

# TMDb movies dataset

What features could influence movie popularity and profit ?

Project Module 2

Orphee

# Data Source n°1

## Drake lyrics

kaggle

Format : This is a dataset consisting of movies released from 1960 to 2016

10K+ Lines and 21 columns originally

<https://www.kaggle.com/juzershakir/tmdb-movies-dataset/code>

< **tmdb\_movies\_data.csv** (6.54 MB)

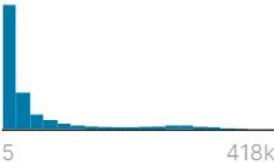

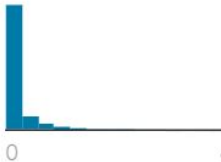


Detail Compact Column

10 of 21 columns ▾

### About this file

CSV file containing 10k+ observations on Movies

id	imdb_id	# popularity	# budget
	<b>10856</b> unique values		

# Challenges

## Challenge 1

**Original dataset columns had weird columns**

Delete unnecessary columns : Homepage, id, keywords, taglines...

## Challenge 2

**Many nulls and duplicates**

About 60% of the dataset contained budget and revenue with a value of zero.

We had to delete them to get better distributions.

We had also to remove duplicates

## Challenge 3

**Add the profit column**

Subtract revenue and profit

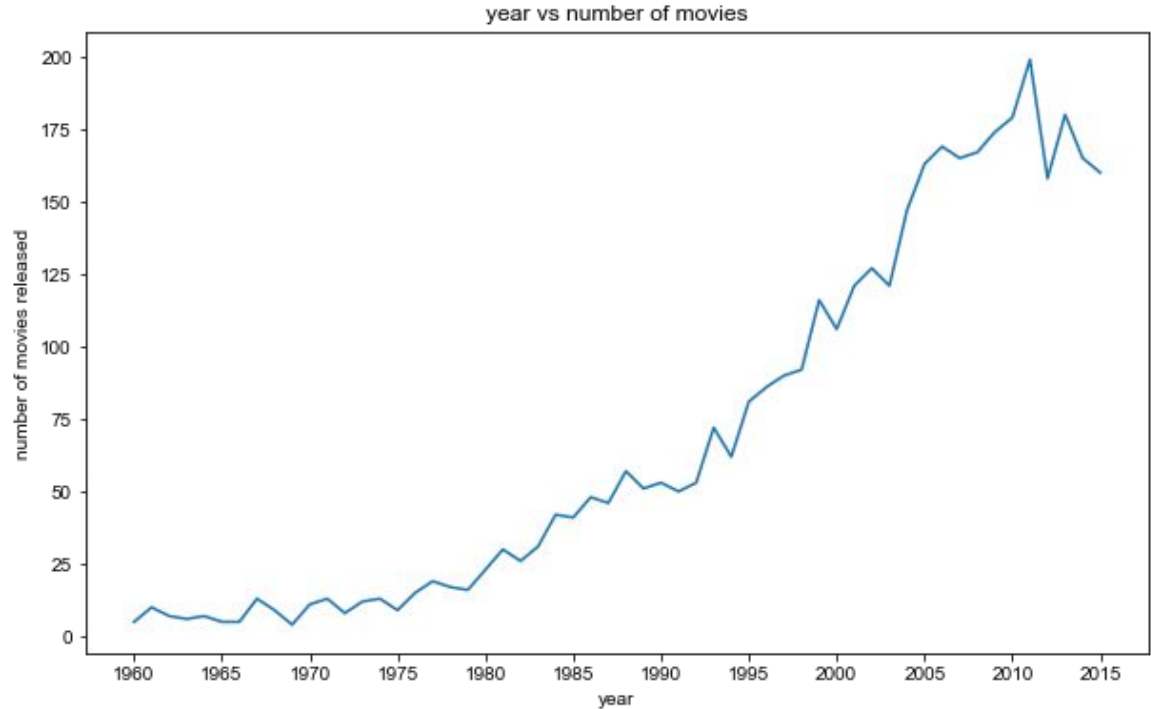
# Findings

Features influencing movies  
popularity and profits



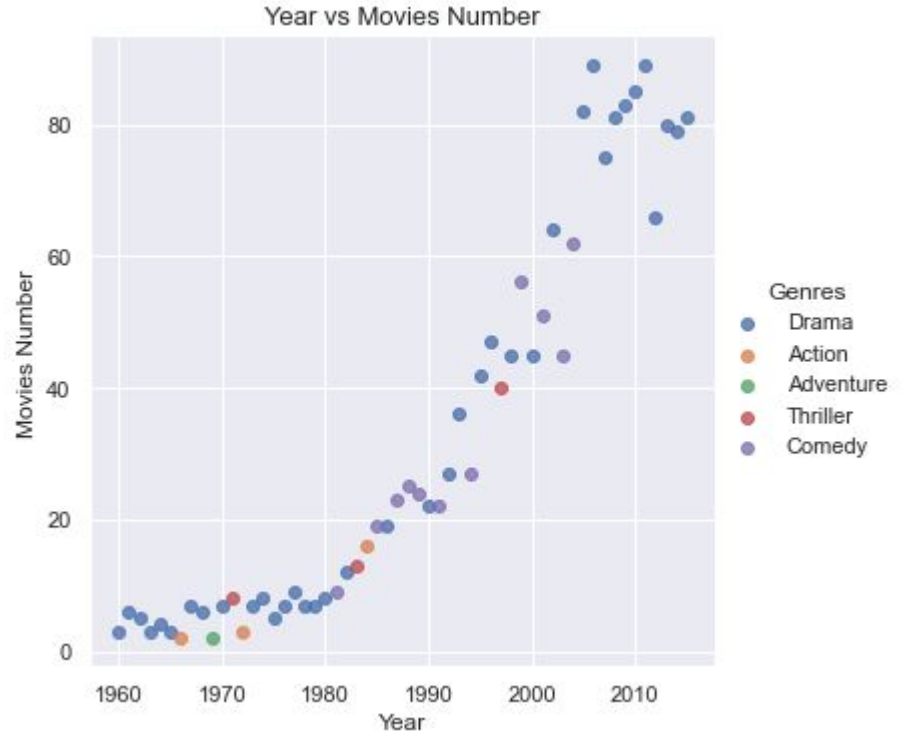
# Findings n°1

The more the years are advanced , the more of the movie production



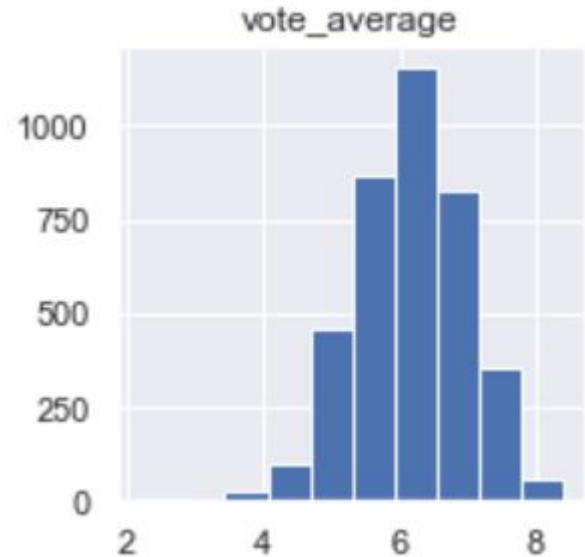
# Findings n°2

The Drama genre is the genre evolving the most through history



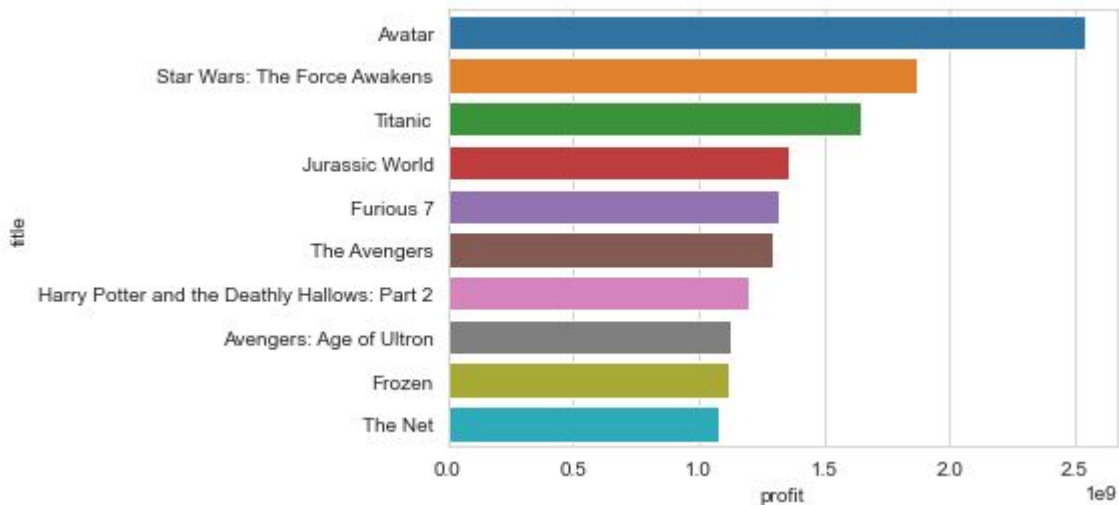
# Findings n°2

The vote average is following a normal distribution



# Findings n°3

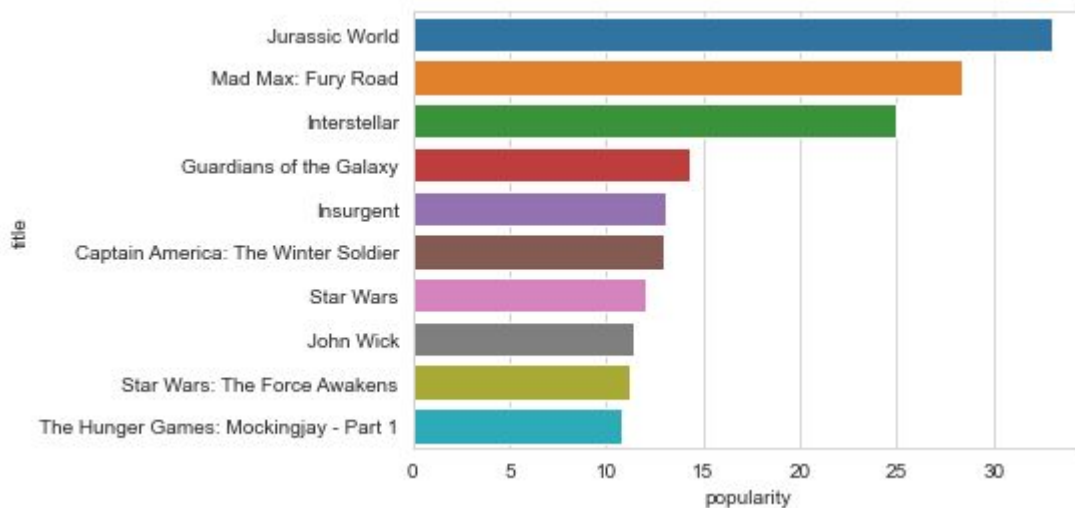
Avatar is the movie that made the most profit with 2.5 Billion dollars (2009 by James Cameron)





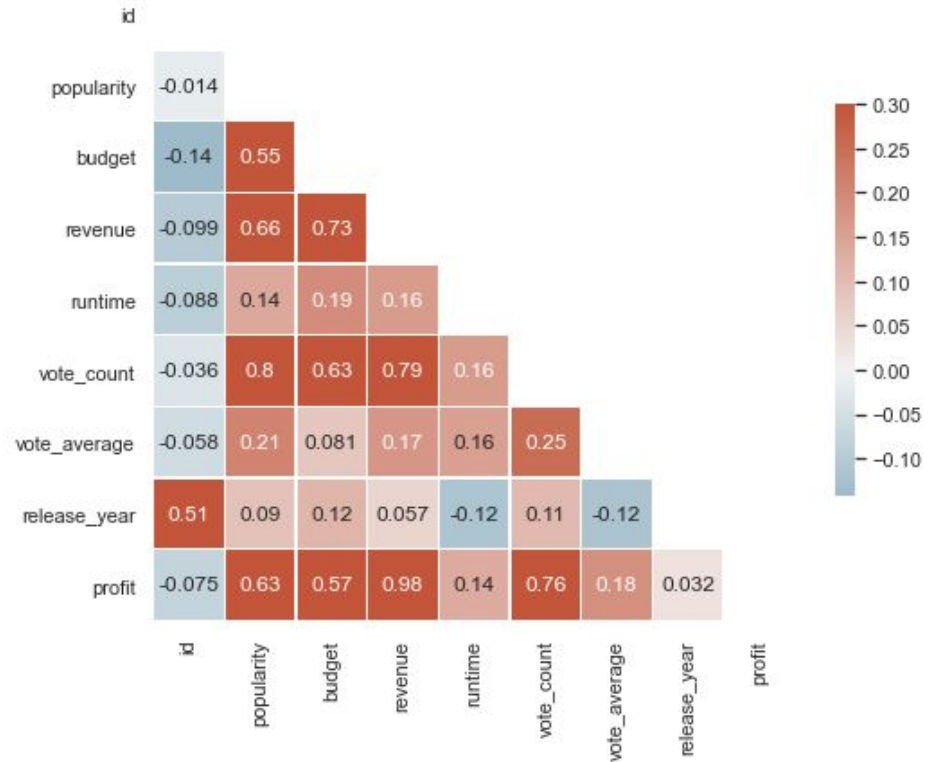
# Findings n°4

Jurassic World is the most popular movie with a score of 32.9 (Colin Trevorrow in 2015)



# Findings n°5

- Movies with high budget have a higher profit and revenue
- Movies with high profit have a higher popularity and revenue
- Movies with high popularity have a higher profit and revenue
- In a nutshell popularity and profit are mainly driven by the revenue of a movie



# Further analysis

-

Getting the relative profit would give more details

Analyzing the popularity and profit by directors

---

# Thanks

Orphee

---