Visualizing Reviewer Trends & Sentiment Over Time

A Technology Review

Background

Reviews on movies are created by critics and have an impact on the success and awards of a given movie. These reviews include chunks of text that describe the movie and the review. We are interested in providing the user information about how a critic rates movies over time so they can see if there is variation in rating and sentiment over the years.

To achieve this goal we will need to use packages for data manipulation, visualization (for exploratory analysis) and learning.

Data

(1)   List of movies that came out from 1903 to 2011
(2)   List of reviews from Rotten Tomatoes from 1929 to 2013 including their score, the reviewer, and the reviewer contents

Data

critic	fresh	imdb	link	publication	quote	review_date	rtid	title
Derek Adams	fresh	114709	http://www.	Time Out	So ingenious	##########	9559	Toy Story
Richard Corli	fresh	114709	http://www.	TIME Magaz	The year's m	##########	9559	Toy Story
David Ansen	fresh	114709	http://www.	Newsweek	A winning ar	##########	9559	Toy Story
Leonard Klad	fresh	114709	http://www.	Variety	The film spo	6/9/08 0:00	9559	Toy Story
Jonathan Ros	fresh	114709	http://onfilr	Chicago Rea	An entertain	##########	9559	Toy Story
Michael Boo	fresh	114709	http://www.	Denver Post	As Lion King	5/3/07 0:00	9559	Toy Story
Geoff Andre	fresh	114709	http://www.	Time Out	The film will	##########	9559	Toy Story
Janet Maslin	fresh	114709	http://movie	New York Tir	Children will	##########	9559	Toy Story
Kenneth Tura	fresh	114709	http://www.	Los Angeles	Although its	##########	9559	Toy Story
Roger Ebert	fresh	114709	http://www.	Chicago Sun-	The result is	1/1/00 0:00	9559	Toy Story
Susan Stark	fresh	114709	http://data.c	Detroit News	Disney's first	1/1/00 0:00	9559	Toy Story
Peter Stack	fresh	114709	http://www.	San Francisc	The script, b	1/1/00 0:00	9559	Toy Story
James Berar	fresh	114709	http://www.	ReelViews	The one big	1/1/00 0:00	9559	Toy Story
Rita Kemple	fresh	114709	http://www.	Washington	It's a nice ch	1/1/00 0:00	9559	Toy Story
	fresh	114709	http://www.	Entertainme	I can hardly	##########	9559	Toy Story
Roger Moore	fresh	114709	http://www.	Orlando Sen	The great vo	##########	9559	Toy Story
Roger Ebert	rotten	113497	http://www.	Chicago Sun-	A gloomy sp	1/1/00 0:00	12436	Jumanji
Richard Schi	fresh	107050	http://www.	TIME Magaz	Walter Matt	##########	10498	Grumpy Old N
Derek Adams	rotten	107050	http://www.	Time Out	Mediocre, re	##########	10498	Grumpy Old N
Caryn James	fresh	107050	http://movie	New York Tir	Just don't ex	##########	10498	Grumpy Old N
James Berar	fresh	107050	http://www.	ReelViews	While it won	1/1/00 0:00	10498	Grumpy Old N
Roger Ebert	rotten	107050	http://www.	Chicago Sun-	The movie is	1/1/00 0:00	10498	Grumpy Old N
Desson Thom	fresh	107050	http://www.	Washington	If you poke tl	1/1/00 0:00	10498	Grumpy Old N
Liam Lacey	rotten	114885	http://www.	Globe and M	Never escap	##########	16697	Waiting to Ex
Kenneth Tura	fresh	114885	http://www.	Los Angeles	A pleasant if	##########	16697	Waiting to Ex
Edward Guth	rotten	114885	http://www.	San Francisc	You want the	1/1/00 0:00	16697	Waiting to Ex
Desson Thom	fresh	114885	http://www.	Washington	This sister-c	1/1/00 0:00	16697	Waiting to Ex
Joan Walsh	rotten	114885	http://www.	Salon.com	The male-ba	1/1/00 0:00	16697	Waiting to Ex
Joe Baltake	rotten	114885	http://www.	Sacramento	With one pos	1/1/00 0:00	16697	Waiting to Ex
Susan Stark	rotten	114885	http://data.c	Detroit News	For all the pl	1/1/00 0:00	16697	Waiting to Ex
James Berar	rotten	114885	http://www.	ReelViews	With the exc	1/1/00 0:00	16697	Waiting to Ex
Roger Ebert	fresh	114885	http://www.	Chicago Sun-	An escapist f	1/1/00 0:00	16697	Waiting to Ex

```
10682 lines (10681 sloc)    510 KB

 1   1::Toy Story (1995)::Adventure|Animation|Children|Comedy|Fantasy
 2   2::Jumanji (1995)::Adventure|Children|Fantasy
 3   3::Grumpier Old Men (1995)::Comedy|Romance
 4   4::Waiting to Exhale (1995)::Comedy|Drama|Romance
 5   5::Father of the Bride Part II (1995)::Comedy
 6   6::Heat (1995)::Action|Crime|Thriller
 7   7::Sabrina (1995)::Comedy|Romance
 8   8::Tom and Huck (1995)::Adventure|Children
 9   9::Sudden Death (1995)::Action
10   10::GoldenEye (1995)::Action|Adventure|Thriller
11   11::American President, The (1995)::Comedy|Drama|Romance
12   12::Dracula: Dead and Loving It (1995)::Comedy|Horror
13   13::Balto (1995)::Animation|Children
14   14::Nixon (1995)::Drama
15   15::Cutthroat Island (1995)::Action|Adventure|Romance
16   16::Casino (1995)::Crime|Drama
17   17::Sense and Sensibility (1995)::Comedy|Drama|Romance
18   18::Four Rooms (1995)::Comedy|Drama|Thriller
19   19::Ace Ventura: When Nature Calls (1995)::Comedy
20   20::Money Train (1995)::Action|Comedy|Crime|Drama|Thriller
21   21::Get Shorty (1995)::Action|Comedy|Drama
22   22::Copycat (1995)::Crime|Drama|Horror|Mystery|Thriller
23   23::Assassins (1995)::Action|Crime|Thriller
24   24::Powder (1995)::Drama|Sci-Fi
25   25::Leaving Las Vegas (1995)::Drama|Romance
26   26::Othello (1995)::Drama
27   27::Now and Then (1995)::Drama
28   28::Persuasion (1995)::Drama|Romance
```
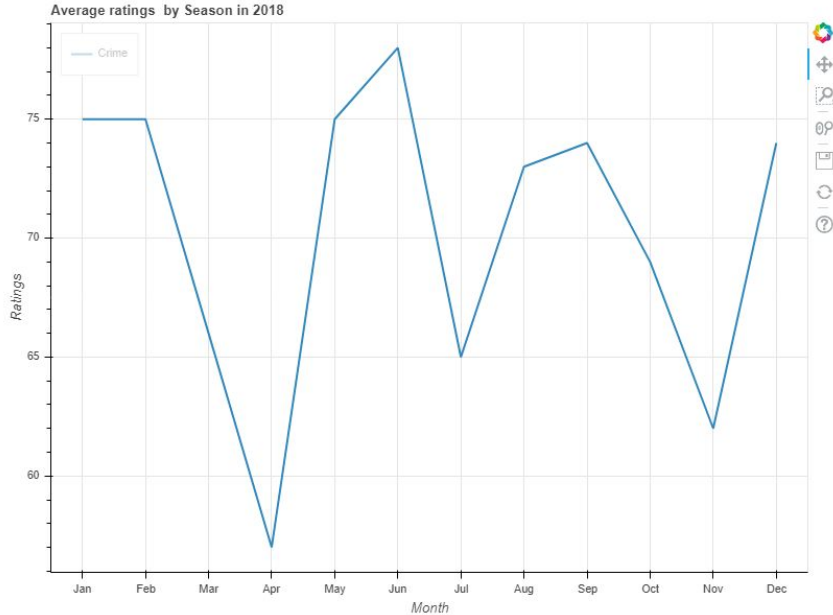
Use Case Example

USER INPUT: User selects the name of a movie reviewer.

SYSTEM: Pulls movie review data for the selected reviewer, visualizes the rating distribution over time

SYSTEM OUTPUT: System shows visualization of rating distribution to the user

SYSTEM ASKS: Would you like to the movies that this critic has reviewed? And the sentiment ?

USER INPUT: Yes or No

SYSTEM OUTPUT: Either displays a list of movies and their sentiment score or ends

Use Case Example - Output 1

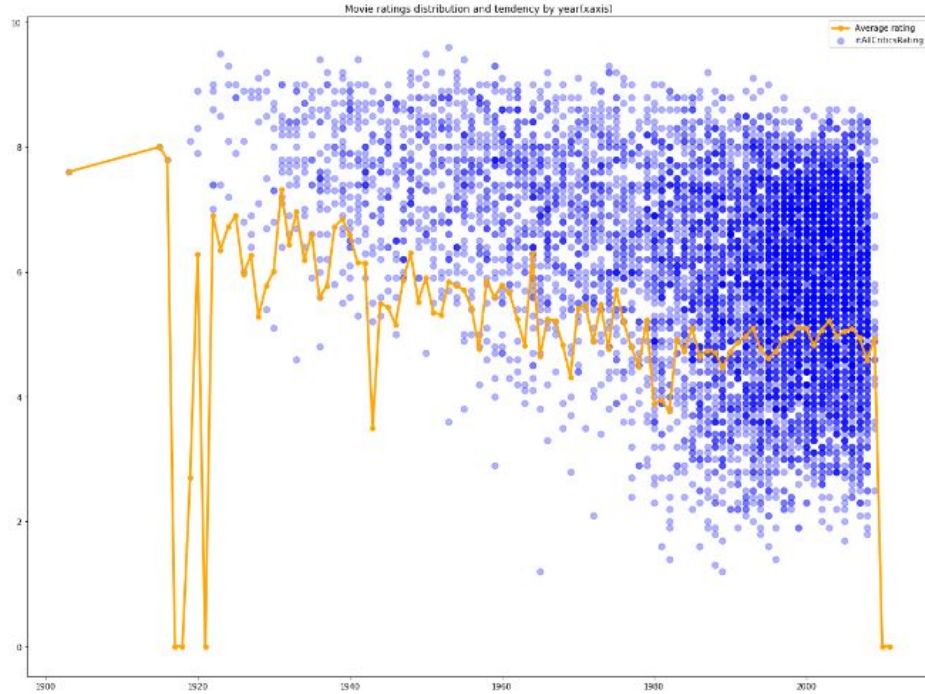Average ratings by Season in 2018

Input: Reviewer Name/ Time Span

Eg: Derek Adams 01/01/2018-12/31/2018

Output: A rating trend line within the time span

Package Usage: Pandas/ Bokeh

Use Case Example - Output 2

Input: All Interested Reviewers /Time Span

Output: Movie Ratings Distribution Over Year

Package Usage: Pandas/ Matplotlib

Use Case Example - Output 3

	quote
So ingenious in concept, design and execution ...	
	Mediocre, regrettably.
It's a Cinderella story that gets turned on it...	
British director Anderson does a fair job with...	
The design and effects teams have lent scale a...	
An unruly orangutan plus an opulent hotel sett...	
It's what Hollywood wanted Woo for: bigger, br...	
Not only does the film bear little resemblance...	
	Insipid songs and not much story.
	Unforgettable.
Has distinct limitations, but the current retu...	
Ephron's timing and sensitivity have deserted ...	

EMOTION	%WORDS
positive	6.16%
eoplesday	
trust	3.08%
anticipation	2.53%
ormatics	
joy	1.76%
ay	
surprise	0.99%
negative	0.88%
sadness	0.55%
disgust	0.44%
fear	0.44%
anger	0.33%

Input: Reviewer Name Eg: Derek Adams

Output: Example Quotes/ Sentiment Score
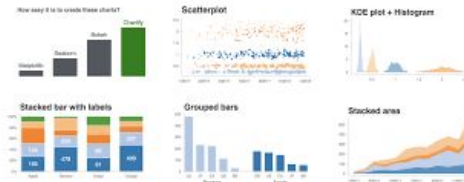
Package Usage: Numpy/ NLTK

Needs

- Data Wrangling Package (pandas and numpy)
- Visualization Package (bokeh or matplotlib)
- Sentiment Analysis/Learning Package (NLTK or SpaCy)

bokeh   matplotlib

Library Packages Options (Visualization)

Package 1 - Bokeh

Pros:
It has high-performance interactivity over very large or streaming datasets

It allows viewers to explore the data themselves for some really complex data which works as a dashboard

Cons:
The interactivity is limited. It only offers a 2-D, grid-like canvas with axes as the baseline

Package 2 - MatPlotLib

Pros:
Good for basic plotting -- bars, pies, lines, scatter plots, etc.

Most popular and easy to use without a lot of code.

Cons:
It doesn't allow any interaction.

It is heavily reliant on other packages.

Library Packages Options (Sentiment Analysis)

Package 1 - NLTK

Pros:

It provides accessible interfaces to a lot of corpora and lexical sources mapped to machine learning algorithms, as well as a robust choice of parsers and utilities.

Cons:
Too many algorithms to choose from for a particular problem which is boon for a researcher but a bane for a developer.

Not efficient compared to SpaCy.

Package 2 - SpaCy

Pros:

It claims to be ''industrial-strength natural language processing. SpaCy keeps the best algorithm for a problem in its toolkit and keep it updated as state of the art improves.

Performance is better.

Cons:

It doesn't support multiple languages.