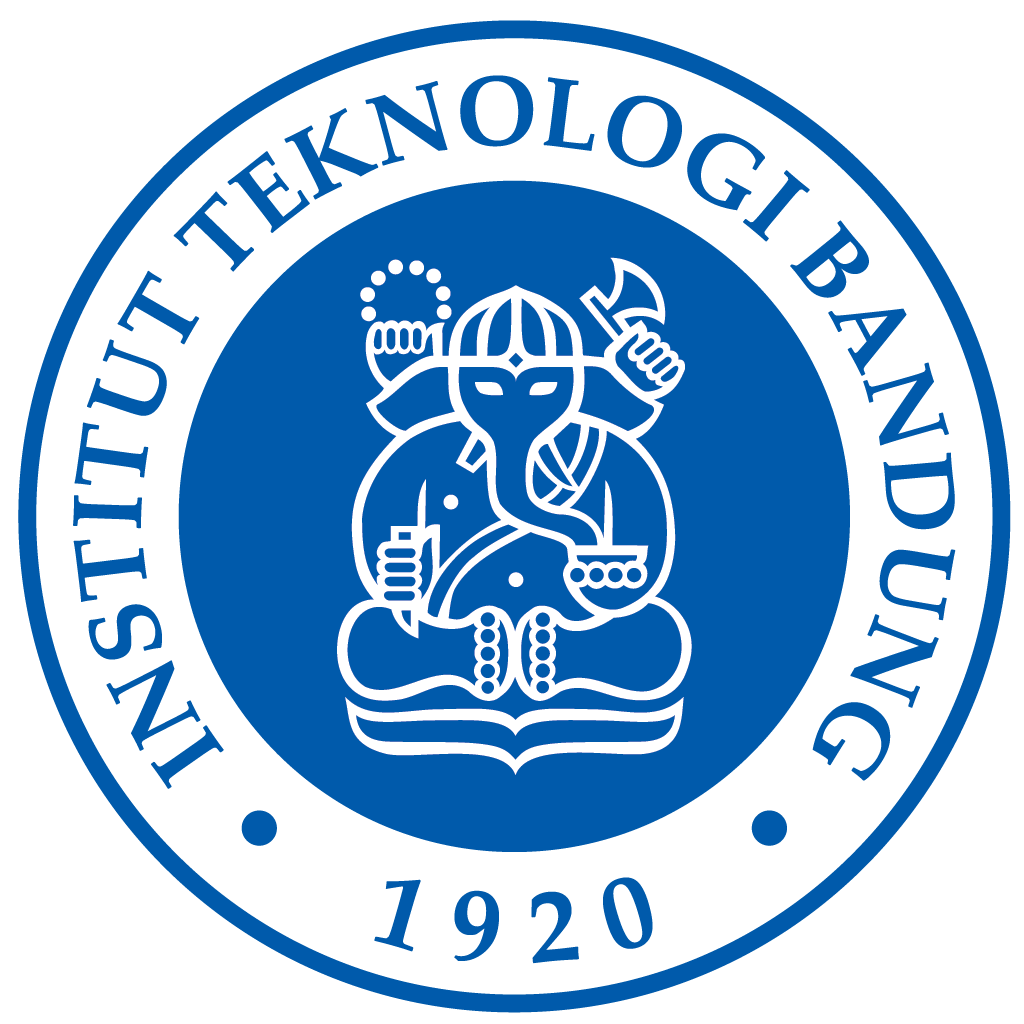
#### **LAPORAN TUGAS BESAR II**

#### **KU1102 PENGENALAN KOMPUTASI**

#### **SEMESTER I 2023-2024**

#### **Analisis Data Penjualan dan Tren Video Game Sepanjang Masa**

#### **(November 2023)**



#### Dosen:

Dr. Fadhil Hidayat, S.Kom., M.T.

#### Disusun oleh:

#### **Yayat Nurhidayat** **16523080**

#### **Nafhan Hadiyan Shafwatudin** **16523213**

#### **Zaka Hanif Nabalah** **19623010**

#### **Atharizza Muhammad Athaya** **19623143**

#### **Ferro Arka Berlian** **19623213**

#### **SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA**

#### **INSTITUT TEKNOLOGI BANDUNG**

#### **2023**

PRAKATA

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa karena berkat rahmat dan karunianya kami dapat menyelesaikan laporan tugas analisis data “Penjualan dan Tren Video Games Sepanjang Masa” secara tepat waktu. Laporan ini kami susun untuk menjelaskan bagaimana cara menganalisis suatu data dan juga bagaimana hasil analisisnya. Adapun tujuan penyusunan laporan ini untuk memenuhi mata kuliah KU1102 Pengenalan Komputasi sebagai tugas besar II semester satu pada perkuliahan di Institut Teknologi Bandung.

Kami menemukan beberapa kendala saat menyusun laporan ini seperti pencarian data yang sesuai dengan syarat, keterbatasan waktu, dan keterbatasan pengetahuan. Namun dengan bantuan dari berbagai pihak, kami dapat menyelesaikan laporan ini tepat pada waktunya. Oleh karena itu, kami mengucapkan terima kasih kepada:

1. Orang tua yang selalu mendukung dan mendoakan kami dalam menyelesaikan laporan ini,
2. Dr. Fadhil Hidayat, S.Kom., M.T., selaku dosen mata kuliah KU1102 Pengenalan Komputasi yang telah membimbing dan memberikan pengetahuannya sehingga kami dapat menyelesaikan tugas dan laporannya dengan baik.

Kami menyadari masih banyak kesalahan dalam penyusunan laporan ini. Oleh karena itu, kami mengharapkan kritik dan saran agar kami dapat menjadi lebih baik kedepannya.

DAFTAR ISI

[PRAKATA 2](#_Toc153825733)

[DAFTAR ISI 3](#_Toc153825734)

[BAB I PENDAHULUAN 4](#_Toc153825735)

[1.1 Latar Belakang 5](#_Toc153825736)

[1.2 Rumusan Masalah 5](#_Toc153825737)

[1.3 Tujuan 5](#_Toc153825738)

[BAB II TEORI DASAR 7](#_Toc153825739)

[2.1 Data Preprocessing 7](#_Toc153825740)

[2.2 Data Analysis 7](#_Toc153825741)

[2.3 Data Visualization 7](#_Toc153825742)

[2.4 Industri Video Game 7](#_Toc153825743)

[BAB III PEMBAHASAN 8](#_Toc153825744)

[3.1 Deskripsi Data dan File 8](#_Toc153825745)

[3.2 Karakteristik Data 8](#_Toc153825746)

[3.2.1 Rank 8](#_Toc153825747)

[3.2.2 Name 8](#_Toc153825748)

[3.2.3 Platform 9](#_Toc153825749)

[3.2.4 Year 9](#_Toc153825750)

[3.2.5 Genre 9](#_Toc153825751)

[3.2.6 Publisher 9](#_Toc153825752)

[3.2.7 NA\_Sales 10](#_Toc153825753)

[3.2.8 EU\_Sales 10](#_Toc153825754)

[3.2.9 JP\_Sales 10](#_Toc153825755)

[3.2.10 Other\_Sales 10](#_Toc153825756)

[3.2.11 Global\_Sales 10](#_Toc153825757)

[3.3 Statistik Data 11](#_Toc153825758)

[3.4 Visualisasi Data 11](#_Toc153825759)

[3.4.1 Distribusi Genre Game Berdasarkan Platform 11](#_Toc153825760)

[3.4.2 Hubungan dan Perbandingan Genre Game dengan Platform 12](#_Toc153825761)

[3.4.3 Jumlah Judul Game Yang Dirlis Dan PenjualannyaBerdasarkan Tahun 13](#_Toc153825762)

[3.4.4 Jumlah Judul Game Yang Dirlis Dan Penjualannya Berdasarkan Platform 14](#_Toc153825763)

[3.4.5 Jumlah Judul Game Yang Dirlis Dan Penjualannya Berdasarkan Publisher 15](#_Toc153825764)

[3.4.6 Jumlah Judul Game Yang Dirlis Dan Penjualannya Berdaasarkan Genre Game 18](#_Toc153825765)

[3.4.7 Persebaran Data Penjualan 20](#_Toc153825766)

[3.5 Korelasi Data 21](#_Toc153825767)

[3.6 Cleansing Data 22](#_Toc153825768)

[BAB IV 24](#_Toc153825769)

[SIMPULAN DAN SARAN 24](#_Toc153825770)

[4.1 Simpulan 24](#_Toc153825771)

[4.2 Saran 24](#_Toc153825772)

[REFERENSI 26](#_Toc153825773)

[LAMPIRAN 27](#_Toc153825774)

BAB I   
PENDAHULUAN

1.1 Latar Belakang

Video game merupakan permainan berbasis elektronik yang melibatkan interaksi antara perengkat lunak permainan dengan pengguna yang memainkannya yang dihubungkan melalui perangkat keras dari permainan tersebut. Awalnya, video game diciptakan sebagai sarana hiburan. Namun seiring berjalannya waktu, video game telah berkembang menjadi alat berkomunikasi, berbisnis, dan edukasi. Video game memiliki berbagai macam genre, mulai dari game berpetualang (*adventure*), teka-teki yang menuntut pemain untuk berpikir keras menyelesaikan masalah, permainan yang melibatkan banyak orang sekaligus (*mmorpg*), hingga game yang sangat digemari pada saat ini yaitu *First* *Person Shooter*.

Dengan berkembangnya industri video game yang semakin pesat, tentunya tren penjualan game juga meningkat dengan pesat. Menurut Statista, di Indonesia sendiri diperkirakan pasar video game akan tumbuh dengan laju pertumbuhan tahunan sebesar 7,64% dan mencapai pendapatan US$1,117.00 juta pada tahun 2023. Pada tahun 2020, pemasukan industri game digital diperkirakan mencapai US$126,5 miliar di seluruh dunia dengan game gratis menumbang 78% dari total pemasukkan industri game digital.

Dari tren penjualan video game, kami ingin melakukan analisis dan visualisasi data terhadap penjualan video game sepanjang masa. Dengan mempelajari data dari Video Game Sales, kami ingin memberikan analisis dan visualisasi,serta mempelajari bagaimana perkembangan dari penjualan game. Data ini mencakup peringkat penjualan game, tahun game dirilis, genre dari game tersebut, penjualan game tersebut pada berbagai negara, dan masih banyak lagi.

Kami menggunakan Phyton sebagai media untuk analisis dan visualisasi data yang kami dapatkan. Phyton merupakan bahasa pemrograman yang dapat digunakan untuk analisis dan visualisasi data. Phyton memiliki banyak *library* yang dapat digunakan untuk membantu dalam mengolah, memvisualisasi, dan menganalisis data dengan mudah.

1.2 Rumusan Masalah

1.2.1 Bagaimana perkembangan penjualan video game dari masa ke masa

1.2.2 Bagaimana hubungan perkembangan penjualan video game dari tahun ke tahun.

1.2.3 Bagaimana visualisasi data dari penjualan video game dari masa ke masa

1.3 Tujuan

1.3.1 Mengetahui bagaimana perkembangan penjualan video game dari masa ke masa

1.3.2 Mengetahui hubungan perkembangan penjualan video game dari tahun ke tahun

1.3.3 Mengetahui bagaimana visualisasi data dari penjualan video game dari masa ke masa

BAB II   
TEORI DASAR

2.1 Data Preprocessing

Data preprocessing adalah proses mengubah data mentah ke dalam bentuk yang lebih mudah dipahami. Proses ini diperlukan untuk memperbaiki kesalahan pada data mentah yang seringkali tidak lengkap dan memiliki format yang tidak teratur.

Data preprocessing melibatkan proses validasi dan imputasi data. Validasi bertujuan untuk menilai tingkat kelengkapan dan akurasi data yang tersaring. Sedangkan imputasi bertujuan memperbaiki kesalahan dan memasukkan nilai yang hilang, baik secara manual atau otomatis melalui program Business Process Automation (BPA).

2.2 Data Analysis

Data analyisis (analisis data) adalah proses pengolahan data dengan tujuan untuk menemukan informasi yang berguna yang dapat dijadikan dasar, dalam pengambilan keputusan untuk solusi suatu permasalahan. Proses analisis ini meliputi kegiatan pengelompokkan data berdasarkan karakteristiknya, melakukan pembersihan data, mentransformasi data, membuat model data untuk menemukan informasi penting dari data tersebut.

2.3 Data Visualization

Data visualization atau visualisasi data adalah proses mengubah data menjadi bentuk visual seperti grafik, diagram, atau peta untuk memudahkan pemahaman dan analisis data. Visualisasi data memungkinkan kita untuk melihat pola, tren, dan anomali dalam data yang mungkin tidak terlihat jika hanya dilihat dalam bentuk tabel atau angka.

2.4 Industri Video Game

Industri video game adalah industri yang berkembang pesat di seluruh dunia. Industri ini mencakup pengembangan, penerbitan, dan penjualan video game. Video game sendiri adalah bentuk hiburan interaktif yang memungkinkan pemain untuk berinteraksi dengan dunia virtual yang diciptakan oleh pengembang game. Industri video game terus berkembang seiring dengan kemajuan teknologi dan meningkatnya minat masyarakat terhadap video game.

Industri video game menghasilkan pendapatan sebesar $152 miliar pada tahun 2019 dan diperkirakan akan terus tumbuh di masa depan. Industri video game juga memberikan dampak positif bagi perekonomian, seperti menciptakan lapangan kerja dan meningkatkan pariwisata. Namun, industri video game juga memiliki dampak negatif, seperti ketergantungan dan kecanduan pada video game.

BAB III  
PEMBAHASAN

3.1 Deskripsi Data dan File

Data yang kami analisis merupakan data penjualan dan tren video game sepanjang masa mulai dari 1980 sampai 2020 dengan salinan mencapai diatas 100.000. Format data ini adalah csv yang diambil dari website kaggle dengan link berikut <https://www.kaggle.com/datasets/gregorut/videogamesales/data>. File data ini berukuran 1,36 MB dengan dimensi data 16598 kolom x 11 baris.

Berikut adalah kode atau program yang kami tulis untuk melakukan loading data dan juga memperoleh informasi lainnya menggunakan bahasa pemrograman python dan library pandas.

import numpy as np

import pandas as pd

import plotly.express as px

import plotly.figure\_factory as ff

from dash import Dash, dcc, html, Input, Output

import torch

df = pd.read\_csv("vgsales.csv")

df

df.info()

df.describe()

3.2 Karakteristik Data

Data ini memiliki 11 kolom yaitu rank, name, platform, year, genre, publisher, NA\_sales, EU\_Sales, JP\_sales, other\_sales, dan global\_sales. Untuk penjelasan makna, jenis, dan karakteristik data setiap kolomnya sebagai berikut

3.2.1 Rank

* + Makna: Peringkat penjualan video game mulai dari penjualan terbanyak.
  + Jenis: -

3.2.2 Name

* + Makna: Nama game yang ada di pasaran.
  + Jenis: -

3.2.3 Platform

* + Makna: Nama game yang ada di pasaran.
  + Jenis: Kategorikal nominal.
  + Nilai unik/terdiri atas: DS, PS2, PS3, Wii, X360, PSP, PS, PC, XB, GBA, GC, 3DS, PSV, PS4, N64, SNES, XOne, SAT, WiiU, 2600, NES, GB, DC, GEN, NG, SCD, WS, 3DO, TG16, GG, PCFX.
  + Data Kosong/Nan(%): 0
  + Bahasa Pemrograman/kode:

print(df['Platform'].unique())

3.2.4 Year

* + Makna: tempat/alat dimana game dimainkan.
  + Jenis: Kuantitatif
  + Range nilai data: 1980-2020
  + Data kosong/NaN(%): 271 (≈1,63%)
  + Bahasa Pemrograman/kode:

print(str(df['Year'].min()) + ' - ' + str(df['Year'].max()))

print(df["Year"].isnull().mean() \* 100)

3.2.5 Genre

* + Makna: Jenis atau tipe dari suatu game.
  + Jenis: Kategorikal nominal.
  + Nilai unik/terdiri atas: Action, Sports, Misc, Role-Playing, Shooter, Adventure, Racing, Platform, Simulation, Fighting, Strategy, Puzzle.
  + Data kosong/Nan(%): 0
  + Bahasa Pemrograman/kode:

print(df['Genre'].unique())

3.2.6 Publisher

* + Makna: Penerbit dari suatu game.
  + Jenis: Kategorikal nominal.
  + Nilai unik/terdiri atas: Electronic Arts, Activision, Namco Bandai Games, Ubisoft, Konami Digital Entertainment, THQ, Nintendo, Sony Computer Entertainment, Sega, Take-Two Interactive, Capcom, Atari, Tecmo Koei, Square Enix, Warner Bros. Interactive Entertainment, Disney Interactive Studios, Unknown, Midway Games, Eidos Interactive, 505 Games, Microsoft Game Studios, D3Publisher, Acclaim Entertainment, Vivendi Games, Codemasters, Idea Factory, Deep Silver, Nippon Ichi Software, Zoo Digital Publishing, Majesco Entertainment, LucasArts, Rising Star Games, Hudson Soft, Banpresto, Bethesda Softworks, Crave Entertainment, dll (terdapat 578 nilai unik jadi tidak ditulis semuanya supaya tidak terlalu banyak).
  + Data Kosong/Nan(%): 58 (≈0,35%)
  + Bahasa Pemrograman/kode:

print(df['Publisher'].unique())

print(df["Year"].isnull().mean() \* 100)

3.2.7 NA\_Sales

* + Makna: Jumlah penjualan suatu game di Amerika Utara dalam juta.
  + Jenis: Kuantitatif.
  + Range nilai data: 0,0 - 41.49
  + Data kosong/NaN(%): 0
  + Bahasa Pemrograman/kode:

print(str(df['NA\_Sales'].min()) + ' - ' + str(df['NA\_Sales'].max()))

3.2.8 EU\_Sales

* + Makna: Jumlah penjualan suatu game di Eropa dalam juta.
  + Jenis: Kuantitatif.
  + Range nilai data: 0,0 - 29,02
  + Data kosong/NaN(%): 0
  + Bahasa Pemrograman/kode:

print(str(df['EU\_Sales'].min()) + ' - ' + str(df['EU\_Sales'].max()))

3.2.9 JP\_Sales

* + Makna: Jumlah penjualan suatu game di Jepang dalam juta.
  + Jenis: Kuantitatif.
  + Range nilai data: 0,0 - 10,22
  + Data kosong/NaN(%): 0
  + Bahasa Pemrograman/kode:

print(str(df['JP\_Sales'].min()) + ' - ' + str(df['JP\_Sales'].max()))

3.2.10 Other\_Sales

* + Makna: Jumlah penjualan suatu game di daerah lain (selain yang telah disebutkan) dalam juta.
  + Jenis: Kuantitatif.
  + Range nilai data: 0,0 - 10,57
  + Data kosong/NaN(%): 0
  + Bahasa Pemrograman/kode:

print(str(df['Other\_Sales'].min()) + ' - ' + str(df['Other\_Sales'].max())).

3.2.11 Global\_Sales

* + Makna: Jumlah penjualan suatu game di seluruh dunia dalam juta.
  + Jenis: Kuantitatif.
  + Range nilai data: 0,01 - 82,74
  + Data kosong/NaN(%): 0
  + Bahasa Pemrograman/kode:

print(str(df['Global\_Sales'].min()) + ' - ' + str(df['Global\_Sales'].max()))

3.3 Statistik Data

Dengan menggunakan command



Kita bisa mendapatkan data statistik penjualan video game sebagai berikut:

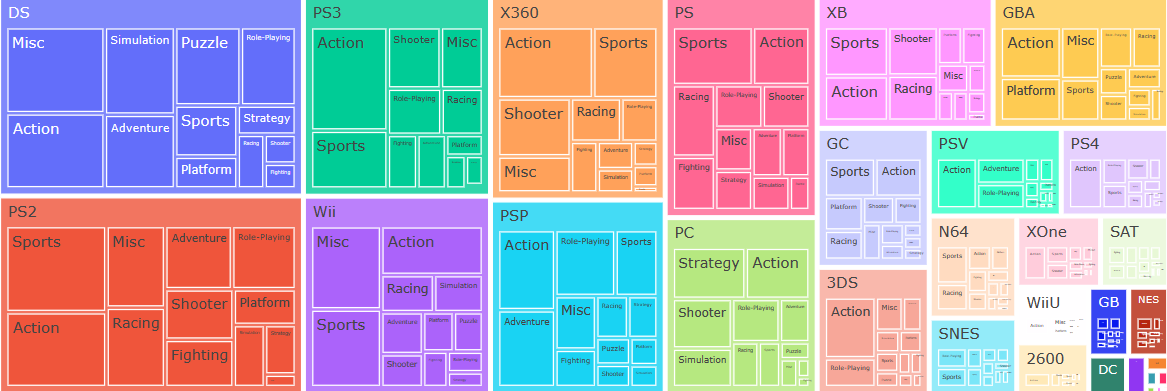


3.4 Visualisasi Data

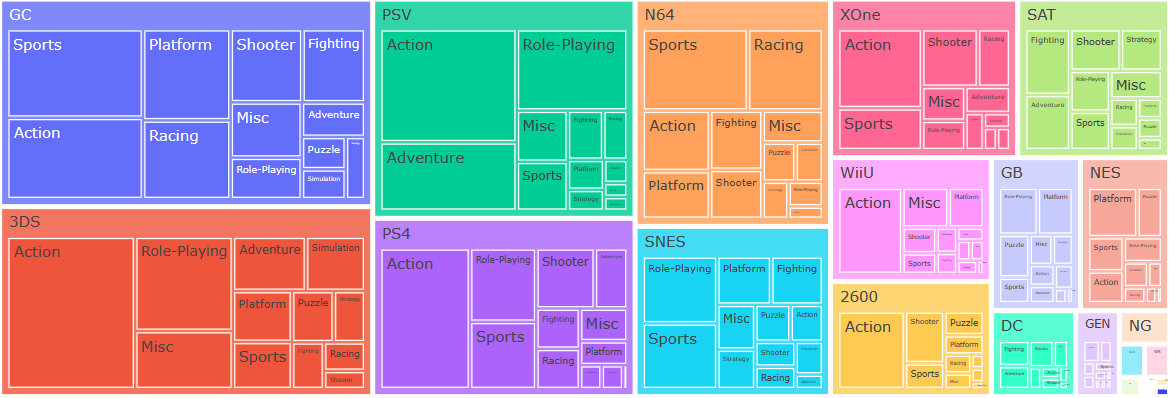
3.4.1 Distribusi Genre Game Berdasarkan Platform

Dengan menggunakan kode

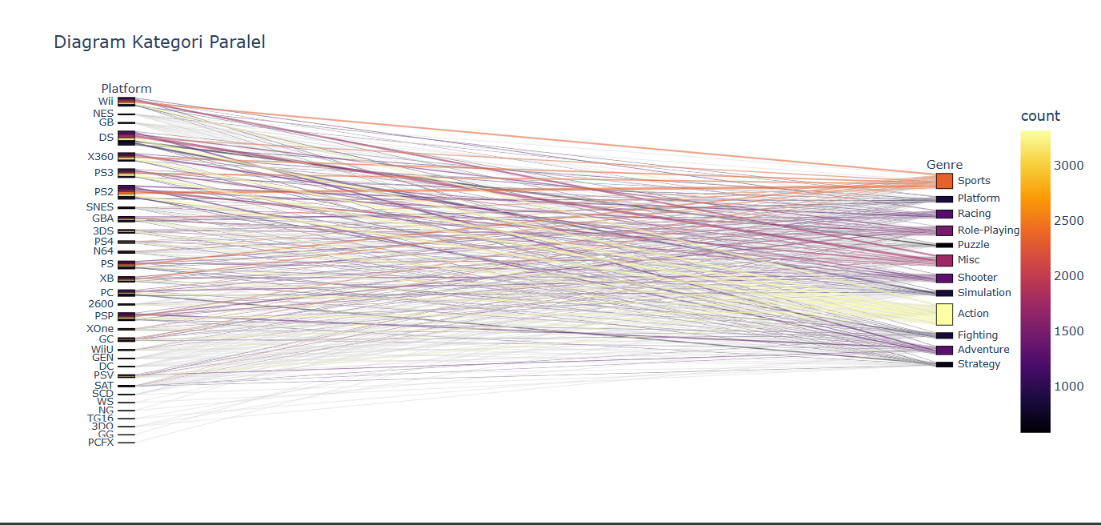
 Didapatkan distribusi genre game berdasarkan platform sebagai berikut:



Adapun terdapat distribusi genre game berdasarkan platform diluar top 10 sebagai berikut:

*Insight: Dari sini bisa kita dapat melihat distribusi genre video games di setiap platformnya. Dapat dilihat bahwa game dengan genre action dan sports mendominasi di banyak platform.*

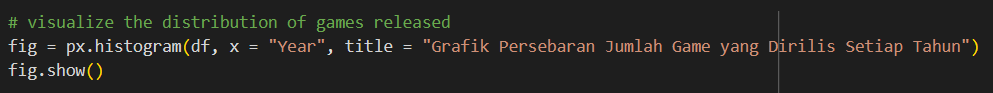
3.4.2 Hubungan dan Perbandingan Genre Game dengan Platform

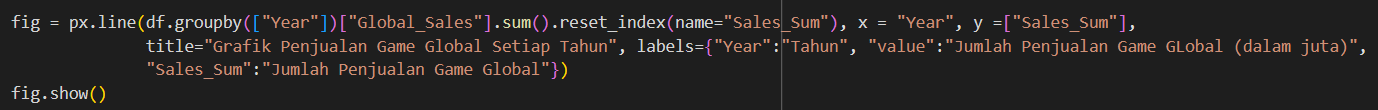


Grafik di atas menunjukkan hubungan sekaligus perbandingan data. Garis yang ditampilkan menunjukkan hubungan data sedangkan warna dan ukuran kotak menunjukkan kuantitas dari data yang terhubung. Dapat diketahui bahwa game dengan genre action memiliki judul game yang dirilis terbanyak dan hubungannya dengan platform yang merilis game tersebut.

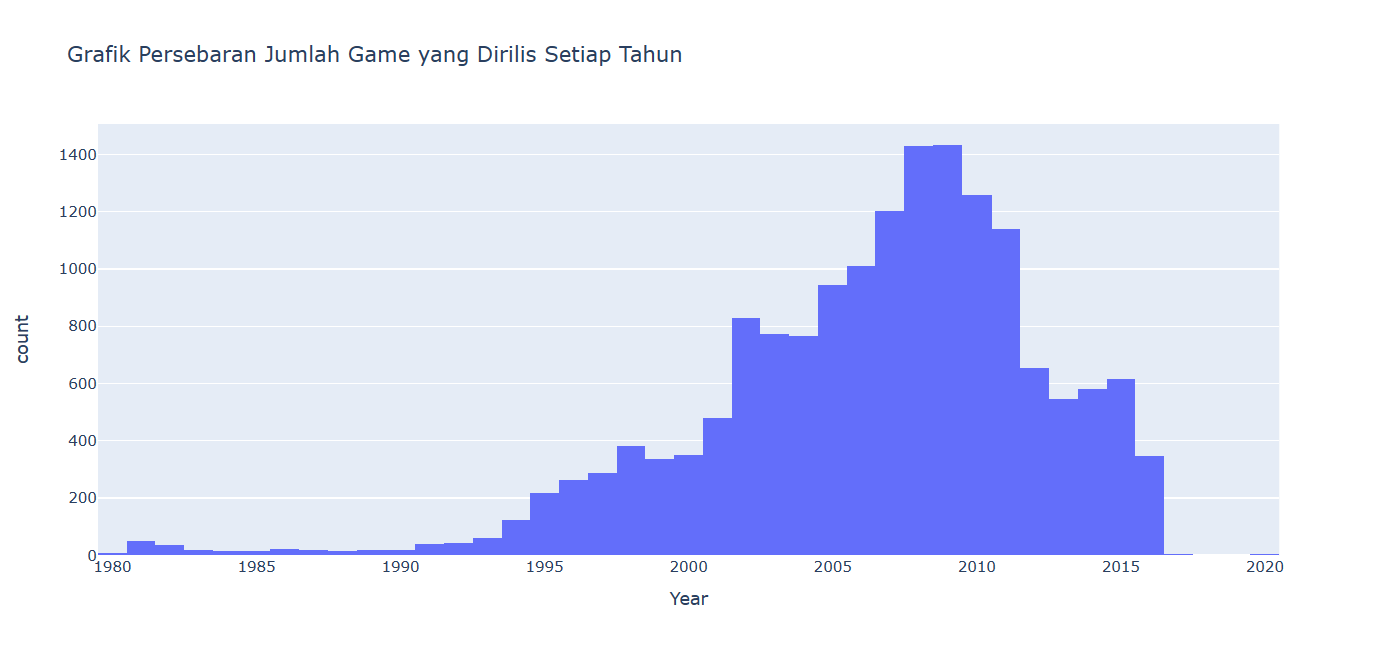
3.4.3 Jumlah Judul Game Yang Dirlis Dan PenjualannyaBerdasarkan Tahun

Dengan menggunakan kode:

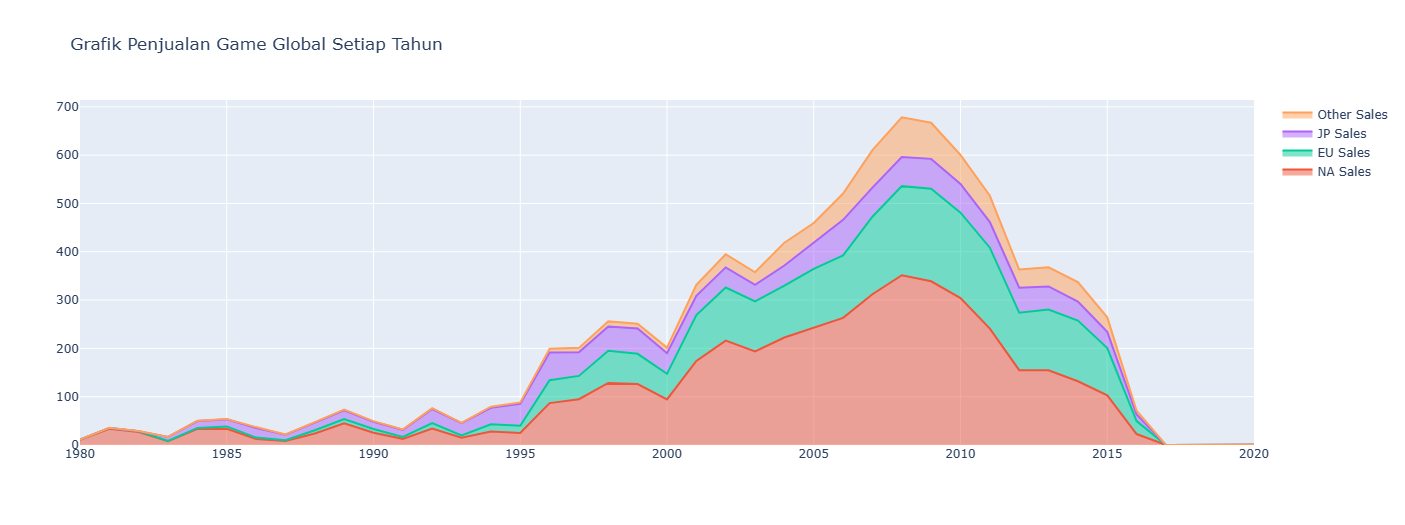
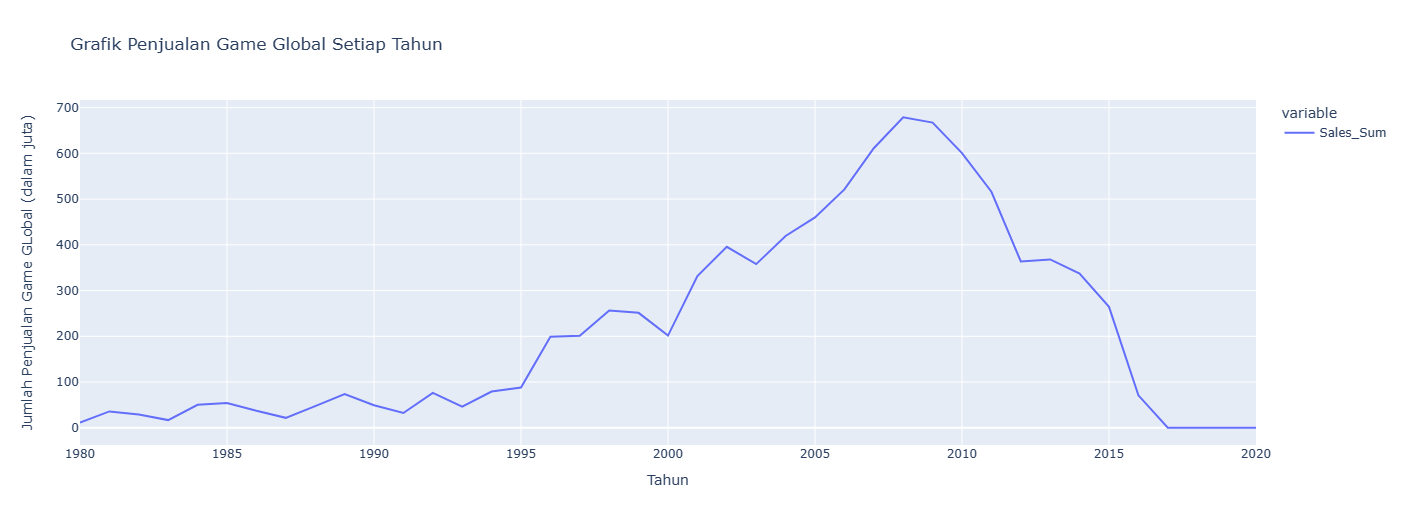




Kita bisa mendapatkan visualisasi data jumlah judul game yang dirilis setiap tahun sebagai berikut:



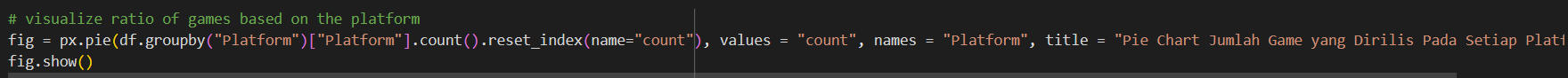
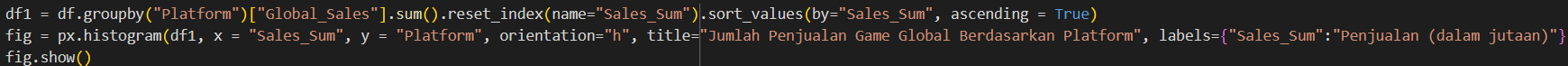
*Insight: Dari grafik persebaran di atas, dapat kita ketahui bahwa jumlah game yang dirilis tertinggi terjadi pada tahun 2008 sampai 2010 dan jumlah game yang dirilis terendah terjadi pada tahun 1983 sampai 1990*

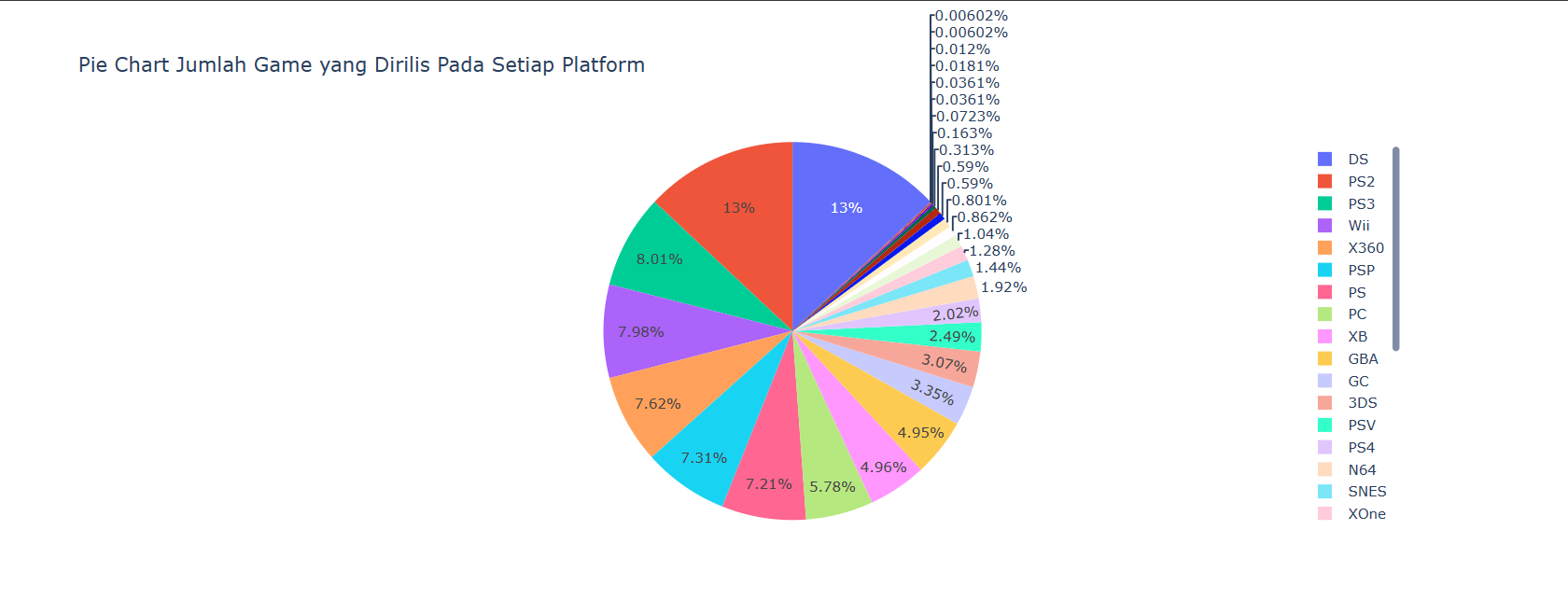


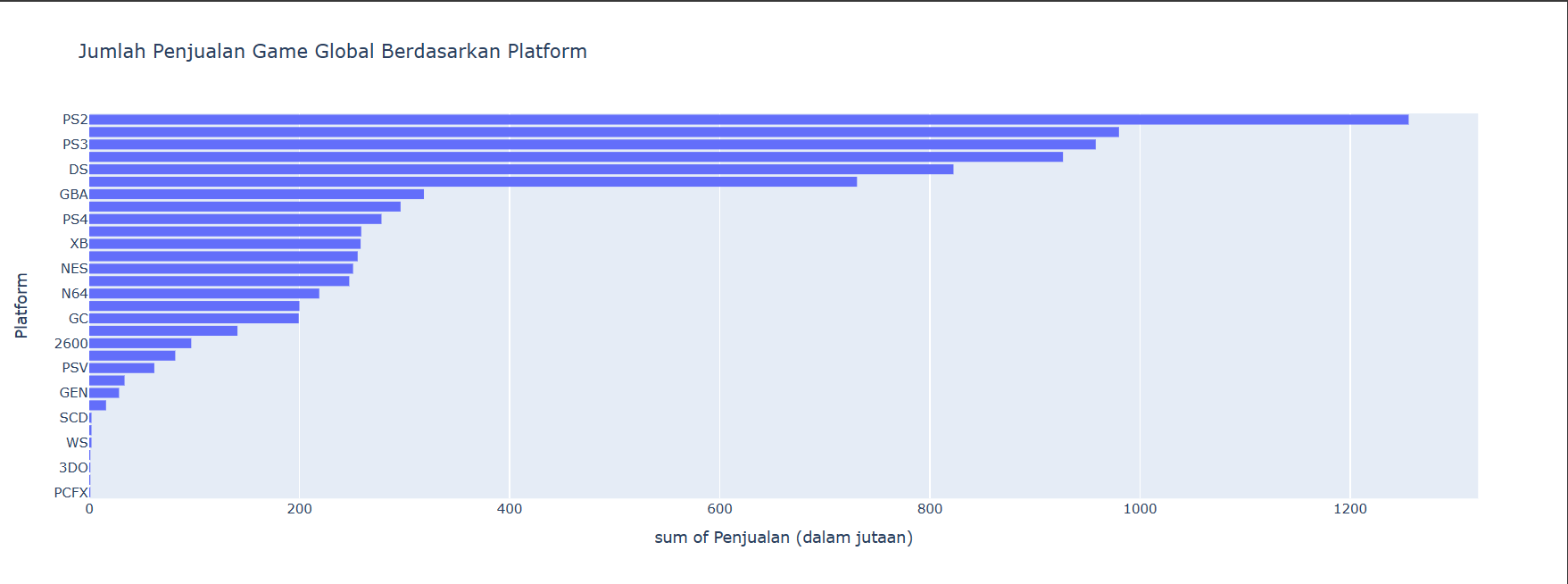
Dari grafik area di atas, dapat kita lihat bahwa distribusi penjualan game setiap tahunnya di masing-masing area mencapai tingkat tertingginya di tahun 2005 sampai dengan tahun 2010.

3.4.4 Jumlah Judul Game Yang Dirlis Dan Penjualannya Berdasarkan Platform

Dengan menggunakan kode:

  Kita bisa mendapatkan visualisasi data perilisan game berdasarkan paltform sebagai berikut:

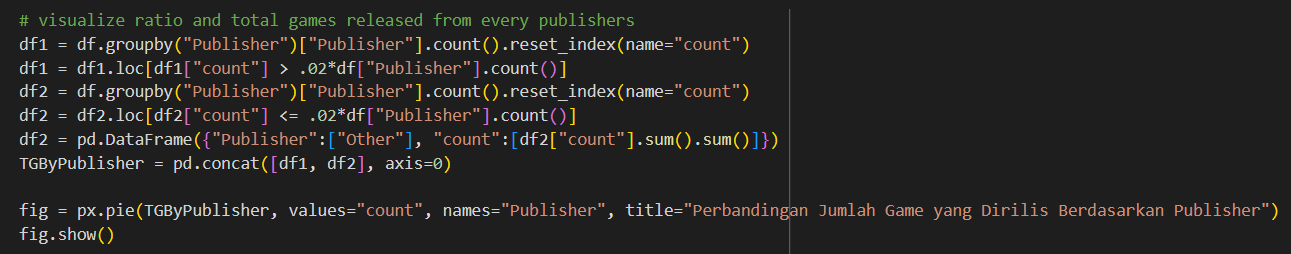
 Kita juga bisa mendapatkan visualisasi data penjualan game berdasarkan platform sebagai berikut:

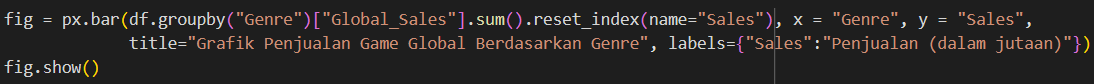


*Insight: Dari grafik pie chart di atas, dapat kita ketahui bahwa DS dan PS2 merupakan platform dengan jumlah game dirilis paling banyak sebesar 13 % sedangkan GG dan PCFX merupakan platform dengan jumlah game dirilis paling sedikit sebesar 0,00602%. Pada grafik bar horizontal di atas, dapat kita ketahui urutan platform dengan penjualan game dari yang terbesar sampai terkecil. PS2 merupakan platform dengan jumlah game dijual paling banyak dan PCFX merupakan platform dengan jumlah game dijual paling sedikit*

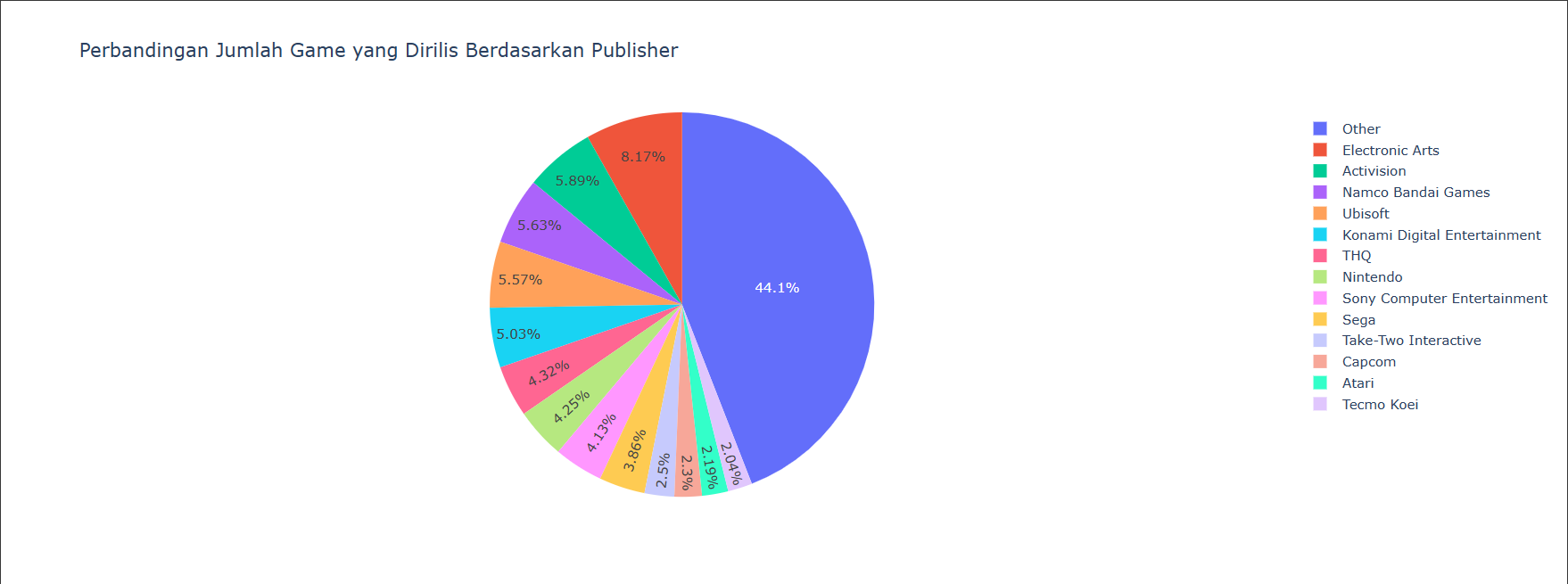
3.4.5 Jumlah Judul Game Yang Dirlis Dan Penjualannya Berdasarkan Publisher

Dengan menggunakan kode



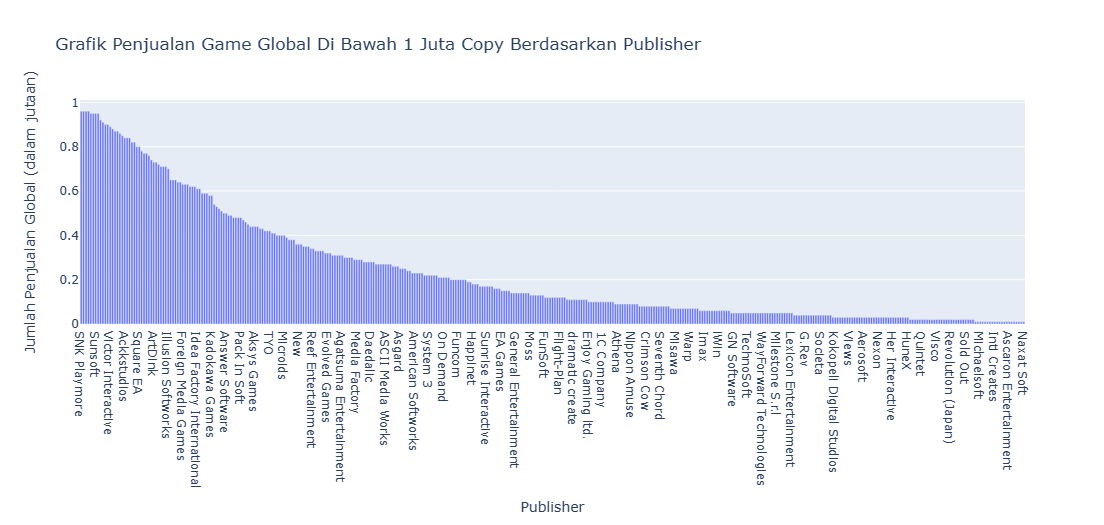


Kita Bisa mendapatkan visualisasi data perilisan game berdasarkan publisher sebagai berikut:

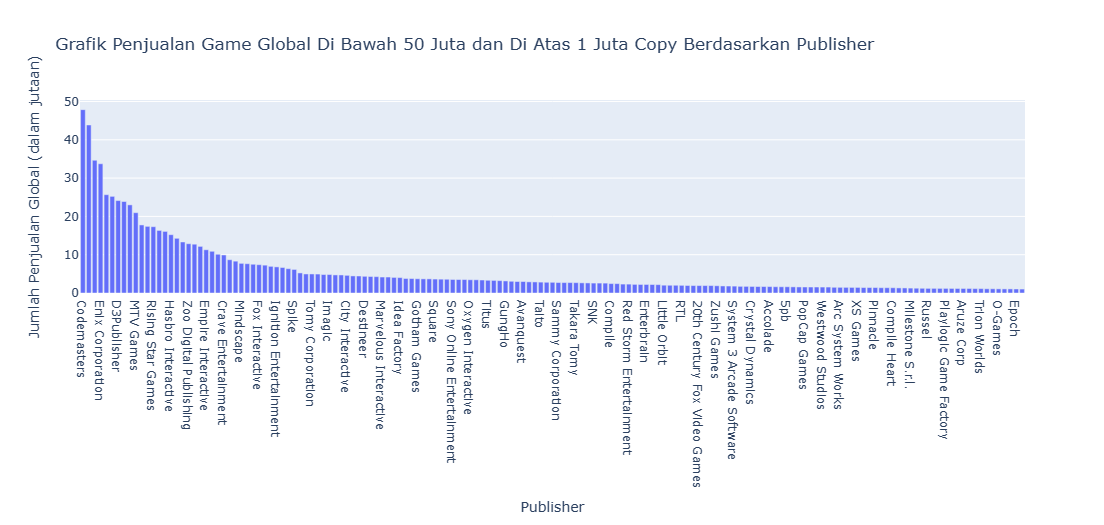
 Kita juga bisa mendapatkan visualisasi data penjualan game berdasarkan publisher sebagai berikut:



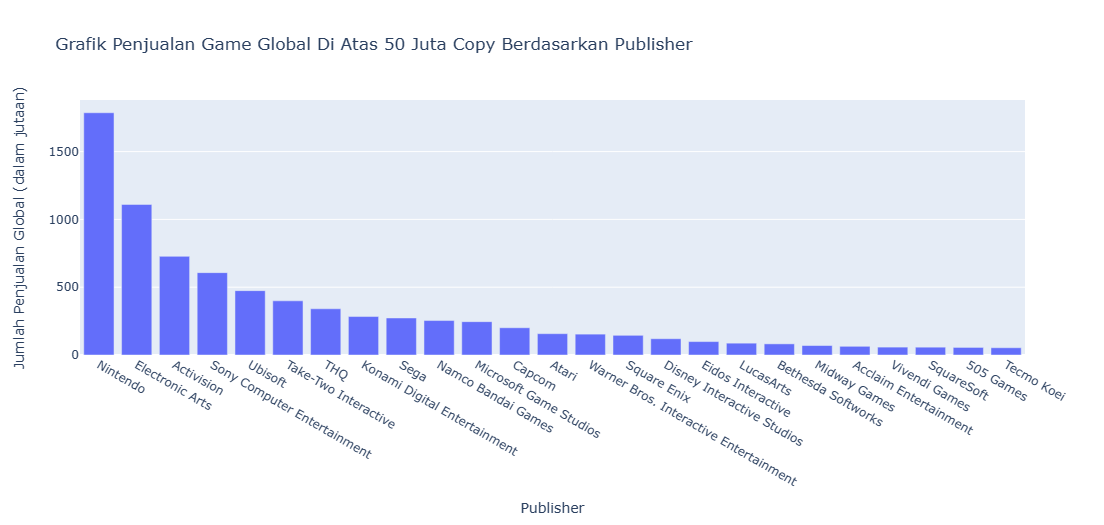
*Insight: Dari grafik pie chart di atas, secara spesifik, dapat kita ketahui bahwa electronic arts merupakan publisher yang merilis game paling banyak, namun jika kita bandingkan dengan semua persentase. Publisher lain-lain diluar data ini memiliki persentase terbesar sebesar 46% dan publisher dengan perilisan game paling sedikit adalah Tecmo Koei. Pada grafik bar chart, kita dapat melihat urutan publisher dengan penjualan game dari yang terbesar hingga terkecil dengan Nintendo sebagai publisher dengan penjualan terbesar sedangkan Commseed sebagai publisher dengan penjualan terkecil*



Data di atas merupakan distribusi penjualan game di bawah satu juta copy yang dibedakan berdasarkan *publisher*-nya. Data di atas diurutkan secara *descending* atau dimulai dari data yang terbesar hingga yang terkecil. Dari data di atas, SNK Playmore menjadi *publisher* dengan penjualan terbanyak di bawah satu juta (sekitar 960 ribu penjualan).



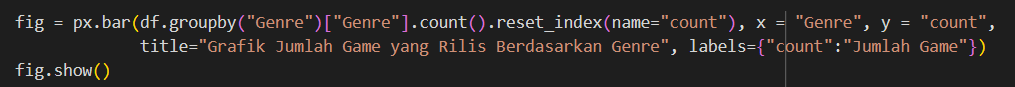
Data di atas merupakan distribusi penjualan game di atas satu juta copy dan di bawah 50 juta copy yang dibedakan berdasarkan *publisher*-nya. Data di atas diurutkan secara *descending* atau dimulai dari data yang terbesar hingga yang terkecil. Dari data di atas, Codemasters menjadi *publisher* dengan penjualan terbanyak di atas satu juta copy dan di bawah 50 juta copy (sekitar 47,87 juta penjualan).

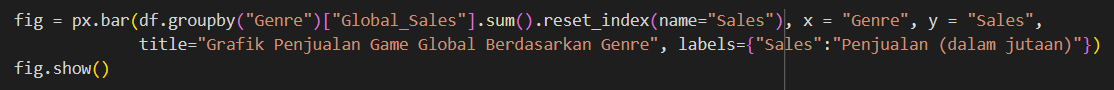


Data di atas merupakan distribusi penjualan game di atas 50 juta copy yang dibedakan berdasarkan *publisher*-nya. Data di atas diurutkan secara *descending* atau dimulai dari data yang terbesar hingga yang terkecil. Dari data di atas, Nintendo menjadi *publisher* dengan penjualan terbanyak di atas 50 juta (sekitar 1,78 milyar penjualan).

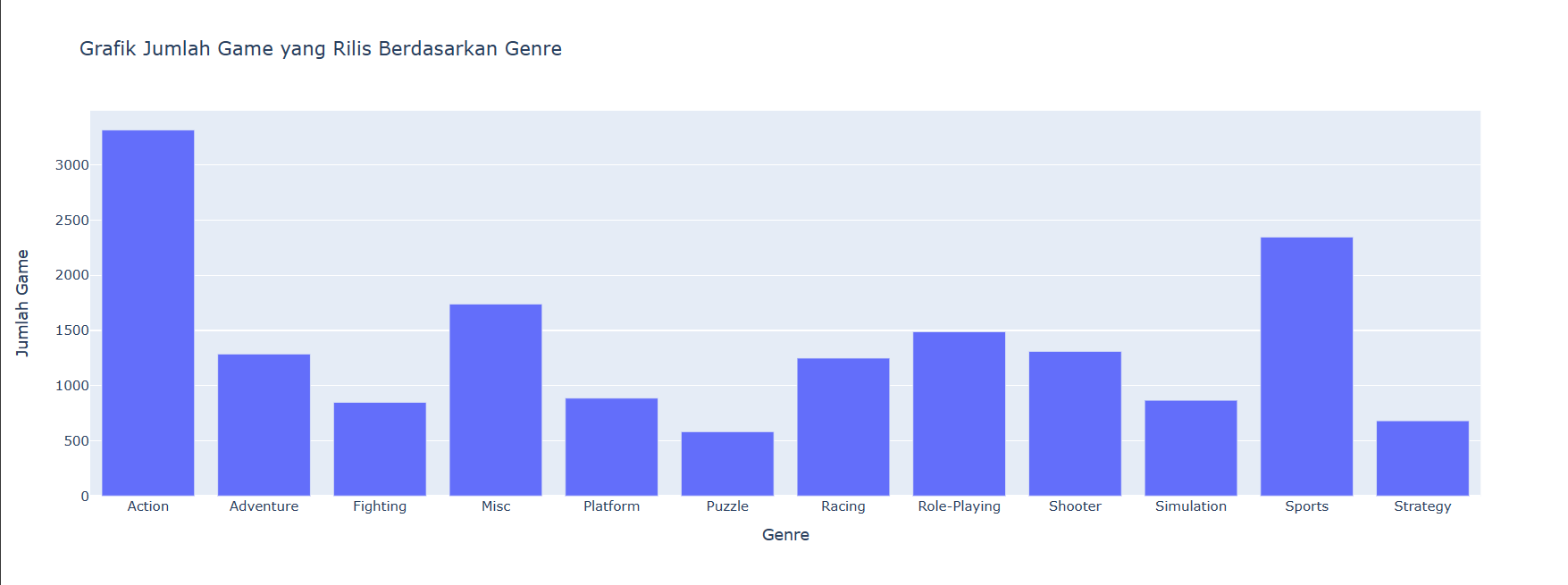
3.4.6 Jumlah Judul Game Yang Dirlis Dan Penjualannya Berdaasarkan Genre Game

Dengan menggunakan kode:

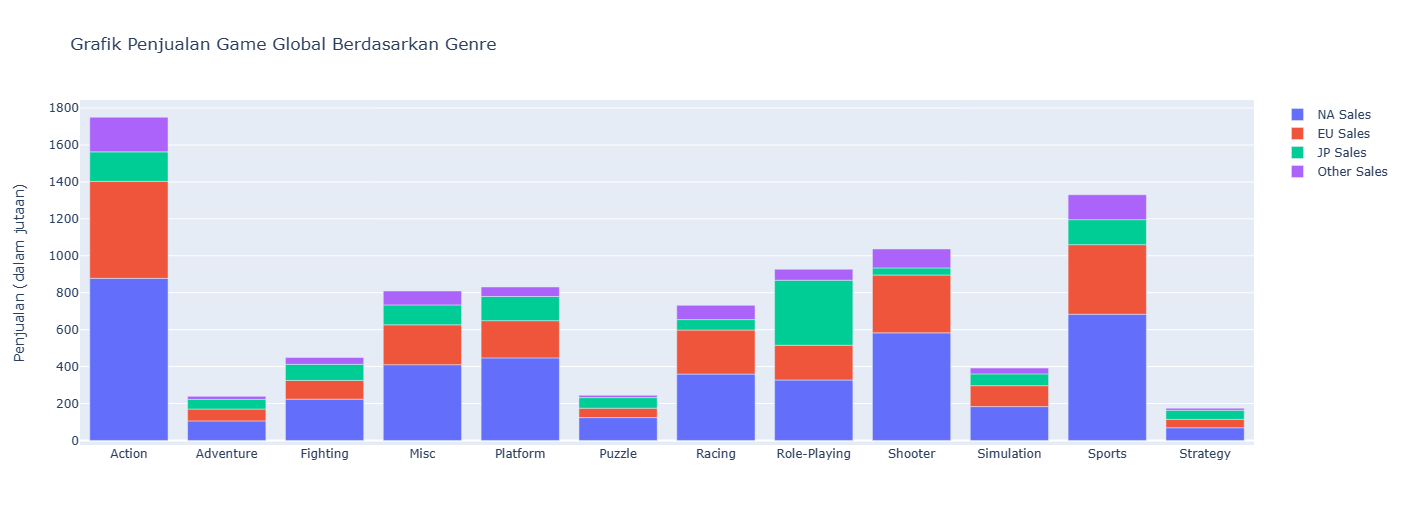




Kita bisa mendapatkan visualisasi data perilisan game berdasarkan genre game sebagai berikut:

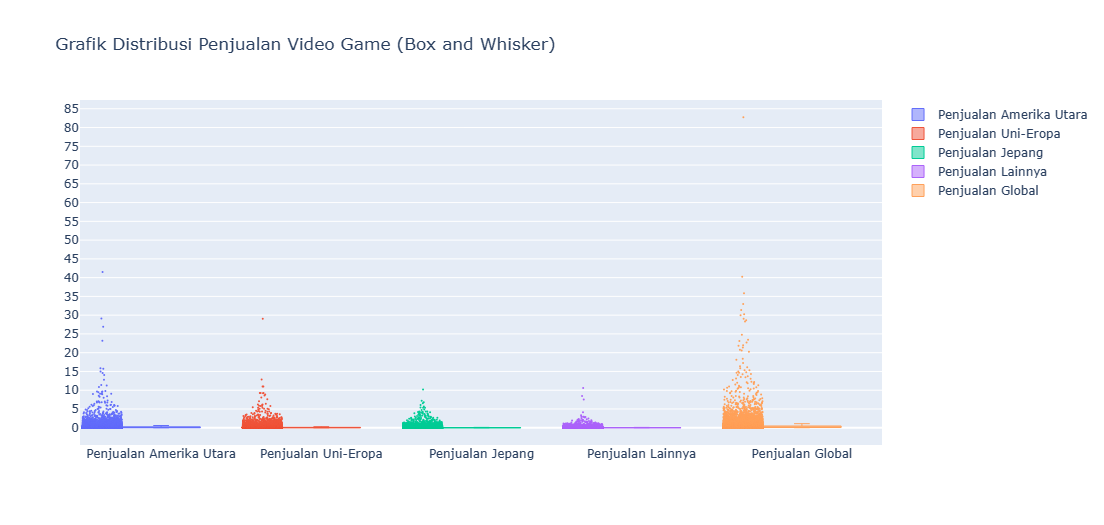


Kita juga bisa mendapatkan visualisasi data penjualan game berdasarkan genre game sebagai berikut:



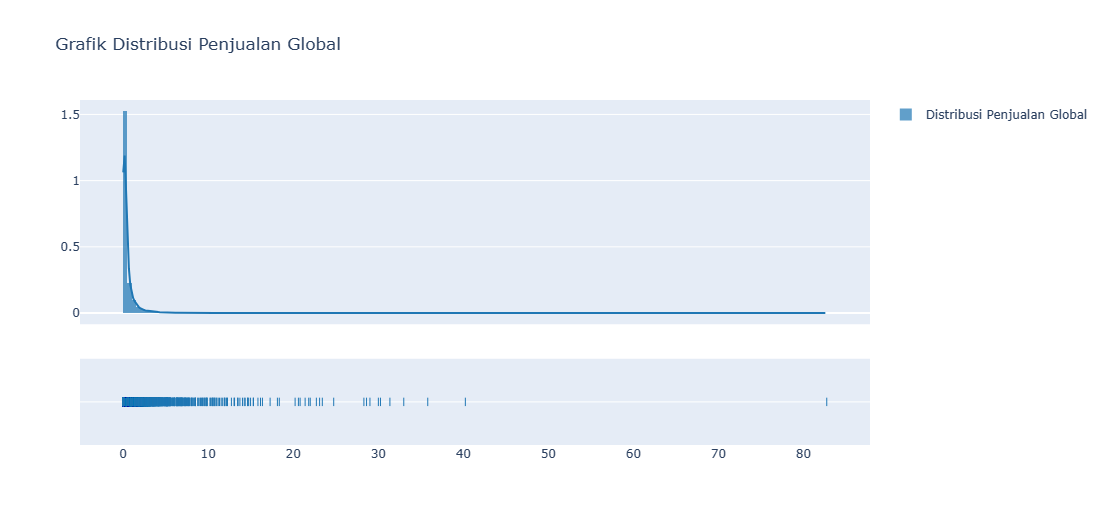
*Insight: Dari grafik bar chart perilisan dapat kita ketahui bahwa game dengan genre action adalah game dengan jumlah game yang dirilis terbanyak sedangkan game dengan genre puzzle adalah game dengan jumlah game yang dirilis paling sedikit. Namun pada grafik bar chart penjualan, dapat kita ketahui bahwa game dengan genre action adalah game dengan jumlah penjualan global terbanyak sedangkan game dengan genre strategy adalah game dengan jumlah penjualan global paling sedikit. Bisa dilihat bahwa terdapat beberapa perbedaan dibandingkan dengan grafik bar chart jumlah game yang dirilis*.

3.4.7 Persebaran Data Penjualan



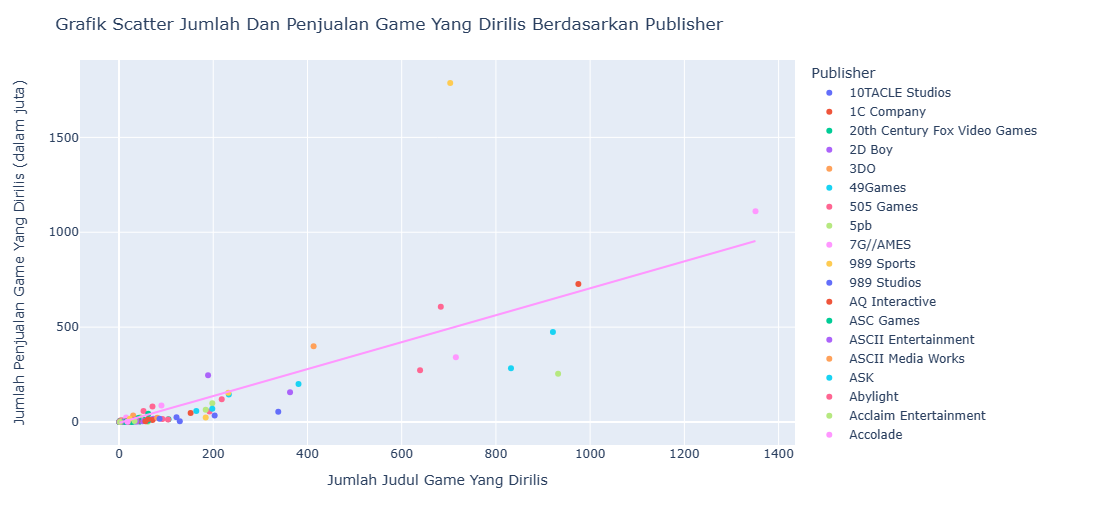
Dari grafik box and whiskers di atas dapat kita lihat bahwa distribusi penjualan global sangat timpang dan didominasi dengan penjualan game global yang memiliki nilai 0 sampai 5 juta *copy.* Grafik diatas adalah grafik interaktif sehingga persebaran data dapat diketahui dengan meng-*hover* bagian box pada grafik.

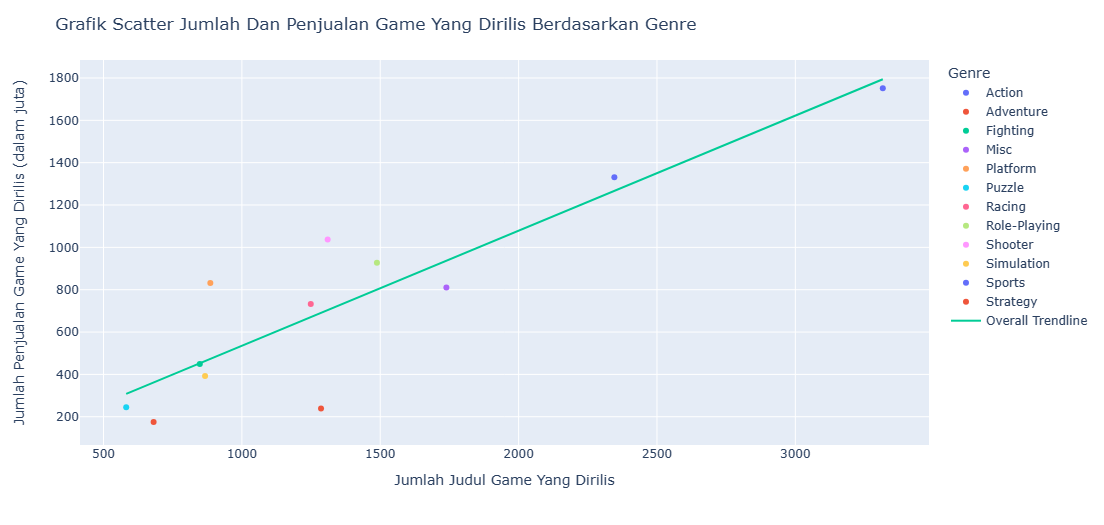
Data juga bisa ditampilkan sebagai distribusi keseluruhan sebagai grafik di bawah:

..

Grafik diatas adalah plot distribusi penjualan global secara keseluruhan. Dapat diketahui bahwa distribusi penjualan game secara keseluruhan (tidak dipisahkan dengan genre atau publisher) padat di angka penjualan 0—10 juta *copy*.

3.5 Korelasi Data



Dari grafik scatter jumlah dan penjualan game yang dirilis berdasarkan publisher di atas, dapat diketahui bahwa jumlah judul game cenderung berbanding lurus dengan jumlah penjuala game yang dirilis dengan nilai koefisien determinasi (R^2) bernilai 0,639834.

Dari grafik scatter jumlah dan penjualan game yang dirilis berdasarkan genre di atas, dapat diketahui lebih jelas bahwa jumlah judul game cenderung berbanding lurus dengan jumlah penjuala game yang dirilis dengan nilai koefisien determinasi (R^2) bernilai 0,790709.

Dengan menggunakan kode:

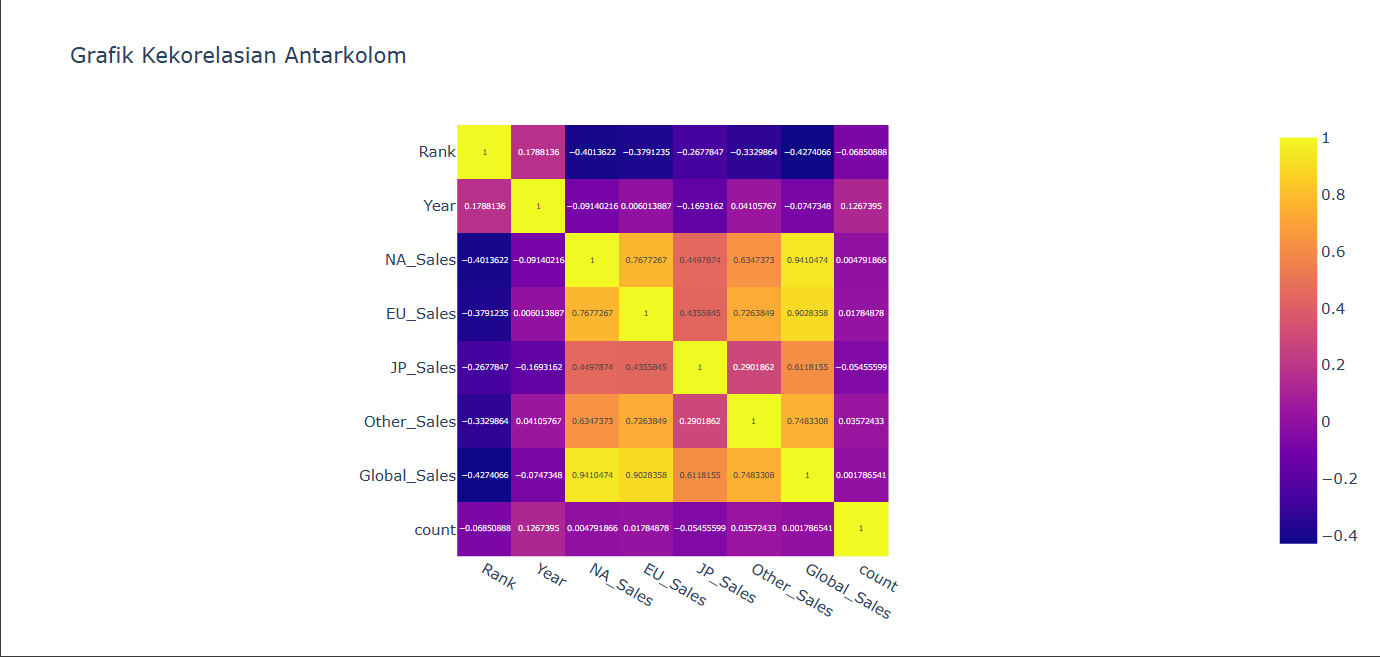


Kita bisa melihat analisis dari korelasi data sebagai berikut:

 Lalu, dengan menggunakan kode:



Kita bisa mendapatkan visualisasi dari analisis korelasi data sebagai berikut

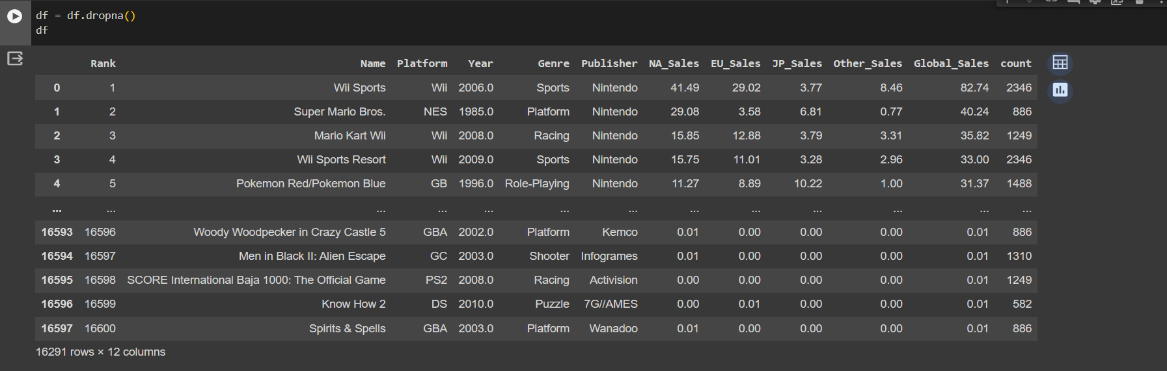


*Insight: Dengan menggunakan heatmap, kita dapat melihat kekorelasian antarkolom yang dibantu oleh penggunaan warna dan angka kekorelasian. Dapat kita simpulkan bahwa data penjualan berkorelasi dengan penjualan lainnya. Sedangkan peringkat dan tahun cenderung tidak berkorelasi dengan data lainnya*

3.6 Cleansing Data

Terdapat data yang kosong pada kolom *Year* dan*Publisher* masing-masing sebanyak 271 dan 58. Kolom tersebut keduanya merupakan data kategorikal nominal. Untuk menangani data kosong tersebut, terdapat dua solusi yang bisa dilakukan:

1. Drop data yang kosong



Data yang didrop akan dihapus dari dataframe sehingga tidak akan memengaruhi analisis data dan visualisasi data. Ini adalah solusi yang paling mudah, tetapi solusi ini harus mengorbankan data yang kosong tersebut.

1. Menggunakan API dan scrapping data di web

Data yang kosong diisi secara otomatis dengan data yang didapat dengan scrapping data di website tertentu. Ini adalah solusi yang paling aman karena data yang kosong dapat terisi sehingga analisis dan visualisasi data lebih akurat, tetapi implementasinya tergolong sulit.

Tidak ada data yang kotor atau kosong pada kolom yang lain termasuk kolom dengan data kuantitatif. Namun, apabila terdapat data yang kosong pada kolom dengan data kuantitatif, data bisa diinterpolasi atau diisi dengan mean dari data kolom tersebut.

BAB IV

SIMPULAN DAN SARAN

4.1 Simpulan

Setelah kami melakukan analisis pada data “Penjualan dan Tren Video Game Sepanjang Masa”, kami mendapatkan beberapa kesimpulan, diantaranya

1. Dengan menggunakan bahasa pemrograman python dan library pandas, melakukan analisis data, terutama data mentah, dapat menjadi lebih mudah.

2. Manfaat dari melakukan analisis data adalah mempermudah pengamatan dan pembacaan bagi semua orang. Beberapa contoh hasil yang kami dapatkan dari analisis data yang kami lakukan seperti berikut

2.1 Pada data yang kami analisis, game dengan genre *action* dan *sports* mendominasi di hampir semua platform.

2.2 Rentang tahun dengan jumlah game rilis tertinggi adalah 2008 – 2010 sedangkan rentang tahun dengan jumlah game rilis terendah adalah 1983 – 1990.

2.3 Platform dengan jumlah game rilis tertinggi adalah *DS* dan *PS2* sedangkan platform dengan jumlah game rilis terendah adalah *GG* dan *PCFX.*

2.4 Platform dengan jumlah game terjual paling tinggi adalah *PS2* sedangkan platform dengan jumlah game terjual paling rendah adalah *PCFX.*

2.5. Publisher yang merilis game paling banyak adalah Electronic Artssedangkan publisher yang merilis game paling sedikit adalah Locus dan kawan-kawan*.*

2.6 Publisher dengan jumlah game terjual paling tinggi adalah Electronic Arts sedangkan publisher dengan jumlah game terjual paling sedikit adalah Naxat Soft dan kawan-kawan.

4.2 Saran

Studi ini masih dapat dikembangkan dengan baik. Beberapa saran yang dapat dilakukan adalah:

1. Mengambil data lebih banyak. Dengan mengambil data lebih banyak, analisi dan visualisasi data akan menjadi lebih akurat.
2. Menggunakan teknik data cleansing yang lebih mutakhir seperti mengisi data kosong dengan menggunakan API yang mengisi data dengan scrapping melalui website tertentu.
3. Menggunakan machine learning untuk menganalisis data lebih dalam dan mutakhir seperti memprediksi penjualan *video game* beberapa tahun ke depan dan lainnya. Penggunaan machine learning juga bisa digunakan untuk mengisi data yang kosong pada data kuantitatif.
4. Membuat aplikasi data yang lebih interaktif. Terdapat *framework* *data apps* yang dapat digunakan dengan python yaitu Dash. Dengan dash kita bisa membuat aplikasi GUI yang sangat baik dan interaktif untuk analisis dan visualisasi data sehingga *user interface* dan *user experience* yang dirasakan pengguna dapat ditingkatkan.

REFERENSI

Dharma, Rifki. 2022. “Data Preprocessing: Pengertian, Manfaat, dan Tahapan Kerjanya”. <https://accurate.id/teknologi/data-preprocessing/>

Kurnia, Firdilla. 2023. “Analisis Data: Definisi, Jenis, Model, Sampai Prosedurnya”. <https://dailysocial.id/post/analisis-data>

Perdana, Arkan. 2022. “Visualisasi Data: Pengertian, Fungsi, dan Tipe-tipenya”. <https://glints.com/id/lowongan/data-visualization-adalah/>

Renaldi, Taufieq. Ferri, Rendika. 2022. “Mengenal Apa itu Video Game dan Sejarah Awalnya”. <https://www.kompas.com/tren/read/2022/06/03/070000165/mengenal-apa-itu-video-game-dan-sejarah-awalnya?page=all#google_vignette>

Jain, Depak. 2023. “Data Processing in Data Mining”. <https://www.geeksforgeeks.org/data-preprocessing-in-data-mining/>

Kelley, Karin. 2023. “What is Data Analysis?: Process, Types, Methods, and Techniques”. <https://www.simplilearn.com/data-analysis-methods-process-types-article>

Lowden, Olivia. 2023. “The video games industry is still growing - but why?”. <https://blog.bccresearch.com/video-games-industry-is-still-growing-but-why>

LAMPIRAN

Tabel pembagian tugas kelompok:

|  |  |  |  |
| --- | --- | --- | --- |
| Tabel Pembagian Kerja Kelompok | | | |
| No | Nama | NIM | Pekerjaan |
| 1 | Yayat Nurhidayat | 16523080 | Melakukan prapemrosesan data menggunakan pandas.  Menganalisis dan memvisualisasikan data dengan pandas dan plotly.  Menginterpretasikan visualisasi data yang telah dibuat sehingga terdapat wawasan yang bisa diketahui pembaca.  Menyusun laporan pada bagian pembahasan visualisasi data (sebagian) dan saran. |
| 2 | Nafhan Hadiyan Shafwatudin | 16523213 | Menyusun laporan bagian deskripsi data, karakteristik data, dan simpulan  Menambahkan sedikit kode untuk prapemrosesan data  Menyunting laporan secara keseluruhan |
| 3 | Zaka Hanif Nabalah | 19623010 | Mencari data set  Membuat sebagian visualisasi data  Editor video presentasi |
| 4 | Atharizza Muhammad Athaya | 19623143 | Membuat Powerpoint  Dokumentasi  Formatting laporan |
| 5 | Ferro Arka Berlian | 19623213 | Mencari teori dasar untuk data analisis  Menyusun sebagian laporan pada bagian pendahuluan, teori, dan pembahasan |

Link google colab analisis dan visualisasi data Video Game Sales: <https://colab.research.google.com/drive/1Nb4L9bshsTD5DclRa9-4lU4rPLshbIKe?usp=sharing>