

Data Visualization CA 2

Submitted by *Vikramaditya Tatke*

Student Number: x17163668

1. Abstract

For the purposes of this infographic we have selected the dataset available at <https://www.kaggle.com/egrinstein/20-years-of-games/version/2>. It consists of 18000 rows with corresponding to different titles and platforms from 1996 to 2016. For example, if a game title was released on 3 different platforms it is accommodated in 3 rows. For every title the dataset contains the release day, month and year specific to each platform the title was released on. Furthermore, a genre associated with the title along with a review score and score phrase given by the popular games reviewing website IGN has been included in the dataset. In addition to this, we can see whether a game was an editor's choice or not. Having such a huge collection of data is extremely useful for not only generating infographics but also to perform data analysis. Well classified data is extremely suitable for creating visualizations, as we can easily extract the information from the source. This also makes it easier to identify trends and make comparisons. As this dataset contains time-series information forming a story, while emphasizing on key elements becomes straightforward, which is the fundamental aspect of an infographic.

2. Process

Firstly, the downloaded data was transferred into a Pandas dataframe to carry out further manipulations. Outliers identified during checking were removed for better fact detection and a few rows containing null values were omitted from further examination. Total number of unique titles released in the past 20 years were calculated. We also determined the market share of various platforms. Then using the Seaborn library, we generated a countplot for all the years to count the number of titles released every year. Then using matplotlib we generated a pie chart showing the popularity of various platforms for gaming. Moving forward, bar graphs were generated to analyze the distribution of genres in the market. Then we generated bar graphs to count the number of masterpieces (according to IGN) released for every platform from 1996 to 2016. Again, by using the seaborn library we generated a heatmap showing the popularity of genres through all the years. The numbers obtained for every generated chart, plot and graph were manually entered in the online Infographic creation tool, Visme, to generate easy-to-understand and vibrant looking infographic. Custom report style templates were generated using Visme.

3. Specification

Our infographic focuses on the history of gaming for the last 2 decades. We can see that there is rapid movement in the gaming sector. Increasing number of platforms have been developed as the years have progressed even though the number of games released every year have no relation with

the number of platforms. The popularity of genres has been fluctuating since the beginning. Only Action, Adventure games have stayed consistently popular since the beginning. Moving on to the platforms, Sony's Playstation 1, 2 and 3 have a combined market share of 29%, while its strongest console competitor Microsoft with its two consoles Xbox and Xbox 360 has a total market share of 18%. Gaming consoles have an average life of 7 or 8 years. Also, Nintendo is very close to Microsoft having a market share of 17% with Nintendo Wii and Nintendo DS. PC has been the single strongest competitor to Microsoft and Sony's gaming consoles with a dominant share of 24%. Mobile gaming has a relatively low credibility with a total market share of just 12%. Furthermore, Playstation 3 has had 9 masterpiece quality titles developed for it, which is by far the highest number, while Playstation 4, Xbox One and PC have only 5 each. In the last two decades gamers have shown exceptional preference for action games with them having a total market share of 27% followed by Sports and Shooters at 13.5% and 11.4% respectively.

4. Technologies

- Python – We have used Python to clean the acquired dataset. To clean and evaluate the data we have made use of Pandas, Scipy and Seaborn libraries.
- Visme – It is an online tool for creation of infographics. This allows us to create creative infographics without the need of photo editing software such as Adobe PS. Visme allows us to customize infographics according to our data and purpose, using various templates. These templates are designed for showing a wide-range of data including comparisons, timelines, processes, reports, etc.

5. Conclusion

Finding data with appropriate parameters and relations was a challenge. Once, data was finalized extracting information through it using Python was comparatively easy. Python with its handy libraries made it simple to form a story out of the extracted data. In contrary to that lack of a proper tool to generate infographics made it difficult to use the graphics I found to be more suitable as they required a premium subscription.

References:

1. Dataset: <https://www.kaggle.com/egrinstein/20-years-of-games/version/2>.
2. Tutorial: <https://www.kaggle.com/ash316/20-years-of-games-analysis>