Nearly New Nautical

Charting a course to the best boats

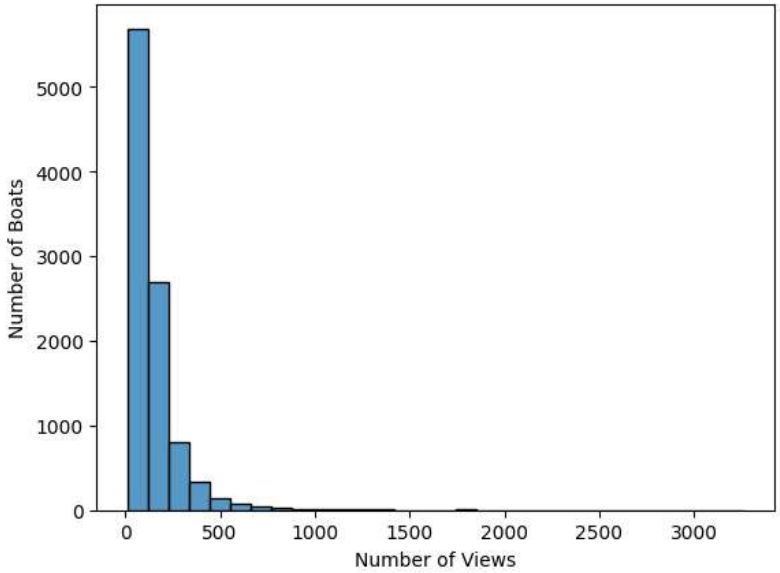
Report prepared by Ryan McLean,
Data Scientist

Nearly New Nautical strives to create a premium platform for boat selling and buying

Users list their boat, providing a range of information

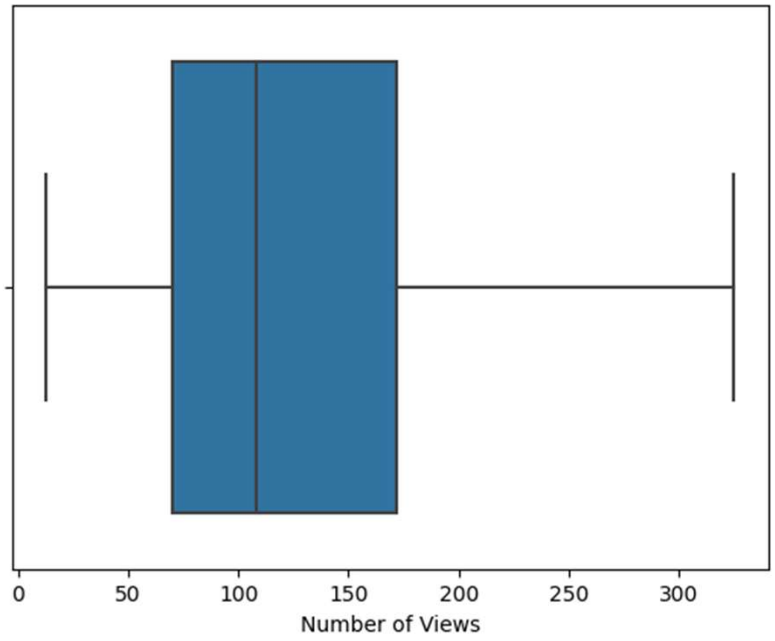
Boats that get lots of views drive traffic

Boats with few views create website noise that weakens the user experience



Right now, the boats on our site generate a wide range of views

50% of our boats are viewed between 70 and 172 times, with a median value of 108



Our plan forward

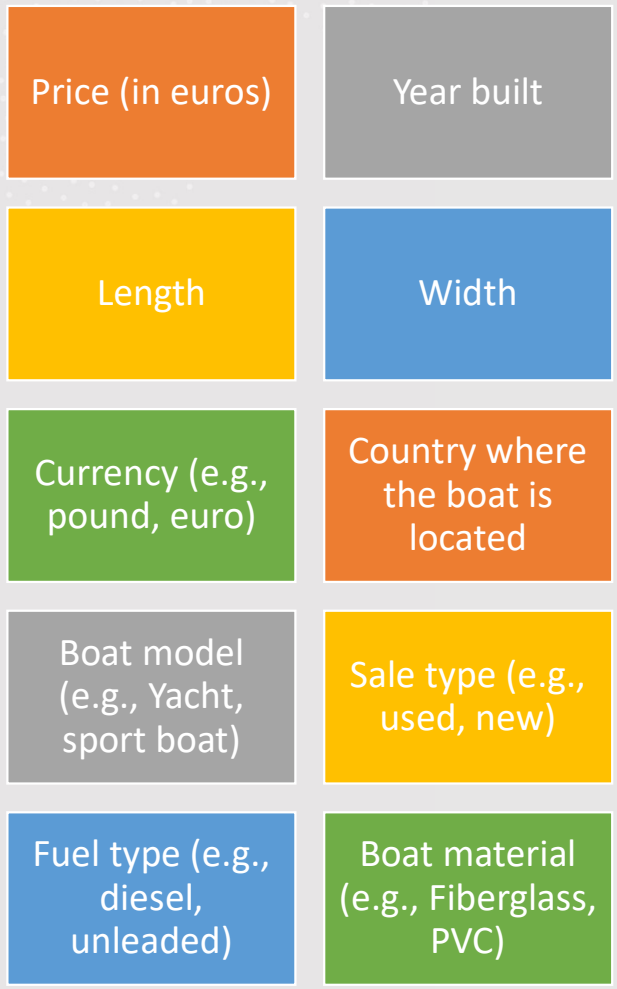
The data and analytics team will use machine learning to predict website traffic, giving us the power to post only boats that will increase traffic

Data Processing Steps

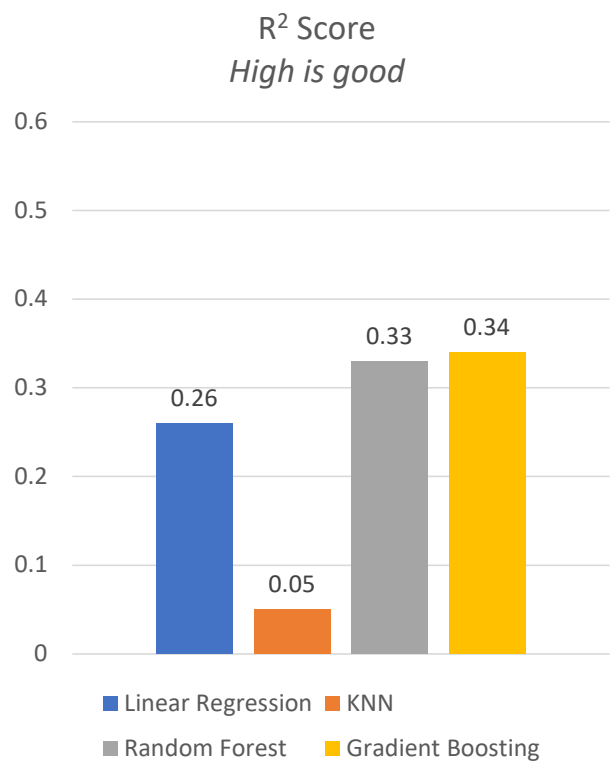
- Outliers: Taking out the most extreme boats so we get a model that fits most boats
- Missing values: To ensure a clean dataset, boats with insufficient information were removed
- Price changes: Since we post across multiple countries, I converted all boat prices to Euros for equal comparison
- Boat model: 56% of our boats fall into 3 different models, and 89% of boats are in just 10 models.
- Sale type: Categorized boat as used, new, or from a display
- Fuel type: Categorized boats into diesel, unleaded, and other energy sources
- Machine Learning Preprocessing: I also performed one-hot dummy encoding and variable scaling to ensure that the information was scaled and interpreted correctly by the mathematics

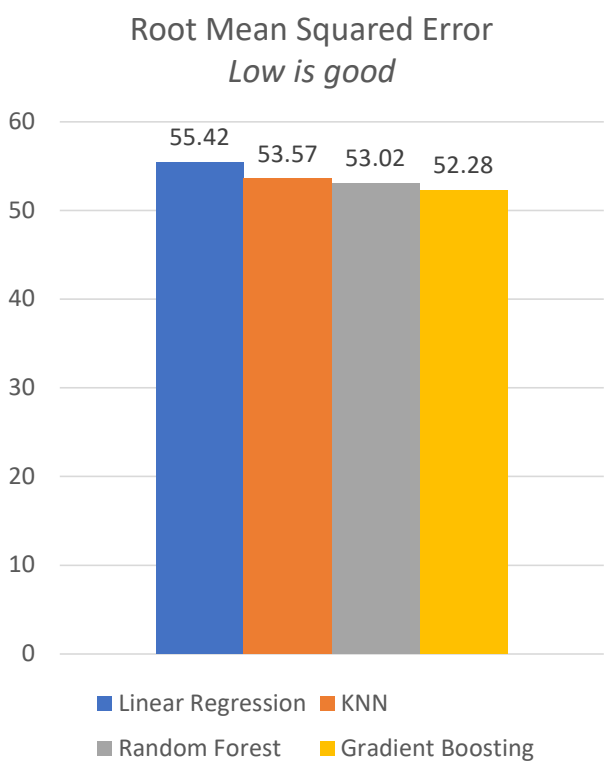
Final list of boat features that predict views

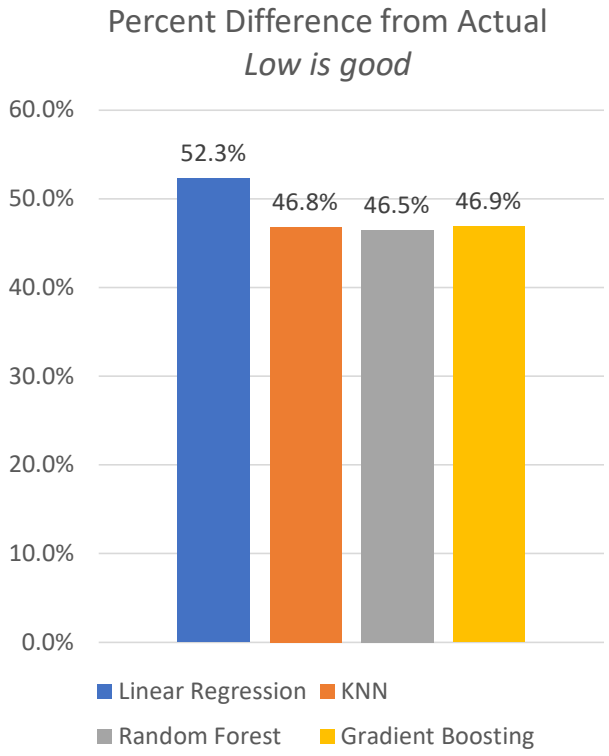
Model trained on 8,476 boats posted on our site



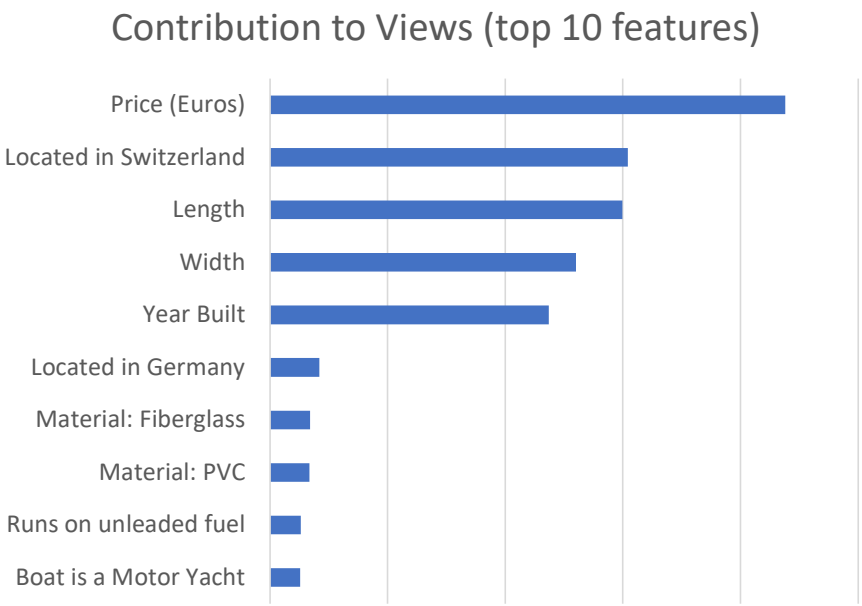
After evaluating multiple metrics, the Random Forest algorithm was chosen for best performance







Price is the strongest predictor of views, followed by location, size, and age.



Listings that are more *expensive*, located in *Switzerland*, have a *larger area*, and were built *more recently* tend to drive website traffic the most

Using this algorithm, we can predict within 46% the number of views a listing will have.

With this information, we can choose to focus our attention on providing the best options for our consumers.

Conclusion

Data Science will help us chart a course to the best boats