

Star War Survey Data Analysis

Wenkai LI



1. Data Preparation

- 1.1 Data Retrieving
- 1.2 Data Check
- 1.3 Check Missing Values
- 1.4 Find and Address the Issues
- 1.5 Data Check after Preparation

2. Data Exploration

- 2.1 How Do People Rank Episode I to VI
- 2.2 Hypothesis Relationships Between Columns
- 2.3 Relationships between Demography and Attitudes
 - 2.3.1 Gender and Attitudes
 - 2.3.2 Age and Attitudes



1. Data Preparation



1.1 Data Retrieving

```
In [3]: #Task 1.1: Data Retrieving. Load data from StarWars.csv.  
#Skip first two header rows. Set the first column index. Rename columns with succinct names.  
starWar = pd.read_csv('StarWars.csv', encoding='ISO-8859-1', skiprows = [0,1],  
index_col=0,header=None, sep = ',', names = ['Seen Star War', 'Star War Fan',  
'Seen Episode 1', 'Seen Episode 2', 'Seen Episode 3', 'Seen Episode 4', 'Seen Episode 5', 'Seen Episode 6',  
'Episode 1 Rank', 'Episode 2 Rank', 'Episode 3 Rank', 'Episode 4 Rank', 'Episode 5 Rank', 'Episode 6 Rank',  
'Attitude to Han Solo', 'Attitude to Luke Skywalker', 'Attitude to Princess Leia Organa', 'Attitude to Anakin Skywalker',  
'Attitude to Obi Wan Kenobi', 'Attitude to Emperor Palpatine', 'Attitude to Darth Vader', 'Attitude to Lando Calrissian',  
'Attitude to Boba Fett', 'Attitude to C-3P0', 'Attitude to R2 D2', 'Attitude to Jar Jar Binks', 'Attitude to Padme Amidala',  
'Attitude to Yoda', 'Character Shot 1st', 'Universe Familiar', 'Universe Fan', 'Star Trek Fan', 'Gender', 'Age', 'Household Income', 'Education', 'Location'])
```



1.2 Data Check

```
In [5]: #Task 1.2: Check data types for each column.  
starWar.dtypes
```

```
Out[5]: Seen Star War          object  
Star War Fan          object  
Seen Episode 1        object  
Seen Episode 2        object  
Seen Episode 3        object  
Seen Episode 4        object  
Seen Episode 5        object  
Seen Episode 6        object  
Episode 1 Rank        float64  
Episode 2 Rank        float64  
Episode 3 Rank        float64  
Episode 4 Rank        float64  
Episode 5 Rank        float64  
Episode 6 Rank        float64  
Attitude to Han Solo  object  
Attitude to Luke Skywalker  object  
Attitude to Princess Leia Organa  object  
Attitude to Anakin Skywalker  object  
Attitude to Obi Wan Kenobi  object  
Attitude to Emperor Palpatine  object  
Attitude to Darth Vader  object  
Attitude to Lando Calrissian  object  
Attitude to Boba Fett  object  
Attitude to C-3P0  object  
Attitude to R2 D2  object  
Attitude to Jar Jar Binks  object  
Attitude to Padme Amidala  object  
Attitude to Yoda  object  
Character Shot 1st  object  
Universe Familiar  object  
Universe Fan  object  
Star Trek Fan  object  
Gender  object  
Age  object  
Household Income  object  
Education  object  
Location  object  
dtype: object
```



1.2.1 Check Missing Values

Seen Star War: has no missing value.
Star War Fan: has 350 missing value. ←
Seen Episode 1: has 513 missing value.
Seen Episode 2: has 615 missing value.
Seen Episode 3: has 636 missing value.
Seen Episode 4: has 579 missing value.
Seen Episode 5: has 428 missing value.
Seen Episode 6: has 448 missing value.
Episode 1 Rank: has 351 missing value.
Episode 2 Rank: has 350 missing value.
Episode 3 Rank: has 351 missing value.
Episode 4 Rank: has 350 missing value.
Episode 5 Rank: has 350 missing value.
Episode 6 Rank: has 350 missing value.
Attitude to Han Solo: has 357 missing value.
Attitude to Luke Skywalker: has 355 missing value.
Attitude to Princess Leia Organa: has 355 missing value.
Attitude to Anakin Skywalker: has 363 missing value.
Attitude to Obi Wan Kenobi: has 361 missing value.
Attitude to Emperor Palpatine: has 372 missing value.
Attitude to Darth Vader: has 360 missing value.
Attitude to Lando Calrissian: has 366 missing value.
Attitude to Boba Fett: has 374 missing value.
Attitude to C-3P0: has 359 missing value.
Attitude to R2 D2: has 356 missing value.
Attitude to Jar Jar Binks: has 365 missing value.
Attitude to Padme Amidala: has 372 missing value.
Attitude to Yoda: has 360 missing value.
Character Shot 1st: has 358 missing value.
Universe Familiar: has 358 missing value.
Universe Fan: has 973 missing value.
Star Trek Fan: has 118 missing value.
Gender: has 140 missing value. ←
Age: has 140 missing value.
Household Income: has 328 missing value.
Education: has 150 missing value.
Location: has 143 missing value.



1.3.2 Check Missing Values

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI		
3.29E+09	Yes	Yes	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	5	6	3	2	1	4	Very favo	Somewh	Somewh	Neither f	Very favo	Somewh	Very favo	Neither f	Neither f	Somewh	Somewh	Very favo	Somewh	Very favo	Han	Yes	Yes	Yes	Male	45-60	\$	
3.29E+09	Yes	Yes	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	5	2	3	6	1	4	Very favo	Very favo	Somewh	Somewh	Very favo	Somewh	Neither f	Unfamili	Neither f	Somewh	Somewh	Somewh	Very favo	Han	Yes	No	Yes	Male	> 60	\$		
3.29E+09	Yes	Yes	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	6	5	4	1	2	3	Somewh	Somewh	Somewh	Neither f	Somewh	Neither f	Neither f	Neither f	Somewh	Somewh	Neither f	Somewh	Somewh	Han	No		Yes	Male	30-44	\$		
3.29E+09	Yes	Yes	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	6	5	4	3	2	1	Very favo	Very favo	Very favo	Somewh	Very favo	Somewh	Very favo	Somewh	Very favo	Somewh	Somewh	Very unf	Neither f	Somewh	Han	Yes	Yes	Yes	Male	30-44	\$	
3.29E+09	Yes	Yes	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	6	5	4	1	2	3	Very favo	Very favo	Very favo	Very favo	Very favo	Somewh	Somewh	Somewh	Neither f	Very favo	Very favo	Very unf	Somewh	Very favo	Greedo	No		Yes	Male	> 60	\$	
3.29E+09	Yes	Yes	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	6	5	4	3	2	1	Very favo	Very favo	Very favo	Very favo	Very favo	Somewh	Very favo	Somewh	Very favo	Very favo	Very unf	Somewh	Very favo	Greedo	No		No	Male	30-44	\$		
3.29E+09	Yes	Yes							Star Wars	Star Wars	3	4	5	6	1	2	Very favo	Very favo	Very favo	Very favo	Very favo	Neither f	Somewh	Somewh	Very favo	Very favo	Very favo	Somewh	Neither f	Neither f	Greedo	No		No	Male	> 60	\$
3.29E+09	Yes	Yes	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	4	5	6	1	3	2	Very favo	Very favo	Very favo	Somewh	Very favo	Neither f	Very favo	Neither f	Neither f	Very favo	Very favo	Somewh	Neither f	Very favo	Han	No		Yes	Male	18-29	\$	
3.29E+09	Yes	Yes	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	4	3	5	6	1	2	Neither f	Somewh	Somewh	Neither f	Somewh	Somewh	Somewh	Neither f	Neither f	Somewh	Somewh	Somewh	Somewh	Somewh	I don't ur	No		No	Male	> 60	\$	
3.29E+09	Yes	Yes	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	6	5	4	1	3	2	Somewh	Very favo	Very favo	Somewh	Somewh	Very unf	Somewh	Neither f	Neither f	Somewh	Very favo	Very unf	Somewh	Very favo	Han	No		Yes	Male	18-29	\$	
3.29E+09	Yes	Yes	Star Wars: Episode I	The Pha	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	6	5	4	2	1	3	Very favo	Very favo	Very favo	Neither f	Very favo	Neither f	Somewh	Somewh	Very favo	Very favo	Very unf	Somewh	Somewh	Han	No		No	Male	30-44	\$		
3.29E+09	Yes	Yes	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	5	3	6	1	2	4	Very favo	Very favo	Very favo	Somewh	Very favo	Neither f	Somewh	Somewh	Very favo	Very favo	Very unf	Somewh	Neither f	Very favo	Greedo	No		Yes	Male	45-60	\$	
3.29E+09	Yes	Yes	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	6	5	3	4	1	2	Somewh	Somewh	Somewh	Somewh	Very favo	Somewh	Somewh	Somewh	Somewh	Somewh	Somewh	Somewh	Somewh	Somewh	Han	No		No	Male	18-29	\$	
3.29E+09	Yes	Yes	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	1	5	6	2	4	3	Somewh	Very favo	Very favo	Somewh	Somewh	Very unf	Somewh	Neither f	Neither f	Somewh	Neither f	Very favo	Very favo	Very favo	Han	No		Yes	Male	45-60	\$	
3.29E+09	Yes	Yes	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	3	4	5	6	1	2	Very favo	Very favo	Very favo	Very favo	Very favo	Somewh	Somewh	Neither f	Neither f	Very favo	Very favo	Very unf	Unfamili	Very favo	Greedo	Yes	Yes	Yes	Male	45-60	\$	
3.29E+09	Yes	Yes	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	5	6	4	1	2	3	Very favo	Very favo	Very favo	Neither f	Very favo	Somewh	Somewh	Neither f	Neither f	Somewh	Somewh	Somewh	Very favo	Very favo	I don't ur	No		Yes	Male	30-44	\$	
3.29E+09	Yes	Yes	Star Wars	Star Wars	Star Wars	Star Wars: Episode	Star Wars	Star Wars	Star Wars	4	5	3	6	1	2	Very favo	Very favo	Very favo	Neither f	Very favo	Somewh	Very unf	Neither f	Neither f	Somewh	Somewh	Neither f	Very favo	Very favo	I don't ur	Yes	No	Yes	Male	> 60	\$	
3.29E+09	Yes	Yes			Star Wars	Star Wars: Episode	Star Wars	Star Wars	Star Wars	5	3	4	6	1	2	Very favo	Very favo	Very favo	Neither f	Very favo	Somewh	Very favorabl	Unfamili	Very favo	Very favo	Very unf	Very favo	Very favo	Han	No		Yes	Male	> 60	\$		
3.29E+09	Yes	Yes	Star Wars: Episode I	The Pha	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	4	5	6	3	1	2	Very favo	Very favo	Very favo	Neither f	Very favo	Somewh	Very favo	Very favo	Somewh	Very favo	Very unf	Somewh	Very favo	Han	No		Yes	Male	45-60	\$		
3.29E+09	Yes	Yess	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	Star Wars	6	5	4	2	1	3	Very favo	Very favo	Very favo	Somewh	Very favo	Somewh	Very favo	Somewh	Somewh	Very favo	Very unf	Very favo	Very favo	Han	Yes	Yes	Yes	Male	18-29	\$		
3.29E+09	No																																	Yes	Male	18-29	\$
3.29E+09	No																																	Yes	Male	30-44	\$
3.29E+09	No																																	No	Male	18-29	\$
3.29E+09	No																																	Yes	Male	30-44	\$
3.29E+09	No																																	No	Male	18-29	\$
3.29E+09	No																																	No	Male	> 60	\$
3.29E+09	No																																	No	Male	> 60	\$
3.29E+09	No																																	No	Male	18-29	\$
3.29E+09	No																																	No	Male	30-44	\$
3.29E+09	No																																	No	Male	30-44	\$
3.29E+09	No																																	No	Male	> 60	\$
3.29E+09	No																																	No	Male	18-29	\$
3.29E+09	No																																	No	Male	30-44	\$
3.29E+09	No																																	No	Male	30-44	\$
3.29E+09	No																																	No	Male	> 60	\$
3.29E+09	No																																	No	Male	30-44	\$
3.29E+09	No																																	No	Male	45-60	\$
3.29E+09	No																																	No	Male	30-44	\$
3.29E+09	No																																	No	Male	30-44	\$

StarWars



1.3.2 Check Missing Values

	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM
1039																						No	Male	> 60	\$100,000 -	Graduate c	West South Centra	
1040																						No	Male	> 60	\$50,000 - \$	Some colle	Mountain	
1041																						No	Male	45-60		Some colle	South Atlantic	
1042																						No	Male	45-60	\$150,000+	Bachelor d	West South Centra	
1043																						No	Male	> 60	\$150,000+	Graduate c	New England	
1044																						No	Male	45-60	\$50,000 - \$	Some colle	Pacific	
1045																						No	Male	> 60	\$0 - \$24,99	Some colle	Pacific	
1046																						No	Male	45-60		Bachelor d	New England	
1047																						No	Male	30-44	\$50,000 - \$	Bachelor d	West South Centra	
1048																						No	Male	45-60	\$100,000 -	Graduate c	Middle Atlantic	
1049	3	4	5	6	Neither fa	Neither fa	Neither fa	Neither fa	Neither fa	Neither fa	Neither fa	Neither fa	Neither fa	Neither fa	Neither fa	Neither fa	Neither fa	I don't und	No			Yes						
1050	5	6	1	2	Very favor	Very favor	Very favor	Very favor	Very favor	Neither fa	Very favor	Unfamiliar	Unfamiliar	Very favor	Very favor	Very favor	Very favor	Somewhat	Neither fa	Somewhat	favorably							
1051	3	4	5	6	Somewhat	Somewhat	Somewhat	Somewhat	Somewhat	Somewhat	Neither fa	Neither fa	Somewhat	Somewhat	Somewhat	Somewhat	Neither fa	Somewhat	Greedo	No		No						
1052	3	4	5	6	Somewhat	Somewhat	Somewhat	Neither fa	Somewhat	Neither fa	Somewhat	Neither fa	Neither fa	Somewhat	Somewhat	Somewhat	Neither fa	Neither fa	Greedo	No		No						
1053	5	6	1	2	Somewhat	Somewhat	Somewhat	Somewhat	Somewhat	Somewhat	Somewhat	Neither fa	Neither fa	Somewhat	Somewhat	Somewhat	Neither fa	Neither fa	Greedo	No		No						
1054	6	3	2	1	Very favor	Somewhat	Somewhat	Somewhat	Very favor	Unfamiliar	Unfamiliar	Very favor	Very favor	Very favor	Very favor	Very favor	Somewhat	Very favor	Greedo	No		No						
1055	6	1	2	3	Very favor	Very favor	Very favor	Somewhat	Neither fa	Neither fa	Neither fa	Somewhat	unfavorab	Somewhat	Somewhat	Somewhat	Somewhat	Very favorably										
1056	6	3	1	2	Very favor	Very favor	Very favor	Somewhat	Very favor	Unfamiliar	Very unfav	Very favor	Neither fa	Very favor	Very favor	Very favor	Very favor	Very favorably										
1057	6	5	1	2	Very favor	Very favor	Very favor	Very favor	Very favor	Very favor	Very favor	Very favor	Very favor	Very favor	Very favor	Very favor	Very favor	Very favorably										
1058	4	2	1	3	Very favor	Somewhat	Very favor	Somewhat	Very favor	Somewhat	Somewhat	Neither fa	Neither fa	Very favor	Somewhat	Unfamiliar	Very unfav	Neither favorably nor unfavorably (neutral)										
1059	3	4	5	6	Somewhat	Somewhat	Somewhat	Somewhat	Somewhat	Somewhat	Somewhat	Somewhat	Somewhat	Somewhat	Somewhat	Somewhat	Somewhat	Somewhat	Han	No		Yes						
1060	2	5	1	4	Very favor	Very favor	Very favor	Very favor	Very favor	Somewhat	Very favor	Somewhat	Somewhat	Very favor	Very favor	Somewhat	Very favor	Very favorably										
1061	5	6	2	3	Very favor	Very favor	Very favor	Very favor	Very favor	Somewhat	Somewhat	Neither fa	Neither fa	Very favor	Very favor	Neither fa	Very favor	Very favor Han	Yes	Yes	Yes							
1062	2	4	5	6	Somewhat	Somewhat	Somewhat	Very favor	Somewhat	Very favor	Very favor	Neither fa	Very favor	Very favor	Somewhat	Neither fa	Very favor	Very favorably										
1063	4	1	2	3	Very favor	Very favor	Very favor	Very favor	Very favor	Very favor	Very favor	Very favor	Somewhat	Very favor	Very favor	Somewhat	Very favor	Very favorably										
1064	4	3	2	1	Very favor	Somewhat	Very favor	Very favor	Very favor	Very unfav	Somewhat	Somewhat	Very favor	Somewhat	Very unfav	Somewhat	Very favor Han	No	Yes									
1065																												
1066																												
1067																												
1068																												

StarWars



1.2.2 Find and Address the Issues

```
Yes      935
No       250
Yes       1
Name: Seen Star War, dtype: int64

Yes      551
No       283
Noo       1
Yess      1
Name: Star War Fan, dtype: int64

Star Wars: Episode I The Phantom Menace    673
Name: Seen Episode 1, dtype: int64

Star Wars: Episode II Attack of the Clones    571
Name: Seen Episode 2, dtype: int64

Star Wars: Episode III Revenge of the Sith    550
Name: Seen Episode 3, dtype: int64

Star Wars: Episode IV A New Hope    607
Name: Seen Episode 4, dtype: int64

Star Wars: Episode V The Empire Strikes Back    758
Name: Seen Episode 5, dtype: int64

Star Wars: Episode VI Return of the Jedi    738
Name: Seen Episode 6, dtype: int64

4.0      237
6.0      168
3.0      130
1.0      129
5.0      100
2.0       71
Name: Episode 1 Rank, dtype: int64
```

```
Very favorably      610
Somewhat favorably  151
Neither favorably nor unfavorably (neutral)    44
Unfamiliar (N/A)    15
Somewhat unfavorably    8
Very unfavorably     1
Name: Attitude to Han Solo, dtype: int64

Very favorably      552
Somewhat favorably  219
Neither favorably nor unfavorably (neutral)    38
Somewhat unfavorably    13
Unfamiliar (N/A)     6
Very unfavorably     3
Name: Attitude to Luke Skywalker, dtype: int64

Very favorably      547
Somewhat favorably  210
Neither favorably nor unfavorably (neutral)    48
Somewhat unfavorably    12
Unfamiliar (N/A)     8
Very unfavorably     6
Name: Attitude to Princess Leia Organa, dtype: int64

Somewhat favorably    269
Very favorably        245
Neither favorably nor unfavorably (neutral)   135
Somewhat unfavorably    83
Unfamiliar (N/A)     52
Very unfavorably     39
Name: Attitude to Anakin Skywalker, dtype: int64
```

```
No      615
Yes     213
Name: Universe Familiar, dtype: int64

No      114
Yes     98
Yess     1
Name: Universe Fan, dtype: int64

No      639
Yes     426
yes      1
Noo      1
no       1
Name: Star Trek Fan, dtype: int64

Female    546
Male      496
female     2
F          1
male       1
Name: Gender, dtype: int64

45-60    291
> 60     269
30-44    268
18-29    217
500       1
Name: Age, dtype: int64
```



1.2.3 Data Check after Preparation

```
YES      820
Name: Seen Star War, dtype: int64

YES      541
NO       279
Name: Star War Fan, dtype: int64

YES      659
NO       161
Name: Seen Episode 1, dtype: int64

YES      560
NO       260
Name: Seen Episode 2, dtype: int64

YES      539
NO       281
Name: Seen Episode 3, dtype: int64

YES      597
NO       223
Name: Seen Episode 4, dtype: int64

YES      745
NO       75
Name: Seen Episode 5, dtype: int64

YES      725
NO       95
Name: Seen Episode 6, dtype: int64

4        233
6        165
3        128
1        123
5         99
2         71
0          1
Name: Episode 1 Rank, dtype: int64

5        295
4        180
2        112
6        101
3        100
1         32
Name: Episode 2 Rank, dtype: int64

6        213
5        200
4        179
3        146
2         45
1          3
0           1
Name: Episode 3 Rank, dtype: int64
```

```
VERY FAVORABLY      554
SOMEWHAT FAVORABLY  179
NEUTRAL              56
UNFAMILIAR (N/A)    10
SOMEWHAT UNFAVORABLY 9
VERY UNFAVORABLY    6
NO ANSWER            6
Name: Attitude to R2 D2, dtype: int64

VERY UNFAVORABLY    203
NEUTRAL              161
SOMEWHAT FAVORABLY  126
VERY FAVORABLY      109
UNFAMILIAR (N/A)    108
SOMEWHAT UNFAVORABLY 98
NO ANSWER            15
Name: Attitude to Jar Jar Binks, dtype: int64

NEUTRAL              202
SOMEWHAT FAVORABLY  179
UNFAMILIAR (N/A)    164
VERY FAVORABLY      163
SOMEWHAT UNFAVORABLY 57
VERY UNFAVORABLY    33
NO ANSWER            22
Name: Attitude to Padme Amidala, dtype: int64

VERY FAVORABLY      596
SOMEWHAT FAVORABLY  141
NEUTRAL              47
UNFAMILIAR (N/A)    10
NO ANSWER            10
SOMEWHAT UNFAVORABLY 8
VERY UNFAVORABLY    8
Name: Attitude to Yoda, dtype: int64

HAN              322
NOT UNDERSTAND  305
GREEDO           193
Name: Character Shot 1st, dtype: int64

NO        608
YES       212
Name: Universe Familiar, dtype: int64

NO ANSWER  608
NO         114
YES         98
Name: Universe Fan, dtype: int64

YES       411
NO        409
Name: Star Trek Fan, dtype: int64

MALE      423
FEMALE    397
Name: Gender, dtype: int64
```

```
45-60      241
30-44      207
> 60       193
18-29      179
Name: Age, dtype: int64

\ $50,000 - \ $99,999      238
\ $25,000 - \ $49,999     147
NO ANSWER                  145
\ $100,000 - \ $149,999   115
\ $0 - \ $24,999          98
\ $150,000+                77
Name: Household Income, dtype: int64

BACHELOR DEGREE      262
SOME COLLEGE OR ASSOCIATE DEGREE 254
GRADUATE DEGREE      226
HIGH SCHOOL DEGREE   71
NO ANSWER             4
LESS THAN HIGH SCHOOL DEGREE 3
Name: Education, dtype: int64

PACIFIC      146
EAST NORTH CENTRAL 134
SOUTH ATLANTIC 133
MIDDLE ATLANTIC  93
WEST SOUTH CENTRAL 76
WEST NORTH CENTRAL 76
MOUNTAIN        68
NEW ENGLAND      60
EAST SOUTH CENTRAL 32
NO ANSWER        2
Name: Location, dtype: int64
```



1.2 Data Check after Preparation

```
In [58]: #After data preparation, check Na again to ensure all missing values are replaced.  
checkNa(starWar)
```

```
Seen Star War: has no missing value.  
Star War Fan: has no missing value.  
Seen Episode 1: has no missing value.  
Seen Episode 2: has no missing value.  
Seen Episode 3: has no missing value.  
Seen Episode 4: has no missing value.  
Seen Episode 5: has no missing value.  
Seen Episode 6: has no missing value.  
Episode 1 Rank: has no missing value.  
Episode 2 Rank: has no missing value.  
Episode 3 Rank: has no missing value.  
Episode 4 Rank: has no missing value.  
Episode 5 Rank: has no missing value.  
Episode 6 Rank: has no missing value.  
Attitude to Han Solo: has no missing value.  
Attitude to Luke Skywalker: has no missing value.  
Attitude to Princess Leia Organa: has no missing value.  
Attitude to Anakin Skywalker: has no missing value.  
Attitude to Obi Wan Kenobi: has no missing value.  
Attitude to Emperor Palpatine: has no missing value.  
Attitude to Darth Vader: has no missing value.  
Attitude to Lando Calrissian: has no missing value.  
Attitude to Boba Fett: has no missing value.  
Attitude to C-3P0: has no missing value.  
Attitude to R2 D2: has no missing value.  
Attitude to Jar Jar Binks: has no missing value.  
Attitude to Padme Amidala: has no missing value.  
Attitude to Yoda: has no missing value.  
Character Shot 1st: has no missing value.  
Universe Familiar: has no missing value.  
Universe Fan: has no missing value.  
Star Trek Fan: has no missing value.  
Gender: has no missing value.  
Age: has no missing value.  
Household Income: has no missing value.  
Education: has no missing value.  
Location: has no missing value.
```



1.2 Data Check after Preparation

```
In [59]: #After data preparation, check data types again for each column.  
starWar.dtypes
```

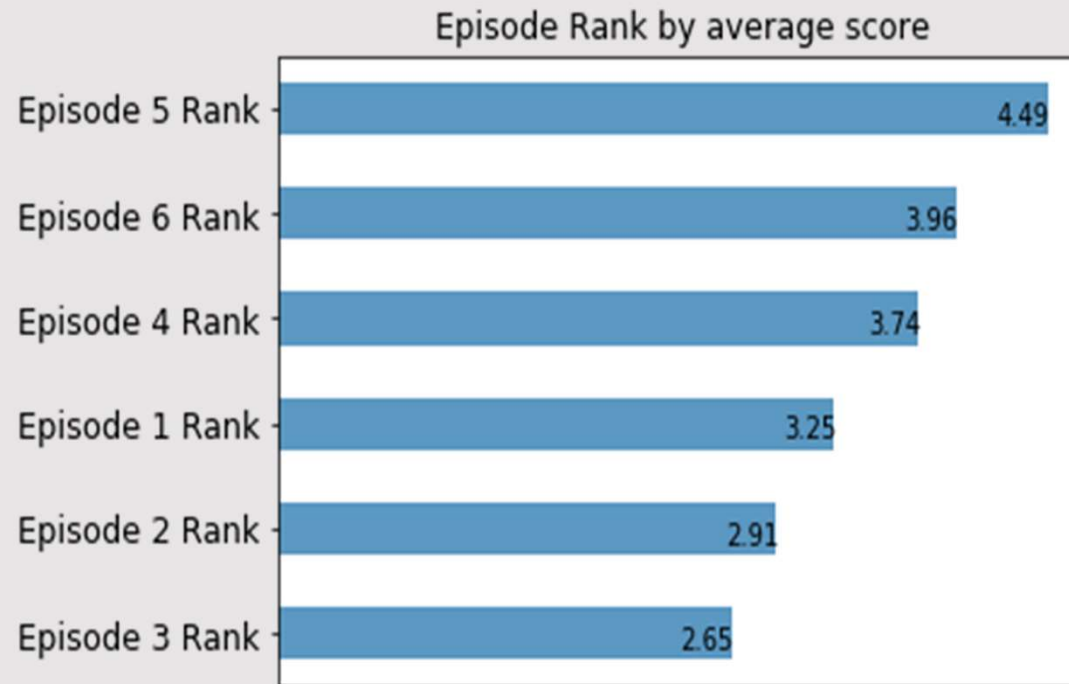
```
Out[59]: Seen Star War          category  
Star War Fan          category  
Seen Episode 1        category  
Seen Episode 2        category  
Seen Episode 3        category  
Seen Episode 4        category  
Seen Episode 5        category  
Seen Episode 6        category  
Episode 1 Rank        category  
Episode 2 Rank        category  
Episode 3 Rank        category  
Episode 4 Rank        category  
Episode 5 Rank        category  
Episode 6 Rank        category  
Attitude to Han Solo  category  
Attitude to Luke Skywalker category  
Attitude to Princess Leia Organa category  
Attitude to Anakin Skywalker category  
Attitude to Obi Wan Kenobi category  
Attitude to Emperor Palpatine category  
Attitude to Darth Vader category  
Attitude to Lando Calrissian category  
Attitude to Boba Fett category  
Attitude to C-3P0     category  
Attitude to R2 D2     category  
Attitude to Jar Jar Binks category  
Attitude to Padme Amidala category  
Attitude to Yoda      category  
Character Shot 1st     category  
Universe Familiar     category  
Universe Fan          category  
Star Trek Fan         category  
Gender                category  
Age                   category  
Household Income      category  
Education              category  
Location              category  
dtype: object
```



2. Data Exploration



2.1 How Do People Rank Episode 1 to 6

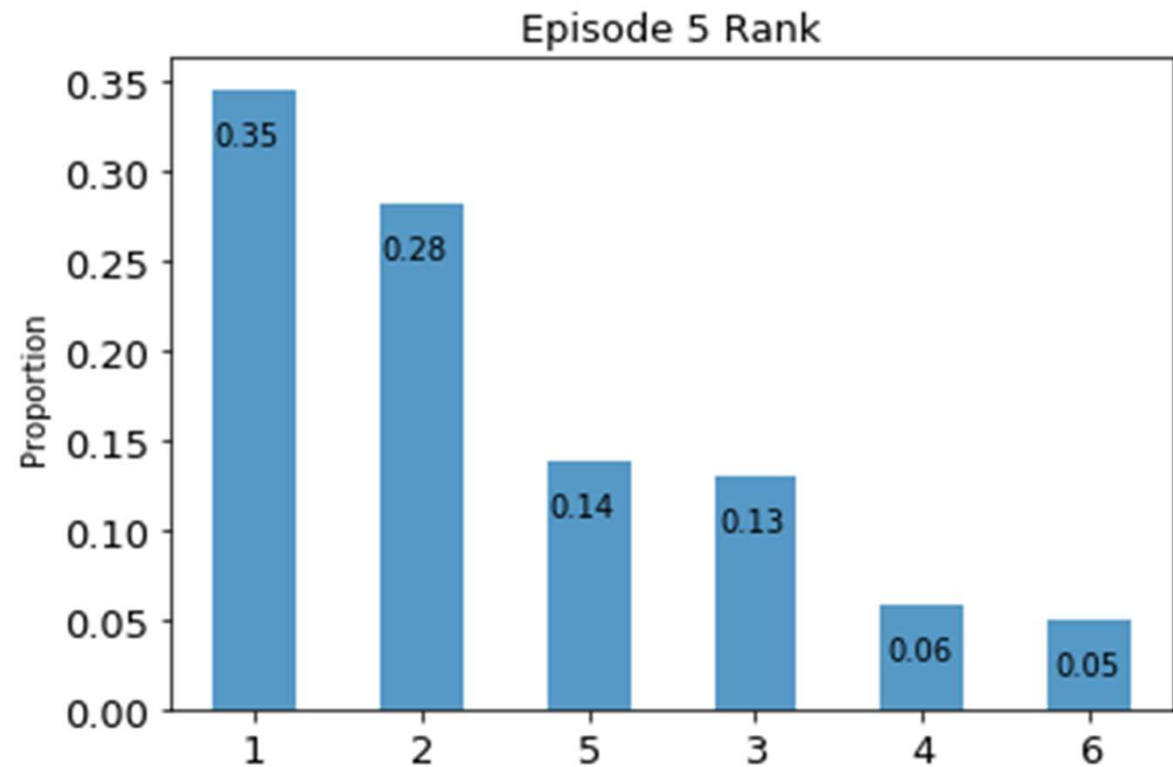


$$\text{avgScore} = \sum_{i=1}^6 (\text{Percentage of Rank } i * \text{Weights})$$



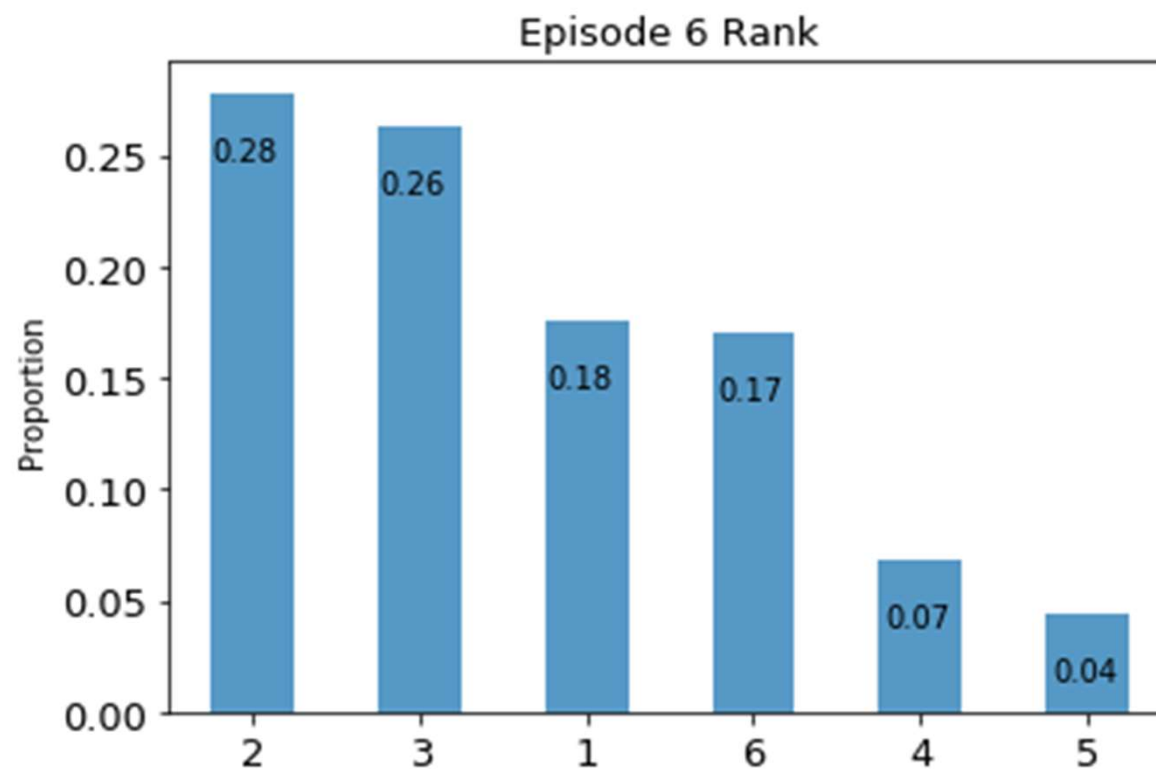


Picture downloaded from <https://www.imdb.com/>





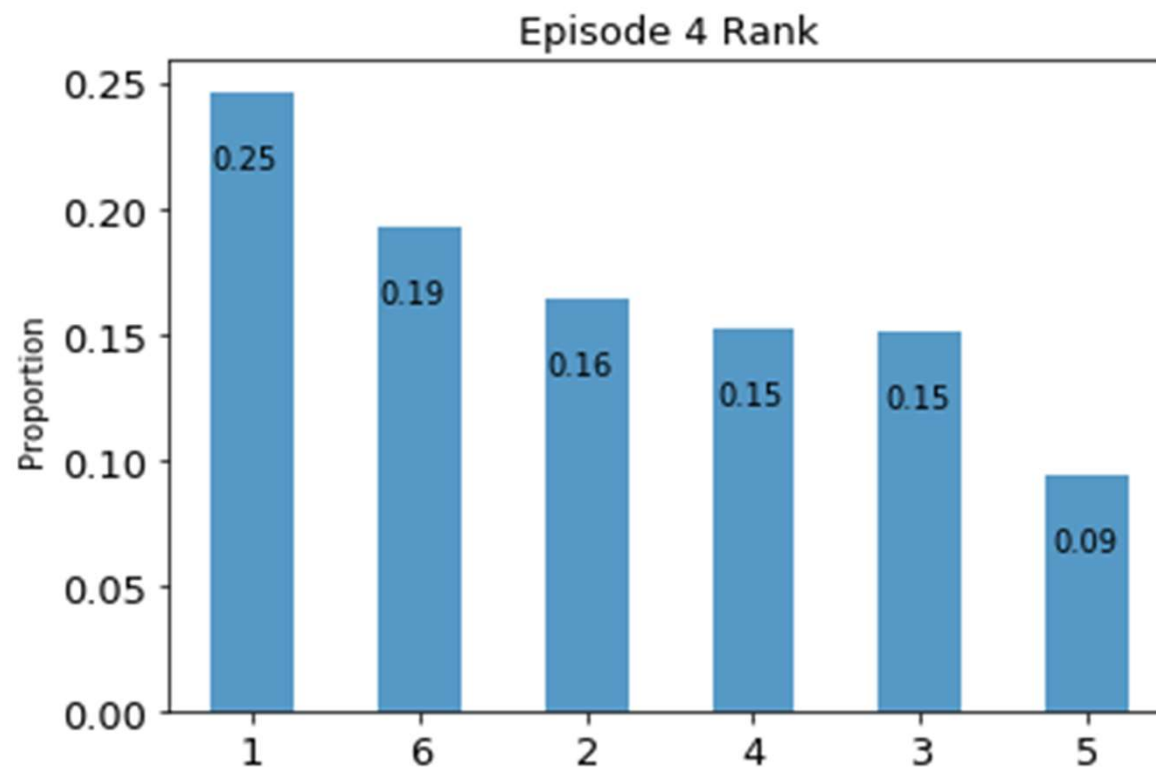
Picture downloaded from <https://www.imdb.com/>





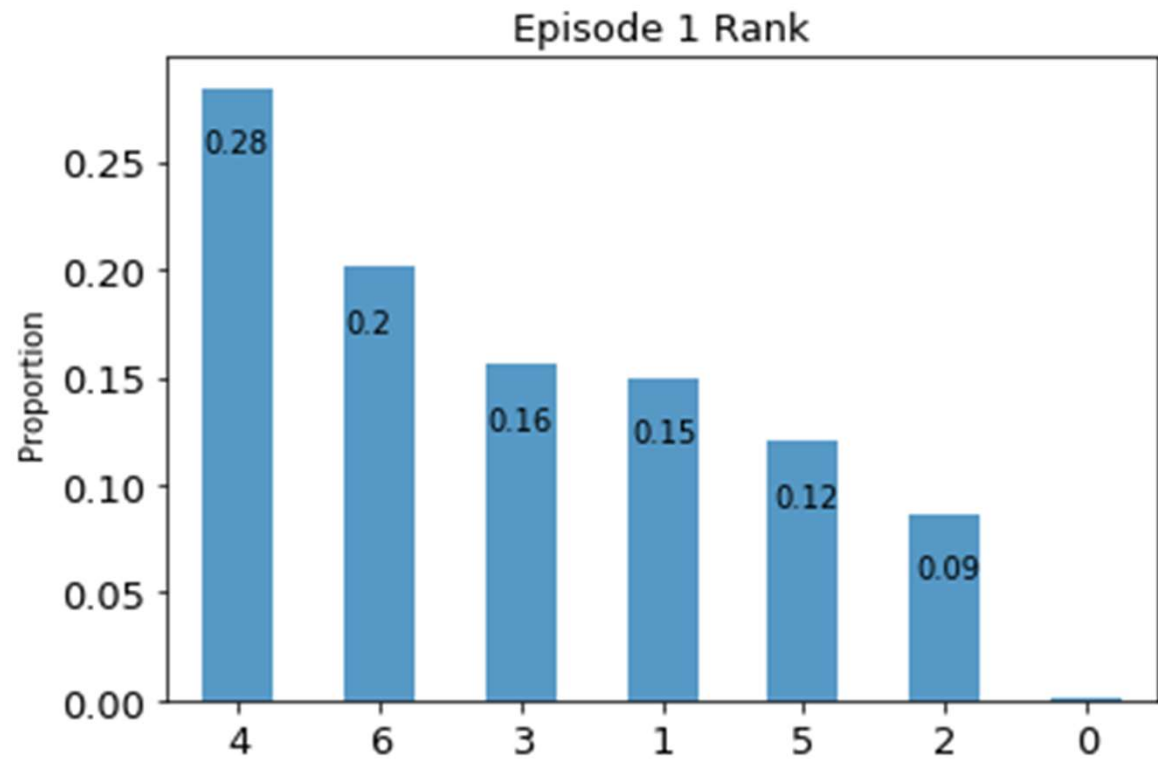
TWENTIETH CENTURY FOX Presents A LUCASFILM LTD PRODUCTION **STAR WARS**
 Starring **MARK HAMILL HARRISON FORD CARRIE FISHER**
PETER CUSHING
 and **ALEC GUINNESS**
 Written and Directed by **GEORGE LUCAS** Produced by **GARY KURTZ** Music by **JOHN WILLIAMS**
PAVAVISION® PRINTS BY DE LUXE® TECHNICOLOR®
 Making Films Sound Better **DOLBY SYSTEM®** Noise Reduction - High Fidelity
 Original Motion Picture Soundtrack on 20th Century Records and Labels

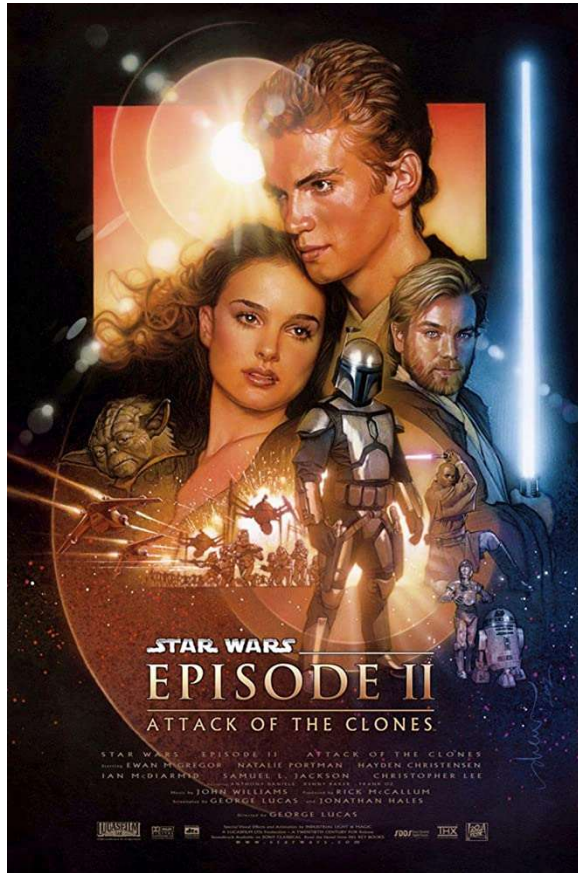
Picture downloaded from <https://www.imdb.com/>



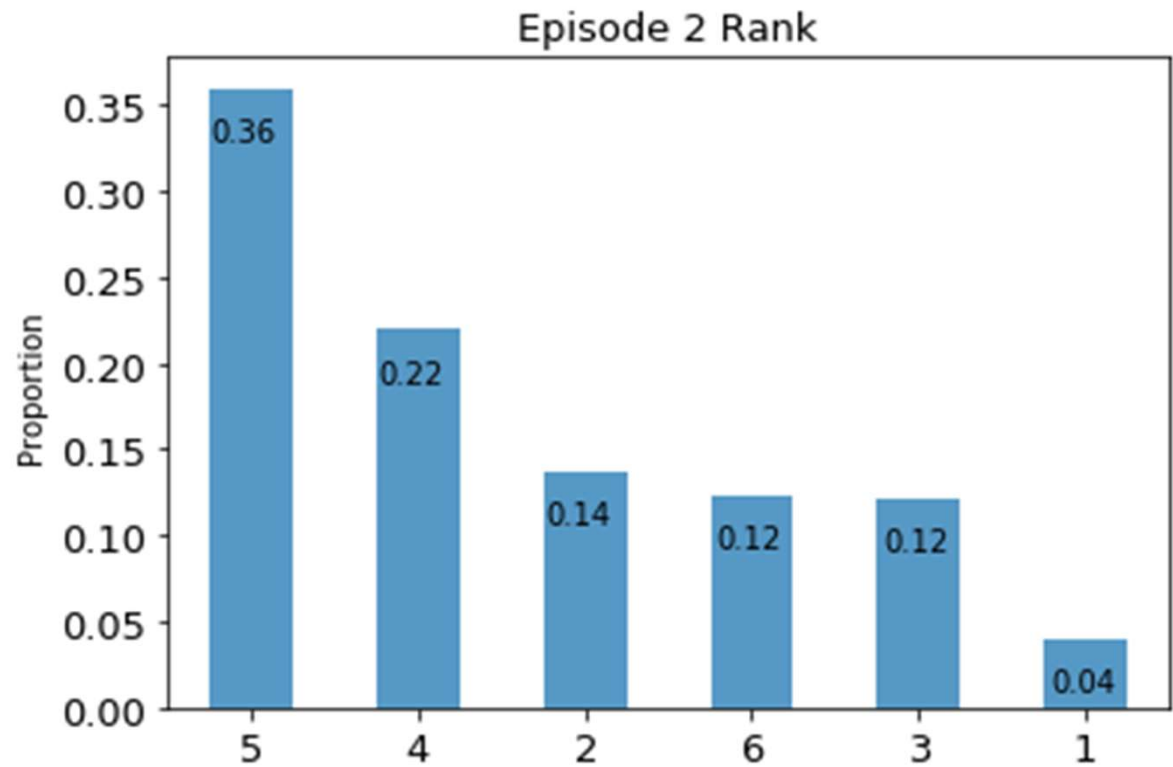


Picture downloaded from <https://www.imdb.com/>



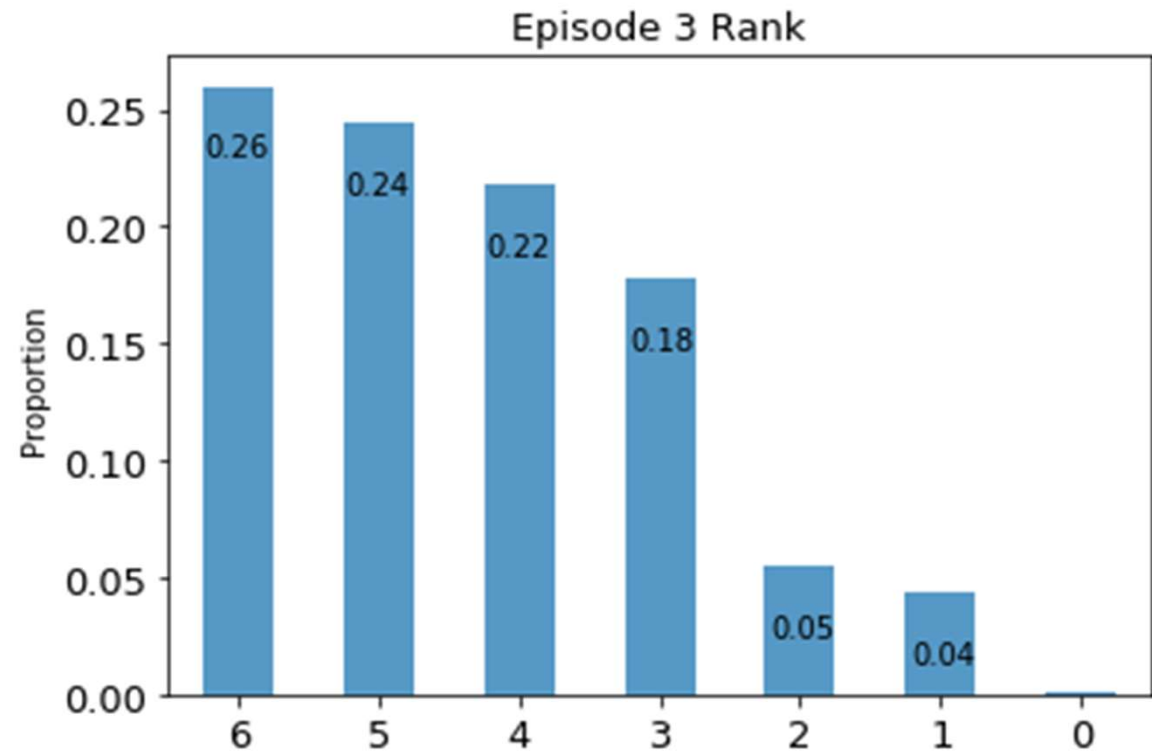


Picture downloaded from <https://www.imdb.com/>





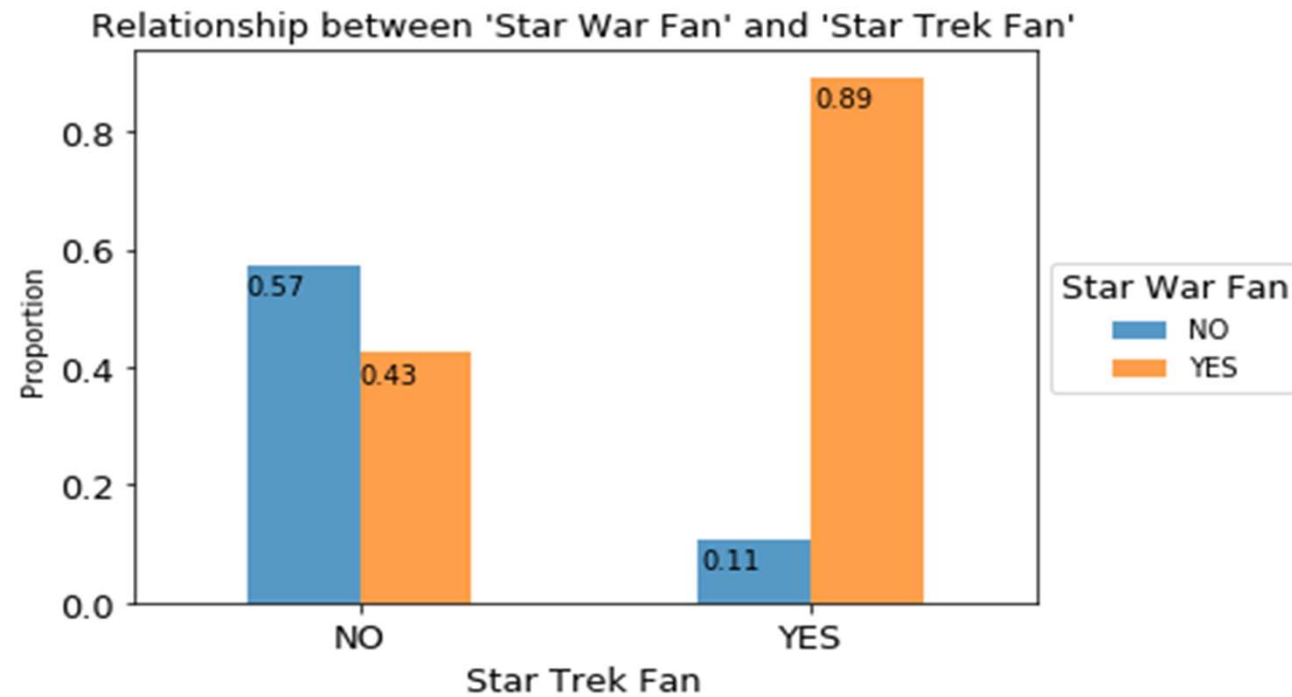
Picture downloaded from <https://www.imdb.com/>



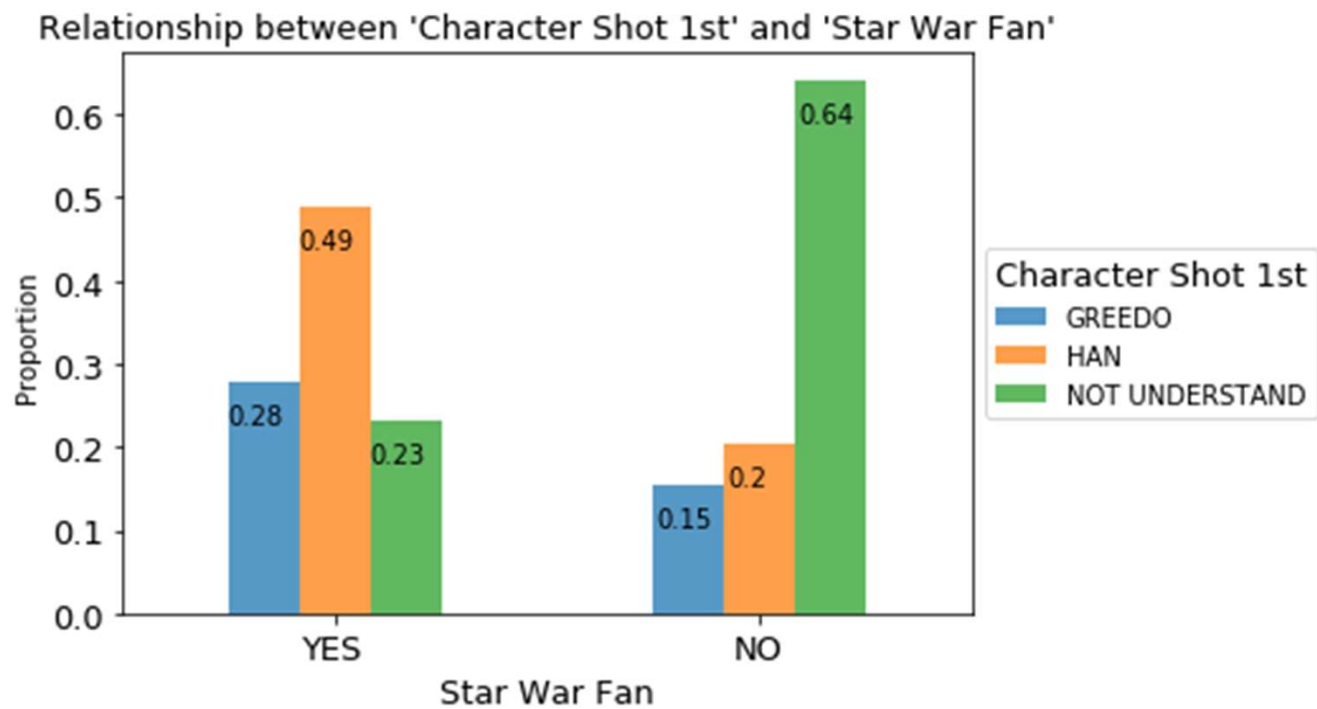
2.2 Hypothesis Relationships Between Columns



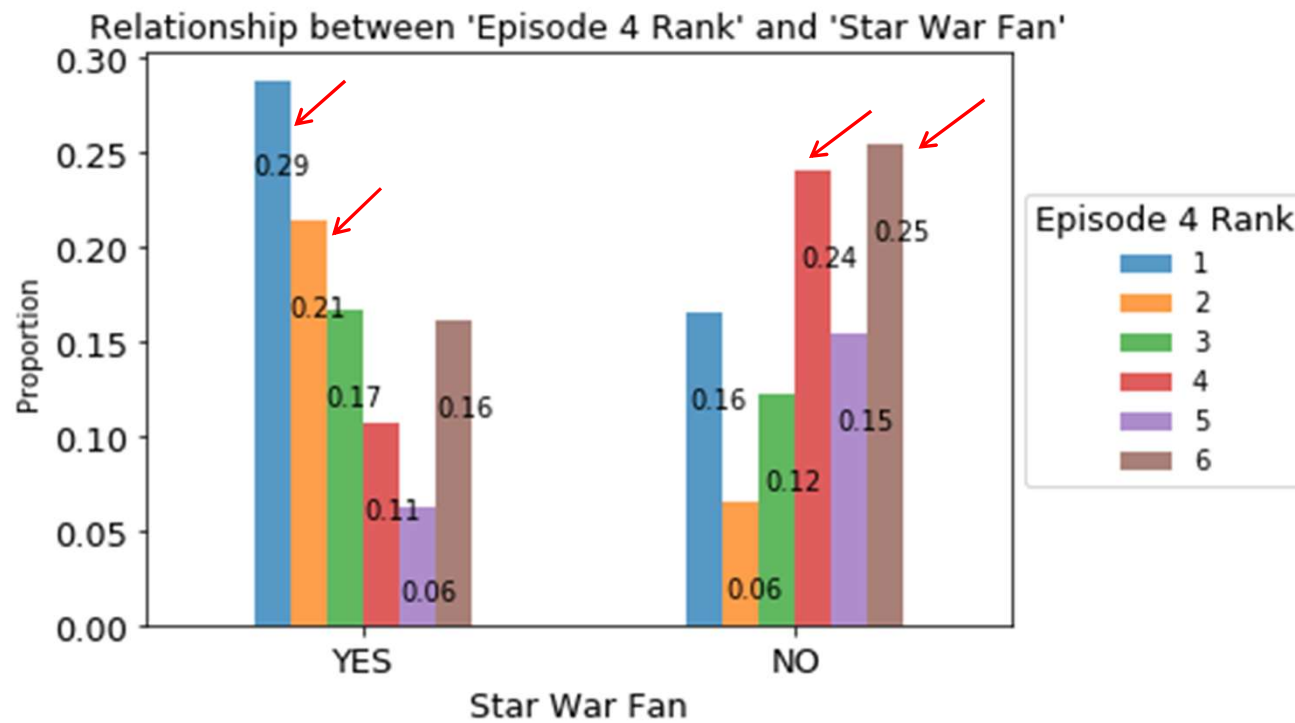
Hypothesis 1: Star Trek fans are probably also Star War fans



Hypothesis 2: Most Star War fans choose Han shot first while most non-fans don't know the question

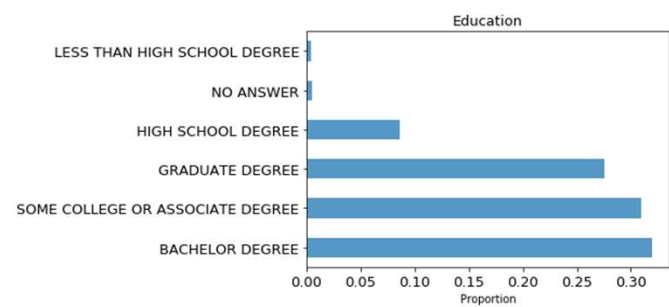
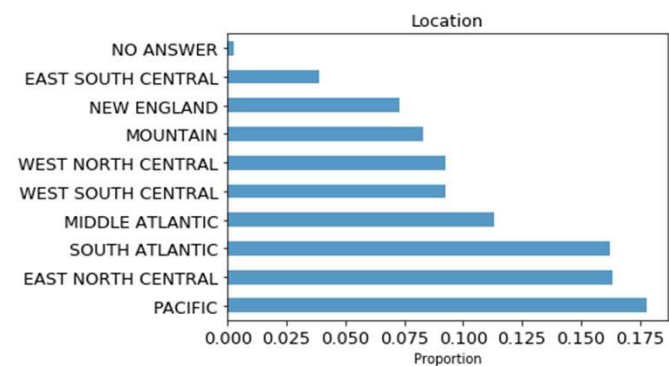
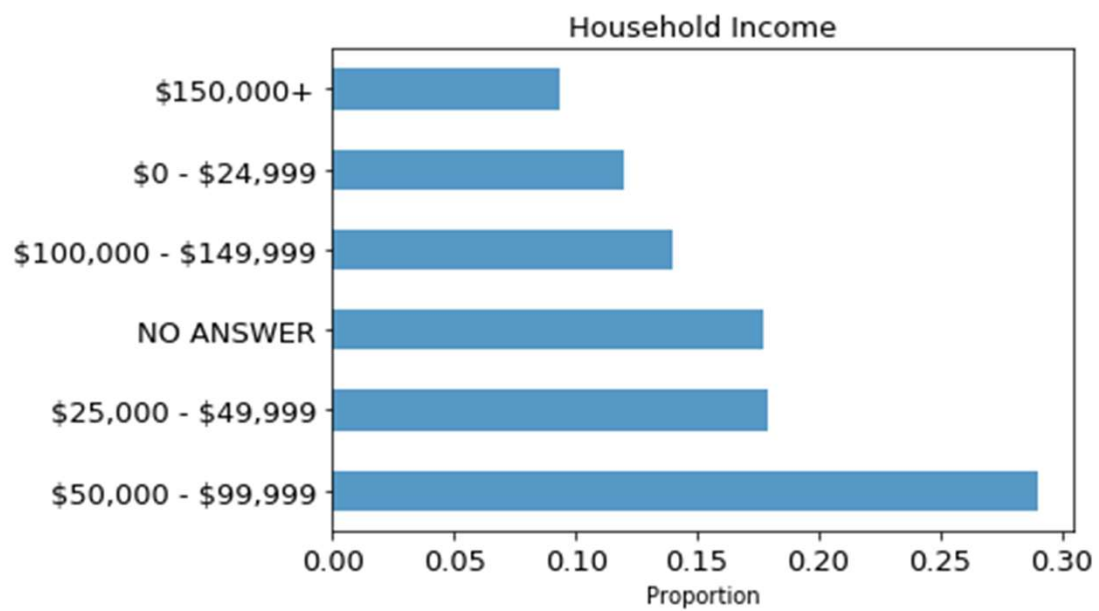


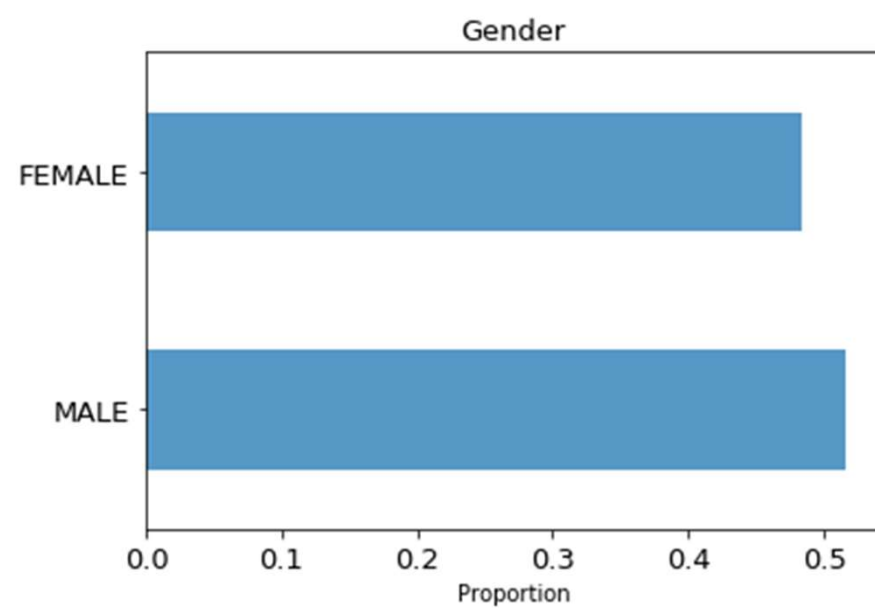
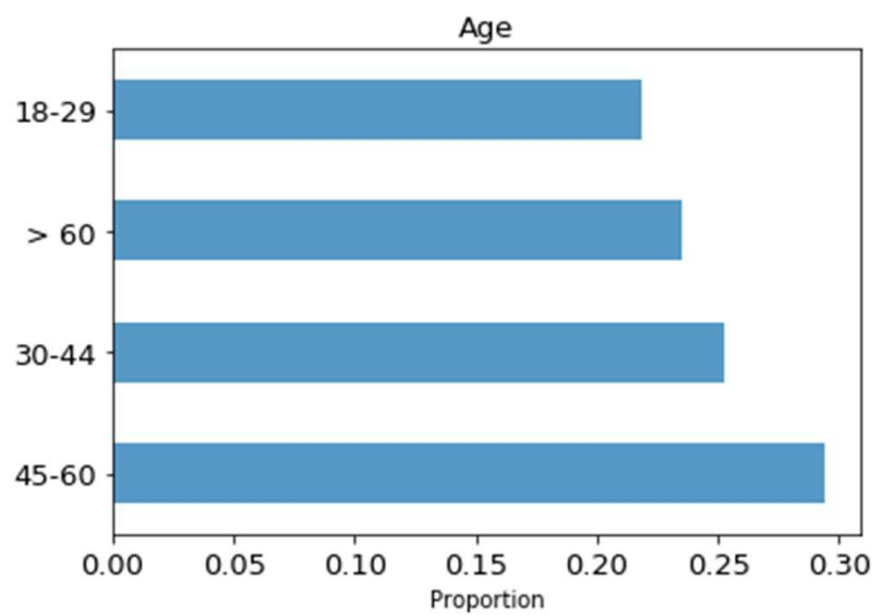
Hypothesis 3: Star War fans rank higher on Episode 4 as it is the first released episode



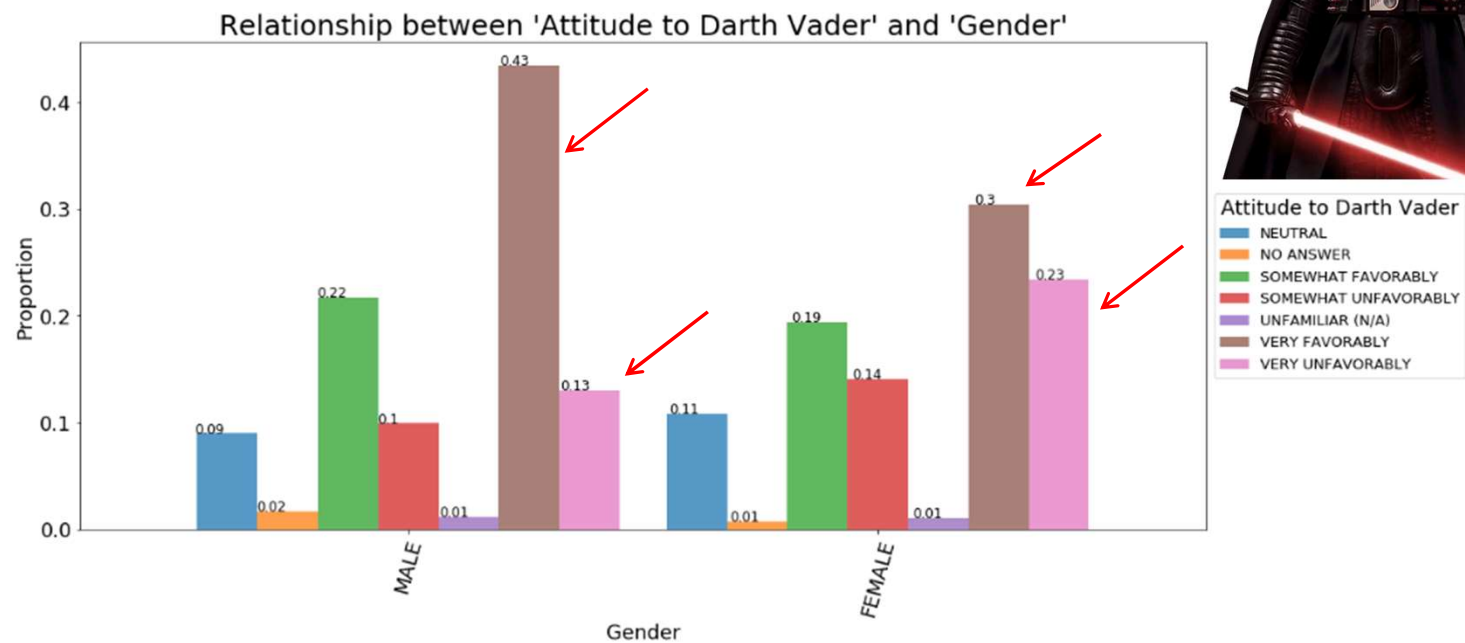
2.3 Relationships between Demography and Attitudes







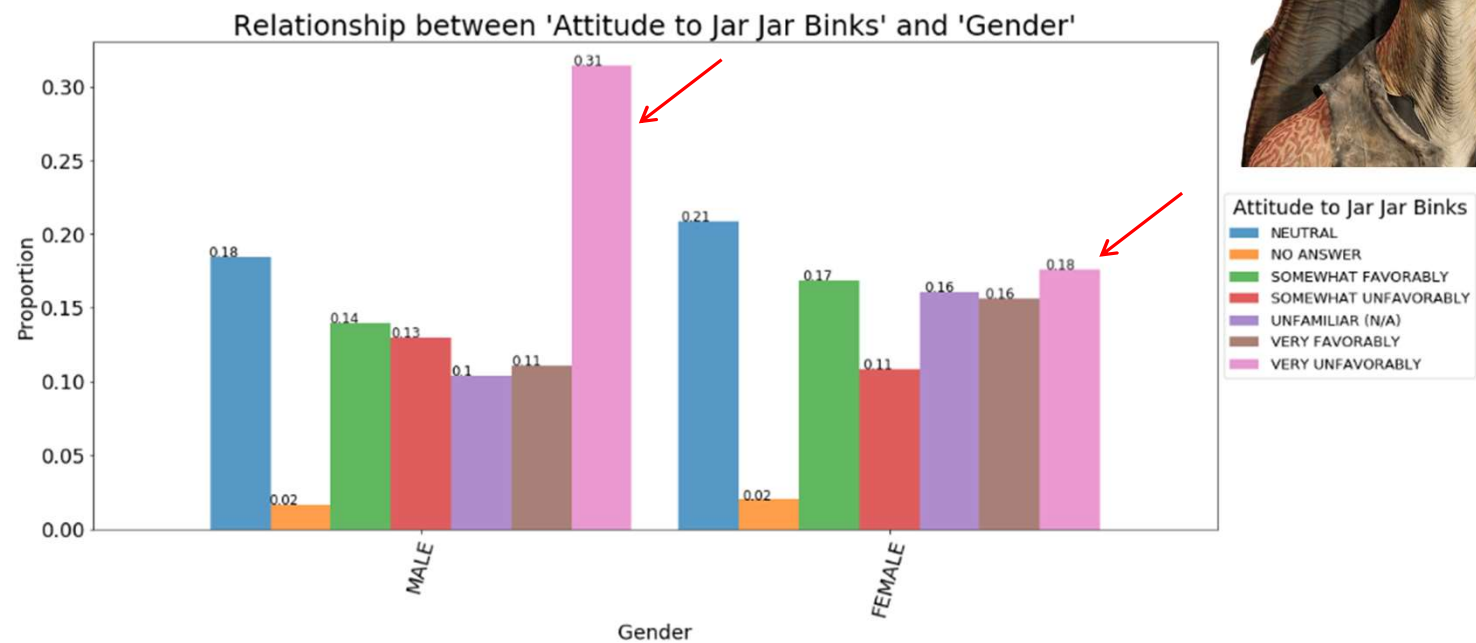
2.3.1 Gender and Attitude



Picture downloaded from <https://toppng.com/>



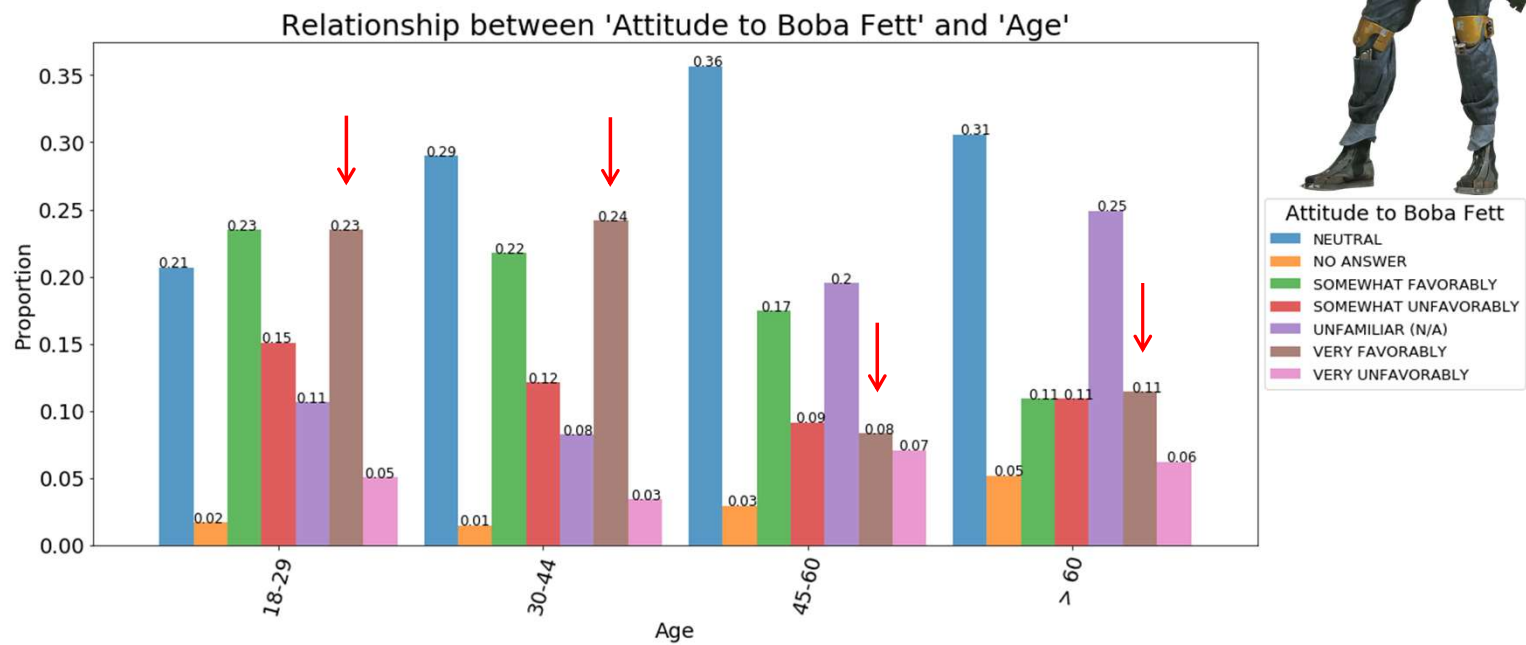
2.3.1 Gender and Attitude



Picture downloaded from <https://toppng.com/>



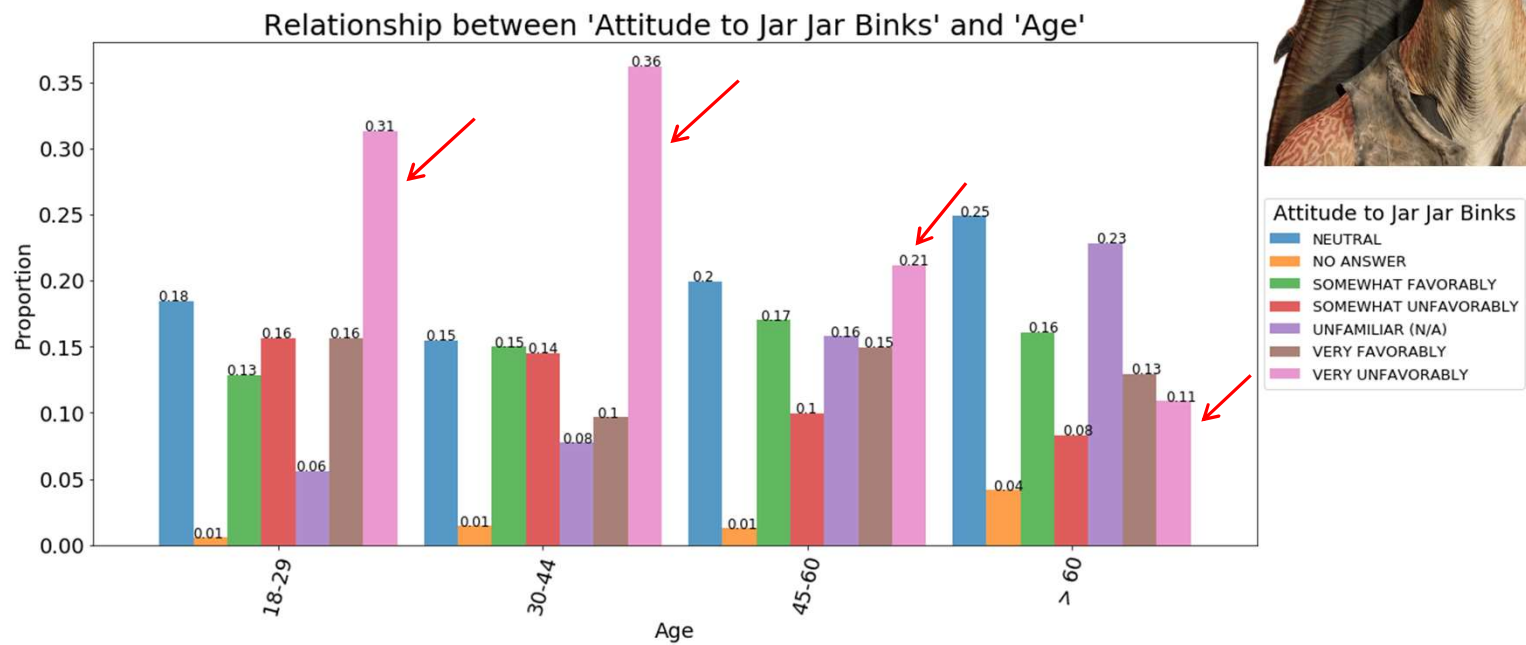
2.3.2 Age and Attitude



Picture downloaded from <https://toppng.com/>



2.3.2 Age and Attitude



Picture downloaded from <https://toppng.com/>



Thanks for watching

