**Group 7 Project 2 - Movies Database**

Members: Lisa Santonato, Minna Wu, Dian Wang, Suchitra Sharma

**Abstract**

The objective of this project is to compare gross movie earnings data with ratings and popularity rankings from various online resources. Our process involves extracting the Movie earnings and rating data from our sources, transform it into a format that is comparative and then load it to the MySQL workbench. The project aims to answer several possible questions as outlined below and will include a summative and comparative analysis.

Possible Research Questions:

* What are the top grossing films per week?
* What are the top ranked films?
* Do the the film rankings differ between websites?
* Are the film ratings for each website normally distributed? (Or skewed/bimodal)
* What does the comparison between critic scores and user scores look like?

Additional Questions May Include:

* What is the distribution of popularity between shows in differing markets?
* What are the most popular TV shows by genre?
* What time slots are most popular, how do time slots affect popularity?
* Look at popularity of shows comparing male or female characters.

**Breakdown of Roles & Tasks:**

Extract:

Lisa: Scrape the-numbers.com for top 100 grossing movies as the base data

WangDian: Scrape Rotten Tomatoes for user vs. critic ratings

Sue: Download IMDB CSV with ratings and use movie IDs to find titles in IMDB API

Minna: Scrape Metacritic for user vs. critic ratings

Transform:

All members will be using Pandas to clean the data

TBA: Merge the data into one table using

Load:

TBA: Load into MySQL