Task 6.1: Sourcing Open Data

**Boat Sales**

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

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

1. characteristics of the most viewed boat listings in the last 7 days
2. is it the most expensive boats that get the most views?
3. Are there common features among the most viewed boats?

**Data Source**

Open sourced data through Kaggle

<https://www.kaggle.com/datasets/karthikbhandary2/boat-sales?resource=download>

**Data Processing**

1. Structured data set to be uniform
   * Price – renamed to Eur\_price
     + Restructured price to be uniform
     + Split price and currency
     + Recalculated price at current EUR exchange rate for each and renamed column Eur\_price
     + Dropped price and currency columns
   * Location – split into country, city, other
     + Dropped Location and other columns
2. Cleaning processes
   * Removed missing data in the following columns:
     + Type
     + Length
     + Country
     + Width
     + City
   * Renamed missing values to none in material and manufacture columns since the volume was over 10%
   * Checked for duplicate data
   * Checked for mixed data types
3. Data profile:
   * Started with 9,888, 10
   * Ended with 8,893,11

**Data Limitations**

1. Locations – not all locations provided city, and country.
2. Currency – updated the currency at the current exchange rate to EUR. Since this can fluctuate, we may want to consider another process

|  |  |  |  |
| --- | --- | --- | --- |
| **Column** | **Description** | **Data Type** | **Time Variable** |
| Boat Type | Type of boat | Qualitative, Nominal | No |
| Type | Type of boat | Qualitative, Nominal | No |
| Manufacturer | Boat manufacturer | Qualitative, Nominal | No |
| Year Built | Year boat was constructed | Quantitative, Discrete | Yes |
| Length | Length of boat in meters | Quantitative, Continuous | No |
| Width | Width of boat in meters | Quantitative, Continuous | No |
| Material | Materials used to build the boat | Qualitative, Nominal | No |
| Number of views last 7 days | Number of times the boat was viewed in last 7 days | Quantitative, Continuous | Yes |
| EUR\_price | Price of boat in EUR | Quantitative, Continuous | No |
| Country | Country boat is currently located | Qualitative, Nominal | No |
| City | City boat is located | Qualitative, Nominal | No |

**Questions to Explore**

1. What characteristics of the most viewed boat listings in the last 7 days
2. Is it the most expensive boats that get the most views?
3. Are there common features among the most viewed boats?