AkzoNobel

HULL PERFORMANCE ANALYSIS

CONTENT

OVERVIEW

- The provided ship monitoring dataset contained information about various factors that affect the performance of the ship.
- The readings are recorded for a single ship at various instances of time.
- The readings are recorded at the end of each hour in most of the cases and in some places there are occational long gaps between the readings.
- The dataset contains information of the ship from July 2018 to February 2024.

FEATURES

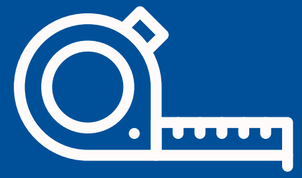
- TIME ATTRIBUTES
- ENGINE ATTRIBUTES
- SPEED ATTRIBUTES
- SHIP STABILITY ATTRIBUTES
- LOCATION ATTRIBUTES
- FUEL ATTRIBUTE

FEATURE ENGINEERING

To make the analysis more efficient, additional features were calculated from the existing features and added to the dataset.

Time elapsed

Time elapsed between the current reading and previous reading.



Distance covered

Distance travelled by the ship between the current reading and previous reading.

Fuel efficiency

Distance the ship has travelled for 1000 kgs of fuel.

Hull stability

Absolute difference between Draft fore and Draft Aft.

DESCRIPTIVE STATISTICS

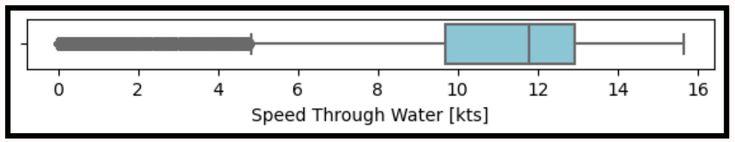
	Speed Through Water [kts]	Shaft Power [kW]	Fuel Oil Cons. [kg/hour]	Shaft Speed [RPM]	Torque [nM]	Wind Speed Abs. [kts]	Speed Over Ground [kts]
count	39518.000000	3.951800e+04	39518.000000	3.951800e+04	39518.000000	39518.000000	39518.000000
mean	9.415576	1.018669e+04	1277.427295	2.299964e+02	1431.284227	15.179964	9.808380
std	4.915067	3.137320e+04	982.687655	3.709839e+04	903.582900	8.213962	5.130566
min	0.000000	0.000000e+00	-8013.666667	-9.288333e+00	-252.833333	0.000000	0.000000
25%	8.966667	4.764021e+03	23.554167	3.797042e+01	666.000000	8.710417	9.368333
50%	11.581667	1.005508e+04	1390.275000	4.800500e+01	1603.725000	14.321667	11.955466
75%	12.815000	1.371635e+04	2066.200000	5.387000e+01	2202.441667	20.514583	13.318333
max	15.643333	4.789570e+06	7279.900000	7.374140e+06	3050.150000	52.316667	17.465000

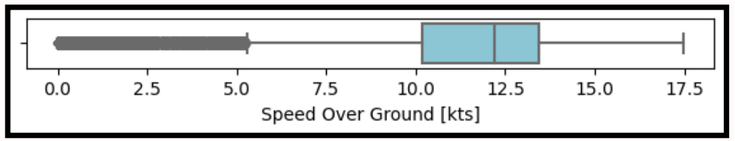
	Draft Aft [m]	Draft Fore [m]	distance covered (nmi)	time elapsed (hrs)	fuel efficiency	hull stability
count	39518.000000	39518.000000	39518.000000	39518.000000	3.231400e+04	39518.000000
mean	16.172651	14.714601	13.017973	1.229187	inf	1.533872
std	4.414605	5.915749	68.510110	8.620931	NaN	1.535312
min	3.900000	0.100000	0.000000	0.000000	-8.906673e+03	0.000000
25%	11.225000	8.001674	0.000000	1.000000	0.000000e+00	0.150000
50%	19.301667	18.920000	0.000000	1.000000	0.000000e+00	0.559167
75%	19.978333	19.825000	22.889399	1.000000	1.665518e+01	3.116667
max	22.993333	21.185000	7050.194086	821.000000	inf	6.920000

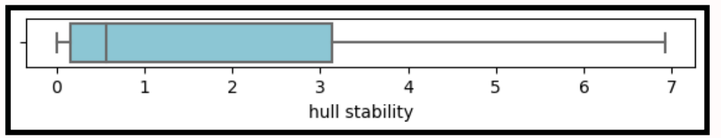
HANDLING ANOMALIES

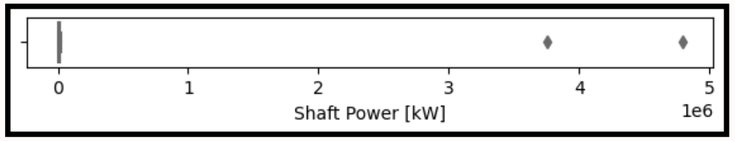
- The provided dataset contains misentries and outliers.
- Misentries are detected using the domain knowledge by analysing certain values of the descriptive statistics like mean, range etc. For example, the rows containing negative values of Torque, Shaft speed and Fuel consumption were removed as their values are physically insignificant.
- Outliers can be detected by visually inspections and using some statistical terminologies like IQR range and Z-scores.
- I have predominantly used box plot to visualize the outliers and used IQR range method to remove the outliers from the dataset.

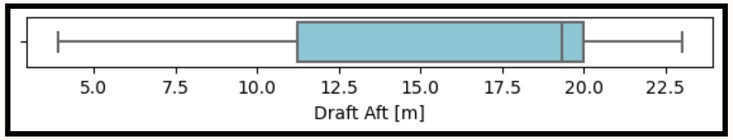
BOX PLOTS

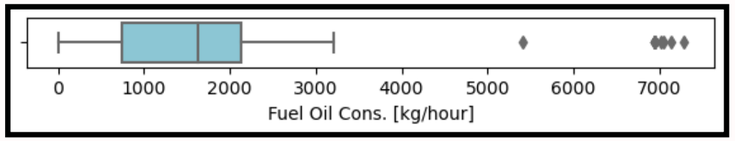

0 2 4 6 8 10 12 14 16
Speed Through Water [kts]

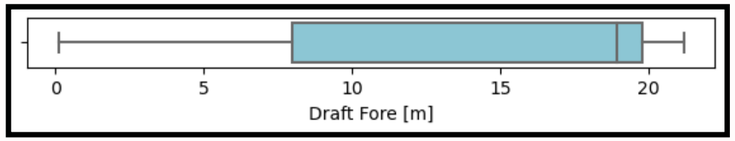

0.0 2.5 5.0 7.5 10.0 12.5 15.0 17.5
Speed Over Ground [kts]

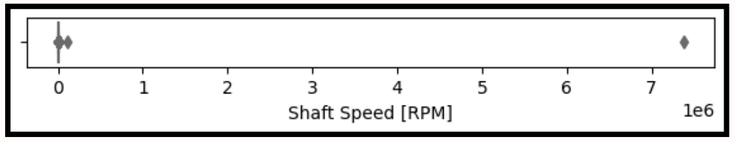

0 1 2 3 4 5 6 7
hull stability

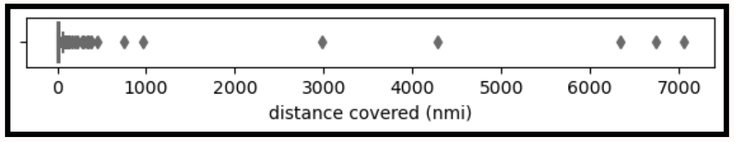

0 1 2 3 4 5
Shaft Power [kW]
1e6

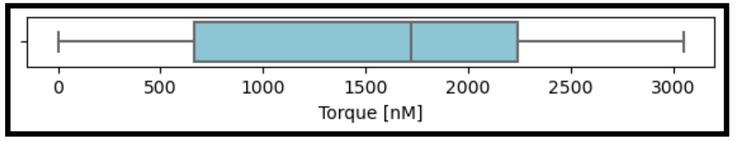

5.0 7.5 10.0 12.5 15.0 17.5 20.0 22.5
Draft Aft [m]

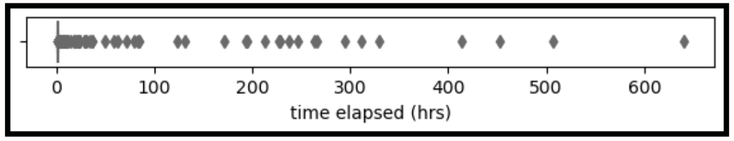

0 1000 2000 3000 4000 5000 6000 7000
Fuel Oil Cons. [kg/hour]

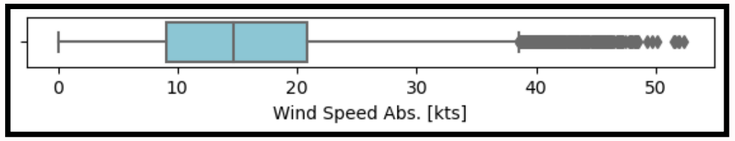

0 5 10 15 20
Draft Fore [m]

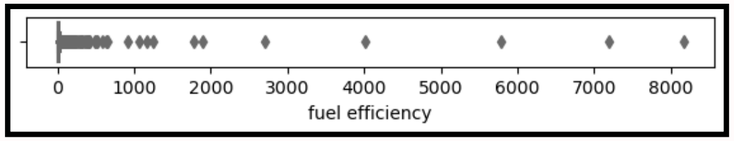

0 1 2 3 4 5 6 7
Shaft Speed [RPM]
1e6


0 1000 2000 3000 4000 5000 6000 7000
distance covered (nmi)


0 500 1000 1500 2000 2500 3000
Torque [nM]


0 100 200 300 400 500 600
time elapsed (hrs)


0 10 20 30 40 50
Wind Speed Abs. [kts]


0 1000 2000 3000 4000 5000 6000 7000 8000
fuel efficiency

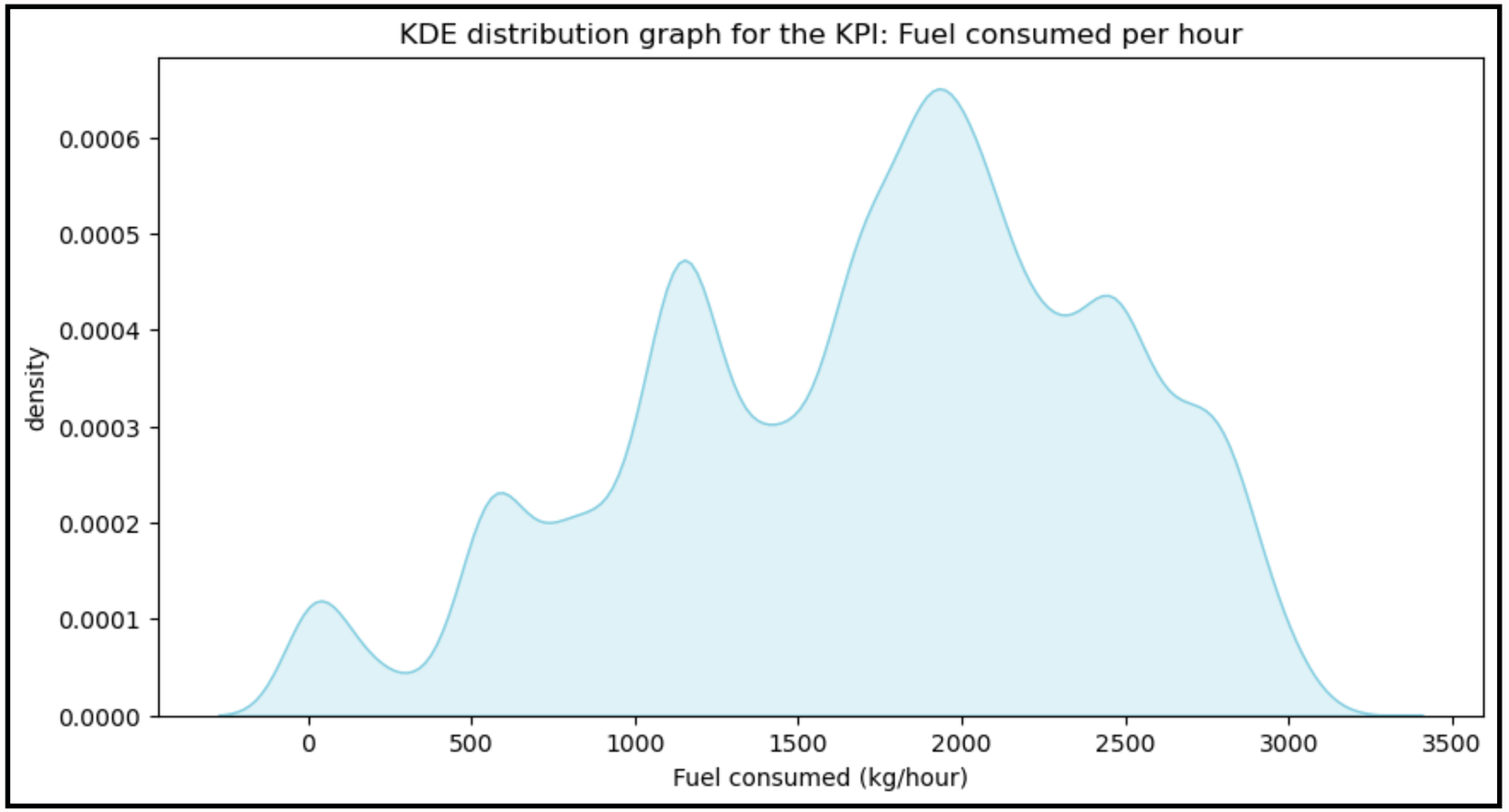
KEY PERFORMANCE INDICATOR

- The **Fuel consumed per hour** by the ship is taken as the key performance indicator for this analysis.

- The datapoints of the variable are distributed with almost equally on both sides of the mean.

- The main reason for choosing this as the primary KPI is that, Akzonobel is more concerned on a sustainable energy plan. The company has produced various strategies to reduce carbon emmision and energy usage to safeguard maritime life and give excellent service to it’s clients. So, it is a necessity to analyse the usage of fuel by the ships.

DISTRIBUTION CURVE OF THE KPI.



ANOVA TEST

- To analyse the significance of difference in the performance of the ship over time, an ANOVA test is performed.
- The dataset was grouped year wise and the significance of differences in the means of fuel efficiency for each year is analysed.
- The null hypothesis is that all means have no significance difference and the alternative hypothesis is atleast one group mean is different.
- Obtained a F-statistics value of 18.94 and p-value of 3.84 x 10^-22 after performing the test.

- 18.95 is the ratio of variance among group mean to variance within group and a very low p-value suggests strong evidence against the null hypothesis.
- Therefore, we can conclude that the difference in performance of the ship over time is statistically significant.
- After obtaining the ANOVA results, a tuckey HSD post hoc test was performed to analyse the pair wise mean differences between groups.
- This shows the pair wise comparison of Ship's performance between different years.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
2018	2019	-0.5728	0.2611	-1.3184	0.1728	False
2018	2020	-0.7607	0.0819	-1.5705	0.049	False
2018	2021	-1.384	0.0	-2.121	-0.647	True
2018	2022	-1.9809	0.0	-2.7284	-1.2334	True
2018	2023	-1.9128	0.0	-2.6763	-1.1492	True
2018	2024	-1.172	0.2007	-2.615	0.2711	False
2019	2020	-0.1879	0.9812	-0.8504	0.4745	False
2019	2021	-0.8112	0.0006	-1.3824	-0.2399	True
2019	2022	-1.4081	0.0	-1.9928	-0.8234	True
2019	2023	-1.34	0.0	-1.9451	-0.7348	True
2019	2024	-0.5992	0.8553	-1.965	0.7667	False
2020	2021	-0.6233	0.0725	-1.276	0.0295	False
2020	2022	-1.2202	0.0	-1.8848	-0.5556	True
2020	2023	-1.152	0.0	-1.8347	-0.4694	True
2020	2024	-0.4113	0.9776	-1.8132	0.9907	False
2021	2022	-0.5969	0.0351	-1.1707	-0.0232	True
2021	2023	-0.5288	0.1192	-1.1233	0.0658	False
2021	2024	0.212	0.9993	-1.1492	1.5732	False
2022	2023	0.0682	0.9999	-0.5393	0.6757	False
2022	2024	0.809	0.5855	-0.558	2.1759	False
2023	2024	0.7408	0.6904	-0.635	2.1166	False

DASHBOARD

A static Power BI dashboard was developed to identify the trends and correlations within the dataset.

AkzoNobel

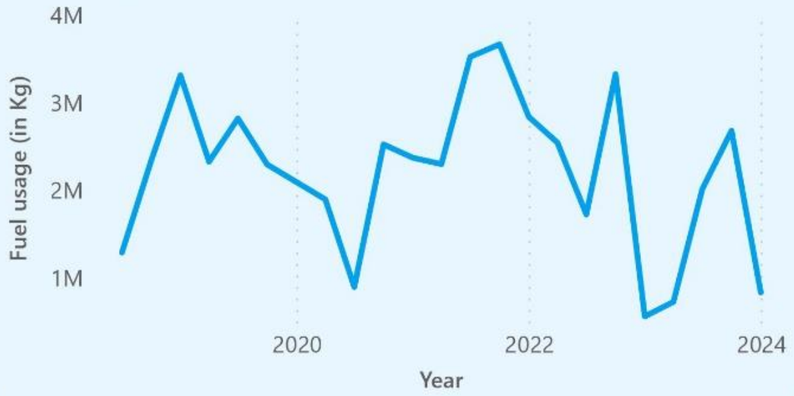
HULL PERFORMANCE ANALYSIS

357.4K
Total voyage distance (nmi)

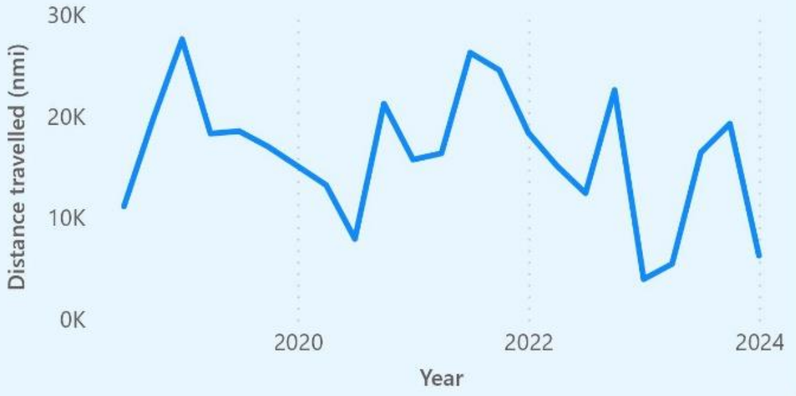
48.8M
Fuel usage (kg)

7.16
Mean fuel efficiency (nmi/tonne)

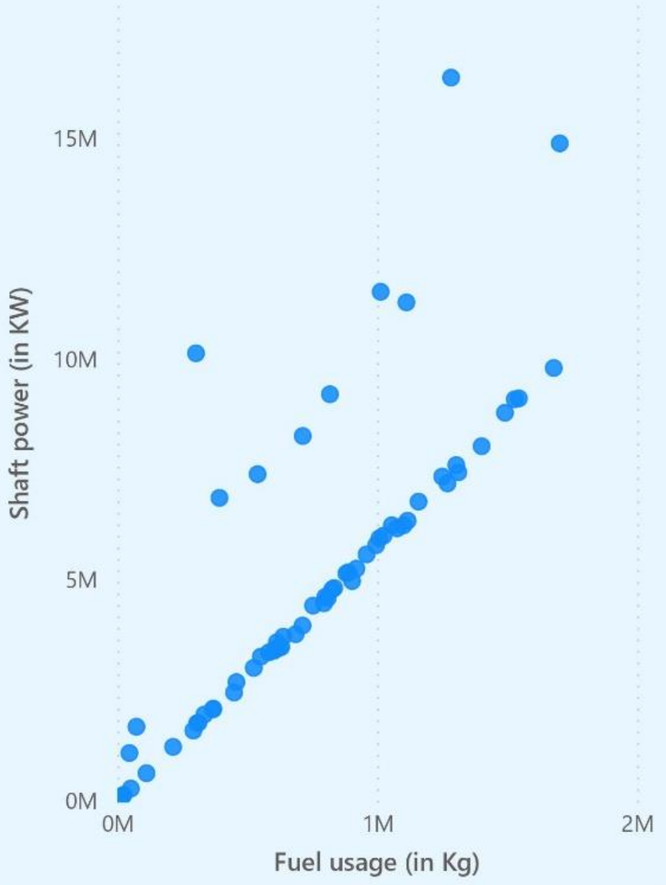
Fuel usage trend



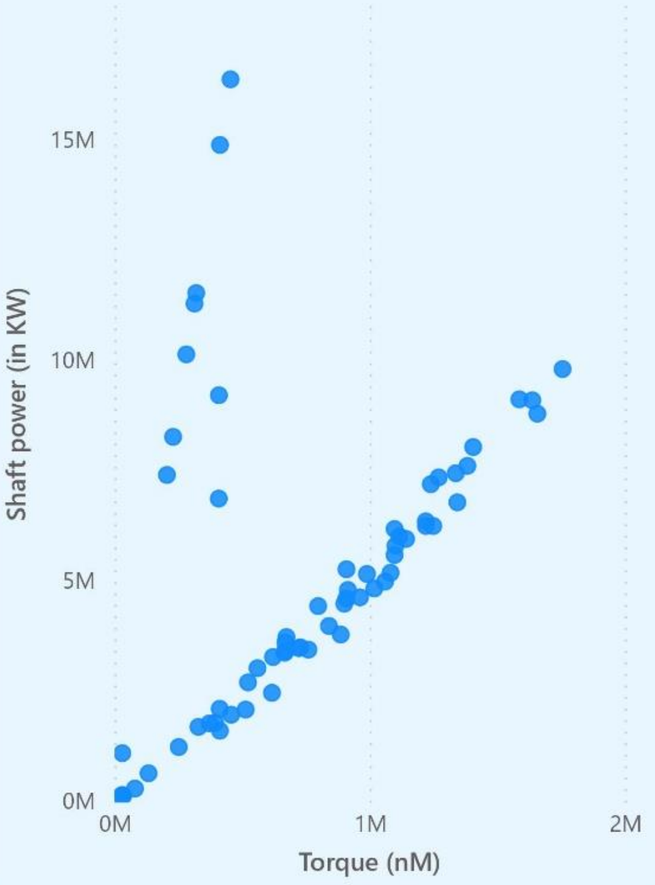
Distance travelled trend



Correlation between power generated and fuel consumed



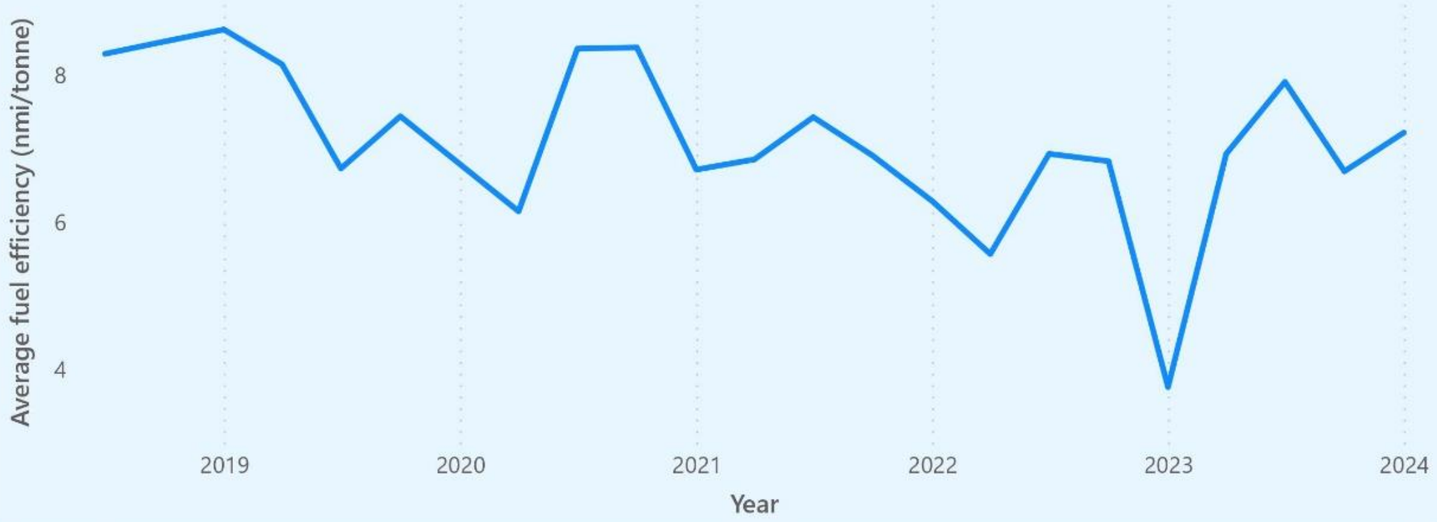
Correlation between power generated and Torque



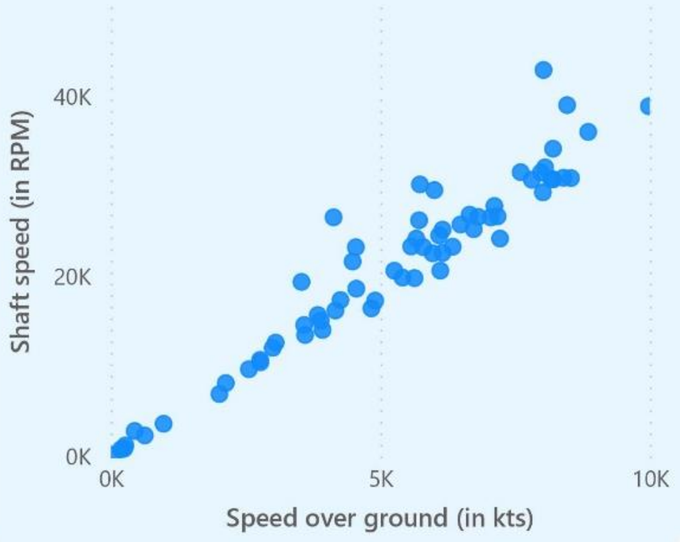
Hull performance table

Year	Quarter	Distance covered (nmi)	Fuel usage (Kg)	Mean fuel efficiency (nmi/tonne)
2018	Qtr 3	11,119.03	1289458.44	8.29
2018	Qtr 4	19,679.47	2339069.61	8.46
2019	Qtr 1	27,587.39	3313928.71	8.62
2019	Qtr 2	18,298.30	2327708.34	8.15
2019	Qtr 3	18,529.01	2821096.83	6.73
2019	Qtr 4	16,994.52	2294513.29	7.44
2020	Qtr 2	13,221.56	1896046.49	6.15
2020	Qtr 3	7,909.62	895053.10	8.36
2020	Qtr 4	21,240.71	2524308.96	8.38
2021	Qtr 1	15,727.41	2371933.15	6.72
2021	Qtr 2	16,351.75	2298965.47	6.86
2021	Qtr 3	26,265.36	3524110.61	7.43
2021	Qtr 4	24,542.22	3667064.26	6.91
2022	Qtr 1	18,378.94	2839213.51	6.29
2022	Qtr 2	15,128.48	2544951.52	5.57
2022	Qtr 3	12,423.23	1724167.79	6.93
2022	Qtr 4	22,580.44	3326883.38	6.83
2023	Qtr 1	3,947.13	559691.53	3.76
2023	Qtr 2	5,458.11	725794.30	6.93
2023	Qtr 3	16,430.91	2014220.65	7.91
2023	Qtr 4	19,271.92	2680869.34	6.69
2024	Qtr 1	6,279.01	833313.37	7.22
Total		**3,57,364.51**	**48812362.63**	**7.16**

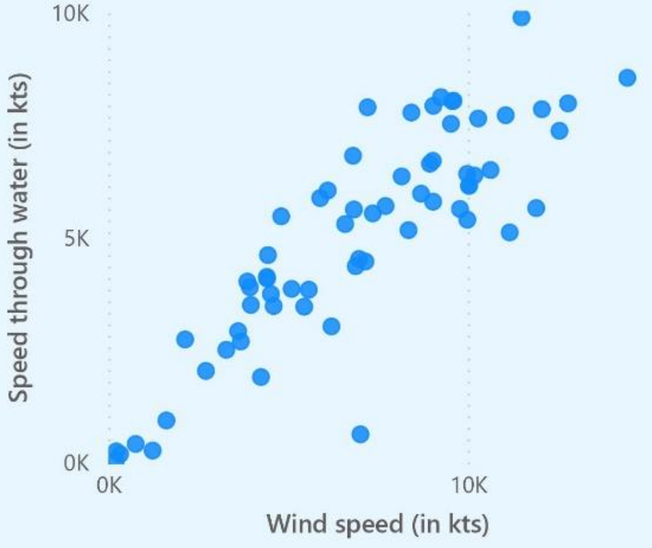
Fuel efficiency trend



Correlation between Shaft speed and Speed over ground

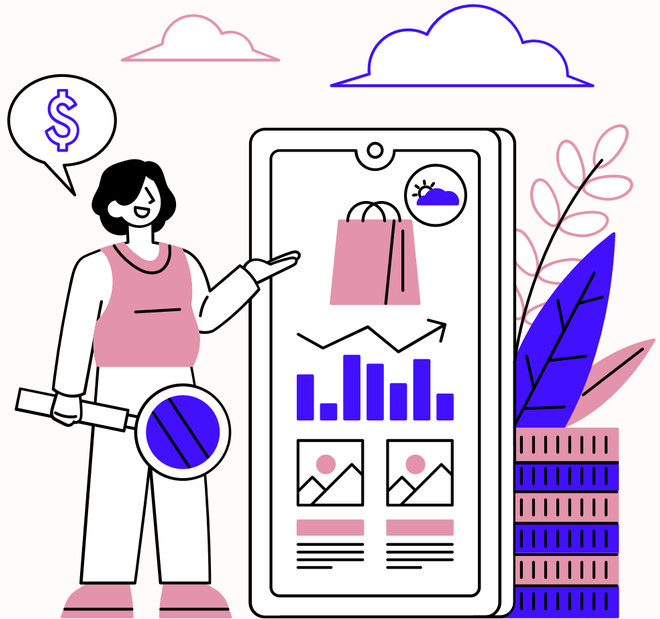


Correlation between Wind speed and Speed through water



TIME SERIES FORECASTING

- To predict the performance of the ship over the period of time, a time series forecasting model with fuel consumption attribute as the target variable was developed.
- ARIMAX model was chosen for this prediciton process.
- ARIMAX is a variation of the Auto regression and Moving averages model along with the consideration of exogeneous variables.
- Here, shaft torque and shaft speed were taken as the exogeneous variables and the fuel consumption was taken as the endogeneous variable.



The hour wise dataset was aggregated by days for clear identification of the correlation of the target variables with it's previous values.

DATA TRANSFORMATION

Autocorrelation function plots and partial auto correlation function plots were plotted to identify the p,d and q values.

COMPUTING THE PARAMETERS OF THE MODEL

STATIONARITY CHECK

The variation of mean and variance of the dataset was checked. For an ARIMAX model the dataset should be stationary.

MODEL VALIDATION

The model is then fitted and the results are evaluated to validate the accuracy of the model.

DATA TRANSFORMATION

- The features that are required for the time series modelling were first selected.

- Shaft torque, shaft speed and fuel consumption are the features selected. They were aggregated datewise with mean as the aggregation type.

- The time difference between the rows in the new transformed dataframe was not uniform. This will definitely affect the performance of the model.

- Therefore the dates that were missing in the dataset were added and the values for the features for these missing dates were filled with zeroes.

STATIONARITY CHECK

- Before fitting the dataset into ARIMAX model, it is essential to check whether the target variable is stationary. That is, the mean and variance of fuel consumption should not vary with time.

- An ADFuller test is performed to identify the stationarity of the dataset.

- We obtain an ADF statistics score, critical values and p-values as a result of this test.

- If the p-value is less than 0.05 and the ADF statistics score is less than the critical values, we can conclude that the dataset is staionary.

- The result of the ADFuller test performed is shown below.

```
ADF statistics : -8.662462288676245
P-value : 4.7815078239015e-14

Critical values

1% : -3.4335946172657774
5% : -2.8629733042364576
10% : -2.567533028727577
```

- Here the p-value is less than 0.05 and the ADF statistics value is less than the critical values. Therefore, we can conclude that the dataset is stationary.

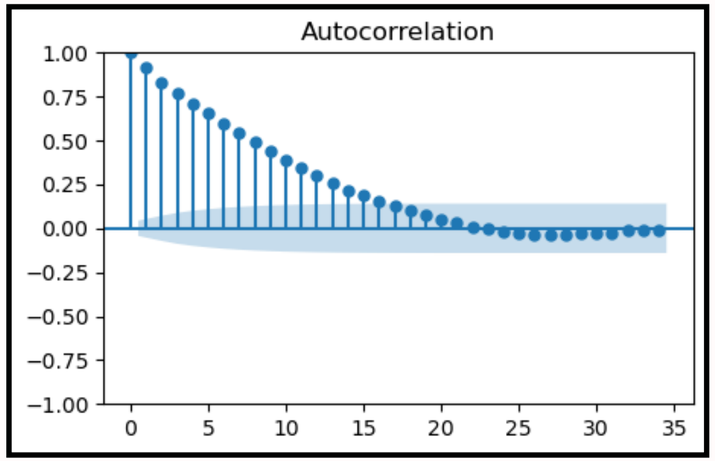
PARAMETERS OF THE MODEL

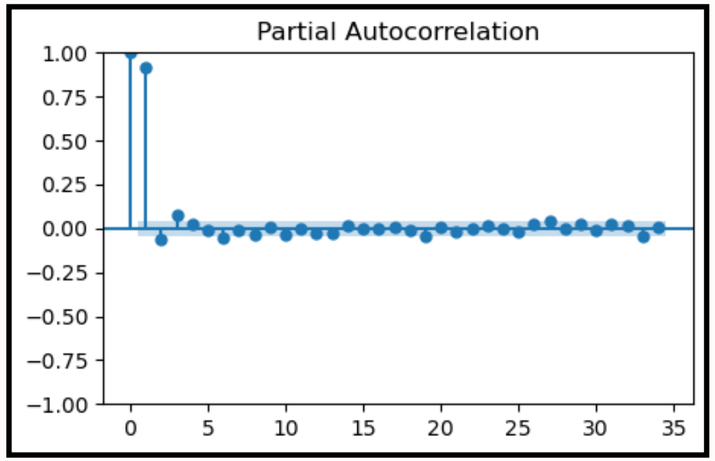
- In an ARIMAX model, the p,d and q values are the additional parameters required for modelling along with the endogeneous and exogeneous variables.
- “p” represents the number of lag values of the time series variable included in the model.
- “q” represents the number of lag values of the forecast errors produced by the moving averages included in the model.
- “d” represents the number of times the target variable is differenced to achieve stationarity. In our case, since we did not do any differencing, the “d” value is 0.
- The “p” and “q” values are identified using ACF and PACF plots.

COMPUTING P AND Q VALUES

- Rule : The lag at which the ACF and PACF plot becomes insignificant or zero is always takes as the “q” and “p” value respectively.

- The ACF and PACF plot for the fuel consumption of the ship is shown below.





- Here, according to the rule p=16 and q=1.

MODEL VALIDATION

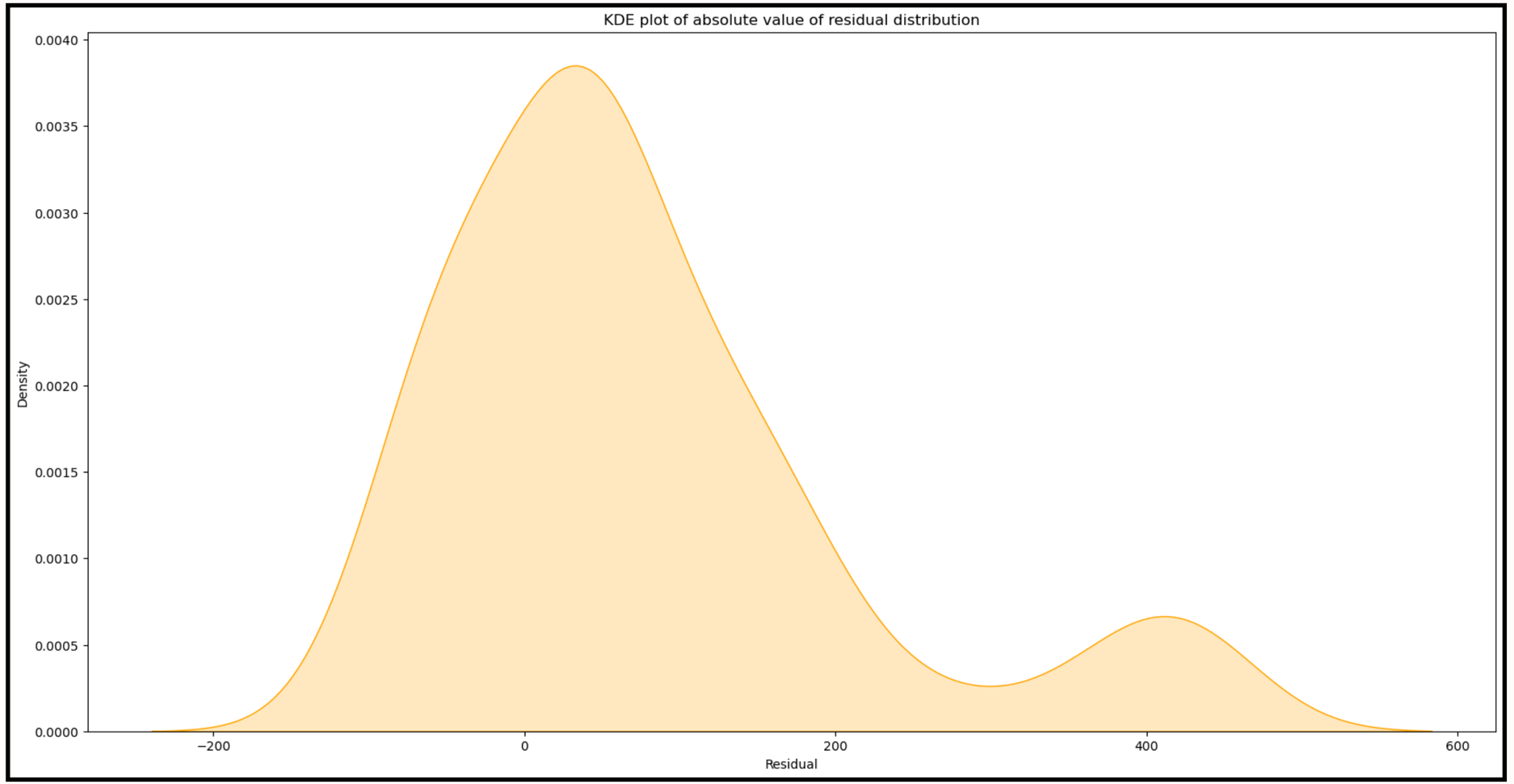
- The dataset was partitioned into a training set, containing information of the ship's performance until the date of August 01, 2023, and a test set, comprising data reflecting the ship's performance subsequent to the mentioned date.

- The model is fitted using the predefined exogeneous variables and endogeneous variables of the train set data along with the order of the ARIMAX model (p,d and q values).

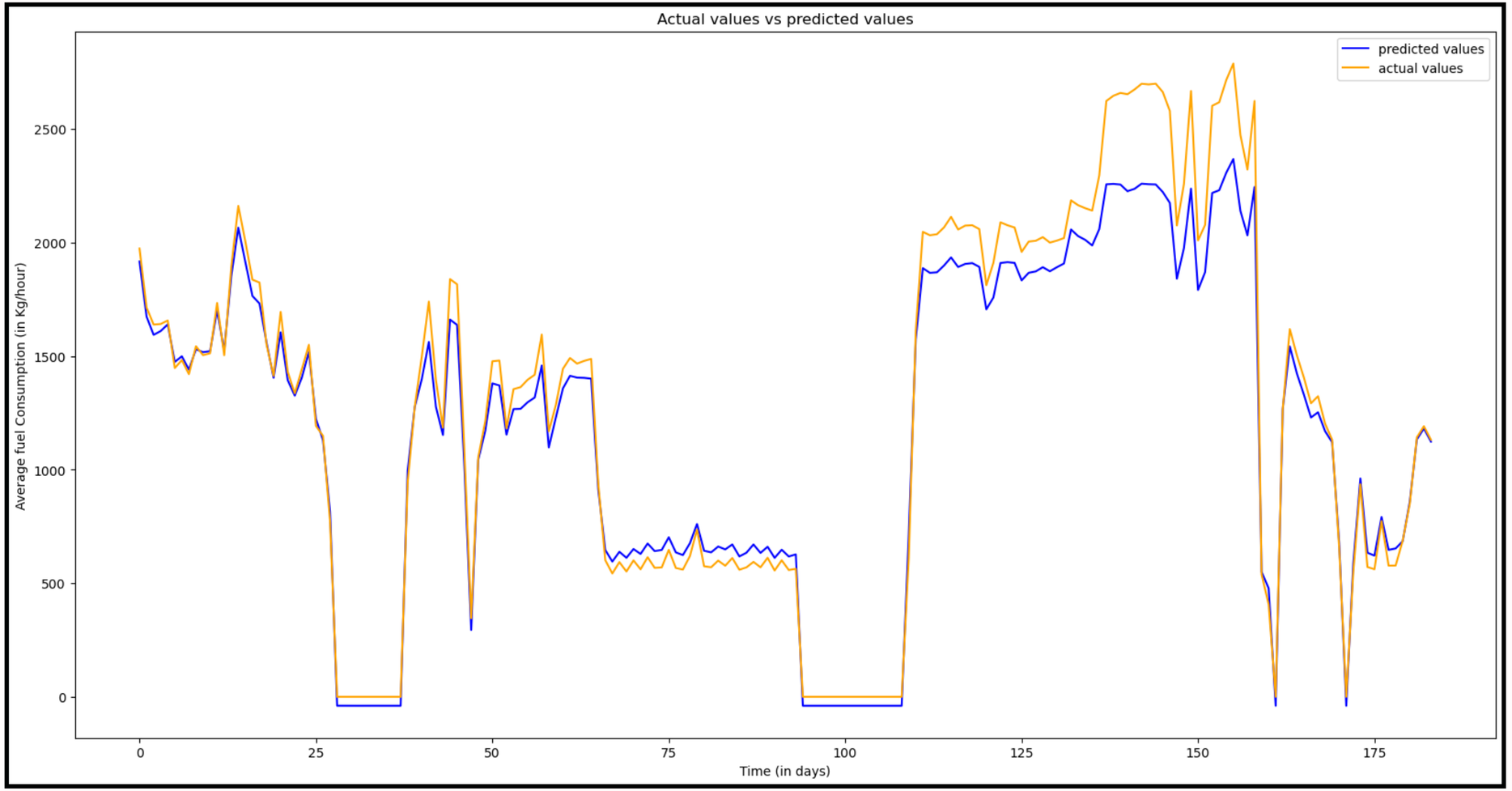
- The fitted model is used to forecast the fuel consumption per hour for the test data.

- The model recorded a MAE of 104.52 and RMSE of 152.68 during the forecast.

- The residual plot of the forecast is shown below.

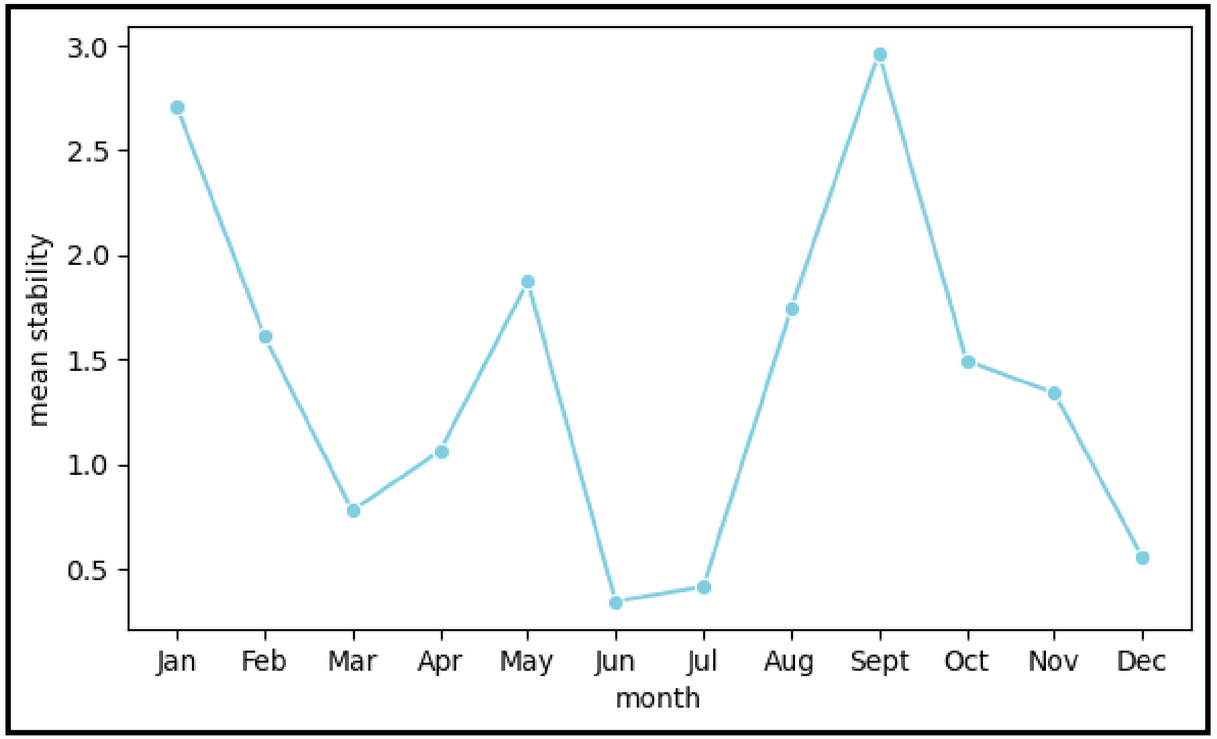


- The variation of the predicted values with respect to the observed value in the test data is shown below. Orange line represents the observed values and blue line represent the predicted values.



ADDITIONAL ANALYSIS

1. The stability of the ship at different time of the year



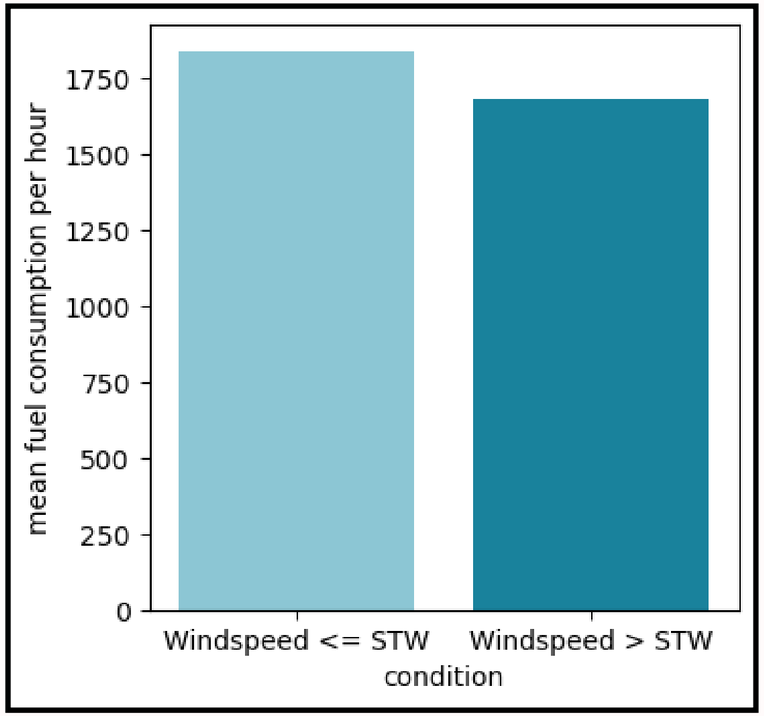
- The average stability of the ship for each month is calculated and the values are plotted in the graph.
- We may see the graph peaking at certain parts of the year, which might be the indication of magnitude of loading and unloading of goods from the ship at those periods.

2. Effect of windspeed on Fuel consumption of the ship.

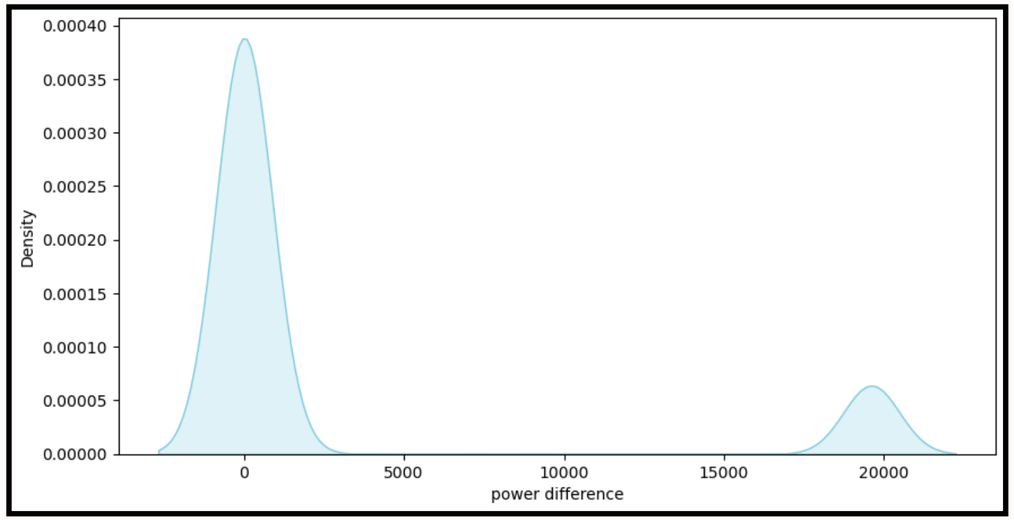
- There was no direct significant correlation between the windspeed and fuel consumption when observed through a correlation plot.

- Therefore, to investigate the influence of wind speed, the dataset was split into two subgroups. One of which contained instances where the wind speed was equal to or less than the ship's speed through water, while the other subset comprised data points where the wind speed exceeded the ship's speed through water.

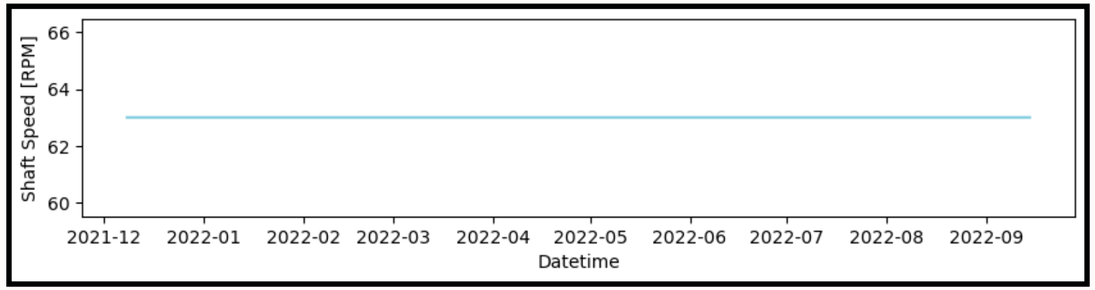
- A comparison bar plot is plotted to show the average fuel consumption in both the cases. We may observe that when the windspeed is greater than the speed through water, the fuel consumed by the ship is relatively low.

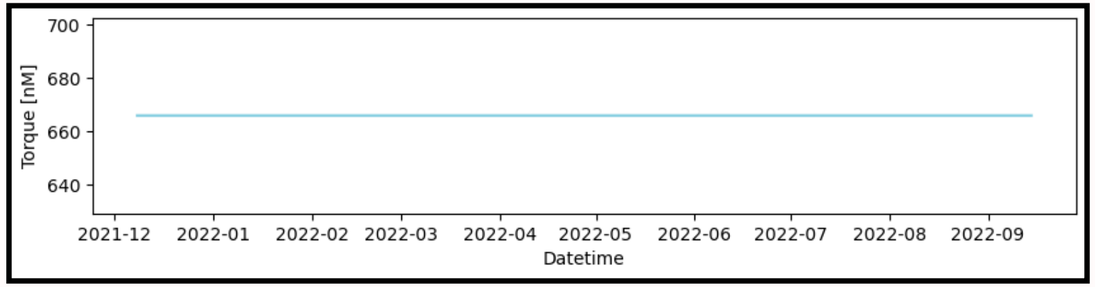


3. Difference between the observed power and calculated power



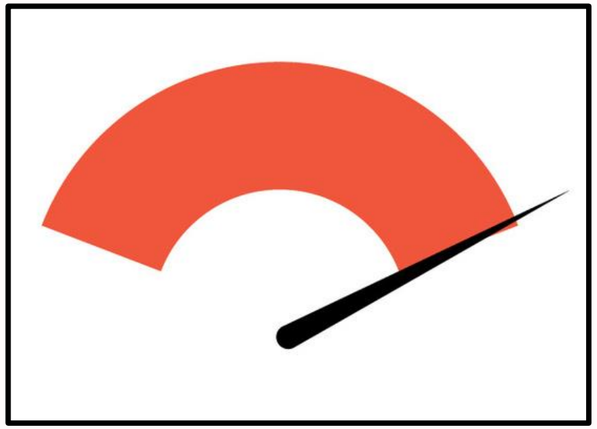
- To investigate any machine errors in the readings in the dataset, I manually computed the Shaft power using the Shaft speed and Shaft Torque.
- The relationship is, Power = (Torque x 2π x speed) / 60
- The difference between calculated power and observed power was calculated, and a distribution graph was plotted.
- In the graph, we may observe that for a certain group of readings, the error values are very high (close to 20000 KW).
- To examine further on these error values, I inspected the Torque and Speed values, associated with these readings



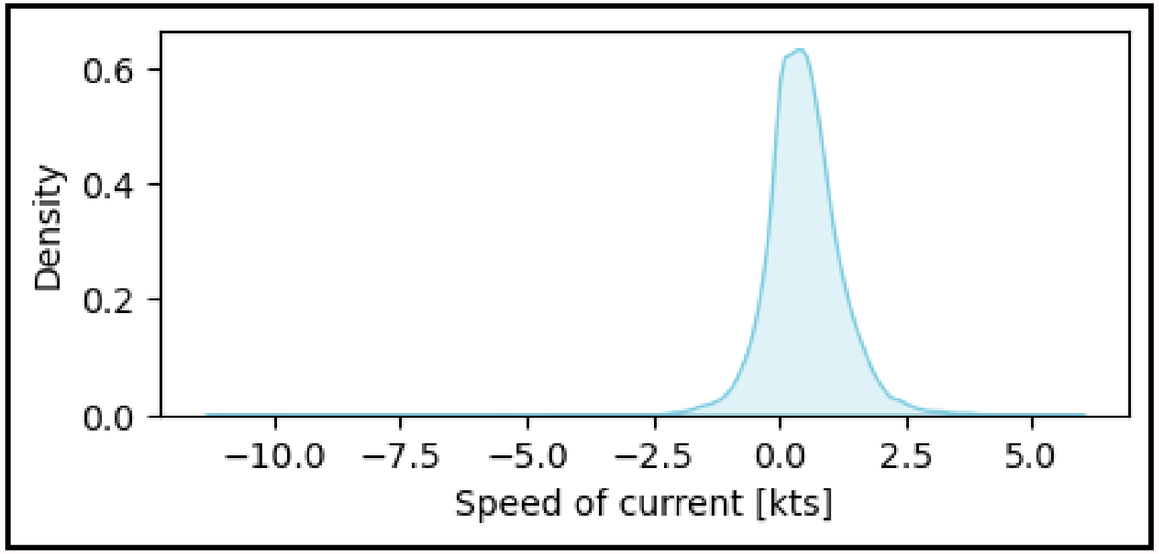


- The values of Torque and Shaft speed was found to be a constant value for all these readings.

- One possible reason behind this could be that these observed torque or speed values may represent the maximum capacity of the meter in which they are recorded. Whenever the ship experiences a value exceeding the meter's range, the machine might have provided the maximum value within its operational limits.



4. The speed of current.



- The graph depicts the distribution curve of the speed of the current.
- The speed of the current is calculated using the difference between the speed of ship over ground and speed of ship relative to water.
- The range of speed of current was observed to be (-11.08, 5.74) in knots.

CONCLUSIONS

- The provided Ship performance dataset is cleaned, transformed and analysed with visualizations and forecasting models.
- The ship has travelled a total voyage distance of more than 357,400 nautical miles consuming over 48.8 million kilograms of fuel.
- The ship has travelled an average of 7.16 nautical miles for every 1000 kg of fuel.
- The difference in performance of the ship over the time is significant.
- The Ship’s performance are heavily influenced by various other factors as well apart from the time factor.
- The fuel efficiency of the ship is gradually reducing as the ship gets older which is a matter of concern.

THANK YOU

GitHub

LinkedIn