CS 353
Spring 2024
Program 2

You are being provided a data set[1] showing global sales of rankings of video games. The format is:

- Index: Row number in data set (Numerical)
- Rank: The rank of the video game based on global sales volume. (Numerical)
- Game Title: The name of the video game. (String)
- Platform: The platform on which the game is available, such as PC, PS4, Xbox One, etc. (Categorical, string)
- Year: The year in which the game was released. (Date)
- Genre: The genre of the game, such as action, adventure, racing, etc. (Categorical)
- Publisher: The company that published the game. (String)
- North America: The number of units sold in North America, in millions. (Numerical)
- Europe: The number of units sold in Europe, in millions. (Numerical)
- Japan: The number of units sold in Japan, in millions. (Numerical)
- Rest of World: The number of units sold in the rest of the world, excluding North America, Europe, and Japan, in millions. (Numerical)
- Global: The total number of units sold worldwide, in millions. (Numerical)
- Review: The review score of the game, on a scale of 1 to 100. (Numerical)

You will write a program reporting various aspects of this data set based on input from the user. Provide a menu function to allow the user to enter up to 3 of the following criteria:

- Name – the name of the game. Spelling counts, but the comparison should be case-insensitive.
- Date – ask for beginning and ending dates of a range. If they are the same, then only games released that year are considered. If the ending date is before the starting date (for example, if the user enters 2006 – 2001 for date range), then reverse them.
- Publisher – This should also be case insensitive, and should accept a partial match. For example, if the user enters “microsoft,” then a game published by Microsoft Game Studio should match. Again, it is not necessary to allow for typical misspellings, or to allow the user to refer to Microsoft by its stock symbol MSFT.
    - Yes, this means that if the user enters “interactive”, then Take Two Interactive and Interactive Games will both match. Your program isn’t responsible for reading the user’s mind.
- Region – North America, Europe, Japan, Rest of World, or Global.
- Genre – as per the categories in the data set.

There are obviously more features we may want to select on, but this will be enough for this assignment. Your program should ask the user to enter up to 3 search criteria, then filter the data set to list the results. Give the user the option whether to display results ordered by sales (rank) or rating (review). Allow the user to repeatedly enter queries until they say they want to quit.

A note on the data set: The original data set on Kaggle had missing years for some games. I filled in the missing dates as best I could. In some cases, it was easy – “Madden NFL 2007” was probably published in 2007, for example—but in other cases it was harder to tell. In those cases, I simply made up a year to avoid the missing data problem. While dealing with this, I noted that there are at least 3

1 Data source: https://www.kaggle.com/datasets/thedevastator/global-video-game-sales-and-reviews

separate entries for Rock Band, all of them missing the year. Any incorrect year information is my responsibility, but isn't relevant for this assignment. As far as I know, there's no other missing data.

Programming notes: Obviously, you're going to be using the filter function heavily in this program. Look for opportunities to use higher-level functions, passing them a lambda function to customize them as needed.

Submit your code or GitHub / GitLab / BitBucket link by the deadline.