

**PROYEK UJIAN AKHIR SEMESTER**

**PEMROSESAN DATA**

***“Analisis Online Game Menggunakan Web Scraping”***

Dosen Pengampu: Alfi Fadliana S.Si., M.Stat.



Disusun oleh:

**Kelompok 10**

Muhammad Riski Alde	(3322600010)
Eky Fernanda Setyawan Putra	(3322600025)
Indah Sulistyowati	(3322600006)

**PROGRAM STUDI D4 SAINS DATA TERAPAN**

**DEPARTEMEN TEKNIK INFORMATIKA DAN KOMPUTER**

**POLITEKNIK ELEKTRONIKA NEGERI SURABAYA**

**2023**

## DAFTAR ISI

DAFTAR ISI.....	ii
DAFTAR GAMBAR .....	iii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
BAB II TINJAUAN PUSTAKA.....	3
2.1 Web Scraping .....	3
2.2 Online Game .....	4
BAB III TAHAPAN ANALISIS DATA .....	5
3.1 Metode Penelitian.....	5
3.2 Sumber Data.....	5
3.3 Langkah – langkah analisis .....	5
BAB IV HASIL DAN PEMBAHASAN .....	7
4.1 Analisis Deskripsi Data .....	7
4.2 Analisis Deskriptif dan Visualisasi Data.....	7
BAB V KESIMPULAN .....	15
DAFTAR PUSTAKA .....	16
LAMPIRAN .....	17

## DAFTAR GAMBAR

<b>Gambar 1.</b> Deskripsi Statistik Data Online Game .....	7
<b>Gambar 2.</b> Box Plot Data Pemrosesan .....	8
<b>Gambar 3.</b> Histogram Data Pemrosesan .....	9
<b>Gambar 4.</b> Heatmap Data Pemrosesan .....	9
<b>Gambar 5.</b> Boxplot Jumlah Pemain Saat ini dan Total Pemain .....	10
<b>Gambar 6.</b> Perbandingan Jumlah Pemain Saat Ini Untuk Setiap Permainan 1.....	11
<b>Gambar 7.</b> Perbandingan Jumlah Pemain Saat Ini untuk Setiap Permainan 2.....	11
<b>Gambar 8.</b> Perbandingan Jumlah Pemain Saat Ini untuk Setiap Permainan 3.....	12
<b>Gambar 9.</b> Perbandingan Jumlah Pemain Saat Ini untuk Setiap Permainan 4.....	12
<b>Gambar 10.</b> Distribusi Harga Game.....	13
<b>Gambar 11.</b> Menampilkan jumlah harga 0 dan lebih dari 0.....	14
<b>Gambar 12.</b> Prediksi jumlah pemain saat harga 200000.....	14

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Data sangat penting bagi bisnis dan organisasi karena membantu dalam membuat keputusan mereka saat ini, sebagian besar data dapat ditemukan di internet. Berbagai organisasi dan bisnis melakukan pengumpulan data melalui internet menggunakan Teknik web scraping untuk membantu mereka melakukan analisa dan membuat keputusan (Almaqabali et al, 2019).

Web scraping adalah proses ekstraksi data dari situs web secara otomatis dengan menggunakan bot atau program komputer. Dalam industri game, web scraping dapat digunakan untuk mengumpulkan data terkait online game, seperti informasi game, ulasan, peringkat, harga, statistik pemain, dan lain-lain dari situs web game. Dalam industri game online, web scraping menjadi alat yang berharga untuk mendapatkan wawasan pasar, melakukan analisis kompetitor, memantau tren, dan mengumpulkan data penting lainnya. web scraping digunakan untuk mengumpulkan informasi tentang game online, seperti judul game, deskripsi, ulasan, peringkat, dan informasi lainnya, dari berbagai sumber online.

Ada beberapa alasan web scraping penting dalam konteks game online yakni pertama, mendapatkan informasi game. Dengan web scraping, kita dapat mengumpulkan informasi tentang game online, seperti judul, genre, peringkat, ulasan, fitur, dan lainnya. Hal ini membantu pengembang dan pemain untuk memahami lebih lanjut tentang game yang ada dan membuat keputusan yang lebih baik. Kedua, Analisis kompetitor dimana Web scraping memungkinkan kita untuk mengumpulkan data tentang kompetitor, seperti daftar game yang mereka tawarkan, harga, ulasan, dan penilaian pengguna. Ini membantu pengembang dan pengecer untuk melacak dan memahami strategi kompetitor, serta membuat keputusan bisnis yang lebih baik.

Ketiga, penelitian pada pasar, dengan web scraping, kita dapat mengumpulkan data tentang preferensi pengguna, tren terkini, dan permintaan pasar terhadap game online. Ini membantu pengembang dalam penelitian pasar dan pengambilan keputusan strategis terkait pengembangan dan pemasaran game baru. Keempat, Penyesuaian harga yakni mengumpulkan data harga game dari situs web, web

scraping dapat membantu pengecer atau platform distribusi game untuk memantau harga game online dan menyesuaikan harga mereka secara kompetitif. Pada penelitian kali ini penulis akan melakukan web scarping dengan sumber web online game guna mengumpulkan dan menganalisa data terkait online game, seperti informasi game, peringkat, ulasan, statistik pemain, harga, dan lain-lain dari situs web game.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Web Scraping**

Web scraping telah menjadi teknik yang penting dalam dunia teknologi modern. Dengan menggunakan alat-alat dan teknik khusus, para pengembang dan peneliti dapat mengumpulkan dan mengolah data dari berbagai sumber secara efisien. Web scraping memungkinkan mereka untuk mengakses informasi yang diperlukan dengan cepat dan otomatis, dari situs web publik hingga platform media sosial, forum, dan banyak lagi. Web Scraping merupakan penggalian informasi dari situs web melalui perangkat lunak komputer, perangkat lunak ini dapat meniru cara manusia menjelajah world wide web (WWW), dengan menjalankan browser web yang lengkap, seperti Mozilla Firefox atau internet explorer, juga cara lain yakni melalui Hypertext Transfer Protocol (HTTP) tingkat rendah.

Web scraping adalah teknik untuk mengubah data web yang tidak terstruktur menjadi terstruktur data yang dapat disimpan dan dianalisis dalam database pusat atau spreadsheet (Sirisuriya, 2015). Tujuan utama dari web scraping adalah untuk mengambil data yang relevan dan berguna dari berbagai situs web secara efisien. Dalam prosesnya, web scraping dapat melibatkan pengambilan teks, gambar, link, tabel, dan data lainnya. Teknik ini telah menjadi penting dalam berbagai bidang seperti riset, analisis pasar, pengumpulan data untuk proyek analitik, dan pengembangan aplikasi web.

Web scraping memiliki banyak aplikasi yang beragam. Misalnya, dalam pengembangan aplikasi web, web scraping digunakan untuk mengumpulkan data dari situs-situs terkait atau mengintegrasikan konten dari sumber eksternal ke dalam aplikasi. Hal ini memungkinkan pengembang untuk menyajikan informasi terkini kepada pengguna dengan mudah.

Selain itu, web scraping juga digunakan dalam penelitian dan analisis data. Para peneliti dapat mengumpulkan data dari berbagai sumber untuk menganalisis tren, pola, dan informasi penting lainnya. Contohnya adalah penggunaan web scraping dalam analisis sentimen media sosial untuk memahami opini publik terkait produk atau layanan tertentu.

## 2.2 Online Game

Perkembangan teknologi dalam bidang online game telah mengalami kemajuan pesat dalam beberapa tahun terakhir. Online game merupakan situs yang menyediakan berbagai jenis permainan yang dapat melibatkan beberapa pengguna internet dari berbagai tempat yang berbeda untuk saling terhubung di waktu yang sama dengan melalui jaringan komunikasi online (Philips, 2009). Inovasi-inovasi baru telah muncul, mengubah lanskap industri game dan menciptakan pengalaman bermain yang lebih menarik dan imersif bagi para pemain. Salah satu perkembangan utama adalah kemajuan dalam teknologi grafis dan visual, yang telah menghasilkan game dengan grafis yang sangat realistis dan detail yang mengagumkan. Para pengembang game juga semakin fokus pada pengembangan sistem fisika yang canggih, sehingga pemain dapat merasakan simulasi yang lebih akurat dan realistis.

Selain itu, teknologi kecerdasan buatan (artificial intelligence) telah memainkan peran penting dalam meningkatkan kecerdasan karakter non-pemain (NPC) dalam game, membuat interaksi dengan karakter dalam game menjadi lebih alami dan menantang. Selain itu, perkembangan dalam teknologi jaringan telah memungkinkan pemain untuk terhubung dan berinteraksi dengan pemain lain di seluruh dunia, membentuk komunitas online yang besar dan aktif.

Melalui fitur-fitur seperti multiplayer online dan fitur berbagai konten, pemain dapat bermain bersama teman-teman mereka atau bahkan mengadu kemampuan dengan pemain yang tidak mereka kenal. Perkembangan teknologi dalam game online juga telah membuka pintu untuk peningkatan dalam permainan mobile, di mana pemain sekarang dapat menikmati pengalaman game yang mendalam dan kompleks melalui perangkat mobile mereka. Semua perkembangan ini mencerminkan komitmen industri game untuk terus berinovasi dan memberikan pengalaman yang lebih baik bagi para pemain, dan dengan teknologi terus berkembang, masa depan game online tampak sangat menjanjikan.

## **BAB III**

### **TAHAPAN ANALISIS DATA**

#### **3.1 Metode Penelitian**

Laporan ini menggunakan metode penelitian deskriptif yang bertujuan untuk mendeskripsikan dan mengidentifikasi karakteristik variable yang berpengaruh, menggambarkan keadaan yang ada, mencatat hubungan antara variabel-variabel, atau menemukan pola atau tren tertentu terhadap online game.

#### **3.2 Sumber Data**

Sumber data yang diambil pada penelitian ini yakni terkait informasi mengenai online game dengan link berikut: <https://store.steampowered.com/>

#### **3.3 Langkah – Langkah Analisis**

Berikut merupakan langkah – langkah dalam melakukan analisis data online game menggunakan teknik web scraping dengan python:

1. Pengumpulan Data

Melakukan pengumpulan data dengan mengakses laman web online game dengan menggunakan teknik web scraping. Mengambil informasi mengenai informasi game terkait tren ranking game, judul game, harga game, pemain yang memainkan game, dan total pemain. Hasil data online game dari web scraping disimpan kedalam bentuk csv.

2. Pemrosesan Data

Menentukan variable – variable pada data hasil online game dengan melakukan pengidentifikasian terkait informasi online game yang selanjutnya dilakukan pemrosesan pada pemrograman python. Melakukan pembersihan data, melakukan pengubahan tipe data, dan menggunakan library BeautifulSoup dan Selenium dalam bahasa pemrograman Python untuk pemrosesan data.

3. Melakukan Visualisasi Data

Menampilkan visualisasi dan membuat statistik data dengan menggunakan library Python seperti Pandas, NumPy, Matplotlib, Seaborn untuk histogram, diagram, boxplot.



4. Melakukan Interpretasi hasil analisis data

Melakukan analisis hasil terkait data online game dengan mengidentifikasi temuan penting, pola atau tren, serta mendapatkan wawasan yang dapat digunakan dalam pengambilan keputusan.

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Analisis Deskripsi Data

Pada data online game dalam analisis laporan ini proses pengambilan data menggunakan teknik web scraping dengan menggunakan variable rangking, harga game, pemain yang memainkan game saat ini, dan total pemain. Variabel – variable tersebut dilakukan pengolahan mengenai pengaruh, hubungan antar variable terkait karakteristik data, dan korelasi pada setiap variable guna mengetahui informasi penting pada analisis data.

#### 4.2 Analisis Deskriptif dan Visualisasi Data

	Rangking	Harga	Pemain saat ini	Total pemain
count	100.000000	1.000000e+02	100.000000	1.000000e+02
mean	50.500000	1.920965e+05	31159.330000	6.150469e+04
std	29.011492	2.312836e+05	96707.961133	1.873928e+05
min	1.000000	0.000000e+00	5995.000000	6.971000e+03
25%	25.750000	0.000000e+00	8639.000000	1.531200e+04
50%	50.500000	1.124990e+05	12860.000000	2.399950e+04
75%	75.250000	2.612492e+05	23002.500000	3.861525e+04
max	100.000000	1.025000e+06	930741.000000	1.718018e+06

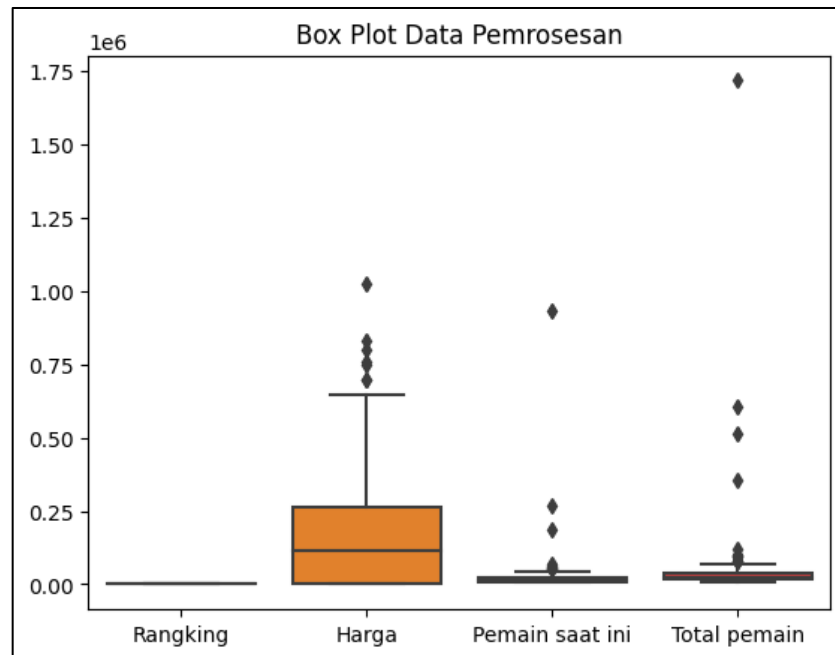
*Gambar 1. Deskripsi Statistik Data Online Game*

Pada statistik deskriptif tersebut dari setiap variabel yang ditampilkan dapat dilihat distribusi harga, jumlah pemain saat ini, dan total pemain dalam dataset. Selain itu, adanya ukuran sebaran data dengan melihat standar deviasi, serta mengetahui nilai-nilai kuartil yang memberikan gambaran tentang letak nilai tengah data.

Data rangking menunjukkan nilai dari 1 hingga 100. Tidak ada statistik deskriptif yang relevan untuk data ini karena hanya mencerminkan posisi peringkat. Harga: Data harga memiliki rata-rata (mean) sebesar 192.096,5 dengan standar deviasi (std) sebesar 231.283,6. Nilai minimal (min) adalah 0, yang menunjukkan adanya harga yang tidak valid atau mungkin ada game yang gratis. Quartil pertama (25%) adalah 0, sedangkan quartil ketiga (75%) adalah 261.249,2. Nilai maksimal (max) adalah 1.025.000.

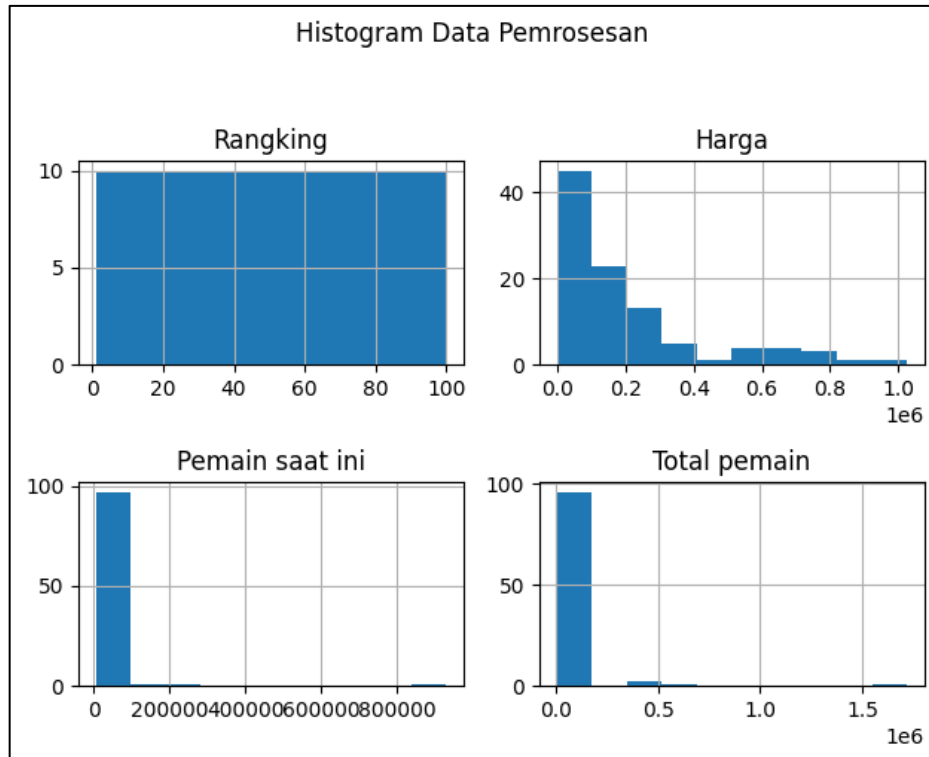
Data pemain saat ini memiliki rata-rata sebesar 31.159,33 dengan standar deviasi sebesar 96.707,96. Jumlah pemain saat ini berkisar dari 5.995 hingga

930.741. Quartil pertama (25%) adalah 8.639, quartil kedua (50%) adalah 12.860, dan quartil ketiga (75%) adalah 23.002,5. Data total pemain memiliki rata-rata sebesar 61.504,69 dengan standar deviasi sebesar 187.392,8. Jumlah pemain berkisar dari 6.971 hingga 1.718.018. Quartil pertama (25%) adalah 15.312, quartil kedua (50%) adalah 23.995, dan quartil ketiga (75%) adalah 38.615,25.



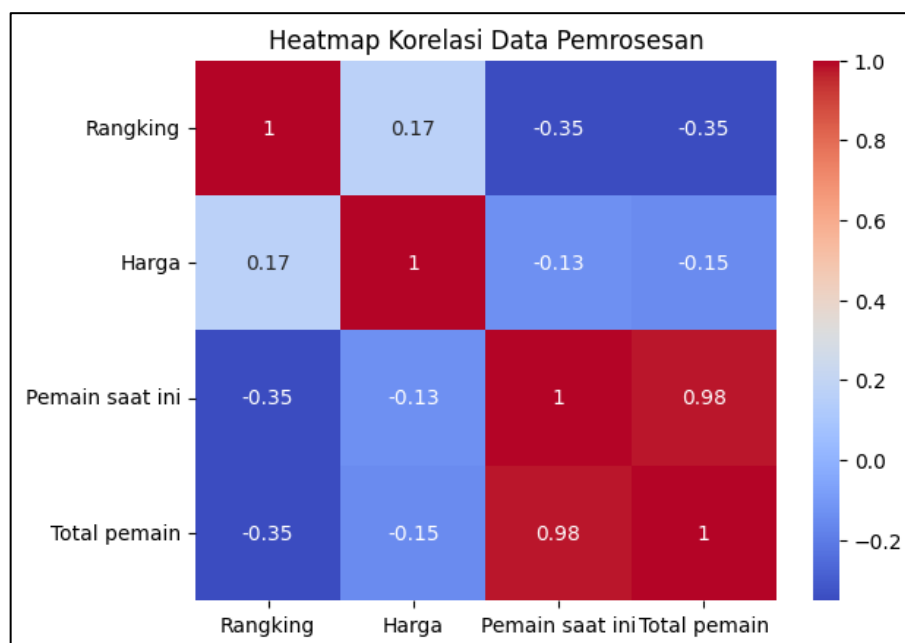
**Gambar 2.** Box Plot Data Pemrosesan

Pada Boxplot gambar 1, tersebut menampilkan persebaran nilai data pada setiap variabel data online game. Diketahui persebaran data yang memiliki nilai terendah ada pada variable *ranging* dengan rentang nilai kurang dari 0,10. Persebaran data pada variable harga memiliki rata – rata yang baik yakni 1.00. Pada variable *total pemain* memiliki nilai outlier paling tinggi diantara yang lainnya sebesar 1.75.



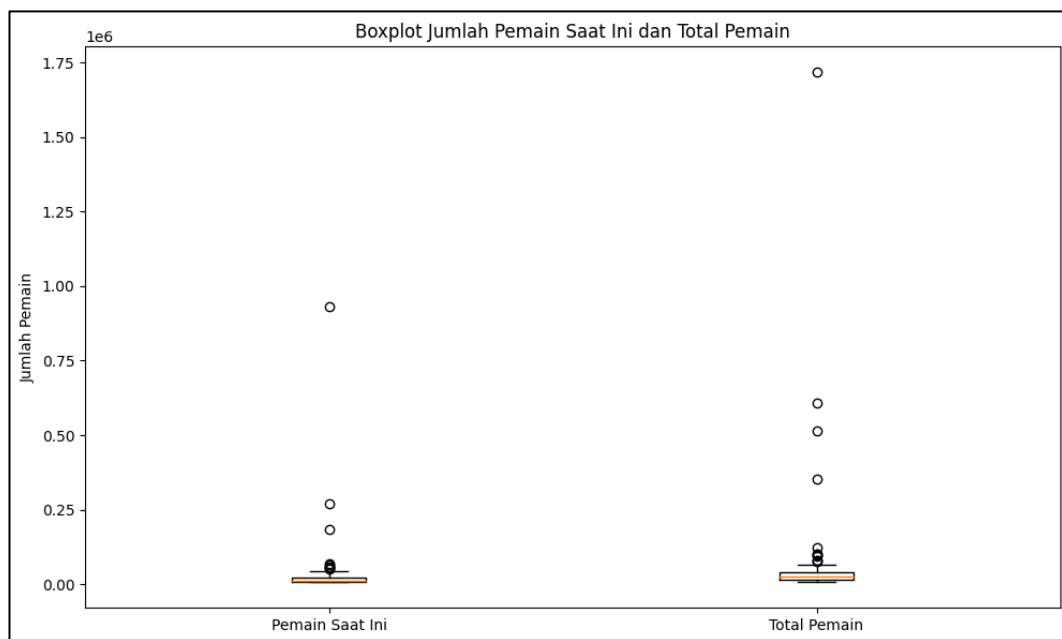
**Gambar 3.** Histogram Data Pemrosesan

Menampilkan Histogram gambar 3, persebaran data pada setiap variable data online game yakni *Rangking*, *Harga*, *Pemain saat ini*, *Total pemain*. Pada setiap variable memiliki rentang nilai yang berbeda – beda tergantung sesuai dengan nilai persebaran data.



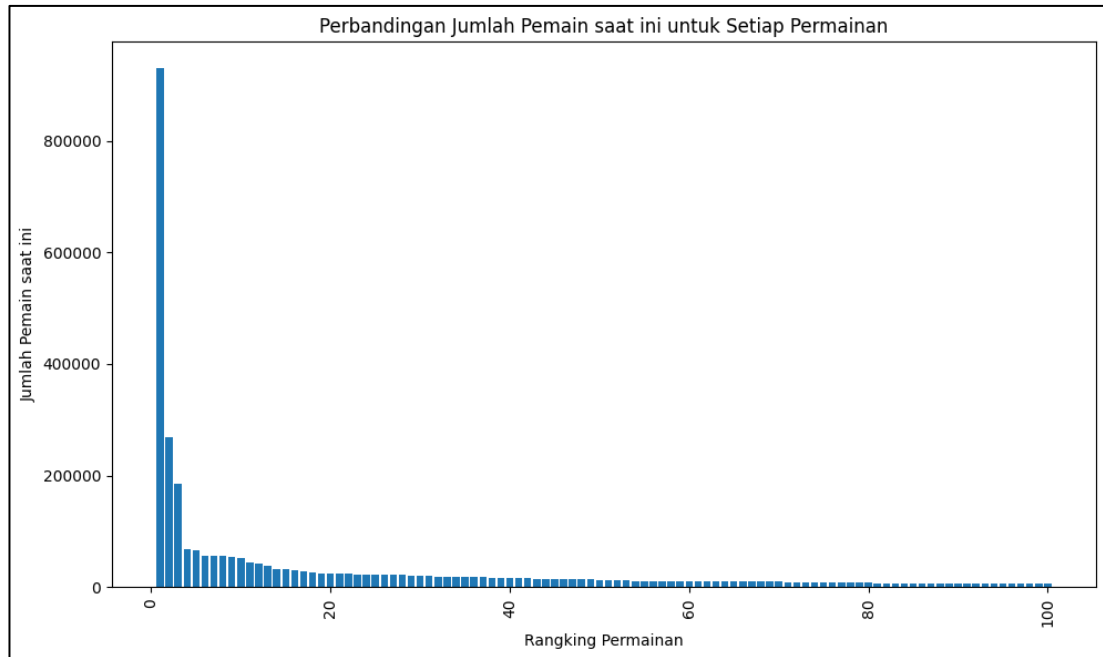
**Gambar 4.** Heatmap Data Pemrosesan

Pada gambar 4, heatmap korelasi menunjukkan pengaruh antar variable, rentang tinggi rendahnya suatu nilai dalam variable ditunjukkan dalam gradasi warna dengan warna merah menunjukkan rentang nilai yang tinggi, warna biru menunjukkan rentang nilai yang rendah. Apabila setiap perbandingan variable bertemu dengan variable dirinya sendiri maka akan menghasilkan nilai yang sama yakni 1, seperti contoh korelasi perbandingan antara *harga* dengan *harga*, *total pemain* dengan *total pemain* menghasilkan nilai yang sama yakni 1.

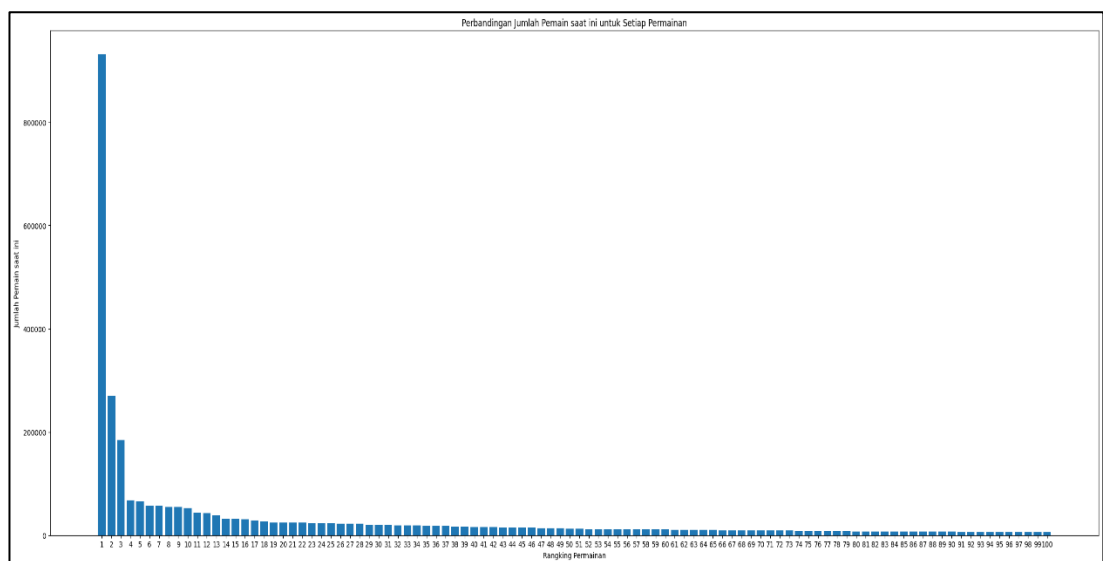


**Gambar 5.** Boxplot Jumlah Pemain Saat ini dan Total Pemain

Boxplot pada gambar 5 menunjukkan perbandingan antara *pemain saat ini* dengan *total pemain* yang ada. Diketahui bahwa rata – rata persebaran nilai *total pemain* memiliki persebaran nilai yang baik dibandingkan dengan *pemain saat ini*, selain itu nilai outlier yang dimiliki *total pemain* juga memiliki nilai yang tinggi sebesar 1.75 dibandingkan dengan *pemain saat ini* yakni kurang dari 1.00.

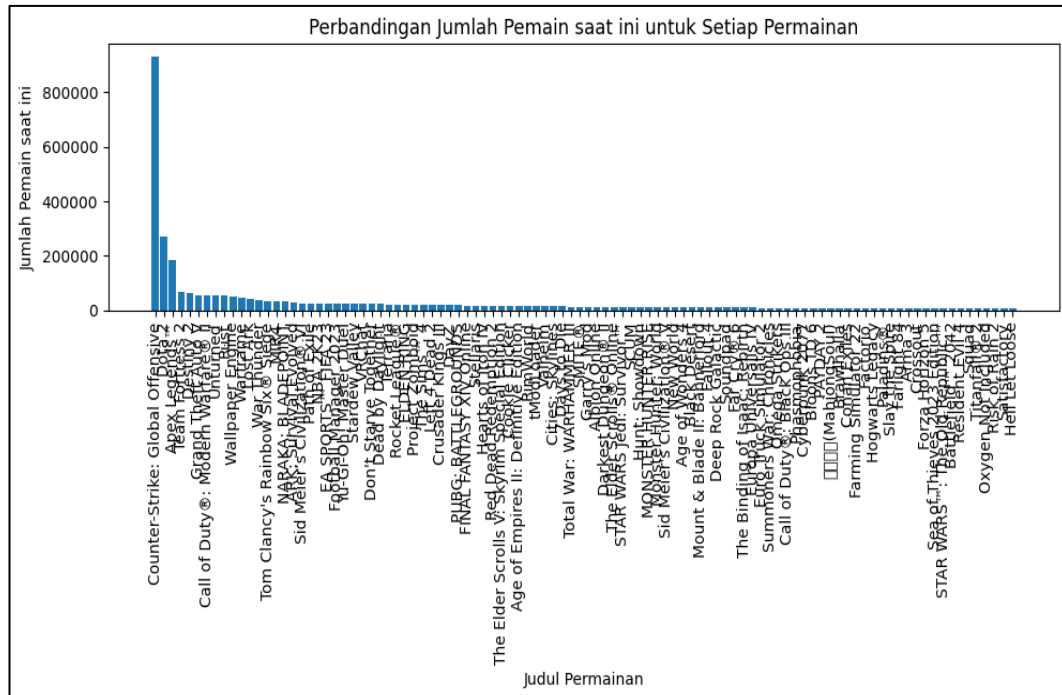


**Gambar 6.** Perbandingan Jumlah Pemain Saat Ini Untuk Setiap Permainan 1

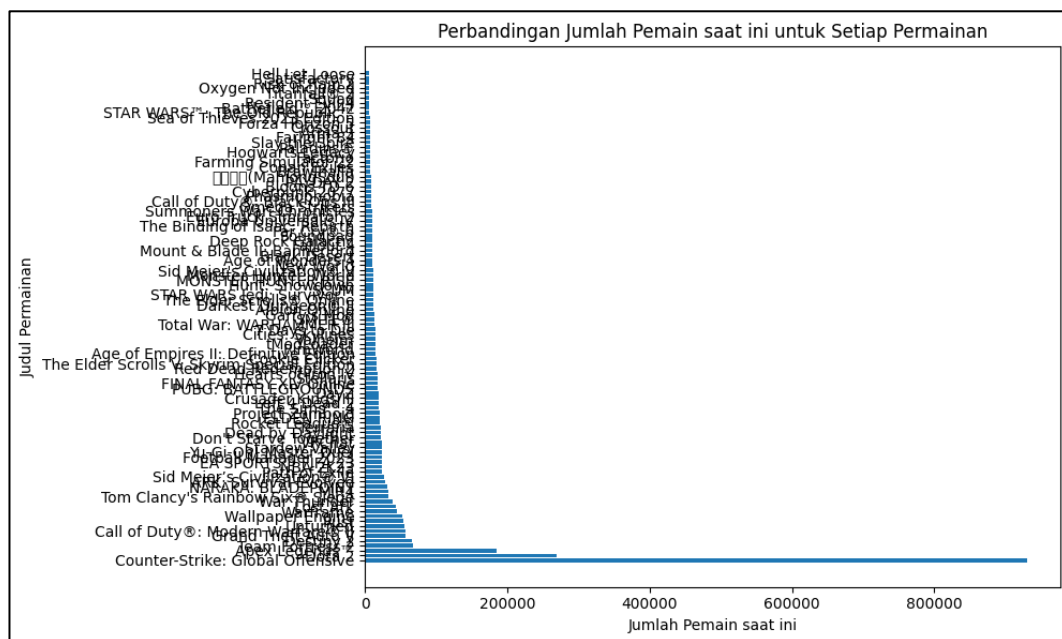


**Gambar 7.** Perbandingan Jumlah Pemain Saat Ini untuk Setiap Permainan 2

Pada gambar 6 dan gambar 7 memiliki hasil keluaran output yang sama, dengan melakukan perbandingan jumlah *pemain saat ini* dengan *rangking* permainan. Pada gambar 6 pada sumbu x *rangking* permainan hanya menampilkan rentang 0 – 100, sedangkan pada gambar 7 pada sumbu x *rangking* permainan menampilkan urutan nilai secara terperinci. Diketahui bahwa rentang jumlah pemain saat ini dengan perbandingan rangking antara 1 – 15 memiliki nilai yang cukup signifikan, terutama pada rangking 1, 2, dan 3. Sedangkan rata – rata nilai lebih dari 15 jumlah pemain saat ini dengan rangking memiliki rata – rata yang sama.



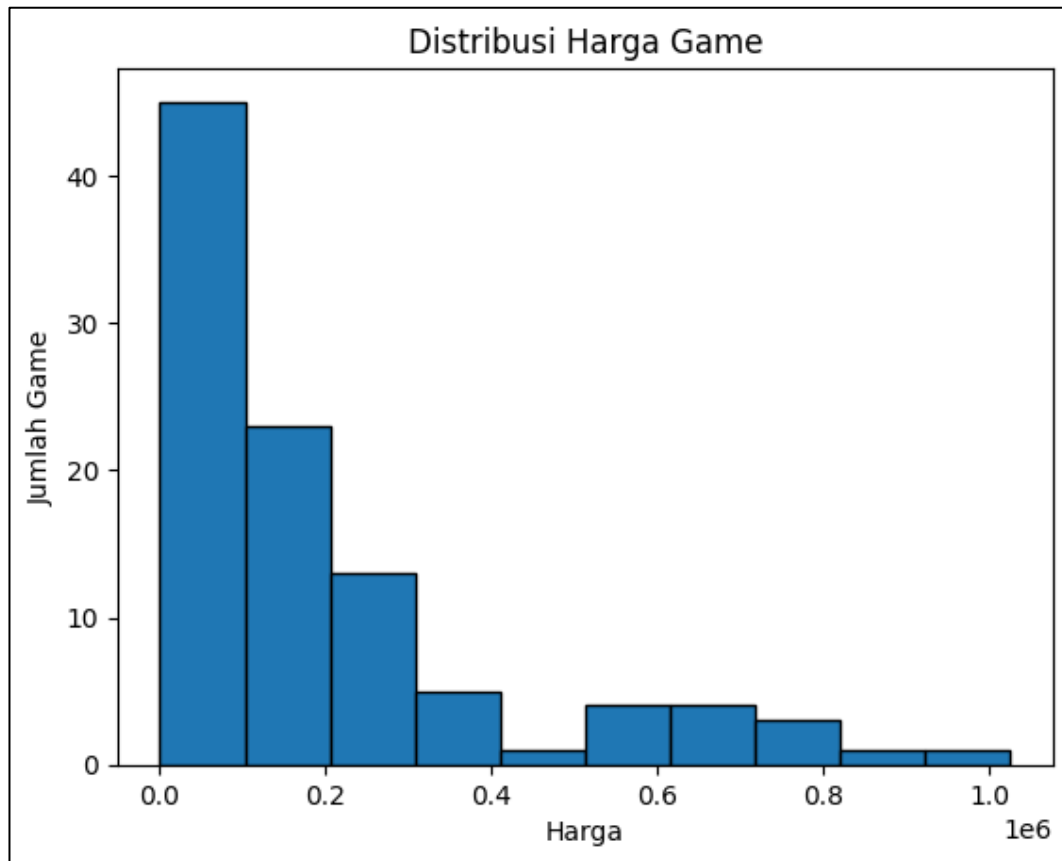
**Gambar 8.** Perbandingan Jumlah Pemain Saat Ini untuk Setiap Permainan 3



**Gambar 9.** Perbandingan Jumlah Pemain Saat Ini untuk Setiap Permainan 4

Gambar 8 dan gambar 9 menampilkan output yang sama, perbedaannya hanya terletak pada posisi penempatan sumbu x dan y yakni secara horizontal dan vertical. Pada histogram ini menampilkan perbandingan antara jumlah *pemain saat ini* dan *judul* permainan yang dimainkan. Judul Counter-Strike: Global Offensive

Memiliki jumlah nilai *pemain saat ini* yang paling tinggi dengan urutan pertama, p ada urutan kedua yakni dengan judul Dota 2, dan urutan ketiga yakni dengan judu l Apex Legends™ , sedangkan judul Hell Let Loose memiliki jumlah nilai pe main saat ini yang paling rendah.



**Gambar 10.** *Distribusi Harga Game*

Pada hisogram gambar 10 mengenai distribusi game menunjukkan game yang memiliki harga 0,0 atau free maka jumlah pemain yang memainkan akan semakin tinggi, berbeda dengan game yang memiliki harga jual yang cukup tinggi maka semakin sedikit pula orang yang memainkan game tersebut. Dibuktikan bahwa orang – orang cenderung memilih game yang memiliki harga jual yang rendah atau free dibandingkan dengan yang berbayar cukup tinggi.



```

1 import pandas as pd
2
3 df = pd.read_csv("C:/Users/ekyfe/Documents/college/Semester 2/Pemrosesan Data/Project_UAS/datapemros.csv")
4 harga_0 = df[df["Harga"] == 0]
5 harga_diatas_0 = df[df["Harga"] > 0]
6 jumlah_harga_0 = len(harga_0)
7 jumlah_harga_diatas_0 = len(harga_diatas_0)
8 print("Jumlah harga 0: ", jumlah_harga_0)
9 print("Jumlah harga di atas 0: ", jumlah_harga_diatas_0)

```

Jumlah harga 0: 29  
Jumlah harga di atas 0: 71

**Gambar 11.** Menampilkan jumlah harga 0 dan lebih dari 0

Hasil output pada gambar 11 menunjukkan jumlah harga yang memiliki nilai 0 sebanyak 29 dan jumlah harga yang memiliki nilai di atas 0 sebanyak 71. Tujuannya untuk mengetahui game yang memiliki harga 0 dan yang memiliki harga lebih dari 0, selain itu informasi ini untuk mengetahui seberapa banyak game yang memainkannya membayar.

```

1 # Memprediksi jumlah pemain saat ini berdasarkan harga
2 harga_prediksi = [[200000]] # Contoh harga yang ingin diprediksi
3 pemain_prediksi = model.predict(harga_prediksi)
4 print("Prediksi jumlah pemain saat harga 200000:", pemain_prediksi)
5

```

Prediksi jumlah pemain saat harga 200000: [168691.44818755]

**Gambar 12.** Prediksi jumlah pemain saat harga 200000

Gambar 11 Untuk memprediksi jumlah pemain saat ini berdasarkan harga. Model menggunakan sebuah contoh harga, yaitu 200000, yang ingin diprediksi jumlah pemainnya. Hasil prediksi yang diperoleh adalah 168691.44818755. Ini berarti, berdasarkan model yang digunakan, diperkirakan jumlah pemain saat harga 200000 akan mencapai sekitar 168691.

## **BAB V**

### **KESIMPULAN**

Web scraping menjadi teknik pengambilan data yang relevan dan berguna dari berbagai situs web secara efisien, salah satunya pengambilan data online game. web scraping digunakan untuk mengumpulkan informasi tentang game online, seperti judul game, deskripsi, ulasan, peringkat, dan informasi lainnya, dari berbagai sumber online. Pengambilan informasi dan kemudian dilakukan pengolahan data lebih lanjut dapat dijadikan analisis dan identifikasi bagi competitor pembuat game guna pengambilan keputusan dengan bijak.

Pada penelitian ini mengambil data online game dengan teknik web scraping dan kemudian dilanjutkan dengan pemrosesan data menggunakan program pemrograman python. Pada pemrosesan data online game ini menghasilkan informasi terkait pengaruh setiap variable data yang dilakukan pemrosesan, sehingga dapat mengetahui hasil analisis sesuai yang selanjutnya dapat dijadikan dalam pengambilan keputusan.

## **DAFTAR PUSTAKA**

- Almaqbal, I. S., Al Khufairi, F. M., Khan, M. S., Bhat, A. Z., & Ahmed, I. (2019). Web Scrapping: Data Extraction from Websites. *Journal of Student Research*.
- Khder, M. A. (2021). Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application. *Int. J. Advance Soft Compu. Appl*, 13(3). Print ISSN: 2710-1274, Online ISSN: 2074-8523. Al-Zaytoonah University of Jordan (ZUJ).
- Phillips dan Young. 2009. *Online Public Relations: A Pratical Guide toDeveloping an Online Strategy in The World of Social Media*. UnitedKingdom: Kogan Page Limited.
- Sirisuriya, D. S. (2015). A comparative study on web scraping. *Proceedings of 8th International Research Conference, KDU*.

## LAMPIRAN

### analisis statistik summary

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv("C:/Users/ekyfe/Documents/college/Semester 2/Pemrosesan Data/Project_UAS/datapemros.csv", sep=',')
print(df)

summary_stats = df.describe()
print(summary_stats)
```

	Rangking	Judul	Harga	Pemain saat ini
\				
0	1	Counter-Strike: Global Offensive	0	930741
1	2	Dota 2	0	269790
2	3	Apex Legends™	0	184496
3	4	Team Fortress 2	0	67588
4	5	Destiny 2	0	65349
..	...			
95	96	Titanfall® 2	37900	6205
96	97	Oxygen Not Included	119999	6146
97	98	Risk of Rain 2	119999	6091
98	99	Satisfactory	139999	6063
99	100	Hell Let Loose	200999	5995

	Total pemain
0	1718018
1	606641
2	514957
3	93972
4	76612
..	...
95	8051
96	9924
97	6971
98	12551
99	10295

[100 rows x 5 columns]

	Rangking	Harga	Pemain saat ini	Total pemain
count	100.000000	1.000000e+02	100.000000	1.000000e+02
mean	50.500000	1.920965e+05	31159.330000	6.150469e+04
std	29.011492	2.312836e+05	96707.961133	1.873928e+05
min	1.000000	0.000000e+00	5995.000000	6.971000e+03
25%	25.750000	0.000000e+00	8639.000000	1.531200e+04
50%	50.500000	1.124990e+05	12860.000000	2.399950e+04
75%	75.250000	2.612492e+05	23002.500000	3.861525e+04
max	100.000000	1.025000e+06	930741.000000	1.718018e+06

```

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv("C:/Users/ekyfe/Documents/college/Semester 2/Pemrosesan Data/Project_UAS/datapemros.csv", sep=',')

summary_stats = df.describe()
print(summary_stats)

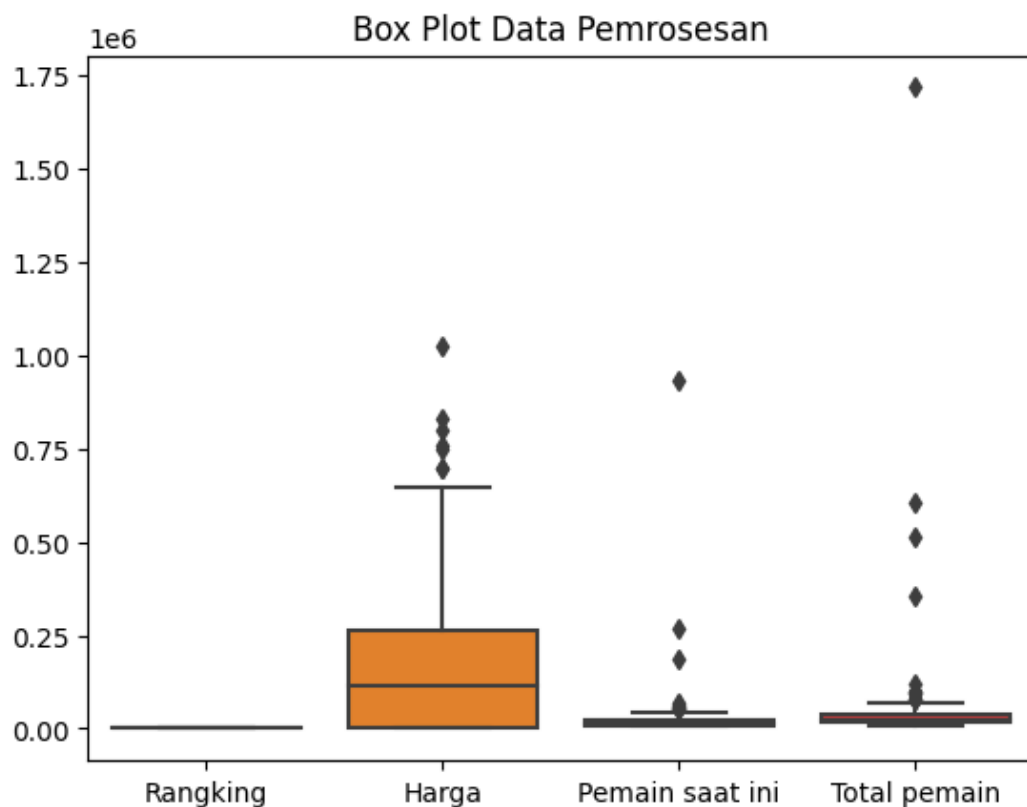
sns.boxplot(data=df)
plt.title("Box Plot Data Pemrosesan")
plt.show()

df.hist()
plt.suptitle("Histogram Data Pemrosesan", y=1.05)
plt.tight_layout()
plt.show()

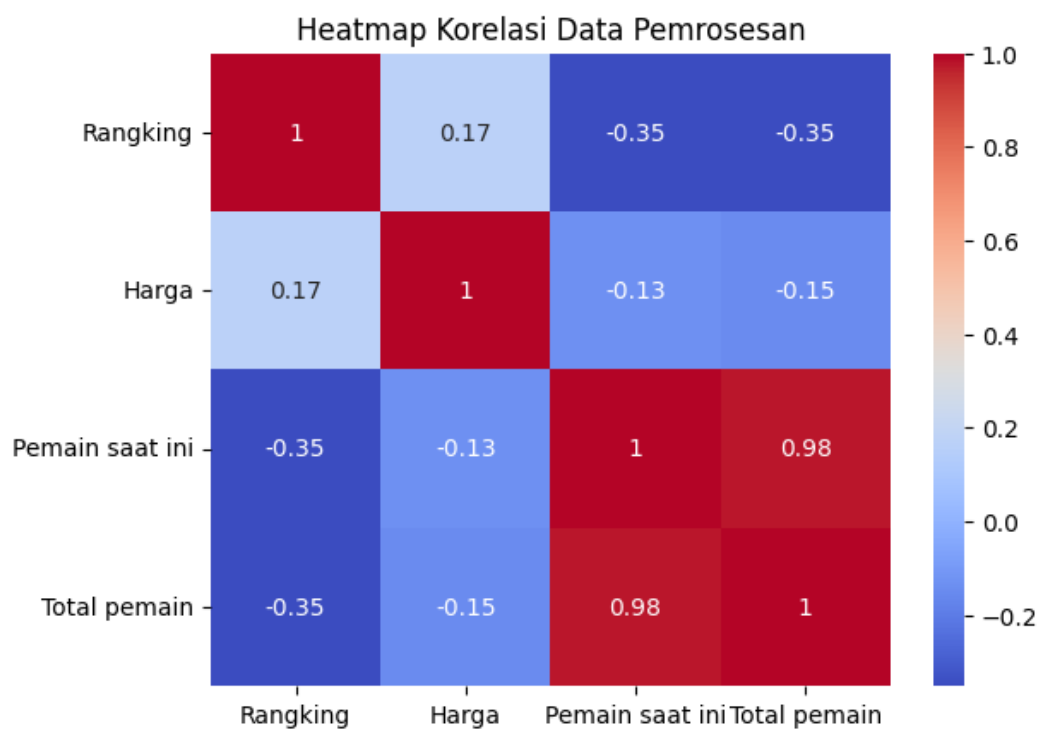
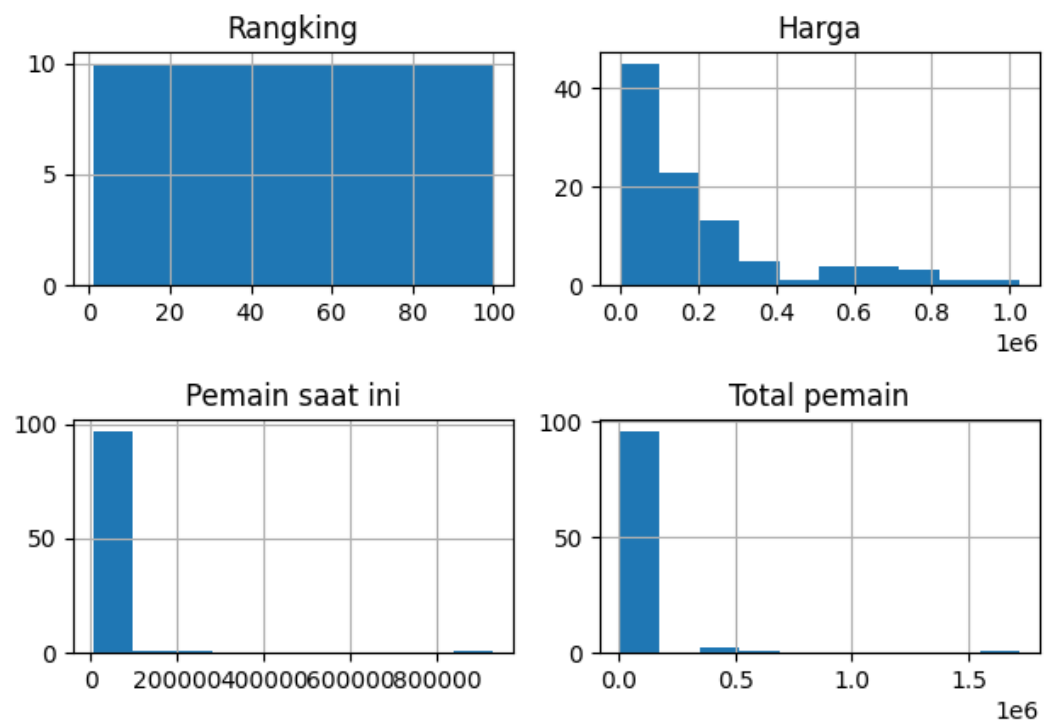
correlation = df.corr()
sns.heatmap(correlation, annot=True, cmap="coolwarm")
plt.title("Heatmap Korelasi Data Pemrosesan")
plt.show()

```

	Rangking	Harga	Pemain saat ini	Total pemain
count	100.000000	1.000000e+02	100.000000	1.000000e+02
mean	50.500000	1.920965e+05	31159.330000	6.150469e+04
std	29.011492	2.312836e+05	96707.961133	1.873928e+05
min	1.000000	0.000000e+00	5995.000000	6.971000e+03
25%	25.750000	0.000000e+00	8639.000000	1.531200e+04
50%	50.500000	1.124990e+05	12860.000000	2.399950e+04
75%	75.250000	2.612492e+05	23002.500000	3.861525e+04
max	100.000000	1.025000e+06	930741.000000	1.718018e+06



## Histogram Data Pemrosesan

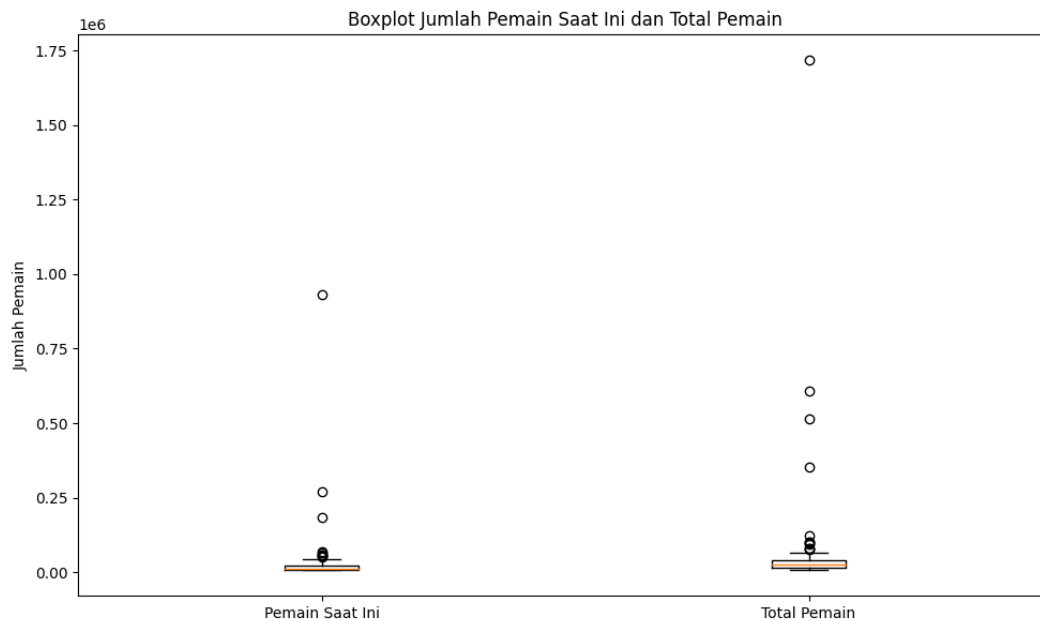


**visualisasi perbandingan antar variable**

```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("C:/Users/ekyfe/Documents/college/Semester 2/Pemrosesan Data/Project_UAS/datapemros.csv")

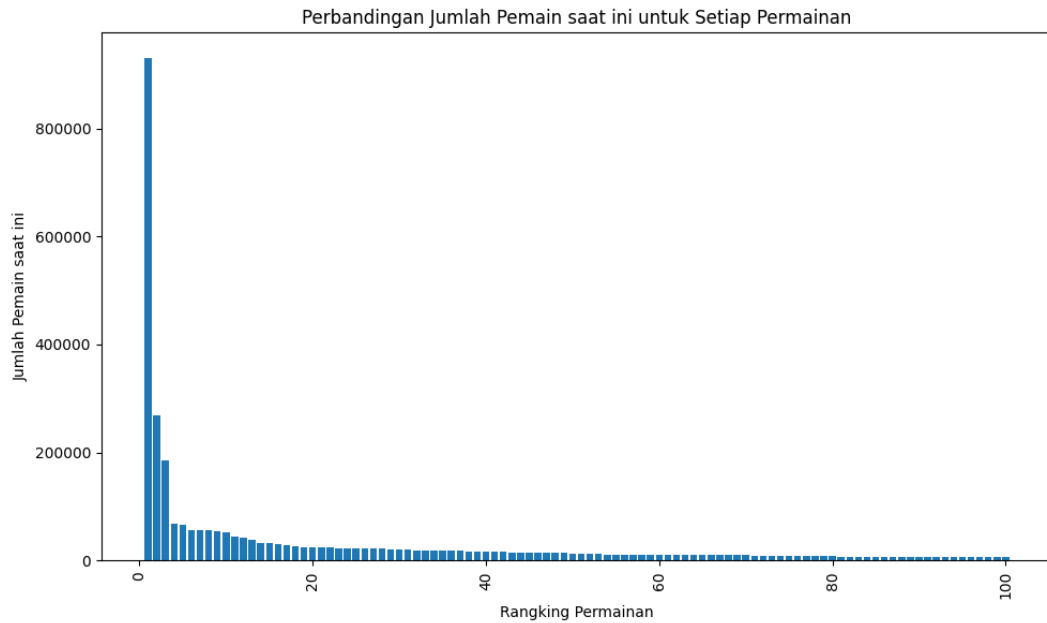
plt.figure(figsize=(10, 6))
plt.boxplot([df["Pemain saat ini"], df["Total pemain"]], labels=["Pemain Saat Ini", "Total Pemain"])
plt.ylabel("Jumlah Pemain")
plt.title("Boxplot Jumlah Pemain Saat Ini dan Total Pemain")
plt.tight_layout()
plt.show()
```



```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("C:/Users/ekyfe/Documents/college/Semester 2/Pemrosesan Data/Project_UAS/datapemros.csv")
print(df.columns)

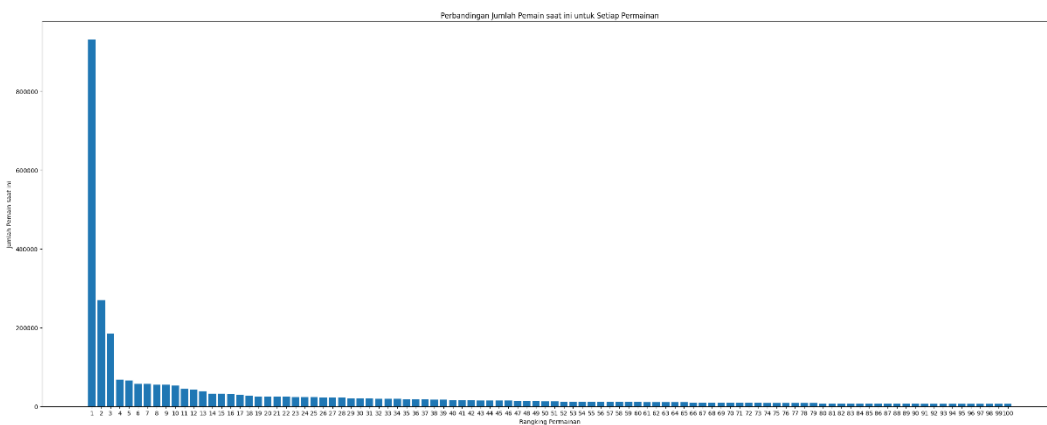
plt.figure(figsize=(10, 6))
plt.bar(df["Rangking"], df["Pemain saat ini"])
plt.xticks(rotation=90)
plt.xlabel("Rangking Permainan")
plt.ylabel("Jumlah Pemain saat ini")
plt.title("Perbandingan Jumlah Pemain saat ini untuk Setiap Permainan")
plt.tight_layout()
plt.show()
```



```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("C:/Users/ekyfe/Documents/college/Semester 2/Pemrosesan Data/Project_UAS/datapemros.csv")

plt.figure(figsize=(25, 10))
plt.bar(df.index + 1, df["Pemain saat ini"])
plt.xticks(df.index + 1, df["Rangking"])
plt.xlabel("Rangking Permainan")
plt.ylabel("Jumlah Pemain saat ini")
plt.title("Perbandingan Jumlah Pemain saat ini untuk Setiap Permainan")
plt.tight_layout()
plt.show()
```



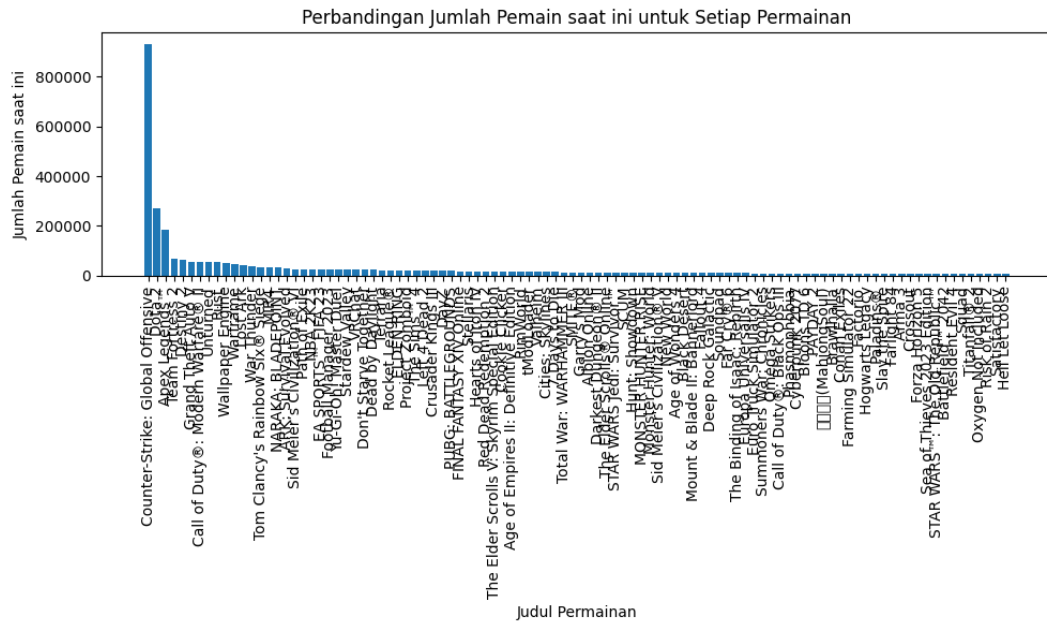
```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("C:/Users/ekyfe/Documents/college/Semester 2/Pemrosesan Data/Project_UAS/datapemros.csv")

print(df.columns)

plt.figure(figsize=(10, 6))
plt.bar(df["Judul"], df["Pemain saat ini"])
plt.xticks(rotation=90)
plt.xlabel("Judul Permainan")
plt.ylabel("Jumlah Pemain saat ini")
plt.title("Perbandingan Jumlah Pemain saat ini untuk Setiap Permainan")
plt.tight_layout()
plt.show()
```

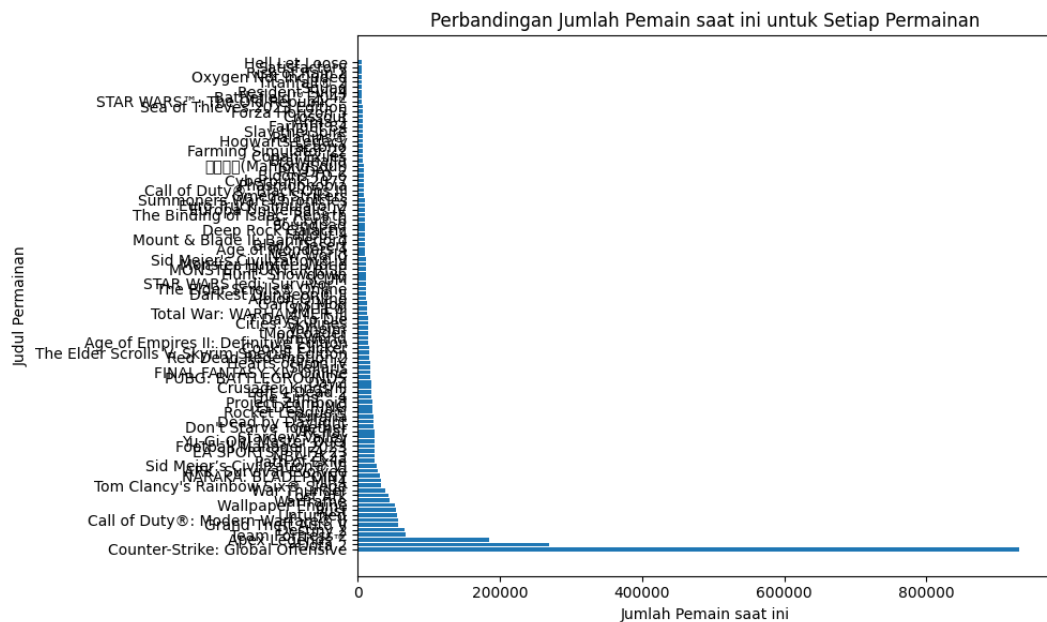




```
import pandas as pd
import matplotlib.pyplot as plt

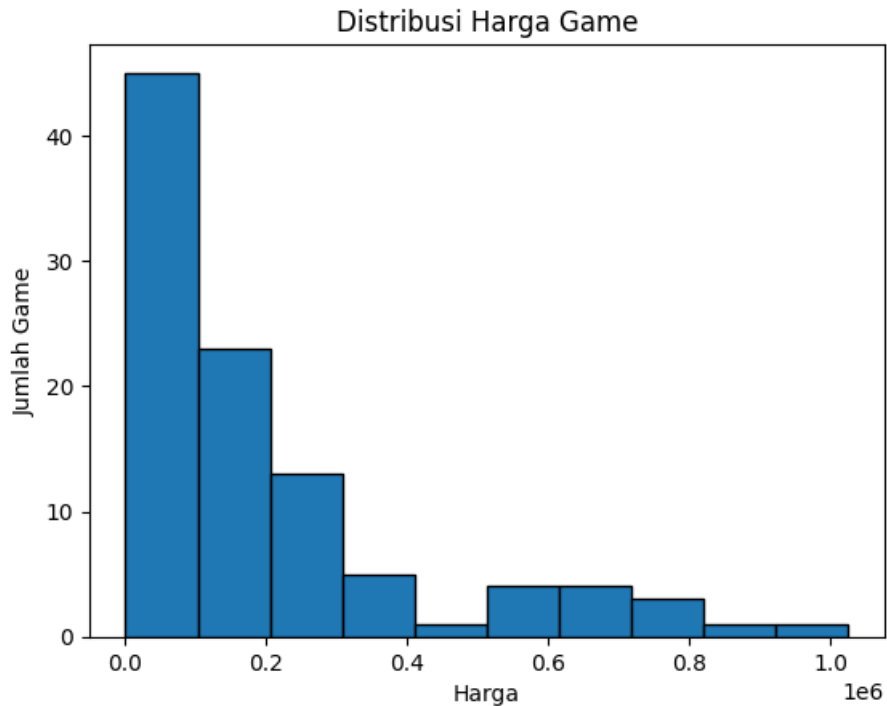
df = pd.read_csv("C:/Users/ekyfe/Documents/college/Semester 2/Pemrosesan Data/Project_UAS/datapemros.csv")

plt.figure(figsize=(10, 6))
plt.barh(df["Judul"], df["Pemain saat ini"])
plt.xlabel("Jumlah Pemain saat ini")
plt.ylabel("Judul Permainan")
plt.title("Perbandingan Jumlah Pemain saat ini untuk Setiap Permainan")
plt.tight_layout()
plt.show()
```



```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("C:/Users/ekyfe/Documents/college/Semester 2/Pemrosesan Data/Project_UAS/datapemros.csv")
harga = df["Harga"]
plt.hist(harga, bins=10, edgecolor='black')
plt.title("Distribusi Harga Game")
plt.xlabel("Harga")
plt.ylabel("Jumlah Game")
plt.show()
```



```
import pandas as pd
df = pd.read_csv("C:/Users/ekyfe/Documents/college/Semester 2/Pemrosesan Data/Project_UAS/datapemros.csv")
harga_counts = df["Harga"].value_counts().sort_index()
print(harga_counts)
```

0	29
37900	1
37999	1
39999	2
42475	1
48999	1
74999	1
77999	1
79999	1
89999	4
90999	1
95999	2
105299	2
105440	1
108999	2
115999	1
119999	4
139999	2
149000	1
154750	1
164750	1
165353	1
165999	1
187999	1
194205	1
194299	1
196934	1

```

199999 1
200999 1
205000 1
209999 1
219999 1
220000 1
227700 1
249999 1
259999 1
265000 1
266000 2
299999 2
300000 1
325285 1
329999 1
334999 1
371999 1
399999 1
494999 1
549999 1
553999 1
599000 1
600000 1
640000 1
648999 1
699000 1
699999 1
749000 1
759000 1
800000 1
830999 1
1025000 1
Name: Harga, dtype: int64

```

```

1 import pandas as pd
2
3 df = pd.read_csv("C:/Users/ekyfe/Documents/college/Semester 2/Pemrosesan Data/Project_UAS/datapemros.csv")
4 harga_0 = df[df["Harga"] == 0]
5 harga_diatas_0 = df[df["Harga"] > 0]
6 jumlah_harga_0 = len(harga_0)
7 jumlah_harga_diatas_0 = len(harga_diatas_0)
8 print("Jumlah harga 0: ", jumlah_harga_0)
9 print("Jumlah harga di atas 0: ", jumlah_harga_diatas_0)

```

Jumlah harga 0: 29  
Jumlah harga di atas 0: 71

## analisis statistik menggunakan regresi linear

```

import pandas as pd
from sklearn.linear_model import LinearRegression

# Membaca data dari file CSV
df = pd.read_csv("C:/Users/ekyfe/Documents/college/Semester 2/Pemrosesan Data/Project_UAS/datapemros.csv")
print(df)

```

	Rangking	Judul	Harga	Pemain saat ini \
0	1	Counter-Strike: Global Offensive	0	930741
1	2	Dota 2	0	269790
2	3	Apex Legends™	0	184496
3	4	Team Fortress 2	0	67588
4	5	Destiny 2	0	65349
..	...	...	...	...
95	96	Titanfall® 2	37900	6205
96	97	Oxygen Not Included	119999	6146
97	98	Risk of Rain 2	119999	6091
98	99	Satisfactory	139999	6063
99	100	Hell Let Loose	200999	5995
Total pemain				
0	1718018			
1	606641			
2	514957			
3	93972			
4	76612			

```

..          ...
95          8051
96          9924
97          6971
98          12551
99          10295

```

[100 rows x 5 columns]

```

1 # Mengambil kolom "Harga" sebagai variabel independen (X) dan kolom "Pemain saat ini" sebagai variabel dependen (y)
2 X = df["Harga"].values.reshape(-1, 1)
3 y = df["Pemain saat ini"].values
4
5 # Membuat model regresi linier
6 model = LinearRegression()
7
8 # Melatih model menggunakan data
9 model.fit(X, y)

```

LinearRegression()

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.  
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

```

1 # Memprediksi jumlah pemain saat ini berdasarkan harga
2 harga_prediksi = [[200000]] # Contoh harga yang ingin diprediksi
3 pemain_prediksi = model.predict(harga_prediksi)
4 print("Prediksi jumlah pemain saat harga 200000:", pemain_prediksi)
5

```

Prediksi jumlah pemain saat harga 200000: [168691.44818755]