

BAIS: 3200

Professor Kristina Bigsby

Final Project Report

Jillian Kozlowski, Julia Pysk, Blake Murray, Andrew Thiele, and Katelyn Bonnett

Introduction

Netflix, a streaming service that virtually almost everyone has at least knows someone who does. Over the course of our lifetime, Netflix has engaged in many different contract negotiations and has annually changed its style and format. In this project, we will apply methods of data analytics to highlight the changes within the application, Netflix. Our database and analytical analysis could be useful to Netflix as a whole company, everyday people, or other parties involved in streaming platforms.

Data

This project delivers data from a 2019 Kaggle dataset about all shows and movies available on Netflix from 2008 to 2021 (<https://www.kaggle.com/shivamb/netflix-shows>). The dataset contains 11 attributes and has 7,785 rows of data. Our data tracks information about show details, production data, genres and ratings. Due to the limited amount of attributes, reducing the data by size was not necessary in this case. Although some titles contained foreign letters, so we had to delete those for our database to work correctly. Table 1 shows the description of the dataset.

Table 1 Data Dictionary:

Data Dictionary						
Attribute	Type	Description				
ShowID	Alpha Numeric	Unique ID to distinguish the Show/movie				
ShowType	Text	Identifies if it is a show or movie				
Title	Text	Title of the show/movie				
Country	Text	Where it is offered it on Netflix				
DateAdded	Date	The date the show/movie was added to Netflix				
ReleaseYear	Numeric	The date the show/movie was released				
Rating	Text	TV rating for the show/movie				
Season	Numeric	How many seasons a Show has				
Duration	Numeric	How long the movie lasts in minutes				
Genre	Text	Which genre the movie belongs to				

The supertype in the database is SHOW, which can be identified by ShowID. ShowID, Title, Release Year, and ShowType are all required in the SHOW entity. Without those attributes, there would not be enough information about each show. DateAdded and Rating are not required in the

SHOW entity and are therefore optional. There are two subtypes, TV and MOVIE. There is complete specialization because each SHOW belongs in either TV or MOVIE and nothing else. There is disjointness because a SHOW cannot be a TV show and a MOVIE at the same time. In addition to the supertype and subtypes, we have two weak entities, COUNTRY and GENRE. These entities are resolved multi-valued attributes and only have partial identifiers. COUNTRY has an Optional-Many to Mandatory-Many relationship with SHOW. This is because SHOW can have 0 or many countries listed, and each COUNTRY has 1 or many movies available in them. GENRE has a Mandatory-Many to Mandatory-Many relationship with SHOW. This is because SHOW has at least 1 or many genres and each GENRE has at least one or many shows. Figure 1 displays the EER diagram for this dataset.

Figure 1 Enhanced Entity Relationship diagram (EERD):

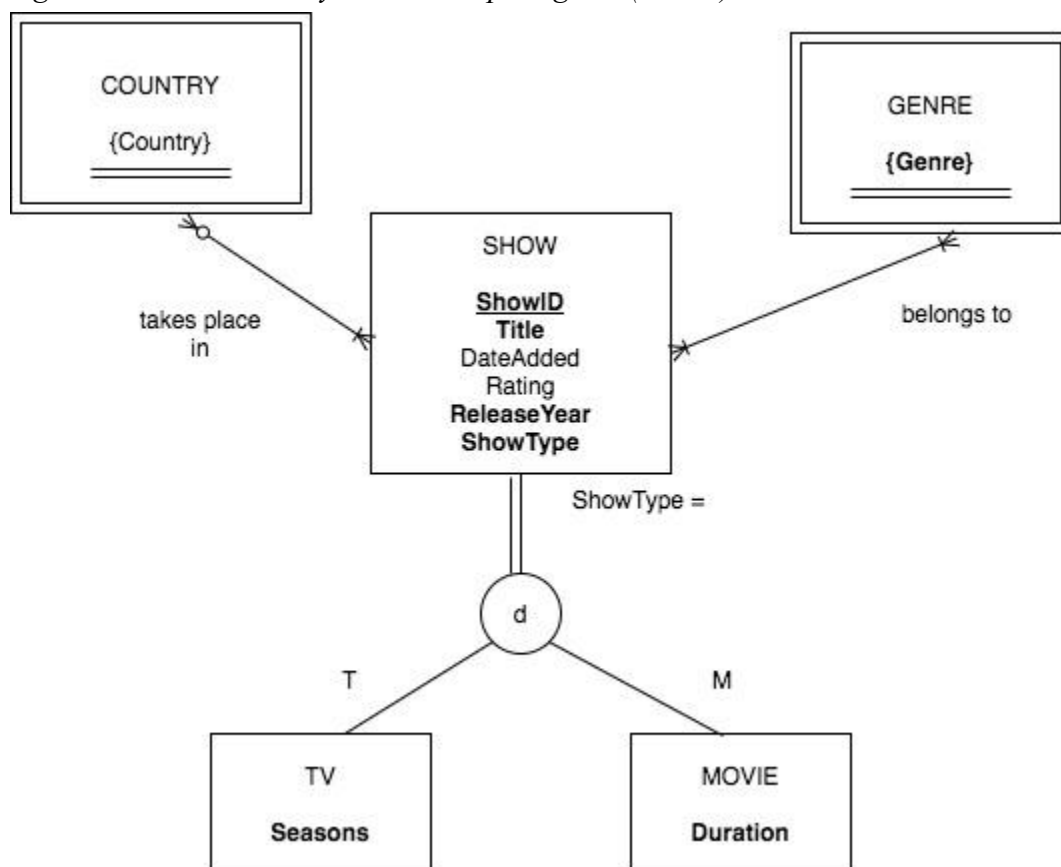
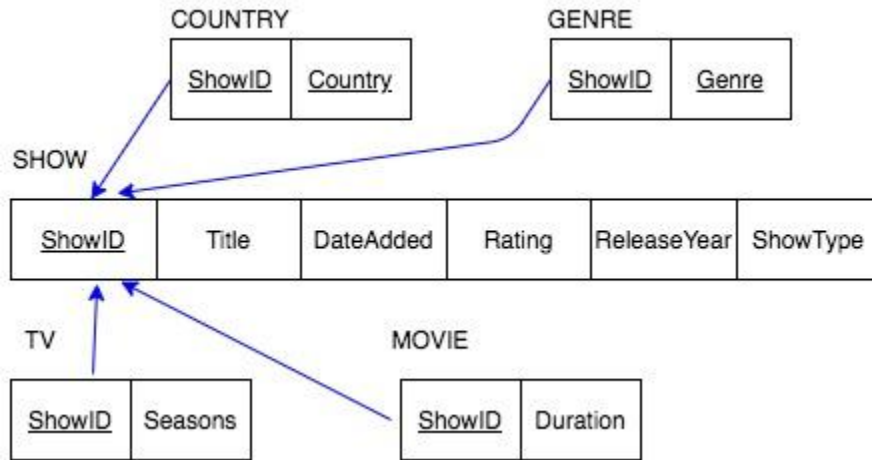


Figure 2 Graphical Relational Schema:



Based on the EER diagram, we normalized the data and created a schema with 5 tables. The figure above shows the graphical relational schema of the database. The SHOW table acts as the parent table with ShowID as the primary key. TV and MOVIE have ShowID as the primary key and inherit all other attributes that are contained in the SHOW entity. GENRE and COUNTRY have primary composite keys composed of their weak entity partial identifier and ShowID.

Database Implementation

To implement the data into our database in Oracle APEX, we have written CREATE TABLE commands as well as insert commands for each table from our graphical relational schema.

SHOW

Because SHOW is the parent table, it was created and populated first in our database.

```
CREATE TABLE SHOW (
    ShowID VARCHAR(10) NOT NULL,
    ShowType VARCHAR(10) NOT NULL,
    Title VARCHAR(250) NOT NULL,
    DateAdded DATE,
    ReleaseYear NUMBER NOT NULL,
    Rating VARCHAR(50),
    CONSTRAINT SHOW_PK PRIMARY KEY (ShowID)
);
```

```
INSERT INTO SHOW (ShowID, ShowType, Title, DateAdded, ReleaseYear, Rating)
VALUES ('s166', 'Movie', 'A Clockwork Orange', 11/1/2020, 1971, 'R');
```

TV

```
1 CREATE TABLE TV (  
2     ShowID VARCHAR(10) NOT NULL,  
3     Seasons NUMBER  
4     CONSTRAINT TV_PK PRIMARY KEY (ShowID)  
5 );
```

```
INSERT INTO TV (ShowID, Seasons)  
VALUES ('s1', '4');
```

MOVIE

```
CREATE TABLE MOVIE (  
    ShowID VARCHAR(10) NOT NULL,  
    Duration NUMBER(4) NOT NULL,  
    CONSTRAINT MOVIE_PK PRIMARY KEY (ShowID)  
);
```

```
INSERT INTO MOVIE (ShowID, Duration)  
VALUES ('s958', '312');
```

COUNTRY

```
CREATE TABLE COUNTRY (  
    ShowID VARCHAR(10) NOT NULL,  
    Country VARCHAR(40) NOT NULL,  
    CONSTRAINT COUNTRY_PK PRIMARY KEY (ShowID, Country),  
    CONSTRAINT COUNTRY_FK FOREIGN KEY (ShowID) REFERENCES SHOW (ShowID)  
);
```

```
INSERT INTO COUNTRY (ShowID, Country)  
VALUES ('s1', 'Brazil');
```

GENRE

```
CREATE TABLE GENRE (
  ShowID VARCHAR (10) NOT NULL,
  Genre VARCHAR(40) NOT NULL,
  CONSTRAINT GENRE_PK PRIMARY KEY (ShowID, Genre),
  CONSTRAINT GENRE_FK FOREIGN KEY (ShowID) REFERENCES SHOW (ShowID)
);
```

```
INSERT INTO GENRE (ShowID, Genre)
VALUES ('s1', 'International TV Shows');
```

Analysis

The purpose of our team's analysis is to inform movie and TV watchers, using the dataset from Kaggle, about each show's characteristics. The intention of this dataset is to inform current and potential Netflix users about which types of tv shows and movies that are available to stream.

Question 1 Length of Shows: What is the title and length of the longest movie available on Netflix? The longest TV show? To answer these questions, we have created a subquery which finds the movie with the longest duration and returns that movie's Title and Duration in minutes. Similarly, for TV Shows we created a subquery that finds the TV show with the largest amount of seasons and returns that TV show's Title and number of Seasons

```
SELECT Title, Duration
FROM MOVIE JOIN SHOW ON MOVIE.ShowID = SHOW.ShowID
WHERE Duration =
(SELECT MAX(Duration) FROM MOVIE);
```

Figure 3 Length of Movies

Results		Explain	Describe	Saved SQL	History
TITLE		DURATION			
Black Mirror: Bandersnatch		312			
1 rows returned in 0.02 seconds		Download			

```
SELECT Title, Seasons
FROM TV JOIN SHOW ON TV.ShowID= SHOW.ShowID
WHERE Seasons =
(SELECT MAX(Seasons) FROM TV);
```

Figure 4 Length of TV Shows

Results Explain Describe Saved SQL History	
TITLE	SEASONS
Grey's Anatomy	16
1 rows returned in 0.02 seconds Download	

The results of this query are shown above in Figure 3 and 4. Out of all movies in this Netflix data, Black Mirror: Bandersnatch was the longest movie with a duration of 312 minutes. Based on Figure 4, Grey's Anatomy was the TV show with the largest number of seasons, that being 16 seasons.

Question 2 Common Genre and Country: Which country has the most amount of shows available in it? What is the most common genre for all shows made in this country? We have created 2 subqueries that returned the country that released the most amount of shows, and the other returned the most popular genre within that country.

```
SELECT DISTINCT Country
FROM COUNTRY
WHERE Country =
    (SELECT Country FROM COUNTRY
    HAVING COUNT (ShowID) =
        (SELECT MAX (COUNT (ShowID) )
        FROM COUNTRY
        GROUP BY COUNTRY)
    GROUP BY COUNTRY) ;
```

Figure 5 Country with Largest Selection

Results Explain Describe Saved SQL History	
COUNTRY	
United States	
1 rows returned in 0.01 seconds Download	

```

SELECT Country, Genre FROM GENRE JOIN COUNTRY ON GENRE.ShowID = COUNTRY.ShowID
WHERE Country =
    (SELECT Country
    FROM COUNTRY
    HAVING COUNT(ShowID) =
        (SELECT MAX(COUNT(ShowID))
        FROM COUNTRY
        GROUP BY Country)
    GROUP BY Country)
HAVING COUNT(Genre)=
    (SELECT MAX(COUNT(Genre))
    FROM GENRE
    JOIN COUNTRY ON GENRE.ShowID=COUNTRY.ShowID
    WHERE Country=
        (SELECT Country
        FROM COUNTRY
        HAVING COUNT(ShowID) =
            (SELECT MAX(COUNT(ShowID))
            FROM COUNTRY
            GROUP BY Country)
        GROUP BY Country)
    GROUP BY Genre)
GROUP BY Country, Genre;

```

Figure 6 Common Genre and Country

Results		Explain	Describe	Saved SQL	History
COUNTRY		GENRE			
United States		Dramas			
1 rows returned in 0.03 seconds		Download			

Based on Figure 5 above, it is clear to see that the United States has the most available shows available on Netflix. After running the second SQL code, the most common genre in the United States is Dramas.

Question 3 Genre Variation: What are all of the genres available for movies on Netflix? More specifically, how many movies are in each genre? For this question, we used a join query to list all of the genres available in the movie category.

```
SELECT DISTINCT Genre, COUNT(MOVIE.ShowID) AS NumberMoviesinGenre
FROM GENRE JOIN MOVIE
ON GENRE.ShowID = MOVIE.ShowID
GROUP BY Genre;
```

Figure 7 Genre Variation

GENRE	NUMBERMOVIESINGENRE
LGBTQ Movies	90
TV Mysteries	1
International Movies	2430
Classic Movies	103
Thrillers	489
Sports Movies	196
Independent Movies	671
Romantic Movies	531
Documentaries	786
Children & Family Movies	530
Faith & Spirituality	56
Sci-Fi & Fantasy	217
Stand-Up Comedy	329
Horror Movies	312
Dramas	2100
Comedies	1469
Crime TV Shows	1
Action & Adventure	719
Anime Features	56
TV Dramas	1
Music & Musicals	321
Movies	56
Cult Movies	59

The results above (Figure 7) return the amount of movies in each genre. The query returned 23 rows of different movie genres in the Netflix streaming service. Out of the 23 rows returned, the highest movie genre is international movies, followed by dramas, then comedies.

Question 4 Movies by Season: During which season are the most shows released on netflix? For this question, we created a CASE query which categorized the DateAdded by season and returned the number of movies released within each season.


```

1  SELECT CASE
2  WHEN EXTRACT(Month from DateAdded) IN (12,1,2) THEN 'Winter'
3  WHEN EXTRACT(Month from DateAdded) IN (3,4,5) THEN 'Spring'
4  WHEN EXTRACT(Month from DateAdded) IN (6,7,8) THEN 'Summer'
5  WHEN EXTRACT(Month from DateAdded) IN (9,10,11) THEN 'Fall'
6  END AS Season, COUNT>ShowID) AS NumberOfMovies
7  FROM SHOW
8  GROUP BY
9  CASE
0  WHEN EXTRACT(Month from DateAdded) IN (12,1,2) THEN 'Winter'
1  WHEN EXTRACT(Month from DateAdded) IN (3,4,5) THEN 'Spring'
2  WHEN EXTRACT(Month from DateAdded) IN (6,7,8) THEN 'Summer'
3  WHEN EXTRACT(Month from DateAdded) IN (9,10,11) THEN 'Fall'
4  END;

```

Figure 8 Movies by Season

SEASON	NUMBEROFMOVIES
Fall	2137
Winter	2058
Summer	1759
Spring	1808

This SQL code returned 4 rows of data for each season. Based on Figure 8 above, we can conclude that the most amount of movies were released in the fall followed by winter.

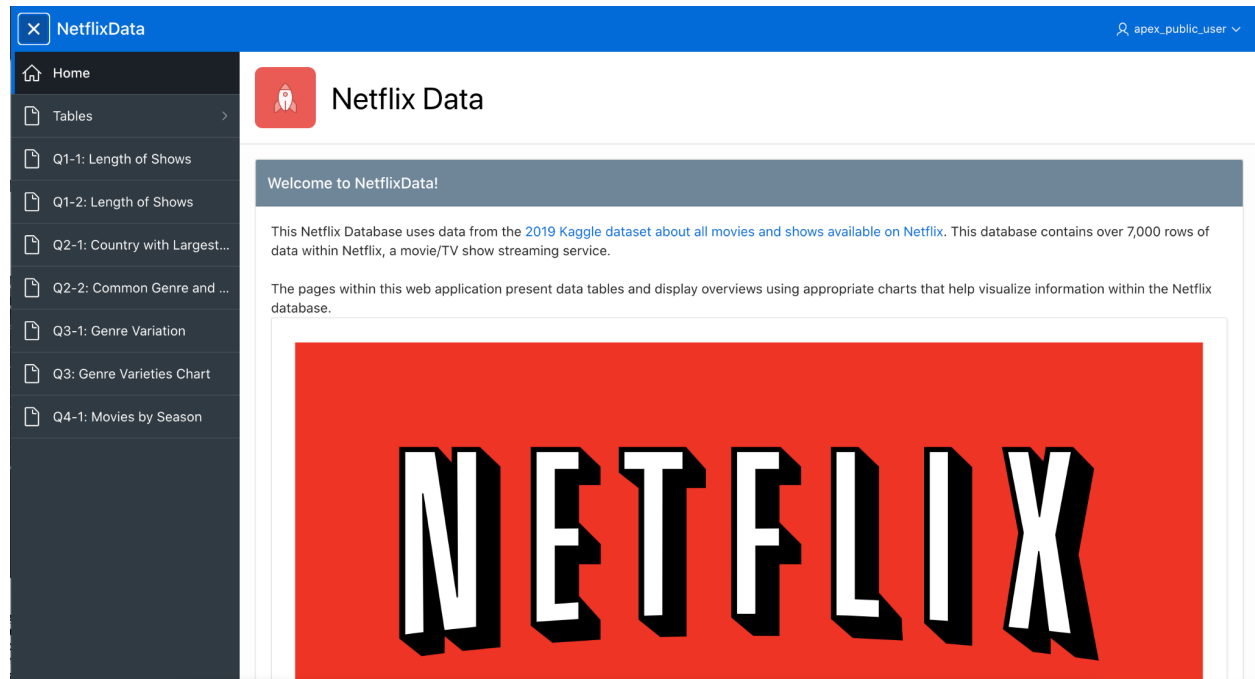
Web Design

<https://apex.oracle.com/pls/apex/dmfinalproject/r/netflixdata/home?session=16928560877871>

Home Page

The homepage to our Netflix database contains a brief description of our project (with a hyperlink of the original data linked above), a navigation menu that is organized with corresponding information, nested tables in the tables page and charts of the data for appropriate queries/questions and an image of the Netflix logo. We used the same color scheme throughout the entire application to match the colors of Netflix's logo. A matching color scheme is a good way to draw the viewer's attention to important content and visuals. Figure 9 below displays a screenshot of our homepage of our web application.

Figure 9 Homepage



Tables

We created an interactive report for all 5 tables. This allows for viewer's to search, filter and group the data (Figures 10-14). Column headings for each table have been added to the report to allow for easy interpretation of the data. Each table also has a text box above containing information pertaining to the data.

Figure 10 SHOW

NetflixData						apex_public_user
Overview						
The SHOW table stores almost 7,774 rows of data. In this data, the SHOW table contains the Show ID, Show Type, Title, Date Added, Release Year and Rating. All fields but DateAdded and Rating are required in this table.						
<div><input type="text" value="Q"/> <input type="button" value="Go"/> <input type="button" value="Actions"/></div>						
Show ID	Show Type	Title	Date Added	Release Year	Rating	
s68	Movie	13TH: A Conversation with Oprah Winfrey & Ava DuVernay	1/26/2017	2017		
s2360	TV Show	Gargantia on the Verdurous Planet	12/1/2016	2013		
s3661	TV Show	Little Lunch	2/1/2018	2015		
s4324	Movie	My Honor Was Loyalty	3/1/2017	2015		
s3738	Movie	Louis C.K.: Hilarious	9/16/2016	2010		
s3739	Movie	Louis C.K.: Live at the Comedy Store	8/15/2016	2015		
s3737	Movie	Louis C.K. 2017	4/4/2017	2017		
s7103	Movie	Tini: The New Life of Violetta	2/19/2017	2016	G	
s7104	Movie	Tinker Bell and the Legend of the NeverBeast	5/17/2015	2014	G	
s6770	Movie	The Polar Express	1/1/2021	2004	G	
s6780	Movie	The Princess and the Frog	7/16/2019	2009	G	
s6840	Movie	The Rugrats Movie	10/1/2019	1998	G	
s7629	Movie	Willy Wonka & the Chocolate Factory	1/1/2020	1971	G	
s7243	Movie	True Grit	1/1/2020	1969	G	
s5989	Movie	Tarzan	6/23/2018	1999	G	

Figure 11 TV

NetflixData

apex_public_user

Overview

The TV table contains exactly 2,404 rows of data meaning there are 2,404 TV shows available on the Netflix database. The fields displayed in the TV table are Show ID and Seasons. The Seasons field pertains to the amount of seasons in each show. Each show must have a corresponding season associated with it.

Q

Go

Actions

Show ID	Seasons
s2539	16
s5913	15
s1472	13
s5138	13
s1538	12
s7170	12
s1301	11
s2264	11
s2679	11
s1578	10
s1598	10
s2289	10

Figure 12 MOVIE

NetflixData

apex_public_user

Overview

MOVIE contains 5,370 rows of data meaning there are 5,370 movies in the Netflix database. This table displays the Show ID and Duration. Duration records the minutes of each movie. Each movie must have both a Show ID and Duration.

Q

Go

Actions

Show ID	Duration
s4754	112
s4810	112
s4850	112
s4862	112
s5136	112
s5159	112
s5160	112
s5434	112
s5441	112
s5561	112
s5773	112
s5937	112

Figure 13 COUNTRY

NetflixData

apex_public_user

Overview

The COUNTRY table contains 9,552 rows of data. The reason why this number exceeds the amount of TV shows and movies there are in the database is because some TV shows/movies may be streamed in more than one country. The COUNTRY table contains two fields: Show ID and Country. Each TV show and movie must have a country associated with it.

Q

Go

Actions

Show ID	Country
s1132	Denmark
s5392	Czech Republic
s5685	Czech Republic
s6034	Czech Republic
s6184	Czech Republic
s6758	Czech Republic
s6777	Czech Republic
s7153	Czech Republic
s7783	Czech Republic
s170	Denmark
s231	Denmark
s766	Denmark

Figure 14 GENRE

NetflixData

apex_public_user

Overview

The GENRE table contains 17,036 rows of data. The reason why this table is significantly greater than the amount of shows and movies offered in the Netflix database is because some shows/movies may fall under multiple genre categories. The GENRE table consists of the Show ID and Genre fields.

Q

Go

Actions

Show ID	Genre
s1	International TV Shows
s1	TV Dramas
s1	TV Sci-Fi & Fantasy
s10	Horror Movies
s10	International Movies
s10	Thrillers
s100	Dramas
s100	International Movies
s100	Sports Movies
s1000	Comedies
s1001	British TV Shows
s1001	Docuseries

Queries

Q1-1: The first question is attached below (Figure 15) which displays the longest movie in the Netflix database. There is also a text box attached at the top of the page to provide information about the question and data returned for the viewer.

Figure 15 Longest Movie on Netflix

NetflixData

apex_public_user

Overview

Question: What is the title and length of the longest movie available on Netflix?

Answer: The longest movie available on Netflix is Black Mirror: Bandersnatch and it is 312 minutes.

Longest Movie

Title ↑	Duration
Black Mirror: Bandersnatch	312

Download

1 - 1

Q1-2: As a follow up question, we asked what the longest TV show in the Netflix database is. Figure 16 below presents the longest TV show that there is on Netflix.

Figure 16 Longest TV Show on Netflix

The screenshot shows the NetflixData application interface. At the top is a blue header bar with a menu icon, the text "NetflixData", and a user profile icon labeled "apex_public_user". Below the header is a red "Overview" section containing a question and an answer. The question is "What is the title and length of the longest TV show available on Netflix?" and the answer is "The longest TV show available on Netflix is Grey's Anatomy and it has 16 seasons." Below this is a section titled "Longest TV Show" which contains a table with two columns: "Title" and "Seasons". The table has one row with "Grey's Anatomy" and "16". Below the table is a "Download" button and a pagination indicator "1 - 1".

Title	Seasons
Grey's Anatomy	16

Q2-1: Our second research question analyzes the country with the most amount of Netflix shows available. The screenshot below displays the most common country within the database (Figure 17).

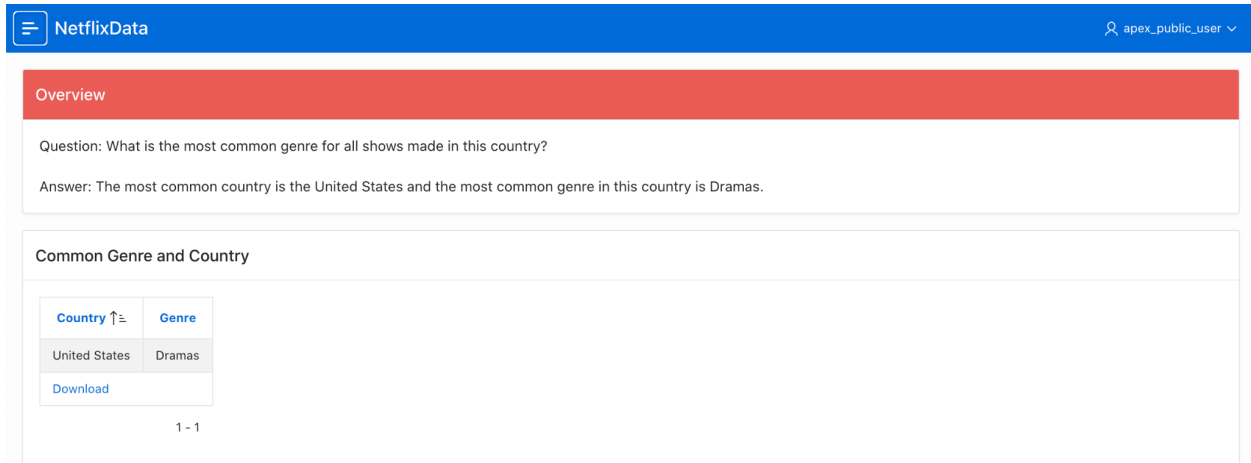
Figure 17 Most Popular Country

The screenshot shows the NetflixData application interface. At the top is a blue header bar with a menu icon, the text "NetflixData", and a user profile icon labeled "apex_public_user". Below the header is a red "Overview" section containing a question and an answer. The question is "Which country has the most amount of shows available in it?" and the answer is "United States has the most amount of shows available in it on Netflix." Below this is a section titled "Country with Largest Selection" which contains a table with one column: "Country". The table has one row with "United States". Below the table is a "Download" button and a pagination indicator "1 - 1".

Country
United States

Q2-2: As a follow up question, we asked what the most popular genre in that most popular country is. The figure below shows what the most popular genre is.

Figure 18 Most Popular Genre



Q3: The third question within this database that we calculated was how many of each genre are there available on the Netflix platform. We counted each genre in the SQL query. We attached a pie chart (figure 20) below to help visualize all the different types of genres along with their popularity in this database. The pie chart is also interactive to show how many genres there are available of each within the Netflix database. Figure 19 below displays each genre within the Netflix streaming service.

Figure 19 Genres in the Netflix Database

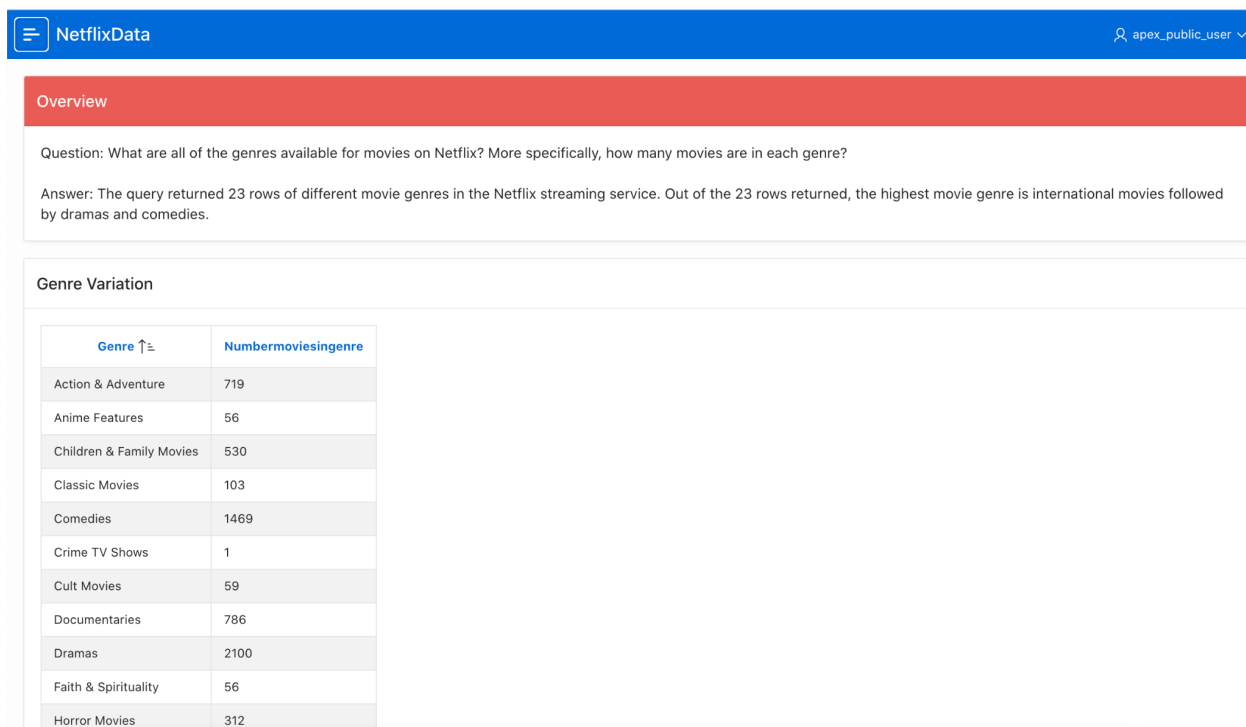
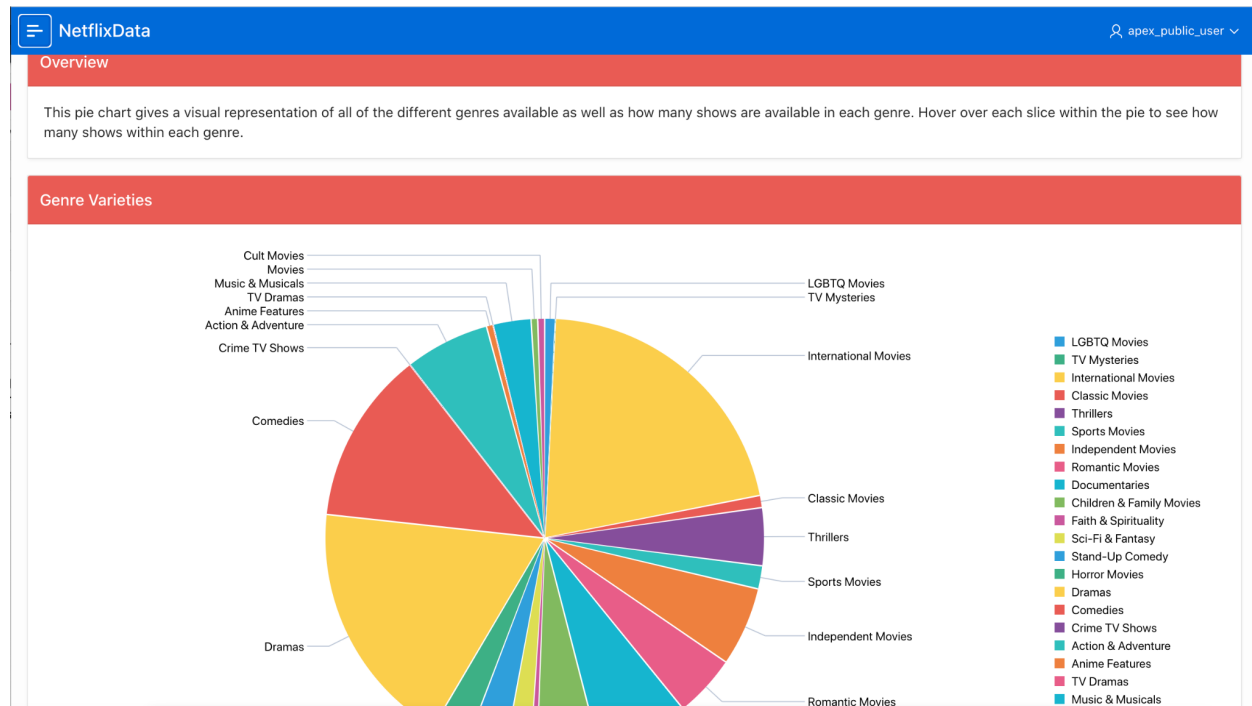


Figure 20 Genre Varieties Chart



Q4: For the final question, we calculated what the most popular season on Netflix was. We created a bar chart to provide a better visual for the viewer. In the figure below, it is clear to see which season released the most amount of shows. The bar chart is also interactive as it states exactly how many shows were released in each season.

Figure 21 Most Popular Season

