1st Capstone; Inferential Statistics Steps.

Title: **“Exploring Netflix’s movie recommendation system”**

All of the code and plots associated with this report can be found in the ‘EDA.ipynb’.

Once the data was successfully wrangled, the first step in my analysis was to look at the distribution of the movie ratings. For this, the data was grouped according to the rating value(1, 2, 3, 4 or 5 stars) using the groupby() method combined with the count( ) function. From this, I learned that the most frequent rating is “4 Stars”.

Next, I grouped the data by user using groupby() along with the mean() function to look at the average rating for each user. The result was plotted with the hist() function from matplotlib. I also used the normaltest() function from scipy.stats to test if the average rating by user was distributed normally. This distribution can be considered, when trying to determine which percentage of users might have a bias (positive or negative) when rating movies.

I also analyzed if the year of released of movie had an effect on the movie rating. For this I exracted all of the rating for movies released on a specific year. Then, I used a one-sample t-test (ttest-1samp() function from scipy.stats) to compare this ratings with the average rating of the entre dataset (~3.6 Stars). In a similar way as explained before, I also tested if the ratings were influenced by the day of the week or the movie’s genre. For the use of this one-sample t-test, I considered a pvalue < 0.001 as indicator of significance.

This analysis allowed me to conclude that factors such as “User specific bias”, “Movie’s release year” and “Movie’s genre” have a signigicant influence in a movie’s rating, thus, they could be incorporated into a predictive model for movie preferences and ratings.

For more details on this statistical analysis, please refere to the following Ipython notebook:

<https://github.com/masaver/springboard/blob/master/capstone1_explring_netlix's_movie_recomender_system/exploratory%20data%20analysis/EDA.ipynb>