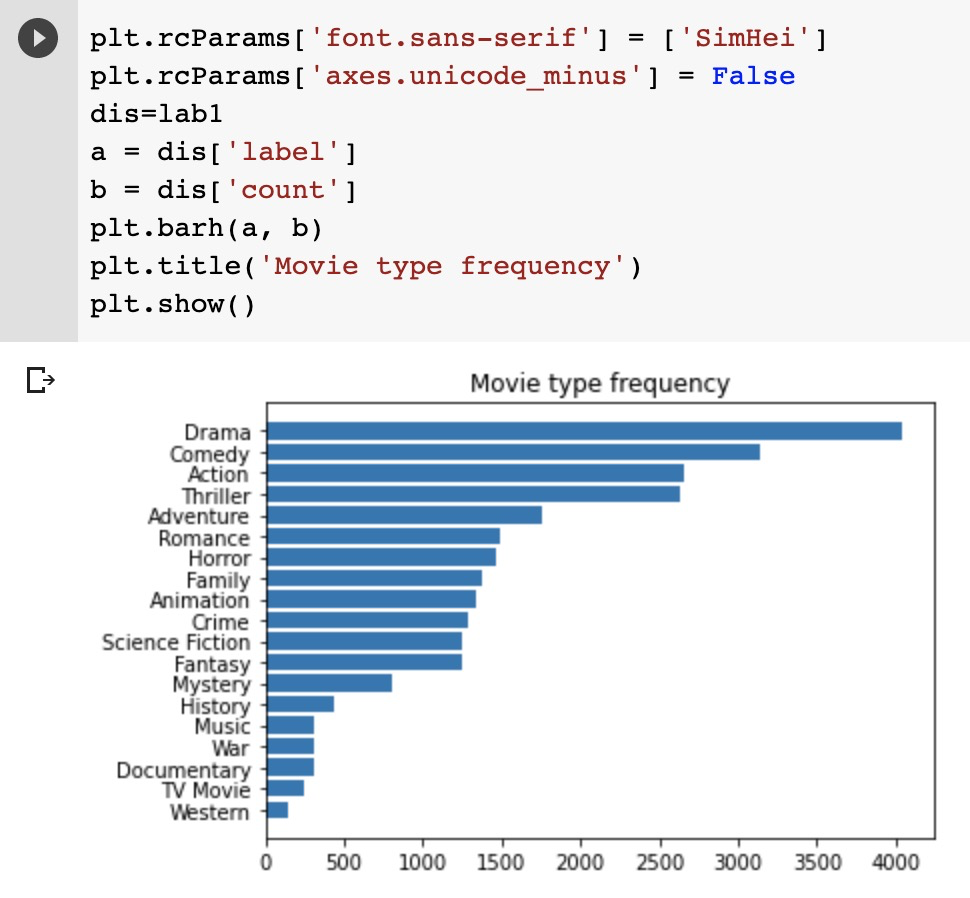
**Movie Recommendation System - Genre**

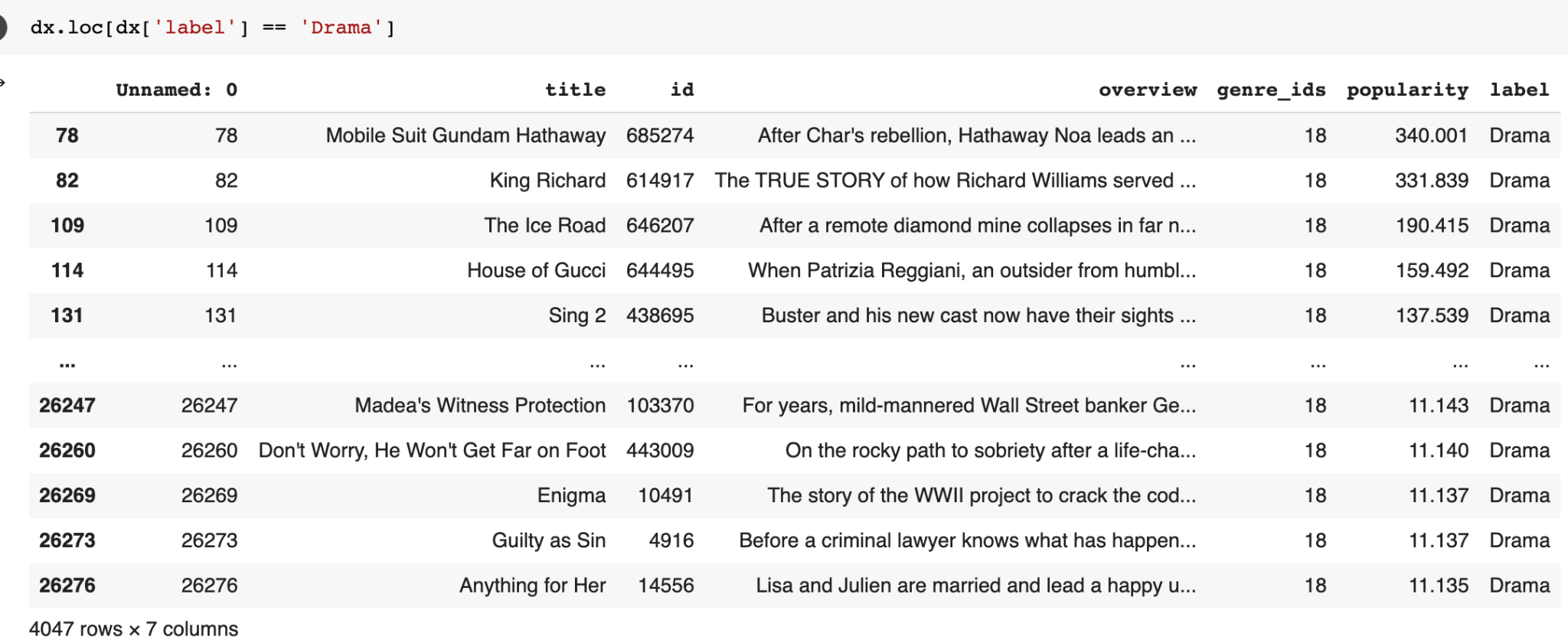
If a movie has multiple genres, then each genre will be calculated once. The following figure sorts and counts the number of occurrences of the genres. Drama has the highest frequency, and it can be known that many movies contain the element of drama.

In AMDb, after the user follows a specific movie genre on the setting page, AMDb will recommend the 20 most popular films of the label for the user on the profile page.



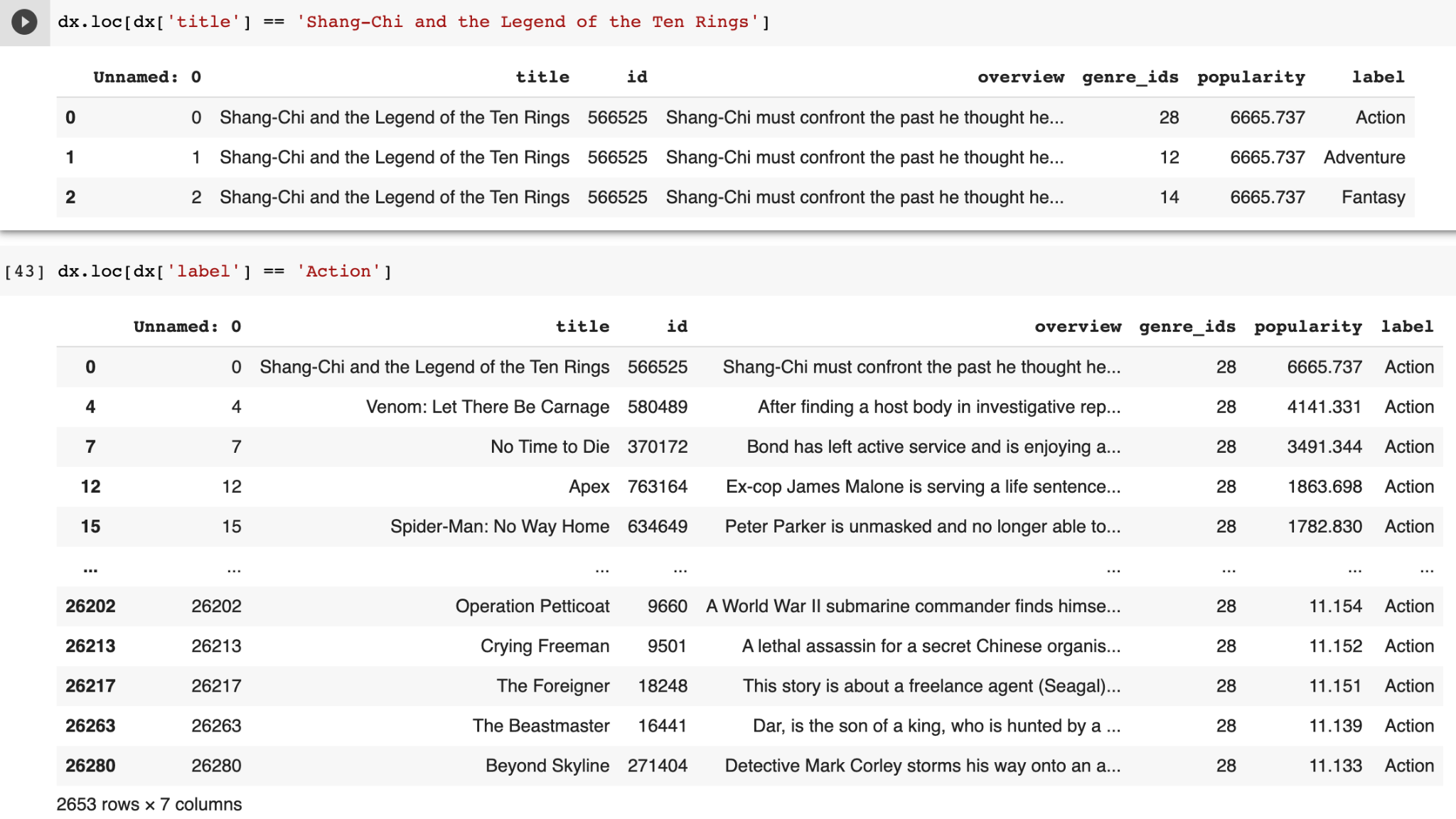
Movie type frequency chart

When the user follow ‘Drama’ lable on the setting page, the number one recommended movie would be ‘Mobile Suit Gundam Hathaway’ which popularity is 340.001.

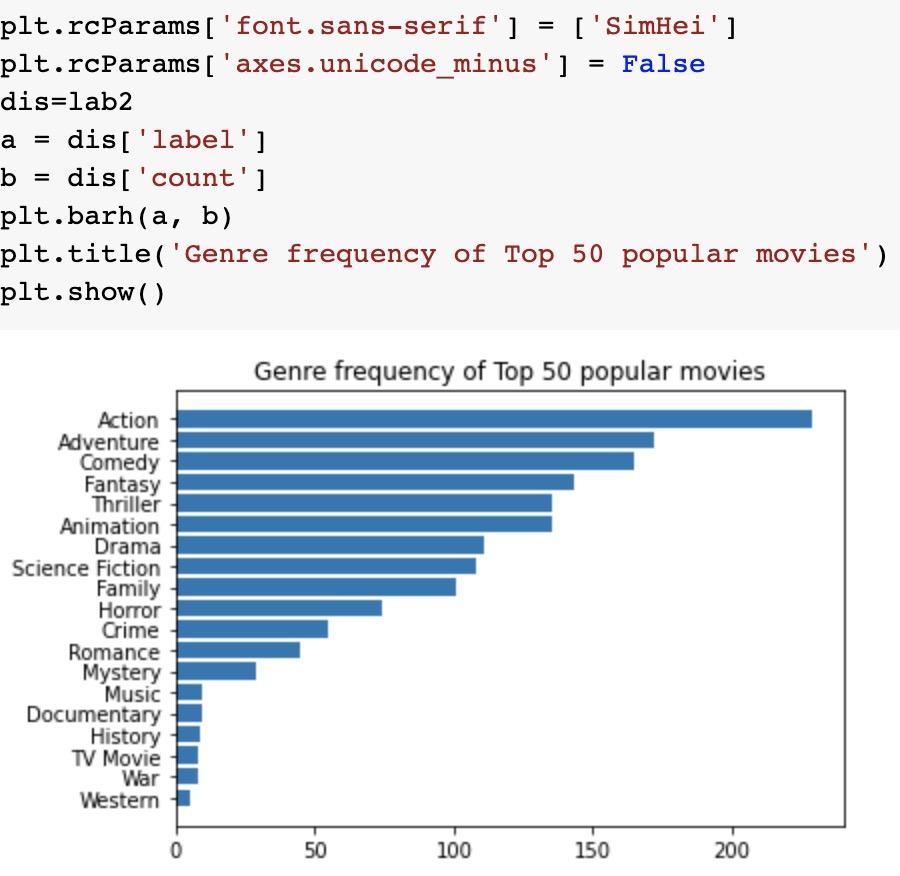


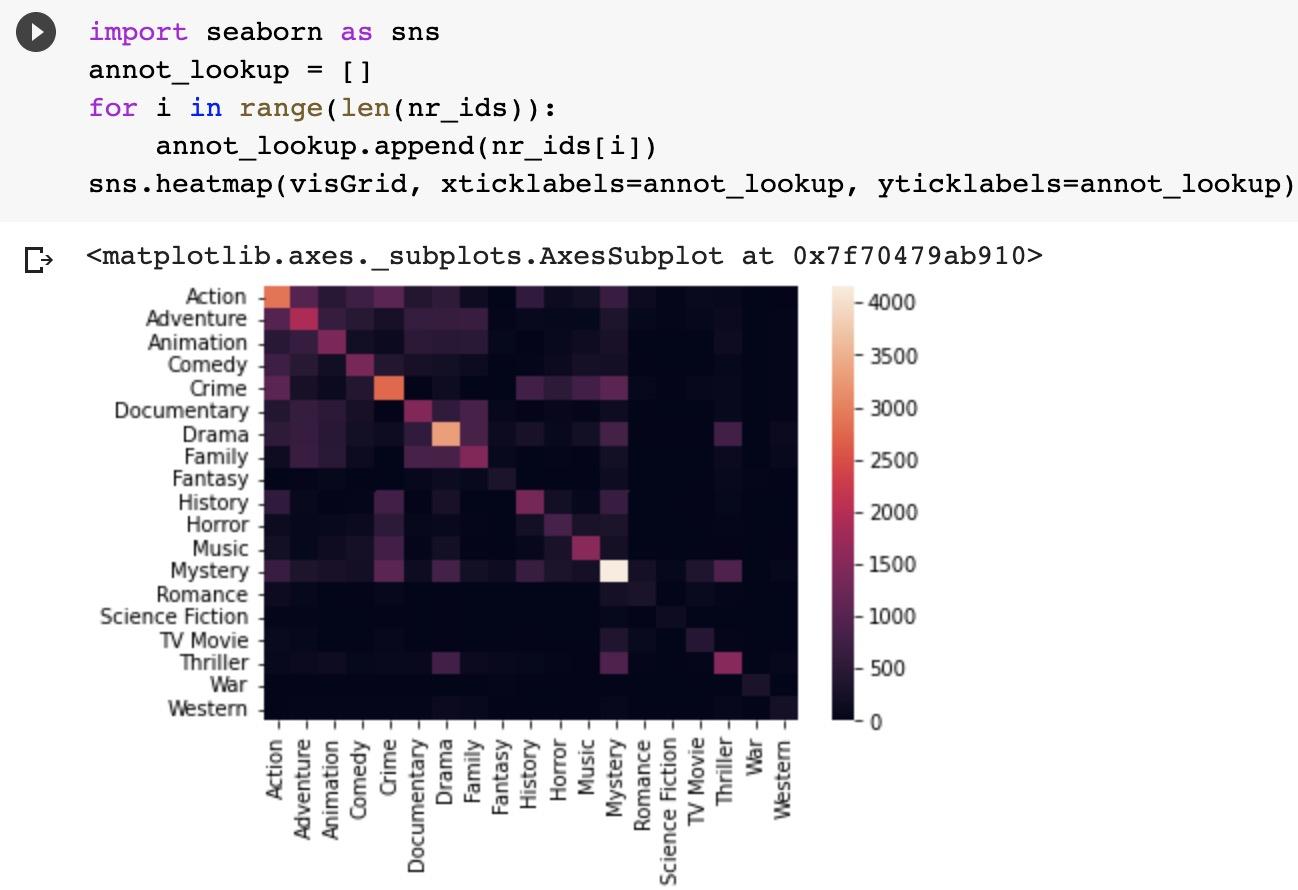
However, when we searched the recent most popular movie, "Shang-Chi and The Legend of The Ten Rings, " its genre included Action, Adventure, and Fantasy. Setting genre is action as the search condition, the popularity of number one movie is as high as 6665.737. The difference in popularity between the No.1 Action movie and No.1 Drama movie is about 20 times.

Because the genre of a movie usually includes at least two or more labels, we are here to consider and analyze two genres of combinations that often appear simultaneously. Next, we consider using the form of heat map to show the distribution of types that often appear at the same time.



Our goal is to recommend the most popular movies to users according the genre label, so we chose to intercept the top 50 most popular movies for analysis. Then make genre statistics for these 50 movies, the statistical results are shown in histogram 2. At this time, it can be seen that the slope between the labels has slowed down, which is helpful for find out popular genres combination.





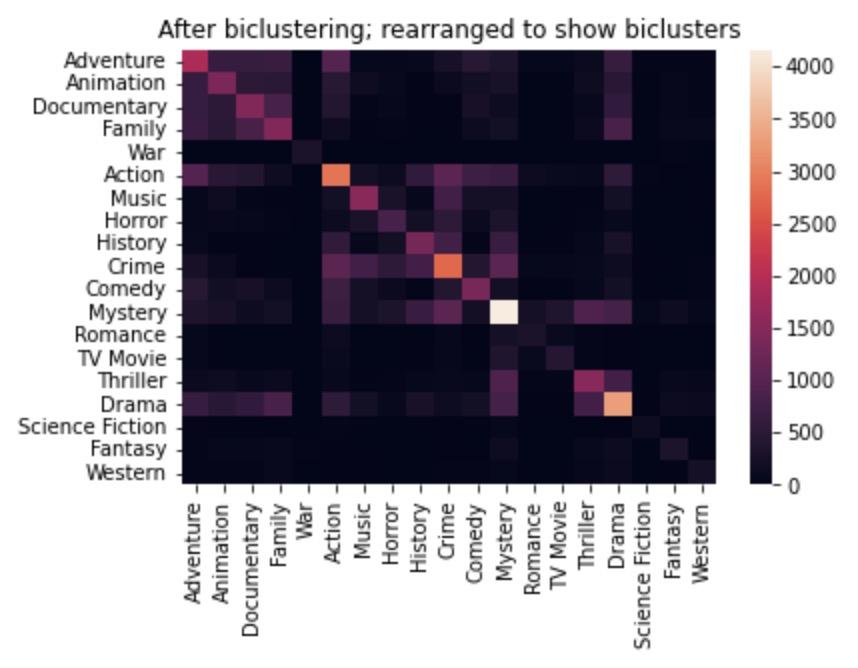
This heat map figure shows the distribution of genres that often appear simultaneously. Among them, it can be found that the diagonal in this heat map is very obvious. The diagonal corresponds to the number of occurrences, such as the only white spot in the whole picture. Its horizontal axis and vertical axis respectively correspond to Mystery. It can be considered that the number of times Mystery and Mystery appear together is most. Then, it can be found that much genre data is concentrated on box area constituteb by Documentary, Drama and Family; and also concentrated on the box area formed by the combination of Action and Adventure.

Because the upper right corner and the lower-left corner are relatively uniform black, it can be inferred that Romance, Science Fiction, TV Movie, Thriller, War and Western, ​​the probability of them appearing in the same movie is relatively low.

At the same time, the two genres of Science Fiction and Western are not widely distributed, because it can be seen from the diagram that their corresponding horizontal and vertical axes are all same colour - black.

The colour blocks in the first heat map above, most of them are distributed in square areas. In other words, some specific genres always appear together simultaneously but rarely interact with other different genres. In a sense, this will highlight and distinguish which genres of movies are more likely to occur together and which are not.

In order to explore in-depth the genres that appear at the same time, bi-clustering analysis is used to analyze this unbalanced data, and then a set of movie genres be obtained, as shown below.



A prominent square can be found in the middle of this heat map; even if there are some black colour blocks in this square, we do not think this will significantly hinder the recommendation result. Since we believe this can bring novelty to the movie recommendation system.

Based on the above analysis, we can classify the data into 3 categories: category A, including Action, Music, Horror, History, Crime, Comedy and Mystery; category B, including Adventure, Animation, Documentary and Family; category C, including Action, Adventure, Comedy and Fantasy.

As long as the user follows one of the genre labels in category A or B on the setting page, in addition to the label selected by the user, AMDb will randomly select any other label in the same category. Recommend movies to users based on these two tags which ensure the novelty and randomness of every time movie recommendation.

For example, when a user selects the ‘Adventure’ label to follow, AMDb may recommend the most popular film among the Adventure and Family at the same time, instead of just recommending movies with an adventure label. When users follow War label, because the movies in this genre are relatively less, the system may recommend movies with label War or Action to users in the meanwhile.

Here it can found that the Mystery genre does not even appear in the top five in the cumulative number of appearances. Then we searched all the movies with the Mystery label in the original data set. It can be found that the popularity of movies with this type of label is not high.

1. Drama
2. Action
3. Science Fiction
4. Excitement (Thriller, Crime, Mystery),
5. Uplifting (Adventure, Fantasy, Animation, Comedy, Romance, Family)
6. Horror
7. History