**Case Study: Optimizing Operational Performance of a Maritime Fleet Using Clustering**

### **Background**

A logistics company manages a diverse fleet of ships transporting goods across various routes. Despite access to detailed operational data, the management lacks clarity on which types of voyages are cost-effective, which need optimization, and how operational patterns relate to ship performance.

### [**Dataset**](https://drive.google.com/file/d/1HkGik5xBe9rRdjjsGksPvXxUKU2ZsdzQ/view?usp=sharing)

### **Business Problem**

**"Can we segment ship voyages into meaningful operational clusters to optimize cost, improve voyage planning, and identify underperforming fleet segments?"**

You are tasked with building a **K-Means clustering model** that helps classify voyages into different performance groups - such as cost-efficient, high-revenue, underperforming, etc. These insights will help the business:

* Reduce operational inefficiencies
* Adjust cargo distribution and route planning
* Recommend ship-specific improvements
* Make data-driven decisions on engine types, routes, and maintenance

### **Dataset Columns Explained**

| **Column** | **Description** |
| --- | --- |
| Date | Date of the voyage |
| Ship\_Type | Type of ship (e.g., Container, Bulk Carrier, Fish Carrier) |
| Route\_Type | Nature of the voyage (Short-haul, Long-haul, Transoceanic) |
| Engine\_Type | Type of engine used (e.g., Diesel, Steam Turbine, HFO) |
| Maintenance\_Status | Current condition of the ship’s maintenance (Good, Fair, Critical) |
| Speed\_Over\_Ground\_knots | Ship’s average speed during voyage |
| Engine\_Power\_kW | Power output of the ship’s engine |
| Distance\_Traveled\_nm | Distance of the voyage in nautical miles |
| Draft\_meters | Vertical distance between waterline and bottom of the hull |
| Weather\_Condition | Weather during voyage (Moderate, Rough, etc.) |
| Cargo\_Weight\_tons | Cargo carried in tons |
| Operational\_Cost\_USD | Cost incurred for the voyage |
| Revenue\_per\_Voyage\_USD | Revenue generated from the voyage |
| Turnaround\_Time\_hours | Total time the ship took for the voyage including port delays |
| Efficiency\_nm\_per\_kWh | Distance efficiency based on power consumed |
| Seasonal\_Impact\_Score | Environmental or seasonal impact score |
| Weekly\_Voyage\_Count | Number of voyages by this ship in a week |
| Average\_Load\_Percentage | Average load percentage vs max capacity |

### **Your Task**

You are to:

1. **Preprocess** the dataset:  
   * Handle categorical variables via encoding.
   * Normalize numerical features.
   * Handle missing values/Duplicates/Errors if any.
2. **Select features** relevant to operational performance (hint: cost, revenue, load, efficiency, cargo weight, route type, engine power).
3. **Apply K-Means Clustering**:  
   * Use the Elbow Method and Silhouette Score to find the optimal number of clusters.
   * Visualize clusters using PCA or t-SNE.
   * Interpret each cluster (e.g., "Cluster A = High Efficiency, Low Cost").
4. **Deliver Insights**:  
   * Which cluster is most cost-effective?
   * Which ship or route types appear in underperforming clusters?
   * How can these insights help optimize future voyages?

### **Bonus Marks**

Add a section in your report titled **"Business Recommendations"** where you suggest:

* Which routes to optimize or discontinue
* Ideal ship configurations for maximum profitability
* Maintenance recommendations based on cluster trends

*This case study mirrors real scenarios in:*

* *Maritime logistics optimization*
* *Fleet management systems*
* *Data-driven supply chain efficiency*
* *Recommender systems for operational decisions*