CA

What is contained within a sheet? make your own notes about the content.

1. There is one sheet.
2. There are 8809 movies/tv shows listed rows.
3. There are 12 attributes of data on each movie/tv show.

**show\_id**: Unique identifier for each title.

**type**: Indicates whether the title is a movie or a TV show.

**title**: Name of the title.

**director**: Name of the director (may be missing for some titles).

**cast**: Names of the main cast members (may also be missing).

**country**: Country of origin.

**date\_added**: When the title was added to Netflix.

**release\_year**: Year the title was originally released.

**rating**: The maturity rating (e.g., PG-13, TV-MA).

**duration**: Duration in minutes for movies or number of seasons for TV shows.

**Unnamed columns (16–25)**: These seem to be unnecessary or contain null values and can likely be ignored or removed during analysis.

What would you like to discover from the data set?

1. When was the earliest recorded TV Show/Movies in the Netflix?
2. Summarize the data: number of movies vs. TV shows, distribution of release years, etc.
3. Top directors, actors, and countries contributing to the platform.
4. What factors influence the popularity of Netflix titles?
5. Is there a trend in content release on Netflix based on genre or country?
6. How does Netflix's content diversity compare to that of other streaming services?

The context of your questions, the industry the dataset is from, typical data for that industry

The dataset likely comes from the **streaming media industry**, which has been rapidly expanding with major players such as Netflix, Hulu, Amazon Prime, and Disney+. A key trend in this industry is the **globalization of content**, where platforms are increasingly offering international content to cater to diverse audiences. Another important trend is the **shift from licensed content to original productions**, as companies seek to retain viewers with exclusive shows and films.

**Typical Data in the Streaming Industry:** Common attributes in datasets from this industry often include:

* **Title**, **Genre**, and **Type** (Movie or TV Show)
* **Release Date**
* **Country of Production**
* **Director and Cast**
* **Duration** or **Seasons**
* **Language** and **Subtitle Options**
* **Viewer Ratings** and **Popularity Metrics** (though this might not always be available publicly)