Project description

Assume I work for the online store Ice, which sells video games all over the world. User and expert reviews, genres, platforms (e.g. Xbox or PlayStation), and historical data on game sales are available from open sources. We need to identify patterns that determine whether a game succeeds or not. This will allow us to spot potential big winners and plan advertising campaigns.

In front of us is data going back to 2016. Let's imagine that it's December 2016 and you're planning a campaign for 2017.

(The important thing is to get experience working with data. It doesn't really matter whether you're forecasting 2017 sales based on data from 2016 or 2017 sales based on data from 2016.)

The dataset contains the abbreviation ESRB. The Entertainment Software Rating Board evaluates a game's content and assigns an age rating such as Teen or Mature.

We have been provided with games datasets which contains the following fields:

—*Name*
—*Platform*
—*Year_of_Release*
—*Genre*
—*NA_sales* (North American sales in USD million)
—*EU_sales* (sales in Europe in USD million)
—*JP_sales* (sales in Japan in USD million)
—*Other_sales* (sales in other countries in USD million)
—*Critic_Score* (maximum of 100)
—*User_Score* (maximum of 10)
—*Rating* (ESRB)