PySpark Dataframe assignment

Note:

- For each question, show 100 rows of the results without truncated.

- Some movies do not have information about year. You can skip these movies.

- A movie might have multiple genres, make sure to separate the genres of each movie so that your calculations are done properly.

Load data from **movies_small.csv** and **ratings_small.csv** to dataframes and perform the following tasks:

1. Show the number of movies made in each year. The results are sorted by year.

Year	Num_of_movies

2. Show the number of movies belonging to each genre made in each year. The results are sorted by year. The movie count of each genre is displayed in each column.

Year	Action	Animation	Comedy	…

3. For each userID, show the average rating of that user for each genre. The results are sorted by userID. The result of each genre is displayed in each column.

User_id	Action	Animation	Comedy	…

4. For each movie, show the name, the year, the number of ratings, and the average rating (from all users) of each movie. The results are sorted by the years and then by the names of the movies.

Year	Movie_name	Num_rating	Average_rating

5. For each user ID, show the genre that received highest average rating from that user and the list of top 5 movies belonging to that genre that receive highest average rating from all user and haven't been rated by that user (For example, 'Action' is the genre that received highest average rating from user ID X. Among 'Action' movies that hasn't been rated by user ID X, you are supposed to show top 5 movies that received highest average rating from all users). The results are sorted by the user ID.

User_id	Highest_rated_genre_name	Top_5_unrated_movies_with_highest_rating

6. For each user ID, show the two genres that received highest rating from that user, and the list of top 5 highest rated movies that have both genres and hasn't been rated by that user. The results are sorted by the user ID.

User_id	Two_highest_rated_genre_names	Top_5_unrated_movies_with_highest_rating

7. Show the years of the first_appearance of each genre.

Genre	First_appearance_year

8. For each user ID, show the list of top 5 movies made after 2000 that received highest rating from that user. The results are sorted by user ID.

User_ID	Top_5_movie_after_2000_rated_by_this_user