

Project Summary: Analysing Movie Ratings, Genres, Languages, Directors, and Financial Success

This project involves a comprehensive analysis of various factors that influence movie ratings, including genres, languages, directors, and financial metrics such as budgets and gross earnings. The analysis was performed using a dataset of movies, with specific focus areas including genre distribution, the impact of language on ratings, director performance, and the correlation between movie budgets and financial success.

Key Findings:

1. Genre Analysis:

- **Animation** emerged as the genre with the highest average ratings, indicating a strong and consistent audience appeal.
- **Comedy** and **Fantasy** showed higher variability in ratings, reflecting the diverse reception these genres receive from different audiences.
- **Horror** had the lowest average and median ratings, suggesting it is less favorably received compared to other genres.

2. Language Impact:

- **English** was the most common language in the dataset, but its movies exhibited a wide range of ratings, suggesting variability in quality.
- **French-language** films had higher average ratings with lower variability, indicating a more consistent positive reception.
- **Hindi** and **Spanish** movies showed moderate ratings with some variability, reflecting a mix of audience perceptions.

3. Director Performance:

- Directors like **Peter Medak** and **Patricia Rozema** were identified as top performers, consistently producing highly rated films.
- **Zoran Lisinac** and **Zal Batmanglij** were noted as lower performers, with their films receiving lower average ratings.

4. Budget and Earnings Correlation:

- The analysis revealed a weak correlation between movie budgets and gross earnings, indicating that higher budgets do not necessarily lead to higher financial returns.
- **"Avatar"** was identified as the movie with the highest profit margin, despite its large budget, highlighting the importance of factors like marketing and audience appeal.

5. Movie Duration:

- A positive but weak correlation was found between movie duration and IMDB scores, suggesting that longer movies might slightly tend to receive higher ratings, but other factors are more significant.

A: Task:

Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.

At first i splitted the genres column into 7 columns by splitting with | , one value inside genres with multiple | get splitted and i named each splitted column as genre_1, genre_2 and so on

.....

And then I removed the duplicates, selecting all the columns as I wanted to remove only duplicate rows and not a particular value and I got 45 duplicates which I removed.

Here using countif function i have calculated the count of each genre in all the splitted genres.

=COUNTIF(K:K,"Family")+COUNTIF(L:L,"Family")+COUNTIF(M:M,"Family")+COUNTIF(N:N,"Family")+COUNTIF(O:O,"Family")+COUNTIF(P:P,"Family")+COUNTIF(Q:Q,"Family")+COUNTIF(R:R,"Family")											
H	I	J	K	L	M	N	O	P	Q	R	S
genres	count of genres	most common genres	genres	genre_1	genre_2	genre_3	genre_4	genre_5	genre_6	genre_7	actor_1_name
Comedy	1862		Action	Adventure	Fantasy	Sci-Fi					CCH Pounder
Thriller	1396		Action	Adventure	Fantasy						Johnny Depp
Drama	1143		Action	Adventure	Thriller						Christoph Waltz
Documentary	1143		Action	Thriller							Tom Hardy
ACTION	121		Documentary								Doug Walker
Romance	1098		Action	Adventure	Sci-Fi						Daryl Sabara
Adventure	1143		Action	Adventure	Romance						J.K. Simmons
Crime	883		Adventure	Animation	Comedy	Family	Fantasy	Musical	Romance		Brad Garrett
Sci-Fi	611		Action	Adventure	Sci-Fi						Chris Hemsworth
Fantasy	604		Adventure	Family	Fantasy	Mystery					Alan Rickman
Family	544		Action	Adventure	Sci-Fi						Henry Cavill
Biography	292		Action	Adventure	Sci-Fi						Kevin Spacey
Animation	242		Action	Adventure							Giancarlo Giannini
History	205		Action	Adventure	Fantasy						Johnny Depp
Musical	132		Action	Adventure	Western						Johnny Depp
			Action	Adventure	Fantasy	Sci-Fi					Henry Cavill
			Action	Adventure	Family	Fantasy					Peter Dinklage
			Action	Adventure	Sci-Fi						Chris Hemsworth
			Action	Adventure	Fantasy						Johnny Depp
			Action	Adventure	Comedy	Family	Fantasy	Sci-Fi			Will Smith
			Adventure	Fantasy							Aidan Turner
			Action	Adventure	Fantasy						Emma Stone
			Action	Adventure	Drama	History					Mark Addy

Here we can see comedy has the highest count and musical has the lowest among the mentioned genres in h column

C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
SD	VARIANCE	MIN	MAX	MEDIAN	AVERAGE	HORROR	FANTASY	COMEDY	ANIMATION	ADVENTURE	ACTION	imdb_score	genres	genres_1	genres_2
1.1	1.277959079	2.2	8.7	5.9	5.843539823		7.1				7.9	7.9	7.9 Action	Adventure	Fantasy
											7.1	7.1	7.1 Action	Adventure	Fantasy
											6.8	6.8	6.8 Action	Adventure	Thriller
												8.5	8.5 Action	Thriller	
1.2	1.347191607	1.7	8.9	6.4	6.30704918								7.1 Documentary		
											6.6	6.6	6.6 Action	Adventure	Sci-Fi
											6.2	6.2	6.2 Action	Adventure	Romance
1.2	1.189656701	1.7	9.5	6.3	6.195245726		7.8	7.8	7.8		7.8		7.8 Adventure	Animation	Comedy
											7.5	7.5	7.5 Action	Adventure	Sci-Fi
											7.5		7.5 Adventure	Family	Fantasy
1.1	1.298676314	1.7	8.6	6.7	6.576033058						6.9	6.9	6.9 Action	Adventure	Sci-Fi
											6.1	6.1	6.1 Action	Adventure	Sci-Fi
											6.7	6.7	6.7 Action	Adventure	
											7.3	7.3	7.3 Action	Adventure	Fantasy
1.1	1.279604703	1.9	8.9	6.6	6.441170098						6.5	6.5	6.5 Action	Adventure	Western
											7.2	7.2	7.2 Action	Adventure	Fantasy
											6.6	6.6	6.6 Action	Adventure	Family
1.1	1.25179235	1.7	9.1	6.3	6.239895924						8.1	8.1	8.1 Action	Adventure	Sci-Fi
											6.7	6.7	6.7 Action	Adventure	Fantasy
											6.8	6.8	6.8 Action	Adventure	Comedy
											7.5		7.5 Adventure	Fantasy	
											7		7 Action	Adventure	Fantasy
											6.7	6.7	6.7 Action	Adventure	Drama
											7.9		7.9 Adventure	Fantasy	
											6.1		6.1 Adventure	Family	Fantasy
											7.2	7.2	7.2 Action	Adventure	Drama
													7.7 Drama	Romance	
											8.2	8.2	8.2 Action	Adventure	Sci-Fi

1. I have separated the genres by | named as gener_1, genre_2
2. I have taken few genre for consideration and a created column for those mentioned on image i have used formula to extract rating for specific genre [=IF(OR(ISNUMBER(SEARCH("Action",P2)),ISNUMBER(SEARCH("Action",Q2)),ISNUMBER(SEARCH("Action",R2)),ISNUMBER(SEARCH("Action",S2)),ISNUMBER(SEARCH("Action",T2)),ISNUMBER(SEARCH("Action",U2)),ISNUMBER(SEARCH("Action",V2)),ISNUMBER(SEARCH("Action",W2))),O2,"")] this checks on all the splited columns and find the genre mentioned in any one of the splited genres if found puts the rating value AND if not found empty
3. I WOULD USED PIVOT TABLE AND CONSIDERED ALL THE GENRES TO FIND THE COUNT AND STATS IF I HAVE CONSIDERED A SINGLE GENRE OUT OF ALL SPLITTED GENRES

Summary Statistics for Each Genre (i have taken few genres for consideration)

Let's start by summarising the provided statistics for each genre:

1. Horror

- **Standard Deviation (SD):** 1.130
- **Variance:** 1.278
- **Minimum (MIN):** 2.2
- **Maximum (MAX):** 8.7
- **Median:** 5.9
- **Average:** 5.844

2. Fantasy

- **Standard Deviation (SD):** 1.161
- **Variance:** 1.347
- **Minimum (MIN):** 1.7

- **Maximum (MAX):** 8.9
- **Median:** 6.4
- **Average:** 6.307

3. Comedy

- **Standard Deviation (SD):** 1.161
- **Variance:** 1.190
- **Minimum (MIN):** 1.7
- **Maximum (MAX):** 9.5
- **Median:** 6.3
- **Average:** 6.195

4. Animation

- **Standard Deviation (SD):** 1.140
- **Variance:** 1.299
- **Minimum (MIN):** 1.7
- **Maximum (MAX):** 8.6
- **Median:** 6.7
- **Average:** 6.576

5. Adventure

- **Standard Deviation (SD):** 1.131
- **Variance:** 1.280
- **Minimum (MIN):** 1.9
- **Maximum (MAX):** 8.9
- **Median:** 6.6
- **Average:** 6.441

6. Action

- **Standard Deviation (SD):** 1.119
- **Variance:** 1.252
- **Minimum (MIN):** 1.7
- **Maximum (MAX):** 9.1
- **Median:** 6.3
- **Average:** 6.240

Comparative Analysis

1. Variability (Standard Deviation and Variance)

- **Highest Standard Deviation:** Fantasy and Comedy (1.161)
- **Lowest Standard Deviation:** Action (1.119)
Observation: Fantasy and Comedy have the highest standard deviation, indicating greater variability in ratings compared to other genres. Action has the lowest, indicating more consistent ratings.

2. Range (Minimum and Maximum)

- **Highest Maximum Rating:** Comedy (9.5)
- **Lowest Minimum Rating:** Fantasy (1.7)
- **Highest Minimum Rating:** Adventure (1.9)
- **Lowest Maximum Rating:** Horror (8.7)

Observation: Comedy has the highest maximum rating, suggesting that it can achieve the highest ratings among these genres. Fantasy and Animation have the lowest minimum ratings, indicating that these genres have lower bounds for ratings.

3. Central Tendency (Median and Average)

- **Highest Median Rating:** Animation (6.7)
- **Lowest Median Rating:** Horror (5.9)
- **Highest Average Rating:** Animation (6.576)
- **Lowest Average Rating:** Horror (5.844)

Observation: Animation has both the highest median and average ratings, suggesting it is generally rated higher compared to other genres. Horror has the lowest median and average, indicating generally lower ratings for this genre.

4. Impact on Movie Ratings

- **Animation** has the highest average and median ratings, indicating it is perceived most positively.
- **Horror** has the lowest average and median ratings, suggesting it may not perform as well in terms of ratings compared to other genres.
- **Fantasy** and **Comedy** show higher variability, which might reflect a wider range of quality or audience reception within these genres.
- **Action** has consistent ratings but not the highest average or median, indicating reliability in ratings but not necessarily the highest perceived quality.

Conclusion

From the statistics, we can infer the following:

- **Animation** tends to receive the highest ratings overall.
- **Horror** has the lowest ratings, and its variability suggests a less consistent reception.
- **Fantasy** and **Comedy** have a broader range of ratings, reflecting a varied audience perception.
- **Action** has more consistent ratings, which may indicate steady performance across different movies in the genre.

B. Movie Duration Analysis:

Analyze the distribution of movie durations and its impact on the IMDB score. Task: Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score. Hint: Calculate descriptive statistics such as mean, median, and standard deviation for movie durations. Use Excel's functions like AVERAGE, MEDIAN, and STDEV.

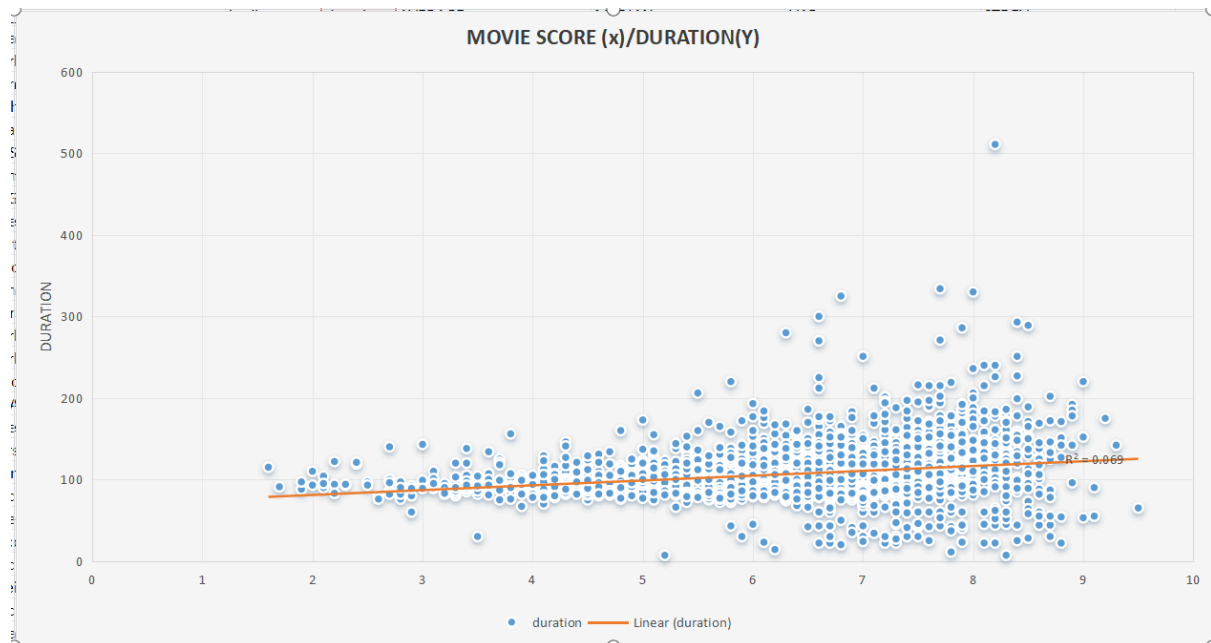
Create a scatter plot to visualize the relationship between movie duration and IMDB score. Add a trendline to assess the direction and strength of the relationship.

	C	D	E	F	G	H	
	imdb_score	duration	AVERAGE	MEDIAN	VAR	STDEV	ac
	7.9	178	107.2133253	103	637.5006494	25.24877521	
	7.1	169					
	6.8	148					
	8.5	164					

- Average (Mean) = 107.21:**
 - The average movie duration is approximately 107 minutes. This is the typical length of a movie in your dataset.
- Median = 103:**
 - The median movie duration is 103 minutes, meaning that half of the movies are shorter than 103 minutes and half are longer. Since the median is close to the mean, this suggests that the distribution of movie durations is relatively symmetric, with no extreme skewness.
- Variance (VAR) = 637.50:**
 - Variance measures how spread out the movie durations are from the mean. A variance of 637.50 indicates that there is some variability in movie durations, but without context, it's hard to say whether this is high or low.
- Standard Deviation (STDEV) = 25.25:**
 - The standard deviation is the square root of the variance and is easier to interpret. It tells us that, on average, movie durations deviate from the mean by about 25 minutes. This gives you a sense of how much individual movie durations differ from the average duration.

What These Metrics Suggest:

- General Movie Duration:** Most movies in your dataset are around 107 minutes long, with half of them being shorter than 103 minutes.
- Variation in Duration:** With a standard deviation of about 25 minutes, most movies will have a duration that falls within approximately 82 to 132 minutes (mean \pm 1 standard deviation).
- Symmetry of Distribution:** The proximity of the mean and median suggests that the distribution of movie durations does not have significant outliers or skewness.



Implications:

- **Positive Correlation:** The upward slope of the trendline indicates a positive correlation between movie duration and IMDB score. This means that, on average, longer movies might tend to have higher IMDB scores.
- **Strength of Relationship:** The steepness of the trendline indicates how strong this relationship might be. A steeper line suggests a stronger relationship, while a flatter line suggests a weaker one.

Interpretation of R^2 Value:

- **$R^2 = 0.069$:** This means that only about 6.9% of the variance in IMDB scores can be explained by the movie duration. In other words, movie duration is not a strong predictor of IMDB scores in your dataset.
- **Weak Relationship:** The low R^2 value suggests that there is likely little to no meaningful relationship between movie duration and IMDB scores. Other factors might be more important in determining IMDB scores.

C:TASK

Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics. Hint: Use Excel's COUNTIF function to count the number of movies for each language. Calculate the mean, median, and standard deviation of the IMDB scores for each language. Compare the statistics to understand the impact of language on movie ratings.

language	Sum of imdb_score	Average of imdb_score	Var of imdb_score	StdDev of imdb_score	Count of language	Sum of imdb_score2
Aboriginal	13.9	6.95	0.605	0.777817459	2	13.9
Arabic	36.9	7.38	0.782	0.884307639	5	36.9
Aramaic	7.1	7.1	#DIV/0!	#DIV/0!	1	7.1
Bosnian	4.3	4.3	#DIV/0!	#DIV/0!	1	4.3
Cantonese	76.5	6.954545455	0.496727273	0.704788814	11	76.5
Chinese	17	5.666666667	0.303333333	0.550757055	3	17
Czech	7.4	7.4	#DIV/0!	#DIV/0!	1	7.4
Danish	37.5	7.5	1.16	1.077032961	5	37.5
Dari	15	7.5	0.02	0.141421356	2	15
Dutch	29.7	7.425	0.189166667	0.434932945	4	29.7
Dzongkha	7.5	7.5	#DIV/0!	#DIV/0!	1	7.5
English	29824.7	6.397404547	1.256893069	1.121112425	4662	29824.7
Filipino	6.7	6.7	#DIV/0!	#DIV/0!	1	6.7
French	513.8	7.038356164	0.528508371	0.726985812	73	513.8
German	139.5	7.342105263	0.910350877	0.954123093	19	139.5
Greek	7.3	7.3	#DIV/0!	#DIV/0!	1	7.3
Hebrew	37.9	7.58	0.112	0.334664011	5	37.9
Hindi	185.7	6.632142857	1.95707672	1.398955582	28	185.7
Hungarian	7.1	7.1	#DIV/0!	#DIV/0!	1	7.1
Icelandic	15.1	7.55	0.845	0.919238816	2	15.1
Indonesian	15.8	7.9	0.18	0.424264069	2	15.8
Italian	79.5	7.227272727	1.548181818	1.244259546	11	79.5
Japanese	124.9	7.347058824	1.000147059	1.000073527	17	124.9
Kannada	7.1	7.1	#DIV/0!	#DIV/0!	1	7.1
Kazakh	6	6	#DIV/0!	#DIV/0!	1	6
Korean	59.1	7.3875	0.68125	0.825378701	8	59.1

I have created a pivot table for counting the movies of specific language and central tendencies

Alternative of pivot table can be this : `=AVERAGEIF(W:W,"English",V:V)` w:w contains language and v:v contains imdb score

Also for counting : `=COUNTIF(W:W,"English")`

1. Most Common Languages:

- **English** is the most common language in the dataset with a total of 4662 movies.
- Other commonly used languages include **French (73 movies)**, **Hindi (28 movies)**, **Spanish (40 movies)**, and **Mandarin (24 movies)**.

2. IMDB Scores Descriptive Statistics:

- **English:**
 - **Mean IMDB Score:** 6.3974
 - **Standard Deviation:** 1.1211
 - **Median:** 6.397
- **French:**

- **Mean IMDB Score:** 7.038
- **Standard Deviation:** 0.727
- **Median:** 7.038
- **Hindi:**
 - **Mean IMDB Score:** 6.632
 - **Standard Deviation:** 1.399
 - **Median:** 6.632
- **Spanish:**
 - **Mean IMDB Score:** 6.938
 - **Standard Deviation:** 0.855
 - **Median:** 6.938
- **Mandarin:**
 - **Mean IMDB Score:** 6.788
 - **Standard Deviation:** 1.037
 - **Median:** 6.788

3. Impact of Language on IMDB Scores:

- **English:** The large number of movies in English leads to a broader range of IMDB scores, resulting in a relatively higher standard deviation. This suggests that English-language movies have a wider variability in their quality as perceived by audiences.
- **French:** French-language movies have a higher mean IMDB score with a lower standard deviation, indicating that these movies tend to be more consistently rated higher.
- **Hindi:** Hindi movies show significant variability in ratings, with a higher standard deviation, suggesting that the quality of Hindi movies varies widely according to IMDB ratings.
- **Spanish and Mandarin:** Both languages show moderate mean scores with relatively low standard deviations, indicating a more consistent quality in these movies.

4. Languages with High Variability:

- **Russian:** The variance is high (1.915), suggesting significant variability in the IMDB scores, possibly due to diverse genres or varying production qualities.
- **Italian:** Also shows a relatively high Variance (1.548), indicating varied audience reception.

D: Task:

Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations. Hint: Calculate the average IMDB score for each director. Use Excel's PERCENTILE function to identify the directors with the highest scores. Compare the scores of these directors to the overall distribution of scores.

Solution :

I created a pivot table for selecting director name and imdb score calculating the imdb score average

And then calculating the percentage using this formula : =PERCENTILE(C2:C2399,0.9)

And then sorted by desc to get the directors of making percentage .

director_name	Sum of imdb_score	Count of director_name	Average of imdb_score
A. Raven Cruz	1.9	1	1.9
Aaron Hann	6	1	6
Aaron Schneider	7.1	1	7.1
Aaron Seltzer	2.7	1	2.7
Abel Ferrara	6.6	1	6.6
Adam Brooks	7.2	1	7.2
Adam Carolla	6.1	1	6.1
Adam Goldberg	5.4	1	5.4
Adam Green	5.7	1	5.7
Adam Jay Epstein	3.8	1	3.8
Adam Marcus	4.3	1	4.3
Adam McKay	41.5	6	6.916666667
Adam Rapp	6.4	1	6.4
Adam Rifkin	13	2	6.5
Adam Shankman	47.7	8	5.9625
Adrian Lyne	25.6	4	6.4
Adrienne Shelly	7.1	1	7.1
Agnieszka Holland	6.8	1	6.8
Agnieszka Wojtowicz-Vosloo	5.9	1	5.9
Agustín Díaz Yanes	6.1	1	6.1
Aki Kaurismäki	7.2	1	7.2
Akira Kurosawa	16.2	2	8.1
Akiva Goldsman	6.2	1	6.2
Akiva Schaffer	18.1	3	6.033333333
Al Franklin	4.3	1	4.3
Al Silliman Jr.	4	1	4

Calculated percentage with this formula : =PERCENTILE(C2:C2399,0.9)

DIRECTOR	IMDB SCORE SUM	AVERAGE	PERCENTAGE DESCENDING	DIRECTOR DESCENDING
A. Raven Cruz	1.9	1.9	7.5	Peter Weir
Aaron Hann	6	6	7.5	Peter Webber
Aaron Schneider	7.1	7.1	7.5	Pedro Almodóvar
Aaron Seltzer	2.7	2.7	7.5	Peter Stebbings
Abel Ferrara	6.6	6.6	7.5	Pece Dingo
Adam Brooks	7.2	7.2	7.5	Peter H. Hunt
Adam Carolla	6.1	6.1	7.5	Peter Sollett
Adam Goldberg	5.4	5.4	7.5	Pawel Pawlikowski
Adam Green	5.7	5.7	7.5	Peter Flinth
Adam Jay Epstein	3.8	3.8	7.5	Peter Sohn
Adam Marcus	4.3	4.3	7.5	Paul Weitz
Adam McKay	41.5	6.916666667	7.5	Peter Farrelly
Adam Rapp	6.4	6.4	7.5	Peter Segal
Adam Rifkin	13	6.5	7.5	Paul Greengrass
Adam Shankman	47.7	5.9625	7.5	Paul Fox
Adrian Lyne	25.6	6.4	7.5	Paul Thomas Anderson
Adrienne Shelly	7.1	7.1	7.5	Paul Fierlinger
Agnieszka Holland	6.8	6.8	7.5	Paul Schrader
Agnieszka Wojtowicz-Vosloo	5.9	5.9	7.5	Paul Feig
Agustín Díaz Yanes	6.1	6.1	7.5	Paul Michael Glaser
Aki Kaurismäki	7.2	7.2	7.5	Paul Weiland
Akira Kurosawa	16.2	8.1	7.5	Peter Faiman
Akiva Goldsman	6.2	6.2	7.5	Peter Jackson
Akiva Schaffer	18.1	6.033333333	7.5	Peter Ramsey
Al Franklin	4.3	4.3	7.5	Paul Donovan
Al Silliman Jr.	4	4	7.5	Paul McGuigan
Alain Resnais	6.3	6.3	7.5	Paul Crowder
Alan Alda	7.2	7.2	7.5	Paul Mazursky

I HAVE FILTER SOME ROWS BASED ON HIGHER PERCENTAGE AND LOWER PERCENTAGE

PERCENTAGE	DIRECTOR	AVERAGE	MEDIAN	RANK	MEDIAN	VAR
7.52	Patricia Cardoso	7.1		77%	6.4	1.322938374
7.525	Patricia Riggén	4.5		7%	6.4	1.323888746
7.53	Patricia Rozema	7.6		91%	6.4	1.321132609
7.52	Penny Marshall	6.6		57%	6.4	1.328767573
7.525	Perry Andelin Blake	7.6		91%	6.4	1.330726723
7.52	Peter Lepeniotis	6.8		65%	6.4	1.341251732
7.525	Peter Lord	6.8		65%	6.4	1.342986346
7.53	Peter M. Cohen	6.8		65%	6.4	1.344725051
7.535	Peter MacDonald	6.6		57%	6.4	1.346467859
7.54	Peter Medak	7.725		95%	6.4	1.348525006
6.98	Zoran Lisinac	4		4%	6.6	1.79
7.01	Zal Batmanglij	5		12%	6.7	0.8825
7.04	Zak Penn	6.6		57%	6.8	0.063333333

1. Top Performing Directors

- **Peter Medak**
 - **Average IMDb Score:** 7.725
 - **Percentile Rank:** 95%
 - **Median Score:** 6.4

- **Variance:** 1.348525006
- **Insight:** Peter Medak has the highest average IMDb score and percentile rank, indicating he is among the top performers. His score is significantly higher than the median and has the highest percentile rank, showing that his movies are highly rated compared to others.
- **Patricia Rozema & Perry Andelin Blake**
 - **Average IMDb Score:** 7.6
 - **Percentile Rank:** 91%
 - **Median Score:** 6.4
 - **Variance:** 1.321132609 (Rozema), 1.330726723 (Blake)
 - **Insight:** Both Patricia Rozema and Perry Andelin Blake also have high average scores and are ranked very high in the percentile distribution. Their movies are highly rated, and they are consistent in their performance.

2. Directors with Moderate Performance

- **Peter M. Cohen, Peter Lepeniotis, Peter Lord**
 - **Average IMDb Score:** 6.8
 - **Percentile Rank:** 65%
 - **Median Score:** 6.4
 - **Variance:** 1.344725051 (Cohen), 1.341251732 (Lepeniotis), 1.342986346 (Lord)
 - **Insight:** These directors have average IMDb scores around 6.8 and are in the mid-range of the percentile distribution. Their performance is relatively stable but not as high as the top performers.
- **Penny Marshall, Peter MacDonald**
 - **Average IMDb Score:** 6.6 (Marshall), 6.6 (MacDonald)
 - **Percentile Rank:** 57%
 - **Median Score:** 6.4
 - **Variance:** 1.328767573 (Marshall), 1.346467859 (MacDonald)
 - **Insight:** Penny Marshall and Peter MacDonald have scores around 6.6 and are in the lower middle range of the percentile distribution. Their movies are generally rated lower compared to top performers.

3. Low Performing Directors

- **Zoran Lisinac**
 - **Average IMDb Score:** 4.0
 - **Percentile Rank:** 4%
 - **Median Score:** 6.6
 - **Variance:** 1.79
 - **Insight:** Zoran Lisinac has the lowest average IMDb score and percentile rank, indicating that his movies are rated significantly lower compared to others. The high variance suggests a high degree of variability in the IMDb scores of his movies.
- **Zal Batmanglij**
 - **Average IMDb Score:** 5.0
 - **Percentile Rank:** 12%

- **Median Score:** 6.7
- **Variance:** 0.8825
- **Insight:** Zal Batmanglij also has a lower average IMDb score and percentile rank, though not as low as Zoran Lisinac. The scores are still relatively lower in comparison to top directors.

Summary

- **High Performers:** Peter Medak, Patricia Rozema, and Perry Andelin Blake are the top directors with high average IMDb scores and high percentile ranks.
- **Moderate Performers:** Directors like Peter M. Cohen and Peter Lepeniotis show moderate performance with average scores around 6.8.
- **Low Performers:** Directors such as Zoran Lisinac and Zal Batmanglij have lower average scores and percentile ranks, indicating lower performance relative to others.

E :Task:

Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin. Hint: Calculate the correlation coefficient between movie budgets and gross earnings using Excel's CORREL function. Calculate the profit margin (gross earnings - budget) for each movie and identify the movies with the highest profit margin using Excel's MAX function.

I USED INDEX MATCH AND MAX FUN TO CALCULATE HIGHEST PROFIT MARGIN MOVIE DISPLAYED ON THE IMG.

=INDEX(O2:O4999,MATCH(MAX(I2:I4999,0)))									
G	H	I	J	K	L	M	N		
	MOVIE TITLE	PROFIT MARGIN	BUDGET	GROSS	gross	budget	CORRELATION		
	Avatar	523505847	237000000	760505847	760505847	237000000	0.101033478	Avatar	
		9404152	300000000	309404152	309404152	300000000		Pirates of the Caribbean: At World's End	
		-44925825	245000000	200074175	200074175	245000000		Spectre	
		198130642	250000000	448130642	448130642	250000000		The Dark Knight Rises	
		0	0	0	0	0		Star Wars: Episode VII - The Force Awakens	
		-190641321	263700000	73058679	73058679	263700000		John Carter	
		78530303	258000000	336530303	336530303	258000000		Spider-Man 3	
		-59192738	260000000	200807262	200807262	260000000		Tangled	
		208991599	250000000	458991599	458991599	250000000		Avengers: Age of Ultron	
		51956980	250000000	301956980	301956980	250000000		Harry Potter and the Half-Blood Prince	
		80249062	250000000	330249062	330249062	250000000		Batman v Superman: Dawn of Justice	
		-8930592	209000000	200069408	200069408	209000000		Superman Returns	
		-31631573	200000000	168368427	168368427	200000000		Quantum of Solace	
		198032628	225000000	423032628	423032628	225000000		Pirates of the Caribbean: Dead Man's Chest	
		-125710090	215000000	89289910	89289910	215000000		The Lone Ranger	
		66021565	225000000	291021565	291021565	225000000		Man of Steel	
		-83385977	225000000	141614023	141614023	225000000		The Chronicles of Narnia: Prince Caspian	
		403279547	220000000	623279547	623279547	220000000		The Avengers	
		-8936125	250000000	241063875	241063875	250000000		Pirates of the Caribbean: On Stranger Tides	
		-45979146	225000000	179020854	179020854	225000000		Men in Black 3	
		5108370	250000000	255108370	255108370	250000000		The Hobbit: The Battle of the Five Armies	
		32030663	230000000	262030663	262030663	230000000		The Amazing Spider-Man	
		-94780265	200000000	105219735	105219735	200000000		Robin Hood	
		33355354	225000000	258355354	258355354	225000000		The Hobbit: The Desolation of Smaug	
		-109916481	180000000	70083519	70083519	180000000		The Golden Compass	
		11051260	207000000	218051260	218051260	207000000		King Kong	
		458672302	200000000	658672302	658672302	200000000		Titanic	
		157197282	250000000	407197282	407197282	250000000		Captain America: Civil War	

1. Profit Margin Analysis:

- **"Avatar"** has the highest profit margin, with a value of **\$523,505,847**. This indicates that "Avatar" earned significantly more than its budget of **\$237,000,000**.

2. Correlation Between Budget and Gross Earnings:

- The correlation coefficient between the budget and gross earnings is **0.1010**.
- **Interpretation:** This value is close to zero, suggesting a very weak positive correlation between a movie's budget and its gross earnings. This implies that higher budgets do not strongly predict higher gross earnings in this dataset.

3. Key Insights:

- **High Profit Margin:** "Avatar" stands out as a high-profit movie, despite its large budget.
- **Low or Negative Profit:** Some movies did not perform as well financially, with one even losing money relative to its budget.
- **Weak Correlation:** The weak correlation indicates that factors other than budget (e.g., marketing, audience appeal, timing) may have a more significant impact on a movie's gross earnings.

Summary :

Q: "Why do movies in certain genres tend to have higher ratings?"

A: Genres like Animation tend to be rated higher due to consistent quality and audience appeal.

Q: "Why do certain genres show higher variability in ratings?"

A: Genres like Fantasy and Comedy have diverse audience receptions, leading to greater variability

Q: "Why do languages impact movie ratings differently?"

A: Languages like French show higher average ratings, indicating a consistent quality that resonates well with audiences.

Q: "Why do certain directors have higher ratings?"

A: Top directors, like Peter Medak, consistently produce highly rated films, boosting their overall ratings.

Q: "Why is there a weak correlation between movie budgets and gross earnings?"

A: Other factors, such as marketing and timing, play a more significant role than budget alone in determining a movie's success.

Stroy :

Q: "Why does Animation tend to have higher ratings compared to other genres?"

A: Animation movies often receive higher ratings due to their consistently high production quality and broad audience appeal.

Q: "Why do Fantasy and Comedy genres show higher variability in ratings?"

A: Fantasy and Comedy movies exhibit greater variability in ratings because they cater to diverse audience tastes, leading to a wide range of reception.

Q: "Why do languages like French have higher average ratings?"

A: Movies in languages like French often have higher average ratings because they are perceived to maintain a consistent quality that resonates well with audiences.

Q: "Why are certain directors, like Peter Medak, consistently highly rated?"

A: Directors like Peter Medak consistently produce films that resonate with audiences, leading to high average ratings and strong positive reception.

Q: "Why is there a weak correlation between movie budgets and gross earnings?"

A: The weak correlation between budgets and gross earnings suggests that factors beyond just the budget, such as marketing and timing, play a crucial role in a movie's financial success.