

Games Luminous- 29_Metagross

Elif Gizem Zedef

191180093

elifgizemzedef@gmail.com

Fatih Talha Tümer

191180081

fatihthalha06@gmail.com

Merjem Kesten

191180094

kestem4.3001.m@gmail.com

Ali Mert Kocaman

191180055

Alimert3215@gmail.com

ABSTRACT



In this report, we analyzed the game sales. While making these analyzes, we visualized the data and conducted research on these visualized data. During this visualization, we used the dataset containing the game sales in period from 1980-2020.

Keywords

“Dataset; game sales; data; visualization; analysis.”

1. INTRODUCTION

Nowadays, video games have become the biggest source of entertainment, so they become an industry and is a subject of deep research. The gaming industry is bigger than the movie and music industries combined, and it's only growing. There are more than two billion players worldwide, although not as much attention as the film and music industries. That is 26% of the world's population.

Sales in the gaming world are increasing day by day, and according to a study conducted in 2016, the video game industry surpassed the movie industry to become the highest-paid industry. GTA V, the game of Rockstar Games, which earned \$815.7 million in its first day alone, grossed \$ 3 billion in 2017, surpassing Avatar, the highest-grossing movie ever, with \$ 2.7 billion. According to the 2016 version of the Global Games Market Report, in 2016 alone the gaming industry brought in \$99.6 billion, while Hollywood generated \$36 billion in revenue.

It's no surprise that companies want a piece of the pie. In 2020, the gaming industry generated \$155 billion in revenue, with analysts predicting the industry will generate more than \$260 billion in revenue by 2025.

2. METHODOLOGY

2.1 Game sales

As you know in this paper we are analyzing video games sales and we imported the necessary libraries (pandas, numpy, sklearn, seaborn etc.) for our visualization to our jupyter notebook. The dataset that we are going to analyze is game_sales which we have imported as csv file. In the first part we obtained how many data and features we have by using 'shape' method. Besides we dropped the Rank column because we won't be using it.

Next, we checked if we have missing values. Since we realized that they are not too much, we dropped those missing values. Later, we realized that 'Year' features type is float. So, we changed its type to integer.

Secondly, we started the exploring part of the game_sales dataset. We created a bar chart which shows us counts of the games according to the 'Year' feature. Then, we created a pie chart that shows percentages of the genres. Later, we created five bar charts and these charts show us which 'Platform' has the highest sales rate in each region. We did the same thing by using 'Genre' feature. Since the number of publishers is very large, we decided to analyze the 20 that published the most games. Afterword in order to compare publishers, we selected 20 games that had the highest sales globally. Later we made heat map with 5 features that represent sales in some regions or countries. The last plot of exploring game_sales dataset part is line plot which shows genre and platform relationship.

The last part of analyzing the game_sales dataset is Regression which consisting of two parts. First one is Linear Regression and here we split the data at a ratio of 3:1 with train_test_split function from Sklearn which is machine learning software library. The target, for both parts, is to make predictions about the sales in Europe, Japan and North America regions. For plotting these predictions we used linear_model also from Sklearn library. In the second part of Regression we did Random Forest Regression by using RandomForestRegression from Sklearn library. One more time we split the data and we obtained the accuracy for train and test sets. Thus completing the analysis of game sales dataset and we started to analyses the Twitch dataset.

2.2 Twitch

Twitch is an interactive livestreaming service for content spanning gaming, entertainment, sports, music, and more[3].

First of all, we used the date and time library and the development environment of the jupyter notebook in addition to the libraries used in game sales. The data we received in the Twitch_game_dataset.csv file was transferred to the pandas library data frame structure. We started with analyzing a set of data with information, shape, and head functions. We then used the isnull function to check if there were zero values in the data set and saw that there was a null value. We deleted this value with the drop function. We later saw that the data in the "Hours of viewing" column of the data set was in the "----- hours" format and we deleted the "hours" section to process this data.

In the second part of our analysis, we used the matplotlib and seaborn libraries to visualize the data. We first visualized overall

In the last part, we perform a regression using the Twitch dataset. Here we start with the effect according to Hours_watched in Streamers. For this, we first split the data in two with the train_test_split function. The main goal is to the Streamers values by looking at the Hours Watched value.

As we mentioned before we fit a linear regression model to predict sales that included sales in Europe, Japan and North America as predictors. We obtained three linear plots (see one of them in Figure6.).

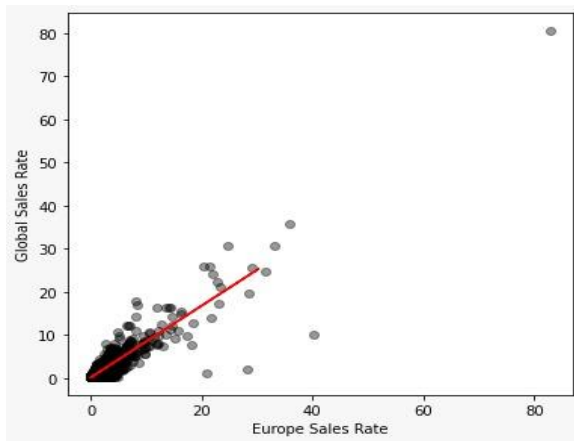


Figure7. Linear regression graph for Europe sales rate

And last but not least are results we got with Random Forest Regression (see Table1.).

Accuracy	Train set	Test set
North America	0.94	0.85
Europe	0.84	0.64
Japan	0.49	0.34

Table1. Accuracy results for train and test sets (Note: values can be changed because the train and test sets are received by random method)

4.2 Twitch

The dataset about Twitch streams and views contains from 1400 rows and 12 features. Exploratory data analysis focused on identifying and testing the effect of the correlated and uncorrelated variables. We have obtained 12 charts in this part (see Figure8, 9, 10, 11 and 12).

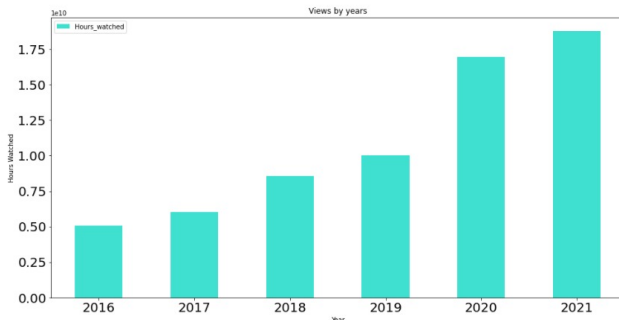


Figure8. Annual total views

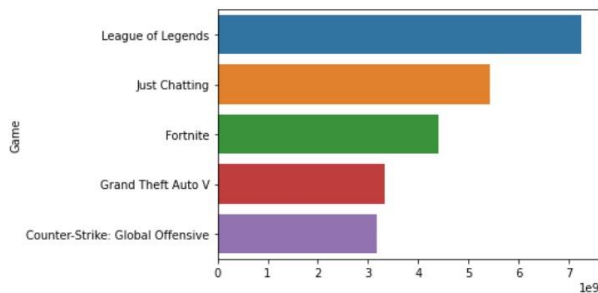


Figure9. Top 5 most watched games

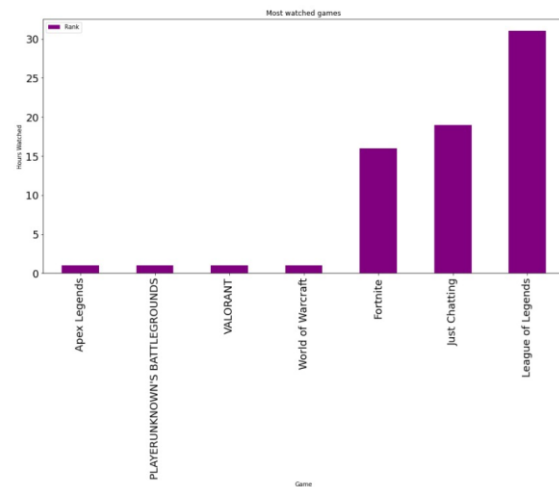


Figure10. Games that are placed in the 1st place the most

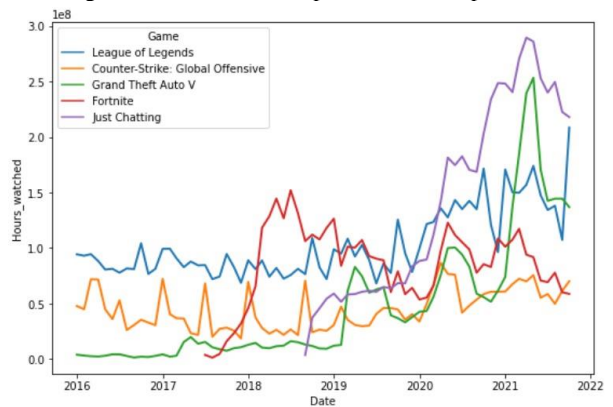


Figure11. Monthly views of the 5 most watched games

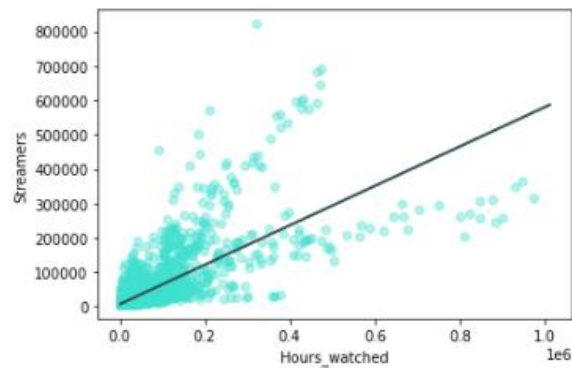


Figure12. Regression of Hours_watched and Streamers

5. CONCLUSION

Our analyses have shown that in dataset from 1980 to 2020, 2009 have been the year of game sales. Genre is a very important feature for us and as we saw in Figure 2, the action genre dominates with about one fifth. But that is not the case when we look at the regions separately. Only region named Other Countries have Action as leader of Genre feature. Considering that this data set is about sales publishers have big

role in it. After our analysis we have found that the EA publisher have published the most games but it is not the one who sold the most games, on that throne is Nintendo. Also one interesting thing we found in Platform feature is that the leading platform is Wii whose games is the first ranked game based on global sales, where the Publisher is Nintendo and the Genre is Sports. The thing we have learned by analyzing the Figure7 is that the Global sales rates and the sales rates in Europe are directly proportional to each other.

In the Twitch dataset, we monitored the viewing rates between 2016 and 2021. Especially between 2020 and 2021, we saw that the total game viewing rates increased. We thought that the pandemic was also effective in this. We looked at the games with the most views and saw that the top three were League of Legends, Just Chatting, and Fortnite. Afterwards, from the graph that calculates how many times the games were ranked 1st, we saw that the most League of Legends games were 1st 31 times, followed by Just Chatting 19 times and Fortnite 16 times. In the Twitch regression, we saw that there was not enough connection between watch hours and Streamers.

6. REFERENCES

- [1] <https://www.kaggle.com/vinodsunny1/insight-s-of-gaming-world/data>
- [2] <https://www.kaggle.com/rankirsh/evolution-of-top-games-on-twitch/version/10>
- [3] <https://www.twitch.tv/>