Process Book

Matthieu Masouyé, Robin Plumey

June 2022

1 Introduction

Video games are still widely considered to be a new form of entertainment, despite existing for more than 50 years, and only recently has it been popularized to a larger and larger extent. For instance, in 2020, the video game market exceeded 300 billion in revenue, up from 8 billion in 2000.

It grew from it's humble beginning in a way few market did, and we want to convey that with our visualizations. We are also interested to see how well received games were by players, and how that shaped this market.

2 Problem Statement

In this project, we want to explore the different relationships between a video game's sales success and its reception by players and critics, and how it evolved over time.

Are the user ratings correlated to the sales success of a video game ? How did the market evolve over time ? Which companies came and went as the years passed ?

We want to show how it evolved, what was successful throughout the years in terms of sales and critical reception, and who were and still are the big players in the video game market.

3 Dataset

Our work focuses on two datasets Video Game Sales — Kaggle and Metacritic Ratings — Kaggle.

The first one, [Video Game Sales — Kaggle](https://www.kaggle.com/datasets/gregorut/videogamesales), comes from a scrape of the vgchartz.com website. It contains 16 598 records with the name, platform, publisher, year, genre and the related sales in europe, north-america, japan and others.

The second, [Metacritic Ratings — Kaggle](https://www.kaggle.com/datasets/xcherry/games-of-all-time-from-metacritic), includes all games from [Metacritics](https://www.metacritic.com/browse/games/ at the time of publication. It contains records for 8831 games, scrapped from

the website, with details about the score assigned by critics and users, platform, genre, developer, and link to the review of the game.

4 Data Pre-Processing

In the end, not a lot of pre-processing was needed on the dataset themselves; The first one on video game sales was already very clean, and while the second one on scores had a couple of problematic columns, it did not pose problem as we simply didn't use them (for instance "Genre" or "Type").

Merging them was a bit more complex: we decided to augment the first dataset with the attributes of the second, which were the Meta and User score, as well as the game's description from Metacritics. To do so, we compared each game name of the second, smaller dataset with the names in the first one using Jaro distance; we would then assign the most likely name to the second dataset's game name, and then merge it with an inner merge.

5 Architecture

We kept the code used to clean and merge the dataset in main, as well as the code used to create the visualizations for the Home page.

The website runs on github pages, and so the code for this is on the branch Page.

6 Home Page

6.1 Graphs

The graphs for the home page were done using the Plotly Python library. We chose this library for several reasons:

- We knew we wanted a slider through to navigate through time, and Plotly allows use to animate a graph in such a way with relative ease
- Plotly allows to export an interactive graph directly to an html folder, making it easy to integrate it into our webpage, instead of for instance having to setup a way for our page and python script to communicate while the page is loading
- This last feature also allows us to precompute the needed graphs, thus making the page load faster

The graphs are done mainly through correctly setting up the Pandas dataframe before giving it as an argument to a Plotly Bar graph: The first obstacle was implementing a sliding window in the dataframe to smooth out irregularities and better represent a video game's lifespan. We solved it by creating a new dataframe, and appending each groups of five years with a different id, such that

we could easily group by that id in the Plotly graph. While not very efficient on resources, as it multiplies the size of the original dataframe by the size of the window, we can get away with this solution as we have a relatiely small dataframe to begin with (a few thousands of entries). Then, it was a matter of sorting this DataFrame, grouping by the element of interest (Games, Publisher or Platform), and taking the top 10 for each sliding window.

Plotly allows us to configure the graph with a lot of tools, for instance we configured the color of the bar to represent an additional information, the Meta or User score. We were able to define custom colors picked to be colorblind friendly . Once we had graphs we were satisfied with, we could integrate them to the main page.

6.2 Page Presentation

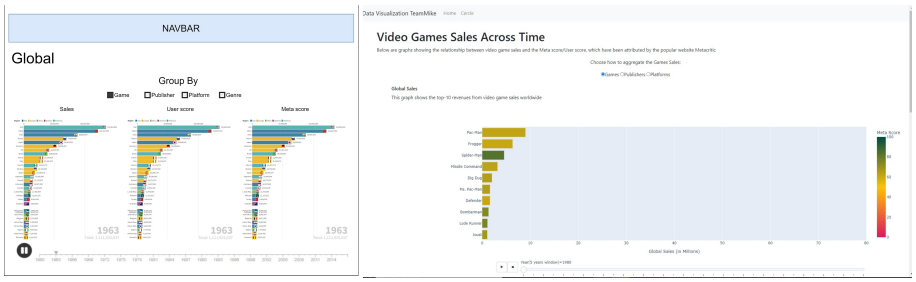

Figure 1: Original sketch vs final page

We stayed close to the original sketch in the end, as we can see in Figure 1. We kept the radio buttons to switch between grouping by Game name, Publisher and Platform. The Plotly graphs also closely resemble the original idea, with a slider on the bottom to choose the timeframe.

However, we wanted to show the graphs in a more compact way; on the original sketch, they are side by side, which wasn't possible with the Plotly graphs as they would be squished way too much for lisibility. Moreover, while having multiple graphs controlled by the same slider is technically feasible, it is very involved deep in the inner workings of Plotly animation to assign each frames to the correct slider.

Our original solution was to go through the three different graphs using buttons, with smooth animations. However, we always ran into some problems, no matter the approach: The graphs wouldn't be clickable anymore, it would break the timeline animation, etc. In the end, we had to leave the page in this rougher state due to time constraints.

7 Circles Page

This visualization was done with D3. We have selected this technology to have a dynamic animation. As we also wanted to be able to navigate in the different bubbles and to see all the different games.

The bubbles are created from a json file containing all the useful information such as the number of sales, the image and score.

The number of sales is represented by the size of the bubbles and the score is used to draw a colored circle according to it. To navigate in this visualization, you just have to click on the bubbles you want to visit and it will display the internal elements.

The major problem of this visualization is that a large part of the images link to the publisher. This is due to our choice of website which was not made for this bus only work with the game icon.

8 Conlusion

Overall, we think we managed to produce an interesting visualization. We didn't manage to add all the feature we wished to do, and there are problems left here and there. We could fix pretty much all of them with more time, but we still managed to produce a viable result.

9 Peer Assessment

As we were only two people, work distribution was facilitated: For the first two milestones, we both worked on it together, writing the reports and doing sketches for the visualizations. We cleaned and did the initial graphs on one dataset each.

For the final Milestone, we shared the workload as such:

- Robin worked on implementing the "Circles" page and visualizations
- Matthieu worked on merging both datasets and did the "Home" page and visualizations