BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
(Empowered Autonomous Institute Affiliated to University of Mumbai)
[Knowledge is Nectar]

Department of Computer Science and Engineering

Advanced Data Visualization

UID	2021700012
Name	Rahul Chalwadi
Class and Batch	BE CSE-DS
Experiment No.	06
Aim	

Link:Animal Dataset

Dataset Description:

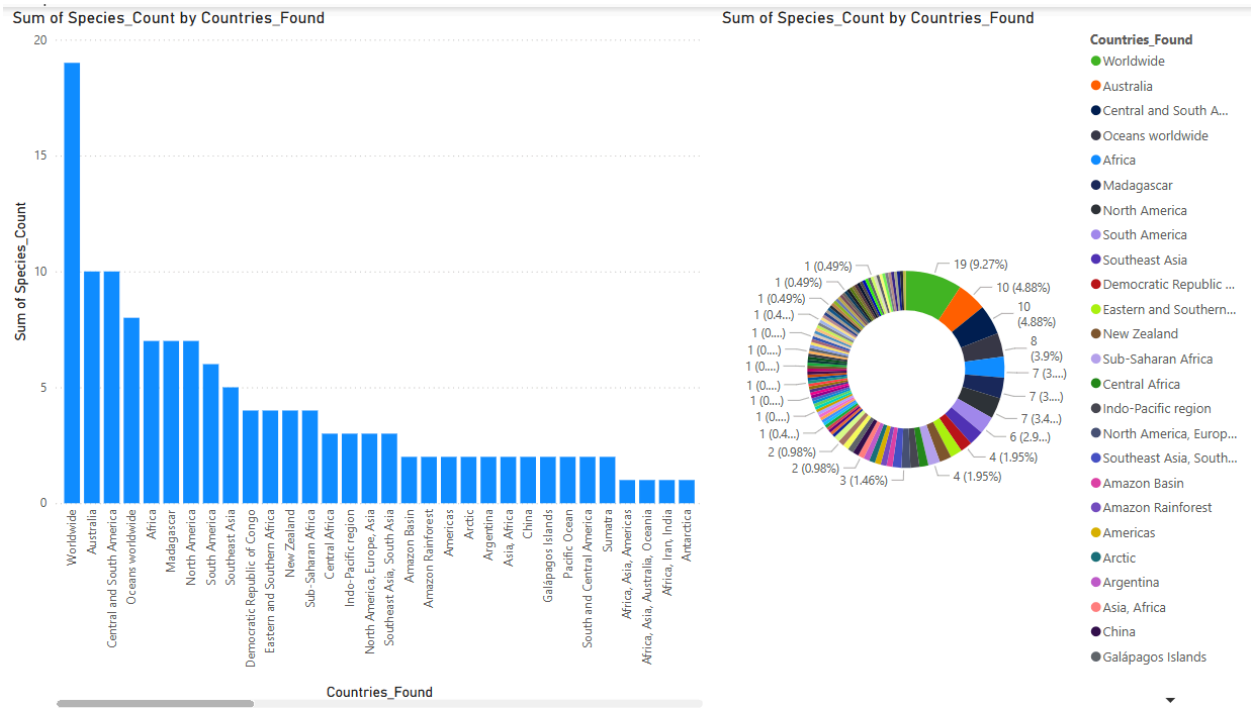
Animal - Name of the animal.
Height (cm) - Height range in centimeters for the animal.
Weight (kg) - Weight range in kilograms for the animal.
Color - Common colors associated with the animal's appearance.
Lifespan (years) - Average lifespan of the animal in years.
Diet - Type of diet the animal primarily follows (e.g., Carnivore, Herbivore).
Habitat - Typical habitat or environment where the animal is found.
Predators - Natural enemies or organisms that prey on the animal.
Average Speed (km/h) - The average speed range the animal can achieve in kilometers per hour.
Countries Found - Countries or regions where the animal is commonly found.
Conservation Status - The conservation status of the animal as per relevant conservation organizations.
Family - Taxonomic family the animal belongs to.
Gestation Period (days) - Range of days representing the gestation or pregnancy period of the animal.
Top Speed (km/h) - The maximum speed the animal can achieve in kilometers per hour.
Social Structure - Information about the social behavior or structure of the animal (e.g., Solitary, Group-based).
Offspring per Birth - The typical number of offspring born per birth or reproduction event for the animal.

Q1.What are the population distributions of various species across different regions?

```
EVALUATE
SUMMARIZE(
    Animal,
    Animal[Countries_Found],        -- Grouping by region/country
    "Species_Count", COUNT(Animal[Animal])  -- Counting the number of species in each region
)
```

Results | Result 1 of 1 | Copy

	final_animal_dataset[Co...	[Total Species]
1	Africa	10
2	Eastern and Southern Af...	4
3	Sub-Saharan Africa	4
4	European Alps	1
5	Amazon Rainforest	2
6	North America	10
7	Central and South Amer...	10
8	Middle East	2
9	Arctic regions	1
10	South America	7
11	Asia (Southeast)	1
12	North Atlantic	1
13	Worldwide	19
14	Australia	13
15	Mexico	1



Observation - Species distribution is highest in Australia and America nearly acounting same as the rest of world category

2.How has the population of specific species changed over time?

Query: This query calculates the average weight of species grouped by their diet type. Analyzing this can help identify correlations between dietary habits and body size.

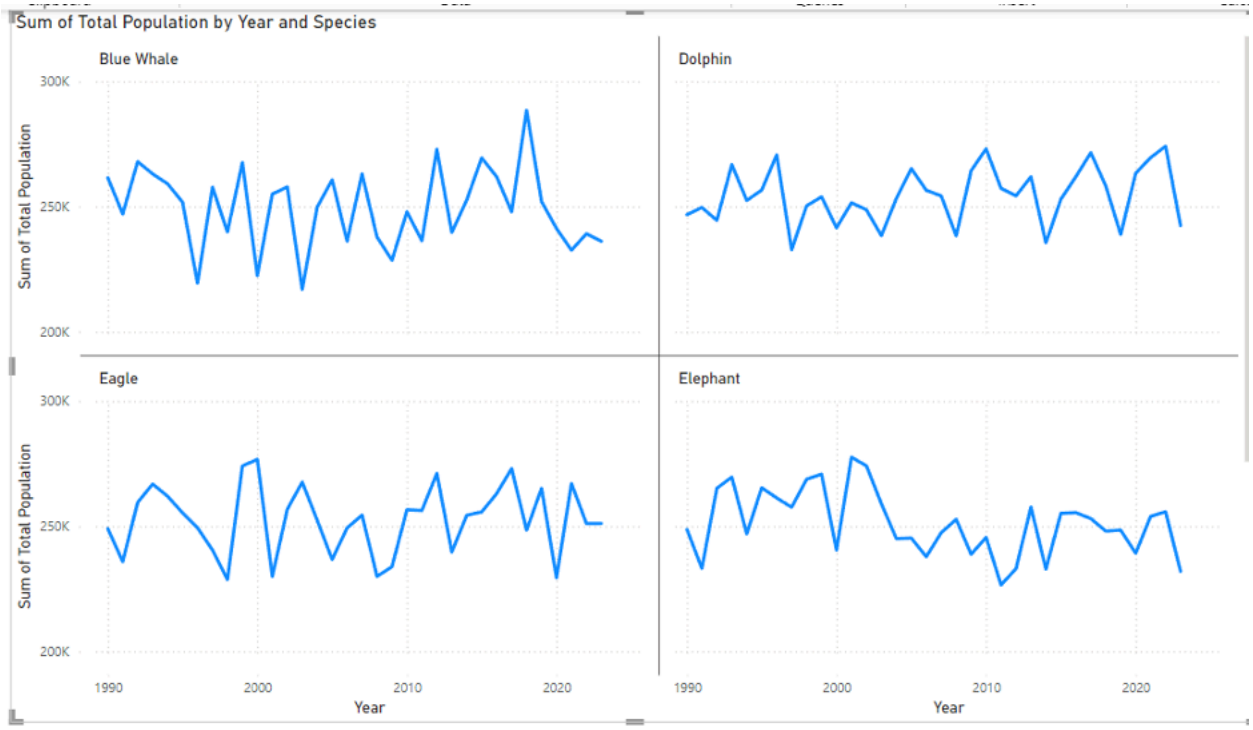
```
EVALUATE
SUMMARIZE(
    'expanded_wildlife_species_population',
    'expanded_wildlife_species_population'[Species],
    'expanded_wildlife_species_population'[Year],
    "Total Population", SUM('expanded_wildlife_species_population'[Population])
)
```

Output:

- All four species show significant population fluctuations over the 30-year period, with no clear consistent upward or downward trend across species.
- The Blue Whale population appears to have the highest peak among the four species, reaching close to 300,000 individuals at its maximum point in the late 2010s.

- The Elephant population seems to show a slight overall decline from 1990 to 2020, ending at a lower point than it started, despite several periods of increase throughout the timeframe.

	expanded_wildlife_speci...	expanded_wildlife_speci...	[Total Population]
1	Elephant	1990	248797
2	Tiger	1990	254472
3	Panda	1990	271331
4	Blue Whale	1990	261584
5	Dolphin	1990	246813
6	Penguin	1990	250353
7	Kangaroo	1990	256807
8	Shark	1990	255108
9	Wolf	1990	267752
10	Eagle	1990	249138
11	Elephant	1991	233269
12	Tiger	1991	245105



Observation - There's no specific trend on population for any particular species, however Blue whale population saw a sudden spike

3.Are there any correlations between environmental factors and species population?

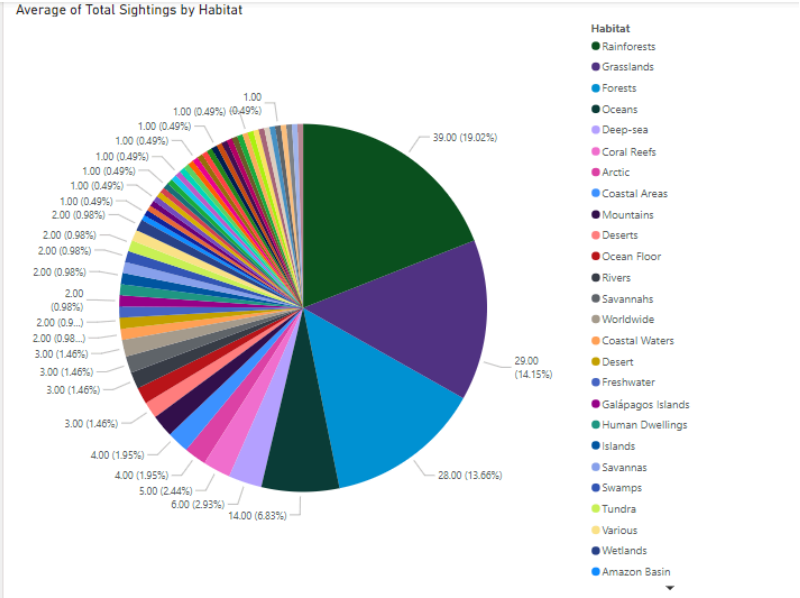
Query: This query summarizes the total sightings of various species by their habitat, helping to identify trends in sightings across different geographic areas.

```
EVALUATE
SUMMARIZE(
    final_animal_dataset,
    final_animal_dataset[Habitat],
    "Total Sightings", COUNT(final_animal_dataset[Animal])
)
```

Result:

Results | Result 1 of 1 | Copy

	final_animal_dataset[Ha...	[Total Sightings]
1	Savannas	2
2	Grasslands	29
3	Savannah	1
4	Savannahs	3
5	Mountains	4
6	Amazon Rainforest	1
7	Middle East	1
8	Desert	2
9	Tundra	2
10	Freshwater Rivers	1
11	North Atlantic	1
12	Oceans	14
13	Coastal Waters	2
14	Lakes	1
15	Rainforests	39
16	Forests	28
17	Deserts	3
18	Deep-sea	6
19	Coastal Areas	4
20	Coral Reefs	5



Observation -

- Sightings are most in Rainforests
- Amazon basins have sightings on lower end, which is not ideally presumable

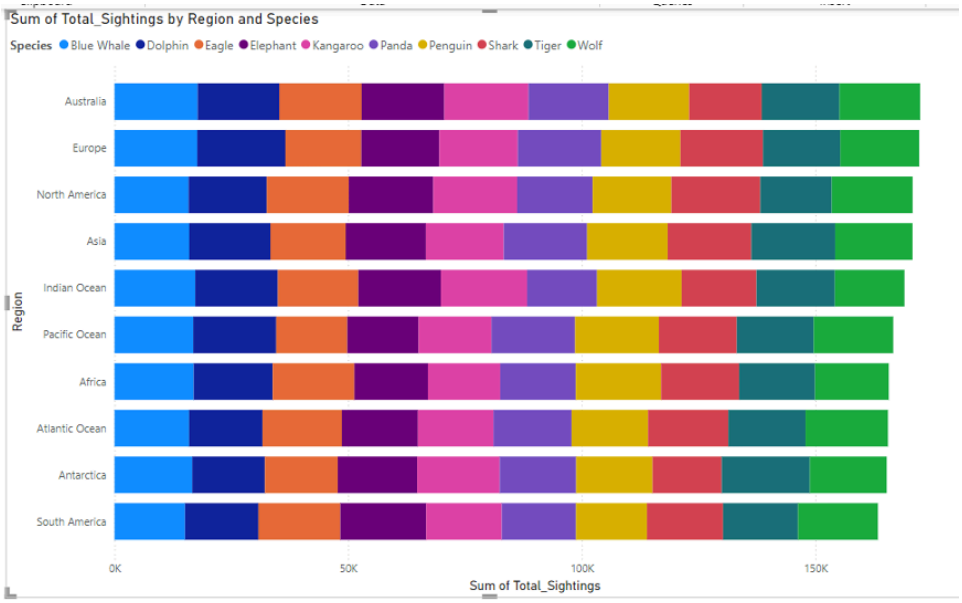
4. What are the trends in animal sightings and marine life in various geographic areas?

Query:

```
EVALUATE
SUMMARIZE(
    expanded_wildlife_species_population,
    expanded_wildlife_species_population[Region],
    expanded_wildlife_species_population[Species],
    expanded_wildlife_species_population[Year],
    "Total_Sightings",
SUM(expanded_wildlife_species_population[Sighting_Count])
)
```

Result:

	expanded_wildlife_speci...	expanded_wildlife_speci...	expanded_wildlife_speci...	[Total_Sightings]
1	Europe	Elephant	1990	541
2	Europe	Tiger	1990	532
3	Europe	Panda	1990	718
4	Europe	Blue Whale	1990	442
5	Europe	Dolphin	1990	349
6	Europe	Penguin	1990	542
7	Europe	Kangaroo	1990	681
8	Europe	Shark	1990	885
9	Europe	Wolf	1990	769
10	Europe	Eagle	1990	496
11	Australia	Elephant	1990	551
12	Australia	Tiger	1990	464



Observation -

- The distribution of species sightings appears relatively consistent across all regions, with each species represented to some degree in every region. This suggests a wide global distribution for most of these species.
- Australia seems to have the highest total number of sightings across all species, as indicated by its bar being the longest. This could be due to factors such as biodiversity, conservation efforts, or more intensive observation programs in the region.
- Aquatic species like Blue Whales, Dolphins, and Penguins have significant sightings across both terrestrial regions and oceanic regions.

5. Are there any significant outliers or anomalies in species population data?

Query:

```
EVALUATE
VAR TotalPopulationTable =
    SUMMARIZE(
        expanded_wildlife_species_population,
        expanded_wildlife_species_population[Species],
        "TotalPopulation",
SUM(expanded_wildlife_species_population[Population])
    )
VAR OverallMean =
    AVERAGEX(TotalPopulationTable, [TotalPopulation])
VAR OverallStdDev =
    STDEVX.P(TotalPopulationTable, [TotalPopulation])
VAR LowerBound = OverallMean - 2 * OverallStdDev
VAR UpperBound = OverallMean + 2 * OverallStdDev
RETURN
    ADDCOLUMNS(
        TotalPopulationTable,
        "Is_Outlier",
        IF(
            [TotalPopulation] < LowerBound ||
            [TotalPopulation] > UpperBound,
            "Yes",
            "No"
        )
    )
```

Result:

	expanded_wildlife_speci...	[TotalPopulation]	[Is_Outlier]
1	Elephant	8544881	No
2	Tiger	8527255	No
3	Panda	8571561	No
4	Blue Whale	8482214	No
5	Dolphin	8651687	No
6	Penguin	8635003	No
7	Kangaroo	8492623	No
8	Shark	8597554	No
9	Wolf	8626399	No
10	Eagle	8593553	No

Observation - No significant outliers