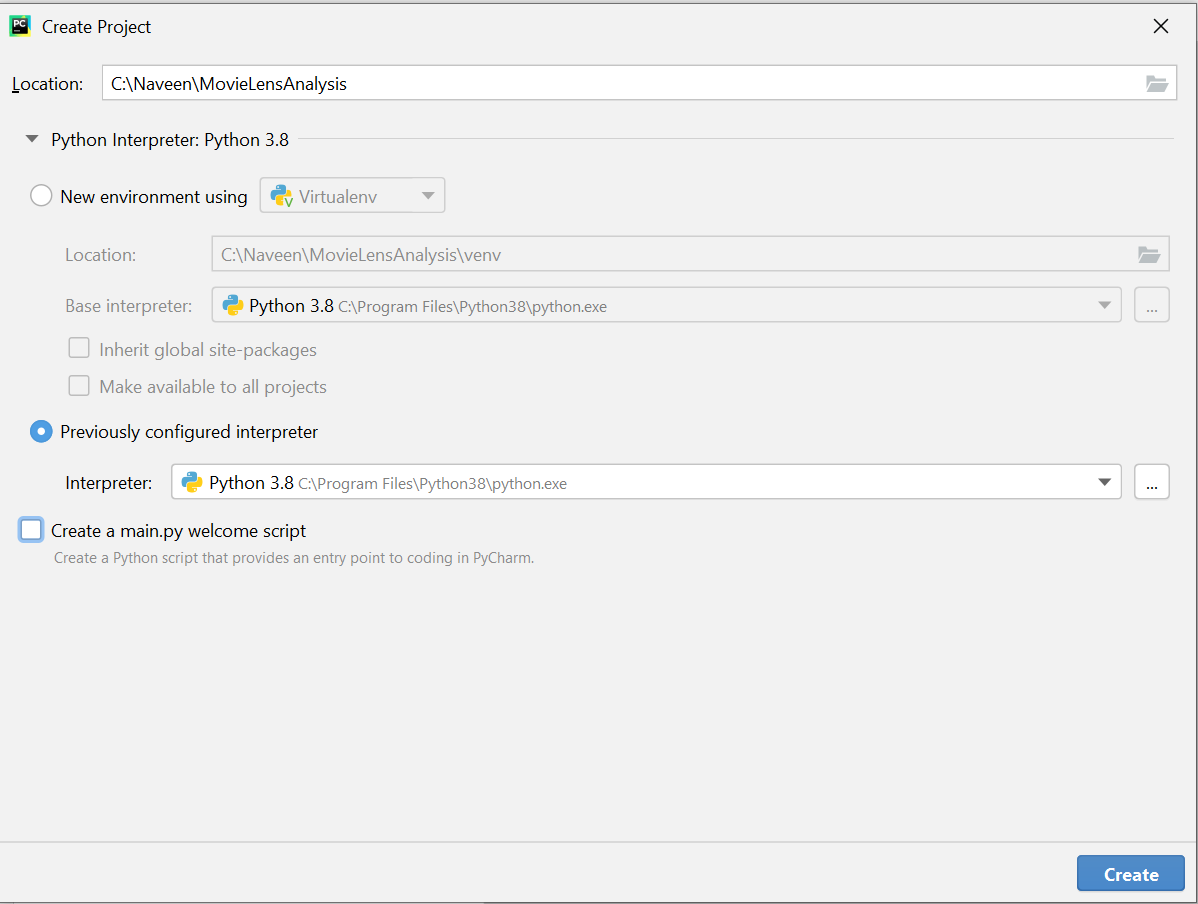
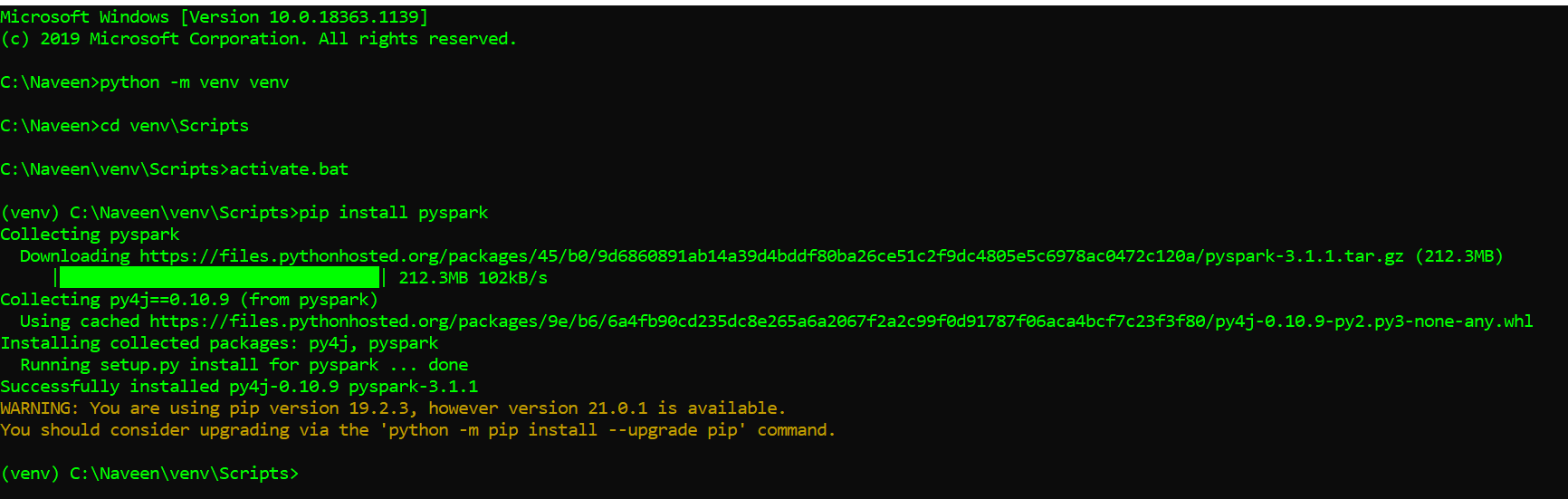
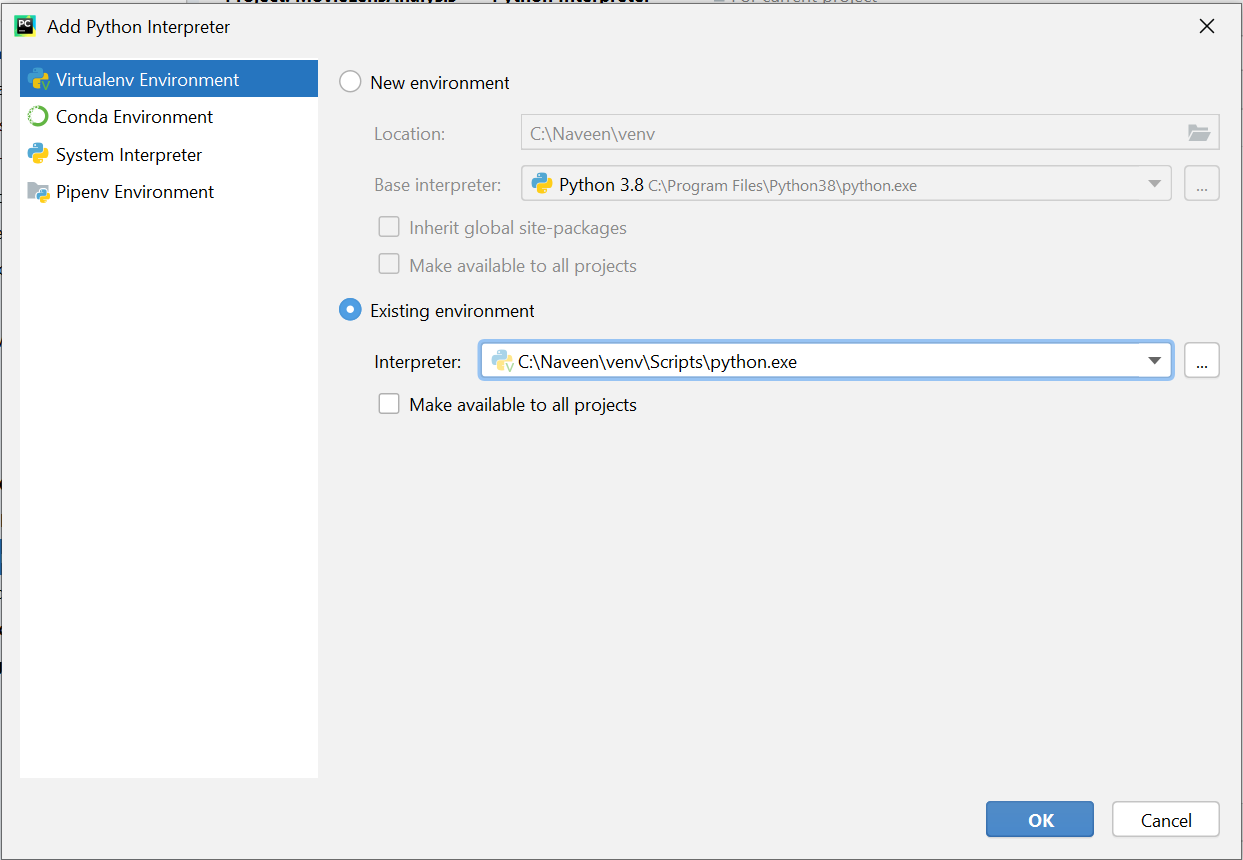
Create a new python project



Create python virtual environment, activate the virtual environment and install pyspark.



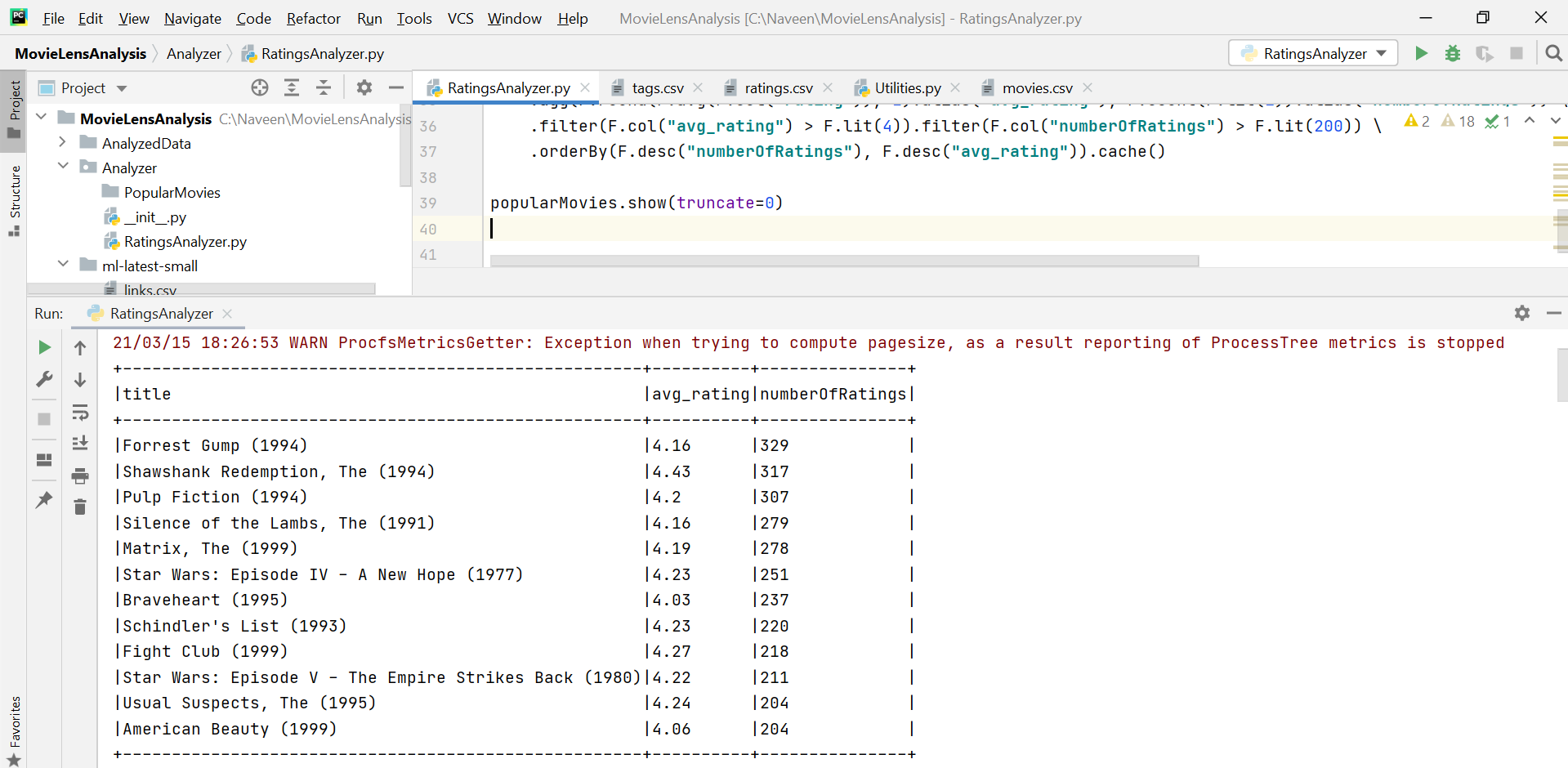
Set the virtual environment to the project



# Analysis on the Movielens Data:

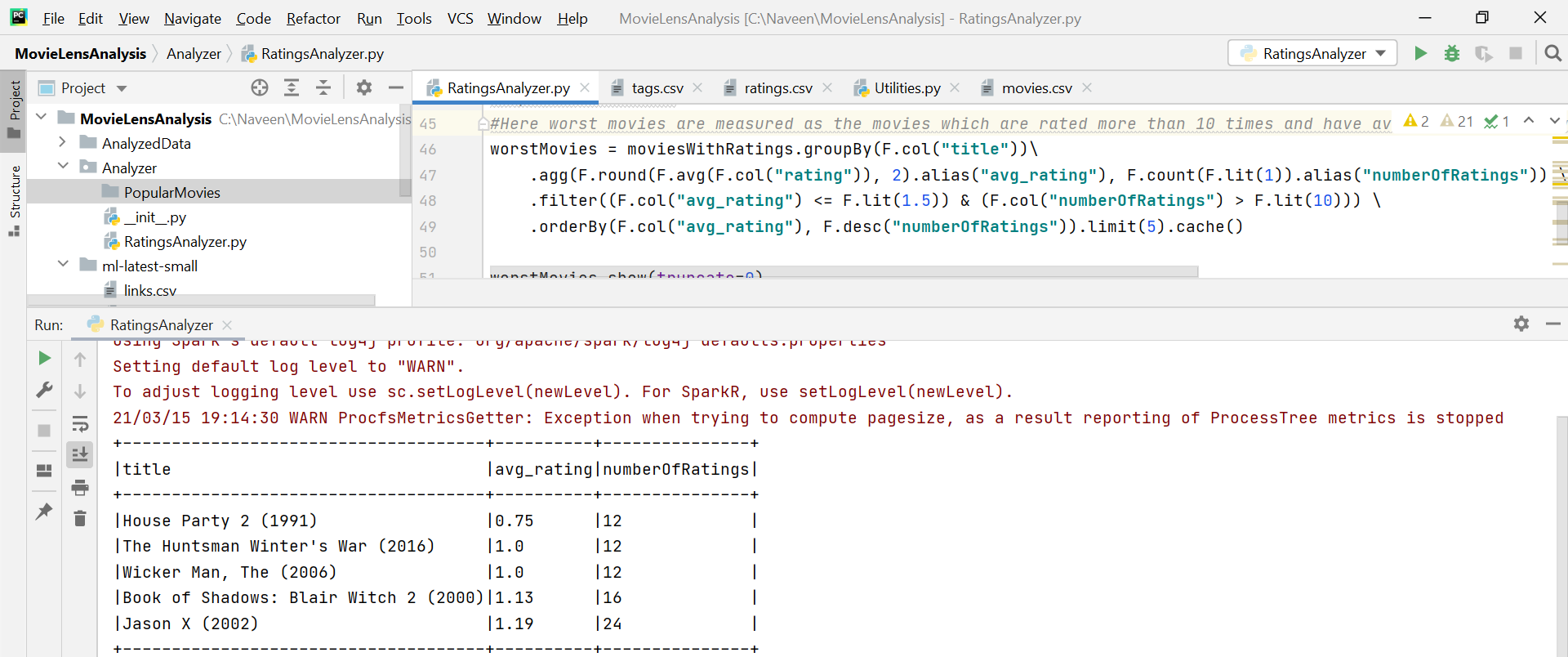
### Most popular movies in the data set:

The movies which are rated more than 200 times and have an average rating above 4 are considered as the popular movies



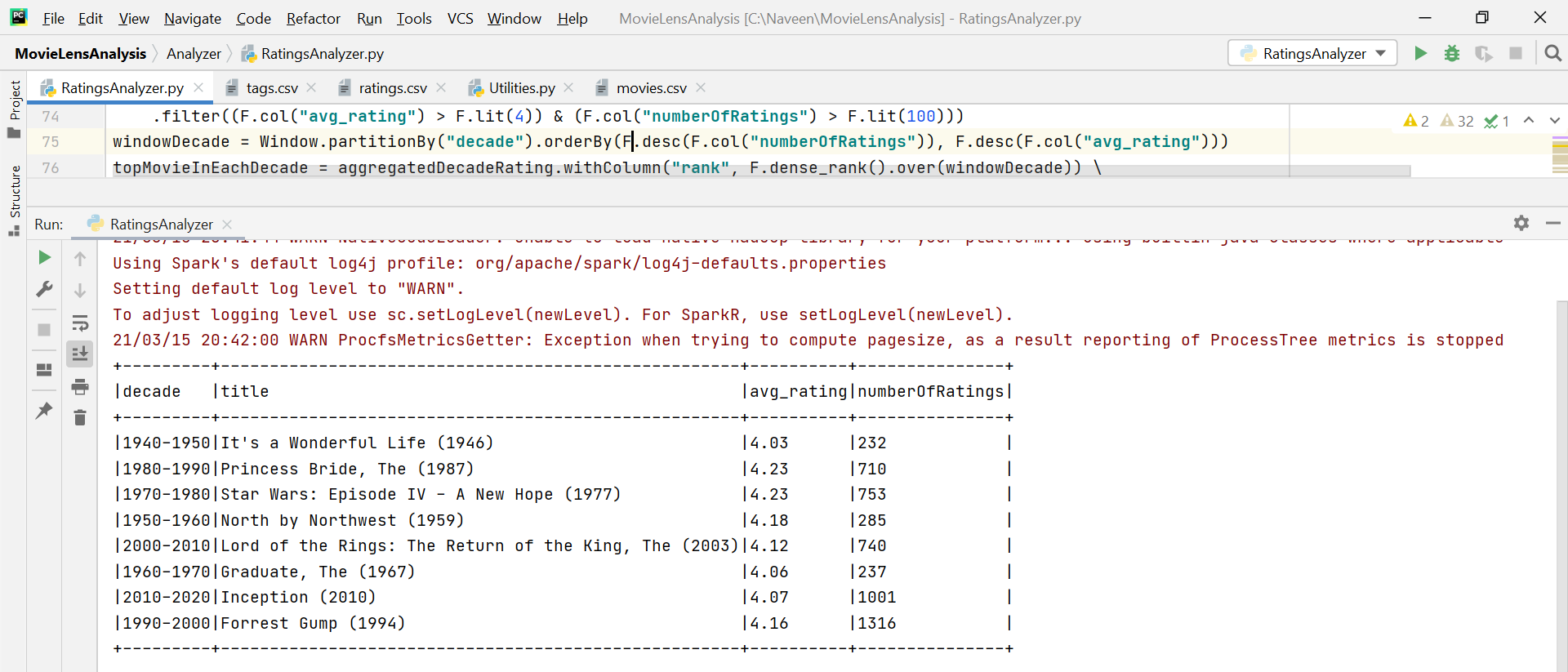
### Top 5 worst movies in the dataset

The worst movies are measured as the movies which are rated by more than 10 users and having rating less than 1.5

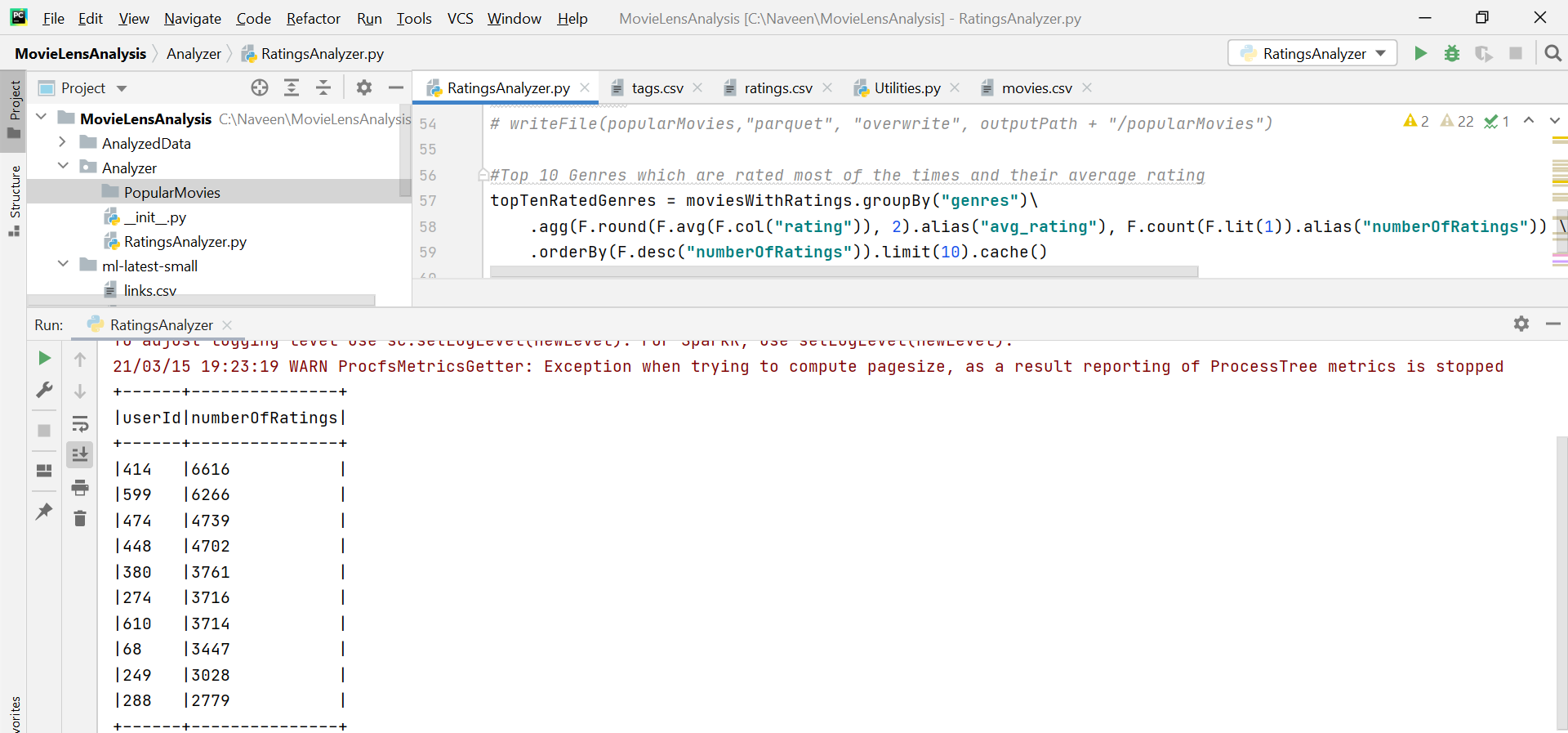


### Movie of the Decade

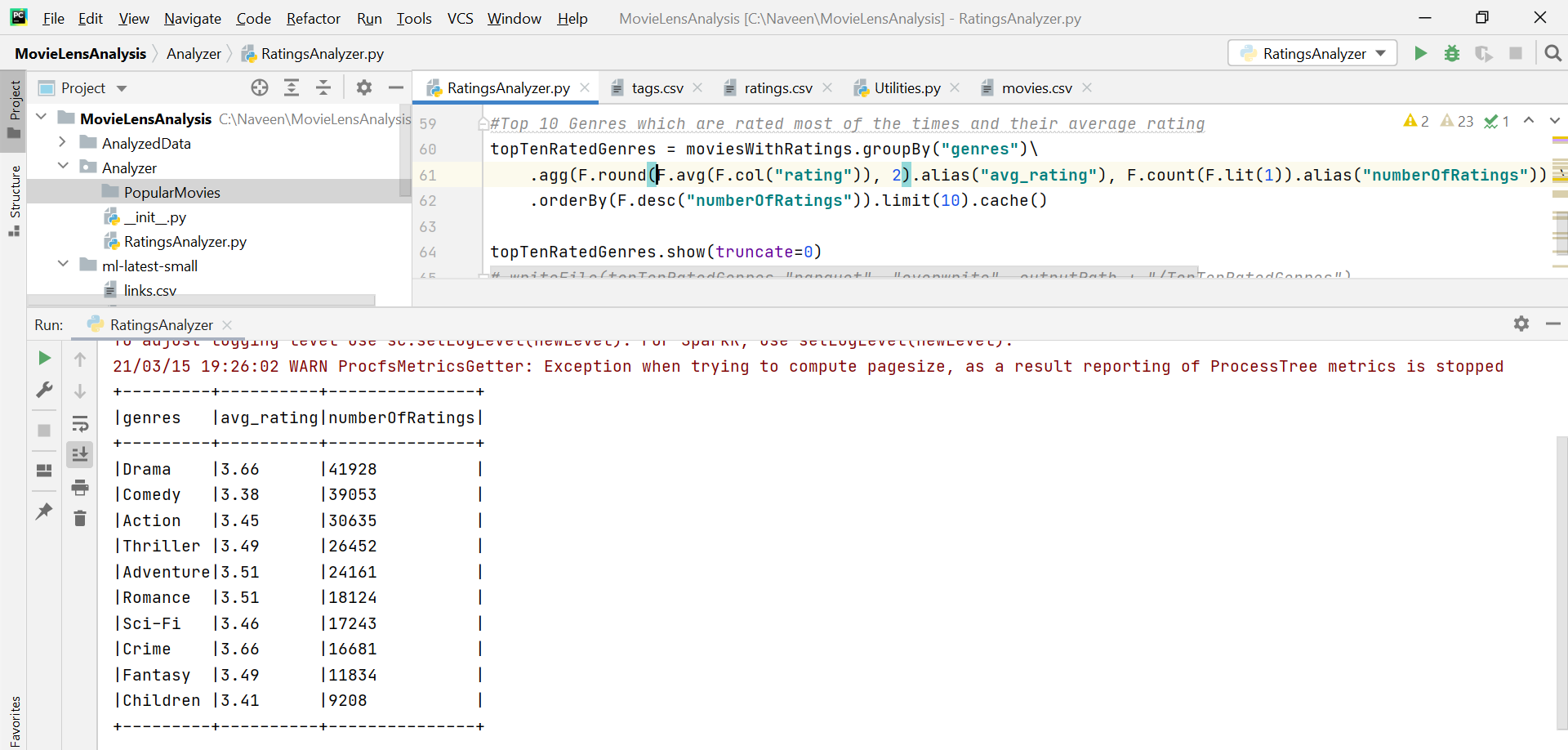
The movie of the decade has been derived as the movie which is rated more than 200 times and has rating more than 4 in each decade.



### Top Ten users who rated most of the movies

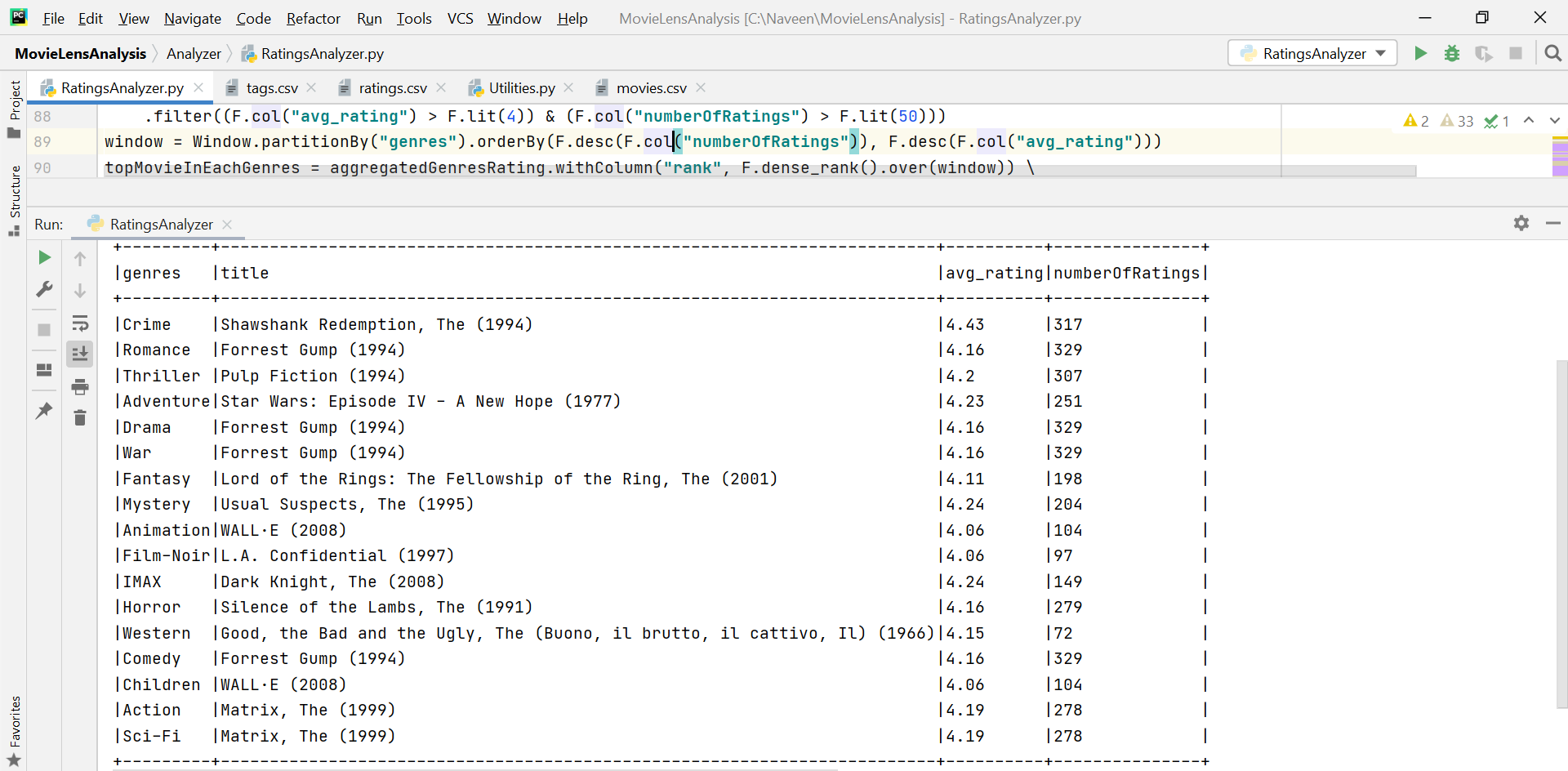


### Genres which are rated most of the time and average genres rating

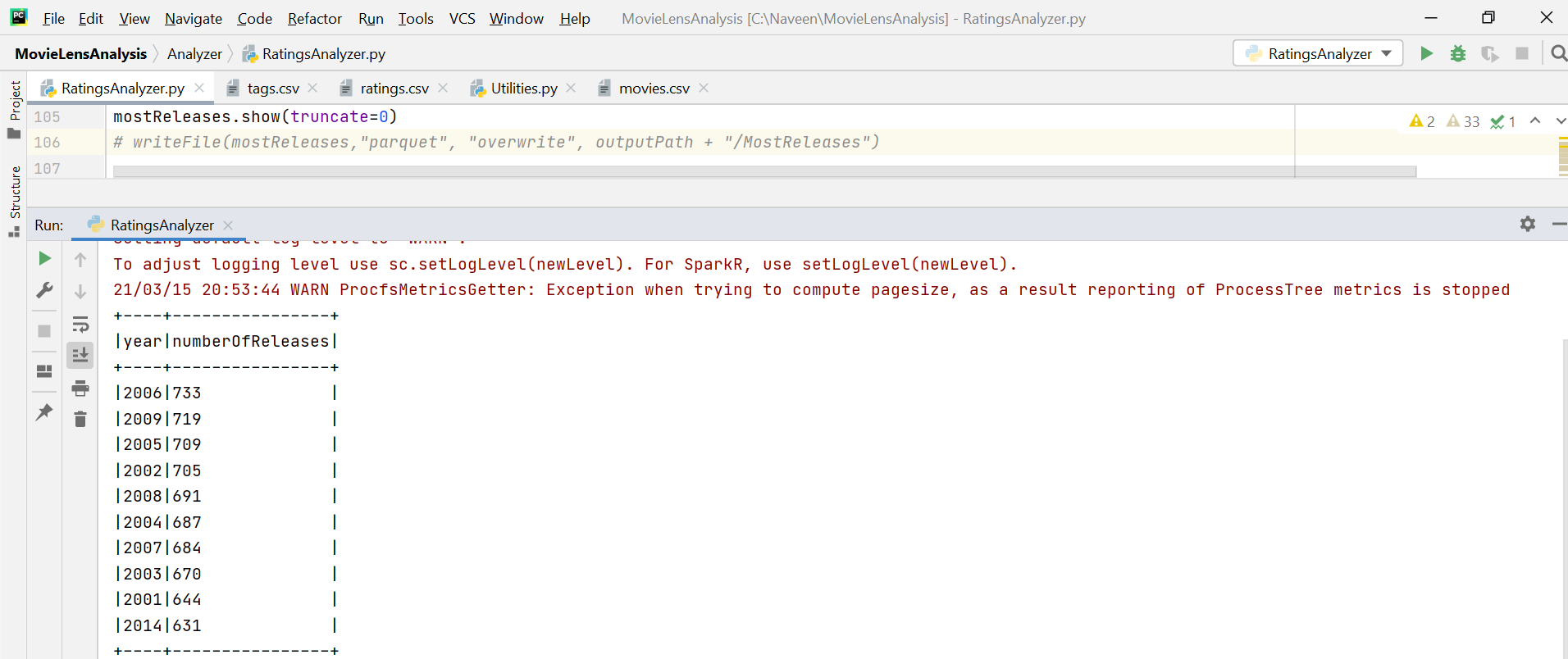


### Popular movie in each Genres

The popular movie in each genres are based on average rating of the film is more than 4 and has been rated atleast 50 times



### Top ten years which had most movie releases



### Top ten movies which are rated most of the time in the same release year

