**Video Game Sales With Ratings Data**

**Exploratory Analysis**

Jaylon Kiper, [jkiper@bellarmine.edu](mailto:jkiper@bellarmine.edu)

1. **INTRODUCTION**

This data set involves video game sales from Vgchartz and corresponds ratings from Metacritic. The data from Vgchartz focuses on the sales from North America, Europe, Japan, etc; based on the year of 2016. The data from Metacritic cover the aggregate score compiled by the staff and score by the website’s subscribers. Also included by Metacritic is the number of critics and users who gave the score to the games. I chose this topic due to my interest in video games and looking at these titles can show how popular some are based on their sales and score.

Source: <https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings>

1. **DATA SET DESCRIPTION**

This data set contains 11563 samples with 15 columns with various data types. A complete listing is shown in **Table 1.** There is some missing information from the sources of Metacritic as they only cover a selection of certain gaming platforms. This would probably include handhelds and older devices like NES, SNES, Gameboy, and the DS/3DS. However, there are some missing with more updated hardware as well. Some years included both 2017 and 2020 releases, but this data set covers only up to 2016, so this may be a mistake on the reviewers part.

**Table 1: Data Types and Missing Data**

|  |  |  |
| --- | --- | --- |
| *Variable Name* | *Data Type* | *Missing Data (%)* |
| Name | nominal/object | 0% |
| Platform | nominal/object | 0% |
| Year\_of\_Release | interval/object | 1.61% |
| Genre | nominal/object | 0% |
| Publisher | nominal/object | 0% |
| NA\_Sales | interval/float64 | 0% |
| EU\_Sales | interval/float64 | 0% |
| JP\_Sales | interval/float64 | 0% |
| Other\_Sales | interval/float64 | 0% |
| Global\_Sales | interval/float64 | 0% |
| Critic\_Score | ordinal/float64 | 51% |
| Critic\_Count | ratio/float64 | 51% |
| User\_Score | ordinal/float64 | 40% |
| User\_Count | ratio/float64 | 55% |
| Developer | nominal/object | 40% |
| Rating | ordinal/object | 40% |

1. **Data Set Summary Statistics**

This section covers the basic statistics of the data set. The first table is a summary of stats based on counting, averages, standard deviations, mins, maxes, etc. The table was created in Python using the describe function grabbing all this data.The next couple of smaller tables show the proportions of each categorical variable. I covered on the frequency of the highest amount of each variable and calculated their proportions based on the number of rows in the data set. The very last table in this section is a correlation matrix of all continuous variables in the data set. This includes the sales by Vgchartz and the scores by Metacritic.

**Table 2: Summary Statistics for IP4 (Video\_Games\_Sales\_2016)**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Variable Name* | *Count* | *Mean* | *Standard Deviation* | *Min* | *25th* | *50th* | *75th* | *Max* |
| NA\_Sales | 16719.00 | 0.263302 | 0.813565 | 0.00 | 0.00 | 0.08 | 0.24 | 41.36 |
| EU\_Sales | 16719.00 | 0.145045 | 0.503359 | 0.00 | 0.00 | 0.02 | 0.11 | 28.96 |
| JP\_Sales | 16719.00 | 0.077625 | 0.308871 | 0.00 | 0.00 | 0.00 | 0.04 | 10.22 |
| Other\_Sales | 16719.00 | 0.047343 | 0.186742 | 0.00 | 0.00 | 0.010 | 0.03 | 10.57 |
| Global\_Sales | 16719.00 | 0.533568 | 1.548125 | 0.01 | 0.06 | 0.17 | 0.47 | 82.53 |
| Critic\_Score | 8137.00 | 68.967679 | 13.938165 | 13.00 | 60.00 | 71.00 | 79.00 | 98.00 |
| Critic\_Count | 8137.00 | 26.360821 | 18.980495 | 3.00 | 12.00 | 21.00 | 36.00 | 113.00 |
| User\_Score | 7590.00 | 7.125046 | 1.500006 | 0.00 | 6.40 | 7.50 | 8.20 | 9.70 |
| User\_Count | 7590.00 | 162.229908 | 561.282326 | 4.00 | 10.00 | 24.00 | 81.00 | 10665.00 |

There should be a table for **EACH** categorical variable.

**Table 3: Proportions for Platform (n=PS2)**

|  |  |  |
| --- | --- | --- |
| *Category* | *Frequency* | *Proportion (%)* |
| *Platform* | *2161* | *12.93%* |

**Proportions for Year of Release (n=2008)**

|  |  |  |
| --- | --- | --- |
| *Category* | *Frequency* | *Proportion (%)* |
| *Year\_of\_Release* | *1427* | *8.54%* |

**Proportions for Genre (n=Action)**

|  |  |  |
| --- | --- | --- |
| *Category* | *Frequency* | *Proportion (%)* |
| *Genre* | *3396* | *20.32%* |

**Proportions for Publisher (n=Electronic Arts)**

|  |  |  |
| --- | --- | --- |
| *Category* | *Frequency* | *Proportion (%)* |
| *Publisher* | *1356* | *8.11%* |

**Proportions for Ubisoft (n=Ubisoft)**

|  |  |  |
| --- | --- | --- |
| *Category* | *Frequency* | *Proportion (%)* |
| *Developer* | *203* | *1.21%* |

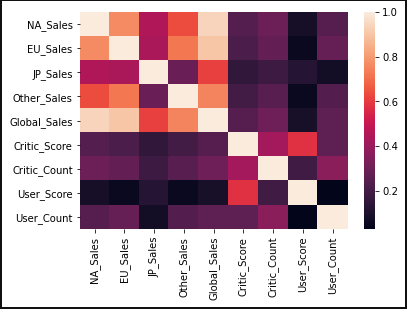
**Proportions for Rating (n=E)**

|  |  |  |
| --- | --- | --- |
| *Category* | *Frequency* | *Proportion (%)* |
| *Rating* | *3990* | *23.87%* |

**Table 4: Correlation Table/Tables**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **NA\_Sales** | **EU\_Sales** | **JP\_Sales** | **Other\_Sales** | **Global\_Sales** | **Critic\_Score** | **Critic\_Count** | **User\_Score** | **User\_Count** |
| **NA\_Sales** | 1.000000 | 0.765336 | 0.449598 | 0.638654 | 0.941010 | 0.240755 | 0.295413 | 0.086200 | 0.246429 |
| **EU\_Sales** | 0.765336 | 1.000000 | 0.435068 | 0.722796 | 0.901239 | 0.220752 | 0.277533 | 0.055337 | 0.283360 |
| **JP\_Sales** | 0.449598 | 0.435068 | 1.000000 | 0.291096 | 0.612300 | 0.152593 | 0.180219 | 0.125598 | 0.075638 |
| **Other\_Sales** | 0.638654 | 0.722796 | 0.291096 | 1.000000 | 0.749242 | 0.198554 | 0.251639 | 0.057119 | 0.238982 |
| **Global\_Sales** | 0.941010 | 0.901239 | 0.612300 | 0.749242 | 1.000000 | 0.245471 | 0.303571 | 0.088139 | 0.265012 |
| **Critic\_Score** | 0.240755 | 0.220752 | 0.152593 | 0.198554 | 0.245471 | 1.000000 | 0.425504 | 0.580878 | 0.264376 |
| **Critic\_Count** | 0.295413 | 0.277533 | 0.180219 | 0.251639 | 0.303571 | 0.425504 | 1.000000 | 0.194133 | 0.362334 |
| **User\_Score** | 0.086200 | 0.055337 | 0.125598 | 0.057119 | 0.088139 | 0.580878 | 0.194133 | 1.000000 | 0.027044 |
| **User\_Count** | 0.246429 | 0.283360 | 0.075638 | 0.238982 | 0.265012 | 0.264376 | 0.362334 | 0.027044 | 1.000000 |

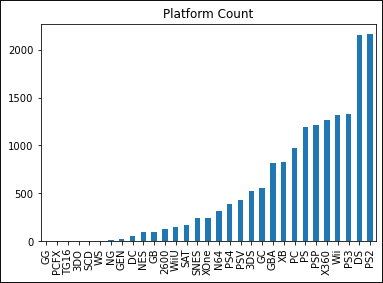
After the table with the raw data, include a heatmap of the correlation matrix as a figure.



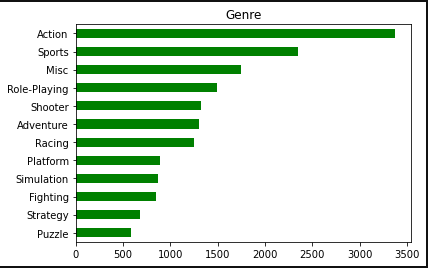
1. **DATA SET GRAPHICAL EXPLORATION**

This section will display various graphs built in Python based on the data set.There are seven graphs in total and all were made using Matplotlib and Seaborn packages. Both categorical and continuous variables were used to form these graphs.

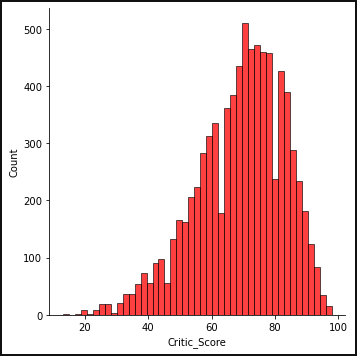
* 1. *Distributions*
  2. *ScatterPlots / Pairwise Plots (continuous variables)*
  3. *Barcharts (categorical variables)*
  4. *Other Plots - don’t skimp – there are likely other plots that would be useful that I haven’t already specified. Include those in this section.*



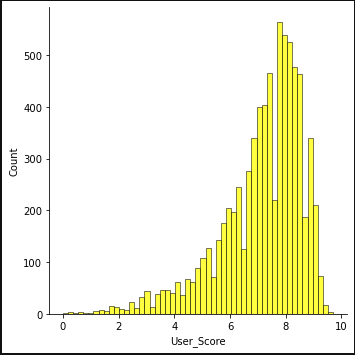
**Figure 1:** Bar chart showing the distinct count for each platform per game.



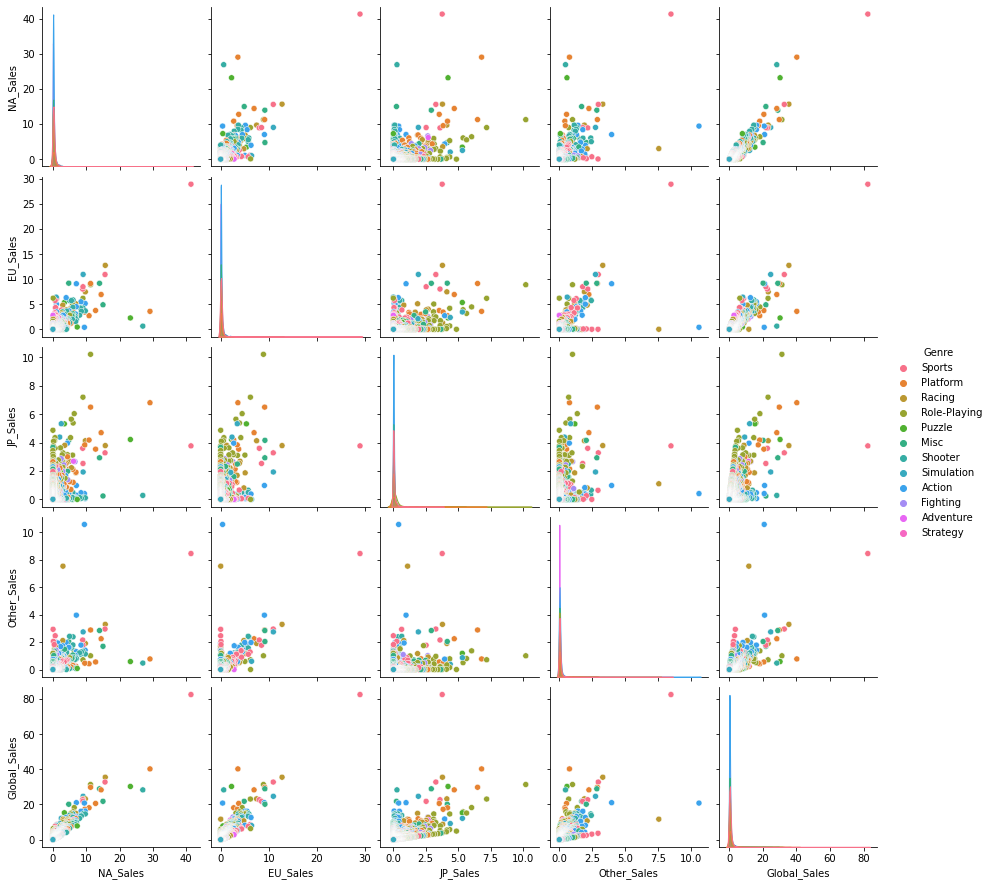
**Figure 2:** Horizontal bar chart showing the distinct count for each genre per game.



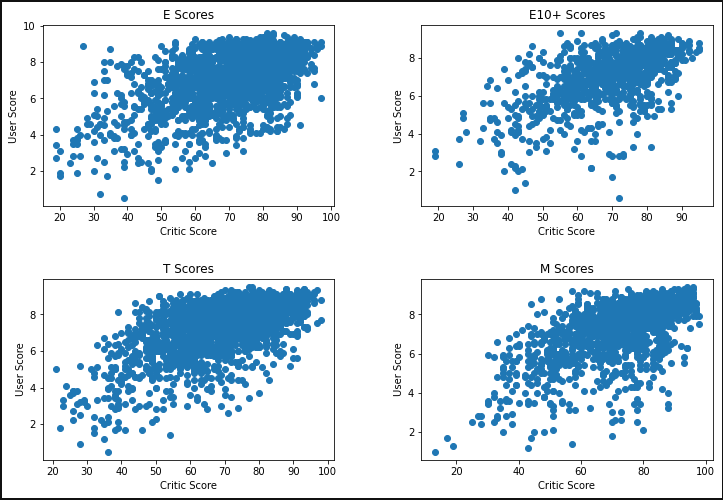
**Figure 3:** Distribution plot with the critic scores per game.



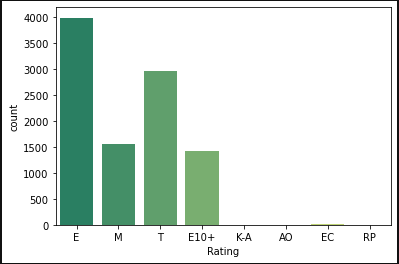
**Figure 4:** Distribution plot with the user scores per game.



**Figure 5:** Pairplot comparing the relationships between all games sales by genre.



**Figure 6:** Scatter subplot comparing Critic and User Score by the games Rating.



**Figure 7:** Countplot with a summer palette with ratings

1. **SUMMARY OF FINDINGS**

While looking into this data set, there are a lot of interesting facts to be told. When getting a count on all the platforms of each game, the Sony PS2 and Nintendo DS take up a significant amount of the set. The PS2 is a very popular system, even considered to be one of the best of all time. The title exclusives including God of War, GTA, and Final Fantasy give the system an edge. The fact that these are exclusive to the console made the PS2 a lot more profitable. The same goes with the DS. Nintendo launching both old and new Pokemon titles on this handheld made the profit and review scores very high. Using distribution plots based on critic and user score, the average number lies between 64 and 78 with critics and between 6.6 and 8.4 with users. This shows what the score for your average video game is like. Scores below 60 or 6.0 are on the tail end of these plots and most games above 90 and 9.0 are very rare to come by in the industry at this time. There is another detail within both sales and review scores. If a game has relatively low sales, it won’t get a review from Metacritic. With the pairplot using all the sales figures grouped by genre, each section can show what's popular. In both NA Sales and EU Sales, sports games take up a high majority. This includes gaming titles like NBA, Madden NFL and MLB The Show in the United States with those sports. The EU Sales definitely are covered with the Fifa titles being popular here. JP Sales tell a different story. The majority here involves role-playing games. JRPG’s are a hit in this category with franchises like Final Fantasy, Persona, and Dragon Quest.