**Title: Movie Success Insights: Statistics, Popularity and Timing Considerations**

**Executive Summary**

In the ever-evolving landscape of the film industry, our comprehensive analysis delves into key factors shaping the success of movies. Through rigorous exploration, we uncover nuanced insights that span across evolving ratings, audience genre preferences, the influence of actors and directors, the delicate balance of budgets, and the strategic interplay of timing.

In essence, our data-driven insights are aimed to provide valuable insights that empower decision-makers in the film industry to navigate the ever-changing landscape and contribute to the creation of memorable and impactful cinematic experiences.

**Data Sources**

The datasets consisted of two files (movies\_cleaned.csv and movies\_data.csv). The datasets included 'Movie\_Name', 'Rating', 'Genre', 'Year', 'Rating\_Score', 'Votes', 'Director', 'Writer', 'Star', 'Country', 'Budget', 'Gross', 'Company', 'Runtime', and 'Release\_date', 'Production', 'Poster', 'Type','Ratings', 'Website', 'totalSeasons', 'Error', 'Response'.

To ensure the data's quality and relevance, we performed data cleaning, handling of missing values, and conversion of specific columns into appropriate formats. Grouping and aggregation were employed for group-level analysis, summary statistics, correlation, regression, independent sample t test, ranges, and plots, following a structured methodology to ensure the data was well-processed and suitable for analysis.

**Data Processing and Pandas Integration**

The Python programming language (version 3.10.13 packaged by Anaconda, Inc.) alongside version 2.0.3 of the Pandas library is utilised within Jupyter Notebook (version 5.3.0) for data processing and analysis. Leveraging Pandas efficiently facilitated the organization and manipulation the information into a structured tabular format, primarily using the Pandas DataFrame as the primary data structure. The DataFrame facilitated data cleaning, handling of missing values, and overall data preparation for a meaningful and insightful analysis. Furthermore, the codes were generated using pandas, numpy, matplotlib, seaborn and scipy.stats documentations (Pandas Documentation, 2023).

**Research Questions**

* OMDB Movie Ratings and Trends (Sunil):
  + *Question:* How have movie ratings evolved over the years, and are there discernible trends within different genres or directors?
* Genre Analysis and Popularity (Aysha):
  + *Question:* Which movie genres reign supreme in popularity, and is there a correlation between a movie's genre and its commercial success?
* Director/Actor Influence on Movie Success (Godswill):
  + *Question:* Do certain actors or directors wield a significant influence on a movie's success or ratings?
* Movie Length and Audience Preference (Ayomide):
  + *Question:* How does the duration of a movie impact audience reception and ratings? Are longer or shorter movies more favorably received?
* Box Office Performance and Budget Analysis (Sadek):
  + *Question:* Is there a relationship between a movie's budget, box office performance, ratings, and how timing influences these dynamics?

**Hypothesis**

1.

2. **Genre Analysis and Popularity**

1. Alternate Hypothesis: The greater the popularity of a movie genre, the more successful it will be.
2. Null Hypothesis: The popularity of a movie genre does not have a significant positive effect on its success.

3. **Director/Actor Influence on Movie Success**

1. Alternate Hypothesis: Directors or actors wield a significant influence on a movie's success.
2. Null Hypothesis: Directors or actors do not wield a significant influence on a movie's success.

5. **Box Office Performance and Budget Analysis**

1. Alternate Hypothesis: there a relationship between a movie's budget, box office performance, and ratings.
2. Null Hypothesis: there is no relationship between a movie's budget, box office performance, and ratings.

**Key Findings and Analysis**

1. **OMDB Movie Ratings and Trends**
2. **Genre Analysis and Popularity**

**Popularity of Genres based on Average Votes and Rating Score**

1. **Average Votes:**  'Mystery' movies received the highest votes on average with over 250,000 votes. The next most popular genre is 'Family' movies with approximately 170,000 votes.
2. **Rating Score:** The movie genre to produce the highest average rating was ‘Biography’. However,  there is not a great deal of variation in the average ratings of each genre; majority of ratings fall between 6 and 7. The small variance indicates that the rating scores tend to be very close to the mean, and to each other, therefore the movie ratings data does not provide substantial evidence to explore the hypothesis.

**Most successful Genre based on Gross Earnings**

The highest grossing genre of movies with over $500 million profit are 'Family' movies, producing almost double than 'Animation' movies.

**Linear Regression Analysis**

1. **Average Votes vs. Average Gross:** The positive correlation (0.51) between the Average Votes and Average Gross per Genre suggests that there is a moderate tendency for genres to be more successful if they have greater votes. However , since it is only a moderate correlation, it does not indicate causation.
2. **Average Rating Score vs. Average Gross:** There is a weak positive correlation (0.32) between the average ratings and gross earnings per Genre. This suggests that the success of a movie genre is not greatly affected by its rating.
3. **Actor/Director Influence**

**Top Directors and Actors Based on Average Ratings:**

1. **Directors:** Roberto Benigni, Tony Kaye, and Nadine Labaki lead in average ratings, indicating consistent high-quality filmmaking.
2. **Actors:** Mark Hamill and Philippe Noiret emerge as top-rated actors, showcasing their impact on the audience's perception of movie quality.

**Total Gross Earnings by Directors and Actors:**

1. **Directors:** Steven Spielberg leads in total gross earnings, followed by Anthony Russo and Peter Jackson. These directors consistently deliver movies that attract significant audiences and generate substantial revenue.
2. **Actors:** Robert Downey Jr. tops the list in total gross earnings, followed by Tom Hanks and Tom Cruise. These demonstrate their ability to draw audiences and contribute to a movie's financial success with impressive cumulative worldwide box office earnings.

**Average Votes:**

1. **Directors:**

The average votes received by directors, with Christopher Nolan leading the pack, indicate their influence on audience engagement. Directors who can capture the audience's attention and appreciation contribute significantly to a movie's overall success.

1. **Actors:**

Actors like Mark Hamill, Ben Burtt, and Alexandre Rodrigues garner high average votes, emphasizing their role in creating movies that resonate with audiences and receive positive feedback.

**Linear Regression Analysis:**

1. **Director's Rating vs. Actor's Rating:** A positive linear relationship is observed, suggesting that higher-rated directors tend to work with higher-rated actors.
2. **Total Gross by Director vs. Total Gross by Actor:** The weak correlation (R=0.08) implies that directorial success does not strongly predict actor-based financial success.
3. **Average Votes by Director vs. Average Votes by Actor:** A moderate correlation (R=0.38) indicates a positive relationship between the average votes received by directors and actors.

**Two-Sample T-Tests:**

1. **Average Votes Comparison:** The t-test fails to reject the null hypothesis, suggesting no significant difference in average votes between directors and actors, as indicated by the high p-value (p = 0.9405).
2. **Total Gross Comparison:** The t-test fails to reject the null hypothesis, suggesting no significant difference in total gross earnings between directors and actors (p = 0.1865).
3. **Average Votes by Director vs. Actor:** The analysis fails to reject the null hypothesis, indicating no significant difference in average votes between directors and actors (p = 0.9405).

**Movie Length and Audience Preference**

**5. Box Office Performance and Budget Analysis**

**Budget vs. Box Office Earnings:**

1. An r-value of 0.5476 indicates a moderate positive linear relationship between the budget and gross of movies. As the budget increases, there tends to be a moderate increase in gross, but it's not as strong as a higher correlation would suggest.
2. The moderate correlation suggests that while there is a tendency for movies with higher budgets to have higher gross earnings, there is significant variability. In other words, there are instances where movies with lower budgets might still perform well in terms of gross earnings, and vice versa.

**Budget vs. Ratings:**

1. An r-value of 0.0051 indicates an almost non-existent linear relationship between the budget allocated for a movie and its rating. Changes in the budget don't show a significant pattern or impact on the movie ratings.
2. It suggests that the budget spent on a movie does not significantly influence or predict the rating it receives. Therefore, when aiming for a higher movie rating, other factors beyond budget allocation need to be considered and prioritized; like quality of the script, acting performances, audience preferences and so on.

**Relationship between Average Budget and Average Gross of Movies over the Years:**

1. A Pearson correlation coefficient of approximately 0.94 suggests a strong positive linear relationship between the average budget and average gross of movies over the years. The p-value associated with this correlation is extremely small (3.26e-19), indicating strong evidence against the null hypothesis that there is no correlation between the variables. This very low p-value suggests that the correlation coefficient is statistically significant.
2. This strong positive correlation over the years could reflect changes in audience preferences, production quality or other factors that have led to increased box office revenues for movies with higher budgets. While a high budget might correlate with higher gross, it also signifies a higher financial risk if the movie doesn't perform well.
3. Although there's a strong correlation, it's important to note that correlation doesn’t imply causation. Other factors, such as marketing strategies, script quality, casting, release timing, etc., also play crucial roles in a movie's success at the box office.

**Recommendations**

**Conclusion**

The analysis underscores the influence of directors and actors on movie success, with some individuals consistently delivering high-rated and commercially successful films.

**References**

Pandas Documentation (2023). <https://pandas.pydata.org/pandas-docs/stable/index.html>. Accessed on November 27, 2023.

Python Documentation (2023). <https://www.python.org/doc/>. Accessed on November 27, 2023. Accessed on November 29, 2023.

Seaborn Documentation (2023). Seaborn: Statistical data visualization. https://seaborn.pydata.org/

Statistical Function Documentation (2023). Retrieved from <https://docs.scipy.org/doc/scipy/reference/stats.html>. Accessed on on November 27, 2023.