Blue Ocean Gear
Exploratory Data Analysis of External Datasets

University of Chicago - Data Science Clinic

Aurora Peng
8th Nov, 2022

Goals of EDA - Analysis of Copernicus Satellite Data

We had 4 main objectives for this analysis:

1. Retrieve Copernicus Data and Merge with Buoy Data
2. Visualize Copernicus Data
3. Analyze Buoy Water Temperature
4. Compare buoy reported temperatures and Copernicus' reported temperatures

# Retrieve Copernicus Data and Merge with Buoy Data

- Data source: Copernicus Satellite Data from E.U. Copernicus Marine Service Information
  - It contains the **hourly** mean surface fields (e.g., temperature, currents, etc) with 10-day forecast
  - Key variable: thetao (**temperature in degree celsius**)
- Merging data
  - Method: By drawing buffers (circles) around each buoy location and calculating the min, max, and average temperature of the satellite data within the buffers
  - Date range: between 2022-04-11 and 2022-10-04 (inclusive) (4201 hours)
  - Volume: 29242 observations
  - Variables: buoy coordinates, buoy reported temperature, temperature (min, max, mean, median) from the Copernicus data, sensor id, fishery id, system_status_interpret, etc

Visualize Copernicus Data

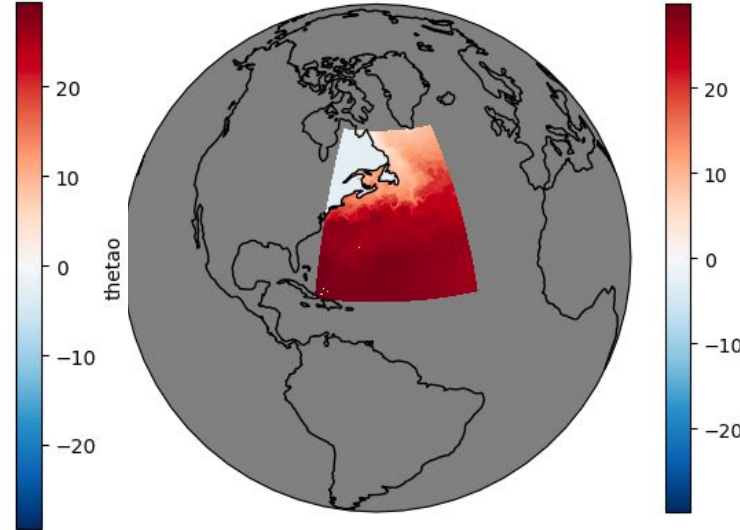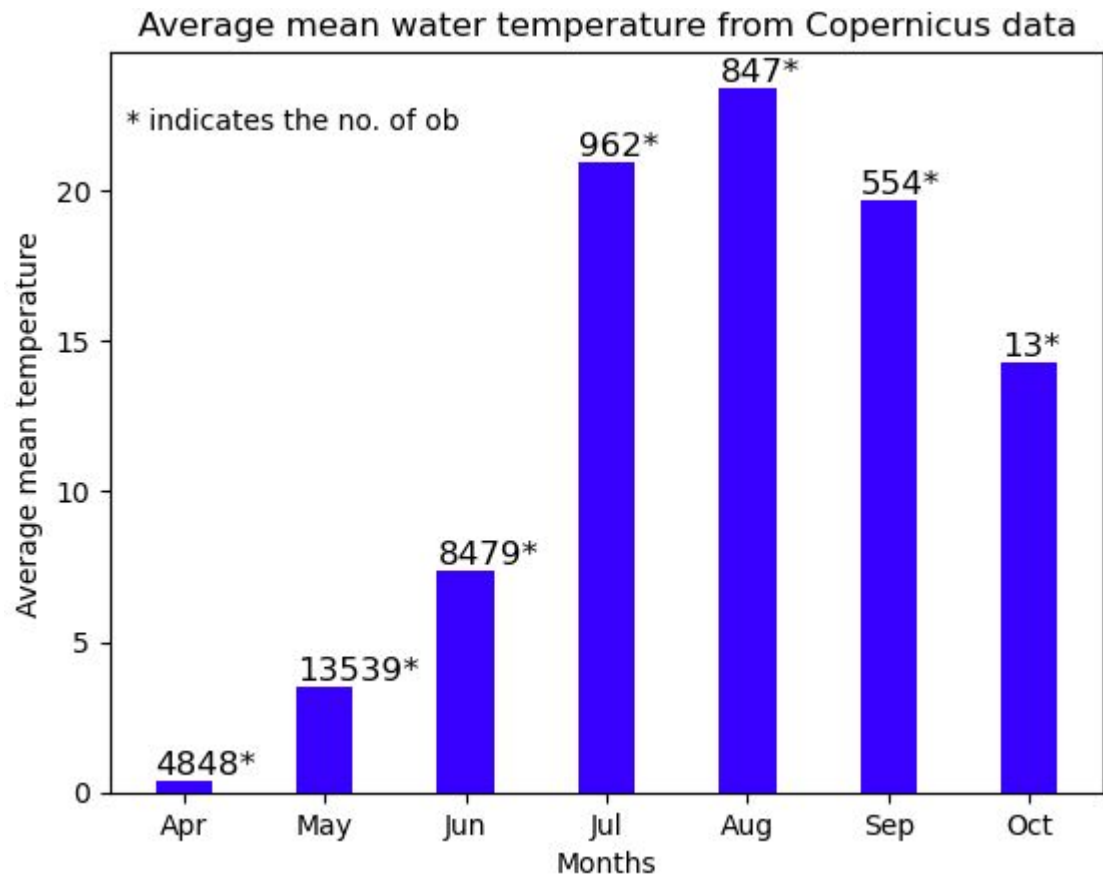time = 2022-04-12T13:30:00, depth = 0.494
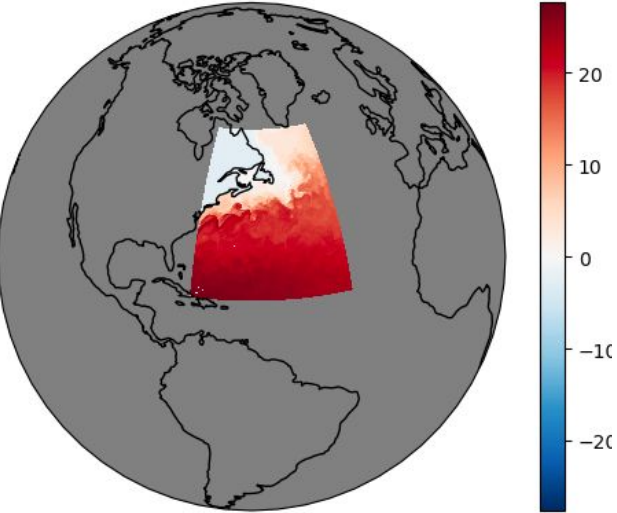
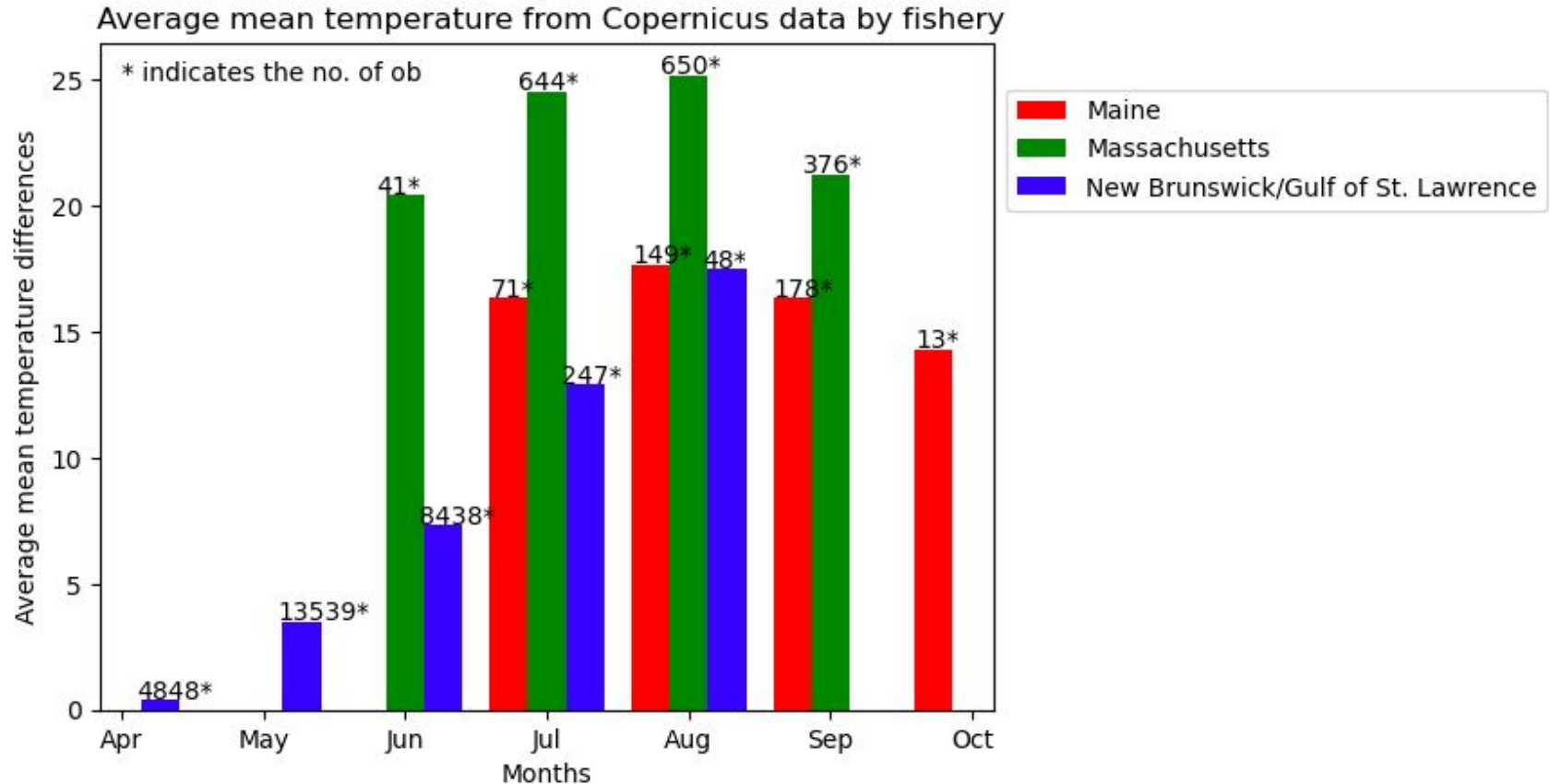time = 2022-07-05T06:30:00, depth = 0.494
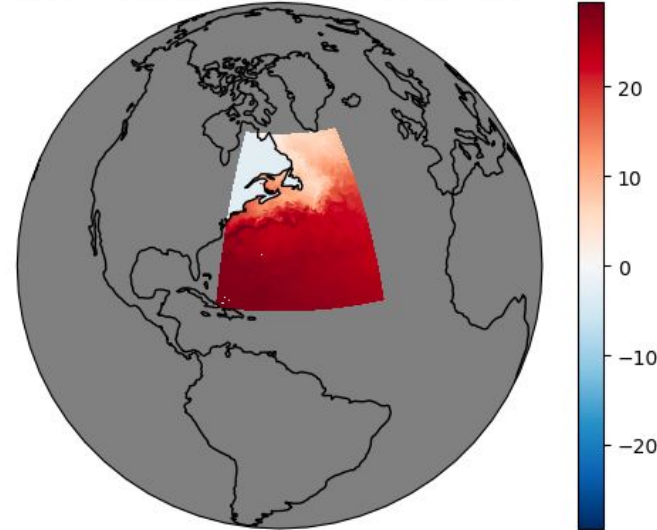
ime = 2022-09-25T07:30:00, depth = 0.494

Visualize Copernicus Data
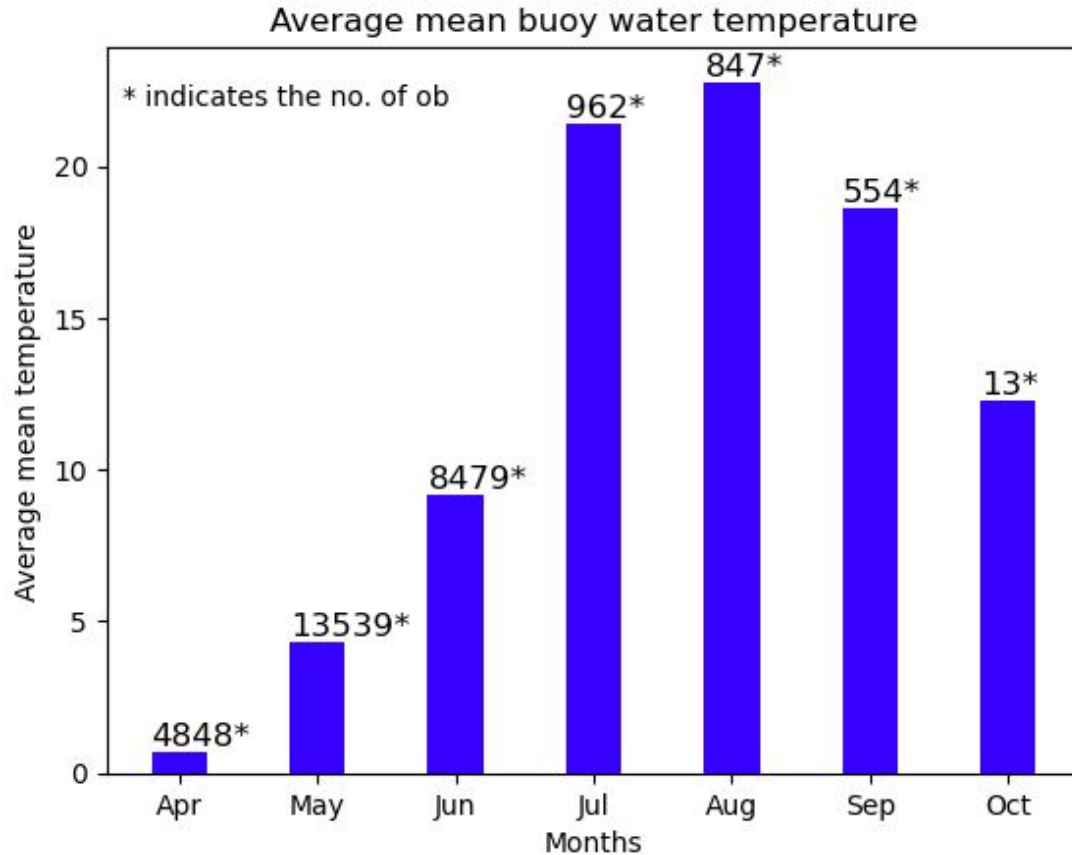

Average mean water temperature from Copernicus data

Visualize Copernicus Data by fishery

Average mean temperature from Copernicus data by fishery

Analyze Buoy Water Temperature

Average mean buoy water temperature

* indicates the no. of ob

4848* 13539* 8479* 962* 847* 554* 13*
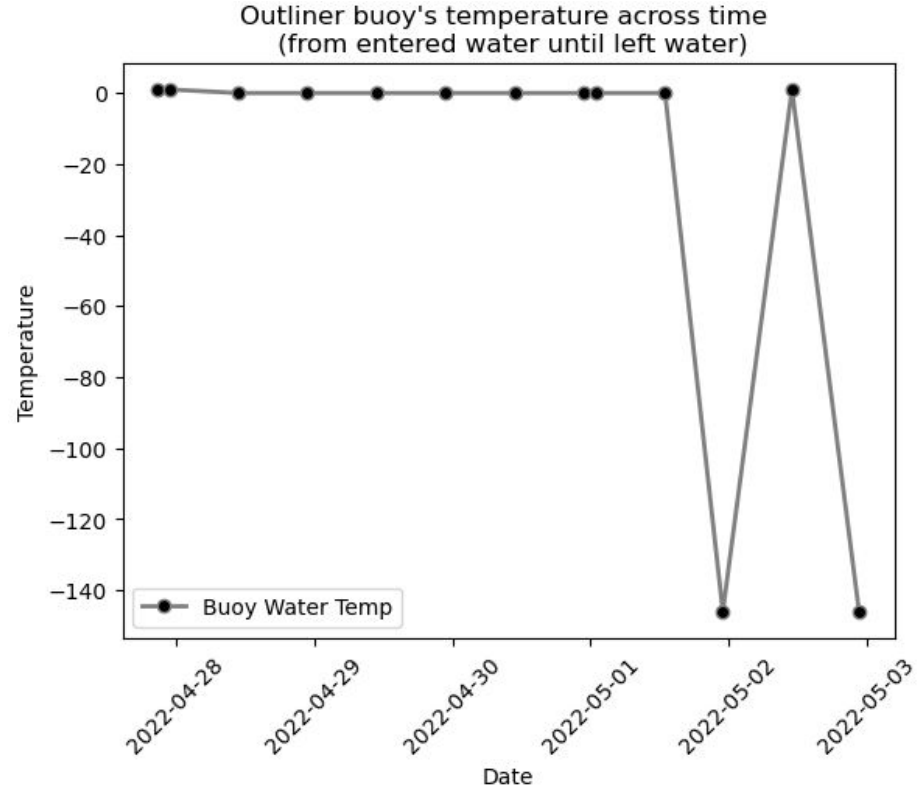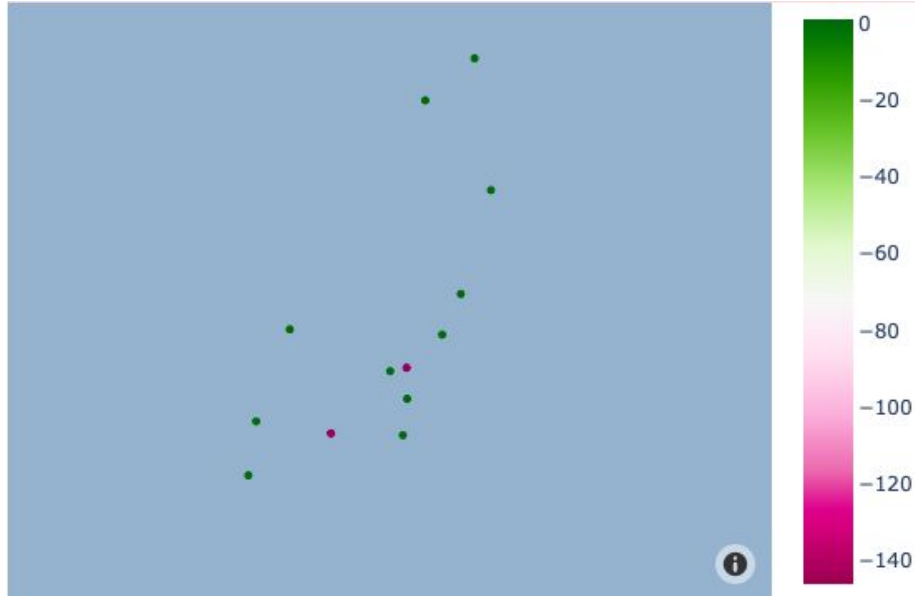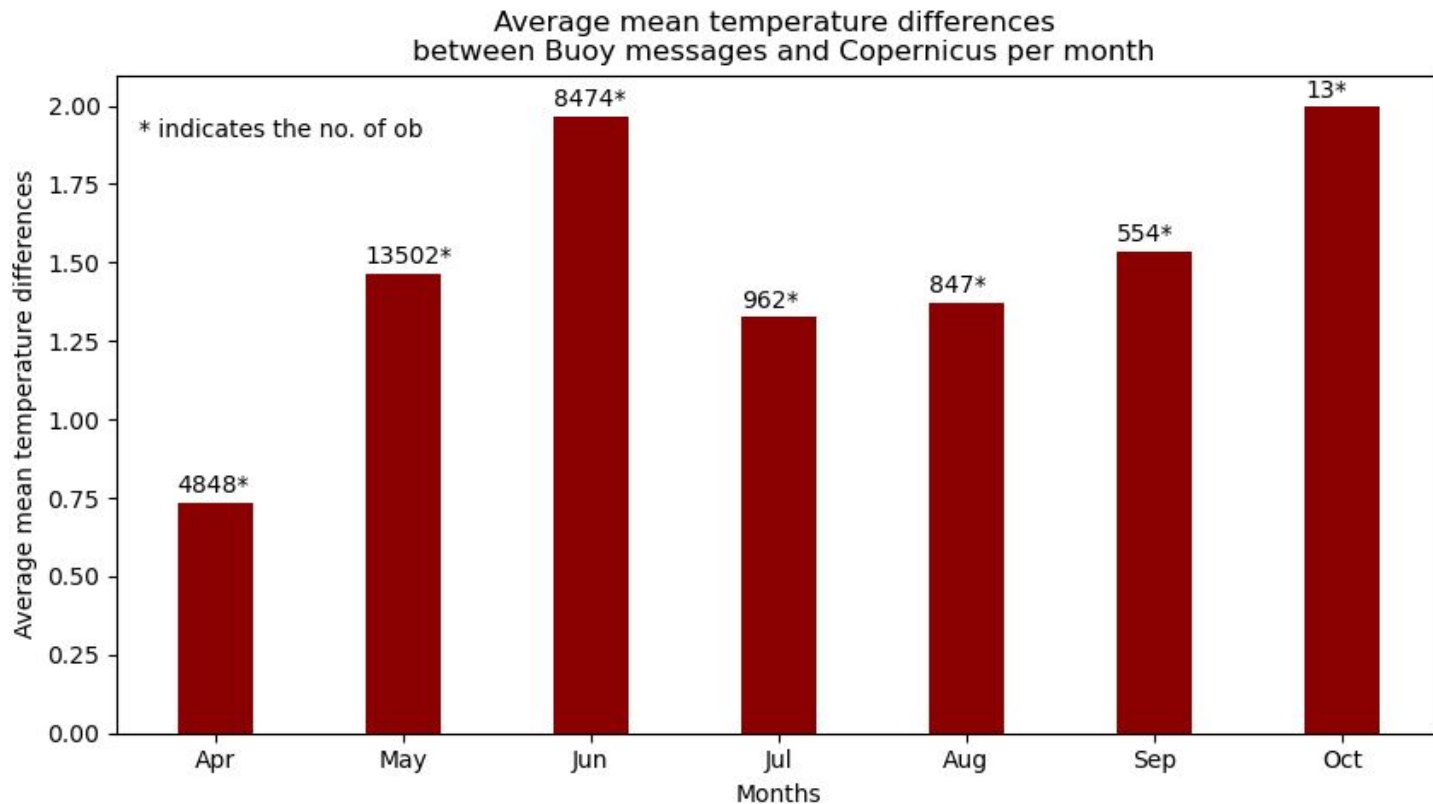Apr   May   Jun   Jul   Aug   Sep   Oct

Months

Average mean temperature

Analyze Buoy Water Temperature - Outlier detection

E.g., Buoy_id = 481

Water_temperature_mean = -146

Outliner buoy's temperature across time
(from entered water until left water)

Compare buoy reported temperatures and Copernicus' reported temperatures across time

Average mean temperature differences between Buoy messages and Copernicus per month
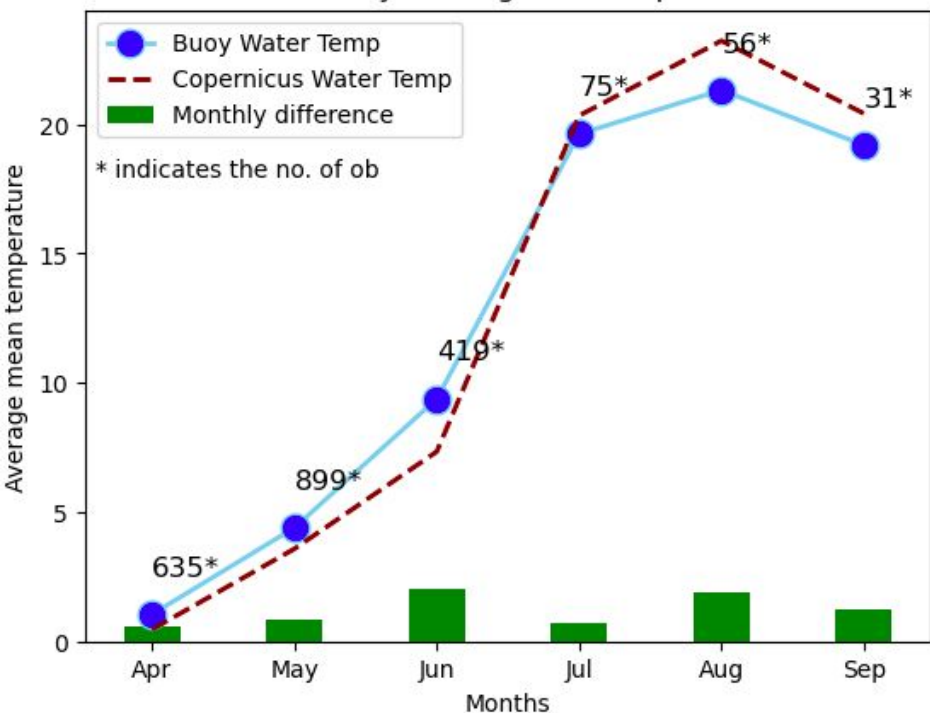
Compare buoy reported temperatures and Copernicus' reported temperatures (by deployment)
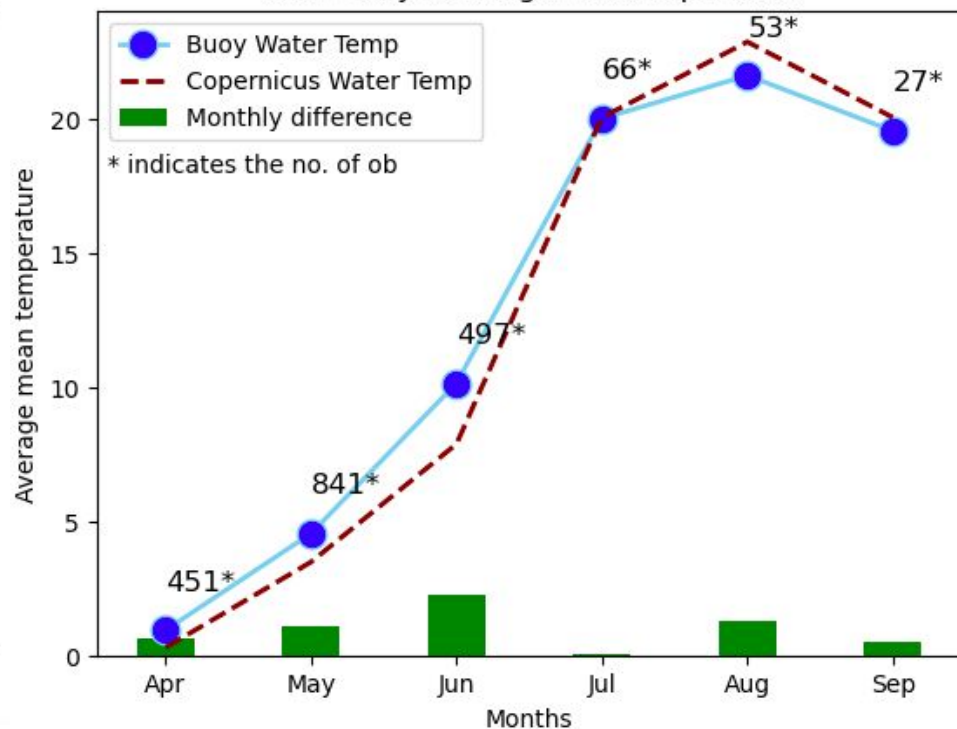
Compare buoy reported temperatures and Copernicus' reported temperatures (by fishery)

Average mean temperature differences between Buoy messages and Copernicus by fishery
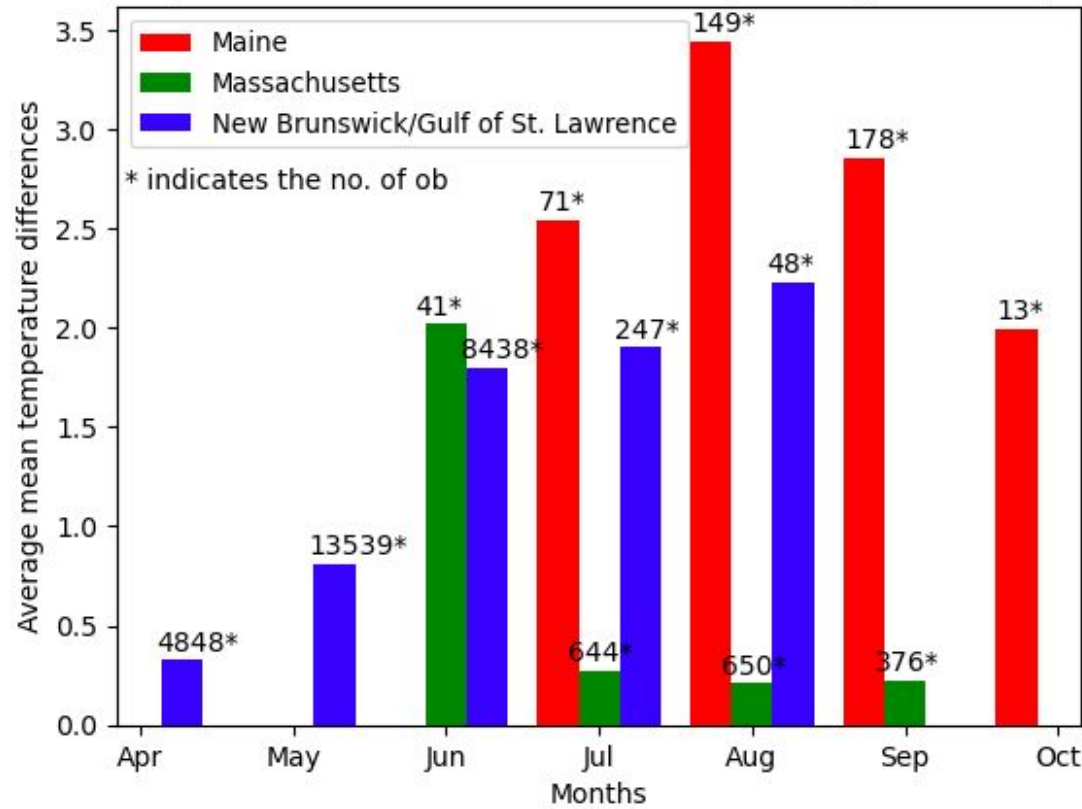
Goals of EDA - Analysis of Canadian Weather Station Data & Buoy Data

We had 4 objectives for this analysis:

1. Retrieve Weather Station Data and Merge with Buoy Data
2. Visualize Weather Station Data
3. Analyze Canadian Water Temperature within each timestep
4. Compare buoy reported temperatures and nearest Canadian Weather Station' reported temperatures

Retrieve Weather Station Data

- Using ERDDAP package to retrieve sea surface temperature data from both [DFO MEDS](#) and [ECCC MSC](#)
- Data source: sea surface temperature data from Realtime data from [Environment and Climate Change Canada buoys](#), [meteorological service of Canada](#) and Blue Ocean Gear buoy data
    - It contains the hourly mean surface fields (e.g., temperature, currents, etc)
    - Key variable: SSTP (sea temperature in given time period)
- Merging data
    - Method: By calculating the distance between the Blue Ocean Gear Data Coordination and the Coordination in Environmental Climate Change Canada, return the nearest station to the buoy in Blue Ocean Gear data with sea temperature
    - Date range to process data: between 2022-09-17 and 2022-09-18 (inclusive)
    - Volume: 31378 observations
    - Variables: buoy coordinates, buoy reported temperature, temperature from the Canadian data

Next steps

- Run significance tests to see if differences between the buoy reported temperatures and Copernicus' reported temperatures are significant

- Run significance tests to see if differences between the buoy reported temperatures and Canadian' reported temperatures are significant

- Segment the timestep in given time slot

Appendix:
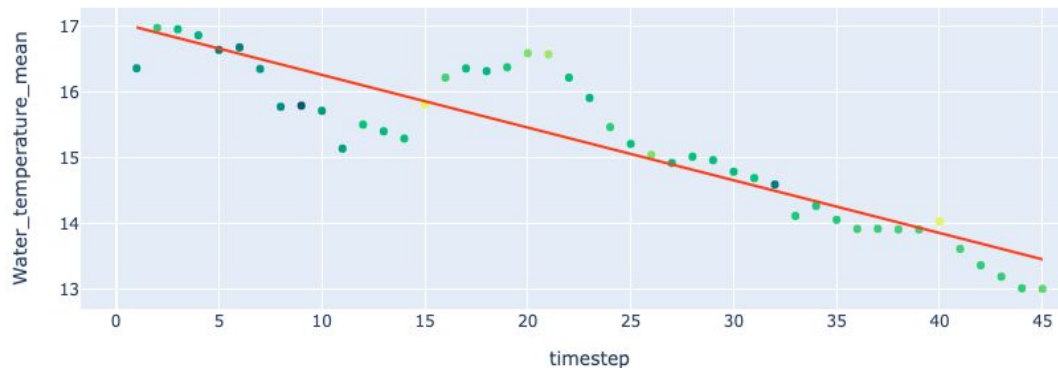# Canadian Sea Surface Temperature Change over Time

Appendix:
Meteorological Service of Canada Sea Surface Temperature Change over Time
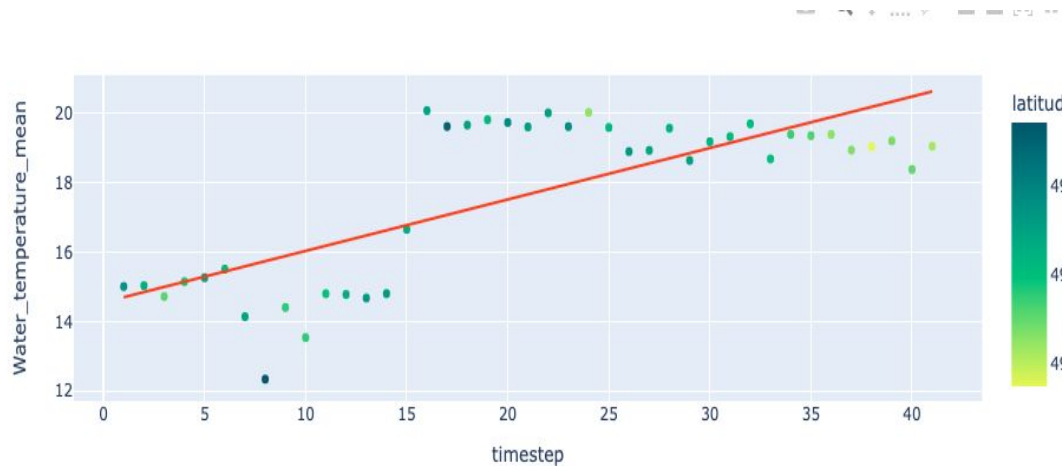
OLS Regression Results

Dep. Variable:	y	R-squared:	0.817
Model:	OLS	Adj. R-squared:	0.813
Method:	Least Squares	F-statistic:	192.5
Date:	Mon, 31 Oct 2022	Prob (F-statistic):	1.76e-17
Time:	17:17:03	Log-Likelihood:	-31.862
No. Observations:	45	AIC:	67.72
Df Residuals:	43	BIC:	71.34
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	17.0606	0.152	111.983	0.000	16.753	17.368
x1	-0.0800	0.006	-13.875	0.000	-0.092	-0.068

Omnibus:	2.169	Durbin-Watson:	0.261
Prob(Omnibus):	0.338	Jarque-Bera (JB):	1.677
Skew:	0.473	Prob(JB):	0.432
Kurtosis:	2.994	Cond. No.	53.8

Appendix:
Visualize the Real time temperature data from Environment and Climate Change Canada buoys

Appendix:
Visualize the Realtime data from Environment and Climate Change Canada buoys (Meteorological Service of Canada)

Appendix:
Merge Buoy Data and Station Data

- calculate each message's nearest weather station and record what the temperature is

[40]:

	datetime	lat	lon	buoy_water_temperature_mean	station_water_temperature_mean
0	2021-03-25 18:01:08	37.471260	-121.940110	23.0	16.3
1	2021-03-25 18:05:25	37.471252	-121.940216	23.0	16.3
2	2021-03-26 00:15:04	37.471207	-121.940300	23.0	16.3
3	2021-03-26 00:19:13	37.471207	-121.940300	24.0	16.3
4	2021-04-01 17:43:23	37.471222	-121.940600	25.0	16.3