

[Get unlimited access](#)[Open in app](#)

Bruno Franco

Aug 12 · 4 min read · [Listen](#)

Save



Boat Sales Case Study

Photo by [Alix Greenman](#) on [Unsplash](#)

Introduction

A boat company has a website called Nearly New Nautical that allows users to advertise and sell their used boats. The company's marketing team is designing a weekly newsletter





Get unlimited access

Open in app

Customer Questions

1. Do the most expensive boats get the most views?
2. Do the most viewed boats share common features?

Data Set

The raw boat data can be found [here](#).

I used R to clean and analyze the data, which contained the following columns:

Column Name	Details
Price	Character, boat price listed in different currencies (e.g. EUR, Â£, CHF etc.) on the website
Boat Type	Character, type of the boat
Manufacturer	Character, manufacturer of the boat
Type	Character, condition of the boat and engine type (e.g. Diesel, Unleaded, etc.)
Year Built	Numeric, year of the boat built
Length	Numeric, length in meter of the boat
Width	Numeric, width in meter of the boat
Material	Character, material of the boat (e.g. GRP, PVC, etc.)
Location	Character, location of the boat is listed
Number of views last 7 days	Numeric, number of the views of the list last 7 days

Image from [DataCamp](#)

Data Analysis

After cleaning the data, I focused on the following variables: price (USD), number of views in the last 7 days, decade built, location, floor area, boat status (new/used), and material.





Get unlimited access

Open in app

```
In [204]: boat_sales_clean <- boat_sales %>%
select(id, Price_USD, Number.of.views.last.7.days, Decade.Built, Location, Floor.Area, Floor.Area,
       Boat.Status, Material)
```

```
In [205]: boat_sales_clean
```

id	Price_USD	Number.of.views.last.7.days	Decade.Built	Location	Floor.Area	Boat.Status	Material
1	3503.85	226	2010s	Switzerland	7.6000	New	
2	3559.80	75	2020s	Germany	6.0000	New	Thermoplastic
4	3626.00	64	2020s	Denmark	3.0000	New	
5	3466.98	58	2010s	Germany	5.1830	New	Aluminium
7	3780.00	474	1990s	Switzerland	14.7560	Used	Aluminium
9	3399.66	45	2010s	Germany	4.9868	New	
10	3366.00	180	2010s	Italy	7.5255	New	
11	3675.00	239	1980s	Switzerland	7.6125	Used	GRP
13	3570.00	69	2000s	Italy	9.4000	Used	GRP
14	4830.00	113	2020s	Switzerland	7.2600	New	Aluminium
15	4725.00	413	1990s	Switzerland	4.9476	Used	Aluminium
18	4200.00	124	1980s	Denmark	4.0000	Used	
19	4080.00	265	1980s	France	22.7200	Used	PVC
20	4080.00	330	2010s	Germany	14.7147	Used	GRP
21	4078.98	69	2010s	Germany	5.1830	New	Aluminium
23	4008.60	90	2020s	Germany	7.0684	New	Aluminium
24	3978.00	98	1970s	Germany	11.7600	Used	GRP
25	3978.00	383	1980s	France	16.1291	Used	
26	4410.00	160	1990s	Switzerland	8.8350	Used	GRP

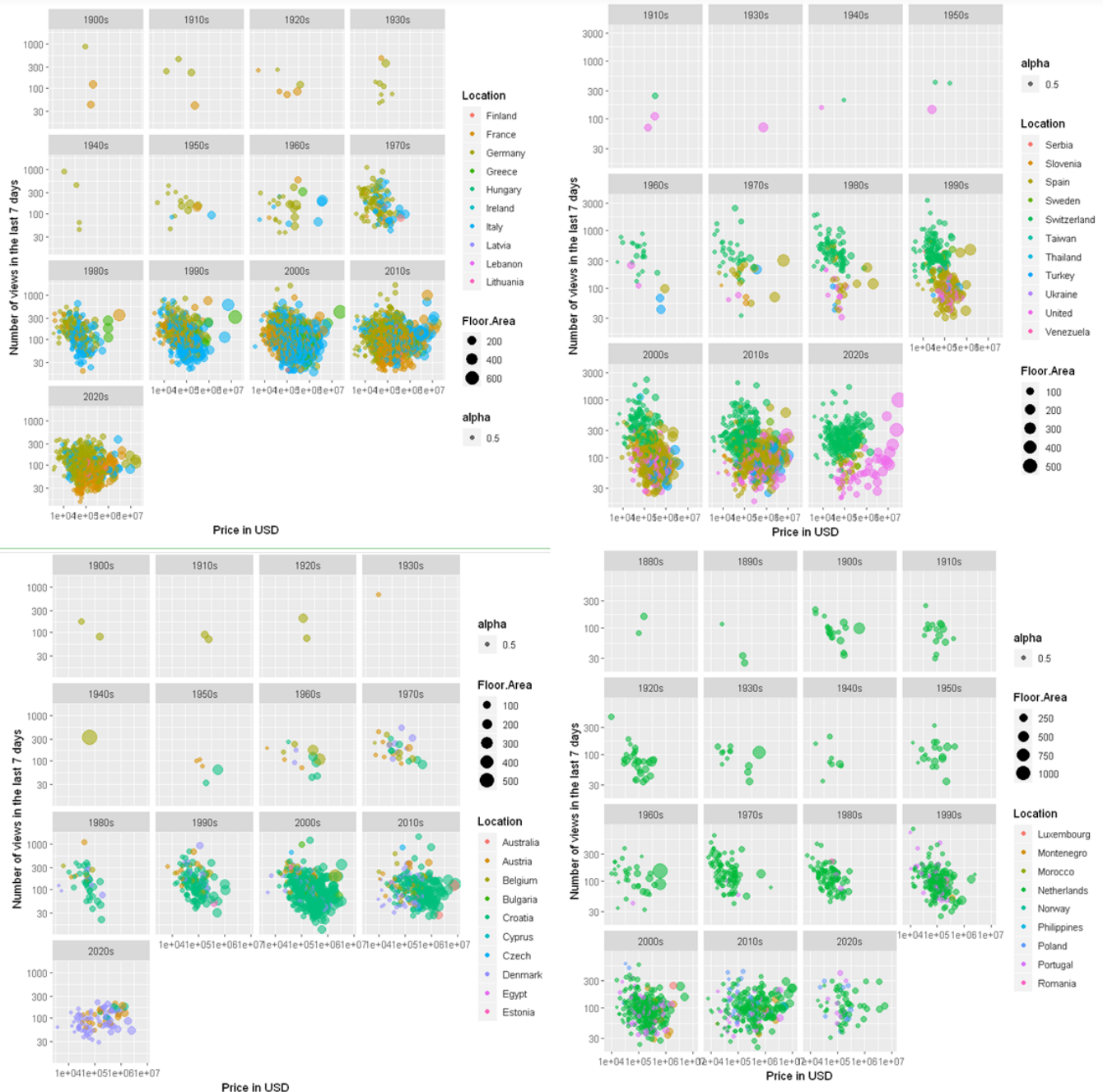
Number of views in the last 7 days versus price in USD





Get unlimited access

Open in app



The graphs are facet wrapped by decade in which the boat was built. The color of the dots represent the location of the boat (country). The size of the dots are based on the floor area of the boat. Each graph focuses on 9–11 locations.

Expensive boats do get a lot of views, but the most expensive boat is not necessarily always the one that gets the most views. The size of the boat also affects the price of the boat, which may contribute to how many views it gets.




[Get unlimited access](#)
[Open in app](#)

respectively, get the most views. Additionally, in boats with documented status, the views are not really affected by whether the boat is new or used.

Location	avg_views	avg_price_usd
Serbia	442.0	12240.0
Czech	401.5	100980.0
Switzerland	273.0	57592.5
Luxembourg	243.0	1938000.0
Romania	225.0	59670.0
Sweden	212.0	302440.0
Belgium	177.0	73134.0
Latvia	150.5	698700.0
Austria	141.0	88230.0
Hungary	141.0	62118.0

Decade.Built	avg_views	avg_price_usd
1970s	167.0	36750
1960s	157.0	62475
1980s	145.0	42840
1990s	127.0	70380
1950s	125.5	81090
1880s	121.5	125970
2020s	110.0	72410
1930s	109.0	50490
2010s	102.0	142290
1910s	96.0	119850
2000s	95.0	153000
1900s	91.0	69870
1940s	87.0	50490
1920s	78.0	79815
1890s	33.0	142290

Boat.Status	avg_views	avg_price_usd
NA	145	69870.0
New	105	66118.5
Used	107	111180.0

The left table arranges the location of the boat by average number of views in descending order. The middle table arranges the decade in which the boat was built by average number of views in descending order. The right table shows the average number of views based on status. All tables also show the average price converted to US dollars.

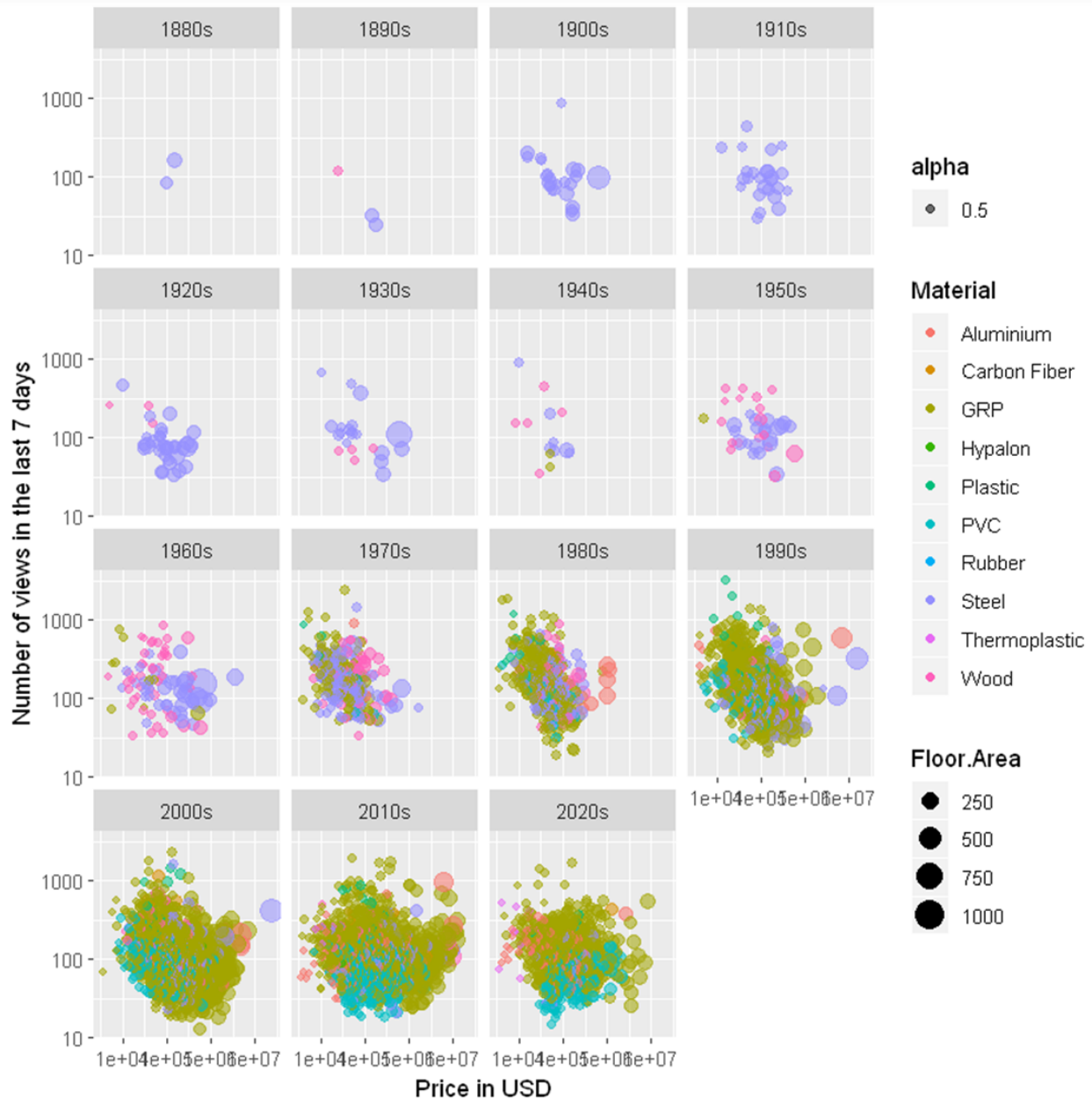
All tables above show that boats that get the most views by location and by decade built are not necessarily the most expensive.





Get unlimited access

Open in app

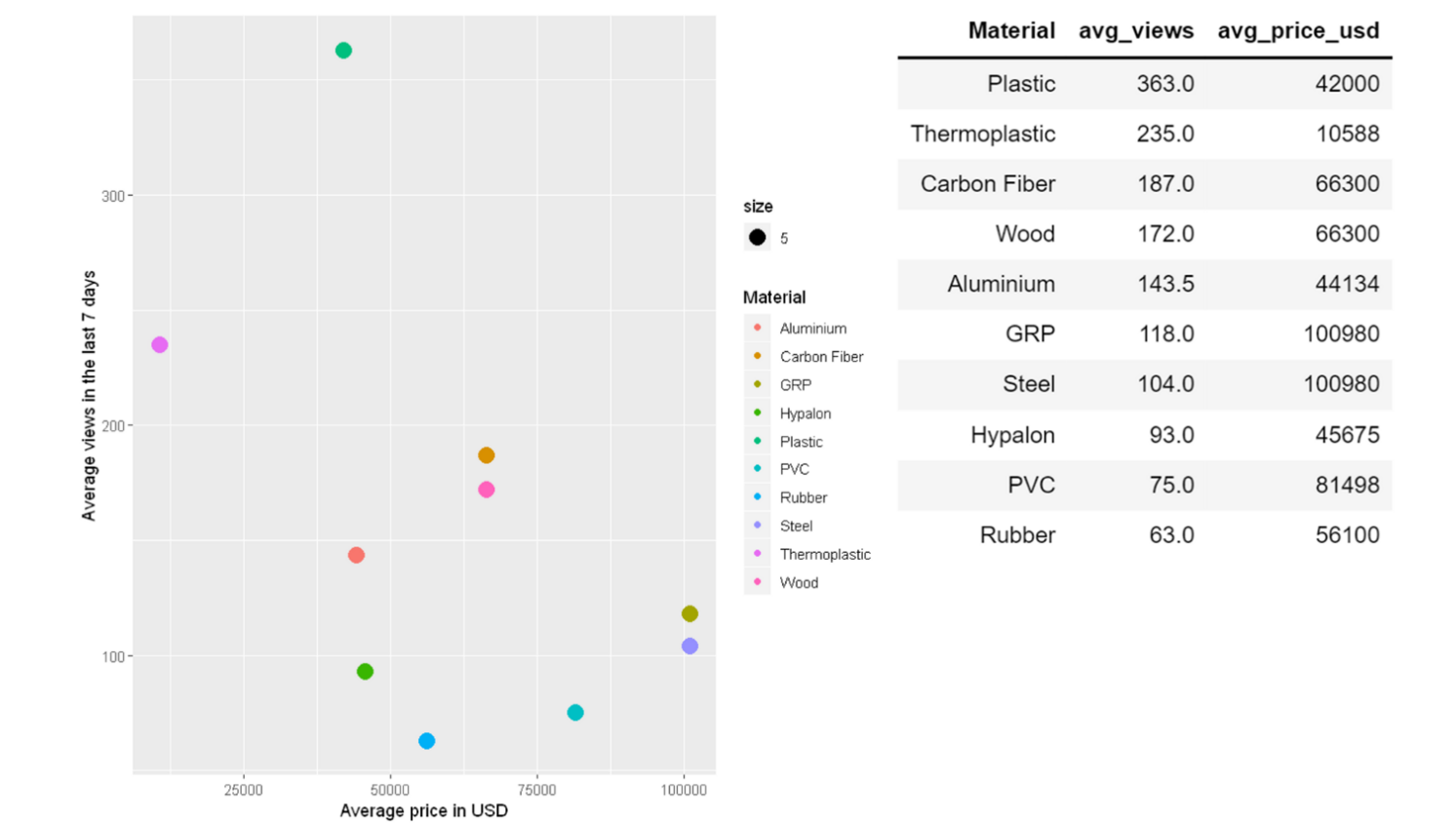


These scatterplots show the number of views in the last 7 days versus price in US dollars faceted wrapped by decade built. The size of the dots are again determined by the floor area. The dot colors are determined by the material of the boat.





The column chart on the left illustrates the average views each material got in the last 7 days. The column chart on the right illustrates the average price in US dollars for each material.



The scatter plot on the left shows the average views in the last 7 days versus average price in US dollars with material of the boat represented by different colors. The table on the right shows the boat material arranged by average views in descending order.



Get unlimited access

Open in app

```
In [54]: cor(boat_sales_clean$Price_USD, boat_sales_clean$Number.of.views.last.7.days)
-0.00506324463573833

In [55]: cor(boat_sales_clean_M_avg_v$avg_views, boat_sales_clean_M_avg_v$avg_price_usd)
#The correlation between average views and price is a stronger negative correlation when taking into account material than wh
-0.492681607960198
```

Correlation between average views and price using all the boat data (In [54]). Correlation between average views and price by material (In [55]).

There is a stronger negative correlation between the average views and price when we focus on material (-0.493) than when we consider all the data (-0.005). This has to do with how plastic, a cheap material, gives boats a lot of views.

Other variables

Manufacturer	avg_views	avg_price_usd	Boat.Type	avg_views	avg_price_usd
Bavaria power boats	192.26486	170125.6	Cabin Boat	225.1282	65750.65
Sea Ray power boats	189.90377	85483.5	Sport Boat	178.6908	79897.17
Princess power boats	162.03734	613476.3	Bowrider	156.7542	57906.58
Cranchi power boats	148.13699	280029.0	Motor Yacht	129.2864	416603.71
Fairline power boats	141.13953	370853.9	Hardtop	128.8817	364082.16
Sunseeker power boats	133.44648	1101672.3	Pilothouse	121.9086	69430.40
Azimut power boats	113.20465	530308.4	Trawler	111.1281	194763.63
Jeanneau power boats	95.10242	144917.3	Flybridge	106.8124	496752.72
BÃfÂ©nÃfÂ©teau power boats	89.08241	150703.6	Deck Boat	104.9399	50337.89
			Center console boat	100.3117	81239.17

Manufacturer (left table) and boat type (right table) arranged by average views in descending order.

Boats manufactured by Bavaria power boats, Sea Ray power boats, and Princess power boats get the most views, respectively. Cabin boats by far get the most views followed by sport boats and bowriders. Average price does not seem to affect the number of views in either of these two cases.

Business Recommendations



[Get unlimited access](#)[Open in app](#)

- Boats from the 1960s, 1970s, and 1980s
- Boats from Serbia and Czech Republic
- Boats manufactured by Bavaria power boats and Sea Ray power boats
- Cabin boats

Number of views are also not affected by whether the boat is documented as new or used. Also, just because a boat is more expensive does not necessarily mean it will get more views. This is especially true considering the average views of plastic boats are greater than those of more expensive materials.

Letting users know which types of boats sell more may increase the use of the website. I recommend adding website features that lets them know which boat features are most popular in order to increase motivation for website use.

All my work from jupyter notebook can be found [here](#).

