

Movie Critic Explorer Front End

Created for SEDS FALL 2019

This project was created to be able to explore movie critic scores and sentiments over time. To use it, use the setup.py file to download all necessary packages. After that, you can use this jupyter notebook as a front-end to access all of the modles and run a sentiment analysis. Throughout the program, you will be asked to specify a time range of interest and a year of interest to create certain visualizations. You can also use main.py for the same thing.

Created by: Melinda Haughey, Olivia Li, and Jessica Liu

Import Relevant Packages

```
In [1]: import altair as alt
from ipywidgets import interact, widgets
import pandas as pd
import nltk
nltk.download('stopwords')
from movieCriticsOverTime import sentiment_analysis, visualization, word_cloud,
clean_data, overall_rating
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]   C:\Users\Jessica\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

Import Data

To use your own data here, simply replace the reference

```
In [3]: MOVIE_DATA_PATH = "./data/movies.dat"
REVIEW_DATA_PATH = './data/reviews.csv'
movies = pd.read_csv(MOVIE_DATA_PATH, delimiter='\t')
reviews = pd.read_csv(REVIEW_DATA_PATH)
```

Clean Data

This puts the data in the correct format for use in the modules

```
In [4]: sub_movies = clean_data.clean_movies(movies)
reviews_clean= clean_data.clean_reviews(reviews)
reviews_merge = clean_data.merge_movies_reviews(reviews_clean, sub_movies)
```

View Overall Shape of Review Scores Over User Specified Time.

This will show you a visualization of the average rating per year. From here, you can find a year of interest to dive further into. If you are running main.py instead of using jupyter notebook as the front end, this will export into your folder as overallratings.html.

Select Year Range of Interest

```
In [5]: print('Please insert the start year of a range of interest between 1903 and 2009')
start_year = int(input())
```

Please insert the start year of a range of interest between 1903 and 2009
1990

```
In [6]: print('Please insert the end year of a range of interest between 1903 and 2009')
end_year = int(input())
```

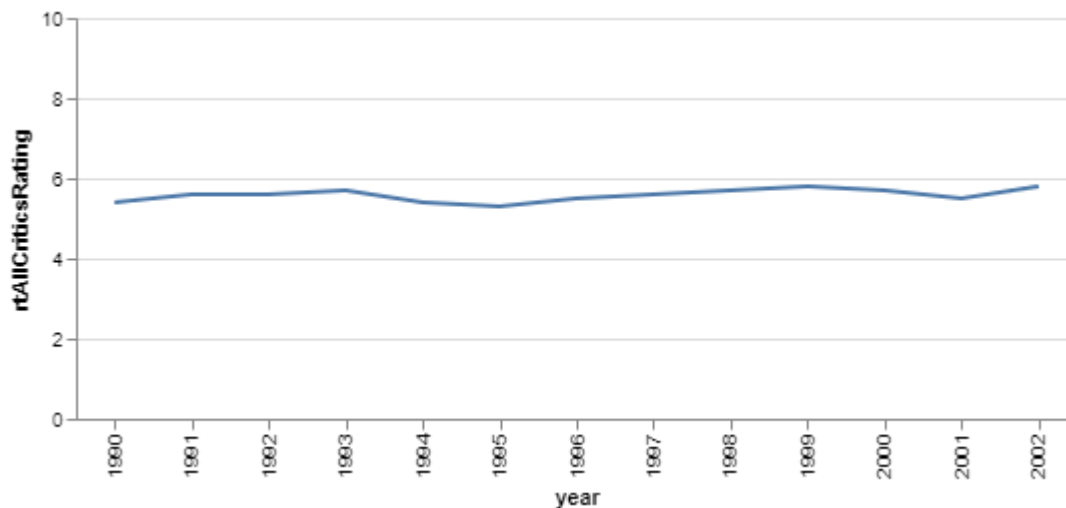
Please insert the end year of a range of interest between 1903 and 2009
2002

```
In [7]: #Initialize ability to see visualization in Jupyter Notebook
alt.renderers.enable('notebook')
```

Out[7]: `RendererRegistry.enable('notebook')`

```
In [8]: overall_rating.draw_trend_line_median_year(sub_movies,start_year,end_year)
```

Out[8]:



Select a Year of Interest to See Top Critics of that Year and Thier Sentiment

Did you see an interesting spike or valley? If so, maybe investigate that more to see what critics could be driving that change!

```
In [9]: print('Please input a year of interest to see critic activity that year: ')
interest_year = int(input())
```

Please input a year of interest to see critic activity that year:
2002

```
In [10]: top_critics = overall_rating.top5_critic_per_year(reviews_merge,interest_year)
print('The top critics that year are:')
pd.DataFrame(top_critics)
#add number of Reviews
```

The top critics that year are:

Out[10]:

	0
0	Mick LaSalle
1	Joe Baltake
2	Mike Clark
3	Eleanor Ringel Gillespie
4	Todd McCarthy

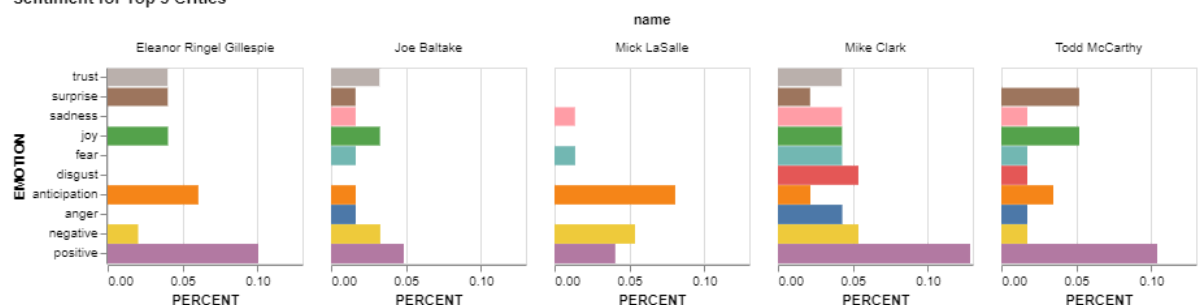
Grab Top Critic Reviews for Year of Interest and Analyze them for Sentiment

```
In [11]: quotes = sentiment_analysis.grab_quotes(reviews_merge,top_critics,interest_year)
emotion_array = sentiment_analysis.analyze_quote(quotes,top_critics)

visualization.visualize(emotion_array)
```

Out[11]:

Sentiment for Top 5 Critics



Word Cloud for Top 5 Critic Words

You can see what words are most commonly used by the critics listed above in the word cloud. This could suggest movies or reasons that may have impacted the overall score for your selected year.

```
In [12]: word_cloud.vis_word_cloud(quotes)
```



```
In [ ]:
```