FINAL REPORT

PREPARED BY
Team 4
Sami Adam, Jake Antmann, Alexander Hall, Drew Novatney

SECTION 1: INTRODUCTION

Our database research will center on LegendScreen, a fictional film studio, as our primary focus.

LegendScreen is a young and thriving film studio renowned for its active engagement in the development, financing, production, distribution, and marketing of a diverse range of films. The company's escalating achievements can be credited to its expertise in launching acclaimed films spanning across the horror, comedy, and drama genres. This expertise is rooted in its exceptional capability to source and train producers who specialize in these genres, surpassing the efficacy of its competitors.

LegendScreen recently entered the Adventure genre following extensive market research. The research highlighted Adventure as the most popular film genre as it boasted the largest potential market share. Furthermore, detailed analysis of recent Adventure film releases showcased the genre's ability to generate a high average return on investment (ROI), making it an appealing opportunity for the studio.

LegendScreen's Adventure operations have faced challenges that threaten long-term viability

Unfortunately, LegendScreen's first three years of operations within the Adventure genre have been notably disappointing. During this period, the studio launched four Adventure films and none managed to meet revenue or box office expectations. Furthermore, critical reviews for each of these movies were predominantly negative, citing a lack of striking visuals and captivating cinematography as main sources of disappointment. Management suspects that the company is not nearly as effective at sourcing and training producers as it is for its portfolio of other genres and that it must develop new best-in-class practices to satisfy the unique needs of Adventure films watchers. LegendScreen must deliver compelling Adventure films that resonate with audiences in the near future, or it risks further financial setbacks, reputational damage, and the potential divestiture of its Adventure operations.

LegendScreen aims to utilize a new database system to strengthen its Adventure operations

LegendScreen wants to compile a database of a variety of Adventure film information so that it may conduct thorough analytics research and draw actionable insights to help generate long-term success for its Adventure operations. By navigating through the relationships for a variety of entities (including but not limited to: MOVIE, ACTOR, DIRECTOR, PRODUCTION STUDIO, VFX STUDIO, COMPOSER, PLATFORM, AWARD), the company should be able to (1) identify trends that correlate with high or low quality or performance (2) determine which attributes best predict financial success and audience satisfaction (3) understand what constitutes effective resource allocation.

While conventional movie databases like IMDb, Rotten Tomatoes, and Metacritic offer valuable insights into films' box office performance and critical reception, our database aims to transcend these conventional metrics. We seek to delve deeper into the multifaceted dynamics of the film industry by exploring a broader range of indicators that contribute to a film's quality and memorability.

Our database will establish connections beyond theatrical success, delving into the film's penetration across various platforms and licensing arrangements. This includes detailing which streaming services or

distribution channels the film is licensed to and for what duration, providing a comprehensive view of its reach beyond the traditional theater experience. Moreover, we want to highlight the interconnectedness of awards not only with films but also with the actors, directors, composers, and other contributors involved. By mapping out these associations, our database will offer clear insights into where talent and value stem from across the film industry's intricate network of collaborators. Our data will even bridge the gap between actors and the characters they portray, as well as the genres they excel in. This will enable us to identify which actors are best suited for specific niche roles, providing valuable insights for casting decisions and genre exploration.

In essence, unlike traditional movie data sources, our database will be more comprehensive, informative, and representative of the nuanced relationships that studios must leverage to release successful films. By capturing the intricate web of connections within the industry, we aim to empower studios with actionable insights for strategic decision-making and creative development.

Leveraging this database system will enable LegendScreen to enjoy plenty of analytics benefits

Utilizing a new database system will allow LegendScreen the ability to identify trends that correlated with high or low box office performances, award recognition, and scope of influence beyond theaters. In doing so, LegendScreen will be able to leverage their operations based on the historical data inferences. This includes identifying common themes, actors, and marketing strategies associated with successful movies. For instance, by identifying successful themes that consistently draw in audience members, LegendScreen will be able to tailor specific projects that align with those preferences. Additionally, by analyzing the combined effects of various contributors to a film, LegendScreen can gain an idea of which contributors produce the best projects with one another, enabling the company to strategically select combinations of inputs into its new films. Using these trends, LegendScreen will also be able to utilize predictive analytics to determine the potential financial success and audience approval of certain projects. Doing so will allow LegendScreen the ability to better choose projects.

A database system that creates clear relationships between producers, films, and other financial data will provide a clear framework in deciding which producers and other contributors to select for specific movies that LegendScreen wants to fund within the Adventure genre. For instance, data linking a producer to multiple projects with similar budgets and cast members to what LegendScreen has in mind can be beneficial, as LegendScreen can then select a producer that has also generated the desired levels of revenue from that field.

The new database will enable LegendScreen to effectively allocate resources

Implementing a new database system will enable the studio to analyze historical data from previous Adventure film productions. Successful film production demands the careful provision of resources across several essential functions including set design/visual effects, promotional marketing, and employee salaries. By examining how previously successful films' have allocated their budget, LegendScreen will be able to identify patterns and establish correlations between budget allocation and box office success. This analytical approach to resource management will serve as a foundation for optimizing LegendScreen's financial planning for future films, thus simplifying the decision-making process and reducing the likelihood of improper resource allocation.

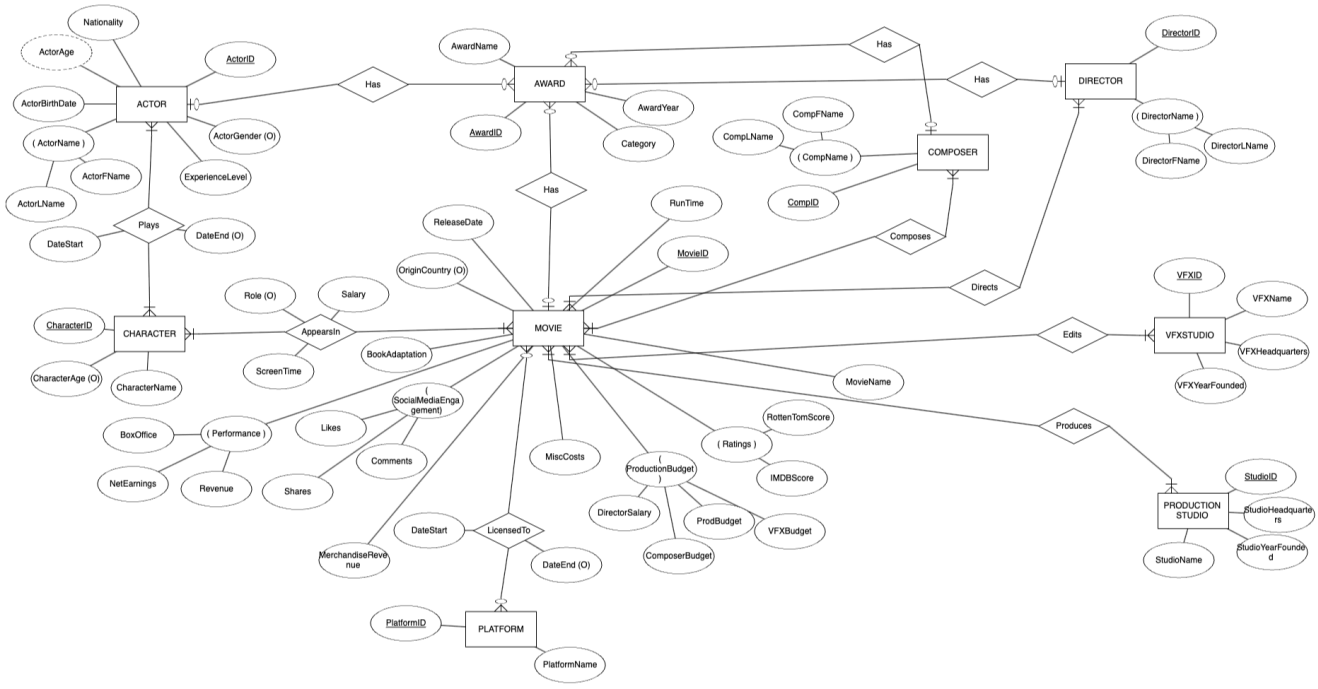
Overall, LegendScreen utilization of the new database will offer analytic capabilities and optimize effective resource allocation. The benefits of a new database will allow LegendScreen the ability to make data-driven decisions based on comprehensive analysis. It will be a key asset in the company's strategic planning initiatives and is a potential competitive advantage over their competitors.

SECTION 2: CONCEPTUAL MODEL

Criteria Used to Draft LegendScreen's ERD

1. Clearly defined entities involved in the data system, including any entities that represent the combined effect of several other entities
2. Clearly defined, original, and relevant attributes associated with each entity (i.e. ExperienceLevel within the ACTOR entity is relevant as it could contribute to predicting the success of a film or combination of films, given that actors with more experience often bring a level of expertise and credibility to their role.)
3. Novel information not found in publicly available databases which can tell a story and provide a competitive edge through unique insights (i.e. Social Media engagement for a specific movie is a proxy for how much "hype" a movie generates pre and post release.)

LegendScreen's ERD



Requirements Used to Draft LegendScreen's ERD

Entities

For each MOVIE: unique MovieID, MovieName, RunTime, OriginCountry, ReleaseDate, Performance (composed of BoxOffice, NetEarnings, and Revenue), Ratings (composed of IMBDScore, RottenTomScore), Productionbudget (composed of DirectorSalary, ComposerBudget, ProdBudget, VFXBudget), MiscCosts, MerchandiseRevenue, BookAdaptation, SocialMediaEngagement (composed of Likes, Shares, Comments)

For each ACTOR: unique ActorID, ActorName (composed of ActorFName and ActorLName), Nationality, ActorBirthdate, ActorAge (derived), ExperienceLevel, ActorGender (optional)

For each CHARACTER: unique CharacterID, CharacterName, CharacterAge (optional)

For each AWARD: unique AwardID, AwardName, Category, AwardYear

For each DIRECTOR: unique DirectorID, DirectorName (composed of DirectorFName and DirectorLName)

For each COMPOSER: unique CompID, CompName (composed of CompFName and CompLName)

For each VFX: unique VFXID, VFXName, VFXHeadquarters, VFXYearFounded

For each STUDIO: unique StudioID, StudioName, StudioHeadquarters, StudioYearFounded

For each PLATFORM (PLATFORM refers to any streaming services or other forms of distribution/consumption a movie might take): unique PlatformID, PlatformName

Relationships

Each actor can have between zero and many awards
Each award can be won by zero or one actor

Each actor can play one or many characters
Each character can be played by one or many actors
For each instance of an actor playing a character, we keep track of the date that the role started and ended (DateStart, DateEnd (optional))

Each character appears in one or many movies
Each movie features one or many characters
For each instance of a character appearing in a movie, we keep track of Salary, ScreenTime, and Role (optional)

Each movie can have between zero and many awards
Each award can be won by zero or one movie

Each director can have between zero and many awards
Each award can be won by zero or one director

Each composer can have between zero and many awards

Each award can be won by zero or one composer

Each movie is directed by one or more directors
Each director can make one or many movies

Each movie is composed by one or more composer
Each composer can make one or many movies

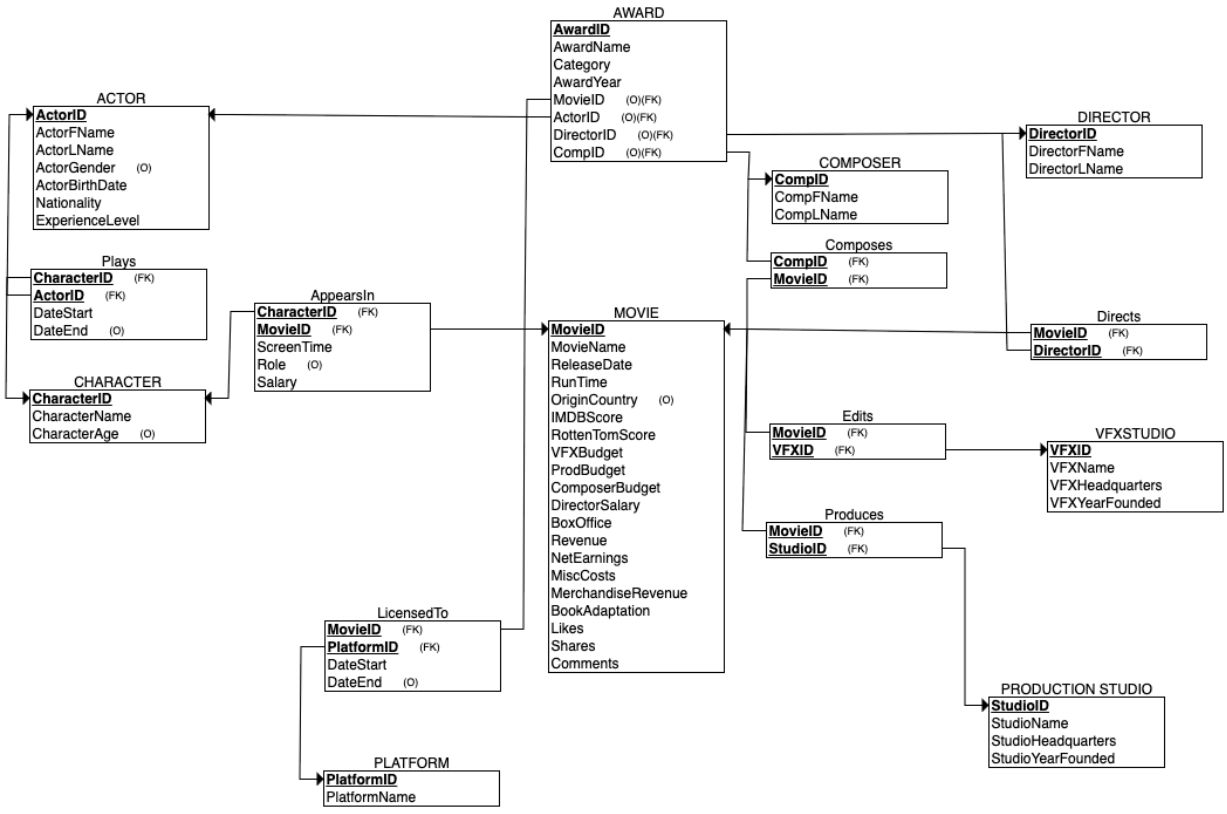
Each movie is edited by one or VFX studios
Each VFX studio can make one or many movies

Each movie is produced by one or more production studios
Each production studio can make one or many movies

Each platform can host zero or many movies
Each movie can be accessible on zero or many platforms

SECTION 3: RELATIONAL MODEL

Below, we have included our normalized relational schema:



SECTION 4: DATABASE CREATION AND DATA ENTRY

To begin, we followed standard procedures to construct our database, ensuring we created tables, attributes, primary and foreign keys, relationships, and appropriate data types. This meticulous approach was guided by our finalized Entity-Relationship Diagram (ERD) and relational schema.

Next, during the data population phase, we conducted extensive research and extracted data from 37 different movies sourced from various platforms on the internet. Our selection criteria aimed to capture iconic classics across diverse categories while ensuring a prominent adventure element or character arc in each entry, aligning with LegendScreen's objectives of enhancing the quality and popularity of its adventure genre releases.

Our entry selections included dramas with adventurous undertones such as "The Godfather" and "The Shawshank Redemption," action-packed adventures exemplified by franchises like James Bond, Mission Impossible, and Kill Bill, as well as sci-fi spectacles like "Inception" and "The Matrix." Additionally, we included fantasy epics like "The Lord of the Rings" and "Pirates of the Caribbean" to fill our database with their adventurous narratives.

Utilizing accessible traditional movie data sources like IMDb, Rotten Tomatoes, and MetaCritic enabled us to extract real and accurate information for our database entries. In instances where specific attributes, such as the duration of licensing agreements with streaming platforms, were unavailable, we inserted null values. For more obscure or confidential aspects, such as salary or budget details and character ages, we attempted to derive reasonable estimates from industry averages on the internet and leverage historical context in an effort to maintain data integrity and reduce potential anomalies. The only time we explicitly generated inaccurate data was when we populated the Awards entity. We gave awards to certain parties that did not receive an award. We did this to ensure the database contained a sufficient amount of test data Overall, our database benefited from the transparency and visibility inherent in the film industry, allowing us to source films with abundant ancillary data.

SECTION 5: DATABASE USE

‘

The first query serves to examine which streaming platforms own the rights to the most movies, and whether those platforms align with higher ratings and more movie awards. A query of this nature will help LegendScreen target specific streaming platforms to launch their films with.

```
SELECT p.PlatformName,
            COUNT(DISTINCT m.MovieID) AS MovieCount,
      AVG(m.IMDBScore) AS AverageIMDBScore,
            AVG(m.RottenTomScore) AS AverageRottenTomScore,
      SUM(CASE WHEN a.MovieID IS NOT NULL THEN 1 ELSE 0 END) AS MovieAwards
FROM movie m
        LEFT JOIN licensedto lt
        ON m.MovieID = lt.MovieID
        LEFT JOIN platform p
        ON lt.PlatformID = p.PlatformID
        LEFT JOIN award a
```

```
        ON m.MovieID = a.MovieID
GROUP BY p.PlatformName
ORDER BY MovieCount DESC;
```

PlatformName	MovieCount	AverageIMDBScore	AverageRottenTomSco...	MovieAwards
Netflix	26	8.03462	87.9231	12
Disney+	16	7.85000	85.0000	7
Amazon Prime	14	7.75714	81.3571	6
Hulu	11	8.00909	81.3636	3
Max	8	8.55000	91.1250	5
Apple TV	3	7.96667	79.6667	0
Paramount	1	7.40000	90.0000	0
Peacock	1	7.40000	90.0000	0

The second query aims to investigate how screen time and salary may be related to actors who are winning awards. From this query, LegendScreen would like to identify actors with relatively low salaries who play major roles in their films and also win awards, because signing affordable, award-winning actors is important to producing profitable films.

```
SELECT a.ActorFName, a.ActorLName, a.ExperienceLevel, c.CharacterName,
AVG(ai.ScreenTime) AverageScreenTime,
            AVG(ai.Salary) AS AverageSalary, (AVG(ai.Salary) / AVG(ai.ScreenTime)) AS
Average_Salary_Per_Minute
FROM actor a
        LEFT JOIN plays p
        ON a.ActorID = p.ActorID
        LEFT JOIN `character` c
        ON p.CharacterID = c.CharacterID
        LEFT JOIN appearsin ai
        ON c.CharacterID = ai.CharacterID
        LEFT JOIN award aw
        ON a.ActorID = aw.ActorID
WHERE aw.ActorID IS NOT NULL
GROUP BY a.ActorFName, a.ActorLName, c.CharacterName
ORDER BY Average_Salary_Per_Minute;
```

ActorFName	ActorLName	ExperienceLevel	CharacterName	AverageScreenTime	AverageSalary	Average_Salary_Per_Minu...
Al	Pacino	Veteran	Michael Corleone	175.0000	35000.0000	200.00000000
Timothee	Chalamet	Intermediate	Paul Atreides	155.0000	300000.0000	1935.48387097
Brendan	Fraser	Veteran	Rick O'Connell	126.0000	3000000.0000	23809.52380952
Taron	Egerton	Intermediate	Gary "Eggsy" Unwin	104.0000	3500000.0000	33653.84615385
Harrison	Ford	Veteran	Indiana Jones	113.5000	5200000.0000	45814.97797357
Uma	Thurman	Veteran	The Bride (Beatrix Kiddo)	111.0000	7000000.0000	63063.06306306
Johnny	Depp	Veteran	Jack Sparrow	143.0000	59000000.0000	412587.41258741
Austin	Butler	Intermediate	Pete "Maverick" Mitchell	130.0000	100000000.0000	769230.76923077

The third query is used by LegendScreen to further identify actors of interest for their films. It selects actors who appear in movies that have above average IMDb ratings, appear in over half the runtime in their respective movies (making them a main character), and also have below average salaries. It orders by box office numbers as well. This query helps discover actors who may not win individual awards, but are consistently playing a major role in films that perform particularly well at the box office as well as with critical reviewers.

```
SELECT a.ActorFName, a.ActorLName, ai. Salary, m.MovieName, m.BoxOffice
FROM movie m
        LEFT JOIN appearsin ai
        ON m.MovieID = ai.MovieID
   LEFT JOIN `character` c
   ON ai.CharacterID = c.CharacterID
   LEFT JOIN plays p
   ON c.CharacterID = p.CharacterID
   LEFT JOIN actor a
   ON p.ActorID = a.ActorID
WHERE m.IMDBScore > (SELECT AVG(IMDBScore)
                                    FROM movie)
AND
        ai.Salary < (SELECT AVG(Salary)
                                FROM appearsin)
AND ai.ScreenTime >= (SELECT AVG(ScreenTime)
                                    FROM appearsin)
ORDER BY m.BoxOffice DESC;
```

ActorFName	ActorLName	Salary	MovieName	BoxOffice
Elijah	Wood	1000000	The Lord of the Rings: The Return of the King	1118000000
Daniel	Craig	17000000	James Bond: Skyfall	1108561007
Elijah	Wood	150000	The Lord of the Rings: The Fellowship of the Ring	891000000
Daniel	Craig	3200000	James Bond: Casino Royale	594200000
Christian	Bale	10000000	Batman: The Dark Knight	534000000
Christian	Bale	15000000	Batman: The Dark Knight Rises	448139099
Timothee	Chalamet	300000	Dune	401700000
Christian	Bale	9000000	Batman: Batman Begins	374893142
Tim	Robbins	500000	The Shawshank Redemption	278000000
Al	Pacino	35000	The Godfather	245000000
Iko	Uwais	4200000	The Raid 2	120000000

The fourth query selects movies and their monetary figures such as box office, net earnings, and merch revenue when the director won Best Director award, and the movie category was specifically “adventure.” This helps LegendScreen additionally identify directors who both perform well in regards to award-winning, in addition to their films being highly profitable.

```
SELECT m.MovieName, m.BoxOffice, m.NetEarnings, m.MerchRevenue, d.DirectorFName,
d.DirectorLName, m.DirectorSalary
FROM movie m
        LEFT JOIN directs ds
        ON m.MovieID = ds.MovieID
        LEFT JOIN director d
        ON ds.DirectorID = d.DirectorID
WHERE d.DirectorID IN (SELECT d.DirectorID
                                    FROM award a
                                            INNER JOIN director d
                                            ON a.DirectorID = d.DirectorID
                                            LEFT JOIN directs ds
                                            ON d.DirectorID = ds.DirectorID
```

```
                                        LEFT JOIN movie m
                                        ON ds.MovieID = m.MovieID
                                WHERE Category = 'Adventure')
ORDER BY m.BoxOffice DESC;
```

MovieName	BoxOffice	NetEarnings	MerchRevenue	DirectorFName	DirectorLNa...	DirectorSalary
James Bond: Skyfall	1108561007	500000000	50000000	Cary Joji	Fukunaga	20000000
James Bond: Spectre	880674609	370000000	40000000	Cary Joji	Fukunaga	20000000
Inception	825000000	400000000	40000000	Christopher	Nolan	20000000
The Batman	770836823	320000000	40000000	Matt	Reeves	20000000
Batman: The Dark Knight	534000000	300000000	40000000	Christopher	Nolan	20000000
Batman: The Dark Knight Rises	448139099	400000000	50000000	Christopher	Nolan	25000000
Dune	401700000	180000000	40000000	Denis	Villeneuve	20000000
Batman: Batman Begins	374893142	140000000	40000000	Christopher	Nolan	15000000

The fifth query is a complex query that aims to bring multiple aspects of the database together to identify specific movies that perform well all-around (awards, box office, and more), such that these films can be examined with further scrutiny and LegendScreen can mimic aspects critical to their success. It acquires award counts for actors, directors, and composers and combines those awards with any movie-level awards won for their respective films, and returns each films total award count.

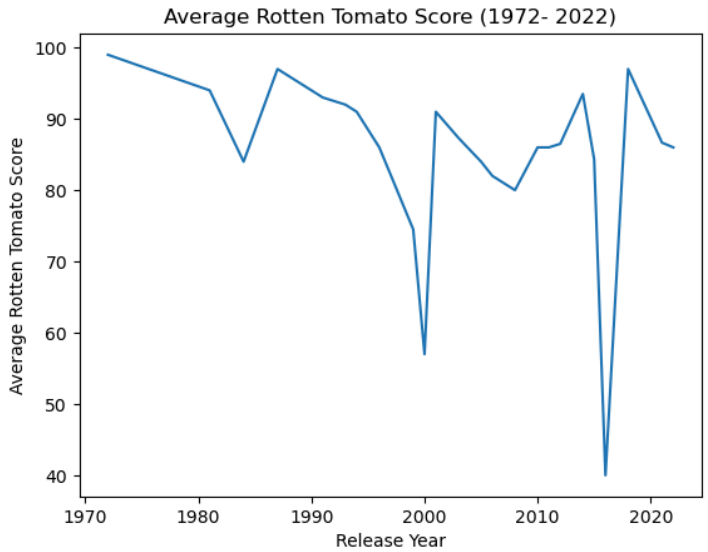
```
SELECT ac.AwardCount + COUNT(ma.AwardID) AS TotalAwardCount, ac.MovieName,
ac.IMDBScore, ac.BoxOffice
FROM (
   SELECT MovieID, MovieName, IMDBScore, BoxOffice, COUNT(*) AS AwardCount
   FROM (

-- Query for Composer awards
      SELECT *
            FROM (
                  SELECT a.AwardID, a.AwardYear, c.CompID,
c.CompFName,c.CompLName, m.MovieID, m.MovieName, m.IMDBScore, m.BoxOffice,
m.ReleaseDate,
                                    ROW_NUMBER() OVER (PARTITION BY a.AwardID
ORDER BY ABS(DATEDIFF(a.AwardYear, m.ReleaseDate))) AS rnk
                  FROM award a
                  INNER JOIN composer c
      ON a.CompID = c.CompID
                  LEFT JOIN composes cs
      ON c.CompID = cs.CompID
                  LEFT JOIN movie m
      ON cs.MovieID = m.MovieID
) AS ranked
WHERE rnk = 1

UNION

-- Query for Director awards
            SELECT *
            FROM (
```

```
                    SELECT a.AwardID, a.AwardYear, d.DirectorID, d.DirectorFName,
d.DirectorLName, m.MovieID, m.MovieName, m.IMDBScore, m.BoxOffice, m.ReleaseDate,
                                ROW_NUMBER() OVER (PARTITION BY a.AwardID
ORDER BY ABS(DATEDIFF(a.AwardYear, m.ReleaseDate))) AS rnk
                    FROM award a
                    INNER JOIN director d
        ON a.DirectorID = d.DirectorID
                    LEFT JOIN directs ds
        ON d.DirectorID = ds.DirectorID
                    LEFT JOIN movie m
        ON ds.MovieID = m.MovieID
) AS ranked
WHERE rnk = 1

UNION

-- Query for Actor awards
                SELECT *
                FROM (
                    SELECT a.AwardID, a.AwardYear, ac.ActorID, ac.ActorFName,
ac.ActorLName, m.MovieID, m.MovieName, m.IMDBScore, m.BoxOffice, m.ReleaseDate,
                                ROW_NUMBER() OVER (PARTITION BY a.AwardID
ORDER BY ABS(DATEDIFF(a.AwardYear, m.ReleaseDate))) AS rnk
                    FROM award a
                    INNER JOIN actor ac
        ON a.ActorID = ac.ActorID
                    LEFT JOIN plays p
        ON ac.ActorID = p.ActorID
                    LEFT JOIN `character` c
        ON p.CharacterID = c.CharacterID
                    LEFT JOIN appearsin ai
        ON c.CharacterID = ai.CharacterID
                    LEFT JOIN movie m
        ON ai.MovieID = m.MovieID
) AS ranked
WHERE rnk = 1
   ) AS all_awards
GROUP BY MovieID
) AS ac
-- Joining on MovieID awards at the end
LEFT JOIN (
   SELECT a.AwardID, a.MovieID
   FROM award a
) AS ma ON ac.MovieID = ma.MovieID
GROUP BY ac.MovieID, ac.AwardCount
ORDER BY TotalAwardCount DESC;
```

TotalAwardCount	MovieName	IMDBScore	BoxOffice
4	Dune	8.1	401700000
3	Raiders of the Lost Ark	8.6	380000000
3	Kill Bill: Vol. 1	8.1	150000000
3	The Godfather	9.2	245000000
2	Batman: The Dark Knight	9.0	534000000
2	Indiana Jones and the Temple of Doom	8.4	300000000
2	Batman: Batman Begins	8.4	374893142
2	James Bond: Casino Royale	8.0	594200000

SECTION 6: DATA ANALYSIS AND VISUALIZATIONS

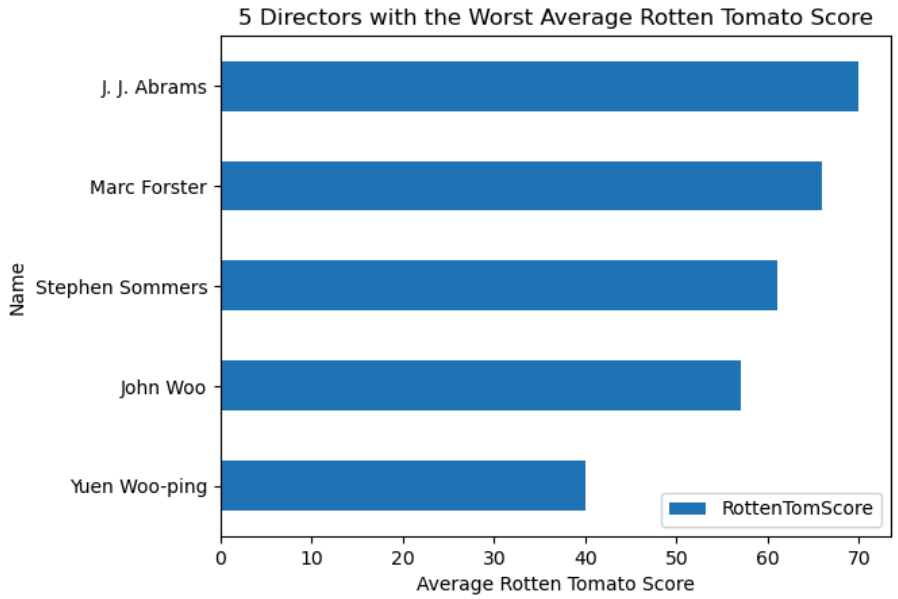
While using the database, we discovered multiple avenues to discover insights. Our first visual compares the average rotten tomato score across 50 years. In order to get this information, we queried the movie table. We selected the average rotten tomato score and year. The important step was that we wrote a group by statement grouping the results by the year that they were released. As seen in the visual below, the average rotten tomato consistently stays above 80, however there have been a couple years where it dipped below. This is an important insight because it affirms our suspicion that adventure movies are appreciated by critics. If they are well liked among critics, we would expect the movies to perform well in the Box Office.



Data Visualization #1

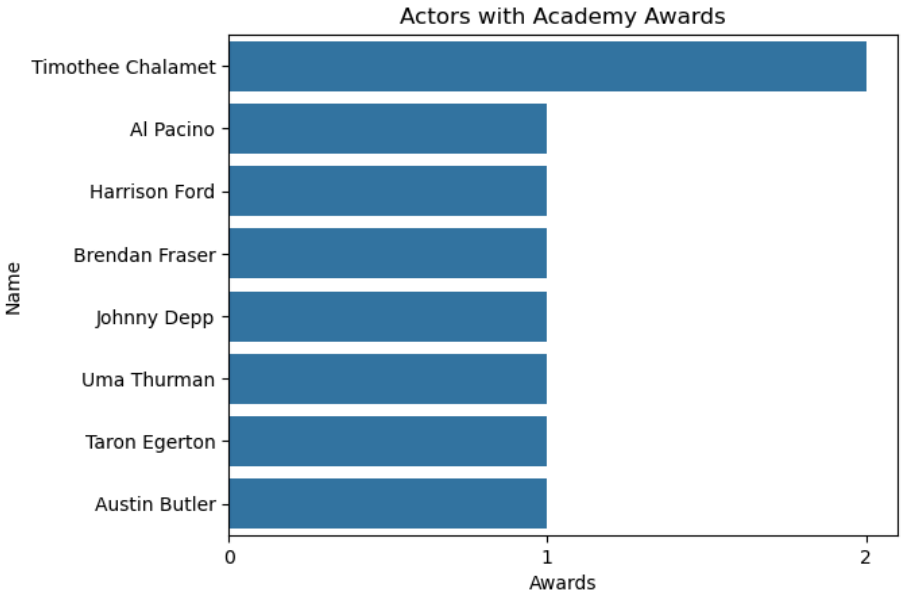
Since we established that the adventure movie category performs well among critics, we decided to see which directors had the lowest average Rotten Tomato score. Finding those directors informs us who to avoid, which is a powerful takeaway. We obtained this information by querying both the movie, directors, and director table. We first inner joined all the columns together, then grouped the values together based on the director name. Then, we selected the average of the movie Rotten Tomato scores and ascendingly

sorted those scores to obtain the list of directors below. The fifth lowest average score was roughly around 70, meaning that any director in our database outside the list of 5 are expected to deliver good movies.



Data Visualization #2

The last step we thought about was who to cast as the main actor/actress. In order to determine who would be the best fit, we looked at the actor and awards tables. We believed that the best actors for adventure movies would be the ones who have won academy awards while performing adventure movies. In order to gather the data, we joined the actor and awards tables. We then grouped the tables together by the actor name and kept track of their count of rows. As seen in the table below, Timothee Chalamet is the only actor in our database that has two academy awards. The selection of actors varies in age, however, there is only one female. This is an important insight because if the lead of our film is a woman, it significantly constricts our options.



Data Visualization #3

SECTION 7: CONCLUSION

LegendScreen aims to compile a comprehensive database of Adventure films and related attributes to draw actionable insights that position it for long-term success in the Adventure film industry. The database includes tables for key entities including actors, directors, composers, production studios, VFX studios, awards, characters, and platforms, among others. It establishes relationships between these entities to analyze trends, reveal predictors of financial success and audience satisfaction, determine resource allocation, and uncover industry dynamics that go beyond conventionally-used metrics like box office performance and critical reception.

While the database is robust in design, it overlooks potentially influential external factors like cultural/economic shifts and technological advancements. These factors can significantly impact audience preferences, industry trends, and film success. For instance, economic downturns, changing societal attitudes towards certain themes, and advancements in special effects technology could all dramatically alter the adventure film industry's landscape over a specific period. However, these influences are not explicitly accounted for in the database. Other limitations include the database's limited scope of data collection and the inherent bias in data selection, leading to the underrepresentation of non-mainstream films in addition to female directors and actresses.

If we continued working on the database for another semester, we'd focus on practical improvements to address the existing limitations. To enhance the database further, we could integrate sentiment analysis for audience reception, involving analyzing reviews and social media comments to more accurately gauge audience sentiments towards films or directors. It would also be beneficial to include data from foreign movies on box office performance and cultural reception, as it would shed light on the role of geography in the perception of movies. Moreover, tracking metrics such as advertising spend and segmenting audiences would provide insights into the effectiveness of promotional efforts, enabling LegendScreen to optimize its marketing campaigns for box office success.