# **Milestone 2 : Project dataset updated**

## Top 10000 Popular Movies Dataset

<https://www.kaggle.com/datasets/omkarborikar/top-10000-popular-movies>

| **Column Name** | **Description** |
| --- | --- |
| id | Every movie has its unique ID. |
| original\_language | There are total 44 languages present in this column. Total 7771 movies with 'English' as original language. Values in this column are ISO 639-1 codes of languages. I.e 'en' for 'English' , 'hi' for 'Hindi' etc. |
| original\_title | Title of the movie. |
| popularity | Popularity of movie. Bigger the number , higher the popularity. |
| release\_date | Release date of the movie. If release date is not present for any movie , then that movie is not released yet. |
| vote\_average | Average of rating/vote for the movie. |
| vote\_count | Number of ratings/vote recorded for the movie. |
| genre | Genre of the movie. |
| overview | Brief description of movie in string format. |
| revenue | Revenue of Movie |
| runtime | Runtime of movie in minutes. |
| tagline | Tagline of the movie |

I am planning to use primary columns and will be discarding the un-necessary columns which I feel, that won’t play an important role as other columns.

Discarding column : overview and tagline