Advanced Data Visualization

Experiment - 6

Hardik Garg

2021300036

BE COMPS A - BATCH H

Aim:

Using DAX queries create Interactive Dashboard for Marinelife /wildlife dataset

Dataset:

https://www.kaggle.com/datasets/anoopjohny/shark-attack-dataset-california?select=Shark+Attacks+-+Sheet1.csv

The dataset "Shark Attacks - Sheet1.csv" includes records of shark attacks with details about the incidents, injuries, and shark species. Here's an overview:

- **IncidentNum**: Unique identifier for each incident.
- **Date**: Date of the shark attack.
- **County**: The county where the attack occurred.
- **Location**: Specific location of the incident.
- **Mode**: The activity the person was engaged in (e.g., Swimming, Surfing).
- **Injury**: Type of injury (e.g., major, minor, fatal).
- **InjuryT**: A more specific description of the injury outcome (e.g., Non-fatal Injury, Fatality).
- **Depth**: The water depth where the incident occurred (most are surface-level).
- **Species**: Type of shark involved (e.g., White Shark).
- **Comment**: Descriptive notes about the attack, often including details about the shark's behavior.
- **Longitude/Latitude**: Coordinates of the attack location.
- **Confirmed**: Source of information, indicating whether the attack has been confirmed.
- **Source**: The reference source for the data.

Key Insights:

1. **Common Shark Species**:
    - The White Shark is by far the most frequent species involved in these attacks (177 out of 202 incidents).
2. **Activity Mode**:
    - The majority of attacks occurred while victims were engaging in *Surfing/Boarding* (80 incidents).
    - Other activities include *Swimming* and *Freediving*.

3. **Injury Severity**:
    - Most recorded injuries were non-fatal, with 108 incidents categorized as *Non-fatal Injury*.
    - Fatalities are rare, with only a small number of incidents marked as *Fatal*.
4. **Location Concentration**:
    - A significant number of incidents (23) occurred in **San Diego County**, making it a hotspot for shark attacks.
    - The most common specific location was **Salmon Creek Beach**, with 9 incidents recorded there.
5. **Geographical Spread**:
    - The dataset includes longitude and latitude coordinates, with attack locations ranging between coordinates like:
        - Longitude: -124.72 to -117.14
        - Latitude: 32.58 to 41.56

This dataset provides a detailed look into shark attack incidents, including geographic distribution and activity patterns, helping identify trends related to shark species, injury severity, and attack hotspots.

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

```
EVALUATE
SUMMARIZE(
    'Shark Attacks',
    'Shark Attacks'[County],          -- Group by Region (County in this case)
    'Shark Attacks'[Species],         -- Group by Species
    "IncidentCount", COUNT('Shark Attacks'[IncidentNum])  -- Count number of incidents
(population proxy)
```

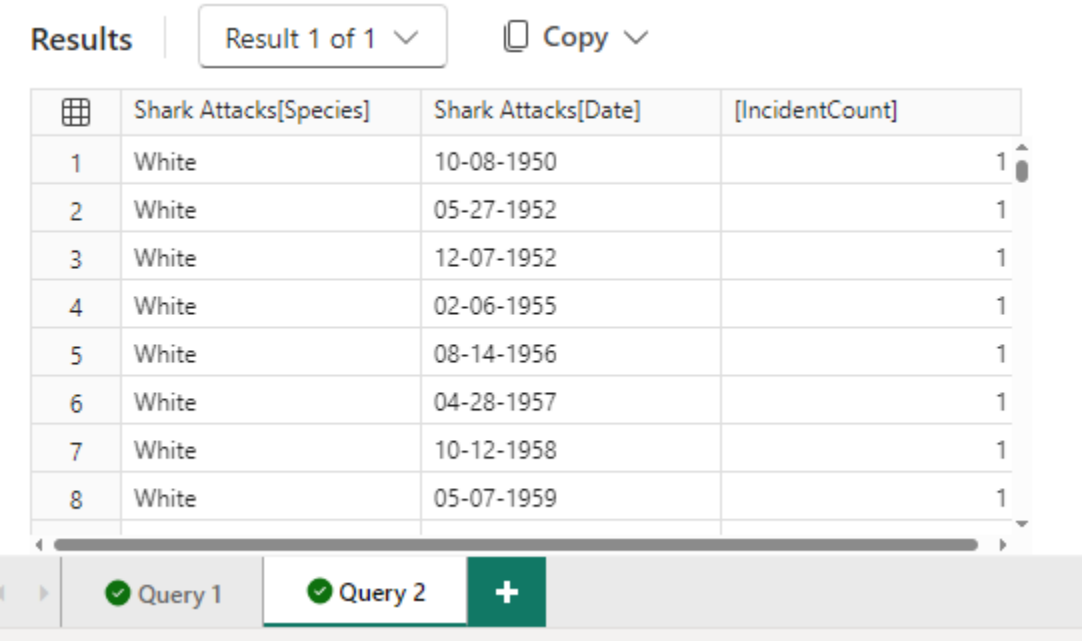
Results | Result 1 of 1 | Copy

	Shark Attacks[County]	Shark Attacks[Species]	[IncidentCount]
1	San Diego	White	15
2	Monterey	White	14
3	San Luis Obispo	White	13
4	San Francisco	White	2
5	Sonoma	White	15
6	Los Angeles	White	4
7	Marin	White	16
8	Santa Cruz	White	14

```
)
```

Q2.How has the population of specific species changed over time?

```
EVALUATE
SUMMARIZE(
    'Shark Attacks',
    'Shark Attacks'[Species],        -- Group by Species
    'Shark Attacks'[Date],     -- Group by Year (extracted from Date column)
    "IncidentCount", COUNT('Shark Attacks'[IncidentNum])  -- Count of incidents (as a proxy for population change)
)
```



Q3.Are there any correlations between environmental factors and species population?

```
EVALUATE
SUMMARIZE(
    'Shark Attacks',
    'Shark Attacks'[Species],        -- Group by Species (shark species)
    'Shark Attacks'[County],         -- Group by Location (County as a proxy for environment)
    'Shark Attacks'[Depth],          -- Group by Depth (as an environmental factor)
    'Shark Attacks'[Longitude],      -- Include Longitude (as a geographic factor)
    'Shark Attacks'[Latitude],       -- Include Latitude (as a geographic factor)
    "IncidentCount", COUNT('Shark Attacks'[IncidentNum])  -- Count of incidents (proxy for species population)
)
```

Results | Result 1 of 1 | Copy

	Shark Attacks[Species]	Shark Attacks[County]	Shark Attacks[Depth]	Shark Attacks[Longitude]	Shark Attacks[Latitude]	[IncidentCount]
1	White	San Diego	surface	-117.15	32.59	1
2	White	San Diego	surface	-117.25	32.59	1
3	White	Monterey	surface	-122.05	36.63	1
4	White	Monterey	surface	-122.15	36.63	1
5	White	San Luis Obispo	surface	-120.65	35.14	1
6	White	San Luis Obispo	surface	-120.88	35.4	1
7	White	San Diego	surface	-117.22	32.69	1
8	White	San Francisco	surface	-122.48	37.79	1

Query 1 | Query 2 | Query 3 | +

Q4. What are the trends in animal sightings and marine life in various geographic areas?

```
EVALUATE
SUMMARIZE(
    'Shark Attacks',
    'Shark Attacks'[County],          -- Group by geographic area (County as a proxy for geographic region)
    'Shark Attacks'[Date],       -- Group by Year (extracted from Date to observe trends over time)
    'Shark Attacks'[Species],          -- Group by Species (shark species)
    "IncidentCount", COUNT('Shark Attacks'[IncidentNum])  -- Count of incidents (proxy for sightings)
)
```

Results | Result 1 of 1 | Copy

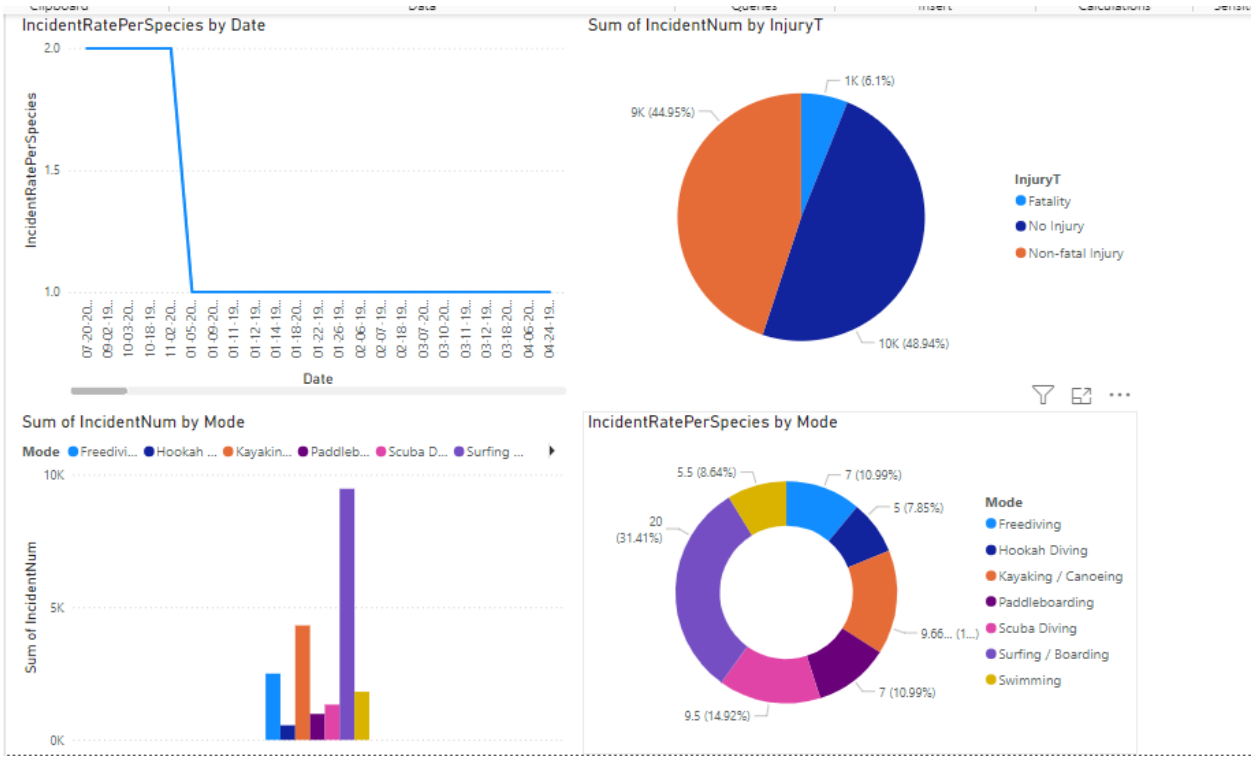
	Shark Attacks[County]	Shark Attacks[Date]	Shark Attacks[Species]	[IncidentCount]
1	San Diego	10-08-1950	White	1
2	San Diego	05-27-1952	White	1
3	Monterey	12-07-1952	White	1
4	Monterey	02-06-1955	White	1
5	San Luis Obispo	08-14-1956	White	1
6	San Luis Obispo	04-28-1957	White	1
7	San Diego	10-12-1958	White	1
8	San Francisco	05-07-1959	White	1

Query 1 | Query 2 | Query 3 | Query 4 | +

Q5. Are there any significant outliers or anomalies in species population data?

```
EVALUATE
FILTER(
    ADDCOLUMNS(
        'Shark Attacks',
        "MeanIncidentCount", CALCULATE(AVERAGE('Shark Attacks'[IncidentNum]),
ALLEXCEPT('Shark Attacks', 'Shark Attacks'[Species])),
        "StdDevIncidentCount", CALCULATE(STDEV.P('Shark Attacks'[IncidentNum]),
ALLEXCEPT('Shark Attacks', 'Shark Attacks'[Species])),
        "UpperThreshold", CALCULATE(AVERAGE('Shark Attacks'[IncidentNum]),
ALLEXCEPT('Shark Attacks', 'Shark Attacks'[Species])) +
                          2 * CALCULATE(STDEV.P('Shark Attacks'[IncidentNum]),
ALLEXCEPT('Shark Attacks', 'Shark Attacks'[Species])),
        "LowerThreshold", CALCULATE(AVERAGE('Shark Attacks'[IncidentNum]),
ALLEXCEPT('Shark Attacks', 'Shark Attacks'[Species])) -
                          2 * CALCULATE(STDEV.P('Shark Attacks'[IncidentNum]),
ALLEXCEPT('Shark Attacks', 'Shark Attacks'[Species]))
    ),
    'Shark Attacks'[IncidentNum] > [UpperThreshold] || 'Shark Attacks'[IncidentNum] <
[LowerThreshold]
)
```

DASHBOARD



Conclusion:

The experiment on creating an interactive dashboard using DAX queries for the Shark Attack dataset provided valuable insights into shark attack patterns and trends. By analyzing factors such as species distribution, geographic concentration, and injury severity, the dashboard effectively visualized key aspects of marine life incidents. The queries revealed significant findings, such as the prevalence of White Shark attacks, the dominance of surfing-related incidents, and geographical hotspots like San Diego County. These visualizations facilitate better understanding of species trends and population dynamics over time, providing a useful tool for both ecological study and public safety efforts.