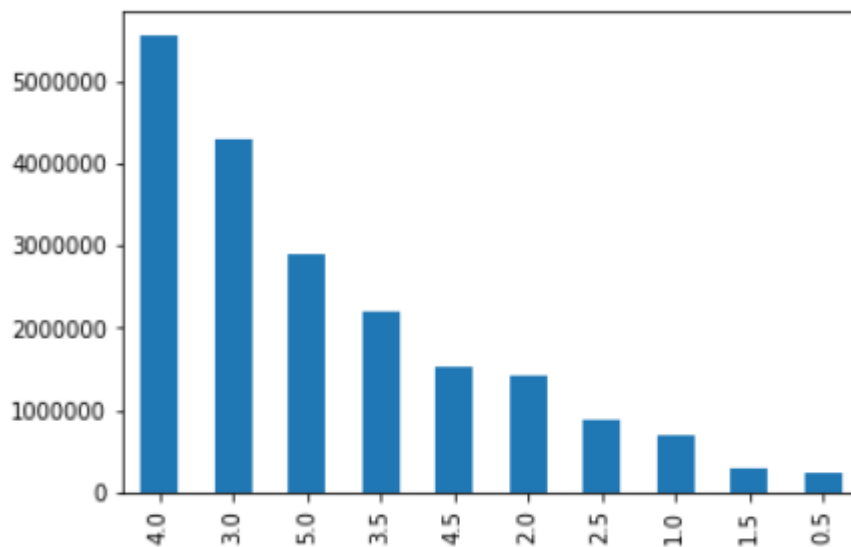


DATA STORY

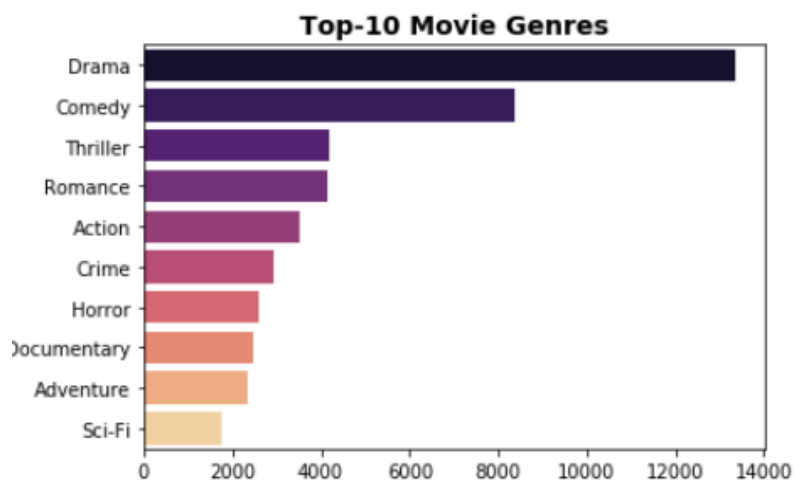
Data storytelling is about communicating your insights effectively, giving your data a voice. Data storytelling is a methodology for communicating information, tailored to a specific audience, with a compelling narrative.

1. Counting the Ratings:



From the above bar graph, we are counting the ratings given by the users for the movies. Most of the users are given rating value as 4. It seems to be good. And the least rating value given by the users are 0.5. From this we conclude that the rating tends to be relatively positive (>3). This may be due to the fact that unhappy customers tend to just leave instead of making efforts to rate. We can keep this in mind - low rating movies mean they are generally really bad

2. Top 10 Movie Genres:



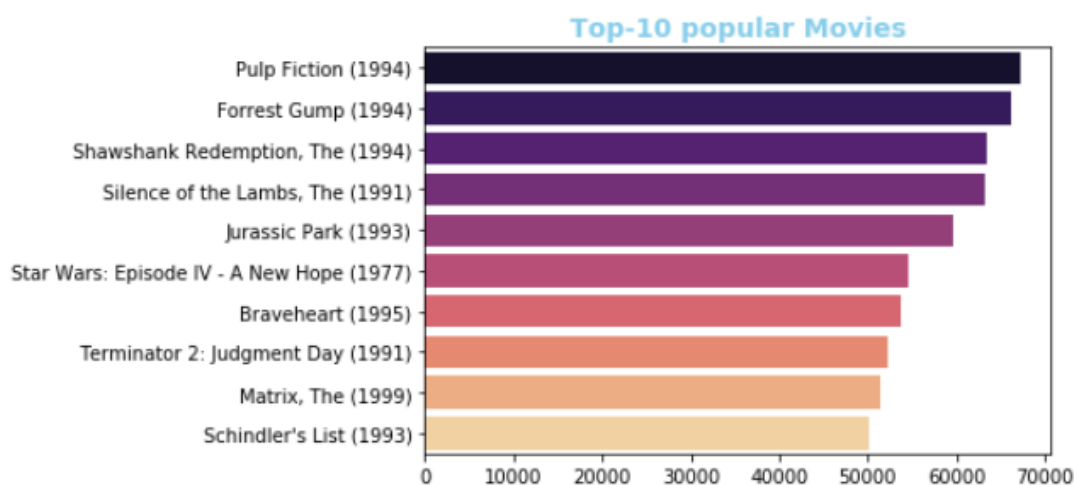
Genre column has various genres associated with that movie but we need to process on each genre separately and hence we will split the genre with the '|' operator and then use explode so that every distinct genre will be in their own row. This will ignore null or empty genres if present in the dataset and if for some reason you want them to persist then you can go for explode outer as it will also store null or empty values.

We have to group the movies by their genre and counted the number of rows to know how many movies are present in different genres. As we can see here drama won the race by some distance.

Drama is the most commonly occurring genre with almost half the movies identifying itself as a drama film. Comedy comes in at a distant second with 25% of the movies having adequate doses of humour. Other major genres represented in the top 10 are Thriller, Romance, Action, Crime, Horror, Documentary, Adventure and Sci-Fi.

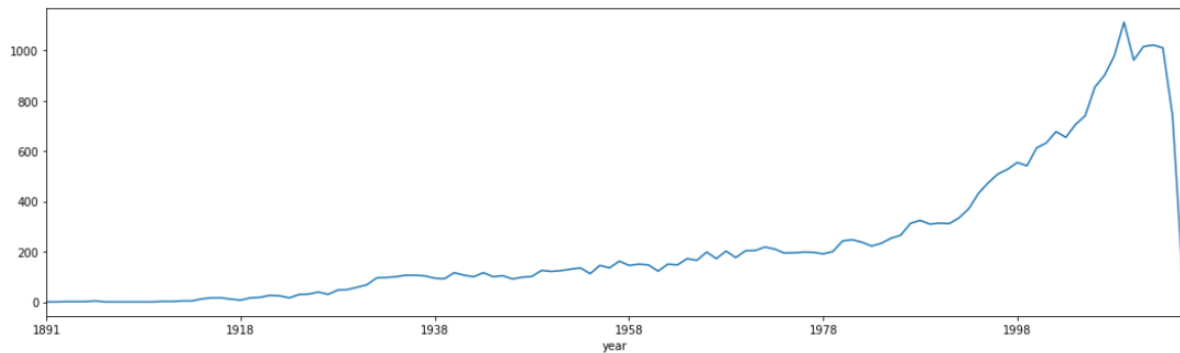
We will consider only those themes that appear in the top 10 most popular genres.

3. Top 10 Popular Movies:



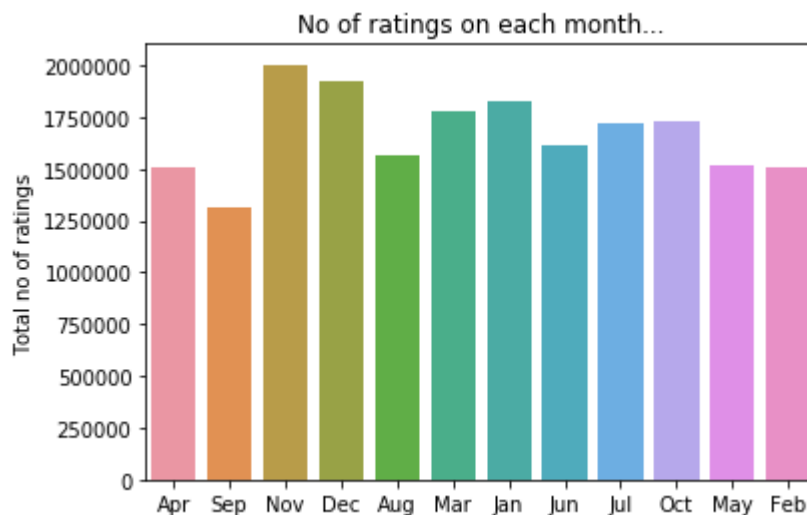
There are almost 27278 movies available with us. From which these are the top 10 popular movies we shown in the above graph. And we notice that Pulp Fiction movie which was released in the year 1994, this is the most popular movie among all the movies. That means this movie is watched by the most number of the users, and then followed by Forrest Gump, Shawshank Redemption etc. From this graph, we conclude that these are the top 10 movies watched by the users.

4. Number of movies released by the year:



From The Movie lens Dataset there are 27278 movies available with us. From that movies we are looking at the number of movies released by the year. We notice that there is a sharp rise in the number of movies starting the 1990s decade.

5. Number of ratings on each month:



From this we wanted to understand that the no of ratings differ by each month. Here, X-axis will be Month and Y-axis will be Total no of ratings. We'll notice that actually September month is far fewer than the other months. And also we noticed that lots of ratings are given by the November month.