

**MAY THE DATA BE WITH
YOU**



By: Neal Gragg

UNCC – ITCS-5156 Applied Machine Learning

April 28th, 2023



Hello There!

- ❖ This project is based on this article from 2014 with data from a survey that had 1,186 responses:
 - ❖ America's Favorite 'Star Wars' Movies (And Least Favorite Characters)
 - ❖ <https://fivethirtyeight.com/features/americas-favorite-star-wars-movies-and-least-favorite-characters/>
 - ❖ Author: Walt Hickey
- ❖ We will look at which films people have seen, what their favorite is, opinions on characters, and more!
- ❖ What can we discern from this data and how does it affect the Star Wars franchise and fan base?



At Last, We Will Reveal the Data to the Jedi



❖ Survey Questions:

- ❖ Have you seen any of the 6 films?
- ❖ Do you consider yourself a fan?
- ❖ Which films have you seen?
- ❖ Rank the films in order of preference: 1-6
- ❖ Favorability of various characters
- ❖ Who Shot First?
- ❖ Are you familiar with the Expanded Universe?
- ❖ Are you a fan of the EU?
- ❖ Are you a Star Trek fan?
- ❖ Gender, Age, Household Income, Education, Location (Region)



This is the Way...to Describe the Data

```
df = pd.read_csv("StarWars.csv")
display(df)

df_describe = df.describe()
df_describe
```



	RespondentID	Film Rank 1	Film Rank 2	Film Rank 3	Film Rank 4	Film Rank 5	Film Rank 6
count	1.186000e+03	835.000000	836.000000	835.000000	836.000000	836.000000	836.000000
mean	3.290128e+09	3.732934	4.087321	4.341317	3.272727	2.513158	3.047847
std	1.055639e+06	1.656122	1.365365	1.400464	1.825901	1.578620	1.666897
min	3.288373e+09	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
25%	3.289451e+09	3.000000	3.000000	3.000000	2.000000	1.000000	2.000000
50%	3.290147e+09	4.000000	4.000000	5.000000	3.000000	2.000000	3.000000
75%	3.290814e+09	5.000000	5.000000	6.000000	5.000000	3.000000	4.000000
max	3.292880e+09	6.000000	6.000000	6.000000	6.000000	6.000000	6.000000

RespondentID	Have you seen any of the 6 films in the Star Wars franchise?	Do you consider yourself to be a fan of the Star Wars film franchise?	Have you seen Episode 1?	Have you seen Episode 2?	Have you seen Episode 3?	Have you seen Episode 4?	Have you seen Episode 5?	Have you seen Episode 6?	Film Rank 1	Jar Jar Binks	Which character shot first?	Are you familiar with the Expanded Universe?	Do you consider yourself to be a fan of the Expanded Universe?
0	3292879998	Yes	Yes	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	Star Wars: Episode III Revenge of the Sith	Star Wars: Episode IV A New Hope	Star Wars: Episode V The Empire Strikes Back	Star Wars: Episode VI Return of the Jedi	3.0	Very favorably	I don't understand this question	Yes
1	3292879538	No	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	3292765271	Yes	No	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	Star Wars: Episode III Revenge of the Sith	NaN	NaN	NaN	1.0	Unfamiliar (N/A)	I don't understand this question	No
3	3292763116	Yes	Yes	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	Star Wars: Episode III Revenge of the Sith	Star Wars: Episode IV A New Hope	Star Wars: Episode V The Empire Strikes Back	Star Wars: Episode VI Return of the Jedi	5.0	Very favorably	I don't understand this question	No
4	3292731220	Yes	Yes	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	Star Wars: Episode III Revenge of the Sith	Star Wars: Episode IV A New Hope	Star Wars: Episode V The Empire Strikes Back	Star Wars: Episode VI Return of the Jedi	5.0	Somewhat favorably	Greedo	Yes
1181	3288388730	Yes	Yes	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	Star Wars: Episode III Revenge of the Sith	Star Wars: Episode IV A New Hope	Star Wars: Episode V The Empire Strikes Back	Star Wars: Episode VI Return of the Jedi	5.0	Very favorably	Han	No
1182	3288378779	Yes	Yes	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	Star Wars: Episode III Revenge of the Sith	Star Wars: Episode IV A New Hope	Star Wars: Episode V The Empire Strikes Back	Star Wars: Episode VI Return of the Jedi	4.0	Very favorably	I don't understand this question	No
1183	3288375286	No	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1184	3288373068	Yes	Yes	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	Star Wars: Episode III Revenge of the Sith	Star Wars: Episode IV A New Hope	Star Wars: Episode V The Empire Strikes Back	Star Wars: Episode VI Return of the Jedi	4.0	Very favorably	Han	No
1185	3288372923	Yes	No	Star Wars: Episode I The Phantom Menace	Star Wars: Episode II Attack of the Clones	NaN	NaN	NaN	Star Wars: Episode VI Return of the Jedi	6.0	Very unfavorably	I don't understand this question	No

1186 rows × 38 columns

Many Analysis Methods, With More on the Way

- ❖ Data Analysis Method to Clone from the Selected Paper:
 - ❖ Percentages of respondents who have seen each film
 - ❖ Of the respondents who have seen all six films, which film ranks the best
 - ❖ How often each film was rated in the top, middle, and bottom third
 - ❖ Character favorability ratings
 - ❖ “Who shot first?” responses
 - ❖ What can we do differently to explore this data?



Time to Compact the Data

❖ Data Preprocessing

- ❖ Convert Yes/No answers to 1/0
- ❖ Convert film titles/blanks under “Have you seen...?” to 1/0
- ❖ Drop responses that did not answer at least 14 questions
- ❖ This will compact the data to 1,058 responses that are now easier to process and analyze



RespondentID	Have you seen any of the 6 films in the Star Wars franchise?	Do you consider yourself to be a fan of the Star Wars film franchise?	Have you seen Episode 1?	Have you seen Episode 2?	Have you seen Episode 3?	Have you seen Episode 4?	Have you seen Episode 5?	Have you seen Episode 6?	Film Rank	...	Jar Jar Binks	Which character shot first?	Are you familiar with the Expanded Universe?	Do you consider yourself to be a fan of the Expanded Universe?	
0	3292879998	1	1.0	1	1	1	1	1	1	3.0	...	Very favorably	I don't understand this question	1.0	0.0
1	3292879538	0	NaN	0	0	0	0	0	0	NaN	...	NaN	NaN	NaN	NaN
2	3292765271	1	0.0	1	1	1	0	0	0	1.0	...	Unfamiliar (N/A)	I don't understand this question	0.0	NaN

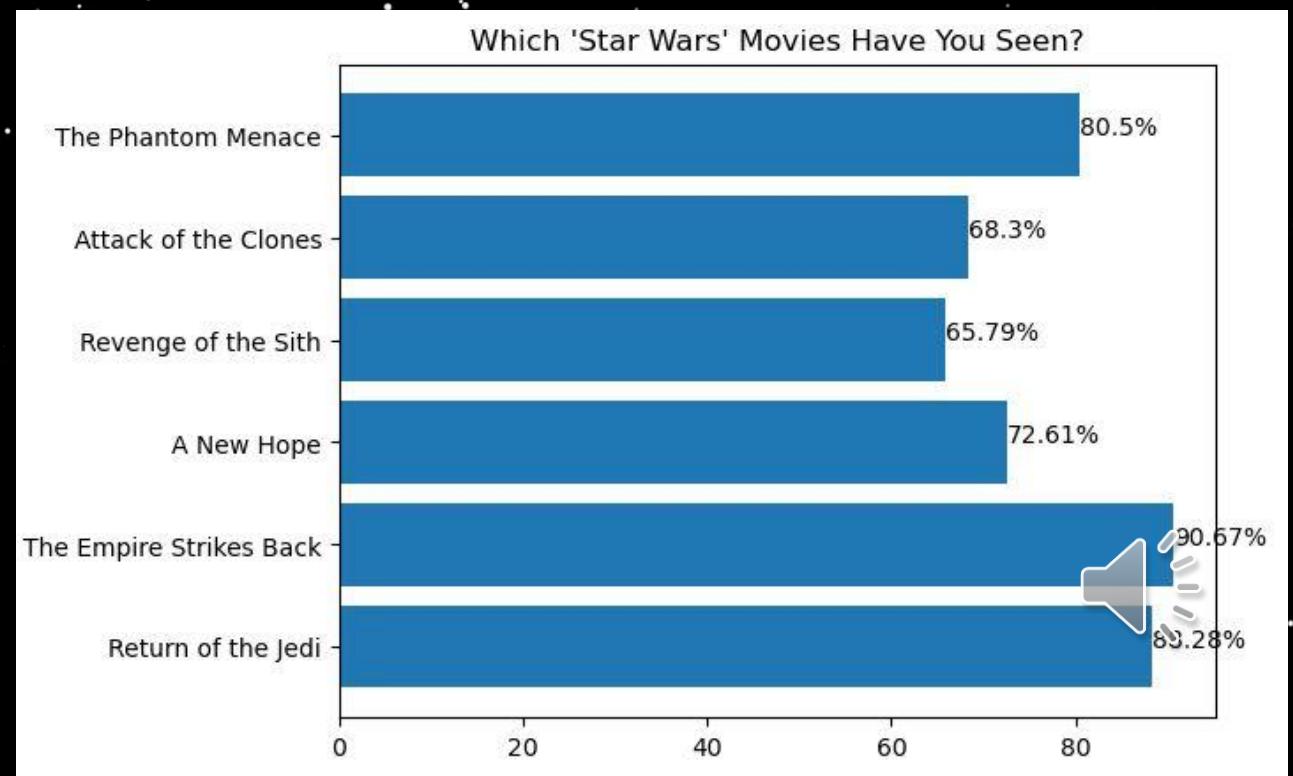


This is Where the Fun Begins!



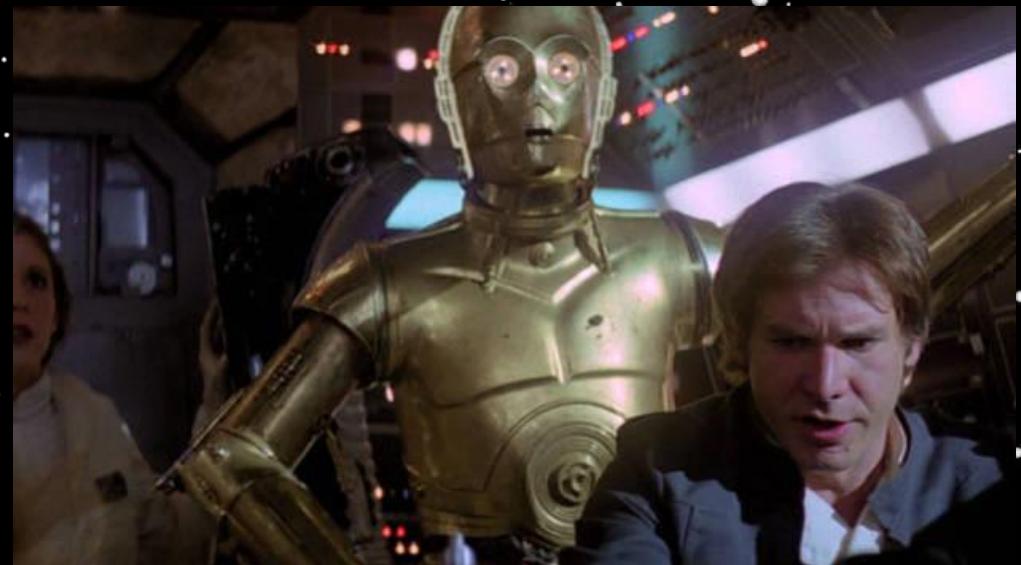
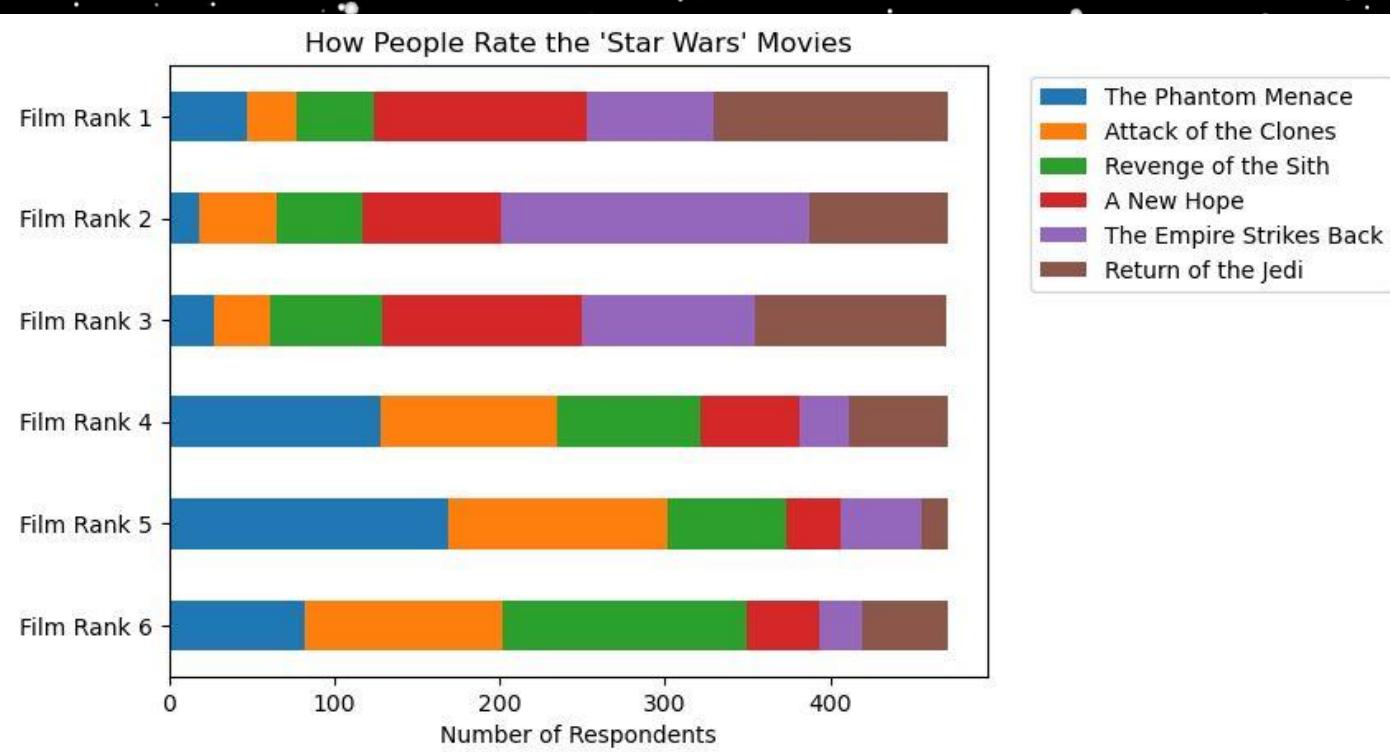
❖ Analysis Cloning – Step 1:

- ❖ Eliminate respondents who have not seen any Star Wars films (836 responses left)
- ❖ Calculate the percentage of respondents who have seen each film and plot results



Never Tell Me the Ranks!

- ❖ Analysis Cloning – Step 2:
 - ❖ Narrow down respondents to those who have seen all 6 films (471 responses left)
 - ❖ Calculate the ranks for each film and plot

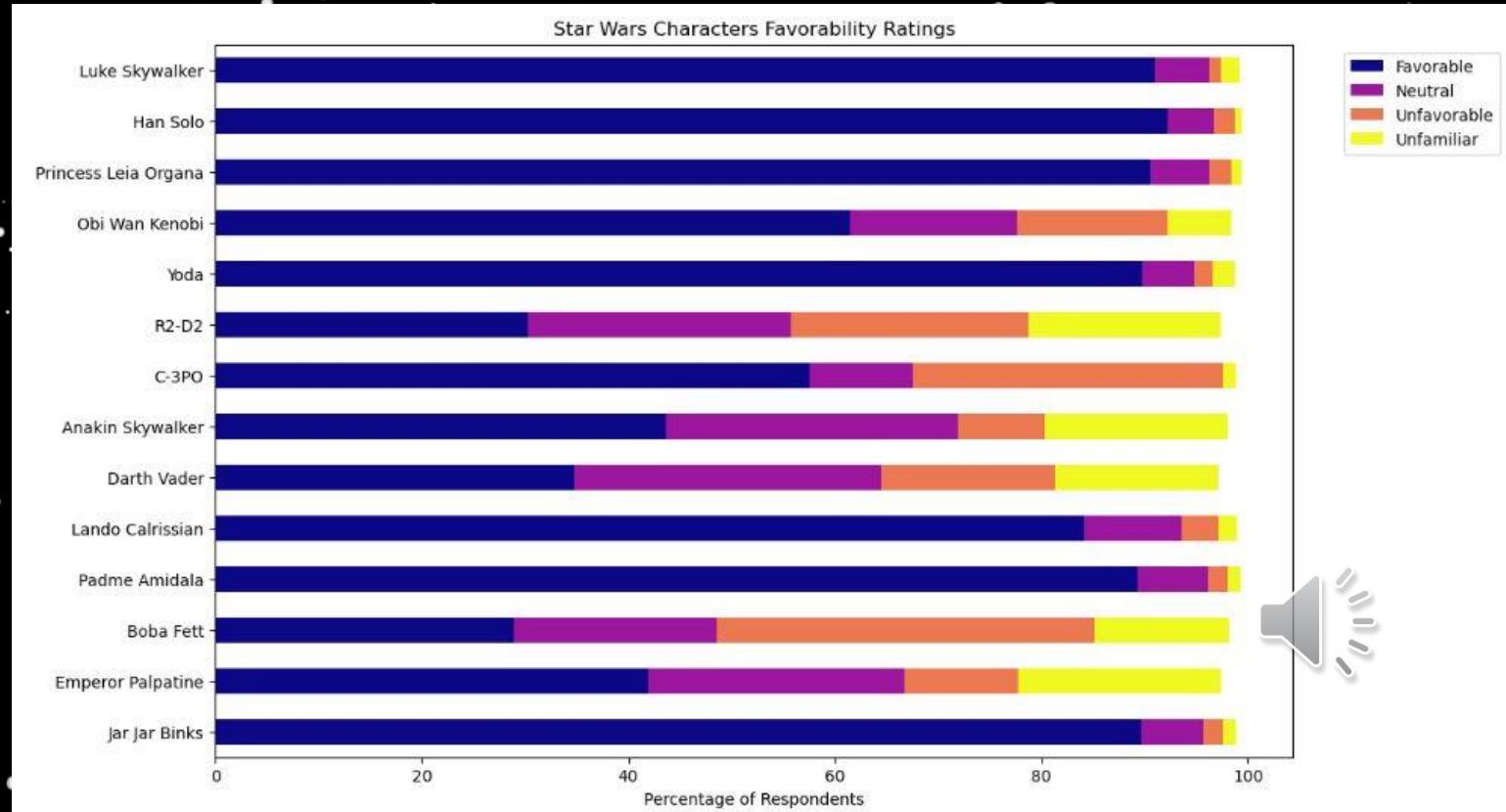


Meesa Big Popular!



- ❖ Analysis Cloning – Step 3:

- ❖ Determine character ratings as percents
- ❖ Create favorability categories
- ❖ Categorize and plot character favorability



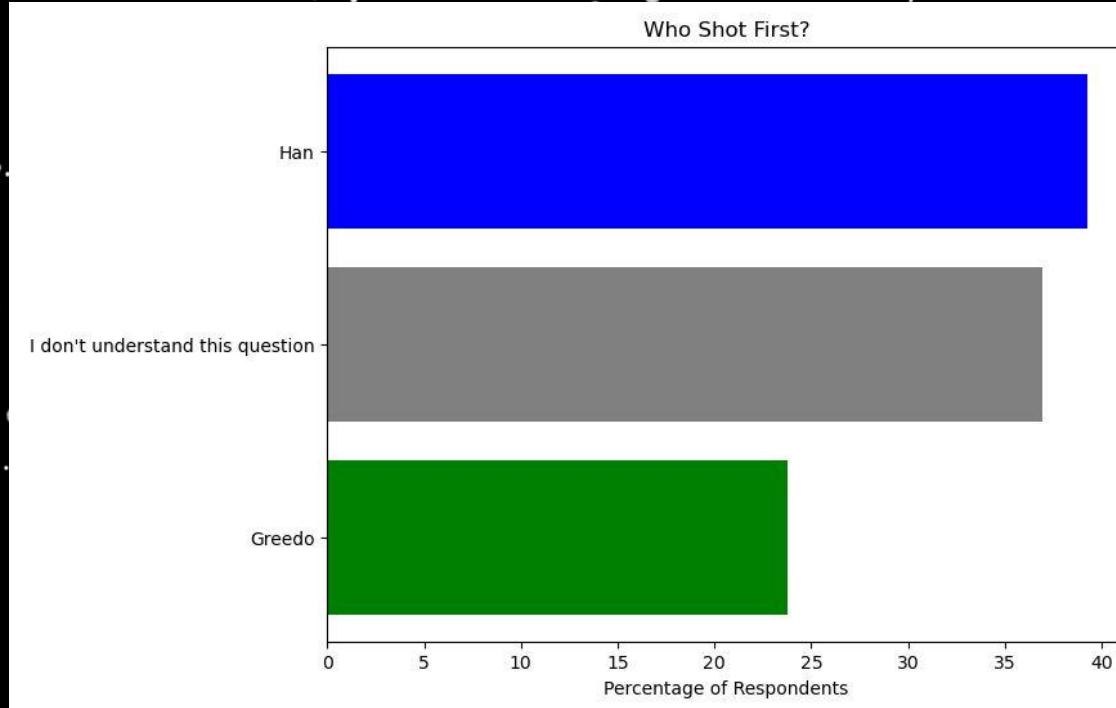
Tell Jabba I've Got His Data

Han



- ❖ Analysis Cloning – Step 4:
- ❖ Calculate “Who Shot First?” results and plot

Greedo



Incomplete, Your Data Analysis Is

- ❖ More Data to Explore:

- ❖ 66% are familiar with the EU
- ❖ Almost 50/50 split on fans of the EU
- ❖ 68% overlap between Star Wars fans and Star Trek fans

- ❖ Always in Motion, the Future Is:

- ❖ Future Work on demographic data compared to Star Wars data
- ❖ Predictions based on the data
- ❖ Focus on what we learn from the fans - original trilogy is still top ranked, Jar Jar is popular?, and Han shot first!
- ❖ How this data will affect marketing and developing of future Star Wars projects



**Written and Directed by
NEAL GRAGG**

