The Hackathon will comprise of two set of questions,
1. Mandatory questions
2. Task based questions

Mandatory questions:

1. Clean your data. Find all the null values in your dataset. Remove columns with categorical data that cannot be substituted. Fill columns with data that can be derived from other attributes.
   Hint: Rows without information about cast or director are not much help. But the date column could be filled if you took advantage of the year column.

2. Use the table below to generalize the ratings for both movies and TV shows. You will be using these generalized ratings in a later task.

Generalized Ratings	Ratings
R	R, TV-MA
PG	PG, TV-PG , TV-14
G	G, TV-Y, TV-G, TV-Y7
NR	NR, UR

3. Generate a visualization to identify the most popular genre in each of the countries in the dataset.

4. Trim your dataset to only shows that originate in the United States

5. Find the outliers for a given numerical attribute

Task Questions (To be applied on the trimmed dataset from q4):

1. Find the director and cast member in the United States who has the most films and identify their first movie/tv show and their favorite genre if it exists.

2. Suppose a US- based producer wants to release a PG rated show on Netflix this year. What month should the producer release the show? Justify your answer using visualizations.

3. Visualize the change in average runtimes for Drama and Comedy movies released in the United States over the past 10 years using a plot of your choice. Try to demonstrate the same in 1 diagram.

4. Visualize the distribution of the genres of all 'zombie' related media in the United States.
    Hint: Search the summary for the word 'zombie'