BEFORE YOU START

Make a Markdown cell with the following statement:

“<Your name> did not use ChatGPT or any other AI/LLM to assist me on this assessment”

1) Create a new dataframe named "movies\_df" consisting of only movies.

2) Create a new dataframe named "movie\_ratings" from movies\_df that consists of

the top5 ratings by count. This dataframe should consist of only two columns,

rating and count.

3) Use the movie\_ratings dataframe to create a bar chart showing count of the

top 5 ratings.

4) Using the movies\_df, subset the data to create a new dataframe named

"year\_duration" that only consists of 2 columns, release\_year and duration.

5) Find min, max, and average duration of movies from the 1975-1984.

6) Create a scatterplot using year\_duration that looks at movie durations

over the years. Do you notice any correlation?

7) Find the actual correlation value, is there any correlation?

8) Use movie\_df to find which year had the most G-rated releases. Can you do

this in a single line of code?

9) Which year had the most movie releases? What percentage of movies that year

had durations that were statistical outliers for that year?

10) Create a new column in movies\_df that catagorizes each movie as either

short, average, or long. Short movies should be 30 mins or less, long movies

should be over 2 hours.

11) Make a pie chart of the above movie lengths.

12) Which movie had the greatest gap from the year it was released to the year

it was added? What is the average?

13) How many movies were added during the first year after the pandemic

lockdown? (Assume the lockdown began on March 16, 2020)

14) Was any movie added on your birthday last year?

15) How many TV shows or Movies have the word 'dog' in the title regardless if

it is capitalized?