

Project 1 - Proposal - Group 7
Cassidy Bell, Gorgina Kareem, Jason Johnson, Erika Pino

1. Dataset-IMDB Movies Dataset

- a. Reason- Given multiple genres to show a dataset that can provide a lot of detail depending on the category.

<https://www.kaggle.com/datasets/harshitshankhdhar/imdb-dataset-of-top-1000-movies-and-tv-shows/data>

2. Inspiration - Movies and TV shows are included and movies are a part of everyday social life along with many avenues of entertainment.

3. Possible Research Questions:

- Q1 - How would the movie runtime affect the rating of the movie itself?
- Q2 - Does the gross outcome of the movie affect the metacore rating?
- Q3 - Does the genre of the movie determine the amount of award nominations? (Boxplot option)
- Q4 - Out of all the movies in the dataset, which director has the most award nominations?-(Pie Chart option)

4. Possible Visualization

- Pie Chart
- Boxplot
- Bar chart or Bar graph
- Scatterplots (Line of best fit if applicable)

5. Color Themes best on Topic

- a. Black & Yellow (IMDB colors)

6. Roles & Responsibilities

- a. Intro - Erika
- b. Objective - Cassidy
- c. Data Engineering - All
- d. Q1 - Jason
- e. Q2 - Cassidy
- f. Q3 - Gorgina
- g. Q4 - Erika
- h. Regression - Gorgina
- i. Bias & Limitations - Cassidy
- j. Conclusion - Erika
- k. Written Report - All