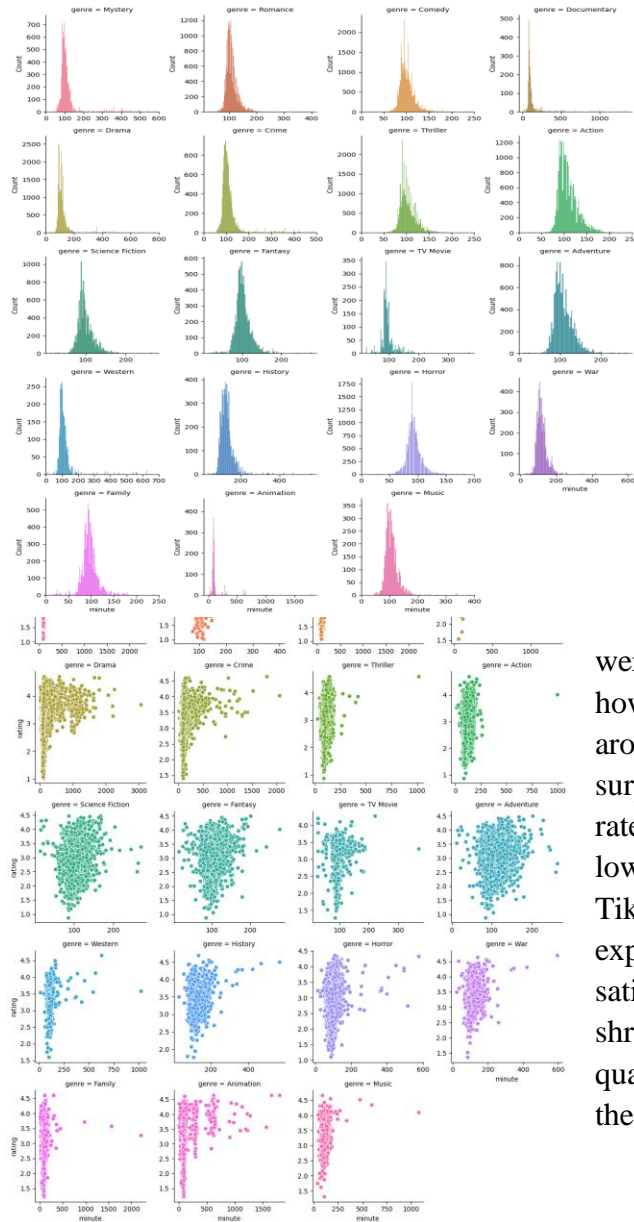


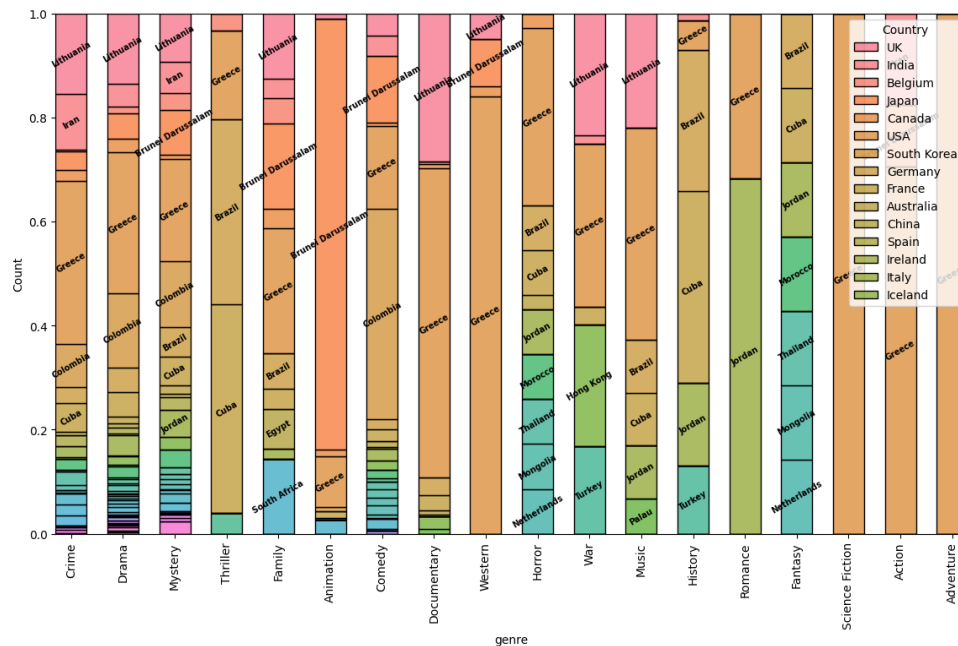
How do movie duration and overall rating differ across country and genre?

The data I analyzed included almost a million observations of films, with information including but not limited to film duration, rating, genre, country and language. While the dataset presented me with a lot of information, I decided to narrow down my focus to analyze whether certain film genres were associated with longer duration and how duration and overall viewer satisfaction were correlated. Since the data was comprised of multiple data frames, each with identifiers that referenced the other data frames, I decided to use DuckDB to create, join, and aggregate the data to answer the questions I formulated.



To begin, I performed some preliminary data analysis to investigate how long each genre of movie was on average. I initially hypothesized that thriller and action films would be among the longest. Contrary to my beliefs, documentaries, action, and drama films were the longest on average. While length varied depending on each film's theme, each distribution was approximately normal, meaning a majority of the films' lengths were concentrated around the average.

After extracting these findings, I decided to take my analysis a step further and examine how movie length and overall rating were correlated. The relationships looked mostly how I predicted, with a lot of variation in ratings around the average film length. However, what surprised me was how longer movies were strictly rated higher across the board, with little to no lower ratings. With the influx of social media like Tik Tok, Instagram reels, and YouTube shorts, I expected longer films to be met with less satisfaction due to people's attention spans shrinking. This led me to further investigate other qualities present in the data that could influence the lack of information available for longer films.



When looking at the available data, the information that stood out to me regarding why longer movies were predominantly rated higher was the country and language of a film. I inferred that information about the origin of a film could explain a large amount of

variation between duration and viewer satisfaction. I merged the countries data frame to the data frame I already had to conduct my analysis. I then visualized the breakdown of country by genre to see which origins are most common for each genre, however I restricted the data to films that were greater than 250 minutes. I found a surprising amount of variety between which countries were more prominent for each genre of film. Based on the given information, I found the average film rating for each relevant country.

	country	av_ratings
0	Jordan	3.728659
1	Brunei Darussalam	3.690000
2	Greece	3.530651
3	Cuba	3.400313
4	Brazil	3.356052
5	Colombia	3.290227
6	Hong Kong	3.224317
7	Lithuania	3.011969

As depicted to the left, each country that tended to have longer film durations for at least one genre had a high rating of at least 3 stars out of 5. This information explains why longer films often observe higher ratings.

Overall, I discovered fascinating patterns of factors that influence viewer satisfaction in films. I realized that contrary to my beliefs, long films can be well-liked as long as the storyline and visuals are well thought out. If I were to redo this analysis, I would delve deeper into other characteristics of

the films, in particular age rating and release type. Genre and country provided me with insightful findings, however the intersection of more variables can provide even more insightful conclusions that explain viewership patterns and help film companies increase their revenue in the long run.