Introduction

Here is the Introduction

Methods

Extract: your original data sources and how the data was formatted (CSV, JSON, pgAdmin 4, etc).

Transform: what data cleaning or transformation was required.

Load: the final database, tables/collections, and why this was chosen.

Data sources:

Different sources have been verified. However, following sources were considered as they have majority of the required information.

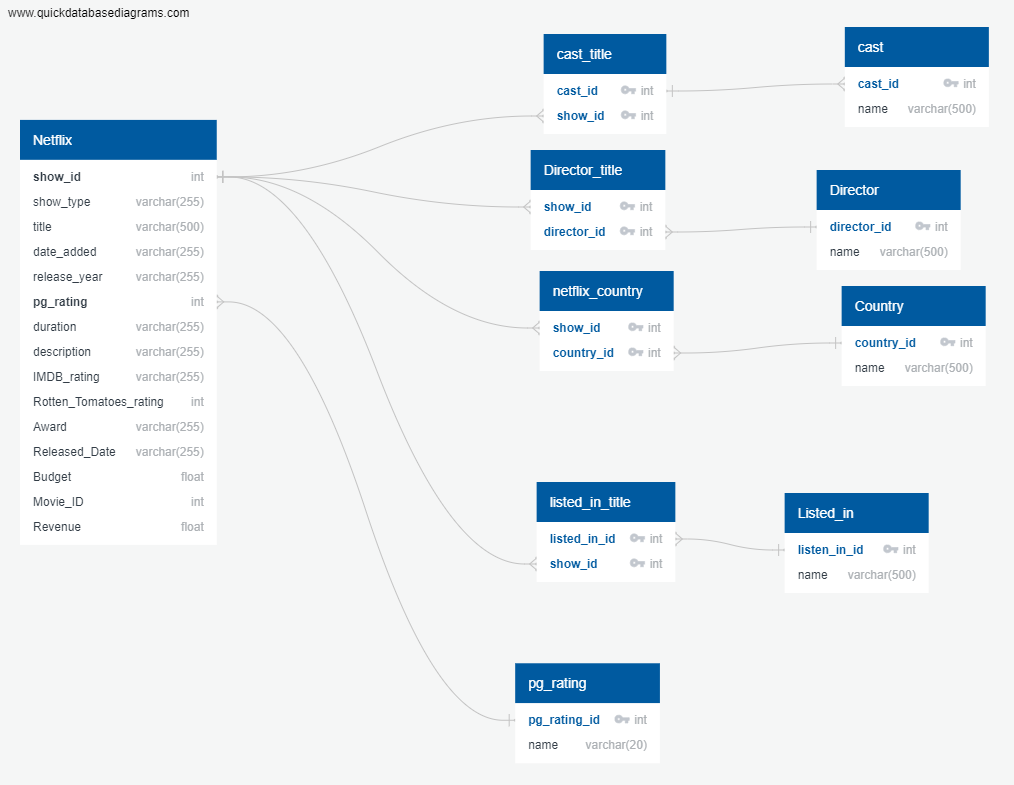
* -https://www.kaggle.com/shivamb/netflix-shows?select=netflix\_titles.csv
* -Public movie APIs:
* OMDB - OMDb to obtain information and metadata about movies.
* TMDb - TMDb for movie discovery.

https://www.kaggle.com/shivamb/netflix-shows?select=netflix\_titles.csv

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Field Name: | Description: | Example: | Data Type: | Unique: | Null: | Comments: |
| show\_id | Unique ID for every Movie / Tv Show | 81145628 | String | Yes | No |  |
| type | Identifier - A Movie or TV Show | Movie | String | No | No |  |
| title | Title of the Movie / Tv Show | Norm of the North: King Sized Adventure | String | No | No |  |
| director | Director of the Movie | Richard Finn, Tim Maltby | String | No | yes | Multiple names in each field |
| cast | Actors involved in the movie / show | Alan Marriott, Andrew Toth, Brian Dobson, Cole Howard, Jennifer Cameron, Jonathan Holmes, Lee Tockar... | String | No | Yes | Multiple names in each field |
| country | Country where the movie / show was produced | United States, India, South Korea, China | String | No | yes | Multiple names in each field |
| date\_added | Date it was added on Netflix | September 9, 2019 | Date | No | yes |  |
| release\_year | Actual Release year of the move / show  text\_formatratingsort  TV Rating of the movie / | 2019 | Integer | No | No |  |
| rating | TV Rating of the movie / show | TV-PG | String | No | Yes |  |
| duration | Total Duration - in minutes or number of seasons | 90 min | String | No | No |  |
| listed\_in | Genre | Children & Family Movies, Comedies | String | No | No | Multiple names in each field |
| description | The summary description | Before planning an awesome wedding for his grandfather, a polar bear king must | String | No | No |  |

This is OMDB API extract:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Field Name: | Description: | Example: | Data Type: | Unique: | Null: | Comments: |
| Title | Title of the Movie / Tv Show | Man of Steel | String | No | No |  |
| Year | Actual Release year of the move / show  text\_formatratingsort  TV Rating of the movie / | 2013 | Integer | No | No |  |
| Rated | TV Rating of the movie / show | PG-13 | String | No | Yes |  |
| Released | Release date | 14 Jun 2013 | String | No | Yes |  |
| Runtime | Duration | 143 min | String | No | Yes |  |
| Genre | Type of movie | Action,Adventure, Sci-Fi | String | No | No |  |
| Director | Director | Zack Snyder | String | No | No |  |
| Writer | Author | David S. Goyer (screenplay)  David S. Goyer (story) | String | No | No |  |
| Actors | Actor and Actress | Henry Cavill  Amy Adams  Michael Shannon  Diane Lane | String | No | No |  |
| Plot | Description | An alien child is evacuated from his dying world and sent to Earth to live among humans. His peace is threatened  when other survivors of his home planet invade Earth | String | Yes | No |  |
| Language |  | English | String | No | No |  |
| Country | Country of producing the product | UA, UK | String | No | No |  |
| Awards | Received Awards for the product | 7 wins & 46 nominations | String | No | Yes |  |
| Poster | create by the studio used to promote or advertise the title | N/A | String | No | No |  |
| Source | IMDb registered users who casts a vote | N/A | String | No | Yes |  |
| Value | Accurate Percentage of the Rating of respected critics | 56% | Float | No | Yes |  |
| Source | Rating of critics | Rotten Tomatoes | String | No | Yes |  |
| Value | Percentage of the Rating of respected critics | 55/100 | Float | No | Yes |  |
| Metascore | Rating of respected critics | 55 | Int | No | Yes |  |
| imdbRating | Rating on IMDB | 7.0 | Int | No | Yes |  |
| imdbVotes | Number of votes for the product | 92 | Int | No | Yes |  |
| imdbID | IMDB Movie ID Number | tt0770828 | String | Yes | No |  |
| Type | Movie or TV-Series | movie | Float | No | No |  |
| DVD | If the product is available on DVD | N/A | String | No | Yes |  |
| BoxOffice | Box Office receives data from a variety of sources, including film studios, distributors, and production companies from around the world | N/A | String | No | Yes |  |
| Production | Production company | Syncopy | String | No | Yes |  |



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Field Name: | Description: | Example | Data Type | Unique | Null | Comments |
|
| backdrop\_path: varchar(255) |  |  |  |  |  |  |
| genre\_ids: Numbers (series) |  |  |  |  |  |  |
| id: varchar (10) |  |  |  |  |  |  |
| original\_language: char (2) |  |  |  |  |  |  |
| original\_title: varchar (255) |  |  |  |  |  |  |
| overview: varchar (1500) |  |  |  |  |  |  |
| popularity: float |  |  |  |  |  |  |
| poster\_path: varchar(255) |  |  |  |  |  |  |
| release\_date: date |  |  |  |  |  |  |
| title: varchar (255) |  |  |  |  |  |  |
| video: TRUE/FALSE |  |  |  |  |  |  |
| vote\_average: float |  |  |  |  |  |  |
| vote\_count: int (edited) |  |  |  |  |  |  |

Extraction:

Extracting & Transformation of Netflix data from [www.kaggle.com](http://www.kaggle.com)

Extracting & Transforming of Budget, Revenue and rating via API to OMDB and TMDB

Extracting & Transformation of Director data

Extracting & Transformation of the Country that movie/TV Series data

Extracting & Transformation of rating Data

Extracting & Transformation of cast Data

Tables:

Table:

Table:

Load on to SQL

Loading ? to SQL Table

Final Database

Findings

## sql db mODEL

The following is the SQL Data Model used in this project:

adult: TRUE/FALSE  
backdrop\_path: varchar(255)  
genre\_ids: Numbers (series)  
id: varchar (10)  
original\_language: char (2)  
original\_title: varchar (255)  
overview: varchar (1500)  
popularity: float  
poster\_path: varchar(255)  
release\_date: date  
title: varchar (255)  
video: TRUE/FALSE  
vote\_average: float  
vote\_count: int (edited)