Rana Alqahtani

Khulud Alshamrani

Ohoud Albabtain

**PROJECT PROPOSAL:**

Using information, we scrape from the web; can we build linear regression models from which we can learn about the movie industry?

We chose to focus on movies "Based on Novels" to determine which features predict success, as measured by Worldwide Gross. Using Beautiful Soup, we scraped movie data from IMDB.com, Boxofficemojo.com. on the following features:

1. MPAA Rating (G/PG/PG-13/R)
2. Runtime
3. IMDB Score
4. Opening # Theaters (Domestic)
5. Opening Gross (Domestic)
6. Total # Theaters (Domestic)
7. Total Gross (Domestic)
8. Budget
9. Genre: Action, Adventure, Biography, Comedy, Crime, Documentary, Drama, Family, History, Horror, Music, Mystery, Romance, Sport, Thriller, War, Western

**Files:**

1. movies.csv contains the data scraped from the web
2. web\_scraping.ipynb shows the process of scraping movie data from the 2 websites
3. model\_testing.ipynb shows the process of evaluating linear regression models and plotting feature relationships