

## Project Overview:

Coastal pollution assessment is a pressing matter as the anthropogenic pressure continues to increase worldwide. A leading approach to assess coastal pollution is using bioindicators. Our goal was to find the meiobenthic composition in each pollution level.

## Data overview:

The dataset provides useful information for the pollution levels, organic matter content, and physical properties of the water and soil samples collected from different shores. The analysis of the data reveals that pollution levels vary across seasons and months, with higher levels in certain months and seasons than others.

**Data source:** [Shore Prediction Dataset](#)

## Data Preparation:

The data was cleaned and transformed in Python notebook. before being imported into Power BI. The steps taken to prepare the data included removing null values, filtering out unnecessary columns, treating outliers and some other preprocessing concepts.

## **Questions we are trying to find its answers:**

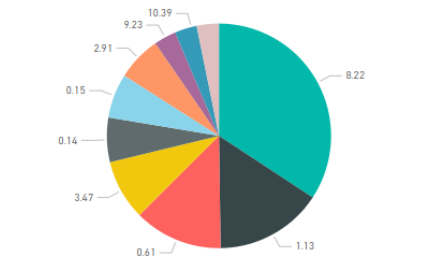
- 1) What are the pollution levels we have and how they are distributed along the shores?
- 2) Are all features affecting the shores in a negative way or some may affect in a positive way?
- 3) Is Water/Soil salinity having a direct relation with pollution level?
- 4) How do total dissolved salts varies from one shore to another?
- 5) Does temperature have a strong influence on pollution?
- 6) How salinity affects PH and Conduction of water?

## **Conclusion and insights we've gained from our analysis:**

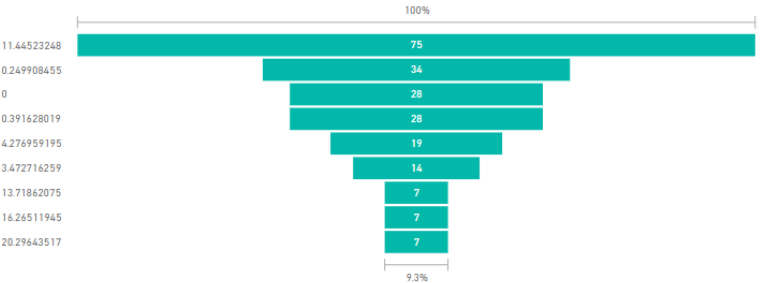
- Shore 1 is less polluted (pollution level = 0)
- Shore 2 is more polluted (pollution level = 1)
- Shore 3 is the most polluted (pollution level = 2)
- Resistance is one of the most affecting features on pollution level.
- Temperature and season vary in all pollution levels.
- Water PH has a strong correlation with conductivity, salinity, and dissolved salts.
- higher conductivity levels corresponding to lower pollution levels.
- there is a significant difference in the organic matter content between different shores, with Shore 1 having the highest organic matter content and Shore 3 having the lowest.

# Visualizations:

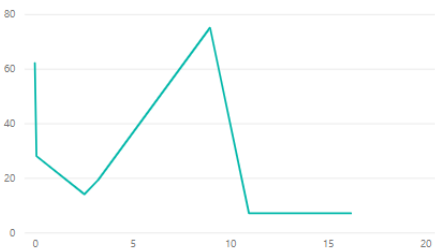
Count of C-A by C-A



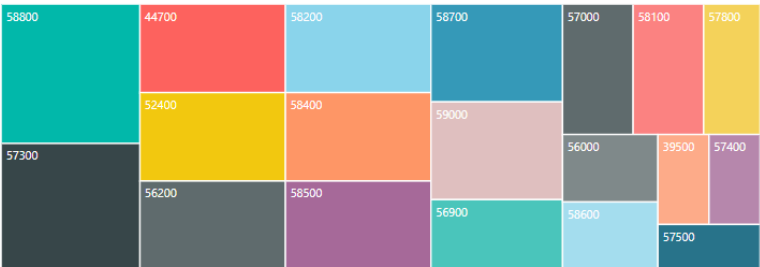
Count of C-C by C-C



Count of C-B by C-B



Count of Conduction by Conduction



Season

Month

Shore

Pollution Level

All

All

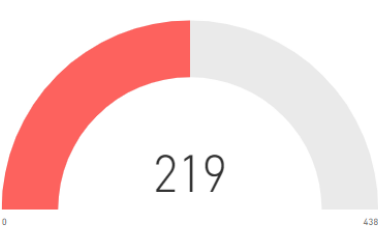
All

All

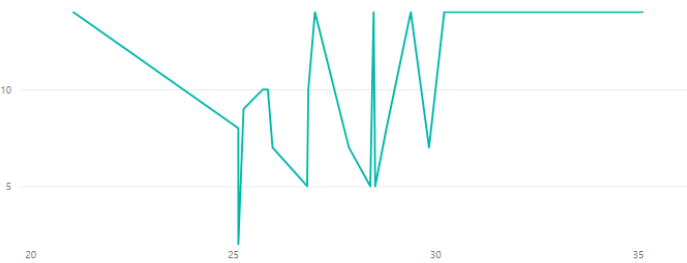
Count of PP by PP



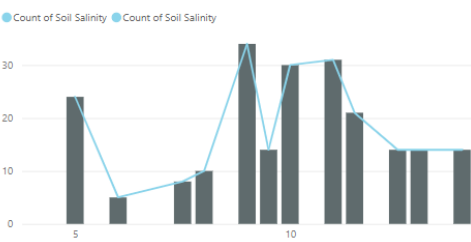
Count of Sample number



Count of P by P



Count of Soil Salinity and Count of Soil Salinity by Soil Salinity



Season

Month

Shore

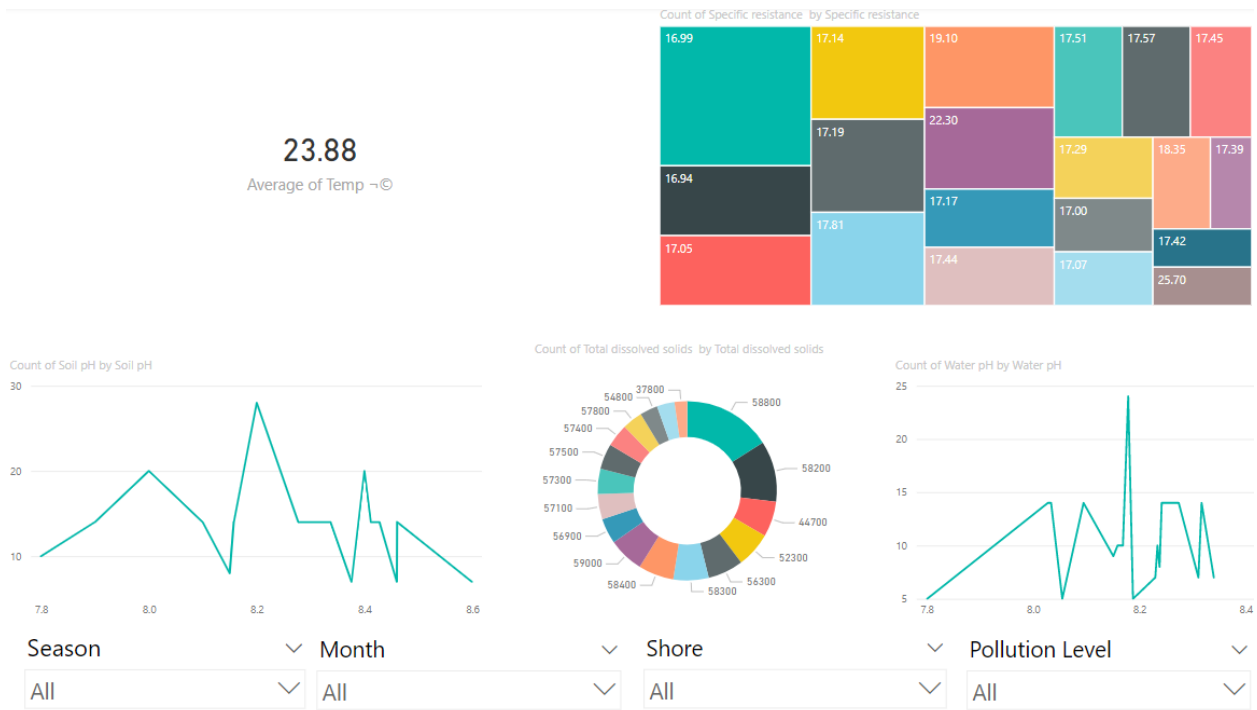
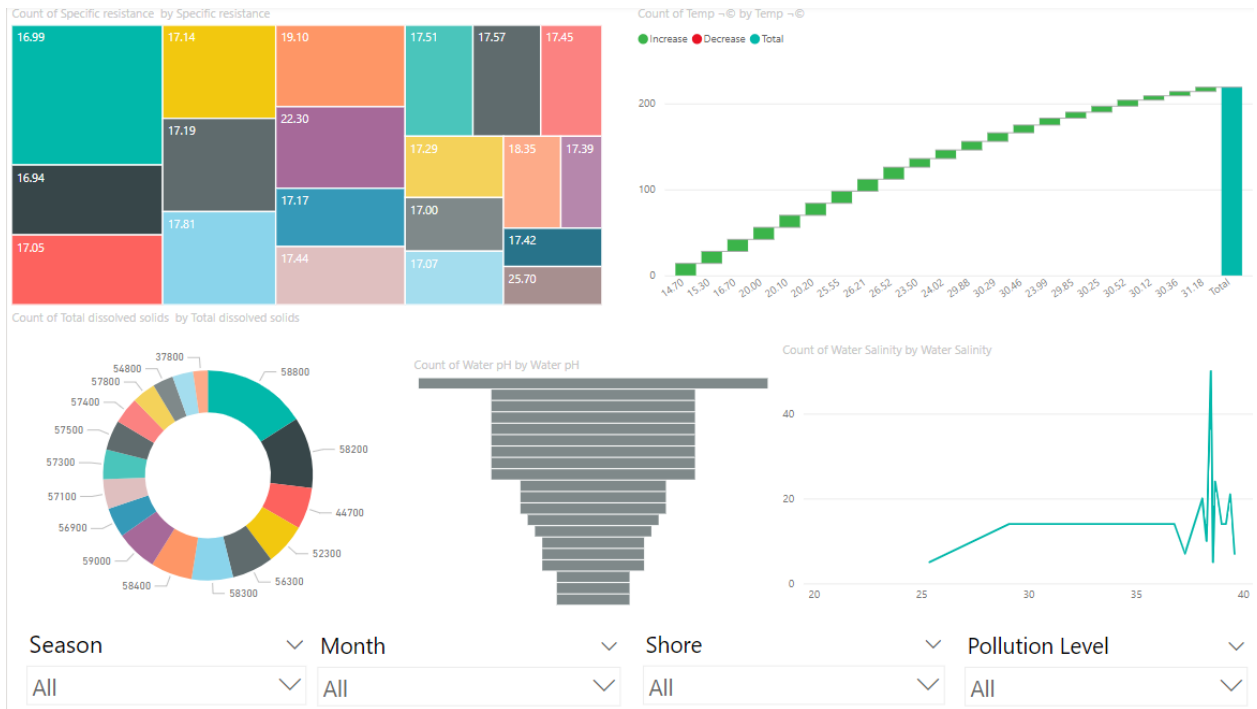
Pollution Level

All

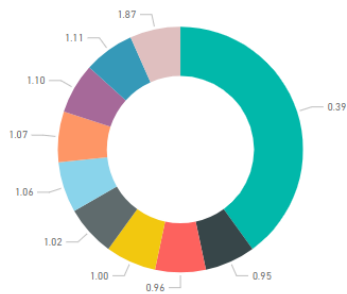
All

All

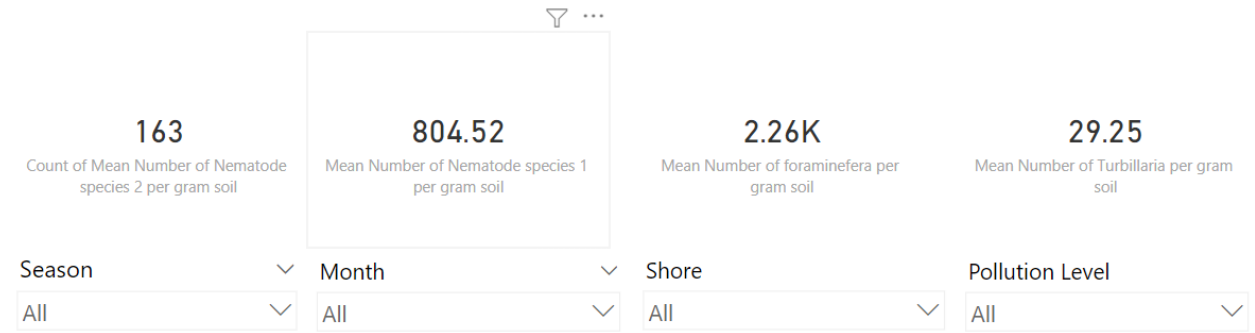
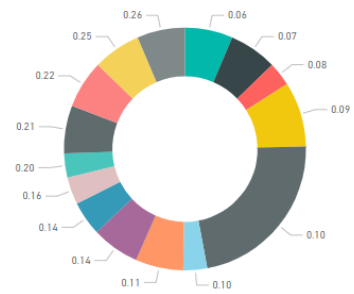
All



top 10 Count of Organic matter% by Organic matter%



Count of H by H



All you need to do is to choose between the season, month, shore and pollution level and you will see how all features behave.

For more visualizations have a look at the notebook: [Shore Prediction Notebook](#)

**All Rights Reserved to [ApplAi](#)**

**Under the supervision of:**

- Director of Machine Learning and Training Department: [Eslam Shouman](#)

**Team Leader:**

- [Nour Ahmed](#)

**Team Members:**

- [Jonathan Mounir](#)
- [Sarah Ahmed](#)
- [Yumna Muhammed](#)
- [Habiba Alaa](#)