A Network Tour of Data Science

Can I predict the success of a new videogame ?

1. Introduction

More and more video game platforms are setting a principle of physical dematerialization of games that only become accessible via download on an online portal. One of these platforms is **Steam**.

Steam is a video game digital distribution service launched as a standalone software client which provides automatic updates for their games. It is today the largest digital distribution platform for PC gaming with over a billion registered accounts with 90 million monthly active users. But each game has its own parameters and thus its own success. *What parameters influence the most its popularity ?* We shall try to answer this question among others.

2. Data acquisition:

In this project, we will attempt to find a way to estimate the success or failure of a released game according to different factors such as the category the game belongs to, its price, the gameplay average time, the release date. In order to do so, we have employed data that was extracted from the Kaggle dataset called "Steam Store Games (Clean dataset)" https://www.kaggle.com/nikdavis/steam-store-games.

3. Data exploration:

We shall use the different subsets available, load and clean the data. We have for the moment started this pre-processing phase and the different graphs are shown in annex.

As constructing a good graph is as important as visualizing it, several construction methods will be explored later in the project and used to extract as much information as possible.

4. Data exploitation

We will focus on understanding how the data is represented, what are the main variables and how each of these are defined. Thus, we will set edges and features of parameter and try to build a list of nodes that represent a certain game. The success will be quantified and measured using grades users have set, the number of people who purchased the game and an analysis of the comment these players leave.

The next step will consist of analyzing and comparing the basic properties of the different factors (price, date, etc…) through different clusters with the graphs. The use of a Laplacian is therefore required for such procedure.

In order to predict the success or failure of the game, we first have to determine what are the popularity parameters. Furthermore, it is necessary to reveal the hidden links that can exist with other success and failure based on the grading and correlate them to the different features.

5. Conclusion

Some basic common knowledge can allow us to anticipate without any data analysis nor machine learning some success factors, for example: action games are more successful than educational game. Will our results allow us to confirm our expectations? Will the success factors we determined allow us to correctly estimate the success of a new game depending on its known characteristics ? We will then discuss these results and summarize our findings.

Bonnesoeur Maxime
Hodara Michael
Kamel Tarik

04/12/19

Annex:

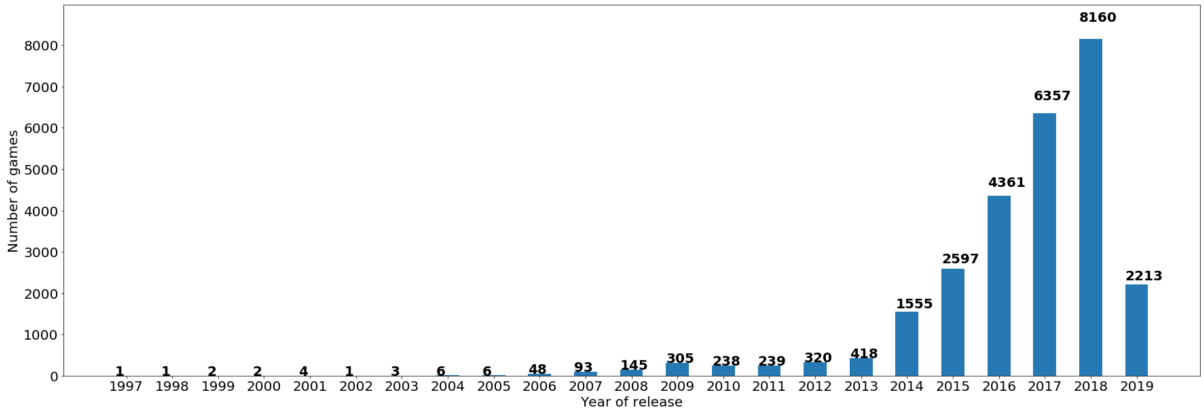

Figure 1 : Breakdown of the number of games presented in our data depending on the year of release

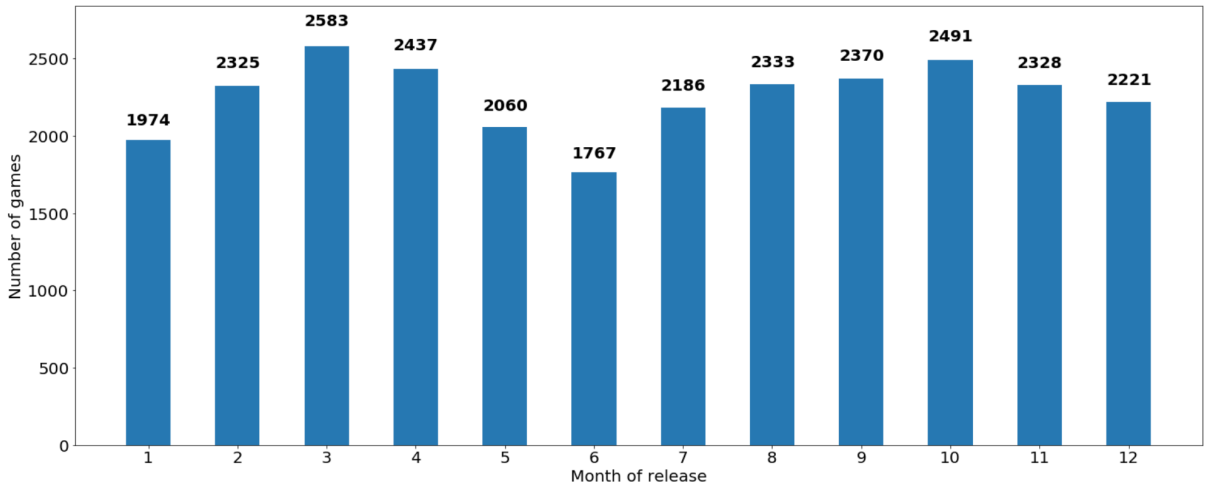

Figure 2 : Number of games released around the year showing us the trends of games release