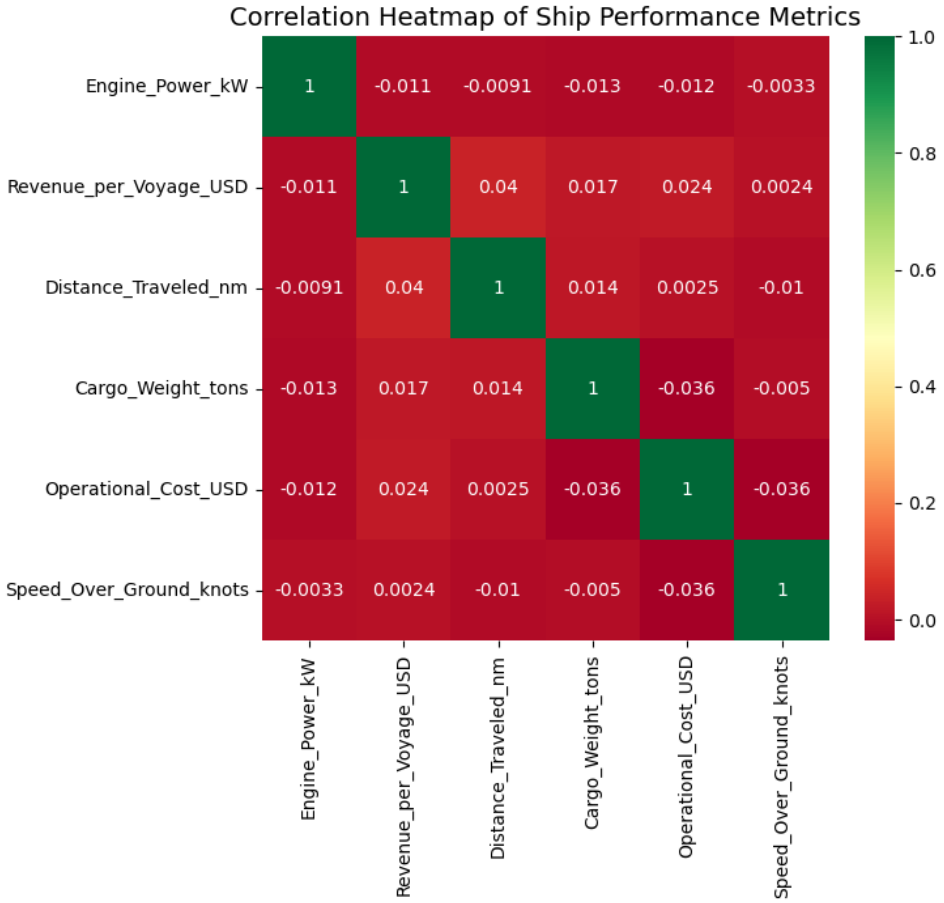
Goal: -

My dataset is primarily focused on the performance of the ships and the factors affecting it. These ships are carrier ships which are used for transporting goods. In this following data visualization assignment, I have tried to visualize the factors affecting revenue. To achieve this, I have used Correlation matrix to compare each element with revenue table.

Image of visualization: -



Insights: -

- Chart Type: Correlation Heatmap.
- **1.0** → Perfect positive correlation (variables increase together), **-1.0** → Perfect negative correlation (one increases, the other decreases).
- Correlation heatmap shows relatively weak correlations between revenue and the other ship performance factors.
- **Revenue_per_Voyage_USD vs. Distance_Traveled_nm (0.04)**: This suggests a very weak positive correlation, means traveling longer distances might have a slight impact on revenue, but it's not a strong factor.
- **Revenue_per_Voyage_USD vs. Cargo_Weight_tons (0.017)**: This suggests a very weak positive relationship, indicating that heavier cargo loads may contribute slightly to revenue but not significantly.

- **Operational_Cost vs. Cargo_Weight vs.Speed_Over_Ground_knots(-0.036)**: This negative correlation suggests that increasing cargo weight and Speed may slightly reduce operational costs, which could be an interesting point for increasing revenue.

Data Abstraction: -

This dataset focuses on analyzing the performance of carrier ships and the factors affecting their efficiency. It includes various operational and environmental factors affecting the ship performance.

- ❖ Key Attributes: -
    - Operational_Cost_USD: - It represent the operational cost to operate each ship which is measured in USD.
    - **Revenue_per_Voyage_USD: -** Contains data of revenue earned by each ship which is measured in USD.
    - **Distance_Traveled_nm: -** This collum represents the distance travelled by each ship.
    - **Cargo_Weight: -** It represent total weight carried by each ship in different time period and weather condition.
    - **Speed_Over_Ground_knots: -** It stores information on speed for each ship**.**
    - **Engine_Power_kW: -** Contains information about power generated by different engine.

Task Abstraction: -

By analyzing the dataset, I have used various attributes which was present in dataset to map out the factors affecting revenue. This allows me to understand various factors affecting revenue and how it will affect if we increase the load or speed and also provide us with important insights on data.

- ❖ **Channels: -**
    I. Position: - Both
    II. Hue
- ❖ **Mark: -**
    I. Area
- ❖ **Datatype: -**
    I. Categorical
    II. Quantitative

Data Source: -

https://www.kaggle.com/datasets/jeleeladekunlefijabi/ship-performance-clustering-dataset

Code: -

```
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns

# Load the dataset
file_path = "Ship_Performance_Dataset.csv"
ship = pd.read_csv(file_path)

#Shows first five rows
ship.head()

#drops the cell with null values
ship.dropna()

# Compute the correlation matrix
selected_columns = ["Engine_Power_kW", "Revenue_per_Voyage_USD", "Distance_Traveled_nm","Cargo_Weight_tons","Operational_Cost_USD","Speed_Over_Ground_knots"]
correlation_matrix = ship[selected_columns].corr()

#To plot the HaetMap
plt.figure(figsize=(7, 6))
plt.title("Correlation Heatmap of Ship Performance Metrics", fontsize=14)
plt.xticks(rotation=45)  # Rotate column names for better readability
plt.yticks(rotation=0)
sns.heatmap(correlation_matrix,cmap="RdYlGn",annot=True)
plt.show()
```