ADVANCED DATA VISUALIZATION
EXPERIMENT 6

NAME: Hrimkar Doshi
ROLL NO: 20217000021
BATCH: L
CSE - DS

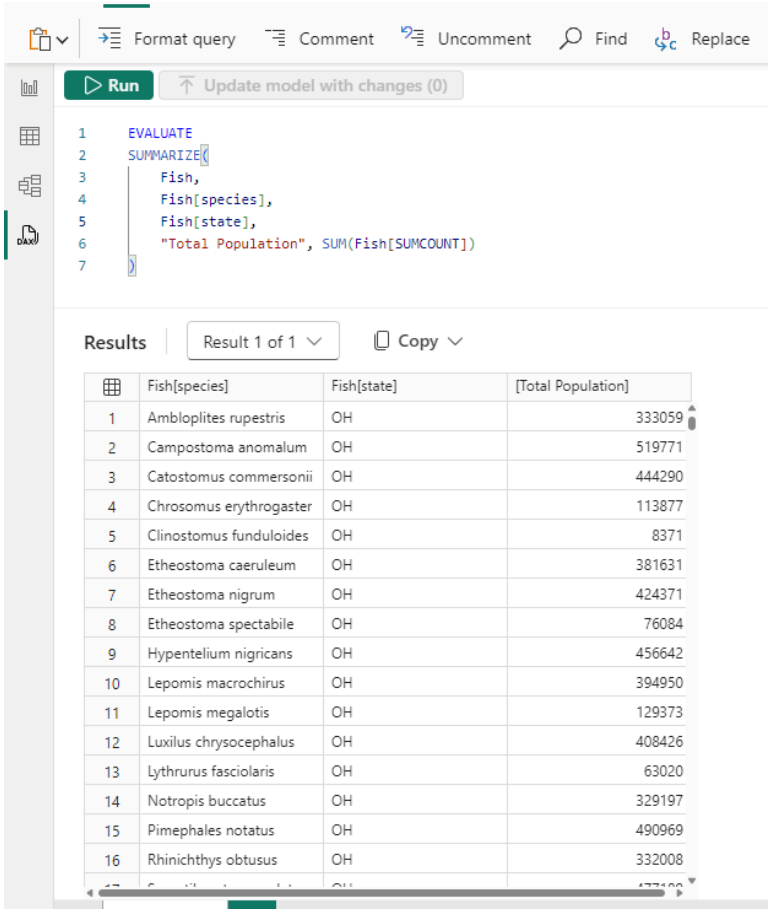
AIM: To design interactive dashboards using Power BI for visualizing and analyzing an Animal/Wildlife/Marine dataset, employing both basic and advanced charts to uncover insights and trends.

DATASET: Marine dataset with different Fish Species distribution
Link -

DESCRIPTION WITH OUTPUT:

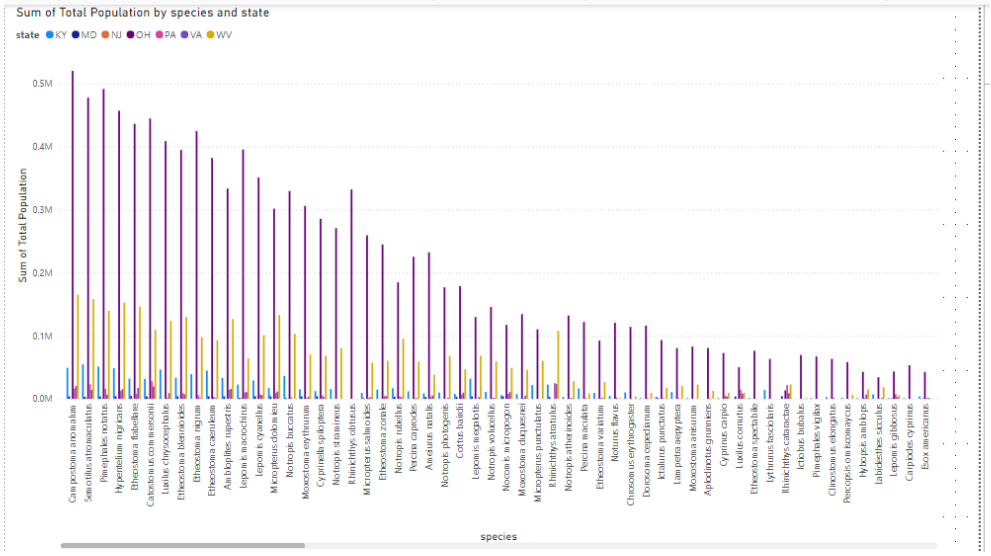
1. **What are the population distributions of various species across different regions?**

Using DAX Summarize to extract the sum of population after grouping the data according to species and state.



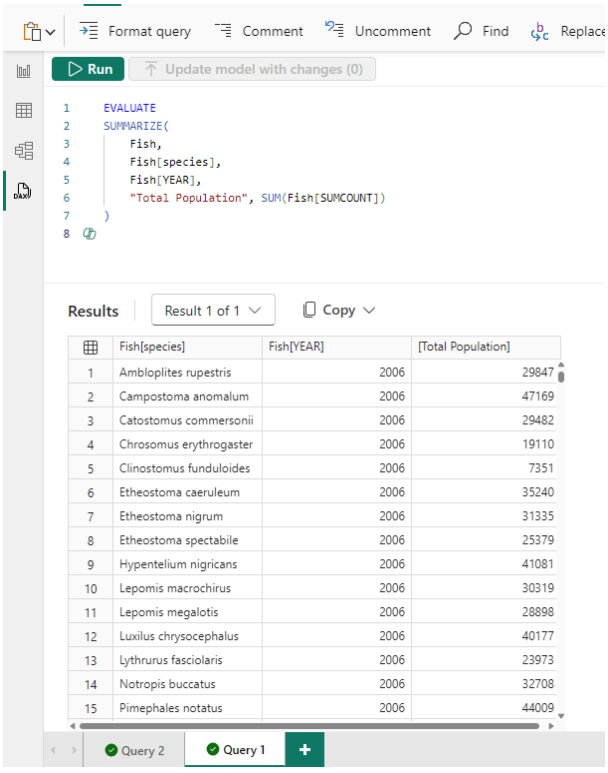
	Fish[species]	Fish[state]	[Total Population]
1	Ambloplites rupestris	OH	333059
2	Campostoma anomalum	OH	519771
3	Catostomus commersonii	OH	444290
4	Chrosomus erythrogaster	OH	113877
5	Clinostomus funduloides	OH	8371
6	Etheostoma caeruleum	OH	381631
7	Etheostoma nigrum	OH	424371
8	Etheostoma spectabile	OH	76084
9	Hypentelium nigricans	OH	456642
10	Lepomis macrochirus	OH	394950
11	Lepomis megalotis	OH	129373
12	Luxilus chrysocephalus	OH	408426
13	Lythrurus fasciolaris	OH	63020
14	Notropis buccatus	OH	329197
15	Pimephales notatus	OH	490969
16	Rhinichthys obtusus	OH	332008

Visualization: Grouped Bar chart showing population of each species where color shows the different regions. I did this by creating a calculated table in the modeling section of powerBI to segregate the required data.



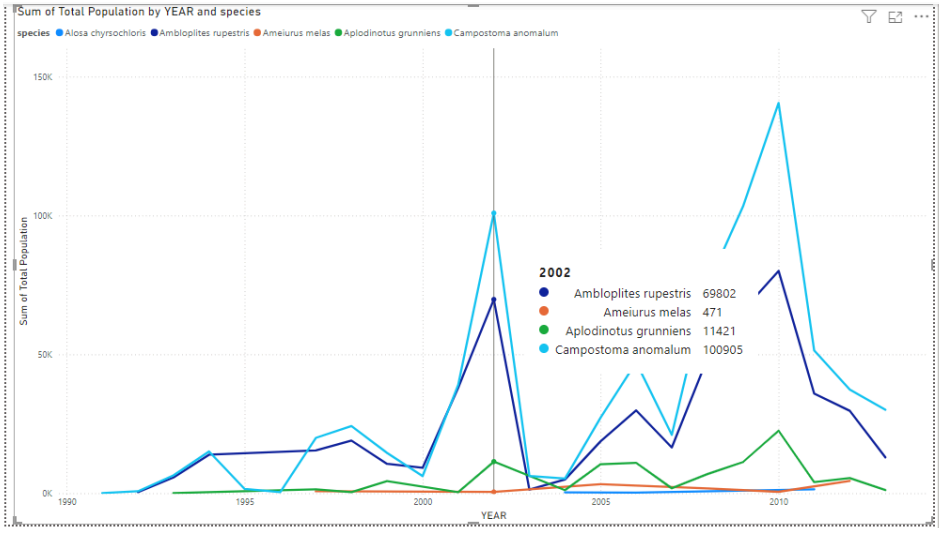
Observation: So we can see that state OH has the highest population when it comes to Campostoma Anomalum species.

2. **How has the population of specific species changed over time?**
   Using DAX Summarize to extract the sum of population for each year to analyse change of trends in species.



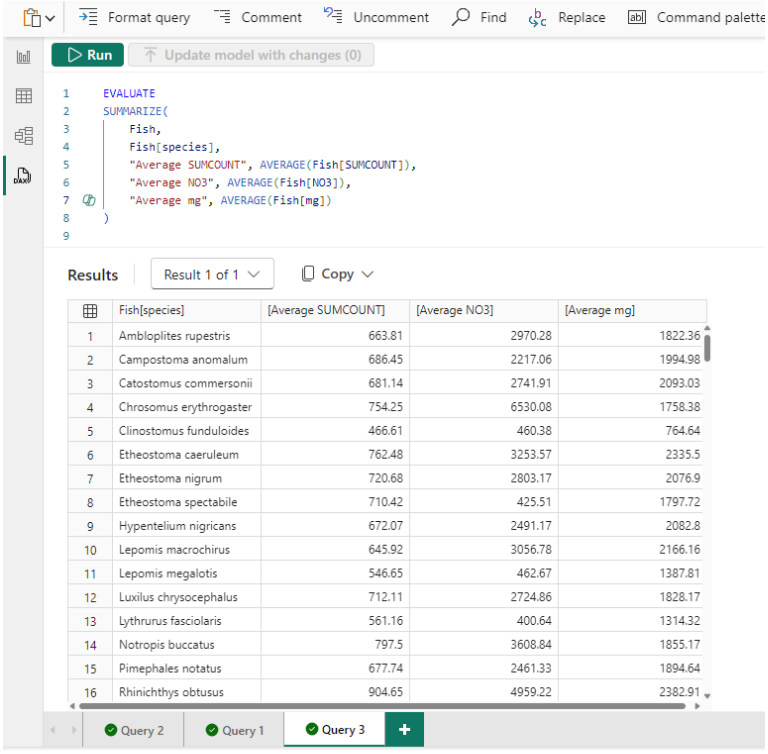
	Fish[species]	Fish[YEAR]	[Total Population]
1	Ambloplites rupestris	2006	29847
2	Campostoma anomalum	2006	47169
3	Catostomus commersonii	2006	29482
4	Chrosomus erythrogaster	2006	19110
5	Clinostomus funduloides	2006	7351
6	Etheostoma caeruleum	2006	35240
7	Etheostoma nigrum	2006	31335
8	Etheostoma spectabile	2006	25379
9	Hypentelium nigricans	2006	41081
10	Lepomis macrochirus	2006	30319
11	Lepomis megalotis	2006	28898
12	Luxilus chrysocephalus	2006	40177
13	Lythrurus fasciolaris	2006	23973
14	Notropis buccatus	2006	32708
15	Pimephales notatus	2006	44009

Visualization: Line chart showing population trends of 5 species over time. I have selected only 5 species to make the chart less cluttered considering there are a huge number of species.



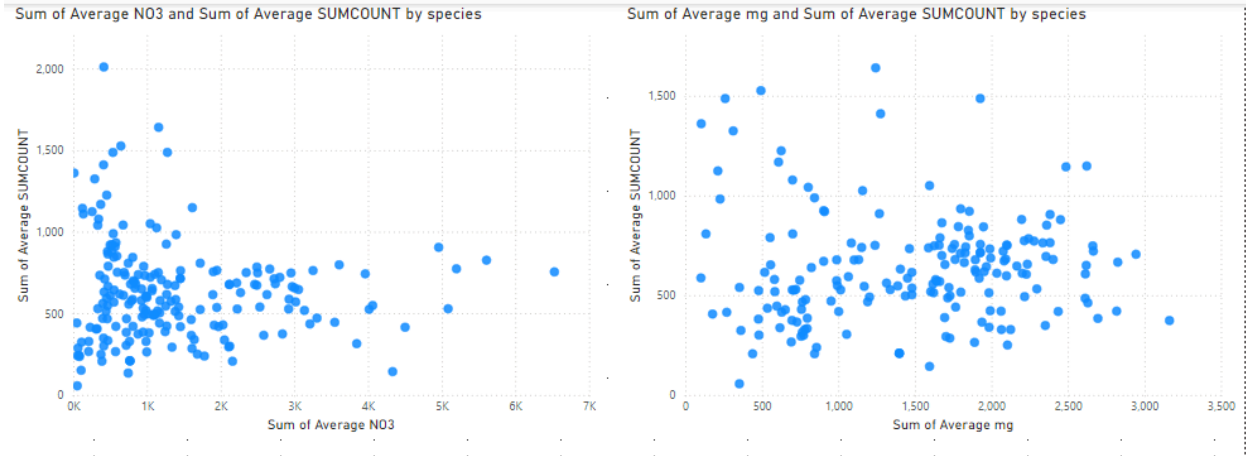
Observation: As we can see, the population of Campostoma Anomalum (light blue line) started very small in 1990 but has continued to grow with 69802 count in 2002, and peaking in 2010 as the highest species recording among these 5 chosen species.

3. **Are there any correlations between environmental factors and species population?**
   For this query I have averaged out the amount of NO3 and Mg present in the environment (water for marine case) to study relation with species population using DAX query AVERAGE function.



	Fish[species]	[Average SUMCOUNT]	[Average NO3]	[Average mg]
1	Ambloplites rupestris	663.81	2970.28	1822.36
2	Campostoma anomalum	686.45	2217.06	1994.98
3	Catostomus commersonii	681.14	2741.91	2093.03
4	Chrosomus erythrogaster	754.25	6530.08	1758.38
5	Clinostomus funduloides	466.61	460.38	764.64
6	Etheostoma caeruleum	762.48	3253.57	2335.5
7	Etheostoma nigrum	720.68	2803.17	2076.9
8	Etheostoma spectabile	710.42	425.51	1797.72
9	Hypentelium nigricans	672.07	2491.17	2082.8
10	Lepomis macrochirus	645.92	3056.78	2166.16
11	Lepomis megalotis	546.65	462.67	1387.81
12	Luxilus chrysocephalus	712.11	2724.86	1828.17
13	Lythrurus fasciolaris	561.16	400.64	1314.32
14	Notropis buccatus	797.5	3608.84	1855.17
15	Pimephales notatus	677.74	2461.33	1894.64
16	Rhinichthys obtusus	904.65	4959.22	2382.91

Visualization: To understand the correlation of species population with environment features I have plotted scatter plots for both average NO3 and Mg.

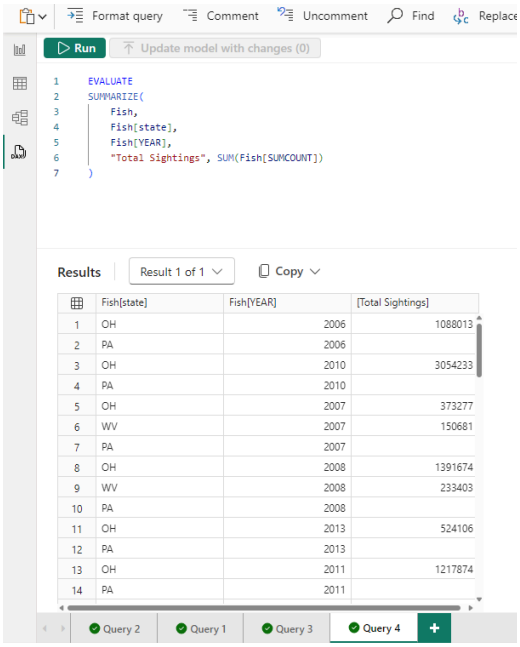


Observation:

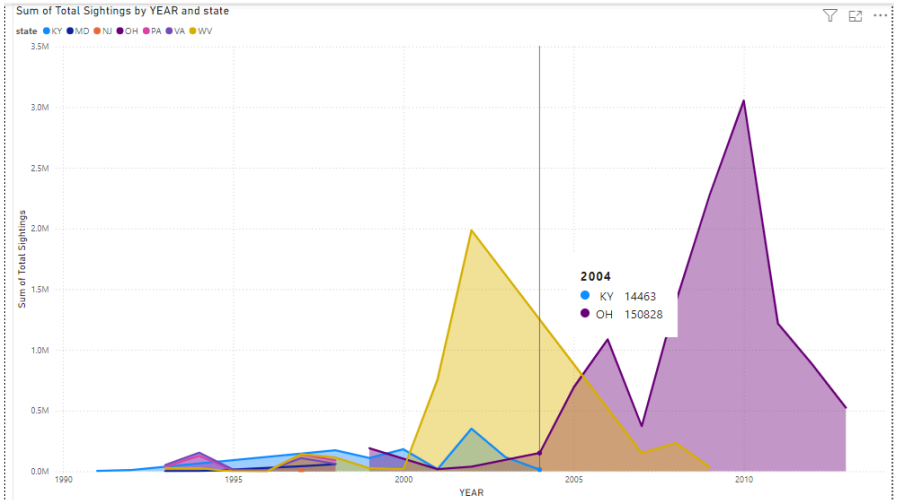
- NO3 vs Species Population: The relationship between the average NO3 levels and species population appears scattered, with no clear linear trend. There are clusters of species populations at lower NO3 values (around 1K-2K), but the population does not increase proportionally with higher NO3 levels.
- Mg vs Species Population: The relationship between the average Mg and species population shows a similarly dispersed pattern, though there is a slightly broader spread at higher Mg levels. No strong positive or negative correlation is evident from the plot.

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

   I have used DAX SUM function to group the sightings by area and year.

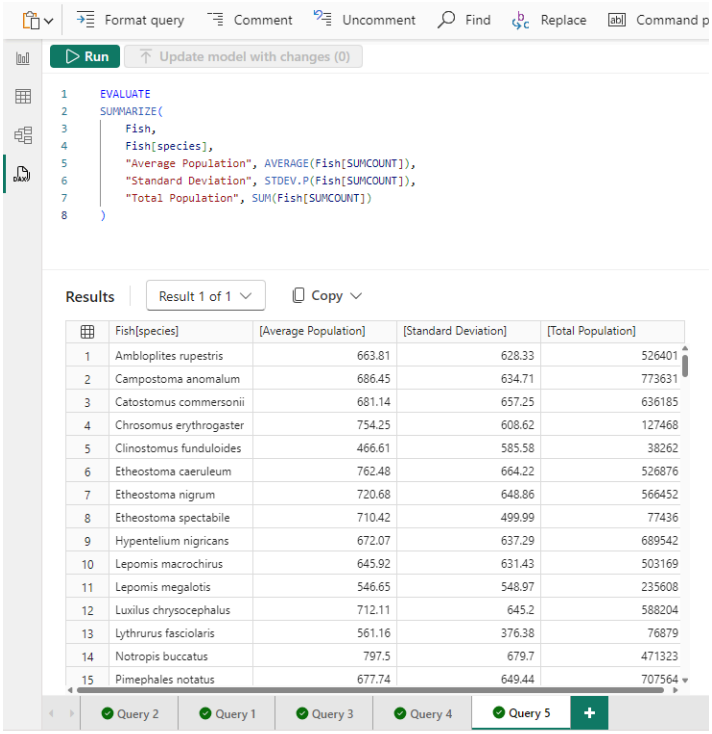


Visualization: I have used Area charts to show the trends in marine life sightings over the years depending on region. The species populations are summed for each year according to state.



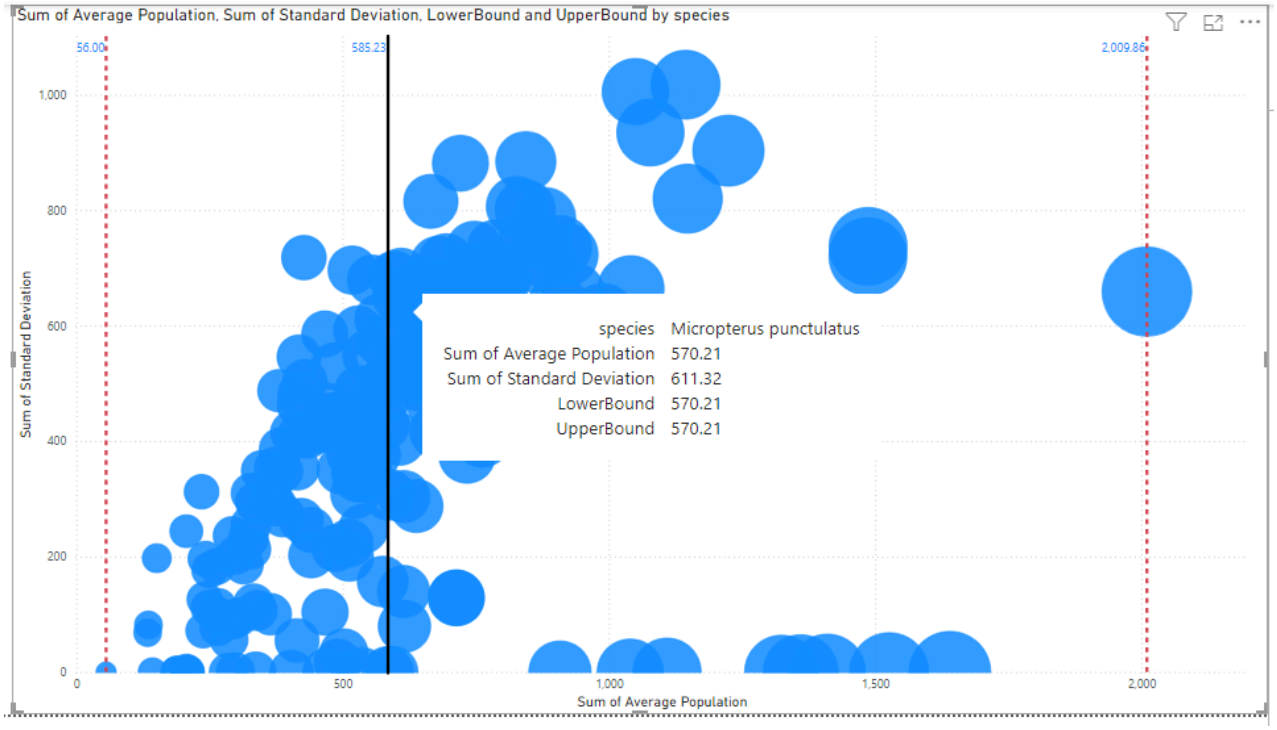
Observation: Using the highlighted point we can observe that in 2004 marine species were sighted only in state codes KY and OH. From the overall graph we can see that for state OH (purple) the population was less initially but has increased in a few years again decreasing at the end.

5. **Are there any significant outliers or anomalies in species population data?**
   For outlier detection I have used AVERAGE, STDEV.P and SUM functions from DAX queries. As box plots are not available in powerBI i have used these statistics to create measures for outlier detection.



	Fish[species]	[Average Population]	[Standard Deviation]	[Total Population]
1	Ambloplites rupestris	663.81	628.33	526401
2	Campostoma anomalum	686.45	634.71	773631
3	Catostomus commersonii	681.14	657.25	636185
4	Chrosomus erythrogaster	754.25	608.62	127468
5	Clinostomus funduloides	466.61	585.58	38262
6	Etheostoma caeruleum	762.48	664.22	526876
7	Etheostoma nigrum	720.68	648.86	566452
8	Etheostoma spectabile	710.42	499.99	77436
9	Hypentelium nigricans	672.07	637.29	689542
10	Lepomis macrochirus	645.92	631.43	503169
11	Lepomis megalotis	546.65	548.97	235608
12	Luxilus chrysocephalus	712.11	645.2	588204
13	Lythrurus fasciolaris	561.16	376.38	76879
14	Notropis buccatus	797.5	679.7	471323
15	Pimephales notatus	677.74	649.44	707564

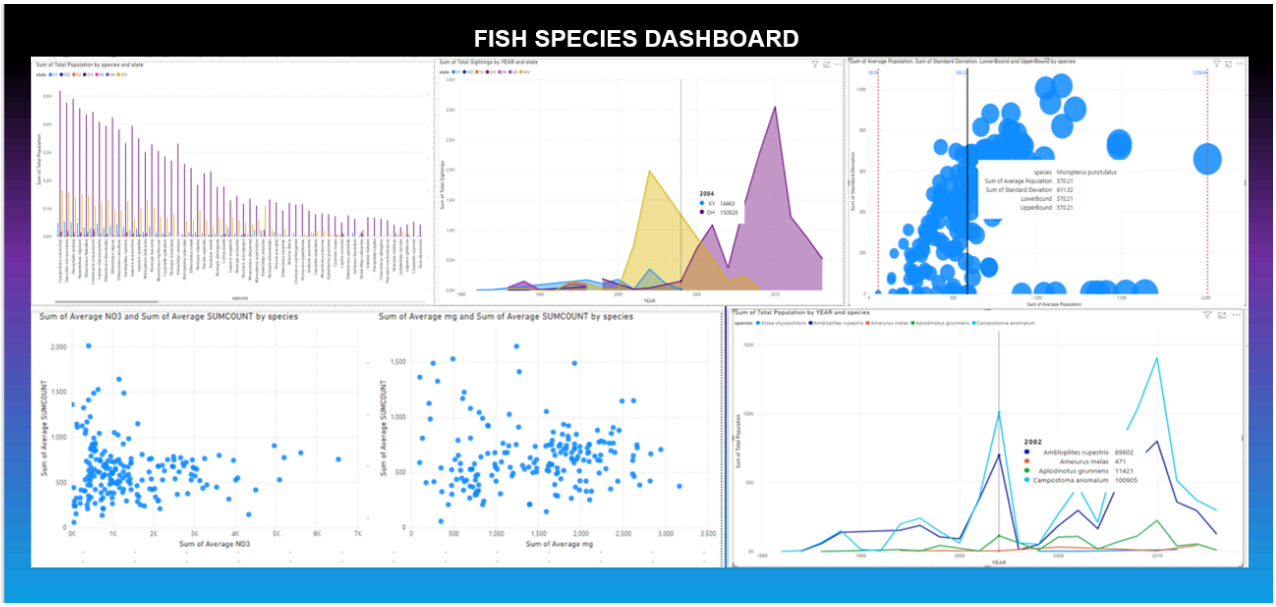
Visualization: Following is a bubble chart showing the standard deviation and mean value for each species, with higher population species having larger bubble size. I have measures to calculate the Upper and Lower bounds along with IQR and median.



Observation:

- Bubble Size: The larger bubbles represent species with higher population.The highlighted species, Micropterus Punctulatus, has an average population around 570, with a standard deviation of 611.32.
- Population Distribution: A dense cluster of species falls within the lower ranges of both population and standard deviation, particularly between 0 to 500 on the x-axis and 0 to 400 on the y-axis. This suggests that many species have a relatively small average population with less variability.
- Outliers: There are a few species with much larger populations, above 1000 on the x-axis, as indicated by the bubbles towards the right. These species are farther apart from the majority, indicating that their populations are not only larger but also potentially more variable, as reflected by their larger standard deviations.
- Thresholds: The red dashed lines show the upper and lower bounds on data, on which species are considered outliers in terms of population size and variation.
- Median Line: I have also plotted the median line to see the spread of the population data. We can clearly observe that there are more outliers on the higher extreme suggesting peak population of some species.

DASHBOARD:


FISH SPECIES DASHBOARD
Sum of Total Population by species and state
Sum of Total Sightings by YEAR and state
2004
KY 14463
OH 150828
Sum of Average Population, Sum of Standard Deviation, LowerBound and UpperBound by species
species Micropterus punctulatus
Sum of Average Population 570.21
Sum of Standard Deviation 611.32
LowerBound 570.21
UpperBound 570.21
Sum of Average NO3 and Sum of Average SUMCOUNT by species
Sum of Average SUMCOUNT
Sum of Average NO3
Sum of Average mg and Sum of Average SUMCOUNT by species
Sum of Average SUMCOUNT
Sum of Average mg
Sum of Total Population by YEAR and species
2002
Ambloplites rupestris 69802
Ameiurus melas 471
Aplodinotus grunniens 11421
Campostoma anomalum 100905
YEAR