# Pandas Practice Tasks for messy_movies_500.csv

This workbook contains 50 Pandas tasks you can perform on the provided messy_movies_500.csv dataset. Use these exercises to practice data cleaning, exploration, and analysis.

## A. Basic Exploration (10 tasks)

1. Load the dataset and display the first 5 rows.
2. Display the shape (rows, columns) of the dataset.
3. List all column names.
4. Display the data types of each column.
5. Show the number of missing values in each column.
6. Show basic statistics for numerical columns.
7. Display unique genres.
8. Count the number of unique movie titles.
9. Show the top 10 directors by the number of movies.
10. Check if there are duplicate rows.

## B. Data Cleaning (15 tasks)

11. Strip extra spaces from title and director columns.
12. Convert title to title case.
13. Convert director to proper case.
14. Replace ',' in genres with '|'.
15. Remove extra spaces in genres.
16. Fill missing runtime with the mean runtime.
17. Fill missing imdb_rating with the median rating.
18. Fill missing vote_count with 0.
19. Drop rows where title is missing.
20. Convert release_year to integer type.
21. Convert runtime, imdb_rating, and vote_count to numeric types.
22. Remove duplicate rows.
23. Standardize genres so each genre is separated by '|' and no extra spaces.
24. Remove leading/trailing spaces in all string columns.
25. Rename all columns to lowercase with underscores.

## C. Filtering & Selection (10 tasks)

26. Select all movies released after 2010.
27. Get all movies with an IMDb rating greater than 8.5.
28. Get all movies directed by Christopher Nolan.
29. Get all Drama movies.
30. Get movies with runtime less than 100 minutes.

31. Get movies with vote_count greater than 1 million.
32. Get movies released between 1990 and 2000.
33. Get movies where genre contains both 'Action' and 'Adventure'.
34. Select only the title and imdb_rating columns.
35. Select the top 5 longest movies.

## D. Grouping & Aggregation (7 tasks)

36. Find the average IMDb rating per director.
37. Find the average runtime per release year.
38. Find the total vote count per genre (split genres first).
39. Find the highest-rated movie per director.
40. Count the number of movies per genre.
41. Find the release year with the most movies.
42. Find the director with the highest average IMDb rating.

## E. Sorting & Ranking (3 tasks)

43. Sort movies by IMDb rating (descending).
44. Sort movies by vote_count and then by imdb_rating.
45. Rank movies based on IMDb rating.

## F. Advanced Tasks (5 tasks)

46. Create a new column decade from release_year.
47. Split genres into separate rows (one genre per row).
48. Create a histogram of IMDb ratings.
49. Create a bar plot of the number of movies per director.
50. Find the correlation between runtime and IMDb rating.