

IMD0033 - Probabilidade

Aula 10 - Análise Exploratória de Dados IV

Ivanovitch Silva
Agosto, 2017



Agenda

- Estudo de caso: analisando filmes
- Motivação: enviesamento de dados
- Histograma e gráfico de caixa

Atualizar o repositório

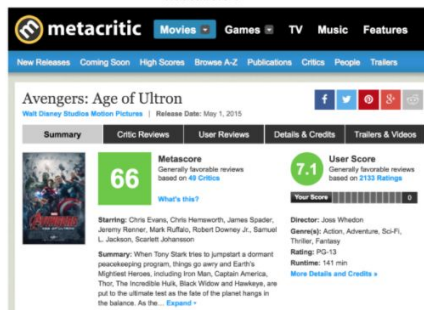
```
git clone https://github.com/ivanovitchm/IMD0033_Probabilidade.git
```

Ou

```
git pull
```

Estudo de caso: avaliando filmes

Metacritic



metacritic Movies Games TV Music Features

New Releases Coming Soon High Scores Browse A-Z Publications Critics People Trailers

Avengers: Age of Ultron

Walt Disney Studios Motion Pictures | Release Date: May 1, 2015

Summary Critic Reviews User Reviews Details & Credits Trailers & Videos

Metascore
66
Generally favorable reviews based on 49 Critics

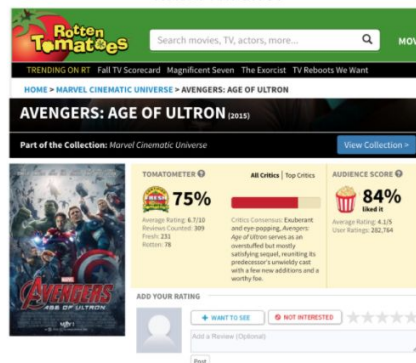
User Score
7.1
Generally favorable reviews based on 2133 Ratings

What's this?

Starring: Chris Evans, Chris Hemsworth, James Spader, Jeremy Renner, Mark Ruffalo, Robert Downey Jr., Samuel L. Jackson, Scarlett Johansson
Director: Josh Whedon
Genre(s): Action, Adventure, Sci-Fi, Thriller, Fantasy
Rating: PG-13
Runtime: 141 min
[More Details and Credits](#)

Summary: When Tony Stark tries to jumpstart a dormant peacekeeping program, things go awry and Earth's Mightiest Heroes, including Iron Man, Captain America, Thor, The Incredible Hulk, Black Widow and Hawkeye, are put to the ultimate test as the fate of the planet hangs in the balance. [Expand](#)

Rotten Tomatoes



Rotten Tomatoes Search movies, TV, actors, more... MOVIE

TRENDING ON RT Fall TV Scorecard Magnificent Seven The Exorcist TV Reboots We Want

HOME > MARVEL CINEMATIC UNIVERSE > AVENGERS: AGE OF ULTRON

AVENGERS: AGE OF ULTRON (2015)

Part of the Collection: Marvel Cinematic Universe [View Collection](#)

TOMATOMETER 75%
Average Rating: 6.7/10
Reviews Counted: 309
Fresh: 233
Rotten: 76

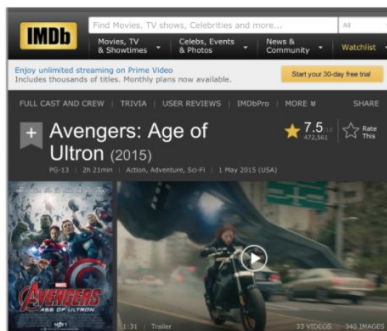
AUDIENCE SCORE 84%
Average Rating: 4.2/5
User Ratings: 282,794

Critic Consensus: Exuberant and eye-popping, *Avengers: Age of Ultron* serves as an overwrought but moodily satisfying sequel, rewarding its predecessor's unwieldy cast with a few new additions and a worthy foe.

ADD YOUR RATING

★★★★★

IMDB



IMDb First Movies, TV shows, celebrities and more... all

Movies, TV & Showtimes Critics, Events & Photos News & Community Watchlist

Enjoy unlimited streaming on Prime Video
Includes thousands of titles. Monthly plans now available. [Start your 30-day free trial](#)

FULL CAST AND CREW TRIVIA USER REVIEWS | IMDbPro | MORE W SHARE

Avengers: Age of Ultron (2015)

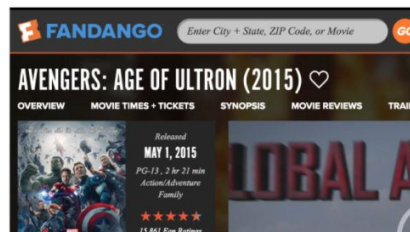
PG-13 | 2h 23min | Action, Adventure, Sci-Fi | 1 May 2015 (USA)

7.5 472,361 Rate This

Trailer

33 VIDEOS 141 PHOTOS

Fandango



FANDANGO Enter City + State, ZIP Code, or Movie GO

AVENGERS: AGE OF ULTRON (2015)

OVERVIEW MOVIE TIMES + TICKETS SYNOPSIS MOVIE REVIEWS TRAILER

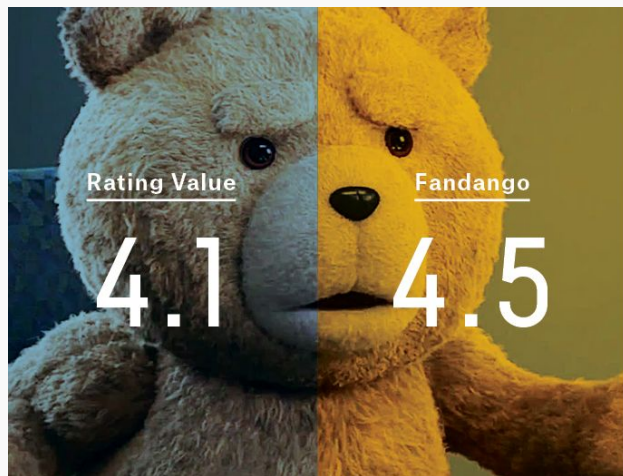
Released MAY 1, 2015

PG-13 - 2 hr 21 min
Action/Adventure
Family

★★★★★
15,861 Fan Ratings

GLOBAL A

Enviesamiento de dados



Conjunto de dados

	FILM	RT_user_norm	Metacritic_user_nom	IMDB_norm	Fandango_Ratingvalue	Fandango_Stars
0	Avengers: Age of Ultron (2015)	4.3	3.55	3.90	4.5	5.0
1	Cinderella (2015)	4.0	3.75	3.55	4.5	5.0
2	Ant-Man (2015)	4.5	4.05	3.90	4.5	5.0
3	Do You Believe? (2015)	4.2	2.35	2.70	4.5	5.0
4	Hot Tub Time Machine 2 (2015)	1.4	1.70	2.55	3.0	3.5

<https://github.com/fivethirtyeight/data/tree/master/fandango>

Distribuição de frequências

Frequency Distribution
(sorted by **frequency** in
descending order)

Value	Frequency
4.1	16
4.2	12
3.9	12
4.3	11
3.7	9
3.5	9
4.5	9
3.4	9
3.6	8
4.4	7
4.0	7
3.2	5
2.9	5
3.8	5
3.3	4
4.6	4
3.0	4
4.8	3
3.1	3
2.8	2
2.7	2

Name: Fandango_Ratingvalue,
dtype: int64

Frequency Distribution
(sorted by **unique value** in
ascending order)

Value	Frequency
2.7	2
2.8	2
2.9	5
3.0	4
3.1	3
3.2	5
3.3	4
3.4	9
3.5	9
3.6	8
3.7	9
3.8	5
3.9	12
4.0	7
4.1	16
4.2	12
4.3	11
4.4	7
4.5	9
4.6	4
4.8	3

Name: Fandango_Ratingvalue,
dtype: int64

Intervalos de agrupamento

Fandango
Frequency
Distribution

		Bins	Count
2.7	2		
2.8	2		
2.9	5	0.0 - 0.5	0
3.0	4	0.5 - 1.0	0
3.1	3	1.0 - 1.5	0
3.2	5	1.5 - 2.0	0
3.3	4	2.0 - 2.5	0
3.4	9	2.5 - 3.0	9
3.5	9	3.0 - 3.5	25
3.6	8	3.5 - 4.0	43
3.7	9	4.0 - 4.5	53
3.8	5	4.5 - 5.0	16
3.9	12		
4.0	7		
4.1	16		
4.2	12		
4.3	11		
4.4	7		
4.5	9		
4.6	4		
4.8	3		

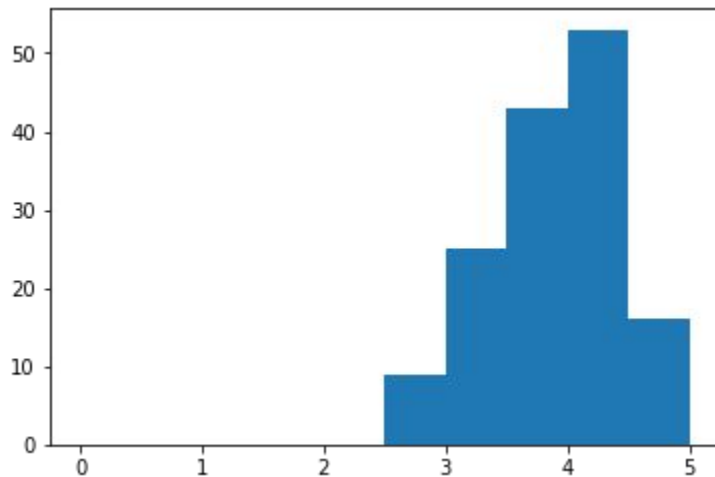
Rotten Tomatoes
Frequency
Distribution

		Bins	Count
2.00	1		
2.10	1		
2.15	1	0.0 - 0.5	0
2.20	1	0.5 - 1.0	0
2.30	2	1.0 - 1.5	0
2.45	2	1.5 - 2.0	0
2.50	1	2.0 - 2.5	8
2.55	1	2.5 - 3.0	20
2.60	2	3.0 - 3.5	50
2.70	4	3.5 - 4.0	58
2.75	5	4.0 - 4.5	10
2.80	2	4.5 - 5.0	0
2.85	1		
2.90	1		
2.95	3		
...	..		
4.00	1		
4.05	1		
4.10	4		
4.15	1		
4.20	2		
4.30	1		

truncated
to save
space

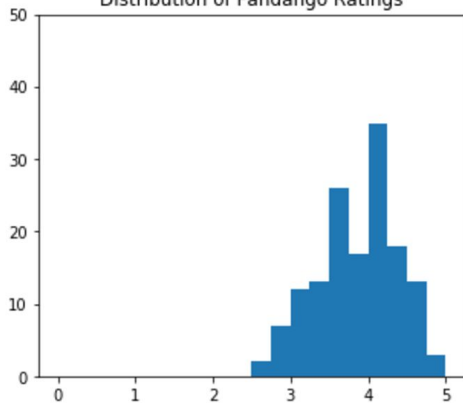
Histograma no Matplotlib

```
ax.hist(norm_reviews['Fandango_Ratingvalue'], range=(0, 5))
```



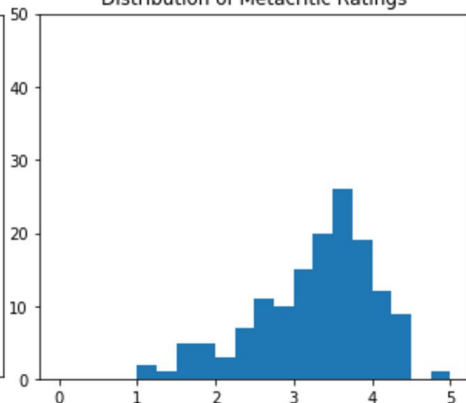
Comparando histogramas

Distribution of Fandango Ratings



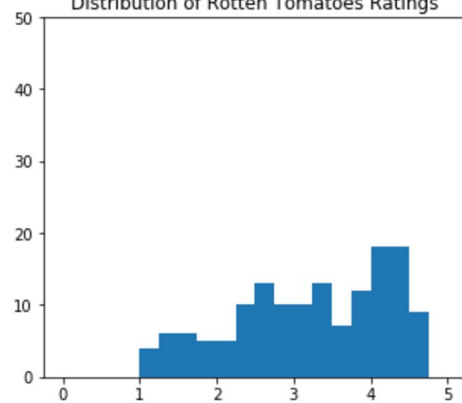
50% das avaliações
estão entre 2 e 4

Distribution of Metacritic Ratings



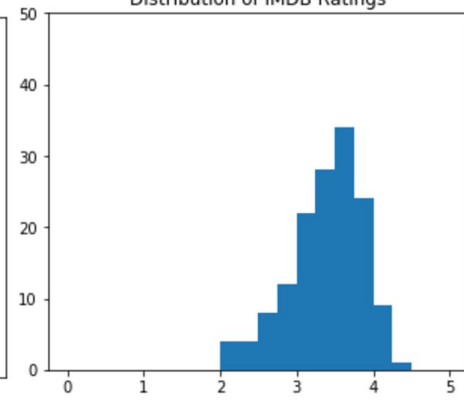
75% das avaliações
estão entre 2 e 4

Distribution of Rotten Tomatoes Ratings



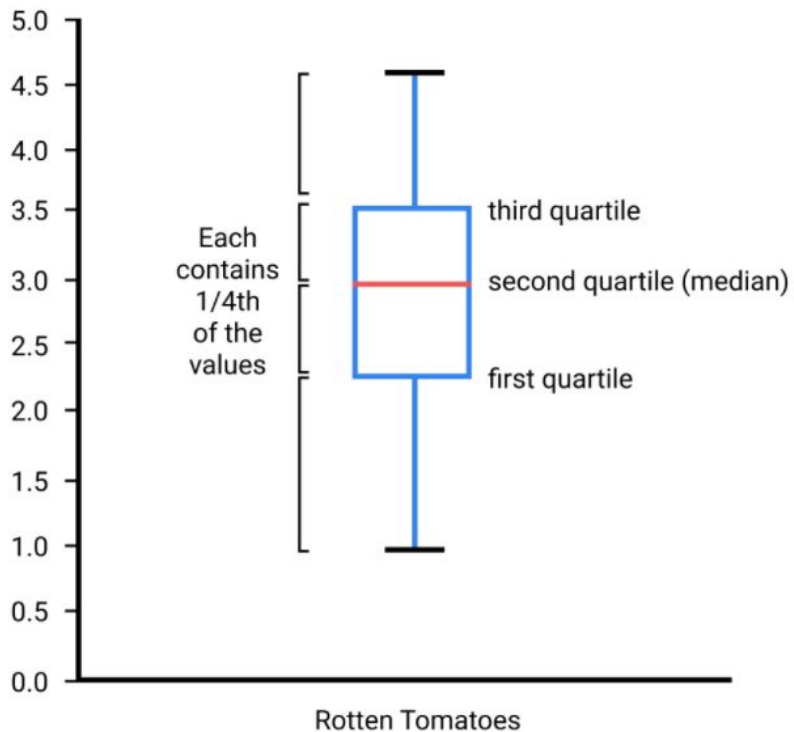
50% das avaliações
estão entre 2 e 4

Distribution of IMDB Ratings



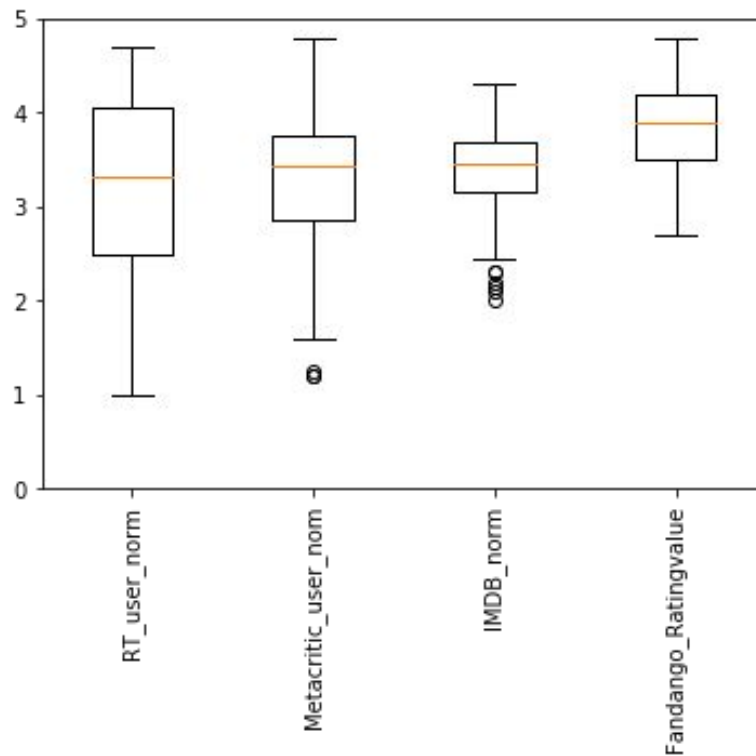
90% das avaliações
estão entre 2 e 4

Quartil e gráfico de caixa



```
ax.boxplot(norm_reviews[ 'RT_user_norm' ] )
```

Múltiplos gráficos de caixa





Exploratory Data Analysis IV - Histogram and Boxplot.ipynb