Netflix_MongoDB_Report

(Netflix Dataset)

📊 Part 1: Data Import Report

✅ Dataset Description

- **Dataset Name:** Netflix Movies & TV Shows
- **Source:** Kaggle - Netflix Shows Dataset
- **Format:** CSV file (`netflix_titles.csv`)
- **Size:** 8,807 records
- **Fields Included:** `type`, `title`, `director`, `cast`, `country`, `date_added`, `release_year`, `rating`, `duration`, `listed_in`, `description`

⚡ Import Setup (Docker)

- MongoDB container setup using:

```
docker run --name mongodb \
  -p 27017:27017 \
  -d \
  -v ~/dsc505:/data/db \
  mongo:5
```

↳ Import Command Used

```
docker exec mongodb mongoimport \
  --db dsc505 \
  --collection netflix \
  --type csv \
  --headerline \
  --file /data/db/netflix_titles.csv
```

✉ Result

- **8,807 documents** successfully imported.
- Verified by running:

```
db.netflix.countDocuments()
```

- Output: `8807`

🚀 Challenges & Resolutions

- Minor adjustments to path during import based on Docker volume
- No major issues faced during import

🔢 Part 2: Analysis Results

✅ 1. Count of All Records

`db.netflix.countDocuments()`

- Result: **8807**

```
dsc505> db.netflix.countDocuments()
8807
```

✅ 2. Sample Record Preview

`db.netflix.findOne()`

- Example Record:
  - o Title: *Dick Johnson Is Dead*
  - o Type: *Movie*
  - o Country: *United States*
  - o Release Year: *2020*
  - o Duration: *90 min*

```
dsc505> db.netflix.findOne()
{
  _id: ObjectId('680829959b44d8a674420746'),
  show_id: 's1',
  type: 'Movie',
  title: 'Dick Johnson Is Dead',
  director: 'Kirsten Johnson',
  cast: '',
  country: 'United States',
  date_added: 'September 25, 2021',
  release_year: 2020,
  rating: 'PG-13',
  duration: '90 min',
  listed_in: 'Documentaries',
  description: 'As her father nears the end of his life, filmmaker Kirsten J
ohnson stages his death in inventive and comical ways to help them both face
 the inevitable.'
}
```

✅ 3. Count by Type (Movie vs TV Show)

```
db.netflix.aggregate([
  { $group: { _id: "$type", count: { $sum: 1 } } }
])
```

- Output:
    - TV Show: 2,676
    - Movie: 6,131

```
dsc505> db.netflix.aggregate([
...    { $group: { _id: "$type", count: { $sum: 1 } } }
... ])
[ { _id: 'TV Show', count: 2676 }, { _id: 'Movie', count: 6131 } ]
dsc505>
```

✔️ **4. Top 5 Countries by Number of Titles**

```
db.netflix.aggregate([
{ $group: { _id: "$country", count: { $sum: 1 } } },
{ $sort: { count: -1 } },
{ $limit: 5 }
])
```

- Output:
    - United States: 2,818
    - India: 972
    - Unknown (empty): 831
    - United Kingdom: 419
    - Japan: 245

```
dsc505> db.netflix.aggregate([
...    { $group: { _id: "$country", count: { $sum: 1 } } },
...    { $sort: { count: -1 } },
...    { $limit: 5 }
... ])
[
  { _id: 'United States', count: 2818 },
  { _id: 'India', count: 972 },
  { _id: '', count: 831 },
  { _id: 'United Kingdom', count: 419 },
  { _id: 'Japan', count: 245 }
]
```

✔️ **5. Earliest Release Year**

```
db.netflix.aggregate([
  { $group: { _id: null, minYear: { $min: "$release_year" } } }
])
```

- Output: 1925

```
dsc505> db.netflix.aggregate([
...     { $group: { _id: null, minYear: { $min: "$release_year" } } }
... ])
[ { _id: null, minYear: 1925 } ]
```

🧑‍🔬 Insights & Summary

- **Movies** constitute the majority of the content on Netflix (about 70%).
- The **United States** and **India** are the most common countries of production.
- Netflix content spans nearly a century, with the earliest title released in **1925**.
- MongoDB queries helped easily group and extract summary statistics from over 8,800 records.