



NETFLIX - EDA

By
Zappcode Academy



CREATE DATABASE

☐ create database netflix;

☐ use netflix;

Upload titles.csv as table for Netflix database

☐ select * from titles;

TOTAL NUMBER OF RECORDS

-- Count total number of records

```
SELECT COUNT(*) AS total_records
```

```
FROM titles;
```

UNIQUE VALUES IN EACH COLUMN

```
SELECT
COUNT(DISTINCT show_id) AS unique_show_ids,
COUNT(DISTINCT type) AS unique_types,
COUNT(DISTINCT title) AS unique_titles,
COUNT(DISTINCT director) AS unique_directors,
COUNT(DISTINCT cast) AS unique_casts,
COUNT(DISTINCT country) AS unique_countries,
COUNT(DISTINCT date_added) AS unique_dates_added,
COUNT(DISTINCT release_year) AS unique_release_years,
COUNT(DISTINCT rating) AS unique_ratings,
COUNT(DISTINCT duration) AS unique_durations,
COUNT(DISTINCT listed_in) AS unique_listed_in,
COUNT(DISTINCT description) AS unique_descriptions
FROM titles;
```

COUNT MISSING VALUES IN EACH COLUMN

```
SELECT  
SUM(CASE WHEN director IS NULL THEN 1 ELSE 0 END) AS missing_directors,  
SUM(CASE WHEN cast IS NULL THEN 1 ELSE 0 END) AS missing_cast,  
SUM(CASE WHEN country IS NULL THEN 1 ELSE 0 END) AS missing_country,  
SUM(CASE WHEN date_added IS NULL THEN 1 ELSE 0 END) AS missing_dates_added,  
SUM(CASE WHEN rating IS NULL THEN 1 ELSE 0 END) AS missing_ratings,  
SUM(CASE WHEN duration IS NULL THEN 1 ELSE 0 END) AS missing_durations  
FROM titles;
```

DISTRIBUTION OF TYPE

-- Distribution of Type

```
SELECT type, COUNT(*) AS count
```

```
FROM titles
```

```
GROUP BY type
```

```
ORDER BY count DESC;
```

DISTRIBUTION OF RATINGS

-- Distribution of Rating

```
SELECT rating, COUNT(*) AS count
```

```
FROM titles
```

```
GROUP BY rating
```

```
ORDER BY count DESC;
```

DISTRIBUTION OF RELEASE YEAR

-- Distribution of Release Year

```
SELECT release_year, COUNT(*) AS count
```

```
FROM titles
```

```
GROUP BY release_year
```

```
ORDER BY release_year;
```


TOP 10 DIRECTORS

-- Top 10 Most Common Directors

```
SELECT director, COUNT(*) AS count
```

```
FROM titles
```

```
GROUP BY director
```

```
ORDER BY count DESC
```

```
LIMIT 10;
```

TOP 10 COUNTRIES

-- Top 10 Countries

```
SELECT country, COUNT(*) AS count
```

```
FROM titles
```

```
GROUP BY country
```

```
ORDER BY count DESC
```

```
LIMIT 10;
```

RECORDS ADDED PER MONTH

-- Records added per month

```
SELECT DATE_FORMAT(STR_TO_DATE(date_added, '%b %d, %Y'), '%Y-%m') AS  
month, COUNT(*) AS count
```

```
FROM titles
```

```
GROUP BY month
```

```
ORDER BY month;
```

RECORDS ADDED EACH YEAR

-- Records added per year

```
SELECT YEAR(STR_TO_DATE(date_added, '%b %d, %Y')) AS year, COUNT(*) AS  
count
```

```
FROM titles
```

```
GROUP BY year
```

```
ORDER BY year;
```

FIND RECORDS FOR DRAMAS

-- Count dramas records

SELECT COUNT(*) AS count

FROM titles

WHERE listed_in LIKE '%dramas%';

AVERAGE DURATION OF MOVIES/SHOWS

-- Average duration of movies/shows

```
SELECT AVG(CAST(SUBSTRING_INDEX(duration, ' ', 1) AS UNSIGNED)) AS  
average_duration
```

```
FROM titles
```

```
WHERE duration REGEXP '^[0-9]+ min$';
```

RECORDS WHERE DIRECTOR DETAILS ARE NULL

-- Find records with No director info

```
SELECT COUNT(*) AS count
```

```
FROM titles
```

```
WHERE director="";
```

RECORDS RELEASED IN LAST 5 YEARS

-- Records released in last 5 Years

SELECT *

FROM titles

WHERE release_year >= YEAR(CURDATE()) - 5;

RECORDS FOR INDIA

-- Records for India Country

SELECT type, title

FROM titles

WHERE country = 'India';

RECORDS WHICH OFFER SEASONS

-- Find records which have seasons

SELECT *

FROM titles

WHERE duration LIKE '%seasons%';

RECORDS WITH MORE THAN 2 SEASONS

-- Records which have more than 2 Seasons

SELECT *

FROM titles

WHERE duration LIKE '%seasons%'

AND CAST(SUBSTRING_INDEX(SUBSTRING_INDEX(duration, ' ', 1), ' ', -1) AS
UNSIGNED) > 2;