

Septa Dwi Cahya

Case Study Data Analytics Mini Course at Revou
“Game Sales”

Introduction Project

— — —

This project discusses about game sales from 1984 to 2021. We will get insight about the oldest and the newest games, publisher/developer who published most of the games, and the series that have the most sales or the most games. For this purpose, the Dataset will be cleaned and analyzed using python then visualized using looker studio.



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

Read Data and Clean

```
[2] # read data from google sheet
sheet_url = 'https://docs.google.com/spreadsheets/d/1cZQvX1GjCGoPZLCUsIHPSBaF35_0tyktV0jpvqvtgY/edit#gid=1485085913'
sheet_url_trf = sheet_url.replace('/edit#gid=', '/export?format=csv&gid=')
df = pd.read_csv(sheet_url_trf)
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] 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
```

```
for col_name in ['Name', 'Sales', 'Series', 'Release', 'Genre', 'Developer',
                 'Publisher']:
    print(col_name, df[col_name].isnull().sum(), '\n')
```

Name 0
Sales 0
Series 36
Release 0
Genre 0
Developer 0
Publisher 0

```
df_cleaned = df.fillna('Unknown Series')
df_cleaned['Release'] = pd.to_datetime(df_cleaned['Release'])
df_cleaned.head()
```

	Name	Sales	Series	Release	Genre	Developer	Publisher
0	PlayerUnknown's Battlegrounds	42.0	Unknown Series	2017-12-01	Battle royale	PUBG Studios	Krafton
1	Minecraft	33.0	Minecraft	2011-11-01	Sandbox, survival	Mojang Studios	Mojang Studios
2	Diablo III	20.0	Diablo	2012-05-01	Action role-playing	Blizzard Entertainment	Blizzard Entertainment
3	Garry's Mod	20.0	Unknown Series	2006-11-01	Sandbox	Facepunch Studios	Valve
4	Terraria	17.2	Unknown Series	2011-05-01	Action-adventure	Re-Logic	Re-Logic

Read and Clean Data

Read dataset from spreadsheet. Then change datatype of Release from 'object' to 'datetime'.

The NaN value in the Series column is filled by Unknown Series.

Question 1: Which game is the oldest and the newest games in that dataset?

```
[10] df_cleaned.sort: df_cleaned.sort_values(by='Release')
df_cleaned.sort.head()
```

	Name	Sales	Series	Release	Genre	Developer	Publisher
133	Hydlide	1.0	Hydlide	1984-12-01	Action role-playing	Technology and Entertainment Software	Technology and Entertainment Software
34	Where in the World Is Carmen Sandiego?	4.0	Carmen Sandiego	1985-06-01	Educational	Broderbund	Broderbund
88	International Karate	1.5	International Karate	1985-11-01	Fighting	System 3	Epyx
162	Tetris	1.0	Tetris	1988-01-01	Puzzle	Spectrum HoloByte	Spectrum HoloByte
22	Last Ninja 2	5.5	The Last Ninja	1988-08-01	Action-adventure	System 3	Activision

```
[11] df_cleaned.sort.tail()
```

	Name	Sales	Series	Release	Genre	Developer	Publisher
139	Mordhau	1.0	Unknown Series	2019-04-01	Action	Tritemion	Tritemion
10	Fall Guys	10.0	Unknown Series	2020-08-01	Battle royale	Mediatonic	Devolver Digital
110	Crusader Kings III	1.0	Crusader Kings	2020-09-01	Grand strategy	Paradox Development Studio	Paradox Interactive
27	Cyberpunk 2077	4.5	Unknown Series	2020-12-01	Action role-playing	CD Projekt Red	CD Projekt
20	Valheim	6.0	Unknown Series	2021-02-01	Survival	Iron Gate	Coffee Stain Publishing

The oldest game is **Hydlide** released on 1st Dec 1984, while the newest game is **Valheim** released on 1st Feb 2021.

Question 2: Which publisher published most of the games?

```
[12] agg_publisher = df_cleaned.groupby('Publisher', as_index=False)['Name'].nunique().sort_values(by='Name', ascending=False)
agg_publisher.head()
```

	Publisher	Name
28	Electronic Arts	19
8	Blizzard Entertainment	8
62	Paradox Interactive	7
2	Activision	6
37	GT Interactive	5

Electronic Arts published most of the games (published 19 games) followed by Blizzard Entertainment, Paradox Interactive, Activision, GT Interactive.

Question 3: Which developer developed most of the games?

```
[13] agg_developer = df_cleaned.groupby('Developer', as_index=False)['Name'].nunique().sort_values(by='Name', ascending=False)
agg_developer.head()
```

	Developer	Name
8	Blizzard Entertainment	8
60	Maxis	6
70	Paradox Development Studio	5
108	id Software	4
106	Westwood Studios	4

Question 4: Which series is the most sales?

```
[14] agg_series = df_cleaned.groupby('Series', as_index=False)['Sales'].sum().sort_values(by='Sales', ascending=False)
agg_series.head()
```

	Series	Sales
83	Unknown Series	156.1
47	Minecraft	33.0
22	Diablo	26.0
75	The Sims	24.0
36	Half-Life	21.0

Blizzard Entertainment developed most of the games (developed 8 games) followed by Maxis, Paradox Development Studio, id Software, Westwood Studios.

Unknown series is the most sales followed by Minecraft, Diablo, The Sims, Half-Life.

Question 5: Which series have the most games?

```
[15] agg_series2 = df_cleaned.groupby('Series', as_index=False)['Name'].nunique().sort_values(by='Name', ascending=False)
agg_series2.head()
```

	Series	Name
83	Unknown Series	36
13	Command & Conquer	5
68	StarCraft	4
2	Age of Empires	4
12	Civilization	4

Unknown series have the most games followed by Command & Conquer, StarCraft, Age of Empires, Civilization.

You can visit this link for full script Python
https://colab.research.google.com/drive/10Ti_PmWBE8X02NvzD_yAeGHk4ppL8FZ1?usp=sharing

Dashboard Games Sales

using Looker Studio

<https://lookerstudio.google.com/reporting/9c939836-8c74-4a34-a5db-07f1b28b324a>

GAME SALES

Genre

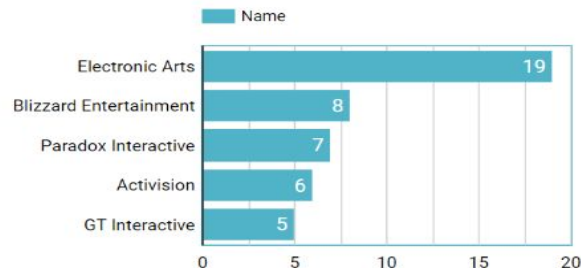
Select date range

Which game is the oldest and the newest games in that dataset?

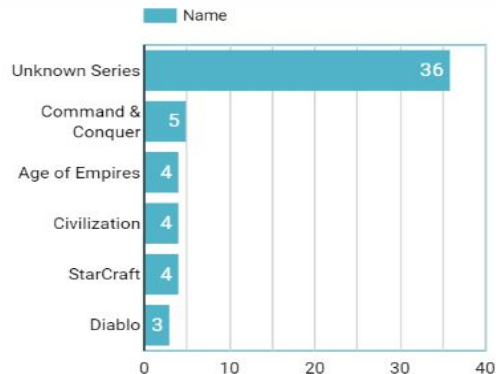
Name	Release
1. Hydlide	Dec 1, 1984
2. Where in the World Is Carmen Sandiego?	Jun 1, 1985
3. International Karate	Nov 1, 1985
4. Tetris	Jan 1, 1988
5. Last Ninja 2	Aug 1, 1988
1 - 5 / 175 < >	

Name	Release
1. Valheim	Feb 1, 2021
2. Cyberpunk 2077	Dec 1, 2020
3. Crusader Kings III	Sep 1, 2020
4. Fall Guys	Aug 1, 2020
5. Mordhau	Apr 1, 2019
1 - 5 / 175 < >	

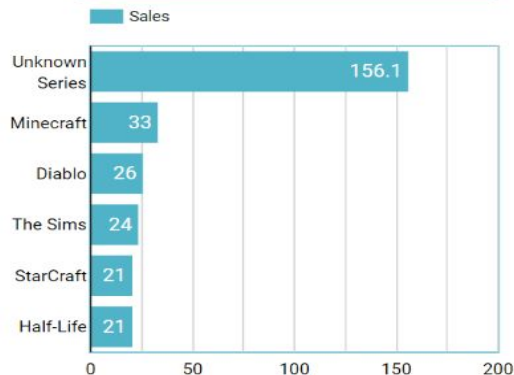
Which publisher published most of the games?



Which series have the most games?



Which series is the most sales?



Which Developer developed most of the games?

