

Video Game Sales Analysis Project

Python Explanatory Data Analysis (EDA)

Objective

The goal of this phase is to conduct a detailed explanatory data analysis using python. This will:

- Handle missing and inconsistent values
- Generate descriptive statistics
- Create more advanced visualisations than Excel allows
- Explore relationships between variables (e.g., critic/user scores vs values)
- Provide insights that will inform the next phases (SQL analysis and Power BI dashboard)

Tools and Libraries

Python 3.x with the following libraries:

- **Pandas:** Data manipulation and cleaning
- **NumPy:** Numerical computations
- **Matplotlib & Seaborn:** Visualisation
- **SciPy/Statsmodels:** Statistical analysis

Tasks

Importing and Cleaning the Data

- Imported the dataset into Python using Pandas with 17,416 rows and 15 columns.
- Scaled User_Score from 0-10 range to 0-100 range for consistency with Critic_Score.
- Handled missing values:
 - Filled missing Critic_Score and User_Score with the median of their respective columns.
 - Filled missing Rating values with "Unknown".
 - Dropped 8 rows missing Years_of_Release
- Final dataset: 17,408 rows.
- Saved a cleaned dataset (vgames_cleaned.csv) for further analysis.

Note:

- Columns Critic_Count and User_Count remain largely missing (>50% each) but were retained for now.
- The cleaned dataset ensures no missing values remain in the key columns for analysis.

Descriptive Statistics

	Year	NA Sales	EU Sales	JP Sales	Other Sales	Global Sales	Critic Score	Critic Count	User Score	User Count
count	17408	17408	17408	17408	17408	17408	17408	8335	17408	7798
mean	2006.63	0.25	0.14	0.08	0.05	0.52	70.00	26.19	73.28	162.67
std	5.90	0.80	0.50	0.30	0.18	1.53	9.72	18.99	10.23	562.84
min	1976	0	0	0	0	0.01	13	3	0	4
25%	2003	0	0	0	0	0.05	71	11	75	10
50%	2008	0.07	0.02	0	0.01	0.16	71	21	75	25
75%	2011	0.23	0.1	0.03	0.03	0.45	71	36	75	81
max	2017	41.36	28.96	10.22	10.57	82.54	98	113	97	10766

Dataset Overview:

- Rows after cleaning 17,408
- Platforms: 31
- Genres: 12
- Publishers: 627
- Ratings: 9
- Years: 1976-2017

Release Years:

- Range: 1976 to 2017
- Median year of release: 2008
- The majority of releases occurred between 2003 and 2011 (based on quartiles).

Sales Distribution:

- Total global sales: ~8,994.6 million units
- Regional contribution:
 - North America ~49%
 - Europe: ~27%
 - Japan: ~15%
 - Other regions: ~9%
- Average global sales per game: ~0.52 million units

Critic & User Scores:

- Critic_Score: Mean = 70, range = 13-98
- User_Score: Mean = 73 (scaled 0-100), Range = 0-97
- Median score for both: 75

Most Common Categories:

- Platforms: DS, PS2, PS3, Wii, PSP
- Genres: Action, Sports, Miscellaneous, Role-Playing, Adventure
- Publishers: Electronic Arts, Activision, Namco Bandai, Ubisoft, Konami

- Ratings: *Unknown*, E, T, M, E10+

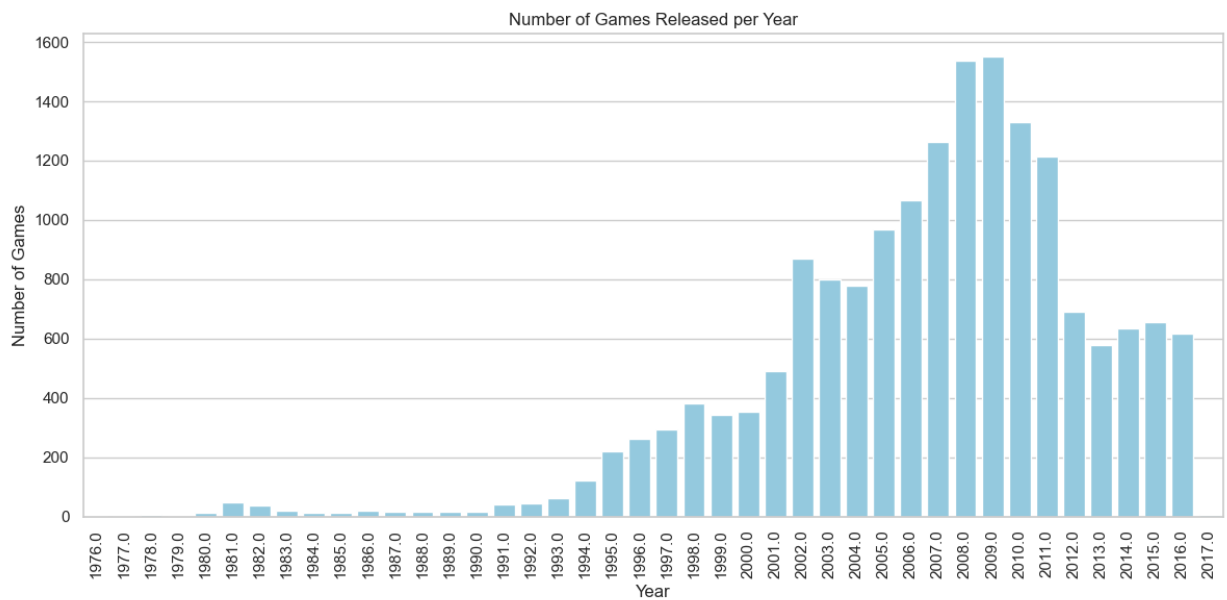
Methodology Note:

These statistics were generated using Python's Pandas .describe() method for numerical columns and value counts for categorical columns. Missing values in key score columns were imputed with median values to maintain data consistency.

Explanatory Visualisations

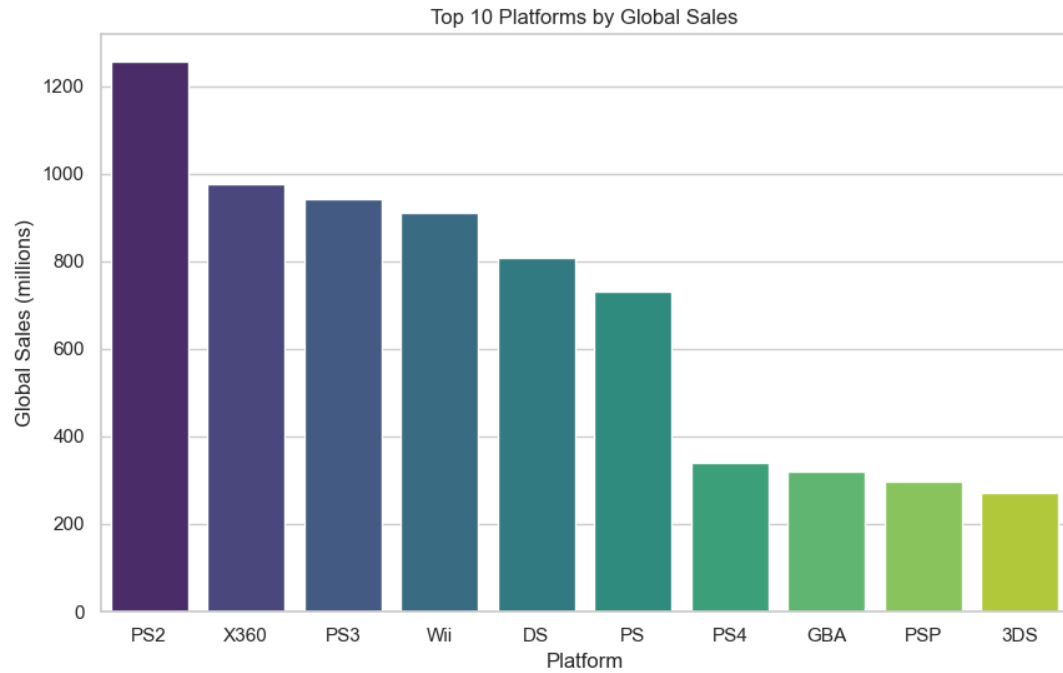
Key Visualisations

1) Number of Games Released per Year



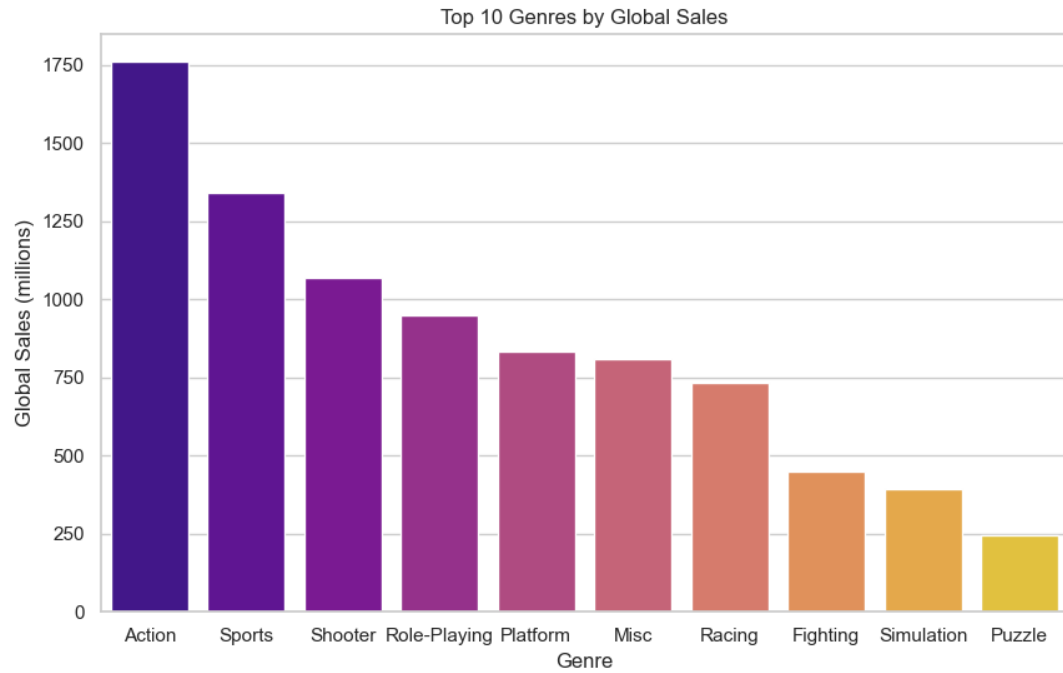
- Sharp growth from 1995 to 2008.
- The peak occurred between 2006 and 2011, followed by a decline after 2011.
- This aligns with the lifecycle of popular consoles like the PlayStation 2, Wii, and Xbox 360, and the digital shift.

2) Top 10 Platforms by Global Sales



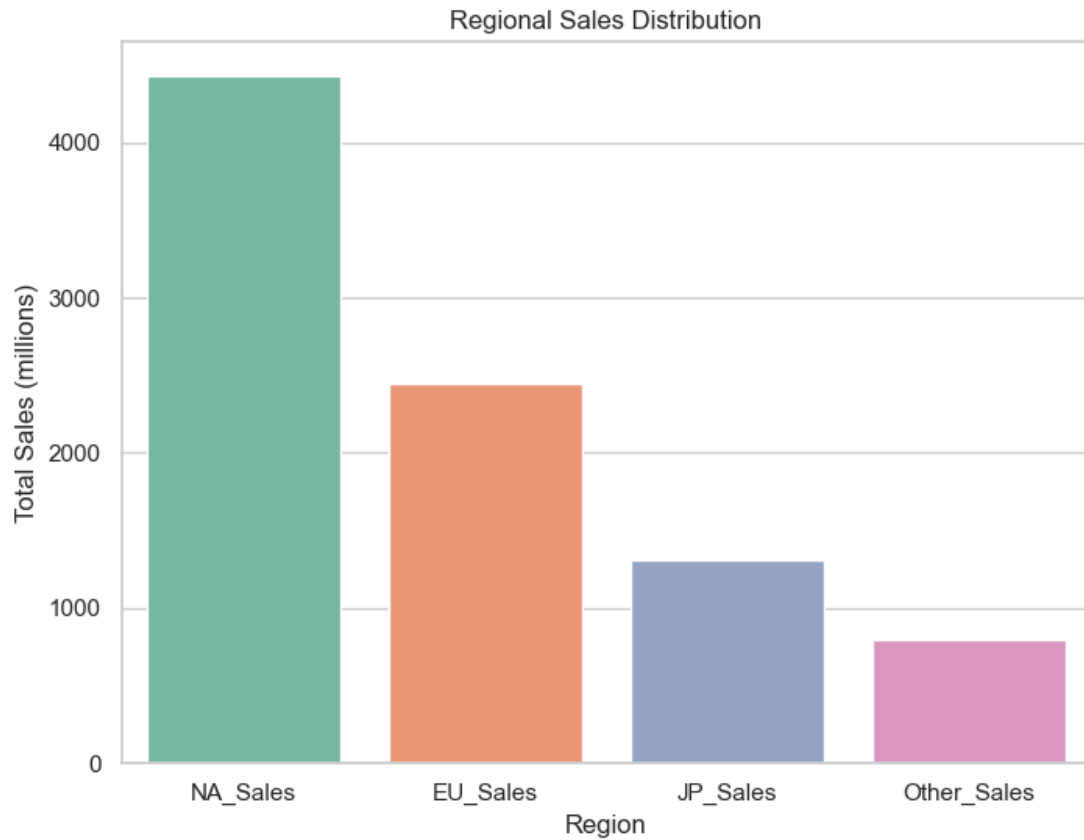
- The PlayStation 2 leads with over 1,250 million units sold globally.
- Other dominant platforms include Xbox 360, PlayStation 3, and Wii.
- Nintendo platforms collectively also make a strong showing (DS, Wii, GBA, 3DS).
- Xbox 360 ranks 2nd in sales despite not being top 5 in number of games.

3) Top 10 Genres by Global Sales



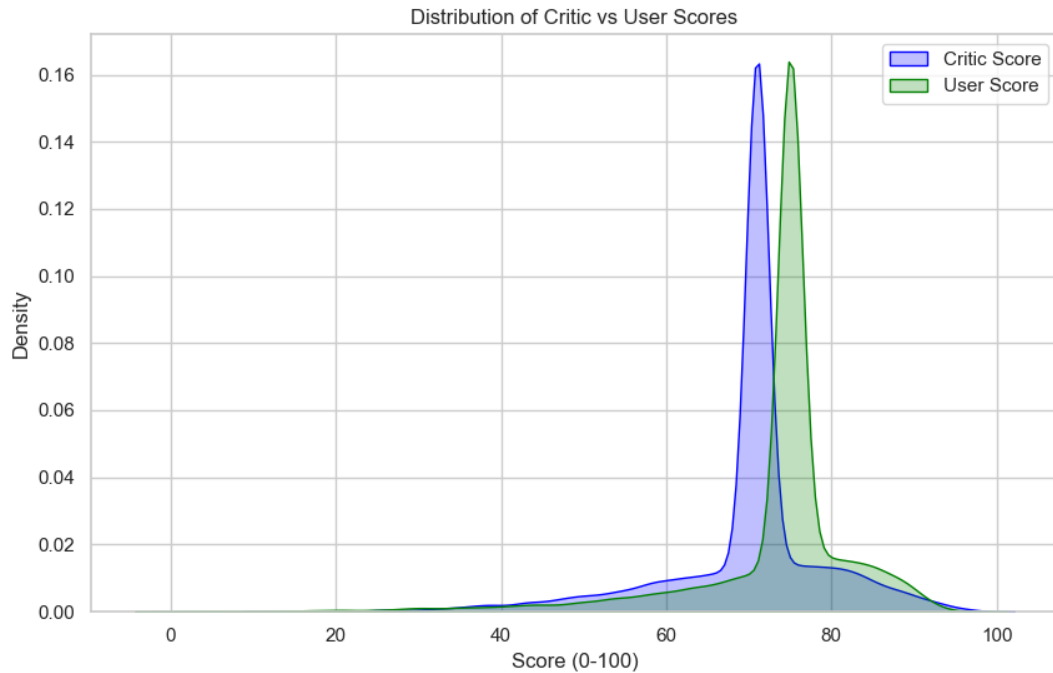
- **Action** games dominate global sales, followed by **Sports** and **Shooter** titles.
- **Shooter** ranks high despite not being top 5 in number of games, showing efficiency in sales.
- **Role-Playing** and **Platform** games also contribute significantly.
- **Puzzle** and **Simulation** games account for smaller shares.
- **Adventure** appears in top 5 by game count but not in sales.

4) Regional Sales Distributions



- **North America** leads with ~49% of total sales.
- **Europe** follows with ~27% and **Japan** accounts for ~15%.
- **Other** regions collectively contribute ~9%.
- This reflects regional market preferences and console availability.
- Japan performs strongly despite being a single country vs. entire continents.

5) Distribution of Critic vs User Scores



- Both critics and users tend to score games around 70-80 out of 100.
- Both distributions peak around 75/100.
- KDE and scatterplots show general agreement, but some outliers diverge strongly.
- The distribution shows that user scores are slightly more spread out, while critics' scores are tighter.
- This suggests critics may be more consistent in their ratings, while users show a wider range of opinions.

Insights

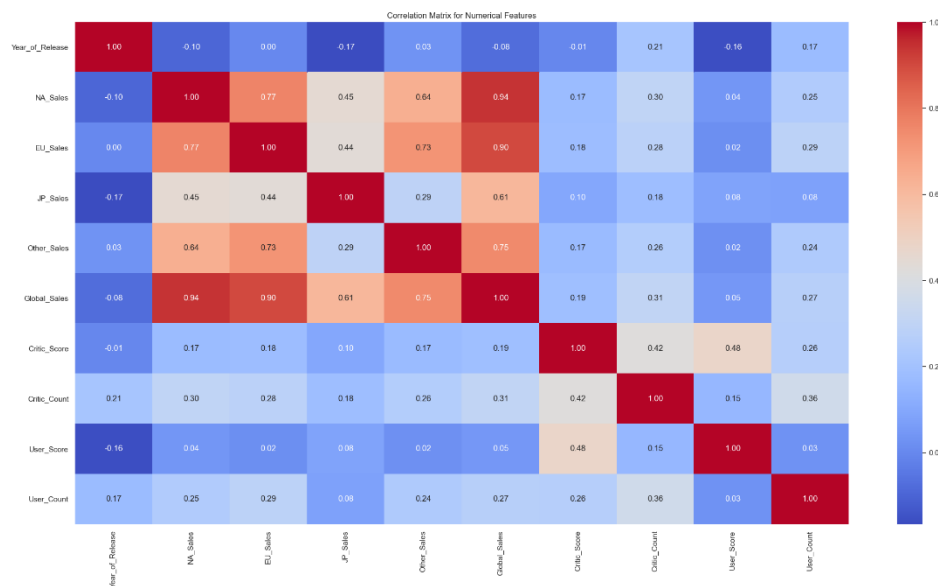
- **Peak Years:** The number of games released peaked between 2006 and 2011, aligning with the global sales peak during those years. This likely reflects the release of multiple major consoles in that period (Xbox 360, PlayStation 3, Nintendo Wii).
- **Xbox 360 Sales Impact:** The Xbox 360 achieved the second-highest global sales despite not being among the top 5 platforms by number of games released, indicating strong performance per title.
- **Shooter Genre Success:** Shooter games rank 3rd in global sales, even though the genre is not in the top 5 number of games, suggesting high demand and profitability for fewer releases.
- **Japan's Market Share:** Despite being a single country compared to multi-country regions like North America and Europe, Japan contributes ~15% of global sales, underlining its strong influence in the gaming market.

- **Nintendo's Collective Dominance:** While the PlayStation 2 is the single highest-selling platform, Nintendo consoles collectively dominate the top-selling platforms list (Wii, DS, GBA, 3DS, NES, SNES, N64, GameCube).
- **Sales Distribution Skew:** The average global sales per game (~0.52 million units) is much lower than the blockbusters in the top 10 list, highlighting that a few hit titles drive a large share of sales.
- **Critic vs User Score Alignment:** Both critics and users cluster around 70-80 out of 100, showing broad agreement, though user scores have more variation (suggesting more diverse opinions).
- **Post-2010 Decline:** After 2010, the number of new game releases and global sales both declined, possibly reflecting the rise of digital distribution and mobile gaming, which are underrepresented in the dataset.

Advanced Insights

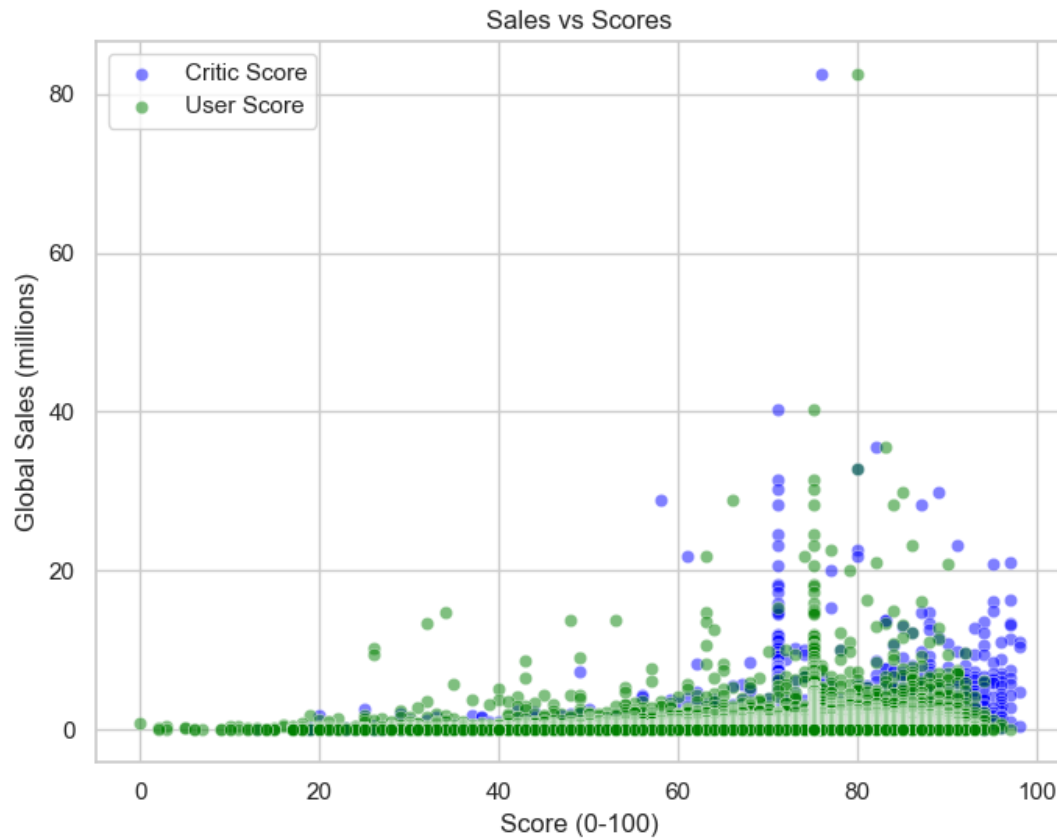
To explore deeper patterns in the dataset, we created advanced visualisations highlighting relationships between sales, genres, and publishers.

Correlation Analysis:



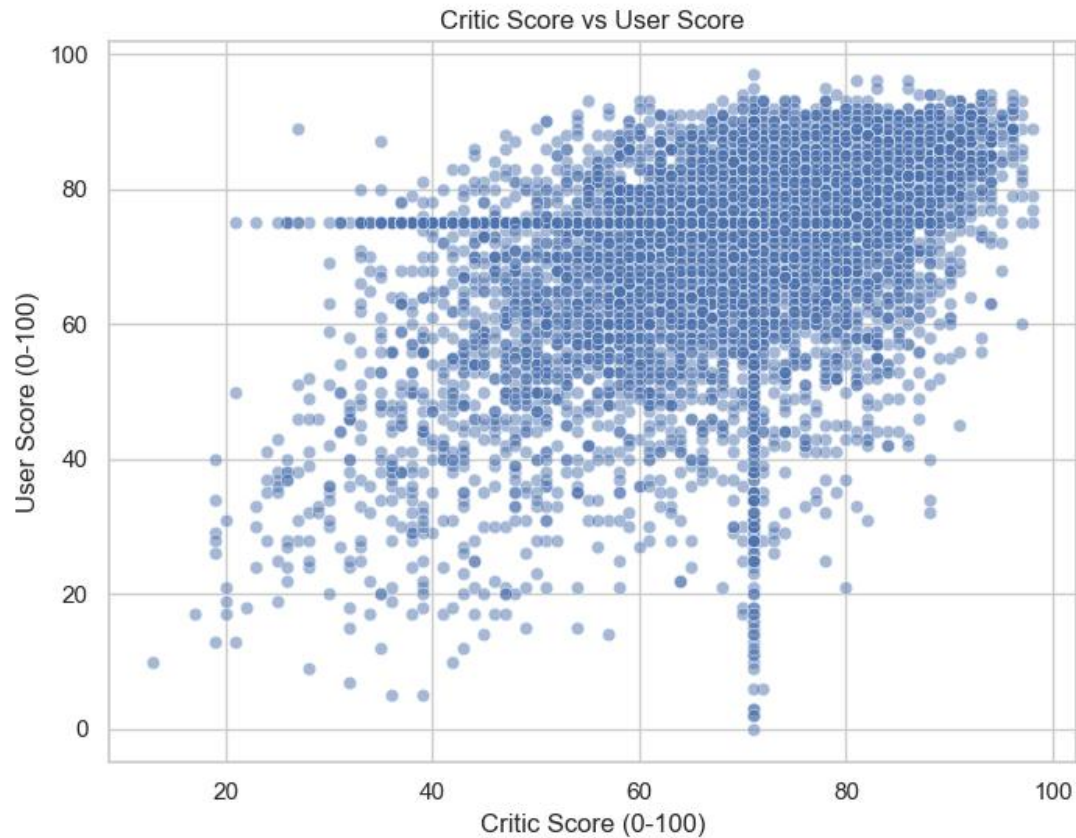
- The correlation matrix revealed strong positive relationships between regional sales and global sales, confirming that performance in North America and Europe largely drives global success.
- Critic_Score and User_Score showed moderate correlation (~0.5), indicating general agreement but with room for divergence.
- Review counts (Critic_Count and User_Count) were positively associated with scores, suggesting more popular titles attract greater critical and user attention.

Scores and Sales:



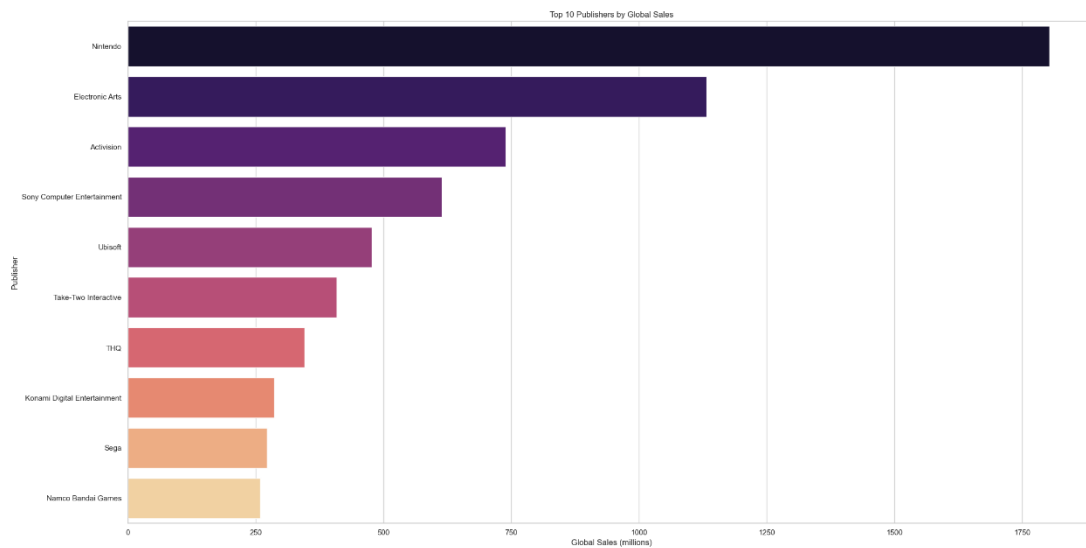
- Scatterplots showed that higher Critic and User Scores generally correspond with higher global sales.
- However, the relationship is not strictly linear, with many top-selling games clustering in the 70-90 score range. This suggests that while strong reviews support commercial success, other factors such as franchise strength, marketing, and platform also play major roles.

Critic vs User Perspective:



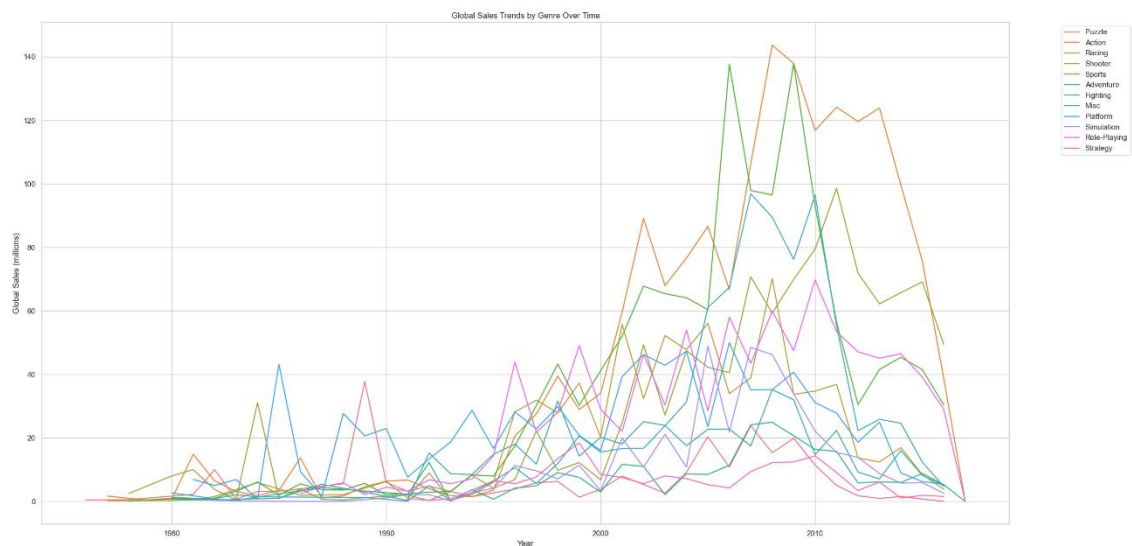
- Comparisons between Critic_Score and User_Score revealed substantial overlap but also some outliers.
- These highlight games that critics praised more than players (or vice versa), underscoring differences in expectations between professional reviewers and the broader player base.

Publisher Performance:



- A bar chart of the top 10 publishers by global sales showed Nintendo as the clear market leader, followed by Electronic Arts, Activision, and Sony. This indicates that a handful of major publishers dominate industry sales, leveraging strong intellectual property portfolio and established reputations.
- Nintendo leads global sales despite not being in top 5 for number of games, showing the dominance of its franchises (e.g., Mario, Pokémon)

Genre Trends Over Time:



- Line plots of genre sales over time highlighted significant growth from 2006 to 2011, coinciding with the release of the PlayStation 3, Xbox 360, and Nintendo Wii.
- Action, Sports, and Shooter genres dominated global sales during this period, while genres such as Role-Playing and Adventure maintained steady but smaller shares.
- Sales decline post-2011 aligns with digital distribution shift.

Key Takeaways

- Regional sales strongly influence global performance, especially in North America and Europe.
- Review scores correlate with sales, but franchise reputation and console availability are equally influential.
- Nintendo leads global sales despite not being among the top 5 publishers by number of games, showing the power of strong franchises.
- Scores and sales correlate positively, but many high-selling games cluster around mid-to-high scores, suggesting factors beyond reviews (e.g., brand, marketing).
- Shooter genre's rise during the 2000s aligns with the Xbox 360/PS3 boom, reinforcing console influence on genre popularity.
- Industry sales peaked between 2006 and 2011, driven by next-generation consoles and popular games like Action and Shooter.
- Post-2011 decline in sales across genres indicates market saturation and the shift to digital distribution models.
- Outliers in critic vs user scores highlight games that were either critical hits but not widely loved by players, or vice versa.
- Japan's relatively small geographic market still delivers significant global sales impact, reflecting cultural influence on gaming.
- PS2 remains the all-time leader in global sales, followed closely by Xbox 360.
- Shooter genres dominate global sales, even when not producing the most games.

Deliverables

- **Jupyter Notebook** (v_games_python_eda.ipynb) including:
 - Cleaned dataset loading and preprocessing
 - Descriptive statistics and value counts
 - Visualisations (line plots, scatterplots, heatmaps, etc.)
 - Written commentary explaining findings
- **Word Report Section:** Screenshots of key charts and written insights