

Data Analysis Project in Excel and Tableau. Reel Insights: Exploring IMDb Ratings

Introduction

In this Data Analysis Project, I delve into the intricate world of movie ratings on IMDb to uncover the factors that influence a film's success, as reflected by its IMDb rating. This exploration holds significance not only for filmmakers striving to create impactful cinema but also for movie enthusiasts seeking to discern the ingredients of a quality film.

Ask

Embarking on this journey, I pose several pertinent questions that guide the analysis:

- 1. What is the distribution of the IMDb Score overall?*
- 2. Does the duration of a movie affect its IMDB rating?*
- 3. Do movies in certain languages or countries tend to receive higher IMDB ratings?*
- 4. Are certain movie genres more likely to receive higher IMDB ratings than others?*
- 5. How have IMDB ratings changed over the years?*
- 6. Is there a relationship between the number of user reviews/votes and IMDB rating?*

Prepare

To navigate through this exploration, I've used a [comprehensive dataset](#) sourced from IMDb. The dataset offers a wealth of information, including:

- Director names
- Movie durations
- Actor names
- Movie genres
- Number of user votes and reviews
- Language and country of origin
- Release years
- IMDb scores

Process

After importing the dataset into Microsoft Excel, meticulous exploration and data preparation were undertaken to ensure the dataset was tidy and primed for analysis. Here are the steps undertaken:

1. Review the Data:

- The dataset was thoroughly reviewed to understand its structure, columns content and its format.

index	director_name	duration	actor_2_name	genres	actor_1_name	movie_title	num_voted_user	actor_3_name	movie_imdb_link	num_user_for_reviews	language	country	title_year	imdb_score
0	James Cameron	178.0	Joel David Moore	Action Adventure Fantasy Sci-Fi	CCH Pounder	Avatar~†	886204	Wes Studi	http://www.imdb.3054.0		English	USA	2009.0	7.9
1	Gore Verbinski	169.0	Orlando Bloom	Action Adventure Fantasy	Johnny Depp	Pirates of the Caribbean: At W	471220	Jack Davenport	http://www.imdb.1238.0		English	USA	2007.0	7.1
2	Sam Mendes	148.0	Rory Kinnear	Action Adventure Thriller	Christoph Waltz	Spectre~†	275868	Stephanie Sigman	http://www.imdb.994.0		English	UK	2015.0	6.8
3	Christopher Nolan	164.0	Christian Bale	Action Thriller	Tom Hardy	The Dark Knight Rises~†	1144337	Joseph Gordon-Levitt	http://www.imdb.2701.0		English	USA	2012.0	8.5
4	Doug Walker		Rob Walker	Documentary	Doug Walker	Star Wars: Episode VII - The Fo	8		http://www.imdb.com/title/tt5289954/?ref_=fn_tt_tt_1					7.1
5	Andrew Stanton	132.0	Samantha Morton	Action Adventure Sci-Fi	Daryl Sabara	John Carter~†	212204	Polly Walker	http://www.imdb.738.0		English	USA	2012.0	6.6
6	Sam Raimi	156.0	James Franco	Action Adventure Romance	J.K. Simmons	Spider-Man 3~†	383056	Kirsten Dunst	http://www.imdb.1902.0		English	USA	2007.0	6.2
7	Nathan Greno	100.0	Donna Murphy	Adventure Animation Comedy Family	Brad Garrett	Tangled~†	294810	M.C. Gainey	http://www.imdb.387.0		English	USA	2010.0	7.8
8	Joss Whedon	141.0	Robert Downey Jr.	Action Adventure Sci-Fi	Chris Hemsworth	Avengers: Age of Ultron~†	462669	Scarlett Johansson	http://www.imdb.1117.0		English	USA	2015.0	7.5
9	David Yates	153.0	Daniel Radcliffe	Adventure Family Fantasy Mystery	Alan Rickman	Harry Potter and the Half-Blood	321795	Rupert Grint	http://www.imdb.973.0		English	UK	2009.0	7.5
10	Zack Snyder	183.0	Lauren Cohan	Action Adventure Sci-Fi	Henry Cavill	Batman v Superman: Dawn of J	371639	Alan D. Purwin	http://www.imdb.3018.0		English	USA	2016.0	6.9
11	Bryan Singer	169.0	Marlon Brando	Action Adventure Sci-Fi	Kevin Spacey	Superman Returns~†	240396	Frank Langella	http://www.imdb.2367.0		English	USA	2006.0	6.1
12	Marc Forster	106.0	Mathieu Amalric	Action Adventure	Giancarlo Giannini	Quantum of Solace~†	330784	Rory Kinnear	http://www.imdb.1243.0		English	UK	2008.0	6.7
13	Gore Verbinski	151.0	Orlando Bloom	Action Adventure Fantasy	Johnny Depp	Pirates of the Caribbean: Dead	522040	Jack Davenport	http://www.imdb.1832.0		English	USA	2006.0	7.3
14	Gore Verbinski	150.0	Ruth Wilson	Action Adventure Western	Johnny Depp	The Lone Ranger~†	181792	Tom Wilkinson	http://www.imdb.711.0		English	USA	2013.0	6.5
15	Zack Snyder	143.0	Christopher Meloni	Action Adventure Fantasy Sci-Fi	Henry Cavill	Man of Steel~†	548573	Harry Lennix	http://www.imdb.2536.0		English	USA	2013.0	7.2
16	Andrew Adamson	150.0	Pierfrancesco Favino	Action Adventure Family Fantasy	Peter Dinklage	The Chronicles of Narnia: Princ	149922	Damiv"n Alcv"zar	http://www.imdb.438.0		English	USA	2008.0	6.6
17	Joss Whedon	173.0	Robert Downey Jr.	Action Adventure Sci-Fi	Chris Hemsworth	The Avengers~†	995415	Scarlett Johansson	http://www.imdb.1722.0		English	USA	2012.0	8.1
18	Rob Marshall	136.0	Sam Claflin	Action Adventure Fantasy	Johnny Depp	Pirates of the Caribbean: On St	370704	Stephen Graham	http://www.imdb.484.0		English	USA	2011.0	6.7
19	Barry Sonnenfeld	106.0	Michael Stuhlbarg	Action Adventure Comedy Family Fai	Will Smith	Men in Black 3~†	268154	Nicole Scherzinger	http://www.imdb.341.0		English	USA	2012.0	6.8
20	Peter Jackson	164.0	Adam Brown	Adventure Fantasy	Aidan Turner	The Hobbit: The Battle of the Fi	354228	James Nesbitt	http://www.imdb.802.0		English	New Zea	2014.0	7.5
21	Marc Webb	153.0	Andrew Garfield	Action Adventure Fantasy	Emma Stone	The Amazing Spider-Man~†	451803	Chris Zylka	http://www.imdb.1225.0		English	USA	2012.0	7.0
22	Ridley Scott	156.0	William Hurt	Action Adventure Drama History	Mark Addy	Robin Hood~†	211765	Scott Grimes	http://www.imdb.546.0		English	USA	2010.0	6.7
23	Peter Jackson	186.0	Adam Brown	Adventure Fantasy	Aidan Turner	The Hobbit: The Desolation of S	483540	James Nesbitt	http://www.imdb.951.0		English	USA	2013.0	7.9
24	Chris Weitz	113.0	Eva Green	Adventure Family Fantasy	Christopher Lee	The Golden Compass~†	149019	Kristin Scott Thomas	http://www.imdb.666.0		English	USA	2007.0	6.1
25	Peter Jackson	201.0	Thomas Kretschmann	Action Adventure Drama Romance	Naomi Watts	King Kong~†	316018	Evan Parke	http://www.imdb.2618.0		English	New Zea	2005.0	7.2
26	James Cameron	194.0	Kate Winslet	Drama Romance	Leonardo DiCaprio	Titanic~†	793059	Gloria Stuart	http://www.imdb.2528.0		English	USA	1997.0	7.7
27	Anthony Russo	147.0	Scarlett Johansson	Action Adventure Sci-Fi	Robert Downey Jr	Captain America: Civil War~†	272670	Chris Evans	http://www.imdb.1022.0		English	USA	2016.0	8.2
28	Peter Berg	131.0	Alexander Skarsgård	Action Adventure Sci-Fi Thriller	Liam Neeson	Battleship~†	202382	Tadanobu Asano	http://www.imdb.751.0		English	USA	2012.0	5.9

2. Data Cleaning:

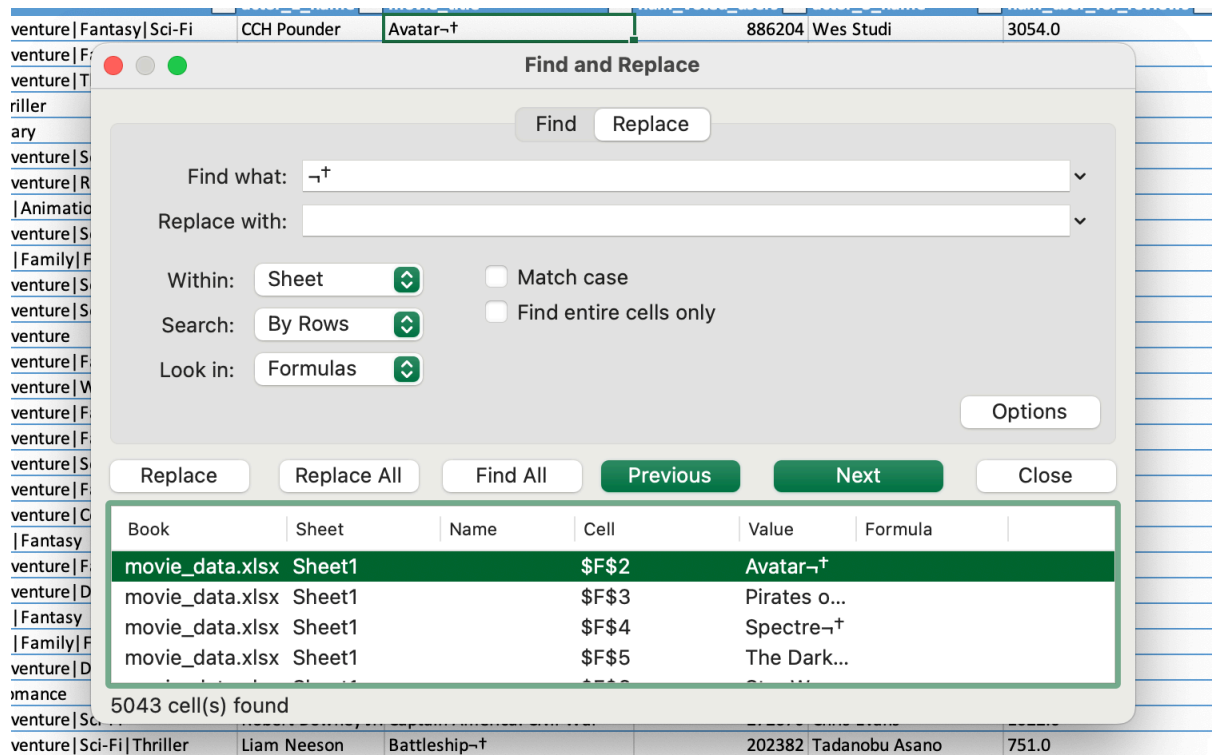
- Columns **index** and **movie_imdb_link** were deleted as they were deemed unnecessary for the analysis.
- Duplicates and rows with empty cells in critical fields such as **duration**, **movie_title**, **language**, **country**, **title_year**, and **imdb_score** were removed to maintain data integrity.

3. Formatting:

- The **title_year** column was converted into date format using the 'yyyy' type for consistency and ease of analysis.

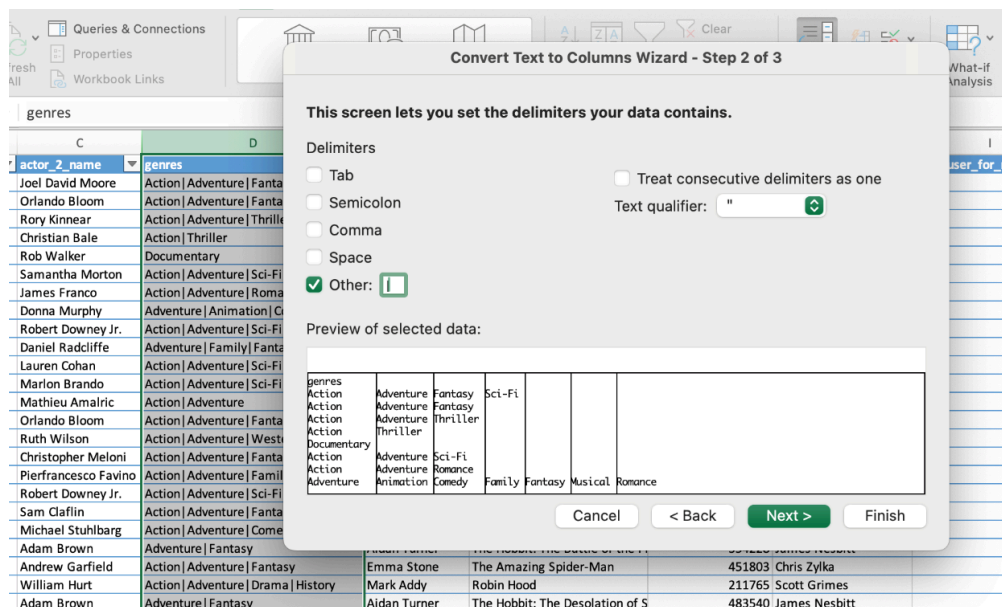
4. Cleaning Columns:

- Unnecessary signs and characters were removed from the dataset to ensure uniformity and clarity of data:



5. Column Transformation:

- The **genres** column was separated into multiple columns, each representing a single genre, to facilitate granular analysis.



6. Analyzing Movie Duration:

- Statistical measures such as MIN(), MAX(), AVERAGE(), MEDIAN(), and STDEV() were utilized to gain insights into movie durations.
- Movie durations were categorized into three groups (Short, Medium, Long) based on mean and standard deviation values to discern patterns in movie length.

Minimum duration:	7
Maximum duration:	330
Average duration:	108,0884897
Median duration:	104
Standard deviation:	22,61070484

- Movie durations were categorized into three groups (Short, Medium, Long) based on mean and standard deviation values to discern patterns in movie length using the formula:

=IF(B2 <= (\$Z\$9-\$Z\$11); "Short"; IF(B2 >= (\$Z\$9+\$Z\$11); "Long"; "Medium")),

where \$Z\$9 is mean and \$Z\$11 is standard deviation:

duration ▼	duration_group ▼
293,00	Long
94,00	Medium
80,00	Short
96,00	Medium
85,00	Short
79,00	Short
123,00	Medium
128,00	Medium

7. Categorizing IMDb Ratings:

- IMDb ratings were divided into five distinct groups (Very Low, Low, Medium, High, Very High) using a formula:

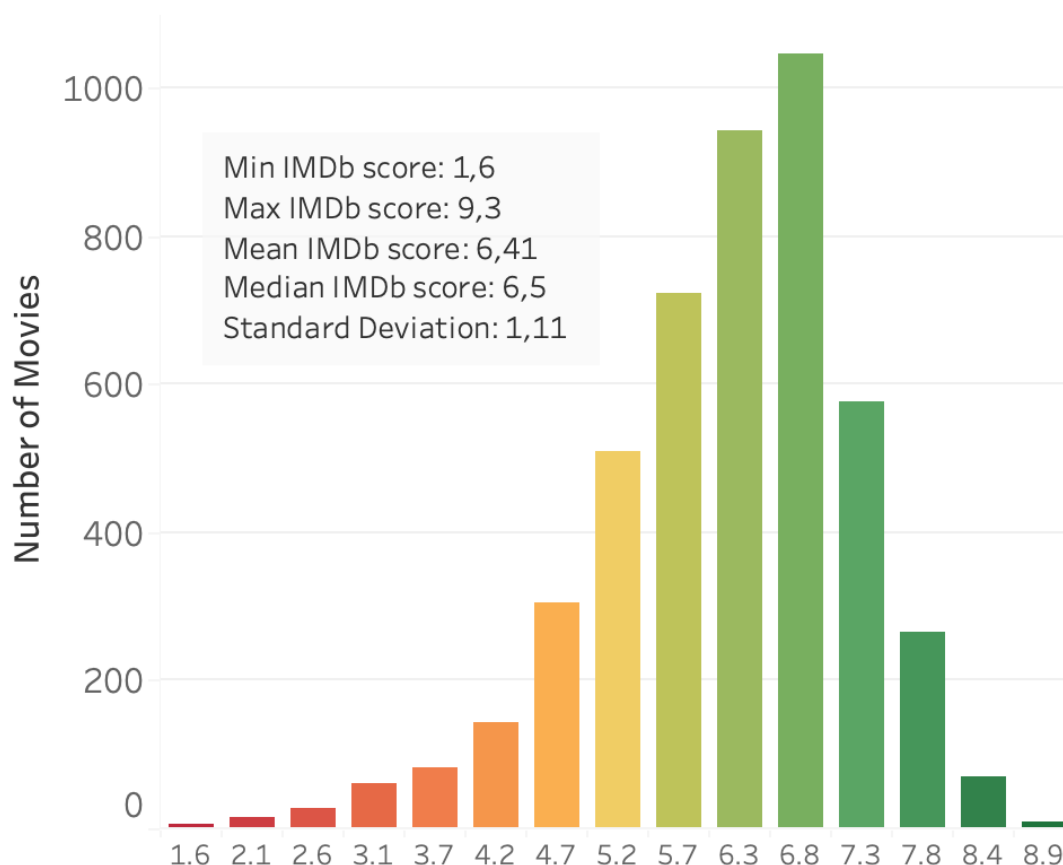
```
=IF(V2 < 3;"Very Low"; IF(V2 >= 3; IF(V2 < 4; "Low";  
IF(V2 >= 4; IF(V2 < 6; "Medium"; IF(V2 >= 6; IF(V2 < 8;  
"High"; "Very High"))))))))
```

imdb_score	imdb_score_groups
8,40	Very High
7,40	High
6,00	High
4,80	Medium
7,20	High
7,10	High
6,80	High
7,50	High
5,80	Medium
7,70	High
6,80	High

Analyse and Share

1. What is the distribution of the IMDb Score overall?

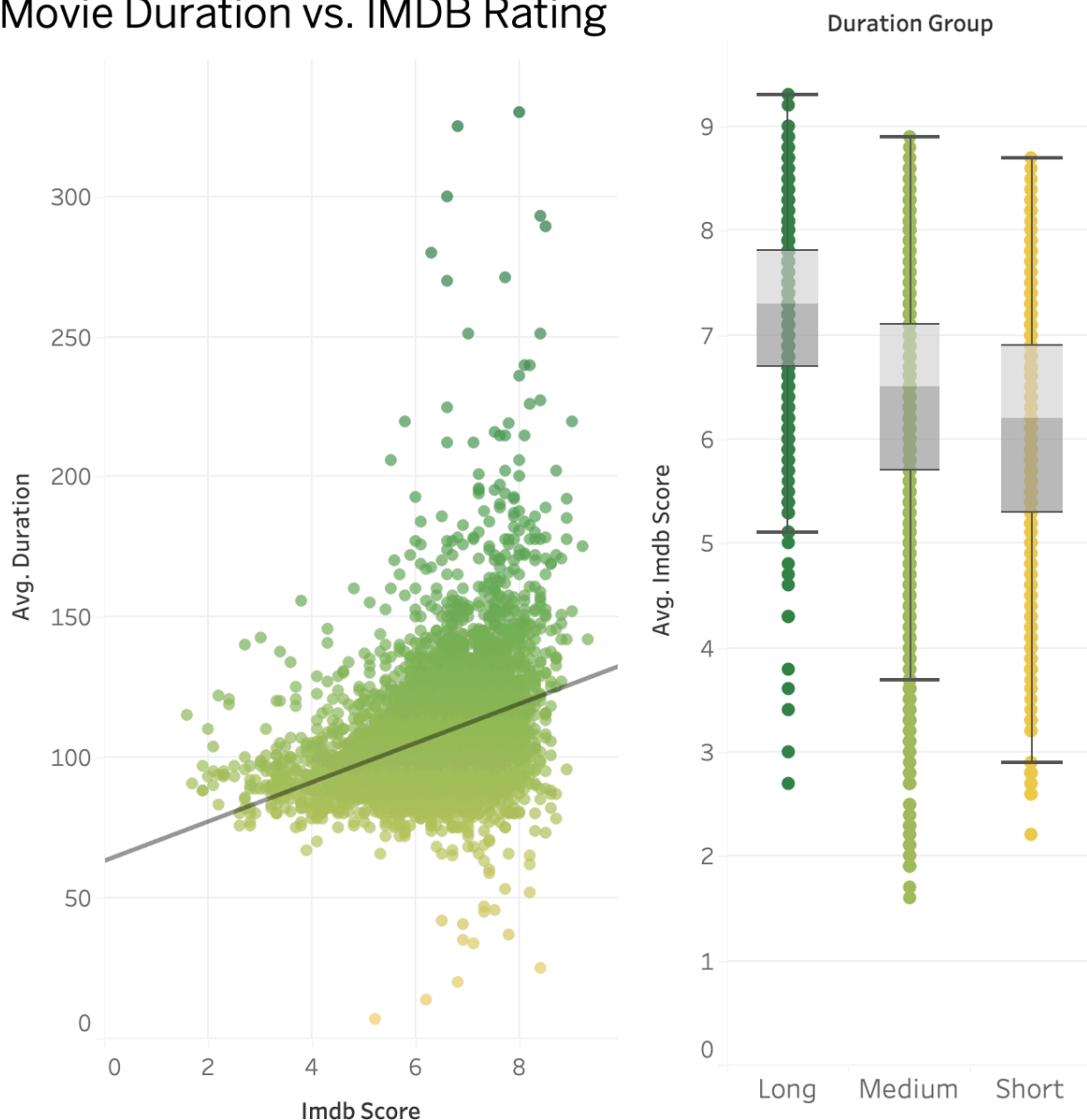
Distribution of the IMDB Rating



The IMDb scores in this dataset show **a range from 1.6 to 9.3**, with an **average score of 6.41**, slightly skewed to the left. The scores are relatively evenly distributed around the mean, with a standard deviation of 1.11 indicating **moderate variability in the ratings**.

2. Does the duration of a movie affect its IMDB rating?

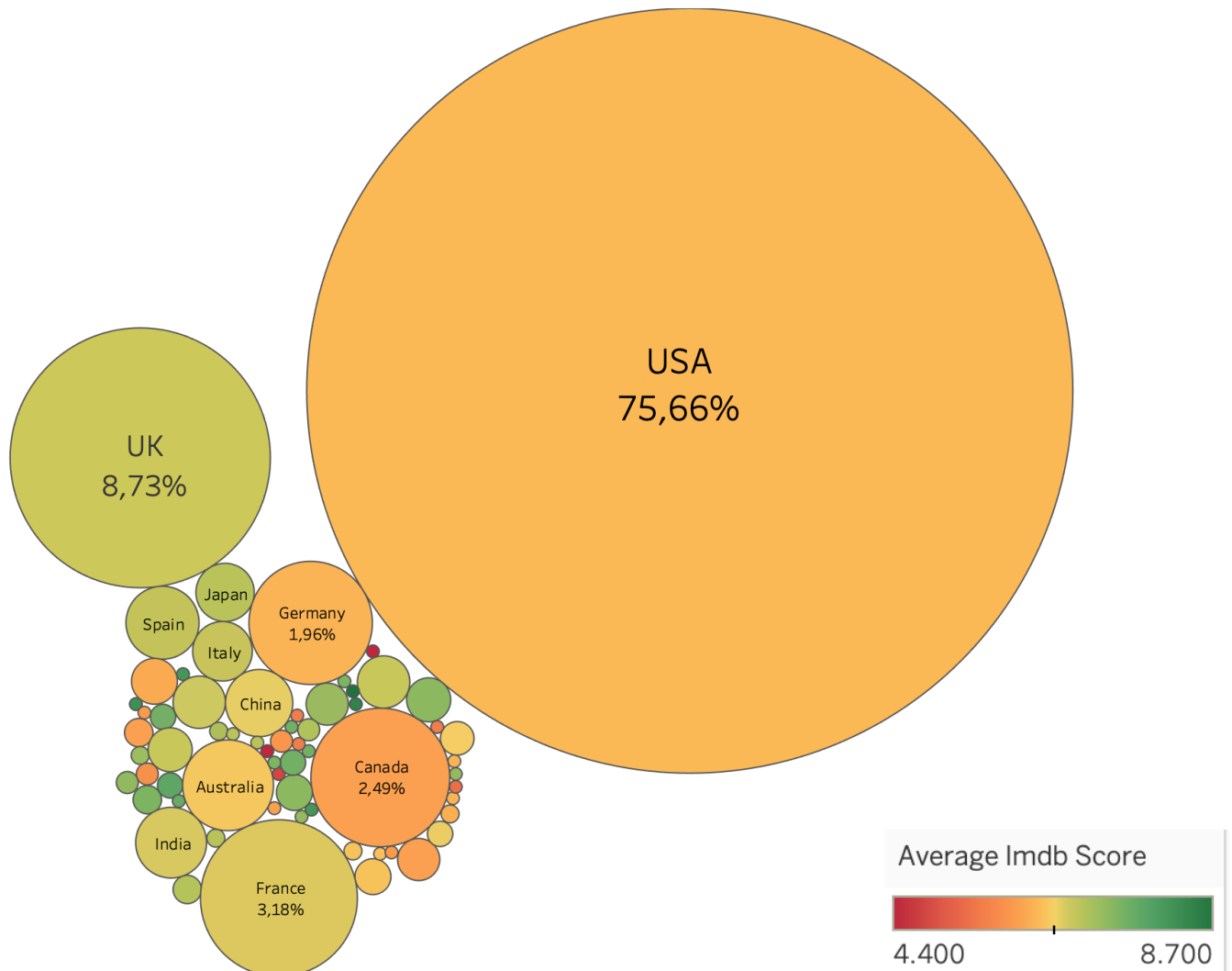
Movie Duration vs. IMDB Rating



The scatter plot and side-by-side boxplot both represent the **tendency of long duration movies to be more highly rated**. Although as the aligned boxplot, there are a lot of low-rated outliers in both Long and Medium groups.

3. Do movies in certain languages or countries tend to receive higher IMDB ratings?

Distribution of the Movies Among Countries



Top 10 Countries by IMDb Rating of Movies
(with 15+ Movies Released)

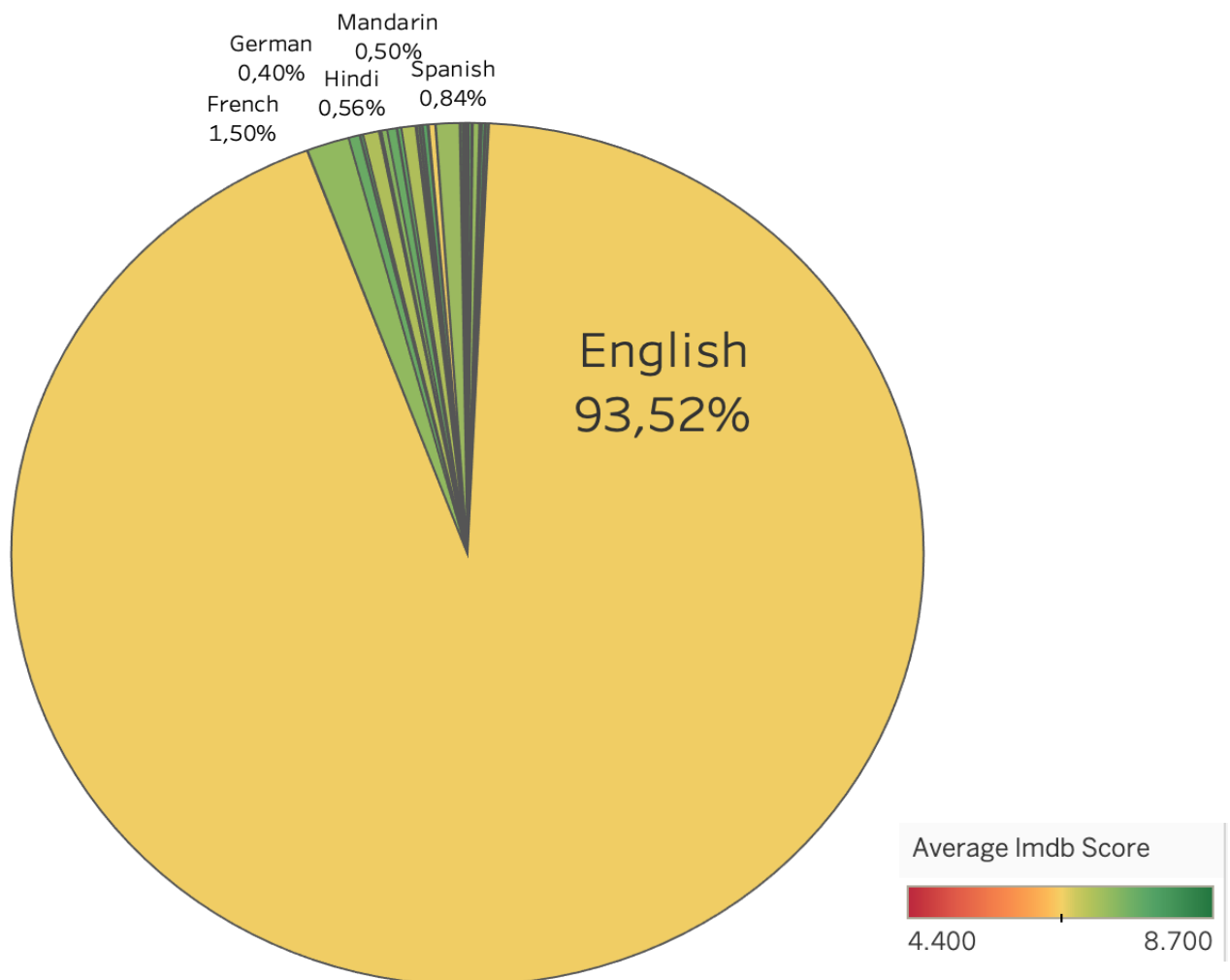


The dataset illustrates a significant disparity among the countries where movies were released, with **nearly 76% of movies originating from the USA**. In the bubble chart

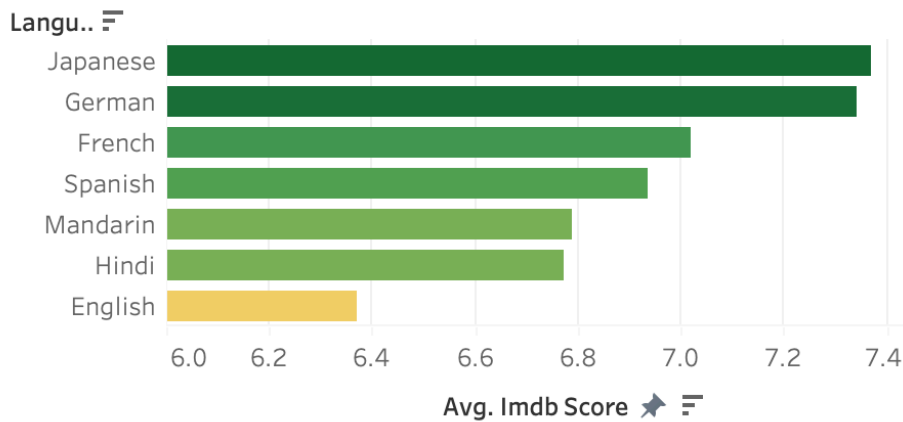
above, the color gradient corresponds to the average rating, where greener hues indicate higher ratings, while yellow and red hues denote medium and low scores, respectively.

On the bar chart depicting the top 10 countries by IMDb rating of released movies, I applied a filter to include only countries with at least 15 movies released. Notably, ***Japan, Spain, Italy, Mexico, the UK, etc, emerge as the countries with the highest-rated movies, while the USA does not appear in this list.***

Movie Language



Top 7 Movie Languages by IMDb Rating (with 15+ Movies Released)

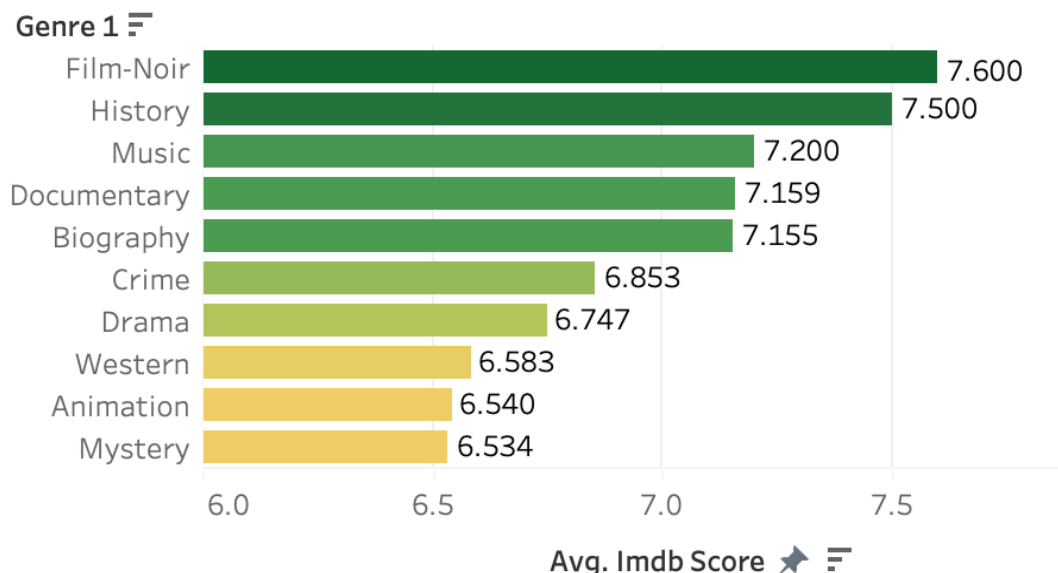


When analyzing the impact of movie languages on IMDb ratings, the following findings emerged: Despite **English being the most popular language** (with a popularity of almost 94%), **it does not consistently stand out** among languages with the highest average rating scores.

This suggests that beyond the origin of a movie or its language there are other factors that might play a significant role in determining movie ratings.

4. Are certain movie genres more likely to receive higher IMDb ratings than others?

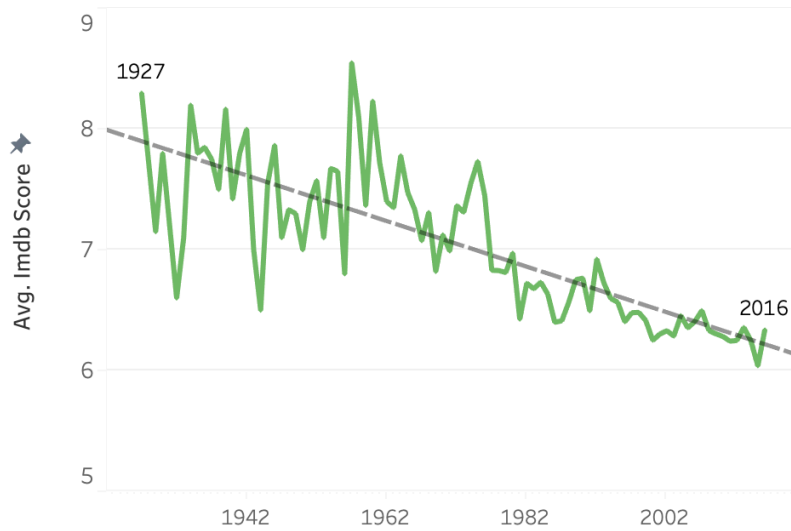
Top 10 Genres with the highest Rating



I compared the average IMDb ratings across primary genres. The bar chart revealed that genres predominantly associated with movies of the **highest ratings include Film-Noir, History, Music, Documentary, Biography, etc.**

5. How have IMDB ratings changed over the years?

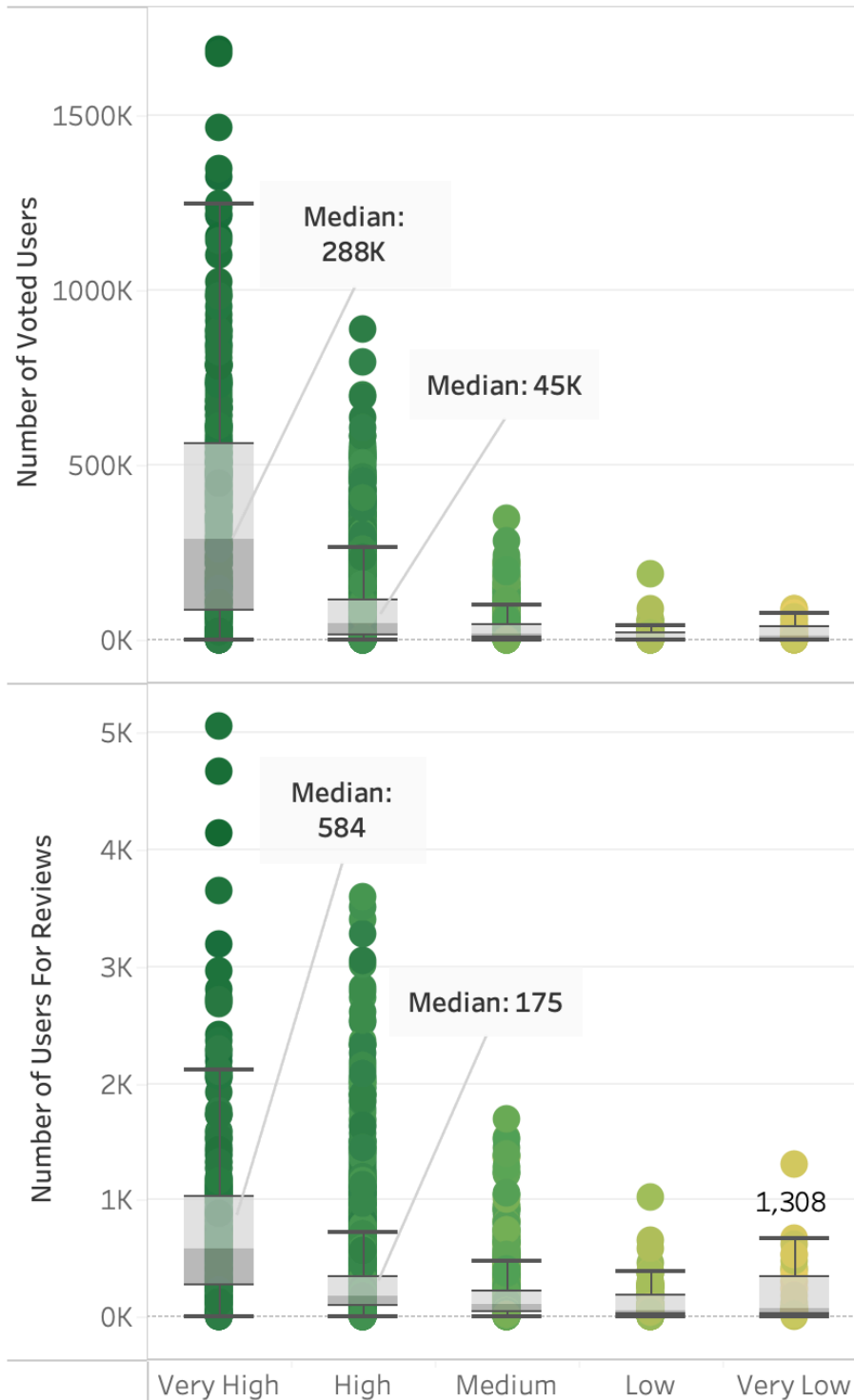
Trends of IMDb Rating Over Time



The trend line of the chart illustrates a ***gradual decrease of the average IMDb rating on movies over the years.***

6. Is there a relationship between the number of user reviews/votes and IMDB rating?

Number of User reviews/votes and IMDB Rating



The side-by-side box plots illustrate that movies categorized as **"Very High" IMDB scores tend to have a significant number of user reviews and/or votes.**

All charts and dashboards can be viewed interactively in my profile in [Tableau](#).

Act:

This comprehensive analysis of IMDb movie ratings sheds light on several key factors influencing a film's reception and success.

- **Overall Distribution of IMDb Scores:** The dataset showcases a diverse range of IMDb scores, with an average rating of 6.41. Scores are moderately variable around the mean.
- **Impact of Movie Duration:** Long-duration movies tend to exhibit higher IMDb ratings, although notable outliers in both long and medium categories suggest that duration alone does not dictate a film's success.
- **Influence of Language and Country:** While certain countries like Japan, Spain, Italy, and the UK consistently produce highly-rated films, the relationship between language and IMDb ratings is more complex, with English not always guaranteeing higher scores. This highlights the multifaceted nature of audience preferences and cultural influences.
- **Association of Genres with IMDb Ratings:** Genres such as Film-Noir, History, Music, Documentary, and Biography stand out for their propensity to garner high IMDb ratings, underscoring the importance of genre-specific appeal in audience reception.
- **Temporal Trends in IMDb Ratings:** Over the years, there has been a gradual decline in the average IMDb rating of movies, reflecting evolving audience tastes and perhaps changes in the cinematic landscape.
- **Relationship between User Engagement and IMDb Ratings:** Movies with "Very High" IMDb scores tend to attract a significant number of user reviews and votes, suggesting a correlation between audience engagement and perceived quality.

In conclusion, while factors such as duration, genre, language, and country of origin contribute to a film's IMDb rating, the nuanced interplay between these variables underscores the complexity of audience preferences and the dynamic nature of cinematic success. This analysis provides valuable insights for filmmakers and movie enthusiasts alike, informing decisions and enhancing understanding of the dynamics of the film industry.