

## Twilight Actors After Twilight

Olga Andreeva

### Research Questions

1. Which actor has acted in the most projects after their appearance in the critically acclaimed movie *Twilight* (2008)?
  - José Zúñiga, who plays their biology teacher.
2. What's the most successful movie that each of the actors have been in since *Twilight*?
  - **Gil Birmingham**: Transformers: The Last Knight (\$605,425,157)
  - **Ned Bellamy**: Django Unchained (\$425,368,238)
  - **Kristen Stewart**: Snow White and the Huntsman (\$396,592,829)
  - **Edi Gathegi**: X: First Class (\$352,616,690)
  - **Anna Kendrick**: Trolls (\$346,864,462)
  - **Jackson Rathbone**: The Last Airbender (\$319,713,881)
  - **Taylor Lautner**: Grown Ups 2 (\$246,984,278)
  - **Kellan Lutz**: Immortals (\$226,904,017)
  - **Rachelle Lefevre**: White House Down (\$205,366,737)
  - **Cam Gigandet**: The Magnificent Seven (\$162,360,636)
  - **Billy Burke**: Lights Out (\$148,868,835)
  - **Robert Pattinson**: Water for Elephants (\$117,094,902)
  - **José Zúñiga**: The Dark Tower (\$113,231,078)
  - **Elizabeth Reaser**: Ouija: Origin of Evil (\$81,705,746)
  - **Ashley Greene**: Bombshell (\$59,636,523)
  - **Justin Chon**: 21 & Over (\$48,065,672)
  - **Peter Facinelli**: Countdown (\$48,021,766)
  - **Gavin Bristol**: Extraordinary Measures (\$15,134,293)
  - **Ayanna Berkshire**: Extraordinary Measures (\$15,134,293)
  - **Bryce Flint-Sommerville**: Extraordinary Measures (\$15,134,293)
  - **Christian Serratos**: Flight 7500 (\$2,871,035)
  - **Michael Welch**: Grace Unplugged (\$2,507,201)
  - **Nikki Reed**: Empire State (\$1,212,500)
  - **Matt Bushell**: Phantom (\$1,197,759)
  - **Sean McGrath**: Recovery (\$985,197)
  - **Sarah Clarke**: Women in Trouble (\$18,000)

## Motivation and Background

Most of the actors from *Twilight* became famous as a result of starring in the movie. However, each actor independently had varying levels of success after starring in this movie. I'm interested in how things like gender/race/character relevance (primary versus secondary characters) impacted the future careers of each character. Another reason I'm interested in this is because Taylor Lautner's career seemed to not be as successful when compared to his co-stars Robert Pattinson and Kristen Stewart. I think it would be particularly interesting to compare the three main character's careers over time.

## Dataset

Scraped by myself:

[https://docs.google.com/spreadsheets/d/17LtaqD5Qyg\\_AQqNtc1kwx6LvHz0oT0j4tsweqFc\\_NnY/edit?usp=sharing](https://docs.google.com/spreadsheets/d/17LtaqD5Qyg_AQqNtc1kwx6LvHz0oT0j4tsweqFc_NnY/edit?usp=sharing)

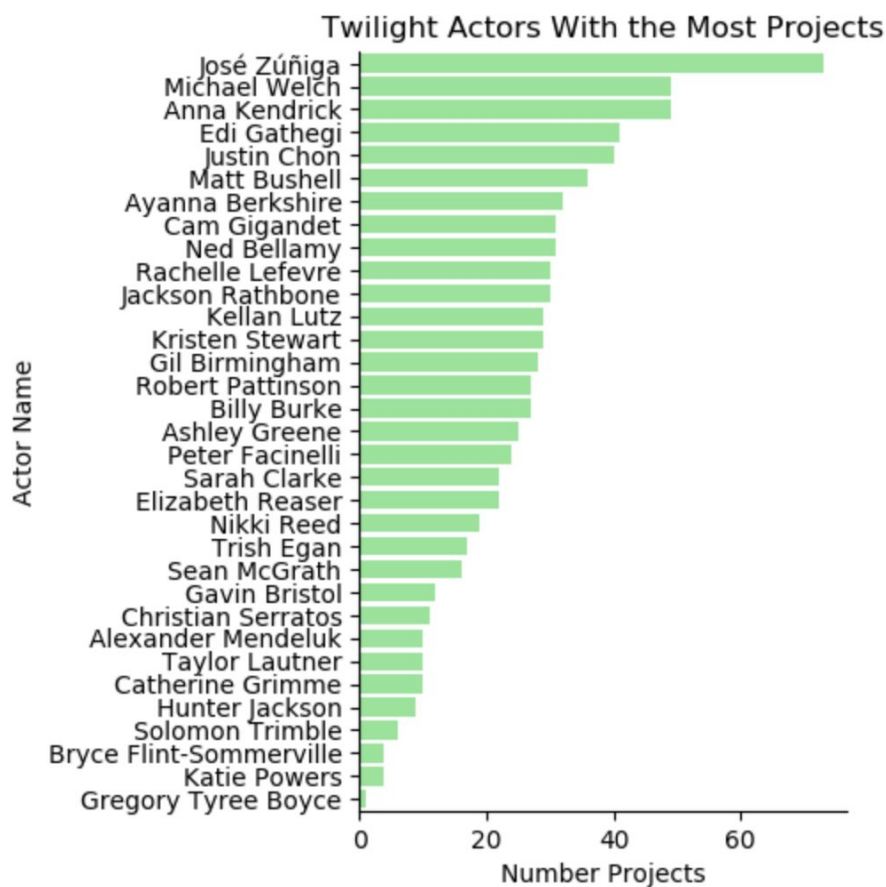
## Methodology

For the first question, I scraped *Twilight*'s IMDB page, grabbed the URL of each actor's IMDB page, and then scraped it for all of the acting projects. To make sure that I'm only looking at acting projects after *Twilight*, I ignored any projects that were released before or during 2008. Then, I added up all the remaining projects for the total number of projects that each actor has starred in since *Twilight* was released. I organized it into a table where each actor has their own row defined as their name and the number of projects they've been in since 2008.

Additionally, I scraped each actor's project's IMDB page for the cumulative worldwide gross. Any project that doesn't have a section for cumulative worldwide gross (either because it's not provided or because it's a TV show) will be represented as a 0. Each of the projects is ranked against one another to find the maximum cumulative worldwide gross and the movie title associated with that maximum. I organized this data into a table where each actor has their own row defined as their name, the top grossing movie they've starred in, and the top grossing dollar amount.

## Results

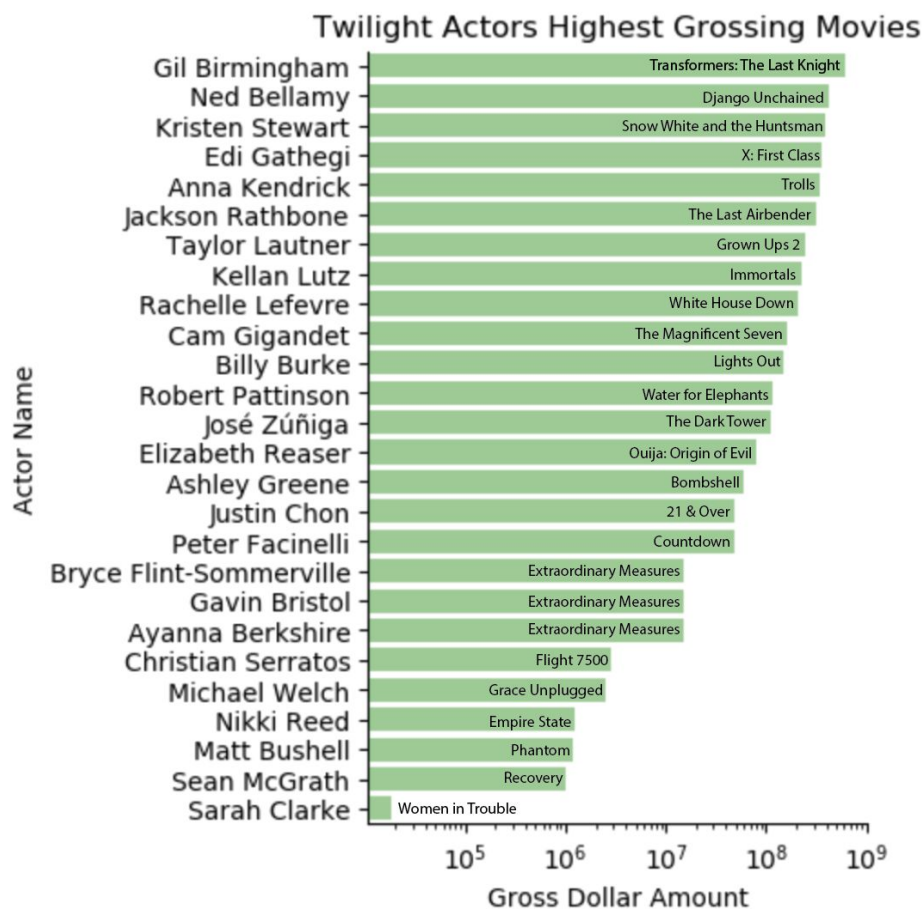
Interestingly enough, none of the trio (Robert Pattinson, Kristen Stewart, or Taylor Lautner) was the actor who acted in the most projects since *Twilight*. This honor went to José Zúñiga, who plays their biology teacher. I would have predicted that it would have been one of the three aforementioned actors, but it makes sense that a secondary character wins this. The actors who gained the most from their roles had their choice of roles after *Twilight*, so they could pick more important roles that required more time to prepare and act in. José Zúñiga, whose IMDB page consists of many singular appearances on different TV shows, would be more open to smaller roles because he can't be as picky as the trio.



The results of the most successful movie that the actors of *Twilight* have been in since 2008 was also surprising. The actors that I expected to have been in blockbusters that made a lot of money were the trio. Kristen Stewart and Taylor Lautner ranked pretty high (3rd and 7th respectfully), but Robert Pattinson was all the way down at 12th. This might be explained by the fact that Robert Pattinson tends to accept roles that are more

“indie” (*The Lighthouse*, *Good Time*, etc) and don’t tend to make as much money. This might change once he becomes the new Batman, however.

The actor that was in the highest grossing movie was Gil Birmingham, Taylor Lautner’s character’s father, who only had about 5 minutes of screentime. The actor that was in the second highest grossing movie was Ned Bellamy, who was killed off after about 3 minutes of screentime. This was interesting because it suggests that another variable separate from their performances in *Twilight* allowed them to be in such high grossing movies. Because there was such a disparity between the top grossing movie (\$605,425,157) and the bottom grossing movie (\$18,000), I had to use a log scale for the x-axis. Anna Kendrick seems to be the overall most successful actor given the parameters I established: she was 3rd for most projects she’s been in since *Twilight* and 5th for highest grossing movie, which is the highest combination of all the actors.



Note: The original seaborn plot for this data didn’t have the names of the movies overlaid because I wasn’t able to do that programmatically, so I added them in using Adobe Illustrator.

## Challenge Goals

I met both of my challenge goals of working with messy data and a new library. I had to build up the data using Beautiful Soup because the dataset that I was looking for didn't exist. Additionally, when I started doing analysis on the data, I had to go back and only look at actors whose characters had names in *Twilight* because otherwise the top n of each result were a bunch of uncredited extras. This was unexpected, but removing them from the input data into my analysis seemed reasonable because most of their future projects were also as extras.

To grab the data I wanted, I learned about how IMDB structures their webpages by examining the HTML files of pages I wanted to scrape so that I could select exactly what I wanted. Originally, I was scraping too much; instead of just selecting the acting roles that each actor had done after *Twilight*, I was also selecting producing/directing/etc roles. By updating my selection and specifying that I only wanted acting roles, I was able to get the data that I wanted.

## Work Plan

- Step 1: Learn how to use a web scraper (1 hour)
  - I've never used a web scraper so this might take more or less time
- Step 2: Scrape IMDB for the necessary data (2 hours)
  - The scraping itself seems like it might take a bit of time, but I also need to find a path for the scraping to be done
- Step 3: Do sanity checks on the data to make sure it seems reasonable ( $\frac{1}{2}$  hour)
  - I will take a look at the data and manually compare it for a couple of actors to make sure the scraper didn't mess up
- Step 4: Clean up the data using pandas ( $\frac{1}{2}$  hour)
  - If there are columns with NaN, I will get rid of them
- Step 5: Do the necessary calculations (2 hours)
  - As described in the methodology step
- Step 6 (optional): Create data visualizations of the data (3+ hours)
  - This is a personal step that I might do if I have the time... :)

My time estimates for everything ended up being more or less accurate. I ended up taking far fewer than 2 hours for the necessary calculations because I realized pandas made it very easy to find the top n given a certain column. I also took a bit more time

scraping IMDB than I allocated because I had some trouble with Beautiful Soup selections that I ended up fixing after staring at my code for too long.

## **Testing**

To make sure that my results were accurate, I went to three different actors' IMDB pages and manually counted up how many projects they had been in. This helped me catch my first bug, which was that I wasn't checking that the project date was after 2008, so I was overcounting because these projects wouldn't have been influenced by the actors' performance in *Twilight*. I also did the same thing for each actor's highest grossing movie. If I did this analysis over again, I would have accounted for TV shows because their success isn't determined by worldwide gross.

## **Collaboration**

I was able to accomplish this assignment on my own for the most part. Hunter's video about Beautiful Soup was very helpful, but I also found myself looking at the Beautiful Soup documentation when I felt like what I was trying to do wasn't working.