

By Dyan Aqlima Febriyanti

Table of contents

01. Dataset Explanation

Description of the dataset, Include details like the source of the data, the variables (columns) etc

02.Importing
Data

Explanation of how the dataset were imported plus including import the library

03.
Data
Preparation

Details on how data were cleaned, involves removing duplicate and changing data types

04.
Data
Visualization

Presentation of visualizations created during the analysis process. This section showcases the graphs or charts

05.

Conclusion

Summarization of the key findings and insights

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]	<pre>df=pd.read_csv('/Games Sales - Games (1).csv')</pre>										
[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
                   object
    Name
    Sales
                  float64
    Series
                  object
    Release
                  object
                  object
    Genre
    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

```
# change data type
df['Release'] = pd.to datetime(df['Release'])
#recheck it
df.dtypes
                     object
Name
Sales
                    float64
                     object
Series
Release
             datetime64[ns]
Genre
                     object
Developer
                     object
Publisher
                     object
dtype: object
```

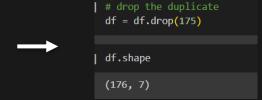
3. Data Preparation

(remove the duplicates)

#checking if there's any duplicate df.loc[df.duplicated()] **Publisher** Sales Series Release Genre Developer StarCraft II: Blizzard Blizzard 175 Heart of the Entertainment Entertainment Swarm StarCraft II: Blizzard Blizzard 176 Legacy of 1.0 StarCraft **Entertainment** Entertainment the Void

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











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')
                                               Top Years Game Released
         2004-03-01 00:00:00
         2017-09-01 00:00:00
         2005-04-01 00:00:00
         2008-09-01 00:00:00
         1997-11-01 00:00:00
         2001-10-01 00:00:00
         2004-09-01 00:00:00
         2004-11-01 00:00:00
         2016-02-01 00:00:00
         2013-03-01 00:00:00
                                   0.5
                                           1.0
                                                   1.5
                                                          2.0
                                                                  2.5
                                                                          3.0
                                                                                 3.5
                                                                                         4.0
```

Count

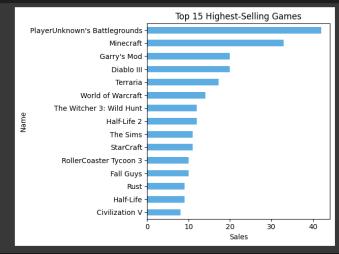
```
[15] ax = df['Developer'].value counts() \
       .head(15).sort values(ascending=True) \
       .plot(kind='barh', title='Top 15 Developer', color='#CD6155')
     ax.set_ylabel('Developer')
     ax.set xlabel('Count')
     Text(0.5, 0, 'Count'
                                                           Top 15 Developer
               Blizzard Entertainment
                              Maxis
         Paradox Development Studio
                              Valve
                      CD Projekt Red
                   Ensemble Studios
                            Capcom
                   Westwood Studios
                         id Software
                       FromSoftware
                   Impressions Game
                           System 3
                            BioWare
                       Firaxis Games
                   Illusion Softworks
                                                                  Count
```

3. Data Visualization

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

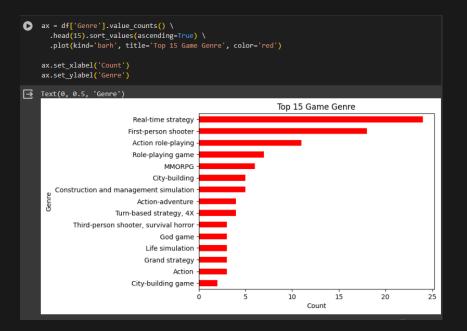
top = sales_by_name.nlargest(15)

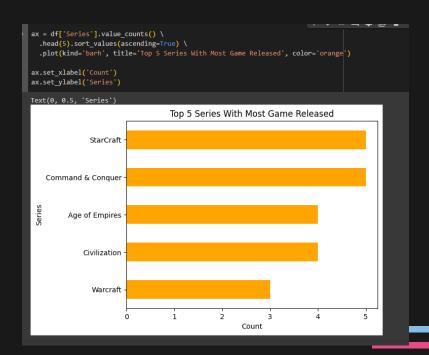
ax = top.sort_values(ascending=True).plot(kind='barh', color='#SDADE2')
ax.set_xlabel('Sales')
ax.set_ylabel('Name')
ax.set_title('Top 15 Highest-Selling Games')
plt.tight_layout()
plt.show()
```



```
[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')
                                                         Top 15 Publisher
                  Electronic Arts
           Blizzard Entertainment
              Paradox Interactive
                       Activision
                  GT Interactive
                        Capcom
                       Microsoft
         Gathering of Developers
                       LucasArts
                            THO
                      CD Projekt
               Virgin Interactive
                         Ubisoft
                   CDV Software
                  Spike Chunsoft
                               0.0
                                        2.5
                                                5.0
                                                        7.5
                                                                10.0
                                                                        12.5
                                                                                15.0
                                                                                       17.5
                                                               Count
```

3. Data Visualization





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





Feedback and suggestions are welcome!



dyanaqlima323@gmail.com



Dyan Aqlima Febriyanti



dyanaqlima323