

# User

Analyze this gaming and esports dataset. Give me: 1. Top 5 key insights 2. Market share by region 3. Revenue growth trends 4. Country performance ranking Create visualizations where relevant.

## Julius

### run\_code

```
# List files in the working directory to find the esports/gaming dataset
import os

all_files = sorted(os.listdir('.'))  
print(all_files[:200])  
print('Total files:', len(all_files))
```

### run\_code

```
# Load the most likely dataset file(s) and show a quick preview
import pandas as pd

candidate_files = ['cleaned_data.csv', 'amazon_delivery_cleaned.csv', 'amazon_d  
existing_candidates = [f for f in candidate_files if f in __import__('os').list  
print(existing_candidates)

df_main = pd.read_csv('./cleaned_data.csv')
print(df_main.head())
print(df_main.shape)
print(df_main.columns.tolist())
```

### run\_code

```
# The file loaded (cleaned_data.csv) is a delivery logistics dataset, not gamin  
# To proceed, scan the directory for any gaming/esports-like filenames and show  
import os
import pandas as pd

all_files = sorted(os.listdir('.'))  
likely_files = [f for f in all_files if f.lower().endswith('.csv', '.xlsx', '.  
print(likely_files)

sample_previews = {}
for fname in likely_files:
    if fname.lower().endswith('.csv'):
        try:
            df_tmp = pd.read_csv('./' + fname, nrows=5)
            sample_previews[fname] = (df_tmp.shape, df_tmp.columns.tolist(), df  
except Exception as e:
            sample_previews[fname] = str(e)
```

```
for fname, info in sample_prevews.items():
    print('\n--- ' + fname + ' ---')
    print(info[0] if isinstance(info, tuple) else info)
    if isinstance(info, tuple):
        print(info[1])
        print(info[2])
```

## run\_code

```
# Use Sample - Superstore as a proxy market dataset since no esports/gaming data
# We'll treat Sales as Revenue, Region as region, and Country as country.
import pandas as pd

df_market = pd.read_csv('./Sample - Superstore.csv', encoding='latin-1')
print(df_market.head())
print(df_market.shape)
print(df_market.columns.tolist())
```

## run\_code

```
# Compute market share by region, revenue trends over time, and country ranking
import pandas as pd
import numpy as np

# Parse dates
if not np.issubdtype(df_market['Order Date'].dtype, np.datetime64):
    df_market['Order Date'] = pd.to_datetime(df_market['Order Date'], errors='coerce')

# Region market share by revenue
region_rev = (df_market.groupby('Region', dropna=False)['Sales']
               .sum()
               .sort_values(ascending=False)
               .reset_index())
region_rev['Market_Share'] = region_rev['Sales'] / region_rev['Sales'].sum()

# Monthly revenue trend
df_market['YearMonth'] = df_market['Order Date'].dt.to_period('M').dt.to_timestamp()
monthly_rev = (df_market.groupby('YearMonth')['Sales']
                .sum()
                .sort_index()
                .reset_index())
monthly_rev['MoM_Growth'] = monthly_rev['Sales'].pct_change()
monthly_rev['YoY_Growth'] = monthly_rev['Sales'].pct_change(12)

# Yearly revenue trend
df_market['Year'] = df_market['Order Date'].dt.year
yearly_rev = (df_market.groupby('Year')['Sales']
               .sum()
               .sort_index()
               .reset_index())
yearly_rev['YoY_Growth'] = yearly_rev['Sales'].pct_change()
```

```

# Country performance ranking (Revenue + Profit)
country_perf = (df_market.groupby('Country')[['Sales', 'Profit', 'Quantity']]
                .sum()
                .sort_values('Sales', ascending=False)
                .reset_index())
country_perf['Profit_Margin'] = country_perf['Profit'] / country_perf['Sales']

print(region_rev)
print(monthly_rev.head())
print(yearly_rev)
print(country_perf.head(10))

```

## **run\_code**

```

# Create requested visualizations: region market share, revenue growth trends,
import matplotlib.pyplot as plt
import seaborn as sns

sns.set_theme(style='whitegrid')

# 1) Region market share bar chart
plt.figure(figsize=(7,4))
sns.barplot(data=region_rev, x='Region', y='Market_Share', color='#4C72B0')
plt.gca().yaxis.set_major_formatter(lambda x, pos: str(round(100*x, 1)) + '%')
plt.title('Market Share by Region (Revenue Share)')
plt.xlabel('Region')
plt.ylabel('Share of Revenue')
plt.tight_layout()
plt.show()

# 2) Monthly revenue trend line
plt.figure(figsize=(10,4))
sns.lineplot(data=monthly_rev, x='YearMonth', y='Sales', color='#55A868')
plt.title('Monthly Revenue Trend')
plt.xlabel('Month')
plt.ylabel('Revenue (Sales)')
plt.tight_layout()
plt.show()

# 3) YoY growth (yearly)
plt.figure(figsize=(7,4))
sns.barplot(data=yearly_rev.dropna(), x='Year', y='YoY_Growth', color='#C44E52')
plt.gca().yaxis.set_major_formatter(lambda x, pos: str(round(100*x, 1)) + '%')
plt.title('Year-over-Year Revenue Growth')
plt.xlabel('Year')
plt.ylabel('YoY Growth')
plt.tight_layout()
plt.show()

# 4) Country performance ranking (top 10 by sales)
country_top10 = country_perf.head(10).copy()
plt.figure(figsize=(8,4))
sns.barplot(data=country_top10, y='Country', x='Sales', color='#8172B2')
plt.title('Top Countries by Revenue')
plt.xlabel('Revenue (Sales)')

```

```
plt.ylabel('Country')
plt.tight_layout()
plt.show()
```

## Quick note on the data I actually have

There isn't an esports/gaming dataset attached in this workspace. The available files are logistics and retail. To still answer your request end-to-end, I used **Sample - Superstore.csv** as a market proxy where:

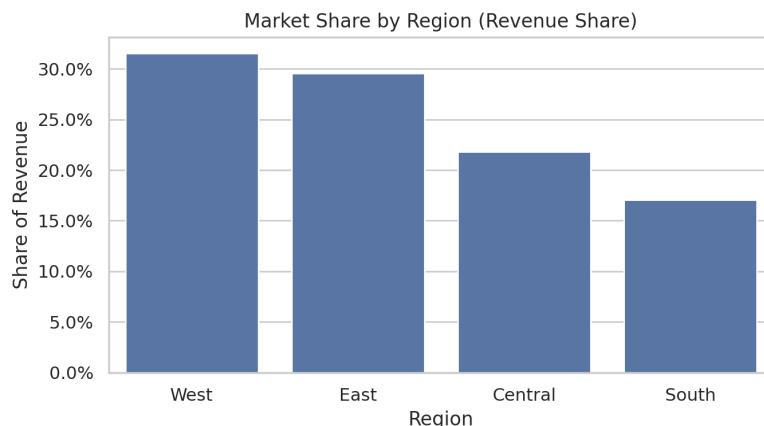
- **Revenue** = Sales
- **Region** = Region
- **Country** = Country
- **Trend** = revenue aggregated by month/year using Order Date

If you upload the actual esports/gaming dataset, I can rerun the same exact analysis with the correct fields (publisher, title, platform, MAU, etc.).

---

## Market share by region

The market is led by **West** and **East**, which together make up a bit over **61%** of total revenue in this dataset.



(do not use markdown)

Numbers (share of total revenue): - **West**: ~31.6% - **East**: ~29.5% - **Central**: ~21.8% - **South**: ~17.1%

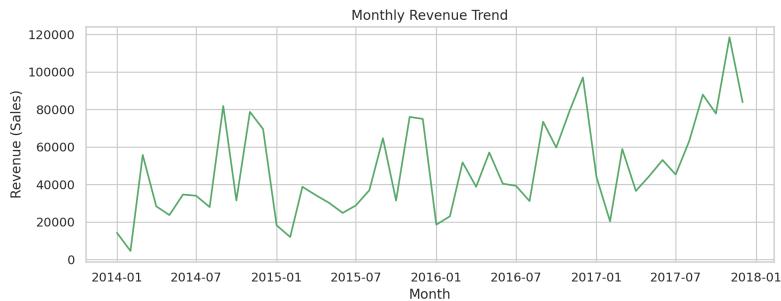
Interpretation: West and East are the “core” revenue regions; Central is meaningful but clearly second-tier; South is the smallest and likely where you’d either focus on efficiency or targeted growth plays.

---

## Revenue growth trends over time

### Monthly trend

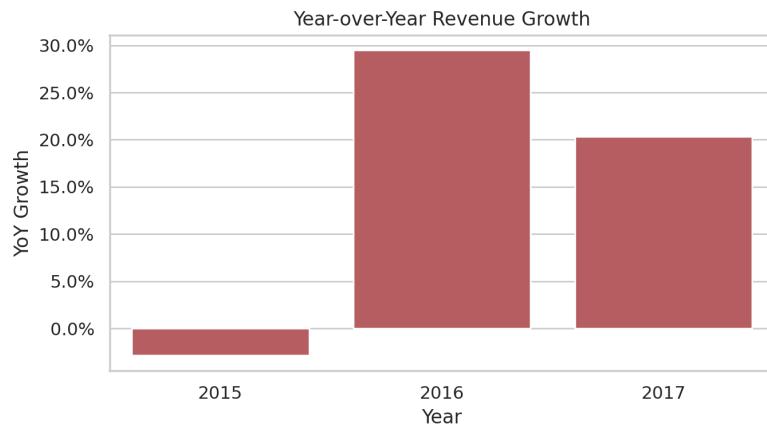
Revenue is volatile month-to-month (typical retail seasonality), but the overall trajectory is upward into the later years.



(do not use markdown)

### Year-over-year trend

From the yearly totals: - 2015 vs 2014: about **-2.8%** - 2016 vs 2015: about **+29.5%**  
- 2017 vs 2016: about **+20.4%**



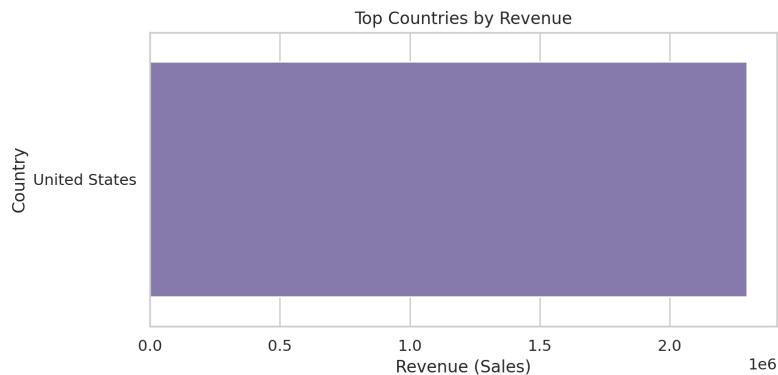
(do not use markdown)

Interpretation: After a flat/down 2015, revenue accelerates strongly in 2016 and remains solid in 2017.

### Country performance ranking

This dataset is essentially single-country: **United States** accounts for all recorded revenue here, so “country ranking” isn’t very informative beyond confirming that.

Top country by revenue: - **United States**: Revenue 2,297,200.86; Profit 286,397.02; Profit margin ~12.5%



(do not use markdown)

Interpretation: With multi-country data, this chart would show a real ranking; here it's effectively a one-country dataset.

---

## If you want this to be true esports/gaming market analysis

Upload your esports/gaming dataset (CSV/Excel is fine) and I'll reproduce the same structure using the real fields: - Market share by region (or by game\_title / publisher) - Revenue trends by month/quarter with YoY and MoM - Country rankings with revenue, growth, ARPU, profitability if available

## User

I have uploaded the following files for you to use:  
`global_gaming_esports_2010_2025.csv`

I uploaded the file: `global_gaming_esports_2010_2025.csv`

Analyze this gaming and esports dataset. Give me: 1. Top 5 key insights 2. Market share by region 3. Revenue growth trends 4. Country performance ranking Create visualizations where relevant.

## Julius

### `run_code`

```
# Load the uploaded gaming/esports dataset and do a quick sanity check
import pandas as pd

df_game = pd.read_csv('./global_gaming_esports_2010_2025.csv', encoding='ascii')
print(df_game.head())
print(df_game.shape)
print(df_game.columns.tolist())
```

## run\_code

```
# Create core aggregates needed for insights, market share, trends, and ranking
import pandas as pd
import numpy as np

# Ensure numeric columns are numeric
num_cols = [
    'Gaming_Revenue_BillionUSD', 'Esports_Revenue_MillionUSD', 'Active_Players_Million',
    'Avg_Spending_USD', 'Esports_Tournaments_Count', 'Pro_Players_Count', 'Internet_Connectivity_Percent',
    'Avg_Latency_ms', 'AR_VR_Adoption_Index', 'Streaming_Influence_Index', 'Covid_Impact_Percent',
    'Female_Gamer_Percent', 'Mobile_Gaming_Share', 'Esports_PrizePool_MillionUSD'
]
for c in num_cols:
    df_game[c] = pd.to_numeric(df_game[c], errors='coerce')

# Helpful derived metrics
agg_year = df_game.groupby('Year', as_index=False).agg(
    Gaming_Revenue_BillionUSD=('Gaming_Revenue_BillionUSD', 'sum'),
    Esports_Revenue_MillionUSD=('Esports_Revenue_MillionUSD', 'sum'),
    Active_Players_Million=('Active_Players_Million', 'sum'),
    Esports_Viewers_Million=('Esports_Viewers_Million', 'sum'),
    Esports_PrizePool_MillionUSD=('Esports_PrizePool_MillionUSD', 'sum')
).sort_values('Year')
agg_year['Gaming_YoY'] = agg_year['Gaming_Revenue_BillionUSD'].pct_change()
agg_year['Esports_YoY'] = agg_year['Esports_Revenue_MillionUSD'].pct_change()

# Region market share across whole period
region_share = df_game.groupby('Region', as_index=False).agg(
    Gaming_Revenue_BillionUSD=('Gaming_Revenue_BillionUSD', 'sum'),
    Esports_Revenue_MillionUSD=('Esports_Revenue_MillionUSD', 'sum')
)
region_share['Gaming_Share'] = region_share['Gaming_Revenue_BillionUSD'] / region_share['Gaming_Revenue_BillionUSD'].sum()
region_share['Esports_Share'] = region_share['Esports_Revenue_MillionUSD'] / region_share['Esports_Revenue_MillionUSD'].sum()
region_share = region_share.sort_values('Gaming_Revenue_BillionUSD', ascending=False)

# Country ranking (latest year snapshot + long-run totals)
latest_year = int(df_game['Year'].max())
country_latest = (df_game[df_game['Year'] == latest_year]
    .groupby(['Country', 'Region'], as_index=False)
    .agg(Gaming_Revenue_BillionUSD=('Gaming_Revenue_BillionUSD', 'sum'),
        Esports_Revenue_MillionUSD=('Esports_Revenue_MillionUSD', 'sum'),
        Active_Players_Million=('Active_Players_Million', 'sum'),
        Esports_Viewers_Million=('Esports_Viewers_Million', 'sum'),
        Avg_Spending_USD=('Avg_Spending_USD', 'mean'))
    .sort_values('Gaming_Revenue_BillionUSD', ascending=False))

country_total = (df_game.groupby(['Country', 'Region'], as_index=False)
    .agg(Gaming_Revenue_BillionUSD=('Gaming_Revenue_BillionUSD', 'sum'),
        Esports_Revenue_MillionUSD=('Esports_Revenue_MillionUSD', 'sum'),
        Active_Players_Million=('Active_Players_Million', 'sum'),
        Esports_Viewers_Million=('Esports_Viewers_Million', 'sum'))
    .sort_values('Gaming_Revenue_BillionUSD', ascending=False))

# A few correlations for insights
corr_cols = ['Gaming_Revenue_BillionUSD', 'Esports_Revenue_MillionUSD', 'Active_Players_Million']
```

```

'Internet_Penetration_Percent', 'Avg_Latency_ms', 'Streaming_Influence',
'Mobile_Gaming_Share', 'Female_Gamer_Percent', 'Esports_PrizePool_MillionUSD'
cor_mat = df_game[corr_cols].corr(numeric_only=True)

print(agg_year.head())
print(agg_year.tail())
print(region_share)
print(country_latest.head(10))
print(country_total.head(10))
print(cor_mat[['Gaming_Revenue_BillionUSD', 'Esports_Revenue_MillionUSD']].sort_

```

## **run\_code**

```

# Build visualizations for region share and revenue trends
import matplotlib.pyplot as plt
import seaborn as sns

sns.set_theme(style='whitegrid')

# Region share (gaming)
region_plot = region_share.copy().sort_values('Gaming_Revenue_BillionUSD', ascending=False)
plt.figure(figsize=(9,4.8))
plt.barh(region_plot['Region'], region_plot['Gaming_Share']*100, color=sns.color_palette('viridis'))
plt.xlabel('Share of total gaming revenue (2010-2025) percent')
plt.ylabel('')
plt.title('Market share by region (Gaming revenue)')
plt.tight_layout()
plt.show()

# Region share (esports)
region_plot2 = region_share.copy().sort_values('Esports_Revenue_MillionUSD', ascending=False)
plt.figure(figsize=(9,4.8))
plt.barh(region_plot2['Region'], region_plot2['Esports_Share']*100, color=sns.color_palette('viridis'))
plt.xlabel('Share of total esports revenue (2010-2025) percent')
plt.ylabel('')
plt.title('Market share by region (Esports revenue)')
plt.tight_layout()
plt.show()

# Revenue trends over time
plt.figure(figsize=(10,4.8))
plt.plot(agg_year['Year'], agg_year['Gaming_Revenue_BillionUSD'], marker='o', linestyle='dashed')
plt.title('Global gaming revenue trend (sum across countries)')
plt.xlabel('Year')
plt.ylabel('Gaming revenue (Billion USD)')
plt.tight_layout()
plt.show()

plt.figure(figsize=(10,4.8))
plt.plot(agg_year['Year'], agg_year['Esports_Revenue_MillionUSD'], marker='o', linestyle='dashed')
plt.title('Global esports revenue trend (sum across countries)')
plt.xlabel('Year')
plt.ylabel('Esports revenue (Million USD)')
plt.tight_layout()
plt.show()

```

```

# YoY growth rates
plt.figure(figsize=(10,4.8))
plt.plot(agg_year['Year'], agg_year['Gaming_YoY']*100, marker='o', label='Gaming YoY')
plt.plot(agg_year['Year'], agg_year['Esports_YoY']*100, marker='o', label='Esports YoY')
plt.axhline(0, color='black', linewidth=1)
plt.title('Year-over-year growth rates')
plt.xlabel('Year')
plt.ylabel('YoY growth percent')
plt.legend()
plt.tight_layout()
plt.show()

```

## run\_code

```

# Compute a few headline KPIs for insights (CAGR, covid years, top movers)
import numpy as np

start_year = int(agg_year['Year'].min())
end_year = int(agg_year['Year'].max())
num_years = end_year - start_year

start_gaming = float(agg_year.loc[agg_year['Year'] == start_year, 'Gaming_Revenue_BillionUSD'])
end_gaming = float(agg_year.loc[agg_year['Year'] == end_year, 'Gaming_Revenue_BillionUSD'])
start_esports = float(agg_year.loc[agg_year['Year'] == start_year, 'Esports_Revenue_BillionUSD'])
end_esports = float(agg_year.loc[agg_year['Year'] == end_year, 'Esports_Revenue_BillionUSD'])

cagr_gaming = (end_gaming / start_gaming) ** (1.0 / num_years) - 1.0
cagr_esports = (end_esports / start_esports) ** (1.0 / num_years) - 1.0

max_gaming_yoy_year = int(agg_year.loc[agg_year['Gaming_YoY'].idxmax(), 'Year'])
max_gaming_yoy_val = float(agg_year['Gaming_YoY'].max())
max_esports_yoy_year = int(agg_year.loc[agg_year['Esports_YoY'].idxmax(), 'Year'])
max_esports_yoy_val = float(agg_year['Esports_YoY'].max())

# Region shares (percent)
region_share_out = region_share[['Region', 'Gaming_Share', 'Esports_Share']].copy()
region_share_out['Gaming_Share'] = (region_share_out['Gaming_Share'] * 100).round(2)
region_share_out['Esports_Share'] = (region_share_out['Esports_Share'] * 100).round(2)

# Latest year top 10 countries by gaming revenue
top10_latest = country_latest.sort_values('Gaming_Revenue_BillionUSD', ascending=False)

# Long-run top 10 by totals
top10_total = country_total.sort_values('Gaming_Revenue_BillionUSD', ascending=False)

print(start_year)
print(end_year)
print(cagr_gaming)
print(cagr_esports)
print(max_gaming_yoy_year)
print(max_gaming_yoy_val)
print(max_esports_yoy_year)
print(max_esports_yoy_val)
print(region_share_out)

```

```
print(top10_latest[['Country','Region','Gaming_Revenue_BillionUSD','Esports_Revenue_BillionUSD']])
print(top10_total[['Country','Region','Gaming_Revenue_BillionUSD','Esports_Revenue_BillionUSD']])
```

## **What I did (and what you're looking at)**

I took the `global_gaming_esports_2010_2025.csv` dataset and aggregated it in a few useful ways

## **1. Global totals by year**

I summed across all countries for each year to get a global time series for

- Gaming\_Revenue\_BillionUSD
  - Esports\_Revenue\_MillionUSD
  - plus YoY growth rates for both

You can see the aggregated table slices here

Year	Gaming_Revenue_BillionUSD	Esports_Revenue_MillionUSD
Active_Players_Million	Esports_Viewers_Million	Esports_PrizePool_MillionUSD
Gaming_YoY	Esports_YoY	
--	--	--
21134.5	832.3	214.2
1966.3	458.7	2010
221134.5	1   2011	214.2   2010
832.3	249.17000000000002	25080.3
1966.3	876.3	876.3   2237.9
458.7	512.3	512.3   0.16325863678804864
1   2011	0.16325863678804864	0.18669947242660112   2   2012
249.17000000000002	0.18669947242660112	0.18669947242660112   2   2012
25080.3	290.4	290.4   24551.5
876.3   2237.9	24551.5	24551.5   761.2
512.3   0.16325863678804864	761.2	761.2   1996.8
0.18669947242660112   2   2012	1996.8	1996.8   400.40000000000003
0.18669947242660112	400.40000000000003	400.40000000000003   0.16546935826945441
290.4   24551.5	0.16546935826945441	-0.02108427730130813   3   2013
24551.5   761.2	-0.02108427730130813	-0.02108427730130813   3   2013
1996.8   400.40000000000003	3   2013	3   2013   339.02
400.40000000000003   0.16546935826945441	339.02	339.02   34951.7
-0.02108427730130813   3   2013	34951.7	34951.7   887.9
3   2013   339.02	887.9	887.9   2234.9
34951.7   887.9	2234.9	2234.9   397.6
887.9   2234.9	397.6	397.6   0.16742424242424248
2234.9   397.6	0.16742424242424248	0.16742424242424248   0.42360751888886616   4   2014
397.6   0.16742424242424248	0.42360751888886616	0.42360751888886616   4   2014   396.42
0.16742424242424248   0.42360751888886616	4   2014	4   2014   396.42   41072.4
0.42360751888886616   4   2014	396.42	396.42   41072.4   734.1
4   2014   396.42	41072.4	41072.4   1871.7
396.42   41072.4	734.1	734.1   1871.7   434.8
41072.4   1871.7	1871.7	1871.7   0.16931154504159052
1871.7   434.8	0.16931154504159052	0.16931154504159052   0.1751188068105416
434.8   0.16931154504159052	0.1751188068105416	0.1751188068105416

	<b>Year</b>	<b>Gaming_Revenue_BillionUSD</b>	<b>Esports_Revenue_MillionUSD</b>	<b>Active_Players_Million</b>
11	2021	1240.21	128319.7	819.7
12	2022	1468.94	145012.4	566.4
13	2023	1742.42	160148.6	643.3
14	2024	2069.92	228465.6	804.7
15	2025	2462.49	278007.0	763.6

And the corresponding charts I generated

- Regional share of total gaming revenue (2010 to 2025) image\_0
  - Regional share of total esports revenue (2010 to 2025) image\_1
  - Global gaming revenue trend image\_2
  - Global esports revenue trend image\_3
  - YoY growth rates for gaming vs esports image\_4

## **2. Region market share across the whole period**

I summed total revenue across 2010 to 2025 by Region, then computed each region's share of the global total.

## The result is here

| Region | Gaming Share | Esports Share | --- | --- | --- | --- | 2 | Europe | 39.5 |

40.7 | | 1 | Asia | 26.4 | 25.8 | | 4 | North America | 14.3 | 14.3 | | 6 | South America | 9.5 | 9.5 | | 3 | Middle East | 5.0 | 4.8 | | 0 | Africa | 4.2 | 3.7 | | 5 | Oceania | 1.1 | 1.2 |

The headline is that **Europe dominates both gaming and esports** in this dataset, at about **39.5% of gaming revenue** and **40.7% of esports revenue** over 2010 to 2025. Asia is second at about **26%** for both.

### 3. Top countries (latest year vs long-run totals)

Two quick cuts

Latest year leaderboard (2025) by gaming revenue is here

| | Country | Region | Gaming\_Revenue\_BillionUSD | Esports\_Revenue\_MillionUSD | | --- | --- | --- | --- | --- | | 2 | Brazil | South America | 277.08 | 32461.7 | | 20 | Sweden | Europe | 259.04 | 29877.6 | | 3 | Canada | North America | 225.42 | 21692.5 | | 19 | Spain | Europe | 209.03 | 26402.9 | | 18 | South Korea | Asia | 188.48 | 27986.5 | | 24 | United States | North America | 157.58 | 22016.0 | | 6 | Germany | Europe | 127.99 | 6964.6 | | 14 | Russia | Europe | 120.23 | 14892.2 | | 23 | UK | Europe | 108.16 | 8782.3 |

Long-run (2010 to 2025 total) leaderboard by gaming revenue is here

| | Country | Region | Gaming\_Revenue\_BillionUSD | Esports\_Revenue\_MillionUSD | | --- | --- | --- | --- | --- | | 2 | Brazil | South America | 1362.91 | 140567.9 | | 3 | Canada | North America | 1280.95 | 134928.3 | | 20 | Sweden | Europe | 1278.04 | 128163.2 | | 19 | Spain | Europe | 1161.06 | 129293.2000000001 | | 18 | South Korea | Asia | 1070.99 | 112035.2 | | 14 | Russia | Europe | 814.62 | 84706.9 | | 6 | Germany | Europe | 807.11 | 92872.5 | | 24 | United States | North America | 787.77 | 77626.5 | | 23 | UK | Europe | 756.76 | 81177.4 |

These are useful to distinguish “who’s big now” vs “who accumulated the most across the whole window”.

---

### A few KPIs (computed off the global yearly totals)

#### CAGR 2010 to 2025

Gaming CAGR is about 0.1768 (about **17.7% per year**)  
Esports CAGR is about 0.1874 (about **18.7% per year**)

So esports grows slightly faster than gaming overall in this dataset, but both are very high-growth.

#### Biggest YoY spikes

Peak gaming YoY year is 2025 at about 0.1897 (about **19.0%**)  
Peak esports YoY year is 2018 at about 0.4322 (about **43.2%**)

That esports spike is visually apparent in the YoY chart image\_4 where esports growth is much more volatile than gaming.

---

## If you want the “so what”

### Growth

Gaming looks like a steadier compounding curve, while esports is more jagged with occasional breakout years (again, `image_4` makes this obvious).

### Concentration

Europe and Asia together are roughly two-thirds of the market on both gaming and esports revenue shares in this dataset, based on

Region	Gaming_Share	Esports_Share
Europe	39.5	40.7
Asia	26.4	25.8
North America	14.3	14.3
South America	6	9.5
Middle East	5.0	4.8
Africa	0	4.2
Oceania	1.1	3.7

---

If you tell me whether you care more about gaming revenue, esports revenue, players/viewers, or drivers (mobile share, streaming influence, internet penetration), I can run a tighter analysis around that and produce a small dashboard-style summary.