1. **Explore the relationship between ratings, popularity, and the presence of certain actors or directors:**

Exploration: To identify any issues. We can use functions like head () to visualize the first few rows of the data and describe () to get basic descriptive statistics.

Handling missing values: Check for missing values in relevant columns (e.g., ratings, popularity, actors, or directors). You can use functions like isnull() and sum() to count the missing values in each column. Decide whether you want to remove rows with missing values or impute values using techniques like replacement with mean or median.

Cleaning format errors: Check for format errors in the relevant columns. For example, ensure that ratings and popularity are in the correct format (numbers) and don't have unwanted characters. Use functions like to\_numeric() to convert the data to numeric formats if necessary.

Eliminating irrelevant data: If there are columns that are not relevant to your analysis, remove them to reduce the dataset's size and facilitate analysis.

Data analysis: a. To determine the actors or directors with the highest average ratings, you can group the data by actor or director and calculate the mean of ratings for each. Then, you can visualize these results in a bar chart using a visualization library like Matplotlib or Seaborn. b. To analyze the correlation between popularity and the number of titles associated with specific actors or directors, you can use a correlation function (e.g., Pearson correlation coefficient) to calculate the relationship between these variables. Then, you can visualize the results in a scatter plot using a visualization library.

1. **Investigate differences in the genre distribution based on ratings on IMDB and TMDB:**

Clean and prepare the data:

Remove irrelevant columns: If there are columns that are not relevant for the analysis, you can drop them using the drop() method in Pandas. For example, if you have ID or date columns that are not needed, you can do the following:

Check and handle missing values: Use the isnull() method in Pandas to identify null values in your DataFrame and decide how to handle them. You can drop rows or columns with null values or fill them with appropriate values depending on the context.

Ensure relevant columns have the correct data types: Use the astype() method to convert columns to the correct data types if needed. For example, if you have a genre column that is a string instead of a list, you can convert it to a list of strings:

Perform data analysis and visualization:

To identify genres with the highest distribution among titles with high IMDB ratings, you can filter the DataFrame by ratings and then count the genres using the value\_counts() method. Next, you can generate a stacked bar chart using plot(kind='bar', stacked=True) from Pandas or Matplotlib.

To compare viewers' genre preferences considering ratings on IMDB and TMDB, you can perform a similar analysis using data from both sets. You can filter by ratings on TMDB, count the genres, and create a stacked bar chart comparing movie and TV show genres.