

# Analysis on History of Video Games — SQL and Interactive Tableau Dashboard



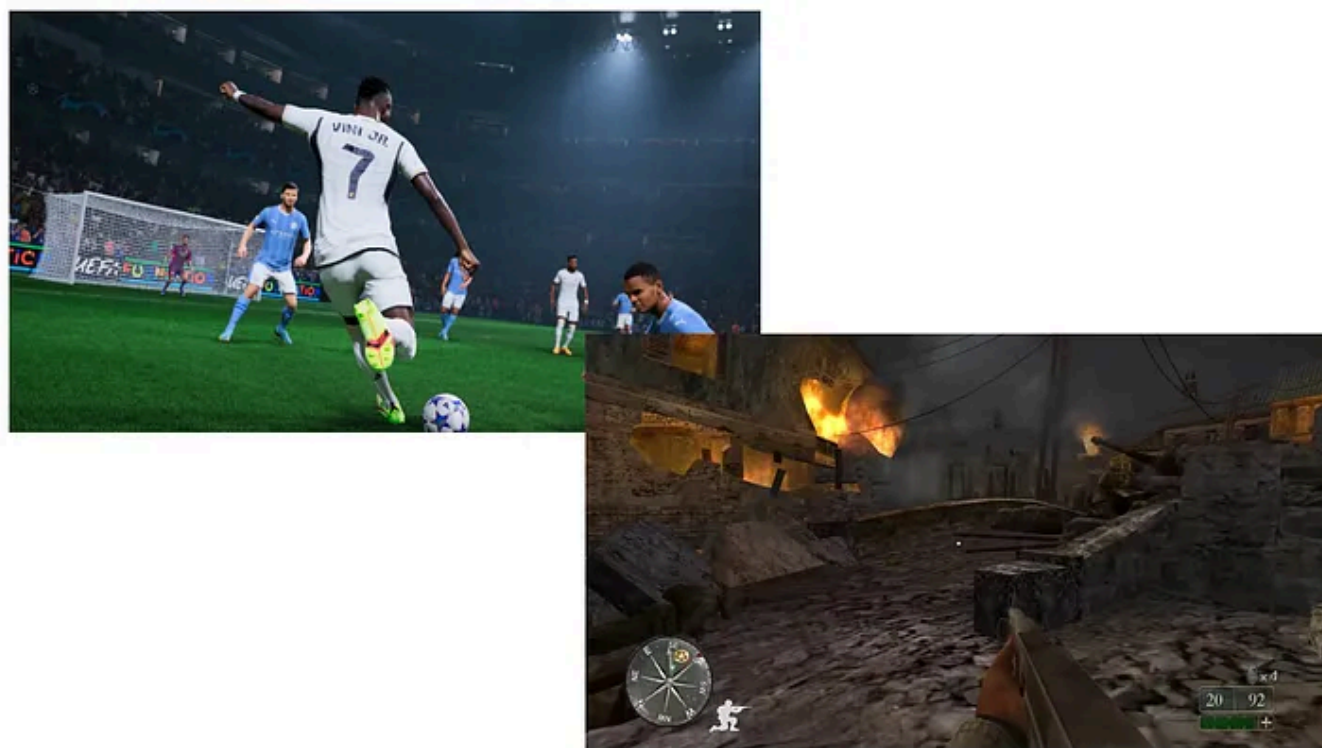
Joe Beadles

7 min read · Aug 10, 2024



Utilizing Excel (Power Query), SQL, and Tableau to process and analyze a historical dataset on the most popular video games released since 1978.

([Tableau Dashboard](#))



via CC-BY

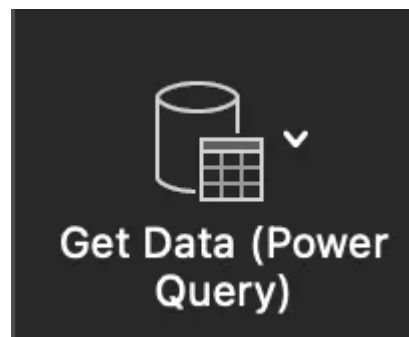
**Overview:** Your assignment is to assist the video game development team in their efforts to create an enjoyable and profitable game. You need to identify the years with the highest game sales and those with the best reception from the audience. Additionally, you should identify a single year that combines high sales with excellent ratings from the gaming fans and critics. In conclusion, you will deliver your findings through a report accompanied by an interactive dashboard. ([DATASET](#))

## Main Objective

Which year is notable for its high sales and excellent game ratings?

## Null Values

Our dataset contained several null values in the critic and user scores columns. These columns are designed to hold double values and therefore nulls are not acceptable in the columns. To address the issue, we utilized the Replace values option in the Power Query Editor in Microsoft Excel. We replaced all our blank/null values in these columns with a 0, noting that this adjustment should be considered when analyzing the scores in the future.



Microsoft Excel Tool

## MySQL

We first need to import our data by utilizing the Data Import Wizard. Next, we will create a table of the first 1000 rows to use for our analysis.

```
CREATE TABLE game_data (  
  SELECT *  
  FROM video_game_data  
  LIMIT 1000  
);
```



Utilizing MySQL Workbench 8.0.38 Community

## Top 15 games by total units sold

We can begin to identify which games were the best selling by accessing our total shipped column.

```
SELECT
    game_rank,
    Name,
    total_shipped AS total_orders
FROM game_data
ORDER BY total_orders DESC
LIMIT 15
```

	game_rank	Name	total_orders
▶	1	Wii Sports	82.9
	2	Super Mario Bros.	40.24
	3	Counter-Strike: Global Offensive	40
	4	Mario Kart Wii	37.32
	5	PLAYERUNKNOWN'S BATTLEGROUNDS	36.6
	6	Minecraft	33.15
	7	Wii Sports Resort	33.13
	8	Pokemon Red / Green / Blue Version	31.38
	9	New Super Mario Bros.	30.8
	10	New Super Mario Bros. Wii	30.3
	11	Tetris	30.26
	12	Call of Duty: Modern Warfare	30.13
	13	Duck Hunt	28.31
	14	Wii Play	28.02
	15	Mario Kart 8 Deluxe	24.77

Through our analysis a recurring theme becomes apparent, games with universal appeal tend to obtain higher sale marks. Games such as Wii Sports and Mario games are loved by all ages and therefore attract a broader audience which results in more sales.

## Biggest Year for Unit sales

Let's find the years that had the most amount of total sales for their games.

```
SELECT
    Year,
    ROUND(SUM(total_shipped),2) AS total_orders
FROM game_data
GROUP BY Year
ORDER BY total_orders DESC
LIMIT 10
```

	Year	total_orders
►	2010	322.87
	2011	294.72
	2009	291.64
	2007	289.96
	2008	280.83
	2006	276.01
	2013	265.78
	2012	228.79
	2014	222.66
	2017	220.75

Returning the top 10 results

2010 was a big year for video games with 322.87 million units being sold within the year. Let's filter our query to access the names of the top selling games during this year.

```
SELECT
    game_rank,
    Name,
    total_shipped
FROM game_data
WHERE Year = 2010
ORDER BY total_shipped DESC
```

	game_rank	Name	total_shipped	Platform
►	6	Minecraft	33.15	PC
	16	Kinect Adventures!	24	X360
	49	Call of Duty: Black Ops	14.74	X360
	68	Pokemon Heart Gold / Soul Silver Version	12.72	DS
	70	Call of Duty: Black Ops	12.67	PS3
	79	Gran Turismo 5	11.95	PS3
	103	Halo: Reach	9.97	X360
	111	Wii Party	9.34	Wii
	141	God of War III	7.6	PS3
	148	Super Mario Galaxy 2	7.41	Wii
	175	Donkey Kong Country Returns	6.53	Wii
	179	Red Dead Redemption	6.5	X360

2010 saw the ever popular Minecraft hit the game shelves

Minecraft was far ahead with 33.15 million sales, a testament to its age-friendly themes and gameplay. Additionally, Call of Duty: Black Ops saw 27.41 million sales in the 2010 across both platforms of Xbox 360 and PS3. It's standard practice for big franchises like Call of Duty to release their new titles on both platforms each year.

To follow this, we can check to see how many of the same games were released on two different platforms.

```
SELECT
    COUNT(*) AS num_of_occurences,
    Name
FROM game_data
WHERE Year = 2010
GROUP BY Name
HAVING COUNT(*) > 1
```

	num_of_occurences	Name
►	2	Call of Duty: Black Ops
	2	Final Fantasy XIII
	3	Fallout: New Vegas
	2	FIFA Soccer 11
	2	Assassin's Creed: Brotherhood
	2	Battlefield: Bad Company 2
	2	Medal of Honor
	2	Need for Speed: Hot Pursuit
	2	Madden NFL 11
	2	LEGO Harry Potter: Years 1-4

Unless it is a platform exclusive, big franchises will always release their titles onto both platforms

## Games released per year

Now that we've looked at the our top selling years, how will this compare to the amount of games released each year?

We simply need utilize the **COUNT()** function in our query and **GROUP BY** year.

```
SELECT
  COUNT(*) AS num_of_games,
  Year
FROM game_data
GROUP BY Year
ORDER BY num_of_games DESC
```

	num_of_games	Year
►	64	2010
	59	2007
	58	2008
	55	2011
	55	2009
	54	2013
	45	2016
	45	2015
	44	2014
	41	2002
	39	2012
	38	2006
	35	2003
	35	2001
	33	2004
	31	2017
	30	2005
	28	2000
	28	2018
	27	1999
	26	1998
	18	1996
	15	1997
	13	2019
	12	1992
	12	1995
	8	1993
	7	1985
	7	1990
	6	1994
	6	1987
	5	1988
	4	1989
	4	1991
	4	1982
	3	2020
	2	1981
	2	1986
	1	1978
	1	1983

1000 games grouped into each year in our dataset

2010 saw the highest number of releases but there were several years that had less than 5 games released, notably 1978, 1981, 1982, 1983, 1986, 1989, 1991, and 2020.

Will the number of games released in a year affect how many units were sold in that same year?

```
SELECT
  COUNT(*) AS num_of_games,
  Year,
  ROUND(SUM(total_shipped),2) AS total_orders
FROM game_data
GROUP BY Year
ORDER BY total_orders DESC
LIMIT 15
```

	num_of_games	Year	total_orders
▶	64	2010	322.87
	55	2011	294.72
	55	2009	291.64
	59	2007	289.96
	58	2008	280.83
	38	2006	276.01
	54	2013	265.78
	39	2012	228.79
	44	2014	222.66
	31	2017	220.75
	45	2015	215.15
	45	2016	190.31
	33	2004	173.02
	30	2005	167.54
	28	2018	166.39

On average, there were 5.04 million sales per game in 2010

```
SELECT
  ROUND((ROUND(SUM(total_shipped),2)/COUNT(*)),2) AS average_sales_per_game,
  Year
FROM game_data
GROUP BY Year
HAVING COUNT
  ((SELECT COUNT(*) FROM game_data)) > 4
ORDER BY average_sales_per_game DESC
LIMIT 15
```

	average_sales_per_game	Year
▶	12.23	1985
	7.85	2019
	7.26	2006
	7.12	2017
	5.94	2018
	5.87	2012
	5.6	1998
	5.58	2005
	5.36	2011
	5.3	2009
	5.27	1990
	5.24	2004
	5.06	2014
	5.04	2010
	4.99	1997

After filtering for years with more than 4 games released, 1985 had the highest average sales per game

After incorporating average sales per game in our analysis, we find that although 2010 had the highest total unit sales, it ranked 14th in terms highest average sales per game list. The year with the highest average sales per



game, which also appeared on the total sales list, was 2006, with an average of 7.26 million sales per game and 276.01 million total unit sales.

## Critic and User Scores

As stated earlier, this dataset came with many nulls in both the user\_score and critic\_score columns. Before finding our highest rated games, we can identify number of missing values in these two columns.

```
SELECT
  COUNT(*)
FROM game_data
WHERE critic_score = 0
```

	COUNT(*)
▶	135

13.5% of our critic scores were null

```
SELECT
  COUNT(*)
FROM game_data
WHERE user_score = 0
```

	COUNT(*)
▶	685

A staggering 68.5% of our user scores were null

In total, 134 or 13.4% of our games had both a missing user score and critic score. We can now continue on into our analysis of these scores. Based on what we now know about our missing user and critic scores, we can incorporate these findings into our analysis.

## Loved by the public and critics?

Which of our games were loved by the public and also by the critics? To find which games appealed to both of our audiences, we need to create two separate tables with the top 15 user scores and the top 15 critic scores. (*Don't forget to filter out our null scores and years that had more than 4 games released!*)

```
CREATE TABLE top_user_score
AS (SELECT Year,
    ROUND(AVG(user_score),2) AS avg_user_score
    FROM game_data
    WHERE user_score != 0
    GROUP BY Year
    HAVING COUNT(*) > 4
    ORDER BY avg_user_score DESC
    LIMIT 15);
```

```
CREATE TABLE top_critic_score
AS (SELECT Year,
    ROUND(AVG(critic_score),2) AS avg_critic_score
    FROM game_data
    WHERE critic_score != 0
    GROUP BY Year
    HAVING COUNT(*) > 4
    ORDER BY avg_critic_score DESC
    LIMIT 15);
```

Now we can utilize an **INTERSECT** to see which years appear in both top 15 lists.

```
SELECT Year FROM top_critic_score
INTERSECT
SELECT Year FROM top_user_score
```

	Year
►	2011
	2001
	2013
	2009
	2014
	2018
	2012
	2016

Let's look further to identify which of our years performed the best

```
SELECT
    c.Year,
    ROUND(((c.avg_critic_score + u.avg_user_score)/2),2) AS overall_avg_score
FROM top_critic_score c JOIN top_user_score u
ON c.Year = u.Year
```

	Year	overall_avg_score
►	2009	8.81
	2011	8.36
	2001	8.18
	2013	8.04
	2012	7.88
	2014	7.77
	2016	7.62
	2018	7.41

2009 led the way by 0.45 points



Although there is no precise measure of comparing the importance or accuracy between user and critic scores, averaging the two scores can give us a quick overview at the gaming environment during those years.

### Is 2009 the pinnacle of gaming thus far?

2009 was a year marked by success in the both the public and critic's eyes. To add to this, the released games benefitted the companies, seeing a total of 291.64 million sales. Despite having a modest 5.30 million sales per game, the combination of high average review scores and substantial amount of total sales marks 2009 as a top year for the gaming industry.



via CC-BY

What games were among the best sellers in 2009?

```
SELECT
    game_rank,
    Name,
    total_shipped
FROM game_data
WHERE Year = 2009
ORDER BY total_shipped DESC
LIMIT 15
```

	game_rank	Name	total_shipped
▶	7	Wii Sports Resort	33.13
	10	New Super Mario Bros. Wii	30.3
	21	Wii Fit Plus	21.13
	57	Call of Duty: Modern Warfare 2	13.53
	93	Call of Duty: Modern Warfare 2	10.61
	138	The Sims 3	7.96
	142	Pokemon Platinum Version	7.6
	167	Uncharted 2: Among Thieves	6.74
	183	Halo 3: ODST	6.35
	217	Assassin's Creed II	5.57
	221	Forza Motorsport 3	5.5
	239	Assassin's Creed II	5.3
	258	Resident Evil 5	5.1
	297	Gran Turismo	4.67
	308	Mario & Luigi: Bowser's Inside ...	4.56

A number of timeless games feature in the 2009 releases

## Conclusion

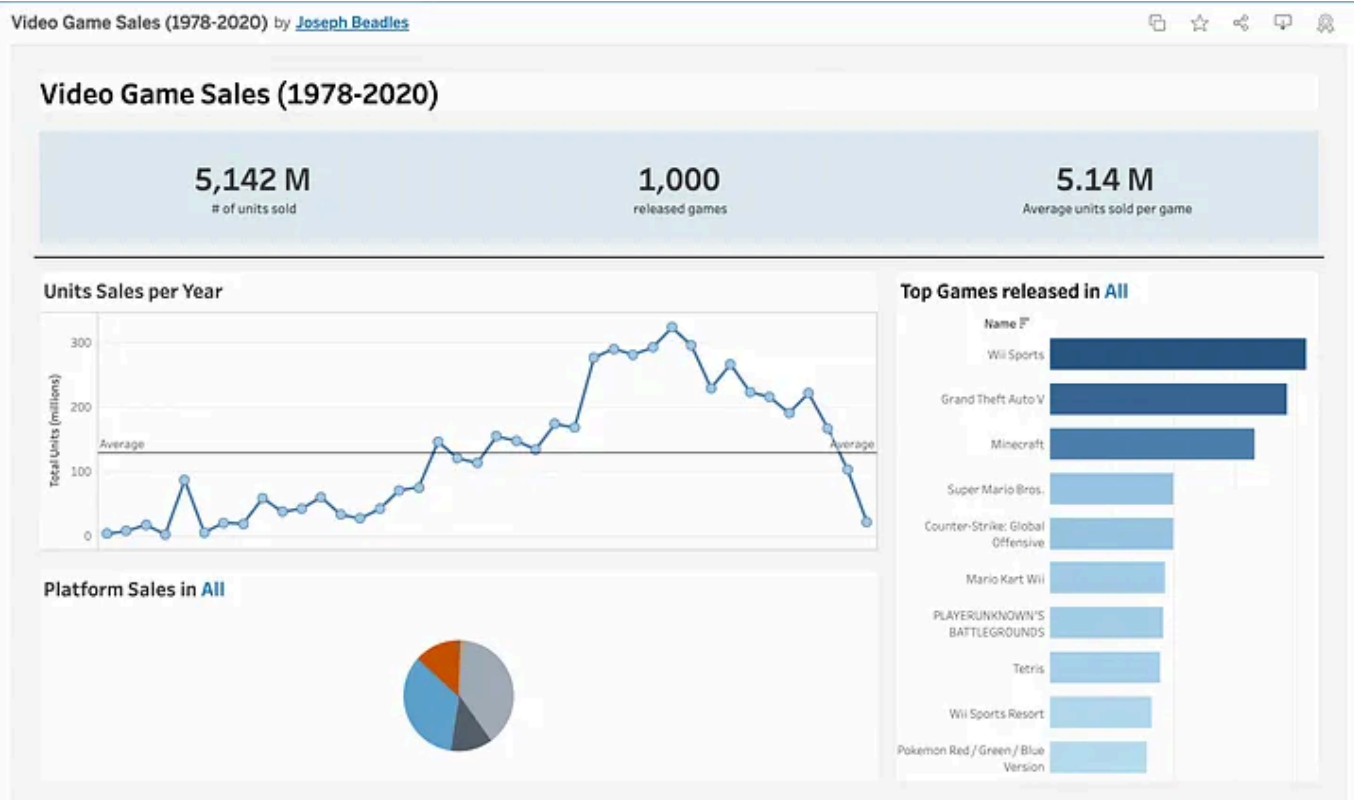
In conclusion, this analysis has provided valuable insights into the sales and popularity of video games over the course of a 42 year period. We learned that 2009 had the highest average review scores whilst also maintaining a high amount of total unit sales. Additionally, we found that a large percentage of the top 15 selling games were centered around a family-oriented experience.

By utilizing these insights, we can provide our development team with the data to make an informed decision about their next game production.

However, it is important to note that user scores significantly affected the large number of null values. Additionally, limiting our analysis to the first 1000 games could have produced a bias towards newer games due to the increase in availability in recent decades.

Considering the effectiveness of a universal appeal in games like Wii Sports or Mario, it is recommended that the video game team prioritizes a fun and inclusive experience for their next game. Top selling titles often include the ability to play with friends or family and adding this element to their production could boost potential sales.

Future research could be aimed at analyzing the variation in sales across the different gaming platforms. Additionally, analyzing a video game sales dataset from a geographic perspective would provide insights into the preferences and habits of specific locations.



[LINK TO DASHBOARD](#)

- Sql
- Data Analysis
- Data Visualization
- Personal Project
- Tableau



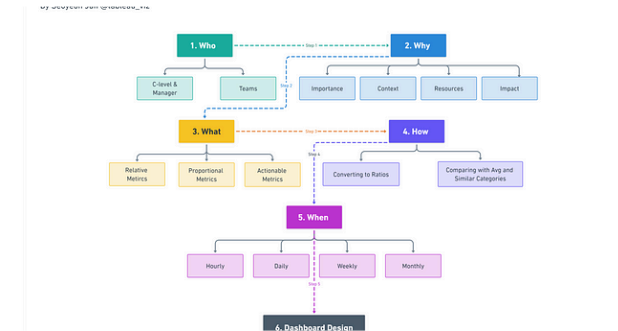
Written by Joe Beadles

[Edit profile](#)

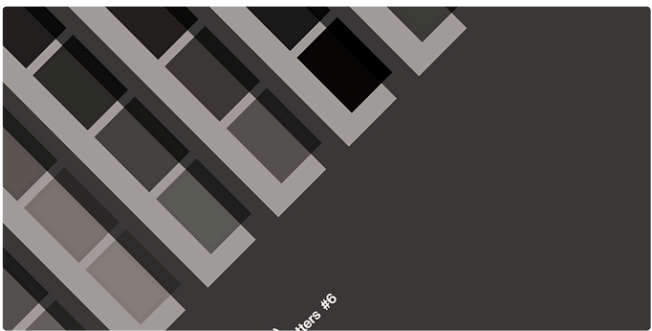
1 Follower

Data Analyst <https://www.linkedin.com/in/josephbeadles/> <https://www.kaggle.com/joebeadles>

## Recommended from Medium



Seoyeon jun



T. from Data Rocks