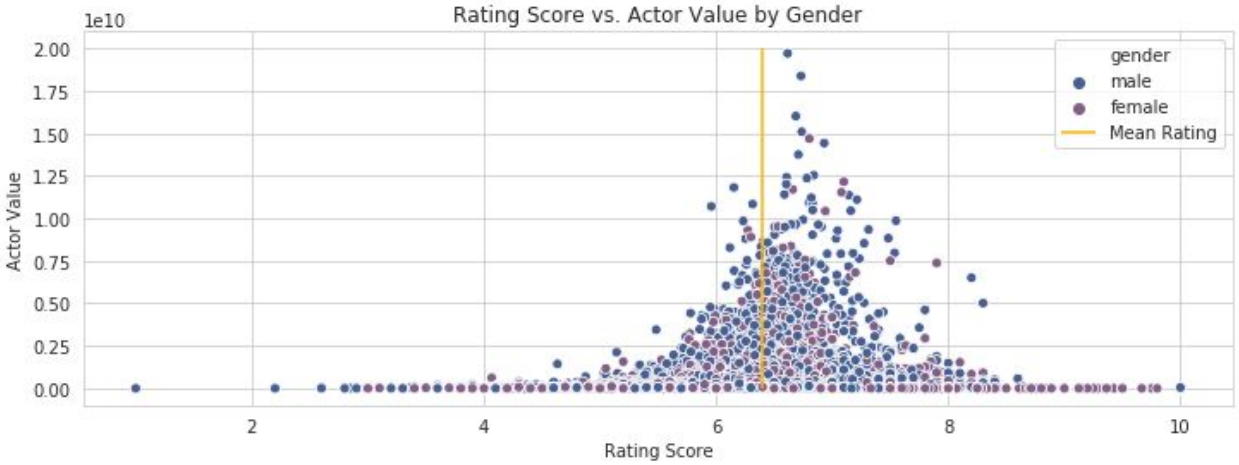
Capstone 1: Data Story

I began to visually analyze the data by searching for patterns in the actors' revenue histories and movie ratings with respect to their ages on the dates their movies were released.

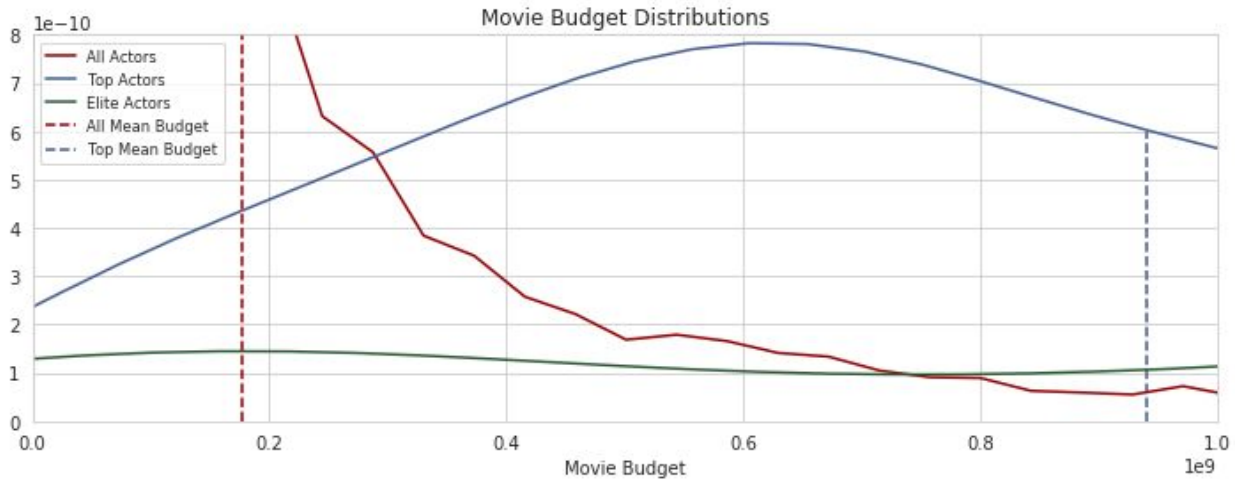


First, it could be seen that the actors who had large revenues were mostly in movies that had above average viewer ratings. This suggested a positive correlation between actors that consistently were in movies that had large ticket sales and the ratings those movies received. Also, most of the actors with high total revenues were male actors.

Next, I chose to group the actors by different tiers with respect to their earning histories. I decided to extract the top 10% and the elite 1% of the actors and to compare them to all of the other actors to identify any trends.

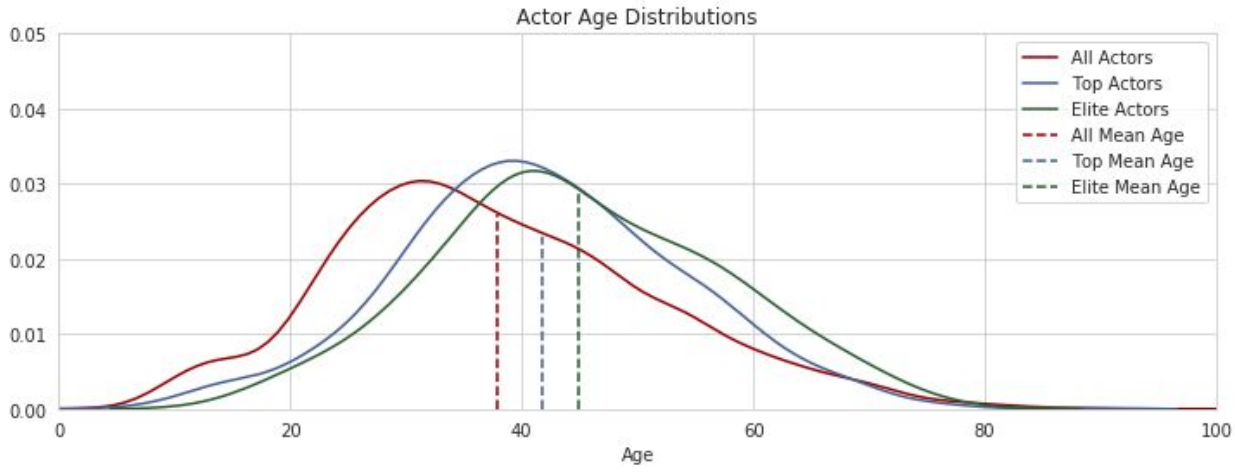
First, I looked for differences between these tiers with respect to the sum of the budgets of their movies.

The distributions showed that as the average movie budget totals grew, the elite actors came to dominate the observations. While this may be attributed to a positive correlation between movie budgets and revenues, it was interesting to observe how the elite tier of actors covered the whole range of budget totals.



On the other hand, the range of budget totals ended abruptly when observing all actors. This would suggest that the movies which featured elite actors were not just limited to movies with the largest budgets. I noted how the elite actor distribution remained horizontal throughout the entire lower range of the budget totals.

Next, I observed the distribution of the actor tiers with respect to the age of the actors on the release dates of their movies.

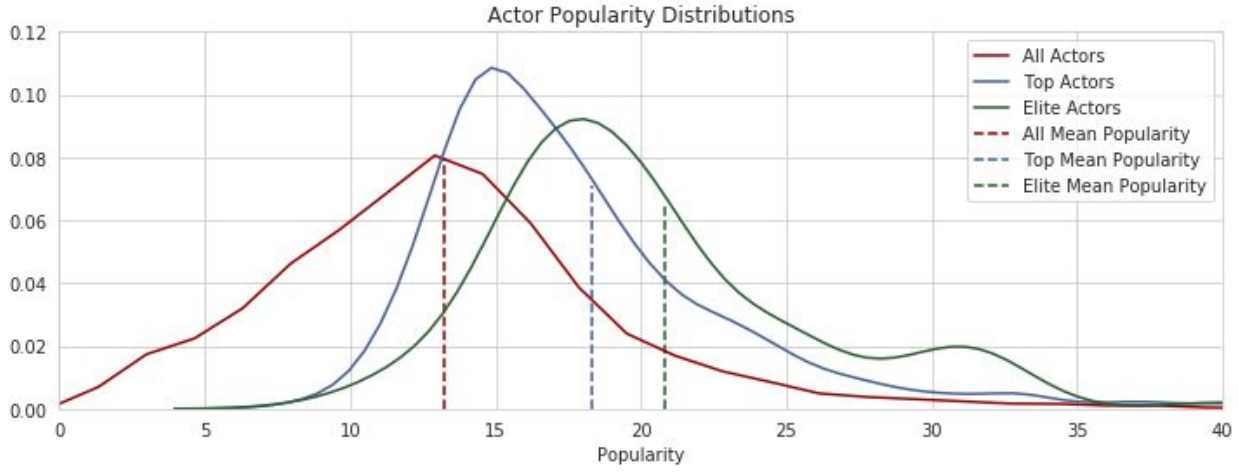


As shown in the plot, the average actors' ages increased as the actors moved up in the tiers. This would be expected, as the tiers were ranked by revenues aggregated over the entire career of the actors. Older actors would have been in more movies, which would only add to this sum.

The next variable that I observed was that of the average billing order spot for each actor. The billing orders represented the actors' rankings in the credits of their movies. Order zero, better known as the lead, represented the top billing of the movie.

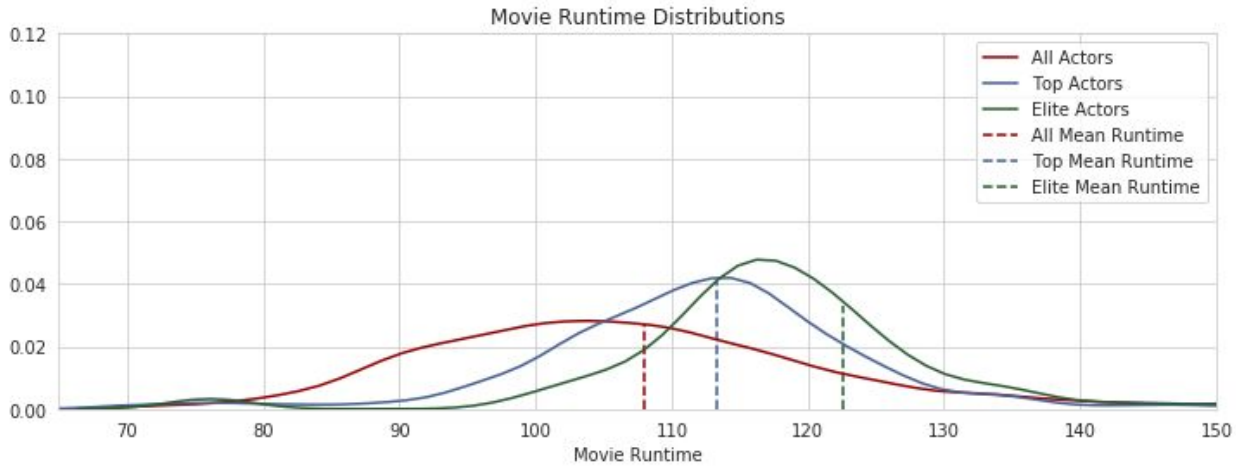
The elite actors tended to have the starring roles. This gave support to the idea of star power. This idea would assume that revenues could be increased by hiring star actors for leading roles. This distribution clearly showed that actors whose movies had the highest revenues held the top billing spots in their movies.

The next variable that I investigated was the actor's average popularity score. This was a TBDb proprietary ranking based on website traffic related to each particular actor.



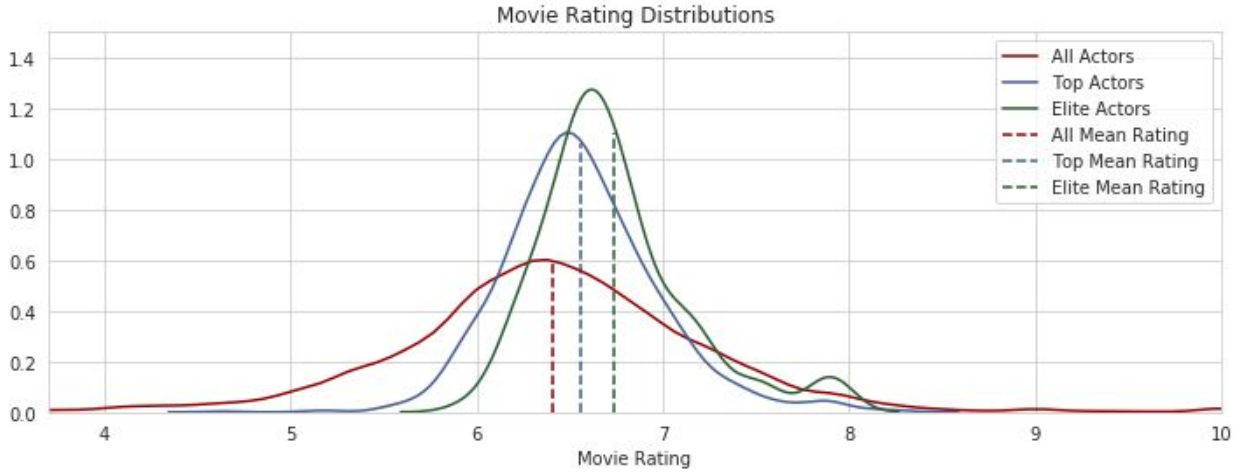
The popularity of an actor rose as the actor's movie revenue totals increased. A second mode of the distribution appeared at the higher end of the popularity scores for elite actors. Was there a super elite group of actors waiting to be discovered? This mystery would have to wait for another day.

The next observations took place on the average runtimes of the actors' movies. Runtimes had been limited to 70 minutes by design.



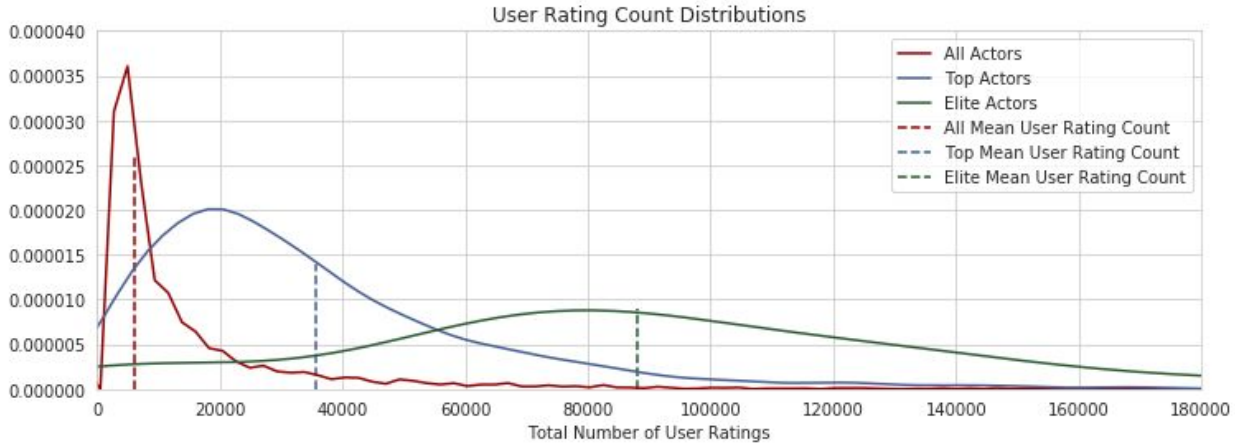
The average runtimes tended to increase with the revenue totals of the actors' movies. I speculated that since budgets had already been shown to positively correlate with revenues, runtimes should as well. This assumed that a longer movie would cost more to make, which seemed logical.

Next I observed the distribution of the average ratings for the actors' movies. These were all ratings conducted on the TMDb website.



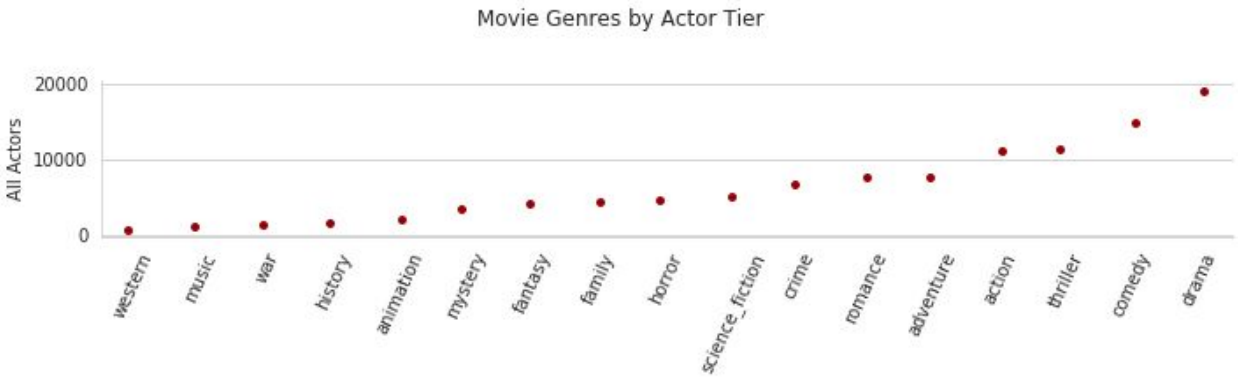
These distributions were very similar to the ones observed for the popularity rating. It should be remembered that these ratings were given to the movies, while the popularity ratings were given to the actors. This strengthened the star power theory, by showing a positive correlation between viewer sentiment for a film and sentiment for the stars in those films. The elite distribution seen in this plot had a similar second mode that was observed in the actors' popularity distribution. Was there such a thing as elite star power? The mystery deepened.

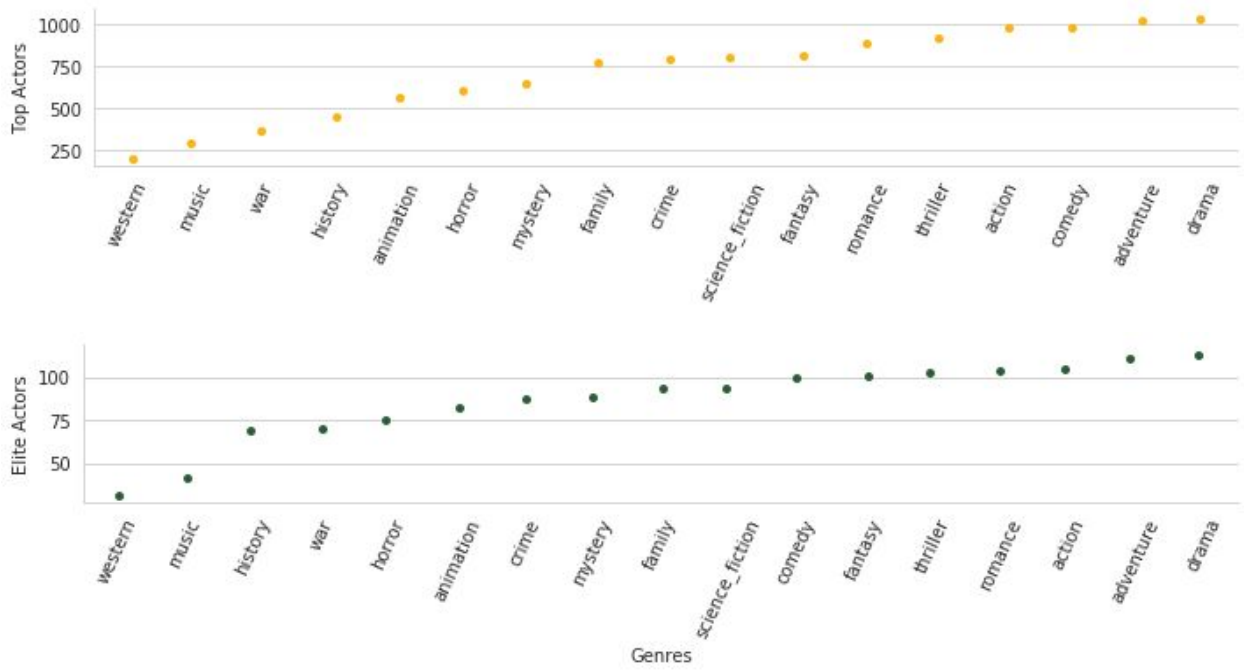
The final distribution that I examined was that of the rating counts. These were the total number of ratings received by all of the movies for each actor. This feature could be thought of as website rating activity related to each actor, regardless of the type of rating the movies received.



These plots appear very similar to the distributions of the total budgets. Even when an actor's movies had low traffic with respect to ratings, the movies of elite actors were still able to show high revenue figures. As higher budget movies would be expected to have bigger promotional campaigns, it would make sense that these actors would be recognizable enough to moviegoers to drive ticket sales without substantial media promotion.

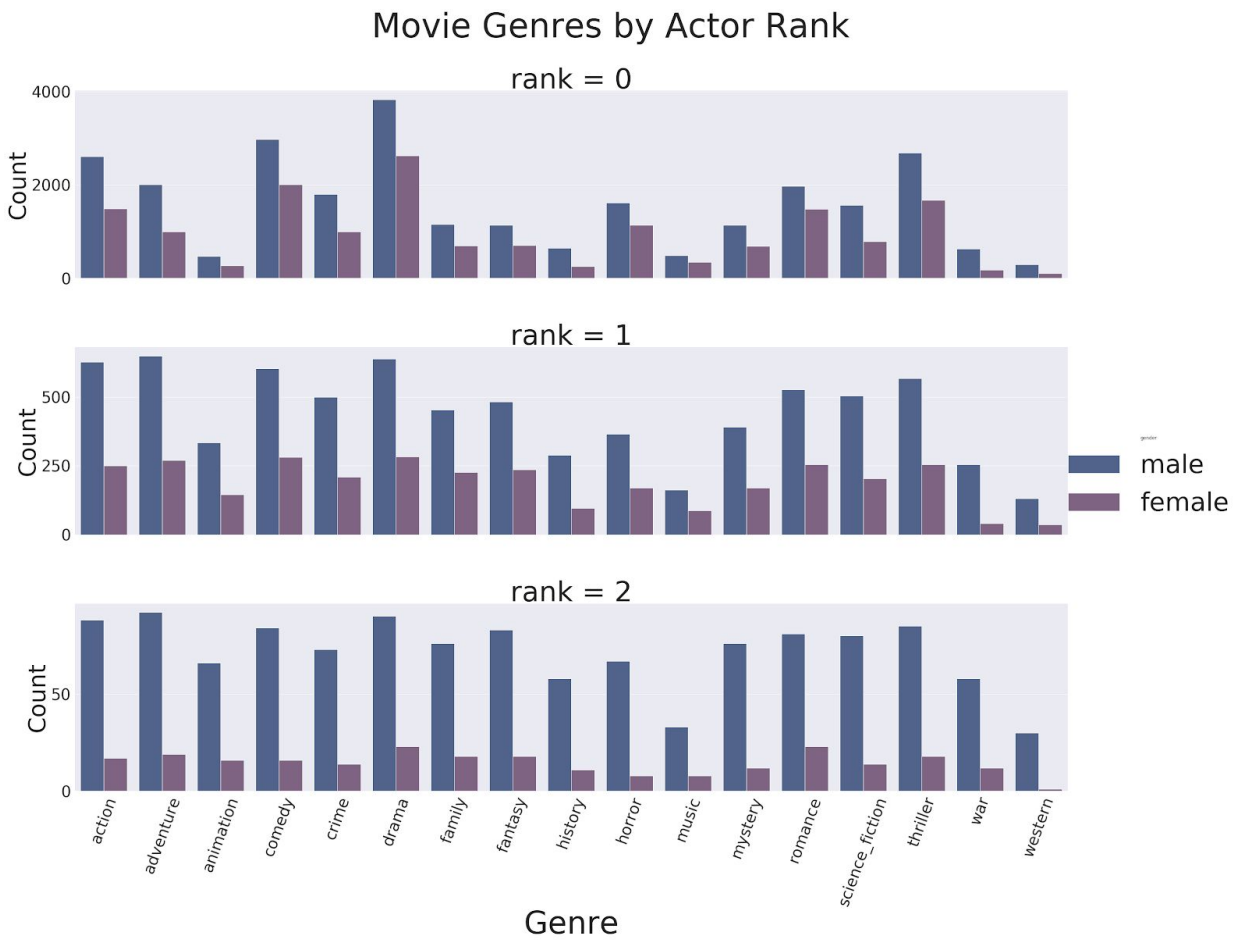
Next, I observed how the movie genres were distributed among the different tiers of actors.





Dramas were consistently the best genre when it came to tapping greater revenue potential. As the focus proceeded up the ranks to the elite tier of actors, the comedy genre declined in importance. Also at the top tier of actors, adventure movies comprised a larger portion of the genres than they did in the other tiers. Western, music, war, and history genres stayed near the bottom group of earners throughout the tiers.

Finally, I ranked the actors according to their tiers. Rank 2 was for the elite 1%, rank 1 was for the top 10%, and all other actors were assigned to rank 0. Then, I plotted the genres by actor ranks and split the distributions by gender.



All of the genres were distributed fairly evenly among the elite actors. Rank 0 actors tended to have their roles in the action, comedy, drama, and thriller movies. By the time the actors were filtered down to rank 2, it was quite apparent that the female actors were sparsely represented across all genres. In fact, male actors comprised almost 80% of that entire group. While the female actors accounted for over 40% of the observations in this dataset, they made up only 20% of the actors who were ranked in the top 1% with respect to total movie revenue.