Project 2

Investigate a Dataset

Introduction:

First, I have investigated **TMDB Movies** dataset.

In investigation I have used Pandas, numpy and matplotlib

Also I have puitted 3 Questions to investigate:

- what is the top 10 movies in 2015?
- What is the most frequent genre?
- What is the most runtime liked by the audience over the popularity?

Data Wrangling:

Here I have checked the data I found null values in some rows, one duplicated row and rows have zero values also I found unnecessary columns.

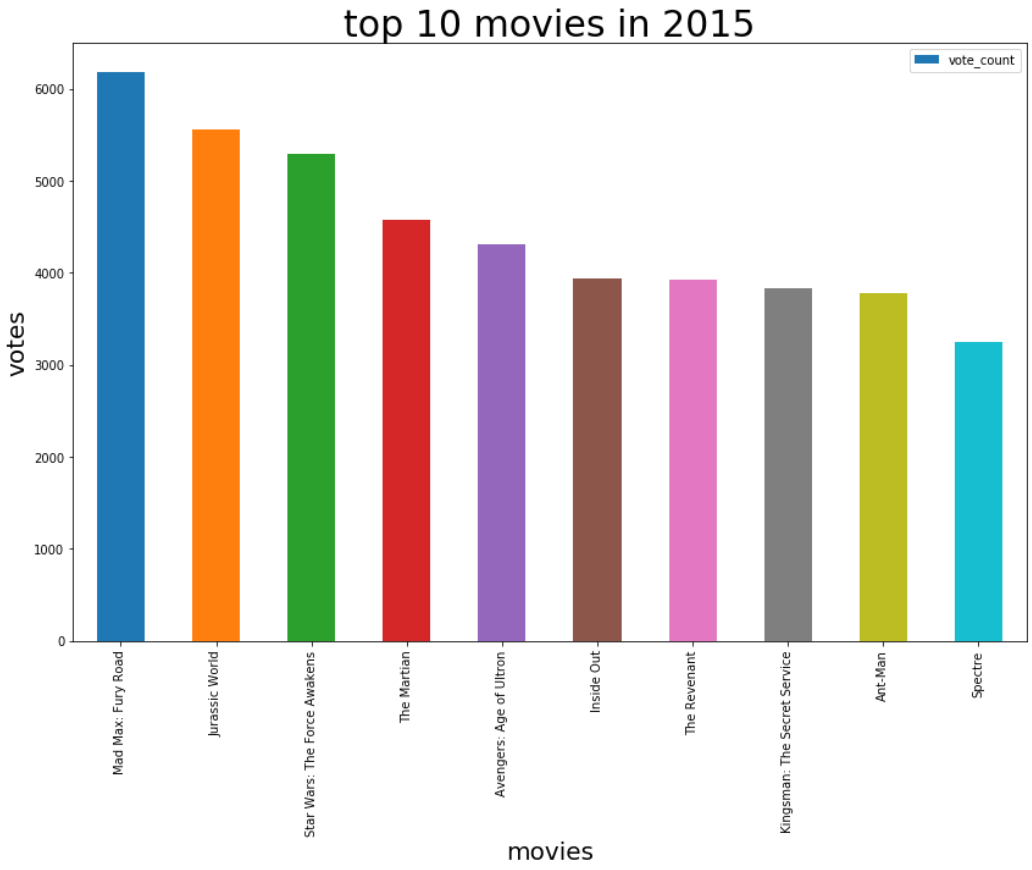
Then I have done some cleaning processes such as deleting unnecessary columns, deleting null values and removing zero values

Exploratory Data Analysis:

After we have done cleaning the data.

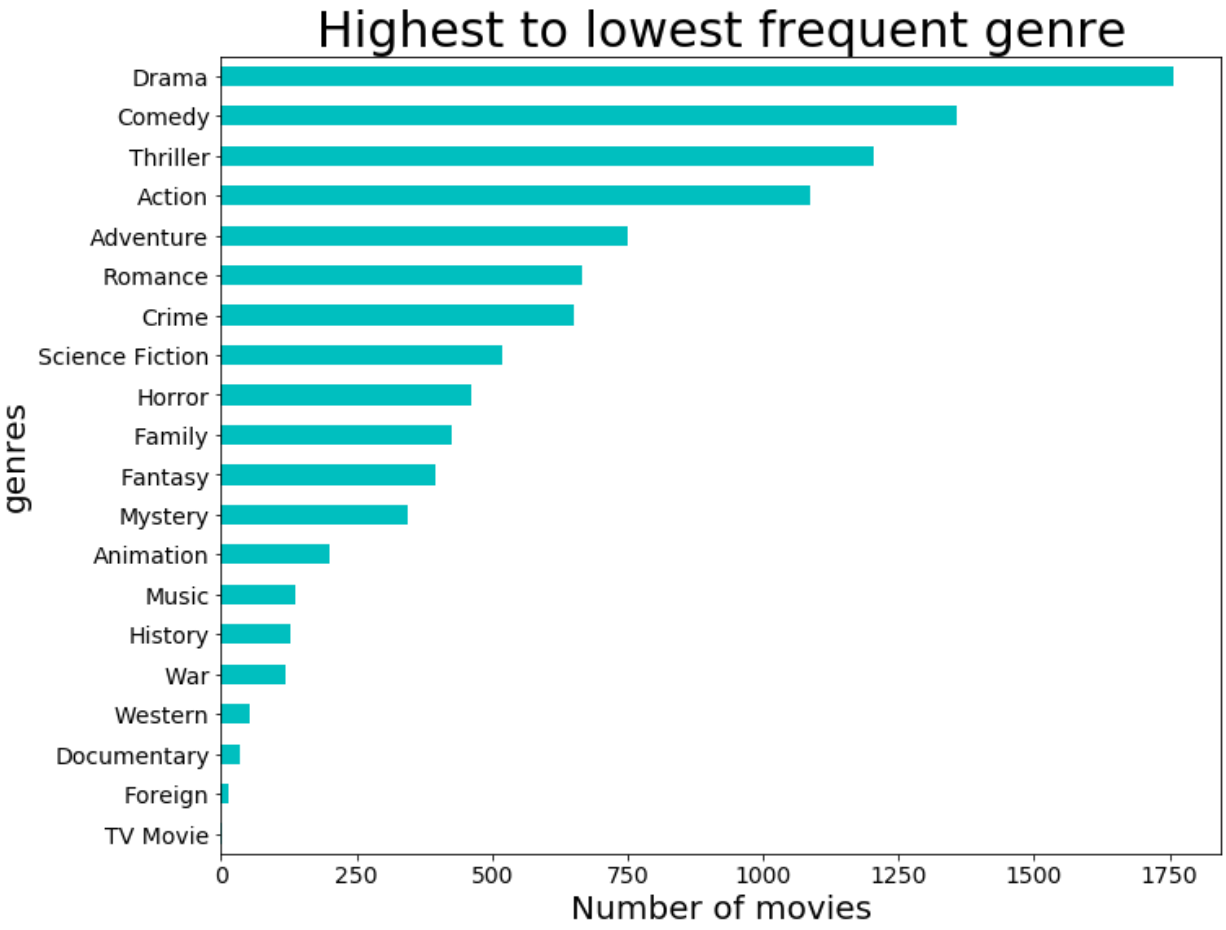
now, we can do some explorations

Research Question 1 (what is the top 10 movies in 2015?



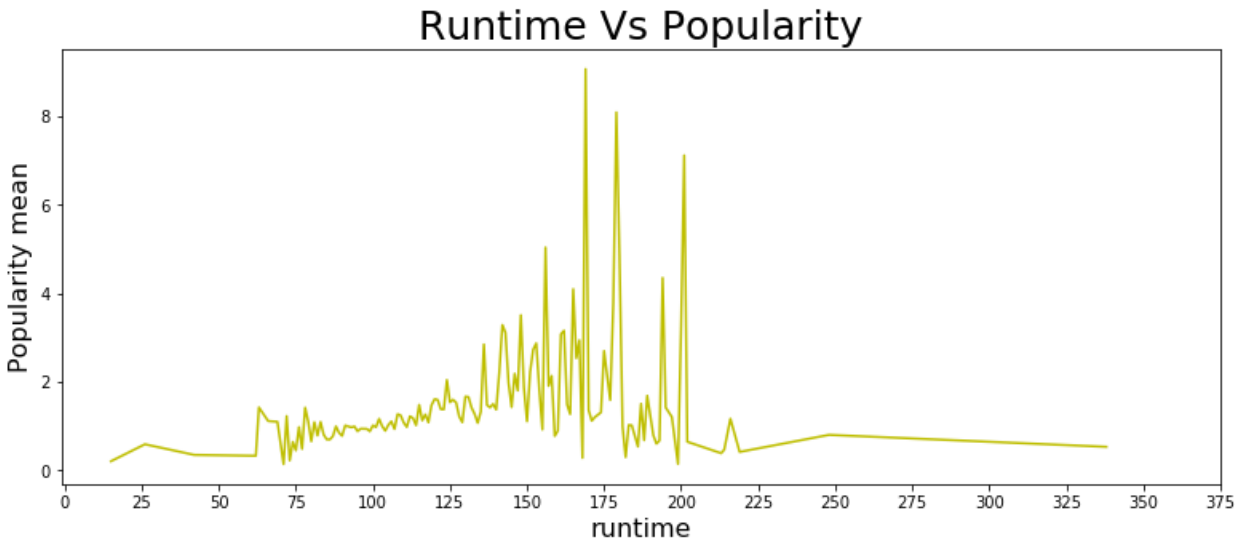
According to the plot we can see that (Mad Max: Fury Road) has more than 6000 votes in 2015.

Research Question 2 (What is the most frequent genre?)



as we can see in this plot (Drama) is the most frequent genre among the other genres

Research Question 3 (What is the most runtime liked by the audience over the popularity?)



the runtime liked by the audience is between 160 and 190 runtimes but approximately most liked is 165.

Conclusions:

1- In Question 1 we can conclude that according to the plot we can see that (Mad Max: Fury Road) has more than 6000 votes in 2015.

2- In Question 2 we conclude that the (Drama) is the most frequent genre among the other genres.

3 - In Question 3 the runtime liked by the audience is between 160 and 190 runtime but approximately most liked is 165.

Limitations:

this dataset is very rich in information it has (10866) rows and (21) columns but it has sum null values and 1 row is duplicated also I found 0 values in some rows so I needed to clean this data by removing the columns that I don't need for analysis and remove rows that has null values and 0 values until the data cleaned it was contains 10866 records after cleaning it become 3855 records.