

LAPORAN PROYEK DATA ANALYST

4143105 - Data Analyst

Analisis Deskriptif dan Visualisasi Data Bisnis Menggunakan Tableau dengan Pendekatan CRISP-DM pada Dataset Video Game Sales 2024



Disusun Oleh:

11423011	Ruth Heppi Evelin Sinambela
11423028	Grace Advryanti Tampubolon
11423006	Samuel Alex Bonaparte Sirait

INSTITUT TEKNOLOGI DEL

FAKULTAS VOKASI

2025

DAFTAR ISI

BAB I.....	3
PENDAHULUAN	3
1.1 Latar Belakang	3
1.2 Tujuan	4
BAB II	5
PEMBAHASAN	5
2.1 Tahapan CRISP-DM.....	5
2.1.1 Business Understanding.....	5
2.1.2 Data Understanding	6
2.1.2.1 Visualisasi Data Mentah	13
2.1.3 Data Preparation.....	17
2.1.3.1 Data Cleaning, Data Transformation, Data Integration	17
2.1.3.1 Descriptive Analysis (Modeling).....	20
2.1.4.1 Hitung Ukuran Statistik Deskriptif	20
2.1.4.1.1 Mean, Median, dan Variance, Mode & Correlation	21
2.1.3.2 Insight Interpretation.....	29
2.1.3.2 Deployment.....	31
DAFTAR PUSTAKA.....	33

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam era transformasi digital yang semakin pesat, kemampuan perusahaan untuk memahami tren pasar dan preferensi konsumen menjadi salah satu faktor penentu keberhasilan dalam persaingan industri hiburan, khususnya industri video game yang terus berkembang secara global. Strategi pemasaran tradisional yang bersifat umum kini dianggap kurang efektif, karena pemain dan konsumen game memiliki karakteristik yang sangat beragam dan cenderung menuntut pendekatan yang lebih personal [1]. Oleh karena itu, penggunaan analisis data dalam pengambilan keputusan bisnis menjadi kebutuhan strategis bagi setiap perusahaan game yang ingin bertahan dan berkembang.

Perusahaan kini banyak mengandalkan analisis data penjualan dan performa produk untuk mendapatkan wawasan mendalam mengenai perilaku konsumen. Salah satu pendekatan yang umum digunakan adalah Descriptive Analytics, yang berfokus pada penggambaran pola dan tren yang terjadi berdasarkan data historis. Melalui analisis deskriptif, perusahaan dapat memahami platform yang paling diminati, genre yang paling populer, wilayah dengan penjualan tertinggi, serta faktor-faktor yang mempengaruhi keberhasilan sebuah game di pasar global [2].

Dalam konteks proyek ini, digunakan dataset Video Game Sales 2024 yang berisi data mengenai penjualan video game di seluruh dunia. Data ini mencakup variabel seperti nama game, platform, tahun rilis, genre, penerbit, pengembang, penjualan di berbagai wilayah (Amerika Utara, Eropa, Jepang, dan wilayah lainnya), serta total penjualan global [3]. Melalui analisis data ini, diharapkan dapat ditemukan hubungan antara faktor-faktor seperti genre, platform, dan wilayah distribusi dengan performa penjualan yang dicapai oleh masing-masing game.

Metodologi yang digunakan dalam penelitian ini adalah CRISP-DM (Cross-Industry Standard Process for Data Mining), sebuah kerangka kerja analisis data yang sistematis dan terstruktur. Tahapan dalam CRISP-DM meliputi Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment. Pendekatan ini memungkinkan proses analisis dilakukan secara bertahap, mulai dari pemahaman permasalahan bisnis hingga interpretasi hasil dan penyusunan rekomendasi berbasis data [4].

Dalam penerapannya, proyek ini menggunakan Tableau sebagai alat bantu visualisasi data untuk melakukan eksplorasi dan analisis deskriptif. Tableau dipilih karena kemampuannya dalam mengolah data secara

interaktif serta menyajikan hasil analisis dalam bentuk visual yang mudah dipahami. Melalui visualisasi interaktif ini, pengguna dapat mengidentifikasi pola-pola tersembunyi seperti perbandingan penjualan antar wilayah, perbedaan performa antar platform, hubungan antara tahun rilis dengan total penjualan, hingga pengaruh genre terhadap tingkat popularitas game [5].

Dengan pendekatan ini, hasil analisis tidak hanya memberikan gambaran umum mengenai kinerja pasar video game global, tetapi juga menghasilkan integration

bisnis yang dapat dijadikan dasar dalam merancang strategi produksi, distribusi, dan pemasaran yang lebih efektif dan tepat sasaran. Selain itu, hasil visualisasi dapat membantu pihak manajemen dalam memahami dinamika pasar industri game secara lebih intuitif, sehingga proses pengambilan keputusan dapat dilakukan secara cepat dan akurat.

Secara keseluruhan, proyek ini bertujuan untuk menunjukkan bagaimana penerapan metodologi CRISP-DM dan analisis deskriptif menggunakan Tableau dapat membantu perusahaan dalam memahami tren penjualan video game, mengidentifikasi faktor keberhasilan produk, serta mendorong pengambilan keputusan berbasis data (data-driven decision making).

1.2 Tujuan

Proyek ini bertujuan untuk menganalisis data penjualan video game menggunakan metode Descriptive Analytics dengan bantuan Tableau. Adapun tujuan khusus dari proyek ini adalah untuk:

1. Memahami faktor-faktor utama yang mempengaruhi tingkat penjualan video game di berbagai wilayah.
2. Mengidentifikasi genre, platform, dan penerbit yang memiliki performa terbaik berdasarkan data penjualan global.
3. Menemukan pola dan tren penjualan yang dapat dijadikan dasar dalam penyusunan strategi produksi, distribusi, dan pemasaran yang lebih efektif.
4. Memberikan insight dan rekomendasi bisnis yang dapat mendukung pengambilan keputusan berbasis data (data-driven decision making) di industri video game.

BAB II

PEMBAHASAN

2.1 Tahapan CRISP-DM

Metodologi CRISP-DM (Cross Industry Standard Process for Data Mining) digunakan sebagai panduan utama dalam pelaksanaan proyek analisis data ini.

2.1.1 Business Understanding

Tahap pertama dalam metodologi CRISP-DM adalah memahami konteks bisnis dan tujuan analisis yang ingin dicapai. Dalam kasus ini, perusahaan yang bergerak di industri video game ingin memahami tren penjualan global serta faktor-faktor yang mempengaruhi keberhasilan suatu game di pasar. Meskipun banyak game telah dirilis setiap tahunnya, perusahaan menghadapi tantangan dalam menentukan platform, genre, atau wilayah distribusi yang paling berkontribusi terhadap penjualan.

Permasalahan utama yang ingin dijawab melalui proyek ini adalah: “Bagaimana memahami faktor-faktor yang mempengaruhi tingkat penjualan video game di berbagai wilayah serta mengidentifikasi genre dan platform dengan performa terbaik untuk strategi bisnis berikutnya?”

Untuk menjawab permasalahan tersebut, ditetapkan beberapa pertanyaan analitik utama, yaitu:

1. Platform apa yang memiliki tingkat penjualan tertinggi di pasar global?
2. Genre apa yang paling diminati oleh konsumen di masing-masing wilayah (Amerika Utara, Eropa, Jepang, dan lainnya)?
3. Apakah terdapat tren tertentu antara tahun rilis game dengan tingkat penjualan globalnya?
4. Penerbit atau pengembang mana yang paling konsisten menghasilkan penjualan tinggi?

Dari pertanyaan tersebut, ditetapkan tujuan bisnis yaitu untuk meningkatkan efektivitas strategi pemasaran dan distribusi game, memahami karakteristik pasar di berbagai wilayah, serta mengoptimalkan pengambilan keputusan berbasis data agar strategi peluncuran dan promosi game menjadi lebih tepat sasaran dan efisien.

2.1.2 Data Understanding

Tahap ini berfokus pada pemahaman terhadap struktur dan karakteristik data yang digunakan. Dataset yang digunakan dalam proyek ini berasal dari platform Kaggle, dengan judul "*Video Game Sales 2024*" Sumber dataset: <https://www.kaggle.com/datasets/asaniczka/video-game-sales-2024>. Dataset ini berisi 39798 baris dan 14 kolom yang dibuat oleh @asaniczka. Analisis Penjualan Video Game merupakan proses untuk memahami pola dan tren penjualan game berdasarkan berbagai faktor seperti platform, genre, penerbit, serta wilayah pasar. Tujuan dari analisis ini adalah membantu perusahaan game, pengembang, maupun distributor dalam menentukan strategi pemasaran, pengembangan produk, dan pengambilan keputusan berbasis data.

Melalui analisis penjualan game, perusahaan dapat mengetahui kategori atau platform mana yang memiliki performa tinggi, serta memahami preferensi pasar di berbagai wilayah. Dengan demikian, perusahaan dapat memfokuskan sumber daya pada genre atau platform yang paling menguntungkan dan diminati pemain.

Dataset "Video Game Sales 2024" mencakup beberapa atribut utama, yaitu:

1) Game Information (Data Utama Game)

1. Title: Menyimpan nama game yang dirilis.
2. Console: Menunjukkan platform atau konsol tempat game tersebut dirilis, misalnya PS3, PS4, Xbox One, PC, atau Nintendo Switch.
3. Release_Date: Berisi tanggal resmi perilisan game dengan format teks "DD/MM/YYYY". Dari kolom ini juga bisa diambil informasi tahun rilis.
4. Genre: Menjelaskan jenis atau kategori game, seperti Action, Adventure, Sports, Role-Playing, dan lain-lain.
5. Publisher: Nama perusahaan penerbit yang merilis game tersebut.
6. Developer: Nama studio atau pengembang yang membuat game.
7. Critic_Score: Nilai atau skor yang diberikan oleh kritikus profesional, bertipe numerik (float) dengan rentang umumnya 0–10.
8. Img: Menyimpan tautan (URL) atau path ke gambar sampul game.
9. Last_Update: Berisi tanggal terakhir kali data game diperbarui, dalam format teks "DD/MM/YYYY", dengan beberapa nilai yang kosong.

2) Sales Performance (Data Penjualan)

1. NA_Sales: Menunjukkan total penjualan game di wilayah Amerika Utara, satuannya dalam juta unit.
2. PAL_Sales: Mencatat total penjualan game di wilayah Eropa dan Australia (PAL region), juga dalam juta unit
3. JP_Sales: Menampilkan total penjualan game di wilayah Jepang, dalam juta unit.
4. Other_Sales: Menunjukkan total penjualan di wilayah lain selain Amerika Utara, Eropa, dan Jepang.
5. Total_Sales: Merupakan total keseluruhan penjualan global game (dalam juta unit).

3) Market Insight (Data Pasar dan Tren)

Data ini memungkinkan analisis mengenai:

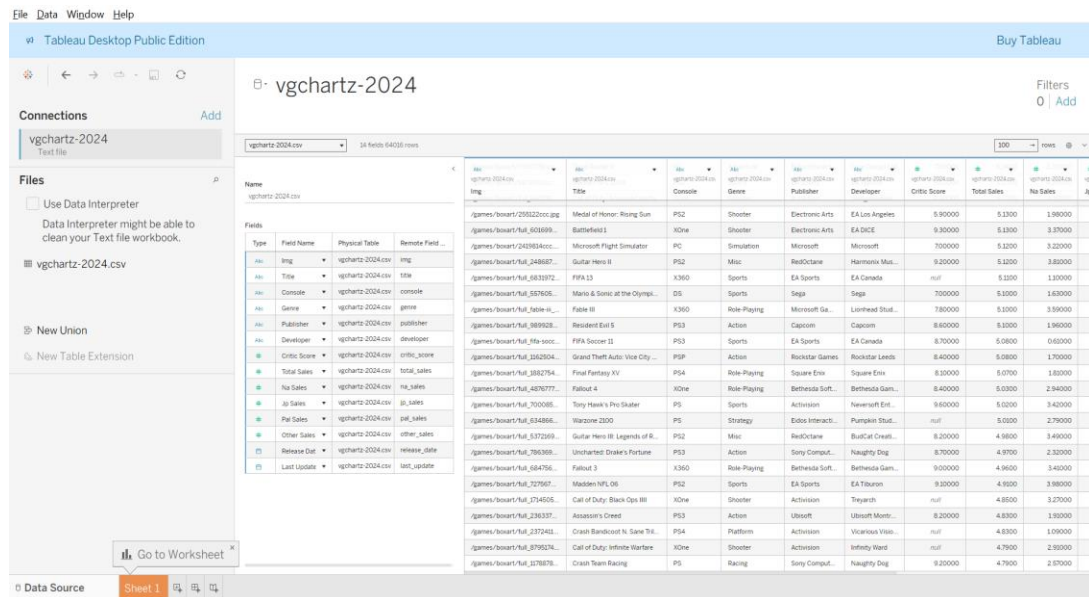
1. Tren penjualan berdasarkan tahun rilis.
2. Platform dengan penjualan tertinggi di tiap wilayah.
3. Genre paling populer di dunia dan di masing-masing kawasan.
4. Performa publisher berdasarkan total penjualan global.

Eksplorasi awal dilakukan untuk menjawab pertanyaan berikut:

1. Apakah ada data hilang (missing values) dalam dataset?
2. Apakah terdapat outlier pada data penjualan (misalnya penjualan global yang terlalu tinggi dibanding rata-rata)?
3. Apakah tipe data pada setiap kolom sudah sesuai (angka, kategori, atau teks)?
4. Apakah terdapat duplikasi data game?
5. Bagaimana distribusi penjualan global dan perbandingan antar wilayah (NA, EU, JP, dan lainnya)?

Langkah – langkah untuk menjawab pertanyaan diatas adalah dengan pengerjaan menggunakan Tools Tableau:

1. Mengimpor dataset melalui menu **Connect** → **Text File** → **vgchartz-2024.csv**.



2. Meninjau tipe data setiap kolom dan memastikan konsistensi (angka, teks, atau tanggal).

No.	Kolom	Tipe Data	Keterangan/Fungsi
1.	img	String (Teks)	URL atau path gambar sampul game.
2.	Title	String (Teks)	Nama game yang dirilis.
3.	Console	Integer (Tahun)	Platform atau konsol tempat game dirilis (misalnya PS3, PS4, Xbox One, PC, Switch, dll).
4.	Genre	String (Teks)	Kategori atau jenis game (misalnya Action, Adventure, Sports, Role-Playing, dll).
5.	Publisher	String (Teks)	Nama perusahaan penerbit game.
6.	developer	String (Teks)	Nama pengembang game.
7.	critic_score	Numeric (Float)	Skor ulasan.
8.	total_sales	Numeric (Float)	Total penjualan global
9.	NA_Sales	Numeric (Float)	Total penjualan game di wilayah Amerika Utara.
10.	JP_Sales	Numeric (Float)	Total penjualan game di wilayah Jepang.

No.	Kolom	Tipe Data	Keterangan/Fungsi
11.	Pal_sales	Numeric (Float)	Total penjualan di wilayah Eropa/Australia
12.	Other_Sales	Numeric (Float)	Total penjualan di wilayah lain selain NA, EU, dan JP.
13.	release_date	Date (Tanggal)	Tanggal resmi rilis game.
14.	last_update	Date (Tanggal)	Tanggal terakhir pembaruan data penjualan atau informasi game.

3. Meninjau kolom yang memiliki missing value.

Berdasarkan hasil eksplorasi awal terhadap dataset *Video Game Sales (VGChartz 2024)* menggunakan Tableau, ditemukan bahwa terdapat beberapa kolom yang memiliki nilai kosong (*missing value*). Tiga kolom yang teridentifikasi memiliki missing value adalah Critic_Score, JP_Sales, dan Last_Update. Kehadiran nilai kosong pada ketiga kolom tersebut menunjukkan bahwa tidak semua data permainan memiliki informasi lengkap mengenai skor dari kritikus, jumlah penjualan di wilayah Jepang, maupun waktu pembaruan terakhir data. Kondisi ini dapat mempengaruhi proses analisis terutama apabila kolom-kolom tersebut digunakan dalam perhitungan statistik atau pemodelan prediktif.

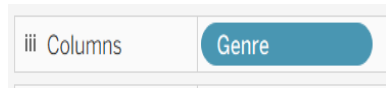
#	#	#
vgchartz-2024.csv	vgchartz-2024.csv	vgchartz-2024.csv
Critic Score	Jp Sales	Last Update
9.60000	0.38000	null
null	0.44000	null
8.10000	0.04000	3/21/2018
8.70000	0.39000	3/21/2018
8.80000	0.08000	10/11/2018
9.80000	0.04000	null
8.40000	0.13000	null
8.00000	0.50000	11/7/2018
9.50000	0.10000	9/12/2018
8.10000	0.10000	4/1/2018
8.80000	0.01000	4/11/2018
9.50000	0.19000	1/14/2018
8.70000	0.27000	3/24/2018
8.30000	0.07000	4/1/2018
10.00000	0.23000	3/24/2018
8.90000	0.13000	4/6/2018
9.50000	null	null
10.00000	0.24000	null
6.90000	0.14000	1/4/2018
7.50000	null	null
9.30000	0.02000	1/14/2018
null	0.06000	null
9.60000	0.29000	3/24/2018
null	0.35000	null
9.30000	0.08000	null
9.00000	0.08000	null

Dengan peninjauan ini, kualitas data dapat ditingkatkan sebelum melanjutkan ke tahap *data cleaning* dan *data visualization*.

4. Meninjau Outlier/nilai ekstrim

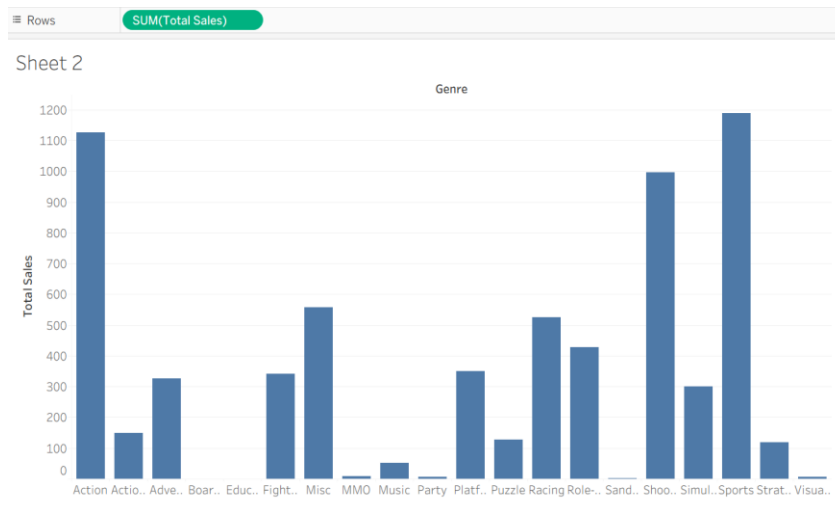
Langkah untuk mendeteksi Outlier di Tableau:

1. Tarik Genre ke area Columns.



2. Tarik Total Sales ke area Rows.

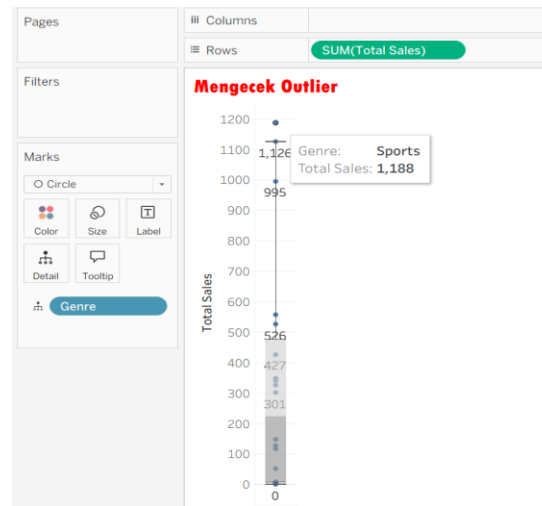
Tableau otomatis membuat *bar chart* awal.



3. Klik menu Show Me di kanan atas layar.



4. Pilih tipe visualisasi Box-and-Whisker Plot (Boxplot).
(Ikonnnya berbentuk kotak dengan garis tengah dan titik di luar batas.)



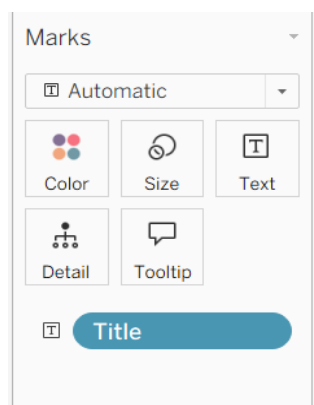
Boxplot di Tableau menunjukkan sebaran dan outlier penjualan tiap genre. Titik di luar kotak menandakan game dengan penjualan ekstrem, kotak menunjukkan variasi utama penjualan, dan garis tengahnya adalah median. Dari sini bisa dilihat genre dengan penjualan tinggi, stabil, atau memiliki game best seller. Misalnya, game dengan penjualan ekstrem tinggi (seperti Shooter genre dengan total sales > 400) menunjukkan produk yang sangat sukses. Ini bukan kesalahan data, tapi fakta menarik yang perlu diperhatikan oleh perusahaan.

5. Mendeteksi Duplikasi Data

Untuk mendeteksi duplikasi ini, digunakan Count Distinct pada kolom Title, karena biasanya setiap baris di dataset punya kolom unik seperti Title.

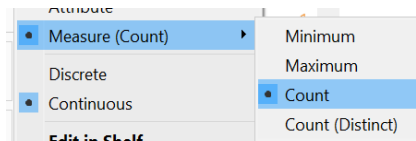
Langkah pengerjaan menggunakan Tableau:

1. Buka Worksheet baru.
2. Tarik Title ke area Text (di Marks).

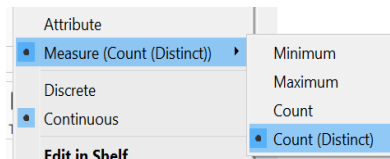


3. Klik tanda panah pada Title → pilih:

- Measure → Count (hitung total semua baris).



- Lalu tambahkan satu lagi Title ke Text → ubah jadi Measure → Count (Distinct).



Sekarang Tableau menampilkan dua nilai:

- Count(Title) = total semua baris.
- CountD(Title) = jumlah judul unik.

4. Untuk menghitung duplikat, buat Calculated Field baru:

IF { FIXED [Title] : COUNT([Title]) } > 1 THEN "Duplicate" ELSE "Unique" END

Duplicate Flag

IF { FIXED [Title] : COUNT([Title]) } > 1 THEN "Duplicate"

5. Interpretasi:

Mengecek Duplikasi

Title	
0 day Attack on Earth	1
0 Kara no Shogi: Shogi Youchien ..	2
0-Ji no Kane to Cinderella: Hallo..	1
0x10c	2
1 on 1	1
1 vs. 100	1
1-2-Switch	1
1... 2... 3... KICK IT! (Drop That Be..	1
1/2 Summer +	1
1nsane	1
1Path	1
2 Days to Vegas	3
2 Fast 4 Gnomz	1

- Jika nilai = 1 → data unik, tidak ada duplikasi.
- Jika nilai > 1 → data duplikat, artinya judul game itu tercatat lebih dari sekali.

Dari hasil pemeriksaan duplikasi di Tableau, terdapat beberapa judul game yang muncul lebih dari satu kali, seperti “0 Kara no Shogi”, “0x10c”, dan “2 Days to Vegas”. Ini menunjukkan

adanya duplikasi data pada kolom Title yang sebaiknya dibersihkan sebelum analisis penjualan dilakukan, agar hasil analisis tidak bias.

2.1.2.1 Visualisasi Data Mentah

Pada tahap ini dilakukan eksplorasi dan visualisasi awal terhadap data mentah (raw data) yang diambil dari dataset *VGChartz 2024.csv*. Tujuan utama dari tahap ini adalah untuk memahami karakteristik awal data sebelum dilakukan proses pembersihan (*data cleaning*), sekaligus mendeteksi adanya anomali, missing value, maupun outlier yang dapat mempengaruhi hasil analisis.

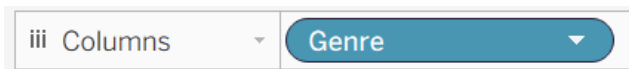
Beberapa jenis visualisasi yang dilakukan pada tahap ini antara lain:

1. Top Genres by Global Sales (Bar Chart)

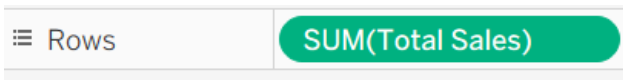
Tujuan: Menunjukkan genre dengan total penjualan tertinggi.

Langkah:

1. Worksheet baru → “Top Genres”.
2. Tarik Genre ke Columns.



3. Tarik Total Sales ke Rows → SUM otomatis.



4. Pilih Bar Chart.
5. Sort Descending → genre terlaris di kiri.
6. Klik Marks → Color → pilih gradasi biru (#a8dadc → #1d3557).
7. Aktifkan Label di ujung bar.
8. Tambahkan judul: “Top Game Genres by Total Sales”.

Interpretasi:

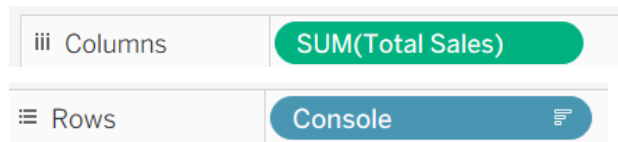


2. Sales by Console (Horizontal Bar Chart)

Tujuan: Bandingkan performa tiap konsol.

Langkah:

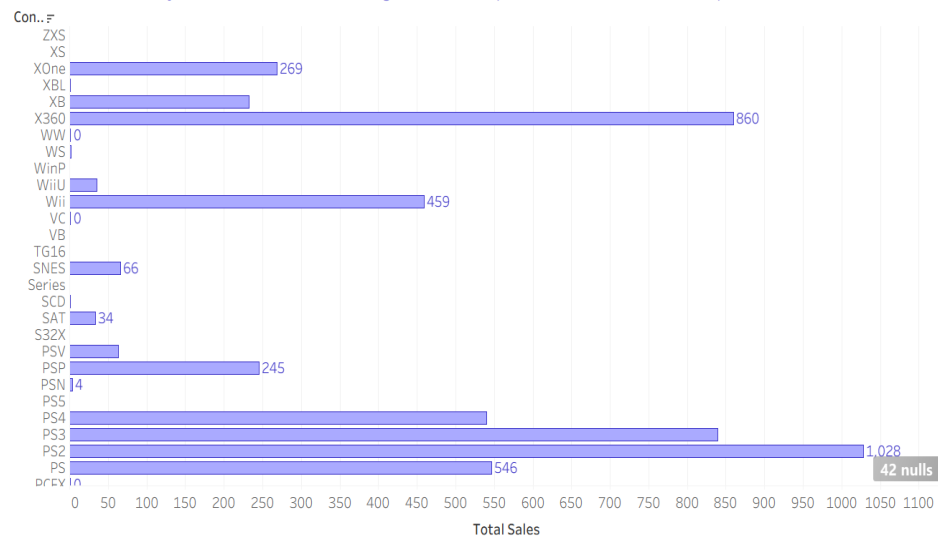
1. Worksheet “Sales by Console”.
2. Tarik Console ke Rows, Total Sales ke Columns.



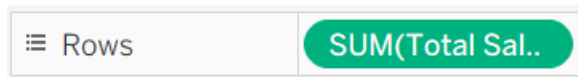
3. Sort → Descending.
4. Color → gunakan palet gradient biru ke ungu (#48cae4 → #3f37c9).
5. Label: tampilkan angka penjualan di kanan bar.

Hasil:

Visualisasi Eksplorasi Awal : Sales by Console (Horizontal Bar Chart)

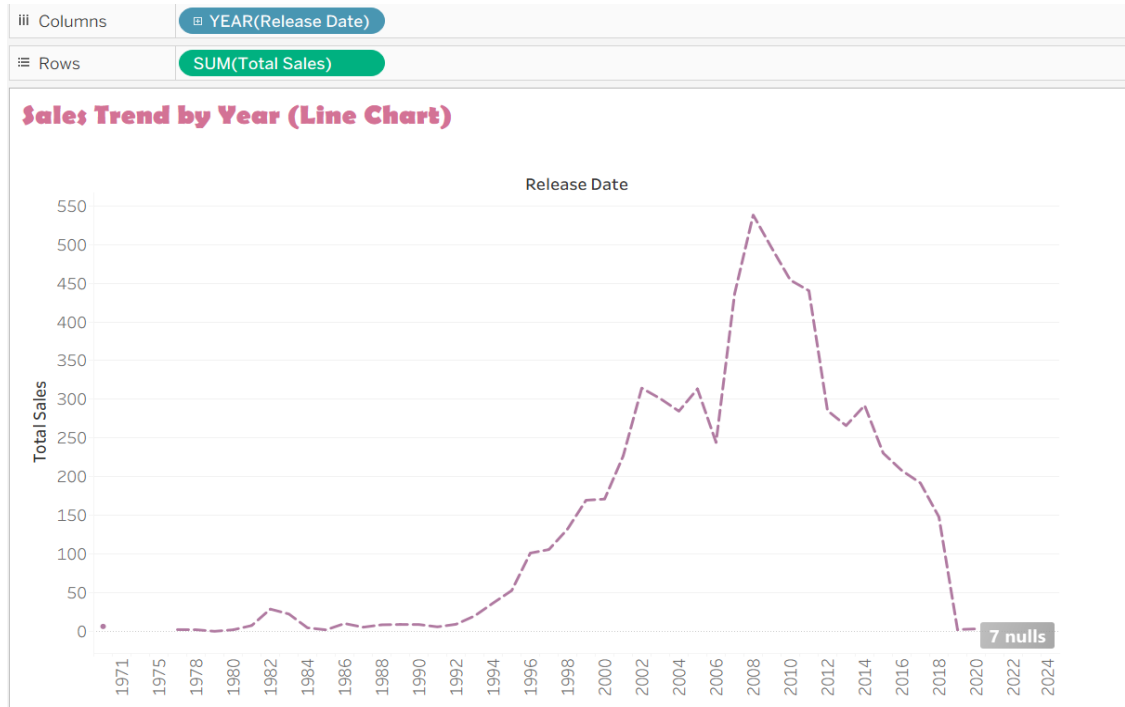


3. Total Sales → Rows.



4. Pilih Line Chart.
5. Warna garis → oranye lembut (#e76f51).
6. Tambahkan titik data (Marks → Line + Circle).

Hasil:



4. Critic Score vs Total Sales (Scatter Plot)

Tujuan: Melihat hubungan skor ulasan dan penjualan.

Langkah:

1. Worksheet "Score vs Sales".
2. Critic Score → Columns.
3. Total Sales → Rows.
4. Genre → Color.
5. Publisher → Detail.
6. Tambahkan Trend Line (klik kanan → Trend Lines → Linear).

Hasil:



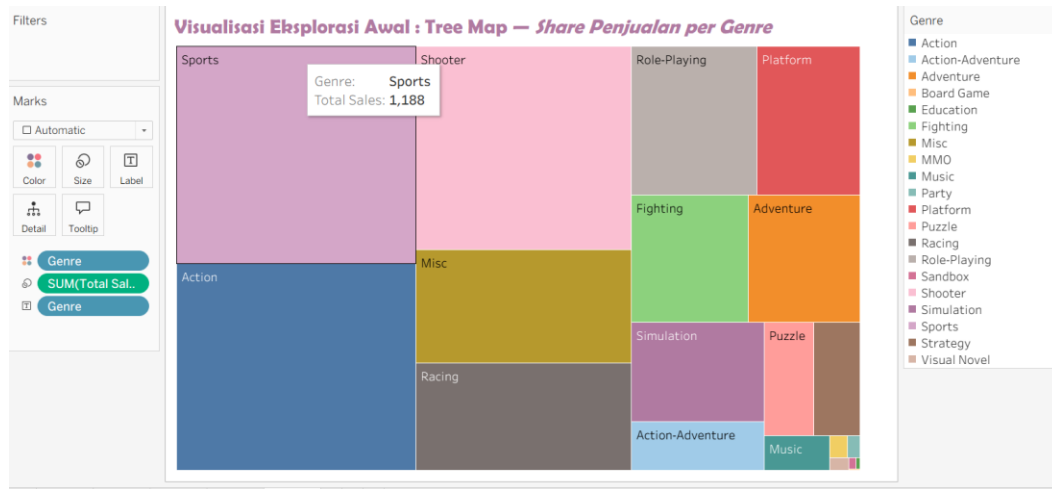
5. Tree Map - *Share Penjualan per Genre*

Tujuan: Alternatif visual dari Pie Chart untuk melihat proporsi.

Langkah:

1. Worksheet baru → “Sales Treemap”.
2. Tarik Genre ke Color dan Label.
3. Tarik Total Sales ke Size. Pilih Marks → Treemap.
4. Klik Label → tampilkan nilai dan % total.

Hasil:



2.1.3 Data Preparation

Tahap Data Preparation merupakan salah satu tahap yang sangat penting dalam metodologi CRISP-DM, karena pada tahap ini dilakukan proses mempersiapkan data agar siap digunakan untuk analisis dan visualisasi. Proses ini mencakup kegiatan pembersihan data (data cleaning), transformasi data (data transformation), serta penggabungan data (data integration) apabila diperlukan.

Dalam proyek ini, dataset yang digunakan adalah “*VGChartz 2024.csv*”, yang berisi informasi penjualan video game dari berbagai platform dan wilayah, termasuk data mengenai nama game, tahun rilis, genre, publisher, dan total penjualan di berbagai kawasan dunia. Pada sub-bab ini akan dijelaskan secara rinci mengenai langkah-langkah data preparation yang dilakukan menggunakan dataset *VGChartz 2024.csv*, dengan tujuan untuk memastikan data dalam kondisi bersih, konsisten, dan siap dianalisis dalam tahap eksplorasi dan visualisasi menggunakan Tableau.

2.1.3.1 Data Cleaning, Data Transformation, Data Integration

Langkah pertama dalam tahap ini adalah membersihkan data (data cleaning) untuk mengatasi permasalahan yang umum ditemukan pada dataset mentah, seperti *missing value*, *duplikasi*, dan *outlier*. Berdasarkan eksplorasi diatas, kami menemukan nilai null yaitu pada kolom Income. Pemeriksaan awal terhadap data dilakukan dengan melihat indikator null pada tampilan *Data Pane* di Tableau. Kolom Income terdeteksi memiliki beberapa nilai kosong (null), yang jika dibiarkan dapat mengganggu analisis statistik seperti perhitungan rata-rata pendapatan atau distribusi pendapatan pelanggan.

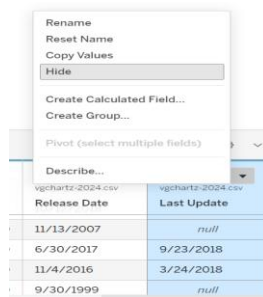
Untuk menangani hal tersebut, dilakukan proses imputasi dengan mengganti nilai kosong menggunakan median pendapatan. Pemilihan median dilakukan karena median lebih tahan terhadap keberadaan *outlier* dibandingkan mean, sehingga tidak menggeser distribusi data secara signifikan.

Langkah:

1. Menghapus Kolom yang Tidak Diperlukan

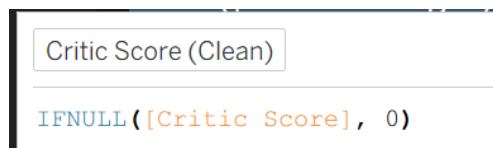
Beberapa kolom mungkin tidak relevan untuk analisis penjualan atau menyebabkan noise.

- Klik kanan nama field di Data Pane → Hide untuk menyembunyikan kolom yang tidak dipakai.



2. Memeriksa dan Isi Nilai Kosong di Critic Score dengan 0 atau “Unknown”

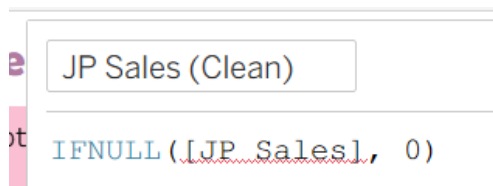
- Klik kanan → Create Calculated Field.
- Beri nama: **Critic Score (Clean)**
- Ketik rumus: **IFNULL([Critic Score], 0)**
- Klik OK.



Baris tanpa skor akan muncul sebagai kategori “Unknown” di visualisasi. Critic Score (clean) ini yang akan digunakan untuk mengganti critic score yang berisi nilai null

3. Memeriksa dan Isi Nilai Kosong di Jp Sales dengan 0

- Buat *Calculated Field* baru bernama: JP Sales (Clean)



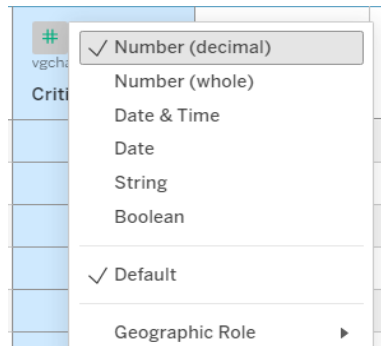
JP Sales (clean) ini yang akan digunakan untuk mengganti nilai null di JP Sales.

4. Menstandarkan Format Data

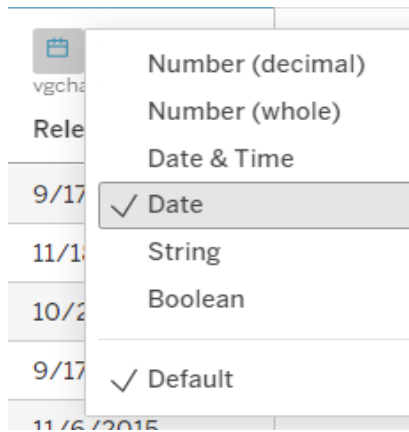
Menstandarkan format data dilakukan untuk memastikan setiap kolom memiliki tipe data yang sesuai, seperti angka pada kolom penjualan, teks pada kolom kategori (genre, publisher, console), dan tanggal pada kolom waktu (release date), sehingga data dapat diproses dan divisualisasikan dengan benar di Tableau.

Langkah:

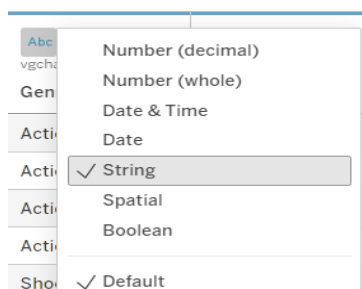
- Total Sales, NA Sales, JP Sales, PAL Sales, Other Sales → tipe Number (Decimal)



- Release Date → tipe Date



- Genre, Publisher, Console, Title → tipe String



5. Menangani Duplikasi Data

Memeriksa apakah ada data game yang muncul lebih dari sekali. Dalam hal ini, outlier digunakan untuk:

- Menunjukkan best-seller games di genre tertentu.
- Menjelaskan perbedaan mean dan median pada grafik Polygon (Mean & Median Total Sales per Genre).

Jadi, tidak perlu dihapus. Setelah tahap ini, data sudah bisa digunakan untuk visualisasi.

Pada tahap integrasi data, dilakukan pemeriksaan untuk memastikan apakah data perlu digabungkan dari beberapa sumber. Karena dataset yang digunakan hanya berasal dari satu sumber (satu file CSV), maka proses integrasi tidak diperlukan. Namun, pemeriksaan struktur dan konsistensi data tetap dilakukan untuk memastikan dataset siap dianalisis.

2.1.3.1 Descriptive Analysis (Modeling)

Analisis deskriptif dilakukan untuk memberikan gambaran umum terhadap data penjualan video game yang terdapat pada dataset *vgchartz-2024.csv*. Analisis ini bertujuan untuk memahami karakteristik data seperti jumlah penjualan total, sebaran penjualan per wilayah (NA, JP, PAL, dan lainnya), serta variasi penjualan berdasarkan genre, console, dan publisher.

Tahapan analisis dilakukan dengan menggunakan Tableau untuk menampilkan visualisasi data berupa grafik batang, pie chart, boxplot, dan scatter plot. Melalui visualisasi ini dapat diketahui genre dan console dengan penjualan tertinggi, proporsi penjualan antar wilayah, serta hubungan antara *critic score* dengan total penjualan. Hasil analisis deskriptif ini membantu dalam memahami tren dan pola pasar industri game secara global pada tahun 2024.

2.1.4.1 Hitung Ukuran Statistik Deskriptif

Pada tahap ini dihitung ukuran seperti mean, median, dan variance dengan menggunakan fitur *Measure* di Tableau. Nilai mean menunjukkan rata-rata dari data, median menggambarkan nilai tengah, dan variance menunjukkan seberapa besar penyebaran data. Perhitungan ini membantu memahami seberapa merata distribusi data dan apakah terdapat nilai ekstrim yang mempengaruhi hasil.

2.1.4.1.1 Mean, Median, dan Variance, Mode & Correlation

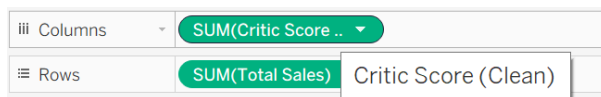
Dengan menghitung ketiga ukuran ini, diperoleh gambaran umum mengenai distribusi data, apakah penjualan game cenderung merata atau terdapat perbedaan yang signifikan antara game dengan penjualan tinggi dan rendah. Hasil analisis statistik deskriptif ini menjadi dasar untuk memahami pola penjualan dan mengidentifikasi kategori game yang berkontribusi besar terhadap total penjualan global.

1. Visualisasi I - Korelasi: Critic Score (Clean) vs Total Sales (Scatter Plot)

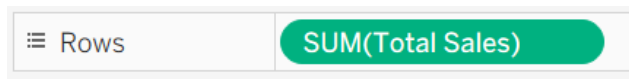
Tujuan: Untuk mengetahui apakah game dengan skor kritik tinggi juga memiliki penjualan tinggi.

Langkah:

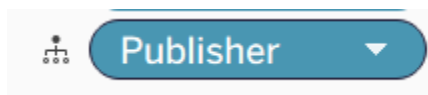
1. Buat *worksheet baru* → beri nama “Korelasi: Critic Score vs Total Sales”.
2. Tarik Critic Score (clean) → ke *Columns*.



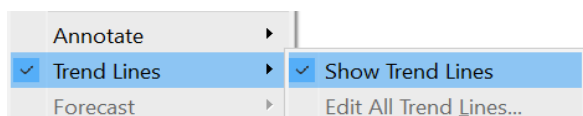
3. Tarik Total Sales → ke *Rows*.



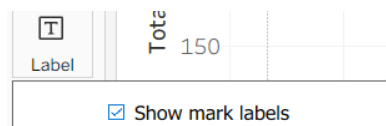
4. Tarik Genre → ke *Color* agar tiap genre punya warna berbeda.
5. Tarik Publisher → ke *Detail* agar data lebih spesifik.



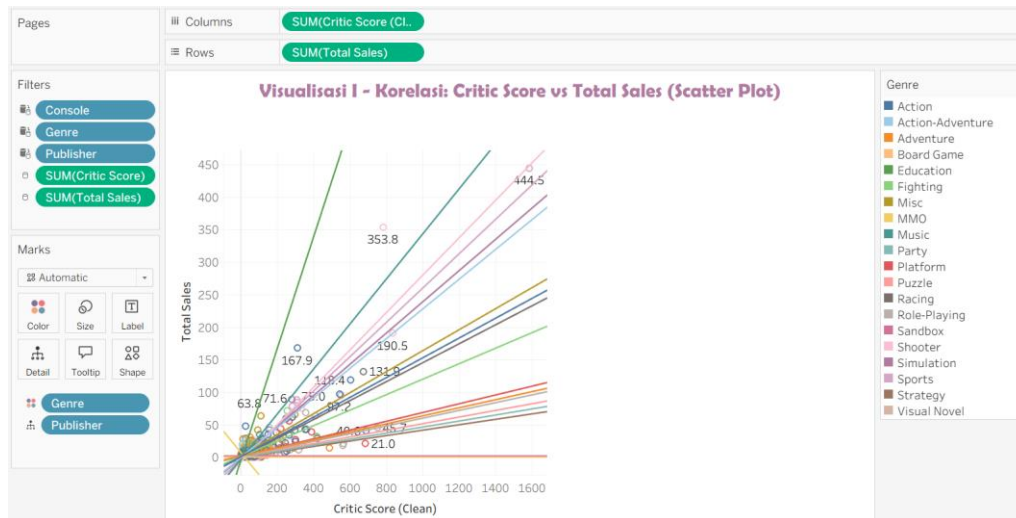
6. Tambahkan *Trend Line* melalui menu kanan → *Add Trend Line*.



7. Tambahkan label *Total Sales* agar mudah membaca nilai.



Hasil:



Alasan pemilihan model visualisasi :

Scatter Plot sangat efektif untuk analisis korelasi karena memberikan gambaran langsung tentang hubungan antar variabel dan membantu dalam pengambilan keputusan berbasis data.

Berdasarkan visualisasi scatter plot "Korelasi: Critic Score(clean) vs Total Sales", terlihat bahwa terdapat korelasi positif antara skor kritik dan total penjualan video game. Titik-titik data tersebar membentuk pola naik, menunjukkan bahwa game dengan skor kritik tinggi cenderung memiliki penjualan yang lebih besar. Beberapa genre seperti *Shooter*, *Role-Playing*, dan *Action* menampilkan titik-titik dengan penjualan sangat tinggi (contoh: 444.5, 353.8, 190.5), yang juga memiliki skor kritik tinggi. Ini mengindikasikan bahwa genre tersebut tidak hanya populer secara gameplay, tetapi juga mendapat penilaian baik dari kritikus.

2. Visualisasi II – Polygon (Mean & Median Total Sales per Genre)

Tujuan: Menunjukkan perbandingan rata-rata dan median penjualan antar genre.

Langkah:

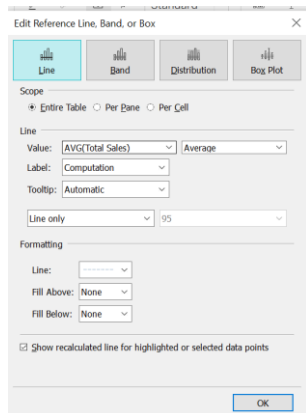
1. Worksheet baru
2. Tarik Genre ke Columns.



3. Tarik Total Sales dua kali ke Rows:
 - Yang pertama ubah jadi Average
 - Yang kedua ubah jadi Median.

☰ Rows	AVG(Total Sales)	MEDIAN(Total Sales)
--------	------------------	---------------------

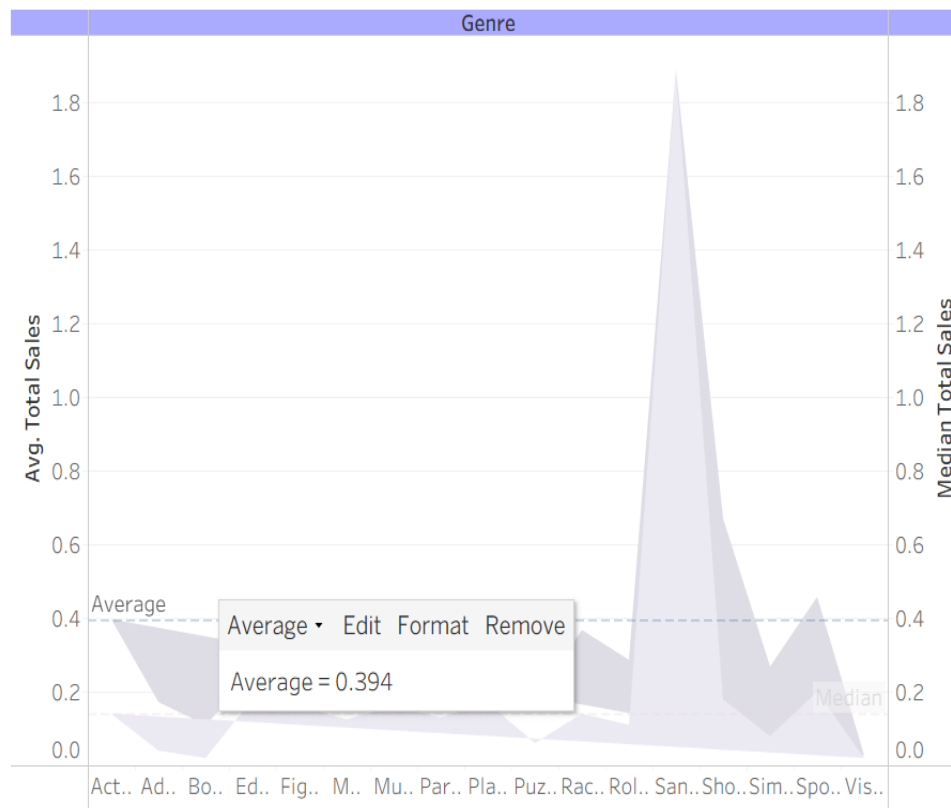
- Gunakan dual axis → klik kanan → *Synchronize Axis*.
- Tambahkan Trend Line untuk Average dan median dengan klik kanan pada sumbu avg dan median → add reference line → show reference → pilih avg dan median.

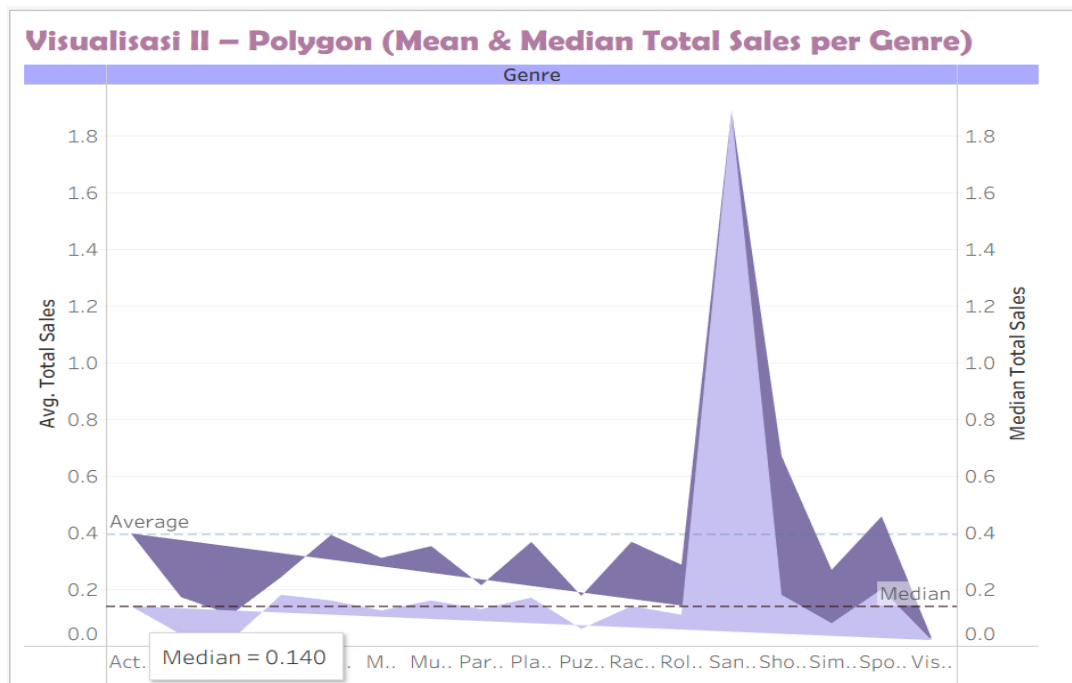


- Tambahkan label rata-rata dan median di atas batang.

Hasil:

Visualisasi II – Polygon (Mean & Median Total Sales per Genre)





Alasan pemilihan model visualisasi :

Polygon Chart sangat efektif untuk analisis penjualan karena menyajikan informasi yang kaya dalam format yang ringkas dan mudah dibaca.

Hasil Deskriptif Analisis Tambahan dari Visualisasi II – Polygon (Mean & Median Total Sales per Genre):

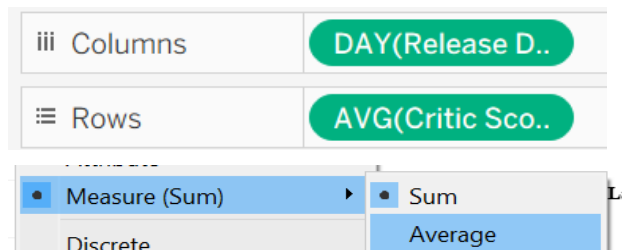
1. Genre Shooter (Sho.) menonjol secara ekstrem, dengan rata-rata dan median penjualan jauh di atas genre lain. Ini menunjukkan bahwa genre ini tidak hanya memiliki beberapa game yang sangat sukses, tetapi juga secara umum stabil dalam performa penjualan.
2. Perbedaan antara Mean dan Median di beberapa genre seperti *Role-Playing*, *Sports*, dan *Action* cukup besar. Ini mengindikasikan adanya outlier—game tertentu yang sangat laku dan menarik rata-rata ke atas, meskipun sebagian besar game dalam genre tersebut memiliki penjualan lebih rendah dari rata-rata.
3. Genre seperti Puzzle (Puz.), Education (Ed.), dan Visual Novel (Vis.) memiliki median yang sangat rendah, bahkan mendekati nol. Ini menunjukkan bahwa mayoritas game dalam genre ini memiliki penjualan yang sangat kecil, meskipun mungkin ada satu atau dua game yang sukses.
4. Garis referensi rata-rata dan median yang ditampilkan di grafik membantu menunjukkan bahwa sebagian besar genre berada di bawah rata-rata total penjualan keseluruhan, menandakan bahwa hanya segelintir genre yang benar-benar mendominasi pasar.

3. Visualisasi III-Rata-Rata skor kritikus (AVG) berdasarkan tanggal rilis game (LineChart)

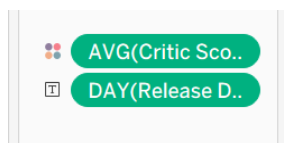
Tujuan: Menunjukkan bagaimana rata-rata skor kritikus berubah seiring waktu, dari tahun 1985 hingga 2018.

Langkah:

1. Buka Tableau Desktop atau Tableau Public.
2. Klik “Connect” → “Text File” dan pilih file vgchartz-2024.csv.
3. Drag Release Date ke Columns → ubah ke DAY(Release Date).
4. Drag Critic Score (Clean) ke Rows → klik kanan → pilih Measure → Average.

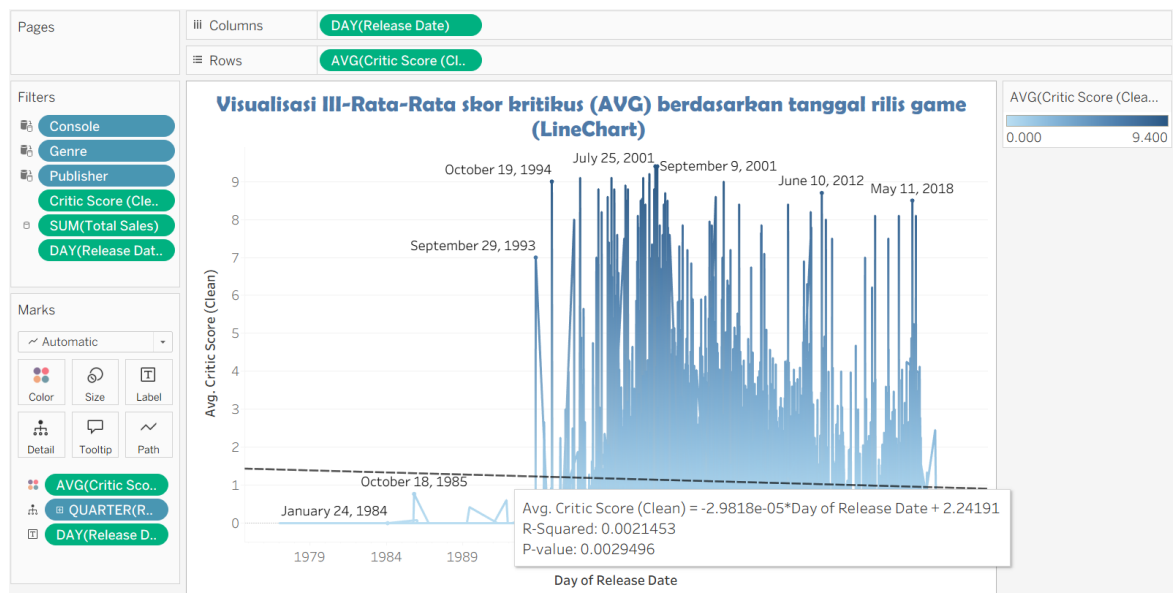


5. Tambahkan AVG kritik Score (clean) ke color, Day(Release date) ke label.



6. Klik kanan pada grafik → Trend Line → Show Trend Line.

Hasil:



Alasan pemilihan model visualisasi :

LineChart sangat cocok untuk analisis mean karena mampu menyoroti perubahan nilai rata-rata secara kronologis dan membantu dalam pengambilan keputusan berbasis waktu.

Hasil analisis deskriptif dari Visualisasi III – Rata-Rata Skor Kritikus (AVG) Berdasarkan Tanggal Rilis Game:

Visualisasi ini menunjukkan tren rata-rata skor kritikus video game berdasarkan tanggal rilisnya. Secara umum, tidak terlihat pola yang konsisten atau peningkatan kualitas dari waktu ke waktu, karena nilai R-squared sangat rendah (0.002), menandakan hubungan yang lemah antara tanggal rilis dan skor kritikus. Beberapa titik data menonjol di tahun-tahun tertentu, namun secara keseluruhan, skor kritikus cenderung fluktuatif dan tidak dipengaruhi secara signifikan oleh waktu rilis. Hal ini mengindikasikan bahwa kualitas game lebih dipengaruhi oleh faktor lain seperti genre, publisher, atau platform, bukan semata-mata oleh era perilisannya.

4. Visualisasi IV - Mode (Genre/Console Paling Sering Muncul) degan TreeMap

Tujuan: Memvisualisasikan popularitas genre secara visual dengan ukuran blok (size) proporsional dengan jumlah game, serta menampilkan kontribusi Publisher melalui skala warna (dari terang ke gelap, mewakili jumlah Publisher).

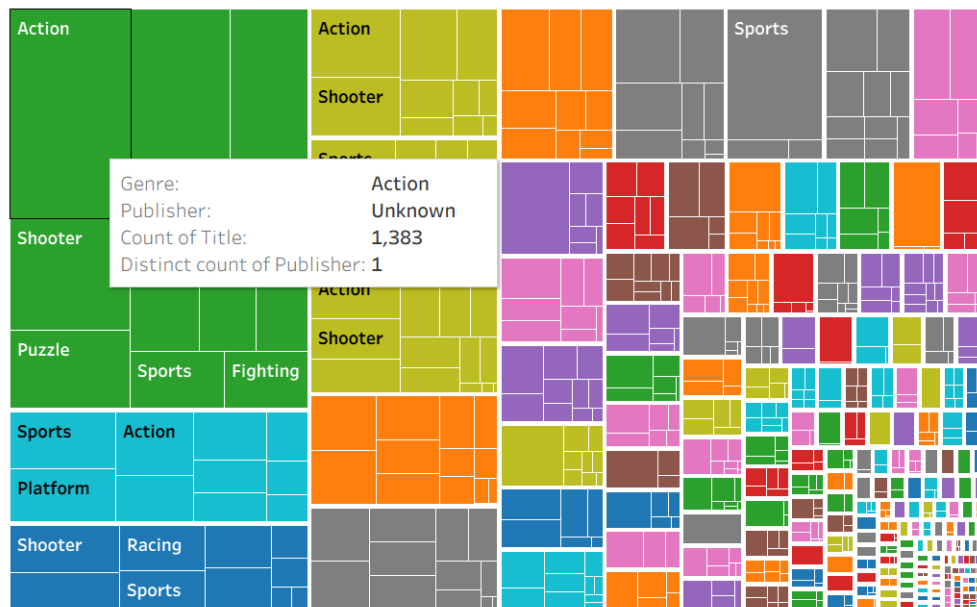
Langkah:

1. Buat Worksheet Baru
2. Buka Tableau → klik “New Worksheet” → Masukkan Data
3. Pastikan dataset sudah di-import ke Tableau (misal: Video Games Sales 2024).
4. Drag dan Drop Fields
 - Columns / Rows: Kosongkan (tidak diperlukan untuk treemap).
 - Marks → Type: Ubah ke Treemap.
5. Atur Ukuran dan Label
 - Size: Tarik COUNT(Title) → ini menentukan besar kotak berdasarkan jumlah game per genre.
 - Label: Tarik Genre → tampilkan nama genre di tiap kotak.
6. Atur Detail & Tooltip
 - Tarik Publisher ke Color → skala warna menunjukkan jumlah Publisher per genre (gelap = banyak Publisher).

- Tarik Publisher juga ke Detail → untuk info tambahan saat hover.
- Tambahkan COUNT(Publisher) atau CNTD(Publisher) ke Tooltip untuk detail jumlah publisher unik.

Hasil:

Visualisasi IV - Mode (Genre/Console Paling Sering Muncul)



Alasan pemilihan model visualisasi :

Treemap dipilih karena mampu menampilkan banyak kategori sekaligus secara ringkas dan memudahkan identifikasi kategori dominan.

Hasil analisis deskriptif dari Visualisasi IV – Mode (Genre/Console Paling Sering Muncul):

Treemap ini secara visual menampilkan popularitas genre video game berdasarkan jumlah judul yang dirilis, serta kontribusi publisher dalam setiap genre. Genre seperti *Action*, *Shooter*, dan *Sports* memiliki kotak terbesar, menandakan bahwa mereka adalah genre yang paling sering muncul dalam dataset baik dari segi jumlah game maupun publisher yang terlibat. Warna pada kotak menunjukkan jumlah publisher unik dalam setiap genre: semakin gelap warnanya, semakin banyak publisher yang berkontribusi.

Sebagai contoh, genre *Action* tidak hanya memiliki jumlah game terbanyak (contoh: 1.383 judul), tetapi juga didominasi oleh satu publisher besar, seperti yang terlihat dari tooltip “Publisher: Unknown” dengan hanya satu publisher yang menguasai genre tersebut. Sebaliknya, genre seperti *Puzzle*,

Platform, dan *Racing* memiliki ukuran kotak yang lebih kecil dan warna yang lebih terang, menandakan bahwa genre tersebut kurang populer dan memiliki keterlibatan publisher yang lebih terbatas.

5. Visualisasi V - Variance Nilai Kritik (Critic Score) per Tahun Rilis dengan BarChart

Tujuan: Menunjukkan seberapa bervariasi (variance) skor ulasan kritikus dari tahun ke tahun.

Langkah:

1. Buka dataset di Tableau.
 - Release Date → ubah ke Date atau Year (klik kanan → *Change Data Type* → *Date* → lalu buat Year(Release Date) di shelf).
 - Critic Score → Measure (angka).
2. Seret Release Date (Year) ke Columns.

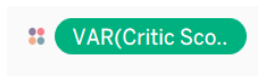


3. Seret Critic Score ke Rows.
4. Klik panah kecil di SUM(Critic Score) → pilih Measure → Variance.

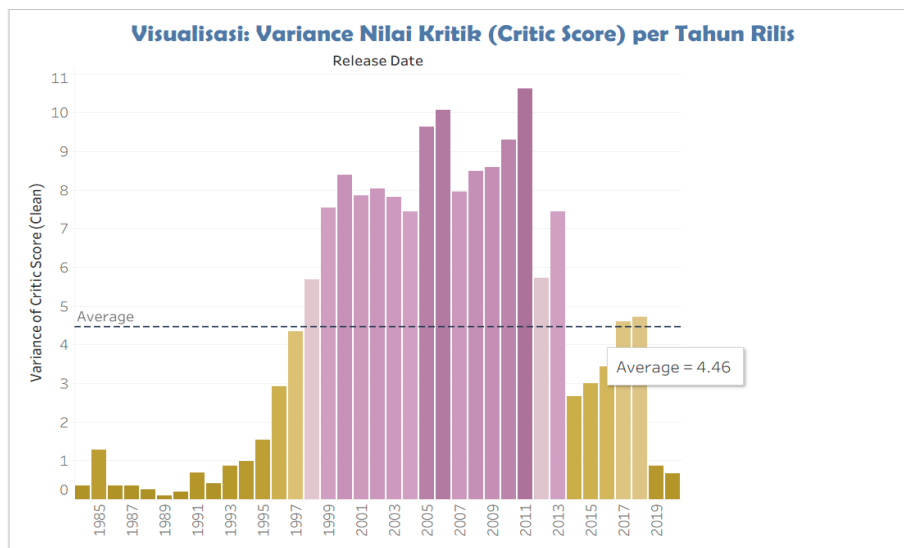


Tableau akan menampilkan grafik batang yang menunjukkan variance skor kritik per tahun.

5. Seret Variance(Critic Score) ke Color → warna makin pekat berarti variasinya makin besar.



Hasil:



Alasan pemilihan model visualisasi :

BarChart dipilih karena mampu menampilkan perbandingan antar tahun secara jelas dan langsung. Dengan tambahan garis rata-rata dan pewarnaan kategori, visualisasi ini memudahkan identifikasi tahun-tahun dengan penyebaran skor yang ekstrem dan membantu dalam analisis tren kualitas game dari perspektif kritikus.

Hasil analisis deskriptif dari *Visualisasi V - Variance Nilai Kritik (Critic Score) per Tahun Rilis*:

Visualisasi ini menunjukkan tingkat variasi skor ulasan kritikus dari tahun ke tahun, mulai dari 1985 hingga 2017. Variance yang tinggi berarti terdapat perbedaan besar antar skor kritik dalam satu tahun menandakan bahwa kualitas game yang dirilis di tahun tersebut sangat beragam, dari yang sangat baik hingga yang sangat buruk. Sebaliknya, variance yang rendah menunjukkan bahwa skor kritik cenderung seragam, artinya game yang dirilis di tahun tersebut memiliki kualitas yang relatif konsisten. Dari grafik, terlihat bahwa tahun-tahun seperti 1998, 2006, dan 2010 memiliki variance tinggi, yang bisa mengindikasikan bahwa banyak game dengan kualitas ekstrem dirilis pada periode tersebut. Sementara tahun-tahun seperti 1985 dan 1993 memiliki variance rendah, menunjukkan bahwa game yang dirilis saat itu mendapat penilaian yang lebih seragam dari para kritikus.

2.1.3.2 Insight Interpretation

Tahap Insight Interpretation merupakan bagian penting dalam proses analisis data, dimana hasil eksplorasi dan visualisasi diinterpretasikan untuk menemukan makna yang relevan bagi kebutuhan bisnis. Tujuan dari tahap ini adalah memberikan pemahaman yang jelas, berbasis data, serta menghasilkan rekomendasi yang dapat mendukung pengambilan keputusan strategis dalam industri video game. Berdasarkan hasil analisis deskriptif dan visualisasi terhadap dataset Video Game Sales 2024, diperoleh insight penting yang menggambarkan tren penjualan, preferensi pasar, serta faktor-faktor utama yang mempengaruhi performa penjualan video game secara global.

1. Korelasi: Critic Score vs Total Sales (Scatter Plot)

Visualisasi ini digunakan untuk melihat hubungan antara skor ulasan (Critic Score) dengan total penjualan (Total Sales) game. Berdasarkan hasil visualisasi, terlihat adanya korelasi positif antara skor kritik dan total penjualan, yang berarti semakin tinggi skor yang diberikan kritikus, semakin besar pula angka penjualan game tersebut. Beberapa genre seperti Shooter, Role-Playing, dan Action menonjol karena memiliki nilai penjualan yang tinggi sekaligus skor kritik yang baik.

Insight:

Temuan ini menunjukkan bahwa kualitas game sangat berpengaruh terhadap tingkat penjualan. Ulasan positif dari kritikus dapat meningkatkan kepercayaan konsumen dan memperkuat minat beli. Oleh

karena itu, perusahaan perlu menjaga kualitas produk dan fokus pada peningkatan pengalaman bermain agar dapat meningkatkan skor ulasan serta penjualan.

2. Mean dan Median Total Sales per Genre

Visualisasi ini bertujuan untuk membandingkan nilai rata-rata (mean) dan median penjualan antar genre game. Hasil analisis menunjukkan bahwa genre Shooter memiliki rata-rata dan median penjualan tertinggi dibandingkan genre lainnya. Hal ini mengindikasikan bahwa genre tersebut memiliki performa penjualan yang stabil dan sangat diminati di pasar. Sementara itu, genre seperti Role-Playing, Sports, dan Action menunjukkan perbedaan yang cukup besar antara nilai rata-rata dan median, yang menandakan adanya outlier, yaitu beberapa game dengan penjualan sangat tinggi yang memengaruhi nilai rata-rata secara keseluruhan. Sebaliknya, genre seperti Puzzle, Education, dan Visual Novel memiliki median penjualan yang rendah, bahkan mendekati nol, sehingga sebagian besar game di kategori tersebut memiliki penjualan yang kecil di pasaran.

Insight:

Hasil ini menunjukkan bahwa genre Shooter dan Action merupakan kategori paling potensial secara komersial, karena memiliki daya tarik besar dan penjualan yang stabil. Dengan demikian, pengembang atau penerbit game dapat memprioritaskan genre ini untuk meningkatkan pendapatan dan meminimalkan risiko kegagalan produk.

3. Rata-Rata Skor Kritikus Berdasarkan Tahun Rilis (Line Chart)

Visualisasi ini digunakan untuk melihat bagaimana perubahan rata-rata skor ulasan kritikus berkembang dari tahun 1985 hingga 2018. Berdasarkan hasil analisis, rata-rata skor kritikus menunjukkan tren penurunan dari waktu ke waktu. Pada periode 1985–2005, sebagian besar game memperoleh skor tinggi di atas 80, namun setelah tahun 2010 terjadi penurunan ke kisaran 60–70, bahkan mendekati 50 menjelang 2018. Penurunan ini dapat disebabkan oleh meningkatnya jumlah game yang dirilis dengan kualitas beragam, perubahan standar penilaian dari kritikus, serta banyaknya variasi genre dan platform yang muncul.

Insight:

Penurunan rata-rata skor kritikus ini menandakan bahwa kualitas game secara keseluruhan mengalami penurunan dalam dua dekade terakhir. Hal ini menjadi peringatan bagi pengembang untuk terus berinovasi dan meningkatkan standar kualitas agar tetap kompetitif di tengah persaingan industri yang semakin ketat.

4. Mode (Genre atau Konsol yang Paling Sering Muncul)

Visualisasi Treemap digunakan untuk melihat genre dan konsol yang paling sering muncul, serta seberapa besar keterlibatan publisher pada masing-masing kategori. Hasilnya menunjukkan bahwa genre Action, Shooter, dan Sports memiliki ukuran kotak paling besar, yang berarti ketiganya adalah genre paling populer dan paling sering diproduksi. Warna yang lebih gelap menunjukkan jumlah publisher yang lebih banyak dalam genre tersebut, menandakan bahwa publisher besar cenderung fokus pada genre-genre dengan potensi keuntungan tinggi. Sementara itu, genre seperti Puzzle, Platform, dan Racing memiliki ukuran dan warna yang lebih kecil, menunjukkan bahwa genre tersebut memiliki pasar yang lebih sempit dan keterlibatan publisher yang terbatas.

Insight:

Temuan ini menggambarkan bahwa publisher besar mendominasi genre populer dengan pasar luas, sementara masih ada peluang bagi pengembang baru untuk menembus pasar dengan kompetisi yang lebih rendah. Dengan strategi yang tepat, genre-genre kecil dapat menjadi area ekspansi baru yang menguntungkan di masa depan.

5. Variasi Nilai Kritik Berdasarkan Tahun Rilis

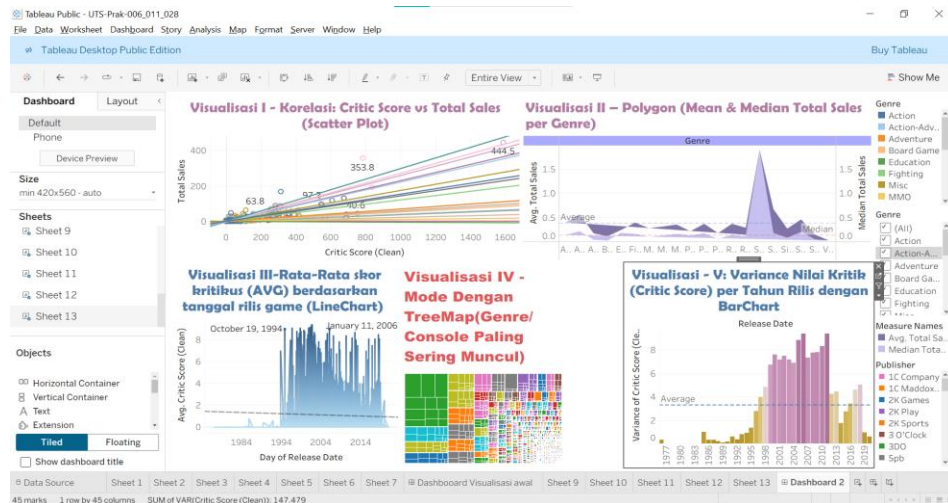
Visualisasi ini menunjukkan tingkat variasi (variance) dari skor ulasan kritikus terhadap game berdasarkan tahun rilis. Hasilnya memperlihatkan bahwa tahun 1998, 2006, dan 2010 memiliki variasi skor tertinggi, yang berarti pada tahun-tahun tersebut terdapat perbedaan kualitas yang signifikan antara game yang dirilis. Sebaliknya, tahun-tahun seperti 1985 dan 1993 memiliki nilai variasi rendah, menandakan bahwa kualitas game pada periode tersebut relatif seragam.

Insight:

Hasil ini menunjukkan bahwa periode dengan variasi skor kritik yang tinggi mencerminkan tingginya tingkat kompetisi dan keberagaman kualitas di pasar game. Perusahaan dapat memanfaatkan informasi ini untuk menentukan waktu peluncuran yang tepat, menghindari periode dengan persaingan kualitas yang ketat, serta merilis produk di waktu yang lebih strategis untuk mendapatkan perhatian pasar.

2.1.3.2 Deployment

Tahap Deployment merupakan tahap akhir dalam metodologi CRISP-DM, di mana hasil analisis data yang telah dilakukan diterapkan dalam bentuk dashboard interaktif menggunakan *Tableau Public*. Dashboard ini dirancang untuk menyajikan hasil analisis deskriptif secara visual, interaktif, dan mudah dipahami oleh pengguna atau pihak pengambil keputusan.



Dashboard berisi lima visualisasi utama, yaitu:

- Visualisasi I - Korelasi: Critic Score vs Total Sales (Scatter Plot)**

Menunjukkan hubungan antara skor kritik dan total penjualan. Terlihat korelasi positif di mana game dengan skor kritik tinggi cenderung memiliki penjualan yang lebih besar.

- Visualisasi II - Polygon (Mean & Median Total Sales per Genre)**

Menampilkan perbandingan nilai rata-rata dan median penjualan pada setiap genre. Hasilnya menunjukkan bahwa genre *Shooter* memiliki performa penjualan paling tinggi dan stabil dibandingkan genre lainnya.

- Visualisasi III - AVG Critic Score Over Time (Line Chart)**

Menggambarkan tren rata-rata skor kritikus dari tahun ke tahun. Ditemukan bahwa skor kritikus cenderung menurun sejak 2010, menandakan variasi kualitas game yang semakin luas.

- Visualisasi IV - Mode dengan Treemap (Genre/Console Paling Sering Muncul)**

Menampilkan genre dan platform yang paling sering muncul dalam dataset. Genre *Action*, *Shooter*, dan *Sports* terlihat mendominasi pasar, menunjukkan tren popularitas yang tinggi.

- Visualisasi V -Variance Nilai Kritik per Tahun Rilis (Bar Chart)**

Menggambarkan seberapa besar variasi penilaian kritikus setiap tahun. Tahun 1998, 2006, dan 2010 menunjukkan variance tertinggi, menandakan adanya perbedaan besar dalam kualitas game yang dirilis pada periode tersebut.

Semua visualisasi tersebut digabung dalam satu dashboard interaktif yang memungkinkan pengguna melakukan eksplorasi data secara dinamis, seperti memfilter berdasarkan *genre*, *publisher*, atau *tahun rilis*. Dashboard ini kemudian dipublikasikan di Tableau Public sebagai media penyajian hasil analisis secara terbuka dan transparan.

DAFTAR PUSTAKA

- [1] I. Wowczko, "A Case Study of Evaluating Job Readiness with Data Mining Tools and CRISP-DM Methodology," *Int. J. Infonomics*, vol. 8, no. 3, pp. 1066–1070, 2015, doi: 10.20533/iji.1742.4712.2015.0126.
- [2] A. Parate, "Integrating Crisp DM Methodology for a Business Using Tableau Visualization," *Researchgate*, no. July, 2020, doi: 10.13140/RG.2.2.36619.31520.
- [3] H. Marastika and S. Nur, "Crisp-Dm and Tableau Visualization in Sentiment Analysis of the," *J. Inf. Technol. Comput. Sci.*, vol. 8, pp. 191–200, 2025.
- [4] M. F. Mustapha, A. N. I. Zulkifli, O. Kairan, N. N. S. M. Zizi, N. N. Yahya, and N. M. Mohamad, "The prediction of student's academic performance using RapidMiner," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 32, no. 1, pp. 363–371, 2023, doi: 10.11591/ijeecs.v32.i1.pp363-371.
- [5] Y. lucky Sukmana, "Implementasi Crisp-Dm Dengan Menggunakan Tableau Untuk Visualisasi Data Dalam Upaya Peningkatan Penerimaan Impor Pada Kpu Bea Cukai Tipe B Batam," *J. Perspekt. Bea Dan Cukai*, vol. 7, no. 1, pp. 19–38, 2023, doi: 10.31092/jpbc.v7i1.2010.