Time Series Wave Height Prediction

By Alex Khater

Why is this relevant ?

- As Climate Change worsens, the ocean's conditions will also alter
- Even a small increase in Wave height can cause massive and dangerous consequences to
  - Coastal Cities
  - Boats
  - Anti-Flood Infrastructure
  - Agriculture
- Being Able to Predict Future Wave Height Can Allow Governments to Prepare

Image Source: https://www.sunshinecoastnews.com.au/2022/07/25/sunshine-coast-record-wave-heights/

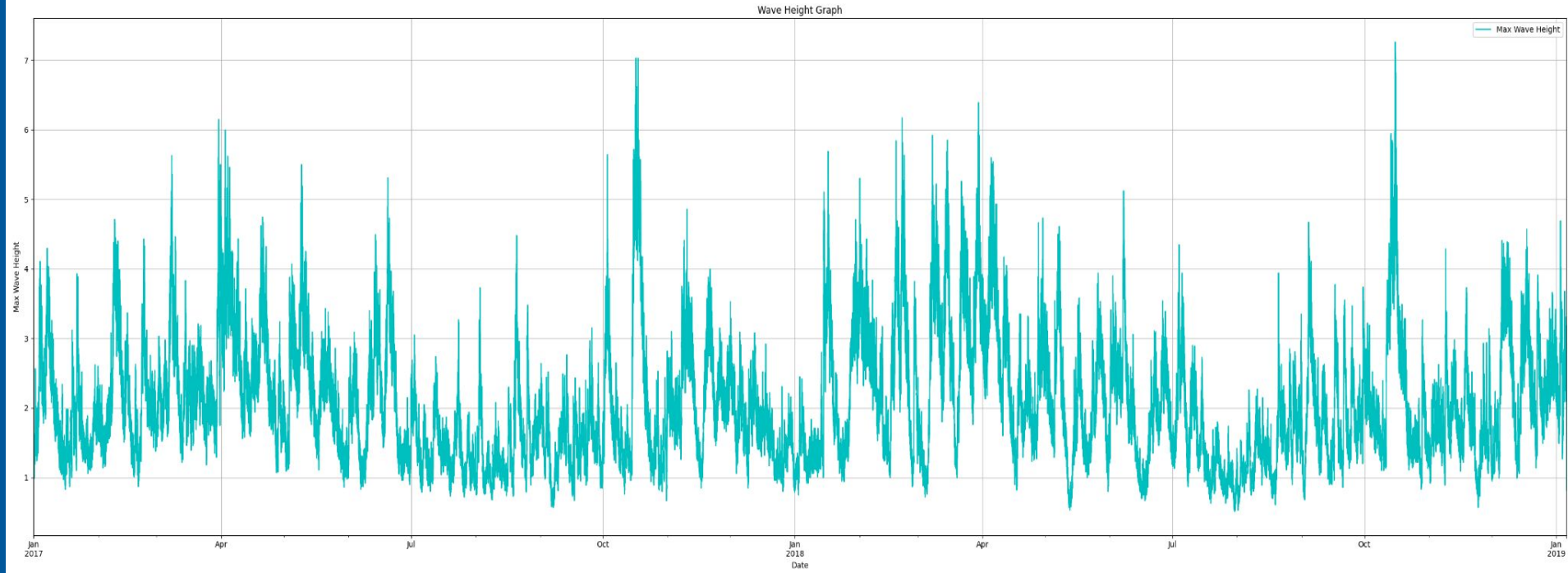
The Dataset

- 3 years of data bi-hourly from the Mooloolaba beach
- 43,000+ Observations
- The data includes features such as:
- Max Wave Height
- Average of the Top 3rd Wave Height
- Sea Temperature
- Zero Period
- Peak Energy
- Direction
- Date and Time Observations

Cleaning Process

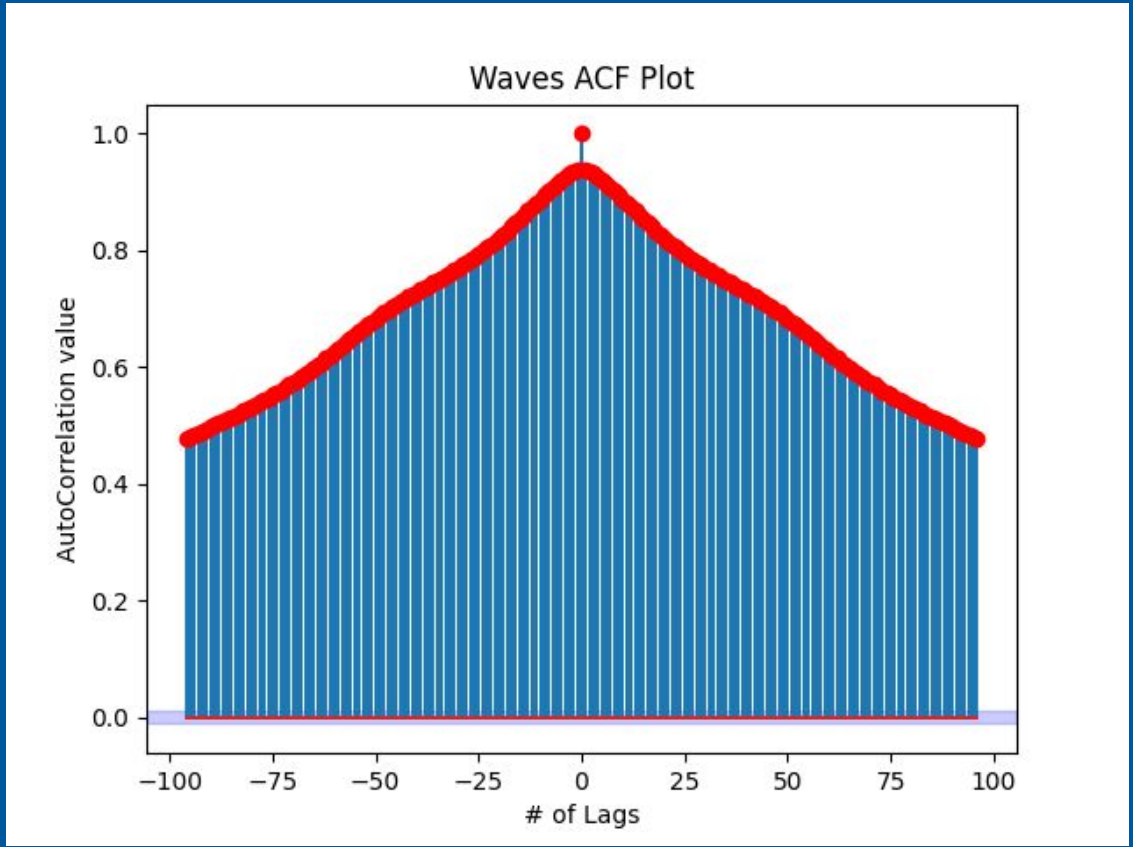
- Used the Drift Method to fill missing days
- Had to cut the data at 2019-01-06 because of uneven sampling
- No categorical Variables to encode
- Final dataset length: 35281

Plotting the Target Variable

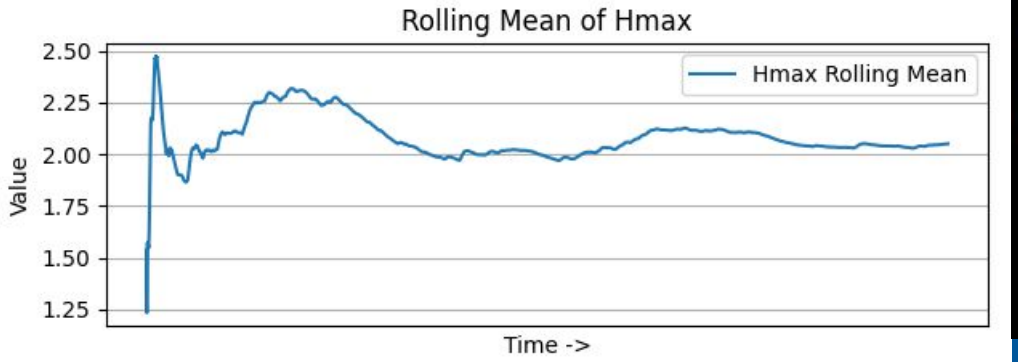


ACF Plot

- Notice the little bump in the ACF values at about lag= +/- 48 ?
- → Daily seasonality



Stationary ?




```
ADF Statistic: -10.664493
p-value: 0.000000
Critical Values:
    1%: -3.431
    5%: -2.862
    10%: -2.567
```

BASIC REGRESSION: VIF

```
Training Set length: 28224
Training Set length: 7057
Iteration 1:
[1.363935808713706, 1.7059569714156988, 1.5900266122963773, 1.1386370781202888, 1.1278098760347293]
MAX VIF IS FEATURE:1
```

All VIF values are below 2, so no dimensionality reduction is required

BASIC REGRESSION: Backwards Stepwise Regression

OLS Regression Results

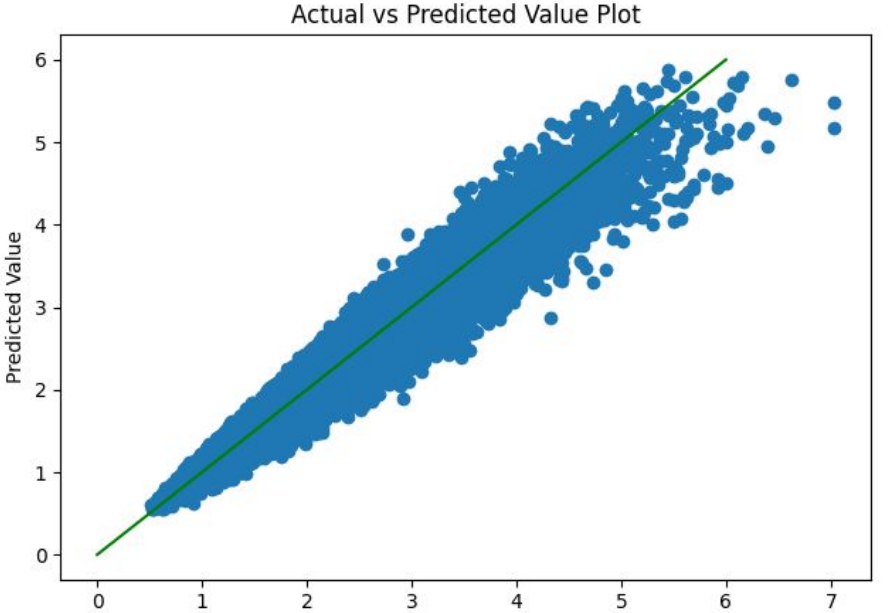
Dep. Variable:	Hmax	R-squared:	0.949
Model:	OLS	Adj. R-squared:	0.949
Method:	Least Squares	F-statistic:	1.748e+05
Date:	Mon, 06 Nov 2023	Prob (F-statistic):	0.00
Time:	12:51:29	Log-Likelihood:	4876.2
No. Observations:	28224	AIC:	-9744.
Df Residuals:	28220	BIC:	-9711.
Df Model:	3		
Covariance Type:	nonrobust		

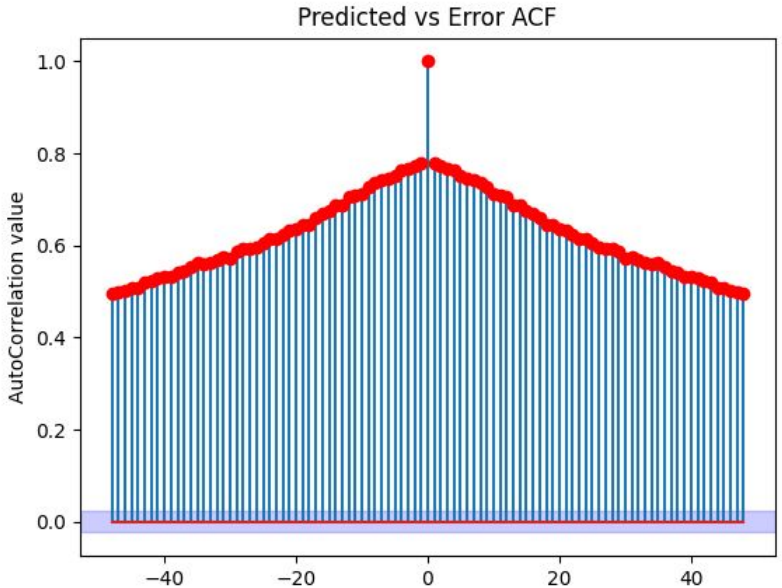
	coef	std err	t	P>|t|	[0.025	0.975]
const	2.0556	0.001	1696.214	0.000	2.053	2.058
Hs	0.8874	0.001	648.921	0.000	0.885	0.890
Tz	-0.0249	0.002	-15.743	0.000	-0.028	-0.022
Tp	0.0031	0.001	2.176	0.030	0.000	0.006

Omnibus:	5967.141	Durbin-Watson:	1.996
Prob(Omnibus):	0.000	Jarque-Bera (JB):	23959.241
Skew:	1.004	Prob(JB):	0.00
Kurtosis:	7.042	Cond. No.	2.15

After 4 rounds:
The Final Model keeps 3 variables: Hs, Tp, and Tz

BASIC REGRESSION: Residual Analysis






```
Variance of Residuals: 0.18128808662321028
Mean of Residuals: -0.9818257081080027
FTEST <F test: F=174757.5974591966, p=0.0, df_denom=2.82e+04, df_num=3>
```

BASIC REGRESSION: Plotting Predictions