

Video Game Sales

Background

You have been hired by Donkey Games, a small games developer in Dundee. They are planning their next release and want to use data to decide what kind of game to make. They have a dataset that contains historical information on the top selling games of all time.

They want to understand what types of games sell a lot of copies. In particular they are looking for analysis that helps them decide which direction to take their company in.

You'll need to think carefully about what information is going to be useful to the client.

Data

There's four datasets you can use. The first is a more up to date dataset with basic information on the top selling games is found in 'sales-2019.csv'. It contains the following variables:

- Rank - Ranking of overall sales
- Name - Name of the game
- Platform - Platform of the game (i.e. PC, PS4, XOne, etc.)
- Genre - Genre of the game
- ESRB Rating - ESRB Rating of the game
- Publisher - Publisher of the game
- Developer - Developer of the game
- Critic Score - Critic score of the game from 10
- User Score - Users score the game from 10
- Total Shipped - Total shipped copies of the game
- Global_Sales - Total worldwide sales (in millions)
- NA_Sales - Sales in North America (in millions)
- PAL_Sales - Sales in Europe (in millions)
- JP_Sales - Sales in Japan (in millions)
- Other_Sales - Sales in the rest of the world (in millions)
- Year - Year of release of the game

The second dataset is slightly out of date, but also has metacritic information. It is contained without 'sales-2016-with-ratings.csv'. It has the following extra variables.

- Critic_score - Aggregate score compiled by Metacritic staff
- Critic_count - The number of critics used in coming up with the Critic_score
- User_score - Score by Metacritic's subscribers
- User_count - Number of users who gave the user_score
- Developer - Party responsible for creating the game
- Rating - The ESRB ratings

You may want to combine the two datasets or use a different dataset for different parts of the analysis.

The last two datasets "xbox-one-game-sales.csv" and "ps4-game-sales.csv" contain information specific to those two consoles. The information may overlap with what is available in the other data sets. You have to decide how to use this data, if you use it at all.