

**LAPORAN TUGAS BESAR 2**  
**Pengenalan Komputasi (KU1102-16)**  
*Data Analysis*  
**“VIDEO GAME SALES”**

Disusun untuk memenuhi nilai tugas besar 2 (dua) Pengenalan Komputasi (KU1102)

Dosen pengampu: Elvayandri, S.Si., M.T.

Disusun oleh:

**KELOMPOK 5 – KELAS 16 STEI**

<b>ALBERT</b>	<b>(16521192)</b>
<b>DAVEN DARMAWAN SENDJAYA</b>	<b>(16521322)</b>
<b>JIMLY FIRDAUS</b>	<b>(16521312)</b>
<b>MUHAMMAD NAUFAL</b>	<b>(16521442)</b>



**PROGRAM TAHAP PERSIAPAN BERSAMA**  
**SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA**  
**INSTITUT TEKNOLOGI BANDUNG**

**2021**

## DAFTAR ISI

<b>DAFTAR ISI</b> .....	<b>i</b>
<b>BAB I: PENDAHULUAN</b> .....	<b>1</b>
1.1. LATAR BELAKANG .....	1
1.2. RUMUSAN MASALAH.....	1
1.3. TUJUAN PENELITIAN.....	1
1.4. METODE YANG DIGUNAKAN.....	2
<b>BAB II: PEMBAHASAN</b> .....	<b>3</b>
<b>2.1. TUGAS 1 (DESKRIPSI DATA DAN FILE)</b> .....	<b>3</b>
2.1.1. Deskripsi Data “video_games_sales” .....	3
2.1.2. Properti File sebelum <i>Data Cleansing</i> .....	3
2.1.3. Properti File setelah Data Cleansing .....	4
<b>2.2. TUGAS 2 (KARAKTERISTIK DATA)</b> .....	<b>5</b>
2.2.1. Kolom, Keterangan, dan Jenis Data.....	5
2.2.2. Kolom dan Karakteristik Data .....	5
<b>2.3. TUGAS 3 (STATISTIK)</b> .....	<b>9</b>
2.3.1. Menampilkan 5 Data Teratas dari Keseluruhan Data .....	9
2.3.2. Menampilkan 3 Game dengan Penjualan Global Tertinggi .....	10
2.3.3. Menampilkan 3 Game dengan Tahun Rilis Terlama .....	10
2.3.4. Sampel 3 Data Kolom “Name”.....	10
2.3.5. Sampel 3 Data Kolom “Platform” .....	10
2.3.6. Sampel 3 Data Kolom “Year_of_Release” .....	10
2.3.7. Sampel 3 Data Kolom “Genre” .....	11
2.3.8. Sampel 3 Data Kolom “Publisher” .....	11
2.3.9. Sampel 3 Data Kolom “Global_Sales” .....	11
2.3.10. Rata-Rata dari Penjualan Global.....	11
2.3.11. Standar Deviasi dari Penjualan Global .....	11
2.3.12. Persentil 10% dari Penjualan Global .....	11
2.3.13. Persentil 25% dari Penjualan Global .....	12
2.3.14. Persentil 50% dari Penjualan Global .....	12
2.3.15. Persentil 75% dari Penjualan Global .....	12
2.3.16. Persentil 90% dari Penjualan Global .....	12
2.3.17. Nilai Maksimum dari Penjualan Global.....	12
2.3.18. Nilai Minimum dari Penjualan Global.....	12
2.3.19. Distribusi Frekuensi dari Publisher.....	13
2.3.20. Informasi/Pengetahuan.....	13
<b>2.4. TUGAS 4 (VISUALISASI)</b> .....	<b>13</b>
<b>2.4.1. Perbandingan Kategori</b> .....	<b>13</b>

2.4.1.1.	Perbandingan Penjualan di Amerika Utara dengan Penjualan Global Berdasarkan Zushi Games .....	13
2.4.1.2.	Perbandingan Penjualan Games Warner Bros. Interactive Entertainment Amerika Utara dengan Eropa.....	14
<b>2.4.2.</b>	<b>Penampilan Perubahan Terhadap Waktu .....</b>	<b>14</b>
2.4.2.1.	Grafik Penjualan Global dari Nintendo dari Tahun 1990-2015 .....	14
2.4.2.2.	Grafik Perubahan Jumlah Peminat Genre “Sports” dari Tahun 2006-2016 ....	15
<b>2.4.3.</b>	<b>Penampilan Hierarki Dan Hubungan Keseluruhan Bagian.....</b>	<b>15</b>
2.4.3.1.	Jumlah Game Terbanyak dari Tahun 1980-2020 .....	15
<b>2.4.4.</b>	<b>Genre Games Terbanyak dari Tahun 1980-2020 .....</b>	<b>16</b>
<b>2.4.5.</b>	<b><i>Plotting Relationships</i> .....</b>	<b>16</b>
2.4.5.1.	Scatter Plot dari Penjualan Global Electronic Arts terhadap Year of Release ...	16
2.4.5.2.	Bubble Plot dari Penjualan Global Activision terhadap Year of Release, dengan <i>Sizes</i> Merupakan Frekuensi <i>Video Games</i> dari <i>Publisher</i> Activision .....	17
<b>2.5.</b>	<b>TUGAS 5 (KORELASI) .....</b>	<b>17</b>
<b>2.6.</b>	<b>TUGAS 6 (DATA CLEANSING).....</b>	<b>19</b>
2.6.1.	Data Mentah .....	19
2.6.2.	Pembersihan DataFrame .....	19
<b>BAB III: PENUTUP .....</b>		<b>20</b>
3.1.	KESIMPULAN DAN <i>LESSON LEARNED</i> .....	20
<b>DAFTAR PUSTAKA .....</b>		<b>21</b>
<b>PEMBAGIAN TUGAS.....</b>		<b>22</b>

## BAB I

### PENDAHULUAN

#### 1.1. LATAR BELAKANG

Secara harfiah, gim merupakan sebuah permainan yang berfungsi sebagai salah satu bentuk hiburan. Pada zaman dahulu, sebuah permainan lazimnya dilakukan secara langsung (dapat diamati secara fisik), misalnya permainan gobak sodor, permainan gasing, atau permainan kartu. Namun, sekarang ini banyak sekali permainan yang bisa dilakukan melalui gawai, misalnya gim video yang tersedia pada *game console* atau yang tersedia di toko aplikasi pada ponsel pintar. Gim tersebut memungkinkan banyak orang untuk bermain bersama tanpa batasan tempat, misalnya seorang pemain dari Indonesia bisa bermain gim secara langsung dengan seorang pemain lain dari negara Amerika Serikat. Selain itu, kemudahan dan kepraktisan dalam memainkan gim membuat banyak masyarakat tertarik untuk menggunakannya untuk mengusir rasa bosan atau menghabiskan waktu senggang. Oleh karena itu, banyak sekali pengembang (*game developer*) yang memproduksi berbagai jenis gim video yang dijual secara masif kepada khalayak umum.

Menurut data yang dihimpun oleh Statista (2021), jumlah pemain gim video di Amerika Serikat mencapai angka 201,7 juta pada tahun 2021. Angka ini diprediksi naik sekitar 5,8%, sehingga total pemain gim video bisa mencapai angka 213,3 juta pada tahun 2025. Selain itu, menurut artikel yang ditulis oleh Anderton (2019) melalui Forbes, pemain gim video rata-rata menghabiskan waktu selama 7 jam 7 menit setiap minggunya untuk bermain gim. Waktu permainan ini meningkat sekitar 20% dari data yang diperoleh pada tahun sebelumnya (2018). Dalam data ini, kelompok usia yang mengalami peningkatan waktu permainan paling besar adalah kelompok 26-35 tahun, yang mengalami peningkatan sebesar 25%. Secara kontras, kelompok usia 18-25 tahun mengalami peningkatan yang paling rendah, yaitu sebesar 9,9%.

Melalui data statistik yang diberikan di atas, jelas peranan gim dalam kehidupan sehari-hari dari masyarakat modern sangatlah besar. Sekarang ini, banyak sekali pemain yang memperoleh penghasilan melalui permainan gim video. Penghasilan tersebut diperoleh melalui hadiah dari turnamen atau pertandingan yang diselenggarakan oleh pihak tertentu. Selain itu, penghasilan juga bisa diperoleh melalui iklan yang ditampilkan saat proses *streaming* yang dilakukan oleh pemain, misalnya di YouTube atau Twitch. Seluruh keuntungan yang didapatkan oleh pemain menyebabkan popularitas gim video sangat meningkat, terutama dikalangan pemain berusia muda.

Dalam tugas besar kedua ini, penulis berminat untuk meneliti angka statistik terkait penjualan (sales) dari gim video. Data yang digunakan oleh penulis dihimpun oleh Gregory Smith dan didapatkan melalui situs kaggle.com. Data yang kami peroleh berisi rincian terkait penjualan 16500 gim. Proses analisis yang akan penulis lakukan dalam tugas besar kedua ini antara lain, deskripsi dari data dan file (penghitungan jumlah kolom dan baris file), menampilkan 5 (lima) data teratas dari file, menampilkan 3 (tiga) data terkait *global sales* tertinggi, tahun rilis gim terlama, dan sampel data kolom nama, *platform*, tahun perilisan, genre, penerbit, dan *global sales*.

Penulis juga akan melakukan penghitungan statistik terkait data *global sales*, misalnya rata-rata, standar deviasi, persentil 10%, persentil 25%, persentil 50%, persentil 75%, persentil 90%, nilai maksimum, dan nilai minimum. Selain itu, penulis juga meneliti distribusi frekuensi dari kolom "publisher". Terkait dengan visualisasi data, penulis meneliti terkait perbandingan kategori, penampilan perubahan terhadap waktu, penampilan hierarki dan hubungan keseluruhan bagian, *plotting relationships*, dan mencari korelasi.

#### 1.2. RUMUSAN MASALAH

1. Bagaimana rincian deskripsi terkait data yang digunakan?
2. Apa format, asal, dan dimensi dari data yang digunakan?
3. Bagaimana karakteristik atau atribut yang tersedia pada data yang digunakan?
4. Bagaimana rincian statistik yang mampu diteliti melalui data yang digunakan?
5. Bagaimana cara memvisualisasikan data yang tersedia dalam file yang digunakan?
6. Bagaimana korelasi antardata yang tersedia pada file yang digunakan?

#### 1.3. TUJUAN PENELITIAN

1. Mengetahui deskripsi terkait data yang digunakan
2. Mengetahui format, asal, dan dimensi dari data yang digunakan
3. Mengetahui karakteristik atau atribut yang tersedia pada data yang digunakan
4. Mempelajari rincian statistik melalui data yang digunakan
5. Mempelajari cara memvisualisasikan data melalui file yang digunakan
6. Mengetahui korelasi antardata yang tersedia pada file yang digunakan

#### 1.4. METODE YANG DIGUNAKAN

1. Pengumpulan data

Pengumpulan data dilakukan dengan mengakses data yang ingin dianalisis, yaitu “Video Game Sales”, melalui situs <https://www.kaggle.com/gregorut/videogamesales>.

2. Pemrograman Python

Proses penelitian terkait data “Video Game Sales” dilakukan melalui kode pemrograman Python melalui *library* “pandas” untuk proses analisis data (mencari deskripsi data, melakukan penghitungan statistik, dan mencari korelasi antardata) dan “matplotlib” untuk visualisasi data.

## BAB II

### PEMBAHASAN

#### 2.1. TUGAS 1 (DESKRIPSI DATA DAN FILE)

##### 2.1.1. Deskripsi Data “video\_games\_sales”

Data “video\_games\_sales” merupakan sebuah data yang dimiliki oleh seorang pemilik akun bernama GregorySmith. Data ini diunggah melalui ke situs penyedia layanan data Kaggle.com. Data ini dapat diakses melalui alamat web <https://www.kaggle.com/gregorut/videogamesales>. Data ini digunakan sebagai acuan analisis terhadap lebih dari 16500 jenis *video games*. Informasi yang bisa dihimpun melalui *dataset* ini, antara lain *video games* dengan penjualan tertinggi, *video games* dengan tahun rilis terlama, serta pengambilan sejumlah sampel data. Selain itu, kita juga bisa mengambil perhitungan statistik untuk mendapatkan beberapa properti data, seperti rata-rata, standar deviasi, persentil 10%, persentil 25%, persentil 50%, persentil 75%, persentil 90%, nilai maksimum, dan nilai minimum dari total penjualan global, serta distribusi frekuensi *publisher* dari *video games*.

Selain pengambilan informasi dan penghitungan statistik, kita juga bisa memvisualisasikan data dari *dataset* “video\_games\_sales” ini. Visualisasi data yang bisa diambil, antara lain perbandingan kuantitas penjualan antara satu tempat dengan tempat lainnya dari salah satu *publisher*, penampilan perubahan terhadap waktu, penampilan hierarki dan hubungan keseluruhan bagian, serta *plotting relationships*. Penghitungan korelasi dari sebuah kolom terhadap kolom lainnya juga bisa dikalkulasi melalui rincian data ini.

Data ini juga bisa menjadi metode pembelajaran bagi penulis terkait metode untuk membersihkan data (*data cleansing*). Dari data ini, penulis belajar untuk menganalisis data-data yang kurang lengkap, lalu menghapusnya sehingga data menjadi lebih rapi dan lengkap.

##### 2.1.2. Properti File sebelum *Data Cleansing*

Format File	Dimensi File (baris x kolom)	Ukuran Data	Ukuran File ( <i>bytes</i> )
<i>Comma-separated value</i> (.csv)	16719 x 16	220788	1618040

Proses pencarian deskripsi terkait file sebelum dilakukan proses *data cleansing* dilakukan melalui kode pemrograman Python melalui *library* “pandas”. Berikut adalah proses pencarian properti data pada file “video\_games\_sales.csv”:

##### Proses Pencarian Jumlah Baris dan Kolom

```
#Jumlah baris
● print("Ada ",len(data)," baris")
#Jumlah kolom
print("Ada",len(data.columns)," kolom")
✓ 0.6s

Ada 16719 baris
Ada 16 kolom
```

### Proses Pencarian Ukuran Data sebelum *Data Cleansing*

```
#ukuran dimensi csv (tanpa data kosong)
data.size - (data.isnull().sum().sum())

✓ 0.1s

220788
```

Ukuran data (total jumlah data yang tersedia) pada file “video\_games\_sales.csv” tercatat memiliki 220.788 data sesuai kode pemrograman yang telah ditunjukkan di atas. Namun, data ini tidak sesuai dengan jumlah baris yang dikalikan dengan jumlah kolom yang hasilnya berjumlah 267.504 data. Hal ini terjadi karena terdapat beberapa data yang bersifat “null” atau kosong sehingga jumlah kolom yang dikalikan oleh baris tidak dapat menggambarkan jumlah keseluruhan data dalam file.csv yang digunakan.

### Proses Pencarian Ukuran File (dalam bytes) sebelum *Data Cleansing*

```
import os
#mencari informasi ukuran file csv
print(os.path.getsize("video_games_sales - Edit.csv"), " bytes")

✓ 0.1s

1618040 bytes
```

#### 2.1.3. Properti File setelah *Data Cleansing*

Format File	Dimensi File (baris x kolom)	Ukuran Data	Ukuran File (bytes)
Comma-separated value (.csv)	6825 x 16	109200	791986

Proses pencarian properti dari file “video\_games\_sales.csv” setelah dilakukan prosedur *data cleansing* dilakukan melalui kode pemrograman Python dengan menggunakan library “pandas”. Berikut adalah proses pengambilan properti terkait data “video\_games\_sales.csv”.

### Proses Pencarian Jumlah Baris dan Kolom

```
#Jumlah baris
print("Ada "+str(len(data))+ " baris")
#Jumlah kolom
print("Ada",len(data.columns)," kolom")

✓ 0.9s

Ada 6825 baris
Ada 16 kolom
```

### Proses Pencarian Ukuran Data setelah *Data Cleansing*

```
#ukuran dimensi csv setelah cleaning
data.size

✓ 0.7s

109200
```

Ukuran data (total jumlah data yang tersedia) pada data “video\_games\_sales.csv” setelah dilakukan proses *data cleansing* berjumlah 109.200 data. Hal ini sesuai dengan penghitungan dari jumlah baris yang dikalikan dengan jumlah kolom yang juga berjumlah 109.200 data. Hal ini bisa terjadi karena proses *data cleansing* yang penulis lakukan menghapus keseluruhan data dari satu baris, jika terdapat data yang kosong pada salah satu kolomnya (proses *data cleansing* bisa dilihat pada Subbab 2.6). Oleh karena itu, data yang penulis peroleh setelah *data cleansing* bersifat lebih rapi dan lengkap dari data yang penulis gunakan sebelum *data cleansing*.

#### Proses Pencarian Ukuran File (dalam bytes) setelah Data Cleansing

```
import os
#mencari informasi ukuran file csv
os.path.getsize("video_games_sales.csv")
✓ 0.1s
791986
```

## 2.2. TUGAS 2 (KARAKTERISTIK DATA)

### 2.2.1. Kolom, Keterangan, dan Jenis Data

Kolom	Keterangan	Jenis Data
Name	Nama dari <i>video games</i>	<i>Categorical - nominal</i>
Platform	Konsol dari <i>video games</i>	<i>Categorical - nominal</i>
Year_of_Release	Tahun rilis <i>video games</i>	<i>Quantitative - continues</i>
Publisher	Pihak yang menerbitkan <i>video games</i>	<i>Categorical - nominal</i>
Genre	Jenis permainan <i>video games</i>	<i>Categorical - nominal</i>
NA_Sales	Penjualan <i>video games</i>	<i>Quantitative - discrete</i>
EU_Sales	Penjualan <i>video games</i> di Europe	<i>Quantitative - discrete</i>
JP_Sales	Penjualan <i>video games</i> di Jepang	<i>Quantitative - discrete</i>
Other_Sales	Penjualan <i>video games</i> di seluruh negara kecuali North America, Europe, dan Jepang	<i>Quantitative - discrete</i>
Global_Sales	Penjualan <i>video games</i> di seluruh dunia	<i>Quantitative - discrete</i>
Critic_Score	Nilai kritik dari <i>video games</i>	<i>Categorical - ordinal</i>
Critic_Count	Jumlah penilai dari informasi Critic_Score	<i>Quantitative - discrete</i>
User_Score	Penilaian dari user terhadap <i>video games</i>	<i>Categorical - ordinal</i>
User_Count	Jumlah pengguna dari <i>video games</i>	<i>Quantitative - discrete</i>
Developer	Perusahaan yang membuat <i>video games</i>	<i>Categorical - nominal</i>
Rating	Penilaian keseluruhan <i>video games</i>	<i>Categorical - ordinal</i>

### 2.2.2. Kolom dan Karakteristik Data

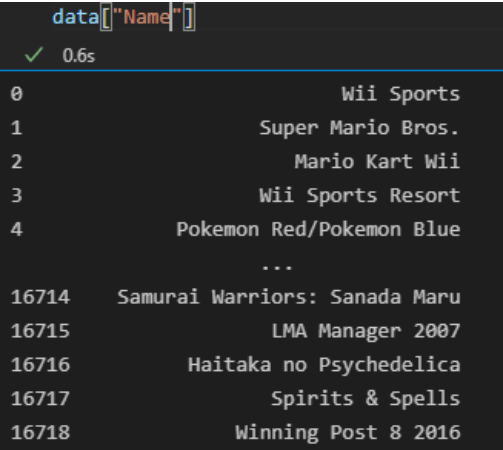
Tabel Karakteristik dalam Setiap Kolom

Kolom	Karakteristik
Name	Wii Sports, Super Mario Bros., Mario Cart Wii, Wii Sports Resort, Pokemon Red/Pokemon Blue, ..., Samurai Warriors: Sanada Maru, LMA Manager 2007, Haitaka no Psychedelica, Spirits & Spells, Winning Post 8 2016  <b>Catatan:</b> Data kolom “Name” tidak bisa ditampilkan semua karena berjumlah sangat banyak.
Platform	PS2, DS, PS3, Wii, X360, PSP, PS, PC, XB, GBA, GC, 3DS, PSV, PS4 N64, XOne, SNES, SAT, WiiU, 2600, NES, GB, DC, GEN, NG, SCD, WS, 3DO, TG16, GG, PCFX
Year_of_Release	1985 - 2016



Publisher	Electronic Arts, Ubisoft, Activision, Sony Computer Entertainment, THQ, ..., DSI Games, Sunsoft, O3 Entertainment, Tetris Online, NewKidCo  <b>Catatan:</b> Data kolom “Publisher” tidak bisa ditampilkan semua karena berjumlah sangat banyak.
Genre	Action, Sports, Misc, Role-Playing, Shooter, Adventure, Racing, Platform, Simulation, Fighting, Strategy, Puzzle
NA_Sales	0 – 41,36 juta
EU_Sales	0 – 28,96 juta
JP_Sales	0 – 10,22 juta
Other_Sales	0 – 10,57 juta
Global_Sales	0 – 82,53 juta
Critic_Score	13 – 98%
Critic_Count	3 – 113 orang
User_Score	0.5 – 9.6%
User_Count	4 – 10665 orang
Developer	EA Canada, EA Sports, Capcom, Ubisoft, Konami, ..., Sanzaru Games, Inc., SCEA, Think and Feel, Ubisoft Annecy, Omega Force, Koei Canada, Atomic Games  <b>Catatan:</b> Data kolom “Developer” tidak bisa ditampilkan semua karena berjumlah sangat banyak.
Rating	T, E, M, E10+, AO, K-A, RP

*Screenshot* Kode Pemrograman untuk Mencari Tabel Karakteristik

Kolom	Gambar
Name	 <pre>data[['Name']] ✓ 0.6s 0      Wii Sports 1      Super Mario Bros. 2      Mario Kart Wii 3      Wii Sports Resort 4      Pokemon Red/Pokemon Blue ... 16714  Samurai Warriors: Sanada Maru 16715  LMA Manager 2007 16716  Haitaka no Psychedelica 16717  Spirits &amp; Spells 16718  Winning Post 8 2016</pre>

Platform	<pre>data["Platform"].value_counts()</pre> <pre>✓ 0.1s</pre> <table> <tr><td>PS2</td><td>1140</td></tr> <tr><td>X360</td><td>858</td></tr> <tr><td>PS3</td><td>769</td></tr> <tr><td>PC</td><td>651</td></tr> <tr><td>XB</td><td>565</td></tr> <tr><td>Wii</td><td>479</td></tr> <tr><td>DS</td><td>464</td></tr> <tr><td>PSP</td><td>390</td></tr> <tr><td>GC</td><td>348</td></tr> <tr><td>PS4</td><td>239</td></tr> <tr><td>GBA</td><td>237</td></tr> <tr><td>XOne</td><td>159</td></tr> <tr><td>3DS</td><td>155</td></tr> <tr><td>PS</td><td>150</td></tr> <tr><td>PSV</td><td>118</td></tr> <tr><td>WiiU</td><td>89</td></tr> <tr><td>DC</td><td>14</td></tr> </table>	PS2	1140	X360	858	PS3	769	PC	651	XB	565	Wii	479	DS	464	PSP	390	GC	348	PS4	239	GBA	237	XOne	159	3DS	155	PS	150	PSV	118	WiiU	89	DC	14
PS2	1140																																		
X360	858																																		
PS3	769																																		
PC	651																																		
XB	565																																		
Wii	479																																		
DS	464																																		
PSP	390																																		
GC	348																																		
PS4	239																																		
GBA	237																																		
XOne	159																																		
3DS	155																																		
PS	150																																		
PSV	118																																		
WiiU	89																																		
DC	14																																		
Year_of_Release	<pre>print("Range:",data1["Year_of_Release"].min()," - ",data1["Year_of_Release"].max())</pre> <pre>✓ 0.1s</pre> <pre>Range: 1985.0 - 2016.0</pre>																																		
Publisher	<pre>data["Publisher"].value_counts()</pre> <pre>✓ 0.1s</pre> <table> <tr><td>Electronic Arts</td><td>944</td></tr> <tr><td>Ubisoft</td><td>496</td></tr> <tr><td>Activision</td><td>492</td></tr> <tr><td>Sony Computer Entertainment</td><td>316</td></tr> <tr><td>THQ</td><td>307</td></tr> <tr><td>...</td><td></td></tr> <tr><td>DSI Games</td><td>1</td></tr> <tr><td>Sunsoft</td><td>1</td></tr> <tr><td>03 Entertainment</td><td>1</td></tr> <tr><td>Tetris Online</td><td>1</td></tr> <tr><td>NewKidCo</td><td>1</td></tr> </table>	Electronic Arts	944	Ubisoft	496	Activision	492	Sony Computer Entertainment	316	THQ	307	...		DSI Games	1	Sunsoft	1	03 Entertainment	1	Tetris Online	1	NewKidCo	1												
Electronic Arts	944																																		
Ubisoft	496																																		
Activision	492																																		
Sony Computer Entertainment	316																																		
THQ	307																																		
...																																			
DSI Games	1																																		
Sunsoft	1																																		
03 Entertainment	1																																		
Tetris Online	1																																		
NewKidCo	1																																		

Genre	<pre>data["Genre"]</pre> <pre>✓ 0.2s</pre> <pre>0      Sports</pre> <pre>1      Platform</pre> <pre>2      Racing</pre> <pre>3      Sports</pre> <pre>4      Role-Playing</pre> <pre>...</pre> <pre>16714  Action</pre> <pre>16715  Sports</pre> <pre>16716  Adventure</pre> <pre>16717  Platform</pre> <pre>16718  Simulation</pre>
NA_Sales	<pre>print("Range:",data1["NA_Sales"].min()," - ",data1["NA_Sales"].max())</pre> <pre>✓ 0.9s</pre> <pre>Range: 0.0 - 41.36</pre>
EU_Sales	<pre>print("Range:",data1["EU_Sales"].min()," - ",data1["EU_Sales"].max())</pre> <pre>✓ 0.5s</pre> <pre>Range: 0.0 - 28.96</pre>
JP_Sales	<pre>print("Range:",data1["JP_Sales"].min()," - ",data1["JP_Sales"].max())</pre> <pre>✓ 0.2s</pre> <pre>Range: 0.0 - 10.22</pre>
Other_Sales	<pre>print("Range:",data1["Other_Sales"].min()," - ",data1["Other_Sales"].max())</pre> <pre>✓ 0.2s</pre> <pre>Range: 0.0 - 10.57</pre>
Global_Sales	<pre>print("Range:",data1["Global_Sales"].min()," - ",data1["Global_Sales"].max())</pre> <pre>✓ 0.2s</pre> <pre>Range: 0.01 - 82.53</pre>
Critic_Score	<pre>print("Range:",data1["Critic_Score"].min()," - ",data1["Critic_Score"].max())</pre> <pre>✓ 0.2s</pre> <pre>Range: 13.0 - 98.0</pre>
Critic_Count	<pre>print("Range:",data1["Critic_Count"].min()," - ",data1["Critic_Count"].max())</pre> <pre>✓ 0.2s</pre> <pre>Range: 3.0 - 113.0</pre>
User_Score	<pre>print("Range:",data1["User_Score"].min()," - ",data1["User_Score"].max())</pre> <pre>✓ 0.1s</pre> <pre>Range: 0.5 - 9.6</pre>
User_Count	<pre>print("Range:",data1["User_Count"].min()," - ",data1["User_Count"].max())</pre> <pre>✓ 0.2s</pre> <pre>Range: 4.0 - 10665.0</pre>

Developer	<pre>data["Developer"].value_counts()</pre> <pre>✓ 0.2s</pre> <table> <tr><td>EA Canada</td><td>149</td></tr> <tr><td>EA Sports</td><td>142</td></tr> <tr><td>Capcom</td><td>126</td></tr> <tr><td>Ubisoft</td><td>103</td></tr> <tr><td>Konami</td><td>95</td></tr> <tr><td>...</td><td></td></tr> <tr><td>Sanzaru Games, Sanzaru Games, Inc.</td><td>1</td></tr> <tr><td>SCEA, Think and Feel</td><td>1</td></tr> <tr><td>Ubisoft Annecy</td><td>1</td></tr> <tr><td>Omega Force, Koei Canada</td><td>1</td></tr> <tr><td>Atomic Games</td><td>1</td></tr> </table>	EA Canada	149	EA Sports	142	Capcom	126	Ubisoft	103	Konami	95	...		Sanzaru Games, Sanzaru Games, Inc.	1	SCEA, Think and Feel	1	Ubisoft Annecy	1	Omega Force, Koei Canada	1	Atomic Games	1
EA Canada	149																						
EA Sports	142																						
Capcom	126																						
Ubisoft	103																						
Konami	95																						
...																							
Sanzaru Games, Sanzaru Games, Inc.	1																						
SCEA, Think and Feel	1																						
Ubisoft Annecy	1																						
Omega Force, Koei Canada	1																						
Atomic Games	1																						
Rating	<pre>data["Rating"].value_counts()</pre> <pre>✓ 0.1s</pre> <table> <tr><td>T</td><td>2377</td></tr> <tr><td>E</td><td>2082</td></tr> <tr><td>M</td><td>1433</td></tr> <tr><td>E10+</td><td>930</td></tr> <tr><td>AO</td><td>1</td></tr> <tr><td>K-A</td><td>1</td></tr> <tr><td>RP</td><td>1</td></tr> </table>	T	2377	E	2082	M	1433	E10+	930	AO	1	K-A	1	RP	1								
T	2377																						
E	2082																						
M	1433																						
E10+	930																						
AO	1																						
K-A	1																						
RP	1																						

## 2.3. TUGAS 3 (STATISTIK)

### 2.3.1. Menampilkan 5 Data Teratas dari Keseluruhan Data

```
#mencetak 5 data pertama
data.head(5)
```

	Name	Platform	Year_of_Release	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	Wii Sports	Wii	2006.0	Sports	Nintendo	41.36	28.96	3.77	8.45	82.53
1	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24
2	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.68	12.76	3.79	3.29	35.52
3	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.61	10.93	3.28	2.95	32.77
4	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.37

### 2.3.2. Menampilkan 3 Game dengan Penjualan Global Tertinggi

```
#mencari data game dengan penjualan global tertinggi
global_sales = data.sort_values("Global_Sales", ascending=False)
#mencetak 3 data pertama
global_sales.head(3)
```

	Name	Platform	Year_of_Release	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	Wii Sports	Wii	2006.0	Sports	Nintendo	41.36	28.96	3.77	8.45	82.53
1	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24
2	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.68	12.76	3.79	3.29	35.52

### 2.3.3. Menampilkan 3 Game dengan Tahun Rilis Terlama

```
#mencari data tahun rilis terlama
year_of_release = data.sort_values("Year_of_Release", ascending=True)
#mencetak 3 data pertama
year_of_release.head(3)
```

	Name	Platform	Year_of_Release	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
262	Asteroids	2600	1980.0	Shooter	Atari	4.00	0.26	0.0	0.05	4.31
5360	Freeway	2600	1980.0	Action	Activision	0.32	0.02	0.0	0.00	0.34
546	Missile Command	2600	1980.0	Shooter	Atari	2.56	0.17	0.0	0.03	2.76

### 2.3.4. Sampel 3 Data Kolom "Name"

```
#mencetak 3 data pertama dari kolom Name
data["Name"].head(3)
```

```
0    Wii Sports
1    Super Mario Bros.
2    Mario Kart Wii
Name: Name, dtype: object
```

### 2.3.5. Sampel 3 Data Kolom "Platform"

```
#mencetak 3 data pertama dari kolom Platform
data["Platform"].head(3)
```

```
0    Wii
1    NES
2    Wii
Name: Platform, dtype: object
```

### 2.3.6. Sampel 3 Data Kolom "Year\_of\_Release"

```
#mencetak 3 data pertama dari kolom Year_of_Release
data["Year_of_Release"].head(3)
```

```
0    2006.0
1    1985.0
2    2008.0
Name: Year_of_Release, dtype: float64
```

### 2.3.7. Sampel 3 Data Kolom “Genre”

```
#mencetak 3 data pertama dari kolom Genre
data["Genre"].head(3)

0    Sports
1    Platform
2    Racing
Name: Genre, dtype: object
```

### 2.3.8. Sampel 3 Data Kolom “Publisher”

```
#mencetak 3 data pertama dari kolom Publisher
data["Publisher"].head(3)

0    Nintendo
1    Nintendo
2    Nintendo
Name: Publisher, dtype: object
```

### 2.3.9. Sampel 3 Data Kolom “Global\_Sales”

```
#mencetak 3 data pertama dari kolom Global_Sales
data["Global_Sales"].head(3)

0    82.53
1    40.24
2    35.52
Name: Global_Sales, dtype: float64
```

Statistik Data Pada Atribut

### 2.3.10. Rata-Rata dari Penjualan Global

```
#mencari rata-rata
data["Global_Sales"].mean()

0.5335426759974459
```

### 2.3.11. Standar Deviasi dari Penjualan Global

```
#mencari standar deviasi
data["Global_Sales"].std()

1.5479351247177218
```

### 2.3.12. Persentil 10% dari Penjualan Global

```
#mencari persentil 10%
data["Global_Sales"].quantile(0.10)

0.02
```

### 2.3.13. Persentil 25% dari Penjualan Global

```
#mencari persentil 25%  
data["Global_Sales"].quantile(0.25)
```

0.06

### 2.3.14. Persentil 50% dari Penjualan Global

```
#mencari persentil 50%  
data["Global_Sales"].quantile(0.50)
```

0.17

### 2.3.15. Persentil 75% dari Penjualan Global

```
#mencari persentil 75%  
data["Global_Sales"].quantile(0.75)
```

0.47

### 2.3.16. Persentil 90% dari Penjualan Global

```
#mencari persentil 90%  
data["Global_Sales"].quantile(0.9)
```

1.2

### 2.3.17. Nilai Maksimum dari Penjualan Global

```
#mencari nilai maksimum  
data["Global_Sales"].max()
```

82.53

### 2.3.18. Nilai Minimum dari Penjualan Global

```
#mencari nilai minimum  
data["Global_Sales"].min()
```

0.01

### 2.3.19. Distribusi Frekuensi dari Publisher

```
#mencari data publisher
publisher = data["Publisher"].value_counts()
publisher.head(5)
```

Electronic Arts	1356
Activision	985
Namco Bandai Games	939
Ubisoft	933
Konami Digital Entertainment	834

Name: Publisher, dtype: int64

Pada laporan ini hanya ditunjukkan perhitungan statistik pada salah satu atribut data karena dengan cara yang sama dapat diperoleh statistik untuk atribut data yang lain.

### 2.3.20. Informasi/Pengetahuan

Dari statistik di atas, ada beberapa informasi dan pengetahuan yang didapatkan.

*Video Games* 3 teratas dengan jumlah penjualan terbanyak berasal dari Publisher yang sama yaitu Nintendo dengan Genre *Sports*, *Platform*, dan *Racing*. *Video Games* dengan penjualan terbanyak juga berasal dari tahun rilis yang relatif lama, berkisar di antara tahun 1985 sampai 2008 dengan Platform Wii dan NES. *Video Games* dengan penjualan terbanyak adalah *Wii Sports* yang rilis di tahun 2006 pada Platform Wii. *Game* yang di-publish Nintendo ini merupakan *game* dengan Genre *Sports* yang terjual sebanyak 82,53 Juta.

*Video Games* pertama kali dijual pada tahun 1980. Ada 3 *Games* yang dijual pada tahun itu, yaitu *Asteroid*, *Freeway*, dan *Missile Command* di Platform 2600. Di antara ketiga *Games* itu, *Asteroid* merupakan yang memiliki penjualan terbanyak hingga tahun 2020 yaitu sebanyak 4,3 juta.

Rata-rata penjualan seluruh *Video Games* di dunia adalah 0,534 juta dengan penjualan terbanyak sebesar 82,53 juta dan penjualan paling sedikit sebesar 0,01 juta. Diketahui nilai standar deviasi sebesar 1,548 juta yang artinya nilai rata-rata merupakan representatif yang buruk dari data penjualan *Video Games* secara keseluruhan.

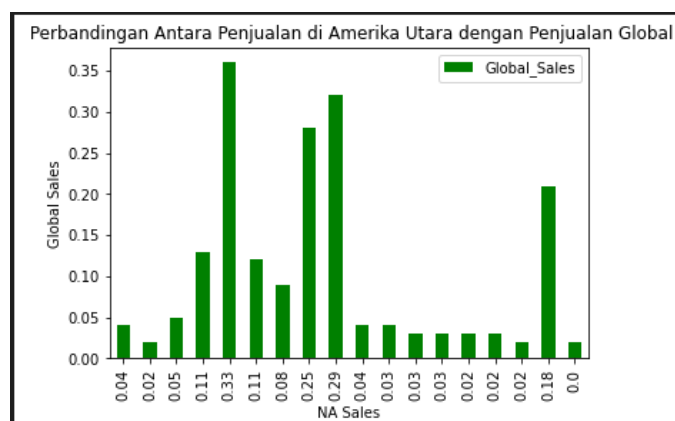
Dari tahun 1980 sampai 2020, Electronic Arts merupakan Publisher dengan rilisan *Video Games* terbanyak dengan total 1356 *games*, disusul oleh Activision sebanyak 985 *games*, Namco Bandai Games sebanyak 939 *games*, Ubisoft sebanyak 933 *games*, dan Konami Digital sebanyak 834 *games*.

## 2.4. TUGAS 4 (VISUALISASI)

### 2.4.1. Perbandingan Kategori

#### 2.4.1.1. Perbandingan Penjualan di Amerika Utara dengan Penjualan Global Berdasarkan Zushi Games

```
import matplotlib.pyplot as plt
#mencari data Zushi Games
data_zhushi_games = data.loc[data["Publisher"] == "Zushi Games"].sort_values("Year_of_Release", ascending=True)
#melakukan plotting
data_zhushi_games.plot(kind="bar", x = "NA_Sales", y = "Global_Sales", title = "Perbandingan Antara Penjualan di Amerika Utara dengan Penjualan Global", color=
plt.xlabel("NA Sales")
plt.ylabel("Global Sales")
plt.show()
#Penggunaan warna hijau mengindikasikan penjualan yang cenderung fluktuatif
```

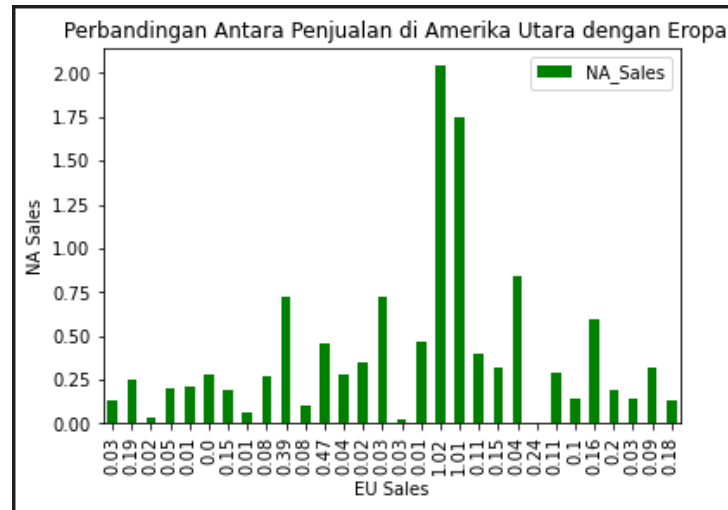




*Insight:* Berdasarkan grafik di atas, penjualan *games* dari Zushi Games di Amerika Utara dan Global memiliki perbandingan yang hampir sama. Hal ini menunjukkan bahwa penjualan *games* ini mayoritas terjual di Amerika Utara. Namun, ada satu *game* yang tidak terjual di Amerika Utara tetapi terjual di negara lainnya.

#### 2.4.1.2. Perbandingan Penjualan Games Warner Bros. Interactive Entertainment Amerika Utara dengan Eropa

```
#mencari data Warner Bros. Interactive Entertainment
data_warner_bros_interactive_entertainment = data.loc[data["Publisher"] == "Warner Bros. Interactive Entertainment"].sort_values("Year_of_Release", ascending=True)
#melakukan plotting
data_warner_bros_interactive_entertainment[:30].plot(kind="bar", x = "EU_Sales", y = "NA_Sales", title = "Perbandingan Antara Penjualan di Amerika Utara dengan Eropa")
plt.xlabel("EU Sales")
plt.ylabel("NA Sales")
plt.show
#Penggunaan warna hijau mengindikasikan penjualan yang cenderung fluktuatif
```

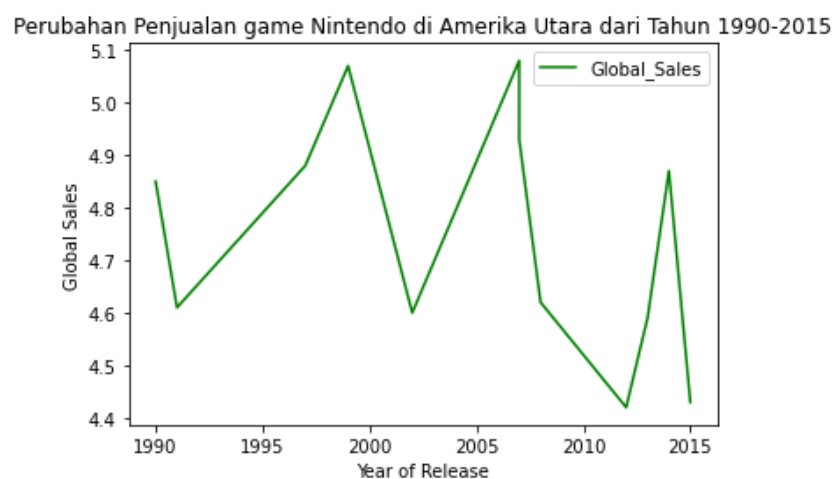


*Insight:* Berdasarkan grafik di atas, jumlah penjualan *games* dari Warner Bros. Interactive Entertainment di Amerika Utara dan Eropa berbeda-beda. Namun, ada beberapa *games* yang penjualan yang memiliki total penjualan yang hampir sama. Hal ini mungkin disebabkan karena minat dan juga jumlah populasi yang berbeda antara Amerika Utara dan Eropa.

#### 2.4.2. Penampilan Perubahan Terhadap Waktu

##### 2.4.2.1. Grafik Penjualan Global dari Nintendo dari Tahun 1990-2015

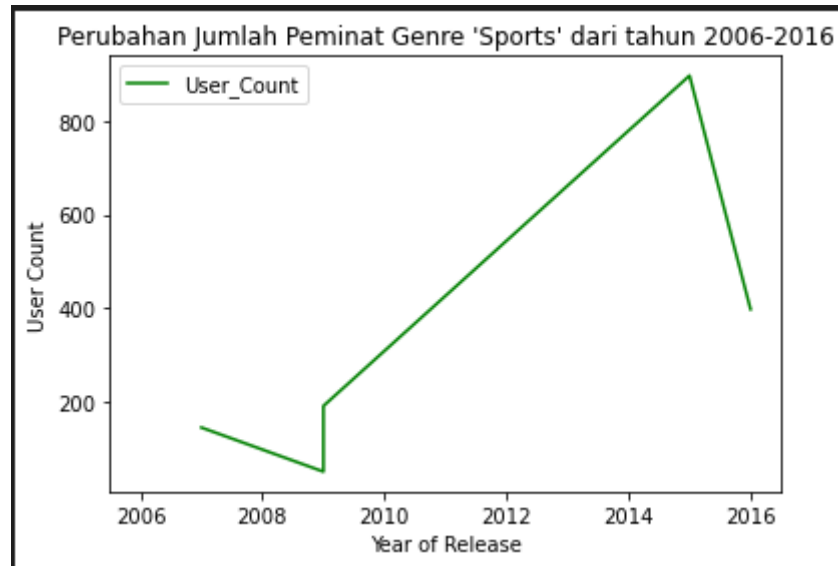
```
#mencari data Nintendo
data_time_series_nintendo = data[200:250].loc[data["Publisher"]=="Nintendo"].sort_values("Year_of_Release", ascending=True)
#melakukan plotting
data_time_series_nintendo.plot(kind="line", x = "Year_of_Release", y = "Global_Sales", title="Perubahan Penjualan game Nintendo di Amerika Utara dari Tahun 1990-2015")
plt.xlabel("Year of Release")
plt.ylabel("Global Sales")
plt.show()
```



*Insight:* Berdasarkan grafik di atas, penjualan game pada konsol Nintendo mengalami naik turun di Amerika Utara. Hal ini mungkin terjadi akibat ketersediaan *game* yang disediakan konsol Nintendo setiap tahunnya berbeda-beda.

#### 2.4.2.2. Grafik Perubahan Jumlah Peminat Genre “Sports” dari Tahun 2006-2016

```
#mencari data dengan genre sports
data_time_series_nintendo = data[100].loc[data["Genre"]=="Sports"].sort_values("Year_of_Release", ascending=False)
#melakukan plotting
data_time_series_nintendo.plot(kind="line", x = "Year_of_Release", y = "User_Count", title="Perubahan Jumlah Peminat Genre 'Sports' dari tahun 2006-2016",
plt.xlabel("Year of Release")
plt.ylabel("User Count")
plt.show()
```

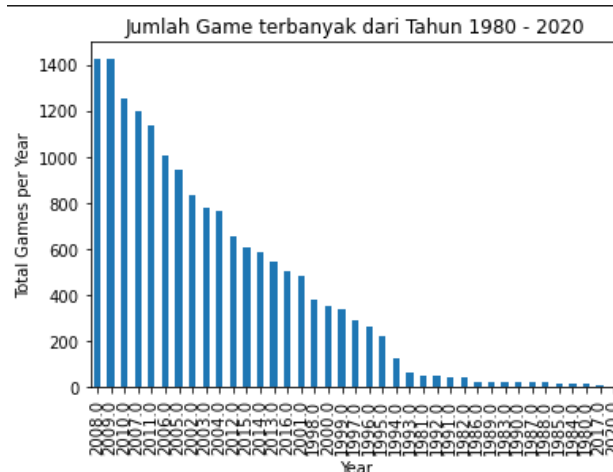


*Insight:* Berdasarkan grafik di atas, terjadi tren menurun dari jumlah peminat *game* dengan genre *Sports* di tahun 2006-2009 dan 2015-2016 dan tren naik di tahun 2009-2015. Tren naik di tahun 2009-2015 ini mungkin terjadi karena perkembangan di dunia olahraga ataupun makin banyak Developer yang membuat *game* dengan genre *Sports*. Tren menurun setelah tahun 2015 mungkin terjadi akibat ukuran file dari *game* yang semakin besar sehingga user-user di tahun sebelumnya perlu meningkatkan konsol yang mereka punya.

#### 2.4.3. Penampilan Hierarki Dan Hubungan Keseluruhan Bagian

##### 2.4.3.1. Jumlah Game Terbanyak dari Tahun 1980-2020

```
#mencari data dari total game yang dirilis tiap tahun
data_total_game = data["Year_of_Release"]
#melakukan plotting
data_total_game.value_counts().plot(kind="bar", stacked = True, title="Jumlah Game terbanyak dari Tahun 1980 - 2020")
plt.xlabel("Year")
plt.ylabel("Total Games per Year")
plt.show()
```



*Insight:* Berdasarkan diagram di atas, dari tahun 2008 sampai 2020, terjadi tren menurun total *videogames* per tahunnya. Hal ini menunjukkan adanya penurunan produksi *videogames* setiap tahunnya. Ini mungkin disebabkan karena biaya produksi *game* yang makin mahal akibat perkembangan teknologi.

#### 2.4.4. Genre Games Terbanyak dari Tahun 1980-2020

```
#mencari data dari total genre games dari tahun 1980-2020
data_total_game = data["Genre"]
#melakukan plotting
data_total_game.value_counts().plot(kind="pie", title="Genre games terbanyak dari Tahun 1980 - 2020")
plt.show()
```

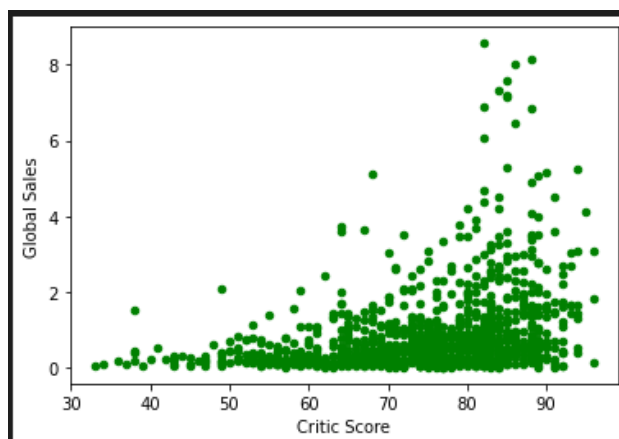


*Insight:* Berdasarkan diagram di atas, dapat diketahui bahwa genre *video games* terbanyak di antara tahun 1980 sampai 2020 adalah genre *Action*. Hal ini menunjukkan bahwa genre *Action* lebih diminati dan lebih banyak terjual dibandingkan genre lainnya.

#### 2.4.5. Plotting Relationships

##### 2.4.5.1. Scatter Plot dari Penjualan Global Electronic Arts terhadap Year of Release

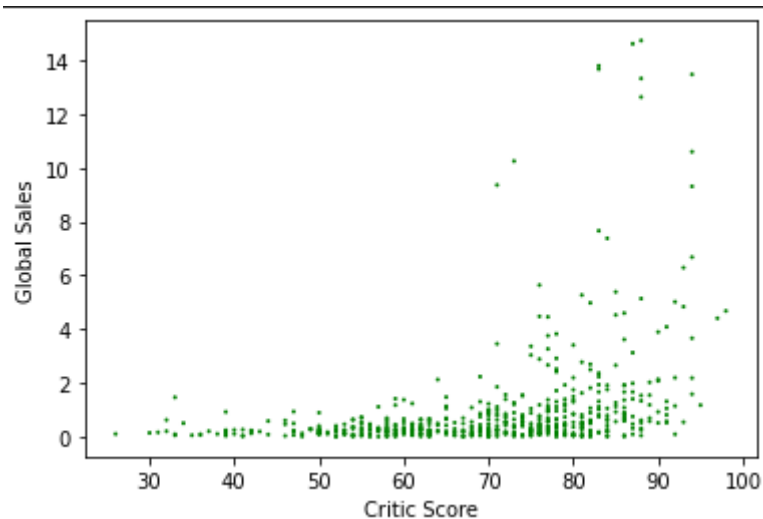
```
#mencari data Electronic Arts
data_electronic_arts = data.loc[data["Publisher"]=="Electronic Arts"].sort_values("Year_of_Release", ascending=True)
#melakukan plotting
data_electronic_arts.plot(kind="scatter", x = "Critic_Score", y = "Global_Sales", color="green")
plt.xlabel("Critic Score")
plt.ylabel("Global Sales")
plt.show()
```



*Insight:* Berdasarkan scatter plot di atas, korelasi antara Critic Score dan Global Sales pada penjualan Global Electronic Arts menunjukkan *Low Positive Correlation* yang artinya korelasi positif yang terlihat samar. Mungkin variabel Critic Score memengaruhi variabel Global Sales, tetapi tingkat pengaruhnya masih diragukan. Ada variabel lain yang perlu dianalisis.

#### 2.4.5.2. Bubble Plot dari Penjualan Global Activision terhadap Year of Release, dengan Sizes Merupakan Frekuensi Video Games dari Publisher Activision

```
#mencari data Activision
data_electronic_arts = data.loc[data["Publisher"]=="Activision"].sort_values("Year_of_Release", ascending=True)
#melakukan plotting
data_electronic_arts.plot(kind="scatter", sizes = data_electronic_arts.value_counts(), x = "Critic_Score", y = "Global_Sales",
plt.xlabel("Critic Score")
plt.ylabel("Global Sales")
plt.show()
```



*Insight:* Berdasarkan Bubble plot di atas, Critic Score, Global Sales, dan frekuensi video games dari Publisher Activision memiliki korelasi yang samar. Hal ini menunjukkan kemungkinan adanya korelasi diantara ketiga variabel, tetapi tingkat pengaruhnya masih diragukan.

## 2.5. TUGAS 5 (KORELASI)

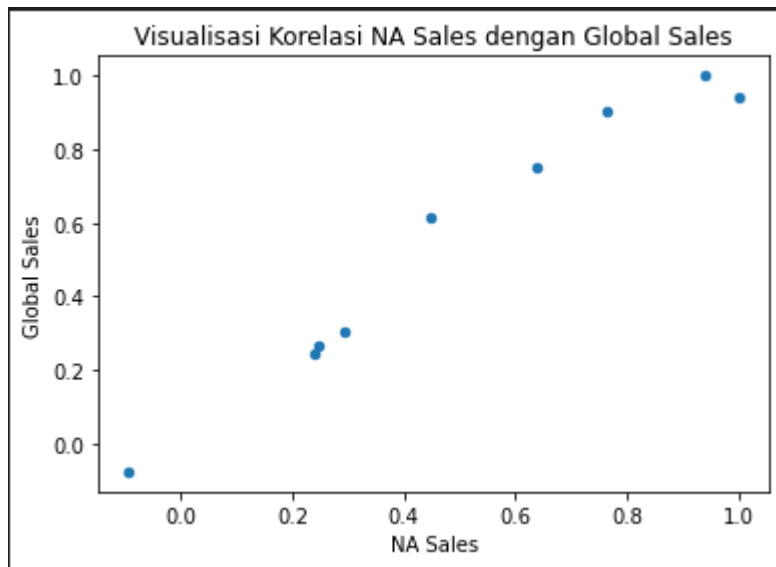
Korelasi antara semua atribut kuantitatif yang ada dalam data

```
print("Korelasi antara semua atribut/data kuantitatif adalah")
#menampilkan korelasi
data.corr(method="pearson")
```

	Year_of_Release	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales	Critic_Score	Critic_Count	User_Count
Year_of_Release	1.000000	-0.092562	0.003842	-0.168386	0.037700	-0.076433	0.011411	0.223407	0.175339
NA_Sales	-0.092562	1.000000	0.765336	0.449598	0.638654	0.941010	0.240755	0.295413	0.246429
EU_Sales	0.003842	0.765336	1.000000	0.435068	0.722796	0.901239	0.220752	0.277533	0.283360
JP_Sales	-0.168386	0.449598	0.435068	1.000000	0.291096	0.612300	0.152593	0.180219	0.075638
Other_Sales	0.037700	0.638654	0.722796	0.291096	1.000000	0.749242	0.198554	0.251639	0.238982
Global_Sales	-0.076433	0.941010	0.901239	0.612300	0.749242	1.000000	0.245471	0.303571	0.265012
Critic_Score	0.011411	0.240755	0.220752	0.152593	0.198554	0.245471	1.000000	0.425504	0.264376
Critic_Count	0.223407	0.295413	0.277533	0.180219	0.251639	0.303571	0.425504	1.000000	0.362334
User_Count	0.175339	0.246429	0.283360	0.075638	0.238982	0.265012	0.264376	0.362334	1.000000

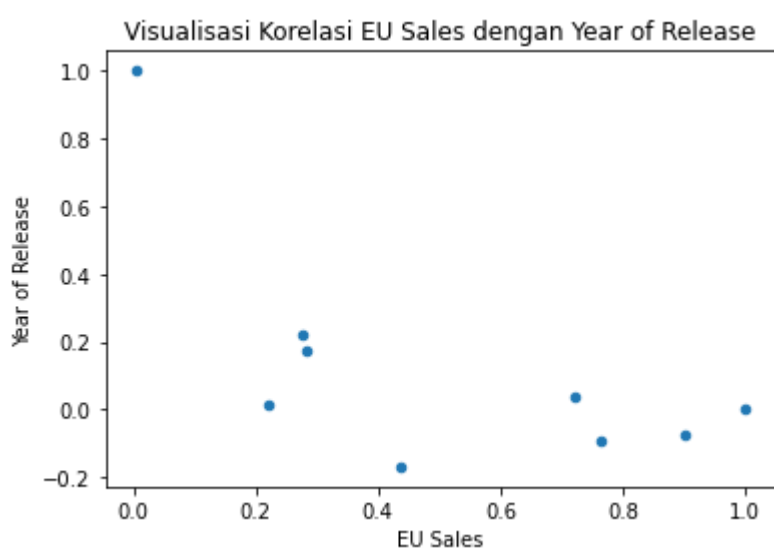
Visualisasi korelasi antara penjualan Amerika Utara dengan penjualan global

```
#mencari data korelasi
correlation = data.corr(method="pearson")
#melakukan plotting
correlation.plot(kind="scatter", x = "NA_Sales", y = "Global_Sales", title="Visualisasi Korelasi NA Sales dengan Global Sales")
plt.xlabel("NA Sales")
plt.ylabel("Global Sales")
plt.show()
```



Visualisasi korelasi antara penjualan Eropa dengan Year of Release

```
#mencari data korelasi
correlation = data.corr(method="pearson")
#melakukan plotting
correlation.plot(kind="scatter", x="EU_Sales", y="Year_of_Release", title="Visualisasi Korelasi EU Sales dengan Year of Release")
plt.xlabel("EU Sales")
plt.ylabel("Year of Release")
plt.show()
```



Kesimpulan dari korelasi yang didapat:

Korelasi untuk setiap data kuantitatif menunjukkan adanya variasi korelasi dalam data. Suatu data memiliki koefisien korelasi yang mendekati 1 dengan data data yang lain menunjukkan bahwa kedua data berkorelasi lurus dan sebaliknya jika koefisien korelasi mendekati  $-1$  menunjukkan bahwa kedua data berkorelasi terbalik. Adapun koefisien korelasi yang bernilai 0 atau mendekati 0 menunjukkan tidak adanya korelasi/hubungan antara kedua data. Hal ini dapat diperjelas dengan visualisasi korelasi antara 2 data di atas. Grafik yang acak menunjukkan kedua data tidak berkorelasi dan untuk grafik yang teratur menunjukkan kedua data memiliki hubungan/korelasi.

## 2.6. TUGAS 6 (DATA CLEANSING)

### 2.6.1. Data Mentah

Banyak data kosong pada dataframe kami

```
# Menghitung jumlah data bernilai Null atau kosong
print("Data kosong : ", data.isnull().sum().sum())
```

✓ 5.6s

Data kosong : 46716

Banyak data yang tidak logis pada dataframe kami

```
# menghitung jumlah data tidak logis
not_logical_data = data.loc[
    (data["Year_of_Release"] < 0) |
    (data["NA_Sales"] < 0) |
    (data["EU_Sales"] < 0) |
    (data["JP_Sales"] < 0) |
    (data["Other_Sales"] < 0) |
    (data["Global_Sales"] < 0) |
    (data["Critic_Score"] < 0) |
    (data["Critic_Count"] < 0) |
    (data["User_Count"] < 0)
]
print('Data yang tidak logis :', len(not_logical_data))
```

✓ 0.5s

Data yang tidak logis : 0

Gambar ini menunjukkan bahwa data kami tidak memiliki data yang bersifat tidak logis. Oleh karena itu data kami hanya terdapat data kosong yang menyebabkan dataframe kami kotor.

### 2.6.2. Pembersihan DataFrame

Untuk membersihkan data yang kosong, kami menggunakan fungsi `dropna()` dalam pandas:

```
#membersihkan data kosong
data.dropna(inplace=True)
```

Parameter “`inplace = Bool`” pada fungsi “`dropna()`” berfungsi untuk mengembalikan dataframe baru yang telah dibersihkan komponen data yang kosong atau bernilai null/nan.

Hasil DataFrame baru:

```
# Menghitung jumlah data bernilai Null atau kosong
print("Data kosong : ", data.isnull().sum().sum())
```

✓ 0.1s

Data kosong : 0

## BAB III

### PENUTUP

#### 3.1. KESIMPULAN DAN LESSON LEARNED

Setelah menyelesaikan tugas ini, kami mendapatkan kesimpulan bahwa data yang kami gunakan untuk menganalisis data merupakan data *video games* dengan penjualan terbanyak di dunia hingga tahun 2020, dengan format csv, dimensi data 16.720 baris dan 16 kolom, serta memiliki ukuran 1.581 KB. Pengumpulan data dilakukan dengan mengakses data yang ingin dianalisis, yaitu “Video Game Sales”, melalui situs <https://www.kaggle.com/gregorut/videogamesales>.

Karakteristik data yang diambil beragam mulai dari *Categorical-nominal*, *Categorical-ordinal*, *Quantitative-continues*, dan *Quantitative-discrete*. Setelah dilakukan pengolahan data maka didapatkan data/rincian statistik seperti rata-rata, standar deviasi, persentile (10%, 25%, 50%, 75%, 90%), nilai maksimum dan minimum serta distribusi frekuensi nilai pada data. Setelah itu, dilakukan visualisasi data dalam bentuk bagan, grafik, dll untuk memudahkan pengambilan informasi. Dari hasil visualisasi data, dapat diketahui perbandingan dari setiap kategori, perbandingan terhadap waktu, serta hierarki dan hubungan keseluruhan bagian. Kami juga mendapatkan korelasi antar data sehingga dapat diketahui apakah data-data berkorelasi lurus, terbalik, ataupun tidak berkorelasi. Selain itu, dikarenakan banyaknya data yang kosong dalam dataset kami, maka dilakukan data cleansing supaya informasi yang didapatkan tidak salah.

Dalam suatu penelitian, seseorang pasti perlu menganalisis dan memvisualisasi data karena banyaknya data yang harus diolah. Dengan adanya perlengkapan dan metode analisis dan visualisasi data ini, akan memudahkan kita dalam mengolah data yang tak terhitung jumlahnya menjadi pengetahuan dan informasi baru yang akurat dan terpercaya. Serta bisa menarik kesimpulan tentang karakteristik yang dicari dalam sebuah data yang lebih simpel dan enak dibaca oleh orang awam.

**DAFTAR PUSTAKA**

- Anderton, Kevin. 2019. "Research Report Shows How Much Time We Spend Gaming [Infographic]." *Forbes*, diperoleh dari <https://www.forbes.com/sites/kevinanderton/2019/03/21/research-report-shows-how-much-time-we-spend-gaming-infographic/?sh=7589a88e3e07>.
- Smith, Gregory. 2016. "Video Game Sales: Analyze sales data from more than 16,500 games." *Kaggle*, diperoleh dari <https://www.kaggle.com/gregorut/videogamesales>.
- Statista Research Department. 2021. "Number of video game users in the United States from 2017-2025 (in millions)." *Statista*, diperoleh dari <https://www.statista.com/forecasts/1277728/physical-or-digital-core-gamers-in-the-us>.



**PEMBAGIAN TUGAS**

<b>Nama</b>	<b>NIM</b>	<b>Pembagian Tugas</b>
Albert	16521192	Membuat PPT dan Bab 3
Daven Darmawan Sendjaya	16521322	Pengerjaan Bab 1 – Bab 2.2, Merapikan Laporan
Jimly Firdaus	16521312	Pembuatan <i>Source Code</i> dan Pengerjaan Analisis Data
Muhammad Naufal	16521442	Informasi Statistik, dan <i>Insight</i> Visualisasi