FINAL CAPSTONE PROJECT: IMDB DATA ANALYSIS WITH SQL #5

▶ Team members:

Rina Rafalski
Intern Code
OGTIPIRDA459

Emem Elice
Intern Code
OGTIPEABA474

▶ Mentor:
Aritri Debnath Ghosh

Oeson
Inspiring generation

Table of contents

Introduction

▶ RSVP Movies, a prominent Indian film production company, has achieved notable success in the domestic market by producing several super-hit films. As they venture into global cinema with an upcoming movie set for a 2024 release, they seek to adopt a data-driven approach to optimize their strategies and ensure international success. Recognizing the need for precise and insightful data analysis, RSVP Movies has enlisted our expertise in data analytics to guide their decision-making process.
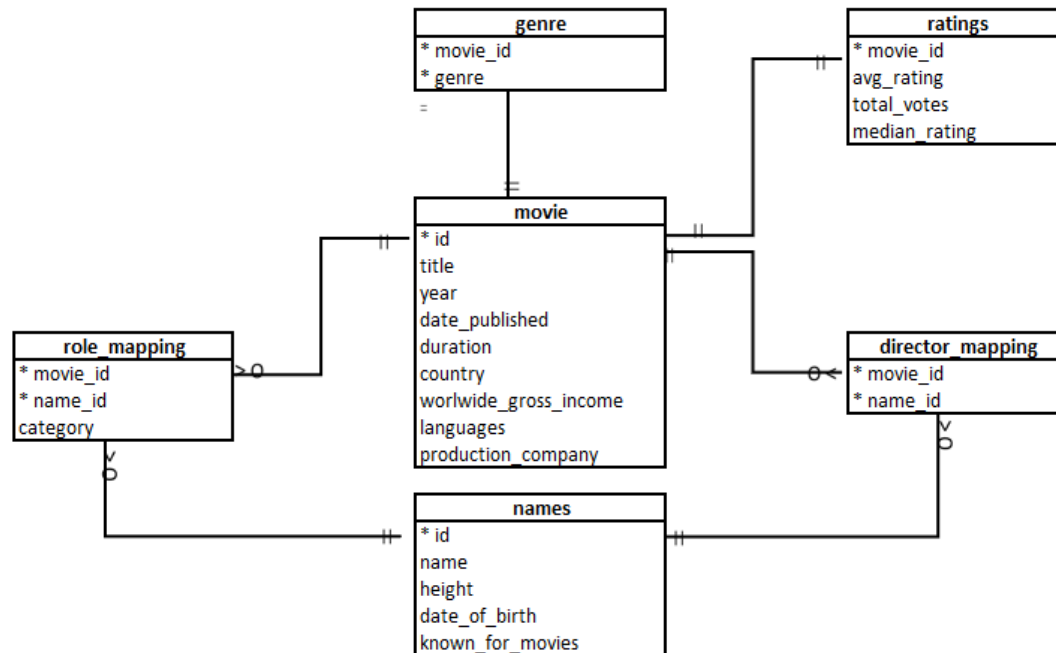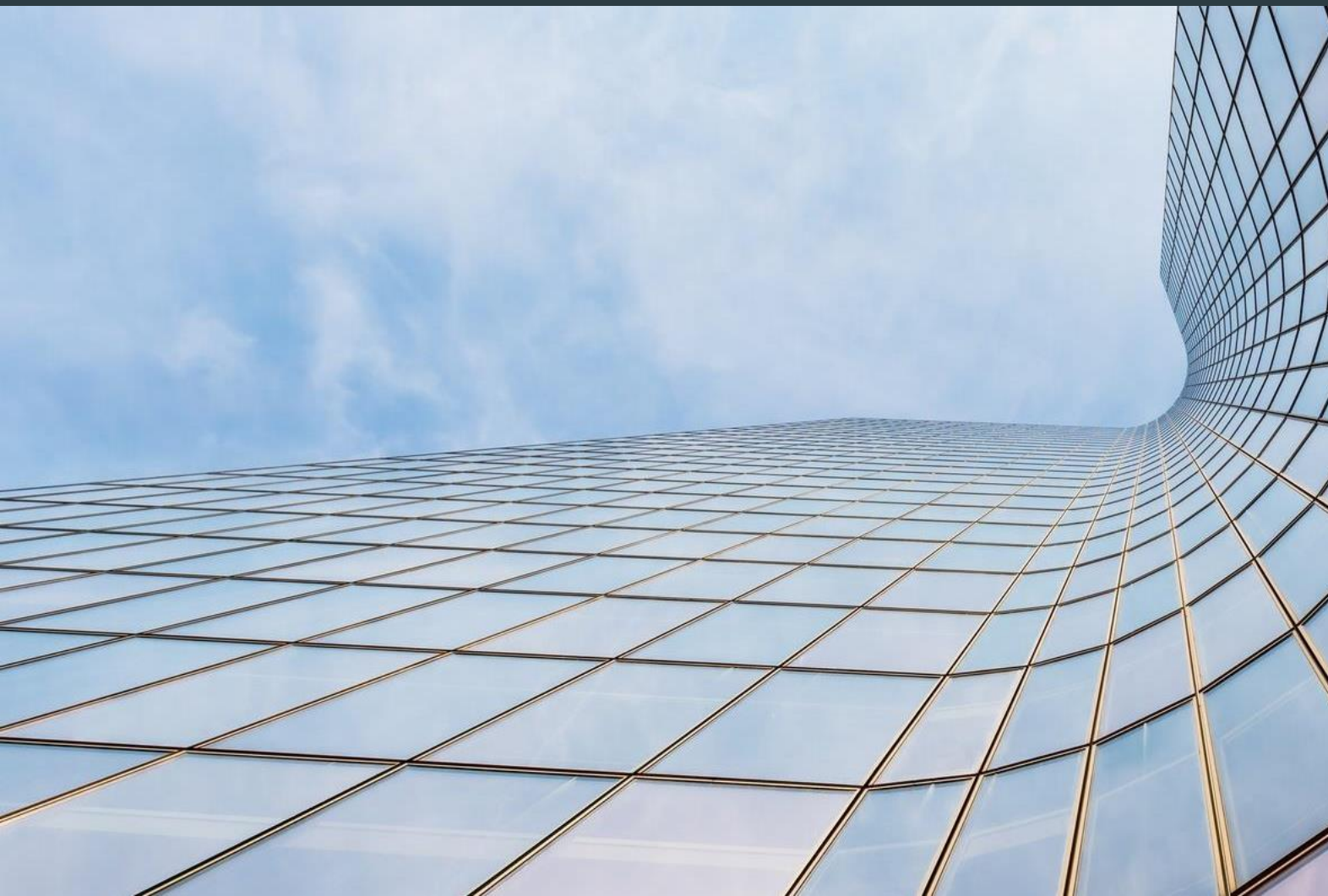
Project Objective

▶ The main goal of this project is to analyze data from films released over the past three years to provide RSVP Movies with actionable insights for their first global film release in 2024.

▶ Using SQL, we will identify key trends and success factors relevant to international audiences. Our findings will inform strategic recommendations, aiming to position RSVP Movies' upcoming film for global success and enhance their presence in the international film industry.

Data overview

IMDb data set consists of 6 tables: director_mapping (3867 rows), genre (14662 rows), movie (7185 rows), names (23683 rows), ratings (8230 rows), role_mapping (14394 rows)
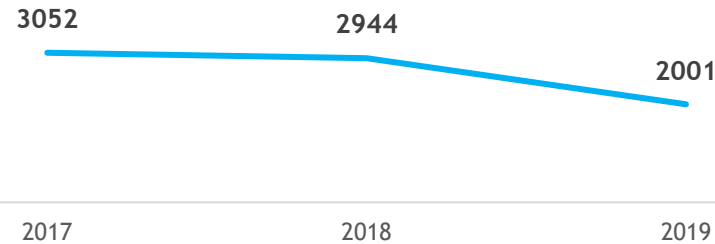
DATA ANALYSIS
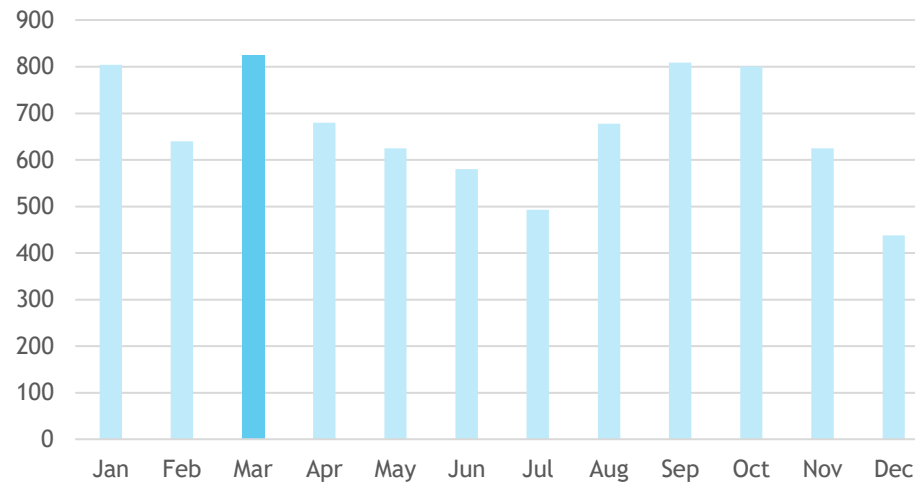Segment 1

Total number of movies released

- **Total number of movies released each year**
  According to the data of 3 years (2017 - 2019) total number of released movies decreases from year to year – from 3052 in the year 2017 to 2001 in the year 2019

- **How does the trend look month wise?**
  The highest number of movies is produced in the month of March.

Movies released year wise

3052   2944

2001

2017   2018   2019

Movies released month wise

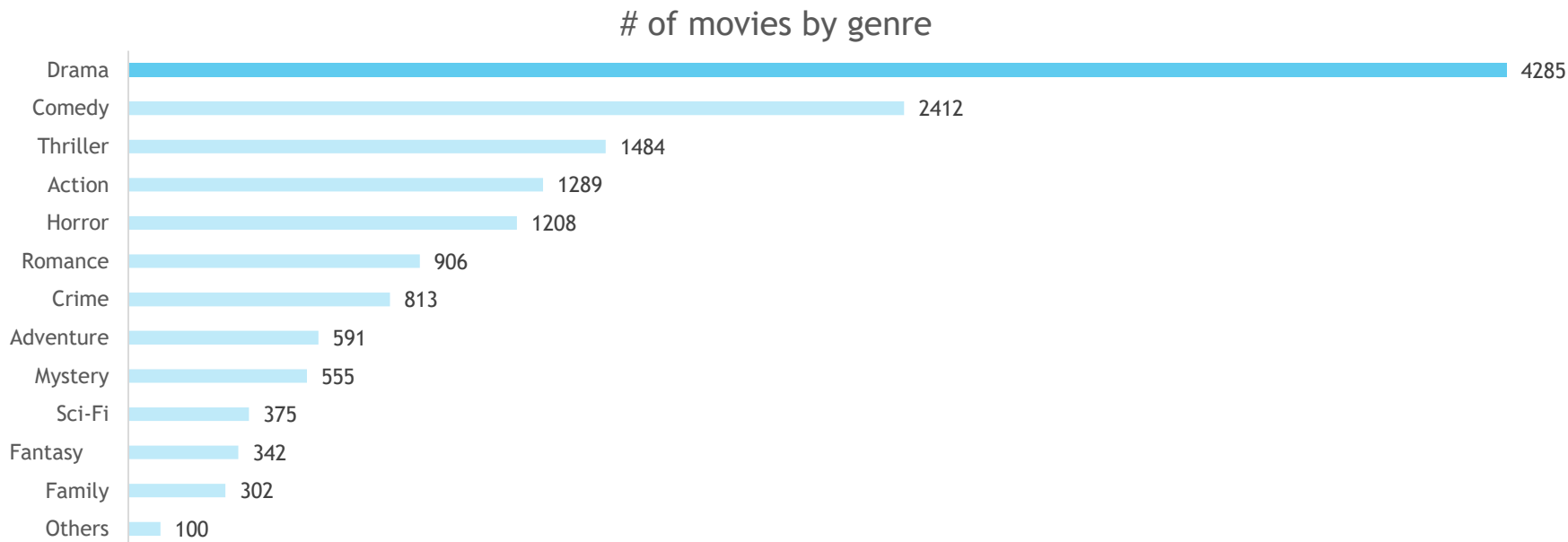| year	number_of_movies
2017	3052
2018	2944
2019	2001

month_num	number_of_movies
1	804
2	640
3	824
4	680
5	625
6	580
7	493
8	678
9	809
10	801
11	625
12	438

Number of movies by genre

- **Number of movies by genre**
  The overall number of drama films produced was 4285, the most of any genre. 3289 movies adhere to only one genre. Therefore, RSVP Movies should focus on the Drama genre.
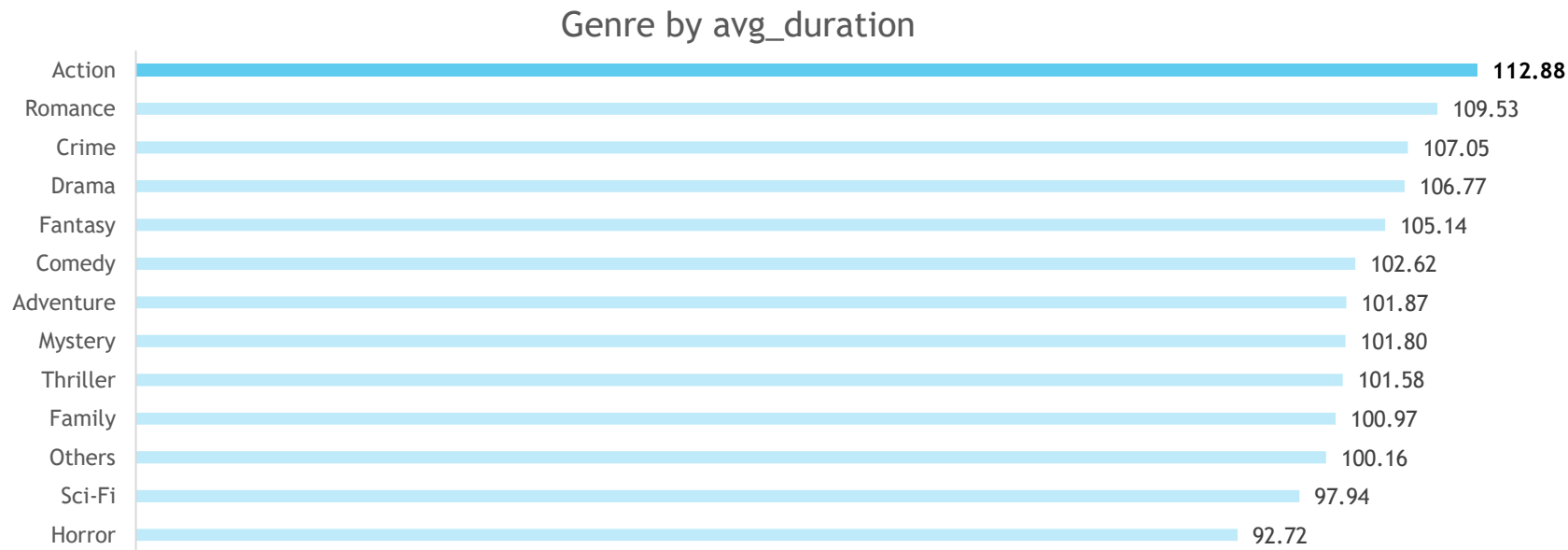
genre	movie_count
Drama	4285
Comedy	2412
Thriller	1484
Action	1289
Horror	1208
Romance	906
Crime	813
Adventure	591
Mystery	555
Sci-Fi	375
Fantasy	342
Family	302
Others	100

of movies by genre

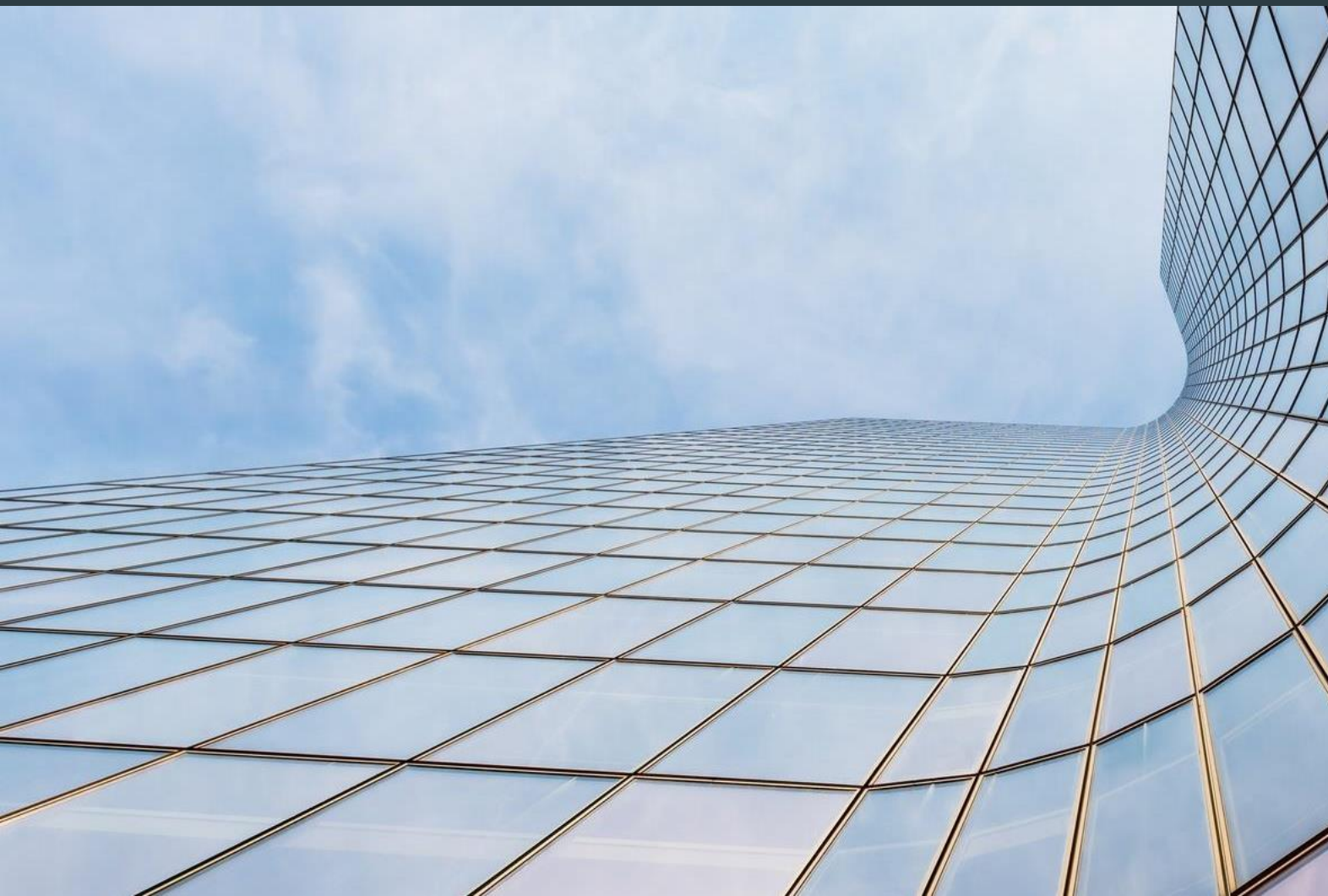| Genre	Count
Drama	4285
Comedy	2412
Thriller	1484
Action	1289
Horror	1208
Romance	906
Crime	813
Adventure	591
Mystery	555
Sci-Fi	375
Fantasy	342
Family	302
Others	100

Genre by average duration

- **Genre by average duration**
  The action genre has the longest duration (112.88 seconds), followed by the romance and crime genres. This provides insight into the likely runtime of 'RSVP Movies' in the upcoming project.

Genre by avg_duration

Genre	avg_duration
Action	112.88
Romance	109.53
Crime	107.05
Drama	106.77
Fantasy	105.14
Comedy	102.62
Adventure	101.87
Mystery	101.80
Thriller	101.58
Family	100.97
Others	100.16
Sci-Fi	97.94
Horror	92.72

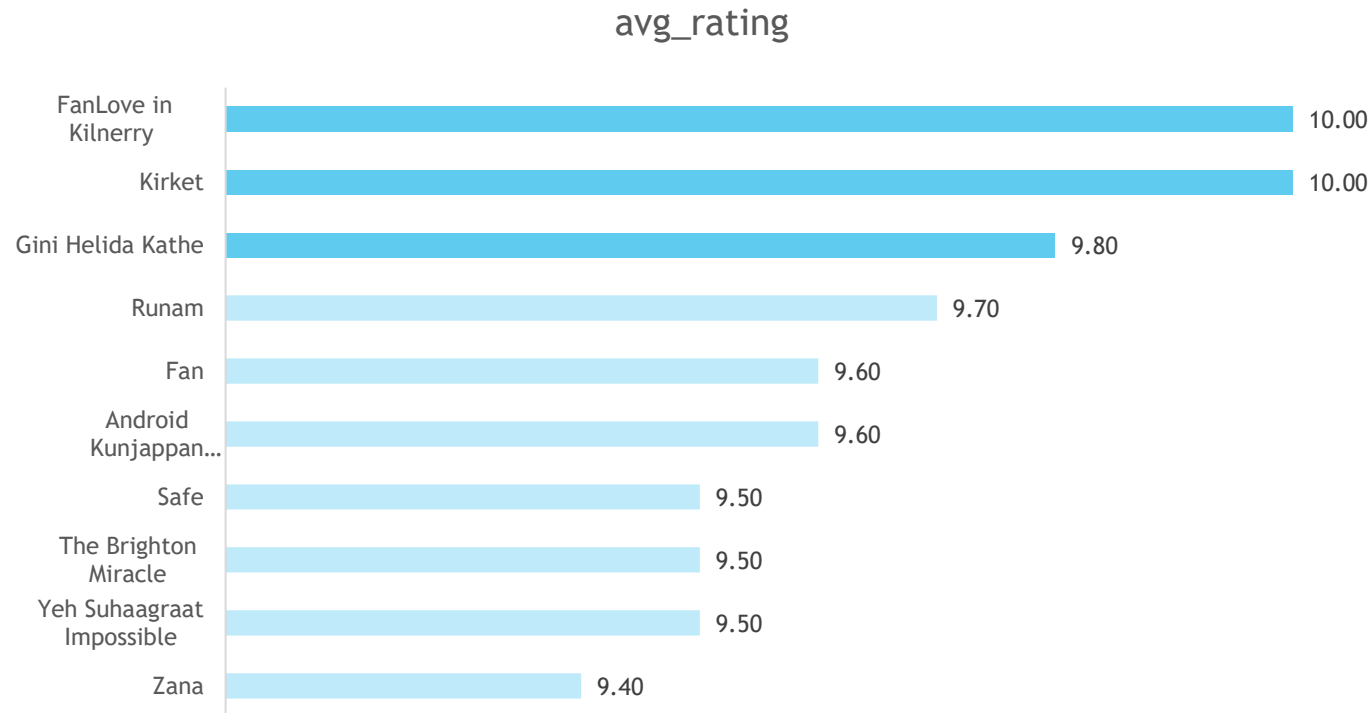| genre	avg_duration
Action	112.8829
Romance	109.5342
Crime	107.0517
Drama	106.7746
Fantasy	105.1404
Comedy	102.6227
Adventure	101.8714
Mystery	101.8000
Thriller	101.5761
Family	100.9669
Others	100.1600
Sci-Fi	97.9413
Horror	92.7243

DATA ANALYSIS
Segment 2

Top 10 movies by average rating

- **Top 10 movies by average rating**
  The first three movies have an **average rating** of more than **9.8**
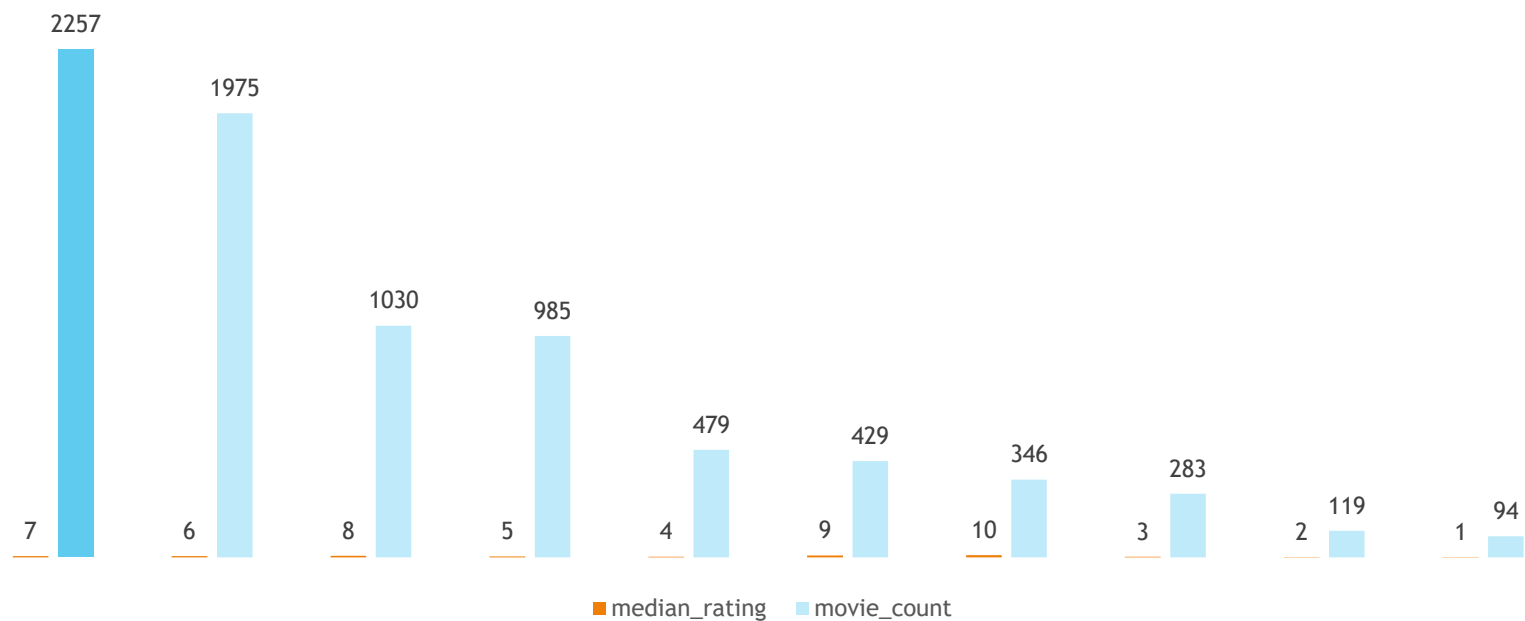
avg_rating

title	average_rating
Love in Kilnerry	10.00000
Kirket	10.00000
Gini Helida Kathe	9.80000
Runam	9.70000
Fan	9.60000
Android Kunjappan Version 5.25	9.60000
Yeh Suhaagraat Impossible	9.50000
Safe	9.50000
The Brighton Miracle	9.50000
Shibu	9.40000

Movie by median rating

- **Movie by median rating**
  The highest movie count of 2257 is found in the median rating of 7.

Movie by median rating

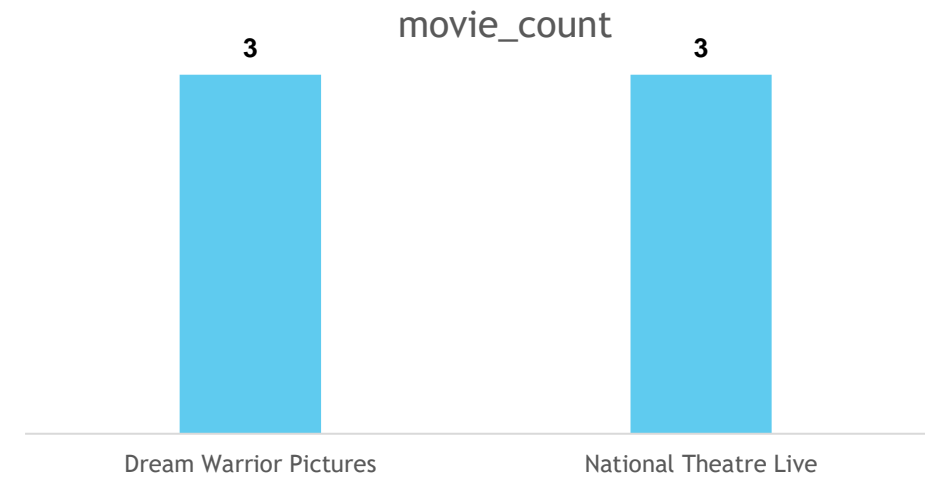| median_rating	movie_count
7	2257
6	1975
8	1030
5	985
4	479
9	429
10	346
3	283
2	119
1	94

Production houses with the most hit movies

- **Production house with the most hit movies**
  The production companies **Dream Warrior Pictures** and **National Theatre Live** have made the most amount of highly rated movies (average rating > 8). Their rank is 1 and their movie count is 3.
  Thus, RSVP Movies can partner with Dream Warrior Pictures and National Theatre Live for its next project

production_company	movie_count	prod_company_rank
Dream Warrior Pictures	3	1
National Theatre Live	3	1

movie_count

3	3
Dream Warrior Pictures	National Theatre Live

DATA ANALYSIS
Segment 3

Talent Analysis for global audience

- **Top 3 directors**
  By analyzing movies with average rating > 8 in top 3 genres we found the top 3 directors: James Mangold, Marianne Elliott, Adesh Prasad.

- **Top 2 actors**
  By analyzing movies with median rating >= 8 we found the top 2 actors: Mammootty, Mohanlal

- The knowledge allows the company to potentially collaborate with these talents for future projects, ensuring high-quality content that resonates well with global audiences.

director_name	movie_count
James Mangold	2
Marianne Elliott	2
Adesh Prasad	1

actor_name	movie_count
Mammootty	8
Mohanlal	5

Top Production Houses by audience votes

- **Top production houses by audience votes**
  These are the top 3 production houses based on the number of votes received by the movies they have produced which can create partnership opportunities for RSVP production. In this case Marvel Studios with a 2656967 votes could be a potential partner

- Partnering with top production houses can facilitate better placement, promotion, and visibility of their movies in the global market, increasing their competitiveness.

production_company	vote_count	prod_comp_rank
Marvel Studios	2656967	1
Twentieth Century Fox	2411163	2
Warner Bros.	2396057	3

Talent Analysis for Local Audience

- Since RSVP Movies is based out of Mumbai, India also wants to woo its local audience. RSVP Movies also wants to hire a few Indian actors for its upcoming project to give a regional feel.

- By ranking actors with movies released in India based on their average ratings that acted in at least 5 Indian movies we found the actors at the top of the list.
  The top actor is Vijay Sethupathi with average rating 8.42

- We found the top 5 actresses in Hindi movies released in India based on their average ratings.
  Taapsee Pannu tops with average rating 7.70

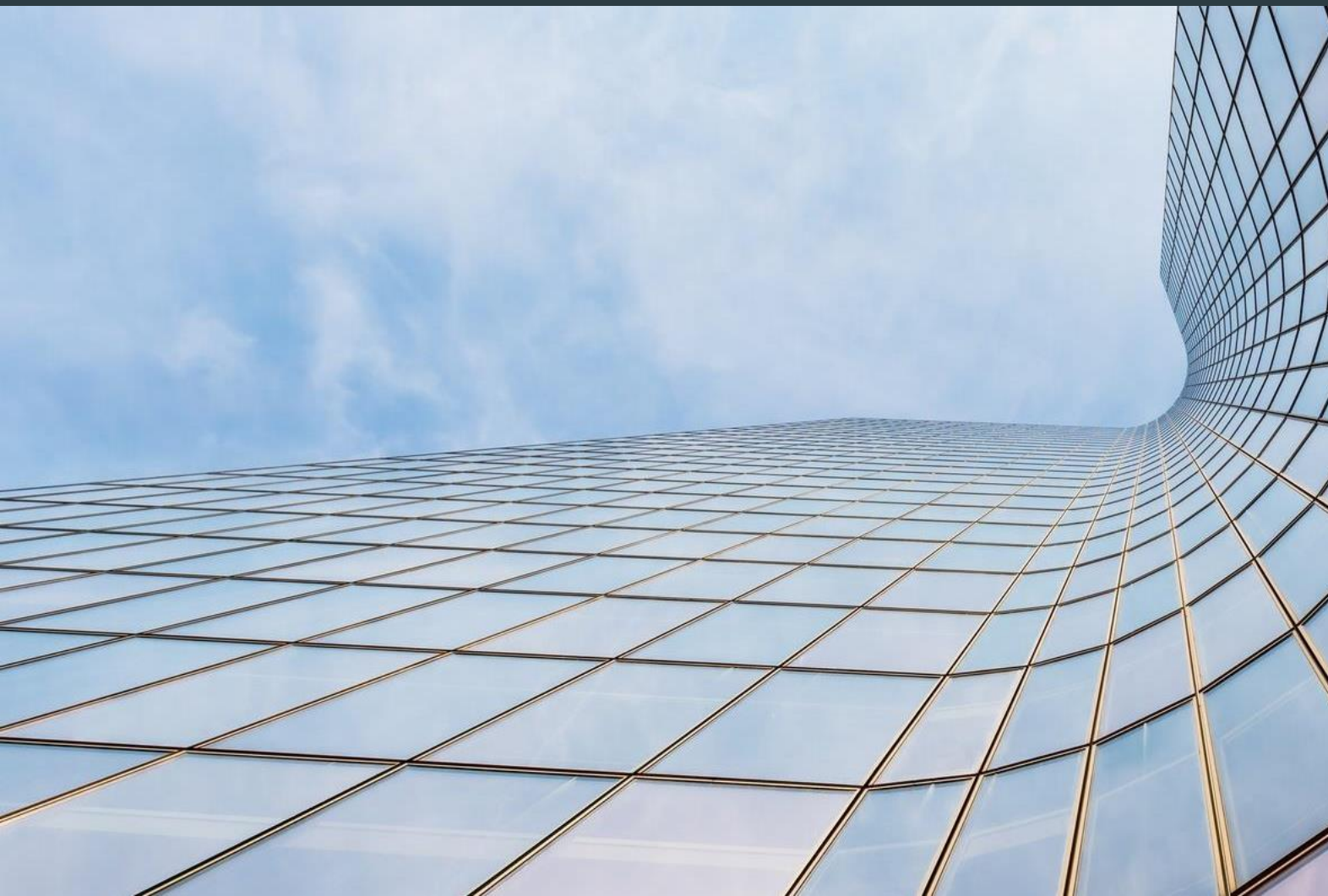| actor_name	total_votes	movie_count	actor_avg_rating	actor_rank
Vijay Sethupathi	23114	5	8.42	1
Fahadh Faasil	13557	5	7.99	2
Yogi Babu	8500	11	7.83	3
Joju George	3926	5	7.58	4
Ammy Virk	2504	6	7.55	5

actress_name	total_votes	movie_count	actress_avg_rating	actress_rank
Taapsee Pannu	18895	5	7.70	1
Raashi Khanna	9816	5	7.01	2
Manju Warrier	11276	5	6.76	3
Nayanthara	8962	6	6.68	4
Sonam Bajwa	2109	5	6.44	5

Thriller movies quality ranking

- We analyzed the quality of thriller movies
  and classified them by average rating
  in the following categories:
  - Rating > 8: Superhit movies
  - Rating between 7 and 8: Hit movies
  - Rating between 5 and 7: One-time-watch
  - Rating < 5: Flop movies

Here we can see the top rank
thriller movies in each category:

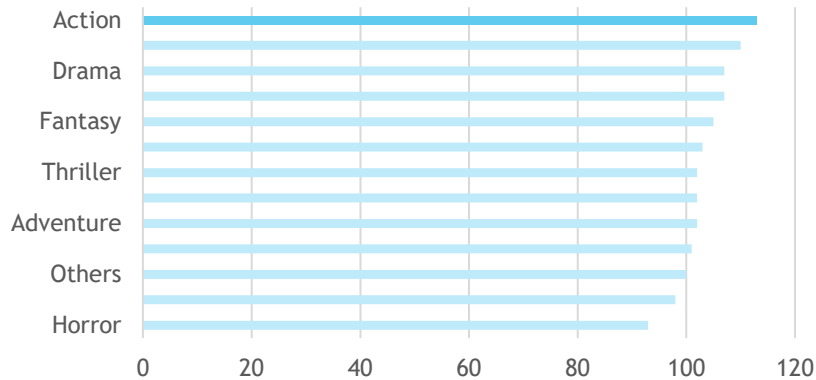| title	avg_rating	avg_rating_category	avg_rating_category_rank
Safe	9.5	Superhit movies	1
Until Midnight	8.0	Hit movies	1
Freaks	6.9	One-time-watch movies	1
Paralytic	4.9	Flop movies	1

DATA ANALYSIS
Segment 4

Movie duration analysis
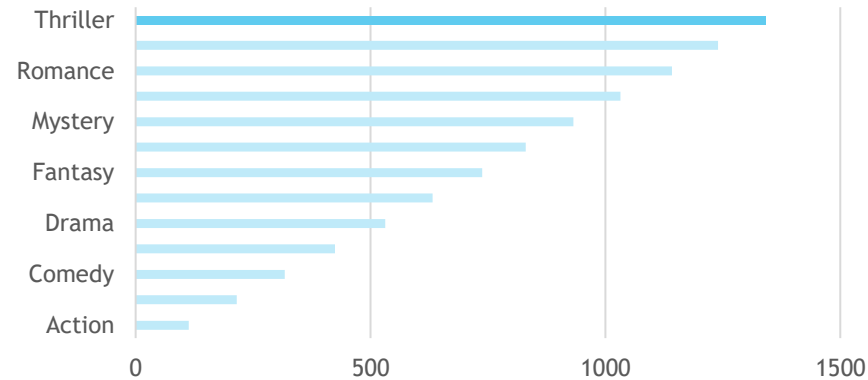
- **Movie duration genre-wise**
  By analyzing movie duration genre-wise we found that movies of genre
  Action have the largest average (static mean) and moving average
  (cumulative mean) duration, movies of genre Thriller have the largest
  running total (cumulative) duration

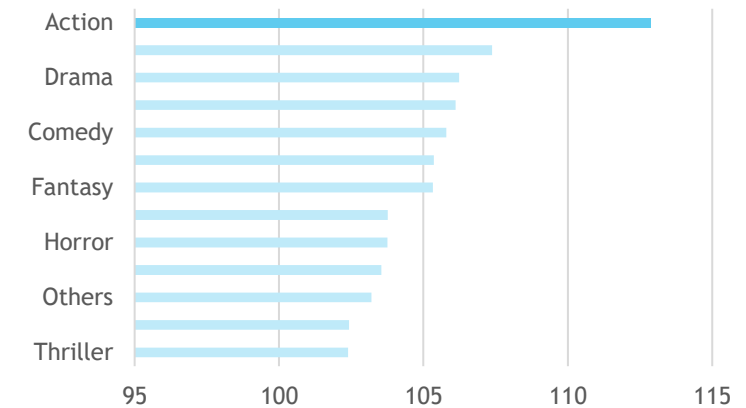| genre	avg_duration	running_total_duration	moving_avg_duration
Action	113	112.88	112.88
Romance	110	1141.51	103.77
Crime	107	424.42	106.11
Drama	107	531.19	106.24
Fantasy	105	737.30	105.33
Comedy	103	317.37	105.79
Adventure	102	214.75	107.38
Mystery	102	931.82	103.54
Thriller	102	1341.03	102.39
Family	101	632.16	105.36
Others	100	1031.98	103.20
Sci-Fi	98	1239.45	102.42
Horror	93	830.02	103.75

avg_duration

running_total_duration

moving_avg_duration

Top Production Houses among multilingual movies

- **Top production houses among multilingual movies**
  We analyzed production houses that have produced the highest number of hits among multilingual movies.

- Top 2 production houses in multilingual cinema are Star Cinema and 20th Century Fox that recognized for their substantial track record of successes, quantified by a median rating of 8 or above.

- Collaborations with leading production houses can enhance a movie's market positioning, promotional reach, and worldwide visibility, thus elevating its global market standing.

production_company	movie_count	prod_comp_rank
Star Cinema	7	1
Twentieth Century Fox	4	2

Talent analysis

- Sangeeta Bhat leads the top 3 actresses in the ranking based on number of Super Hit movies (average rating >8) in drama genre

actress_name	total_votes	movie_count	actress_avg_rating	actress_rank
Sangeetha Bhat	1010	1	9.6	1
Fatmire Sahiti	3932	1	9.4	2
Adriana Matoshi	3932	1	9.4	2

- Andrew Jones leads the top 9 directors in the ranking based on number of movies

director_id	director_name	number_of_movies	avg_inter_movie_days	avg_rating	total_votes	min_rating	max_rating	total_duration	director_rank
nm2096009	Andrew Jones	5	191	3.02	1989	2.7	3.2	432	1
nm1777967	A.L. Vijay	5	177	5.42	1754	3.7	6.9	613	2
nm6356309	Özgür Bakar	4	112	3.75	1092	3.1	4.9	374	3
nm2691863	Justin Price	4	315	4.50	5343	3.0	5.8	346	4
nm0814469	Sion Sono	4	331	6.03	2972	5.4	6.4	502	5
nm0831321	Chris Stokes	4	198	4.33	3664	4.0	4.6	352	6
nm0425364	Jesse V. Johnson	4	299	5.45	14778	4.2	6.5	383	7
nm0001752	Steven Soderbergh	4	254	6.48	171684	6.2	7.0	401	8
nm0515005	Sam Liu	4	260	6.23	28557	5.8	6.7	312	9

CONCLUSIONS AND RECOMMENDATONS
Summary

Conclusions and recommendations

- Major number of movies were released in the month of March, thus releasing in **March** might not be much profitable due to strong competition.

- Studying yearly and monthly movie release data helps grasp global market trends, anticipate prime release times, and smartly roll out Indian films for optimal impact.

- Finding top-rated movies by average ratings assists producers in comparing quality, gauging performance against industry benchmarks, and making informed production decisions.

- The company should focus more on releasing/making movie for **Drama** genre as it is the most popular genre and most produced genre in terms of number of movies produced.

- Perhaps considering the **Drama** as the key genre to focus upon, the **average duration** should be approximately **107 minutes**.

- RSVP Movies should partner with **Dream Warrior Pictures** or **National Theatre Live** production houses as they have produced the most number of hit movies with average rating greater than **8.**

Conclusions and recommendations

- The most popular director **James Mangold**, acclaimed in the top 3 genres with ratings over 8, is the prime directorial candidate for RSVP's film project

- **Mammootty** or **Mohanlal** could be hired as actors based on the median ratings and number of movies. As they were the top actors from the desired insights.

- The actor **Vija Sethupathi** and the actress **Tapsee Pannu** could be hired for the next movie project as they were the most popular based on the average rating and total votes.

- RSVP can opt to partner with production house **Marvel Studios** as their global partners based on highest votes received.

- RSVP should consider hiring **Andrew Jones** as he is the top ranked director.

- For multilingual movies, RSVP can partner with **Star Cinema** or **20th Century Fox** as they are top 2 production houses in terms of multilingual movies.

THANK YOU

CONTACT

in [Rina Rafalski](#)
@ [rinaraf@gmail.com](#)

in [Emem Alice](#)
@ emem.alice@yahoo.com