

CASE STUDY

GAME SALES



Revou Data Analytics
Mini Course



by Nugroho Adi Wirapratama

TABLE OF CONTENT

01

PROBLEM QUESTION

02

TOOLS

03

DATA EXTRACTION AND
CLEANING (GOOGLE COLAB)

04

DATA VISUALIZATION

05

CONCLUSION





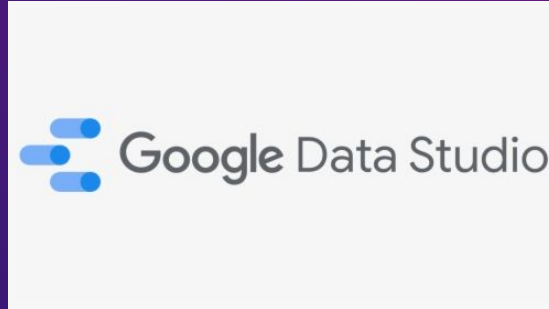
01

QUESTIONS

1. Which game is the oldest and the newest games in that dataset ?
2. Which publisher published most of the games ?
3. Which developer developed most of the games ?
4. Which series is the most sales ?
- 5. Which series have the most games ?



02 TOOLS



You can check my google colab [CLICK HERE](#)



03

DATA EXTRACTION

&

DATA CLEANING



Data Extraction & Data Cleaning



```
# Import library
import pandas as pd
import numpy as np
```



```
# Import Dataset from Google Drive
with open('/content/drive/MyDrive/Mini Course Revou/05 10 2022/Case Study/Games Sales - Games.csv','r') as dataset:
    df = pd.read_csv(dataset)
# Read Dataset
df
```



	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
...
172	Who Wants to Be a Millionaire?	1.0	NaN	11/1/1999	Trivia game	Jellyvision	Disney Interactive Studios
173	Wing Commander 3: Heart of the Tiger	1.0	Wing Commander	12/1/1994	Space combat simulation	Origin Systems	Electronic Arts
174	Zoo Tycoon	1.0	Zoo Tycoon	10/1/2001	Business simulation	Microsoft	Blue Fang Games
175	StarCraft II: Heart of the Swarm	1.0	StarCraft	3/1/2013	Real-time strategy	Blizzard Entertainment	Blizzard Entertainment
176	StarCraft II: Legacy of the Void	1.0	StarCraft	11/1/2015	Real-time strategy	Blizzard Entertainment	Blizzard Entertainment

177 rows × 7 columns



Data Extraction & Data Cleaning

```
# information dataset check  
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 177 entries, 0 to 176  
Data columns (total 7 columns):  
#   Column      Non-Null Count  Dtype    
---  ---        
0   Name        177 non-null    object   
1   Sales       177 non-null    float64  
2   Series      141 non-null    object   
3   Release     177 non-null    object   
4   Genre       177 non-null    object   
5   Developer   177 non-null    object   
6   Publisher   177 non-null    object   
dtypes: float64(1), object(6)  
memory usage: 9.8+ KB
```

```
# sum of data null in dataset  
df.isna().sum()
```

```
Name      0  
Sales     0  
Series    36  
Release   0  
Genre     0  
Developer 0  
Publisher 0  
dtype: int64
```

```
# sum of Data Duplicates  
print("Sum Data Duplicates =", df.duplicated().sum())  
  
# column table containing duplicate data  
df[df.duplicated()]
```

```
Sum Data Duplicates = 2
```

	Name	Sales	Series	Release	Genre	Developer	Publisher
175	StarCraft II: Heart of the Swarm	1.0	StarCraft	3/1/2013	Real-time strategy	Blizzard Entertainment	Blizzard Entertainment
176	StarCraft II: Legacy of the Void	1.0	StarCraft	11/1/2015	Real-time strategy	Blizzard Entertainment	Blizzard Entertainment

Have 3 issues that need to be cleaned

1. Change the 'Release' column data type to datetime
2. There are 2 duplicates data that appears
3. In 'Series' row, there are 36 null rows



Solve 3 Issues

```
df_cleaned = df.copy()

# Convert Release to Date format = (yyyy-dd-mm)
df_cleaned['Release'] = pd.to_datetime(df_cleaned['Release'])

# Change null row in Series Column by '-'
df_cleaned['Series'] = df_cleaned['Series'].fillna('-')

# Remove Duplicated Dataset
df_cleaned = df_cleaned.drop_duplicates(keep='first')
```

```
# Check Dataset After Drop the Duplicated Data
print('Sum Data Duplicates =', df_cleaned.duplicated().sum())
df_cleaned[df_cleaned.duplicated()]
```

```
Sum Data Duplicates = 0
```

Name	Sales	Series	Release	Genre	Developer	Publisher
------	-------	--------	---------	-------	-----------	-----------

```
df_cleaned.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 175 entries, 0 to 174
Data columns (total 7 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Name        175 non-null   object
1   Sales       175 non-null   float64
2   Series      175 non-null   object
3   Release     175 non-null   datetime64[ns]
4   Genre       175 non-null   object
5   Developer   175 non-null   object
6   Publisher   175 non-null   object
dtypes: datetime64[ns](1), float64(1), object(5)
memory usage: 15.0+ KB
```



04

DATA VISUALIZATION



```
df_cleaned.to_csv("Game Sales Cleaned.csv", index=False)
```

export cleaned dataset and move to GDS for visualization

OLDEST AND NEWEST GAMES

```
release_sort = df_cleaned.sort_values(by='Release')
games_oldes_newest = pd.concat([release_sort.head(1), release_sort.tail(1)])
games_oldes_newest
```

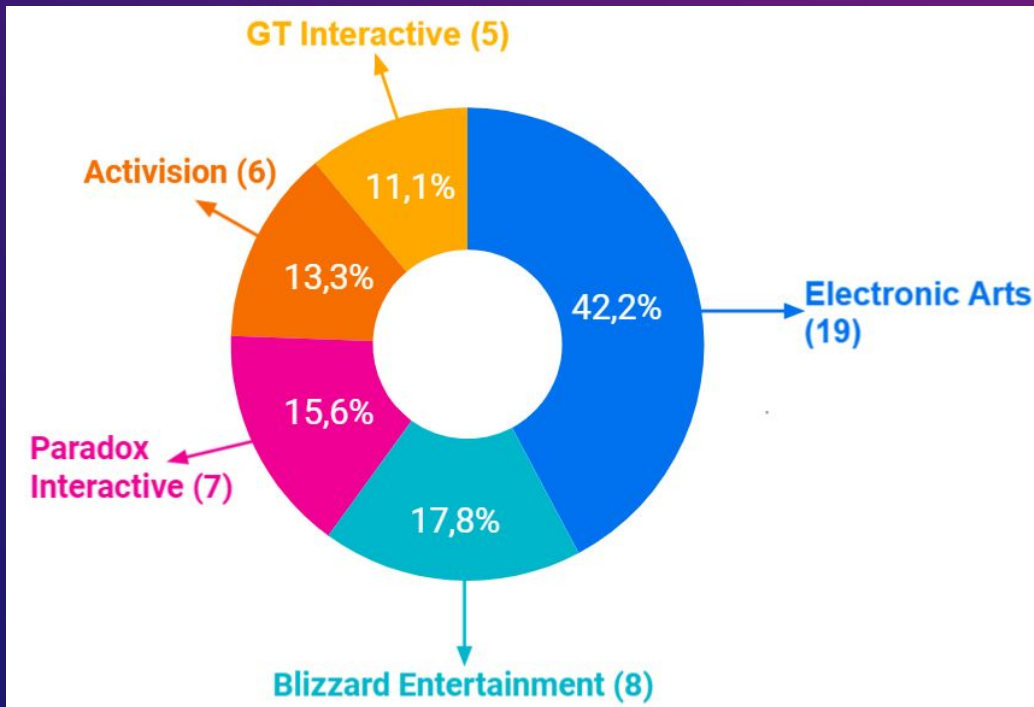
	Name	Sales	Series	Release	Genre	Developer	Publisher
133	Hydlide	1.0	Hydlide	1984-01-12	Action role-playing	Technology and Entertainment Software	Technology and Entertainment Software
20	Valheim	6.0	-	2021-01-02	Survival	Iron Gate	Coffee Stain Publishing

Oldest Game = Hydlide (**December 1, 1984** Release Date)

Newest Game = Valheim (**February 1, 2021** Release Date)



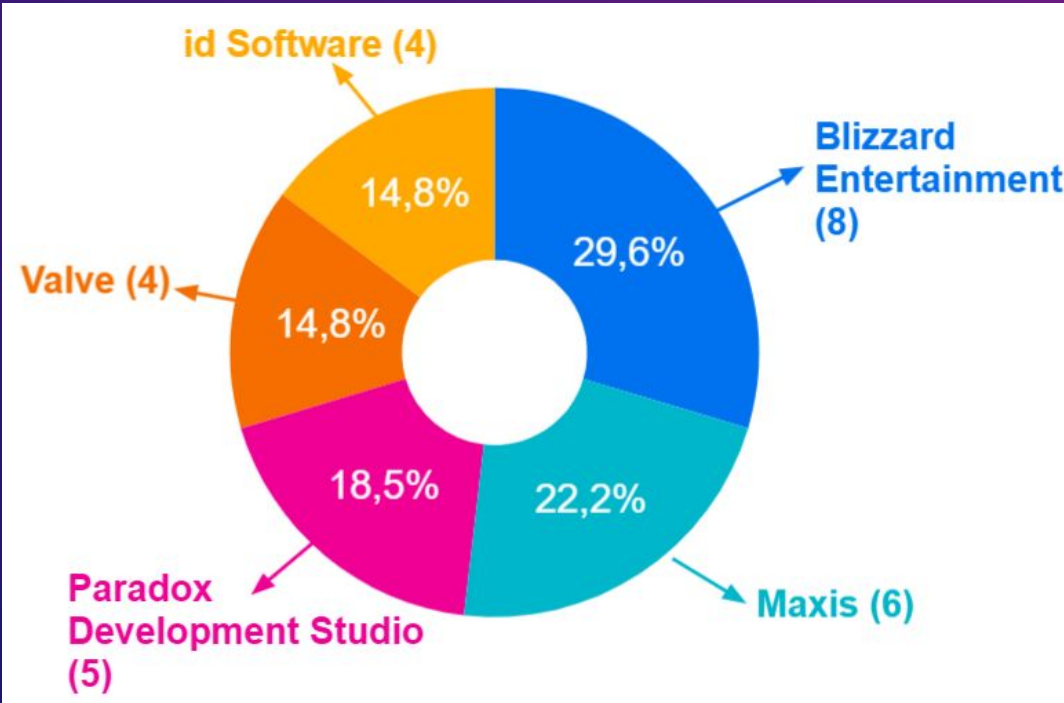
PUBLISHER WITH THE MOST GAMES PUBLISHED



Electronic arts publisher is the most games published with a total of 19 games



DEVELOPER WITH THE MOST DEVELOPED GAMES



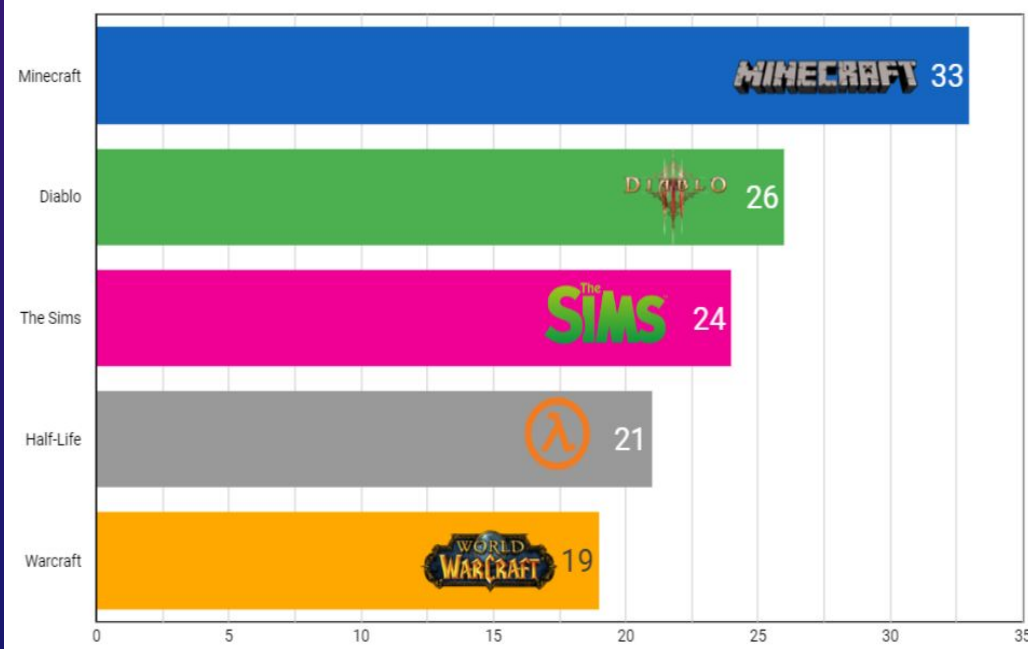
Blizzard Entertainment
Developer is the most developed games with a total of 8 games



SERIES BASED ON THE MOST SALES

Minecraft is the Series with the most number of sales with a total of 33

MINECRAFT





SERIES BASED ON MOST GAMES

Command & Conquer Series is the series with the most games with a total of 5 games

**COMMAND
&
CONQUER®**



04 CONCLUSION



OLDEST & NEWEST GAMES

Oldest Game = Hydlide
(December 1, 1984 Release Date)
Newest Game = Valheim
(February 1, 2021 Release Date)

PUBLISHER WITH THE MOST GAMES PUBLISHED

Electronic Arts, with a total of 19 games

DEVELOPER WITH THE MOST DEVELOPED GAMES

Blizzard Entertainment, with a total of 8 games

SERIES BASED ON THE MOST SALES **MINECRAFT**

Minecraft, with a total of 33 sales

SERIES BASED ON MOST GAMES **COMMAND & CONQUER**

Command & Conqueror, with a total of 5 games



Please keep this slide for attribution

