Assignment 5: DC vs. Marvel

30 Points scaled to 20 Points

Introduction

This assignment makes use of a dataset that is made available on Kaggle (https://www.kaggle.com/datasets/jcraggy/marvel-vs-dc-imdb-rotten-tomatoes), which consists of data relating to DC and Marvel movies. In this assignment, you will only use movies that were released during or after the year 2000. So, you will need to filter the dataset prior to answering the questions. Your goal is to compare the DC and Marvel franchises by answering the 15 posed questions. Note that the questions assist you in determining what fields to use. The "entity" field differentiates the two groups: "MARVEL" and "DC".

Objectives

- *Query and filter datasets*
- *Obtain summary statistics*
- *Compare groups*

Deliverables

- *Jupyter Notebook (Python) or R Markdown file (R) with all code and answers to the questions embedded. Files can be rendered to HTML webpages if your instructor requires this. Questions should be stated and be answered within Markdown cells.*

Questions

This assignment can be conducted using either Python or R, whichever you prefer or whichever you instructor requires. Generate code to perform the following tasks or answer the questions posed.

*Only use movies released during or after the year 2000 in this assignment.

Question 1. How many of the movies in the original table were released during or after 2000 ("year")? (2 Points)

Question 2. How many movies did DC release during or after 2000? How many movies did Marvel release during or after 2000? (2 Points)

Question 3. Does DC or Marvel have the highest average rating on IMDB ("imdb_rating")? (2 Points)

Question 4. Was the lowest rated movie on IMDB ("imdb_rating") a DC or a Marvel movie? (2 Points)

Question 5. Was the highest rated movie on IMDB ("imdb_rating") a DC or a Marvel movie? (2 Points)

Question 6. Does DC or Marvel have the highest average critic's rating on Rotten Tomatoes ("tomato_meter")? (2 Points)

Question 7. Does DC or Marvel have the highest average audience rating on Rotten Tomatoes ("tom_aud_score")? (2 Points)

Question 8. Which movie had the largest difference between the critic's score ("tomato_meter") and audience score ("tom_aud_score") on Rotten Tomatoes in which the audience score was higher than the critic's score? (2 Points)

Question 9. Which movie had the largest difference between the critic's score ("tomato_meter") and audience score ("tom_aud_score") on Rotten Tomatoes in which the audience score was lower than the critic's score? (2 Points)

Question 10. From the movies included in the database and those released during or after 2000, did DC or Marvel have the highest total gross earnings ("imdb_gross")? (2 Points)

Question 11. From the movies included in the database and those released during or after 2000, did DC or Marvel have the highest earning movie and what was the movie ("imdb_gross")? (2 Points)

Question 12. Does DC or Marvel movies have a longer average runtime ("runtime")? (2 Points)

Question 13. Does DC or Marvel have the larger total runtime for all movies released during or after 2000 ("runtime")? (2 Points)

Question 14. Calculate the gross earnings per runtime for DC and Marvel (total gross earnings ("imdb_gross") divided by total runtime ("runtime")). Does DC or Marvel have the highest gross earnings per unit runtime? (2 Points)

Question 15. Does Marvel of DC have a higher percentage of movies with a PG-13 rating vs. all other ratings ("mpa_rating")? (2 Points)