Netflix Data Analysis Project Presentation

1: Title

Title: Netflix Data Analysis Project
Subtitle: Cleaning, Analyzing & Visualizing Netflix Data using MySQL & Excel
Presented by: Shivam Patel

2: Introduction

- Netflix is a leading global streaming service with a vast content library.
- This project aims to analyze Netflix data by:
    - Cleaning raw data using **MySQL**.
    - Identifying duplicate and missing values.
    - Performing data analysis to extract insights.
    - Visualizing findings in **Excel**.

3: Problem Statement

- The dataset contains:
    - Missing values in fields like **director** and **country**.
    - Duplicate records that can affect analysis accuracy.
    - Inconsistent formats in date, duration, and rating columns.
- Goals:
    - **Clean and preprocess** the data using MySQL.
    - **Analyze and visualize** trends in Netflix content.

4: Data Cleaning Using SQL Queries

Handling Missing Values.

- Standardized date formats for consistency.

SQL Query:

```
CREATE TABLE netflix_cleaned AS
SELECT
    show_id,
    type,
    title,
    director,
    country,
    STR_TO_DATE(date_added, '%m/%d/%Y') AS date_added,
    release_year,
    rating,
    Coalesce(CASE
                WHEN duration LIKE '%min' THEN
                CAST(SUBSTRING_INDEX(duration, ' ', 1) AS
                UNSIGNED)
                ELSE NULL
            END, 0) AS duration_min,
    Coalesce(CASE
                WHEN duration LIKE '%Season%' THEN
                CAST(SUBSTRING_INDEX(duration, ' ', 1) AS
                UNSIGNED)
                 ELSE NULL
            END, 0) AS num_seasons,
    listed_in
FROM netflix1;
```

5: Checking for Duplicate Values

Finding Duplicate Entries

SQL Query:

```
SELECT show_id, title, COUNT(*) AS duplicate_count
FROM netflix_cleaned
GROUP BY show_id, title
HAVING COUNT(*) > 1;
```

- No duplicate records were found in the dataset.

6: Data Analysis Using SQL Queries

Top 10 Countries with Most Content

```sql
SELECT country, COUNT(*) AS total_shows
FROM netflix_cleaned
GROUP BY country
ORDER BY total_shows DESC
LIMIT 10;
```

Most Popular Genres

```sql
SELECT listed_in, COUNT(*) AS genre_count
FROM netflix_cleaned
GROUP BY listed_in
ORDER BY genre_count DESC
LIMIT 10;
```

Trend of New Content Over the Years

```sql
SELECT release_year, COUNT(*) AS total_releases
FROM netflix_cleaned
GROUP BY release_year
ORDER BY release_year ASC;
```

Average Movie Duration

```sql
SELECT ROUND(AVG(duration_min), 2) AS avg_movie_duration
FROM netflix_cleaned
WHERE type = 'Movie';
```

7: Data Visualization in Excel

Key Visualizations Created:

✅**Bar Chart:** Top 10 countries with the most Netflix content.
✅**Bar Chart:** Distribution of Netflix content by genre.
✅**Line Chart:** Trend of new Netflix content releases.
✅**Bar Chart:** Average movie duration.

8: Key Insights & Results

- **USA leads** with the highest number of Netflix shows.
- **Drama and Comedy** are the most common genres.
- **Increase in Netflix content** releases after 2015.
- **Average movie duration** is around **99.73 minutes**.


9: Conclusion

- Netflix content is highly concentrated in specific regions and genres.
- **SQL helped clean and analyze** the data efficiently.
- **Excel provided clear visual insights** into content trends.
- This project showcases skills in **data cleaning, SQL analysis, and Excel visualization**.

Country	Total_shows
United States	68
United Kingdom	7
South Africa	3
Pakistan	124
Not Given	20
Japan	12
India	12
Germany	4
France	4
China	2
Grand Total	**256**

Total_shows

Top 10 Countries Producing Netflix Content
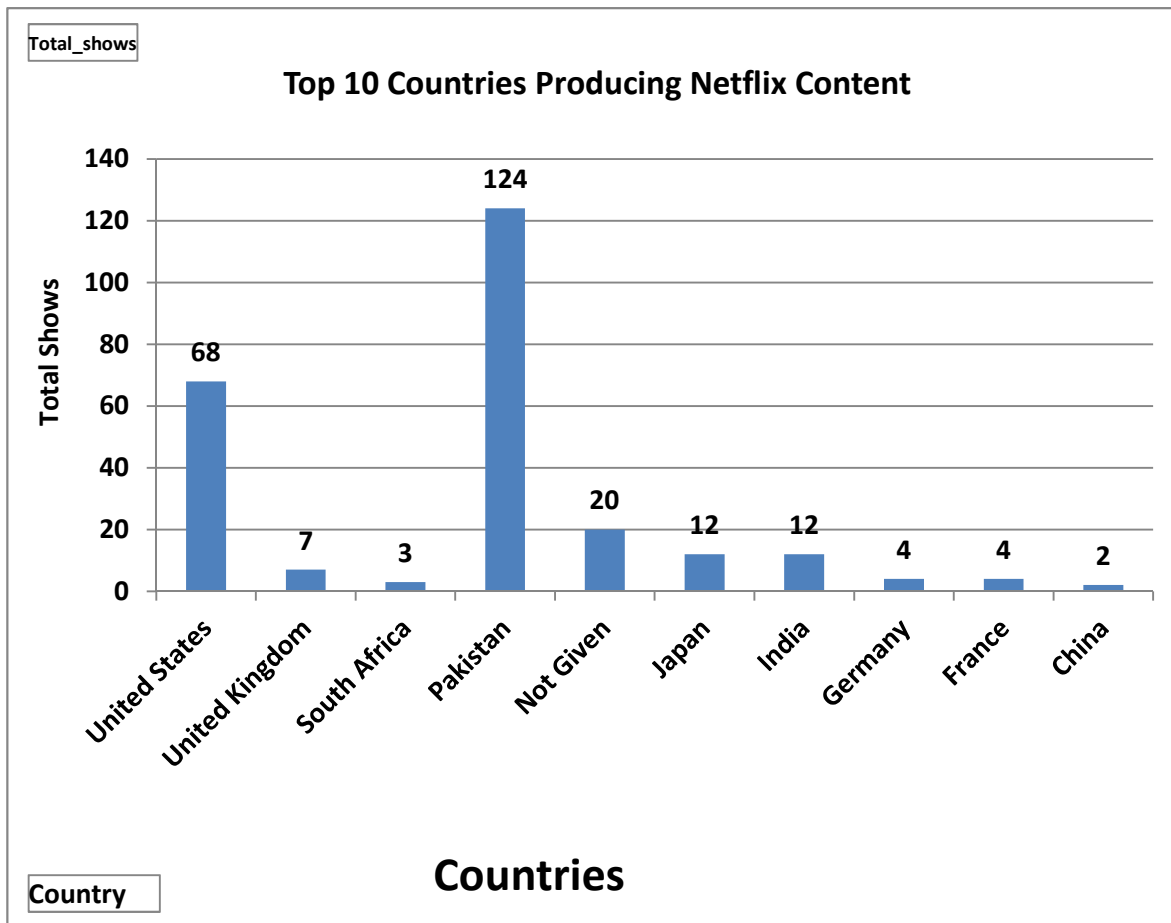
Countries

Country

Listed_in	genre_count
Reality TV	8
Kids' TV, TV Comedies	8
Kids' TV	12
International TV Shows, Romantic TV Shows, TV Comedies	8
Docuseries, International TV Shows	6
Docuseries	6
Documentaries	7
Children & Family Movies, Comedies	9
Children & Family Movies	7
Action & Adventure, Anime Features, International Movies	12
Grand Total	**83**

genre_count

Most Popular Netflix Genres

listed_in
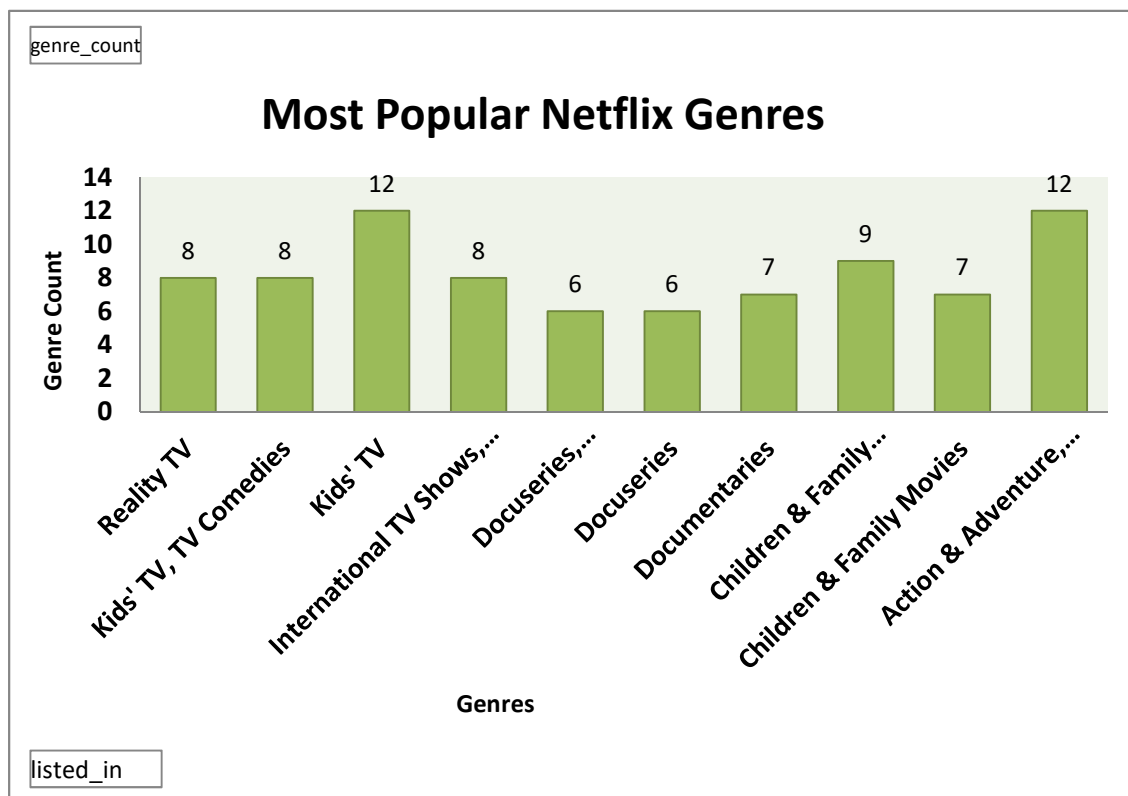
Release Year	Count of New Release
2011	2
2012	1
2013	5
2015	4
2016	5
2017	8
2018	12
2019	13
2020	30
2021	117
Grand Total	**197**

Count of New Release

Trends of New Content Over the Year

Count of New Release

140

120

100

80

60

40

20

0

117

30

2   1   5   4   5   8   12   13

2011 2012 2013 2015 2016 2017 2018 2019 2020 2021

Release Year

release_year
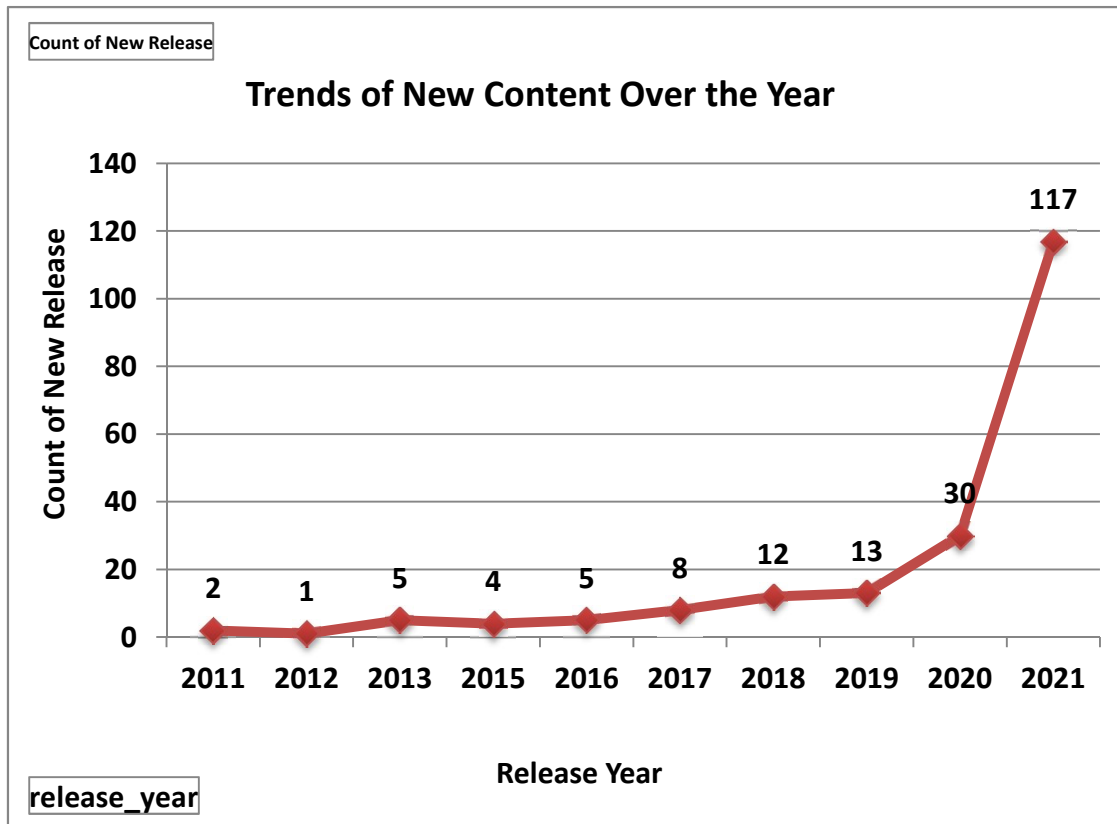
type	Movie

Average of duration_min

99.73

type

Average of duration_min

Average Movie Duration

99.73

Duration min	
120 min	
100 min	
80 min	
60 min	
40 min	
20 min	
0 min	

Total

Movie