ANALYSING MOVIE SCRIPTS
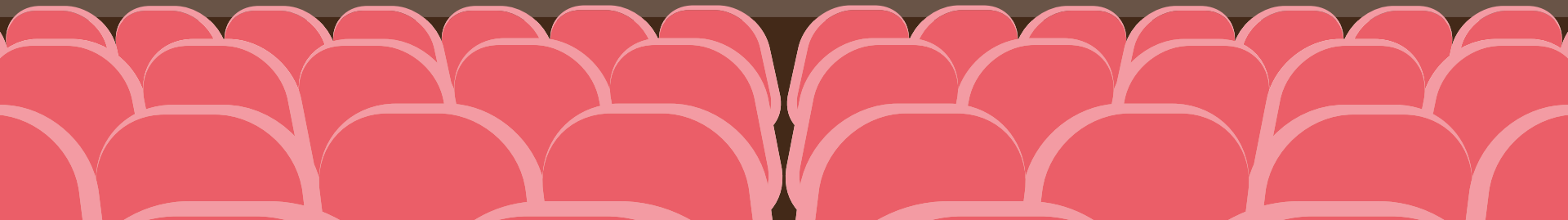
Using NLP to interpret movie scripts and make recommendations
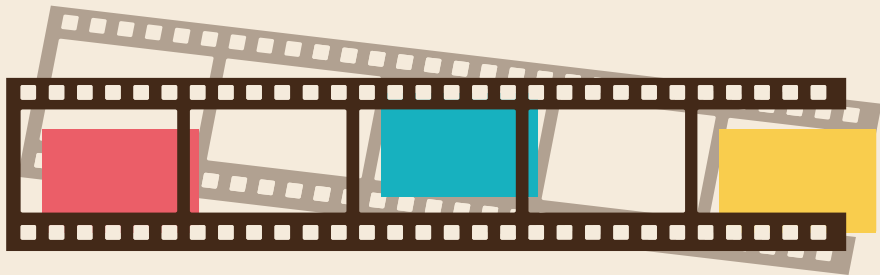
By Sam Ouimet

DATA AND PROCESS

- Scraped 70 full length scripts from IMSDB.com
  - 95-125 pages each
  - Split each into 100 chunks
- Tools
  - NLTK
  - NMF (11 topics)
  - Matplotlib/seaborn/plotly
- Recommender

FADE IN:

1   INT. DEPT. OF HEALTH, OFFICE - MORNING

CLOSE ON JOKER (30's), tears in his eyes from laughing so
hard. He's trying to get it under control. His greasy, black
hair is matted down. He's wearing an old, faded red hooded
zip-up sweatshirt, a threadbare gray scarf, thin from years
of use, hangs loosely around his neck.

WE NOTICE TWO FADED OLD SCARS cut at the corners of his
mouth. Almost forming a smile.

He's sitting across from an overworked SOCIAL WORKER (50's),
African American. Her office is cramped and run-down in a
cramped and run-down building. Stacks of folders piled high
in front of her.

She just sits behind her desk, waiting for his laughing fit
to end, she's been through this before. Finally it subsides.

Joker takes a deep breath, pauses to see if it's over.

Beat.

                    JOKER
            --is it just me, or is it getting
            crazier out there?

Despite the laughter, there's real pain in his eyes.
Something broken in him. Looks like he hasn't slept in days.

                    SOCIAL WORKER
            It's certainly tense. People are
            upset, they're struggling. Looking
            for work. The garbage strike seems
            like it's been going on forever.
            These are tough times.
                (then)
            How 'bout you. How's the job? Still
            enjoying it?

                    JOKER
        Yeah, I   mean, it's different each
        day, so   I really like that. I don't
        think I   could ever work in an
        office.   Behind a desk.
                    (MORE)
                                            2.

TOPICS

01. HORROR/ SUSPENSE

Turns, suddenly, slowly, dead, runs, looking

02. WAR/ VIOLENCE

German, private, nazi, firing, captain, soldier

03. ACTION IN ANY GENRE

Continuous, smashes, complex, hits, jumps, grabs

04. CRUDE COMEDY/ DRAMA

Dude, f***, s***, joint, mean, talking, huh

05. FRIENDLY/HAPPIER THEME

Well, really, good, right, little, great, okay, big

JOKER (2019)

"Arthur makes one bad decision that brings about a chain reaction of escalating events in this gritty character study." - Rotten Tomatoes

How do the topics in Joker change over time?

JOKER (2019)

RECOMMENDER

```
2
3   def recommend_movie(fav_movie):
4       sort_input = distances_named[fav_movie].sort_values(ascending=False)
5       mask = sort_input.index.isin([fav_movie])
6       all_results = sort_input[~mask]
7
8       return all_results[:4] # return 4 most similar movies
9
10  # run recommeder function
11  recommend_movie('Joker')
```

executed in 7ms, finished 19:36:35 2020-02-26

```
Scream            0.988154
Insidious         0.979025
Book-of-Eli,-The  0.973320
Get-Out           0.973310
Name: Joker, dtype: float64
```

FRIDAY THE 13TH VS. SUPERBAD
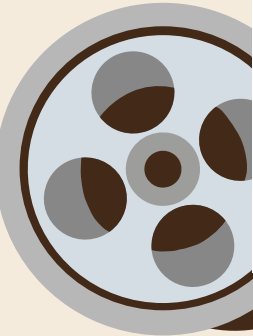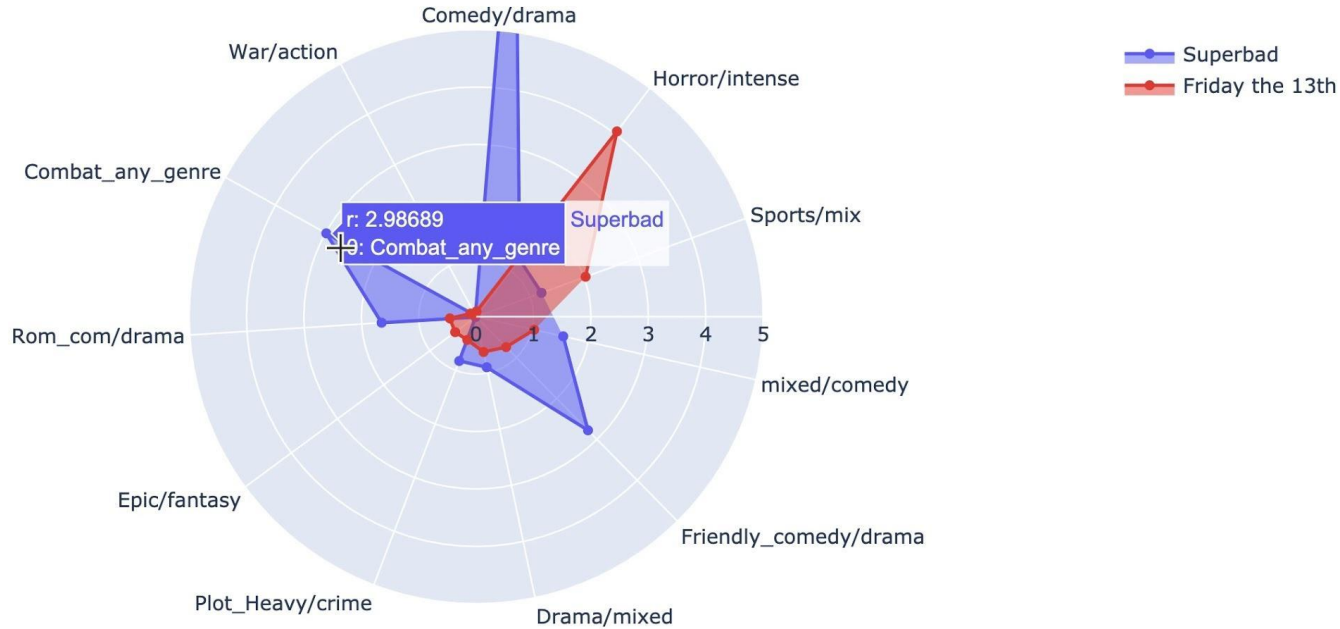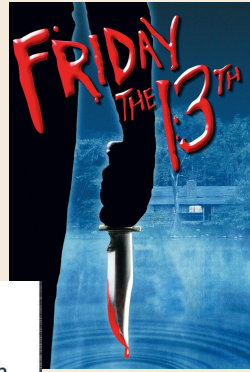
FUTURE WORK

More Movies

Expand dataset to potentially thousands of movies
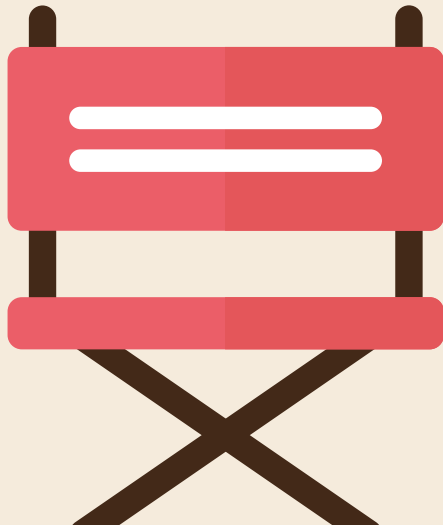
Movie/User Info

Movie rating, box office, director, actors, etc.
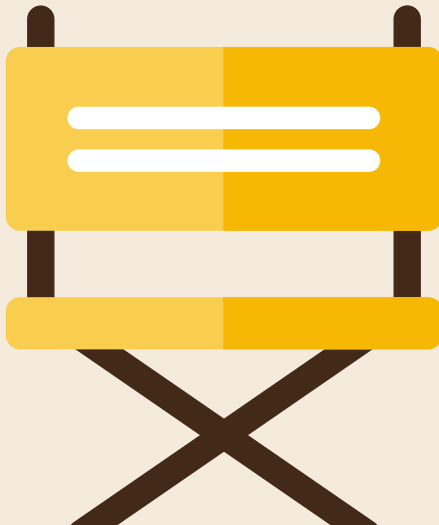
Web App
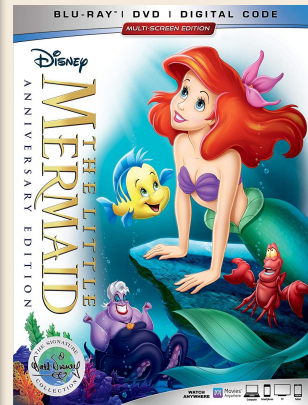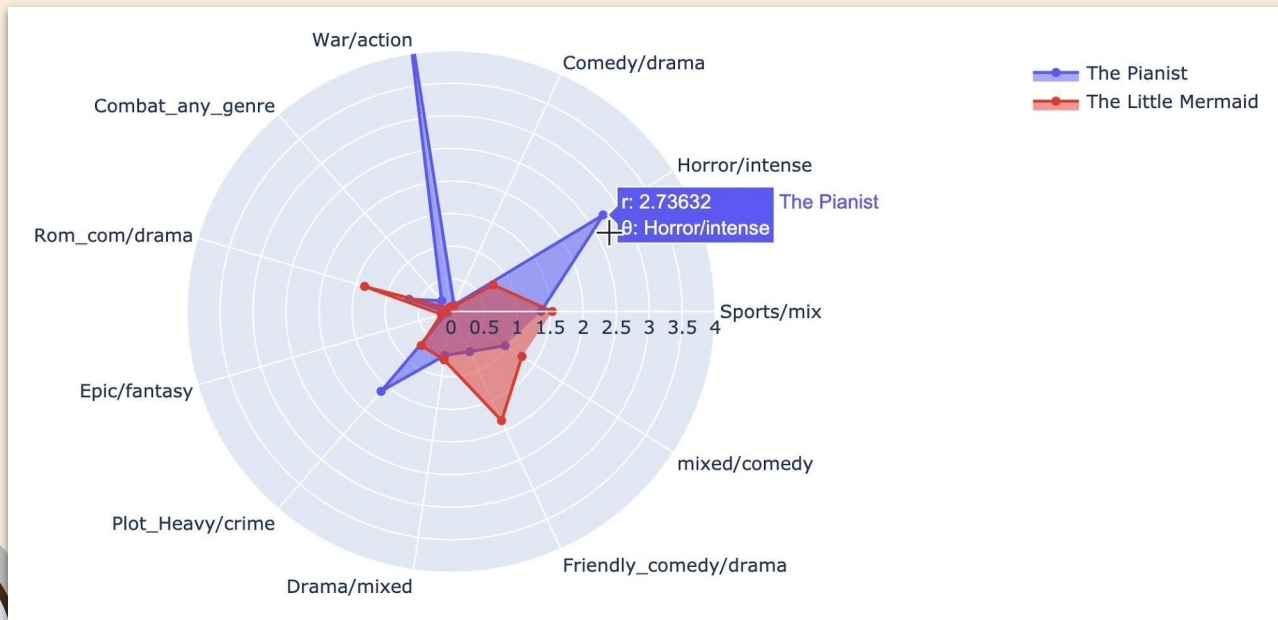
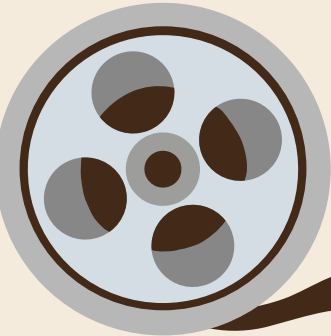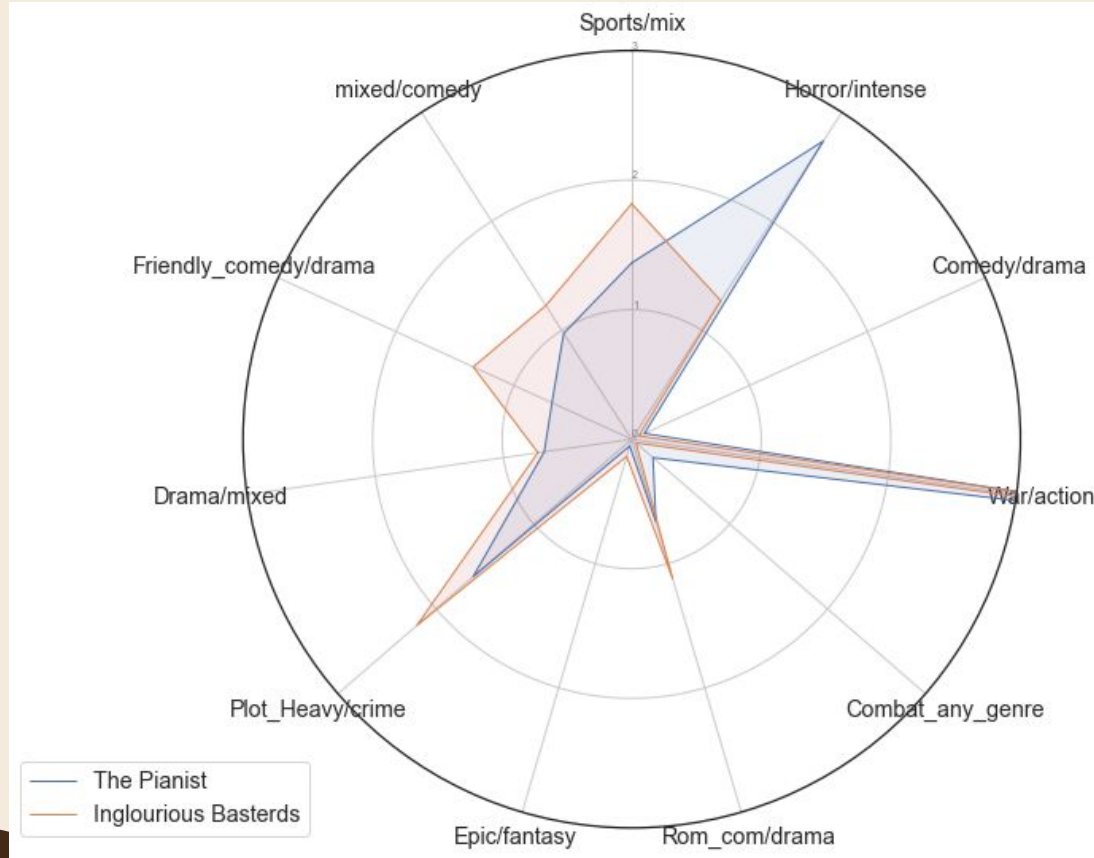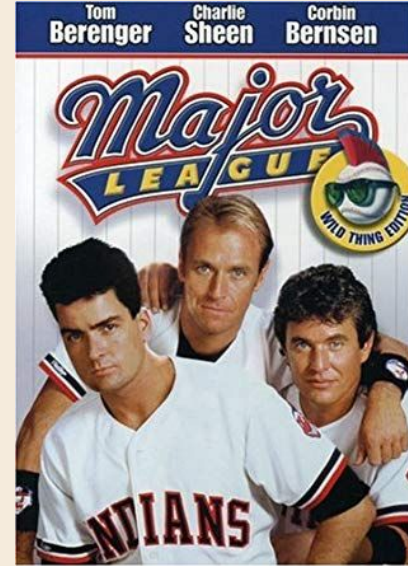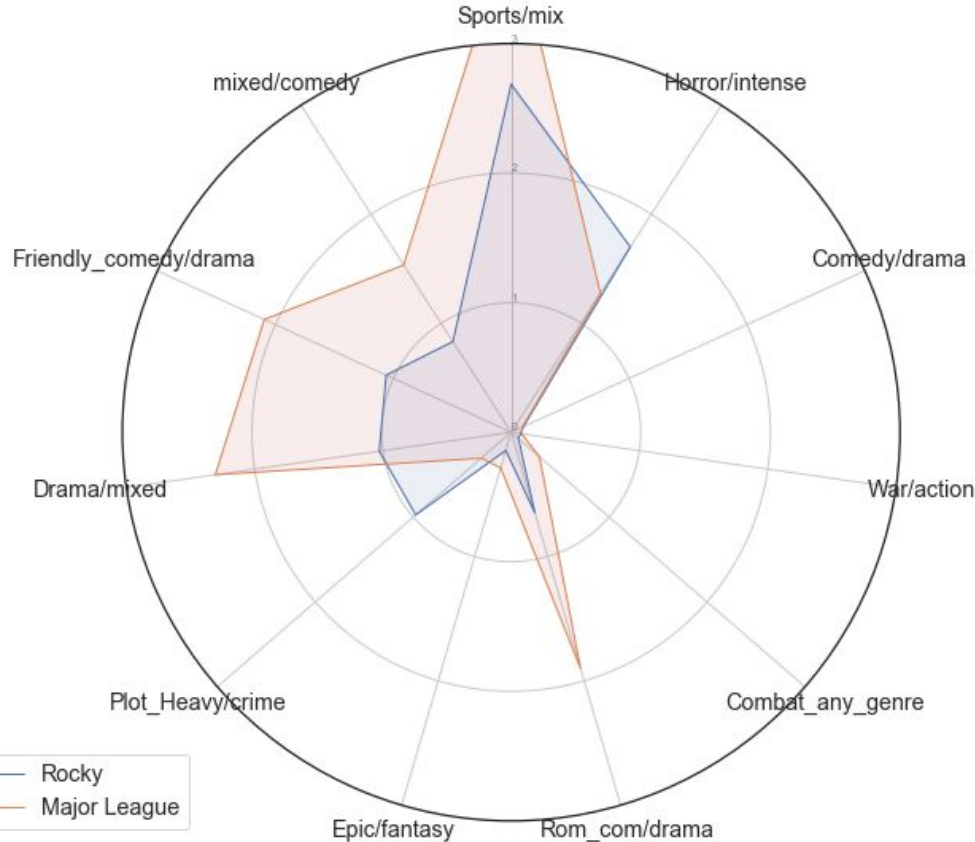Bring model to production with a usable web app
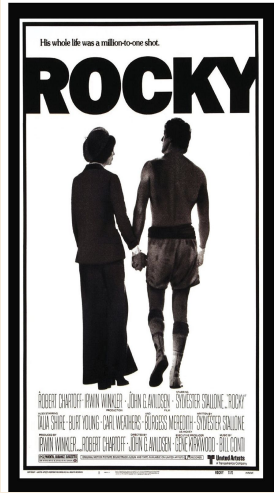
THANK YOU!
QUESTIONS?
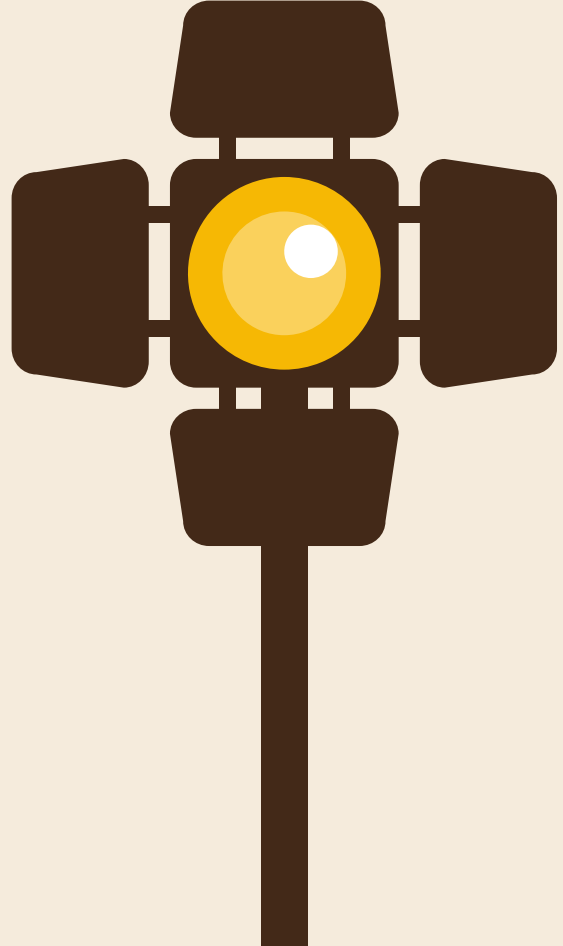
APPENDIX

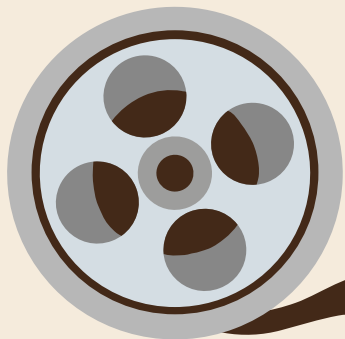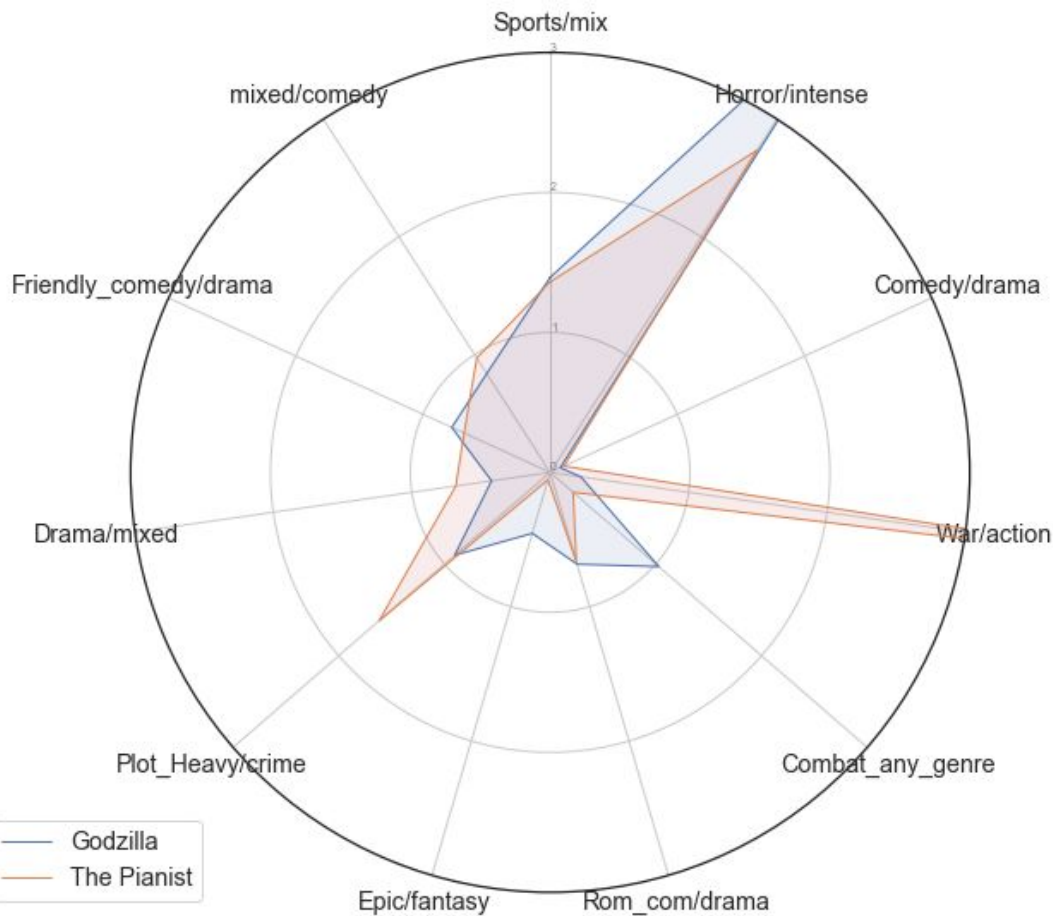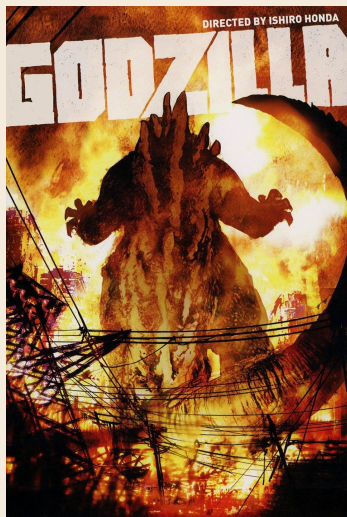# THE PIANIST VS. THE LITTLE MERMAID
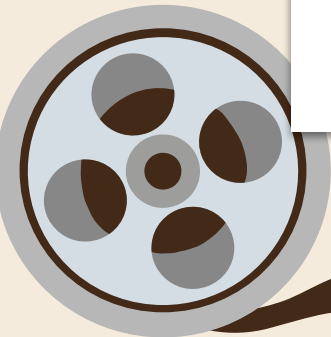
THE PIANIST VS. INGLOURIOUS BASTERDS

ROCKY VS. MAJOR LEAGUE

GODZILLA VS. THE PIANIST

PROCESS

The average feature screenplay, traditionally, is between **95 and 125 pages long...**

1. Split each script into **100 chunks**
2. Manually removed a lot of character names
3. Kept just adjectives and verbs (NLTK)
4. NMF for topic modeling