Success Code of Netflix Original Movies

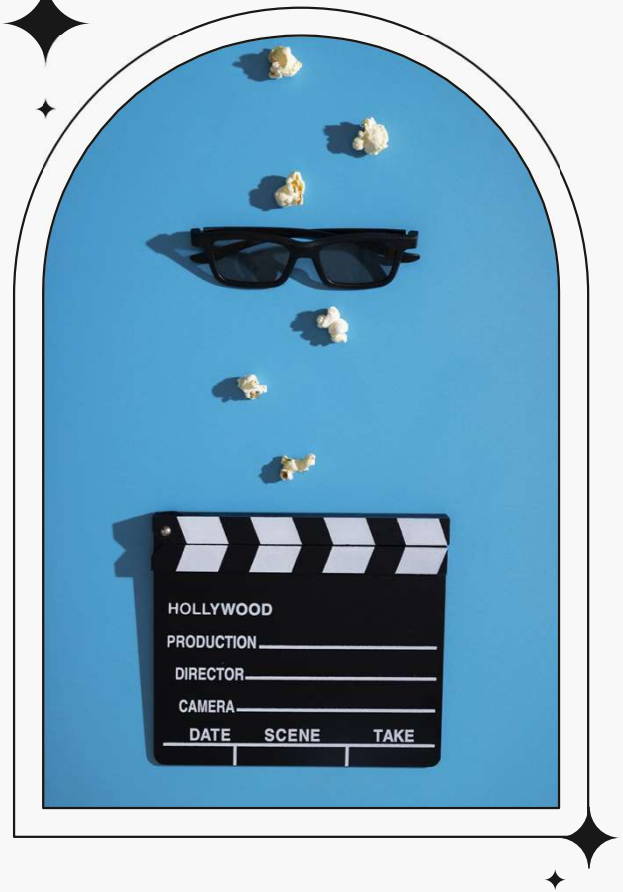

Table of contents

Introduction

This research aims to uncover what makes Netflix original movies successful, considering genre, runtime, IMDB score, and language. Insights gained will assist Netflix in creating more appealing content, ultimately attracting a larger audience and enhancing its position in the OTT platform market.

Project Scope

1. Analysis will focus on Netflix original movies across diverse genres, languages, and regions.

2. Attributes such as genre, runtime, IMDB score, and language will be examined to identify characteristics associated with success.

3. The analysis will utilize a comprehensive dataset covering a specified time period to ensure robust insights.

4. Stakeholder engagement will be included to align findings with organizational goals and facilitate informed decision-making within Netflix.

5. The scope will encompass understanding audience preferences and viewing behaviors to optimize content planning and production strategies.

Goals & KPIs

1. **Goal:** Enhance Viewer Engagement and Retention

   KPI: Average View Duration

   Explanation: This goal aims to improve viewer engagement and retention by measuring the average duration of views for Netflix original movies. A higher average view duration indicates that viewers are more engaged with the content, which can lead to increased retention and satisfaction.

2. **Goal:** Optimize Content Quality and Appeal

   KPI: Average IMDB Score

   Explanation: This goal focuses on optimizing the quality and appeal of Netflix original average IMDB score indicates that the content is well-received by viewers and has a positive impact on audience perception and satisfactional movies by tracking the average IMDB score.

3. **Goal:** Expand Market Reach and Diversity

   KPI: Language Diversity

   Explanation: This goal aims to expand the market reach and diversity of Netflix original movies by monitoring language diversity. The KPI measures the number of languages available for content, reflecting Netflix's efforts to cater to a global audience and increase accessibility to diverse linguistic communities.

Insights

1. **What is the average IMDB score for each genre?**
   The SQL query calculates the average IMDB score for each primary genre of Netflix original movies. It utilizes a subquery to extract the primary genre from the 'Genre' column, employing string manipulation functions. The results are then grouped by the primary genre using the GROUP BY clause, allowing for the calculation of average IMDB scores within each genre category.

```sql
--Q1 What is the average IMDB score for each genre?
SELECT ISNULL(LEFT(genre, CHARINDEX('/', genre + '/') - 1), genre) AS primary_genre,
       ROUND(AVG(IMDB_Score),1) AS Avg_IMDB_Score
FROM dbo.NetflixOriginals
GROUP BY ISNULL(LEFT(genre, CHARINDEX('/', genre + '/') - 1), genre)
ORDER BY Avg_IMDB_Score DESC;
```

121 %

Results | Messages

	primary_genre	Avg_IMDB_Score
1	Anthology	7.6
2	Concert Film	7.6
3	Making-of	7.5
4	Action-adventure	7.3
5	Coming-of-age comedy-drama	7.2
6	Drama-Comedy	7.2
7	Historical drama	7.2
8	Mentalism special	7.1
9	One-man show	7.1
10	War drama	7.1
11	Christmas	7

2. **Which genre has the highest average runtime?**

   The SQL query retrieves the primary genre with the highest average runtime among Netflix original movies. It utilizes a subquery to extract the primary genre from the 'Genre' column and calculates the average runtime. The results are then grouped by primary genre, ordered by average runtime in descending order, and the top result is selected using the TOP 1 clause.

```sql
--Q2 Which genre has the highest average runtime?
SELECT TOP 1 ISNULL(LEFT(   genre, CHARINDEX('/', genre +'/') -1), genre) AS primary_genre,
       AVG(Runtime) AS Average_Runtime
FROM dbo.NetflixOriginals
GROUP BY ISNULL(LEFT(   genre, CHARINDEX('/', genre +'/') -1), genre)
ORDER BY Average_Runtime DESC;
```

121 %

Results | Messages

	primary_genre	Average_Runtime
1	Anthology	149

3. **Are longer movies rated higher on IMDB?**

   The SQL query selects the 'Runtime' and 'IMDB_Score' columns from the 'NetflixOriginals' table where the 'Runtime' is greater than or equal to 94 minutes. The results are then ordered by 'Runtime' in descending order. This reveals that shorter movies with higher IMDB scores exist, indicating that movie length does not necessarily correlate with higher ratings on IMDB.

```sql
--Q3 Are longer movies rated higher on IMDB?
SELECT Runtime, AVG(IMDB_Score) AS Average_Runtime
FROM dbo.NetflixOriginals
GROUP BY Runtime
ORDER BY Average_Runtime DESC;
```

21 %

Results | Messages

	Runtime	Average_Runtime
1	153	8.5
2	51	8.3
3	12	7.8
4	209	7.8
5	135	7.7
6	137	7.5
7	15	7.5
8	71	7.5
9	26	7.4
10	136	7.35
11	24	7.3

4. **What is the distribution of movies across different languages?**

The SQL query calculates the count of movies for each language in the 'NetflixOriginals' table and aliases it as 'MOVIE_COUNT'. It then groups the results by language and orders them by movie count in descending order. This provides insight into the distribution of movies across different languages, revealing which languages have the highest number of movies available on Netflix.

```sql
--Q4 What is the distribution of movies across different languages?
SELECT Language, COUNT(*) AS Movie_Count
FROM dbo.NetflixOriginals
GROUP BY Language
ORDER BY Movie_Count DESC;
```

121 %

Results | Messages

	Language	Movie_Count
1	English	401
2	Hindi	33
3	Spanish	31
4	French	20
5	Italian	14
6	Portuguese	12
7	Indonesian	9
8	Japanese	6
9	Korean	6
10	German	5
11	English/Spanish	5

5. **What is the trend in the number of releases over the years?**

   The SQL query utilizes the COUNT() function to calculate the total number of releases and the YEAR() function to extract the year from the 'Premiere' column. It groups the results by year of premiere, providing insight into the distribution of Netflix original releases over time. This analysis helps understand the yearly trend of new releases, facilitating strategic planning and decision-making regarding content production and scheduling.

```sql
--Q5 What is the trend in the number of releases over the years?
SELECT TOP 1 YEAR(Premiere) AS Years, Count(Premiere) AS Num_Of_Release
FROM dbo.NetflixOriginals
GROUP BY YEAR(Premiere)
ORDER BY Num_Of_Release DESC;
```

121 %

Results | Messages

	Years	Num_Of_Release
1	2020	183

6. **Which year had the highest average IMDB score?**

   The SQL query calculates the average IMDB score for Netflix original releases each year using the YEAR() function on 'Premiere' and the AVG() function on 'IMDB_Score'. It groups the results by year of premiere and orders them by average IMDB score in descending order. The TOP 1 clause selects the year with the highest average IMDB score, providing insight into the year with the most positively rated Netflix original releases.

```sql
--Q6 Which year had the highest average IMDB score?
SELECT TOP 1 YEAR(Premiere) AS Years, ROUND(AVG(IMDB_Score),1) AS Average_IMDB_Score
FROM dbo.NetflixOriginals
GROUP BY YEAR(Premiere)
ORDER BY Average_IMDB_Score DESC;
```

121 %

Results | Messages

	Years	Average_IMDB_Score
1	2015	6.9

7. **Are movies in certain languages rated higher on average?**

- The SQL query calculates the average IMDB score for Netflix original movies, grouping them by their primary language. It utilizes a subquery to extract the primary language from the 'Language' column, considering multiple languages if present. The results are then ordered by average IMDB score in descending order, revealing the highest-rated languages for Netflix original content.

```sql
--Q7 Are movies in certain languages rated higher on average?
SELECT TOP 5 ISNULL(LEFT(Language, CHARINDEX('/', Language + '/')-1), Language) AS primary_language,
       ROUND(AVG(IMDB_Score),1) AS Average_IMDB_Score
FROM dbo.NetflixOriginals
GROUP BY ISNULL(LEFT(Language, CHARINDEX('/', Language + '/')-1), Language)
ORDER BY Average_IMDB_Score DESC;
```

121 %

Results | Messages

	primary_language	Average_IMDB_Score
1	Khmer	7.2
2	Tamil	7.2
3	Bengali	7.1
4	Georgian	6.8
5	Thia	6.7

8. **What is the distribution of IMDB scores?**

- The SQL query categorizes Netflix original movies into IMDb score ranges using the CASE statement. It then counts the number of movies within each range and groups them accordingly. The results are ordered by IMDb score range, providing insights into the distribution of movie counts across different IMDb score brackets.

```sql
--Q8 What is the distribution of IMDB scores?
SELECT CONCAT(FLOOR(IMDB_Score - 0.5), '-', FLOOR(IMDB_Score + 0.5)) AS Score_range,
       COUNT(*) AS Movie_Count
FROM dbo.NetflixOriginals
GROUP BY CONCAT(FLOOR(IMDB_Score - 0.5), '-', FLOOR(IMDB_Score + 0.5))
```

121 %

Results | Messages

	Score_range	Movie_Count
1	2-3	5
2	3-4	20
3	4-5	82
4	5-6	213
5	6-7	212
6	7-8	49
7	8-9	3

9. **What is the most common genre on Netflix?**

   The SQL query identifies the most common primary genre among Netflix original movies by extracting and grouping genres. It uses a subquery to isolate the primary genre from the 'Genre' column, considering multiple genres if present. The TOP 1 clause selects the most frequent primary genre, providing insight into the predominant genre within the NetflixOriginals dataset.

```sql
--Q9 What is the most common genre on Netflix?
SELECT TOP 1 ISNULL(LEFT(genre, CHARINDEX('/', genre + '/') - 1), genre) AS primary_genre, COUNT(*) AS Movie_Count
FROM dbo.NetflixOriginals
GROUP BY ISNULL(LEFT(genre, CHARINDEX('/', genre + '/') - 1), genre)
ORDER BY Movie_Count DESC;
```

121 %

Results | Messages

	primary_genre	Movie_Count
1	Documentary	159

10. **What is the average runtime of movies released in different months?**

- The SQL query calculates the average runtime of Netflix original movies for each month of their premiere dates. It utilizes the MONTH() function to extract the month from the 'Premiere' column and the AVG() function to compute the average runtime. The results are grouped by month of premiere, providing insights into the seasonal trends of movie runtimes throughout the year.

```sql
--Q10 What is the average runtime of movies released in different months?
SELECT MONTH(Premiere) AS months, AVG(Runtime) AS Average_Runtime
FROM dbo.NetflixOriginals
GROUP BY MONTH(Premiere);
```

121 %

Results | Messages

	months	Average_Runtime
1	1	94
2	2	96
3	3	94
4	4	94
5	5	92
6	6	101
7	7	96
8	8	86
9	9	87
10	10	94
11	11	95

Thanks!