Nama : Brian Filbert

NIM : 2802504844

Final Project UAS Data Analytics

---

1. **Review & Evaluation of Data Quality**
   a. Summarize the main data quality issues identified during the Mid Project (UTS).

      Pada saat UTS, terdapat beberapa quality issues yaitu:

      - Missing values pada Publisher
      - Missing values pada Year
      - Outlier pada game dengan penjualan sangat tinggi seperti Wii Sports

Kolom	**Missing**
Year	271
Publisher	58

      Selain itu ditemukan:

      - Distribusi Global Sales sangat skewed (skewness = 17.32)
      - Terdapat 1,827 outlier (11.19% data)
      - Tidak ditemukan duplicate records.

   b. Evaluate how effective your data wrangling process was in improving data quality.

      Missing Year:

```
[115]  df = df.dropna(subset=['Year'])
```

      271 data dihapus

      Missing Publisher:

```
[116]  df['Publisher'] = df['Publisher'].fillna('Unknown')
```

      58 data diperbaiki

      Type Conversion:

```
[120]  df['Year'] = df['Year'].astype('Int64')
```

      Validasi Konsistensi untuk Memastikan bahwa

NA + EU + JP + Other = Global

```
df['check_global'] = (
    df['NA_Sales'] +
    df['EU_Sales'] +
    df['JP_Sales'] +
    df['Other_Sales']
)

(df['check_global'] - df['Global_Sales']).abs().max()
```

[113]

```
np.float64(0.020000000000000462)
```

Selisih maksimum hanya: 0.02

c. Indikator Keberhasilan

Metric	**Before**	**After**
Missing Year	271	0
Missing Publisher	58	0
Duplicate Rows	0	0
Data Consistency Error	Max 0.02	Validated

Data wrangling berhasil menghilangkan seluruh missing values dan meningkatkan konsistensi data sehingga dataset siap digunakan untuk advanced analysis.

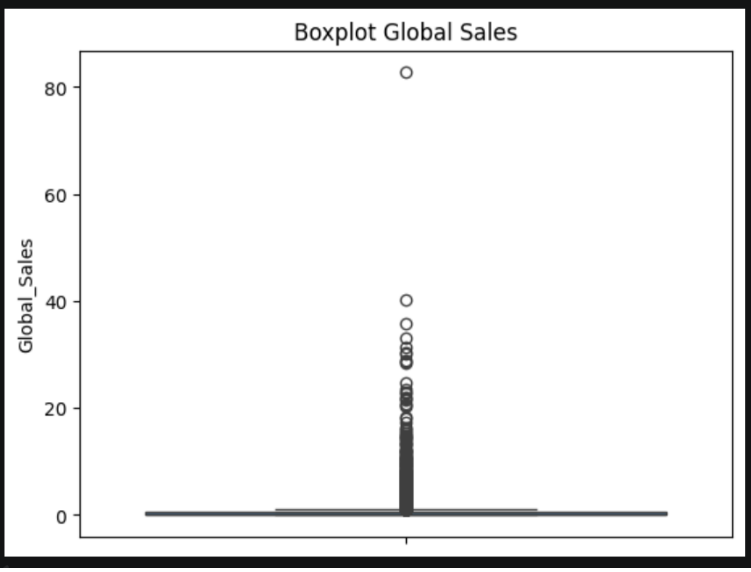
2. Advanced Pattern, Trend, Relationship Analysis

a. Advanced Analysis

I) Global Sales Analysis

```
sns.boxplot(df_eda['Global_Sales'])
plt.title('Boxplot Global Sales')
plt.show()
```

[21] ✓ 0.1s



```
skewness = df_eda['Global_Sales'].skew()

print(f"Skewness: {skewness:.2f}")
```

[22] ✓ 0.0s

```
Skewness: 17.32
```

```
Q1 = df_eda['Global_Sales'].quantile(0.25)
Q3 = df_eda['Global_Sales'].quantile(0.75)
IQR = Q3 - Q1

outliers = df_eda[(df_eda['Global_Sales'] < Q1 - 1.5*IQR) |
                  (df_eda['Global_Sales'] > Q3 + 1.5*IQR)]

len(outliers)
```

[23] ✓ 0.0s

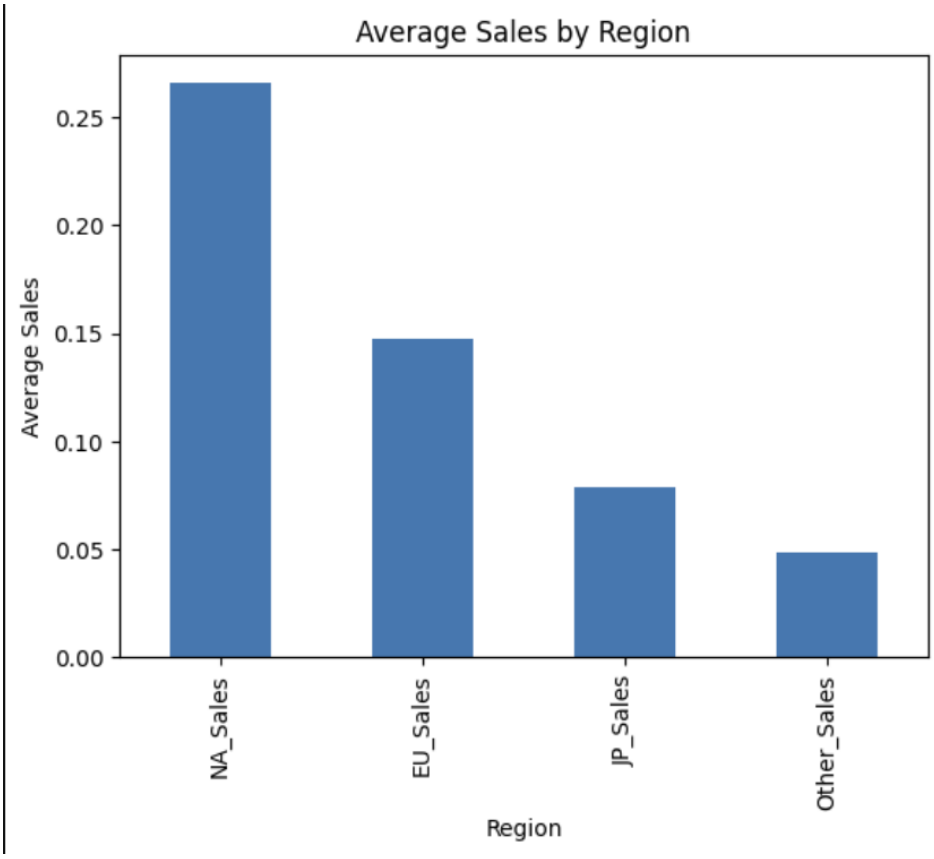
```
1827
```

```
total = len(df_eda)
outlier_count = len(outliers)

print("Outliers:", outlier_count)
print("Percentage: ", (outlier_count / total) * 100)
```

[24] ✓ 0.0s

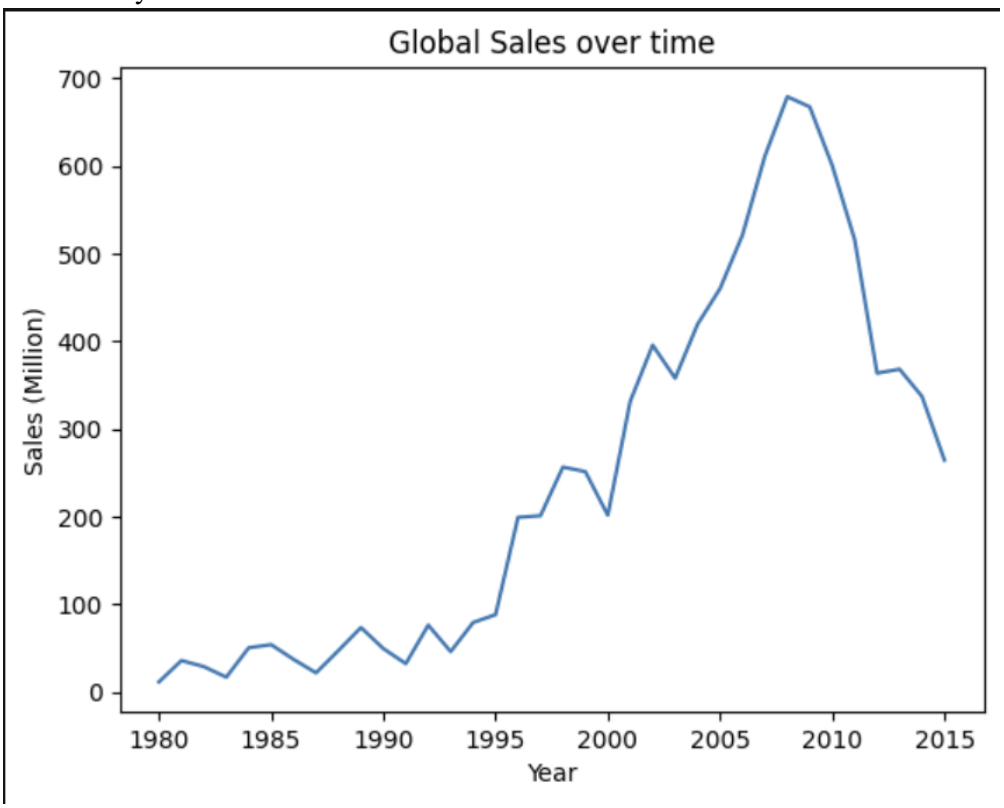
```
Outliers: 1827
Percentage:  11.190053285968029
```

Dari Boxplot keliatan bahwa data tidak simetris, banyak nilai kecil dan ada nilai yang sangat besar, jadinya outlier banyak karena distribusi skewed, kemungkinan karena 90% game merupakan game yang peminatnya lebih kurang daripada 10% game seperti GTA, FIFA, Mario, dll.



Dapat dilihat bahwa NA_Sales mendominasi penjualan dibandingkan sales lainnya.

II) Trend Analysis

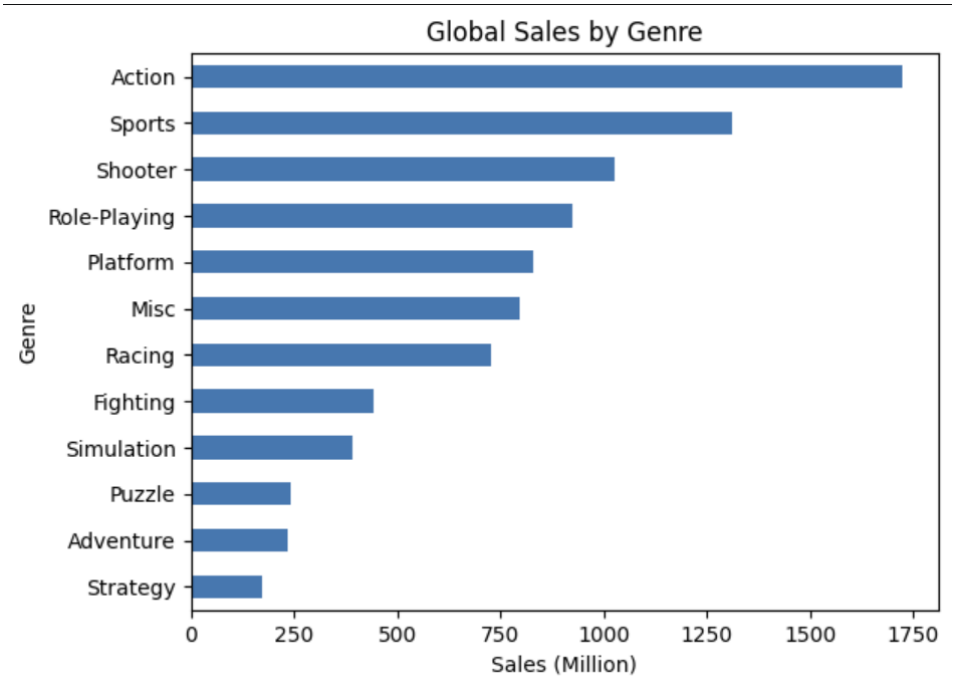


Dapat dilihat bahwa sales naik sampai tahun 2008, yang kemudian turun setelah sekitar

2009, hal ini juga karena kemungkinan disebabkan oleh rilisnya PS2, Xbox 360, PS3 dna Wii di market, yang di mana PS2 rilis pada tahun 2000, Xbox 360 pada tahun 2005, kemudian PS3 dan Wii yang rilis pada tahun 2006.

III) Genre vs Global Sales Analysis

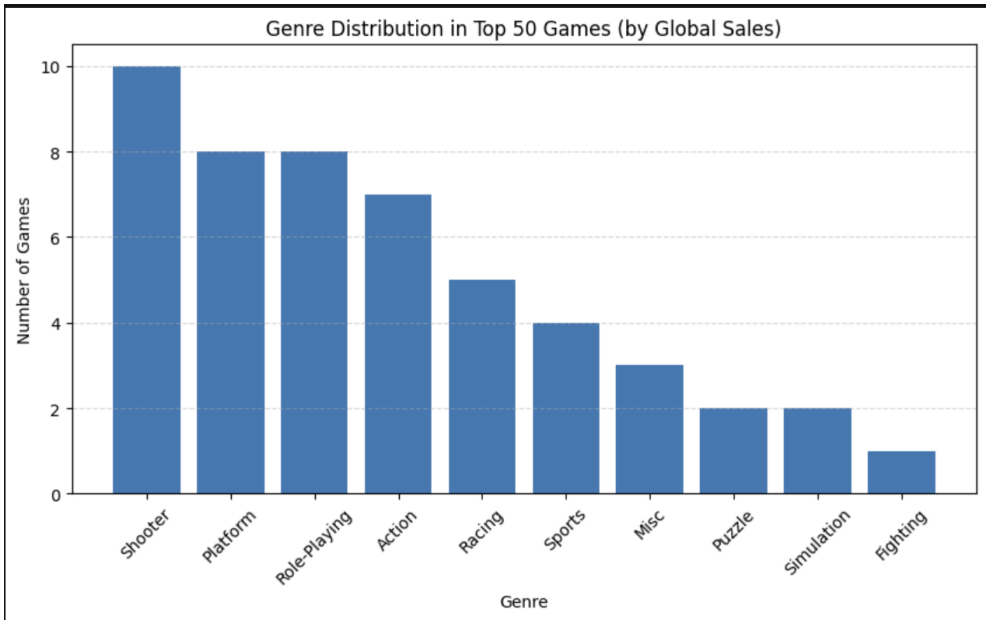
```
genre_sales = df_eda.groupby('Genre')['Global_Sales'].sum().sort_values()

genre_sales.plot(kind='barh')
plt.title('Global Sales by Genre')
plt.xlabel('Sales (Million)')
plt.ylabel('Genre')
plt.show()
```



Insight pertama yang di dapat dari ini adalah Genre Action menghasilkan total penjualan global tertinggi.

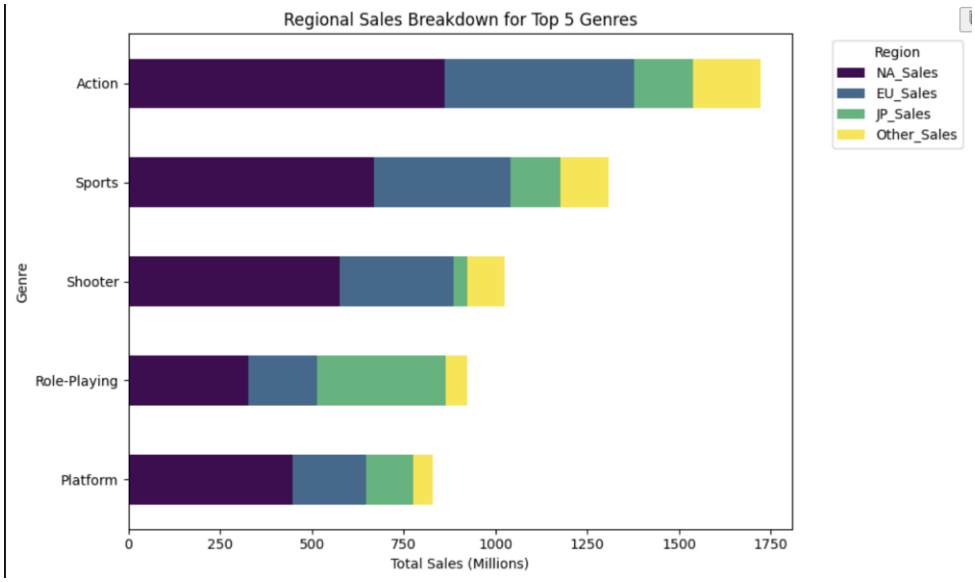
Namun, hasil ini perlu diinterpretasikan lagi karena distribusi Global Sales sangat right-skewed (skewness = 17.32), dan dari analisis, dapat dilihat bahwa Action memang tinggi dikarenakan jumlah banyak gamenya.

```
top_games = df_eda.nlargest(50, 'Global_Sales')
genre_counts = top_games['Genre'].value_counts().reset_index()
genre_counts.columns = ['Genre', 'Count']

plt.figure(figsize=(10,5))

plt.bar(genre_counts['Genre'], genre_counts['Count'])

plt.title('Genre Distribution in Top 50 Games (by Global Sales)')
plt.xlabel('Genre')
plt.ylabel('Number of Games')

plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--', alpha=0.5)

plt.show()
```



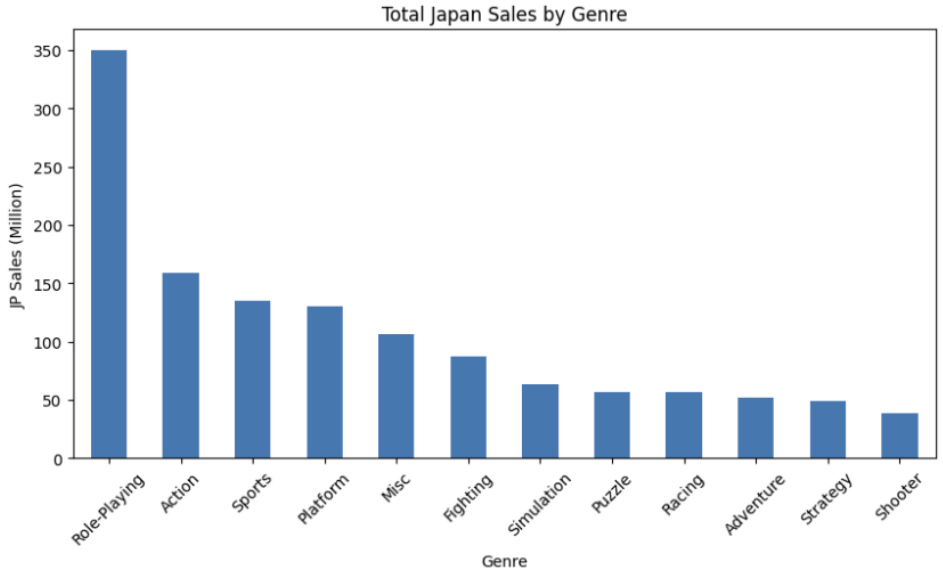
Dapat dilihat bahwa Genre Shooter merupakan Genre dengan Tingkat sales yang banyak terutama di Top 50 sebagai blockbuster game.

IV) Regional Analysis

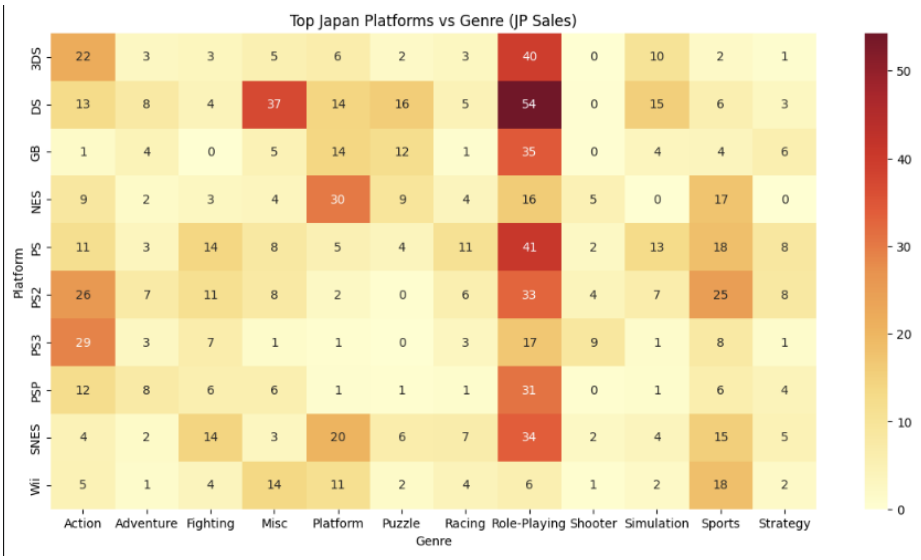


Disini juga dapat dilihat bahwa selain preference yang standard, JP_Sales memliki Tingkat penjualan Role-Playing yang tinggi, dan memiliki sales yang lebih tinggi di RPG ketimbang Sales lainnya.

i) JP_Sales Deeper Analysis



Grafik Japan Sales by Genre menunjukkan bahwa Role-Playing memiliki total penjualan tertinggi di Jepang dibanding genre lain. Hal ini menunjukkan bahwa pasar Jepang memiliki preferensi yang kuat terhadap genre RPG.

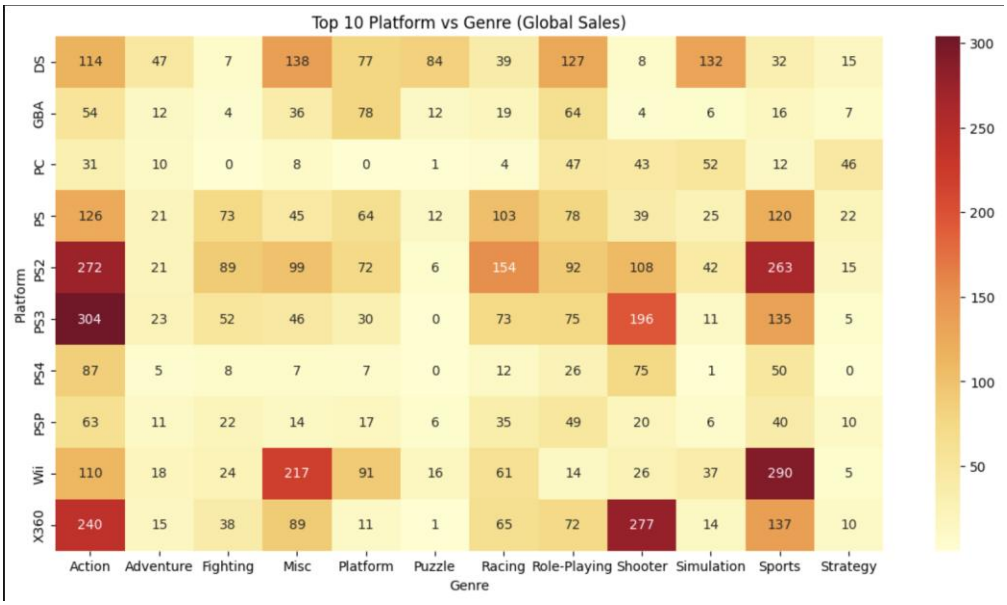


Heatmap Platform vs Genre berdasarkan JP Sales menunjukkan bahwa penjualan RPG tinggi pada banyak platform, terutama DS, PS, 3DS, GB, SNES, PS2, dan PSP. Hal ini menunjukkan bahwa tingginya penjualan RPG di Jepang tidak hanya berasal dari satu platform tertentu, tetapi didukung oleh ekosistem RPG yang kuat pada platform Nintendo dan PlayStation.

V) Platform Analysis

Dari hasil yang didapat diatas, dapat kita ketahui bahwa PS2 jadi yang paling laku sepanjang masa.

Selain itu, terlihat juga bahwa Platform baru belum tuntu unggul, seperti yang dilihat, PS4 lebih rendah dari PS2 dan PS3, yang dimana kemungkinan besarnya adalah antara Datasetnya belum mencakup tahun terbaru sepenuhnya atau waktu rilis lebih baru membuat data belum akumulatif.

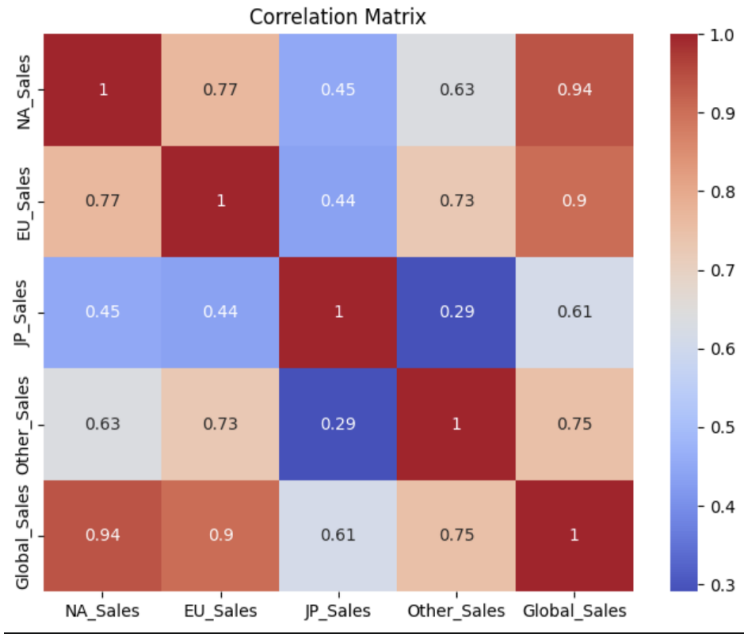


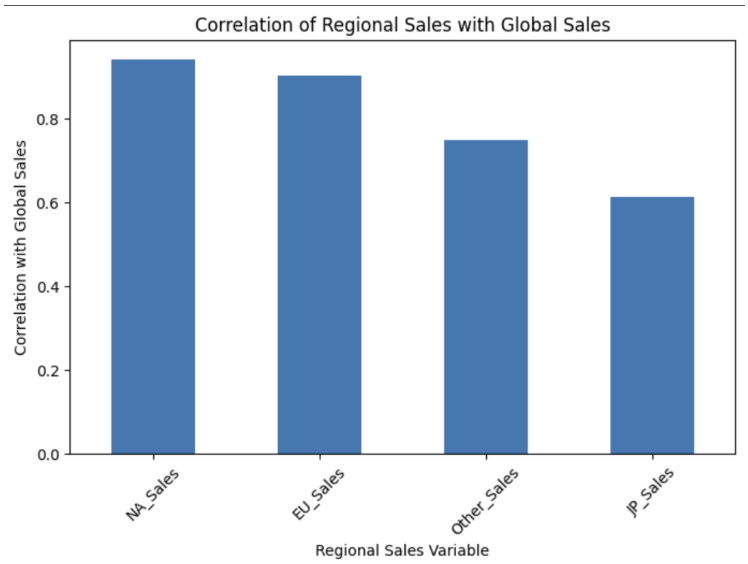
Jika dibandingkan dengan heatmap global, penjualan global lebih banyak didominasi oleh kombinasi Action, Shooter, dan Sports, seperti PS3-Action, X360-Shooter, dan Wii-Sports. Sebaliknya, heatmap khusus Jepang menunjukkan konsentrasi yang lebih

kuat pada Role-Playing di berbagai platform. Hal ini memperkuat bahwa Jepang memiliki pola preferensi genre yang berbeda dari pasar global.

b. Identify the most influential variables related to the problem context.

I) Correlation Matrix





Berdasarkan hasil analisis, variabel yang paling berpengaruh terhadap penjualan video game adalah Region Sales, Genre, dan Platform. Region Sales berpengaruh karena correlation matrix menunjukkan bahwa NA_Sales memiliki hubungan paling kuat dengan Global_Sales sebesar 0.941, diikuti oleh EU_Sales, Other_Sales, dan JP_Sales. Hal ini menunjukkan bahwa kontribusi penjualan dari setiap wilayah memiliki hubungan yang kuat terhadap performa penjualan global.

Selain Region Sales, Genre juga menjadi variabel penting karena performa penjualan berbeda antar genre dan antar wilayah. Contohnya, genre Role-Playing memiliki JP_Sales tertinggi di Jepang, sedangkan genre lain seperti Action, Shooter, dan Sports

lebih dominan pada pasar global. Hal ini menunjukkan bahwa preferensi konsumen berbeda berdasarkan region.

Platform juga menjadi variabel yang berpengaruh karena heatmap Platform vs Genre menunjukkan bahwa penjualan RPG di Jepang banyak didukung oleh platform Nintendo dan PlayStation. Dengan demikian, performa suatu genre tidak hanya dipengaruhi oleh genre itu sendiri, tetapi juga oleh platform tempat game tersebut dirilis.

c. Explain why these relationships are meaningful using analytical evidence

Hubungan antara Region, Genre, dan Platform bermakna karena ketiganya menunjukkan bahwa penjualan video game tidak dapat dijelaskan hanya dari total Global_Sales saja. Berdasarkan hasil correlation matrix, NA_Sales memiliki hubungan paling kuat dengan Global_Sales sebesar 0.941, sehingga kontribusi wilayah memiliki peran penting terhadap performa penjualan global. Namun, hubungan ini tetap perlu diinterpretasikan secara hati-hati karena Global_Sales merupakan hasil penjumlahan dari seluruh regional sales.

Selain itu, hubungan antara Genre dan Region juga penting karena setiap wilayah memiliki preferensi genre yang berbeda. Analisis JP_Sales by Genre menunjukkan bahwa Role-Playing merupakan genre dengan penjualan tertinggi di Jepang, sedangkan pasar global lebih banyak didominasi oleh genre seperti Action, Sports, dan Shooter. Hal ini menunjukkan bahwa strategi pasar tidak dapat disamaratakan untuk semua region.

Hubungan antara Platform dan Genre juga bermakna karena heatmap Platform vs Genre menunjukkan bahwa performa suatu genre dapat dipengaruhi oleh platform tempat game tersebut dirilis. Pada pasar Jepang, RPG memiliki penjualan tinggi pada beberapa platform seperti DS, PS, 3DS, GB, SNES, PS2, dan PSP. Hal ini menunjukkan bahwa tingginya penjualan RPG di Jepang tidak hanya disebabkan oleh preferensi genre, tetapi juga didukung oleh ekosistem platform Nintendo dan PlayStation.

Dengan demikian, hubungan Region, Genre, dan Platform penting karena dapat membantu publisher dan developer membuat keputusan yang lebih spesifik. Misalnya, RPG dapat lebih diprioritaskan untuk pasar Jepang dan platform yang memiliki basis RPG kuat, sedangkan strategi untuk pasar global atau wilayah lain dapat menyesuaikan genre dan platform yang lebih sesuai dengan preferensi masing-masing pasar.

3. Evaluation of Analytical Methods and Metrics

a. Describe the analytical techniques or models used in your project (e.g., correlation analysis, statistical comparison, simple predictive or scoring approach if applicable).

Dalam project ini, beberapa teknik analisis digunakan untuk mengevaluasi pola penjualan video game. Teknik pertama adalah descriptive analysis, yang digunakan untuk merangkum dataset melalui total sales, average sales, median sales, distribusi sales, dan jumlah game pada setiap kategori. Teknik ini membantu mengidentifikasi pola umum seperti genre, platform, dan region yang paling dominan.

Teknik kedua adalah correlation analysis, yang digunakan untuk melihat hubungan antara variabel regional sales dan Global_Sales. Correlation matrix digunakan untuk mengidentifikasi region sales mana yang memiliki hubungan paling kuat dengan penjualan global.

Teknik ketiga adalah comparative analysis, yang digunakan untuk membandingkan performa penjualan berdasarkan genre, platform, dan region. Contohnya adalah analisis JP_Sales by Genre, Platform vs Genre, dan kontribusi sales tiap region. Teknik ini membantu mengidentifikasi perbedaan preferensi pasar, seperti kuatnya performa Role-Playing games di Jepang.

Teknik keempat adalah hypothesis testing, yang digunakan untuk mengevaluasi secara statistik apakah perbedaan performa penjualan signifikan. Independent T-Test digunakan untuk membandingkan rata-rata Global_Sales antara genre Action dan Sports, sedangkan One-Way ANOVA digunakan untuk menguji apakah rata-rata Global_Sales berbeda secara signifikan antar semua genre.

Teknik kelima adalah simple linear regression, yang digunakan untuk mengevaluasi seberapa kuat NA_Sales dapat menjelaskan variasi Global_Sales. Teknik ini membantu memvalidasi hubungan antara regional sales dan global sales performance.

Analytical Technique	Purpose in Project	Example in Analysis
Descriptive Analysis	Summarize sales patterns	Total sales by genre, platform, region
Distribution Analysis	Evaluate sales distribution	Skewness, outlier detection
Correlation Analysis	Measure relationship between numeric variables	NA_Sales vs Global_Sales
Comparative Analysis	Compare groups	JP_Sales by Genre, Platform vs Genre
Hypothesis Testing	Test statistical significance	T-Test and ANOVA
Linear Regression	Measure explanatory power	NA_Sales predicting Global_Sales

b. Select and explain appropriate performance or evaluation metrics used to assess the effectiveness of your analysis.

Metrics yang digunakan adalah:

- Pearson Correlation Coefficient
- P-value
- T-statistic
- F-statistic
- R^2
- Skewness & Outlier Percentage\

#Hypotesis Testing

H0 : Tidak ada perbedaan rata-rata Global Sales antara genre Action dan Sports

H1 : Ada perbedaan rata-rata Global Sales antara genre Action dan Sports

```
from scipy.stats import ttest_ind

action_sales = df_eda[df_eda['Genre']=='Action']['Global_Sales']

sports_sales = df_eda[df_eda['Genre']=='Sports']['Global_Sales']

t_stat, p_value = ttest_ind(
    action_sales,
    sports_sales,
    equal_var=False
)

print("T-Statistic :", t_stat)
print("P-Value :", p_value)
```

[117] ✓ 0.1s

```
T-Statistic : -0.7982983070671815
P-Value : 0.42475483355072835
```

```
alpha = 0.05

if p_value < alpha:
    print("Reject H0")
    print("Terdapat perbedaan signifikan antara genre Action dan Sports")
else:
    print("Fail to Reject H0")
    print("Tidak terdapat perbedaan signifikan")
```

[118] ✓ 0.0s

```
Fail to Reject H0
Tidak terdapat perbedaan signifikan
```

```
from scipy.stats import f_oneway

groups = [
    group['Global_Sales']
    for _, group in df_eda.groupby('Genre')
]

f_stat, p_value = f_oneway(*groups)

# Output hasil
print(f"F-statistic: {f_stat:.4f}")
print(f"P-value: {p_value:.10f}")

alpha = 0.05

if p_value < alpha:
    print("\nDecision: Reject H0")
    print("Conclusion: Ada perbedaan signifikan terhadap average global sales di antara genre.")
else:
    print("\nDecision: Fail to Reject H0")
    print("Conclusion: Tidak ada perbedaan signifikan terhadap average global sales di antara genre..")
```

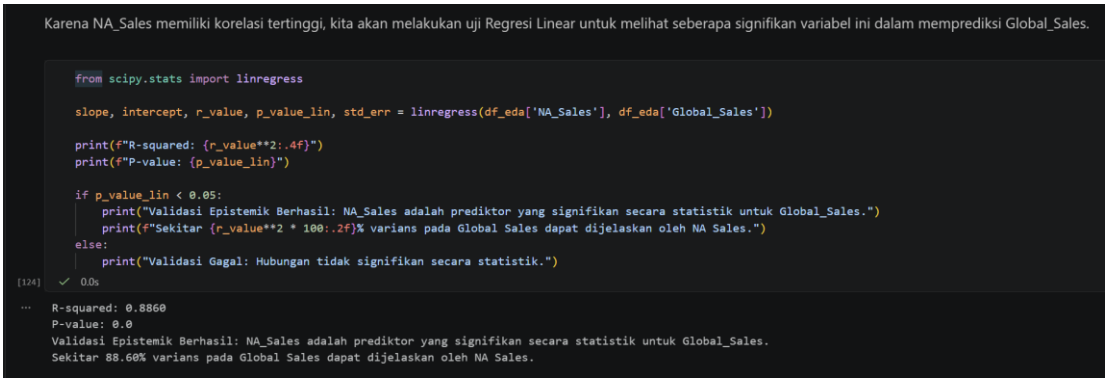
[122] ✓ 0.0s

```
F-statistic: 18.2415
P-value: 0.0000000000

Decision: Reject H0
Conclusion: Ada perbedaan signifikan terhadap average global sales di antara genre.
```

```
global_corr = corr['Global_Sales'].drop('Global_Sales')

highest_region = global_corr.idxmax()
highest_value = global_corr.max()

print("Region dengan korelasi tertinggi:", highest_region)
print("Nilai korelasi:", round(highest_value,3))
```

[123] ✓ 0.0s

```
Region dengan korelasi tertinggi: NA_Sales
Nilai korelasi: 0.941
```



Method	**Evaluation Metric**	**Result**	**Interpretation**
Correlation Analysis	Pearson Correlation	NA_Sales = 0.941, EU_Sales = 0.903, Other_Sales = 0.748, JP_Sales = 0.613	NA_Sales has the strongest relationship with Global_Sales
Independent T-Test	p-value	0.4247	No significant difference between Action and Sports average sales
One-Way ANOVA	F-statistic and p-value	F = 18.2415, $p < 0.001$	There is a significant difference in average sales across genres
Linear Regression	R^2 and p-value	$R^2 = 0.886$, p = 0.000	NA_Sales explains 88.6% of the variance in Global_Sales
Distribution Analysis	Skewness and Outlier Percentage	Skewness = 17.32, Outliers = 11.19%	Sales distribution is highly right-skewed and affected by blockbuster games

Beberapa metrik evaluasi digunakan untuk menilai efektivitas analisis. Pertama, Pearson correlation coefficient digunakan untuk mengukur kekuatan hubungan linear antara regional sales dan Global_Sales. Hasil menunjukkan bahwa NA_Sales memiliki korelasi tertinggi terhadap Global_Sales sebesar 0.941, diikuti oleh EU_Sales sebesar 0.903, Other_Sales sebesar 0.748, dan JP_Sales sebesar 0.613.

Kedua, p-value digunakan dalam hypothesis testing untuk menentukan apakah perbedaan penjualan antar kelompok signifikan secara statistik. Pada Independent T-Test antara Action dan Sports, p-value sebesar 0.4247 lebih besar dari 0.05, sehingga gagal menolak H0. Artinya, tidak terdapat perbedaan signifikan antara rata-rata Global_Sales genre Action dan Sports.

Ketiga, One-Way ANOVA dievaluasi menggunakan F-statistic dan p-value. Hasil ANOVA menunjukkan F-statistic sebesar 18.2415 dengan p-value kurang dari 0.001. Karena p-value lebih kecil dari 0.05, maka H0 ditolak. Artinya, terdapat perbedaan signifikan rata-rata Global_Sales antar genre.

Keempat, R^2 digunakan untuk mengevaluasi kekuatan linear regression antara NA_Sales dan Global_Sales. Nilai R^2 sebesar 0.886 menunjukkan bahwa sekitar 88.6% variasi Global_Sales dapat dijelaskan oleh NA_Sales.

Terakhir, skewness dan outlier percentage digunakan untuk mengevaluasi distribusi Global_Sales. Nilai skewness sebesar 17.32 dan outlier sebesar 11.19% menunjukkan bahwa distribusi penjualan sangat right-skewed dan dipengaruhi oleh sejumlah kecil game blockbuster.

c. Discuss the limitations of your analysis or models based on these metrics.

Walaupun analisis memberikan insight yang berguna, terdapat beberapa keterbatasan berdasarkan metrik evaluasi. Pertama, variabel Global_Sales memiliki distribusi yang sangat right-skewed dengan nilai skewness 17.32 dan outlier sebesar 11.19%. Artinya, dataset sangat dipengaruhi oleh sejumlah kecil game blockbuster dengan penjualan sangat tinggi. Akibatnya, metrik seperti mean, korelasi, regresi, T-Test, dan ANOVA dapat terpengaruh oleh nilai ekstrem tersebut.

Kedua, hasil correlation analysis perlu diinterpretasikan dengan hati-hati. Correlation matrix menunjukkan bahwa NA_Sales memiliki hubungan paling kuat dengan Global_Sales sebesar 0.941. Namun, Global_Sales merupakan hasil penjumlahan dari NA_Sales, EU_Sales, JP_Sales, dan Other_Sales. Oleh karena itu, korelasi tinggi tersebut sebagian memang wajar dan tidak membuktikan bahwa NA_Sales secara langsung menyebabkan peningkatan Global_Sales.

Ketiga, linear regression antara NA_Sales dan Global_Sales menghasilkan R^2 sebesar 0.886. Namun, karena Global_Sales sudah mengandung NA_Sales sebagai salah satu komponennya, model ini lebih tepat dipahami sebagai hubungan asosiasi, bukan model prediksi yang benar-benar independen.

Keempat, Independent T-Test antara Action dan Sports menghasilkan p-value 0.4247, sehingga tidak terdapat perbedaan signifikan antara rata-rata Global_Sales kedua genre tersebut. Namun, T-Test ini hanya membandingkan dua genre dan tidak mewakili seluruh genre. Selain itu, data yang sangat skewed dapat memengaruhi keandalan perbandingan berbasis rata-rata.

Kelima, One-Way ANOVA menunjukkan adanya perbedaan signifikan antar genre dengan $F = 18.2415$ dan $p < 0.001$. Namun, ANOVA hanya menunjukkan bahwa setidaknya ada satu genre yang berbeda secara signifikan, tetapi tidak menunjukkan genre mana yang berbeda tanpa post-hoc test. Selain itu, keberadaan outlier dapat memengaruhi hasil ANOVA.

Metric / Method	**Limitation**
Skewness & Outlier Detection	Global_Sales is highly right-skewed and affected by blockbuster games
Correlation Analysis	High correlation does not imply causation, especially because Global_Sales is the sum of regional sales

Linear Regression	NA_Sales is part of Global_Sales, so the model is not fully independent
T-Test	Only compares Action and Sports, not all genres
ANOVA	Shows that differences exist, but does not identify which genres differ without post-hoc testing
Mean-based Analysis	Mean sales can be biased by extreme blockbuster games

Secara keseluruhan, keterbatasan ini menunjukkan bahwa hasil analisis perlu diinterpretasikan secara hati-hati. Meskipun correlation, regression, T-Test, dan ANOVA memberikan bukti statistik yang berguna, distribusi data yang sangat skewed dan adanya outlier dapat memengaruhi hasil analisis. Oleh karena itu, kesimpulan tidak boleh hanya bergantung pada total sales atau average sales, tetapi juga perlu mempertimbangkan median sales, jumlah game, perbedaan region, serta hubungan antara platform dan genre agar interpretasi yang dihasilkan lebih seimbang.

4. Exploratory Data Analysis (EDA)

a. Propose 1-2 actionable recommendations based on your insights.

I) Rekomendasi 1: Prioritaskan Strategi RPG untuk Pasar Jepang

Berdasarkan hasil analisis, publisher sebaiknya memprioritaskan game bergenre Role-Playing untuk pasar Jepang, terutama pada platform yang menunjukkan performa RPG kuat seperti DS, PS, 3DS, GB, SNES, PS2, dan PSP. Analisis JP_Sales by Genre menunjukkan bahwa Role-Playing memiliki total penjualan tertinggi di Jepang, sedangkan heatmap Platform vs Genre berdasarkan JP_Sales menunjukkan bahwa penjualan RPG didukung kuat oleh platform Nintendo dan PlayStation.

Rekomendasi ini menunjukkan bahwa publisher dapat berfokus pada pengembangan game RPG, lokalisasi bahasa Jepang, pemasaran berbasis cerita, serta perilisan pada platform yang memiliki ekosistem RPG kuat di Jepang.

II) Rekomendasi 2: Terapkan Regional Pricing dan Lokalisasi untuk Meningkatkan Other_Sales

Untuk meningkatkan Other_Sales, publisher sebaiknya menerapkan strategi regional pricing dan localized marketing pada pasar di luar NA, EU, dan Jepang. Meskipun Other_Sales memiliki rata-rata kontribusi paling rendah dibanding wilayah lain, correlation matrix menunjukkan bahwa Other_Sales memiliki korelasi sebesar 0.748 terhadap Global_Sales, lebih tinggi dibanding JP_Sales yang sebesar 0.613. Hal ini menunjukkan bahwa Other_Sales tetap memiliki hubungan yang bermakna terhadap performa penjualan global dan dapat menjadi peluang pertumbuhan strategis.

Regional pricing dapat membantu membuat game lebih terjangkau di pasar dengan daya beli yang berbeda, sedangkan localized marketing dapat meningkatkan relevansi

produk melalui penyesuaian bahasa, promosi, dan strategi distribusi sesuai perilaku konsumen lokal.

b. Explain how these recommendations could support decision making for relevant stakeholders.

Rekomendasi ini dapat membantu proses pengambilan keputusan bagi publisher, developer, tim marketing, dan platform partner. Untuk pasar Jepang, strategi fokus pada RPG membantu stakeholder mengalokasikan sumber daya pengembangan dan pemasaran pada genre yang sudah terbukti memiliki permintaan tinggi. Karena RPG memiliki performa kuat di Jepang dan didukung oleh platform Nintendo serta PlayStation, publisher dapat membuat keputusan yang lebih tepat terkait pemilihan genre, platform rilis, dan investasi lokalisasi.

Untuk Other_Sales, strategi regional pricing dan lokalisasi dapat membantu publisher memperluas pasar. Karena Other_Sales memiliki kontribusi rata-rata yang rendah tetapi korelasinya terhadap Global_Sales masih cukup kuat, peningkatan pada wilayah ini berpotensi membantu pertumbuhan penjualan global. Regional pricing membantu menyesuaikan harga dengan daya beli lokal, sementara lokalisasi pemasaran dapat meningkatkan penerimaan konsumen di pasar yang lebih beragam.

c. Discuss potential risks or considerations if the recommendations are implemented.

Terdapat beberapa risiko dan pertimbangan dalam menerapkan rekomendasi ini. Pertama, terlalu fokus pada RPG di Jepang dapat menyebabkan ketergantungan pada satu genre. Meskipun RPG saat ini memiliki performa kuat di Jepang, preferensi konsumen dapat berubah seiring waktu sehingga publisher perlu tetap memantau tren genre.

Kedua, penerapan regional pricing dapat menimbulkan tantangan seperti perbedaan harga antar wilayah, risiko grey-market reselling, atau ketidakpuasan konsumen jika mereka membandingkan harga antar region. Oleh karena itu, regional pricing harus diterapkan secara hati-hati dengan segmentasi pasar dan kontrol distribusi yang jelas.

Ketiga, dataset yang digunakan tidak mencakup variabel penting seperti harga game, penjualan digital, tingkat pendapatan konsumen, data negara secara detail dalam Other_Sales, biaya marketing, atau demografi pemain. Oleh karena itu, rekomendasi regional pricing belum dapat divalidasi secara langsung dari dataset ini. Rekomendasi tersebut lebih tepat dipahami sebagai strategi bisnis berdasarkan temuan bahwa Other_Sales memiliki kontribusi rata-rata rendah tetapi tetap memiliki korelasi yang cukup kuat terhadap Global_Sales.

Terakhir, karena dataset memiliki distribusi yang sangat skewed dan dipengaruhi oleh game blockbuster, rekomendasi tidak boleh hanya bergantung pada total sales. Publisher juga perlu mempertimbangkan median sales, rata-rata sales per game, jumlah game, ketersediaan platform, serta tren pasar terbaru sebelum menerapkan strategi tersebut.