

Data Analysis of Steam DataSet

Hazar Belge
Galatasaray University, Computer Engineering
Istanbul, Turkey
Introduction to Data Analysis

I. INTRODUCTION

Steam is a video game digital distribution service by Valve. It was launched as a standalone software client in September 2003 as a way for Valve to provide automatic updates for their games and, expanded to include games from third-party publishers. Steam has also expanded into an online web-based and mobile digital storefront.

In this paper, we will look into a dataset about different video games with their different attributes like price, owner count, playtime, achievements count etc.

After having introduced the data throughly we will ask a research question and technical questions to help answer it. In the end we will have an explanation of the methods we predict on using to answer the questions that we previously posed.

II. QUESTIONS ABOUT OUR DATA

To analyse our data, first we need a high level question. Upon our introduction of the dataset, we have a much clearer picture regarding what we can focus on.

A. The High Level Question: What are the attributes that affect game sales the most?

While developing video games, the sales numbers of the game and the income we can generate from it are very important. That's why we want our game to sell well, but it is not certain what are the attributes that make this sale big. There are many factors that will affect game sales. While Flappy Bird with almost no budget makes millions, the game with a budget of millions of dollars may not be liked at all and may not make a profit. Which brings us to our high-level research question that may reveal the solution to this problem. "What are the attributes that affect game sales the most? Of course, we also need subquestions to answer this question.

B. Sub Questions

- Does having more reviews achievements affects the game sales positively?
- Do higher age ratings affect the game sales negatively?
- Does having english support affects the game sales positively?
- Do genres affect the game sales?
- Do categories affect the game sales?
- Does developers' name affects the game sales?

- Does having more platform support affects the game sales positively?
- Do higher prices affect the game sales negatively?
- Does publishers' name affects the game sales?
- Do release dates affect the game sales?
- Does having more reviews affects the game sales positively?

III. STEAM DATASET

This dataset is published by Nik Davis on Kaggle. It includes 27075 different video games and 18 different attributes for each of them from Steam. The dataset was last updated in May 2019, so I removed the games that were released in 2019 to prevent them from misleading the year-to-year comparison. After this removing process we have 24862 video games in total. So, let's take a look at some of attributes of dataset.

A. Achievements

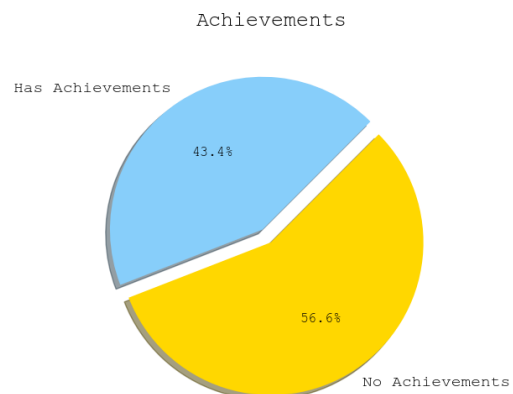


Fig. 1. How many games have achievements?

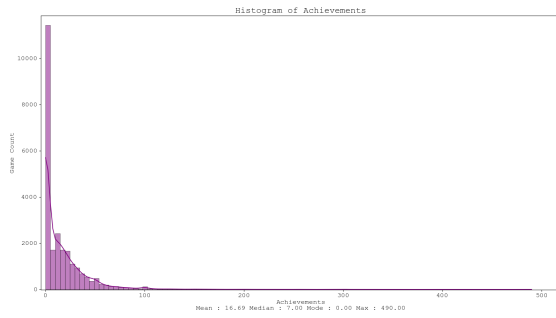


Fig. 2. Histogram of Achievements.

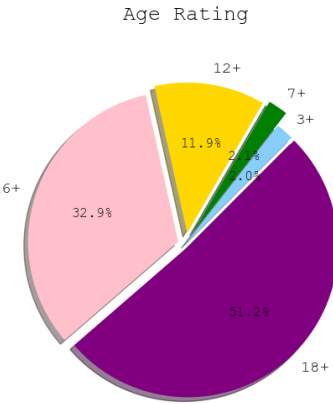


Fig. 4. Age Ratings Pie Chart.

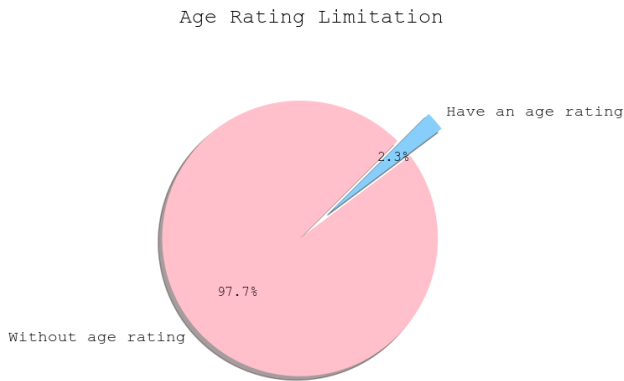


Fig. 3. How many games have age rating?

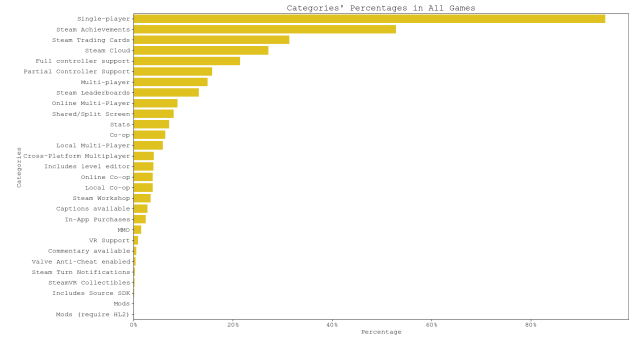


Fig. 5. Categories' percentages of all games.

- B. Age Rating
- C. Category
- D. English Support
- E. Genre
- F. Owner Count
- G. Platform
- H. Positive Rate
- I. Price
- J. Released Date

IV. CONSTRUCTING HYPOTHESIS TESTS AND THE METHODS

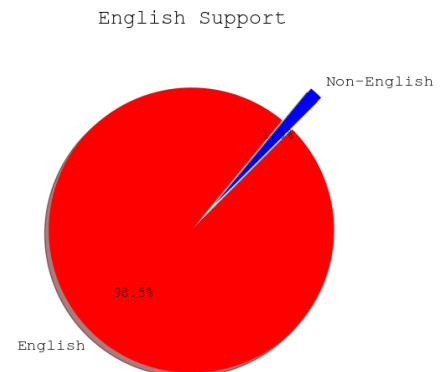


Fig. 6. How many games have English support?

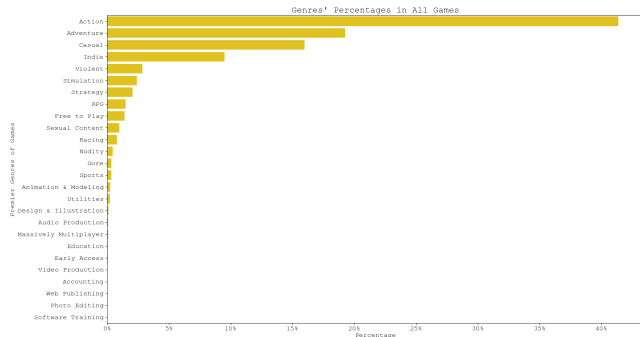


Fig. 7. Genres' percentages of all games.

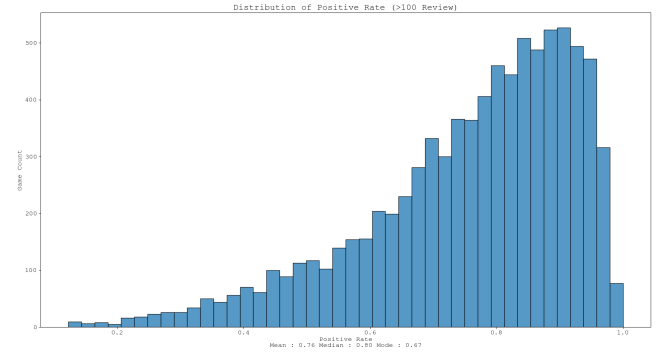


Fig. 10. Histogram of Positive Rate.

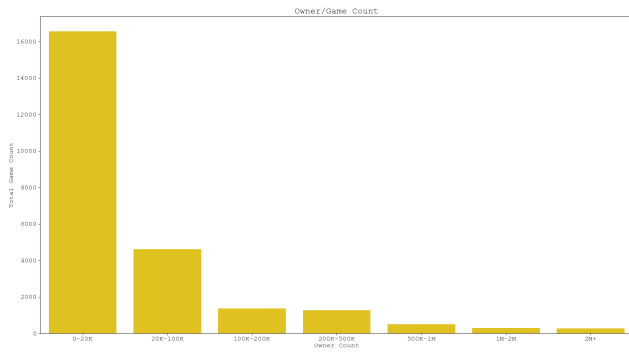


Fig. 8. Distribution of owners count in all games.

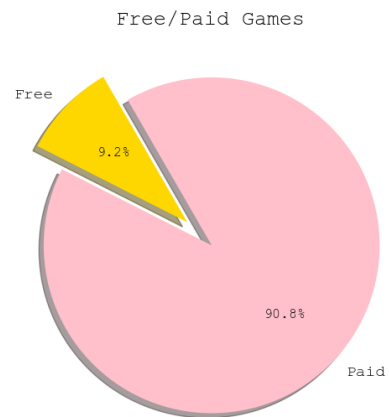


Fig. 11. How many games are paid?

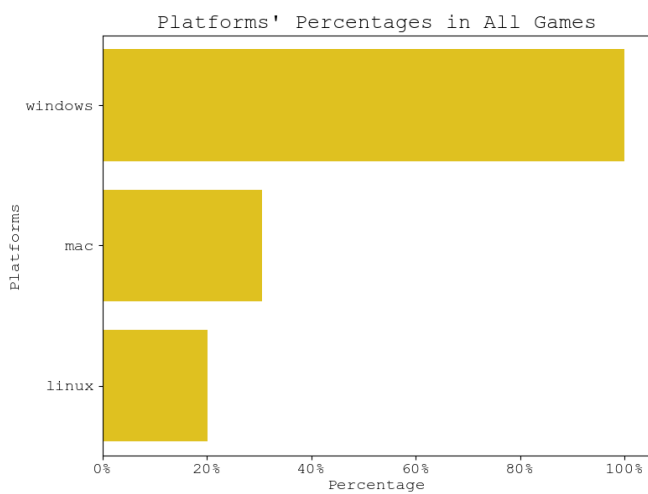


Fig. 9. Platforms' percentages of all games.

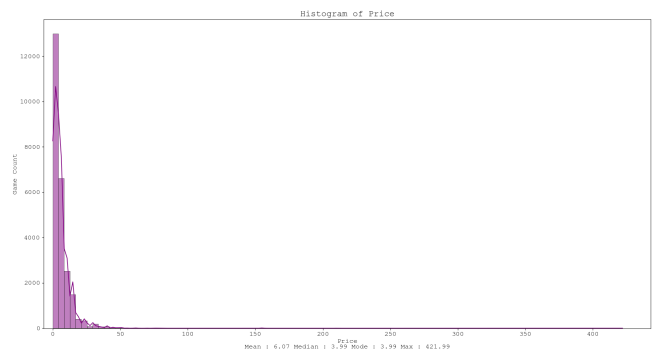


Fig. 12. Histogram of Price.

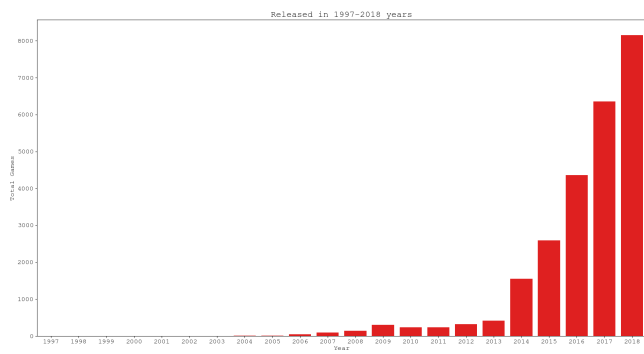


Fig. 13. Video Games Per Year.