Video Game Sales Analysis

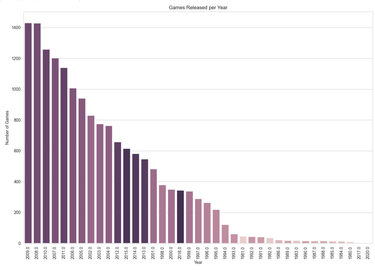
Dataset Link: https://www.kaggle.com/datasets/gregorut/videogamesales/data

In [ ]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
plt.figure
import numpy as np
sns.set_style("whitegrid")
df = pd.read_csv("C:\Users\kevin\OneDrive\Documents\Code\Dataset\vgsales.csv")
```

Number of games published per year

In [ ]:
```
plt.figure(figsize=(15, 10))
sns.barplot(x="Year", y=df.Year.value_counts().index, data=df, order = df["Year"].value_counts().index, hue="Year", legend=False)
plt.xticks(rotation=90)
plt.title("Games Released per Year")
plt.ylabel("Number of Games")
```

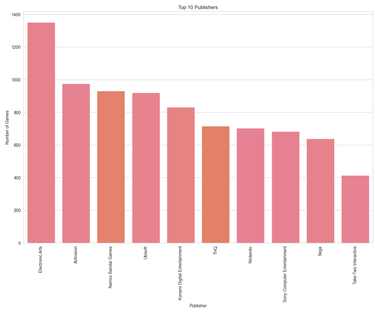
Out[ ]:
```
Text(0, 0.5, 'Number of Games')
```



Top 10 publishers by number of games published

In [ ]:
```
plt.figure(figsize=(15, 10))
sns.barplot(x="Publisher", y=df.Publisher.value_counts().head(10).index, data=df, order = df["Publisher"].value_counts().head(10).index, hue="Publisher", legend=False)
plt.xticks(rotation=90)
plt.title("Top 10 Publishers")
plt.ylabel("Number of Games")
```

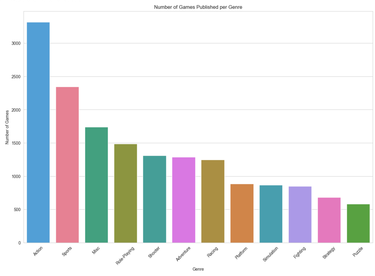
Out[ ]:
```
Text(0, 0.5, 'Number of Games')
```



Number of games published by each genre

In [ ]:
```
plt.figure(figsize=(15, 10))
sns.barplot(x="Genre", y=df.Genre.value_counts().index, data=df, order = df["Genre"].value_counts().index, hue="Genre")
plt.xticks(rotation=90)
plt.title("Number of Games Published per Genre")
plt.ylabel("Number of Games")
```

Out[ ]:
```
Text(0, 0.5, 'Number of Games')
```



Total Sales by Region

In [ ]:
```
region_sales = df[["NA_Sales", "EU_Sales", "JP_Sales", "Other_Sales"]].sum().reset_index()
region_sales = region_sales.rename(columns={"index": "Region", 0: "Total Sales"})
region_sales
```

Out[ ]:

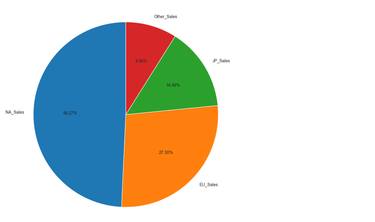
	Region	Total Sales
0	NA_Sales	4392.95
1	EU_Sales	2434.13
2	JP_Sales	1291.02
3	Other_Sales	797.75

In [ ]:
```
plt.figure(figsize=(15, 10))

labels = region_sales["Region"]
sizes = region_sales["Total Sales"]

plt.pie(x=sizes, labels=labels, autopct="%.1f%%", startangle=90)
plt.title("Regional Sales")
```

Out[ ]:
```
Text(0.5, 1.0, 'Regional Sales')
```

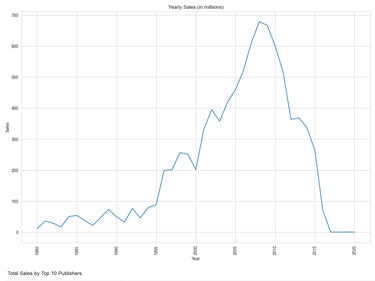


Total Sales by Year

In [ ]:
```
total_year_sales = df.groupby("Year")["Global_Sales"].sum().reset_index()

plt.figure(figsize=(15, 10))
sns.lineplot(x="Year", y="Global_Sales", data=total_year_sales)
plt.xticks(rotation=90)
plt.title("Yearly Sales (in Millions)")
plt.ylabel("Sales")
```

Out[ ]:
```
Text(0, 0.5, 'Sales')
```

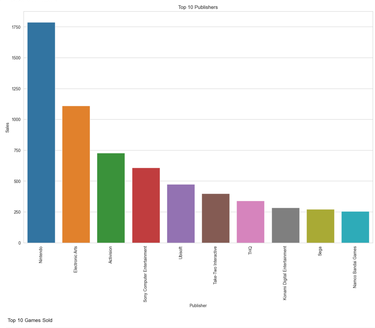


Total Sales by Top 10 Publishers

In [ ]:
```
sales_pub = df.groupby("Publisher")["Global_Sales"].sum().sort_values(ascending=False).head(10).reset_index()

plt.figure(figsize=(15, 10))
sns.barplot(x="Publisher", y="Global_Sales", data=sales_pub, hue="Publisher")
plt.xticks(rotation=90)
plt.title("Top 10 Publishers")
plt.ylabel("Sales")
```

Out[ ]:
```
Text(0, 0.5, 'Sales')
```

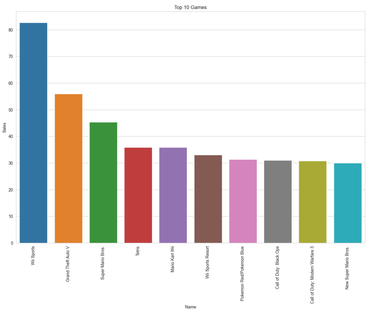


Top 10 Games Sold

In [ ]:
```
games_sales = df.groupby("Name")["Global_Sales"].sum().sort_values(ascending=False).head(10).reset_index()

plt.figure(figsize=(15, 10))
sns.barplot(x="Name", y="Global_Sales", data=games_sales, hue="Name")
plt.xticks(rotation=90)
plt.title("Top 10 Games")
plt.ylabel("Sales")
```

Out[ ]:
```
Text(0, 0.5, 'Sales')
```


Top 10 Games
Sales
0
10
20
30
40
50
60
70
80
Wii Sports
Grand Theft Auto V
Super Mario Bros.
Tetris
Mario Kart Wii
Wii Sports Resort
Pokemon Red/Pokemon Blue
Call of Duty: Black Ops
Call of Duty: Modern Warfare 3
New Super Mario Bros.
Name