

**TUGAS PERTEMUAN 3**  
**MATA KULIAH STATISTIKA KOMPUTASI (A)**



**DISUSUN OLEH:**

- |                         |               |
|-------------------------|---------------|
| 1. Winna Dhestyan Putri | (22083010015) |
| 2. Dwi Amalia Putri     | (22083010042) |

**DOSEN PENGAMPU:**

Trimono, S.Si., M.Si

**FAKULTAS ILMU KOMPUTER**  
**PROGRAM STUDI SAINS DATA**  
**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR**  
**2023**

A. Informasi tentang situs/web yang kami gunakan untuk *web scrapping*

1. Nama situs : transfermarkt.co.id
2. Link web : [https://www.transfermarkt.co.id/spieler-statistik/wertvollstespieler/marktwertetop?land\\_id=0&ausrichtung=alle&spielerposition\\_id=alle&altersklasse=alle&jahrgang=0&kontinent\\_id=0&plus=1](https://www.transfermarkt.co.id/spieler-statistik/wertvollstespieler/marktwertetop?land_id=0&ausrichtung=alle&spielerposition_id=alle&altersklasse=alle&jahrgang=0&kontinent_id=0&plus=1)
3. Gambaran umum web. Web berisi pemain paling berharga jika dilihat dari harga pasarannya di seluruh dunia. Pada web berisi tabel yang berisi 15 variabel, meliputi variabel nama pemain, posisi pemain, umur, negara asal, klub, harga pasaran, banyaknya pertandingan, banyaknya gol, gol bunuh diri dan umpan gol, banyak pemain mendapat kartu kuning, kartu kuning kedua, dan kartu merah, status pemain pada pertandingan (pemain pengganti masuk, pemain pengganti keluar)
4. Pada web kami hanya akan menggunakan x variabel, yaitu nama pemain,

B. Langkah langkah web scraping

1. Install package rvest dan xml2, dengan cara `install.packages("nama_package")`. Jika sudah terinstall, panggil library dengan code : `library(package)`

```
library(xml2)
library(rvest)
```

Library tersebut berfungsi untuk dapat menjalankan fungsi `read_html`

2. Import data dari web ke Rstudio

```
> #Import data web ke R
> alamatweb = 'https://www.transfermarkt.co.id/spieler-statistik/wertvollstespieler/marktwertetop?land_id=0&ausrichtung=alle&spielerposition_id=alle&altersklasse=alle&jahrgang=0&kontinent_id=0&plus=1'
> lamanweb = read_html(alamatweb)
> lamanweb
{html_document}
<html lang="id">
[1] <head>\n<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">\n<script type="text/javascri ...
[2] <body>\n<div class="werbung werbung-header">\n<!-- Start GPT Tag -->\n<script as ...
```

Variabel “alamatweb” digunakan untuk menampung URL target yang akan kita gunakan untuk mengumpulkan data, memanggil server web dan memarsing dari data tersebut. Selanjutnya URL tersebut digunakan fungsi `read_html` dan ditampung pada variabel `lamanweb`.

3. Mengambil data pada variabel tertentu dengan `selectorgadget`

Untuk mengambil suatu informasi dari web (web scrapping), perlu download terlebih dahulu `selectorgadget` pada google. Selanjutnya tekan pada area, teks atau informasi yang akan dimasukkan ke R. Kemudian copy tulisan yang muncul di pojok kanan bawah. Code script untuk meng-import data sebagai berikut:

```
nama_data=html_nodes(variabel_penampung, “.class deskriptornya”)
```

Untuk melakukan ekstraksi node dari XML objek, gunakan `html_nodes()`, dan diikuti dengan `.` untuk menandakan *class* deskriptornya (didapat ketika menekan informasi dengan selector gadget di pojok kanan bawah). Output yang akan dihasilkan adalah list dari semua node yang ditemukan

```
nama_data_fix=html_text(variabel_penampung, “. class deskriptornya”)
```

Gunakan `html_text()` pada node yang sudah kita temukan untuk mengekstraksi *tagged data*,

Selain cara diatas, juga dapat menggunakan “cara langsung” sebagai berikut

```
nama_data =variabel_penampung %>% html_nodes(“.class deskriptornya”) %>% html_text()
```

- Variabel 1, nama pemain

```
> namapemain_data <- html_nodes(lamanweb, '.inline-table a')
> nama_pemain <- html_text(namapemain_data)
> nama_pemain
 [1] "" "Kyllian Mbappé" "" "Erling Haaland" ""
 [6] "Vinicius Junior" "" "Jude Bellingham" "" "Phil Foden"
[11] "" "Pedri" "" "Jamal Musiala" ""
[16] "Bukayo Saka" "" "Federico Valverde" "" "Gavi"
[21] "" "Aurélien Tchouameni" "" "Harry Kane" ""
[26] "Rafael Leão" "" "Rodrygo" "" "Declan Rice"
[31] "" "Rodri" "" "Dusan Vlahovic" ""
[36] "Christopher Nkunku" "" "Bernardo Silva" "" "Joshua Kimmich"
[41] "" "Mohamed Salah" "" "Kevin De Bruyne" ""
[46] "Antony" "" "Luis Díaz" "" "Josko Gvardiol"
> nama_pemain <- nama_pemain[nama_pemain!=""]
> nama_pemain
 [1] "Kyllian Mbappé" "Erling Haaland" "Vinicius Junior" "Jude Bellingham" "Phil Foden"
 [6] "Pedri" "Jamal Musiala" "Bukayo Saka" "Federico Valverde" "Gavi"
[11] "Aurélien Tchouameni" "Harry Kane" "Rafael Leão" "Rodrygo" "Declan Rice"
[16] "Rodri" "Dusan Vlahovic" "Christopher Nkunku" "Bernardo Silva" "Joshua Kimmich"
[21] "Mohamed Salah" "Kevin De Bruyne" "Antony" "Luis Díaz" "Josko Gvardiol"
```

Setelah mengekstrak data menggunakan `html_text`, ada data yang tidak bisa langsung digunakan, contohnya pada variabel pertama. Pada variabel "nama pemain" ini, terdapat data kosong jika tidak dihilangkan akan mengganggu posisi data lain di data frame, oleh karena itu data yang kosong (“”) dihilangkan dari list

- Variabel ke-2, posisi

```
> posisipemain_data <- html_nodes(lamanweb, '.inline-table tr+ tr td')
> posisi_pemain <- html_text(posisipemain_data)
> posisi_pemain
 [1] "Depan-Tengah" "Depan-Tengah" "Sayap Kiri" "Gel. Tengah" "Sayap Kiri" "Gel. Tengah" "Gel. Serang"
 [8] "Sayap Kanan" "Gel. Tengah" "Gel. Tengah" "Gel. Bertahan" "Depan-Tengah" "Sayap Kiri" "Sayap Kanan"
[15] "Gel. Bertahan" "Gel. Bertahan" "Depan-Tengah" "Depan-Kedua" "Gel. Serang" "Gel. Bertahan" "Sayap Kanan"
[22] "Gel. Serang" "Sayap Kanan" "Sayap Kiri" "Bek-Tengah"
```

Pada variabel kedua, setelah diekstrak data bisa langsung digunakan

- Variabel ke-3, Umur

```
> umur_pemain_data <- html_nodes(lamanweb,'td:nth-child(3)')
> umur_pemain <- html_text(umur_pemain_data)
> umur_pemain
[1] "24" "22" "22" "19" "22" "20" "20" "21" "24" "18" "23" "29" "23" "22" "24" "26" "23" "25" "28" "28" "30" "31"
[23] "23" "26" "21"
> umur_pemain <- as.numeric(umur_pemain)
> umur_pemain
[1] 24 22 22 19 22 20 20 21 24 18 23 29 23 22 24 26 23 25 28 28 30 31 23 26 21
```

Dari gambar di atas, setelah diekstrak menggunakan `html_text`, variabel “umur” yang seharusnya bertipe integer atau numerik namun masih dalam bentuk string. Oleh karena itu, data pada variabel umur diubah dari tipe string atau karakter ke numerik dengan fungsi `as.numeric(nama_data)`

- Variabel ke-4, harga pasaran pemain

```
> harga_pasaran_data <- html_nodes(lamanweb,'.rechts.hauptlink')
> harga_pasaran <- html_text(harga_pasaran_data)
> harga_pasaran
[1] "Rp3.128,70Mlyr. " "Rp2.954,88Mlyr. " "Rp2.085,80Mlyr. " "Rp1.911,98Mlyr. " "Rp1.911,98Mlyr. "
[6] "Rp1.738,16Mlyr. " "Rp1.738,16Mlyr. " "Rp1.738,16Mlyr. " "Rp1.738,16Mlyr. " "Rp1.564,35Mlyr. "
[11] "Rp1.564,35Mlyr. " "Rp1.564,35Mlyr. " "Rp1.477,44Mlyr. " "Rp1.390,53Mlyr. " "Rp1.390,53Mlyr. "
[16] "Rp1.390,53Mlyr. " "Rp1.390,53Mlyr. " "Rp1.390,53Mlyr. " "Rp1.390,53Mlyr. " "Rp1.390,53Mlyr. "
[21] "Rp1.390,53Mlyr. " "Rp1.390,53Mlyr. " "Rp1.303,62Mlyr. " "Rp1.303,62Mlyr. " "Rp1.303,62Mlyr. "
> harga_pasaran <- substring(harga_pasaran,3,7) #Menghilangkan Rp dan Mlyr
> harga_pasaran <- as.numeric(harga_pasaran)
> harga_pasaran <- harga_pasaran*1000
> harga_pasaran
[1] 3128 2954 2085 1911 1911 1738 1738 1738 1738 1564 1564 1564 1477 1390 1390 1390 1390 1390 1390 1390 1390 1390
[23] 1303 1303 1303
```

Setelah diekstrak, data pada variabel harga pasaran pemain masih terdapat poin yang harus dihilangkan seperti “Rp” dan “Mlyr” dan perubahan bentuk format angka. Pada Rstudio simbol “,” (koma) tidak bernilai pada bilangan, sementara simbol “.” (titik), menunjukkan float pada bilangan.

Untuk menghilangkan kata “Rp” dan “Mlyr”, kami mengambil poin atau digit pada urutan ke 3 sampai urutan ke 7, dengan menggunakan fungsi `substring(nama_data, urutan awal, urutan akhir)`. Karena data masih bertipe string (“”) yang seharusnya bertipe numerik, maka diubah dengan fungsi `as.numeric(nama_data)`. Untuk mengubah format bilangan yang sesuai dengan nominal uang, yang semula x.xxx dalam satuan miliar dikalikan 1000, menjadi xxxx miliar

- Variabel ke-5, banyak pertandingan yang pernah diikuti pemain

```
> banyak_pertandingan_data <- html_nodes(lamanweb,'.hauptlink+ .zentriert')
> banyak_pertandingan <- html_text(banyak_pertandingan_data)
> banyak_pertandingan <- as.numeric(banyak_pertandingan)
> banyak_pertandingan
[1] 33 39 38 35 34 32 36 38 39 40 33 38 37 37 33 40 25 32 42 37 37 37 29 14 30
```

Setelah diekstrak data perlu dijadikan ke dalam bentuk integer dengan menggunakan fungsi `as.numeric`

- Variabel ke-6, banyak gol

```
> gol=lamanweb %>% html_nodes('td:nth-child(8)') %>% html_text() %>% as.numeric()
> gol
[1] 29 39 18 10 10 7 14 10 12 4 0 21 9 10 1 2 12 17 3 6 21 6 8 4 2
```

Karena data butuh diubah dari tipe string menjadi numerik maka ditambah kode %>% as.numeric() pada cara langsung

- Variabel ke-7, banyak umpan sehingga dapat mencetak gol

```
> umpan_gol=lamanweb %>% html_nodes('td:nth-child(10)') %>% html_text() %>% as.numeric()
> umpan_gol
[1] 6 5 9 7 5 0 12 9 4 5 2 3 10 8 3 6 7 7 8 9 10 19 3 3 0
```

- Variabel ke-8, akumulasi perolehan kartu kuning

```
> kartu_kuning=lamanweb %>% html_nodes('td:nth-child(11)') %>% html_text() %>% as.numeric()
> kartu_kuning
[1] 4 3 9 7 2 3 1 6 4 7 3 5 5 1 3 7 0 2 6 7 0 1 4 2 1
```

- Setelah mengambil, mengekstrak dan memperbaiki format data, selanjutnya data data tersebut dikumpulkan dalam bentuk dataframe dengan menggunakan fungsi data.frame(nama\_variabel(yang akan muncul sebagai judul kolom)=nama\_data)  
Selanjutnya untuk melihat 6 baris data teratas dari dataset, perlu menggunakan fungsi head(nama\_dataset)

```
> dataset_pemain_bernilai <-data.frame>Nama_Pemain = nama_pemain, Posisi = posisi_pemain, Umur = umur_pemain,
+ Harga_Pasaran_M= harga_pasaran, Pertandingan=banyak_pertandingan,total_Gol=gol,
+ Total_UmpaGol=umpangol,total_Kartukuning=kartukuning)
> head(dataset_pemain_bernilai)
  Nama_Pemain Posisi Umur Harga_Pasaran_M Pertandingan total_Gol Total_UmpaGol total_Kartukuning
1 Kylian Mbappé Depan-Tengah 24 3128 33 29 6 4
2 Erling Haaland Depan-Tengah 22 2954 39 39 5 3
3 Vinicius Junior Sayap Kiri 22 2085 38 18 9 9
4 Jude Bellingham Gel. Tengah 19 1911 35 10 7 7
5 Phil Foden Sayap Kiri 22 1911 34 10 5 2
6 Pedri Gel. Tengah 20 1738 32 7 0 3
```

## C. Analisis Pendahuluan

- Informasi umum mengenai dataset

```
> str(dataset_pemain_bernilai) #menampilkan jumlah observasi dan variabel pada data frame yang telah kita buat
'data.frame': 25 obs. of 8 variables:
 $ Nama_Pemain : chr "Kylian Mbappé" "Erling Haaland" "Vinicius Junior" "Jude Bellingham"
...
 $ Posisi : chr "Depan-Tengah" "Depan-Tengah" "Sayap Kiri" "Gel. Tengah" ...
 $ Umur : num 24 22 22 19 22 20 20 21 24 18 ...
 $ Harga_Pasaran_M : num 3128 2954 2085 1911 1911 ...
 $ Pertandingan : num 33 39 38 35 34 32 36 38 39 40 ...
 $ total_Gol : num 29 39 18 10 10 7 14 10 12 4 ...
 $ Total_UmpaGol : num 6 5 9 7 5 0 12 9 4 5 ...
 $ total_Kartukuning: num 4 3 9 7 2 3 1 6 4 7 ...
```

Dengan menggunakan fungsi str(nama dataset), akan muncul rincian tipe data dari setiap variabel dan juga jumlah observasi (baris data) dan banyak variabel (kolom data)

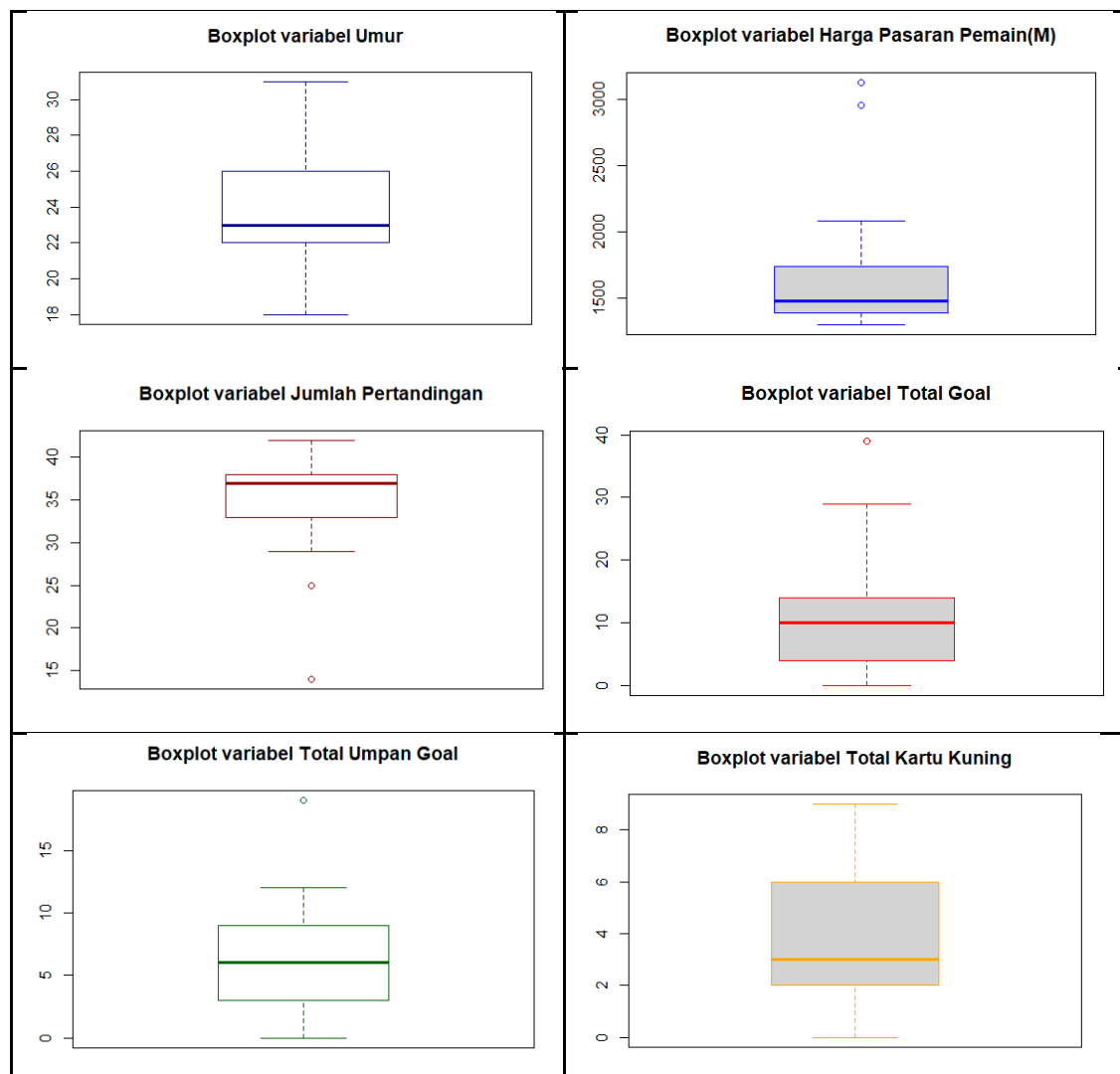
- Ada tidaknya nilai missing value pada data

```
> colSums(is.na(dataset_pemain_bernilai))
  Nama_Pemain      Posisi      umur  Harga_Pasaran_M  Pertandingan  total_Gol  Total_UmpaGol
1           0           0           0           0           0           0           0
```

Untuk mengetahui secara rinci ada tidaknya missing value pada setiap kolom, dengan menggunakan kode `colSums(is.na(nama_dataset))`

### 3. Ada tidaknya nilai outlier

```
boxplot(dataset_pemain_bernilai$umur, main='Boxplot variabel Umur',
        col='white',border='darkblue')
boxplot(dataset_pemain_bernilai$Harga_Pasaran_M, main='Boxplot variabel Harga Pasaran Pemain(M)',
        col='lightgrey',border='blue')
boxplot(dataset_pemain_bernilai$Pertandingan, main='Boxplot variabel Jumlah Pertandingan',
        col='white',border='darkred')
boxplot(dataset_pemain_bernilai$total_Gol, main='Boxplot variabel Total Goal',
        col='lightgrey',border='red')
boxplot(dataset_pemain_bernilai$total_Umpun_Gol, main='Boxplot variabel Total Umpun Goall',
        col='white',border='darkgreen')
boxplot(dataset_pemain_bernilai$total_KartuKuning, main='Boxplot variabel Total Kartu Kuning',
        col='lightgrey',border='orange')
```



Terdapat nilai outlier pada beberapa variabel, tepatnya 2 data pada variabel harga pasaran pemain, 2 data pada jumlah pertandingan, 1 data pada variabel total goal, dan 1 data pada variabel total umpun goal. Ada tidaknya nilai outlier ditandai dengan ada

tidaknya bulatan kecil di atas atau bawah batas data. Menurut kami, nilai pada kasus/data ini tidak perlu untuk dihilangkan atau diganti nilainya dengan nilai median, batas atas/bawah dan modus. Nilai outlier ini ada karena data yang kami peroleh bukan data keseluruhan pemain (hanya ada 25 baris data yang dapat menggunakan fitur selectorgadget), disamping itu faktor profesionalitas dan juga pengalaman masing masing pemain di dunia sepak bola yang juga diperhitungkan.

#### 4. Menghitung korelasi antar variabel

##### a. Hubungan variabel harga pasaran pemain dalam miliar dengan variabel umur

```
> cor.test(dataset_pemain_bernilai$Harga_Pasaran_M, dataset_pemain_bernilai$umur, method="pearson")

Pearson's product-moment correlation

data: dataset_pemain_bernilai$Harga_Pasaran_M and dataset_pemain_bernilai$umur
t = -1.3806, df = 23, p-value = 0.1807
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.6055728  0.1330375
sample estimates:
      cor 
-0.2766353
```

Nilai korelasi antara variabel harga pasaran dengan variabel umur adalah -0,277 yang berarti hubungan antara variabel lemah (rendah). Minus (-) dalam nilai korelasi ini menunjukkan bahwa semakin tua umur pemain maka semakin rendah harga pasarannya, dan sebaliknya

##### b. Hubungan variabel harga pasaran pemain dalam miliar dengan banyaknya pertandingan yang telah diikuti

```
> cor.test(dataset_pemain_bernilai$Harga_Pasaran_M, dataset_pemain_bernilai$Pertandingan, method="pearson")

Pearson's product-moment correlation

data: dataset_pemain_bernilai$Harga_Pasaran_M and dataset_pemain_bernilai$Pertandingan
t = 0.96013, df = 23, p-value = 0.347
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.2155450  0.5488626
sample estimates:
      cor 
0.1963049
```

Nilai korelasi antara variabel harga pasaran dengan variabel banyak pertandingan adalah 0,196 yang berarti hubungan antara variabel lemah (sangat rendah). Nilai korelasi yang positif, menunjukkan bahwa semakin banyak pertandingan yang pernah diikuti pemain maka semakin tinggi harga pasarannya, dan sebaliknya.

##### c. Hubungan variabel harga pasaran pemain dalam miliar dengan variabel total gol

```
> cor.test(dataset_pemain_bernilai$Harga_Pasaran_M,dataset_pemain_bernilai$total_Gol,method="pearson")

Pearson's product-moment correlation

data: dataset_pemain_bernilai$Harga_Pasaran_M and dataset_pemain_bernilai$total_Gol
t = 5.6193, df = 23, p-value = 1.018e-05
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5225672 0.8886767
sample estimates:
      cor
0.7606394
```

Nilai korelasi antara variabel harga pasaran dengan variabel banyak total gol pemain adalah 0,760 yang berarti hubungan antara variabel sangat kuat (kuat). Nilai korelasi yang positif, menunjukkan bahwa semakin banyak total gol yang dicetak oleh pemain maka semakin tinggi harga pasarannya, dan sebaliknya

- d. Hubungan variabel harga pasaran pemain dalam miliar dengan variabel total\_umpan\_gol

```
> cor.test(dataset_pemain_bernilai$Harga_Pasaran_M,dataset_pemain_bernilai$Total_UmpaGol,method="pearson")

Pearson's product-moment correlation

data: dataset_pemain_bernilai$Harga_Pasaran_M and dataset_pemain_bernilai$Total_UmpaGol
t = -0.19233, df = 23, p-value = 0.8492
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.4284183 0.3607724
sample estimates:
      cor
-0.04007064
```

Nilai korelasi antara variabel harga pasaran dengan variabel banyak total umpan gol adalah -0,040 yang berarti hubungan antara variabel kurang berarti (sangat rendah). Minus (-) dalam nilai korelasi ini menunjukkan bahwa semakin banyak total umpanan pemain yang menghasilkan gol maka semakin rendah harga pasarannya, dan sebaliknya.

- e. Hubungan variabel harga pasaran pemain dalam miliar dengan variabel total\_KartuKuning

```
> cor.test(dataset_pemain_bernilai$Harga_Pasaran_M,dataset_pemain_bernilai$total_KartuKuning,method="pearson")

Pearson's product-moment correlation

data: dataset_pemain_bernilai$Harga_Pasaran_M and dataset_pemain_bernilai$total_KartuKuning
t = 0.80173, df = 23, p-value = 0.4309
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.2462936 0.5257608
sample estimates:
      cor
0.1648835
```

Nilai korelasi antara variabel harga pasaran dengan variabel banyak perolehan kartu kuning adalah 0,165 yang berarti hubungan antara variabel lemah (sangat rendah). Nilai korelasi yang positif, menunjukkan bahwa semakin banyak perolehan kartu kuning pemain maka semakin tinggi harga pasarannya, dan sebaliknya.



## D. Analisis Deskriptif

1. Mencari rangkuman statistika deskriptif (seperti mean, median, standardevisi, nilai max, nilai min, Q1, Q3, dll) dari data dengan library summarytools

```
> library(summarytools)
> descr(dataset_pemain_bernilai)
Non-numerical variable(s) ignored: Nama_Pemain, Posisi
Descriptive Statistics
dataset_pemain_bernilai
N: 25
```

	Harga_Pasaran_M	Pertandingan	total_Gol	total_Kartukuning	Total_UmpaGol
Mean	1661.16	34.60	11.00	3.72	6.40
Std. Dev	468.15	5.80	9.20	2.53	4.09
Min	1303.00	14.00	0.00	0.00	0.00
Q1	1390.00	33.00	4.00	2.00	3.00
Median	1477.00	37.00	10.00	3.00	6.00
Q3	1738.00	38.00	14.00	6.00	9.00
Max	3128.00	42.00	39.00	9.00	19.00
MAD	128.99	4.45	8.90	2.97	4.45
IQR	348.00	5.00	10.00	4.00	6.00
CV	0.28	0.17	0.84	0.68	0.64
Skewness	1.99	-1.82	1.29	0.29	0.90
SE. Skewness	0.46	0.46	0.46	0.46	0.46
Kurtosis	3.30	3.92	1.42	-1.10	1.42
N. Valid	25.00	25.00	25.00	25.00	25.00
Pct. Valid	100.00	100.00	100.00	100.00	100.00

Table: Table continues below

```
-----
              umur
-----
Mean          23.76
Std. Dev       3.42
Min            18.00
Q1             22.00
Median         23.00
Q3             26.00
Max            31.00
MAD            2.97
IQR            4.00
CV             0.14
Skewness       0.47
SE. Skewness   0.46
Kurtosis      -0.69
N. Valid      25.00
Pct. Valid    100.00
```

2. Mencari nilai modus dari data

```
> modus<-function(x){
+   uniqx<-unique(x)
+   uniqx[which.max(tabulate(match(x,uniqx)))]
+ }
> modus(dataset_pemain_bernilai$Posisi)
[1] "Depan-Tengah"
> modus(dataset_pemain_bernilai$umur)
[1] 22
> modus(dataset_pemain_bernilai$Harga_Pasaran_M)
[1] 1390
> modus(dataset_pemain_bernilai$Pertandingan)
[1] 37
> modus(dataset_pemain_bernilai$total_Gol)
[1] 10
> modus(dataset_pemain_bernilai$Total_UmpaGol)
[1] 3
> modus(dataset_pemain_bernilai$total_Kartukuning)
[1] 3
```

3. Menghitung banyaknya data (count)

```

> count(dataset_pemain_bernilai, Posisi)
  Posisi n
1 Bek-Tengah 1
2 Depan-Kedua 1
3 Depan-Tengah 4
4 Gel. Bertahan 4
5 Gel. Serang 3
6 Gel. Tengah 4
7 Sayap Kanan 4
8 Sayap Kiri 4
> # mencari jumlah pemain dengan umur kurang dari rata rata
> count(dataset_pemain_bernilai, Umur<mean(Umur))
  Umur < mean(Umur) n
1 FALSE 11
2 TRUE 14
> # mencari jumlah pemain dengan harga pasaran > harga pasaran rata rata
> count(dataset_pemain_bernilai, Harga_Pasaran_M>mean(Harga_Pasaran_M))
  Harga_Pasaran_M > mean(Harga_Pasaran_M) n
1 FALSE 16
2 TRUE 9

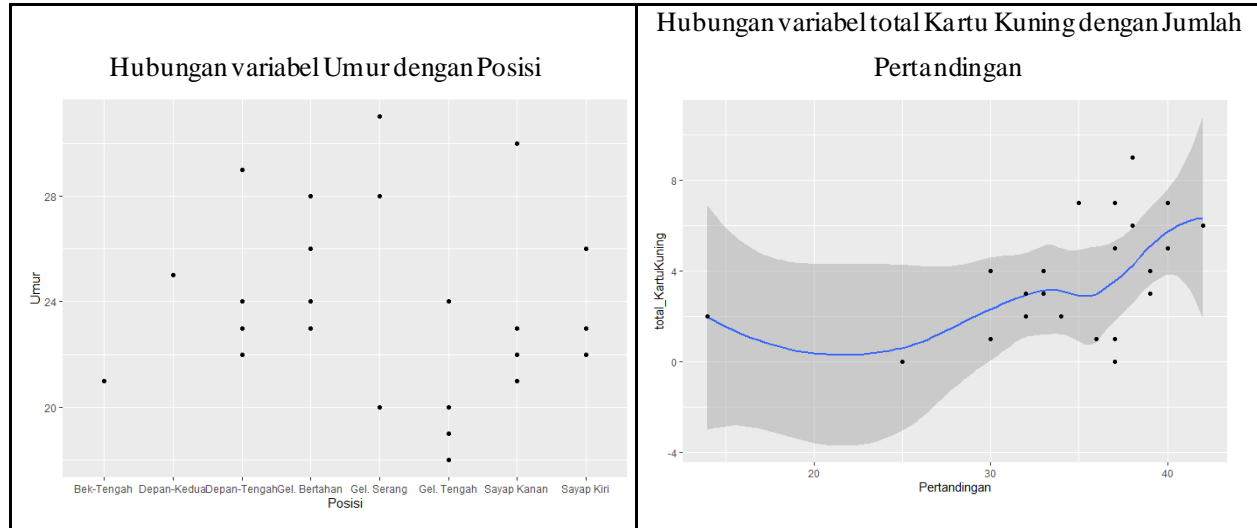
```

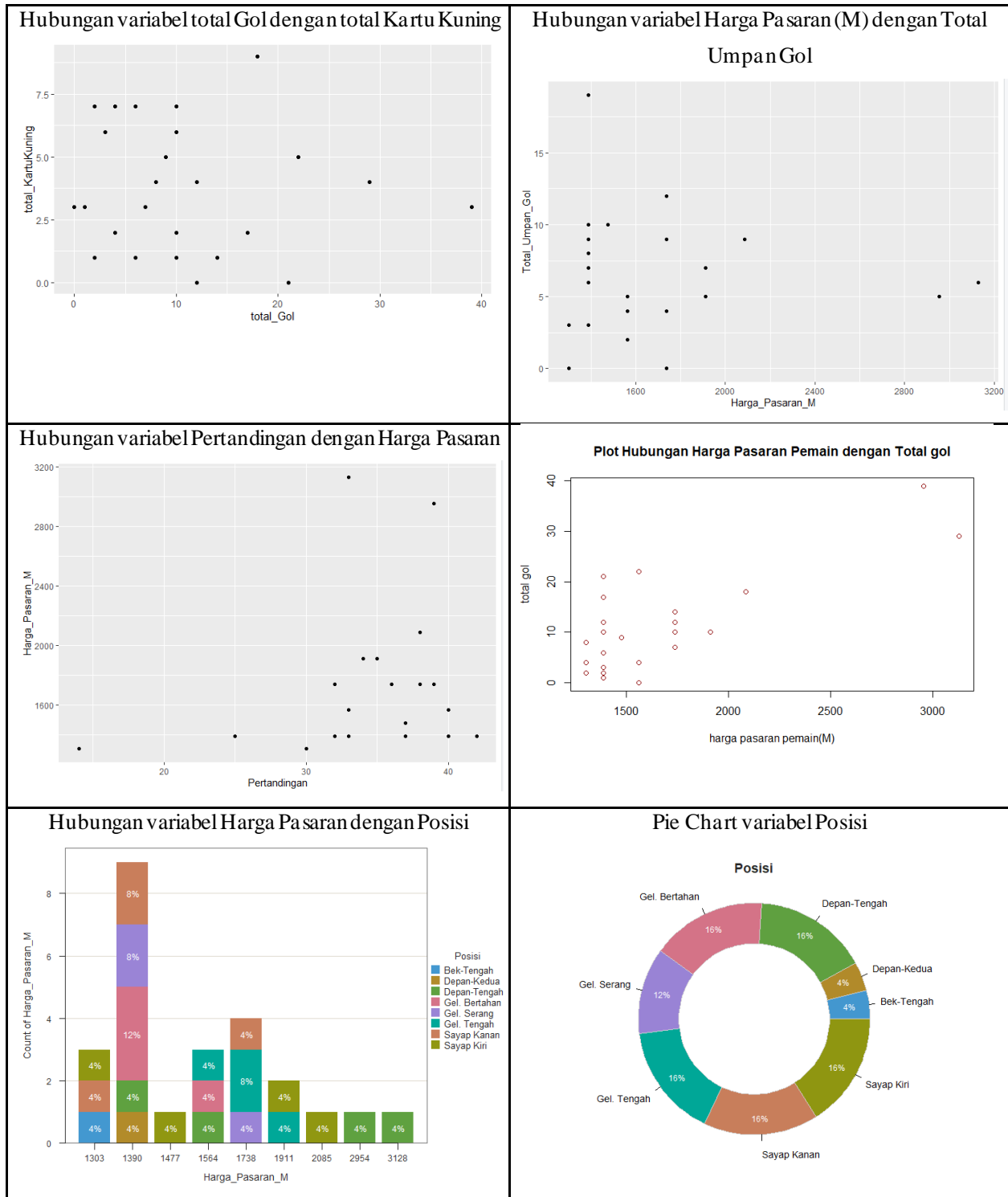
## E. Visualisasi data

```

#visualisasi data
library(ggplot2)
ggplot(dataset_pemain_bernilai,aes(x=Posisi,y=Umur))+geom_point()
ggplot(dataset_pemain_bernilai,aes(x=Pertandingan,y=total_KartuKuning))+geom_smooth()+geom_point()
ggplot(dataset_pemain_bernilai,aes(x=Pertandingan, y=Harga_Pasaran_M))+geom_point()
library(lessR)
plot(dataset_pemain_bernilai$Harga_Pasaran_M,dataset_pemain_bernilai$total_gol,
      main='Plot Hubungan Harga Pasaran Pemain dengan Total gol',
      xlab='harga pasaran pemain(M)',ylab='total gol',
      col='darkred')
BarChart(Harga_Pasaran_M,data=dataset_pemain_bernilai, by=Posisi)
PieChart(Posisi, data=dataset_pemain_bernilai)

```





## F. Analisis deskriptif terhadap data tersebut dan Interpretasinya

- Variabel pertama, yaitu nama pemain. Pada variabel ini, data bertipe karakter. Variabel ini berisi 25 daftar nama pemain yang dianggap bernilai di dunia sepak bola. Kylian

Mbappé merupakan pemain dengan jumlah bayaran (harga pasaran) tertinggi pada dataset kami. Erling Haaland adalah pemain pencetak gol terbanyak. Pemain sepak bola tertua dari dataset yang kami himpun atas nama Kevin De Bruyne.

- b) Variabel kedua, yaitu posisi pemain. Dalam dataset yang kami gunakan ada 9 jenis posisi pemain yang terdiri dari posisi bek-tengah, sayap kiri, sayap kanan, gel-serang, gel-bertahan, depan-kedua, depan tengah, gel-tengah, depan-tengah. Modus dari variabel posisi ini adalah Depan-Tengah, Gel. Bertahan, Gel. Tengah, Sayap Kanan, Sayap Kiri, masing masing jenis muncul 4 kali pada dataset
- c) Variabel ketiga, yaitu umur. Dalam dataset yang kami gunakan, rentang umur pemain dari 18 sampai 31 tahun, dengan rata rata umur 23,76 tahun. Nilai tengah dari variabel umur 23, 22 merupakan nilai Q1, dan 26 merupakan nilai Q3. Modus atau data yang sering muncul yaitu umur 22 tahun. Jumlah umur pemain yang umurnya kurang dari rata rata sebanyak 14 pemain. Usia paling muda (18 tahun) yaitu Gavi yang berperan sebagai gelandang tengah, telah bertanding sebanyak 40 kali.
- d) Variabel keempat, yaitu harga pasaran. Dalam dataset yang kami gunakan rentang harga pasaran yang tertera dari harga Rp. 1303 milyar sampai dengan Rp. 3128 milyar. Pada variabel ini rata rata harga pasaran pemain 1661,16 milyar, modus dari data harga pasaran pemain adalah 1390 milyar, dengan nilai tengah 1477. Kuartil bawah dari variabel ini 1390 dan kuartil atasnya 1738. Sebanyak 9 pemain dengan harga pasaran diatas rata rata, dengan Kylian Mbappé sebagai pemain dengan harga pasaran tertinggi dan Josko Gvardiol, Antony, Luis Díaz sebagai pemain dengan harga pasaran terendah
- e) Variabel kelima, yaitu banyak pemain bertanding sepak bola. Dalam dataset yang kami gunakan pemain bertanding paling sedikit sebanyak 14 kali dan jumlah pertandingan paling banyak sebanyak 42 kali. Rata rata setiap pemain bertanding sebanyak 34,60 kali, dengan banyak pertandingan yang sering muncul dalam data sebanyak 37. Q1 dari variabel ini 33, nilai Q2 variabel ini adalah 37, dan nilai Q3 dari variabel ini adalah 38. Pemain “Luis Díaz” merupakan pemain dengan riwayat pertandingan paling sedikit dan pemain “Bernardo Silva” merupakan pemain yang bertanding paling banyak yaitu 24 kali.
- f) Variabel keenam, yaitu total gol. Dalam dataset yang kami gunakan total gol yang diperoleh paling sedikit 0 gol oleh pemain ” Aurélien Tchouameni” dan jumlah gol

tertinggi sebanyak 39 kali gol oleh pemain “Erling Haaland”. Modus dari variabel total gol ini adalah 10 dan median (Q2) sebanyak 10. Q1 dari variabel ini sebanyak 4 dan Q3 sebanyak 14.

- g) Variabel ketujuh, yaitu total umpan gol. Dalam dataset yang kami gunakan terdapat paling sedikit yaitu 0 kali dan paling banyak yaitu 19 kali. Q1 dari data adalah 3, Q2 sebesar 6, dan Q3 sebesar 9. Rata rata pemain pernah melakukan umpan sehingga dapat tercetak gol sebanyak 6,40. Pemain “Kevin De Bruyne” yang berposisi sebagai Gelandang serang merupakan pemain dengan umpan gol paling banyak. Modus dari variabel ini adalah 3.
- h) Variabel terakhir, yaitu total kartu kuning yang diperoleh setiap pemain. Dalam dataset yang kami gunakan yaitu terdapat paling sedikit mendapatkan kartu kuning sebanyak 0 kali dan paling banyak mendapatkan kartu kuning sebanyak 9 kali oleh “Vinicius Junior”. Berdasarkan data, pemain mendapatkan kartu kuning dengan rata rata 3,72 kali. Modus dari variabel ini adalah 3, artinya banyak pemain yang pernah mendapatkan kartu kuning sebanyak 3 kali. Nilai median dari variabel ini adalah 3.

#### Analisis keseluruhan data

1. Berdasarkan nilai korelasi, Hubungan variabel harga pasaran pemain dalam miliar dengan variabel total gol merupakan hubungan yang terkuat pada dataset ini dengan nilai korelasi sebesar 0,760. Sedangkan hubungan yang paling lemah ditunjukkan oleh Hubungan variabel harga pasaran pemain dalam miliar dengan variabel total\_umpan\_gol dengan nilai korelasi hanya sebesar -0,040. Dengan nilai korelasi, kita dapat mengetahui tingkat hubungan setiap variabel dengan variabel harga pasaran, dan juga dapat menunjukkan faktor atau variabel mana saja yang mempengaruhi besar kecilnya variabel harga pasaran pemain.
2. Dari visualisasi data di atas, diperoleh informasi bahwa variabel posisi tidak memiliki hubungan yang kuat dengan variabel umur, hampir pada masing masing posisi memiliki rentang umur yang beragam, dan **tidak dapat disimpulkan** jika semakin muda atau tua umur pemain maka berkesempatan menjadi posisi x. Namun jika diperiksa secara detail dengan memahami pola bagan, di samping adanya perbedaan skill dan keterampilan pemain, pemain yang berposisi sebagai “Gel-tengah”

didominasi oleh pemain muda ( $\leq 20$ ) sementara posisi "Depan tengah" didominasi oleh pemain dengan umur  $> 20$  dan  $\leq 24$ . Posisi "gelandang serang" memiliki sebaran umur yang paling beragam.

3. Pada bagan gabungan `geom_point()` dengan `geom_smooth()` diatas mengenai Hubungan antara variabel total Kartu Kuning dengan Jumlah Pertandingan. Sebagian besar point atau titik menunjukkan kecenderungan semakin ke kanan semakin ke atas, artinya semakin banyak riwayat pertandingan pemain maka semakin tinggi/banyak pemain mendapat kartu kuning. Dari data ini, dapat diperkirakan jika semakin lama atau sering pemain bertanding maka semakin besar kemungkinan pemain mendapatkan kartu kuning.
4. Tidak terdapat hubungan yang kuat antara variabel total Gol dengan total Kartu Kuning. Namun terlihat pada bagan, jika semakin sedikit perolehan kartu kuning pemain maka semakin banyak gol yang tercipta. Semakin banyak akumulasi perolehan kartu kuning maka semakin sedikit gol yang tercipta. Walaupun beberapa titik tidak menggambarkan demikian, terdapat total kartu kuning yang sedikit dan gol yang tercipta juga sedikit
5. Tidak terdapat hubungan yang kuat antara variabel Harga Pasaran (M) dengan Total Umpan Gol. Hal ini juga terlihat pada nilai korelasi antara keduanya yang lemah
6. Seperti nilai korelasi sebelumnya, Hubungan variabel Pertandingan dengan Harga Pasaran berbanding lurus, sama halnya hubungan variabel total gol dengan harga pasaran hak tersebut berarti semakin tinggi harga pasaran pemain maka semakin besar total riwayat pertandingan pemain dan semakin besar total gol yang tercipta.