

Video Games Analysis

Capstone Project 2
By Michelle Ang

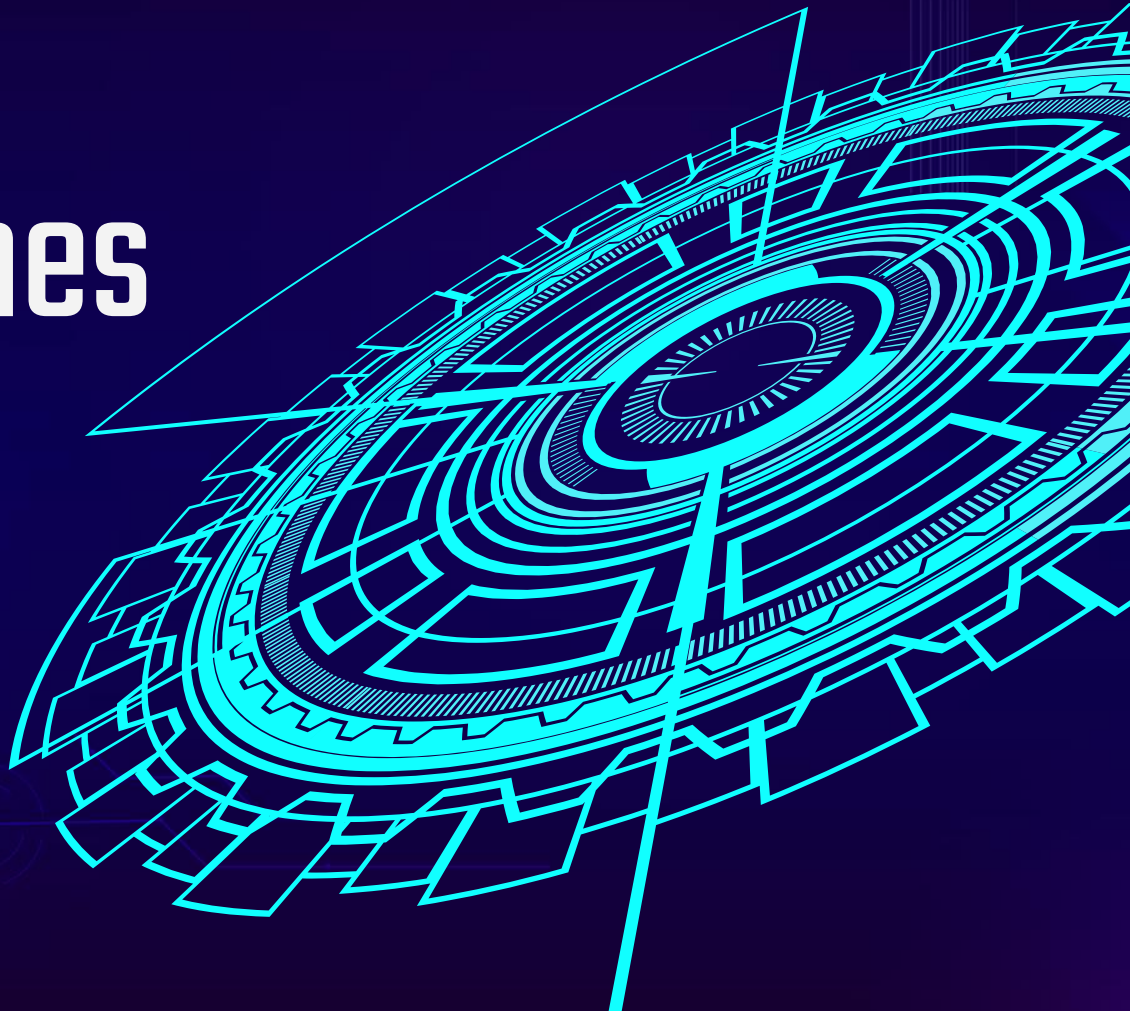




TABLE OF CONTENTS

01

Problem Statement

02

ER diagram & Schema

03

Data Cleaning & Preparation

04

Data Analysis & Dashboard

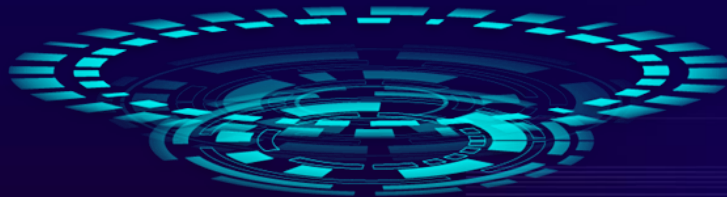
05

Conclusion

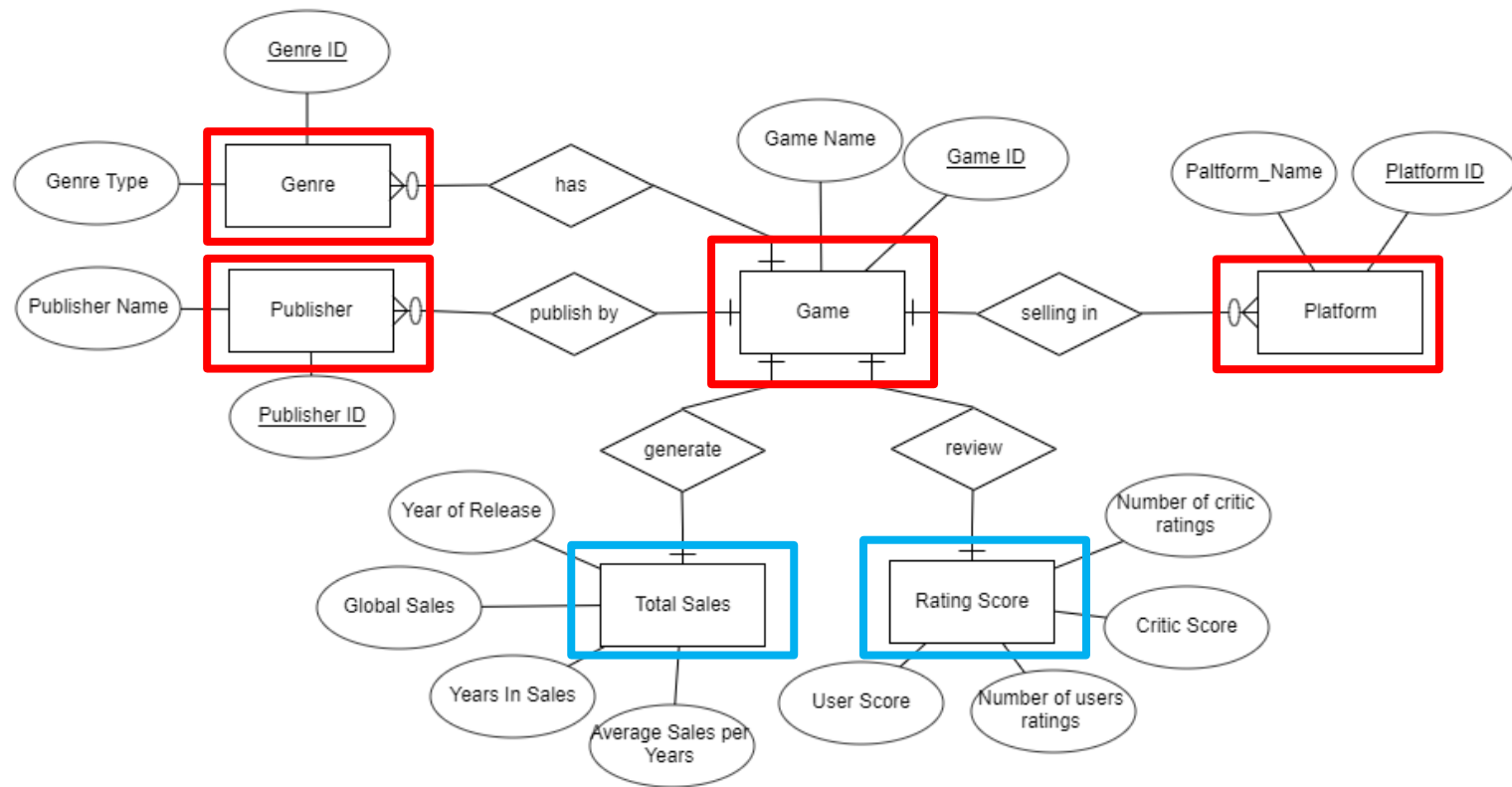


Problem Statement

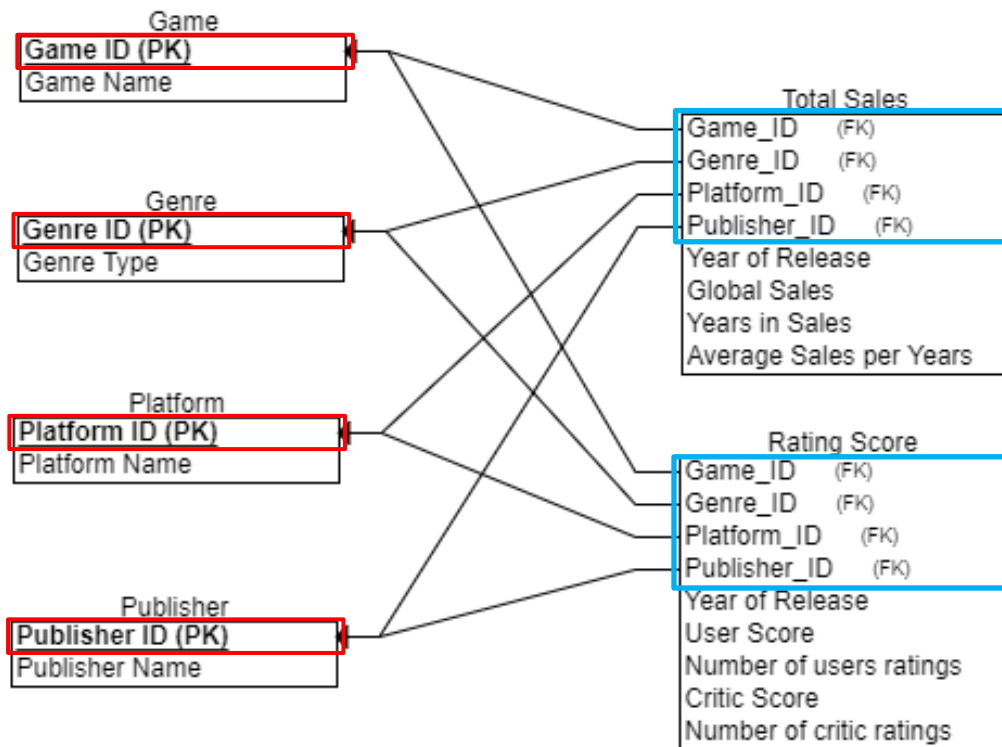
A startup game seller company intends to find out the top few most popular video games on which genre and gaming platform to be selling in their global online gaming store.



ER Diagram



Database Schema



Data Cleaning

1. Empty cell on the name column: 2 records were dropped due to incomplete information.

	A	B	C	D	E
	Name ▾	Platform ▾	Year_of_Release ▾	Genre ▾	Publisher ▾
		GEN	1993		Acclaim Entertainment
48		GEN	1993		Acclaim Entertainment
91					

2. Removed unwanted column on North America, Europe, Japan, Other Sales Column.

Normalisation

Original csv file in Kaggle does not have ID number, using the SQL queries to find out the different distinct game name, platform, genre and publisher.

Generating random unique IDs to each game name, platform, genre and publisher.

```
select distinct Name from [dbo].[videogamesales]  
select distinct Platform from [dbo].[videogamesales]  
select distinct Genre from [dbo].[videogamesales]  
select distinct Publisher from [dbo].[videogamesales]
```

Normalisation

Game

ID	Game_Name
1	'98 Koshien
2	.hack//G.U. Vol.1//Rebirth
3	.hack//G.U. Vol.2//Reminisce
4	.hack//G.U. Vol.2//Reminisce (jp sales)
5	.hack//G.U. Vol.3//Redemption
6	.hack//Infection Part 1
7	.hack//Link
8	.hack//Mutation Part 2
9	.hack//Outbreak Part 3
10	.hack//Quarantine Part 4: The Final Chapter

Publisher

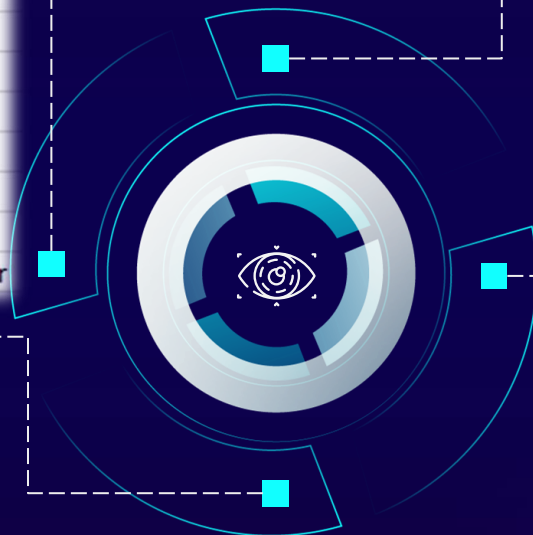
ID	Publisher
H001	10TACLE Studios
H002	1C Company
H003	20th Century Fox Video Games
H004	2D Boy
H005	3DO
H006	49Games
H007	505 Games
H008	5pb
H009	7G//AMES
H010	989 Sports

Genre

ID	Genre
G001	Puzzle
G002	Adventure
G003	Sports
G004	Shooter
G005	Racing
G006	Strategy
G007	Action
G008	Fighting

Platform

ID	Platform
P001	XB
P002	GC
P003	PS3
P004	TG16
P005	X360
P006	XOne
P007	SAT
P008	WS



Adding Primary Key And Foreign Key

Game, Genre, Platform and Publisher

```
ALTER TABLE [dbo].[Game] ADD PRIMARY KEY(ID);  
ALTER TABLE [dbo].[Platform] ADD PRIMARY KEY(ID);  
ALTER TABLE [dbo].[Genre] ADD PRIMARY KEY(ID);  
ALTER TABLE [dbo].[Publisher] ADD PRIMARY KEY(ID);
```

Rating

```
ALTER TABLE videogamesrating  
add foreign key(game_id) references game(id)  
ALTER TABLE videogamesrating  
add foreign key(platform_id) references platform(id)  
ALTER TABLE videogamesrating  
add foreign key(genre_id) references genre(id)  
ALTER TABLE videogamesrating  
add foreign key(publisher_id) references publisher(id)
```

Sales

```
ALTER TABLE videogamesalesvalue  
add foreign key(game_id) references game(id)  
ALTER TABLE videogamesalesvalue  
add foreign key(platform_id) references platform(id)  
ALTER TABLE videogamesalesvalue  
add foreign key(genre_id) references genre(id)  
ALTER TABLE videogamesalesvalue  
add foreign key(publisher_id) references publisher(id)
```

Data Preparation



Global Sales/Year Column

Added 2 Columns for analysing the top 10 purchased games

- 1) Years in Sales: How many years has the video games is in sales?

Years in Sales = 2016 (ie.The year that the data was collected) – Year of Release + 1 (ie.current year)

- 2) Average Sales per Year: Average out the Global sales through the years from the video games are released.

Game_ID	Game_Name	Platform_ID	Year_of_Release	Genre_ID	Publisher_ID	Global_Sales	Years_in_Sales	Average Sales_per_Year
1	'98 Koshien	P009	1998	G003	H300	0.41	19	0.0216
2	.hack//G.U. Vol.1//Rebirth	P027	2006	G012	H353	0.17	11	0.0155
3	.hack//G.U. Vol.2//Reminisce	P027	2006	G012	H353	0.23	11	0.0209
4	.hack//G.U. Vol.2//Reminisce (jp sales)	P027	2006	G012	H353	0.16	11	0.0145
5	.hack//G.U. Vol.3//Redemption	P027	2007	G012	H353	0.17	10	0.0170
6	.hack//Infection Part 1	P027	2002	G012	H054	1.27	15	0.0847
7	.hack//Link	P014	2010	G012	H353	0.14	7	0.0200
8	.hack//Mutation Part 2	P027	2002	G012	H054	0.68	15	0.0453

Limitation & Assumption

- 1) Sales per year = Year of release to compare all games sales on the same scale. Assume games released at start of year.
- 2) The games will still be relevant within a period of 10 years.
- 3) If the game has other variations, it will be treated as another unique game.



Data Analysis

What are the best performing Platforms and how many games are in each platform?

```
select pl.Platform_Name, count(*) as 'No of Games', Round(sum(vgid.Global_sales),2) as 'Global Sales' from videogamesalesvalue vgid
inner join Platform pl
on pl.id = vgid.Platform_ID
where vgid.Year_of_Release between 2006 and 2017
group by pl.Platform_Name
order by sum(vgid.Global_sales) desc;
```

	Platform_Name	No of Games	Global Sales
1	X360	1216	953.18
2	PS3	1306	931.15
3	Wii	1286	891.74
4	DS	1981	656.24
5	PS4	393	314.23
6	PS2	807	268.84
7	3DS	512	257.92
8	PSP	1085	239.6
9	PC	777	164.38
10	XOne	247	159.44
11	WiiU	147	82.16
12	PSV	431	53.85
13	GC	44	11.56
14	XB	66	10.88
15	GBA	45	8.78
16	DC	2	0.06

Data Analysis

What are the best performing Genre and how many games are in each genre?

```
select G.Genre_Type, count(*) as 'No of Games', Round(sum(vgid.Global_sales),2) as 'Global Sales' from videogamesalesvalue vgid
inner join Genre G
on g.id = vgid.Genre_ID
where vgid.Year_of_Release between 2006 and 2017
group by G.Genre_Type
order by sum(vgid.Global_sales) desc;
```

	Genre_Type	No of Games	Global Sales
1	Action	2384	1117.62
2	Sports	1266	795.05
3	Shooter	737	717.3
4	Misc	1253	555.12
5	Role-Playing	998	522.29
6	Racing	564	299.24
7	Platform	340	279.87
8	Simulation	607	221.17
9	Fighting	414	191.35
10	Adventure	1023	141.45
11	Puzzle	366	90.21
12	Strategy	393	73.34

Data Analysis

How many games within the period from 2006 to 2017?

```
select count(distinct Game_Name) as 'No of Games' from videogamesalesvalue
where Year_of_Release between 2006 and 2017
```

Results	
No of Games	
1	6806

What are the Top 10 games with the best performance based on sales?

```
select top 10 g.Game_Name, ge.Genre_Type, round(sum(vgid.Average_Sales_Per_Year),2) as 'Sum of Global Sales per year' from videogamesalesvalue vgid
inner join game g
on g.id=vgid.game_id
inner join Genre Ge
on ge.id=vgid.Genre_ID
where vgid.Year_of_Release between 2006 and 2017
group by g.Game_Name, ge.Genre_Type
order by sum(vgid.Average_Sales_Per_Year) desc
```

	Game_Name	Genre_Type	Sum of Global Sales per year
1	Grand Theft Auto V	Action	15.94
2	Call of Duty: Black Ops 3	Shooter	12.84
3	FIFA 17	Sports	11.47
4	FIFA 16	Sports	8.15
5	Minecraft	Misc	7.74
6	Wii Sports	Sports	7.5
7	Call of Duty: Advanced Warfare	Shooter	7.33
8	Pokemon Sun/Moon	Role-Playing	7.14
9	Call of Duty: Infinite Warfare	Shooter	7.02
10	Call of Duty: Ghosts	Shooter	6.85

Data Analysis

What are the Top 10 games with the best performance based on user rating?

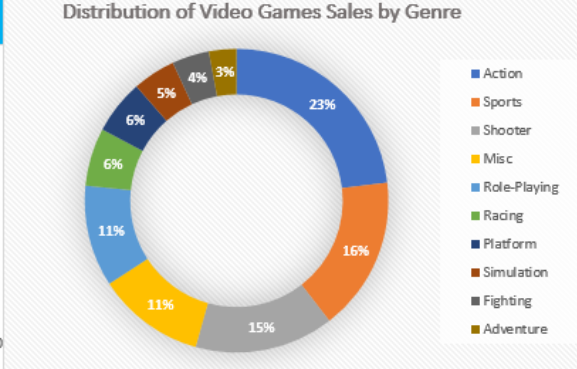
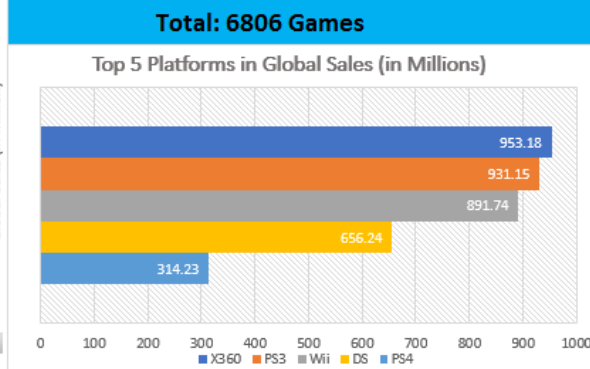
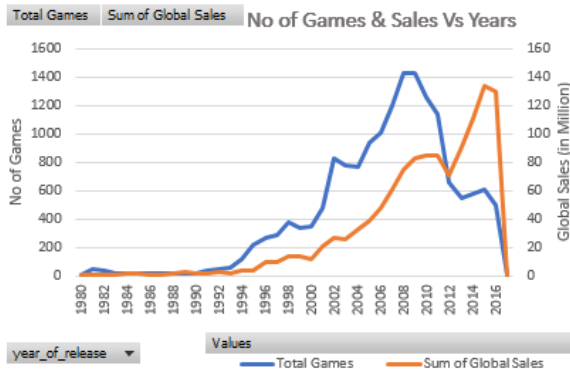
```
select top 10 g.Game_Name, ge.Genre_Type, round(avg(vgr.user_score),2) as 'Avg User Score' from videogamesrating vgr
inner join game g
on g.id=vgr.game_id
inner join Genre Ge
on ge.id=vgr.Genre_ID
where vgr.Year of Release between 2006 and 2017 and user_count>1000
group by g.Game_Name, ge.Genre_Type
order by avg(vgr.user_score) desc
```

	Game_Name	Genre_Type	Avg User Score
1	Cory in the House	Action	9.5
2	The Orange Box	Shooter	9.3
3	The Witcher 3: Wild Hunt	Role-Playing	9.23
4	Persona 4: Golden	Role-Playing	9.2
5	The Last of Us	Action	9.1
6	Super Mario Galaxy 2	Platform	9.1
7	Mario Kart 8	Racing	9.1
8	Fire Emblem: Awakening	Role-Playing	9.1
9	The Legend of Zelda: Twilight Princess	Action	9
10	Super Mario 3D World	Platform	9

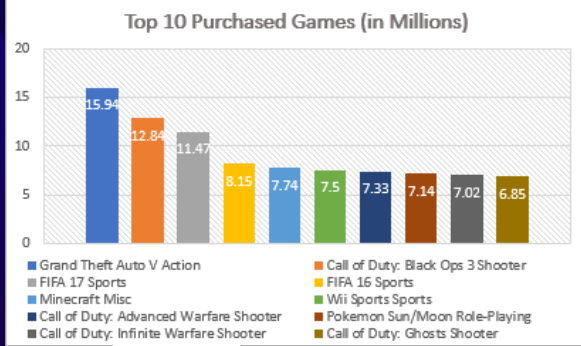
Dashboard



Video Game Popularity Analysis



Year	
1980	
1981	
1982	
1983	
1984	
1985	
1986	
1987	
1988	
1989	
1990	
1991	
1992	
1993	
1994	
1995	
1996	
1997	
1998	
1999	
2000	
2001	
2002	
2003	
2004	
2005	
2006	
2007	
2008	
2009	
2010	
2011	
2012	
2013	
2014	
2015	
2016	
2017	



Conclusion

Platform & Genre

Further prediction on what genre of new video games to notice.

Sales

The most popular video games (ie. Top 10) that users like directly affect the sales.



Critic Rating

Before any video games are released, a user will look at those games with higher critic rating before deciding to buy the video games.

User Rating

After playing the video games, the user will give a user rating. High user rating mean that they have liked the games.

Recommended Games

	Game Name	Genre Type
Top 10 Sales	Grand Theft Auto V	Action
	Call of Duty: Black Ops 3	Shooter
	FIFA 17	Sports
	FIFA 16	Sports
	Minecraft	Misc
	Wii Sports	Sports
	Call of Duty: Advanced Warfare	Shooter
	Pokemon Sun/Moon	Role-Playing
	Call of Duty: Infinite Warfare	Shooter
	Call of Duty: Ghosts	Shooter
Top 10 User Rating	Cory in the House	Action
	The Orange Box	Shooter
	The Witcher 3: Wild Hunt	Role-Playing
	Persona 4: Golden	Role-Playing
	The Last of Us	Action
	Super Mario Galaxy 2	Platform
	Mario Kart 8	Racing
	Fire Emblem: Awakening	Role-Playing
	The Legend of Zelda: Twilight Princess	Action
Top 10 Critic Rating	Super Mario 3D World	Platform
	Grand Theft Auto IV	Action
	Super Mario Galaxy	Platform
	Uncharted 2: Among Thieves	Action
	The Elder Scrolls V: Skyrim	Role-Playing
	Red Dead Redemption	Action

THANKS

Does anyone have any questions? ▶

References

<https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings>