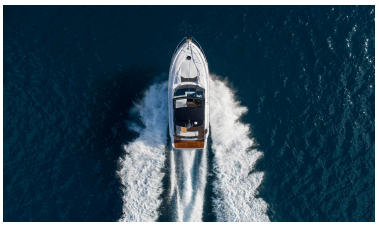
Business Intelligence project - BI &DBMS

- **Business Idea:** **Boat Selling Company**

- **GitHub Link:** https://github.com/Yosri-zayani/Boats_BI_project.git
- **Project Outline:**



1. Introduction
2. Main Phases
3. Conclusion

1. Introduction:

Boat sales have consistently constituted a thriving billion-dollar industry for many years.
Imagine being able to analyze the sales patterns for different features for most boats and approach the company with a thorough report.
The datasets we used contain information regarding the sales of boats across various territories, while also giving information regarding the Boat types, features, and external factors that may affect sales.
→This report will support the HR department in making strategic decisions about its workforce.

After our thorough analysis, we will answer the following questions:

- What are the boat types that are making top 5 sales?
- Are there underperforming boats that need improvement?
- Which territory has performed best in terms of sales?
- Do we need to open new stores or close or transfer poorly established ones?
- Does the building year of a boat affect its sales?
- Should we end our business relationships with certain manufacturers?

- Is there a positive correlation between the GDP of a country and its presence in the boating market?

2. Phases Implemented :

- **Phase1 :**

Data Gathering:

We gathered our data from the website kaggle.com.
The dataset belongs to a boat-selling company called AquaDreams and contains the sales history of the company.
We gathered two types of datasets: external and internal.
The external dataset contains information about the external factors that may affect the company's sales across various regions.
We collected the data in its original raw format: JSON and CSV.

- **Phase 2:**

Extract/Transform/Load Process:

The extract/transform/load process was done using TALEND.
We integrated the necessary information to answer our previous questions.
We used different functions to transform our data accurately: Tmap, Aggregate, Tsort...

- **Phase 3:**

Data Storage/Modeling :

In this phase, we focused on conceptualizing our data and the relationships between different data elements.
We identified the facts, dimensions, and attributes.
We used Power BI Desktop to come up with the star schema of our model.
We opted for multidimensional modeling for its scalability and enhanced analysis capabilities.

- **Phase 4:**

Dashboard using PowerBI :

In this final phase, our objective is to transform vast and intricate data from prior stages into an intuitive, actionable format. To achieve this, we opted for Power BI as our specialized Business Intelligence tool.

- **Phase 5:**

Conclusions and Recommendations :

→In the conclusions and recommendations phase of our boat-selling data analysis portfolio, we distill key insights gleaned from the comprehensive analysis of our datasets. By leveraging advanced analytics and visualization tools, we uncover patterns, trends, and performance indicators that inform our understanding of market dynamics, customer behavior, and sales effectiveness. With these insights in hand, we formulate strategic recommendations aimed at optimizing our boat-selling operations. Whether identifying untapped market segments, refining marketing strategies, or enhancing inventory management, our conclusions guide actionable decisions. This phase serves as a crucial bridge between data analysis and informed decision-making, empowering our boat-selling company to adapt, thrive, and excel in a dynamic market landscape.

3. Thorough Steps with Code:

- **Phase1: Gather data**

Data sources(Boat.csv,GPDapi)

- **Phase2: ETL**

1. Extract data
2. Data cleaning
    - Handling missing data (remove empty rows OR fill missing values with specific values): Location(blanks)- Material(blanks)- Width(blanks)- Length(blanks)- Year built (0 values)- Manufacturer(blanks)- Type(blanks) :
    - print(df.isnull())--> to check null values
    - print(df.isnull().sum()) →Number of null values per column
    - df=df.dropna() →remove rows with missing values df=df[df[‘Year Built’]] != 0] → remove rows that contain 0 values
    - Handling duplicates (remove duplicate rows → df=df.drop_duplicates()
    - Dropping unnecessary columns → df=df.drop(columns=’Number of views last 7 days’) print(df))

- Text cleaning (remove leading and trailing whitespaces(/,..,-,) from strings and lowercasing the ‘type’ column→ df[‘Type’]=df[‘Type’].str.lower()
- Display cleaned data frame→ df.to_csv(r'boat_data.csv')

- **Phase3: Data Storage and Modeling**

We stored our Data in an SQL Database.
We used a star schema for the modeling.

- **Phase4: Data Analysis**

We analyzed the output and concluded.
We then provided significant recommendations .

Key Performance Indicators :

- **Boat type**
  1. Sales per boat type (what are the most used boats, Top sellers' models and what are the underperforming boats that need improvement).
  2. Revenue per boat type.
- **Manufacturer**
  1. Are there outperforming manufacturers (manufacturers that contribute the most to our sales)?
  2. Should we keep our business relationships with certain manufacturers?
- **Year built**
  1. Analyze the age of the inventory (does the age of the inventory affect the sales).
- **Material**
  1. Sales per material (Most preferred boat materials).
- **Location**
  1. Sales per country (identify high-performing regions, which territory has performed best in sales?).

- **External factors affecting sales**
    1. GDP per capita in each country.