**Netflix Userbase Dataset**

**About Dataset**

The dataset provides a snapshot of a sample Netflix userbase, showcasing various aspects of user subscriptions, revenue, account details, and activity. Each row represents a unique user, identified by their User ID. The dataset includes information such as the;

* **User's subscription type** (Basic, Standard, or Premium)
* **User ID** (Users identification numbers)
* **Monthly revenue** generated from their subscription
* **Join Date** The date they joined Netflix
* **Last Payment Date** (date of their last payment)
* **Country** in which they are located.

Additional columns have been included to provide insights into user behavior and preferences. These columns include

* **Device Type** (e.g., Smart TV, Mobile, Desktop, Tablet)
* **Age**
* **Gender** (male & female)
* **Plan duration**

The dataset serves as a synthetic representation and does not reflect actual Netflix user data. It can be used for analysis and modeling to understand user trends, preferences, and revenue generation within a hypothetical Netflix userbase.

**Objective**

The objective of the analysis is aimed at understanding user behavior, preferences, and subscription patterns.

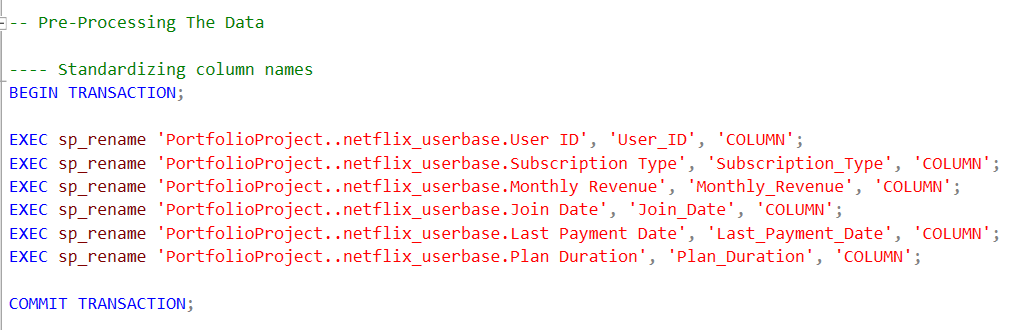
**Analysis Questions**

1. What is the distribution of Subscription type
2. Run a user demographic analysis by; Country, Age and Gender
3. What is the range of subscription duration among users
4. Which device is used mostly to view Netflix contents
5. Is there a correlation between number of users and monthly revenue

**Analysis Tool Used: MS SQL Server**

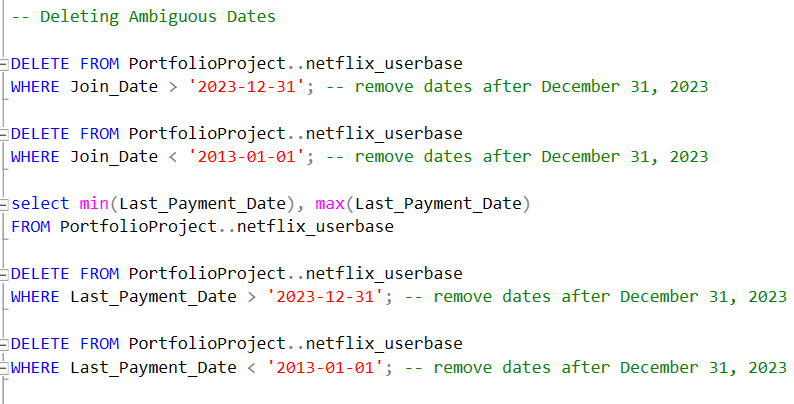
**Data Pre-Processing**

Upon acquiring the dataset, it was discovered that the data columns weren’t in a standard format so, a query was ran in order to standardize column names in order to ease analysis.

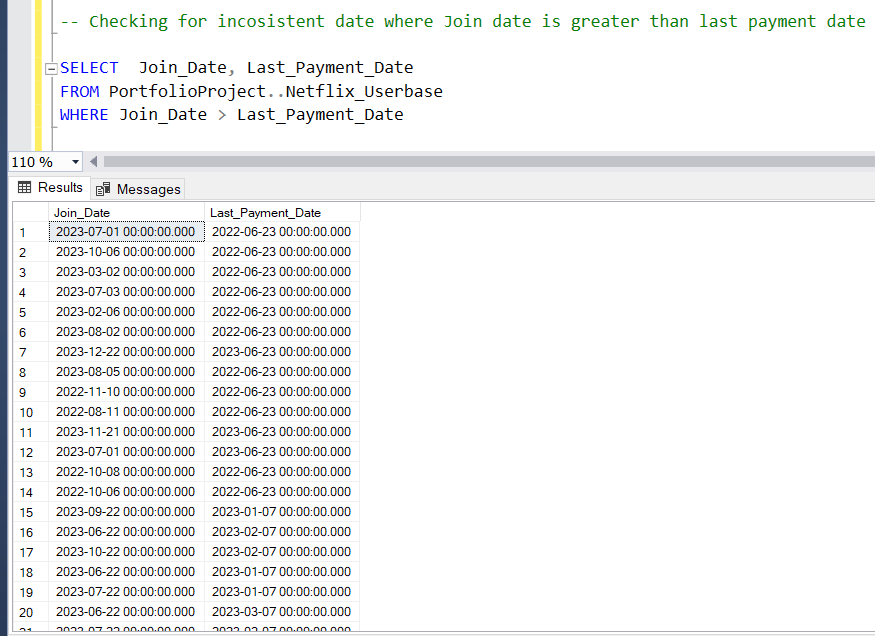


**Fixing inconsistent dates**

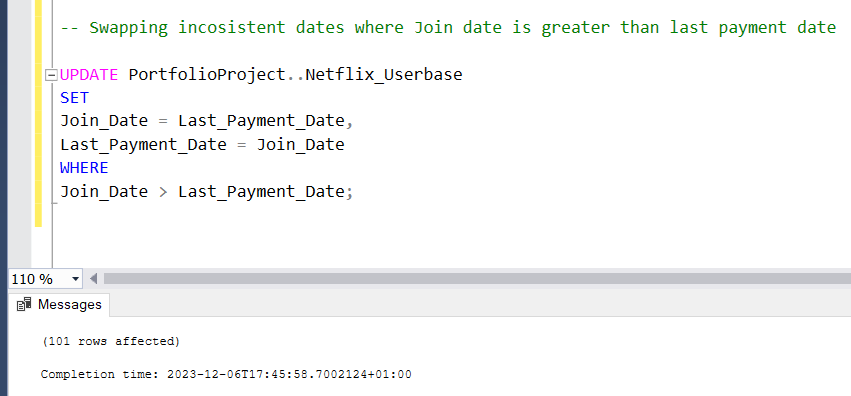
Also, it was discovered that the dataset contained ambiguous dates before Netflix was founded as well containing dates past the present year 2023. These inconsistent dates were removed also.



Still on inconsistent dates, there were issues whereby a total of 101 join dates were greater than last payment dates.



I decided to swap the join dates that were greater than last payment dates in order to ensure data validity and consistency.

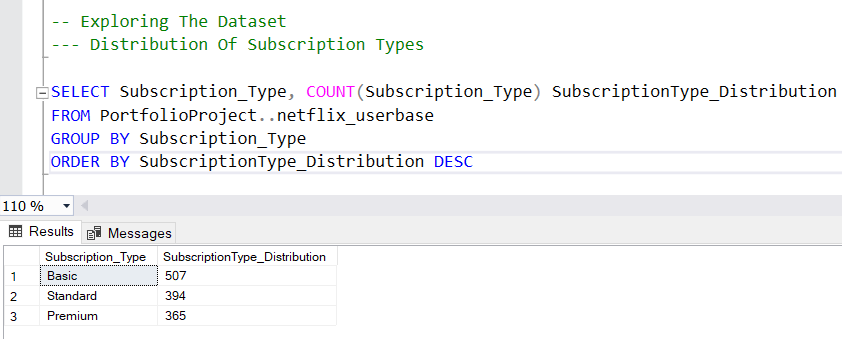


Missing values were also removed from the dataset.

**Data Exploration**

The following are the queries, results and insights to each analysis questions in order:

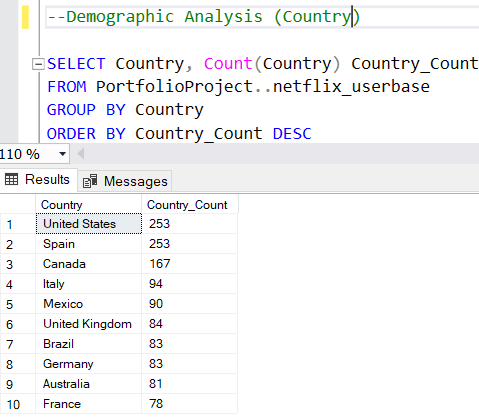
1. What is the distribution of Subscription type

****

**Insight:** Users subscribed more to the Basic subscription time which totaled 507. This could be as a result of the Basic subscription type being cheaper than Standard which had 394 subscriptions followed by Premium 365.

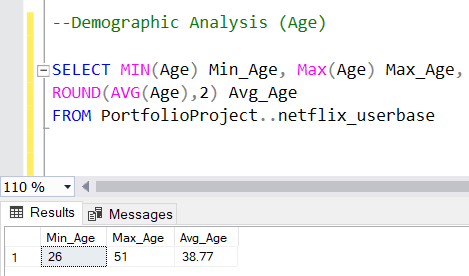
1. **Run a user demographic analysis by; Country, Age and Gender**

* **Geographical Analysis(Country)**

****

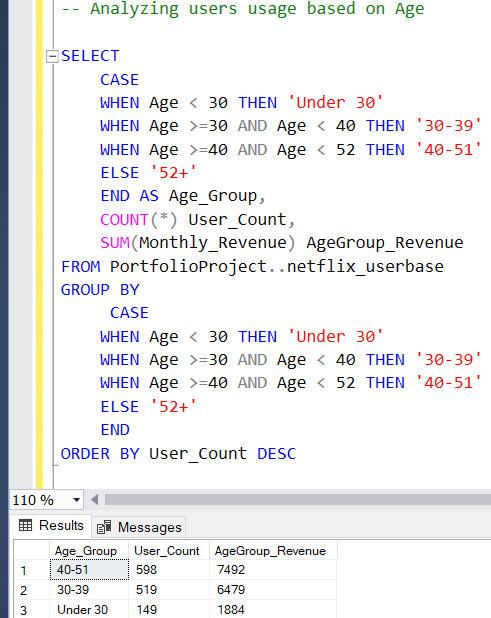
**Insight:** United States and Spain had the most significant number of users (253) 0n the Netflix platform, followed by Canada which had a number of 167 users. On the other hand, countries like Germany, Australia and France amongst others had the least number of users 83, 81 and 78 respectively.

* **Demographic analysis(Age)**



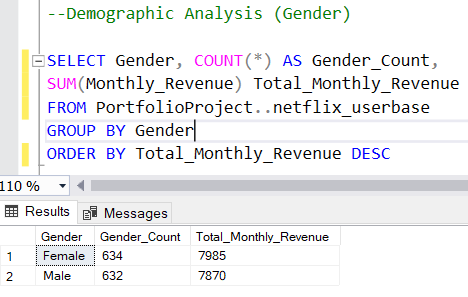
**Insight:** Theresult above shows the minimum age of users to be 26 and maximum age 51. On average, users have an age of approximately 38.77.

* **Users usage based on Age**

****

Insight: From the above query and result, users aged 40-50 generated the most revenue totaling 7492 followed by those within the age bracket of 30-39 with a total revenue of 6479. Users aged Under 30 generated the least revenue 1884. It is a bit surprising that users above 30years of age generated the most revenue although, not too surprising as they could be more of those in the working class.

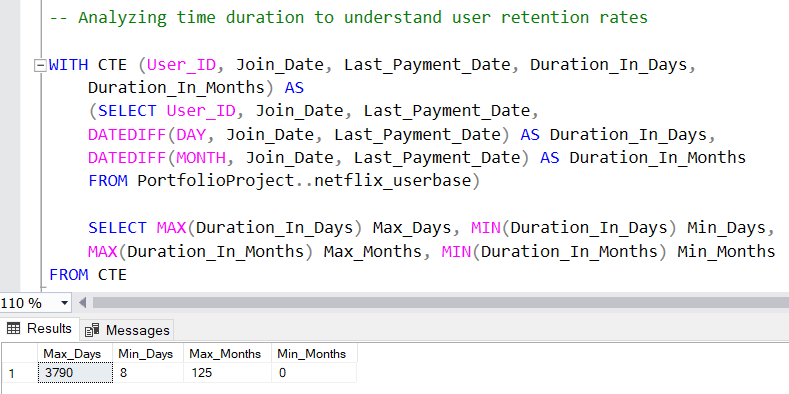
* **Demographic analysis(Gender)**



There are more female users with a total count of 634 females generating a total monthly revenue of 7985 as compared to the male counterparts who have a total user count of 632 generating a total monthly revenue of 7870.

1. What is the range of subscription duration among users

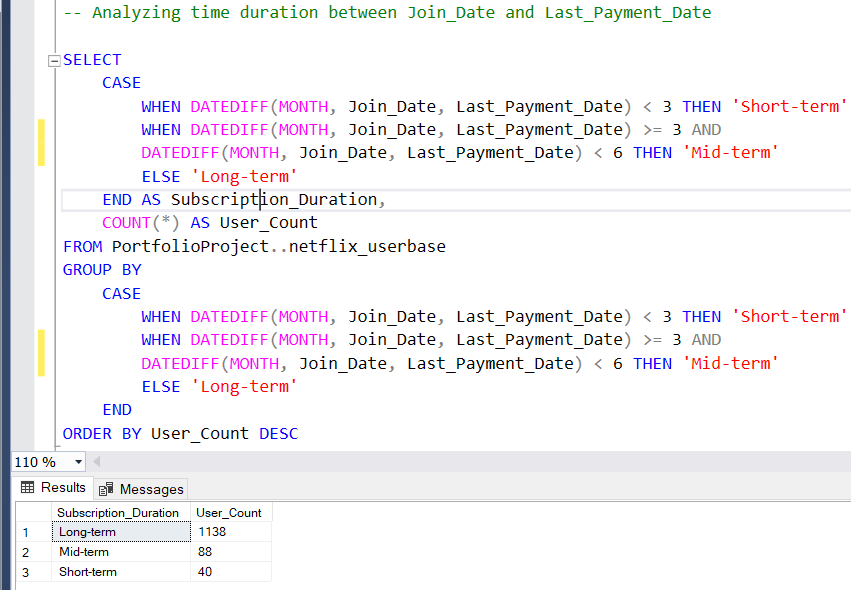
* Time duration to understand user retention rates



From the above, Max\_Days value of 3790 days indicates the longest duration a user stayed subscribed between their joining date and their last payment date and Min\_Days value indicates the shortest duration users remained subscribed between their join and last payment dates.

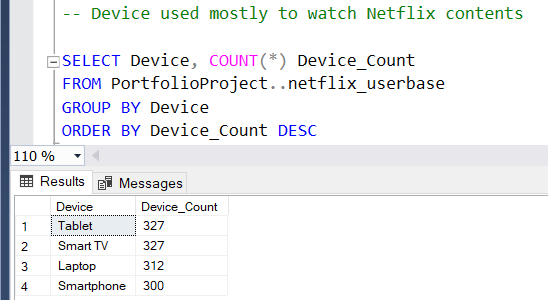
On the other hand, the maximum duration in months between Join\_Date and Last\_Payment\_Date among users is 125 months representing the longest subscription period in terms of months with a minimum duration of 0 months. This is an indication that some records might have 'Join\_Date' and 'Last\_Payment\_Date' on the same month, resulting in a 0-month duration.

* Distribution of users across different subscription duration categories



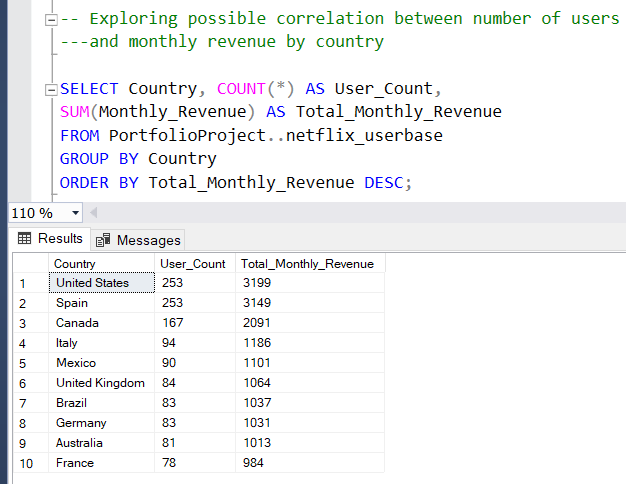
From the result above, the Long-term category represents users who have a subscription duration of over 6 months. The count of users in this category is 1138. On the other hand, the Mid-term' category represents users whose subscription duration falls between 3 to 6 months indicating that users have subscriptions that are longer than the Short-term group but shorter than the 'Long-term' group. The count of users in this category is 88. Lastly, Users categorized as Short-term have a subscription duration of less than 3 months. These users subscribed for a brief period, and the count of users in this category is 40.

1. Which device is used mostly to view Netflix contents



With a count of 327 each, Tablets and Smart TVs’ are the devices used mostly by users to view Netflix contents, these are followed by Laptop (312) and Smartphone (300) respectively.

1. Is there a correlation between number of users and monthly revenue



The top 3 countries generating the most revenues are; United States (3199), Spain (3149) and Canada (2091) meanwhile, Germany (1031), Australia (1013) and France (984) are the bottom 3 countries generating the least monthly revenue.

There seems to be a correlation between number of users and total monthly revenue because, as the number of users increase, so does the total monthly revenue too.