

# Project: Investigate a Dataset - [TMDb Movie Dataset]

## Table of Contents

- [Introduction](#)
- [Data Wrangling](#)
- [Exploratory Data Analysis](#)
- [Conclusions](#)

## Introduction

### Dataset Description

The TMDb dataset. This data set contains information about 10,000 movies collected from The Movie Database (TMDb), including user ratings and revenue.

### Question(s) for Analysis

1. Which movies are most popular from year to year?
2. What kinds of properties are associated with movies that have high revenues?
3. What are the properties of movies that have high average votes?
4. Is the budget any inclination of expected revenue?
5. What movie had the highest and lowest profits for each year?

## Data Wrangling

Here I load the selected data, check for cleanliness and then trim and clean the data to prepare it for analysis

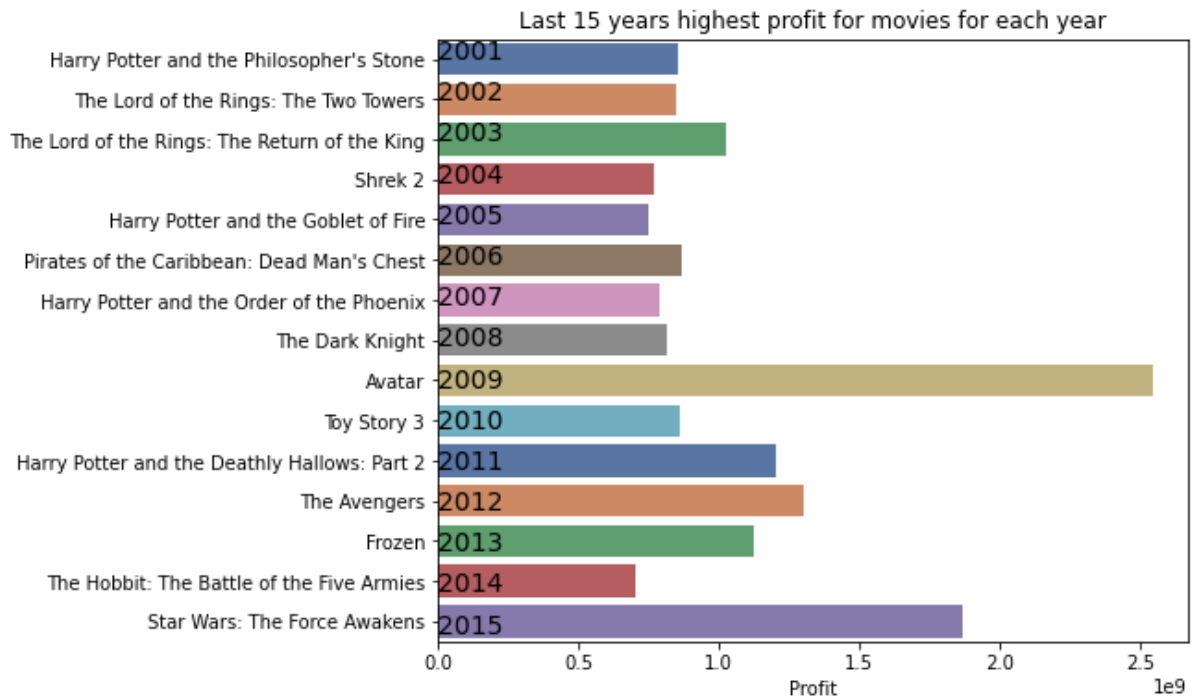
- I checked for the shape of the dataset. There were 10866 rows and 21 columns

- There were some columns that i concluding would not be useful in the exploration of the data so i deleted those columns and trimmed the data
- Of the needed columns, the null values were removed and the only duplicated row was also dropped
- “Release\_date” column was converted to datetime format
- After cleaning the dataset, there were a total of 15 columns and 9772 rows

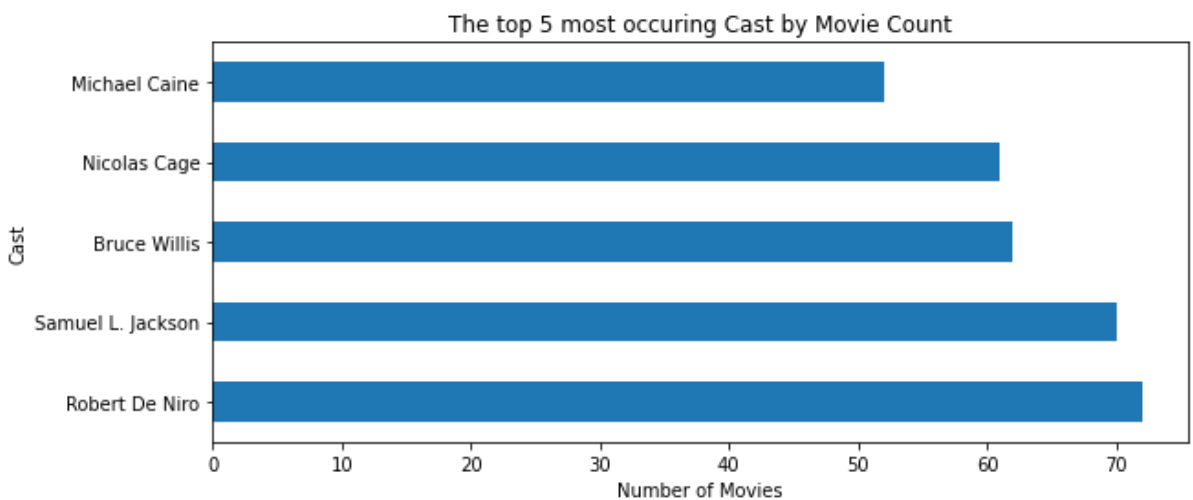
## Exploratory Data Analysis

Here i first created funtions that will be used in answering the questions posed before delving into exploring the data. The functions helped to prevent the repetition of code.

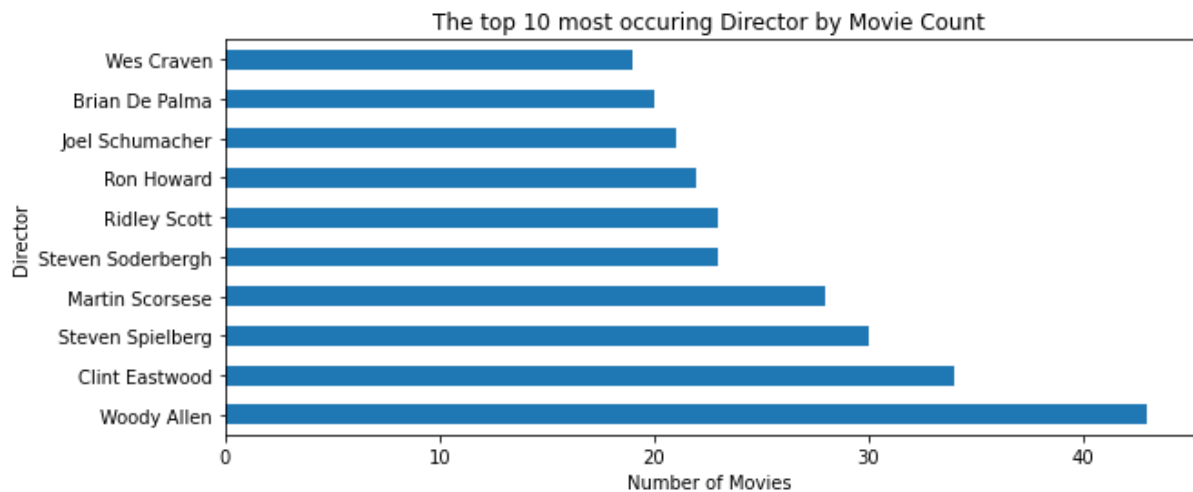
- The most profitable movie is Avatar and filmed in 2009. Star Wars: The Force Awakers is second, and Titanic is the third one.
- The least profitable movie is The Warrior’s Way and this movie also has the highest budget.
- The most profitable genre is: Action|Adventure|Fantasy|Science Fiction
- The genre with the lowest profit is adventure |Fantasy |Action | Western|Thriller
- The most popular genres was Drama, Comedy, and Action.
- The highest profiting movie for each year:



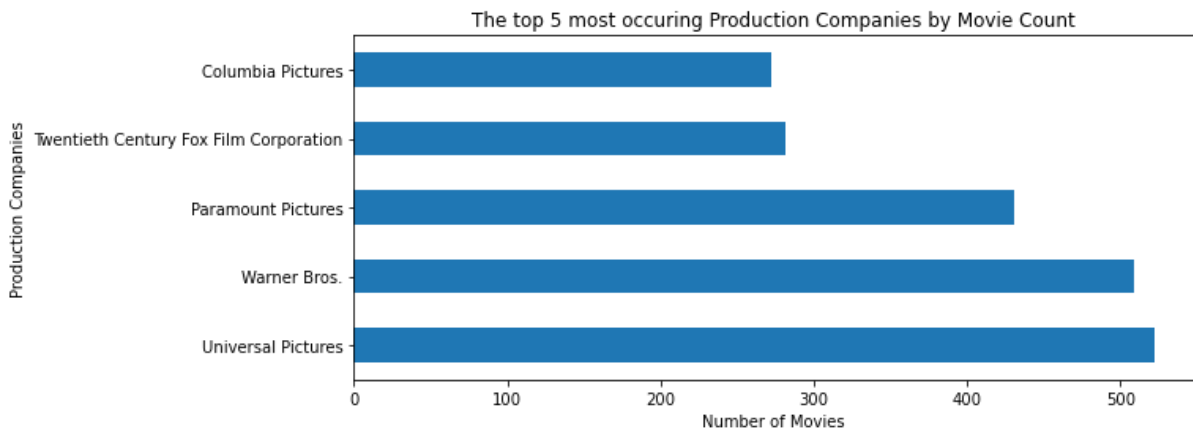
- The five most filmed actors were Robert De Niro, Bruce Willis and Samuel L. Jackson, Nicholas Cage and Micheal cane



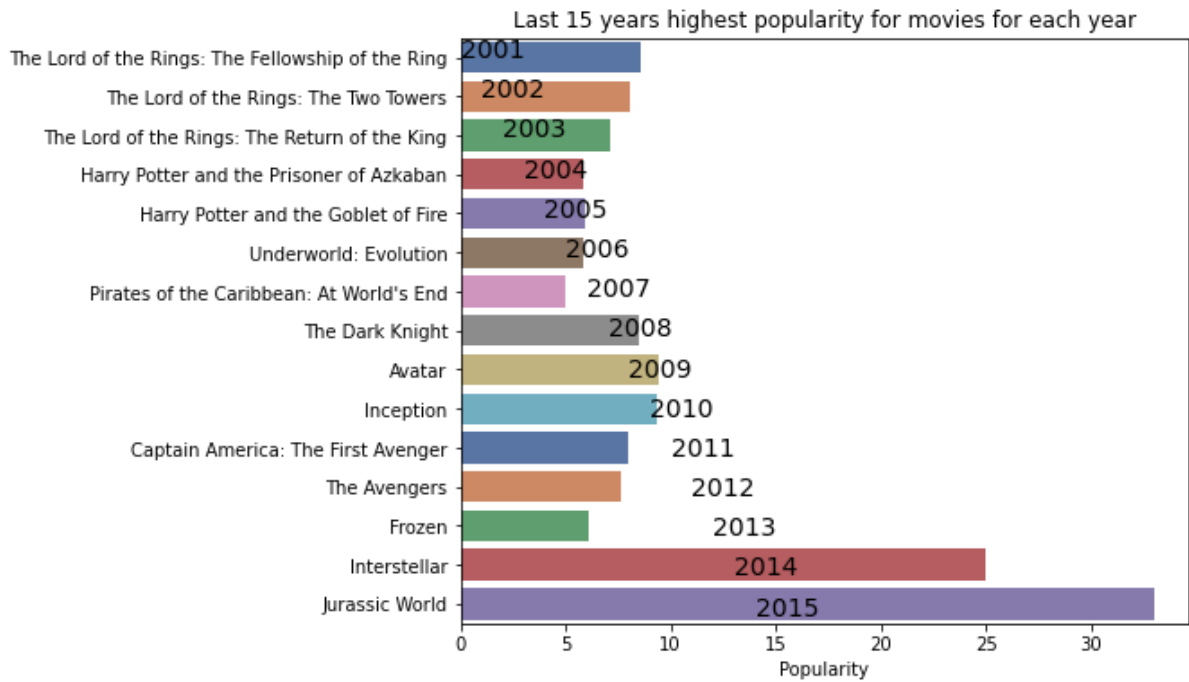
- The most filmed directors were Steven Spielberg, Clint Eastwood, and Ridley Scott.



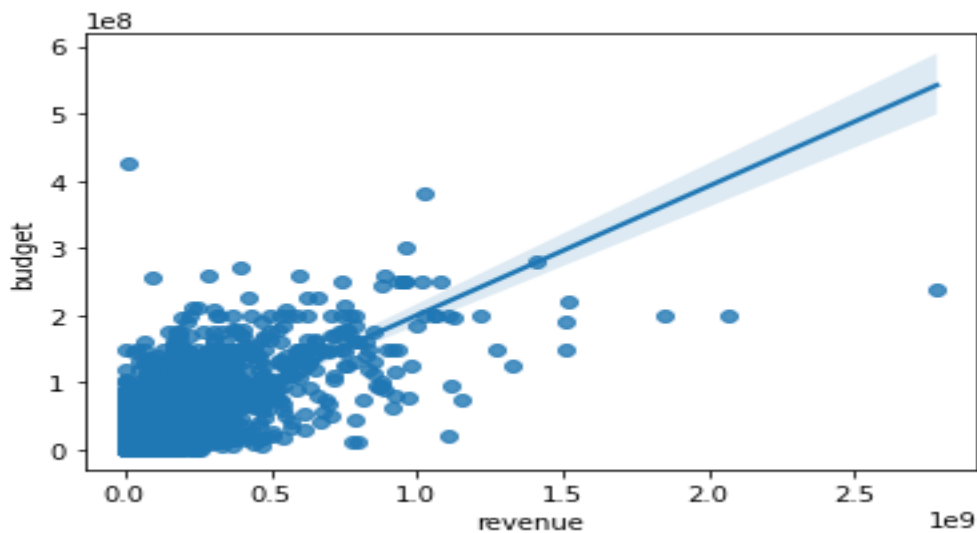
- The most filmed production company was Universal Pictures, Warner Bros, and Paramount Pictures.



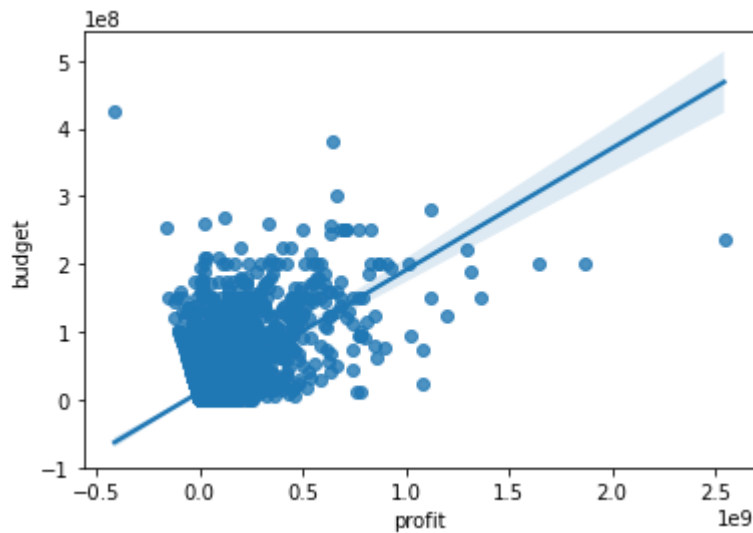
- most popular movie for each year Jurassic in 2015 and inception in 2010



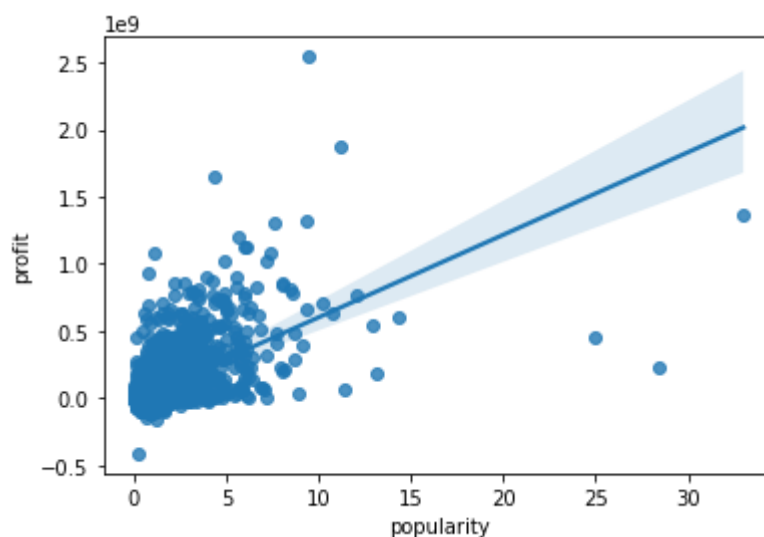
- Revenue and budget both have a positive correlation between them. This means there is a good possibility that movies with higher investments result in better revenues.



- Profit Vs Budget: Profit And Budget both have a positive correlation between them. Means there is a good possibility that movies with higher investments result in better profit.



- **Popularity Vs Profit:** Popularity and profit have a positive correlation. It means that movie with high popularity tends to earn high profit.



## Conclusion

Limitations that prevent a broader analysis of the data set are:

There were some missing values in the dataset affecting our results.

Incorrect datatypes had to be fixed

