

Proposal for Project Luther

Predicting Used Sailboat Prices

Sean Davern

Seattle, Fall Cohort, weeks 2-3

[Boats.com](https://boats.com), founded in 1999, is the largest global search engine for boats in the leisure marine market. It is common for owners to need estimates of the current value of their vessels. Obviously, someone wishing to sell their boat or buy a specific boat will want to know its value. In this situation a detailed analysis and, likely, a survey would be in order. In other situations someone might wish to have a quicker, rougher estimate. For example, annual insurance renewals or when considering significant maintenance or refurbishment projects one might desire a less rigorous estimate. Separately, one might also want an estimate that compares across manufactures, models, age, location, or other factors with a model that facilitates real-time interactive exploration and might enable quick relative comparisons or graphical presentations.

One such method for obtaining estimates is the reputable [NADA Guides](#). This source requires one-off manual estimations of each boat. Note also, that the NADA values are based on historical data and are only sales via brokers and dealers, not private parties.¹

So, this project proposes to scrape listing prices and boat detail and evaluate the quality of regression model to predict list prices.

An MVP would scrape the roughly 1,000 of the 3,245 sailboats currently on Boats.com in the length range of 40-50 ft ([example](#)) for their listing prices and other data (below) and regress a model predicting list price. The model would enable assessment of prediction uncertainty to compare again the estimated 10% variance in NADA valuations.¹ The listing prices are not anticipated to reliably reproduce the NADA valuations since purchase prices may expected to be lower than listing prices. However, the relative magnitudes should be similar and if the relative uncertainties were similar that would be a significant success. Additionally, the MVP will not include processing of text information which could include listed electrical equipment, electronics, installed equipment, sails and other extras that can be scraped.

Data

Variable	Type	Description	Used for Model
List Price	float	Selling owners requested price in US\$	Target
Manufacturer	string	Boat manufacturer (Make)	Y
Model	string	Boat model (correlated with manufacture, length and displacement)	N
Year	int	Model year of boat	Y
LOA	float	length overall in ft	Y
Displacement	float	Displacement in lb	Y
Fuel type	string	Diesel or gasoline	Y
Power	float	Size of engine in hp	Y
Engine Hours	float	Measure of engine usage in hrs	Y
Engine Year	int	Year engine was manufactured	Y
State/Region	string	State the boat is located in	Y
Features	txt	Covers, electrical equipment, electronics, inside equipment, outside equipment/Extras, rigging, sails	N
Description	txt	A detailed text description of the vessel	N

Known Unknowns:

- Extras (captured in "Features" above) can add very significantly to a boats value. The NADA site allows for adding certain high value extras. Without parsing the extras lists can a model adequately model prices. Perhaps a simple use of the length of this text information could be an adequate predictor.
- Boats listed may be by both dealers/brokers and private individuals which can affect sale and asking price.