WATER QUALITY PREDICTOR

A Group 9 Data Analytics Production

Rubric

Segment 1

Content

Team members have drafted their project, including the following:

✓ Selected topic

✓ Reason why they selected their topic

✓ Description of their source of data

✓ Questions they hope to answer with the data

Note: The content does not yet need to be in the form of a presentation; text in the README.md works as well.

Topic

Water Quality Predictor

Why do we need a Water Quality Predictor?

We all love recreation activities in and around water. We always wonder where we should go, and which locations have the safest water for recreational activities, as well as which locations have potable water.

Where do we get data to predict water quality?

King County, Washington Lake Services and Information website:

https://green2.kingcounty.gov/lakes/Query.aspx

What questions do we hope to answer?

- Is the water safe for recreational activities (e.g. swimming, boating, water skiing, etc.)?
- Is this water safe to drink?
- What is the water temperature?

Rubric

Segment 2

Content

✓ Description of the data exploration phase of the project

✓ Description of the analysis phase of the project

Description of the data exploration phase of the project

- As a team we:
  - downloaded a .csv file of one testing station from Lake Sammamish from the King County, Washington Lake Services and Information website: https://green2.kingcounty.gov/lakes/Query.aspx
  - imported the .csv file using Python & Pandas in Jupyter Notebook
  - created a dataframe from the imported .csv file data
  - explored the data by checking the shape and count of values in each column
  - cleaned the data by
    - remove unneeded columns
    - remove null values
    - reformat data types

Description of the analysis phase of the project

- We realized that the formula for Water Quality Index (WQI), needed to be calculated and added to the dataframe
  - Ritu created a MS-Excel spreadsheet/formula to calculate the WQI from the data
  - Nicole translated the MS-Excel formula into a python calculation and added a new column to the dataframe to save that calculation for each row of data
- Once the data was cleaned, we each chose one type of machine learning model to experiment with
  - Erica and Ritu tried Neural Network models
  - Nicole and Ling tried regression models
- We agreed as a team that a regression model would be the simplest to use
- Ritu & Ling worked to optimize the model
- Erica created an ERD and database
- Ling created a draft of the dashboard containing charts and reports to visually display our data
- Nicole created a draft of the presentation
- Nicole created a simple webpage to display the WQI predictor values