RevoU
Mini Course

Data Analytics Case Study :
Games Sales

Febe Jovita

Table of Interest

```
sheet_url = 'https://docs.google.com/spreadsheets/d/10poofg-l8DMdtUgGy8mOpra2IHmA9EQC7drmF9AyYHA/edit#gid=1485085913'
sheet_url_trf = sheet_url.replace('/edit#gid=', '/export?format=csv&gid=')
df = pd.read_csv(sheet_url_trf)
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

Questions to be answered

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

- Which publisher published most of the games ?

- Which developer developed most of the games ?

- Which series is the most sales ?

- Which series have the most games ?

Tools

Grand Theft Auto V

Data Cleaning

- Drop Duplicates

  df = df.drop_duplicates()

  df.info()

- Check Empty Values

  df.isna().sum()

- Change Datatype

  df['Release'] = pd.to_datetime(df['Release'], format='%m/%d/%Y')

  df.info()

  For more details, you can view Jupiter Notebook here
  https://colab.research.google.com/drive/10gBDuwZ8Fhl6KwzcLBwRt
  UCM2GU42evV#scrollTo=POwsD2kOoyRn

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

The oldest games is Hydlide and the newest game is Valheim.

```
[ ] df[['Name','Release']]
```

	Name	Release
0	PlayerUnknown's Battlegrounds	2017-12-01
1	Minecraft	2011-11-01
2	Diablo III	2012-05-01
3	Garry's Mod	2006-11-01
4	Terraria	2011-05-01
...	...	...
170	Vietcong	2003-03-01
171	Warhammer Online: Age of Reckoning	2008-09-01
172	Who Wants to Be a Millionaire?	1999-11-01
173	Wing Commander 3: Heart of the Tiger	1994-12-01
174	Zoo Tycoon	2001-10-01

175 rows × 2 columns

```
[ ] #Oldest Games
    df[['Name','Release']].sort_values('Release').head(1)
```

	Name	Release
133	Hydlide	1984-12-01

```
[ ] #Newest Games
    df[['Name','Release']].sort_values('Release', ascending=False).head(1)
```
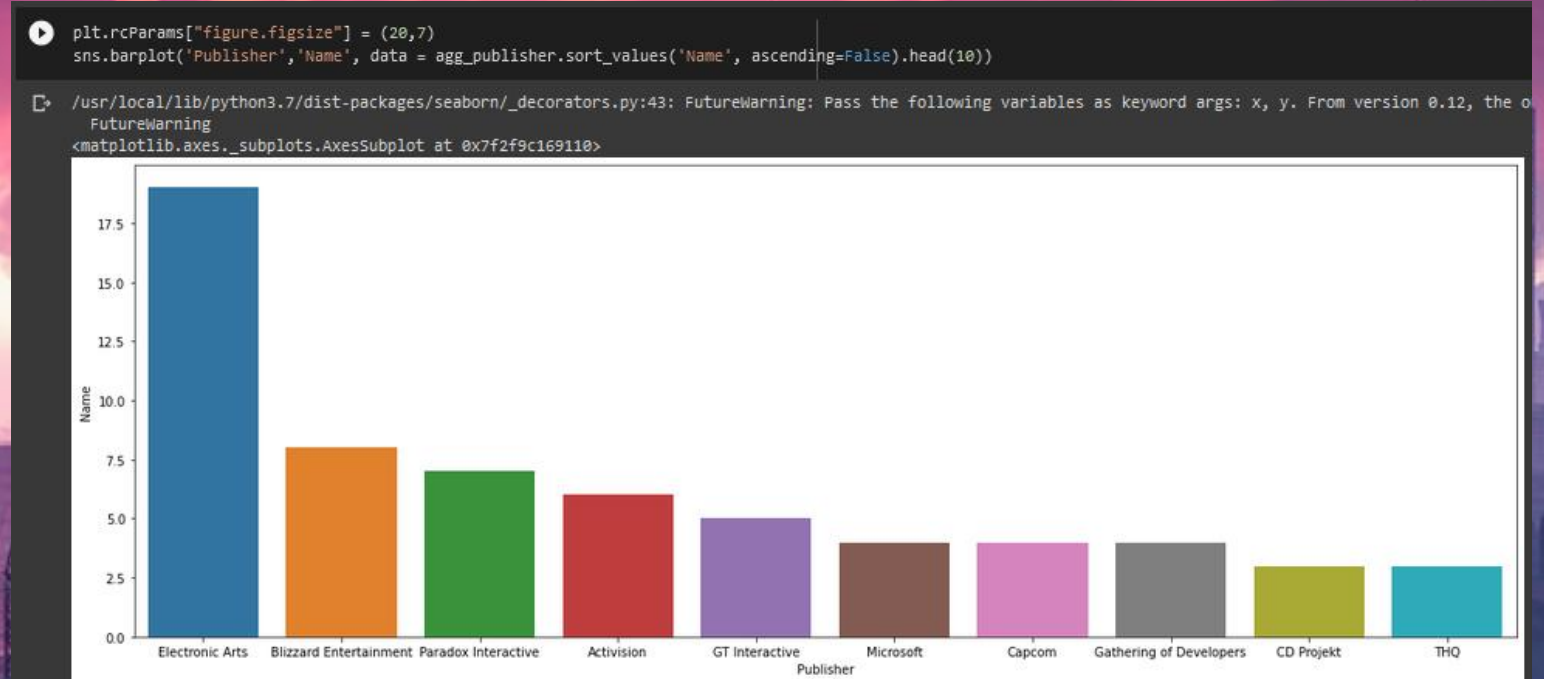
	Name	Release
20	Valheim	2021-02-01

Which publisher published most of the games ?

Electronic Arts with 19 published games.

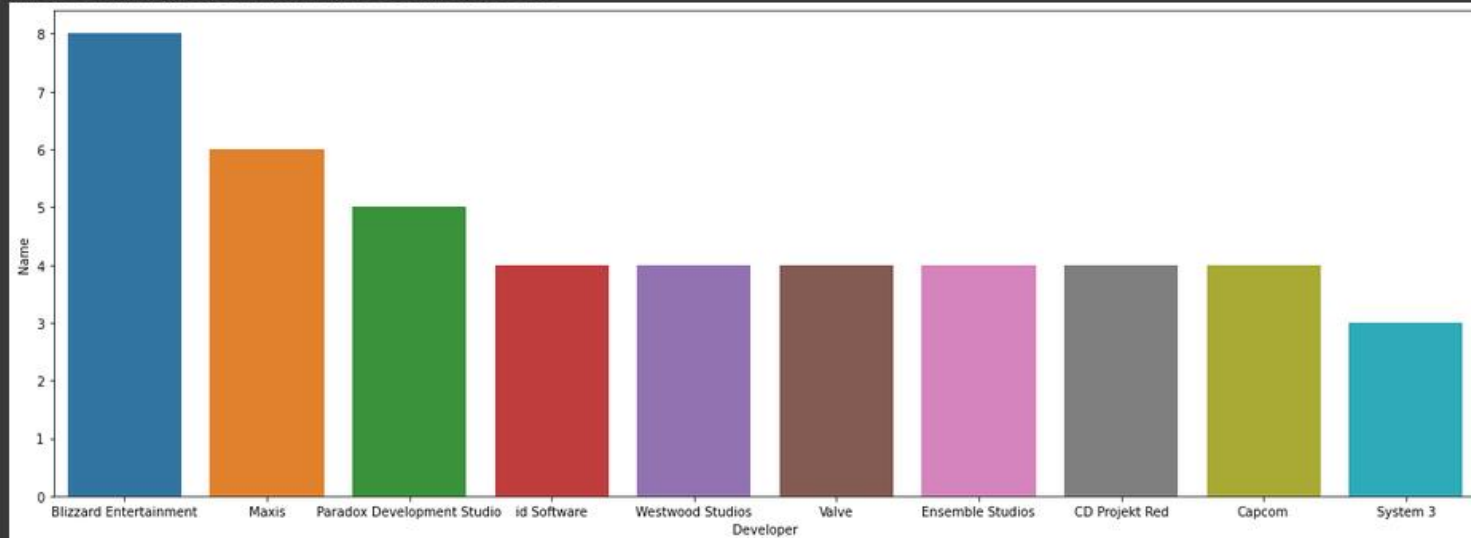# Which developer developed most of the games ?

Blizzard Ent.

```
[ ] agg_dev = df.groupby('Developer', as_index=False)['Name'].nunique()
    agg_dev.sort_values('Name', ascending=False)
```

	Developer	Name
6	Blizzard Entertainment	8
60	Maxis	6
70	Paradox Development Studio	5
108	id Software	4
106	Westwood Studios	4
...	...	...
48	Introversion Software	1
47	Infocom	1
44	Heuristic Park	1
42	Gray Matter Interactive	1
54	Kojima Productions	1

109 rows × 2 columns

```
sns.barplot('Developer','Name', data = agg_dev.sort_values('Name', ascending=False).head(10))
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the
  FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f2f9c02b710>
```


For more details, you can view Jupiter Notebook here
https://colab.research.google.com/drive/10gBDujwZ8FhI6KwzcLBwRt
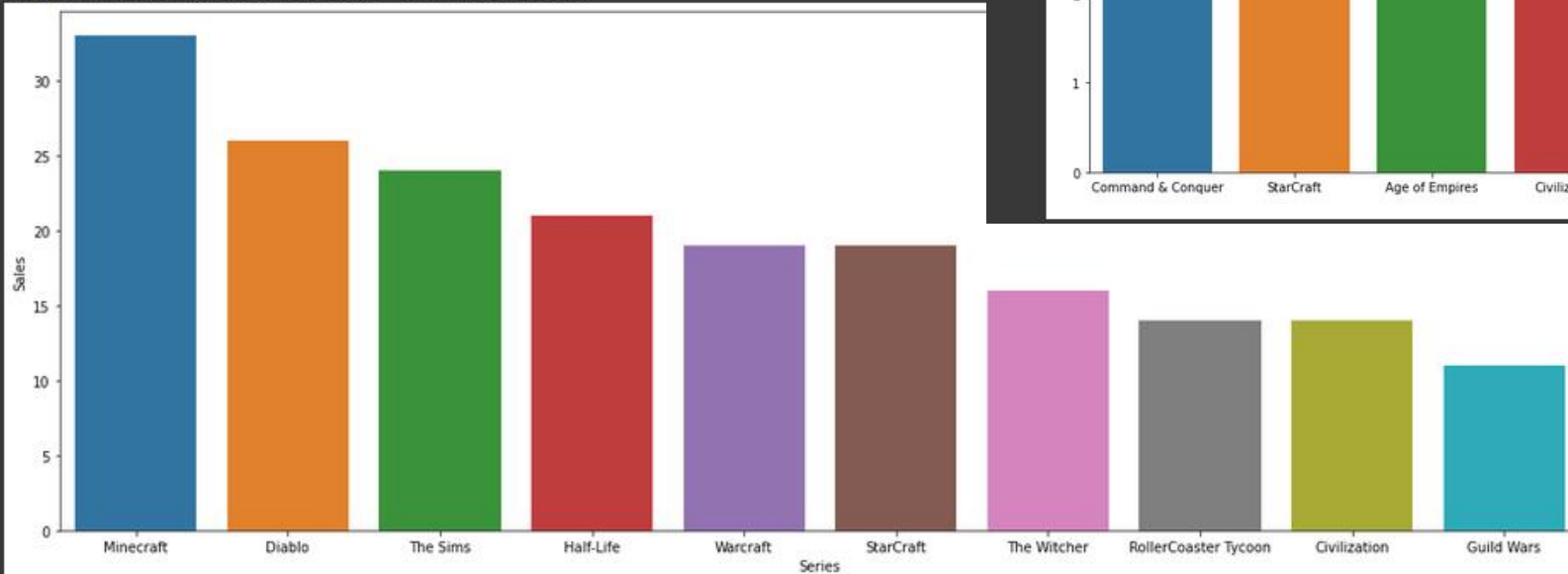UCM2GU42evV#scrollTo=POwsD2kQoyRn

Which series is the most sales ?
Which series have the most games ?

Series with most sales is **Minecraft** and series with most games is **Command & Conquer.**

```
sns.barplot('Series','Name', data = agg_series.sort_values('Name', ascending=False).head(10))
```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the
    FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f2f9be8dc90>
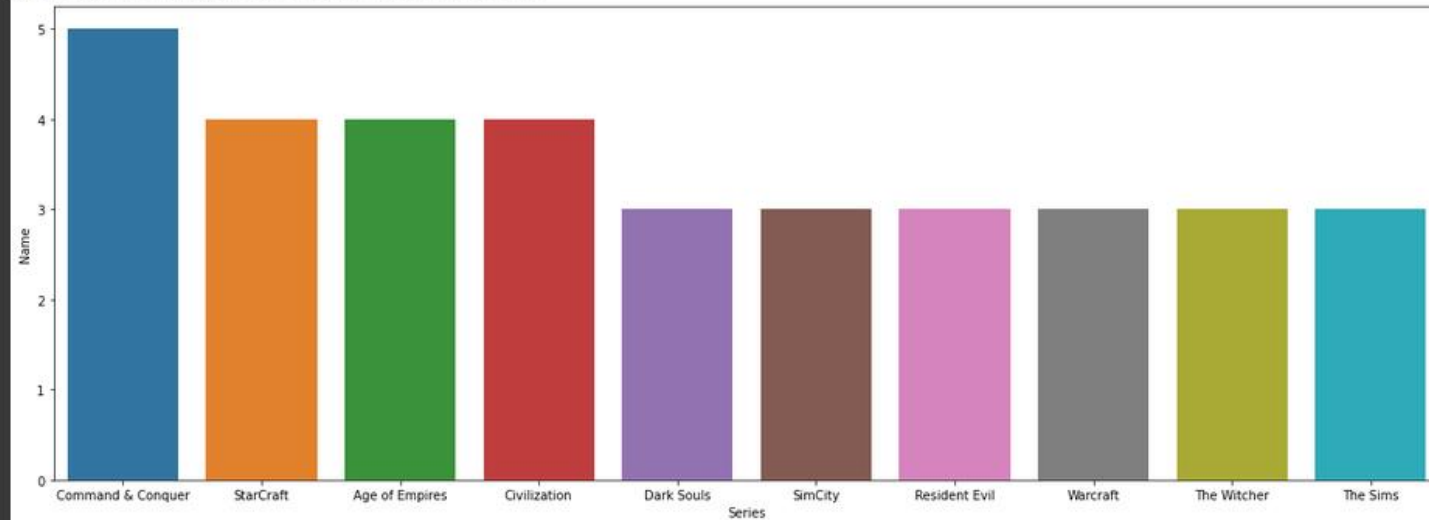


```
[ ] sns.barplot('Series','Sales', data = agg_series.sort_values('Sales', ascending=False).head(10))
```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the follow
    FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f2f9bf4e450>

Conclusion

1. The oldest games is Hydlide and the newest game is Valheim.

2. The publisher published most of the games is Electronic Arts with 19 published games.

3. The developer developed most of the games is Blizzard Ent.

4. Series with most sales is Minecraft and series with most games is Command & Conquer.

Recommendations

- Promotion : give attractive promotions to get more gamers.

- Championship : organize championships to increase the spirit of gamers.

- Sustainable : make many game variants for the best-selling series.

- Collaboration : enhance collaboration among developers, publishers and influencers.

- Innovation : make many games in metaverse.

Thank You

LinkedIn : www.linkedin.com/in/febe-jovita-7b1572246

Email : febejovita@gmail.com

Instagram : @fjovitaj