Project Report: Maritime Carbon Footprint and Fleet Efficiency Analysis

TASK ID: EDA-6

---

1. Problem Statement and Objectives

Problem Statement

The maritime shipping industry is a major contributor to global carbon dioxide (CO_2) emissions. While regulatory frameworks such as the EU Monitoring, Reporting, and Verification (MRV) system ensure transparency, the reported data is primarily compliance-oriented and not directly actionable. The key challenge is transforming raw regulatory data into analytical insights that enable **comparative vessel efficiency assessment** and **predictive emissions forecasting**.

This project addresses this challenge by developing an end-to-end analytical pipeline that ranks vessel efficiency, predicts emissions, and presents insights through an interactive dashboard.

Project Objectives

Category	Objective	Status
Data Analysis	Rank individual vessels and fleet segments using a normalized efficiency metric (CO_2 per nautical mile).	✅ Completed
ML / Prediction	Develop and validate a regression model to forecast annual CO_2 emissions.	✅ Completed
Delivery	Build an interactive Streamlit dashboard for analysis and model validation.	✅ Completed

---

2. Dataset Details

A. Primary Data Source

- **Source:** European Union Monitoring, Reporting, and Verification (EU MRV) System
- **Format:** Cleaned and aggregated CSV (eu_mrv_cleaned_sample.csv)
- **Coverage:** Over **1,200 unique vessels**
- **Key Features:**
    - IMO Number
    - Ship Type

- Distance Sailed (Nm)
- Deadweight Tonnage (DWT)
- Fuel Consumed (tonnes)
- CO_2 Emissions (tonnes)

B. Secondary Data Products (Generated Outputs)

File Name	Purpose	Key Metric
step9_full_ship_efficiency_ranking.csv	Stores derived efficiency scores for all vessels	CO_2 per nautical mile
step10_model_predictions.csv	Stores model validation results	Actual vs. Predicted CO_2

---

3. Methodology and Implementation

A. Data Processing & Feature Engineering

- **Aggregation:** Data aggregated at the **IMO (ship) level** to ensure consistent annual performance metrics.
- **Efficiency Metric:**

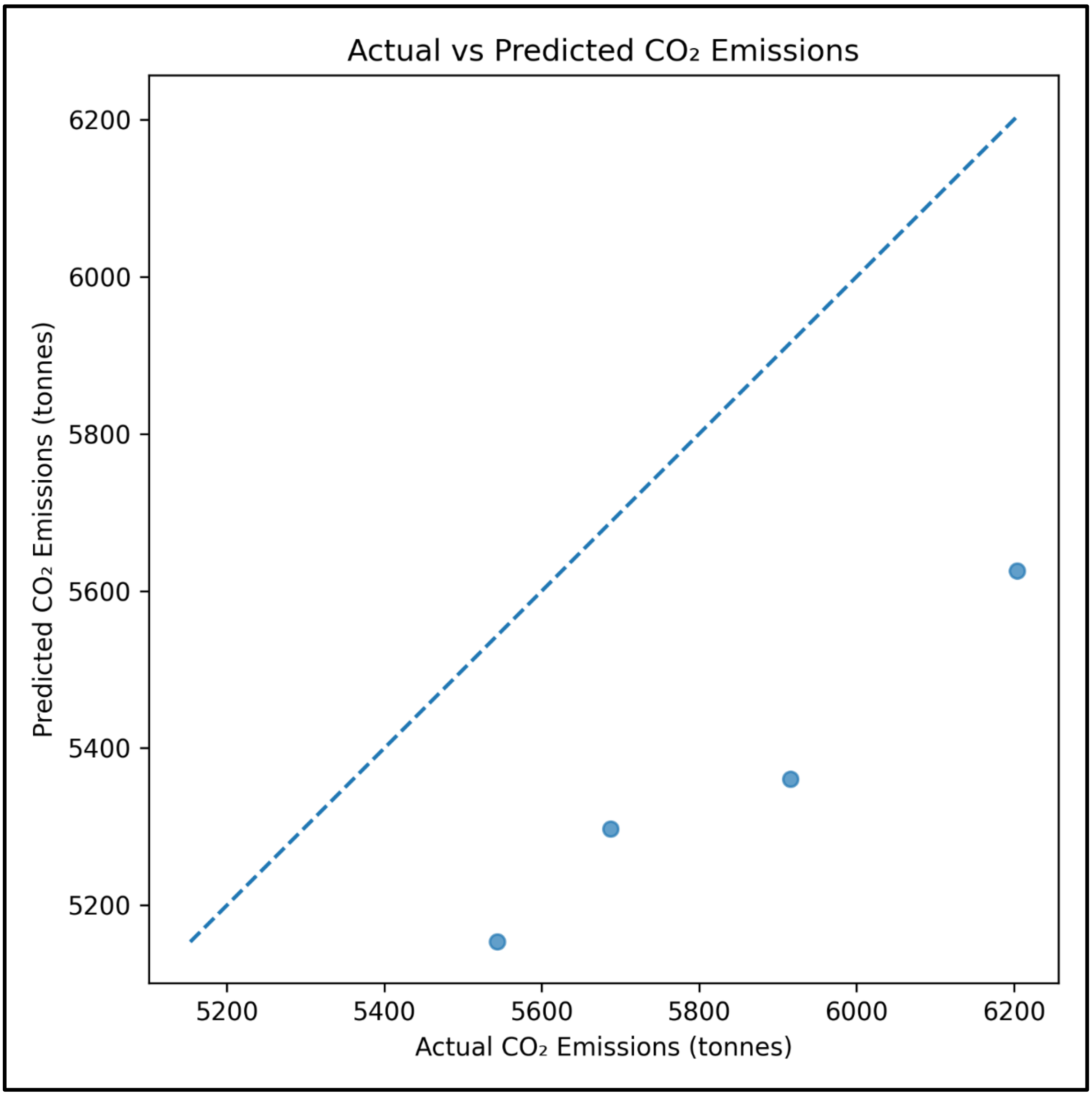
$$CO_2 \text{ per nm} = \frac{\text{Annual } CO_2 \text{ Emissions (tonnes)}}{\text{Distance Sailed (Nm)}}$$

- **Preprocessing:** Null handling, numeric type enforcement, and categorical encoding of ShipType.

B. Predictive Modeling (Linear Regression)

- **Model:** Linear Regression (scikit-learn)
- **Rationale:** Chosen for interpretability and its suitability for strongly linear relationships between fuel consumption and CO_2 emissions.
- **Features:**
  - DistanceNm
  - DWT
  - Fuel_tonnes
  - One-hot encoded ShipType

- **Validation Strategy:** Train-test split with performance evaluated on held-out test data.



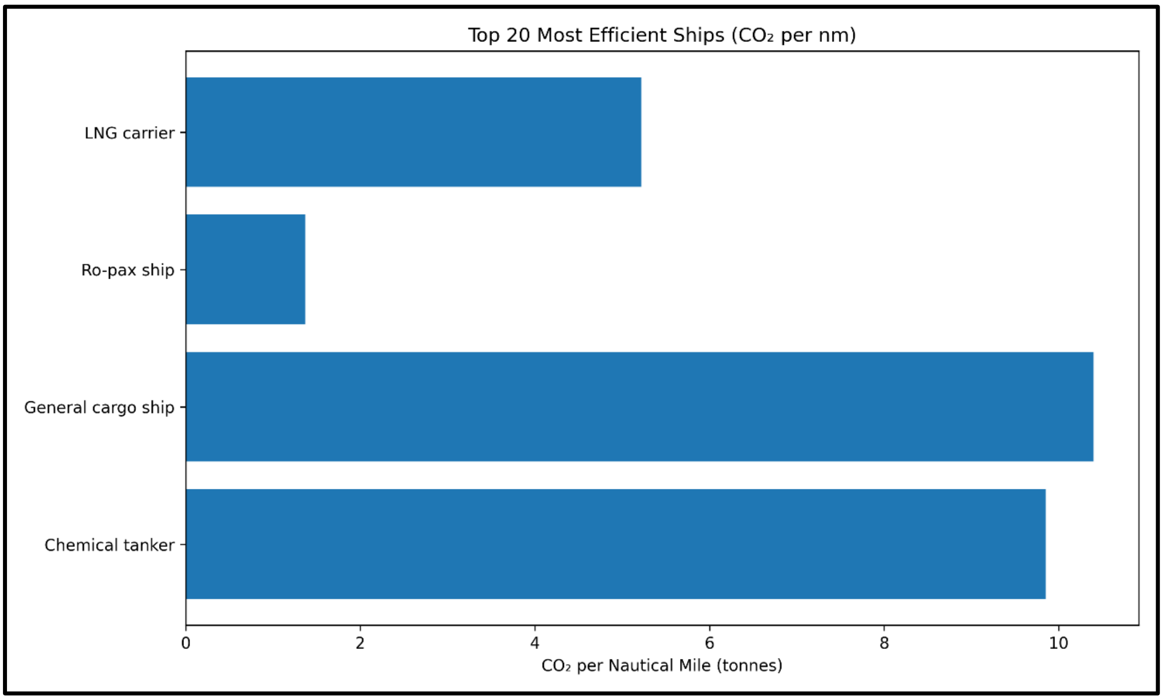
C. Dashboard Implementation

- **Framework:** Streamlit + Plotly
- **Layout:**
    - KPI summary row
    - 2×2 analytical chart grid
- **Interactivity:**
    - Ship-type filtering via sidebar
    - Data scope toggle (full fleet vs. ML validation subset)
- **Objective:** Provide a unified interface for efficiency benchmarking and model evaluation.

4. Results and Discussion

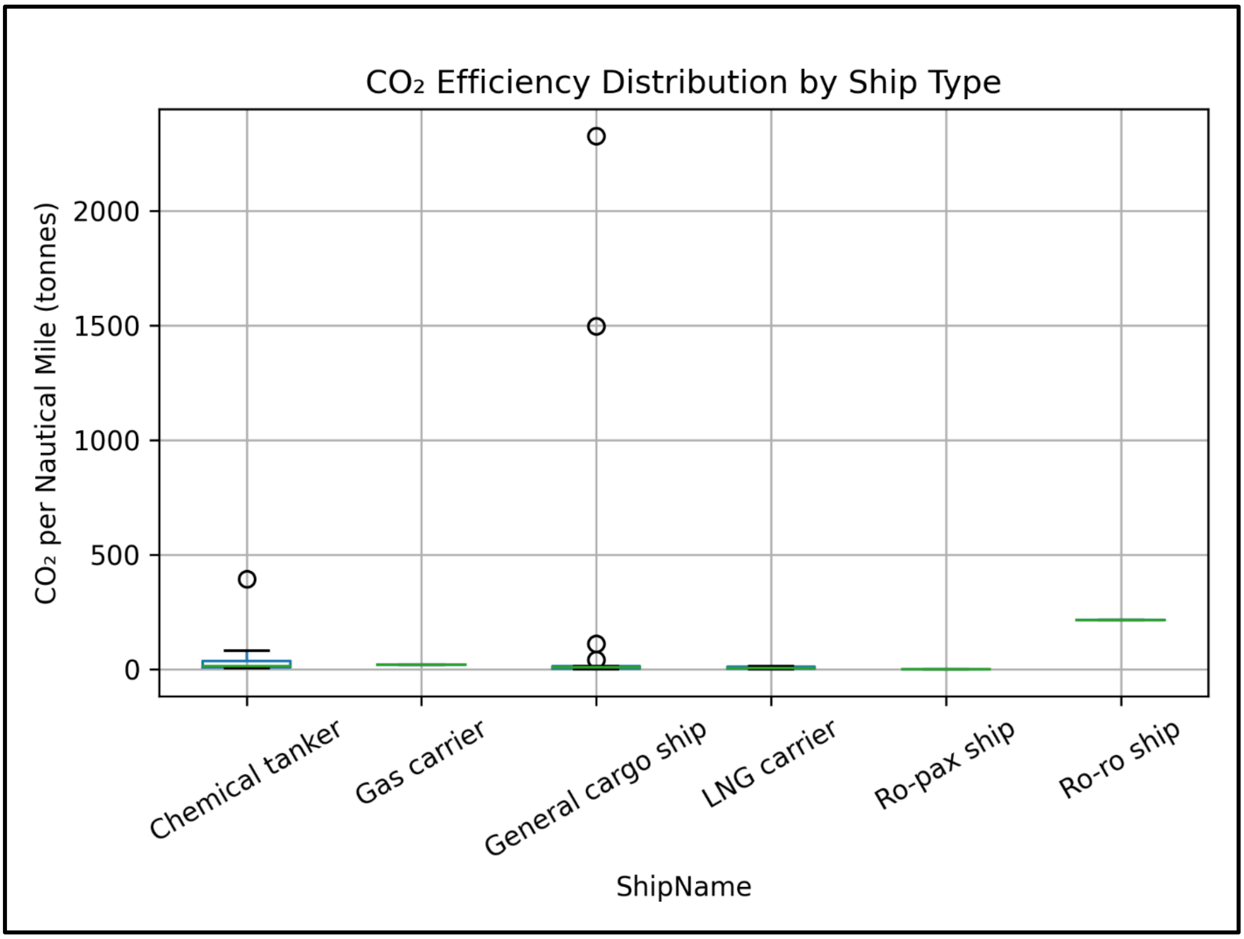
A. Efficiency Ranking Insights

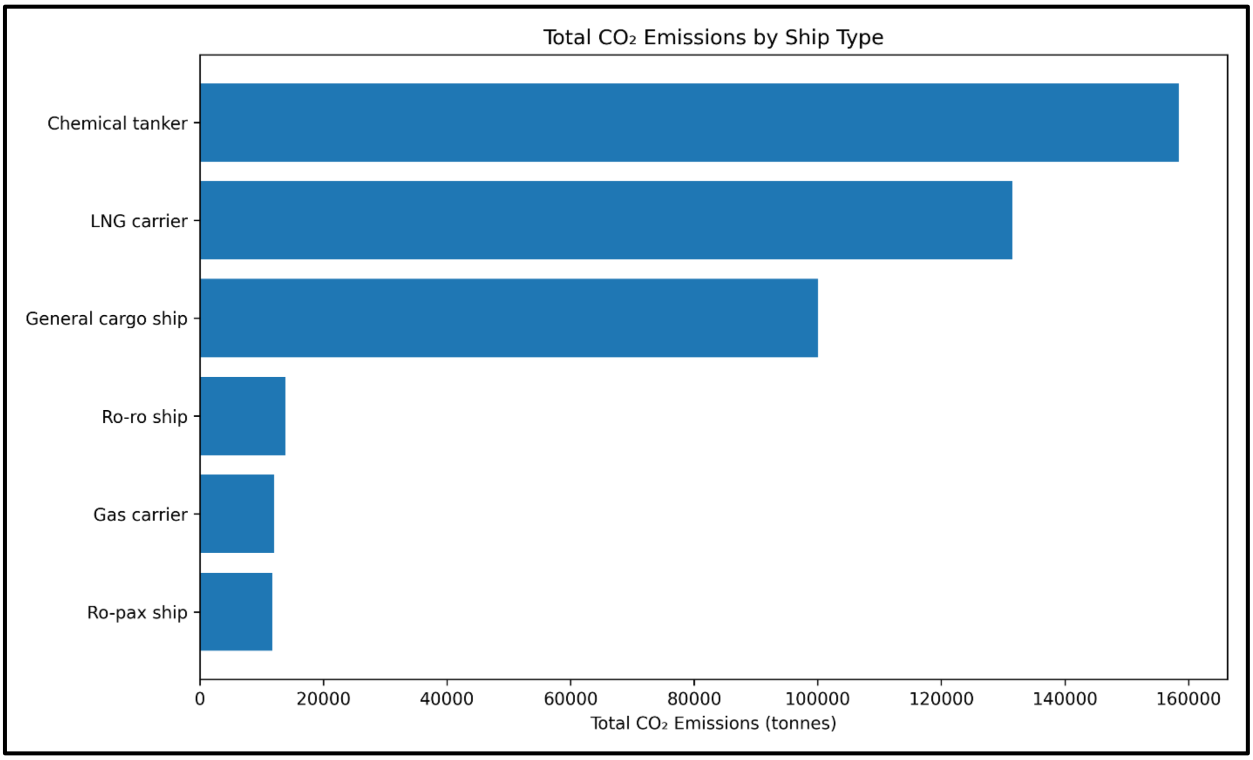
- **Top Performers:** Several vessels exhibit significantly lower CO_2 per nm, indicating optimized operations and/or modern propulsion systems.
- **Segment Variability:** Box-plot analysis reveals wide efficiency dispersion within certain ship types (notably chemical tankers and general cargo vessels), suggesting strong potential for targeted decarbonization efforts.



B. Linear Regression Model Performance

Metric	Value	Interpretation
R^2 Score	**0.6149**	Model explains ~61% of variance in CO_2 emissions
Mean CO_2 (Test Set)	**4,109 tonnes**	Baseline emissions level
RMSE	**1,664.8 tonnes**	Average prediction deviation
Relative Error	**40.51%**	Indicates scope for improvement
Prediction Accuracy	**≈ 59.49%**	Derived from RMSE-to-mean ratio


CO₂ Efficiency Distribution by Ship Type
CO₂ per Nautical Mile (tonnes)
0
500
1000
1500
2000
Chemical tanker
Gas carrier
General cargo ship
LNG carrier
Ro-pax ship
Ro-ro ship
ShipName


Total CO₂ Emissions by Ship Type
Chemical tanker
LNG carrier
General cargo ship
Ro-ro ship
Gas carrier
Ro-pax ship
0
20000
40000
60000
80000
100000
120000
140000
160000
Total CO₂ Emissions (tonnes)

Validation Analysis:

Despite the limited validation sample caused by ship-level aggregation, the Actual vs. Predicted CO_2 scatter plot shows clustering around the ideal $y = x$ line, confirming that the model captures the dominant emissions drivers.

---

5. Conclusion and Future Scope

Conclusion

This project successfully transforms static maritime compliance data into actionable analytical intelligence. By combining vessel-level efficiency ranking, regression-based emissions prediction, and an interactive Streamlit dashboard, the solution provides a practical foundation for fleet monitoring, regulatory analysis, and strategic decarbonization planning.

Future Scope

Area	Recommendation	Rationale
Feature Expansion	Integrate real-time operational data (speed, draught, weather)	CO_2 emissions are highly sensitive to operational conditions
Advanced Modeling	Implement panel or hierarchical regression models	Better captures vessel- and segment-level heterogeneity
UX Enhancement	Add threshold-based alerts for inefficient vessels	Improves usability for compliance and operations teams