



Universidad
de Navarra

Hercul
Wing

Analyzing Ocean Dynamics

By Miguel Díaz P.d.J and Dante Schrantz

Note: All the videos included in this presentation were created by Miguel Díaz for his project Herculwing, which you can explore [here](#). © All rights reserved. Unauthorized use, reproduction, or distribution of this content is strictly prohibited.

Our Own Database



The data used in this project comes from a buoy located in the sea, which has been collecting measurements over the past year. This buoy recorded various environmental and oceanographic data.



GOBIERNO
DE ESPAÑA

MINISTERIO
DE TRANSPORTES, MOVILIDAD
Y AGENDA URBANA

Puertos del Estado



A Coruña



Lon:

-8.383

Lat:

43.413

Google Cloud My First Project Search (

Explorer + ADD K

Search BigQuery resources ?

Show starred only

Queries Shared queries

- Updating Currents Table Format
- Updating Salinity Table Format
- Updating WaterTemp Table format + variables
- Updating Wave Table Format + variables
- Updating Wind Table Format

Notebooks Data canvases Data preparations Workflows External connections

boyacoruna Models (1)

- automl_model
- BoyaAll
- Currents
- Salinity
- WaterTemp

Tables Formatting

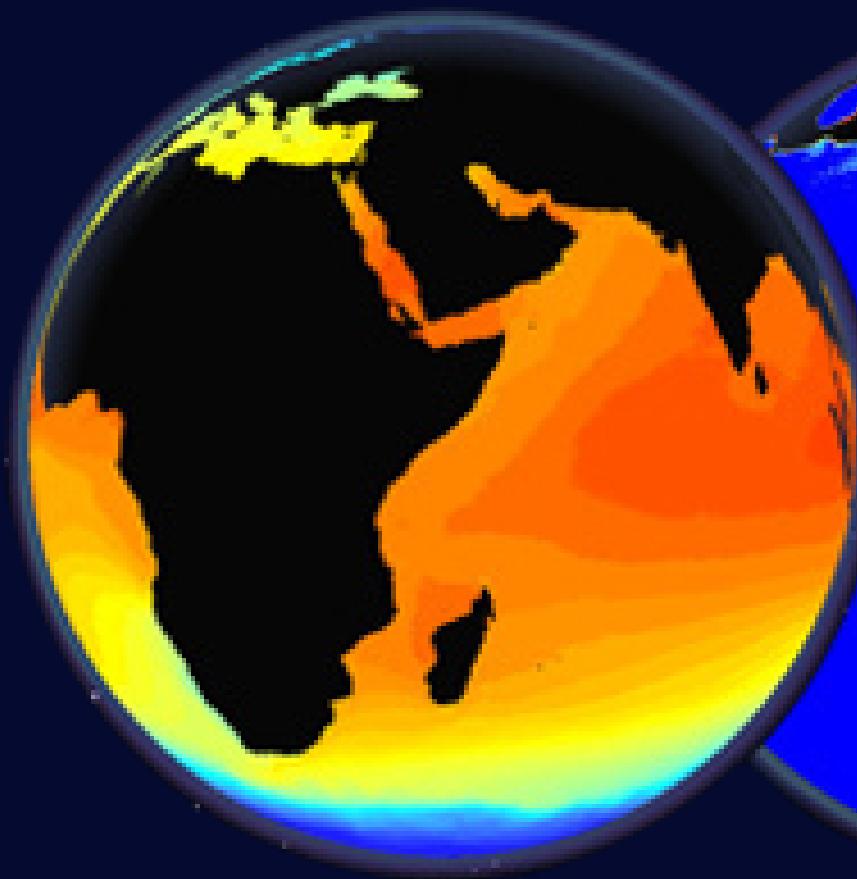
Updating... les X + EJECUTAR

Updating Wave Table Format + variables

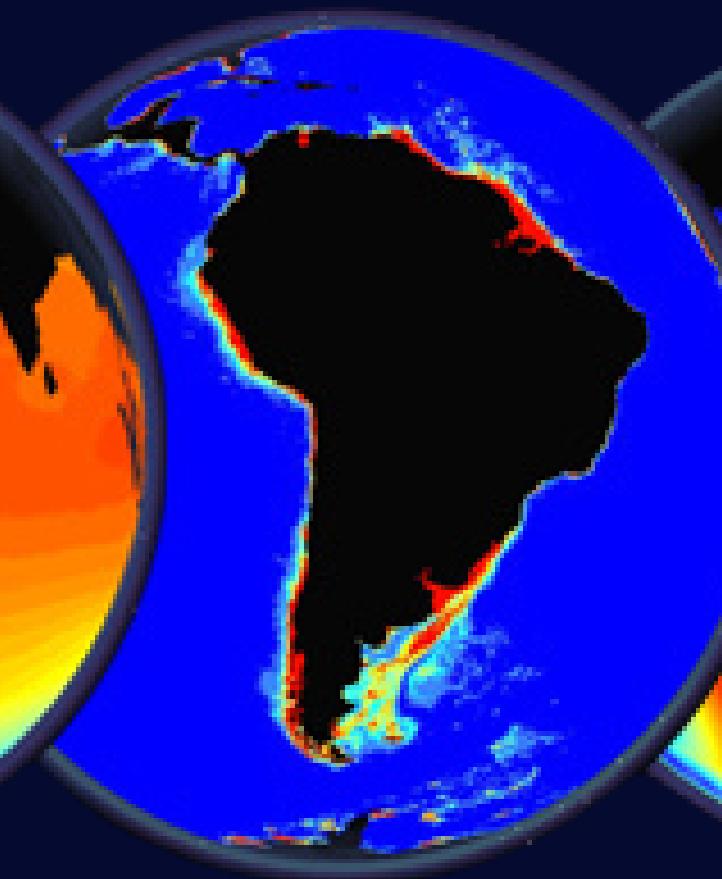
```
1 CREATE OR REPLACE TABLE `lofty-gravity-420212.boyacoruna.Wave` AS
2 WITH tabla_limpia AS (
3     SELECT
4         string_field_0 AS date_gmt,
5         SAFE_CAST(string_field_1 AS FLOAT64) AS significant_wave_height_m,
6         SAFE_CAST(string_field_2 AS FLOAT64) AS mean_period_s,
7         SAFE_CAST(string_field_3 AS FLOAT64) AS peak_period_s,
8         SAFE_CAST(string_field_4 AS INT64) AS mean_wave_direction
9     FROM `lofty-gravity-420212.boyacoruna.Wave`
10    WHERE string_field_0 NOT IN ('Valor nulo: -9999.9', 'Fecha (GMT)')
11 )
12     SELECT *
13     FROM tabla_limpia
14    WHERE significant_wave_height_m IS NOT NULL;
```

Relational Database

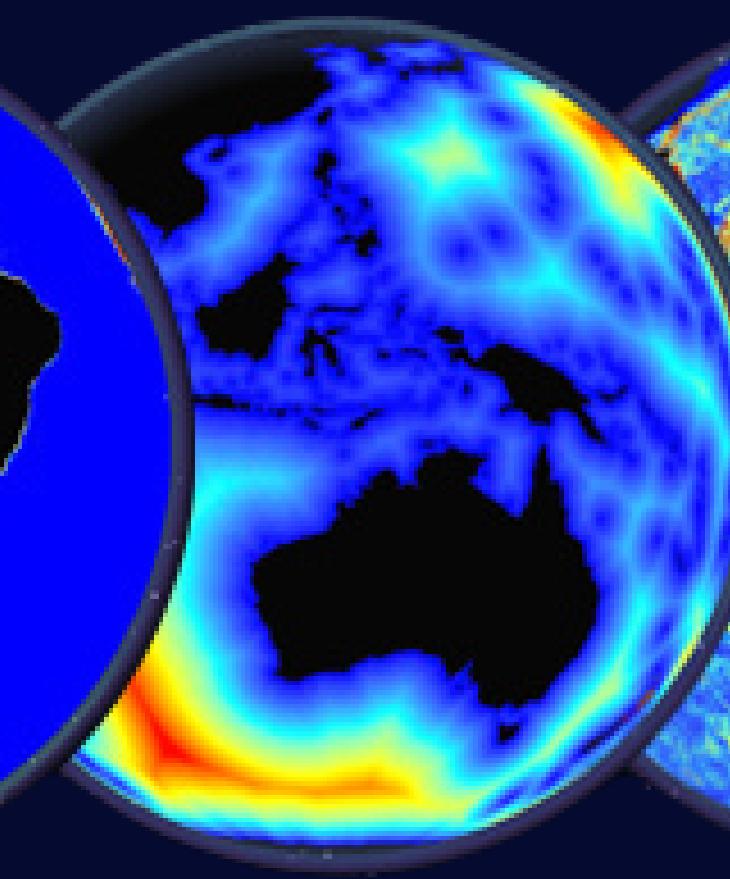
<https://gmed.auckland.ac.nz/index.html>



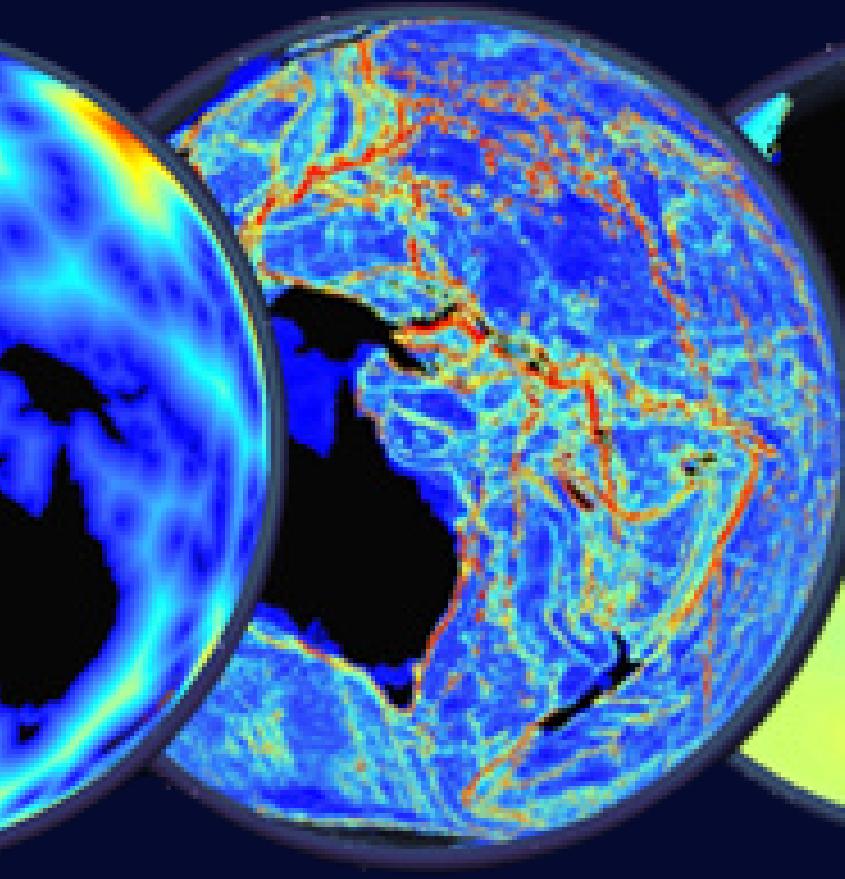
Temperature



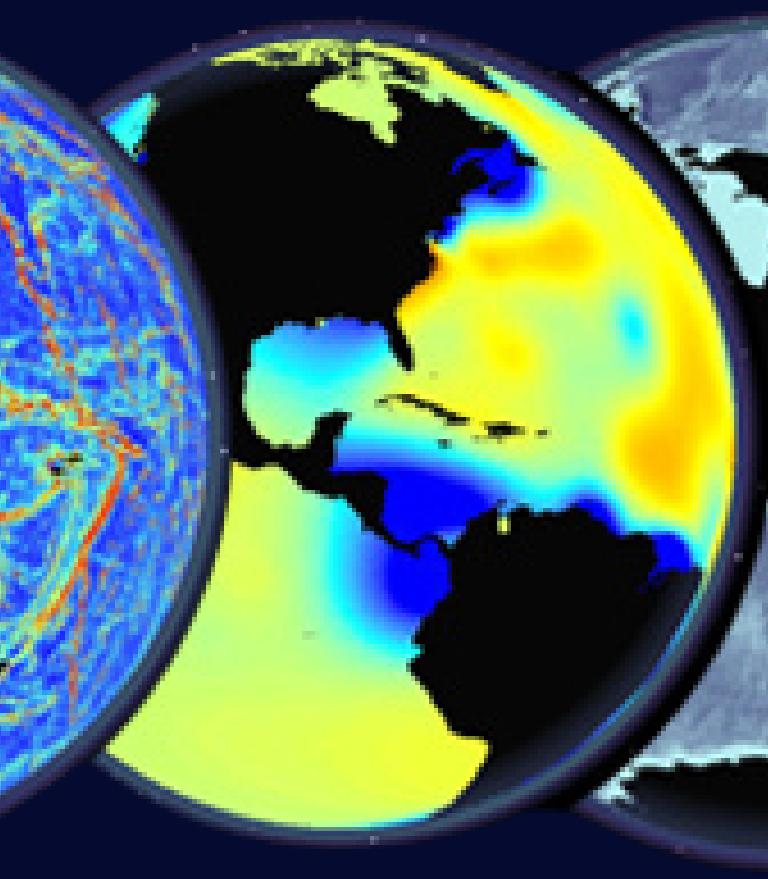
Waves



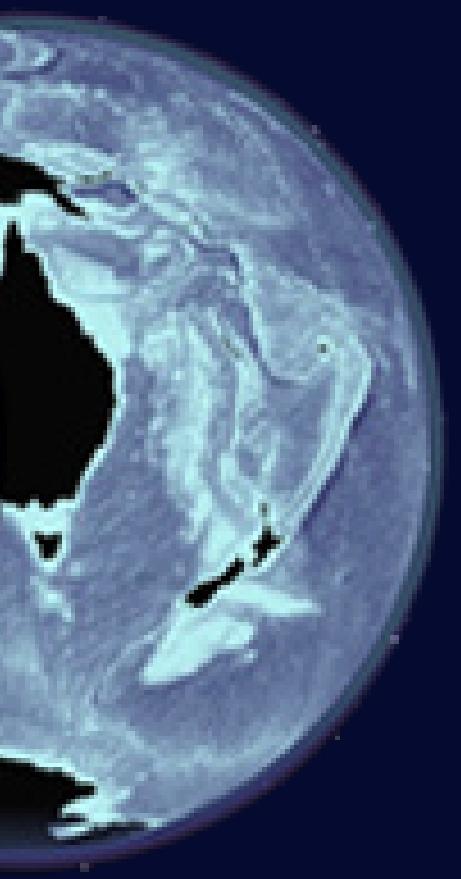
Currents

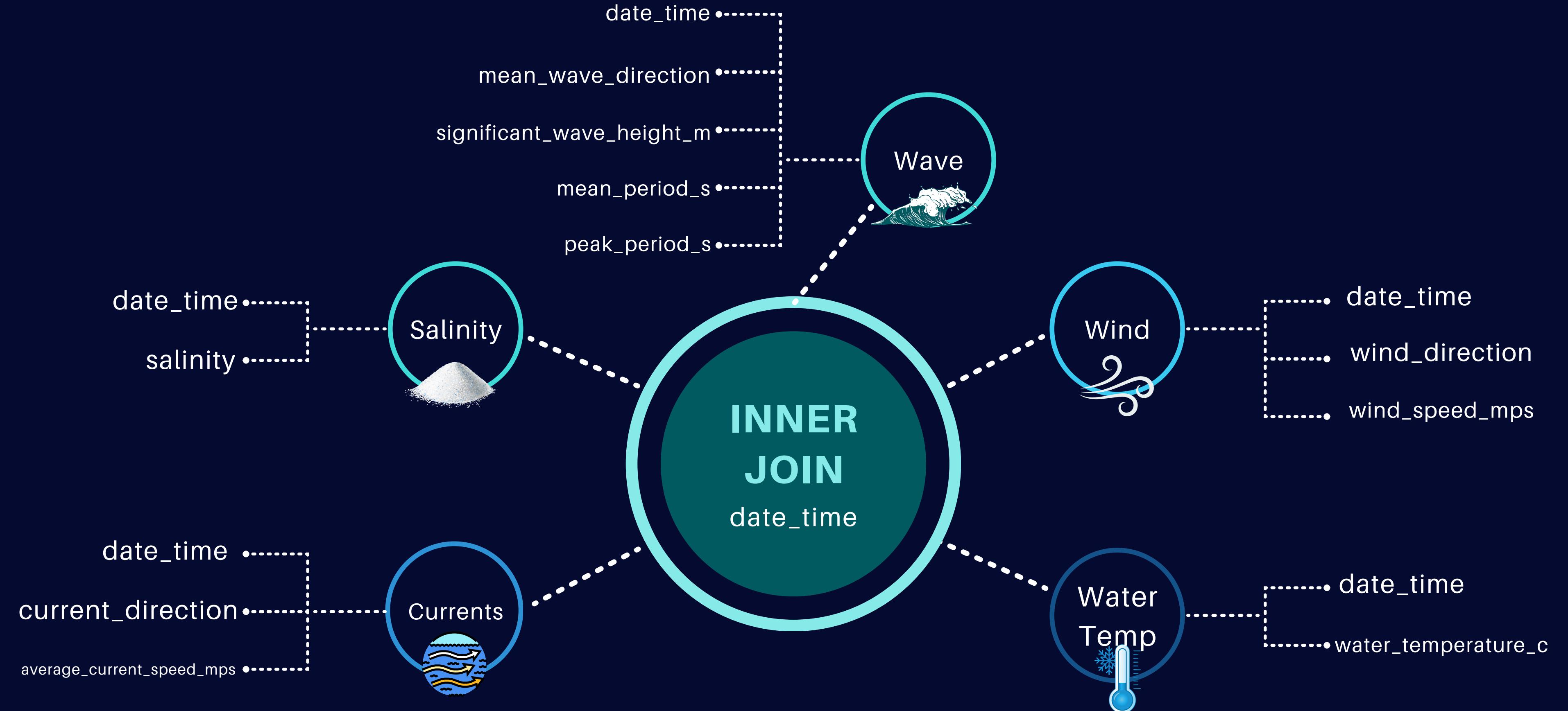


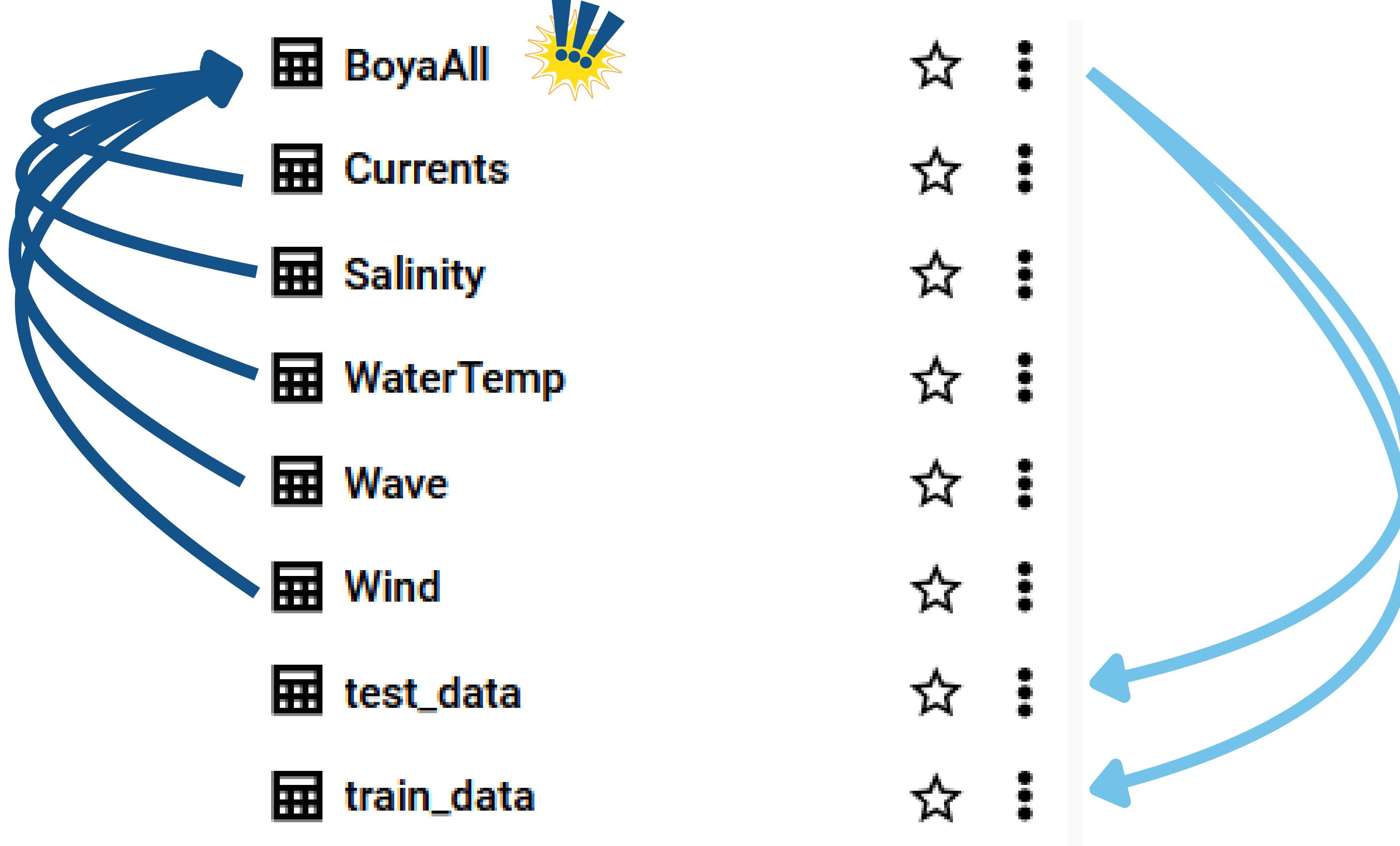
Wind



Salinity

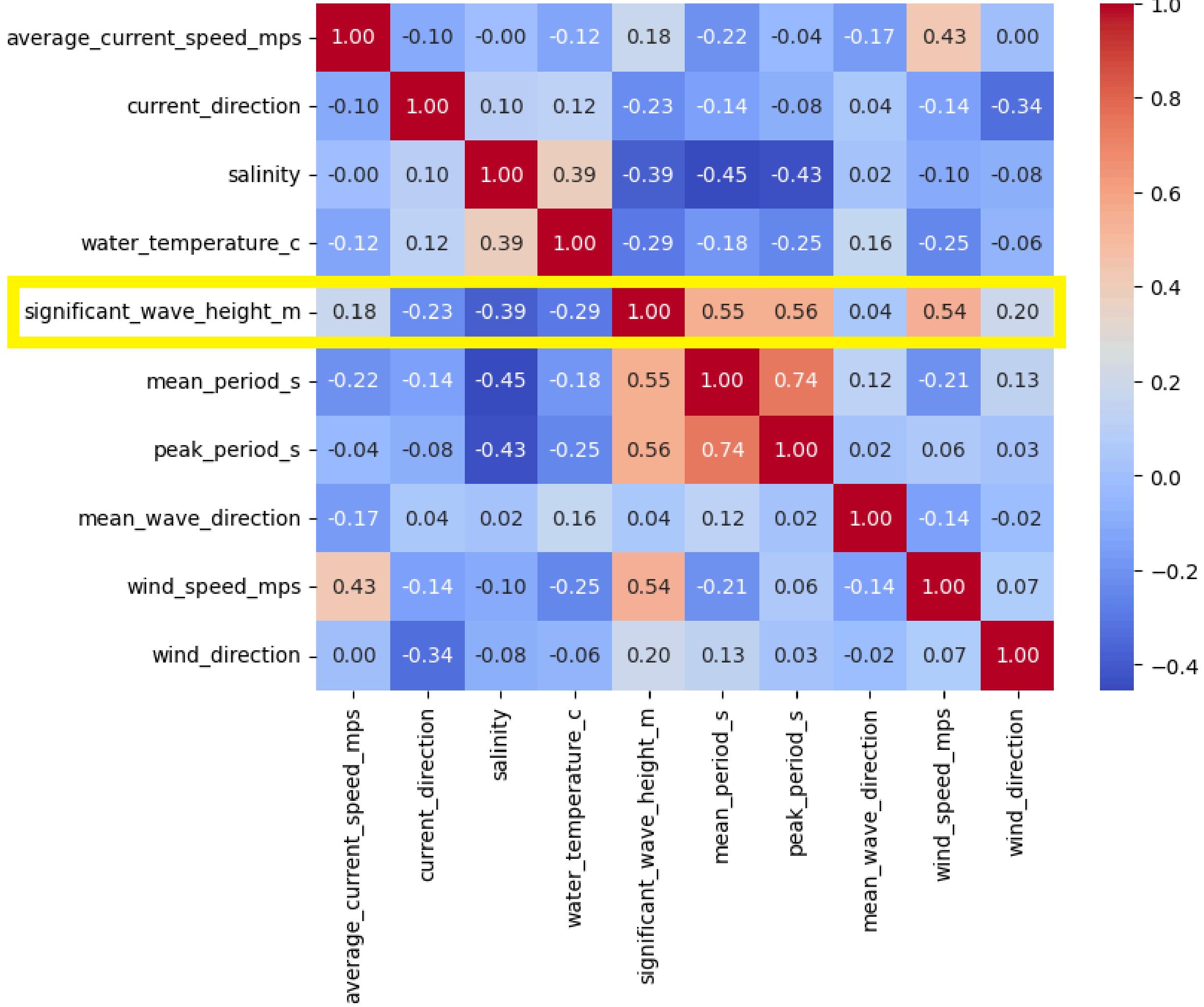






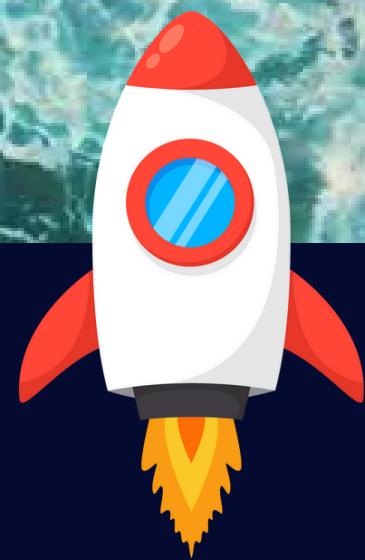


Correlation Matrix

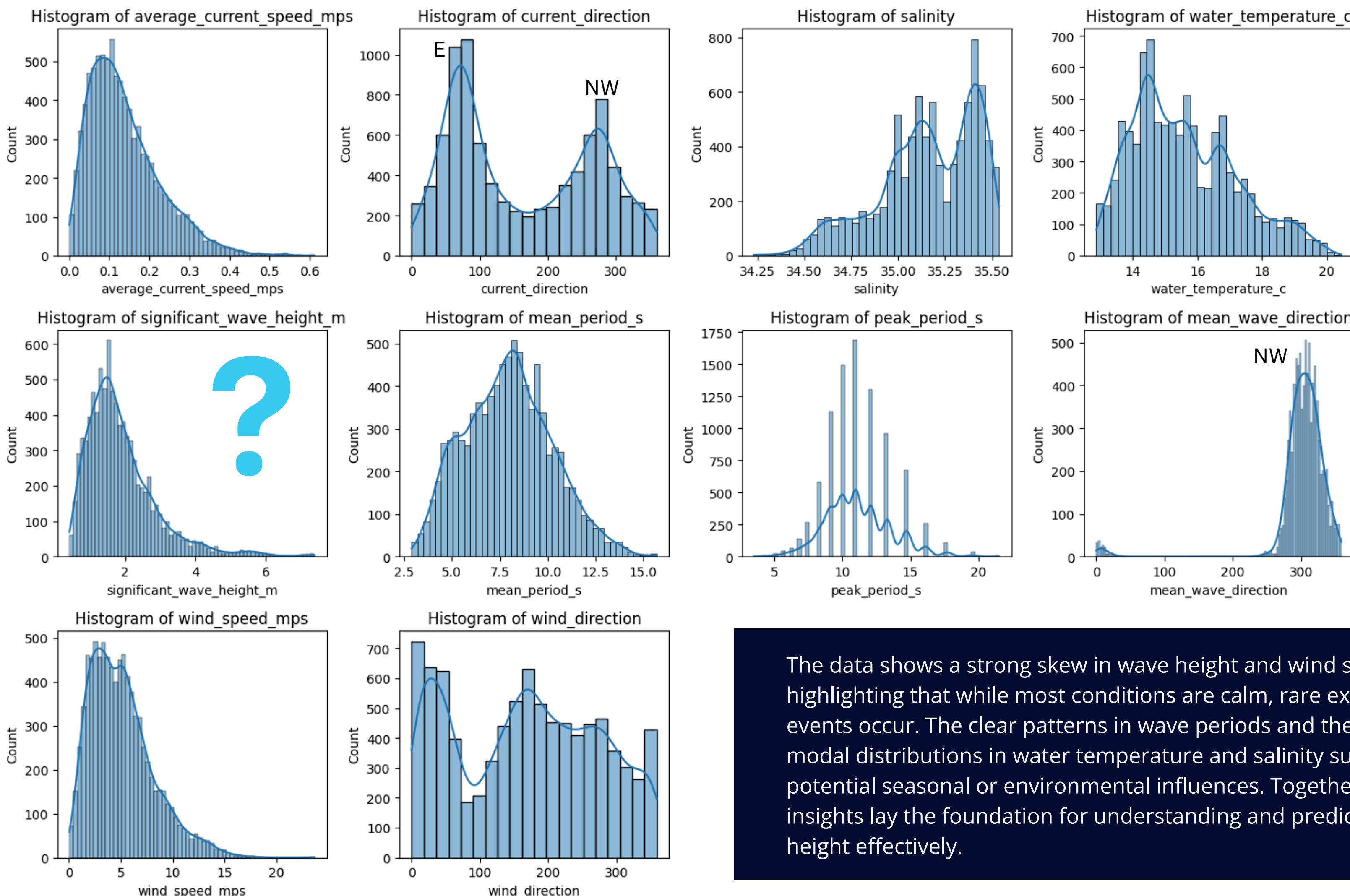


Goal

The goal of this project is to predict **significant wave height (m)** based on a variety of environmental and oceanographic features.



Supervised Regression



The data shows a strong skew in wave height and wind speed, highlighting that while most conditions are calm, rare extreme events occur. The clear patterns in wave periods and the multimodal distributions in water temperature and salinity suggest potential seasonal or environmental influences. Together, these insights lay the foundation for understanding and predicting wave height effectively.

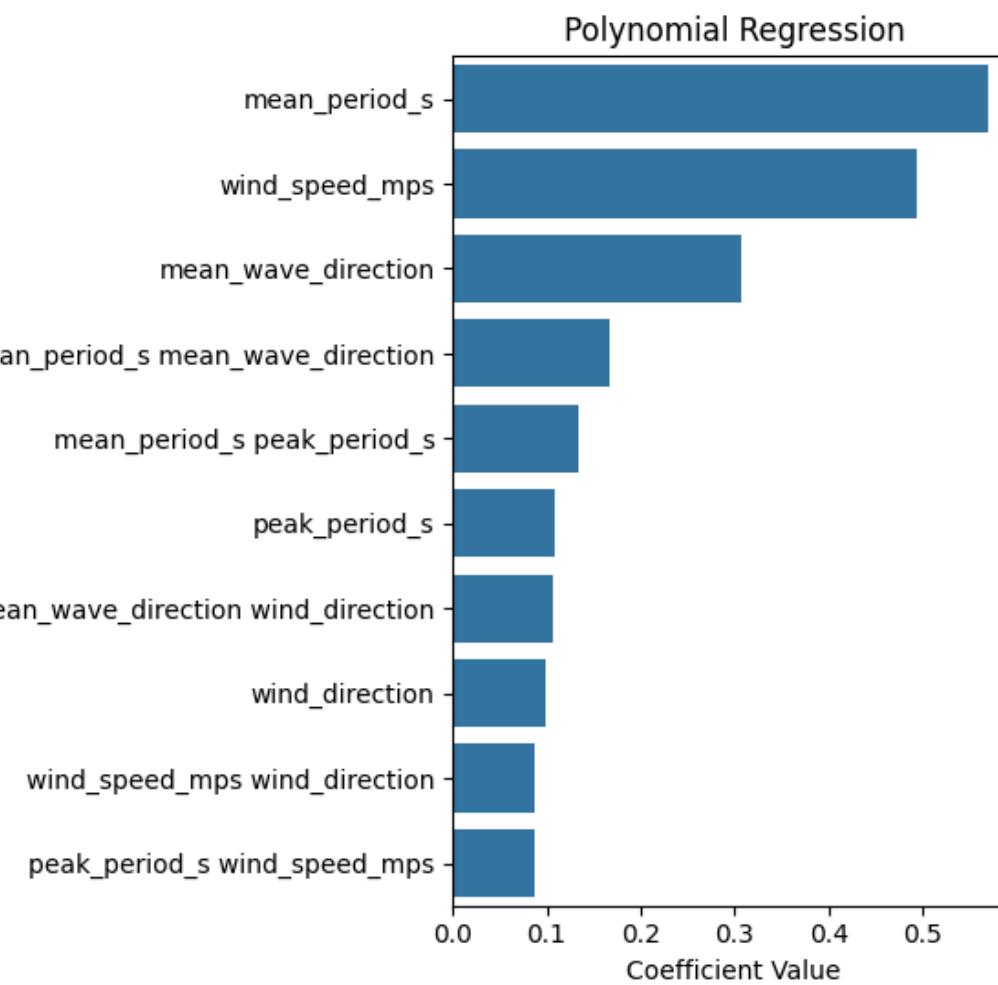
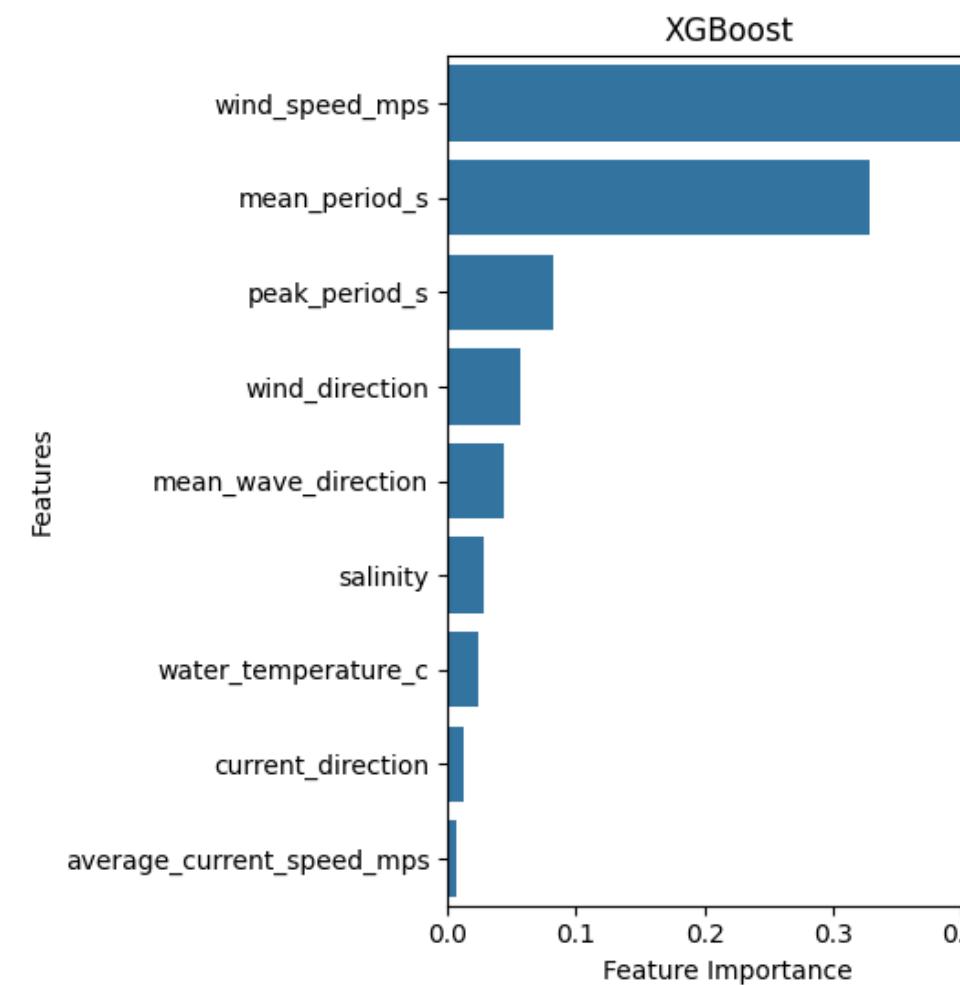
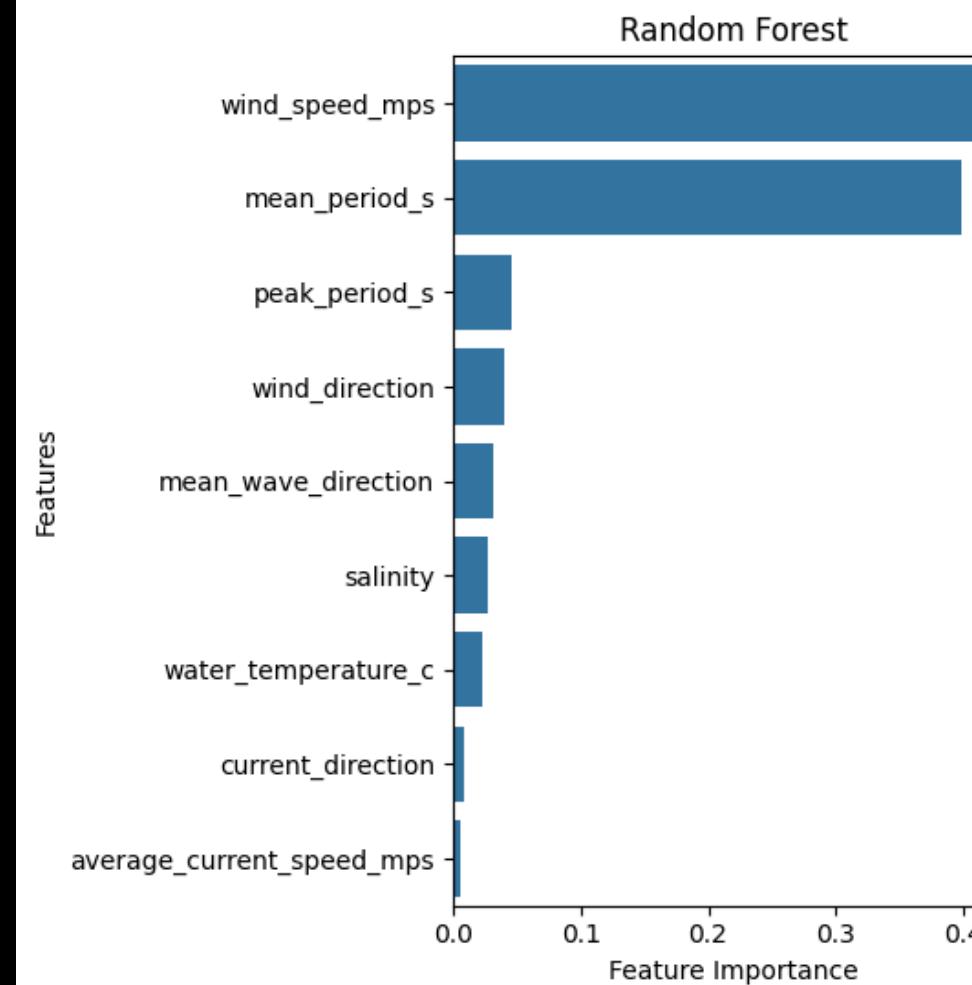
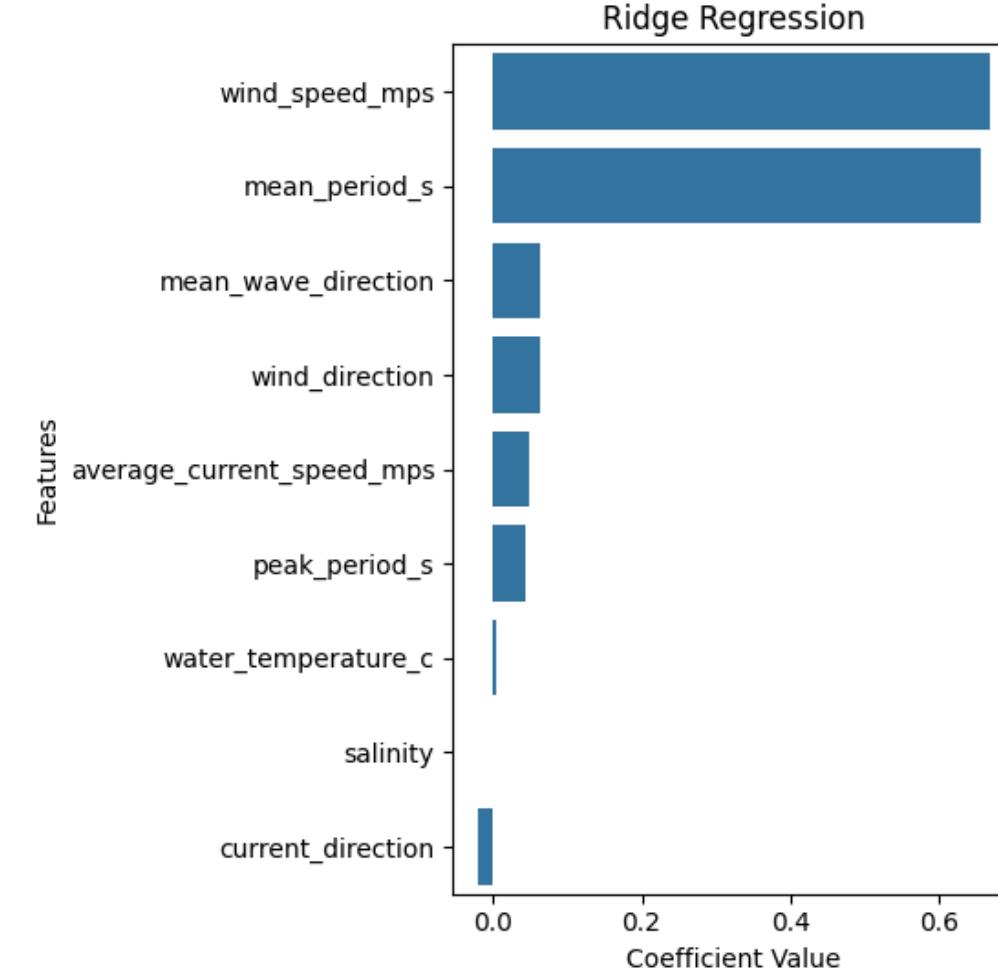
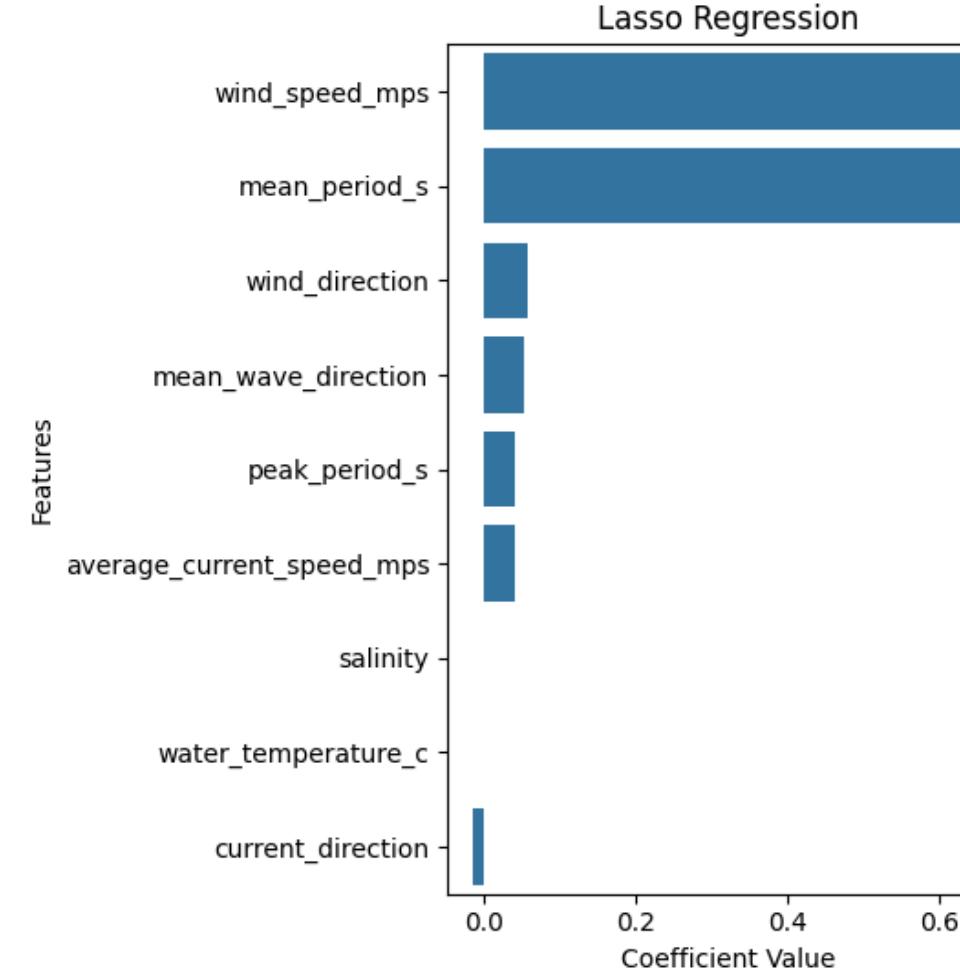
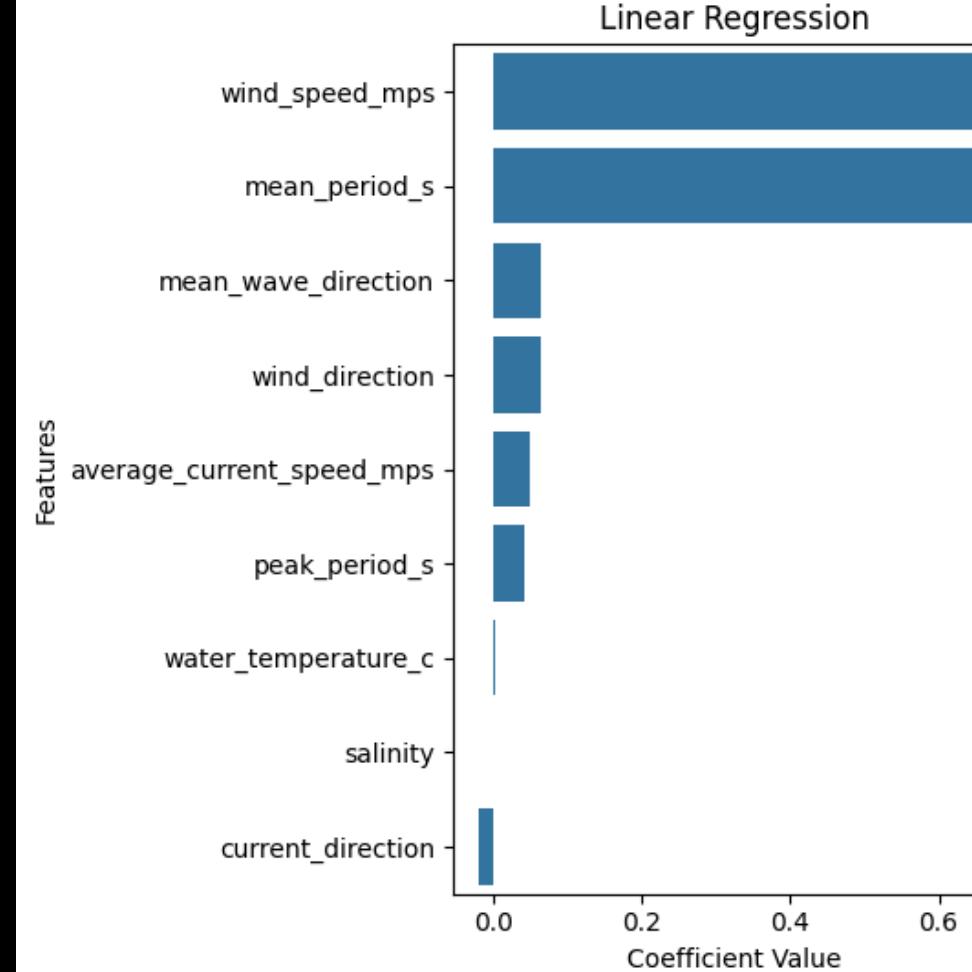
Predictors

BigQuery table
visualization

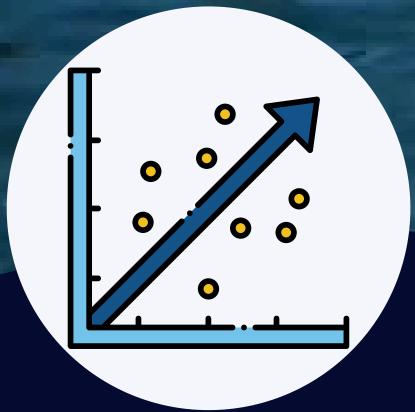
Predictors

date_time	average_current	current_direction	salinity	water_temperature	significant_wave	mean_period_s	peak_period_s	mean_wave_dir	wind_speed_mph	wind_direction
2024-04-23T12:00:00	0.2	236.3	35.432	13.644	0.82	3.2	3.5	352	7.19	7.0
2024-09-23T21:00:00	0.026	42.13	35.387	15.912	0.81	4.0	7.5	291	7.65	202.0
2024-06-12T13:00:00	0.121	264.55	35.534	14.483	1.14	4.0	7.5	352	9.06	3.0
2024-06-12T14:00:00	0.138	230.9	35.536	14.614	1.24	4.0	7.5	353	10.58	5.0
2024-06-12T15:00:00	0.189	216.75	35.533	14.43	1.29	4.0	7.5	354	11.12	7.0
2024-09-25T14:00:00	0.366	73.7	35.309	15.268	1.54	4.0	7.5	259	11.34	213.0
2024-09-16T21:00:00	0.168	296.5	35.371	16.311	0.88	4.0	7.5	358	5.21	51.0
2024-09-16T20:00:00	0.277	286.59	35.373	16.787	0.88	4.0	7.5	357	5.98	49.0
2024-09-16T19:00:00	0.385	278.78	35.353	16.968	0.88	4.0	7.5	355	5.64	43.0
2024-06-01T19:00:00	0.264	272.95	35.373	14.106	1.49	4.5	7.5	10	9.33	53.0
2024-07-14T09:00:00	0.138	244.39	35.399	17.3	0.46	4.5	7.5	308	5.31	173.0
2024-07-14T08:00:00	0.158	200.7	35.398	17.309	0.48	4.5	7.5	309	4.48	181.0
2024-05-20T21:00:00	0.041	195.73	34.816	15.948	0.8	4.5	7.5	327	3.77	326.0
2024-06-01T21:00:00	0.235	279.39	35.391	13.858	1.39	4.5	7.5	9	7.59	54.0
2024-05-31T19:00:00	0.296	269.51	35.246	15.259	1.47	4.5	7.5	6	9.07	52.0
2024-06-01T18:00:00	0.293	272.3	35.37	14.176	1.53	4.5	7.5	10	9.39	55.0
2024-06-02T20:00:00	0.166	287.67	35.442	13.709	1.34	4.5	7.5	8	7.21	54.0
2024-06-01T22:00:00	0.242	282.62	35.394	13.812	1.34	4.5	7.5	8	6.99	56.0

Feature Importance for All Models



Models

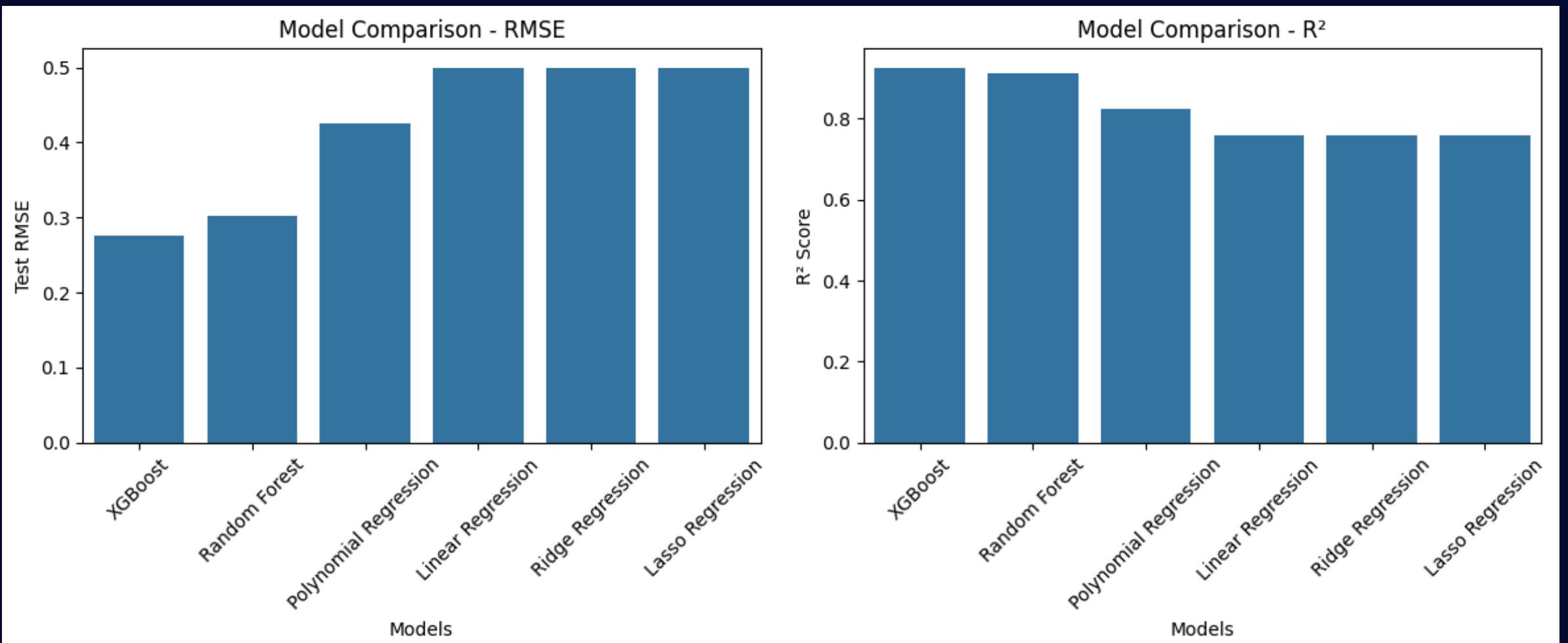


$$ax^2 + bx + c$$



Model Performance Comparison:

	Model	Train RMSE	Test RMSE	R^2
0	Linear Regression	0.506491	0.499066	0.758828
1	Ridge Regression	0.506491	0.499074	0.758820
2	Lasso Regression	0.507018	0.500151	0.757778
3	Random Forest	0.213899	0.302263	0.911533
4	XGBoost	0.189743	0.275249	0.926639
5	Polynomial Regression	N/A	0.424685	0.825359



XGBoost performs best with the lowest Test RMSE (0.275) and highest R² (0.93).

Random Forest is the second-best, with solid performance.

Linear, Ridge, and Lasso Regression perform similarly (R² ~0.76), showing limitations with nonlinearity.



Explorer + ADD <

Search BigQuery resources ?

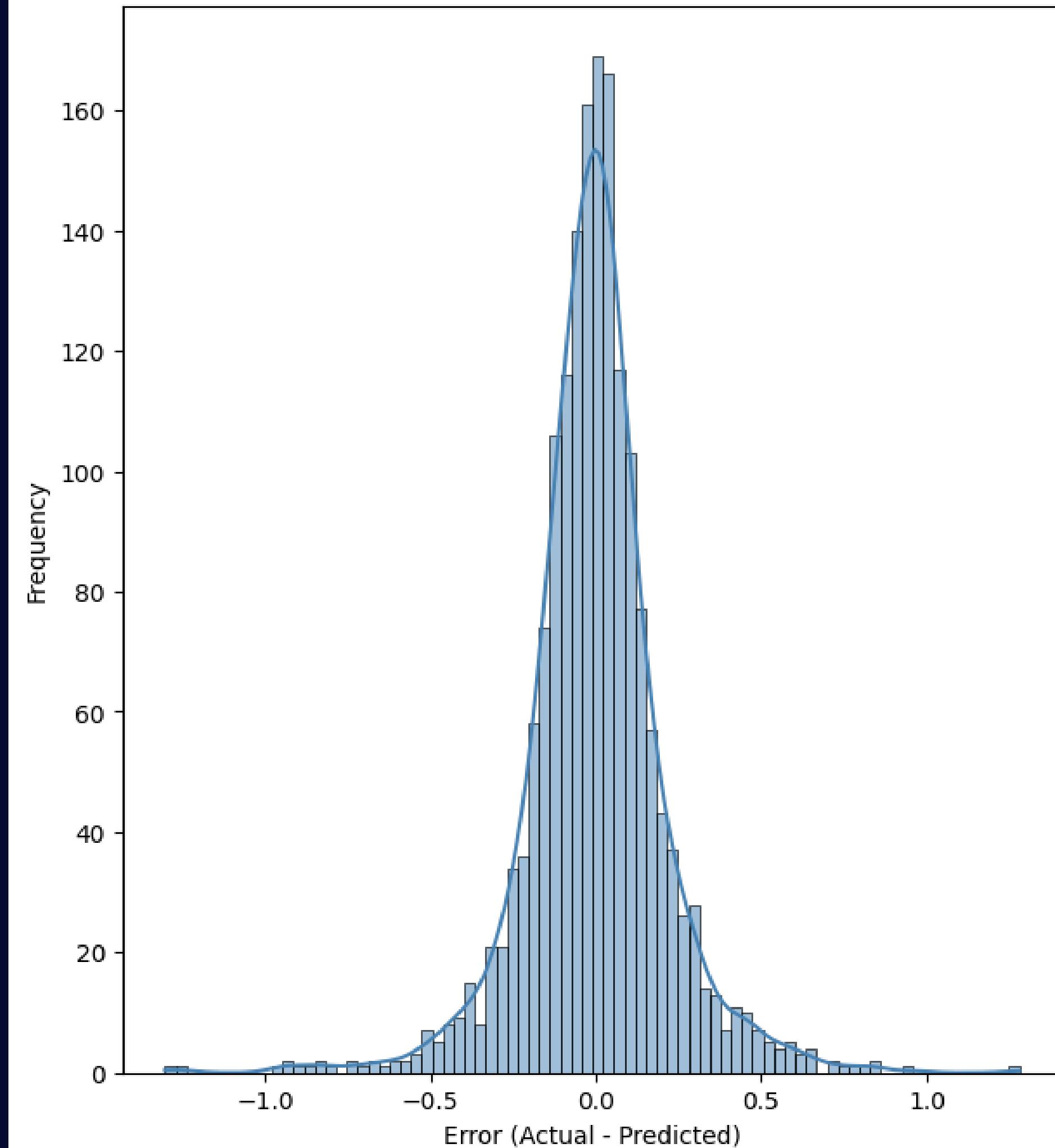
Show starred only

- ▼ Queries ⋮
 - ▼ Shared queries ⋮
 - Updating Currents Table Format
 - Updating Salinity Table Format
 - Updating WaterTemp Table format + variables
 - Updating Wave Table Format + variables
 - Updating Wind Table Format
 - ▶ Notebooks ⋮
 - ▶ Data canvases ⋮
 - ▶ Data preparations ⋮
 - ▶ Workflows ⋮
 - ▶ External connections ⋮
- ▼ boya_acoruna ⋮
 - ▼ Models (1) ⋮
 - automl_model ⋮
 - BoyaAll
 - Currents
 - Salinity
 - WaterTemp

Improving XGBOOST

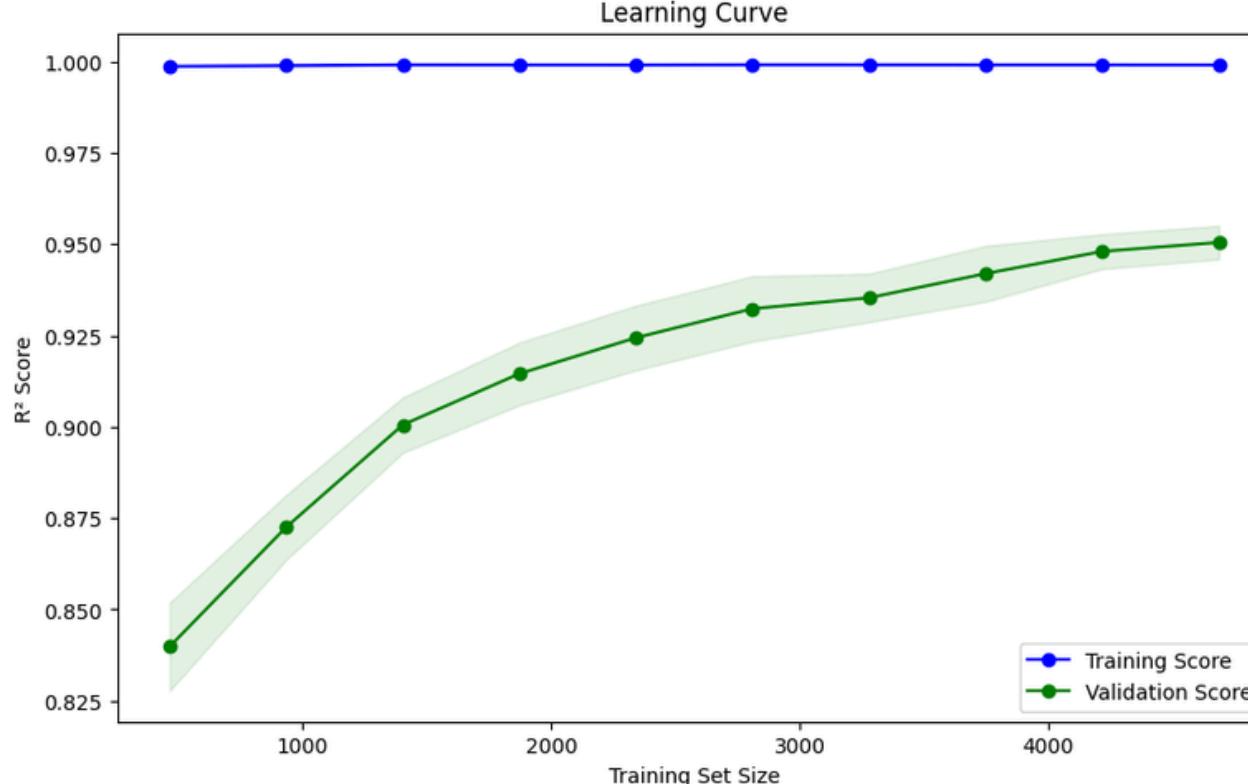
Time to improve our
highest performing model

Distribution of Prediction Errors



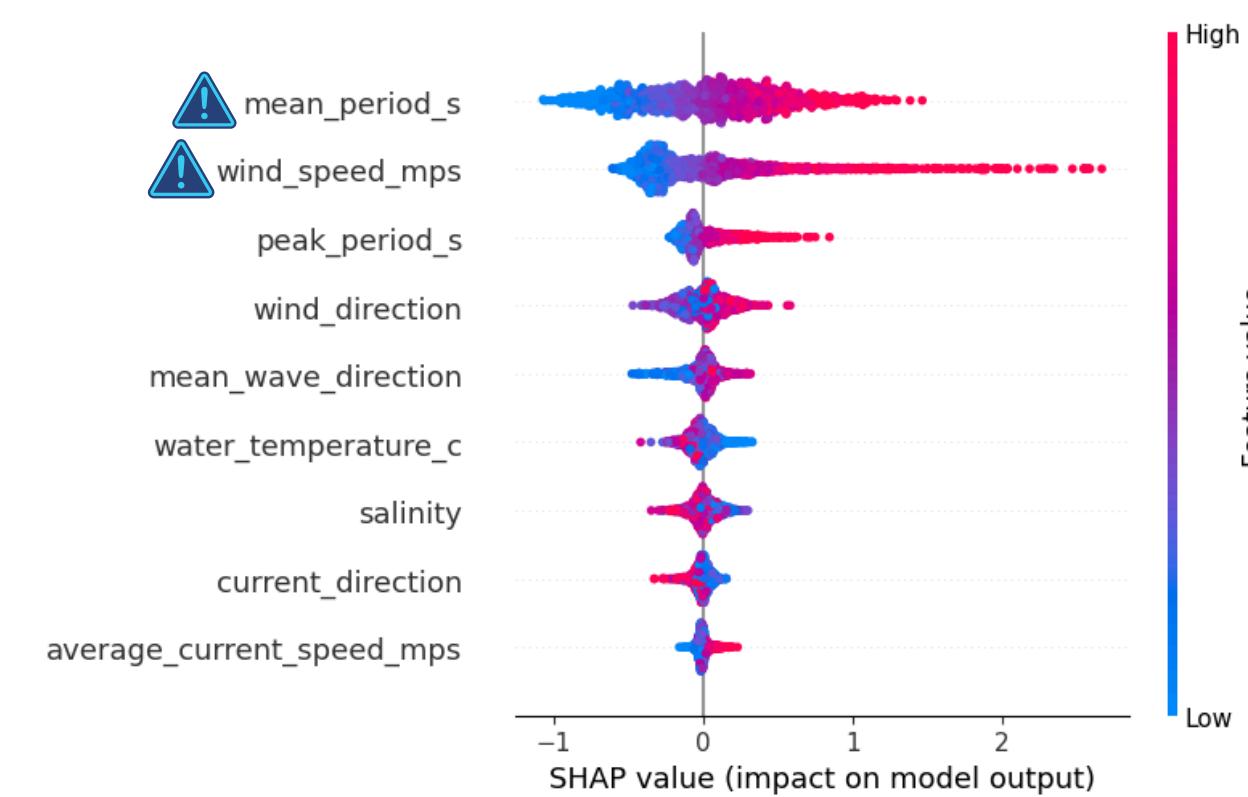
PERFORMANCE

How well did our final XGBoost model perform?

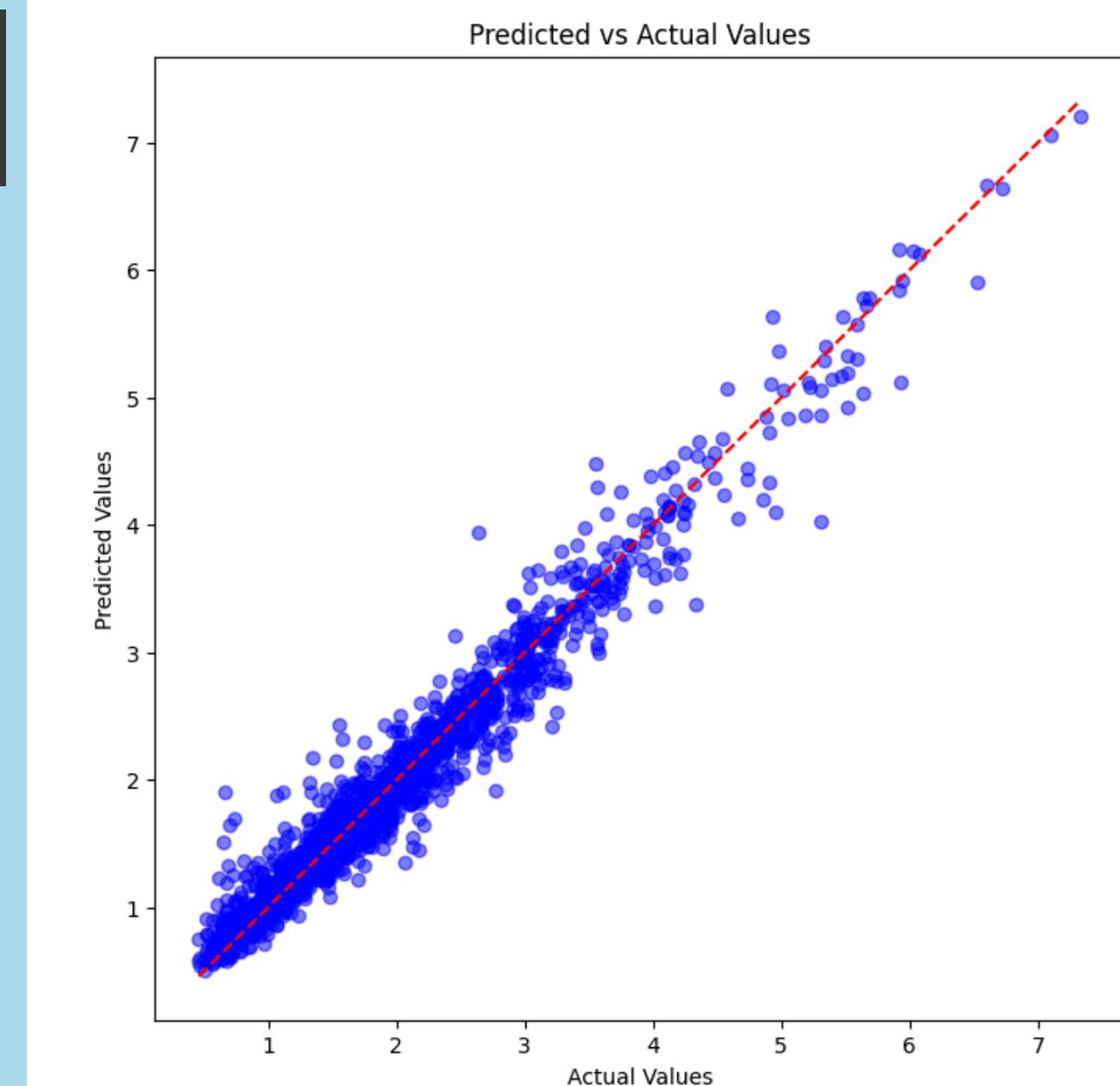
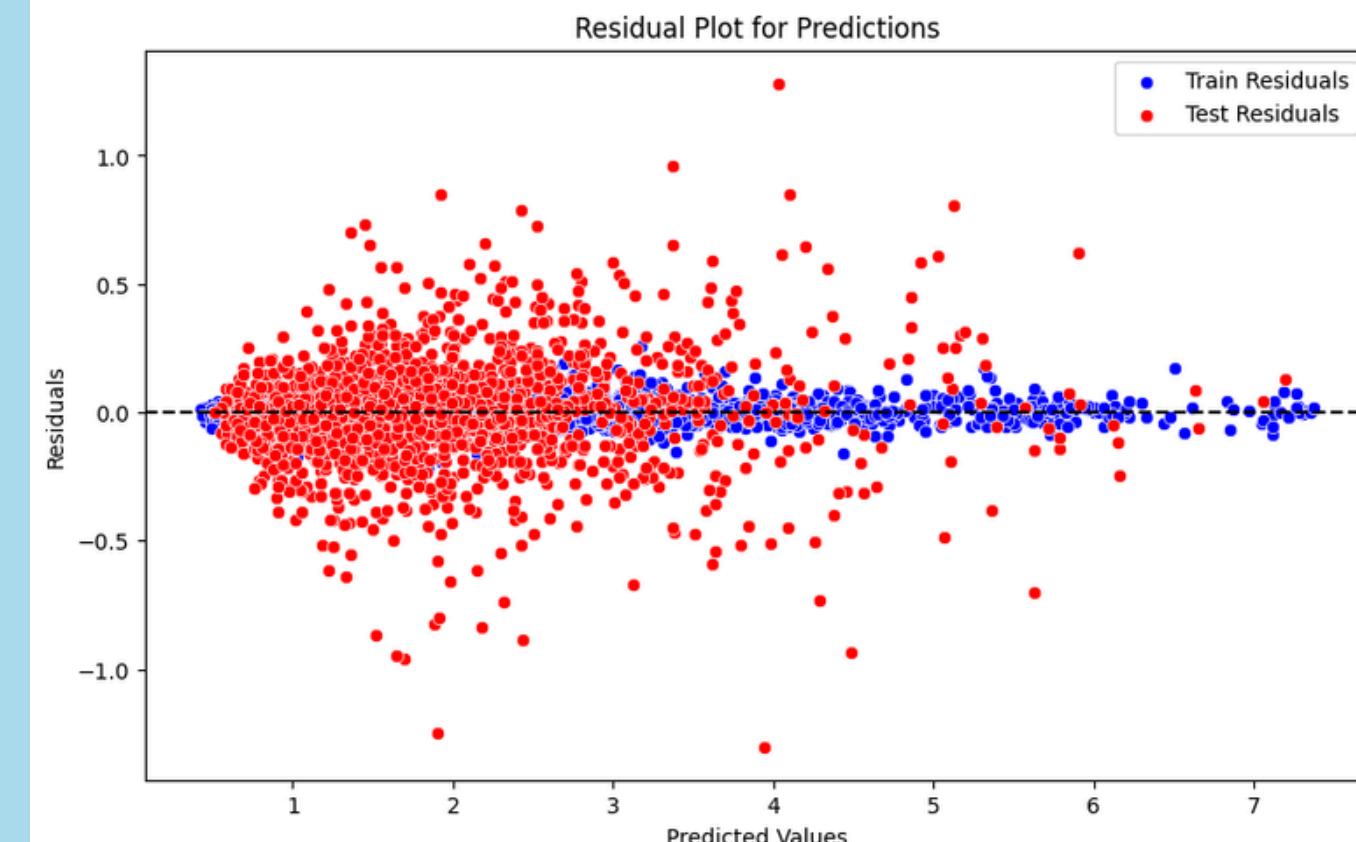


Training Final XGBoost Model...

Final XGBoost Results:
Train RMSE: 3.6738, Test RMSE: 14.3574, R²: 0.9528



Very good results in the train data and performed fairly well in the test data as well. High explanation of variance with an R² of 0.95.



Thank You!

Visit our Gitub profiles for more information:



miguel Diaz pdj - Overview

i code sometimes @Data-Drivers-Unav.
miguel Diaz pdj has 9 repositories available. Follow their code on GitHub.



Dante Sc03 - Overview

I am a third year student at the University of Navarra majoring in Business Administration with Data Analytics. I am working @CadeiraAI and...



PdE - Solicitud de descarga aceptada

Externo

Recibidos x



noreply@puertos.es

para mí ▾

(hace 9 horas)



Le informamos de que su petición de datos a Puertos del Estado nº 20241215192441 ha sido aceptada. En la parte inferior de este mensaje encontrará los botones de acceso a la descarga de los datos solicitados. Los datos estarán disponibles para dicha descarga durante 10 días, en concreto hasta la fecha [REDACTED] después serán eliminados.

Los ficheros cuentan con una primera línea a modo de cabecera donde se indica la variable correspondiente a cada columna y sus unidades. Para saber más sobre las características de los datos, acceda al documento disponible en el botón "+ info"

Datos del solicitante

Miguel Diaz - [n](#)

Organización: Universidad de Navarra
Motivo de la solicitud: Trabajo académico
Descripción: Para la asignatura de Large Scale Computing del grado de ADE + Data Analytics



Datos de la solicitud - Peticiones

Puertos del Estado

We obtained the data for our project after contacting Puertos del Estado and submitting a formal request for access to information from the buoy located in the maritime area of La Coruña.