

IMDB DATASET SUMMARY

After analysing the dataset, insights are as follows:

- There is no information of Worldwide gross income for 3724 movies.
- Most number of movies were produced in 2017 and quantity of movies produced is decreased in following years.

```
SELECT year,  
       Count(title) AS number_of_movies  
FROM   movie  
GROUP BY year;
```

- Most number of movies were most likely to be released in March. This might help in predicting the competition.

```
SELECT Month(date_published) AS month_num,  
       Count(title)          AS number_of_movies  
FROM   movie  
GROUP BY Month(date_published)  
ORDER BY Month(date_published);
```

- Overall highest number of movies produced belongs to 'Drama' Genre followed by 'Comedy' & 'Thriller' and hence, RSVP Movies should primarily focus on the 'Drama' genre.

```
SELECT genre,  
       Count(movie_id) AS no_of_movies  
FROM   movie AS m  
       INNER JOIN genre AS g  
           ON m.id = g.movie_id  
GROUP BY genre  
ORDER BY no_of_movies DESC  
LIMIT 1;
```

- As per 2019 data, 'Action' & 'Drama' genres have average duration of 113 & 107 minutes respectively. This might decide the duration of future movie to be produced.

```
SELECT genre,  
       Round(Avg(duration), 2) AS avg_duration  
FROM   movie AS m  
       INNER JOIN genre AS g  
           ON m.id = g.movie_id  
GROUP BY genre  
ORDER BY avg_duration DESC;
```

- Dream Warrior Pictures (Rank 1), National Theatre Live (Rank 2) or both can be considered by RSVP Movies as their production company as they have produced the greatest number of hit movies with an average rating of more than 8.

```
SELECT      production_company,
            Count(id)                                AS movie_count,
            Dense_rank() OVER (ORDER BY Count(id) DESC) AS prod_company_rank
FROM        movie                                    AS m
INNER JOIN  ratings                                  AS r
ON          m.id = r.movie_id
WHERE       avg_rating > 8
GROUP BY   production_company limit 5;
```

- Based on the number of votes received - Marvel Studios, Twentieth Century Fox & Warner Bros. are the top three production houses that can be considered by RSVP.

```
SELECT      m.production_company,
            Sum(r.total_votes)                        AS vote_count,
            Dense_rank() OVER (ORDER BY Sum(r.total_votes) DESC) AS prod_comp_rank
FROM        movie                                    AS m
INNER JOIN  ratings                                  AS r
ON          m.id = r.movie_id
GROUP BY   m.production_company limit 3;
```

- If multi-lingual movie is being considered, then Star Cinema and Twentieth Century Fox can be considered as production houses as they produced the highest number of hits with an average rating of more than 8

```
SELECT      production_company,
            Count(movie_id)                          AS movie_count,
            Dense_rank() OVER(ORDER BY Count(movie_id) DESC) AS prod_comp_rank
FROM        movie                                    AS m
JOIN        ratings                                  AS r
ON          m.id = r.movie_id
WHERE       r.median_rating >= 8
AND         m.production_company IS NOT NULL
AND         position(',', IN m.languages) > 0
GROUP BY   m.production_company limit 2;
```

- James Mangold, Anthony Russo and Soubin Shahir are the top three directors in the top three genres whose movies have an average rating > 8. Hence one of these can be hired as the director for RSVP's next project.

```
WITH top_3_genres AS
(
    SELECT    genre,
              Count(g.movie_id) AS movie_counts
    FROM      genre           AS g
    JOIN      ratings         AS r
    ON        g.movie_id = r.movie_id
    WHERE     avg_rating > 8
    GROUP BY  g.genre
    ORDER BY  movie_counts DESC limit 3 )
SELECT      n.NAME           AS director_name,
              Count(g.movie_id) AS movie_count
FROM        names            AS n
JOIN        director_mapping AS d
ON          n.id = d.name_id
JOIN        genre AS g
ON          d.movie_id = g.movie_id
JOIN        ratings AS r
ON          r.movie_id = g.movie_id,
            top_3_genres
WHERE       g.genre IN (top_3_genres.genre)
AND         avg_rating > 8
GROUP BY    director_name
ORDER BY    movie_count DESC limit 3;
```

- Mammooty and Mohanlal are the top two actors based on median rating. Hence hiring one of them favours RSVP.

```
SELECT DISTINCT name AS actor_name,
                Count(r.movie_id) AS movie_count
FROM ratings AS r
    INNER JOIN role_mapping AS rm
        ON r.movie_id = rm.movie_id
    INNER JOIN names AS n
        ON rm.name_id = n.id
WHERE r.median_rating >= 8
      AND rm.category = 'actor'
GROUP BY n.name
ORDER BY movie_count DESC
LIMIT 2;
```

- Based on the movie count, average rating and Indian as criteria, Vijay Sethupathi can be hired as the additional actor for the next project.

```
WITH indian_actor_rank
AS (SELECT n.NAME,
          Sum(total_votes) AS
          total_votes,
          Count(r.movie_id) AS
          movie_count,
          Round(Sum(avg_rating * total_votes) / Sum(total_votes), 2) AS
          actor_avg_rating
FROM movie AS m
    JOIN ratings AS r
        ON m.id = r.movie_id
    JOIN role_mapping AS rm
        ON m.id = rm.movie_id
    JOIN names AS n
        ON rm.name_id = n.id
WHERE country LIKE 'India'
      AND category = 'actor'
GROUP BY NAME)
SELECT *,
       Rank()
       OVER (
           ORDER BY actor_avg_rating DESC, total_votes DESC) AS actor_rank
FROM indian_actor_rank
WHERE movie_count >= 5;
```

- Based on movie count, average rating and Indian as criteria, Taapsee Pannu can be hired as the actress for the next project.

```
SELECT  NAME AS actress_name,
        total_votes,
        Count(m.id)                                AS movie_count,
        Round(Sum(avg_rating*total_votes)/Sum(total_votes),2) AS actress_avg_rating,
        Dense_rank() OVER (ORDER BY Avg(avg_rating) DESC, total_votes DESC) AS actress_rank
FROM    names AS n
JOIN    role_mapping AS rm
ON      n.id = rm.name_id
JOIN    ratings AS r
ON      rm.movie_id = r.movie_id
JOIN    movie AS m
ON      r.movie_id = m.id
WHERE   category = "actress"
AND     country LIKE "%India%"
AND     languages LIKE "%Hindi%"
GROUP BY NAME
HAVING  Count(m.id) >= 3 limit 5;
```

- Parvathy Thiruvothu, Susan Brown & Amanda Lawrence are the top 3 actresses respectively based on number of Super Hit movies (average rating >8) in 'Drama' genre. If movie is planned to be 'Drama' genre, then considering one of them will be a plus for RSVP movies.

```
WITH top AS
(
    SELECT  n.NAME AS actress_name,
            Sum(total_votes) AS total_votes,
            Count(m.id) AS movie_count,
            r.avg_rating
    FROM    names AS n
    INNER JOIN role_mapping ro
    ON      n.id = ro.name_id
    INNER JOIN ratings r
    ON      ro.movie_id = r.movie_id
    INNER JOIN movie m
    ON      m.id = r.movie_id
    INNER JOIN genre g
    ON      m.id = g.movie_id
    WHERE   category = 'Actress'
    AND     genre = 'Drama'
    AND     avg_rating > 8
    GROUP BY actress_name )

SELECT  *,
        Dense_rank () OVER ( ORDER BY movie_count DESC) AS actress_rank
FROM    top LIMIT 3;
```