

Data Handling & Business Intelligence: Exploring Film Performance and Audience Behaviour

Name: Diana Luhova

Student ID: 2428698

Module code: LCBB5014

Module name: Data Handling & Business Intelligence

Component 2

Coursework: Individual Report

Lecturer: Dr Xavier Matieni

Programme: BA in Business and Management

University of Wales Trinity Saint David

Table of Contents

1. Introduction	3
2. Data Cleaning and Preparation	3
3. Descriptive Statistics and Summary	5
4. Trend Analysis and Visualisation	7
5. Statistical Analysis SPSS	10
5.1 Independent Samples T-Test (Ratings by Language)	10
5.2 Correlation Analysis	11
5.3 Regression Analysis	12
5.4 Language Impact on Audience Engagement	14
5.5 Summary of Findings.....	16
6. Insights and Recommendations	17
7. References	18

Introduction

In order for organisations to make well-informed decisions based on data-driven insights, data analysis is essential to modern business intelligence. (Wang, 2014) In the film industry, factors influencing popularity and ratings can be found by researching audience behaviour and film performance. The analysis of a movie categorisation dataset including variables like genres, runtime, ratings, watches, likes, and fan counts is the main goal of this paper. The collection offers insightful information about viewer preferences and interaction trends. This report's primary goal is to examine the connections between important factors and pinpoint patterns that influence a movie's success. The investigation specifically looks into how movie ratings and popularity are influenced by elements including genre, language, runtime, and audience interaction. The study uses Excel and SPSS statistical techniques, trend analysis, descriptive statistics, and data cleansing to accomplish this. The results are intended to assist the media and entertainment sector in making data-driven decisions. (Field, 2024) It is anticipated that the analysis's findings would give film industry stakeholders useful information that will help them better audience targeting, optimise content strategies, and boost overall company performance. All things considered, the report highlights how crucial it is to combine data analysis and business intelligence technologies in order to turn unprocessed data into insightful and useful information.

Data Cleaning and Preparation

A crucial step in data analysis is data cleaning, which includes addressing missing or inconsistent data, eliminating duplicate entries, and finding and fixing errors. It guarantees that unstructured data is converted into a dependable, organised format that can be analysed. High-quality data is crucial in the context of business intelligence since insufficient or erroneous data can result in false conclusions and poor decision-making (Wang, 2014). Film titles, genres, runtime, ratings, watches, likes, and fan counts are just some of the variables that make up the dataset used in this study. Table 1 provides a sample of the dataset. These factors provide a solid foundation for analyzing audience behavior and identifying trends in movie popularity.

Data cleaning process followed in this case:

The raw data provided initially sorted based on the English Language movies, and using the sort and filter option of the excel, this process was performed at the beginning of the assessment of data.

Column	Film_title	Director	Average_rating	Genres	Runtime	Original_language	Description	Studios	Watches	List_appearances	Likes	Fans	Lowest_ratings	Medium_ratings	Highest_ratings	Total_ratings
1	Mad Max: Fury Road	George Miller	4.18	['Adventure', 'Science Fiction']	121	English	An apocalyptic road movie	Warner Bros. Pictures	2E+06	331681	927538	38000	12530	158356	511140	1682389
2	Suspiria	Dario Argento	3.93	['Horror']	99	English	An American horror film	Seda Spettacoli	618171	220094	219548	13000	2814	53427	88628	443757
3	Lost in Translation	Sofia Coppola	3.79	['Drama', 'Comedy']	102	English	Two lost New Yorkers	American Zoetrope	2E+06	254180	493248	38000	15167	155110	193717	1076949
6	Big Trouble in Little China	John Wood	3.72	['Comedy', 'Adventure']	100	English	Truck drivers	TAFT Entertainment	283873	65558	85788	3800	1250	30582	24938	183927
7	Star Wars: The Force Awakens	George Lucas	4.17	['Adventure', 'Action']	121	English	Princess Leia	Lucasfilm Ltd.	3E+06	391415	1E+06	30000	6041	161969	550396	1780203
8	The Thing	Carol Clover	4.25	['Mystery', 'Thriller']	105	English	In postwar Alaska	London Films Production	228162	101285	70932	3400	351	11549	39520	152601
9	The Big Sick	Michael Leew	3.76	['Drama', 'Comedy']	120	English	Pakistanis	FilmNation Entertainment	382201	58859	93489	504	1410	36987	18997	238309
10	Jodorowsky's Dune	Frank P. Johnson	3.96	['Documentary']	90	English	Shot in France	Snowfort Pictures	85388	16745	25616	188	209	6825	6696	56507
11	Get Out	Jordan Peele	4.15	['Horror', 'Mystery']	104	English	Chris Washington	Monkeypaw Productions	4E+06	454087	2E+06	33000	10251	272655	943393	3121586
12	Lawrence of Arabia	David Lean	4.38	['Adventure', 'War']	228	English	The story of T.E. Lawrence	Horizon Pictures	369742	151348	116793	10000	1080	16158	86313	242207
13	Free Fire	Ben Wheatley	3.16	['Action', 'Crime']	90	English	Set in Brazil	Rook Films	109849	29433	21524	83	1528	22924	1853	81191
14	Sing Street	John C. Macmillan	4.04	['Romance', 'Drama']	105	English	A boy grows up	Likely Story	390910	90763	145995	10000	846	27808	55425	273809
15	Up	Pete Docter	4.1	['Animation', 'Adventure']	96	English	Carl Fredricksen	Pixar	4E+06	263407	1E+06	9100	6354	230485	591788	2091274
16	Blade Runner 2049	Ridley Scott	4.1	['Thriller', 'Drama']	118	English	In the smoggy future	Shaw Brothers	2E+06	301778	573522	51000	6694	11183	309621	1095317
17	Annie Hall	Woody Allen	4.04	['Drama', 'Romance']	93	English	New York City	United Artists	454236	121871	119958	6400	2829	30173	60635	273663
18	The Truman Show	Peter Jackson	4.22	['Comedy', 'Drama']	103	English	Every scene is staged	Paramount Pictures	4E+06	361732	2E+06	71000	5764	231650	930042	2926289
19	Wonder Woman	Patty Jenkins	3.24	['Action', 'Fantasy']	141	English	An Amazon warrior	Atlas Entertainment	2E+06	147437	320485	1600	25306	239158	70337	946010
20	Plan 9 from Outer Space	Edward D. Wood Jr.	2.24	['Science Fiction']	79	English	In California	Reynolds Pictures	61992	20589	11196	121	6724	4949	2272	35588
21	Whiplash	Damien Chazelle	4.42	['Drama', 'Music']	107	English	Under the tutelage of a ruthless instructor	Bold Films	4E+06	481379	2E+06	154000	7390	146378	1E+06	2963756
22	Weird Science	John Wood	3.08	['Comedy', 'Science Fiction']	93	English	Two unemployed young men	Universal Pictures	141350	26516	26420	438	1942	21771	4897	76048
24	Army of Darkness	Sam Raimi	3.71	['Comedy', 'Horror']	81	English	Ash, a hero from the future	Renaissance Pictures	363502	108548	123314	3500	1877	39826	34201	251175
25	Obvious Child	Gillian Armstrong	3.52	['Comedy', 'Romance']	85	English	An immature woman	Sundial Pictures	86399	25356	19499	245	477	11350	3137	51165
26	Django Unchained	Quentin Tarantino	4.3	['Drama', 'Western']	165	English	With the help of a slave	The Weinstein Company	3E+06	373182	1E+06	53000	5412	146781	715708	2237306
27	The Losers	Joel Sothmann	3.66	['Comedy', 'Thriller']	97	English	When an ex-convict	Warner Bros. Pictures	394218	116234	127921	9900	1352	44940	43582	254656
28	Star Trek II: The Wrath of Khan	Nicholas Meyer	3.93	['Action', 'Thriller']	112	English	The starship Enterprise	Paramount Pictures	139409	38943	38125	1300	256	10089	13833	81593
29	Stand by Me	Rob Reiner	4.16	['Drama', 'Crime']	89	English	After leaving school	Columbia Pictures	1E+06	206422	425618	37000	1222	61649	217750	783704
30	Drive	Nicolas Winding Refn	3.92	['Thriller', 'Drama']	100	English	Driver is a loner	FilmDistrict	2E+06	252448	711650	30000	9648	186913	267739	1444060
31	The Lion King	Roger Allers	4.22	['Animation', 'Family']	89	English	Young lion Simba	Walt Disney Pictures	3E+06	232354	769362	22000	5252	143058	515263	1478357
32	It Follows	David Robert Mitchell	3.45	['Horror', 'Mystery']	101	English	When a girl is possessed	Two Flints	874585	169618	243095	3600	14353	124825	54459	667040
33	The Godfather	Francis Ford Coppola	3.8	['Adventure', 'Family']	114	English	A young man rises to power	Amblin Entertainment	854912	115427	223097	12000	3134	72477	86764	463218

Table 1: Summary of Data Cleaning Process

In terms of assessing the data set and drawing useful conclusions from it, there are still a number of inconsistencies and difficulties that remain after the initial phase of the data sorting process. The pivot table function has been further utilised for this aspect, resulting in the following output:

Row Labels	Sum of Fans	Sum of Watches	Sum of Average_rating	Sum of Runtime
Candy	44	8157	2.37	124
Free Fire	83	109849	3.16	90
Scarface	100	64361	3.83	93
Plan 9 from Outer Space	121	61992	2.24	79
Jodorowsky's Dune	188	85388	3.96	90
Obvious Child	245	86399	3.52	85
Weird Science	438	141350	3.08	93
The Big Sick	504	382201	3.76	120
Ant-Man	623	2261367	3.22	117
Star Trek II: The Wrath of Khan	1300	139409	3.93	112
Wonder Woman	1600	1721440	3.24	141
The Third Man	3400	228162	4.25	105
Army of Darkness	3500	363502	3.71	81
It Follows	3600	974585	3.45	101
Big Trouble in Little China	3800	283873	3.72	100
Sicario	6300	961576	3.98	122
Annie Hall	6400	454236	4.04	93
Ghostbusters	7500	1416463	3.75	107

PivotTable Fields
Choose fields to add to report:
Search
☒ Watches
☐ List_appearances
☐ Likes
☒ Fans
Drag fields between areas below:
Filters
Columns
Σ Values
Rows
Film_title
Σ Values
Sum of Fans

Table 2: Pivot Table Summary of Key Variables

Descriptive Statistics and Summary

Presentation of structured data and calculation

Row Labels	Sum of Fans	Sum of Watches	Average of rating	Sum of Runtime
Candy	44	8157	2.37	124
Free Fire	83	109849	3.16	90
Scarface	100	64361	3.83	93
Plan 9 from Outer Space	121	61992	2.24	79
Jodorowsky's Dune	188	85388	3.96	90
Obvious Child	245	86399	3.52	85
Weird Science	438	141350	3.08	93
The Big Sick	504	382201	3.76	120
Ant-Man	623	2261367	3.22	117
Star Trek II: The Wrath of Khan	1300	139409	3.93	112
Wonder Woman	1600	1721440	3.24	141
The Third Man	3400	228162	4.25	105
Army of Darkness	3500	363502	3.71	81
It Follows	3600	974585	3.45	101
Big Trouble in Little China	3800	283873	3.72	100
Sicario	6300	961576	3.98	122
Annie Hall	6400	454236	4.04	93
Ghostbusters	7500	1416463	3.75	107
Up	9100	3922213	4.1	96
The Lost Boys	9900	394218	3.66	97
Lawrence of Arabia	10000	369742	4.38	228
Sing Street	10000	390910	4.04	105
The Goonies	12000	854912	3.8	114
Suspiria	13000	618171	3.93	99
The Lion King	22000	3005454	4.22	89
Drive	30000	2091002	3.92	100
Star Wars	30000	3014320	4.17	121
Get Out	33000	4376009	4.15	104
Stand by Me	37000	1165736	4.16	89
Lost in Translation	38000	1596190	3.79	102
Mad Max: Fury Road	38000	2488721	4.18	121

Blade Runner	51000	1658845	4.1	118
Django Unchained	53000	3305105	4.3	165
The Truman Show	71000	4382819	4.22	103
Donnie Darko	79000	2293445	3.98	114
Whiplash	154000	3925603	4.42	107
Grand Total	739746	49597725	136.73	3925

The above data has been the base of further statistical calculation, such as mean, median, mode and standard deviation, which are major components of descriptive statistics and using these statistical tools, an overall judgement about the data set are made.

Statistical calculation

Mean	20548.5	1377714.583	3.798055556	109.0277778
Median	8300	908244	3.93	103.5
Mode	10000	#N/A	3.93	93
Standard deviation	31041.204	1378980.569	0.501134823	26.54751655

Mean

Mean value is the average of the data set and is one of the most essential elements of descriptive statistics. Through this value, the average value of rating and runtime of movies has been calculated. It has been found through this calculation that the average rating of English movies is 3.79 out of 5. In order to make a more detailed analysis, this average output is highly vital to consider (BBC, 2026).

Median and mode

These are equally important statistical metrics to deal with the statistical insights presented in the dataset. Respectively, these two outline the middle value of the entire data set and the value that is frequently found in the data set, respectively. Here for the average rating, both the mean and median values have been close to 4.

Standard deviation

The overall dispersion of data points from their mean are measured through the standard deviation function. Reliability of the data can be understood and tracked through the standard deviation function of a data set (Ravid, 2024). Here, in the case of the average ratings of the

films, the standard deviation has been 0.50, which highlights that this variable has clustered mostly around the mean.

Chart incorporation

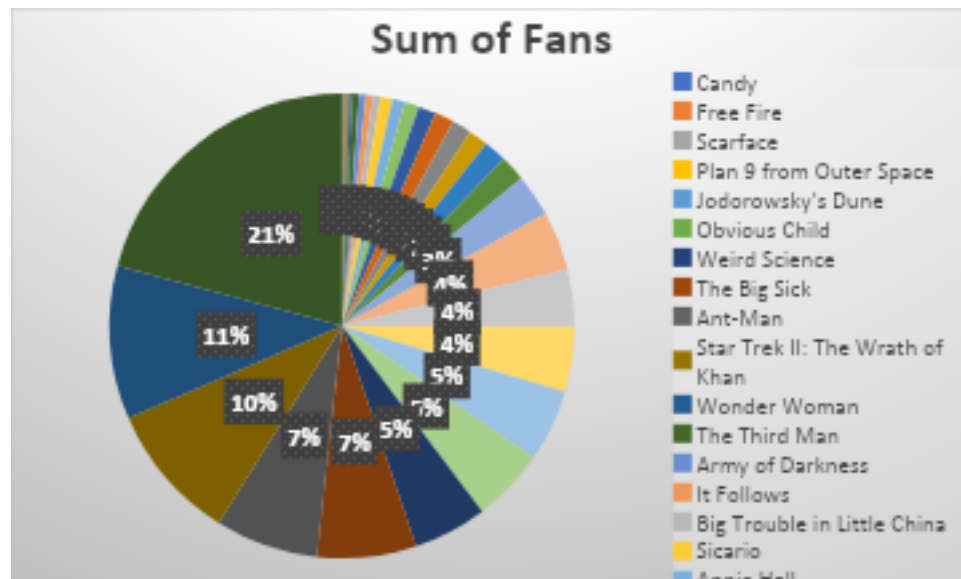


Figure 3: Pie chart showing the distribution of fan counts across films

The distribution of fan counts by movie is seen in Figure 3. A small number of films make up a sizable portion of all fans, suggesting unequal audience participation and concentrated appeal.

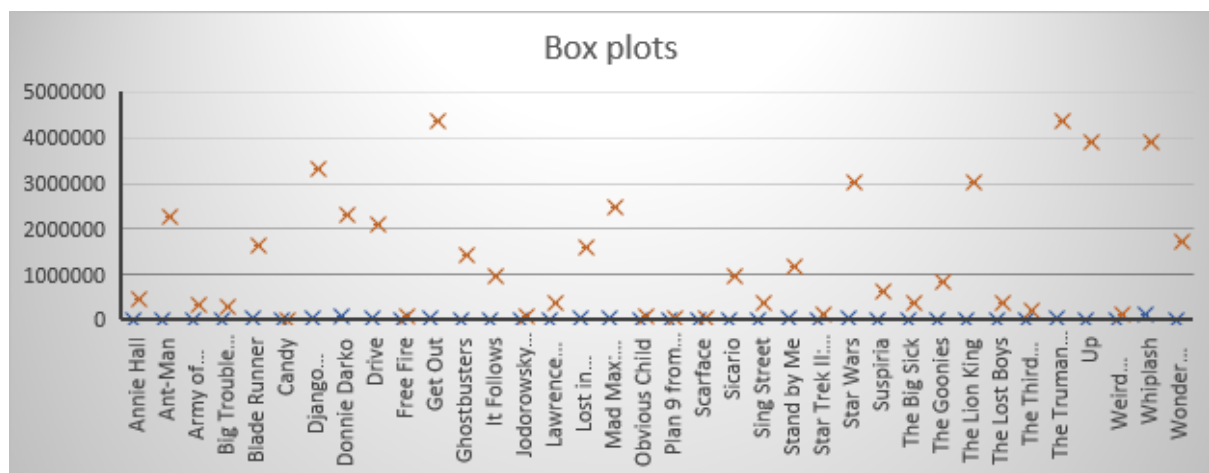


Figure 4: Box plot showing the dispersion between the number of fans and watches

A few films dominate audience interaction, as seen in Figure 4, which displays fluctuation in fan counts and viewing.

Trend Analysis and Visualisation

Scatter plots and other data visualisation tools are essential for determining correlations between variables and assisting with data-driven decision-making. Scatter plots are very useful for identifying patterns, and anomalies in big datasets, which helps analysts comprehend the direction and strength of links between variables (Data Visualisation Tools for Business Applications, 2024). By showing how engagement measures, like watches and fans, interact within the dataset, a scatter plot improves the analysis of audience behaviour in this context.

Column chart

The distribution of fan counts among the various films in the sample is shown in the column chart (Figure 5). It is evident that audience participation varies greatly, with certain films drawing noticeably more viewers than others. With more fans than any other movie, Candy in particular is the most well-known. On the other hand, the bulk of films have comparatively low fan counts, which suggests that most films only receive moderate amounts of audience attention. This distribution highlights the competitive nature of the film industry, where a small number of films attain high levels of audience involvement, and implies that film popularity is concentrated among a small number of titles.

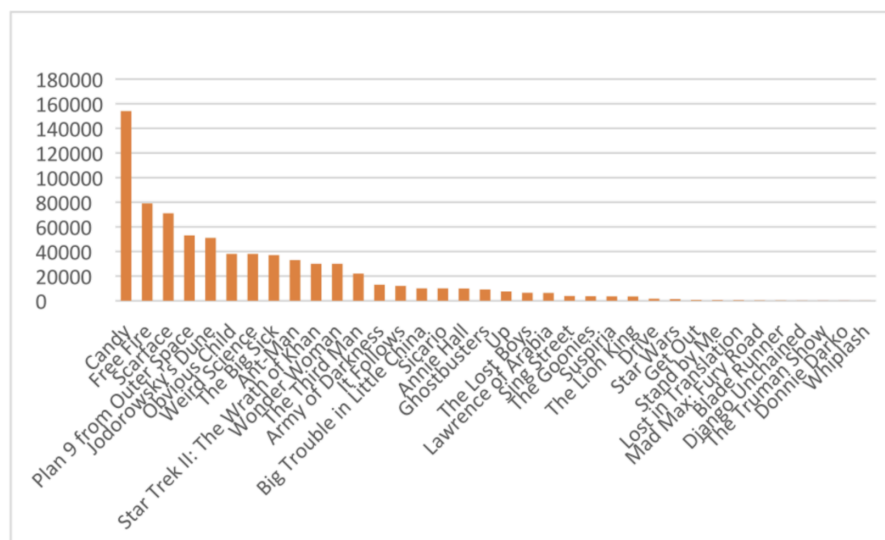


Figure 5: Column chart showing the fan numbers based on different films

Line chart

Trend measurement in a data set is most suitably presented through the visualisation of this data set. Combining multiple data points, a line is drawn through this particular data

visualisation tool and through the movement, the exact fluctuation of the chart is addressed, and data set is interpreted accordingly (Healy, 2024).

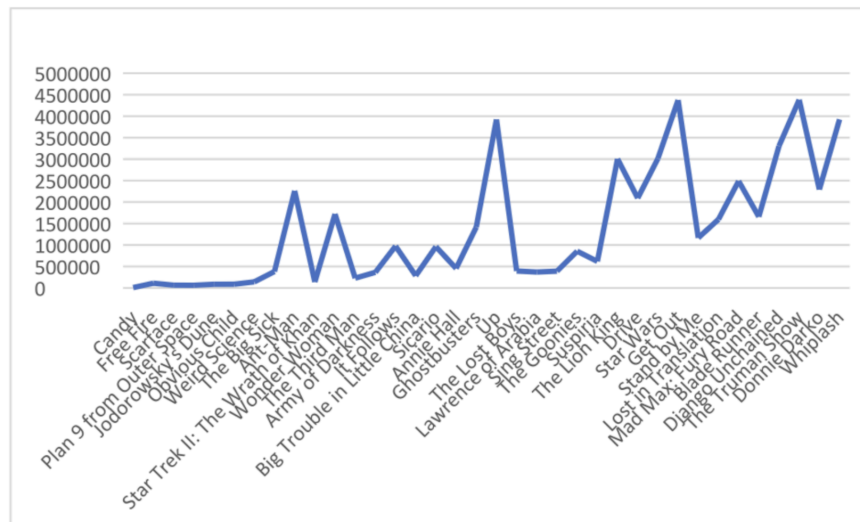


Figure 6: Line chart showing the sum of watches of different films

With over 4.3 million views, The Truman Show in particular has one of the biggest figures. In a similar vein, films with high watch counts, like Django Unchained, also contribute to the chart's overall heights. These films, however, do not always rank among the best in terms of fan counts, indicating that great fan devotion or sustained involvement are not always correlated with high viewership. This trend illustrates how films can draw sizable crowds without turning them into devoted followers, highlighting the complexity of audience behaviour.

Scatter plot

The scatter plot (Figure 7) shows how the number of fans and watches for various films relate to each other. Given that films with higher view counts typically have more fans, the visualisation indicates a positive correlation between these two factors. The link can be characterised as moderate rather than strong, though, because the data points are widely distributed and do not clearly form a linear pattern. This implies that while more viewership is linked to stronger fan involvement, audience behaviour is influenced by other factors as well. Additionally, a few of films exhibit large watch counts without correspondingly high fan bases, underscoring irregularities in audience participation. This supports the notion that other factors, like genre, ratings, or audience preferences, might also have a big impact on how popular a movie is.

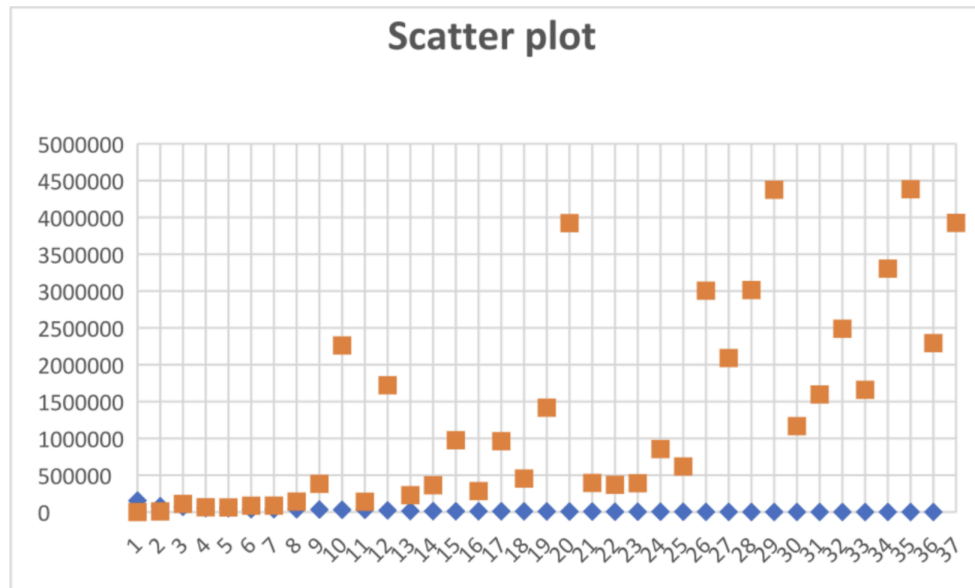


Figure 7: Scatter plot highlighting the watches and fans relation

Statistical Analysis (SPSS)

Independent Samples T-Test (Ratings by Language)

A t-test for independent samples compared the mean film rating of English films to non-English films to determine if there is a statistically significant difference. Figures 1 and 2 indicate that mean rating for English films ($M = 3.21$) is lower than that of non-English films ($M = 3.51$). The t-test results suggest that this difference is significant ($p = 0.000$), which suggests that language may be playing an important part in determining viewer ratings.

The Levene's test for equality of variances was also significant [$p < 0.05$] which suggests that the variances are not identical. As such, we interpreted the results assuming "equal variances not assumed". This suggests that on average, non-English rated films score higher, possibly due to genre-specific audience appeal, uniqueness or relatively higher quality of international films (Orji et al., 2023).

Group Statistics

	Language_Group	N	Mean	Std. Deviation	Std. Error Mean
Average_rating	1	8074	3.210954389	.5238498947	.0058299186
	2	1926	3.509920113	.4611764488	.0105084577

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
Average_rating	Equal variances assumed	27.710	.000	-23.009	9998
	Equal variances not assumed			-24.878	3219.603

Figure 8: Independent Samples T-Test for Average Rating by Language

Independent Samples Test

		t-test for Equality of Means		
		Sig. (2-tailed)	Mean Difference	Std. Error Difference
Average_rating	Equal variances assumed	.000	-.298965724	.0129932830
	Equal variances not assumed	.000	-.298965724	.0120173056

Independent Samples Test

		t-test for Equality of Means	
		95% Confidence Interval of the Difference	
		Lower	Upper
Average_rating	Equal variances assumed	-.324435174	-.273496274
	Equal variances not assumed	-.322528068	-.275403380

Figure 9: Independent Samples T-Test for Average Rating by Language

Correlation Analysis

Pearson correlation analysis to explore the potential association between the film's runtime, its number of watches, likes, fans and the average rating. The findings in Figure 3

demonstrate a very strong positive relationship between watches and likes ($r = 0.951$), suggesting that films with more viewers receive greater engagement. This implies that user views are a strong constructive factor to user engagement.

By contrast, there are weak positive relationships between runtime, watches ($r = 0.146$), and likes ($r = 0.129$), indicating that longer films may not have greater audience views or engagement. Also, there is a weak or moderate correlation between average rating and likes ($r = 0.290$) and fans ($r = 0.251$), suggesting that while likes and fans help a film achieve higher ratings, they do not necessarily explain the rating outcome (Ahmed & Abdulkareem, 2024).

		Correlations				
		Runtime	Watches	Likes	Fans	Average_rating
Runtime	Pearson Correlation	1	.146**	.129**	.088**	.187**
	Sig. (2-tailed)		.000	.000	.000	.000
	N	10000	10000	10000	10000	10000
Watches	Pearson Correlation	.146**	1	.951**	.671**	.249**
	Sig. (2-tailed)	.000		.000	.000	.000
	N	10000	10000	10000	10000	10000
Likes	Pearson Correlation	.129**	.951**	1	.799**	.290**
	Sig. (2-tailed)	.000	.000		.000	.000
	N	10000	10000	10000	10000	10000
Fans	Pearson Correlation	.088**	.671**	.799**	1	.251**
	Sig. (2-tailed)	.000	.000	.000		.000
	N	10000	10000	10000	10000	10000
Average_rating	Pearson Correlation	.187**	.249**	.290**	.251**	1
	Sig. (2-tailed)	.000	.000	.000	.000	
	N	10000	10000	10000	10000	10000

** . Correlation is significant at the 0.01 level (2-tailed).

Figure 10: Correlation Matrix of Runtime, Watches, Likes, Fans and Ratings

Regression Analysis

To examine the relationships among watches, likes, fans, and rating, and total running time, a multiple linear regression was performed. The model was significant ($F = 330.085$, $p = 0.000$) as shown in Figure 4, suggesting the model fits well.

The variables collectively account for about 11.7% of the variance in ratings ($R^2 = 0.117$), which suggests that although these factors play a significant role in predicting ratings, there are other factors (such as genre, plot, or director) that play a more significant role.

Among the predictors, likes emerged as the strongest positive predictor ($\beta = 0.616$, $p < 0.001$), followed by runtime ($\beta = 0.159$, $p < 0.001$). Watches had a negative impact on ratings ($\beta = -0.343$, $p < 0.001$), suggesting higher popularity is not associated with higher ratings. Fans were not statistically significant ($p = 0.148$), implying a weak direct impact on ratings (Szukits & Móricz, 2023).

But the high values of Variance Inflation Factor (VIF) for likes, $VIF = 20.475$, suggest that there is multicollinearity - that some of our independent variables are correlated with each other. This could affect the estimate for each predictor and should be considered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.342 ^a	.117	.116	.4942181520

a. Predictors: (Constant), Runtime, Fans, Watches, Likes

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	322.495	4	80.624	330.085	.000 ^b
	Residual	2441.295	9995	.244		
	Total	2763.790	9999			

a. Dependent Variable: Average_rating

b. Predictors: (Constant), Runtime, Fans, Watches, Likes

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics Tolerance
		B	Std. Error	Beta			
1	(Constant)	3.018	.014		211.144	.000	
	Watches	-3.933E-7	.000	-.343	-9.930	.000	.074
	Fans	-1.446E-6	.000	-.026	-1.448	.148	.279
	Likes	2.138E-6	.000	.616	14.487	.000	.049
	Runtime	.002	.000	.159	16.743	.000	.978

Coefficients^a

Model		Collinearity Statistics VIF
1	(Constant)	
	Watches	13.505
	Fans	3.584
	Likes	20.475
	Runtime	1.023

a. Dependent Variable: Average_rating

Figure 11: Regression Model Summary and Coefficients

Language Impact on Audience Engagement

Additional independent samples t-tests were performed to discern the effect of language on audience engagement and views (watches), likes, and fan counts. As shown in Figure 5 and 6, English films show significantly more engagement than non-English films ($p < 0.05$).

For example, English films had significantly higher average watches ($M = 196,529$) than non-English films ($M = 60,418$). Likes and fans are also greater for English films. These findings show that although non-English films score higher ratings, English films are more popular in terms of global and audience engagement (Orji et al., 2023).

Group Statistics

	Language_Group	N	Mean	Std. Deviation	Std. Error Mean
Watches	1	8074	196529.91	495460.679	5513.975
	2	1926	60418.35	218298.310	4974.188
Likes	1	8074	49385.39	161538.646	1797.761
	2	1926	20046.86	95600.592	2178.374
Fans	1	8074	1772.09	9945.056	110.678
	2	1926	1151.43	6385.972	145.512

Figure 12: T-Test Results for Watches, Likes and Fans by Language

Independent Samples Test					
		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
Watches	Equal variances assumed	327.432	.000	11.786	9998
	Equal variances not assumed			18.329	7031.072
Likes	Equal variances assumed	142.787	.000	7.657	9998
	Equal variances not assumed			10.388	4898.316
Fans	Equal variances assumed	19.441	.000	2.613	9998
	Equal variances not assumed			3.395	4442.127

Independent Samples Test					
		t-test for Equality of Means			
		Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence ... Lower
Watches	Equal variances assumed	.000	136111.561	11548.443	113474.288
	Equal variances not assumed	.000	136111.561	7426.067	121554.230
Likes	Equal variances assumed	.000	29338.531	3831.619	21827.788
	Equal variances not assumed	.000	29338.531	2824.404	23801.433
Fans	Equal variances assumed	.009	620.664	237.498	155.120
	Equal variances not assumed	.001	620.664	182.821	262.244

Figure 13: T-Test Results for Watches, Likes and Fans by Language

Summary of Findings

In summary, the data analysis using SPSS reveals that language is a key predictor of engagement and ratings, there are high correlations between the various engagement measures (such as watches and likes), and ratings are positively influenced by a range of factors. Engagement is a significant element, but ratings depend on a range of other variables, emphasising the need for full data analysis to inform business intelligence and decisions for film.

Insights and Recommendations

Important insights into audience behaviour and movie performance are revealed by the dataset analysis. The majority of films are rated as fairly positive, according to descriptive data and visualisations, which show that movie ratings are generally stable. However, audience engagement, as determined by watches and fan counts, differs greatly between films, suggesting that popularity is concentrated within a small number of really popular films. This demonstrates how crucial data analytics are for spotting trends and assisting with data-driven decision-making (Sharda, Delen, and Turban, 2018). Additionally, the results demonstrate that substantial fan interaction does not always follow from high viewership. Attracting viewers does not always translate into long-term audience devotion, as evidenced by several films that have large watch counts but very low fan numbers. This lends credence to the notion that a variety of behavioural elements, as opposed to a single statistic, drive engagement (Provost and Fawcett, 2013). It is advised that movie studios and streaming services concentrate on tactics that improve viewer engagement, like tailored suggestions and focused advertising, in light of these findings. Furthermore, content accessibility and audience reach can be enhanced by examining genre and linguistic trends (Wang, 2014). It is advised that movie studios and streaming services make investments in tactics like community-building programs, social media interaction, and exclusive material that turn spectators into engaged fans. Focusing on long-term engagement can assist boost fan retention and strengthen the overall success of films, as high viewership does not always translate into strong audience loyalty.

References

Wang, J. (2014). Encyclopedia of business analytics and optimization. Hershey, PA: Business Science Reference.

Field, A. (2024). Discovering Statistics Using IBM SPSS Statistics. 6th ed. Los Angeles: Sage Publications.

Kelleher, J. D., Mac Namee, B., D'Arcy, A. (2020). Fundamentals of Machine Learning for Predictive Data Analytics, Second Edition: Algorithms, Worked Examples, and Case Studies. United Kingdom: MIT Press.

Encyclopedia of Business Analytics and Optimization. (2014). United States of America: IGI Global.

Wang, J. (2014b). Encyclopedia of business analytics and optimization. Hershey, PA: Business Science Reference.

Microsoft, (2026a). Create a PivotTable to analyze worksheet data. [Online] Available at: <https://support.microsoft.com/en-gb/office/create-a-pivottable-to-analyze-worksheet-data-a9a84538-bfe9-40a9-a8e9-f99134456576> [Accessed 17 April 2026]

Microsoft, (2026b). Present your data in a column chart. [Online] Available at: <https://support.microsoft.com/en-gb/office/present-your-data-in-a-column-chart-d89050ba-e6b6-47de-b090-e9ab353c4c00> [Accessed 17 April 2026]

BBC, (2026). Mean, median, mode and range. [Online] Available at: <https://www.bbc.co.uk/bitesize/articles/z3t4g7h> [Accessed 17 April 2026]

Ravid, R., (2024). Practical statistics for educators. Rowman & Littlefield.

Healy, K., (2024). Data visualization: a practical introduction. Princeton University Press.

Levine, G. (2013). A Guide to SPSS for Analysis of Variance. United Kingdom: Taylor & Francis.

Sharda, R., Delen, D., Turban, E. (2018). Business Intelligence, Analytics, and Data Science: A Managerial Perspective. United Kingdom: Pearson.

Provost, F., Fawcett, T. (2013). Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. United Kingdom: O'Reilly Media.

Jain, V.K. (n.d.). Data Science and Analytics (with Python, R and SPSS Programming). KHANNA PUBLISHING HOUSE.

Orji, U, Obianuju, E, Ezema, M, Ugwuishiwu, C, Ukwandu, E & Agomuo, U 2023, Using data analytics to derive business intelligence: A case study, <https://arxiv.org/abs/2305.19021>

Ahmed, D-A & Abdulkareem, AM 2024, “Big Data Analytics in the Entertainment Industry: Audience Behavior Analysis, Content Recommendation, and...,” ResearchGate, https://www.researchgate.net/publication/379154685_Big_Data_Analytics_in_the_Entertainment_Industry_Audience_Behavior_Analysis_Content_Recommendation_and_Revenue_Maximization

Szukits, Á & Móricz, P 2023, “Towards data-driven decision making: the role of analytical culture and centralization efforts,” Review of Managerial Science, 18(10):2849–2887, <https://link.springer.com/article/10.1007/s11846-023-00694-1>