Data Engineering Capstone Project Plan

Name: Justina Konopackaja

### Intended Data Source: **Marine Weather API ->** [**http://api.weatherapi.com/v1/marine.json?key=2cb550d930e043de8c7120907242606&q=Bermuda&days=1**](http://api.weatherapi.com/v1/marine.json?key=2cb550d930e043de8c7120907242606&q=Bermuda&days=1)

# Define the Objective

Marine weather API method returns up to next 7-day marine and sailing weather forecast and tide data as Json or xml. The data is returned as a Marine Object which contains astronomy data, day weather forecast and hourly interval weather information and tide data for a given sea/ocean point. This project aims to create an automated system to dynamically fetch Marine Objects from Hamilton, Bermuda as they update to create a database. This data will then be visualised to show weather, astro and tide changes. The goal of this project is to capture weather trends in one of the famously known mysterious Bermuda Triangle destinations.

# Technical Plan

Marine weather API can be requested openly after acquiring a key. A Python script will run on a CRON job hosted on the Digital Futures server, collecting the data every 30 minutes and updating it into a SQL database with DBeaver database management software. A Streamlit database, hosted on Streamlit community cloud, will read the SQL database and visualise the trend information.

# Potential Risks

I acquire no previous experience on working with a frequently updated API, containing this large amount of data. Efficient code will have to be implemented in a script to deal with big loads of data. I also have limited knowledge on data visualisation methods on Streamlit and how to connect it to SQL database. Big amount of research and work will have to be completed in short time to meet the deadline.

# Timeline & Goals

By End of Wednesday 26th of June -> choose a frequently updated website or API and finish writing a Data Engineering Capstone Project Plan

By End of Monday 1st of July -> Complete a script that obtains the data

By End of 2nd of July -> Data collection & database updates completed.

By End of 3rd of July -> Data Visualisation Completed

By End of Wednesday 4th of July -> Diagram and presentation completed

# Stretch goals and next steps

Bermuda Triangle, also known as the Devil's Triangle, is region in the western part of the North Atlantic Ocean which associated with numerous aircraft and ship disappearance under mysterious circumstances. It has a total area which ranges between 1,300,000 and 3,900,000 square kilometres and can not be defined by a forecast of just one location. An extension of this project would be to also collect data from other islands and include more geological and seismographic data to learn what might have influenced this area to become known for being so mysterious and dangerous.