Exploring Games Sales Trends

A Data Analysis Project

By Dyan Aqlima Febriyanti

Table of contents

1. Dataset Explanation

The game sales dataset encompasses total sales, release dates, developers, and genres, facilitating diverse analysis in the gaming industry. It allows insights into sales trends, temporal patterns, developer impacts, and genre-specific market dynamics, despite potential data limitations.

2. Importing Data

IMPORT THE LIBRARY

```
[1]  import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt
     import seaborn as sns
     %matplotlib inline
```

READ THE CSV FILE FROM LOCAL COMPUTER USING PANDAS

```
[2]  df=pd.read_csv('/Games Sales - Games (1).csv')
```

```
[3]  df.head()
```

	Name	Sales	Series	Release	Genre	Developer	Publisher
0	PlayerUnknown's Battlegrounds	42.0	NaN	12/1/2017	Battle royale	PUBG Studios	Krafton
1	Minecraft	33.0	Minecraft	11/1/2011	Sandbox, survival	Mojang Studios	Mojang Studios
2	Diablo III	20.0	Diablo	5/1/2012	Action role-playing	Blizzard Entertainment	Blizzard Entertainment
3	Garry's Mod	20.0	NaN	11/1/2006	Sandbox	Facepunch Studios	Valve
4	Terraria	17.2	NaN	5/1/2011	Action-adventure	Re-Logic	Re-Logic

3. Data Preparation
(Information about the dataset)

```
[5]  df.shape
```

```
(177, 7)
```

```
[6]  df.columns
```

```
Index(['Name', 'Sales', 'Series', 'Release', 'Genre', 'Developer',
       'Publisher'],
      dtype='object')
```

```
[7]  df.dtypes
```

```
Name          object
Sales        float64
Series        object
Release       object
Genre         object
Developer     object
Publisher     object
dtype: object
```

3. Data Preparation
(changing data type)

The 'Release' Column's Data Type was initially object; Hence, Conversion to *datetime* is required as it represents dates

```
[8]  # change data type
     df['Release'] = pd.to_datetime(df['Release'])
```

```
[9]  #recheck it
     df.dtypes
```

```
Name                    object
Sales                  float64
Series                  object
Release         datetime64[ns]
Genre                   object
Developer               object
Publisher               object
dtype: object
```

3. Data Preparation
(remove the duplicates)

```
#checking if there's any duplicate
df.loc[df.duplicated()]
```

	Name	Sales	Series	Release	Genre	Developer	Publisher
175	StarCraft II: Heart of the Swarm	1.0	StarCraft	2013-03-01	Real-time strategy	Blizzard Entertainment	Blizzard Entertainment
176	StarCraft II: Legacy of the Void	1.0	StarCraft	2015-11-01	Real-time strategy	Blizzard Entertainment	Blizzard Entertainment

```
#check the duplicate in Name columns
df.query('Name == "StarCraft II: Heart of the Swarm" ')
```

	Name	Sales	Series	Release	Genre	Developer	Publisher
157	StarCraft II: Heart of the Swarm	1.0	StarCraft	2013-03-01	Real-time strategy	Blizzard Entertainment	Blizzard Entertainment
175	StarCraft II: Heart of the Swarm	1.0	StarCraft	2013-03-01	Real-time strategy	Blizzard Entertainment	Blizzard Entertainment

Due to a Duplicate Entry in the 'Name' Column for 'Starcraft II' at line 175, so it has been drop from the dataset, and following the removal, the dataset now consists of 176 entries in the 'Name' Column

```
# drop the duplicate
df = df.drop(175)
```

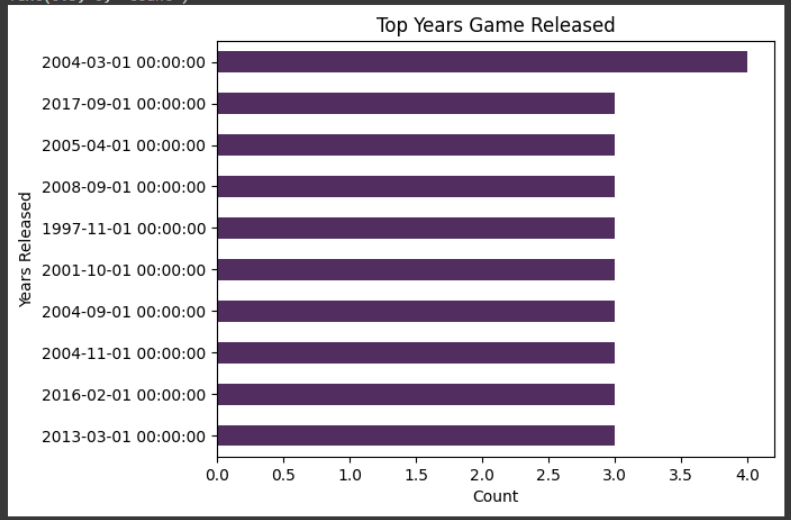
```
df.shape
```

```
(176, 7)
```

3. Data Visualization

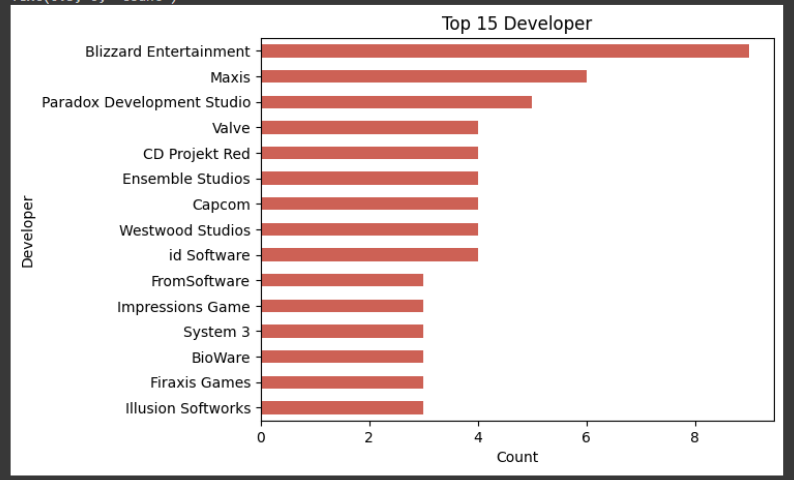
```
[14] ax = df['Release'].value_counts() \
       .head(10).sort_values(ascending=True) \
       .plot(kind='barh', title='Top Years Game Released', color='#512E5F')

     ax.set_ylabel('Years Released')
     ax.set_xlabel('Count')
```

```
Text(0.5, 0, 'Count')
```



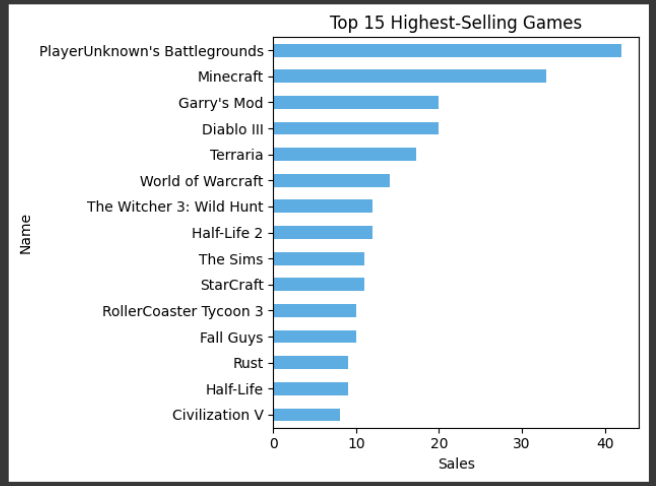

```
[15] ax = df['Developer'].value_counts() \
       .head(15).sort_values(ascending=True) \
       .plot(kind='barh', title='Top 15 Developer', color='#CD6155')

     ax.set_ylabel('Developer')
     ax.set_xlabel('Count')
```

```
Text(0.5, 0, 'Count')
```



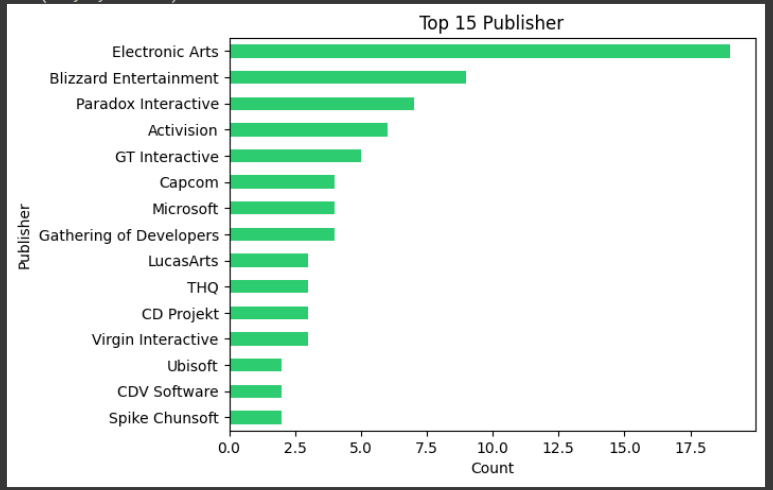
3. Data Visualization

```
[17] sales_by_name = df.groupby('Name')['Sales'].sum()

     top = sales_by_name.nlargest(15)

     ax = top.sort_values(ascending=True).plot(kind='barh', color='#5DADE2')
     ax.set_xlabel('Sales')
     ax.set_ylabel('Name')
     ax.set_title('Top 15 Highest-Selling Games')
     plt.tight_layout()
     plt.show()
```




```
[16] ax = df['Publisher'].value_counts() \
       .head(15).sort_values(ascending=True) \
       .plot(kind='barh', title='Top 15 Publisher', color='#2ECC71')

     ax.set_ylabel('Publisher')
     ax.set_xlabel('Count')
```

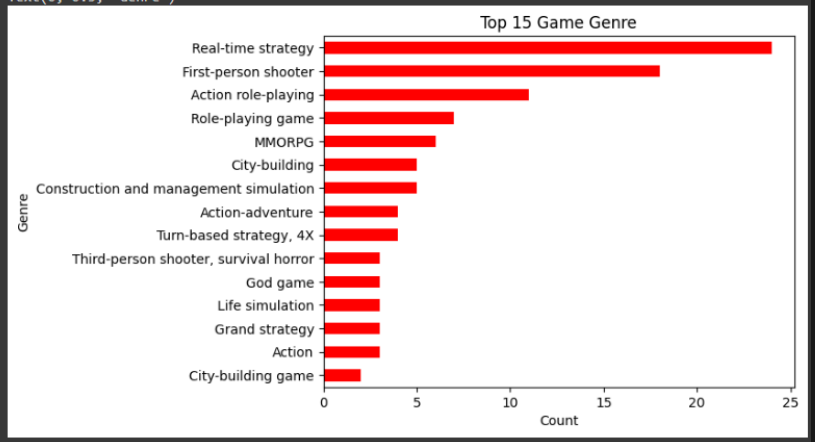
```
Text(0.5, 0, 'Count')
```



3. Data Visualization

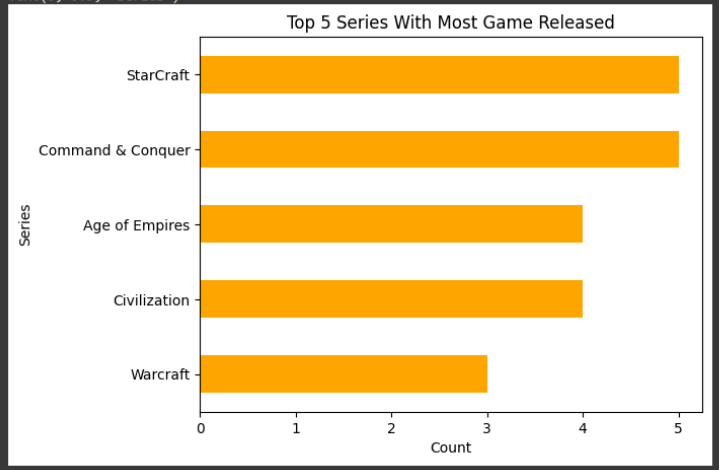
```
ax = df['Genre'].value_counts() \
  .head(15).sort_values(ascending=True) \
  .plot(kind='barh', title='Top 15 Game Genre', color='red')

ax.set_xlabel('Count')
ax.set_ylabel('Genre')
```

```
Text(0, 0.5, 'Genre')
```




```
ax = df['Series'].value_counts() \
  .head(5).sort_values(ascending=True) \
  .plot(kind='barh', title='Top 5 Series With Most Game Released', color='orange')

ax.set_xlabel('Count')
ax.set_ylabel('Series')
```

```
Text(0, 0.5, 'Series')
```



4. Conclusion

Blizzard Ent. Is The Top Developer

PUBG Is The Highest-Selling Games

Electronic Arts Is The Top Publisher

Real-Time Strategy: The Dominant Force in Gaming Genres

Starcraft Is Series With The Most Game Released

THANKS!

Feedback and suggestions are welcome!


dyanaqlima323@gmail.com


Dyan Aqlima Febriyanti


dyanaqlima323