**Capstone Project**

**Dataset Description**

This data set contains information about +7000 movies extracted from TMDB API.

**Columns Descriptions**

**Release\_Date:** Date when the movie was released.

**Title:** Name of the movie.

**Overview:** Brief summary of the movie.

**Popularity:** It is a very important metric computed by TMDB developers based on the number of views per day, votes per day, number of users marked it as "favorite" and "watchlist" for the data, release date and more other metrics.

**Vote\_Count:** Total votes received from the viewers.

**Vote\_Average:** Average rating based on vote count and the number of viewers out of 10.

**Original\_Language:** Original language of the movies. Dubbed version is not considered to be original language.

**Genre:** Categories the movie it can be classified as.

**Poster\_Url:** Url of the movie poster.

**Techniques to be used:**

* Data Wrangling
* Data Cleaning
* Data Visualization

**EDA Questions**

Q1: What is the most frequent genre in the dataset?

Q2: What genres has highest votes?

Q3: What movie got the highest popularity? what's its genre?

Q4: Which year has the most filmmed movies?