**Analysis Netflix Movies and TV Shows**

1. **Netflix Movies and TV Shows** (Overview)

**About this Dataset:** [*Netflix*](https://en.wikipedia.org/wiki/Netflix)*is one of the most popular media and video streaming platforms. They have over 8000 movies or tv shows available on their platform, as of mid-2021, they have over 200M Subscribers globally. This dataset consists of listings of all the movies and tv shows available on Netflix, along with details such as - cast, directors, ratings, release year, duration, etc.*

**Data analysis & cleaning &** **the software I used it**

* For CleaningExcel& SQL Server **(SSMS).**
* Transform CSV File to Excel File by **(Excel)** to import in

SQL Server **(SSMS)** to Keep or Save Data Type in

SQL Server.

* Normalization by SQL Server **(SSMS).**
* For Coding SQL Server **(SSMS).**
* Cleaning after Coding by **(Power Query).**
* Calculations & Mensuration by **(DAX)**
* Visualization by **(Power BI).**
* Documentation by **(Word).**
* Presentation by **(PowerPoint).**

**-Interesting Task Ideas for Netflix**

1. Understanding what content is available in different countries
2. Identifying similar content by matching text-based features
3. Network analysis of Actors / Directors and find interesting insights
4. Does Netflix has more focus on TV Shows than movies in recent years.
5. **Understanding what content is available in different countries**

**Detailed explanation:**

This query calculates the number of movies and TV shows **(type)** available for each **country**.

It uses the **COUNT()** analytical function with **PARTITION BY** to split the results by **country** and type (movie or TV show).

The results are **sorted by country** and number of contents **in descending order.**

**Insight for Business:**

This query allows Netflix to understand the distribution of available content in each **country**, and **can help identify countries** that may need additional content or focus on optimizing their content library.

The company can focus on increasing the types of content in countries that lack diversity, enhancing the appeal of the service in these regions.

|  |
| --- |
| SELECT  c.country,  n.type,  COUNT(n.show\_id) OVER (PARTITION BY c.country, n.type) AS content\_count  FROM  Netflix\_Movies AS n INNER JOIN Netflix\_country AS c  ON n.show\_id = c.show\_id  ORDER BY  country, content\_count DESC; |

1. **Identifying similar content by matching text-based features (matching titles)**

**Detailed explanation:**

This query uses self-join to join the same table **(Netflix\_Movies)** to search for content with the same title.

The show identifiers **(show\_id)** are checked to ensure there is no duplication between the same content.

Displays fields that identify similar content such as **listed\_in (genre)**, **release\_year (release year), and country (country)**.

**Insight for Business:**

Netflix can use this analysis to understand which content is duplicated or repeated in different countries or within different genres.

It may help make smarter recommendations to viewers by promoting shows that are similar to what they've already watched.

|  |
| --- |
| SELECT  N1.title AS Title\_1,  N2.title AS Title\_2,  L1.listed\_in AS Genre\_1,  L2.listed\_in AS Genre\_2,  N1.release\_year AS Year\_1,  N2.release\_year AS Year\_2,  C1.country AS Country\_1,  C2.country AS Country\_2  FROM  Netflix\_Movies N1 INNER JOIN Netflix\_listed\_in L1  ON N1.show\_id = L1.show\_id INNER JOIN Netflix\_country C1  ON N1.show\_id = C1.show\_id  INNER JOIN  Netflix\_Movies N2 INNER JOIN Netflix\_listed\_in L2  ON N2.show\_id = L2.show\_id INNER JOIN Netflix\_country C2  ON N2.show\_id = C2.show\_id  ON N1.show\_id = N2.show\_id  AND N1.title = N2.title  ORDER BY  N1.title; |

1. **Network analysis of Actors/Directors: How often do actors and directors collaborate**

**Detailed explanation:**

This query **counts** the number of **collaborations** between a given director and an actor or group of actors **(cast).**

**COUNT(\*)** is used with **PARTITION BY** on director and cast to calculate the number of collaborations for each duo **(director - actor).**

**Insight for Business:**

This analysis allows Netflix to understand how successful collaborations between specific directors and actors are.

This data can be used to identify which directors and actors are most effective together and monetize more productions that bring them together.

|  |
| --- |
| SELECT --DISTINCT  d.director,  c.cast,n.type,  COUNT(\*) OVER (PARTITION BY d.director, c.cast) AS collaboration\_count  FROM  Netflix\_Movies AS n INNER JOIN Netflix\_director AS d  ON n.show\_id = d.show\_id INNER JOIN Netflix\_cast AS c  ON n.show\_id = c.show\_id  ORDER BY  collaboration\_count DESC; |

**4. Actors appearing in multiple shows (find the most frequent actors)**

**Detailed explanation:**

This query **displays the number of shows** each actor has appeared in using the **COUNT()** function with **PARTITION BY** on the **cast** column to count the number of appearances of the **actors**.

**Insight for Business:**

Netflix can utilize this information to see which actors are the most popular and who works on more than one show.

This can help build marketing strategies around which actors are most popular.

|  |
| --- |
| SELECT DISTINCT TOP 5  c.cast,n.type,  COUNT(n.show\_id) OVER (PARTITION BY c.cast) AS show\_count  FROM  Netflix\_cast AS c INNER JOIN Netflix\_Movies AS n  ON c.show\_id = n.show\_id  ORDER BY  show\_count DESC; |

**5. Directors who have worked on the most shows**

**Detailed explanation:**

The query **is similar to the previous** query but focuses on the number of shows each director has directed.

**COUNT()** is used with **PARTITION BY** on the **director** column to calculate the number of works for **each director**.

**Insight for Business:**

Netflix can use this analysis to identify directors who have extensive experience in producing content and direct more productions to them.

This helps improve the quality of shows by collaborating with experienced directors.

|  |
| --- |
| SELECT DISTINCT TOP 5  d.director,n.type,  COUNT(n.show\_id) OVER (PARTITION BY d.director) AS show\_count  FROM  Netflix\_Movies AS n INNER JOIN Netflix\_director AS d  ON N.show\_id = d.show\_id  ORDER BY  show\_count DESC; |

**6. Is Netflix more focused on TV Shows than movies in recent years? (Yearly analysis)**

**Detailed explanation:**

This query analyzes the distribution of content based on its type **(movie or TV show)** over the past five years.

The number of shows is calculated using **COUNT()** with the results broken down by year **(release\_year)** and genre **(type).**

**Insight for Business:**

This analysis allows Netflix to determine whether it has put more emphasis on TV shows or movies recently.

The company can use this data to optimize future content strategies based on viewer preferences.

|  |
| --- |
| SELECT DISTINCT  release\_year,  type,  COUNT(show\_id) OVER (PARTITION BY release\_year, type) AS content\_count  FROM  Netflix\_Movies --Amazon\_Prime  WHERE  release\_year >= (select max(release\_year)  from Netflix\_Movies) - 5  ORDER BY  release\_year DESC, content\_count DESC; |

**7. Find the top 5 genres that have the most content across all countries**

**Detailed explanation:**

This query calculates the number of contents per type **(listed\_in)** across all countries and **returns the 5 most popular content types**.

**Business insight:**

Netflix can use this data to determine which genres are more popular and which are more popular and which to invest in.

It can help guide production and promotion strategies.

**-Using common table expression (CTE)**

|  |
| --- |
| WITH GenreCounts AS (  SELECT  listed\_in,  COUNT(show\_id) OVER (PARTITION BY listed\_in) AS genre\_count  FROM  Netflix\_listed\_in  )  SELECT DISTINCT  listed\_in,  genre\_count  FROM  GenreCounts  ORDER BY  genre\_count DESC  OFFSET 0 ROWS FETCH NEXT 5 ROWS ONLY; |

**8. Determine the earliest and latest show added by country**

**Detailed explanation:**

This query displays **the first and last** show added to the Netflix platform for **each country** using the **MIN() and MAX() functions.**

**Insight for Business:**

This analysis helps in understanding the dynamics of content addition in each country.

It can be useful for determining launch and promotion strategies based on the timing of content addition.

|  |
| --- |
| SELECT DISTINCT  c.country,  MIN(n.date\_added) OVER (PARTITION BY c.country) AS earliest\_show,  MAX(n.date\_added) OVER (PARTITION BY c.country) AS latest\_show  FROM  Netflix\_Movies AS n INNER JOIN Netflix\_country AS c  ON n.show\_id = c.show\_id  ORDER BY  c.country; |

**9. Actors who appear in the most diverse genres**

**Detailed explanation:**

The query analyzes the diversity of actors in their roles based on the genres they appear **in (listed\_in).**

**Insight for Business:**

This data can help identify actors who are more diverse and who can attract a larger audience through diverse roles.

|  |
| --- |
| SELECT DISTINCT  c.cast,  COUNT(l.listed\_in) OVER (PARTITION BY c.cast) AS genre\_variety  FROM  Netflix\_cast AS c INNER JOIN Netflix\_listed\_in AS l  ON c.show\_id = l.show\_id  ORDER BY  genre\_variety DESC; |

**10. Which directors have the most variety in content (genres)?**

**Detailed explanation:**

This query is similar to the previous query but focuses on directors and their diversity in working on different types of content.

**Insight for Business:**

This data can help identify directors who have the ability to produce diverse content in different genres, making them a strong option for future investment.

|  |
| --- |
| SELECT DISTINCT  d.director,  COUNT(l.listed\_in) OVER (PARTITION BY d.director) AS genre\_variety  FROM  Netflix\_director AS d INNER JOIN Netflix\_listed\_in AS l  ON d.show\_id = l.show\_id  ORDER BY  genre\_variety DESC; |