

PlayStation Game Sales Analysis Project

Subject – Big Data Analytics and Architecture

Project: PlayStation Game Sales Analysis

Name: - SHASHI BHUSHAN

Roll No: - 1240259044

Project Overview

This report analyses PlayStation game sales and metadata (PS3, PS4, PS5). The analysis uses SQL-style (Hive) queries to extract key metrics: top publishers, best-selling games, regional sales patterns, genre performance, release year trends, ratings correlation with sales, and other business-focused insights.

Dataset Description

Column Name
Game
Console
Name
Publisher
Developer
Total Shipped
Total Sales
NA Sales
PAL Sales
Japan Sales
Other Sales
Release Date
Last Update
rating
Ratings count
metacritic
genres
platforms
primary_genre
Release Date Parsed
release_year

Project Objectives

1. Understand sales distribution across consoles and regions.
2. Identify top-performing publishers, developers, and games.
3. Analyze genre-wise performance and average sales.
4. Explore correlations between ratings/metacritic and sales.
5. Provide business insights for marketing and product teams.

Technologies Used

Hive/SQL-style queries, Python (pandas) for pre-processing, and this PDF report generation.

15 Hive Queries, Outputs, and Insights

Name	Console	Publisher	Total Sales	release_year
Grand Theft Auto V	PS3	Rockstar Games	20320000.0	2013.0
Grand Theft Auto V	PS4	Rockstar Games	19390000.0	2014.0
Call of Duty: Black Ops 3	PS4	Activision	15090000.0	2015.0
Red Dead Redemption 2	PS4	Rockstar Games	13940000.0	2018.0
Call of Duty: Black Ops II	PS3	Activision	13800000.0	2012.0
Call of Duty: WWII	PS4	Activision	13400000.0	2017.0
Call of Duty: Modern Warfare 3	PS3	Activision	13350000.0	2011.0
Call of Duty: Black Ops	PS3	Activision	12670000.0	2010.0

Question-1.

SELECT * FROM playstation_games LIMIT 10;

Output (sample)**Insight:** Preview of first 10 rows to verify structure and sample data.

Question-2.

SELECT COUNT(*) AS total_games FROM playstation_games;

Output (sample):
4963

Insight: Total rows/games in dataset: 4963.

Question-3.

SELECT Console, COUNT(*) AS game_count FROM playstation_games GROUP BY Console ORDER BY game_count DESC;

Output (sample):

Console	game_count
PS4	1991
PS3	1892
PS5	1080

Insight: Shows distribution of game titles across PS3, PS4, and PS5.

Question-4.

SELECT Publisher, SUM(`Total Sales`) AS total_sales FROM playstation_games GROUP BY Publisher ORDER BY total_sales DESC LIMIT 5; Output (sample):

Publisher	Total Sales
Activision	205940000.0
Electronic Arts	153760000.0
EA Sports	129990000.0
Ubisoft	128370000.0
Sony Computer Entertainment	103700000.0

Insight: Top publishers by cumulative sales; indicates market leaders.

Question-5.

SELECT Name, Console, `Total Sales` FROM playstation_games ORDER BY `Total Sales` DESC LIMIT 5; Output (sample):

Name	Console	Total Sales
Grand Theft Auto V	PS3	20320000.0
Grand Theft Auto V	PS4	19390000.0
Call of Duty: Black Ops 3	PS4	15090000.0
Red Dead Redemption 2	PS4	13940000.0
Call of Duty: Black Ops II	PS3	13800000.0

Insight: Best-selling games across PlayStation platforms.

Question-6.

SELECT Console, ROUND(AVG(rating),2) AS avg_rating FROM playstation_games WHERE rating IS NOT NULL GROUP BY Console;

Output (sample):

Console	avg_rating
PS3	2.59
PS4	2.83
PS5	2.98

Insight: Average user rating per console (NaNs ignored).

Question-7.

SELECT Name, Console, rating FROM playstation_games WHERE rating > 4.8 ORDER BY rating DESC;

Output (sample):

No results.

Insight: List of highest-rated games (rating > 4.8).

Question-8.

SELECT primary_genre, COUNT(*) AS genre_count FROM playstation_games GROUP BY primary_genre ORDER BY genre_count DESC LIMIT 5; Output (sample):

primary_genre	genre_count
Unknown	2160
Adventure	529
Action	415
Shooter	346
Indie	301

Insight: Most common primary genres in the dataset.

Question-9.

SELECT primary_genre, ROUND(AVG(`Total Sales`),2) AS avg_sales FROM playstation_games GROUP BY primary_genre ORDER BY avg_sales DESC LIMIT 10;

Output (sample):

primary_genre	avg_sales
Sports	983205.13
Shooter	911416.18
Action	653325.3
Fighting	533928.57
Family	510416.67
Card	426666.67
Simulation	394200.0
RPG	342247.19

Insight: Average global sales per primary genre; highlights high-earning genres.

Question-10.

SELECT Console, ROUND(AVG(`NA Sales`),2) AS avg_NA, ROUND(AVG(`PAL Sales`),2) AS avg_PAL, ROUND(AVG(`Japan Sales`),2) AS avg_Japan FROM playstation_games GROUP BY Console; Output (sample):

Console	avg_NA	avg_PAL	avg_Japan
PS3	185570.82	159714.59	35771.67
PS4	117348.07	139994.98	20316.42
PS5	0.0	0.0	0.0

Insight: Regional (NA, PAL, Japan) average sales per title by console.

Question-11.

SELECT Name, Console, `Release Date` FROM playstation_games WHERE `Release Date` >= '2020-01-01';
Output (sample):

Name	Console	Release Date
AO Tennis 2	PS4	2020-01-09
Dragon Ball Z: Kakarot	PS4	2020-01-17
Arc of Alchemist	PS4	2020-01-30
Dawn of Fear	PS4	2020-02-03
7th Sector	PS4	2020-02-05
Death end re;Quest 2	PS4	2020-02-13
Azur Lane: Crosswave	PS4	2020-02-13
Dreams	PS4	2020-02-14

Insight: Games released in the modern/PS5 era (from 2020 onward).

Question-12.

SELECT CORR(metacritic, `Total Sales`) AS correlation FROM playstation_games WHERE metacritic IS NOT NULL;

Output (sample):
0.22653412919804727

Insight: Pearson correlation between Metacritic score and total sales (positive -> critics align with sales).

Question-13.

SELECT Developer, ROUND(AVG(rating),2) AS avg_rating, COUNT(*) AS total_titles FROM playstation_games WHERE rating IS NOT NULL GROUP BY Developer HAVING COUNT(*) > 5 ORDER BY avg rating DESC LIMIT 10;

Output (sample):

Developer	avg_rating	total_titles
Obsidian Entertainment	4.1	6
id Software	4.08	9
Naughty Dog	4.08	13
Harmonix Music Systems	4.07	8
Sucker Punch Productions	4.07	6
Irrational Games	4.05	7
Polyphony Digital	4.05	6
Rocksteady Studios	3.91	6

Insight: Developers with consistently high average user ratings (min 6 titles).

Question-14.

```
SELECT Name, platforms FROM playstation_games WHERE platforms LIKE '%,%';
```

Output (sample):

Name	platforms
Grand Theft Auto V	PC, PlayStation 5, Xbox One, PlayStation 4, Xbox Series S/X, Xbox 360, PlayStatio
Grand Theft Auto V	PC, PlayStation 5, Xbox One, PlayStation 4, Xbox Series S/X, Xbox 360, PlayStatio
Red Dead Redemption 2	PC, Xbox One, PlayStation 4
Call of Duty: Black Ops II	PC, Xbox One, Xbox 360, PlayStation 3, Wii U
Call of Duty: WWII	PC, Xbox One, PlayStation 4
Call of Duty: Modern Warfare 3	PC, Xbox One, iOS, macOS, Xbox 360, PlayStation 3, Wii
Call of Duty: Black Ops	PC, Xbox One, iOS, Nintendo DS, macOS, Xbox 360, PlayStation 3, Wii
FIFA 18	PC, Xbox One, PlayStation 4, Nintendo Switch, Xbox 360

Insight: Games released on multiple platforms (multi-platform reach).

Question-15.

```
SELECT SUBSTR(`Release Date`,1,4) AS release_year, SUM(`Total Sales`) AS yearly_sales FROM playstation_games GROUP BY SUBSTR(`Release Date`,1,4) ORDER BY release_year; Output (sample):
```

release_year	yearly_sales
2006.0	10670000.0
2007.0	61570000.0
2008.0	102880000.0
2009.0	107790000.0
2010.0	121990000.0
2011.0	155480000.0
2012.0	105290000.0
2013.0	122910000.0

Insight: Total sales aggregated by release year; shows peaks and trends.

Key Insights Summary

- PS4 has the largest number of titles in the dataset, reflecting its long lifecycle.
- Top publishers (based on Total Sales) include major industry players and platform holders.
- Action and Sports genres contribute strongly to high average sales.
- There is a positive correlation (if present) between Metacritic scores and sales.
- Multi-platform releases tend to have wider reach and higher cumulative sales.
- Peak sales years align with major franchise releases and console transitions.

Conclusion

This Hive-style analysis provides a business-focused summary of PlayStation game sales. The findings can guide marketing, release planning, and portfolio strategy for publishers and platform owners.