SQL DATA EXPLORATION

WHEN WAS THE GOLDEN AGE OF VIDEO GAMES

BACKGROUND TO THE REPORT

- Video games are big business: the global gaming market is projected to be worth more than $300 billion by 2027 according to Mordor Intelligence. With so much money at stake, the major game publishers are hugely incentivized to create the next big hit. But are games getting better, or has the golden age of video games already passed?

- In this project, we'll explore the top 400 best-selling video games created between 1977 and 2020. We'll compare a dataset on game sales with critic and user reviews to determine whether or not video games have improved as the gaming market has grown.

- Our database contains two tables. We've limited each table to 400 rows for this project, but you can find the complete dataset with over 13,000 games on Kaggle.

DATA TYPES OF THE GAME SALES TABLE

column	type	meaning
game	varchar	Name of the video game
platform	varchar	Gaming platform
publisher	varchar	Game publisher
developer	varchar	Game developer
games_sold	float	Number of copies sold (millions)
year	int	Release year

DATA TYPES OF THE REVIEW TABLE

column	type	meaning
game	varchar	Name of the video game
critic_score	float	Critic score according to Metacritic
user_score	float	User score according to Metacritic

SQL QUERIES - TOP TEN BEST-SELLING GAMES

✓ Select all information for the top ten best-selling games

✓ Order the results from best-selling game down to tenth best-selling

SELECT *

FROM game_sales

ORDER BY games_sold DESC

LIMIT 10;

SQL QUERIES - TOP TEN BEST-SELLING GAMES RESULT

game	platform	publisher	developer	games_sold	year
Wii Sports for Wii	Wii	Nintendo	Nintendo EAD	82.90	2006
Super Mario Bros. for NES	NES	Nintendo	Nintendo EAD	40.24	1985
Counter-Strike: Global Offensive for PC	PC	Valve	Valve Corporation	40.00	2012
Mario Kart Wii for Wii	Wii	Nintendo	Nintendo EAD	37.32	2008
PLAYERUNKNOWN'S BATTLEGROUNDS for PC	PC	PUBG Corporation	PUBG Corporation	36.60	2017
Minecraft for PC	PC	Mojang	Mojang AB	33.15	2010
Wii Sports Resort for Wii	Wii	Nintendo	Nintendo EAD	33.13	2009
Pokemon Red / Green / Blue Version for GB	GB	Nintendo	Game Freak	31.38	1998
New Super Mario Bros. for DS	DS	Nintendo	Nintendo EAD	30.80	2006
New Super Mario Bros. Wii for Wii	Wii	Nintendo	Nintendo EAD	30.30	2009

2. Missing review scores

✓ Wow, the best-selling video games were released between 1985 to 2017! That's quite a range; we'll have to use data from the reviews table to gain more insight on the best years for video games.

✓ First, it's important to explore the limitations of our database. One big shortcoming is that there is not any reviews data for some of the games on the game_sales table.

SQL QUERIES – MISSING REVIEW SCORES

✓ Join games_sales and reviews
✓ Select a count of the number of games where both critic_score and user_score
  are null

SELECT COUNT(game_sales.game)
FROM game_sales
LEFT JOIN reviews
ON reviews.game=game_sales.game
WHERE critic_score is NULL AND user_score IS NULL;

Count: 31

3. Years that video game critics loved

✓It looks like a little less than ten percent of the games on the game_sales table don't have any reviews data.

✓That's a small enough percentage that we can continue our exploration, but the missing reviews data is a good thing to keep in mind as we move on to evaluating results from more sophisticated queries.

✓There are lots of ways to measure the best years for video games! Let's start with what the critics think

SQL QUERIES – YEARS THAT VIDEO GAME CRITICS LOVED

✓ Select release year and average critic score for each year, rounded and aliased
✓ Join the game_sales and reviews tables
✓ Group by release year

✓ Order the data from highest to lowest avg_critic_score and limit to 10 results

```
SELECT year, ROUND(AVG(critic_score),2) AS avg_critic_score
FROM game_sales
INNER JOIN reviews
ON reviews.game=game_sales.game
GROUP BY year
ORDER BY avg_critic_score DESC
LIMIT 10;
```

SQL QUERIES – YEARS THAT VIDEO GAME CRITICS LOVED

year	avg_critic_score
1990	9.80
1992	9.67
1998	9.32
2020	9.20
1993	9.10
1995	9.07
2004	9.03
1982	9.00
2002	8.99
1999	8.93

%%sql-- Select release year and average critic score for each year, rounded and aliased-- Join the game_sales and reviews tables-- Group by release year-- Order the data from highest to lowest avg_criti

SQL QUERIES - DATA EXPLORATION

4. WAS 1982 REALLY THAT GREAT?

✓ The range of great years according to critic reviews goes from 1982 until 2020: we are no closer to finding the golden age of video games!

✓ Hang on, though. Some of those avg_critic_score values look like suspiciously round numbers for averages. The value for 1982 looks especially fishy. Maybe there weren't a lot of video games in our dataset that were released in certain years.

✓ Let's update our query and find out whether 1982 really was such a great year for video games.

SQL QUERIES - WAS 1982 REALLY THAT GREAT?

✓Paste your query from the previous task; update it to add a count of games released in each year called num_games
✓Update the query so that it only returns years that have more than four reviewed games

**SELECT year, ROUND(AVG(critic_score),2) AS avg_critic_score,
COUNT(game_sales.game) AS num_games
FROM game_sales
INNER JOIN reviews
ON reviews.game=game_sales.game
GROUP BY year
HAVING COUNT(game_sales.game)>4
ORDER BY avg_critic_score DESC
LIMIT 10;**

SQL QUERIES - WAS 1982 REALLY THAT GREAT?
Outcome

year	avg_critic_score	num_games
1998	9.32	10
2004	9.03	11
2002	8.99	9
1999	8.93	11
2001	8.82	13
2011	8.76	26
2016	8.67	13
2013	8.66	18
2008	8.63	20
2012	8.62	12

5. Years that dropped off the critics' favorites list

✓That looks better! The num_games column convinces us that our new list of the critics' top games reflects years that had quite a few well-reviewed games rather than just one or two hits.

✓But which years dropped off the list due to having four or fewer reviewed games? Let's identify them so that someday we can track down more game reviews for those years and determine whether they might rightfully be considered as excellent years for video game releases!

Set theory skills application here. To get started, I created tables with the results of our previous two queries:

top_critic_years

column	type	meaning
year	int	Year of video game release
avg_critic_score	float	Average of all critic scores for games released in that year

top_critic_years_more_than_four_games

column	type	meaning
year	int	Year of video game release
num_games	int	Count of the number of video games released in that year
avg_critic_score	float	Average of all critic scores for games released in that year

SQL QUERIES – YEARS THAT DROPPED OFF THE LIST OF CRITICS' FAVOURITE

✓Select the year and avg_critic_score for those years that dropped off the list of critic favorites
✓Order the results from highest to lowest avg_critic_score

SELECT year, avg_critic_score
FROM top_critic_years
WHERE year NOT IN (
SELECT year FROM top_critic_years
INTERSECT
SELECT year FROM top_critic_years_more_than_four_games
GROUP BY year
HAVING num_games > 4)
ORDER BY avg_critic_score DESC;

SQL QUERIES – YEARS THAT DROPPED OFF THE LIST OF CRITICS' FAVOURITE

year	avg_critic_score
1990	9.80
1992	9.67
2020	9.20
1993	9.10
1995	9.07
1982	9.00

SQL QUERIES – YEARS VIDEO GAME PLAYERS LOVED

6. Years video game players loved

✓Based on our work in the task above, it looks like the early 1990s might merit consideration as the golden age of video games based on critic_score alone, but we'd need to gather more games and reviews data to do further analysis.

✓Let's move on to looking at the opinions of another important group of people: players! To begin, let's create a query very similar to the one we used in Task Four, except this one will look at user_score averages by year rather than critic_score averages.

SQL QUERIES – YEARS VIDEO GAME PLAYERS LOVED

✔ Select year, an average of user_score, and a count of games released in a given year, aliased and rounded
✔ Include only years with more than four reviewed games; group data by year
✔ Order data by avg_user_score, and limit to ten results

SELECT year, ROUND(AVG(user_score),2) AS avg_user_score,
COUNT(game_sales.game) AS num_games
FROM game_sales
INNER JOIN reviews
ON reviews.game=game_sales.game
GROUP BY year
HAVING COUNT(game_sales.game) > 4
ORDER BY avg_user_score DESC
LIMIT 10;

SQL QUERIES – YEARS VIDEO GAME PLAYERS LOVED

year	avg_user_score	num_games
1997	9.50	8
1998	9.40	10
2010	9.24	23
2009	9.18	20
2008	9.03	20
1996	9.00	5
2005	8.95	13
2006	8.95	16
2000	8.80	8
1999	8.80	11

7. Years that both players and critics loved

✓ Alright, we've got a list of the top ten years according to both critic reviews and user reviews. Are there any years that showed up on both tables? If so, those years would certainly be excellent ones!

✓ Recall that we have access to the top_critic_years_more_than_four_games table, which stores the results of our top critic years query from Task 4:

SQL QUERIES – YEARS THAT BOTH PLAYERS AND CRITICS LOVED

`top_critic_years_more_than_four_games`
We've also saved the results of our top user years query from the previous task into a table

column	type	meaning
year	int	Year of video game release
num_games	int	Count of the number of video games released in that year
avg_critic_score	float	Average of all critic scores for games released in that year

SQL QUERIES – YEARS THAT BOTH PLAYERS AND CRITICS LOVED

✓Select the year results that appear on both tables

**SELECT year
FROM top_critic_years_more_than_four_games
INTERSECT
SELECT year
FROM top_user_years_more_than_four_games;**

SQL QUERIES – YEARS THAT BOTH PLAYERS AND CRITICS LOVED

year
1998
2008
2002

8. Sales in the best video game years

✓ Looks like we've got three years that both users and critics agreed were in the top ten! There are many other ways of measuring what the best years for video games are, but let's stick with these years for now. We know that critics and players liked these years, but what about video game makers? Were sales good? Let's find out.

✓ This time, we haven't saved the results from the previous task in a table for you. Instead, we'll use the query from the previous task as a subquery in this one! This is a great skill to have, as we don't always have write permissions on the database we are querying.

SQL QUERIES – SALES IN THE BEST VIDEO GAME YEARS

✓ Select year and sum of games_sold, aliased as total_games_sold; order results by total_games_sold descending. Filter game_sales based on whether each year is in the list returned in the previous task:

SELECT year, SUM(games_sold) AS total_games_sold

FROM game_sales

WHERE year IN (

SELECT year

FROM top_critic_years_more_than_four_games

INTERSECT

SELECT year

FROM top_user_years_more_than_four_games)

GROUP BY year

ORDER BY total_games_sold DESC;

SQL QUERIES – SALES IN THE BEST VIDEO GAME YEARS

year	total_games_sold
2008	175.07
1998	101.52
2002	58.67

THANK YOU