**Team Members:**

Chloe Adams, Jestin Bray, Mollie Cox, Trevor Lammers, Rachel Redeker

**Introduction**

Netflix is a subscription streaming service that allows subscribers to watch movies and TV shows. It is said to be the “single most important company that rewrote television in the twenty-first century” [Netflix, Inc. and Online Television.](https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119269465.ch7) Our group will determine the connections between shows and various other fields using the data shown below. This analysis may be used by individuals in the media entertainment industry such as television streaming services or even current or potential Netflix subscribers wanting to gain more insight into the streaming service.

**Data**

Our project uses data provided by a 2021 Kaggle dataset consisting of movies and TV shows on Netflix <https://www.kaggle.com/datasets/shivamb/netflix-shows>. On the streaming service, Netflix has a library of around 8000 movies and television shows and around 220.6 million subscribers to the platform as of 2022 <https://www.businessofapps.com/data/netflix-statistics/>. The dataset consists of 8807 rows of unique data across 11 fields. Table 1 provides a description of the data fields used in this project.

*Table 1 Data Dictionary*

Table

Description automatically generated

Our ERD consists of 4 entities, 2 strong entities and 2 weak entities. Our data has many multivalued attributes that will be resolved by creating new tables. In the ERD Diagram, the COUNTRY and GENRE multivalued attributes are treated as weak entities because they have partial identifiers. The entity SHOW is related to both weak entities with a one-to-many relationship. The modality of these relationships is required on the SHOW side of the relationship and optional for the COUNTRY but required for the GENRE side of the relationship. We created a surrogate identifier for the Director table to make the analysis easier resulting in DIRECTOR being a strong entity with a many to many relationship with SHOW. Figure 1 displays the ERD for this data.

*Fig. 1 Entity relationship diagram (ERD)*

Diagram

Description automatically generated

Using the ERD, we normalized the data and created a relational schema consisting of 5 tables as seen in Figure 2. The Show table acts as the parent table to the Country, Genre, and DirectedBy tables. The Director table also acts as the parent table to the DirectedBy table. Each child table stores the parent table’s primary key as a foreign key which is part of their composite key. The primary key for the Country and Genre tables consists of ShowID and the weak entity’s partial identifier. The primary key of DirectedBy is a composite key consisting of ShowID and DirectorID.

*Fig. 2 Relational Schema Diagram (ERD)*

Diagram

Description automatically generated

**Database Implementation**

We wrote CREATE Table commands for each of the tables presented in our relational schema to create all the table structures in the APEX database. We also wrote an INSERT INTO command to provide an example of populating data values into the data.

SHOW

As the parent table, SHOW was populated first.

CREATE TABLE SHOW(

SHOWID VARCHAR(10) NOT NULL,

TYPE VARCHAR(10) NOT NULL CHECK (TYPE IN ('TV Show', 'Movie')),

TITLE VARCHAR(150) NOT NULL,

DATEADDED DATE,

RELEASEYEAR NUMBER(4,0) NOT NULL,

RATING VARCHAR(10),

MINUTES NUMBER(3,0),

SEASONS NUMBER(2,0),

CONSTRAINT SHOW\_PK PRIMARY KEY (SHOWID));

INSERT INTO SHOW (ShowID, Type, Title, DateAdded, ReleaseYear, Rating, Minutes)

VALUES (‘s1’, ‘Movie’, ‘Dick Johnson Is Dead’, ‘9/25/2021’, 2020, ‘PG-13’, 90);

COUNTRY

CREATE TABLE COUNTRY(

COUNTRY VARCHAR(125) NOT NULL,

SHOWID VARCHAR(7) NOT NULL,

CONSTRAINT COUNTRY\_PK PRIMARY KEY (COUNTRY, SHOWID),

CONSTRAINT COUNTRY\_FK FOREIGN KEY (SHOWID)

REFERENCES SHOW (SHOWID));

INSERT INTO COUNTRY VALUES (‘s1’, ‘United States’);

GENRE

CREATE TABLE GENRE(

GENRE VARCHAR(30) NOT NULL,

SHOWID VARCHAR(7) NOT NULL,

CONSTRAINT GENRE\_PK PRIMARY KEY (GENRE, SHOWID),

CONSTRAINT GENRE\_FK FOREIGN KEY (SHOWID)

REFERENCES SHOW (SHOWID));

INSERT INTO GENRE VALUES (‘s1’, ‘Documentaries’);

DIRECTOR

As a parent table to DirectedBy, Director was populated before DirectedBy.

CREATE TABLE DIRECTOR(

DIRECTORID CHAR(4) NOT NULL,

FIRSTNAME VARCHAR(50) NOT NULL,

MIDDLENAME VARCHAR(50),

LASTNAME VARCHAR(50),

SUFFIX VARCHAR(50),

CONSTRAINT DIRECTOR\_PK PRIMARY KEY (DIRECTORID));

INSERT INTO DIRECTOR (DirectorID, FirstName, LastName) VALUES (‘s1’, ‘2233’, ‘Kirsten’, ‘Johnson’);

DIRECTEDBY

CREATE TABLE DIRECTEDBY(

DIRECTORID CHAR(4) NOT NULL,

SHOWID VARCHAR(7) NOT NULL,

CONSTRAINT DIRECTEDBY\_PK PRIMARY KEY (DIRECTORID, SHOWID),

CONSTRAINT DIRECTEDBY\_FK1 FOREIGN KEY (SHOWID)

REFERENCES SHOW (SHOWID),

CONSTRAINT DIRECTEDBY\_FK2 FOREIGN KEY (DIRECTORID)

REFERENCES DIRECTOR (DIRECTORID));

INSERT INTO DIRECTEDBY VALUES(‘0001’, ‘s3538’);

**Analysis**

This Netflix dataset found on Kaggle allows users to evaluate and analyze relationships between shows, genres, countries, and directors in Netflix’s current library. This analysis is intended to spot trends and insights from Netflix’s shows and their characteristics. We are particularly interested in the popularity of directors, movies that belong to varying genres, the longest running TV shows, top publishing countries, and ratings in particular genres. Research and analysis questions are found below.

Question 1: Director Popularity

Which directors are the most popular in Netflix’s library? Specifically, which directors have directed the most shows? In solving this question, we implemented a join query that counts the number of shows each director has directed and ranks the directors by order of their total shows directed. We were particularly interested in the top 10 most popular directors in the database.

SELECT FIRSTNAME||' '||LASTNAME AS FULLNAME, COUNT(SHOW.SHOWID) AS SHOWCOUNT

FROM SHOW JOIN DIRECTEDBY

ON SHOW.SHOWID=DIRECTEDBY.SHOWID

JOIN DIRECTOR ON DIRECTEDBY.DIRECTORID=DIRECTOR.DIRECTORID

GROUP BY FIRSTNAME||' '||LASTNAME

ORDER BY SHOWCOUNT DESC

FETCH FIRST 10 ROWS ONLY;

The results of this query are shown below in Figure 3. Rajiv Chilaka was the most popular director in the database having directed 22 shows in Netflix’s library. He was closely followed by Raúl Campos and Jan Suter who directed 19 and 18 shows respectively.

*Fig. 3 Director Popularity*

A screenshot of a computer

Description automatically generated with medium confidence

Question 2: Classic and Children & Family Movies

Classic and Children & Family Movies seem to operate in different spheres, but are there any movies in the database that belong to both genres? To solve this problem, we used a compound query that finds classic movies and intersects them with Children & Family Movies to find movies that are listed in both genres.

SELECT TITLE, RELEASEYEAR

FROM SHOW JOIN GENRE

ON SHOW.SHOWID=GENRE.SHOWID

WHERE TYPE= 'Movie' and GENRE='Children & Family Movies'

INTERSECT

SELECT TITLE, RELEASEYEAR

FROM SHOW JOIN GENRE

ON SHOW.SHOWID=GENRE.SHOWID

WHERE TYPE= 'Movie' and GENRE='Classic Movies';

The results of this query are shown below in Figure 4. There are a total of 9 movies listed in both genres of Classic and Children & Family Movies. Further, all the movies listed in both genres were all released before 1993 showing a possible association between genre and release year.

*Fig. 4 Classic and Children & Family Movies*

*Background pattern

Description automatically generated*

Question 3: Longest Running Netflix Shows

Especially in the past few years, Netflix US has grown as a production company. What are some of the longest running Netflix licensed TV shows in the US since 2019? To answer this, we created a join query that joined the SHOW and COUNTRY tables to find the longest running licensed shows in the US since 2019 and ordered the results by number of seasons.

SELECT SHOW.SHOWID, TITLE, DATEADDED, RELEASEYEAR, SEASONS

FROM SHOW

JOIN COUNTRY ON COUNTRY.SHOWID = SHOW.SHOWID

WHERE COUNTRY.COUNTRY = 'United States'

AND EXTRACT(YEAR FROM DATEADDED) = RELEASEYEAR

AND TYPE = 'TV Show'

AND RELEASEYEAR >= 2019

AND SEASONS >= 6

ORDER BY SEASONS DESC;

The results of this query are shown below in Figure 5. This query returned 14 rows of data showing that the longest running licensed TV show in the US since 2019 is COMEDIANS of the world with a total of 13 seasons.

*Fig. 5 Longest Running Netflix Shows*

A screenshot of a computer

Description automatically generated with medium confidence

Question 4: Top Publishing Countries

What is the most common publishing country for the shows in Netflix’s dataset? Specifically, which countries have produced the most shows in the catalog? To determine this, we made a join query that combined the SHOW table with the COUNTRY table to find the 20 most common Netflix producing countries and their respective total number of projects. We created a new column using a case statement displaying whether the countries had produced more than 100 shows.

SELECT COUNTRY, COUNT(DISTINCT SHOW.SHOWID) AS TOTALPROJECTS,

CASE WHEN COUNT(DISTINCT SHOW.SHOWID) >= 100 THEN 'Y' ELSE 'N' END AS Y\_OR\_N

FROM COUNTRY

LEFT JOIN SHOW ON COUNTRY.SHOWID = SHOW.SHOWID

GROUP BY COUNTRY

ORDER BY TOTALPROJECTS DESC

FETCH FIRST 20 ROWS ONLY;

The results of this query are shown below in Figure 6. This query returns 20 rows which are displayed below. As seen in the query results, the United Sates is the top publishing country with 3,690 projects, more than triple any other country. India and the United Kingdom come in second and third with 1,046 and 806 shows respectively. Looking further into the results, 17 different countries had more than 100 projects in Netflix’s library. As Netflix continues to grow as an international brand, we can only expect these numbers to increase substantially.

*Fig. 6 Top Publishing Countries*

*A screenshot of a computer

Description automatically generated with medium confidence*

Question 5: Comedy Ratings

What are the most common content ratings in the comedy genre? To answer this question, we created a join query that combines the SHOW table with the GENRE table to find the rating count for each content rating in the comedy genre.

SELECT RATING, COUNT(RATING) AS RATINGCOUNT

FROM SHOW JOIN GENRE

ON SHOW.SHOWID=GENRE.SHOWID

WHERE GENRE= 'Comedies'

GROUP BY RATING

ORDER BY RATINGCOUNT DESC;

The results of this query are shown below in Figure 7. TV-14 is the most common rating for Comedy shows with TV-MA being a close second with respective totals of 465 and 431 shows belonging to each category. Comedy shows are very versatile when it comes to content rating and comprise all the top rating types.

*Fig. 7 Comedy Ratings*

Background pattern

Description automatically generated

**Web Design**

Database Project website link:

<https://apex.oracle.com/pls/apex/r/netflixdatabaseproject/netflix-project202818/home?session=9667979175057>

Home page

The home page of our application provides background information on Netflix, a description of our project, and information on where we collected our data to allow users to better understand our analyses. We included a navigation menu on the side of the homepage for easy access to the rest of our project pages. Throughout the entire application, we utilized the accent color of red to draw users’ eyes to important ideas and relate back to Netflix’s company color. A screenshot of the homepage is listed below in Figure 8.

*Fig. 8 Home Page*

Graphical user interface, text, application

Description automatically generated

Tables

We further created an interactive report for each of our database tables, allowing users the ability to easily find information through searching, sorting, filtering, and grouping the data. We provided a description of the table above the specified table.

*Fig. 9 SHOW*

Graphical user interface, application, table

Description automatically generated

*Fig. 10 DIRECTOR*

Graphical user interface, application, table

Description automatically generated

*Fig. 11 GENRE*

Graphical user interface, application, table, Word

Description automatically generated

*Fig. 12 COUNTRY*

Graphical user interface, application, table, Word

Description automatically generated

*Fig. 13 DIRECTEDBY*

Graphical user interface, application, table, Word

Description automatically generated

Queries

For each of our analysis questions, we created a page on our application to effectively display the results either in reports, graphs, or both! Each page contains a description of the question and answer at the top of the page to enhance understanding. All the analysis questions and pages are listed below.

**Question 1:**

For the first research question on director popularity, we presented the results in a classic report and a bar chart to be able to visualize the results as seen below in Figure 14. The report and chart show the top directors in terms of shows produced.

*Fig. 14 Director Popularity*

Graphical user interface, application, table, Excel

Description automatically generated

**Question 2:**

For the second research question regarding movies that belong to both the Classic and Children & Family Movies, we presented the results in a classic report which is shown below in Figure 15. The report shows the movies belonging to both these categories and shows their respective release dates.

*Fig. 15 Classic and Children & Family Movies*

Graphical user interface, application, Word

Description automatically generated

**Question 3:**

For the third research question on the longest running Netflix TV shows, we presented the results in a classic report and a bar chart to be able to understand and visualize the results as seen below in Figure 16. The report and chart show the longest running licensed TV shows since 2019 in terms of the number of seasons.

*Fig. 16 Longest Running TV Shows*

Graphical user interface, application, table

Description automatically generated

**Question 4:**

For our fourth research question regarding top publishing countries, we presented the results in a bar chart to visualize the results as seen below in Figure 17. The bar graph shows the top publishing countries in terms of the number of shows each country has produced.

*Fig. 17 Top Publishing Countries*

A picture containing application

Description automatically generated

**Question 5:**

For our fifth research question on the common rating types in the comedy genre, we created a classic report showing the frequency of each rating type in the comedy genre. We complimented this report with a bar chart showing content rating categories and their frequency in Netflix’s library as seen below in Figure 18.

*Fig. 18 Comedy Ratings*

Graphical user interface, application

Description automatically generated