Success Code of Netflix Original Movies

Problem & Background:

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.

Solution:

1. **Data Analysis:**
   In this phase, we will gather and analyze data on various attributes of Netflix original movies including genre, runtime, IMDB score, and language. This will involve data collection, cleaning, and exploratory analysis to identify patterns and correlations among these attributes.

2. **Data-Driven Insights:**
   Based on the analysis conducted, we will derive actionable insights that highlight the characteristics of successful Netflix original movies. These insights will help in understanding the preferences of the audience, thereby aiding Netflix in strategic decision-making related to content planning and production.

3. **Stakeholder Management:**
   Effective communication of findings and insights to relevant stakeholders within Netflix, including content producers, directors, and executives, will be crucial. Stakeholder management involves presenting the analysis results in a clear and concise manner, fostering discussion, and aligning strategies with organizational goals to drive success in the OTT platform industry.

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.

Methodology:

1. **Data Sources:**
    i. **Kaggle:** Utilizing a dataset obtained from Kaggle, a popular platform for sharing datasets and machine learning projects. This dataset contains valuable information on Netflix original movies, including attributes such as genre, runtime, IMDB score, and language.
    ii. **SQL:** Querying data from existing SQL databases containing information on Netflix original movies.

2. **Data Wrangling:**
    i. **Data understanding:** Inspecting the structure and contents of tables in the database to understand the available attributes and their formats.
    ii. **Data cleaning:** Writing SQL queries to handle missing values, inconsistencies, and errors in the data within SQL tables.

iii. **Data manipulation:** Using SQL UPDATE, INSERT, DELETE, or other manipulation commands to transform and reformat data as needed within SQL tables**.**

3. **Data Analysis:**
   i. **Finding trends and patterns:** Executing SQL queries with aggregation functions (e.g., COUNT, SUM, AVG) and GROUP BY clauses to analyze trends and patterns in the data.

GOALS and 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 movies by tracking the average IMDB score. A higher average IMDB score indicates that the content is well-received by viewers and has a positive impact on audience perception and satisfaction.

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.

Technical Processes

1. Employing SQL for project data analysis.
2. Utilizing functions like COUNT(), AVG(), MAX(), MONTH(), and YEAR() to aggregate and analyze data. Employing GROUP BY and ORDER BY clauses for data grouping and sorting. Additionally, applying the TOP() function for limiting query results.

Business Concepts Used

In the project, several key business concepts are utilized to analyze and improve Netflix's original content strategy. Here's how each concept is applied:

1. **Market Understanding:** The project extensively analyzes market trends, competitors' strategies, and audience preferences within the OTT platform industry. This understanding helps Netflix make informed decisions about content creation, distribution, and marketing to stay competitive and attract a larger audience.

2. **Customer Demographic:** By segmenting the audience based on demographics such as age, gender, location, and viewing habits, Netflix gains insights into the preferences and interests of different customer groups. This segmentation informs content development and customization strategies to better cater to diverse audience demographics.

3. **Customer Behavior:** Analyzing customer behavior, including viewing patterns, preferences, and engagement metrics, provides valuable insights into how audiences interact with Netflix original content. By understanding customer behavior, Netflix can optimize content recommendations, user interface design, and content delivery to enhance the overall user experience and increase viewer engagement**.**

4. **Customer Retention:** Customer retention strategies aim to keep existing subscribers engaged and satisfied to reduce churn rates. Netflix achieves this by offering a diverse range of high-quality original content, personalized recommendations, and seamless viewing experiences. By prioritizing customer retention, Netflix ensures long-term loyalty and recurring revenue streams.

5. **New Customer Acquisition:** Netflix employs various tactics to attract new subscribers and expand its customer base. This includes targeted marketing campaigns, partnerships with content creators, and innovative features to differentiate its service from competitors. By focusing on new customer acquisition, Netflix sustains growth and maintains its position as a leading player in the OTT platform industry.

Recommended Analysis:

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.

```
SELECT primary_genre, AVG(IMDB_Score) AS Avg_IMDB_Score
FROM (SELECT
        CASE
        WHEN CHARINDEX('/', Genre) > 0 THEN SUBSTRING(Genre, 1, CHARINDEX('/', Genre) - 1)
        ELSE Genre
        END AS primary_genre, IMDB_Score
      FROM dbo.NetflixOriginals) AS subquery
GROUP BY primary_genre;
```

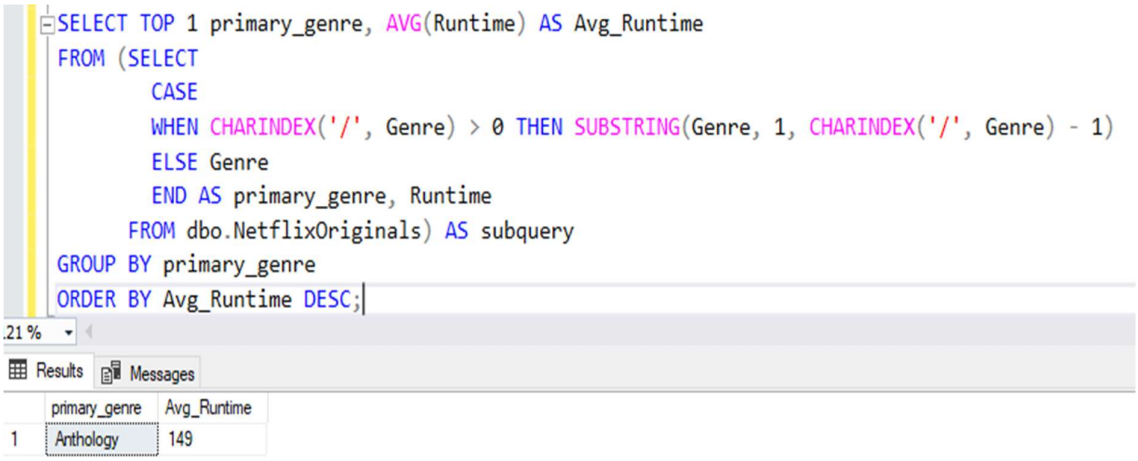
121 %

Results | Messages

	primary_genre	Avg_IMDB_Score
1	Action	5.411111
2	Action comedy	5.420000
3	Action thriller	6.400000
4	Action-adventure	7.300000
5	Action-thriller	6.133333
6	Adventure	6.033333
7	Adventure-romance	6.100000
8	Aftershow	6.750000
9	Animated musical comedy	6.700000
10	Animation	6.756250
11	Anime	6.150000

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.



```
SELECT TOP 1 primary_genre, AVG(Runtime) AS Avg_Runtime
FROM (SELECT
        CASE
        WHEN CHARINDEX('/', Genre) > 0 THEN SUBSTRING(Genre, 1, CHARINDEX('/', Genre) - 1)
        ELSE Genre
        END AS primary_genre, Runtime
      FROM dbo.NetflixOriginals) AS subquery
GROUP BY primary_genre
ORDER BY Avg_Runtime DESC;
```

21 %

Results | Messages

	primary_genre	Avg_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.

```
SELECT Runtime, IMDB_Score
FROM dbo.NetflixOriginals
WHERE Runtime > = 94
ORDER BY Runtime DESC;
```

121 %

Results | Messages

	Runtime	IMDB_Score
1	209	7.8
2	155	6.5
3	153	8.5
4	151	6.2
5	151	6.3
6	149	3.7
7	149	7.3
8	149	7.6
9	148	5.9
10	147	3.5
11	144	4.3

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.

```
SELECT COUNT(*) AS MOVIE_COUNT, Language
FROM dbo.NetflixOriginals
GROUP BY Language
ORDER BY  MOVIE_COUNT DESC;
```

21 %

Results | Messages

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

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
SELECT COUNT(*) AS Num_Of_Releases, YEAR(Premiere) as Years
FROM dbo.NetflixOriginals
GROUP BY YEAR(Premiere)
```

121 %

Results | Messages

	Num_Of_Releases	Years
1	1	2014
2	9	2015
3	30	2016
4	66	2017
5	99	2018
6	125	2019
7	183	2020
8	71	2021

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
SELECT TOP 1 YEAR(Premiere) AS Years, AVG(IMDB_Score) AS Avg_IMDB_Score
FROM dbo.NetflixOriginals
GROUP BY YEAR(Premiere)
ORDER BY Avg_IMDB_Score DESC;
```

121 %

Results | Messages

	Years	Avg_IMDB_Score
1	2015	6.877777

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.

```
SELECT primary_language, AVG(IMDB_Score) AS AVG_IMDB_Score
FROM (SELECT
        CASE
        WHEN CHARINDEX('/', Language) > 0 THEN SUBSTRING(Language, 1, CHARINDEX('/', Language) - 1)
        ELSE Language
        END AS primary_language, IMDB_Score
      FROM dbo.NetflixOriginals) AS subquery
GROUP BY primary_language
ORDER BY AVG_IMDB_Score DESC;
```

21 %

Results | Messages

	primary_language	AVG_IMDB_Score
1	Khmer	7.200000
2	Tamil	7.200000
3	Bengali	7.100000
4	Georgian	6.800000
5	Thia	6.700000
6	Japanese	6.400000
7	English	6.391408
8	Spanish	6.314705
9	Portuguese	6.216666
10	Marathi	6.066666
11	Hindi	5.981818

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
SELECT
    CASE
    WHEN IMDB_Score >= 0 AND IMDB_Score < 2 THEN '0-1.9'
    WHEN IMDB_Score >= 2 AND IMDB_Score < 4 THEN '2-3.9'
    WHEN IMDB_Score >= 4 AND IMDB_Score < 6 THEN '4-5.9'
    WHEN IMDB_Score >= 6 AND IMDB_Score < 8 THEN '6-7.9'
    WHEN IMDB_Score >= 8 THEN '8-10'
    END AS IMDb_Score_Range, COUNT(*) AS Movie_Count
FROM dbo.NetflixOriginals
GROUP BY
    CASE
    WHEN IMDB_Score >= 0 AND IMDB_Score < 2 THEN '0-1.9'
    WHEN IMDB_Score >= 2 AND IMDB_Score < 4 THEN '2-3.9'
    WHEN IMDB_Score >= 4 AND IMDB_Score < 6 THEN '4-5.9'
    WHEN IMDB_Score >= 6 AND IMDB_Score < 8 THEN '6-7.9'
    WHEN IMDB_Score >= 8 THEN '8-10'
    END
ORDER BY IMDb_Score_Range;
```

121 %

Results | Messages

	IMDb_Score_Range	Movie_Count
1	2-3.9	9
2	4-5.9	198
3	6-7.9	360
4	8-10	17

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
SELECT TOP 1 primary_genre
FROM (SELECT
        CASE
        WHEN CHARINDEX('/', Genre) > 0 THEN SUBSTRING(Genre, 1, CHARINDEX('/', Genre) - 1)
        ELSE Genre
        END AS primary_genre
      FROM dbo.NetflixOriginals) AS subquery
GROUP BY primary_genre;
```

21 %

Results | Messages

	primary_genre
1	Documentary

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
SELECT MONTH(Premiere) AS Month, AVG(Runtime) AS Avg_Runtime
FROM dbo.NetflixOriginals
GROUP BY MONTH(Premiere);
```

121 %

Results | Messages

	Month	Avg_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