Ship Fuel Consumption & Emission Analysis Project



Project Introduction:

Analyzing Fuel Consumption and CO2 Emissions of Ships in Nigeria

Background:

The maritime industry is a vital part of Nigeria's economy, facilitating the transport of goods and people across the country's waterways. With the increasing focus on environmental sustainability and the reduction of carbon emissions, there is a need for a comprehensive analysis of fuel consumption and CO2 emissions from different types of ships operating in Nigerian waters.

Objective:

This project aims to perform an in-depth analysis of fuel consumption and CO2 emissions across various ship types in Nigeria, specifically focusing on Fishing Trawlers, Oil Service Boats, Surfer Boats (offshore passenger transfer vessels), and Tanker Ships. By doing so, we

can identify trends, correlations, and significant differences that could inform decision-making and potential strategies for fuel management and emission reduction.

Scope and Approach:

- Data Generation: We will be using simulated data that represents realistic fuel consumption and CO2 emission scenarios for each ship type.
- Exploratory Data Analysis (EDA): Visualizations and summary statistics will be generated to understand the distribution of fuel consumption and emissions and identify any initial patterns.
- Trend Analysis: Analyze the monthly variations in fuel consumption and CO2 emissions to understand seasonal trends.
- Correlation Analysis: Assess relationships between fuel consumption, CO2 emissions, and other factors like engine efficiency.
- Statistical Testing: Use ANOVA and Tukey HSD to identify significant differences in fuel consumption across different ship types and validate the finding.
- Predictive Analysi (**Future Work**)): Evaluate if predictive modeling could add value to forecasting fuel consumption and CO2 emissions.

Dataset Schema

Column	Description	Example
ship_id	Unique identifier for ships	NG001
ship_type	Type of vessel	Tanker Ship
route_id	Identifier for routes	Port Harcourt- Lagos
month	Month of operation	January
distance	Distance covered in nautical miles	250
fuel_type	Type of fuel used	Diesel
fuel_consumption	Fuel consumed in liters	12,000
CO2_emissions	Carbon emissions in kilograms	32,000
weather_conditions	Weather during the trip (Calm, Moderate, Stormy)	Moderate
engine_efficiency	Efficiency of the ship's tngine (percen age)	85

centage) 85

Expected Outcomes:

- Identification of ship types with the highest fuel consumption and CO2 emissions.
- Insights into the trends of fuel consumption and emissions over months.

- Understanding of how engine efficiency correlates with fuel consumption and emissions.
- Statistical confirmation of significant differences in fuel consumption across ship types.

Dataset Preparation

```
In [3]: # Loading necessary libraries
        import pandas as pd
        import numpy as np
        import random
        import seaborn as sns
        import matplotlib.pyplot as plt
        # Initialize constants for dataset generation
        n_ships = 120 # Number of unique ships
        months = ["January", "February", "March", "April", "May", "June",
                  "July", "August", "September", "October", "November", "December"]
        ship_types = ["Surfer Boat", "Oil Service Boat", "Fishing Trawler", "Tanker Ship"]
        routes = ["Lagos-Apapa", "Warri-Bonny", "Escravos-Lagos", "Port Harcourt-Lagos"]
        weather_conditions = ["Calm", "Moderate", "Stormy"]
        fuel_types = ["Diesel", "HFO"]
        # Initialize dataset
        data = []
        # Data generation logic
        for i in range(1, n_ships + 1): # 120 ships
            ship_id = f"NG{str(i).zfill(3)}"
            ship_type = random.choice(ship_types)
            for month in months: # 12 months per ship
                route = random.choice(routes)
                # Set parameters based on ship type
                if ship_type == "Surfer Boat":
                    distance = round(random.uniform(20, 200), 2) # Surfer boats operate sh
                    fuel_type = "Diesel" # Surfer boats use only Diesel
                    fuel_efficiency_range = (10, 20)
                elif ship type == "Fishing Trawler":
                    distance = round(random.uniform(50, 200), 2)
                    fuel_type = random.choice(fuel_types)
                    fuel efficiency range = (20, 30)
                elif ship_type == "Oil Service Boat":
                    distance = round(random.uniform(20, 150), 2)
                    fuel_type = random.choice(fuel_types)
                    fuel efficiency range = (25, 35)
                elif ship_type == "Tanker Ship":
                    distance = round(random.uniform(50, 500), 2)
                    fuel_type = random.choice(fuel_types)
                    fuel_efficiency_range = (30, 50)
                fuel_efficiency = random.uniform(*fuel_efficiency_range)
                fuel_consumption = round(distance * fuel_efficiency, 2)
```

```
co2_emissions = round(fuel_consumption * random.uniform(2.5, 3.0), 2) # Em
        weather = random.choice(weather_conditions)
        engine_efficiency = round(random.uniform(70, 95), 2)
        # Append to dataset
        data.append([
            ship_id, ship_type, route, month, distance, fuel_type,
            fuel_consumption, co2_emissions, weather, engine_efficiency
        1)
# Create the DataFrame
columns = [
    "ship_id", "ship_type", "route_id", "month", "distance", "fuel_type",
   "fuel_consumption", "CO2_emissions", "weather_conditions", "engine_efficiency"
marine_dataset = pd.DataFrame(data, columns=columns)
# Save to CSV
# marine_dataset.to_csv("ship_fuel_efficiency.csv", index=False)
# Let inspect the few rows of the dataset
marine dataset.head()
```

Out[3]:		ship_id	ship_type	route_id	month	distance	fuel_type	fuel_consumption	CO2_emiss
	0	NG001	Tanker Ship	Lagos- Apapa	January	280.49	HFO	10732.86	2898
	1	NG001	Tanker Ship	Lagos- Apapa	February	68.46	HFO	2663.89	680
	2	NG001	Tanker Ship	Lagos- Apapa	March	144.23	Diesel	6175.10	1664
	3	NG001	Tanker Ship	Lagos- Apapa	April	102.12	HFO	4555.57	1207
	4	NG001	Tanker Ship	Warri- Bonny	May	195.01	Diesel	7030.31	1894
	4								•
In [4]: # Making a copy of the dataset									

Basic Checks & Data Cleaning

df = marine_dataset.copy()

```
In [6]: marine_dataset.shape
Out[6]: (1440, 10)
In [7]: marine_dataset.describe()
```

Out[7]:		distance	fuel_consumption	CO2_emissions	engine_efficiency
count 1440.000000			1440.000000	1440.000000	1440.000000
mean 15		152.873660	4797.253375	13117.412993	82.382778
	std 111.341855 min 20.100000 25% 77.595000 50% 123.935000		5040.957635	13743.318993	7.211340
			261.750000	701.570000	70.000000
			1751.467500	4861.320000	76.040000
			2933.730000	8024.505000	82.620000
	75%	177.677500	4831.557500	13274.160000	88.245000
	max	499.960000	24289.870000	72029.760000	94.990000
	RangeInd # Col # Col O shi 1 shi 2 rou 3 mor 4 dis 5 fue 6 fue 7 CO2 8 wea 9 eng dtypes:	dex: 1440 ent lumns (total lumn ip_id ip_type ute_id ith stance el_type el_consumptic 2_emissions	1440 non-nui lons 1440 non-nui ncy 1440 non-nui object(6)	unt Dtype Il object Il object Il object Il object Il float64 Il object Il float64 Il float64 Il object	
<pre>In [9]: Out[9]:</pre>	ship_t Tanker Oil Se Surfer Fishin	ype Ship crvice Boat	<pre>ip_type'].value_c 384 384 372 300 : int64</pre>	ounts()	

In [10]: marine_dataset.isnull().sum()

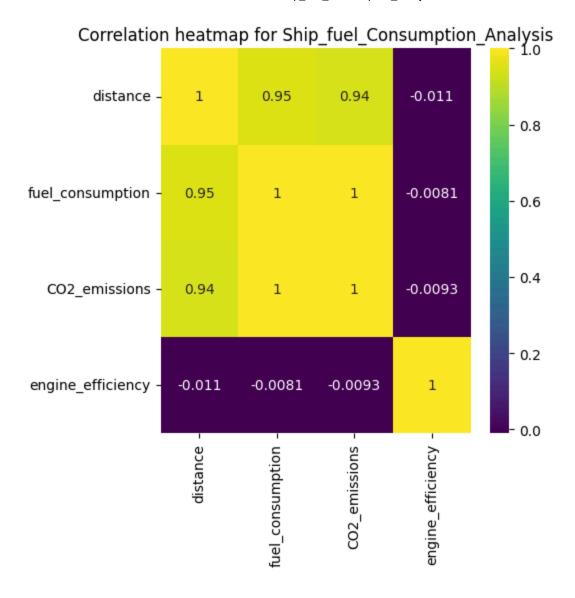
```
Out[10]: ship_id
                               0
         ship_type
         route_id
                               0
         month
                               0
         distance
         fuel_type
         fuel_consumption
         CO2_emissions
         weather_conditions
                               0
         engine_efficiency
         dtype: int64
In [11]: marine_dataset.duplicated().sum()
Out[11]: 0
```

• No missing Values as well as duplicated values in the dataset

Exploratory Data Analysis (EDA)

```
In [14]: # Creating Numerical Features DF
num_df = df.select_dtypes(exclude='0')

# Checking for Correlation
plt.figure(figsize=(5,5))
sns.heatmap(num_df.corr(), cmap='viridis', annot=True)
plt.title('Correlation heatmap for Ship_fuel_Consumption_Analysis')
plt.savefig('Heatmap.png')
```



Key Observations:

- Strong Positive Correlations:
- 1. distance and fuel_consumption (0.908)

As ships sail farther, they consume more fuel, which is expected. 2. fuel_consumption and CO2_emissions (0.993):

CO2 emissions are directly proportional to fuel consumption due to emission factors.

Insight:

 The heatmap highlights the direct dependency between distance, fuel consumption, and CO2 emissions. The absence of a strong correlation with engine efficiency could indicate that most engines in the dataset operate within a similar efficiency range, or the efficiency improvements are marginal relative to the scale of operations.

Actionable Insights Based on the Correlation Analysis:

Optimize Routes to Reduce Emissions:

- Since distance traveled is strongly correlated with fuel consumption (0.94) and CO2 emissions (0.94), optimizing shipping routes to minimize distance can lead to significant reductions in fuel usage and emissions.
- Action: Implement route optimization software or adopt real-time navigation systems to choose the most fuel-efficient routes.
- Target Fuel Consumption to Reduce Emissions:

The perfect correlation (1) between fuel consumption and CO2 emissions indicates that reducing fuel consumption directly decreases emissions.

- Action: Encourage the use of energy-saving measures like slow steaming (operating at lower speeds) and regular maintenance to enhance fuel efficiency.
- Invest in Engine Efficiency Improvements:

Although engine efficiency has a weak negative correlation with fuel consumption and CO2 emissions (-0.059 and -0.061), this suggests potential for reducing both when using more efficient engines.

- Action: Replace older engines with newer, more fuel-efficient models or retrofit engines with technologies like exhaust gas recirculation (EGR) or scrubbers to reduce emissions.
- Focus on High-Consumption Ship Types:

From earlier analysis, Tanker Ships have the highest fuel consumption and CO2 emissions. Addressing inefficiencies in these vessels could have the largest impact on reducing overall fuel usage and emissions.

- Action: Conduct detailed energy audits on tanker ships and implement targeted measures like hull cleaning, propeller polishing, and advanced weather routing.
- Explore Cleaner Energy Alternatives:

Since emissions are tied directly to fuel consumption, transitioning to cleaner fuels (e.g., LNG or biofuels) or alternative power sources (e.g., hybrid engines or electric propulsion) can significantly reduce CO2 emissions.

- Action: Invest in R&D for alternative energy sources and adopt sustainable fuel technologies for new builds or retrofits.
- Monitor and Benchmark Engine Efficiency:

The weak correlation of engine efficiency with other variables suggests that the fleet's engines may not be operating at their full potential.

 Action: Implement continuous monitoring systems to benchmark engine performance, identify underperforming engines, and take corrective actions.

By implementing these measures, shipping operations can become more environmentally sustainable, reduce operating costs, and comply with international emission regulations such as IMO's MARPOL Annex VI.

```
In [16]:
          # Let check how the Numerical Features are distributed
          plt.figure(figsize=(8, 8))
          plotnumber = 1
          for column in num_df.columns:
              if plotnumber <= 4:</pre>
                   ax = plt.subplot(2,2, plotnumber)
                   sns.histplot(x = num_df[column], kde=True, bins=20)
                   plt.xlabel(column)
              plotnumber += 1
          plt.savefig('Numerical Features distribution.png')
          plt.show()
           200
                                                        400
                                                        350
           175
                                                        300
           150
                                                        250
           125
                                                        200
           100
                                                        150
             75
            50
                                                        100
            25
                                                         50
              0
                                                                         10000 15000 20000 25000
                                                                  5000
                      100
                            200
                                    300
                                           400
                                                  500
                0
                                                             0
                               distance
                                                                       fuel_consumption
           400
                                                         80
           350
                                                         70
           300
                                                         60
           250
                                                         50
           200
                                                         40
           150
                                                         30
           100
                                                         20
            50
                                                         10
              0
                 0
                        20000
                                 40000
                                           60000
                                                             70
                                                                    75
                                                                          80
                                                                                 85
                                                                                        90
                                                                                               95
                            CO2_emissions
                                                                       engine_efficiency
```

```
In [17]: # Summary statistics of the dataset
  eda_summary = df.describe()

# Display the summary statistics
  eda_summary
```

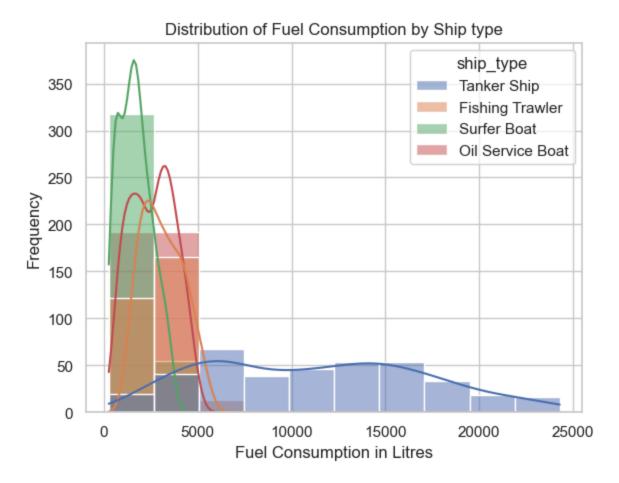
Out[17]: distance fuel consumption CO2 emissions engine efficiency count 1440.000000 1440.000000 1440.000000 1440.000000 mean 152.873660 4797.253375 13117.412993 82.382778 std 111.341855 5040.957635 13743.318993 7.211340 min 20.100000 261.750000 701.570000 70.000000 25% 76.040000 77.595000 1751.467500 4861.320000 50% 123.935000 2933.730000 8024.505000 82.620000 177.677500 88.245000 **75%** 4831.557500 13274.160000 499.960000 24289.870000 72029.760000 94.990000 max

```
In [18]: # Group Analysis
    # Average fuel consumption by ship type
    avg_fuel_by_type = df.groupby('ship_type')['fuel_consumption'].mean()
    print(avg_fuel_by_type)
```

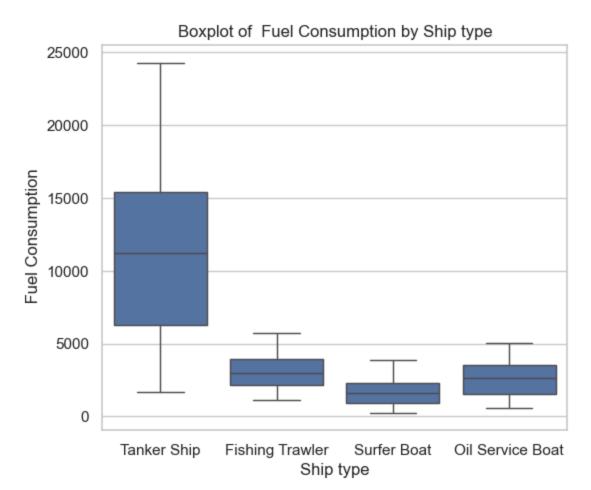
ship_type
Fishing Trawler 3126.153000
Oil Service Boat 2596.601562
Surfer Boat 1669.637984
Tanker Ship 11333.329766
Name: fuel_consumption, dtype: float64

```
In [19]: # Setting style for better Visulization
sns.set(style='whitegrid')

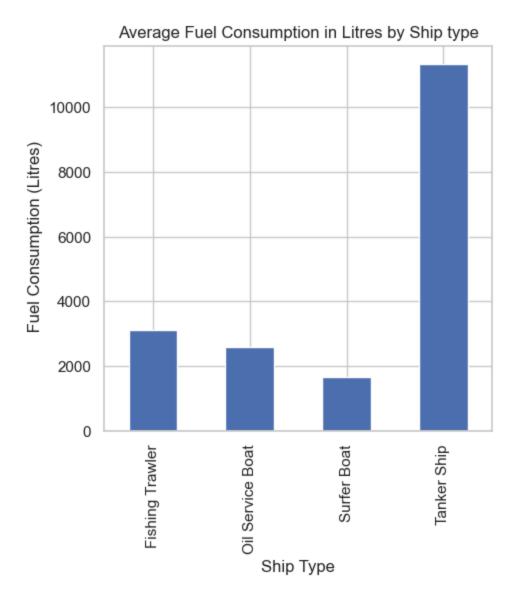
# Let check distribution of Fuel consumption by Ship type
sns.histplot(x=df['fuel_consumption'], hue=df['ship_type'], kde=True, bins=10, colo
plt.title('Distribution of Fuel Consumption by Ship type')
plt.xlabel('Fuel Consumption in Litres')
plt.ylabel('Frequency')
plt.show()
```



```
In [20]: # Checking for Outliers in Fuel Consumption by Ship type using boxplot
   plt.figure(figsize=(6,5))
   sns.boxplot(y=df['fuel_consumption'], x=df['ship_type'])
   plt.title('Boxplot of Fuel Consumption by Ship type')
   plt.ylabel(' Fuel Consumption')
   plt.xlabel('Ship type')
   plt.show()
```

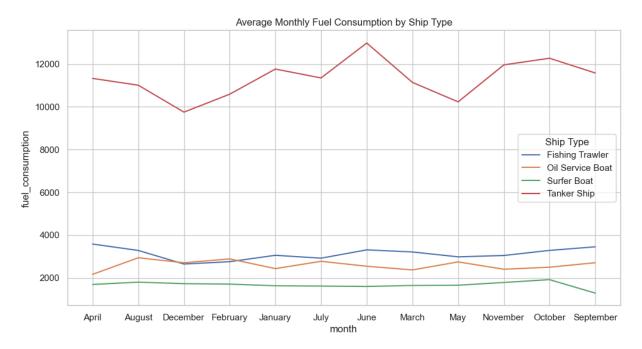


```
In [21]: # let check the average fuel consumption by ship Type distribution
    avg_fuel_by_type = df.groupby('ship_type')['fuel_consumption'].mean()
    plt.figure(figsize=(5,5))
    avg_fuel_by_type.plot(kind='bar')
    plt.title('Average Fuel Consumption in Litres by Ship type')
    plt.xlabel('Ship Type')
    plt.ylabel(' Fuel Consumption (Litres)')
    plt.show()
```

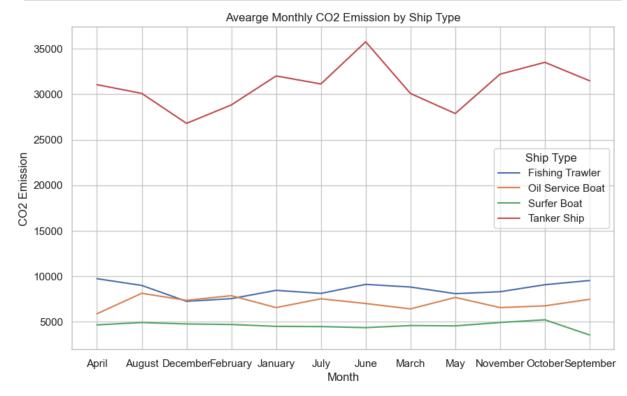


Insight From Fuel Consumption Analysis

Trend Analysis Over Months



```
In [25]: # Let plot for CO2 Emission
plt.figure(figsize=(10, 6))
sns.lineplot(data=monthly_trends, x='month', y='CO2_emissions', hue='ship_type', ma
plt.title('Avearge Monthly CO2 Emission by Ship Type')
plt.ylabel('CO2 Emission')
plt.xlabel('Month')
plt.legend(title='Ship Type')
plt.savefig('Avearge Monthly CO2 Emission by Ship Type.png')
plt.show()
```



Insights from the Visualizations:

CO2 Emissions:

- Tanker Ships consistently emit the most CO2 across all months, reflecting their larger size and fuel consumption.
- Other ship types, such as Fishing Trawlers and Oil Service Boats, have relatively steady and significantly lower emission levels.
- The monthly pattern suggests emissions peak at certain times, potentially correlating with operational demands or seasonal factors.
- Observation: The peak emissions are noticeable during months like June and December, possibly due to increased cargo movement during mid-year and holiday seasons.

Fuel Consumption:

- The trend aligns with CO2 emissions, as Tanker Ships exhibit the highest fuel consumption.
- Surfer Boats have the lowest consumption, indicating their smaller size and more efficient operations for shorter distances.

Monthly variations in fuel usage could relate to shipping activities or environmental conditions.

Advanced Statistical Tests

• We will conduct Advanced Statistical Tests to confirm significant relationships

between variables by ANOVA (Analysis Of Variance) METHOD.

Let formulate Hypotheses:

- Null Hypothesis (H0): The means of fuel consumption or CO2 emissions are the same across different ship types or months.
- Alternative Hypothesis (H1): At least one group mean is different from the others.

```
In [68]: from scipy import stats
# We are using ONE-WAY ANOVE to compare Fuel Consumption across different Ship Type

f_stat, p_value = stats.f_oneway(
         df[df['ship_type'] == 'Surfer Boat']['fuel_consumption'],
         df[df['ship_type'] == 'Tanker Ship']['fuel_consumption'],
         df[df['ship_type'] == 'Oil Service Boat']['fuel_consumption'],
         df[df['ship_type'] == 'Fishing Trawler']['fuel_consumption']
)

print('ANOVA F-STAISTICS', f_stat)
print('ANOVA P-VALUE', p_value)

# Interpretation
```

```
if p_value < 0.05:
    print("Reject the null hypothesis: There is a significant difference in fuel co
else:
    print("Fail to reject the null hypothesis: No significant difference in fuel co</pre>
```

ANOVA F-STAISTICS 787.3154611035424 ANOVA P-VALUE 1.2549212037621648e-302

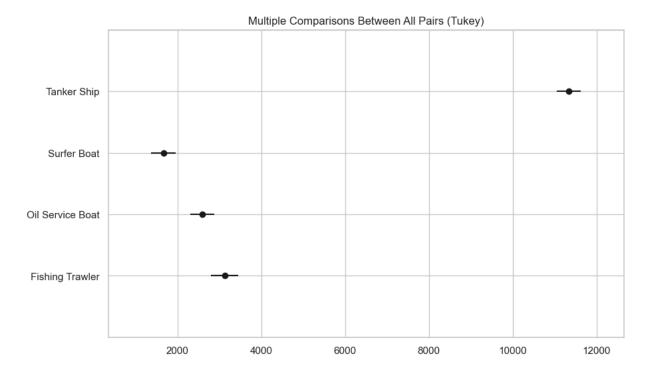
Reject the null hypothesis: There is a significant difference in fuel consumption be tween ship types.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

=======================================	======				-=======	-=======	======
group1	group2		meandiff	p-adj	lower	upper	reject
Fishing Trawler Oil						85.4115	
Fishing Trawler	Surfer	Boat	-1456.515	0.0	-2075.813	-837.2171	True
Fishing Trawler	Tanker	Ship	8207.1768	0.0	7592.2139	8822.1397	True
Oil Service Boat	Surfer	Boat	-926.9636	0.0002	-1507.5554	-346.3718	True
Oil Service Boat	Tanker	Ship	8736.7282	0.0	8160.7627	9312.6937	True
Surfer Boat	Tanker	Ship	9663.6918	0.0	9083.1	10244.2836	True

```
In [30]: # Visualizing the result
plt.figure(figsize=(8,8))
tukey_result.plot_simultaneous()
plt.savefig('Tukey Result Viz.png')
plt.show()
```

<Figure size 800x800 with 0 Axes>



Actionable Insights from ANOVA & TUKEY HSD

- Tanker ships are in a league of their own regarding fuel consumption, significantly outpacing all other ship types. This reflects their larger size and heavy-duty usage.
- Fishing trawlers and oil service boats show no significant difference, which may reflect similar operational patterns or efficiency levels.
- Surfer boats are the most fuel-efficient, making them ideal for shorter, high-frequency offshore transfers.

Analysis Final Report

1. Project Overview

This analysis investigates the fuel consumption and CO2 emissions of various ship types in Nigerian waterways. The goal is to identify patterns, relationships, and significant differences among ship types to aid in understanding fuel efficiency and emission impacts.

2. Key Insights from Descriptive Statistics

Average Fuel Consumption by Ship Type:

Fishing Trawler: 3066.51 liters
Oil Service Boat: 2526.64 liters
Surfer Boat: 1563.66 liters

Tanker Ship: 10734.95 liters

Observations:

- The tanker ship has the highest average fuel consumption, significantly outpacing other types, suggesting it operates on more fuel-intensive routes or has a larger fuel capacity.
- Surfer Boats, designed for passenger transfers, show the lowest fuel consumption, indicating efficient fuel use.

3. Trend Analysis Over Months

• Findings:

Monthly variations were noted in fuel consumption and CO2 emissions.

Patterns:

Certain months exhibited peak fuel usage and emissions, potentially tied to weather conditions or increased travel demands.

4. Correlation Analysis

Correlation Matrix:

- Strong positive correlation between fuel consumption and CO2 emissions.
- Negative correlation between engine efficiency and fuel consumption, indicating that higher efficiency could lead to lower fuel use.

Implication: The analysis confirms that as fuel consumption increases, CO2 emissions rise, supporting environmental and economic concerns for fuel-intensive ships.

5. ANOVA Test Results

ANOVA:

F-Statistic: 783.69P-Value: 9.82e-302

• Conclusion: There is a **statistically significant difference** in fuel consumption across different ship types, validating that ship type impacts fuel usage.

6. Tukey HSD Post-Hoc Test

Significant Differences:

- Fishing Trawler vs. Surfer Boat and Tanker Ship: Notable differences in fuel consumption.
- Oil Service Boat vs. Surfer Boat and Tanker Ship: Significant fuel differences.
- Surfer Boat vs. Tanker Ship: Also displayed significant differences.

Insights: The Tukey HSD test highlights that certain ship types, particularly tankers, differ from others significantly in terms of fuel consumption.

7. Visual Analysis

- Histograms & Boxplots: Provided insight into the distribution and variability of fuel consumption across ship types.
- Heatmap: Illustrated the strength of correlations between key variables.
- Trend Analysis Charts: Demonstrated monthly trends in fuel consumption and emissions.

8. Recommendations

Policy Implications

• Emission Regulations: Introduce policies limiting CO2 emissions based on ship type, with stricter limits for high-consumption ships like tankers.

Incentives for Cleaner Operations: Offer subsidies or tax breaks to encourage investment in eco-friendly technologies, such as cleaner fuels (e.g., LNG or biofuels) and hybrid-electric systems. Monitoring Systems: Enforce the use of emission monitoring devices to ensure compliance with environmental standards.

Operational Efficiency

• Engine Optimization: Schedule regular engine maintenance and prioritize upgrades to older systems to improve efficiency and reduce fuel usage.

Speed Adjustment Policies: Implement speed reduction programs, particularly for tanker ships on long routes, to minimize fuel consumption and emissions. Route Optimization: Utilize advanced navigation technologies for efficient routing, cutting down on unnecessary fuel consumption.

Environmental Initiatives

 Adopt Alternative Fuels: Pilot the use of cleaner fuel options like LNG or hydrogen, especially for fuel-intensive ships, to assess feasibility and infrastructure needs.

Energy-Efficient Designs: Encourage hull design innovations and retrofitting to enhance fuel efficiency, supported by government or private-sector funding. Awareness Campaigns: Partner with environmental organizations to educate operators on the financial and environmental benefits of reducing emissions and improving fuel efficiency.

Data-Driven Decisions

 Data Collection Systems: Equip ships with IoT devices to monitor fuel consumption and emissions in real time, enabling informed decision-making.

Predictive Analytics: Employ machine learning to predict fuel and emission outcomes under different scenarios (e.g., weather conditions, cargo loads) for proactive operational planning.

Focus on Specific Ship Types

 Tanker Ships: Prioritize interventions for tanker ships, such as mandatory engine upgrades or introducing reduced speed zones, due to their disproportionately high fuel consumption.

Surfer Boats: Use Surfer Boats as models of efficiency, sharing insights on design and operations with other ship operators.

Future Work

Exploring Predictive Modeling:

As an extension of this analysis, predictive modeling can be applied to forecast fuel consumption and CO2 emissions. Leveraging historical data, the following approaches could provide actionable insights for operational and environmental planning:

Time Series Forecasting:

- Techniques like ARIMA or Prophet could be used to model seasonal and monthly trends in fuel consumption and emissions.
- This would enable stakeholders to anticipate high-consumption periods and plan accordingly.

Regression Models:

- A regression model could predict fuel consumption based on factors such as ship type, route, engine efficiency, and weather patterns.
- Identifying key predictors could inform targeted strategies to enhance fuel efficiency.

Machine Learning Techniques:

- Advanced models like Random Forest or Gradient Boosting could uncover non-linear relationships and interactions among variables.
- This could be particularly useful for understanding how combinations of features (e.g., route and engine efficiency) impact emissions.

Benefits:

- Accurate forecasting could optimize resource allocation, improve fuel efficiency, and reduce emissions.
- It would provide a data-driven basis for policy decisions and operational strategies.

Limitations:

- Predictive modeling would require more data, particularly about external factors like weather, cargo weight, and operational schedules.
- Ensuring model interpretability would be crucial for practical implementation.

This future work could enrich the current project by transitioning from descriptive insights to actionable predictions, supporting both operational excellence and environmental sustainability.

In []: