

Link to dataset: <https://www.kaggle.com/gregorut/videogamesales>

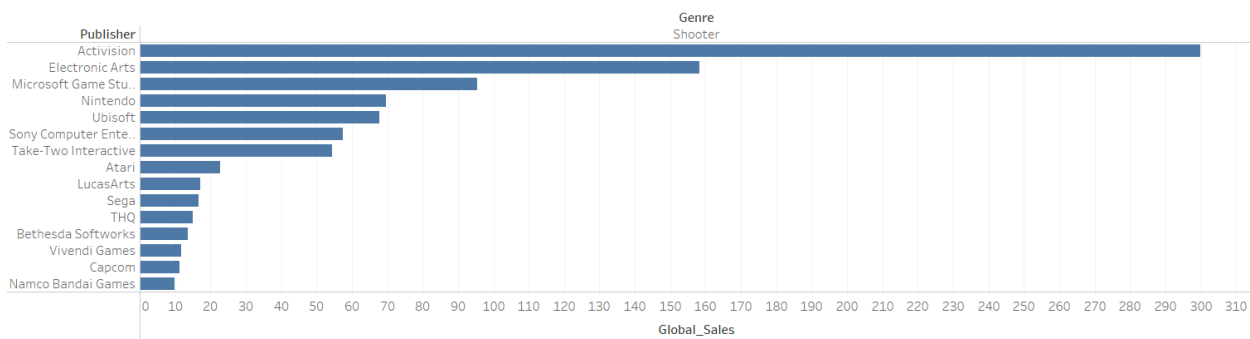
The linked dataset contains a list of video games that sold more than 100,000 copies, generated from a scrape of vgchartz.com. The fields (columns) of the dataset are 1) Rank – according to global sales, 2) Name – the title of the game, 3) Platform – PC, X360, DS, etc., 4) Year – year of game’s release, 5) Genre – genre of the game, 6) Publisher – company that published the game, 7) NA_Sales – millions of copies sold in North America, 8) EU_Sales – millions of copies sold in Europe, 9) JP_sales – millions of copies sold in Japan, 10) Other_Sales – millions of copies sold elsewhere in the world, and 11) Global_Sales – total worldwide sales; sum of previous 4 fields. The Rank column was deleted from the CSV along with the field name row.

Bullfrog Entertainment, a fledgling video game developer, is currently working on a new ‘shooter’ game. Bullfrog needs to partner with a publisher so that it can bring its game to market. Its managers are interested in looking at games that fit the shooter genre according to global sales (across all platforms) and the publisher associated with the game.

The first mapper filtered out all games that did not belong to the Shooter genre and created key-value pairings where the key is the name of the video game, and the value is the publisher and global sales. Because each platform edition is treated as a separate release, the global sales needed to be added up across all platforms to get each game’s total copies sold (in millions). This was the role of the first reducer. The second mapper-reducer pair turns total copies sold into the key and the new values are Name-Publisher pairings.

Black Ops (2010) from Activision sold the most copies (31.03 million) worldwide. Activision’s name also appears often in this list, which seems appropriate given it is one of the largest players in the industry. The final output is a bit messy because the Publisher name often appears multiple times for a game. However, this may be beneficial, as the number of times a Publisher’s name appears for a game indicates how many platforms the game was released on.

Millions of Copies Sold in Shooter Genre for Top 15 Publishers



Sum of Global_Sales for each Publisher broken down by Genre. The view is filtered on Genre and Publisher. The Genre filter keeps Shooter. The Publisher filter keeps 15 of 579 members.