

# Project Writeup - Star War Survey

## Objectives

Below is the list of questions that we are looking to answer as a final outcome of this project:

1. Does the rest of America realize that “The Empire Strikes Back” is clearly the best of the bunch?
2. What are the gender specific movie likes/dislikes in the US?

## Goal Significance

The first and foremost question is what the significance of the objectives mentioned above. In other words, why does the list of questions above as a part of the objectives matter? Let’s enlist the reasons that clarifies the importance of our objectives for this project.

- This information will provide very useful information to the movie makers to identify the gender specific market potentials.
- The information would also reveal a very vital information about the gender specific likes and dislikes for the entertainment topics.

## Data

### Data Source

The data for this project is collected from the **fivethirtyeight** github repository. This data contains data behind the story America’s Favorite ‘Star Wars’ Movies (And Least Favorite Characters). The data can be downloaded from this location:

['Star Wars' Movies and Least Favorite Characters](#)

### Data Lists

The data is available in the .csv file format with “ISO-8859-1” encoding.

Data file name: `StarWars.csv`

## Data Sampling Methods

### Data Extraction Details

The data contains response of around 835 survey participants to 37 close ended questions. The discrete information reveals the history of respondents’ ‘Star Wars’ movie watch and their preferences towards different movie characters. The data also contains some limited demographic details of the respondents.

## How does the analysis progress?

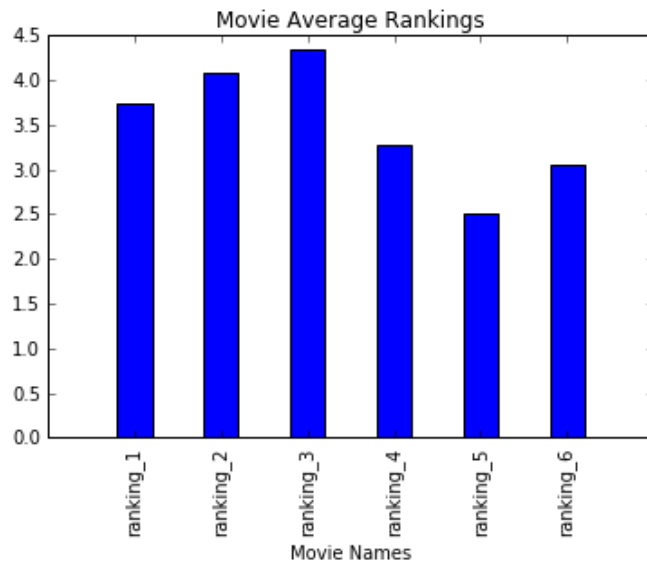
After reviewing the available info, we thoroughly clean the data to the extent so that we can extract necessary info and manipulate it to get useful inferences to server our objectives.

## Project Writeup - Star War Survey

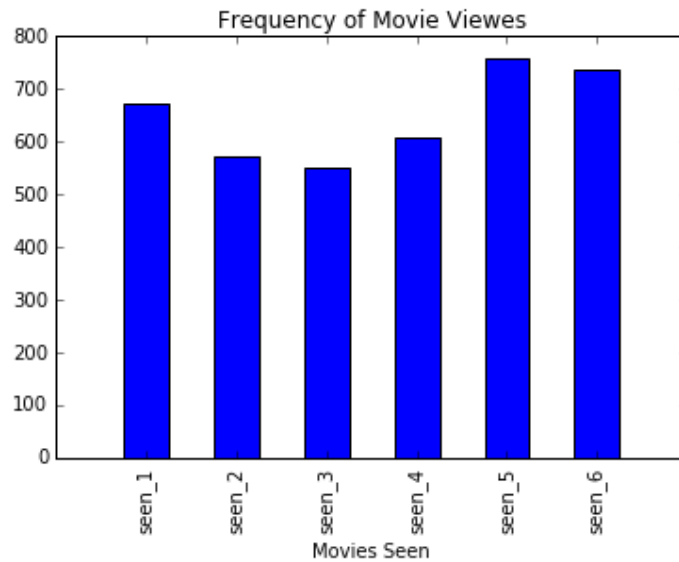
### Results

Visualize the results.

- **Average Ranking of the movies:**

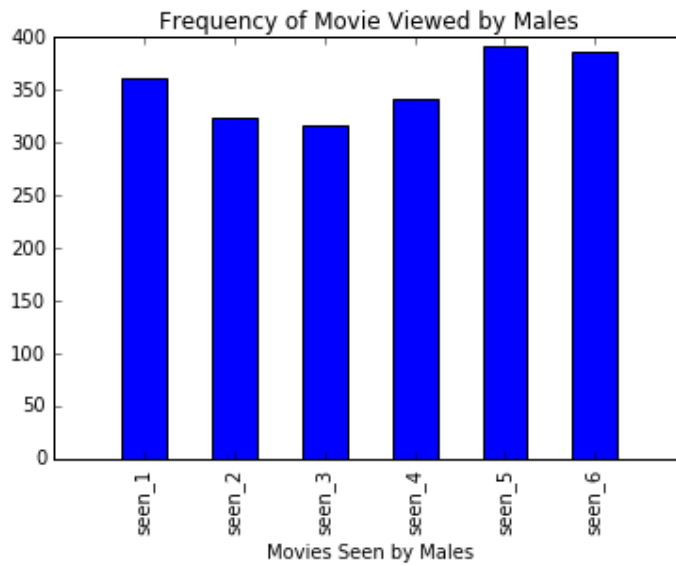


- **Frequency of Movie Views**

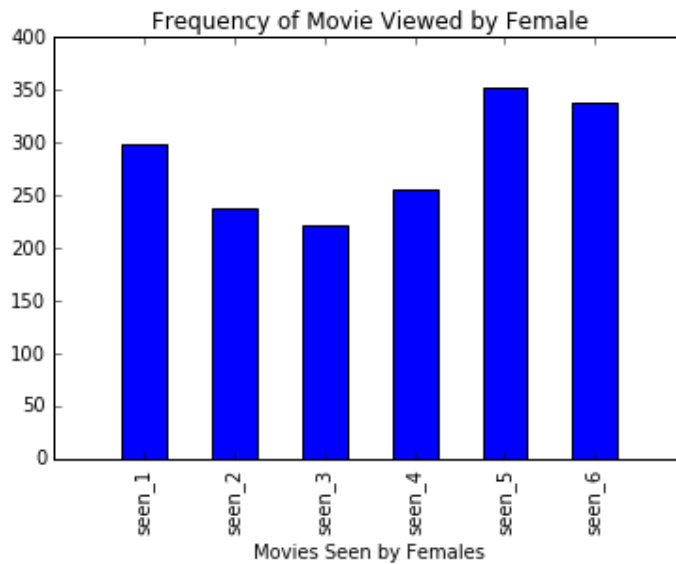


- **Movies Seen by Males**

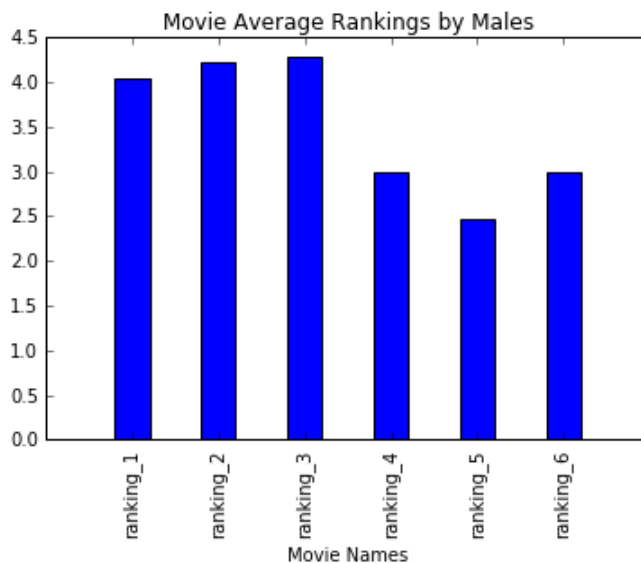
## Project Writeup - Star War Survey



- **Movies Seen by Females**

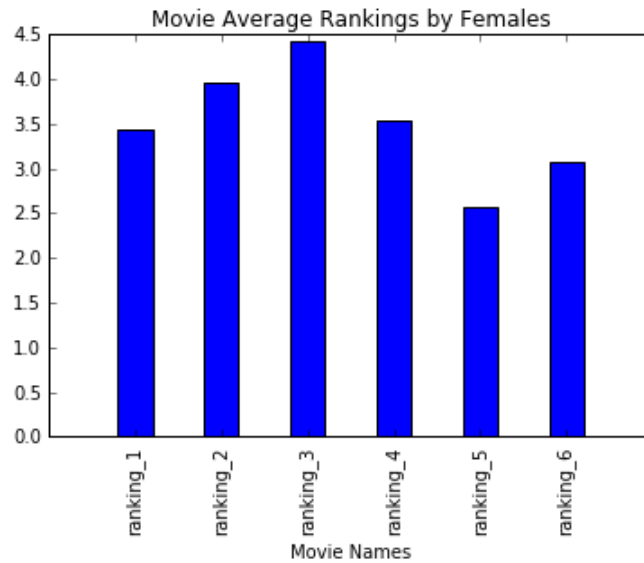


- **Average Movie Rankings by Males**

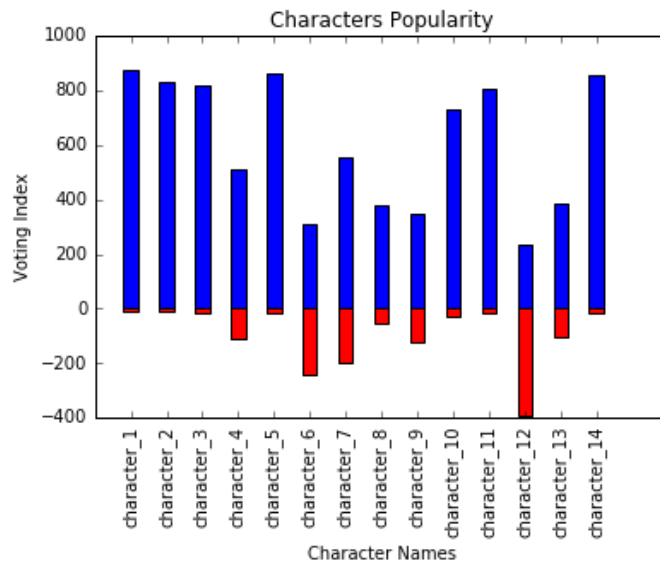


## Project Writeup - Star War Survey

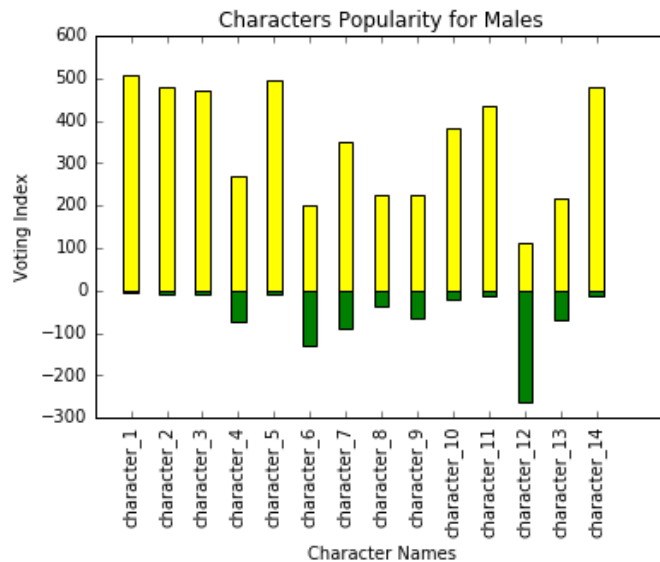
- **Average Movie Rankings by Females**



- **Respondents' Overall Character Preference Pattern**

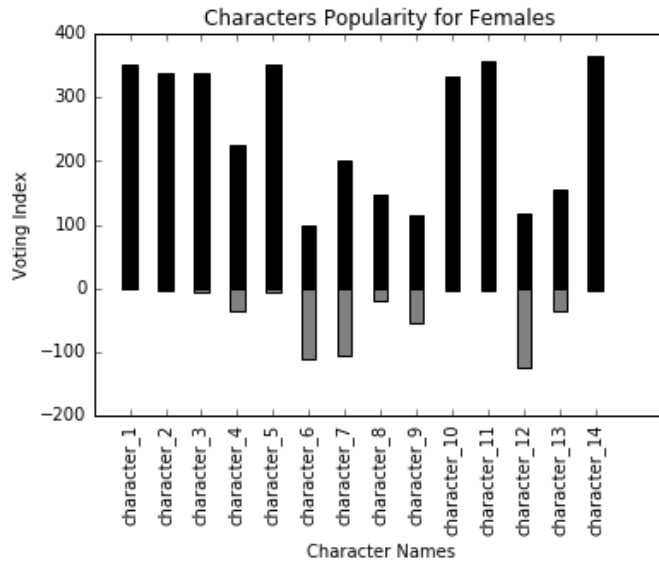


- **Male Respondents' Character Preference Pattern**



## Project Writeup - Star War Survey

- Female Respondents' Character Preference Pattern



Explain the results in a simple, concise and easy way. (non-technical audience)

- The results inform us about the like and dislike for 'Star Wars' movies as well as the same for various characters.
- We also learn about the gender specific movie watch frequency, their ranking patterns and particular scene popularity.
- We can also understand the popularity controversy for a specific movie character, if exists, based on the thin marginal spread between its "Like" and "Dislike" votes.

Explain the results in the most technical way. (technical, data-scientist audience)

- Earlier movies appear to be more popular since the respondents tend to see more original movies than the new movies.
- On an average more males watched the scenes 1 thru 3 than the female. However, females liked episodes\_3 more than the males.
- The respondents overall like character\_1 the most.
- The respondents overall dislike character\_12 the most.
- The average weightage of women's characters preference is around 72% of that of male's character preferences.
- Considering thin marginal spread between the likes and dislikes for character\_6; respondents' preference for this character appear to be controversial.

## Conclusion

What we learn from this outcome. (non-technical audience)

People more like the original/old sci-fi movies than the new releases of the 'Star Wars' series. Males get attracted more than the females for the fiction movies. Character\_1 is the overall favorite character while Character\_12 is the least like character of the movie series across the entire audience spectrum.

## Project Writeup - Star War Survey

Technical significance of the results. (technical, data-science audience)

- The results give insights to the new sci-fi movie maker about people's likes and dislikes about the fiction plots. People likes the original/old movie themes than the new ones.
- The preferential likes and dislikes for the given set of characteristics reveal the upcoming trend of viewers' character choice. Character\_1 is the most popular character while Character\_12 is the most disliked character of "Star Wars" movies.
- The frequency of movie watching for males and females gives ideas about the gender proportion of the audience as well as the gender specific preferences for the movies. The average movie watching frequency of males is more than the females.

## Suggestion for Further Development

How differently you would have done this project, if get a chance to do it again.

Why?

I could have involved other untouched data such as "Age", "Household Income", "Education", and "Location". It would be interesting to see how these factors impact the sci-fi movie audience and the character likes/dislikes pattern.

Your suggestions for someone if s/he wants to further continue this work.

Someone could pick one or more of the untouched data fields and continue this journey further to see the correlations between these factors and sci-fi movie watching pattern, likes/dislikes and the favorite characters.