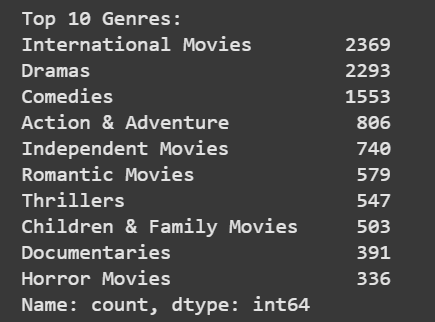
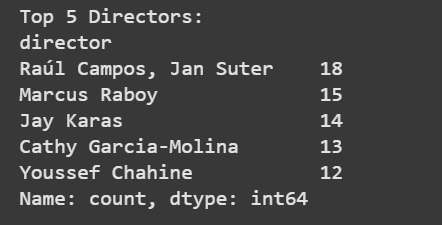
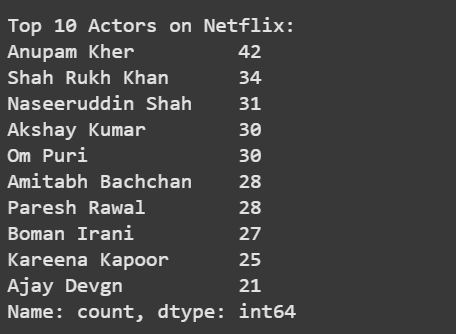
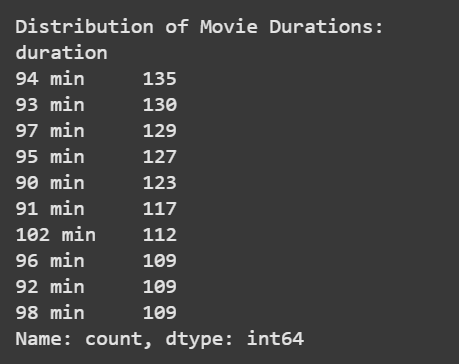
**Netflix Insights Report**

**Hemavarshini S**

**1. Data Cleaning Steps (using Pandas)**

* Removed duplicate rows.
* Converted the date\_added column to proper datetime format.
* Filled missing values:
  + country → "Unknown"
  + director → "Not Available"
  + cast → "Not Available"
  + rating → "Not Rated"
  + duration → "0"
* Removed unnecessary spaces from categorical columns.
* Removed special characters from country.
* Dropped rows with missing title.
* Saved a cleaned dataset (NETFLIX\_INF\_CLEANED.csv).

**2. Metrics and Insights**

* **Total records after cleaning**: 5332
* **Movies vs TV Shows**:
  + Movies = 5185
  + TV Shows = 147
* **Top 10 countries with most content**:  
  
* **Top 10 genres**:  
  
* **Top 5 directors**:  
  
* **Content added per year**:  
  
* **Top 10 actors**:  
  
* **Distribution of Movie Durations**:  
  

**3. Encoding**

To prepare the dataset for future machine learning tasks, I also applied encoding:

* **Label Encoding** → Converted rating column into numeric codes.
* **Frequency Encoding** → Applied to country and primary\_genre.
* **One-Hot Encoding** → Expanded listed\_in column into multiple binary genre columns.