks
- **Regression:**
 - How does the size of the park affect its number of visitors?
 - Consider number of parks in that state.
 - Consider park exclusivity.
- **Color Palette**
 - Color palette taken from this photo: [Yellowstone National Park](#)
 - <https://coolers.co/096ec8-3f4c27-e1cbb9-39534e-427937-e48904>
- **Roles and Responsibilities**

- Everyone does data cleaning, research questions, and leads on slides for their driving question.
 - Questions 1: Revano
 - Question 2: William
 - Question 3: Shahla
 - Regression: Kerry
- **Further Questions**
 - How does the visitor count change over time?
 - How does this data change post covid?