**Tugas Besar 2**

**Pengenalan Komputasi**

**Data Analisis dan Visualisasi**

**November 2019**

**Kelompok 4**

|  |  |
| --- | --- |
| **Nama** | **NIM** |
| Frederic Ronaldi | 16519422 |
| Gavriel Benny | 16519032 |
| Aulia Adila | 16519152 |
| Abu Hanif Muhammad Syarubani | 16519282 |

**Tema: Penjualan Video Games Beserta Ratingnya**

**Dataset yang digunakan:**

1. Penjualan video games

Data diambil dari : <https://www.kaggle.com/ashaheedq/video-games-sales-2019/data#vgsales-12-4-2019.csv>

Format : CSV

1. Penilaian video games dari Metacritic dan pemain

Data diambil dari : <https://www.kaggle.com/destring/metacritic-reviewed-games-since-2000>

Format : CSV

**Data yang akan dianalisis dan divisualisasi:**

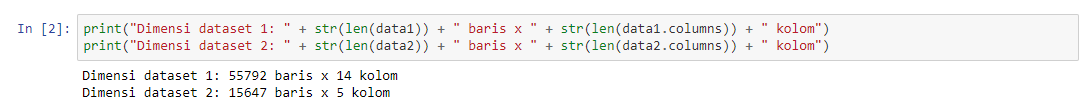
1. Game-game dengan penjualan terbanyak
2. Platform (seperti PS4, Wii, Nintendo DS, dsb.) dengan produksi game terbanyak
3. Perkembangan produksi game pada masing-masing platform sejalan dengan waktu
4. Publisher dengan game terbanyak
5. Developer dengan game terbanyak
6. Perkembangan penjualan game setiap tahun
7. Genre dengan produksi game terbanyak
8. Genre yang paling disuka dalam masing-masing tahun produksi
9. Banyaknya game sesuai kategori umur (ESRB Rating)
10. Relasi antara penjualan video games dengan penilaian video games dari pemain dan Metacritic
11. Relasi antara penilaian dari pemain dan Metacritic.

Pendahuluan

**Import dan membaca data:**



**Mencari dimensi dataset:**

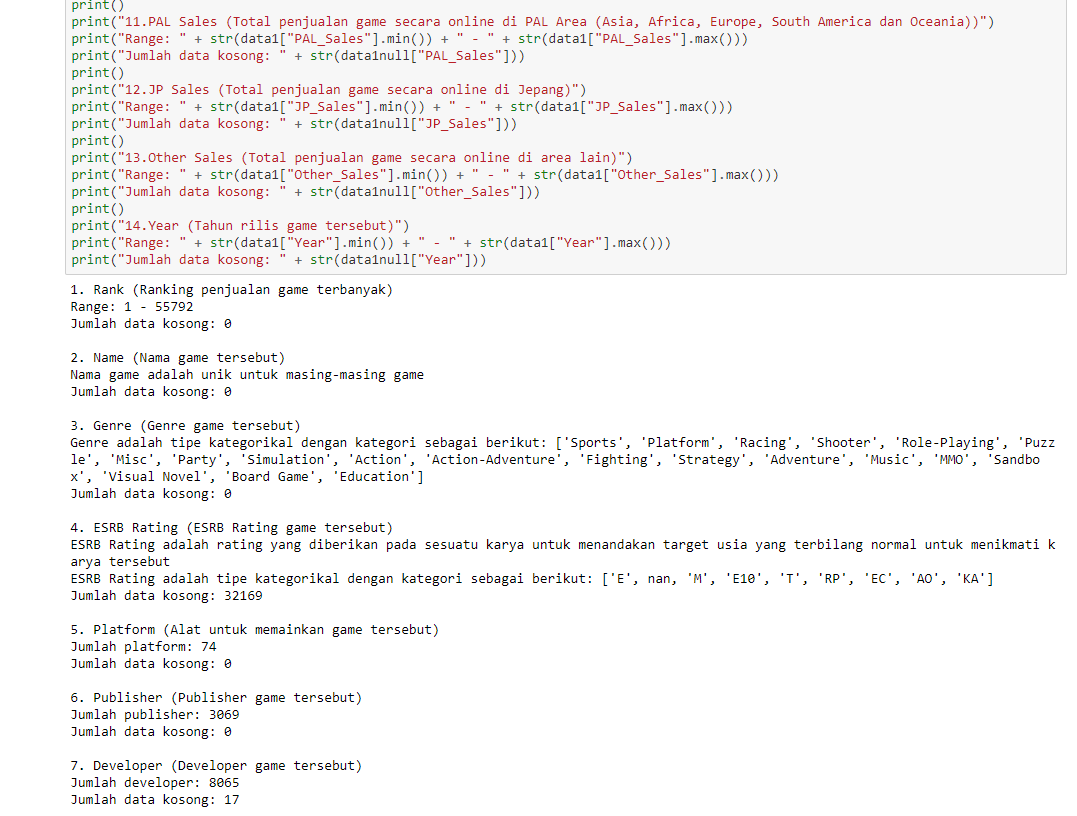
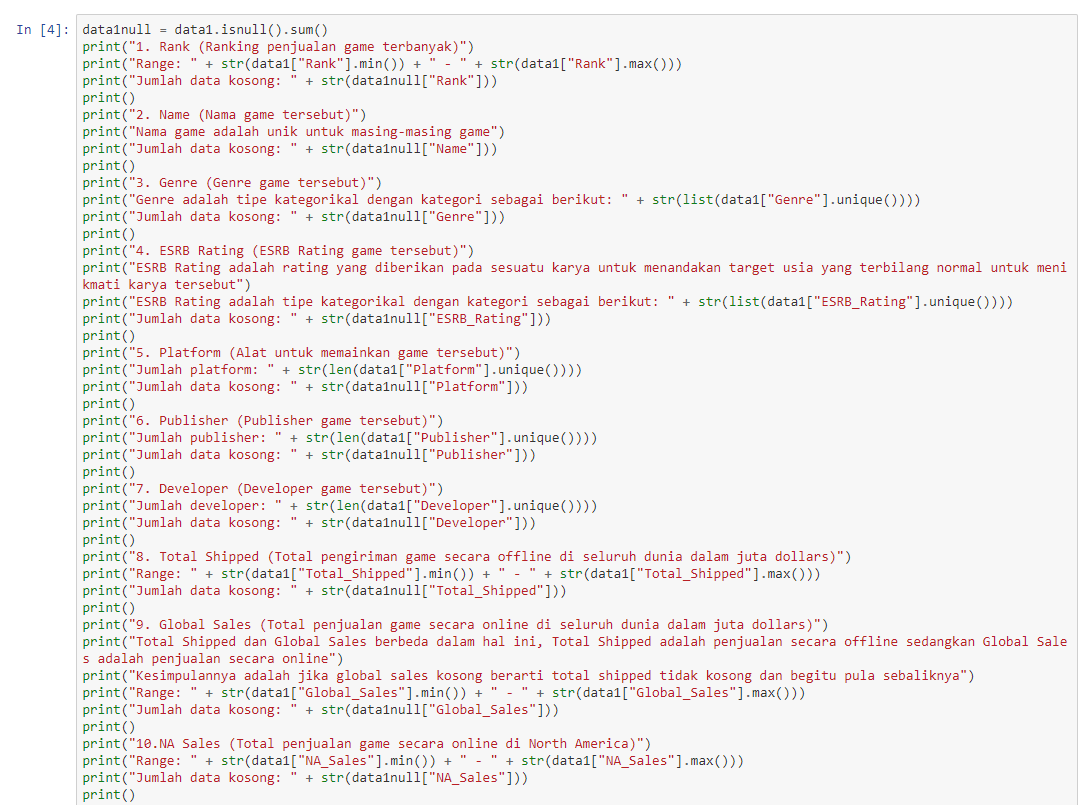


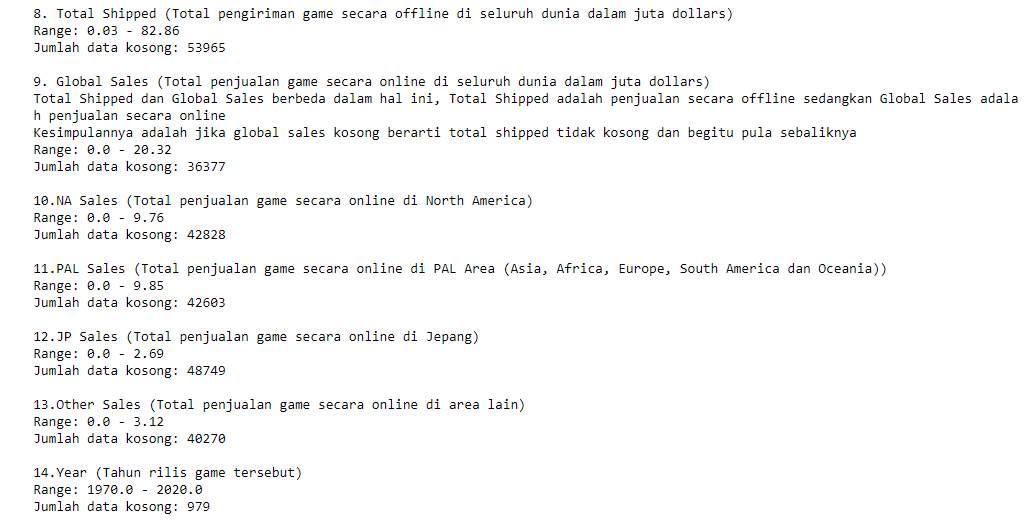
**Kolom yang ada dalam dataset serta jenisnya:**

****

**Kesimpulan:** *Terlihat bahwa jenis int64 adalah jenis integer, object dapat berupa string atau tipe kategorical dan float adalah angka tidak bulat.*

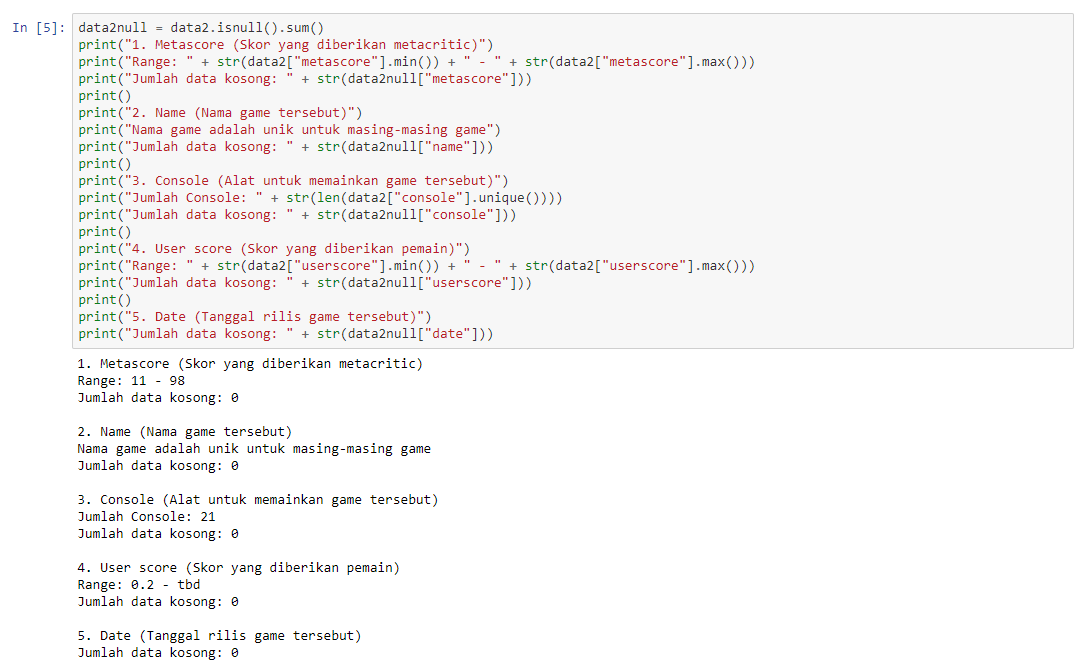
**Karakteristik data untuk tiap kolom:**

**Dataset 1: **

****

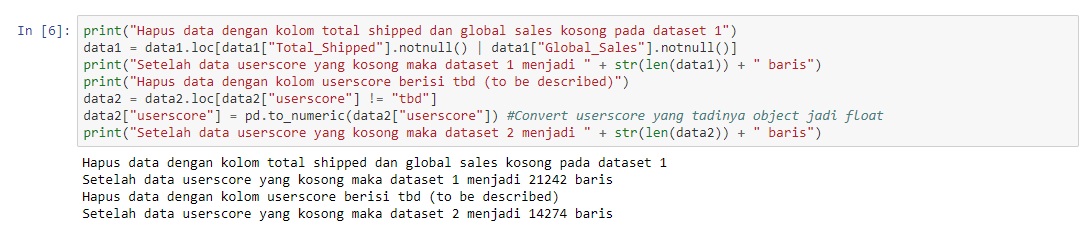
**Kesimpulan:** *Terlihat bahwa Global Sales dan Total Shipped banyak yang kosong datanya (salah satu dari kolom tersebut pasti ada yang kosong dalam satu baris) sehingga perlu dilakukan data cleaning dimana data dengan kedua kolom tersebut kosong harus dihapus.*

***Dataset 2:***

**

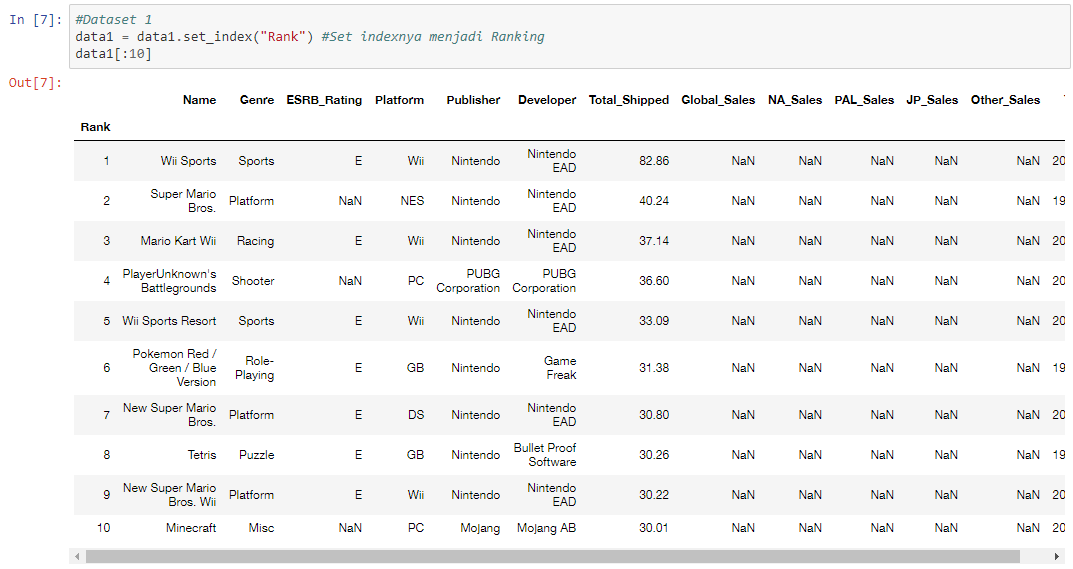
**Kesimpulan*:*** *Terlihat bahwa user score ada yang tbd (to be described) atau belum dihitung sehingga perlu dilakukan data cleaning*

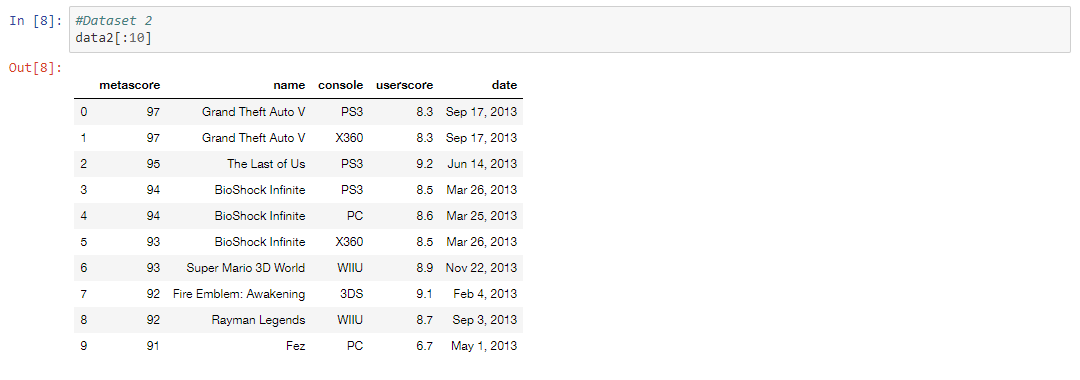
Data Cleansing



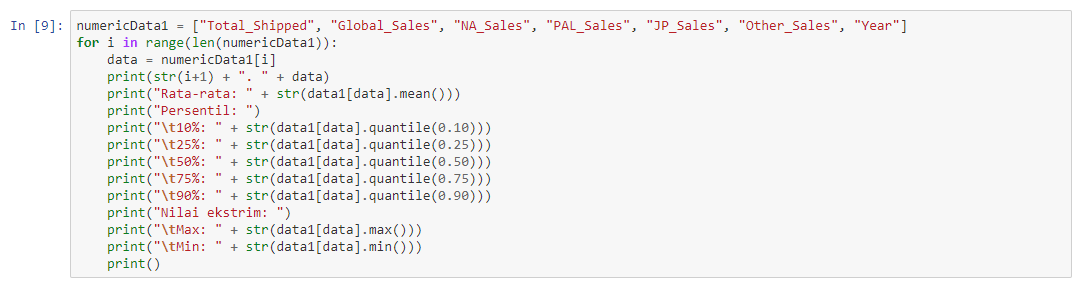
Statistics

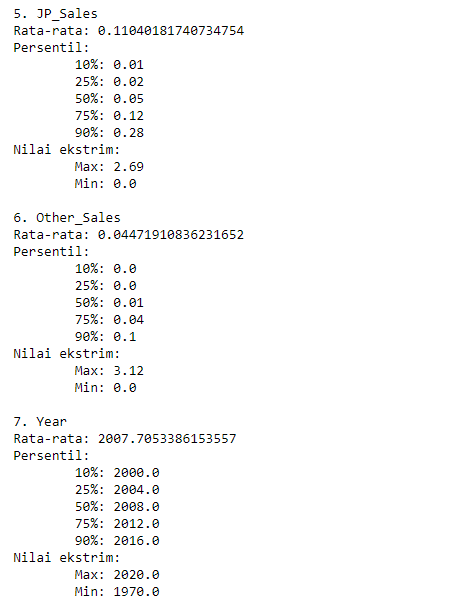
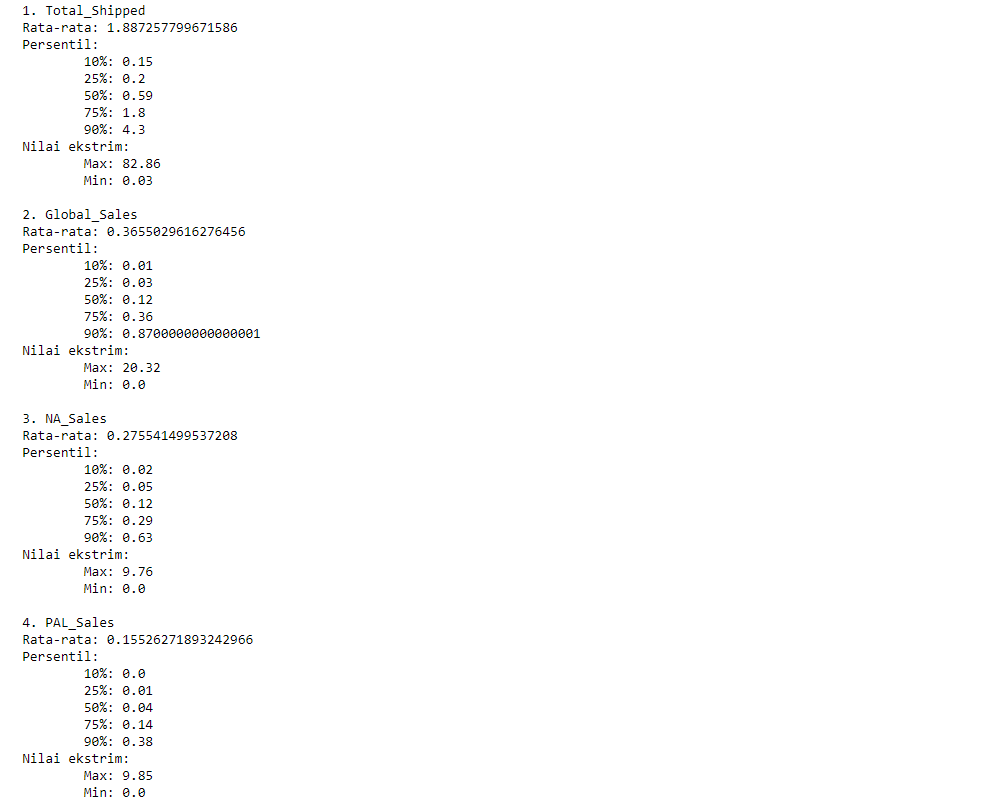
**Sampel Data**

****

****

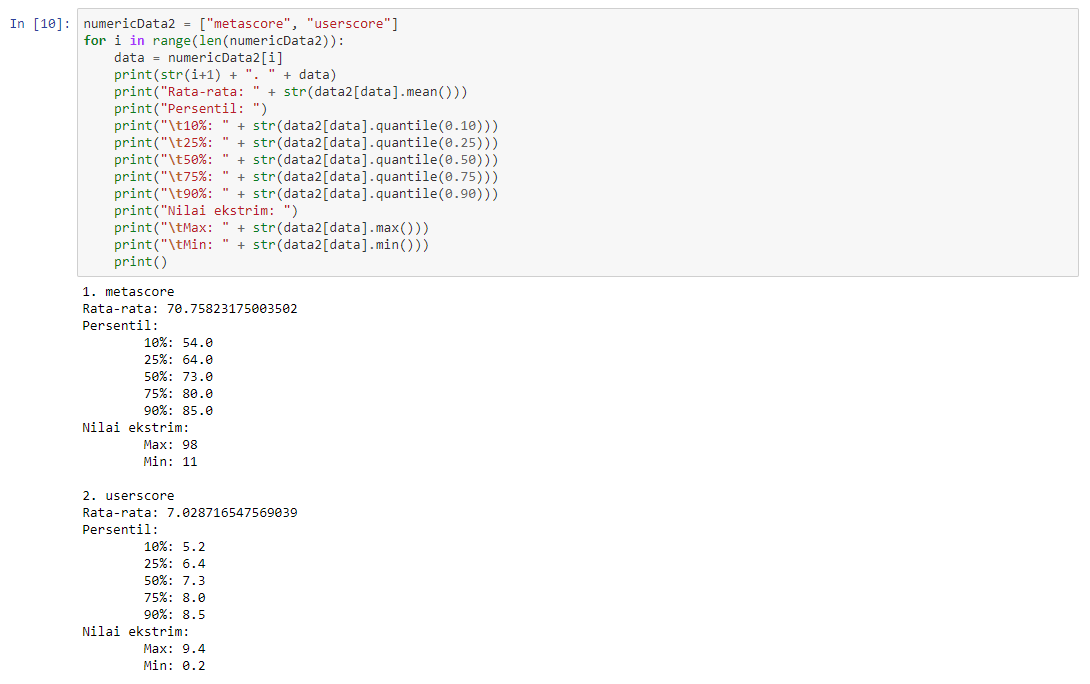
**Statistik pada dataset 1**

****

****

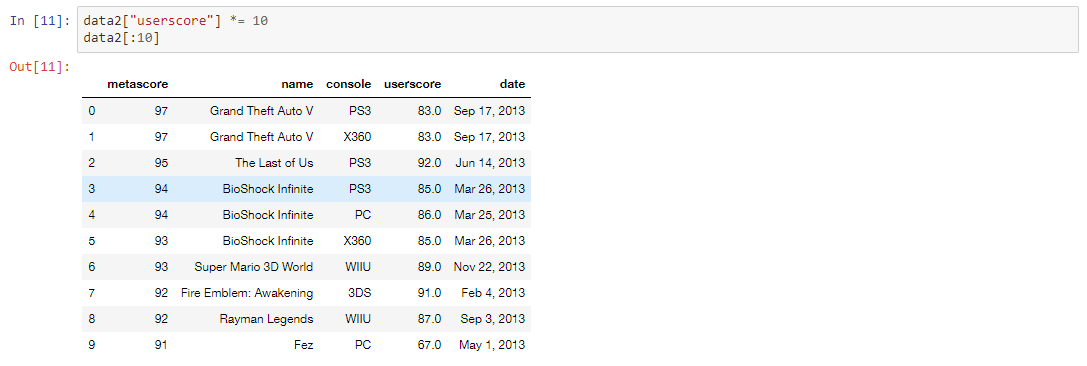
**Kesimpulan:** *Rata-rata penjualan game menghasilkan uang sekitar 1.8 juta dollar dimana uang sejumlah itu tergolong cukup banyak. Terlihat bahwa walaupun Jepang memiliki rata-rata yang lebih kecil daripada PAL area (gabungan beberapa negara dan kontingen) namun Jepang memiliki rata-rata yang mendekati PAL area dimana hal itu menunjukkan bahwa Jepang memiliki ketertarikan yang lebih pada game, namun tetap lebih sedikit daripada Amerika Utara. Game yang tercatat paling tua di dataset ini adalah 1970, dan ada game yang akan dirilis 2020 (belum rilis)*

**Statistik pada dataset 2**

****

**Kesimpulan:** *Terlihat bahwa userscore skornya ber-range dari 0 hingga 10 dan metascore ber-range dari 0 hingga 100. Sehingga kita dapat mengnormalisasikan data dengan mengkali userscore dengan 10.*

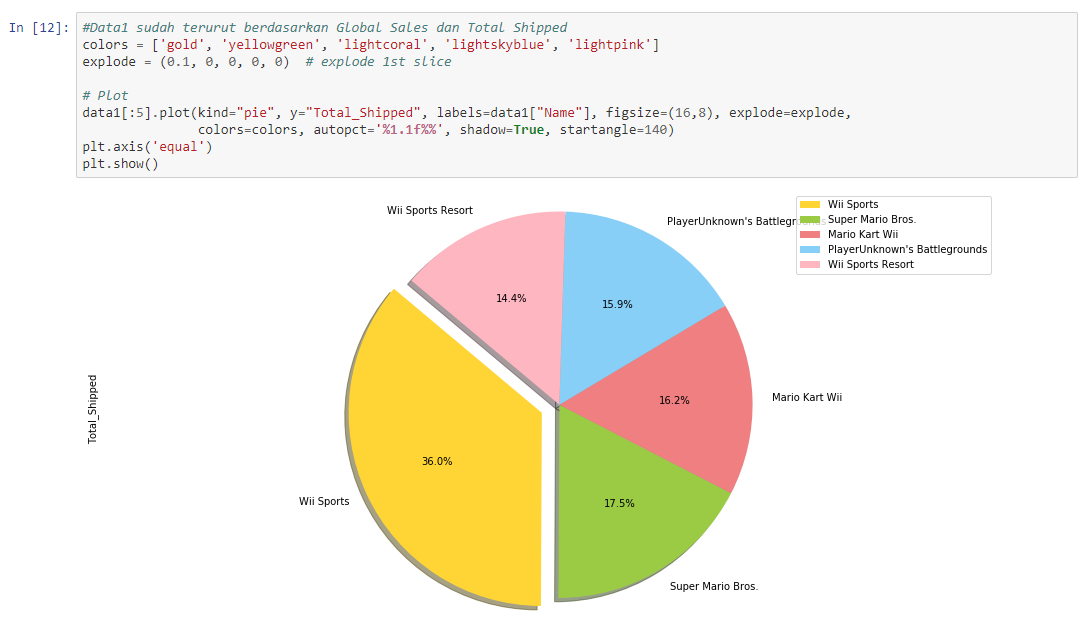
**Normalisasi userscore pada dataset 2**

****

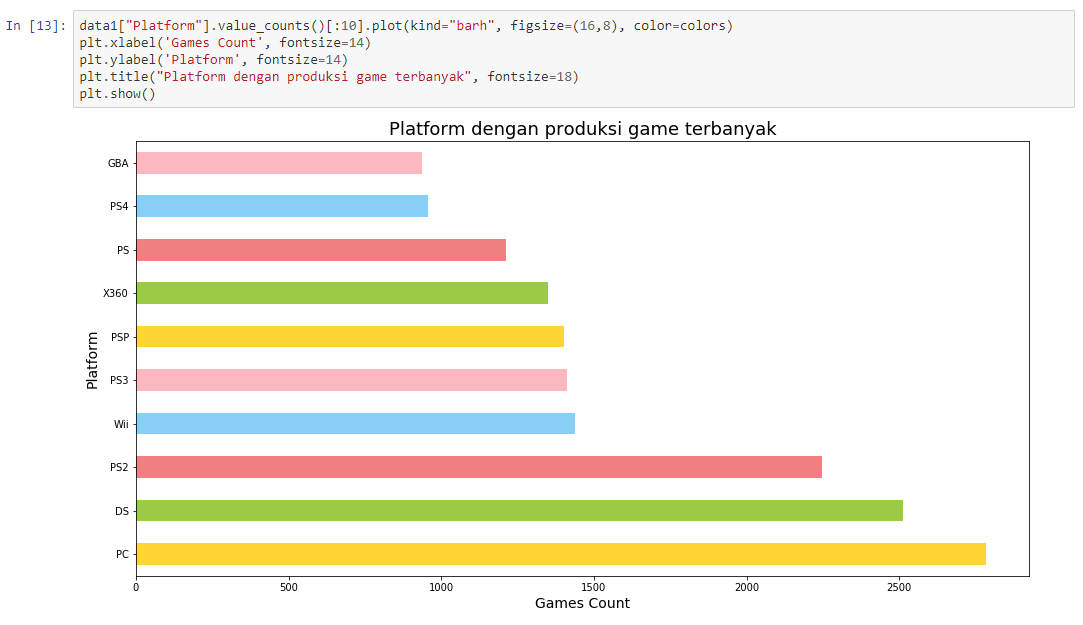
**Kesimpulan:** *Userscore sudah sebanding dengan metascore*

Visualisasi

1. **Game-game dengan penjualan terbanyak**

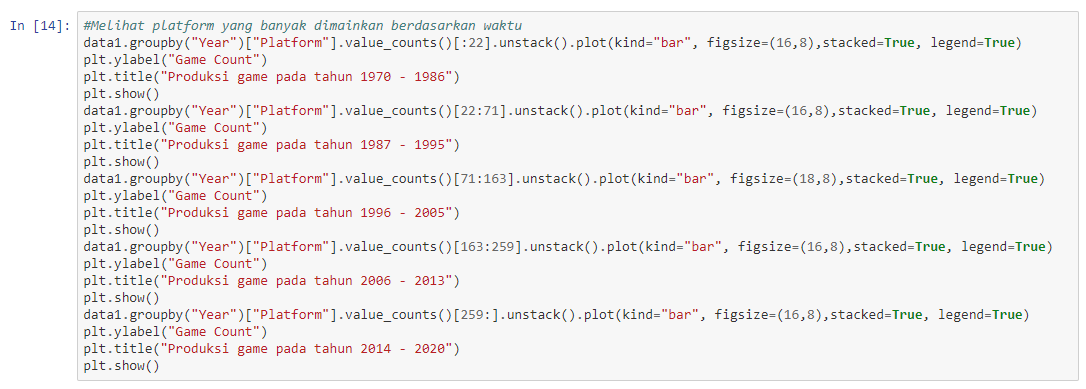
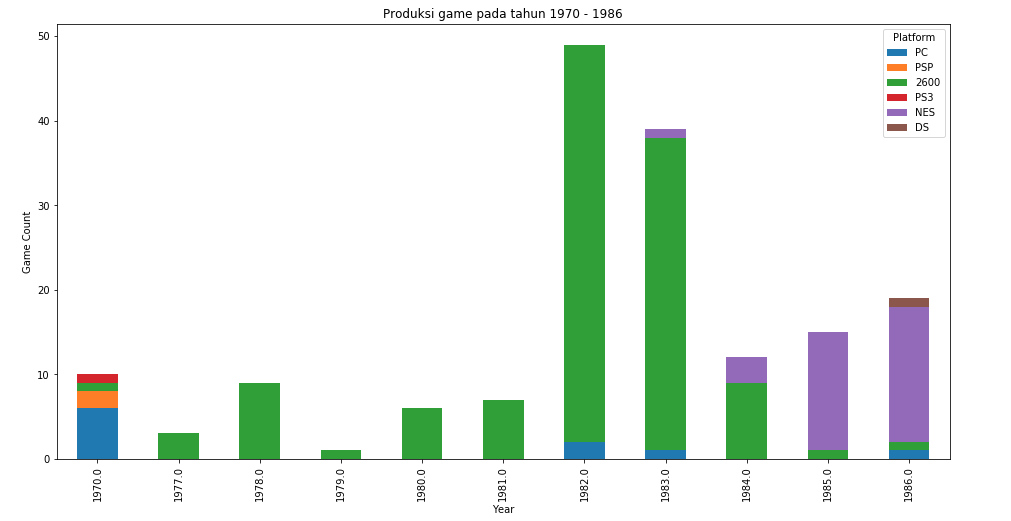
****

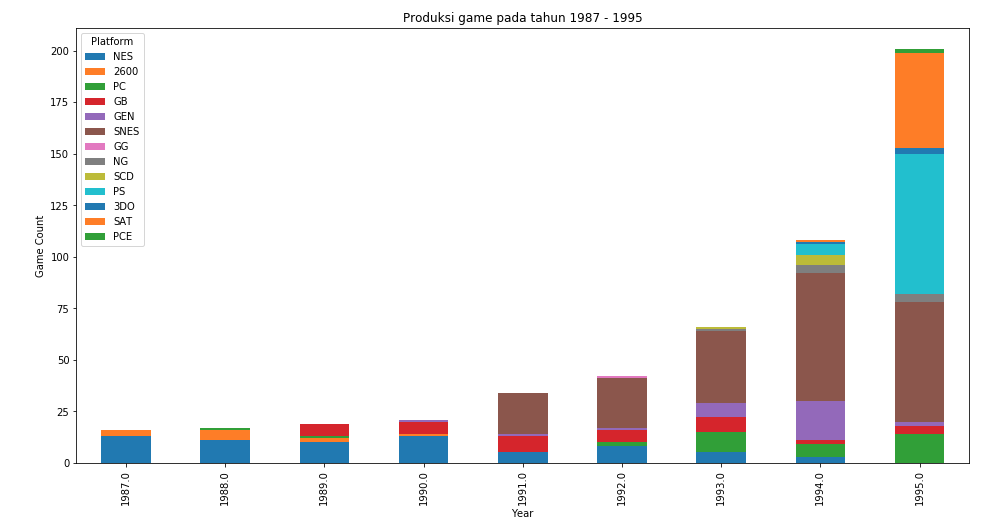
**Kesimpulan:** *Terlihat bahwa Wii Sports mendapatkan penjualan terbanyak diikuti oleh Super Mario Bros, Mario Kart Wii, PlayerUnknown's Battleground dan Wii Sports Resort.*

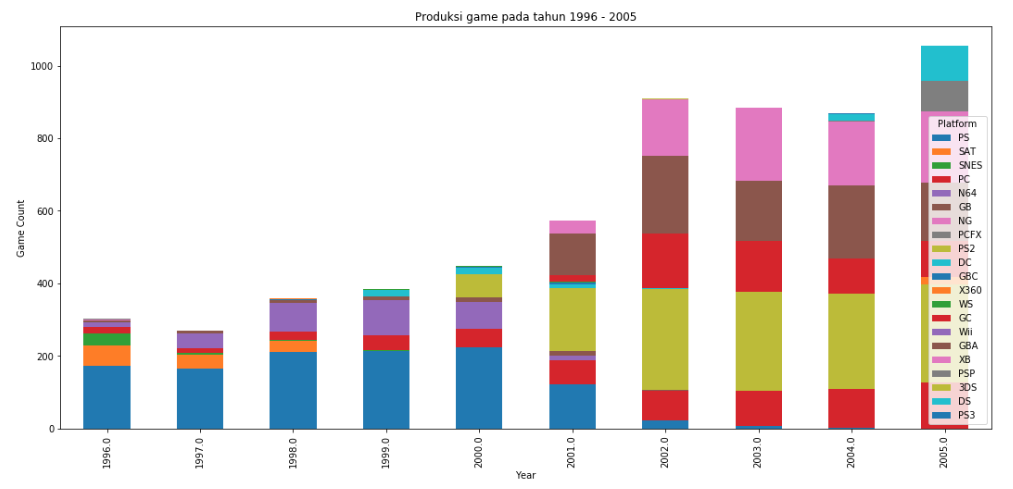
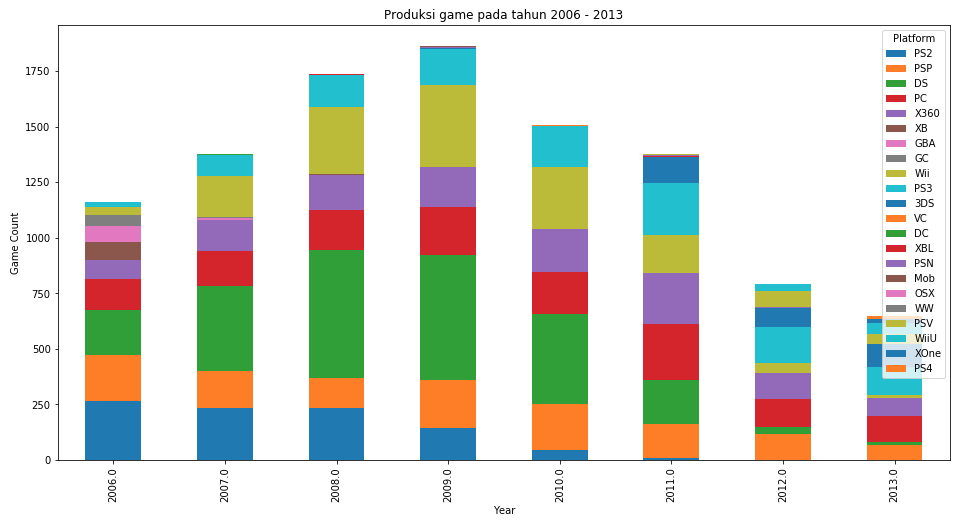
1. **Platform dengan produksi game terbanyak**

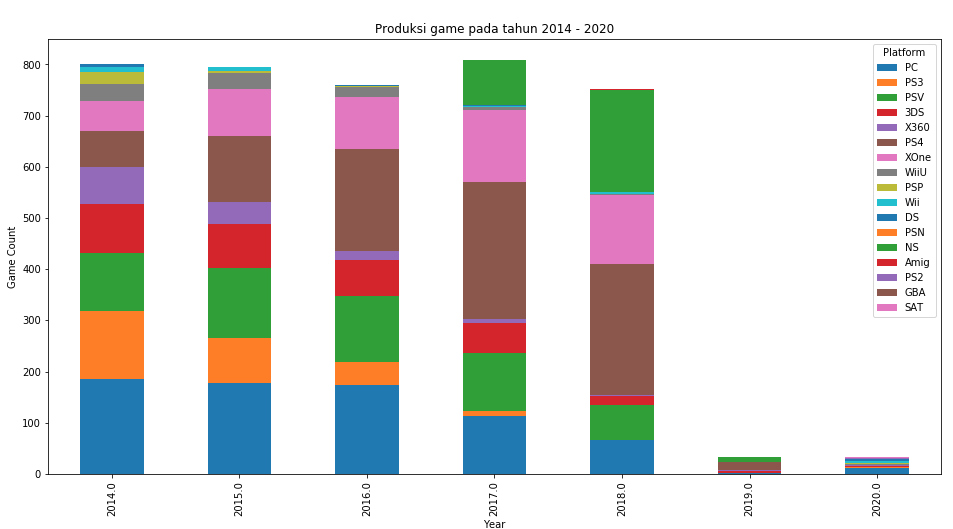
**Kesimpulan:** *Terlihat bahwa walaupun game-game dengan penjualan terbanyak berasal dari Wii, namun jumlah game terbanyak tetap dimenangkan oleh game PC*

1. **Perkembangan produksi game pada masing-masing platform sejalan dengan waktu**

****

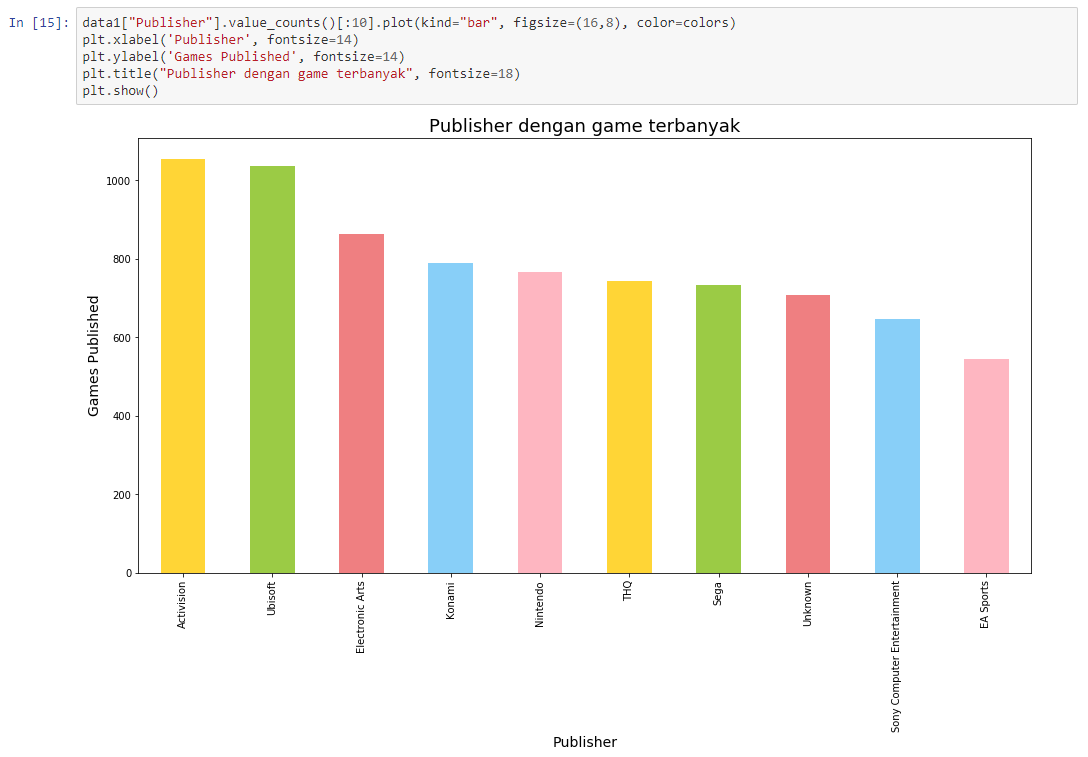
****

****

****

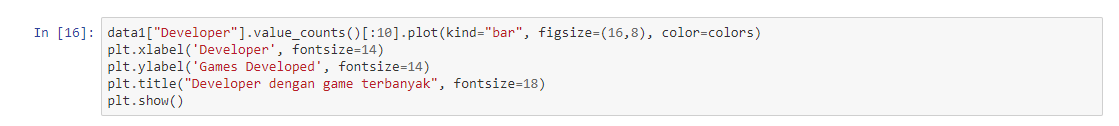
**Kesimpulan:** *Melihat dari data tersebut dapat disimpulkan bahwa game mulai berkembang sangat pesat pada tahun 1991 dan aneka macam platform bermunculan seiringan dengan berjalannya waktu dan mulai kembali terpusat ke beberapa platform setelah memasuki abad ke 21. Pada tahun 2019 dan 2020, data yang diambil kurang karena data ini diambil pada awal tahun 2019 yang menyebabkan data kurang lengkap.*

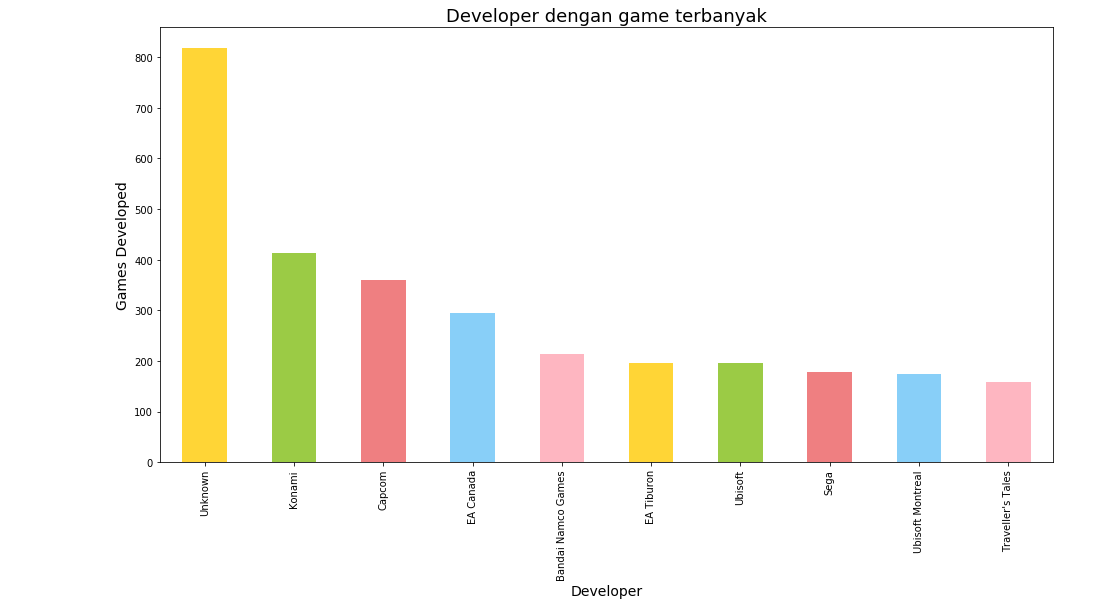
1. **Publisher dengan game terbanyak**

****

**Kesimpulan:** *Activision mengungguli setiap publisher lain dengan mempublish game terbanyak, diikuti oleh Ubisoft dan Electronic Arts.*

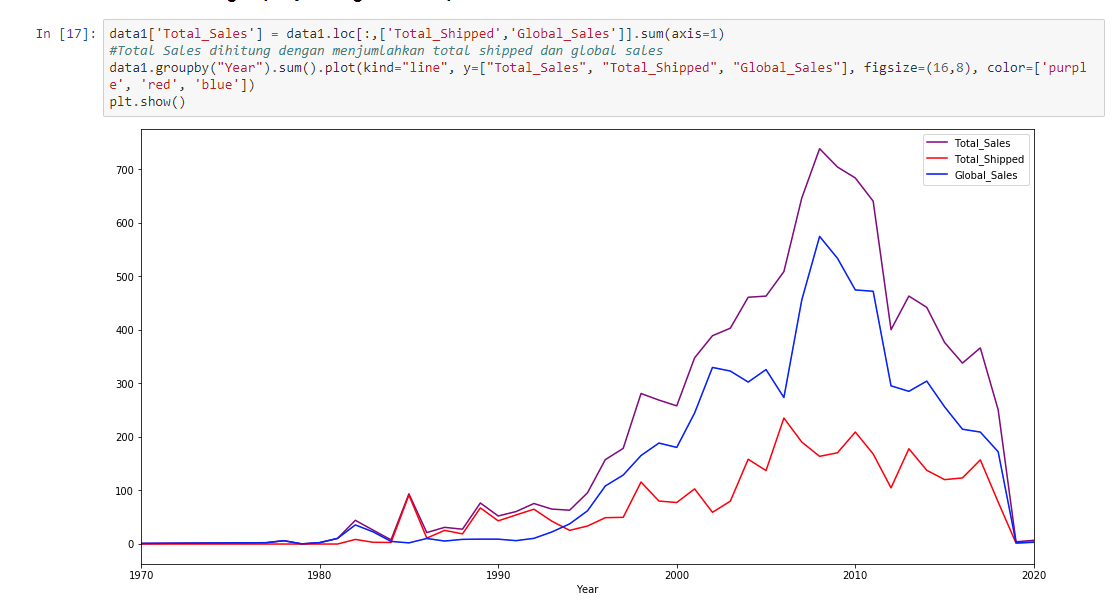
1. **Developer dengan game terbanyak**

****

****

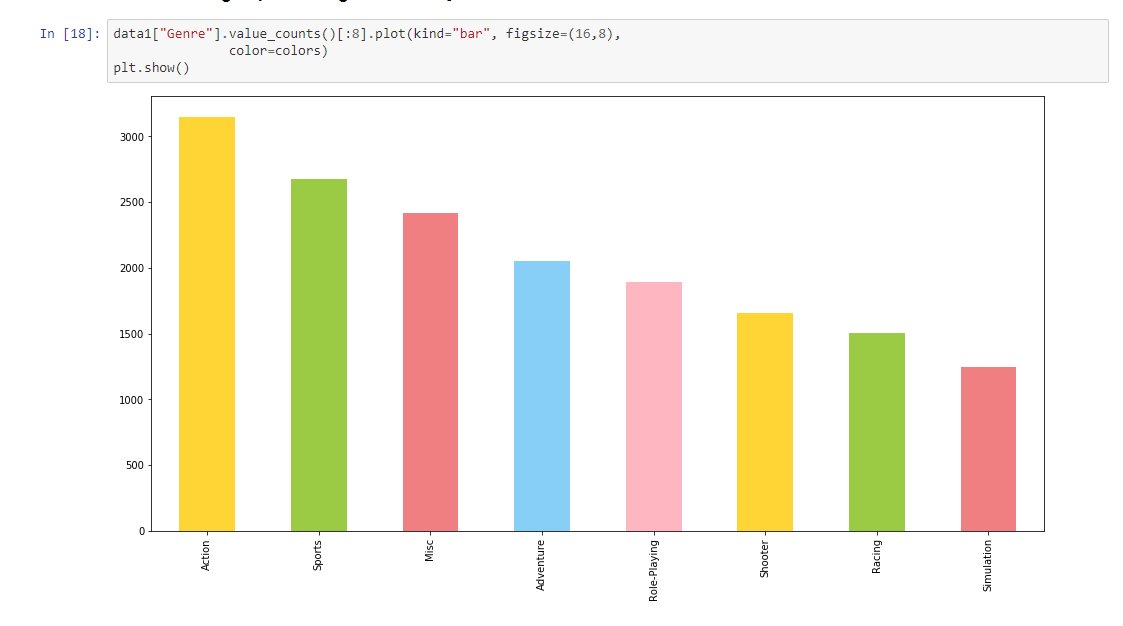
**Kesimpulan:** *Dengan melihat publisher dengan game terbanyak adalah Activision dan dengan melihat grafik developer dengan game terbanyak, maka dapat disimpulkan bahwa banyaknya game developer dan publisher tidak saling berkaitan.*

1. **Perkembangan penjualan game setiap tahun**

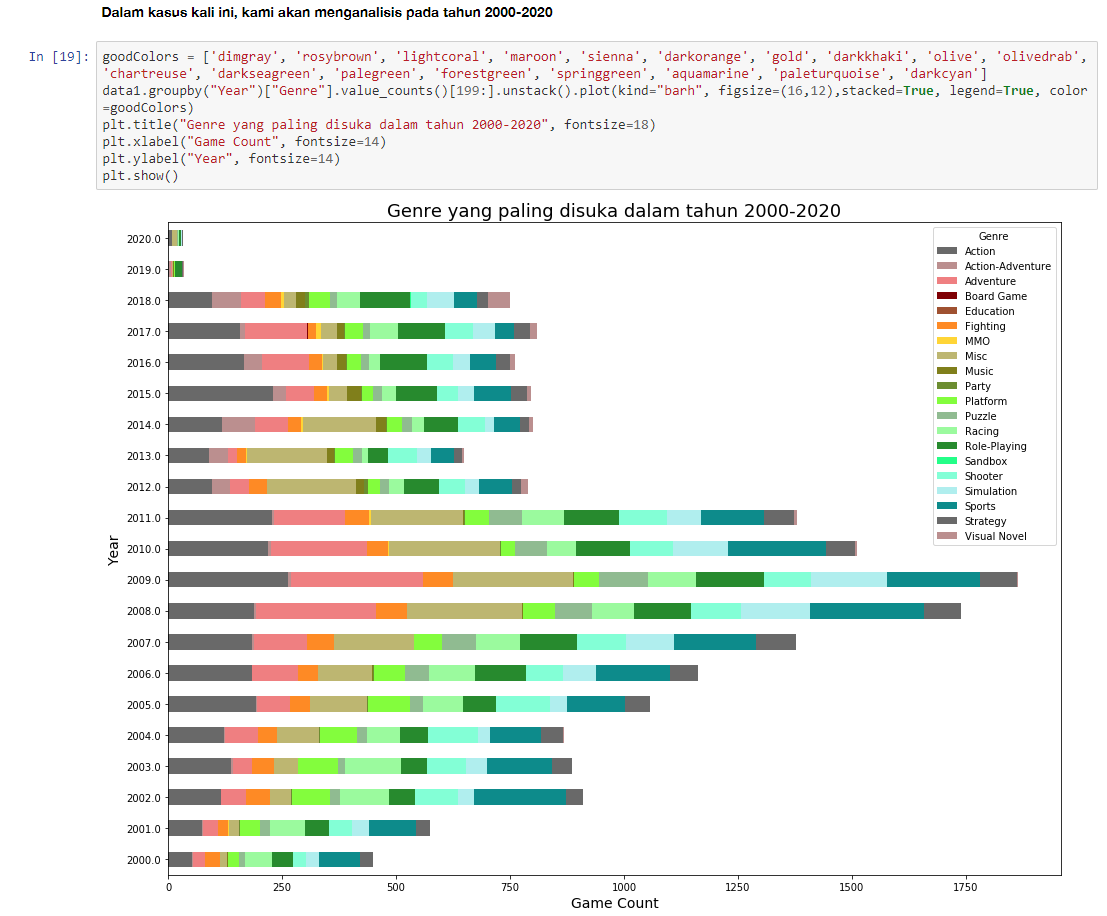
****

**Kesimpulan:** *Dapat disimpulkan kembali bahwa pada tahun 1991 penjualan terus meningkat dengan tajam hingga pada tahun 2010 penjualan mulai kembali menurun.*

1. **Genre dengan produksi game terbanyak**

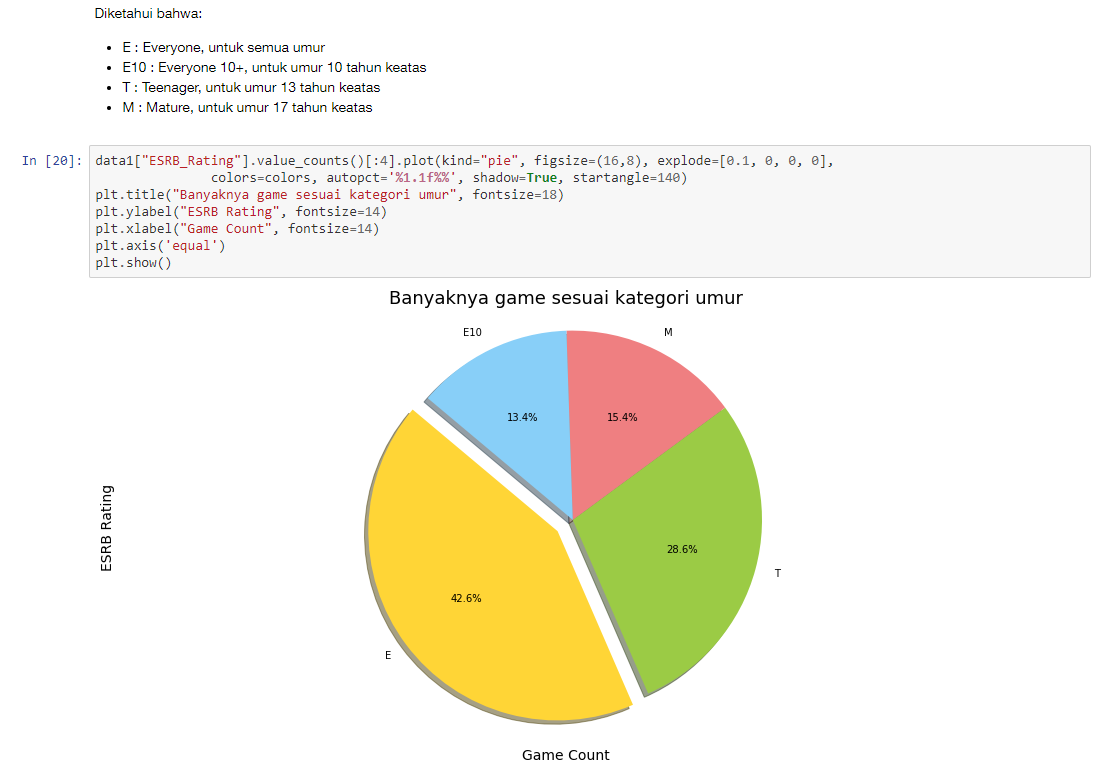
****

**Kesimpulan:** *Terlihat bahwa Action adalah genre yang paling disukai selama ini*

1. **Genre yang paling disukai dalam masing-masing tahun produksi**

**Kesimpulan:** *Terlihat bahwa dalam tahun 2012 dan 2013 peminat untuk game Adventure dan Action berkurang, sedangkan pada tahun-tahun lain yang biasanya memimpin adalah Action, Adventure, Role-Playing atau Sports*

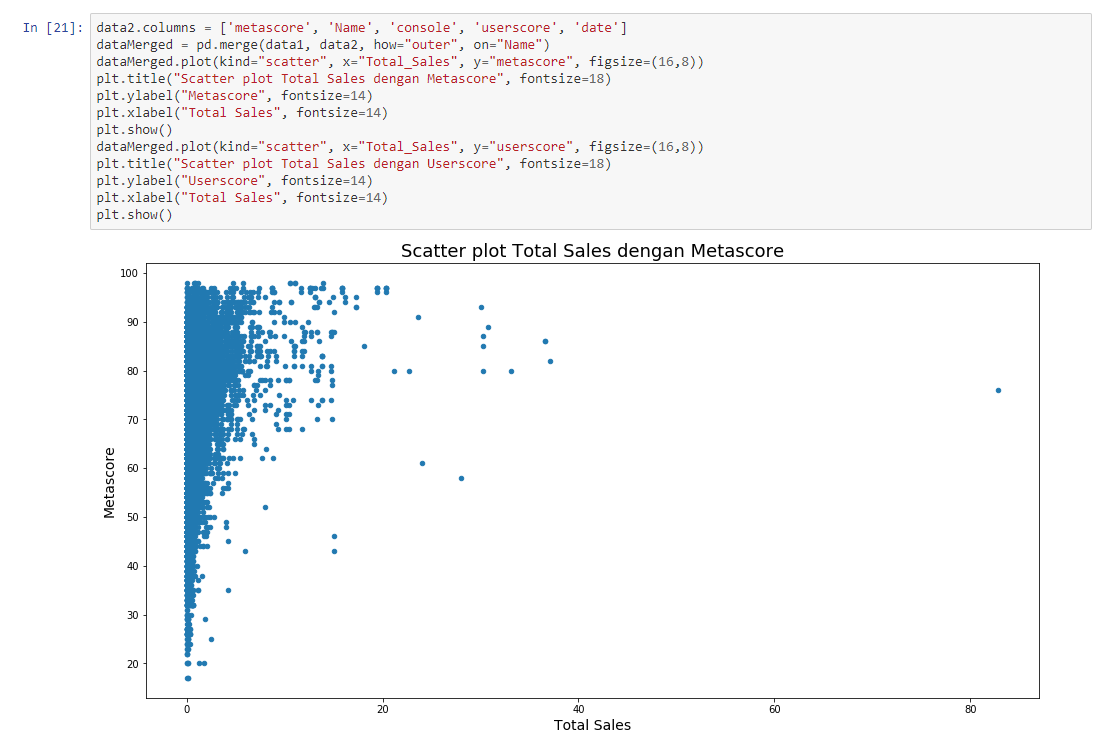
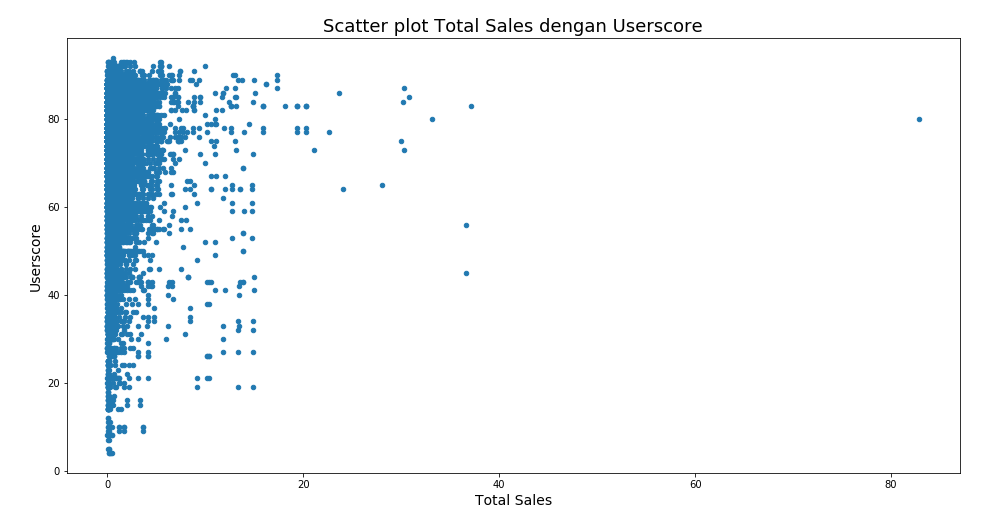
1. **Banyaknya game sesusai kategori umur (ESRB Rating)**

****

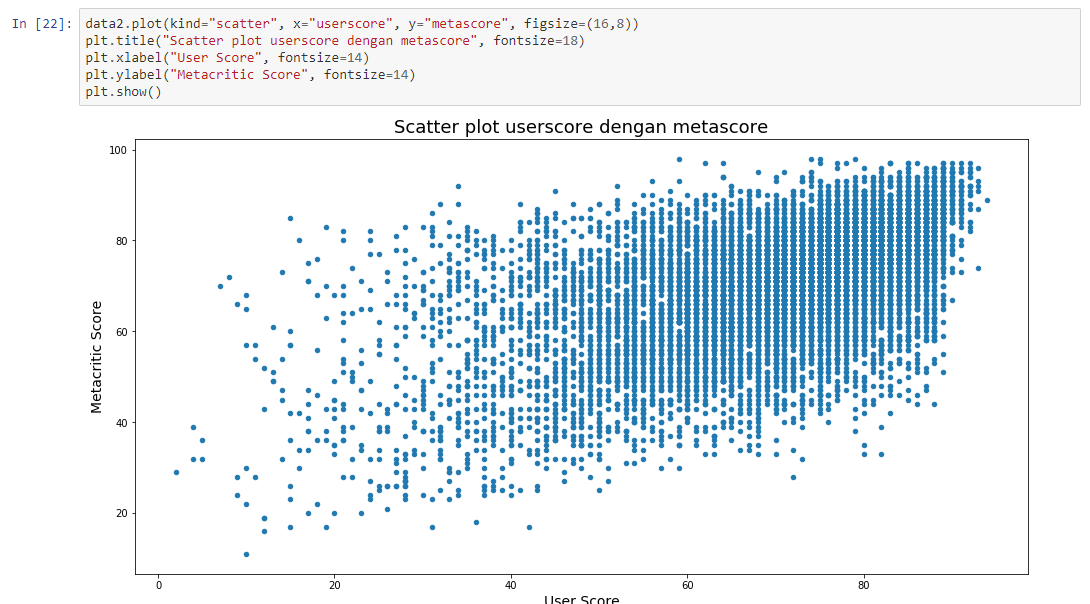
**Kesimpulan**: Terlihat dengan jelas bahwa game paling banyak tersedia untuk semua usia dan selanjutnya diikuti oleh teenager/remaja dan game-game untuk 17 tahun keatas yang mengandung kekerasan, darah, konten dewasa, dan bahasa yang keras masih mengungguli dari game 10 tahun keatas yang mana lebih banyak masuk ke kategori teenager.

Korelasi

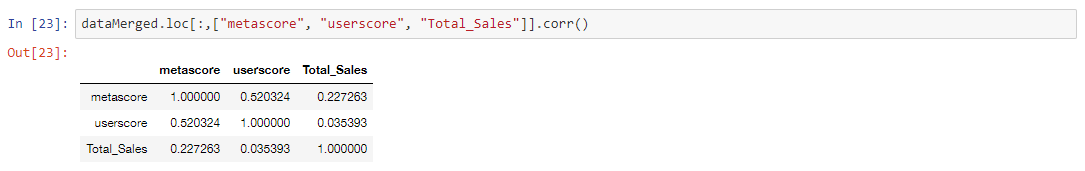
1. **Relasi antara penjualan video games dengan penilaian video games dari pemain dan Metacritic**

****

1. **Relasi antara penilaian dari pemain dan Metacritic**

****

**Korelasi antara Metascore, Userscore, Total Sales**

****

**Kesimpulan:** *Kita dapat melihat bahwa korelasi antara metascore, userscore dan total sales beragam. Pada baris pertama kita dapat menyimpulkan bahwa userscore dan metascore cukup berbanding lurus. Hal itu menyebabkan bahwa jika metascore tinggi cukup tinggi juga kemungkinan userscore tinggi. Kita dapat melanjutkan ke analisis selanjutnya, Total Sales dan Metascore kurang berbanding lurus, sehingga jika penjualan game besar belum tentu metascore besar. Hal ini menurut kami disebabkan oleh kemungkinan orang sudah membeli game tersebut sebelum mengetahui bahwa game itu tidak terlalu sesuai harapannya. Hal ini juga berlaku untuk Total Sales dan Userscore bahkan tidak berbanding lurus sama sekali.*