

## **Exploration Summary**

- We have a data frame consisting of 9827 rows and 9 column.
- Our dataset looks a bit tidy with no Nan and duplicate values.
- ReleaseDate column needs to be casted into date time and extract only the year.
- Overview, Original\_Language and Poster-Url wouldn't be so useful during analysis.
- There are noticeable outliers in Popularity column.
- Vote\_Average better be categorized for proper analysis.
- Genre column has comma separated values and white space that needs to be handle.

## **Conclusion**

### **Q1. What is the most frequent genre in the dataset?**

Ans:- Drama genre is the most frequent genre in the dataset and has appeared more than 14% of the times among 19 other genres.

### **Q2. What genre has the highest votes?**

Ans:- We have 25.5% of our dataset with popular vote(6520 rows). Drama again gets the highest popularity among fans by having more than 18.5% of movies popularity.

### **Q3. What movie got the highest popularity? What's its genre?**

Ans:- Spider-Man: No Way Home has the highest popularity rate in our dataset and it has genres of action, Adventure and Science Fiction.

### **Q4. What movie got the lowest popularity? What's its genre?**

Ans:- The united states, thread' has the highest lowest rate in our dataset and it has genres of music , drama , 'war', 'sci-fi' and history` .

### **Q4: Which year has the most filmed movies?**

Ans:- year 2020 has the highest filming rate in our dataset.