

# Predicting Movie Scores



by Shi Yunfan




# Web Scraping

- movie names
- year
- # of reviews
- type
- scores
- actors/actress

**Most Voted Titles Released 2017-01-01 to 2017-12-31**

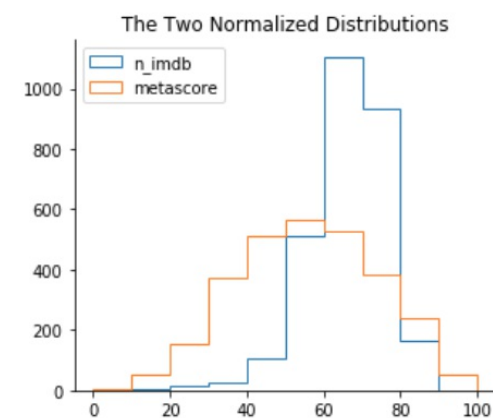
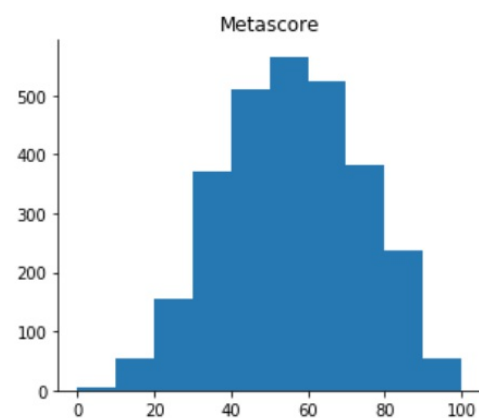
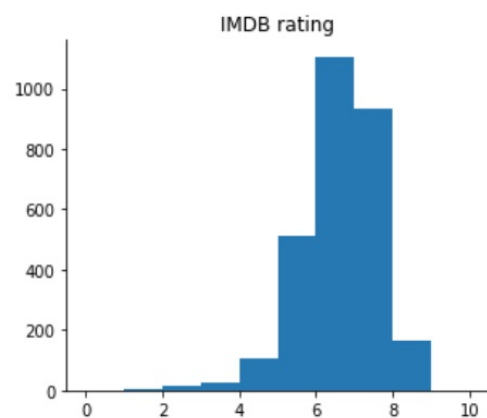
1 to 50 of 117,062 titles | Next » View Mode: Compact Detailed

Sort by: Popularity | Alphabetical | IMDb Rating | **Number of Votes** | US Box Office | Runtime | Year | Release Date

-  **1. Logan** (2017)  
R | 137 min | Action, Drama, Sci-Fi  
★ **8.3** ☆ Rate this **77** Metascore  
In the near future, a weary Logan cares for an ailing Professor X somewhere on the Mexican border. However, Logan's attempts to hide from the world and his legacy are upended when a young mutant arrives, pursued by dark forces.  
Director: James Mangold | Stars: Hugh Jackman, Patrick Stewart, Dafne Keen, Boyd Holbrook  
Votes: 309,245 | Gross: \$226.23M
-  **2. Guardians of the Galaxy Vol. 2** (2017)  
PG-13 | 136 min | Action, Adventure, Sci-Fi  
★ **8.1** ☆ Rate this **67** Metascore  
The Guardians must fight to keep their newfound family together as they unravel the mystery of Peter Quill's true parentage.  
Director: James Gunn | Stars: Chris Pratt, Zoe Saldana, Dave Bautista, Vin Diesel  
Votes: 169,093 | Gross: \$369.87M
-  **3. Beauty and the Beast** (2017)  
PG | 129 min | Family, Fantasy, Musical  
★ **7.5** ☆ Rate this **65** Metascore  
An adaptation of the fairy tale about a monstrous-looking prince and a young woman who fall in

# Initial Data Analysis

	imdb	metascore
min	1.6	7.0
max	9.0	100.0



# Multiple Linear Regression

```
Level2R      -7.461e+01  1.246e+00 -59.896 < 2e-16 ***
Score2       5.439e-01  8.578e-02   6.341 2.32e-10 ***
Votes        8.952e-04  5.006e-06 178.804 < 2e-16 ***
```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 93.61 on 28057 degrees of freedom  
Multiple R-squared: 0.7327, Adjusted R-squared: 0.7318  
F-statistic: 801 on 96 and 28057 DF, p-value: < 2.2e-16

# movie market gross prediction using multiple linear regression

**Select Votes:**

10,000 600,000

0 60,000 120,000 240,000 360,000 480,000 600,000

**Select Movie Score:**

10 60 100

10 19 28 37 46 55 64 73 82 91 100

**Select Movie Year:**

1,990 2,010 2,019

1,990 1,996 2,002 2,008 2,014 2,019

**Select movie level:**

☒ not R

☐ R

data

22.25