Name: Rashmi Rajeshirke
UID: 2021700051
Class: CSE-DS BE
Batch: A

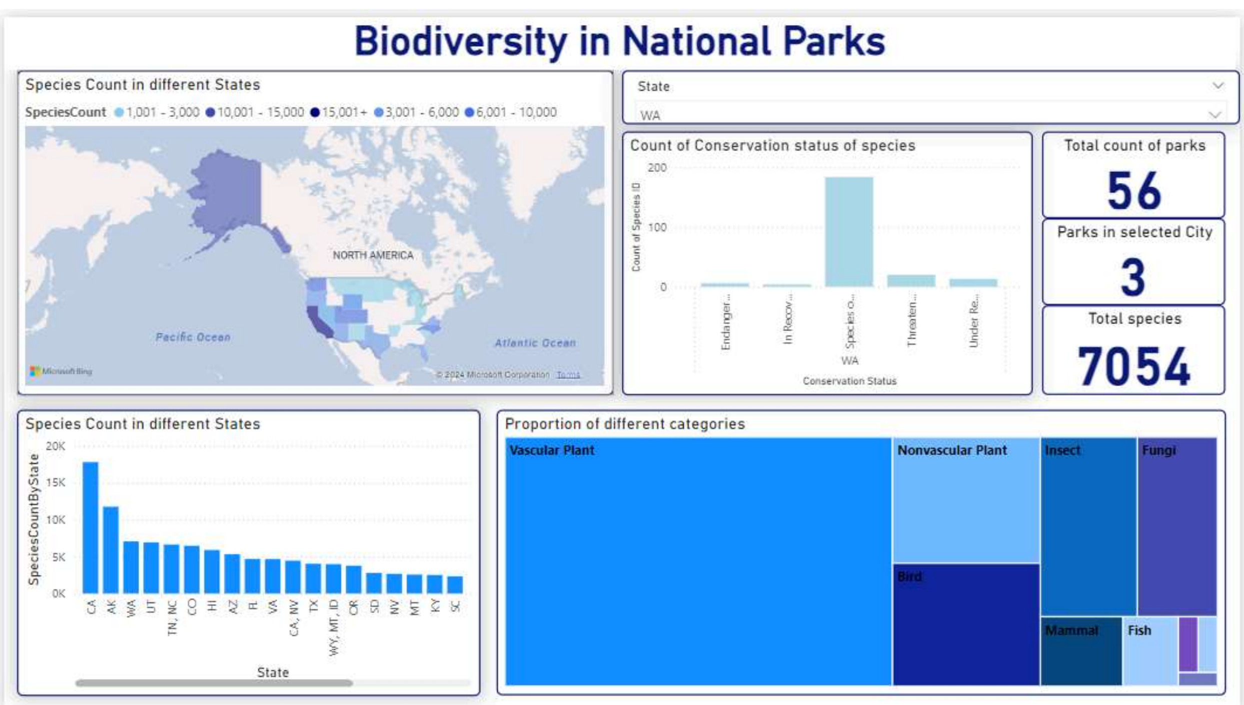
Lab6

Aim: To design interactive dashboards using Power BI for visualizing and analyzing an Animal/Wildlife/Marine dataset, employing both basic and advanced charts to uncover insights and trends.

Objectives:

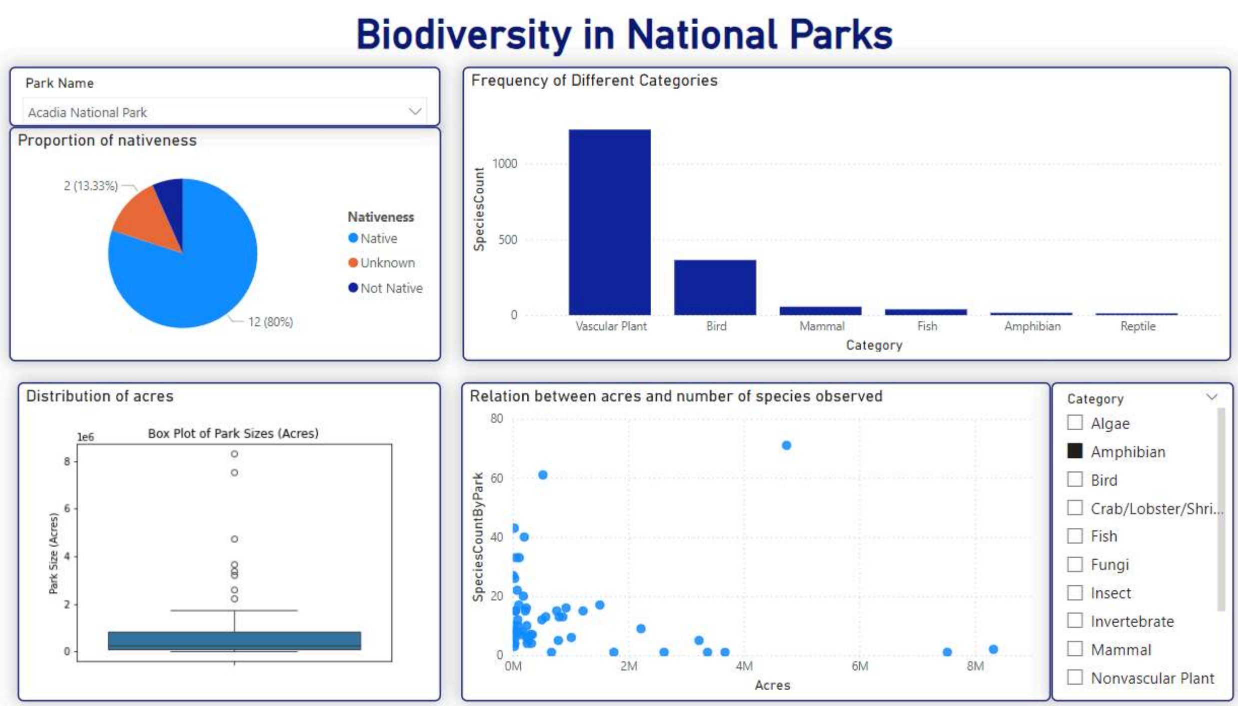
1. To create visually appealing and interactive dashboards that provide insights into the dataset.
2. To explore the distribution, trends, and relationships within the dataset using various types of visualizations.
3. To enable data-driven storytelling by highlighting key patterns, anomalies, and correlations.

Power Bi Dashboard:



1) The Species Count by states (Map) visualizes species distribution across various parks, highlighting Alaska (AK), California (CA), and Utah (UT) as regions with the highest biodiversity. The darker shades indicate areas of concentration, primarily in western U.S. states, giving a clear insight into the regions where biodiversity is particularly rich.

2) The Count of Conservation of species (Bar chart) reveals that in Washington (WA), the majority of species fall under "Species of Concern," with smaller numbers in the "Endangered", “under review”, “in recovery” and "Threatened" categories. This highlights the focus areas for conservation efforts in the state, showing that many species are vulnerable and require monitoring to ensure ecological balance.
3) In the Species Count in different StatesState (Bar Chart), states like California (CA), Alaska (AK), and Utah (UT) lead with the highest species count, while states like South Carolina (SC) and Kentucky (KY) have significantly fewer species. This chart reveals patterns of biodiversity across the nation and flags potential outliers, suggesting that states with lower species counts may need more research or conservation focus.
4) The Proportion of different categories chart (Tree Map) demonstrates that vascular plants are the most dominant species type, followed by non-vascular plants, insects, fungi, and birds. This distribution suggests that plant species are the backbone of these ecosystems, while certain groups like mammals and fish are less represented, likely due to habitat limitations or specialized ecological niches.
5) The Total Parks, Selected Parks, and Total Species section provides an overview, indicating there are 56 parks with a total of 7,054 species across the dataset, and 3 parks in the selected region of Washington



6) In the Proportion of Nativeness by park(Pie Chart) for Acadia National Park, 80% of species are native, while a smaller portion is non-native or unknown. This indicates a predominantly healthy ecosystem with a few potential risks posed by non-native species, emphasizing the importance of monitoring ecological balance in national parks.
7) The Species Count by Category (Bar Chart) in Acadia National Park illustrates the species count in Acadia National Park categorized by type, showing that vascular plants

are overwhelmingly dominant, followed by birds and mammals. The chart suggests a diverse ecosystem where plant life dominates, while animals like amphibians and reptiles are less represented, indicating a potential area for increased conservation focus.

8) This box plot representing Distribution of acres presents the distribution of national park sizes, with most parks falling between 0 and 2 million acres. However, there are several large parks that are notable outliers, with one park exceeding 8 million acres. This suggests that while most parks are of moderate size, a few very large parks contribute significantly to the overall land area. The presence of these outliers may indicate that larger parks could potentially house a wider variety of species, though this visualization does not directly link park size to biodiversity.
9) This scatter plot examines the relationship between park size (acres) and species count, showing that smaller parks tend to have species counts ranging from 10 to 60, but there is no clear trend that links larger parks to higher species counts. Some smaller parks have higher species counts, which could be an indication of habitat richness or better management practices. Conversely, larger parks do not always have a proportionally larger number of species, suggesting that park size alone is not a strong indicator of biodiversity.

Questions answered:

1. Population distributions across regions:
   The Species Count by States (Map) and Bar Chart show that Alaska, California, and Utah have the highest species diversity, while states like South Carolina and Kentucky have fewer species. This highlights regional differences in biodiversity, with western U.S. states having richer ecosystems.
2. Correlations between environmental factors and species population
   The Scatter Plot (Species Count by Park by Acres) suggests park size isn't the only factor affecting species richness, pointing to other environmental factors. The Conservation Count (Bar Chart) in Washington shows many "Species of Concern," indicating a correlation between environmental stressors and species vulnerability.
3. Trends in animal sightings and marine life
   The Category Bar Chart and Tree Map show vascular plants dominating species counts, while sightings of mammals, fish, and marine life are lower. This suggests trends in animal sightings are influenced by habitat availability and park ecosystems.
4. Significant outliers or anomalies
   The Box Plot of Park Sizes and Scatter Plot show outliers in park size and species counts, with some small parks having high biodiversity. The Map highlights states with unusually low species counts, suggesting research or conservation gaps.

Dax Commands used:

1. For world map showing Species count in different states:

   Two new columns were created using dax:

```
BinnedSpeciesCount =
    SWITCH(
        TRUE(),
        [SpeciesCountInParkState] <= 1000, "0 - 1,000",
        [SpeciesCountInParkState] <= 3000, "1,001 - 3,000",
        [SpeciesCountInParkState] <= 6000, "3,001 - 6,000",
        [SpeciesCountInParkState] <= 10000, "6,001 - 10,000",
        [SpeciesCountInParkState] <= 15000, "10,001 - 15,000",
        [SpeciesCountInParkState] > 15000, "15,001+"
    )
```

This command uses SWITCH to categorize species counts into bins. It groups parks based on the number of species in them, making the map more readable by color-coding ranges like 0-1,000 or 10,001-15,000 species.

```
   SpeciesCountInParkState =
CALCULATE(
    COUNT(species[Species ID]),
    FILTER(
        species,
        RELATED(parks[State]) = parks[State]
    )
)
```

This calculates the total species count for a park in a state. It uses CALCULATE and FILTER to count species records that are related to the specific state for each park.

2. Bar chart for species count in different states

```
SpeciesCountByState = COUNTROWS(
      RELATEDTABLE(species)
   )
```

The COUNTROWS(RELATEDTABLE(species)) counts how many species are related to each state in the dataset, providing the total number of species for each state.

3. Bar chart: Frequency of different categories

```
SpeciesCount = COUNT(species[Species ID])
```

This uses COUNT(species[Species ID]) to calculate the total number of species by counting the Species ID in the dataset, used for bar chart visualization of species categories.

4. Scatter plot : Relation between acres and number of species observed

```
SpeciesCountByPark =
    COUNTROWS(
        RELATEDTABLE(species)
    )
```

Similar to the earlier one, COUNTROWS(RELATEDTABLE(species)) counts the number of species related to each park. This is used for the scatter plot to show how the number of species relates to park size in acres.

Conclusion:
In this lab experiment, I was able to explore species distribution, biodiversity patterns, and the relationship between park size and species counts. By visualizing data through maps, scatter plots, and bar charts, I identified regional biodiversity hotspots, correlations between environmental factors and species vulnerability, and outliers that suggest areas needing further research. These insights emphasize the importance of targeted conservation efforts and the role of environmental conditions in shaping species populations across national parks.