Coding challenge

Expectations

- **State your assumptions** - Anywhere you feel that the requirements are unclear please make an assumption and document that assumption.
- **Describe Trade-offs** - When you're making a decision about using one design/approach vs. another try to make a quick note about why you made the choice you did.
- **Provide tests** - You should provide unit tests for the code that you write. The choice of testing tools is up to you.
- Consider resource constraints - Focus on a solution that can be run on a “regular pc” with up to 16GB of memory.

We prefer solutions written in Scala or Python, however you can use other programming language, that you feel most comfortable with, to implement the solution. Please submit your solution as a **.zip** file. We prefer a simple, yet well tested and well explained solution that could be considered "production-ready".

Challenge

For this challenge you’ll be working with the IMDB movie dataset. You can check the provided link for the description of the dataset. The link to the actual files should forward you here.

Implement a solution that answers the following questions:

- How many movies are made each year?
    - Show the distribution over the past 100 years
- Which actor-director pairs have the most collaborations?
    - Show the top 10
- Find the genre(s) that an actor has mostly worked in, i.e. the number of movies and tv shows per genre. Note that the *knownForTitles* column in the *name.basic.tsv* file **does not** contain the whole filmography of an actor.
    - Show this for Omar Sy, Frances McDormand, and Saoirse Ronan

Your code should be able to parse the relevant files and compute the results to the questions. However, you don’t have to include the files in your submission. We will download the same files to check the solution.