deo-game-sales-engagement-analysis

January 8, 2026

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
plt.style.use("default")
```

[6]:
```
from google.colab import files

uploaded = files.upload()
```

```
<IPython.core.display.HTML object>

Saving games.csv to games.csv
Saving vgsales.csv to vgsales.csv
```

[7]:
```
games_df = pd.read_csv("games.csv")
sales_df = pd.read_csv("vgsales.csv")
```

[8]:
```
import os
os.listdir()
```

[8]:
```
['.config', 'games.csv', 'vgsales.csv', 'sample_data']
```

[9]:
```
games_df.head()
```

[9]:

	Unnamed: 0	Title	Release Date	Team	Rating	Times Listed
0	0	Elden Ring	Feb 25, 2022	['Bandai Namco Entertainment', 'FromSoftware']	4.5	3.9K
1	1	Hades	Dec 10, 2019	['Supergiant Games']	4.3	2.9K
2	2	The Legend of Zelda: Breath of the Wild	Mar 03, 2017	['Nintendo', 'Nintendo EPD Production Group No…	4.4	4.3K
3	3	Undertale	Sep 15, 2015	['tobyfox', '8-4']	4.2	3.5K
4	4	Hollow Knight	Feb 24, 2017	['Team Cherry']	4.4	3K

```
  Number of Reviews                                                Genres  \
0              3.9K                                ['Adventure', 'RPG']
1              2.9K          ['Adventure', 'Brawler', 'Indie', 'RPG']
2              4.3K                                ['Adventure', 'RPG']
3              3.5K  ['Adventure', 'Indie', 'RPG', 'Turn Based Stra…
4                3K                ['Adventure', 'Indie', 'Platform']

                                              Summary  \
0  Elden Ring is a fantasy, action and open world…
1  A rogue-lite hack and slash dungeon crawler in…
2  The Legend of Zelda: Breath of the Wild is the…
3  A small child falls into the Underground, wher…
4  A 2D metroidvania with an emphasis on close co…

                                              Reviews Plays Playing Backlogs  \
0  ["The first playthrough of elden ring is one o…   17K    3.8K     4.6K
1  ['convinced this is a roguelike for people who…   21K    3.2K     6.3K
2  ['This game is the game (that is not CS:GO) th…   30K    2.5K       5K
3  ['soundtrack is tied for #1 with nier automata…   28K     679     4.9K
4  ["this games worldbuilding is incredible, with…   21K    2.4K     8.3K

  Wishlist
0     4.8K
1     3.6K
2     2.6K
3     1.8K
4     2.3K
```

[10]:
```
sales_df.head()
```

[10]:
```
   Rank                     Name Platform    Year         Genre Publisher  \
0     1                Wii Sports      Wii  2006.0        Sports  Nintendo
1     2         Super Mario Bros.      NES  1985.0      Platform  Nintendo
2     3            Mario Kart Wii      Wii  2008.0        Racing  Nintendo
3     4         Wii Sports Resort      Wii  2009.0        Sports  Nintendo
4     5  Pokemon Red/Pokemon Blue       GB  1996.0  Role-Playing  Nintendo

   NA_Sales  EU_Sales  JP_Sales  Other_Sales  Global_Sales
0     41.49     29.02      3.77         8.46         82.74
1     29.08      3.58      6.81         0.77         40.24
2     15.85     12.88      3.79         3.31         35.82
3     15.75     11.01      3.28         2.96         33.00
4     11.27      8.89     10.22         1.00         31.37
```

[11]:
```
print("Games Shape:", games_df.shape)
print("Sales Shape:", sales_df.shape)
```

```
print("\nGames Columns:\n", games_df.columns)
print("\nSales Columns:\n", sales_df.columns)
```

```
Games Shape: (1512, 14)
Sales Shape: (16598, 11)

Games Columns:
 Index(['Unnamed: 0', 'Title', 'Release Date', 'Team', 'Rating', 'Times Listed',
       'Number of Reviews', 'Genres', 'Summary', 'Reviews', 'Plays', 'Playing',
       'Backlogs', 'Wishlist'],
      dtype='object')

Sales Columns:
 Index(['Rank', 'Name', 'Platform', 'Year', 'Genre', 'Publisher', 'NA_Sales',
       'EU_Sales', 'JP_Sales', 'Other_Sales', 'Global_Sales'],
      dtype='object')
```

[12]:
```
games_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1512 entries, 0 to 1511
Data columns (total 14 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Unnamed: 0         1512 non-null   int64
 1   Title              1512 non-null   object
 2   Release Date       1512 non-null   object
 3   Team               1511 non-null   object
 4   Rating             1499 non-null   float64
 5   Times Listed       1512 non-null   object
 6   Number of Reviews  1512 non-null   object
 7   Genres             1512 non-null   object
 8   Summary            1511 non-null   object
 9   Reviews            1512 non-null   object
 10  Plays              1512 non-null   object
 11  Playing            1512 non-null   object
 12  Backlogs           1512 non-null   object
 13  Wishlist           1512 non-null   object
dtypes: float64(1), int64(1), object(12)
memory usage: 165.5+ KB
```

[13]:
```
sales_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16598 entries, 0 to 16597
Data columns (total 11 columns):
 #   Column        Non-Null Count  Dtype
```

```
---  ------         --------------  -----
 0   Rank           16598 non-null  int64
 1   Name           16598 non-null  object
 2   Platform       16598 non-null  object
 3   Year           16327 non-null  float64
 4   Genre          16598 non-null  object
 5   Publisher      16540 non-null  object
 6   NA_Sales       16598 non-null  float64
 7   EU_Sales       16598 non-null  float64
 8   JP_Sales       16598 non-null  float64
 9   Other_Sales    16598 non-null  float64
 10  Global_Sales   16598 non-null  float64
dtypes: float64(6), int64(1), object(4)
memory usage: 1.4+ MB
```

[14]:
```
games_df.isnull().sum()
```

[14]:
```
Unnamed: 0            0
Title                 0
Release Date          0
Team                  1
Rating               13
Times Listed          0
Number of Reviews     0
Genres                0
Summary               1
Reviews               0
Plays                 0
Playing               0
Backlogs              0
Wishlist              0
dtype: int64
```

[15]:
```
sales_df.isnull().sum()
```

[15]:
```
Rank             0
Name             0
Platform         0
Year           271
Genre            0
Publisher       58
NA_Sales         0
EU_Sales         0
JP_Sales         0
Other_Sales      0
Global_Sales     0
dtype: int64
```

[16]:
```
games_df['Rating'] = games_df['Rating'].fillna(games_df['Rating'].median())
games_df['Plays'] = games_df['Plays'].fillna(0)
games_df['Wishlist'] = games_df['Wishlist'].fillna(0)
games_df['Backlogs'] = games_df['Backlogs'].fillna(0)
```

[18]:
```
games_df.columns
```

[18]:
```
Index(['Unnamed: 0', 'Title', 'Release Date', 'Team', 'Rating', 'Times Listed',
       'Number of Reviews', 'Genres', 'Summary', 'Reviews', 'Plays', 'Playing',
       'Backlogs', 'Wishlist'],
      dtype='object')
```

[19]:
```
games_df.drop(columns=['Unnamed: 0'], inplace=True)
```

[20]:
```
games_df.columns = (
    games_df.columns
    .str.strip()
    .str.lower()
    .str.replace(' ', '_')
)

games_df.columns
```

[20]:
```
Index(['title', 'release_date', 'team', 'rating', 'times_listed',
       'number_of_reviews', 'genres', 'summary', 'reviews', 'plays', 'playing',
       'backlogs', 'wishlist'],
      dtype='object')
```

[21]:
```
text_cols = ['title', 'team', 'genres']

for col in text_cols:
    games_df[col] = games_df[col].astype(str).str.strip().str.lower()
```

[22]:
```
games_df['rating'] = games_df['rating'].fillna(games_df['rating'].median())
games_df['plays'] = games_df['plays'].fillna(0)
games_df['wishlist'] = games_df['wishlist'].fillna(0)
games_df['backlogs'] = games_df['backlogs'].fillna(0)
```

[23]:
```
games_df.head()
```

[23]:

	title	release_date	\
0	elden ring	Feb 25, 2022	
1	hades	Dec 10, 2019	
2	the legend of zelda: breath of the wild	Mar 03, 2017	
3	undertale	Sep 15, 2015	
4	hollow knight	Feb 24, 2017	

```
                                                team  rating times_listed  \
0     ['bandai namco entertainment', 'fromsoftware']     4.5         3.9K
1                               ['supergiant games']     4.3         2.9K
2  ['nintendo', 'nintendo epd production group no…     4.4         4.3K
3                                 ['tobyfox', '8-4']     4.2         3.5K
4                                    ['team cherry']     4.4           3K

  number_of_reviews                                              genres  \
0             3.9K                                  ['adventure', 'rpg']
1             2.9K           ['adventure', 'brawler', 'indie', 'rpg']
2             4.3K                                  ['adventure', 'rpg']
3             3.5K  ['adventure', 'indie', 'rpg', 'turn based stra…
4               3K                 ['adventure', 'indie', 'platform']

                                              summary  \
0  Elden Ring is a fantasy, action and open world…
1  A rogue-lite hack and slash dungeon crawler in…
2  The Legend of Zelda: Breath of the Wild is the…
3  A small child falls into the Underground, wher…
4  A 2D metroidvania with an emphasis on close co…

                                             reviews plays playing backlogs  \
0  ["The first playthrough of elden ring is one o…   17K    3.8K     4.6K
1  ['convinced this is a roguelike for people who…   21K    3.2K     6.3K
2  ['This game is the game (that is not CS:GO) th…   30K    2.5K       5K
3  ['soundtrack is tied for #1 with nier automata…   28K     679     4.9K
4  ["this games worldbuilding is incredible, with…   21K    2.4K     8.3K

  wishlist
0     4.8K
1     3.6K
2     2.6K
3     1.8K
4     2.3K
```

[25]:
```
games_df.head()
```

[25]:
```
                                  title  release_date  \
0                            elden ring  Feb 25, 2022
1                                 hades  Dec 10, 2019
2  the legend of zelda: breath of the wild  Mar 03, 2017
3                             undertale  Sep 15, 2015
4                         hollow knight  Feb 24, 2017

                                                team  rating times_listed  \
0     ['bandai namco entertainment', 'fromsoftware']     4.5         3.9K
1                               ['supergiant games']     4.3         2.9K
```

```
2  ['nintendo', 'nintendo epd production group no…     4.4         4.3K
3                               ['tobyfox', '8-4']     4.2         3.5K
4                                  ['team cherry']     4.4           3K

  number_of_reviews                                              genres  \
0             3.9K                                 ['adventure', 'rpg']
1             2.9K           ['adventure', 'brawler', 'indie', 'rpg']
2             4.3K                                 ['adventure', 'rpg']
3             3.5K  ['adventure', 'indie', 'rpg', 'turn based stra…
4               3K                ['adventure', 'indie', 'platform']

                                              summary  \
0  Elden Ring is a fantasy, action and open world…
1  A rogue-lite hack and slash dungeon crawler in…
2  The Legend of Zelda: Breath of the Wild is the…
3  A small child falls into the Underground, wher…
4  A 2D metroidvania with an emphasis on close co…

                                             reviews plays playing backlogs  \
0  ["The first playthrough of elden ring is one o…   17K    3.8K     4.6K
1  ['convinced this is a roguelike for people who…   21K    3.2K     6.3K
2  ['This game is the game (that is not CS:GO) th…   30K    2.5K       5K
3  ['soundtrack is tied for #1 with nier automata…   28K     679     4.9K
4  ["this games worldbuilding is incredible, with…   21K    2.4K     8.3K

  wishlist
0    4.8K
1    3.6K
2    2.6K
3    1.8K
4    2.3K
```

[26]:
```
sales_df.head()
```

[26]:
```
   Rank                     Name Platform    Year         Genre Publisher  \
0     1               Wii Sports      Wii  2006.0        Sports  Nintendo
1     2         Super Mario Bros.      NES  1985.0      Platform  Nintendo
2     3            Mario Kart Wii      Wii  2008.0        Racing  Nintendo
3     4         Wii Sports Resort      Wii  2009.0        Sports  Nintendo
4     5  Pokemon Red/Pokemon Blue       GB  1996.0  Role-Playing  Nintendo

   NA_Sales  EU_Sales  JP_Sales  Other_Sales  Global_Sales
0     41.49     29.02      3.77         8.46         82.74
1     29.08      3.58      6.81         0.77         40.24
2     15.85     12.88      3.79         3.31         35.82
3     15.75     11.01      3.28         2.96         33.00
4     11.27      8.89     10.22         1.00         31.37
```

[27]:
```
print("Games dataset shape:", games_df.shape)
print("Sales dataset shape:", sales_df.shape)

print("\nGames columns:\n", games_df.columns)
print("\nSales columns:\n", sales_df.columns)
```

```
Games dataset shape: (1512, 13)
Sales dataset shape: (16598, 11)

Games columns:
 Index(['title', 'release_date', 'team', 'rating', 'times_listed',
       'number_of_reviews', 'genres', 'summary', 'reviews', 'plays', 'playing',
       'backlogs', 'wishlist'],
      dtype='object')

Sales columns:
 Index(['Rank', 'Name', 'Platform', 'Year', 'Genre', 'Publisher', 'NA_Sales',
       'EU_Sales', 'JP_Sales', 'Other_Sales', 'Global_Sales'],
      dtype='object')
```

[33]:
```
sales_df.columns
```

[33]:
```
Index(['Rank', 'Name', 'Platform', 'Year', 'Genre', 'Publisher', 'NA_Sales',
       'EU_Sales', 'JP_Sales', 'Other_Sales', 'Global_Sales'],
      dtype='object')
```

[35]:
```
sales_df.columns = (
    sales_df.columns
    .str.strip()
    .str.lower()
    .str.replace(' ', '_')
)

sales_df.columns
```

[35]:
```
Index(['rank', 'name', 'platform', 'year', 'genre', 'publisher', 'na_sales',
       'eu_sales', 'jp_sales', 'other_sales', 'global_sales'],
      dtype='object')
```

[36]:
```
text_cols = ['name', 'platform', 'genre', 'publisher']

for col in text_cols:
    sales_df[col] = sales_df[col].astype(str).str.strip().str.lower()
```

[37]:
```
sales_df['year'] = sales_df['year'].fillna(sales_df['year'].median())
sales_df['publisher'] = sales_df['publisher'].fillna('unknown')
```

[38]:
```
merged_df = pd.merge(
    games_df,
    sales_df,
    left_on='title',
    right_on='name',
    how='inner'
)
```

[39]:
```
merged_df.shape
```

[39]:
```
(1384, 24)
```

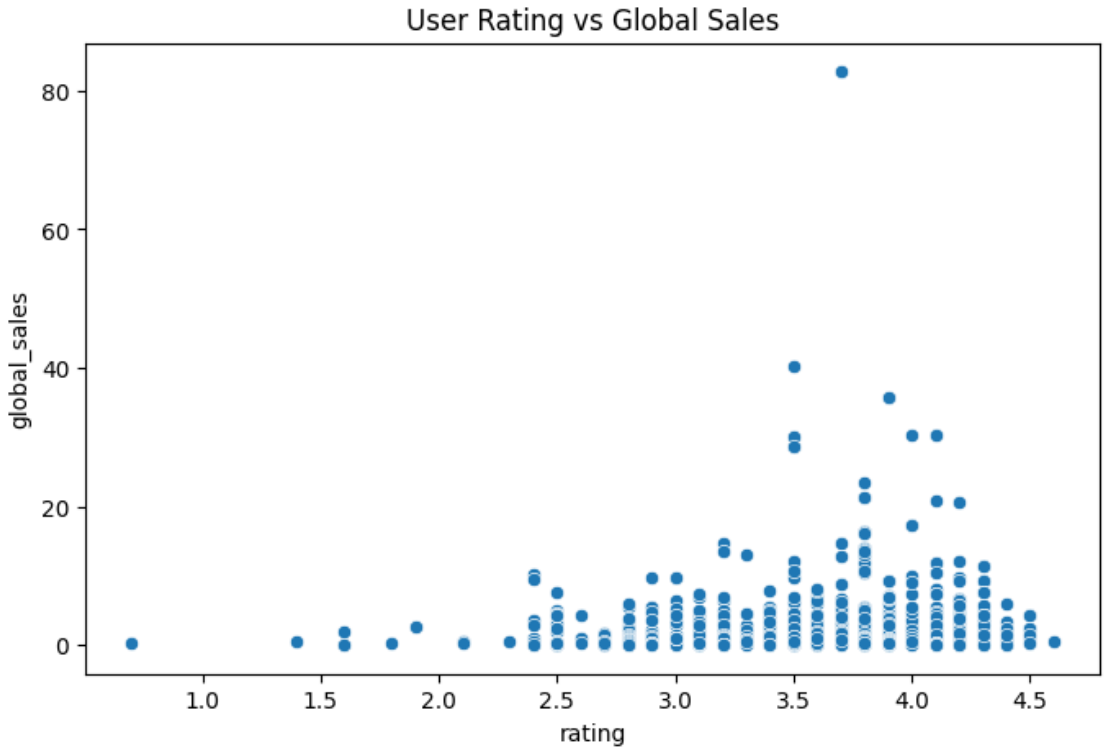
[40]:
```
merged_df.head()
```

[40]:

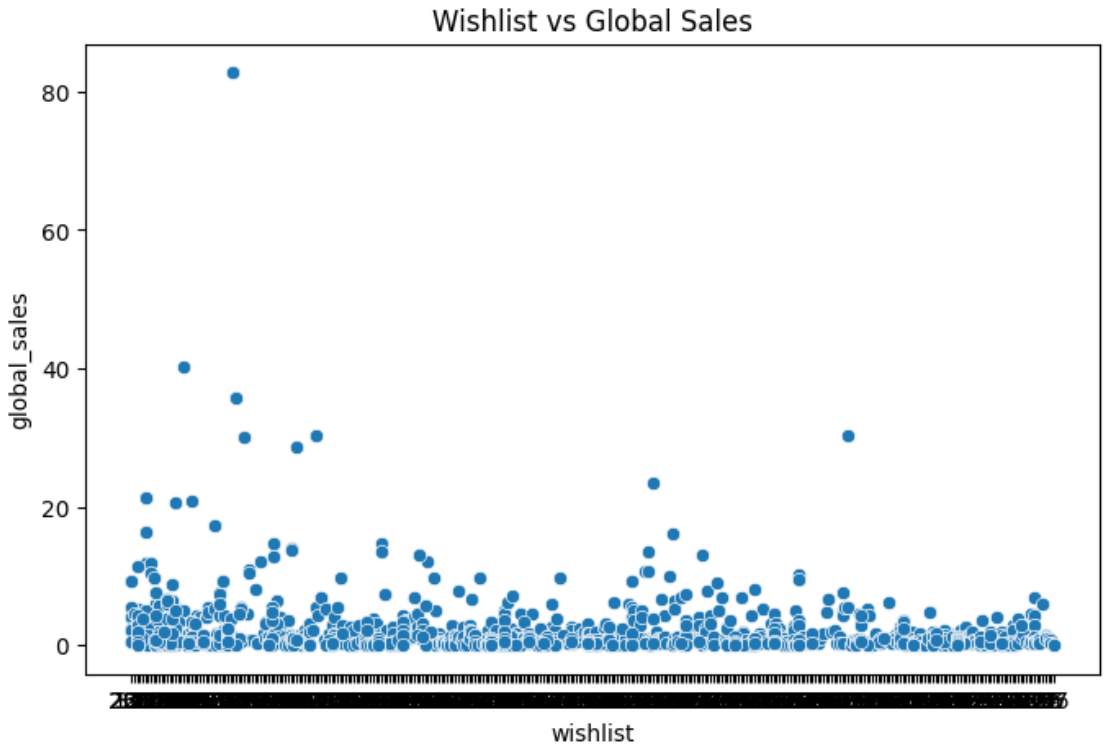
	title	release_date	team	rating	times_listed	number_of_reviews	genres	summary	reviews	plays	…	name	platform	year	genre	publisher	na_sales
0	minecraft	Nov 18, 2011	['mojang studios']	4.3	2.3K	2.3K	['adventure', 'simulator']	Minecraft focuses on allowing the player to ex…	['Minecraft is what you make of it. Unfortunat…	33K	…	minecraft	x360	2013.0	misc	microsoft game studios	5.58
1	minecraft	Nov 18, 2011	['mojang studios']	4.3	2.3K	2.3K	['adventure', 'simulator']	Minecraft focuses on allowing the player to ex…	['Minecraft is what you make of it. Unfortunat…	33K	…	minecraft	ps3	2014.0	misc	sony computer entertainment	1.97
2	minecraft	Nov 18, 2011	['mojang studios']	4.3	2.3K	2.3K	['adventure', 'simulator']	Minecraft focuses on allowing the player to ex…	['Minecraft is what you make of it. Unfortunat…	33K	…	minecraft	ps4	2014.0	misc	sony computer entertainment europe	1.38
3	minecraft	Nov 18, 2011	['mojang studios']	4.3	2.3K	2.3K	['adventure', 'simulator']	Minecraft focuses on allowing the player to ex…	['Minecraft is what you make of it. Unfortunat…	33K	…	minecraft	xone	2014.0	misc	microsoft game studios	1.43
4	minecraft	Nov 18, 2011	['mojang studios']	4.3	2.3K	2.3K	['adventure', 'simulator']	Minecraft focuses on allowing the player to ex…	['Minecraft is what you make of it. Unfortunat…	33K	…	minecraft					

```
4      psv  2014.0   misc  sony computer entertainment europe     0.28

   eu_sales jp_sales other_sales  global_sales
0      2.83     0.02        0.77          9.20
1      2.51     0.00        0.94          5.42
2      1.87     0.12        0.65          4.02
3      0.76     0.00        0.22          2.41
4      0.79     0.87        0.32          2.25

[5 rows x 24 columns]
```

[41]:
```
import seaborn as sns
import matplotlib.pyplot as plt

plt.figure(figsize=(8,5))
sns.scatterplot(x='rating', y='global_sales', data=merged_df)
plt.title("User Rating vs Global Sales")
plt.show()
```



[42]:
```
plt.figure(figsize=(8,5))
sns.scatterplot(x='wishlist', y='global_sales', data=merged_df)
plt.title("Wishlist vs Global Sales")
plt.show()
```



[43]:
```
merged_df.groupby('genre')['global_sales'].sum().sort_values(ascending=False).
↪head(10)
```

[43]:
```
genre
action          984.78
platform        651.05
shooter         532.48
role-playing    397.69
sports          279.60
racing          198.01
puzzle          135.33
misc            133.87
fighting         89.26
adventure        65.04
Name: global_sales, dtype: float64
```

[44]:
```
merged_df.to_csv("merged_games_sales.csv", index=False)
```

[45]:
```
from google.colab import files
files.download("merged_games_sales.csv")
```

```
<IPython.core.display.Javascript object>
```

```
<IPython.core.display.Javascript object>
```

[46]:
```
# Save cleaned datasets
games_df.to_csv("clean_games.csv", index=False)
sales_df.to_csv("clean_sales.csv", index=False)
```

[47]:
```
from google.colab import files

files.download("clean_games.csv")
files.download("clean_sales.csv")
```

```
<IPython.core.display.Javascript object>
```

```
<IPython.core.display.Javascript object>
```

```
<IPython.core.display.Javascript object>
```

```
<IPython.core.display.Javascript object>
```

[ ]: