PySpark Dataframe assignment

Rules:

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

- Teams with 4 members are required to complete all **8** questions in the assignment. Teams with 3 or fewer members are required to complete questions from **1 to 6** in the assignment.

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.

2. Show the number of movies belonging to each genre made in each year. The results are sorted by year. The results of each genre are supposed to be displayed in a column.

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

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

5. For each user ID, show the genre that received highest rating from that user and the list of top 5 rated movies belong to that genre that hasn't been rated by that user. The results are shorted by the user ID.

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

----

7. Show the years of the first_appearance of 'Fantasy' and 'Sci-Fi' movies

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