



29/11/2023

# The Performance of Hollywood Movies

Python & Power BI Assignment



Noelli Sipolatti

DATA TECHNICIAN SKILLS BOOTCAMP – JUST IT

## Table of Contents

The performance of Hollywood Movies.....	2
Scenario.....	2
Python.....	2
Power BI .....	3
Final Considerations.....	4
Sources.....	4

# The performance of Hollywood Movies

## Scenario

In this assignment, I will investigate into an ample analysis of the performance of Hollywood movies, examining various key factors that contribute to their performance. The dataset under analysis contains essential information such as movie titles, genres, studios, profitability, and ratings, covering the period from 2007 to 2011.

Through a careful exploration of these elements, we target to collect valuable insights into the dynamic landscape of the film industry during this specific timeframe.

## Python

In the initial exploratory analysis phase, I choose to utilized Python programming language and a Jupyter notebook within the Google Colab environment, to create a script that efficiently loaded and processed the data. This step involved running exploratory commands, checking for missing values, and performing necessary data cleaning and transformation tasks, such as changing data types and renaming columns. The clean data was then exported for further visualization.

The process unfolded as follows:

- Creation of a New Script.
- Loading the Dataset.
- Execution of Exploratory Commands, such as checking information about the dataset, visualizing head, and bottom rows, checking column names, and examining data types.
- Data Cleaning and Transformation: checking and removing missing values, changing data types, renaming column name, and rounding two of the columns to two decimal places.
- Finally, exporting the Clean Data.

```
[70] #Importing Panda
import pandas as pd

[71] #Loading the dataset
df=pd.read_csv('HollywoodMostProfitableStories.csv', encoding='latin1')
df

   Film      Genre  Lead Studio  Audience score %  Profitability  Rotten Tomatoes %  Worldwide Gross  Year
0  27 Dresses    Comedy        Fox             71.0      5.343622             40.0      160.308654  2008.0
1  (500) Days of Summer  Comedy        Fox             81.0      8.096000             87.0      60.720000  2009.0
2  A Dangerous Method    Drama    Independent             89.0      0.448645             79.0       8.972895  2011.0
3  A Serious Man        Drama    Universal             64.0      4.382857             89.0      30.680000  2009.0
4  Across the Universe    Romance    Independent             84.0      0.652603             54.0      29.367143  2007.0
...
74  The Hangover    Comedy         NaN             NaN             NaN             NaN             NaN             NaN
75  Avatar        Fantasy         NaN             NaN             NaN             NaN             NaN      2009.0
76  Black Swan        Drama         NaN             NaN             NaN             NaN             NaN      2010.0
77  Hugo           NaN         NaN             NaN             NaN             NaN             NaN      2011.0
78  The Help        Drama         NaN             NaN             NaN             NaN             NaN      2011.0
79 rows x 8 columns

[72] #Checking informations about the dataset
df.info

<bound method DataFrame.info of
0  27 Dresses    Comedy        Fox             71.0
1  (500) Days of Summer  Comedy        Fox             81.0
2  A Dangerous Method    Drama    Independent             89.0
3  A Serious Man        Drama    Universal             64.0
4  Across the Universe    Romance    Independent             84.0
...
74  The Hangover    Comedy         NaN             NaN
75  Avatar        Fantasy         NaN             NaN
76  Black Swan        Drama         NaN             NaN
77  Hugo           NaN         NaN             NaN
78  The Help        Drama         NaN             NaN
79 rows x 8 columns>
```

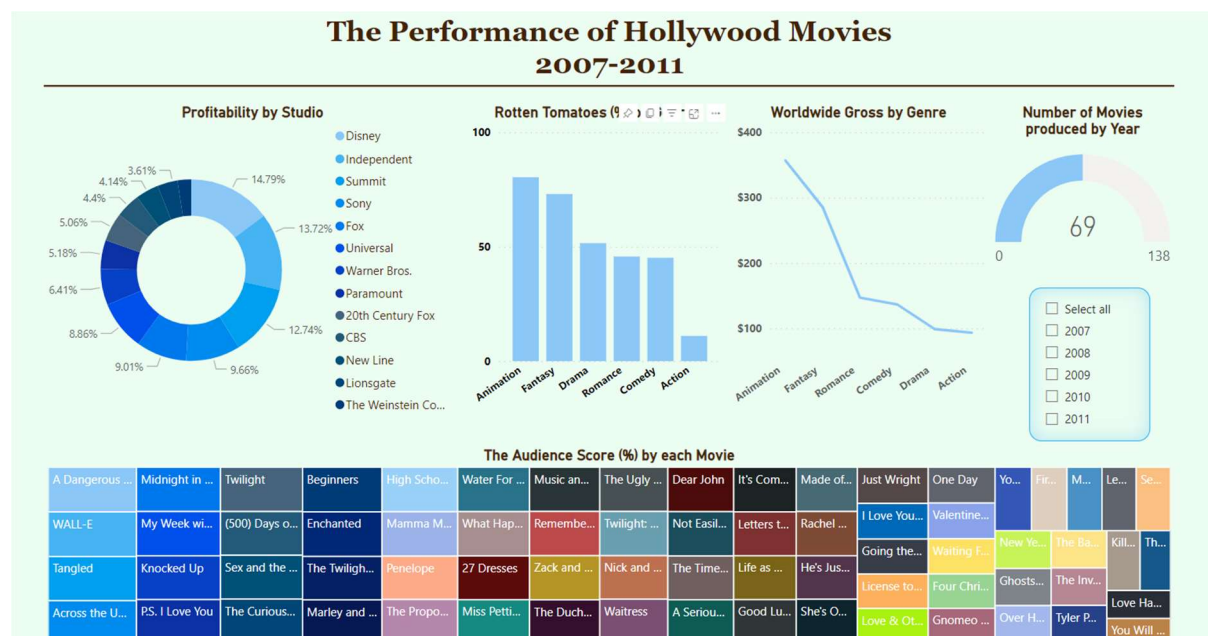
The link to this Python Script can be found here:

<https://colab.research.google.com/drive/1R2oazAH-bTT2KubpMWwbQGTqYJ66kGde?usp=sharing>

## Power BI

The subsequent step involved the creation of a Power BI Dashboard, aligning with the client's preferences and requirements. The client expressed a desire to see key metrics, including the average Rotten Tomatoes ratings by genre, the number of movies produced per year, audience scores for each film, profitability per studio, and worldwide gross per genre. The challenge was not only to meet these specifications but also to present the information in a visually appealing and insightful manner.

Throughout the creation of the Power BI Dashboard, the incorporation of the company's brand colours – blue, green, and brown – added an extra layer of customization. The use of light and dark shades of these colours allowed for a harmonious and branded presentation.



The link to the Power PI Dashboard can be found here:

[https://app.powerbi.com/links/ 3-nwvN-Sn?ctid=6efd0f20-57c8-4447-b53f-00d4992ca50b&pbi\\_source=linkShare](https://app.powerbi.com/links/ 3-nwvN-Sn?ctid=6efd0f20-57c8-4447-b53f-00d4992ca50b&pbi_source=linkShare)

## Final Considerations

Completing this assignment has been a fascinating and challenging journey into the world of Hollywood movies' performance analysis.

Bellow some Analysis Points:

- Animation and Fantasy emerge as the top-performing genres, boasting the highest audience scores and subsequently achieving greater worldwide gross revenue. However, despite its high score, the drama genre does not fare as well in the worldwide gross chart.
- When examining movie profitability by studio, there is a range of around 12%. Disney leads the pack with a profitability of 14.79%, while The Weinstein Company is the last one with only 2.44% of the total profits.
- The trend in worldwide gross revenue reveals a shift over the years, transitioning from Comedy dominating in the earlier years to Animation taking the lead in the later years (specifically in 2010 and 2011).

In conclusion, this assignment not only honed my technical skills in data analysis and visualization but also emphasized the importance of presenting findings in a meaningful and engaging way.

This experience deepened my appreciation for the dynamic and multifaceted nature of the entertainment industry, showcasing how data analytics can unlock valuable insights. The culmination of this assignment is a comprehensive Power BI Dashboard that not only meets the client's expectations but also goes beyond to deliver a visually compelling narrative of Hollywood movie performance during the specified period.

## Sources

- <https://informationisbeautiful.net/data/>
- <https://learn.microsoft.com/en-us/power-bi/create-reports/desktop-report-themes>
- Data Bootcamp Just IT materials from Weeks 3 and 5