**Motivation and Purpose**

The marketing team is preparing a weekly newsletter for boats owners. The newsletter is designed to help sellers to get more views of their boats, as well as stay on top of market trends.

They would like to take a look at the recent data and get some insights. The possible questions that we can ask ourselves is:

- Characteristics of the most viewed boat listing in the last 7 days

- What is the price of boats that get the most views?

-  Are there common features among the most viewed boats?

- What is the location of the most viewed boats?

**Overview**

A Total of **6,001 boats and yachts** have been registered in **50 countries** and territories, over 90% of which are in Europe.

**Germany** ranks first with **1209 boats and yachts**, followed by **Italy** with **918 boats** and **Switzerland** with **900 boats and yachts**.

In fact, all the top 10 countries by the number of registered boats are in Europe.

This can be explained by the high income level of Europeans, access to the sea and a love of adventure and relaxation and sailing (Switzerland for example).

**Exploratory analysis**

For the sake of exploration we began by looking at linear relationships between variables.

We found there is **no linear dependence between year built and view** and also **no linear relationship between length/ width and views**. We therefore attampt to see how a combination of variables may affect the number of views.

This graph represents the relationship between boat length and price. There seems to be a **connection between length and price**.

The upwards trend leads us to the following **hypothesis** - **the larger the length of the boat the higher the price.**

**Regression analysis**

***The larger the length of the boat the higher the price.***

To test this hypothesis we conducted **a linear regression**.

The results of this regression showed that the length only contributes to around 38% of the trend in the data meaning the relationship is not entirely linear. There are many points that fall beyond the regression line especially as the boat size reaches the larger end of the spectrum.

This regression alone is not enough to explain the trends in the data so we try another approach.

**cluster analysis.**

Another approach used to try and explain the trends in data was cluster analysis.

Cluster analysis groups data into clusters. We can compare the groups to discover new patterns.

Our cluster analysis yeilded 5 distinct groups of data points, whis you can see represented in different colours on this plot, **(white, steel blue, light blue, mid blue and dark blue).**

The cluster 2 returned best stats out of the lots. The **median number of lengths**, **width** and **price is highest** and significantly larger than the other four. This is interest has length and width would determine the size of a boat and this could be proportionate to price.

I say a minimal impact as this relationship is very unclear in the visualistion without looking at it in depth. The 3nd claster has the most views in the last 7 days while the boats greater in length (2 and 4) have a tendancy to the lesser amount of views.

**How big are they?**

Average Long-Width of Boats and Number of Views

The most viewed boats in terms of length and width have nothing to do with its dimensions

**Most vieved - most expensive**

The material of boats or yachts with the highest view is the Glass Reinforced Plastic (GRP) with over 700 thousand views. This can be as a result of its strong, extremely light and highly versatile nature. At a distance place is Polyvinyl Chloride (PVC), a synthetic polymer of plastic with over 84 thousand views and Steel with about 60 thousand views.

**conclusion**

When creating a listing the follwoing should be considered:

The price of the boat will be somewhat dependent on the size (larger the boat higher the price). A combination of price and size will only have a limited impact on the number of views.

There are a number of variables we did not analyse that would impact the number of views a listing gets. These variables include the material, boat type and manufacturer. Further anlysis on these variables would have to be conducted to have a more complete picture of what users of the boat selling platform could do to their listings to get maximum views.

1. **Germany, Switzerland, Italy, France, Netherlands** are countries with the most boats, making up more than 50% of all listings. Therefore the newsletter should be catered for these regions.
2. **Motor yachts** and **sport boats** are the most popular amongst boat types.
3. **Most** popular manufacturers are **Beneteau**, **Jeanneau** and **Sunseeker.**
4. **Cheaper** and **smaller** boats tend to get more views.

Caveats:

* Number of views in last 7 days may not be an accurate measure of popularity for all listings. It’s possible that a very popular boat gets sold in less than 7 days, thus having a lower view count, therefore under counting the views.
* Cheaper boats get more views because they are more affordable and more frequently purchased. However, more expensive boats are more representative of the latest boat technology and design, so its possible they will attract more readers in a newsletter.
* Other unknown factors may have a strong influence on views, and hence popularity, but is not recorded in the dataset. e.g. mileage of a boat
* **Everything was done in Python**
  + Pandas and NumPy for data cleaning and data wrangling
  + Matplotlib and Seaborn for data visualization
* **Initial data cleaning and wrangling**
  + The price column was in string format and in different currencies
    - Created a new column that has all the prices in Euros and in integer format
  + Year column had 0 for NaN values
  + Location column had non-English characters
    - The strings were replaced, and the countries data were extracted from this column
  + Boat Type had redundant values

Price had to be separated into 10 percentile groups

The longer the boat, the higher the price.

- Slight correlation between newer boat models costing higher than older boats.

- Sport Boats and Yachts with GRP material are the most viewed boats.

- Switzerland and Germany together constitute 56% of total views from all countries

**Limitations**

- As there are no actual sale values, we are unaware if higher views relates to actual sales.

- Therefore a recommendation on change in stock or marketing campeigns cannot be made.