

# 202480-Fall

## 2024-ITCS-3162-051-Introduction to Data Mining: Project 1

Sean Kiser

9/5/2024

### **Kaggle Dataset Link:**

<https://www.kaggle.com/datasets/nayanack/online-game-dataset>

### **Introduction to Online Game Dataset:**

I am using the data set Online Game Dataset. This data set has 16598 rows and 11 columns. Some of its key columns are Name, Platform, Year, Genre, and Sales. This dataset will primarily be using sales for visualizing data and finding correlations. This data set also separates its sales into 3 regions: North America, Europe, and Japan. It will be very interesting to see how sales differ based on different locations.

### **Questions to Answer:**

One of the questions that I am most interested in answering is seeing how sales change depending on location and genre. Primarily I am interested in the shooter genre. I would like to see how popularity in the shooter genre differs from North America and Japan. Another question that I have is how popularity in games changes over time. I would like to see the trend of global sales change from year to year. Another problem I would like to solve is to create a table that lists the top selling game from any given year. Users could use this table for inspiration when picking a game to play.

**Advantages and Disadvantages of Data:**

This dataset I think could be very beneficial in looking at relationships when it comes to video games and violence. These findings would be based on preconceived knowledge that gun violence is worse in North America than gun violence in Japan. Some disadvantages of this are uncontrolled compounding variables, bias, and limited data. Using this dataset to create a table of the most popular games from each year would be useful for users picking out new games to play. In a general sense people care about popularity and what's being played. People who are looking for a blast to the past and want to enjoy a new game could use this table for such purposes. An issue with the trending of video games overtime is that as time passes the population increases. So my initial thought would be that obviously over time more and more video games are being bought. One way to combat this would be to use another dataset that looks at population each year and compares it to global sales of video games.

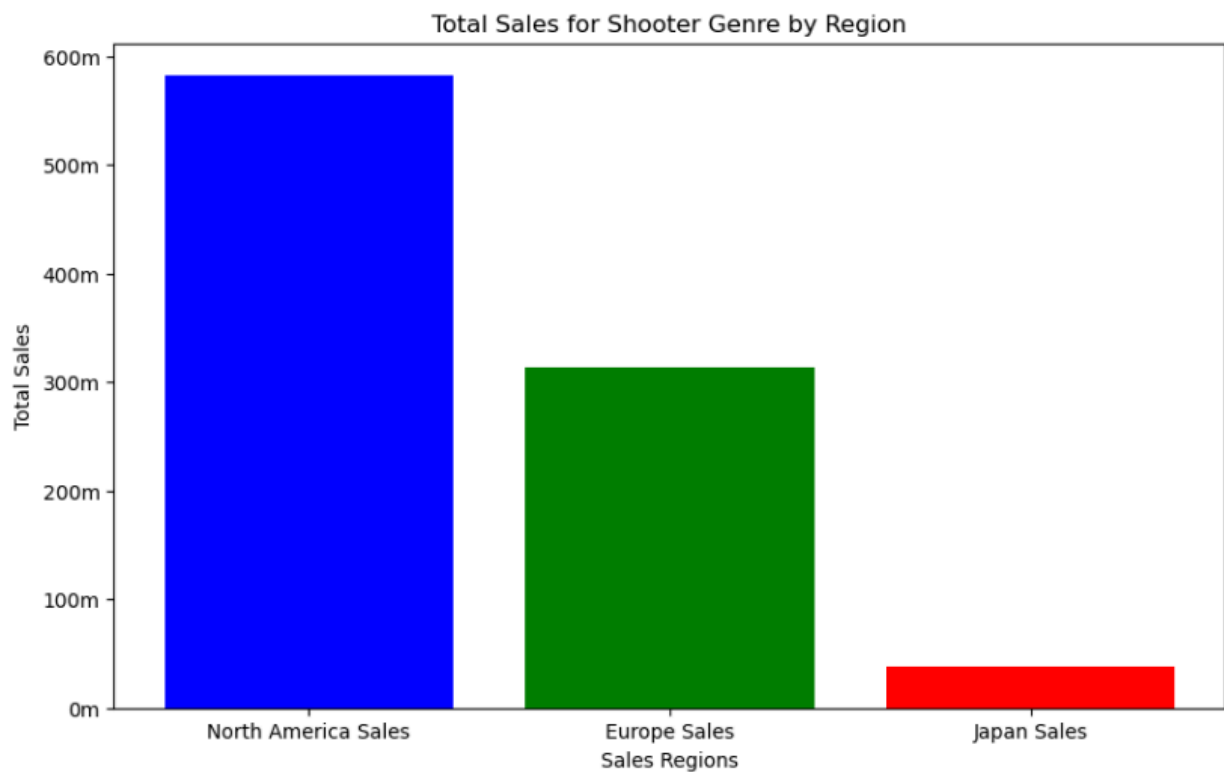
**Pre-Processing the Data:**

For this particular data set there are three things that need to be considered when cleansing this data. There are two columns in this data set that have null values. The “year” column has 271 rows that are null. So any visuals that will use the year column, these rows will have to be dropped. Also the “publisher” columns have 58 columns that are null. So if there are any relationships that I want to regard that use the “publisher column” I will also have to drop. Finally this data set has duplicate rows when it comes to video game names. Any video game that was released on different platforms has repeated rows. So I will need to sum sales for any visuals that will use sales based off of video game names.

### Visualization #1

Figure one looks at the amount of total sales from each region of video games with the genre “Shooter”. This figure shows that North America has significantly more video games with the genre “shooter” than Japan. North America had 582.60 million total sales and Japan had 38.28 million sales. This is particularly interesting because of the level of gun violence in North America vs. Japan. Obviously this diagram does little to prove anything regarding the correlation between video games and gun violence, but it is cool to see the different values culturally that North American and Japan have.

**Figure 1.**



### Visualization #2

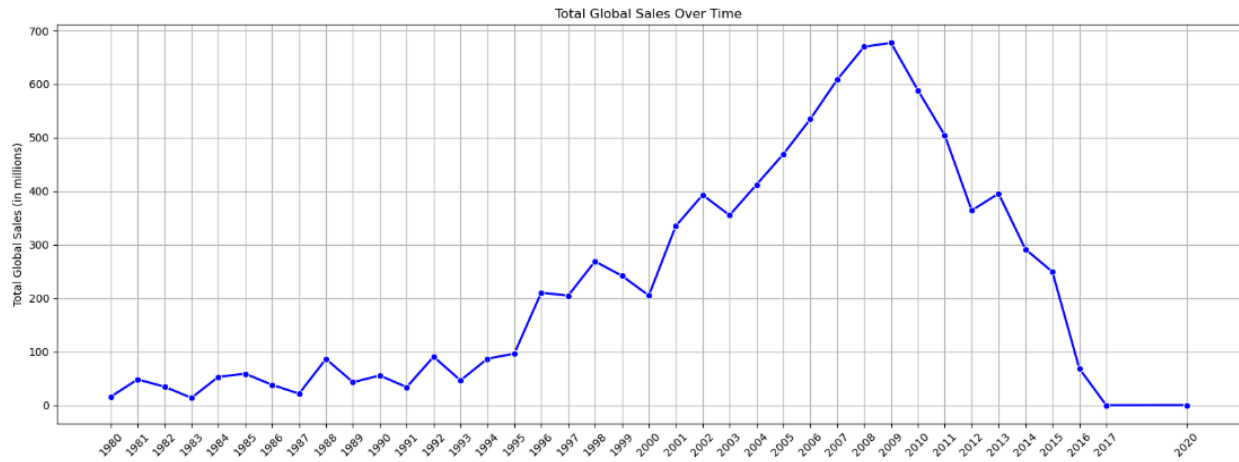
Figure 2 looks at the top sold game from each year from 1980 to 2016. This figure can be used as inspiration to find old or new games that were once very popular. It also is helpful to show extreme outliers such as Wii Sports which sold 82 million copies. People care about popularity so now this table can help people find popular games from the past.

**Figure 2.**

	Name	Rank	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	Asteroids	259	2600	1980.0	Shooter	Atari	4.64	0.69	0.00	0.12	5.46
1	Frogger	279	PS	1981.0	Action	Hasbro Interactive	5.85	0.39	0.00	0.13	6.36
2	Pac-Man	90	2600	1982.0	Puzzle	Atari	7.55	0.53	0.85	0.10	9.03
3	Baseball	422	NES	1983.0	Sports	Nintendo	1.39	0.37	3.00	0.05	4.81
4	Duck Hunt	10	NES	1984.0	Shooter	Nintendo	26.93	0.63	0.28	0.47	28.31
5	Super Mario Bros.	2	NES	1985.0	Platform	Nintendo	32.48	4.88	6.96	0.99	45.31
6	The Legend of Zelda	128	NES	1986.0	Action	Nintendo	3.74	0.93	1.69	0.14	6.51
7	Zelda II: The Ad...	252	NES	1987.0	Adventure	Nintendo	2.19	0.50	1.61	0.08	4.38
8	Tetris	6	GB	1988.0	Puzzle	Nintendo	26.17	2.95	6.03	0.69	35.84
9	Super Mario Land	22	GB	1989.0	Platform	Nintendo	10.83	2.71	4.18	0.42	18.14
10	Super Mario World	19	SNES	1990.0	Platform	Nintendo	15.99	4.86	4.49	0.75	26.07
11	The Legend of Ze...	232	SNES	1991.0	Action	Nintendo	4.17	1.43	1.48	0.23	7.31
12	Super Mario Land...	51	GB	1992.0	Adventure	Nintendo	6.16	2.04	2.69	0.29	11.18
13	Super Mario All-...	58	SNES	1993.0	Platform	Nintendo	5.99	2.15	2.12	0.29	10.55
14	Donkey Kong Country	72	SNES	1994.0	Platform	Nintendo	6.60	2.91	3.59	0.43	13.53
15	Super Mario Worl...	283	SNES	1995.0	Platform	Nintendo	3.40	1.06	2.38	0.18	7.03
16	Pokemon Red/Poke...	5	GB	1996.0	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.37
17	Gran Turismo	53	PS	1997.0	Racing	Sony Computer En...	4.02	3.87	2.54	0.52	10.95
18	Pokémon Yellow: ...	31	GB	1998.0	Role-Playing	Nintendo	5.89	5.04	3.12	0.59	14.64
19	Pokemon Gold/Pok...	13	GB	1999.0	Role-Playing	Nintendo	9.00	6.18	7.20	0.71	23.10
20	Tony Hawk's Pro ...	226	PS	2000.0	Sports	Activision	4.49	1.88	0.02	0.23	6.62
21	Gran Turismo 3: ...	29	PS2	2001.0	Racing	Sony Computer En...	6.85	5.09	1.87	1.16	14.98
22	Grand Theft Auto...	25	PS2	2002.0	Action	Take-Two Interac...	8.41	5.52	0.47	1.79	16.19
23	Need for Speed U...	105	PS2	2003.0	Racing	Electronic Arts	5.26	3.98	0.09	1.14	10.46
24	Grand Theft Auto...	18	PS2	2004.0	Action	Take-Two Interac...	10.77	1.96	0.41	10.72	23.86
25	Nintendogs	11	DS	2005.0	Simulation	Nintendo	9.07	11.00	1.93	2.75	24.76
26	Wii Sports	1	Wii	2006.0	Sports	Nintendo	41.49	29.02	3.77	8.46	82.74
27	Wii Fit	14	Wii	2007.0	Sports	Nintendo	8.94	8.03	3.60	2.15	22.72
28	Mario Kart Wii	3	Wii	2008.0	Racing	Nintendo	15.85	12.88	3.79	3.31	35.82
29	Wii Sports Resort	4	Wii	2009.0	Sports	Nintendo	15.75	11.01	3.28	2.96	33.00
30	Call of Duty: BL...	32	X360	2010.0	Shooter	Activision	17.59	9.50	0.59	3.36	31.03
31	Call of Duty: Mo...	30	X360	2011.0	Shooter	Activision	15.58	11.29	0.62	3.35	30.83
32	Call of Duty: BL...	35	PS3	2012.0	Shooter	Activision	14.08	11.05	0.72	3.88	29.72
33	Grand Theft Auto V	17	PS3	2013.0	Action	Take-Two Interac...	23.46	23.04	1.39	8.03	55.92
34	Call of Duty: Ad...	94	PS4	2014.0	Shooter	Activision	10.44	8.18	0.34	2.93	21.90
35	Call of Duty: BL...	34	PS4	2015.0	Shooter	Activision	11.89	9.56	0.43	3.42	25.32
36	FIFA 17	222	PS4	2016.0	Sports	Electronic Arts	0.50	5.48	0.07	0.85	6.91

### Visualization #3

This visualization shows the trend line of total global sales of video games. As suspected the trend line in general goes upward. This shows that over time video games have become more and more popular. Something to note about this visualization is that near the end the table starts to decline. What causes this decline though? Is it because this data set is only physical copies sold? This would make a lot of sense because now most video games are sold virtually. This trend line in general does show that over time video games have become more popular.



### Impact

This project has looked at how video games sales have trended over time. Figure 1 looks at how video games with the genre 'shooter' sell depending on region. Interestingly there are significantly more sales in the United States than in Japan. This result can be harmful though as incorrect correlations can be made using this data. Instead, this data does a good job of showing cultural differences between the United States and Japan. Figure 2 is a table that shows the top selling game from each year. This table could be useful for nostalgic gamers trying to find a popular game from the past. Something to note about this data set is that there could be some missing data for some of the newer video games. For some reason total sales went down after the year 2009 which is strange because it would be expected to keep going up. Figure 3 shows the trending of global sales over time. This graph shows that video games have steadily become more and more popular. But once again this data set looks like it is missing data because global sales started to decline after 2009.