# Mod20\_Segment\_1

By the end of this module segment, we will have created the foundation for the final project. By defining roles between team members and establishing a communication structure. Additionally, Decide on a topic for the project—think of a question that can be answered using data. Create a repository for the project and invite the other team members to join. Source a dataset that will suit your needs (you can even use multiple datasets if applicable). Begin to clean, organize, and perform exploratory data analysis on your datasets so that they're ready for analysis. Include mockups of a machine learning model and a database.

# Purpose

First Segment: Sketch It Out: Decide on your overall project, select your question, and build a simple model. You'll connect the model to a fabricated database, using comma-separated values (CSV) or JavaScript Object Notation (JSON) files, to prototype your idea.

# Presentation

\* Selected topic:

Predicted when polar ice will be gone in the Artic and Antarctic's

\* Reason we selected the topic:

Sea ice, which is comprised of frozen ocean water, is found only in the Arctic and Antarctic. The amount of sea ice at each pole changes throughout the year, growing in the winter months and shrinking during the summer months. This visualization on Google Earth shows the minimum sea ice measured for each year from 1979 to 2018. And could expand this to 2022. For the Antarctic, this measurement is taken in March after summer in the Southern Hemisphere. Sea ice has an impact on the global climate from helping to regulate global temperature to affecting ocean currents and providing a habitat for wildlife. We want to explore the catastrophes and aftermath that may occur after that date, and provide useful information related to the expected extinction of species to help navagate hard times in the future if we don't do something now.

## Description of the source of data:

1. National Snow and Ice Data Center. NSIDC data on Google Earth. Digital media. http://nsidc.org/data/google\_earth. Accessed March 1, 2019.

2. wildfinder Database https://www.worldwildlife.org/publications/wildfinder-database

1. National Snow and Ice Data Center. NSIDC data on Google Earth. Digital media. http://nsidc.org/data/google\_earth. Accessed March 1, 2019.

2. wildfinder Database https://www.worldwildlife.org/publications/wildfinder-database

**## Questions we hope to answer with the data:**

1. Predicting when the polar ice will be gone in the Artic and Antarctic's

2. Predicting Instinction of which animals due to polar ice melt

# Github

\* Description of the communication protocols:

We created a message group for direct messages with team members in Slack. Established a daily meetup time for everyone and a communication protocol in case of an emergency as follows:

![D2a](https://github.com/ALIYA2Group/Mod20\_Segment\_1/blob/main/Pictures/D1a.PNG)

\* Individual Branches:

There is one branch for each team member as follows:

![D2b](https://github.com/ALIYA2Group/Mod20\_Segment\_1/blob/main/Pictures/D2b.PNG)

\* Each team member has at least four commits for the duration of the first segment:

# Machine Learning Model

We will present a provisional machine learning model that stands in for the final machine learning model and accomplishes the following:

\* Takes in data from the provisional database

\* Outputs label for input data

Questions

• Which model did you choose and why?

• How are you training your model?

• What is the model's accuracy?

• How does this model work?

# Database

Present a provisional database that stands in for the final database and accomplishes the following:

\* Sample data that mimics the expected final database structure or schema

\* Draft machine learning model is connected to the provisional database

# Technologies Used

• Which tools are the best fit for your project?

• What will be used for each section?

• For example, data cleaning and data storage will be completed using two different tools.

• How will the dashboard be built?

## Data Cleaning and Analysis

ex. Pandas will be used to clean the data and perform an exploratory analysis. Further analysis will be completed using Python.

## Database Storage

ex. Mongo is the database we intend to use, and we will integrate Flask to display the data.

## Machine Learning

ex. SciKitLearn is the ML library we'll be using to create a classifier. Our training and testing setup is \_\_\_. Extra ML verbiage here.

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated

# Dashboard

ex. In addition to using a Flask template, we will also integrate D3.js for a fully functioning and interactive dashboard. It will be hosted on \_\_\_.

There are no deliverables associated with the dashboard for this segment.

Table

Description automatically generated