**VIDEO GAME SALES ANALYSIS**

An SQL project analyzing video game sales data to uncover key insights into sales trends, console preferences, and regional popularity using python for visualizations.

**TABLE OF CONTENTS**

Introduction

Objectives

Use Case

Dataset Description

Data Cleaning and Processing

Data Analysis

Data Visualization

Recommendation

Conclusion

**INTRODUCTION**

The video game industry has grown significantly over the years, with multiple platforms (consoles) and genres catering to different audiences worldwide. Understanding sales trends can help developers, publishers, and marketers make informed decisions about game development and distribution.

This project analyzes video game sales data to uncover key insights into sales trends, console preferences, and regional popularity.

### ****OBJECTIVES****

The main goals of this analysis are:

1. Identify the **best-selling games worldwide**.
2. Determine which **year had the highest sales** and analyze the industry's growth over time.
3. Explore whether **certain consoles specialize in specific genres**.
4. Identify **titles that performed well in one region but flopped in another**.
5. Provide recommendations based on the findings.

### ****DATASET DESCRITION****

The dataset for this project is sourced from Maven Analytics website designed specifically for practice purposes. It is presented in an Excel file with one table having 14 columns and 64,016 rows.

The dataset contains the following columns:

* img – Image link for the game
* title – Name of the game
* console – The platform it was released on (e.g., PS4, Xbox, Switch)
* genre – The game’s genre (e.g., Action, RPG, Sports)
* publisher – The company that published the game
* developer – The studio that developed the game
* critic\_score – Review score given by critics
* total\_sales – Total worldwide sales (in millions)
* na\_sales, jp\_sales, pal\_sales, other\_sales – Regional sales breakdown
* release\_date – The date the game was released
* last\_update – The most recent update in the dataset

### ****DATA CLEANING AND PREPROCESSING****

* Handle missing values (e.g., missing critic\_score or release\_date).
* Convert release\_date to a proper DATE format.
* Standardize genre names to avoid inconsistencies (e.g., "Action" vs. "Action Game").

**DATA ANALYSIS USING SQL**

1. **Which Titles Sold the Most Worldwide?**

Select Top 1 Title,

Round(Sum (total\_sales), 0) As Total\_Sales

From vgchartz\_2024

Group by Title

Order by Sum (total\_sales) DESC

This query identifies the highest-selling games of all time.

1. **Which Year Had the Highest Sales? Has the Industry Grown?**

Select Top 1 release\_date As release\_year,

Round(Sum (total\_sales), 0) As Total\_Sales

From vgchartz\_2024

Group by release\_date

Order by Sum (total\_sales) DESC

This query finds the **peak year** for video game sales.

The query below sort’s release\_year chronologically to help visualize the **growth trends of the industry**.

SELECT release\_date AS release\_year,

Round(Sum(total\_sales), 0) AS total\_sales

FROM vgchartz\_2024

WHERE release\_date IS NOT NULL

GROUP BY release\_date

ORDER BY release\_year ASC;

### ****Insight Gained****:

* Allows visualization of the industry growth or decline.
* A steady increase suggests industry expansion, while a decline could indicate market saturation or changing trends.

1. **Do Consoles Specialize in a Particular Genre?**

Find the Number of Games per Console and Genre

SELECT console, genre, COUNT(\*) AS game\_count

FROM vgchartz\_2024

GROUP BY console, genre

ORDER BY console, game\_count DESC;

### ****Insight Gained****:

* This query shows how many games exist for each genre per console.
* Helps identify if a console has a strong focus on a specific genre

sql

CopyEdit

WITH GenreRank AS (

SELECT console, genre, COUNT(\*) AS game\_count,

RANK() OVER (PARTITION BY console ORDER BY COUNT(\*) DESC) AS rank

FROM your\_table

GROUP BY console, genre

)

SELECT console, genre, game\_count

FROM GenreRank

WHERE rank = 1;

* + This identifies the **most common genre** for each console.

1. **What Titles Were Popular in One Region but Flopped Elsewhere?**

sql

CopyEdit

SELECT title,

na\_sales, jp\_sales, pal\_sales, other\_sales,

(na\_sales - (jp\_sales + pal\_sales + other\_sales)) AS na\_bias,

(jp\_sales - (na\_sales + pal\_sales + other\_sales)) AS jp\_bias,

(pal\_sales - (na\_sales + jp\_sales + other\_sales)) AS pal\_bias

FROM your\_table

ORDER BY na\_bias DESC, jp\_bias DESC, pal\_bias DESC

LIMIT 10;

* + This query highlights **games that were successful in one region but underperformed in others**.

### ****Findings & Insights****

* **Best-Selling Games**: The top games are usually from well-known franchises like Grand Theft Auto, Call of Duty, and Pokémon.
* **Industry Growth**: The gaming industry saw a **major increase in sales around the mid-2000s**, potentially due to the rise of online multiplayer gaming and digital sales.
* **Console-Genre Trends**:
  + **PlayStation & Xbox**: Dominated by Action and Shooter games.
  + **Nintendo Consoles**: Have a strong presence in Platformers and Family-friendly games.
  + **Handheld Consoles**: Tend to have more RPG and Adventure games.
* **Regional Popularity Differences**:
  + **JRPGs** tend to sell well in Japan but underperform in North America.
  + **Sports games** are popular in North America and Europe but struggle in Japan.
  + **First-person shooters** perform best in North America.

**DATA VISUALIZATION**

Using Python (Matplotlib & Seaborn) to make the insights more engaging. Below are some visualization ideas along with the relevant Python code.

## ****1. Most Sold Video Games Worldwide (Bar Chart)****

**Insight:** Displays the top 10 best-selling games globally.

### ****Code:****

python

CopyEdit

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# Load dataset

df = pd.read\_csv("your\_dataset.csv")

# Select top 10 best-selling games

top\_games = df.nlargest(10, "total\_sales")

# Plot

plt.figure(figsize=(12, 6))

sns.barplot(x="total\_sales", y="title", data=top\_games, palette="viridis")

plt.xlabel("Total Sales (Millions)")

plt.ylabel("Game Title")

plt.title("Top 10 Best-Selling Video Games Worldwide")

plt.show()

### ****Expected Output:****

A **horizontal bar chart** showing the **top-selling games and their total sales**.

## ****2. Yearly Sales Trend (Line Chart)****

**Insight:** Shows whether the gaming industry is growing over time.

### ****Code:****

python

CopyEdit

# Convert release\_date to datetime format

df["release\_date"] = pd.to\_datetime(df["release\_date"], errors='coerce')

# Extract year and sum sales per year

yearly\_sales = df.groupby(df["release\_date"].dt.year)["total\_sales"].sum().reset\_index()

# Plot

plt.figure(figsize=(12, 6))

sns.lineplot(x="release\_date", y="total\_sales", data=yearly\_sales, marker="o", color="b")

plt.xlabel("Year")

plt.ylabel("Total Sales (Millions)")

plt.title("Yearly Video Game Sales Trend")

plt.grid(True)

plt.show()

### ****Expected Output:****

A **line graph** showing the **rise and fall of total sales per year**, indicating industry growth.

## ****3. Console Specialization in Genres (Heatmap)****

**Insight:** Identifies which consoles dominate specific genres.

### ****Code:****

python

CopyEdit

# Create a pivot table of console vs genre counts

console\_genre\_pivot = df.pivot\_table(index="console", columns="genre", aggfunc="size", fill\_value=0)

# Plot

plt.figure(figsize=(15, 6))

sns.heatmap(console\_genre\_pivot, cmap="coolwarm", linewidths=0.5, annot=True, fmt="d")

plt.xlabel("Genre")

plt.ylabel("Console")

plt.title("Console Specialization in Game Genres")

plt.show()

### ****Expected Output:****

A **heatmap** where **dark colors indicate high game counts** for a particular **console-genre combination**.

## ****4. Regional Sales Comparison (Stacked Bar Chart)****

**Insight:** Shows sales distributions across different regions for top games.

### ****Code:****

python

CopyEdit

# Select top-selling games

top\_games = df.nlargest(10, "total\_sales")

# Plot

top\_games.plot(

x="title",

y=["na\_sales", "jp\_sales", "pal\_sales", "other\_sales"],

kind="bar",

stacked=True,

figsize=(12, 6),

colormap="tab10"

)

plt.xlabel("Game Title")

plt.ylabel("Total Sales (Millions)")

plt.title("Regional Sales Distribution of Top 10 Games")

plt.legend(["NA Sales", "JP Sales", "PAL Sales", "Other Sales"])

plt.xticks(rotation=45)

plt.show()

### ****Expected Output:****

A **stacked bar chart** showing **how different regions contribute to total sales** for top-selling games.

## ****5. Popularity in One Region vs. Another (Scatter Plot)****

**Insight:** Identifies games that are **hits in one region but flops in another**.

### ****Code:****

python

CopyEdit

# Plot sales in NA vs Japan

plt.figure(figsize=(8, 6))

sns.scatterplot(x=df["na\_sales"], y=df["jp\_sales"], alpha=0.6)

plt.xlabel("North America Sales (Millions)")

plt.ylabel("Japan Sales (Millions)")

plt.title("Regional Popularity: NA vs. Japan")

plt.axhline(y=df["jp\_sales"].mean(), color="r", linestyle="--", label="JP Avg Sales")

plt.axvline(x=df["na\_sales"].mean(), color="b", linestyle="--", label="NA Avg Sales")

plt.legend()

plt.show()

### ****Expected Output:****

A **scatter plot** showing **games that are popular in North America but not in Japan**, and vice versa.

### ****RECOMMENDATIONS****

* **Game Developers**: Target specific consoles based on genre success.
* **Publishers**: Market region-specific games effectively (e.g., JRPGs in Japan).
* **Retailers**: Stock games based on **regional preferences** to maximize sales.

### ****CONCLUSION****

* **Sales trends** show that certain games and genres are more successful in specific regions.
* The **gaming industry has grown over time**, with spikes in certain years.
* Some **consoles are better suited for specific game genres**.