**Star Wars Movie Survey**

**About Dataset:**

This dataset is about online survey of Star Wars Movie Episodes. In which they have received 1187 total response that include unknown response too.

The data has several columns, including:

* RespondentID - An anonymized ID for the respondent (person taking the survey)
* Gender - The respondent's gender
* Age - The respondent's age
* Household Income - The respondent's income
* Education - The respondent's education level
* Location (Census Region) - The respondent's location
* Have you seen any of the 6 films in the Star Wars franchise? - Has a Yes or No response
* Do you consider yourself to be a fan of the Star Wars film franchise? - Has a Yes or No response.

There are several other columns, which involve questions about the Star Wars movies. Some questions involved checkboxes, where someone was asked which of several options they liked, and to check all the ones they did like.

**General Approach For Analysis:**

* Data Preparation
* Data Cleaning and Mapping
* Exploratory Analysis and Visualisation

**Data Preparation:**

Begin with the reading of Csv file that is encoded and observing the first couple of rows of dataset that look alike dirtier and need some cleaning for further analysis.

**Data Cleaning and Mapping:**

This type of data is difficult to represent in columnar format. Most of the columns are unnamed and rows are without respondent id. First we cleaned the Nan values in respondent Id, then we looks at the other columns:

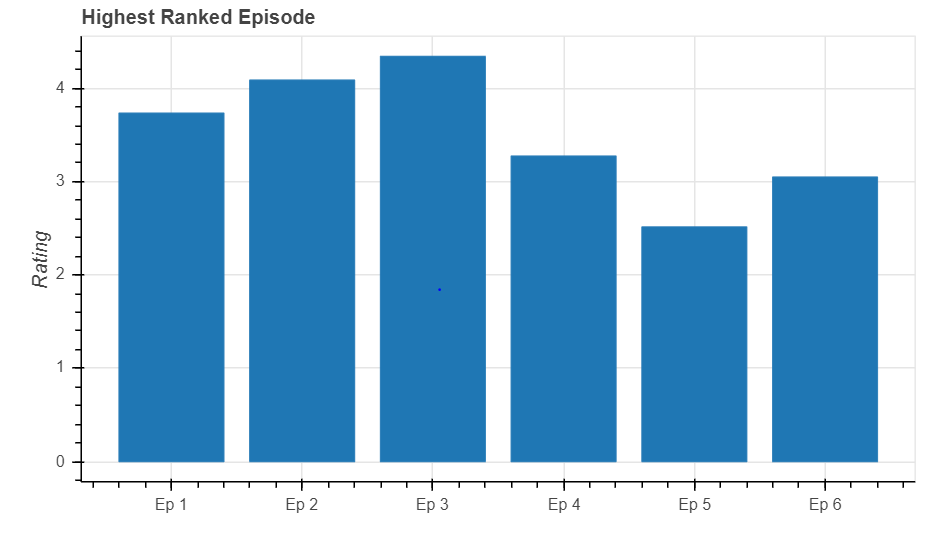
Some columns have string types, because the main values they contain are Yes and No. We can make the data a bit easier to analyze down the road by converting each column to a Boolean having only the values True, False, and NaN. Booleans are easier to work with because we can select the rows that are True or False without having to do a string comparison and this process is known as Mapping.

After Mapping I rename the unnamed columns to make it more intuitive.

**Exploratory Analysis and Visualisation**

Creating Simple Plot for graphical representation to have some clear insight of Dataset.

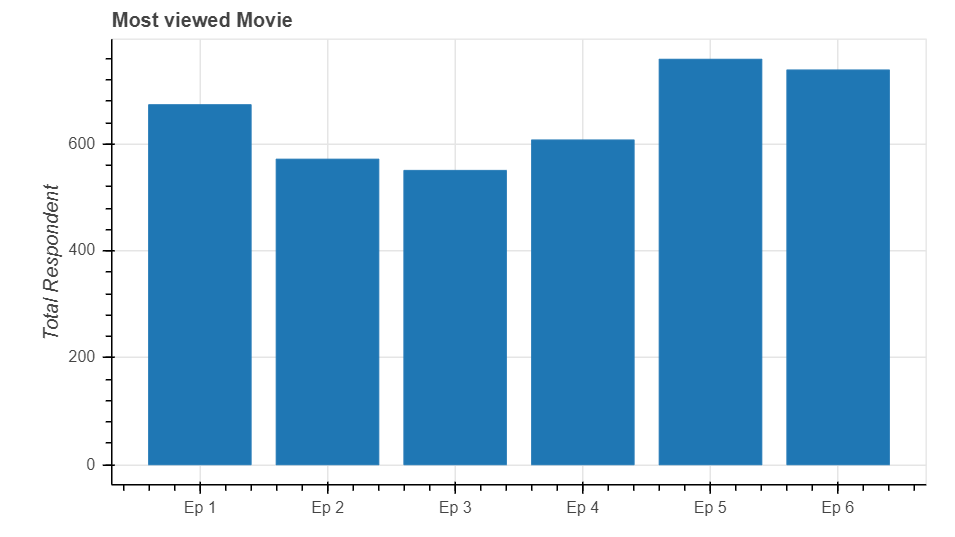
**Highest Ranked Movie:**

****

From the above Bar chart we can see that Star Wars Episode 5 has the highest rating (Lower Rating is better as mentioned in Dataset).

Column 10 in dataset contains the following string: 'Please rank the Star Wars films in order of preference with 1 being your favourite film in the franchise and 6 being your least favourite film. So columns with lower ranking values are considered better by the survey respondents. From the chart, it looks like the movies(4-5-6) have higher rankings than the others.

**Most Viewed Movie**

****

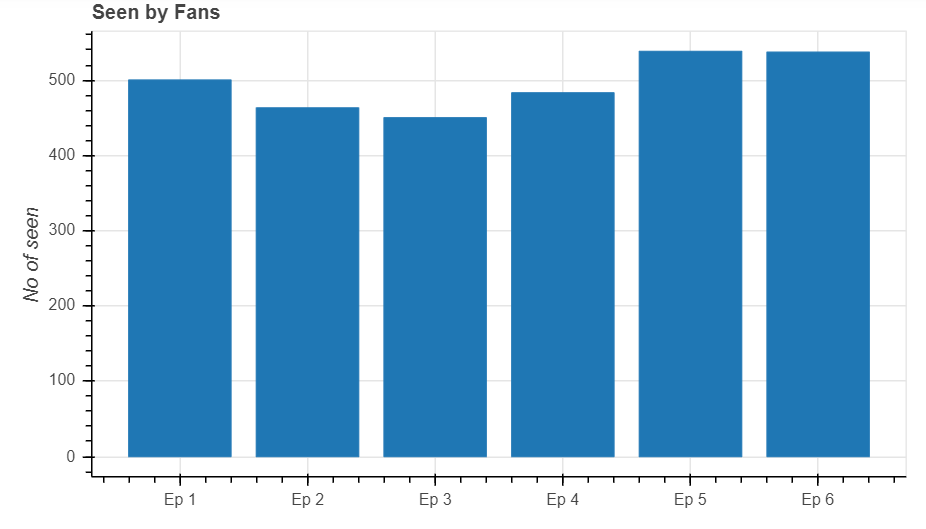
Same thing here, more respondents saw the movies(4-5-6), as they ranked higher. Keep in a mind a lower value for average ranking means the respondent liked the movie more. People are likely to watch those movies that have better Ranking.

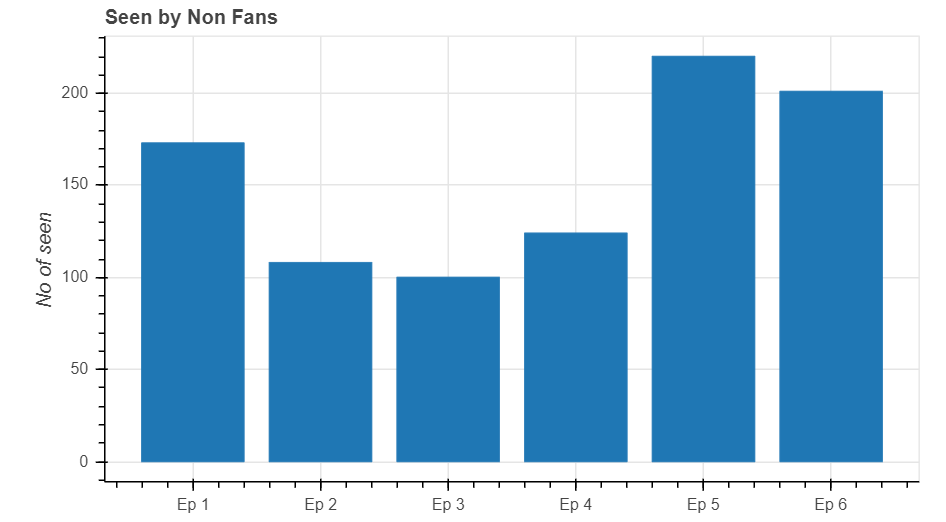
**Segmentation of Respondents by Fans, Non fans and unknown**

**A screenshot of a cell phone

Description automatically generated**

We can see that more than 50% respondents are fans and About 30% of the respondents who did not specify whether they were fans or not

****

****

From the above two chart we can observe that respondents who consider themselves as fans of Star Wars saw each episode, whereas non fans of star wars didn't watch all the episode.

**Exploring Data by Gender Segmentation**

**A screenshot of a cell phone

Description automatically generated**

As compared to females more number of males have watched the movies, and in case of reviews given by females are more as compared to male by this we can say that most of the females who have watched the movie have given the review

**Segmentation of data by Education and Location**

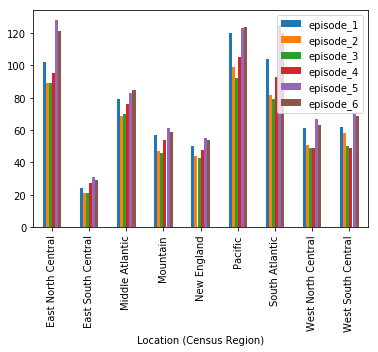
**By Education**

**A picture containing screenshot

Description automatically generated**

From this categorical Bar Chart we can see People watching star wars movies have educational background related to Bachelor Degree,Graduate degree,Some college or Associate degree. Most people lie in the category of some college and associate degree.

**By Location:**

****

Highest star wars movie watchers belong are from East North Central, Pacific, South Atlantic. The least star movie watchers are from East South Central.