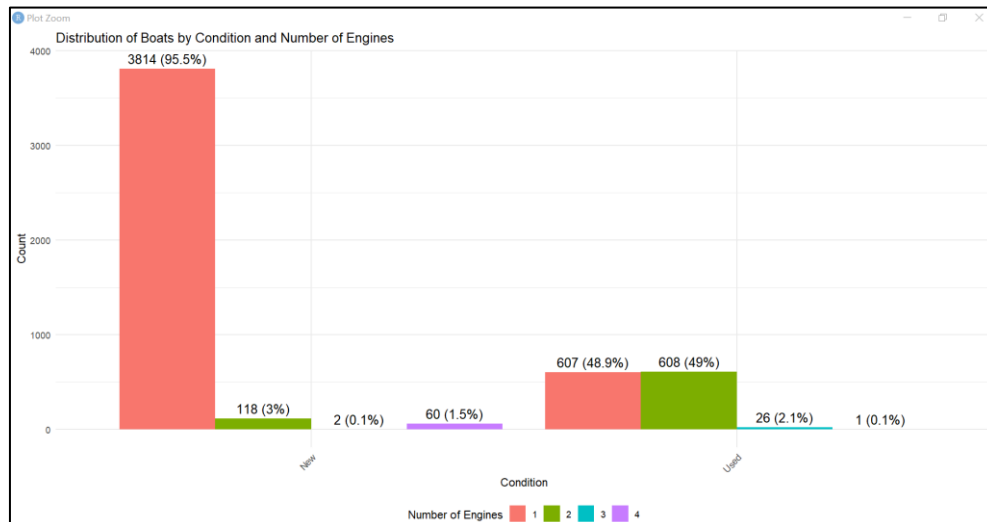


## Data Analysis

### 1. How many of these boats are new (0) or used (1)?

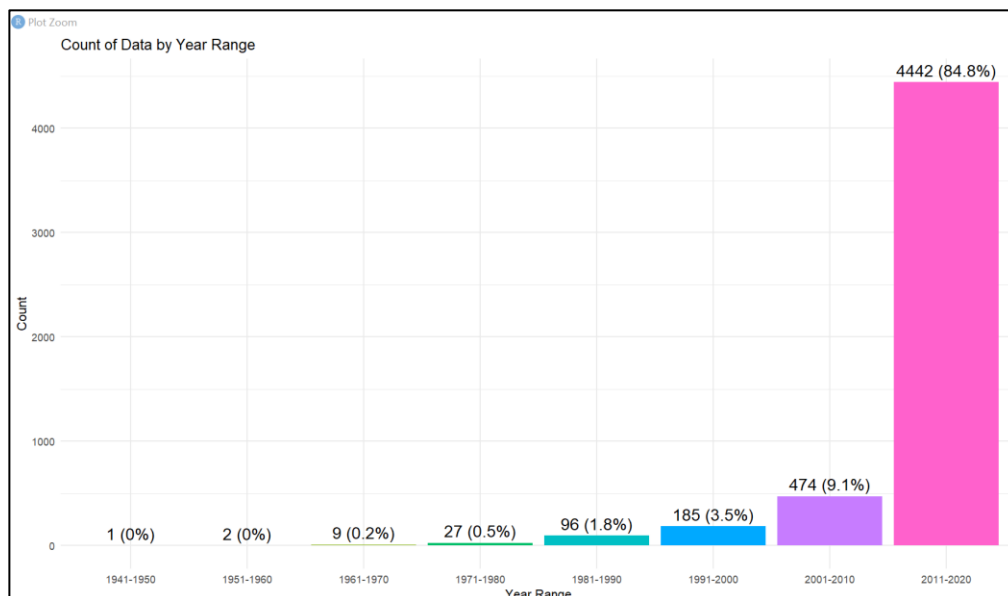


Grafik di atas menunjukkan jumlah kapal baru dan bekas, serta jumlah mesin yang dimilikinya. Berikut adalah rinciannya:

- **Kapal baru (New)** memiliki 3814 kapal dengan 1 mesin, 118 kapal dengan 2 mesin, 2 kapal dengan 3 mesin, dan 60 kapal dengan 4 mesin.
- **Kapal bekas (Used)** memiliki 607 kapal dengan 1 mesin, 608 kapal dengan 2 mesin, 26 kapal dengan 3 mesin, dan 1 kapal dengan 4 mesin.

Mayoritas kapal baru memiliki 1 mesin dan mayoritas kapal bekas juga memiliki 1 mesin, dapat disimpulkan bahwa mesin tunggal adalah yang paling umum untuk kapal, baik baru maupun bekas.

*Also the number of boats manufactured in 10 years span?*



Grafik di atas menunjukkan jumlah kapal berdasarkan kategori **tahun pembuatan**, dimana mayoritas kapal sebanyak 84,8% diproduksi antara tahun 2011 dan 2020. Di sisi lain, periode awal, terutama dari tahun 1941 hingga 1980, menyumbang persentase relatif yang lebih kecil dari total kapal. Distribusi ini menunjukkan pergeseran yang signifikan dalam tren manufaktur kapal dari waktu ke waktu, dengan terlihatnya konsentrasi yang mencolok dari kapal-kapal baru dalam beberapa dekade terakhir.

## 2. Model to predict a boat's price.

```
Call:
lm(formula = price ~ year + condition + length_ft + beam_ft +
    dryweight_lb + numEngines + totalHP, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-1022224  -14981   -1348    9466  2759633

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.648e+07  5.015e+05 -32.859 < 2e-16 ***
year         8.129e+03  2.476e+02  32.826 < 2e-16 ***
condition    2.382e+04  4.458e+03   5.342 9.57e-08 ***
length_ft    3.211e+03  3.964e+02   8.100 6.79e-16 ***
beam_ft     -1.874e+03  9.724e+02  -1.927  0.054 .
dryweight_lb  5.728e+00  1.802e-01  31.781 < 2e-16 ***
numEngines   1.805e+04  4.259e+03   4.239 2.28e-05 ***
totalHP      1.962e+02  8.462e+00   23.187 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 86810 on 5228 degrees of freedom
Multiple R-squared:  0.7007,    Adjusted R-squared:  0.7003
F-statistic: 1748 on 7 and 5228 DF,  p-value: < 2.2e-16
```

Hasil regresi semua variabel

Dalam analisis ini, perlu diperhatikannya **t-statistik** dan **nilai p (p-value)** untuk setiap koefisien dalam model regresi. Nilai p yang lebih rendah menunjukkan bahwa variabel tersebut secara statistik signifikan dalam menjelaskan variasi dalam variabel dependen (harga), dengan asumsi tingkat signifikansi tertentu (0,05).

- **Year:** T-statistiknya adalah 32,83, dan nilai p-nya sangat rendah. Ini menunjukkan bahwa variabel "tahun" secara **sangat signifikan** dalam menjelaskan variasi harga.
- **Condition:** T-statistiknya adalah 5,34, dan nilai p-nya signifikan, meskipun tidak serendah variabel "tahun". Namun, ini menunjukkan bahwa variabel secara statistik signifikan dalam menjelaskan variasi harga.
- **Length\_ft:** T-statistiknya adalah 8,10, dan nilai p-nya signifikan, menunjukkan bahwa "length\_ft" secara signifikan dalam menjelaskan variasi harga.
- **Beam\_ft:** T-statistiknya adalah -1,93, dan nilai p-nya relatif tinggi, di atas tingkat signifikansi yang biasanya digunakan yaitu 0,05. Ini menunjukkan bahwa "beam\_ft" mungkin tidak secara statistik signifikan dalam menjelaskan variasi harga, meskipun masih mendekati.
- **DryWeight\_lb:** T-statistiknya adalah 31,78, dan nilai p-nya sangat rendah, menunjukkan bahwa "dryWeight\_lb" secara **sangat signifikan** dalam menjelaskan variasi harga.
- **NumEngines:** T-statistiknya adalah 4,24, dan nilai p-nya signifikan, menunjukkan bahwa "numEngines" secara statistik signifikan dalam menjelaskan variasi harga.
- **TotalHP:** T-statistiknya adalah 23,19, dan nilai p-nya sangat rendah, menunjukkan bahwa "totalHP" secara **sangat signifikan** dalam menjelaskan variasi harga.

Berdasarkan analisis ini, variabel "**tahun**", "**dryWeight\_lb**", dan "**totalHP**" lebih baik dalam menjelaskan variasi harga dibandingkan dengan variabel lain dalam model, karena memiliki nilai p yang sangat rendah, menunjukkan signifikansi statistik yang tinggi.

```

Residuals:
    Min       1Q   Median       3Q      Max
-1126733  -12912    -3819     5704 2669276

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.341e+07  3.600e+05  -37.25  <2e-16 ***
year         6.639e+03  1.784e+02   37.21  <2e-16 ***
totalHP      2.650e+02  5.518e+00   48.02  <2e-16 ***
dryWeight_lb 6.150e+00  1.488e-01   41.33  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 88040 on 5232 degrees of freedom
Multiple R-squared:  0.6919,    Adjusted R-squared:  0.6917
F-statistic: 3916 on 3 and 5232 DF,  p-value: < 2.2e-16

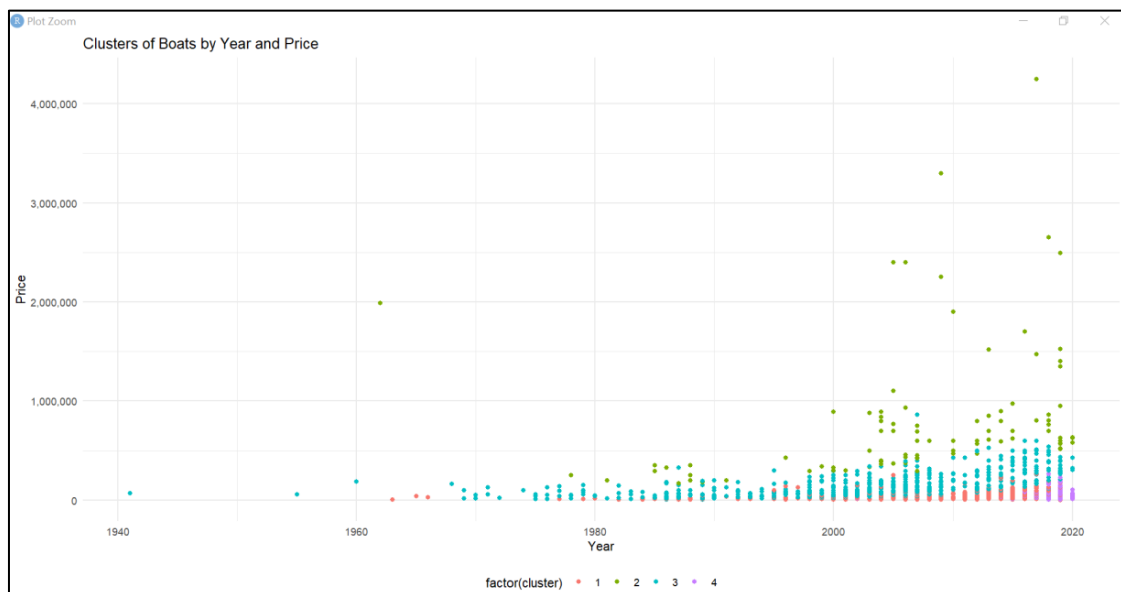
```

Hasil Regresi Terbaru

Maka persamaannya adalah sebagai berikut:

$$\text{Price} = -13410000 + 6639 \times \text{year} + 265 \times \text{totalHP} + 6.15 \times \text{dryWeight\_lb}$$

### 3. Clustering model with four clusters



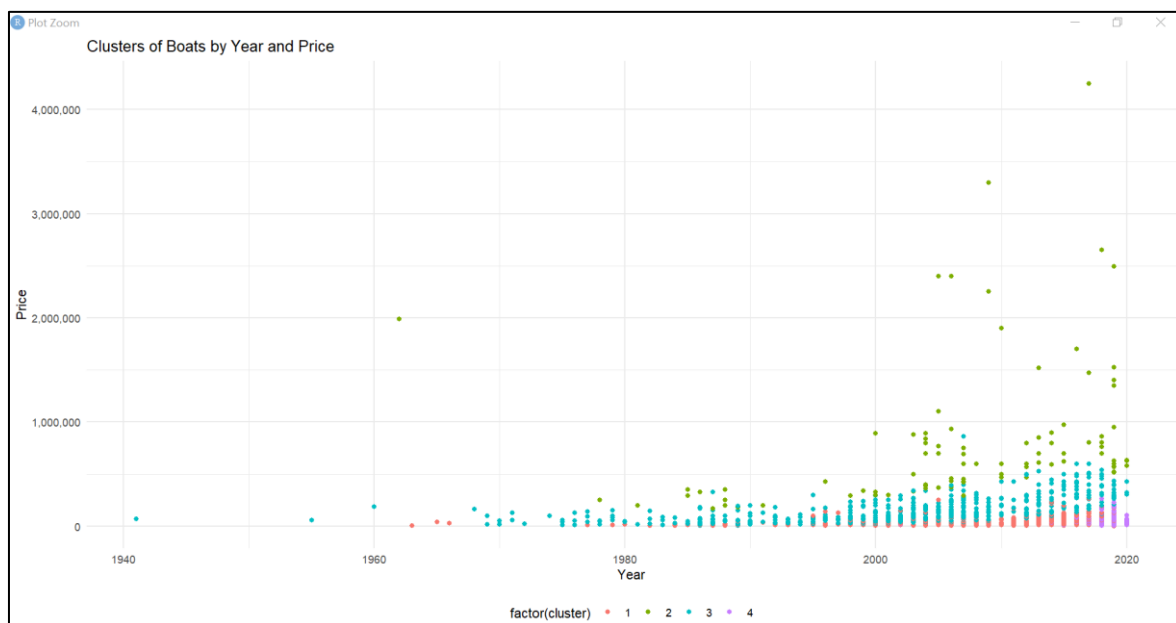
Analisis yang dilakukan di sini menggunakan **metode clustering K-means** untuk mengelompokkan data kapal ke dalam **empat klaster** berdasarkan tahun pembuatan dan harga. Pertama, data diubah menjadi skala yang seragam menggunakan fungsi skala dalam R-Studio. Kemudian, model K-means diterapkan pada data yang telah diubah skala, dengan spesifikasi empat pusat klaster.

Setelah model terlatih, setiap baris data diberi label klaster yang sesuai. Setelah klaster ditentukan, data divisualisasikan menggunakan ggplot2, menampilkan hubungan antara tahun pembuatan kapal (sumbu x) dan harga (sumbu y), dengan warna titik data menunjukkan klaster yang ditetapkan sebelumnya. Hal ini agar dapat melihat pola atau kelompok dalam data, memperoleh pemahaman tentang variasi harga berdasarkan tahun pembuatan kapal.

Cluster	Tahun Rata-rata	Kondisi	Kisaran Ukuran (ft)	Kisaran Berat (lb)	Kisaran Harga
1	2008	Bekas (599 item)	Bervariasi	Bervariasi	Sampai dengan \$294,900
2	2011	Baru (68 item), Bekas (80 item)	Ukuran lebih besar	Lebih berat	\$17,742 hingga \$4,250,000
3	2003	Mayoritas Bekas (563 item), Beberapa Baru (60 item)	Bervariasi, Beberapa besar	Bervariasi	\$4,900 hingga \$859,750
4	2018	Baru (3866 item)	Bervariasi, Umumnya lebih ringan	Bervariasi	\$618 hingga \$259,999

Tabel diatas adalah hasil interpretasi cluster penjualan Boats Victory yang sudah di export dari R-Studio. Cluster 1 terdiri dari kapal bekas yang dibuat sekitar tahun 2008, dengan harga maksimal sekitar \$294,900. Cluster 2 memiliki kapal baru dan bekas yang dibuat sekitar tahun 2011, menonjol dengan ukuran dan berat yang lebih besar, serta kisaran harga yang luas, dari \$17,742 hingga \$4,250,000. Cluster 3, dengan kapal-kapal yang paling tua rata-rata dari tahun 2003, mayoritas bekas dan beberapa baru, menawarkan variasi dalam ukuran dan berat, dengan harga berkisar antara \$4,900 dan \$859,750. Terakhir, Cluster 4 mencakup kapal baru dari tahun 2018, umumnya lebih ringan, dengan harga yang berkisar dari \$618 hingga \$259,999, menandakan segmentasi kapal baru yang lebih terjangkau.

4. *The employee who suggests you analyze the company's data brings up several questions that are usually asked by the customers. It would be beneficial to have the answers ready so that the sales team can communicate with the customers better.*

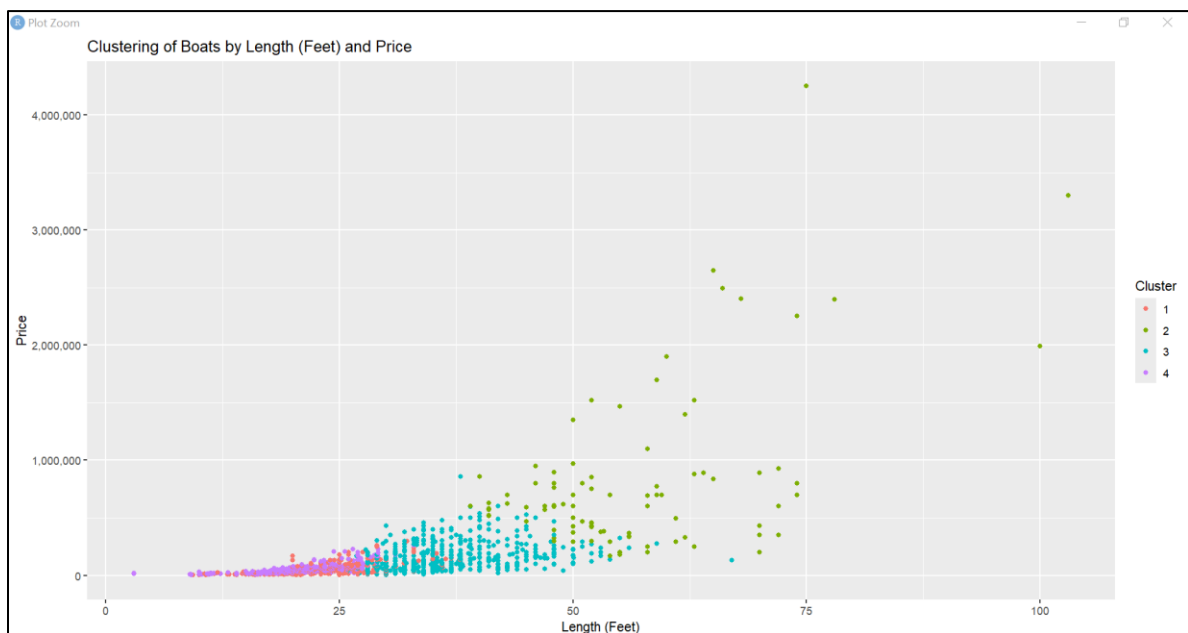


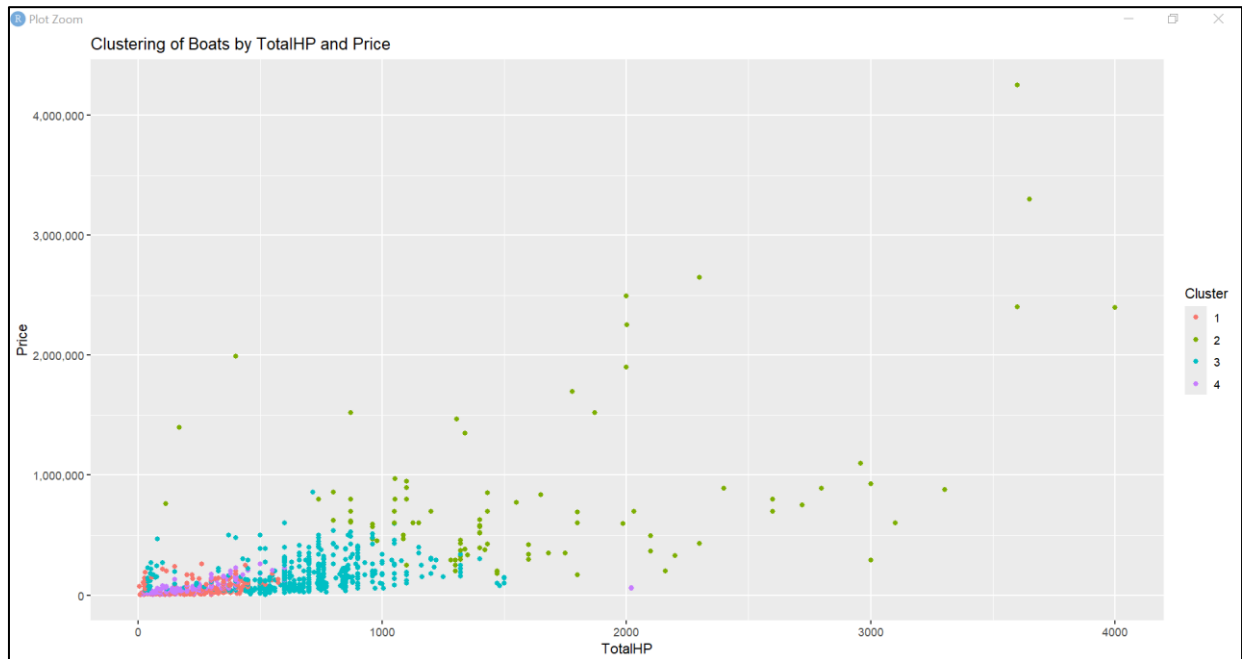
- a. Jika pelanggan menginginkan kapal yang cukup baru, disarankan untuk memilih **klaster 4** dalam analisis gambar. Klaster 4 mewakili kelompok kapal dengan tahun pembuatan lebih modern, yang sesuai dengan preferensi pelanggan untuk mencari kapal yang tidak terlalu tua. Selain itu, untuk perkiraan biaya, pelanggan sebaiknya mempersiapkan biaya di bawah 500.000. Ini disesuaikan dengan rata-rata harga kapal dalam klaster 4.

- b. Jika pelanggan memiliki biaya sekitar \$200.000, rekomendasikan untuk pelanggan tersebut adalah mempertimbangkan kapal dalam klaster 1 dan klaster 4. **Klaster 1** menawarkan kapal dengan rentang harga sekitar \$200.000, yang umumnya diproduksi antara tahun 1963 hingga 2019. Kapal dalam klaster ini umumnya dalam kondisi bekas pakai (used) dan biasanya dilengkapi dengan 1 hingga 2 mesin. Sementara itu, **klaster 4** juga menyediakan kapal dengan harga sekitar \$200.000, namun kapal-kapal ini memiliki tahun pembuatan lebih baru, umumnya antara tahun 2002 hingga 2020. Kapal-kapal dalam klaster 4 cenderung dalam kondisi baru (new) dan juga biasanya memiliki 1 hingga 2 mesin.

Namun rekomendasi yang cocok untuk pelanggan tersebut adalah **klaster 4**, yang menyajikan opsi menarik bagi pelanggan untuk mencari kapal yang relatif baru dengan kondisi yang prima dan spesifikasi modern. Dengan memilih kapal dari klaster 4, pelanggan dapat memiliki keyakinan tambahan dalam memperoleh kapal yang memenuhi kebutuhan mereka dengan baik, serta memastikan investasi mereka memberikan nilai yang optimal dalam jangka panjang.

- c. Interpretasi panjang (Length (Feet)) dan tenaga kuda (TotalHP) berdasarkan klaster kapal Boats Victory menunjukkan **klaster 1** umumnya terdiri dari kapal yang lebih kecil dengan tenaga kuda rendah, yang cenderung menunjukkan kapal tua atau yang dirancang untuk kegiatan santai. Sebaliknya, **Klaster 2** menampilkan kapal yang lebih besar dengan tenaga kuda tinggi, mengindikasikan desain yang mewah atau berperforma tinggi. **Klaster 3** memiliki variasi yang lebar dalam hal panjang dan tenaga kuda, yang dapat mencakup kapal klasik atau spesialisas tertentu. Terakhir, **Klaster 4** dominan dengan kapal baru yang memiliki ukuran dan tenaga kuda yang bervariasi, cenderung mengarah ke model yang lebih modern dan efisien. Berikut coding dan grafik terkait *Clustering of Boats by Length, TotalHP, and Price*:





Jika seorang pelanggan menginginkan kapal dengan panjang total 80 kaki dan total tenaga kuda sebesar 2000, berdasarkan grafik dibawah, kapal tersebut cenderung termasuk dalam **klaster 2**. Dalam klaster ini, kapal-kapal umumnya memiliki panjang sekitar 80 kaki dan total tenaga kuda sekitar 2000. Berdasarkan data, perkiraan biaya untuk kapal ini adalah **sekitar 2.000.000 hingga 3.000.000**.