

Task description

You have the two following .csv files:

- boats: each row is a boat with its specific features and location. Additionally, a charter ID is provided that links the boat to a charter
- charters: each row is a charter with its specific location and type of charter

Given the previous files and information, perform the following tasks using python (the packages pandas, numpy, and matplotlib/seaborn should be enough):

- Import the two files into two distinct data frames
- Explore both data frames by inspecting basic statistics about columns (pandas should be sufficient)
- Eventually, deal with null values (justify why you choose to do so, use pandas)
- Use specific columns (for example: boat type, location of boats and charters) to group the data and compare the different groups (use pandas and check for basic stats)
- Perform a visual analysis of both data frames (by using the full data and also inspecting different boat types/models) and point out if you see any interesting differences/patterns (use matplotlib/seaborn)
- Join the two data frames into a new one and rank the charters by number of boats in ascending order, showing the top 10

Optional:

- By using the boats dataset, try to see if by using specific columns (or all of them) you can infer the boat type. Hint: this is a classification problem, use sklearn and use a train set to validate your model

Output:

- Share the exercise as notebook/PDF/html file