Coding Exercises – Set 2

There is one dataset that has been provided

- imdb.csv

Questions 1 – 8: Utilize the imdb data set

Data Description: The imdb data contains the rating and other information related to movies and episodes across a lot of genres and years.

*HINT: The file is correct and can be loaded without any issues. The error that you might receive when loading the file is intentional. "escape chars" is the key term to solve the issue.

1. Get the imdb rating for fifth oldest movie in the data set.
2. Return titles of movies with shortest and longest run time.
3. Sort the data frame by the order of released date and ratings, Hint: release_date (earliest) and Imdb rating (highest to lowest)
4. What is the name of earliest video.episode available in the data set? Which year was it released?
5. Generate a report that tracks the various Genre combinations for each type year on year. The result data frame should contain type, genre_combo, year, avg_rating, min_rating, max_rating, total_run_time_mins.
6. Is there a relation between the length of a movie title and the ratings? Generate a report that captures the trend of the number of letters in movie titles over years. Create a cross tab between the year of the release and the quantile of the movie title length. The results should contain year, min_length, max_length, num_videos_less_than25Percentile, num_videos_25_50Percentile, num_videos_50_75Percentile, num_videos_greaterthan75Precentile.
7. Generate a report that tracks the average imdb rating quarter on quarter, in the last 10 years, for movies that are top performing. You can take the top 10% grossing movies every quarter. Add the number of top performing movies under each genre in the report as well.
8. Bucket the movies into deciles using the duration. Generate the report that tracks various features like nominations, wins, count, top 3 genres in each decile.