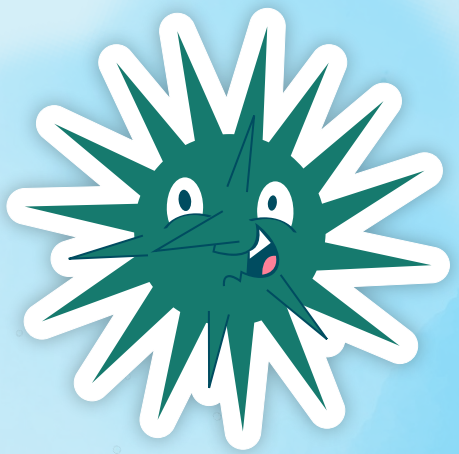
Understanding Used Sailboat Prices

Math 42 Final Project

Melissa Chang, Ruei-yu (Michelle) Chang,
Khang Thai, Kenneth Chow, Charles Wei

Background

- The second-hand sailboat market has been significantly growing throughout the years.
- Determining used sailboat prices has been challenging due to factors such as regional economic and climate conditions have made it difficult to accurately determine the price for each sailboat variant.
- Creating a model that includes various factors that influence used sailboat prices can give us a better understanding and provide a more precise estimate of sailboat values.

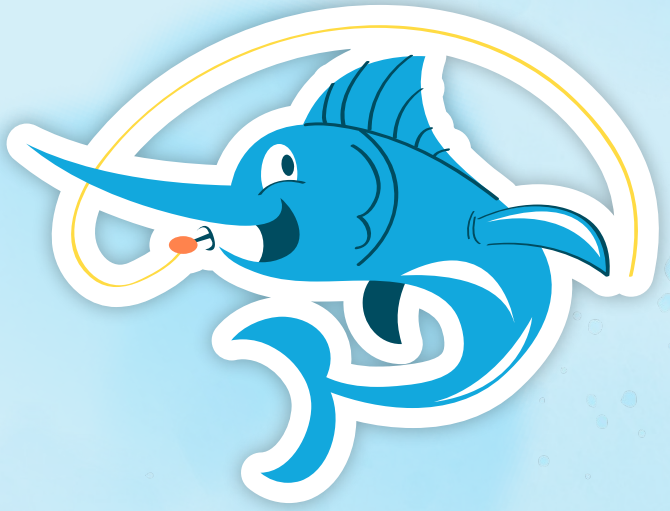


Problem

- Develop a mathematical model that explains the listing prices of each of the sailboats
- Include additional features such as Make, Variant, Length, Geographic Region, Listing Price, Year, Beam, Draft, Displacement, and Fuel Capacity to construct a regression model with multiple independent variables.
- Impact of geographic regions on the pricing of different types of sailboats based on the established model.
- Use the model to explain the importance of the given geographic regions in the Hong Kong market.
- Find additional interesting and informative inferences from the dataset.

Assumptions

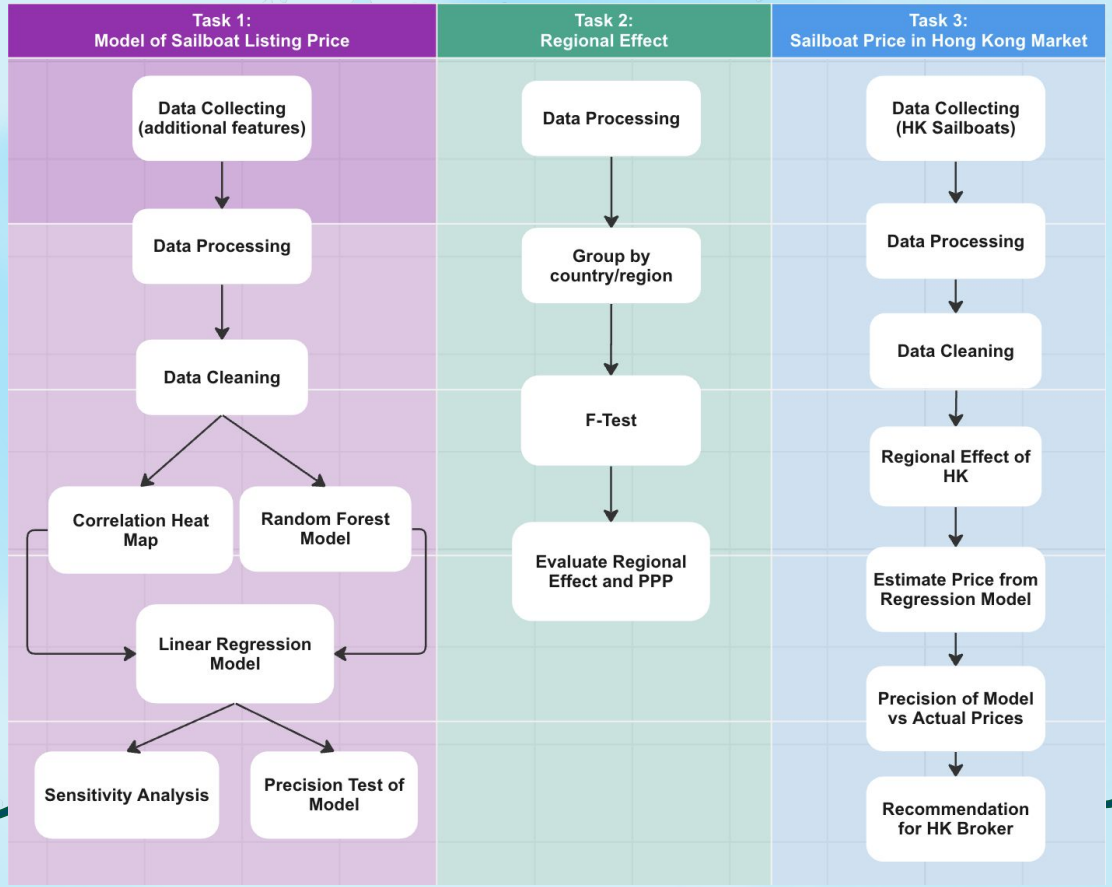
- Features that affect price are limited to **Make, Variant, Length, Region, Country, Year, Beam, Draft, Displacement, and Fuel Capacity** and are not affected by human factors.
- We assume all **market conditions** are stable, and factors such as inflation or financial crisis will not affect the price of used sailboats.
- For economic metrics, we will classify Europe as the European Union and UK.
- We assume all **sellers list the price rationally**, based on their consideration of the condition of the used sailboat itself.



Monohulled Sailboat

Catamaran

Modeling Approach



Data processing

Make	Variant	Length (ft)	Geographic Region	Country/Region/State	Listing Price (USD)	Year	Average PPP ($)
Alubat	Ovni 395	41	Europe	France	$267,233	2005	38,770
Bavaria	38 Cruiser	38	Europe	Croatia	$75,178	2005	23,716
Bavaria	38 Cruiser	38	Europe	Croatia	$66,825	2005	23,716
Bavaria	38 Cruiser	38	Europe	Croatia	$54,661	2005	23,716
Bavaria	38 Cruiser	38	Europe	Croatia	$53,447	2005	23,716
Bavaria	38 Cruiser	38	Europe	Greece	$91,101	2005	27,542
Bavaria	39 Cruiser	39	Europe	Greece	$66,748	2005	27,542
Bavaria	42 Match	41	Europe	Croatia	$78,945	2005	23,716
Bavaria	42 Match	41	Europe	Croatia	$58,297	2005	23,716

Make	Variant	Length (ft)	Geographic Region	Country/Region/State	Catamaran Listing Price (USD)	Year	Average PPP ($)
Lagoon	380	38	Caribbean	Martinique	$204,921	2005	14400
Lagoon	380	38	Caribbean	Guadeloupe	$200,071	2005	38,770
Lagoon	380	38	USA	Florida	$219,000	2005	52322
Fountaine Pajo	Lavezzi 40	39	Caribbean	Mexico	$210,000	2005	17418
Leopard	40	39	Caribbean	Panama	$200,000	2005	23807
Nautitech	40	39.5	Europe	Croatia	$188,252	2005	23,716
Nautitech	40	39.5	Europe	Croatia	$188,131	2005	23,716
Lagoon	410	40.5	Caribbean	Grenada	$225,000	2005	13000
Lagoon	410-S2	40.5	Europe	Spain	$303,395	2005	34099

Data processing

Make	Variant	Length (ft)	Geographic Region	Country/Region/State	Listing Price (USD)	Year	Beam (ft)	Draft (ft)	Fuel Capacity (L)	Regional PPP ($)
Bavaria	38 Cruiser	38	Europe	Croatia	$75,178	2005	12.80	6.46	150	23716
Bavaria	38 Cruiser	38	Europe	Croatia	$66,825	2005	12.80	6.46	150	23716
Bavaria	38 Cruiser	38	Europe	Croatia	$54,661	2005	12.80	6.46	150	23716
Bavaria	38 Cruiser	38	Europe	Croatia	$53,447	2005	12.80	6.46	150	23716
Bavaria	38 Cruiser	38	Europe	Greece	$91,101	2005	12.80	6.46	150	27542
Bavaria	39 Cruiser	39	Europe	Greece	$66,748	2005	13.00	6.08	210	27542
Bavaria	42 Match	41	Europe	Croatia	$78,945	2005	12.25	7.07	230	23716
Bavaria	42 Match	41	Europe	Croatia	$58,297	2005	12.25	7.07	230	23716
Bavaria	42 Cruiser	42	Europe	Croatia	$112,906	2005	13.09	5.92	230	23716

Make	Variant	Length (ft)	ographic Regi	untry/Region/S	sting Price (US	Year	Displacement (lb)	Fuel Capacity (L)	Beam (Ft)	Draft (ft)	Average PPP
Lagoon	380	38	Caribbean	Martinique	$204,921	2005	16,005.00	200	21.417	3.833	14,400
Lagoon	380	38	Caribbean	Guadeloupe	$200,071	2005	16,005.00	200	21.417	3.833	38,770
Lagoon	380	38	USA	Florida	$219,000	2005	16,005.00	200	21.417	3.833	52,322
Fountaine Pajot	Lavezzi 40	39	Caribbean	Mexico	$210,000	2005	13,228.00	250	21.333	3.583	17,418
Leopard	40	39	Caribbean	Panama	$200,000	2005	16,821.00	350	22.000	4.417	23,807
Nautitech	40	39.5	Europe	Croatia	$188,252	2005	16,314.00	270	22.667	4.417	23,716
Nautitech	40	39.5	Europe	Croatia	$188,131	2005	16,314.00	270	22.667	4.417	23,716
Lagoon	410	40.5	Caribbean	Grenada	$225,000	2005	15,961.00	200	23.333	3.917	13,000
Lagoon	410-S2	40.5	Europe	Spain	$303,395	2005	20,282.53	200	23.333	3.917	34,099

First attempt: Correlation Heat Map

Correlation Heat Maps

Correlation heat map for monohulled with 3 features

Correlation Heat Maps

Correlation heat map for catamaran with 3 features

Correlation Heat Maps

Correlation heat maps for sailboats with additional features

Second attempt:
Random Forest Regression Analysis

Random Forest Results

Random Forest Regression Analysis for Monohulled with 3 Features:

- Length: **42.3%** importance
- Year: **30.7%** importance
- Average PPP: **27.0%** importance

Random Forest Regression Analysis for Catamarans with 3 Features:

- Length: **61.94%** importance
- Year: **27.98%** importance
- Average PPP: **10.08%** importance

Random Forest Results

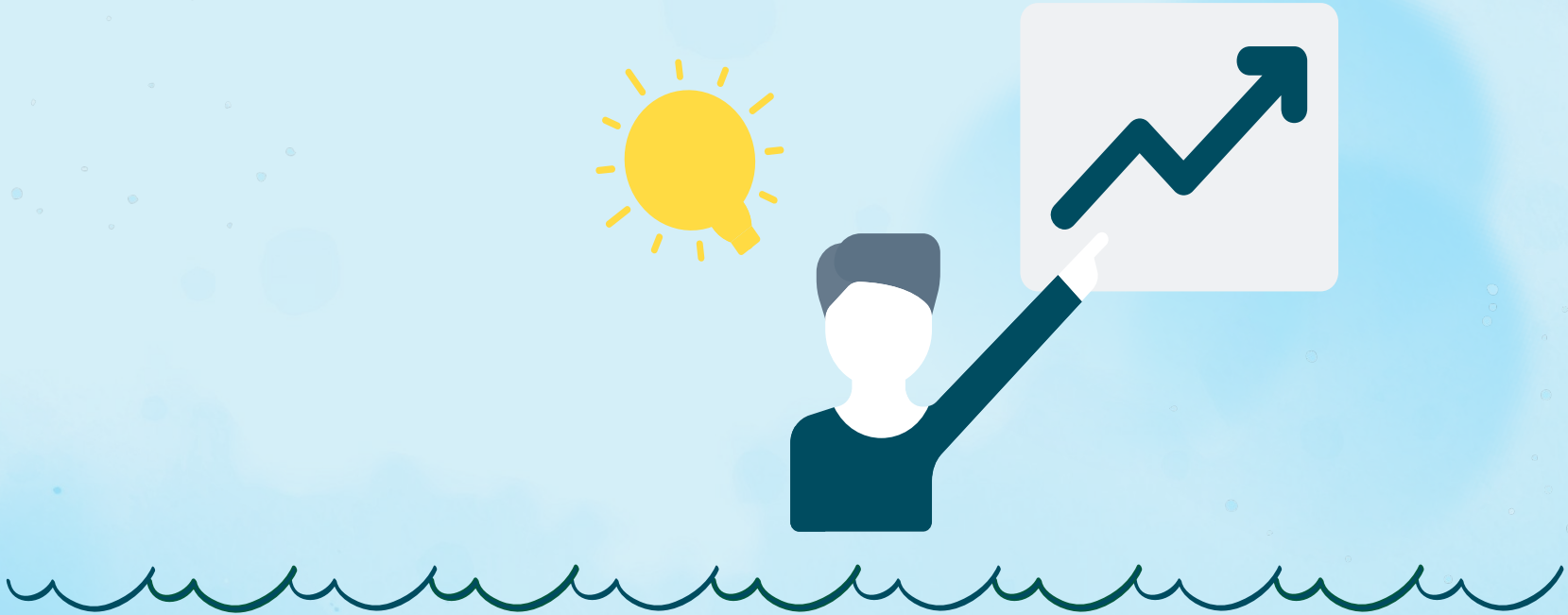
Random Forest Analysis for Monohulled with Additional Features:

- Fuel Capacity: **56.9%** importance
- Year: **20.5%** importance
- Length: **6.7%** importance
- Beam: **6.6%** importance
- Draft: **4.7%** importance
- Average PPP: **4.6%** importance

Random Forest Regression Analysis for Catamarans with Additional Features:

- Length: **57.1%** importance
- Beam: **17.1%** importance
- Displacement: **6.7%** importance
- Year: **5.9%** importance
- Average PPP: **5.8%** importance
- Fuel Capacity: **4.4%** importance
- Draft: **3.0%** importance

Model Evaluation



Exploring Relationships

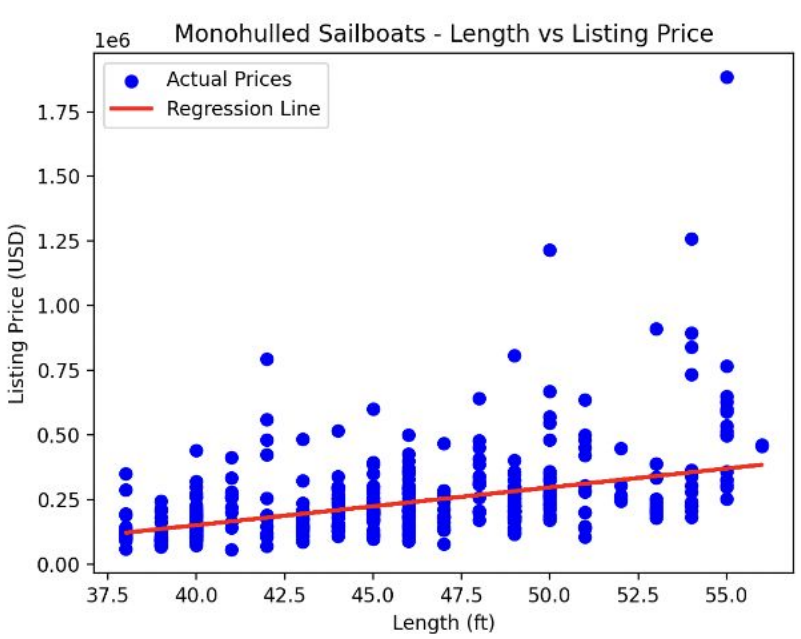
Correlation Coefficient:

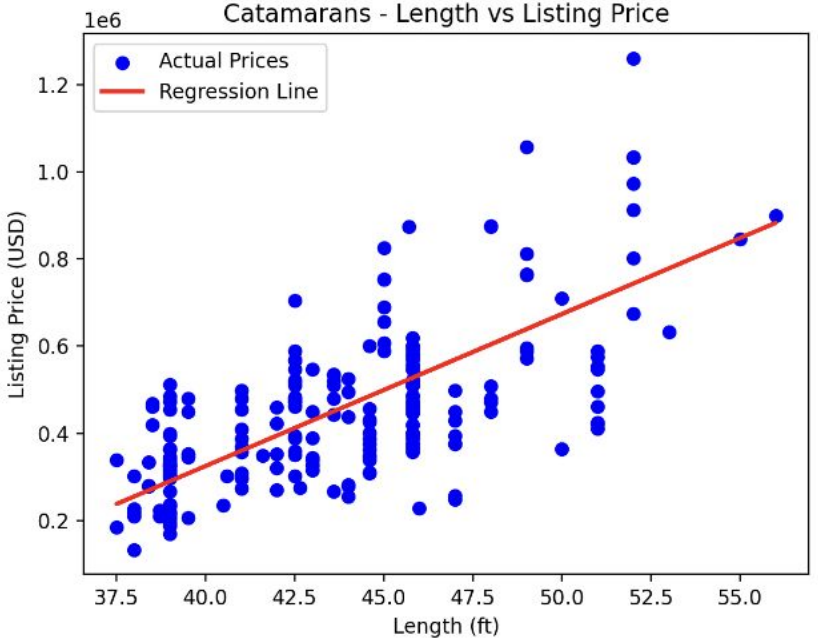
0.476923323910931O6

<

Correlation Coefficient:

0.6839815819823276





Regression Model

P = Price, x1 = Length (ft), x2 = Year, x3 = Average PPP ($USD)

Monohull Sailboats:

$$P = 15187.13 * x_1 + 12450.62 * x_2 + 3.81 * x_3 - 25625815.47$$

$$R^2 = 0.40463728584057 21$$

Catamarans:

$$P = 34204.91 * x_1 + 22645.16 * x_2 + 1.86 * x_3 - 46691648.54$$

$$R^2 = 0.731602780282924$$

Precision of Model

Chi-Square Test

Monohull Sailboats:

- Chi-Square: 1464.04448022151
- P-value: 2.041233544232998e-277

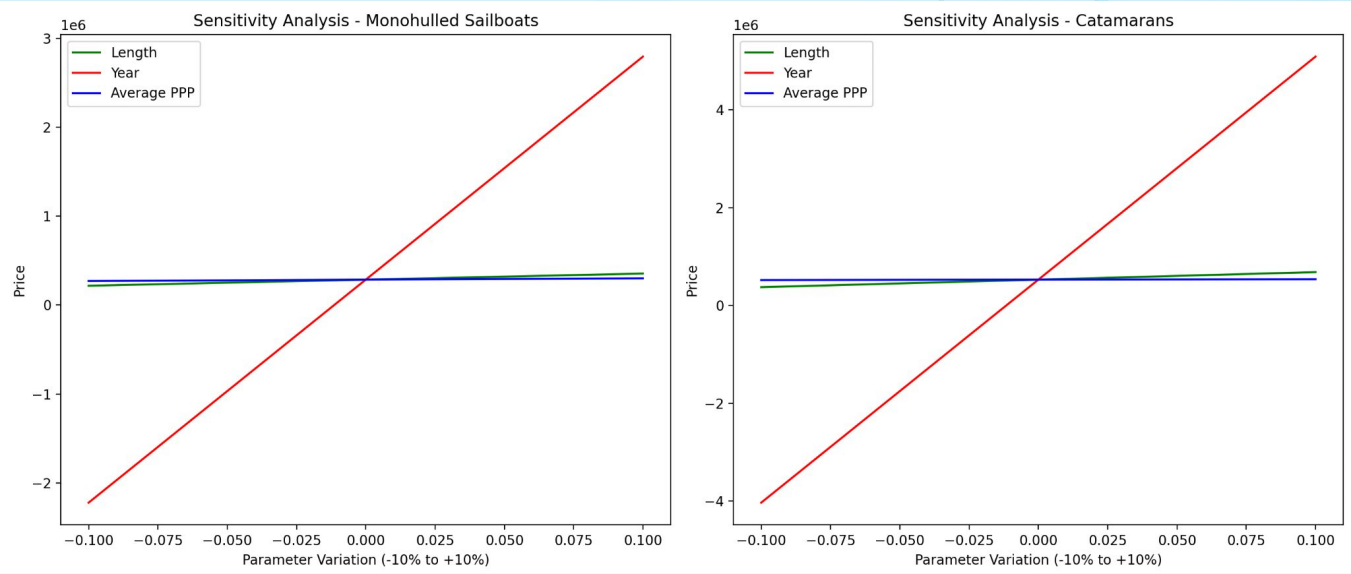
Catamarans:

- Chi-Square: 433.99564306721896
- P-value: 1.1125282562717235e-68

Monohulled Sailboats' Dataset: 2347 observations

Catamarans' Dataset: 1146 observations

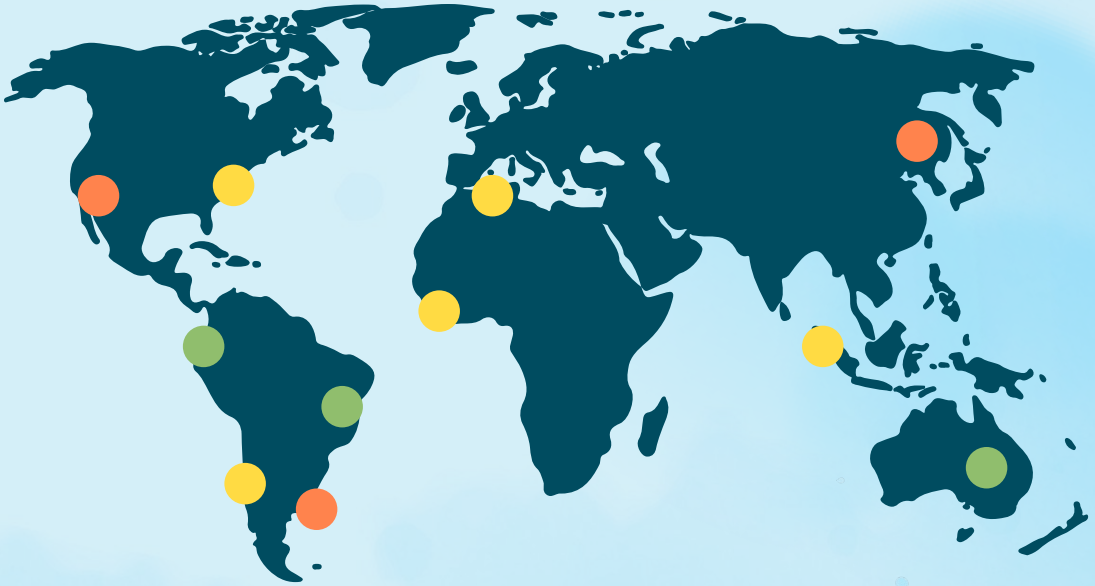
Sensitivity Analysis



Change parameters from -10% to +10%

Fluctuation in Year, but not Length and Average PPP

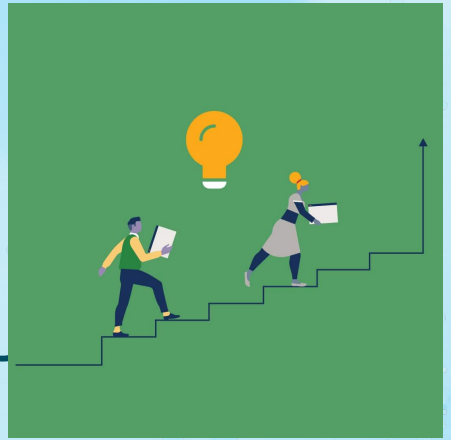
Regional Impacts



Purpose and setting up

- To analyze if each region changes the listing price of each variants

- Separating data of catamarans and monohulls sailboats
- Check the significant relationship between region and listing price using the f-test



F-Test Data - Region

- P-value < 0.05 significance value
- F-value > 1, large variation of difference in pricing

Table 1.1 F-Test Based on Region

Sailboats	F-Value	P-Value
Catamarans	1.844	0.001
Monohulls	1.949	0.000006

F-Test Data - Average PPP

- P-value, one > 0.5 and one < 0.5
- F-value > 1, large variation of difference in pricing

Table 1.2 - F-Test Based on Average PPP

Sailboats - PPP	F-Value	P-Value
Catamarans	1.842	0.0726
Monohulls	19.595	0.0003

Interpretation from data

- There is enough evidence that shows region affecting listing price (change based on different structure and usage)

- Not enough evidence to tell PPP affects listing price for catamarans however, there is enough evidence for monohulls.

Hong Kong Analysis

Precision of Model (Hong Kong Dataset)

Monohull Sailboats:

$$P = 15187.13 * x_1 + 12450.62 * x_2 + 3.81 * x_3 - 25625815.47$$

$$R^2 = 0.06272794359222178$$

Catamarans:

$$P = 34204.91 * x_1 + 22645.16 * x_2 + 1.86 * x_3 - 46691648.54$$

$$R^2 = 0.3736596691640808$$

Chi-Square Test

Monohull Sailboats:

- Chi-Square: 1464.04448022151
- P-value: 2.041233544232998e-277

Catamarans:

- Chi-Square: 433.99564306721896
- P-value: 1.1125282562717235e-68

Regional Effect on Hong Kong

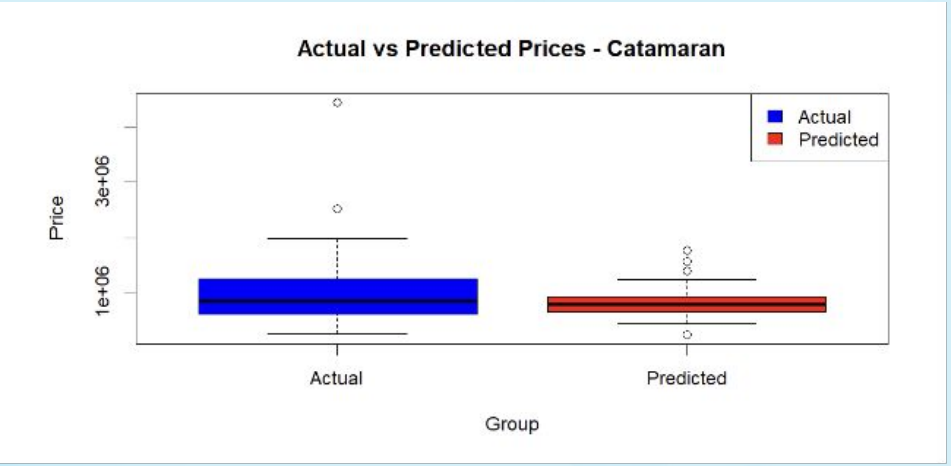
Main question: How does regional effect change listing price in Hong Kong?

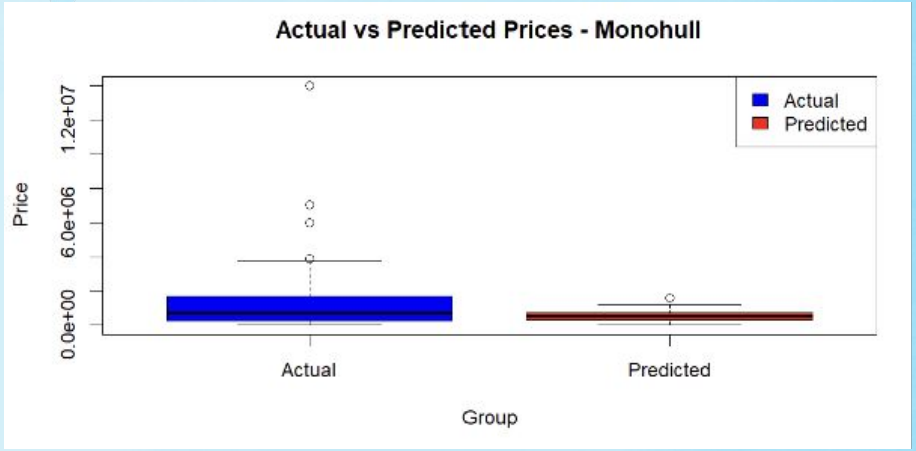
Setting up

How do we do this?

1. Clean the data and create predicted values based on mathematical model for listing price.
2. Combine the predicted values with the actual listing priced from a Hong Kong sailboat dataset.
3. We know that sailboat prices is based on region from previous analysis, so we can analyze the difference in prices based in Hong Kong compared to other regions.

Some Visuals





T - test data

- P - value > 0.5 and < 0.5
- Observe that from the t-value, actual listed prices are higher,
- Monohull have a higher effect in region

Table 1.3 - T-Test of Estimated Price and Actual Listing Price

Sailboats	T - Value	P - Value
Catamarans	1.4181	0.1648
Monohulls	2.7910	0.0072

Analysis

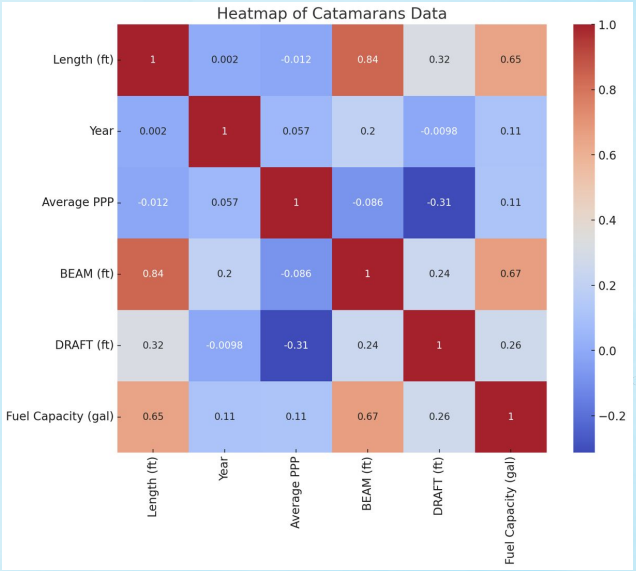
- In general, different region, means change in listing price

- Why may monohulls be more affected and why is price higher?
  - Hong Kong is surrounded by sea
  - Monohulls are more versatile and hence more in demand

Additional Findings

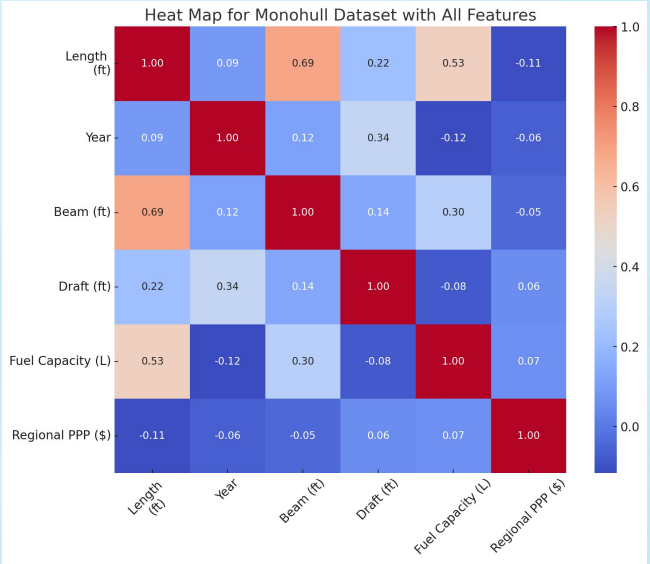
Heatmap Correlation

- Heatmap for both Catamarans and Monohull sailboats was used to find the characteristic correlation.
- Factors between length, year, average ppp, beam, draft, and fuel capacity.
- Strong correlation relationship between length and beam and length and fuel capacity.
- Typically have a larger hull and have a higher structural requirement.
- Width and length of the boat should be more matched to maintain better structural stability.



Heatmap Correlation

- Strong correlation relationship between length and beam.
- Overall shows less of a correlation relationship between each of the independent variables.
- Correlation with other factors that are not included in the heatmap.

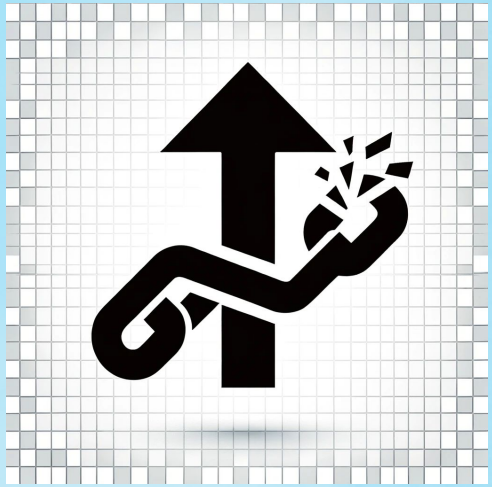


Strengths

- **2 different methods**: heat map and random forest regression using machine learning
- **Linear regression model** provides clear understanding of relationships between Year, Length, Average PPP, and Listing Price → easy for brokers to predict sailboat prices
- Used **PPP** to account for the varying levels of economic development and wealth
- Analyzing **Rˆ2 values**, **chi-squared values**, **p-values**, and created a **sensitivity graph**

Limitations

- **Only three variables** were used (Length, Year, Average PPP)
- Our linear regression model assumes that the relationship between the variables and price is linear. However, a more **complex regression model**
- We did not analyze the **variability of prices** in each specific region and its underlying factors; we looked at average PPP as a whole



Conclusion

- Developed a model for second-hand sailboat pricing that includes both monohull and catamaran sailboats
- Collected data from the used sailboat market with factors that has influences on the price such as Make, Variant, Length, Region, Country, Year, Beam, Draft, Displacement, and Fuel Capacity,
- Created a regression model for monohull and catamaran sailboats and analyze the impact of geographic regions on pricing.
- Used the model to analyzed the regional impact for used sailboats in Hong Kong.

Thank You!