

**TASK**

**Exploratory Data Analysis on the Movies Data Set**

[](https://www.hyperiondev.com/)

**Introduction**

Summary of the data set

The dataset consists of the following information about movies:

budget, genres, homepage, id, keywords, original language, original title, overview, popularity, production companies, production countries, release date, revenue, runtime, spoken languages, status, tagline, title, vote average, and vote count.

**DATA CLEANING**

THIS IS A SUMMARY OF THE METHODS AND VISUALISATIONS DONE DURING DATA CLEANING.

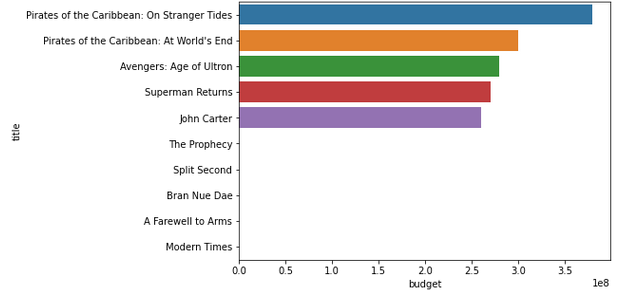
* There were some redundant columns that had to be removed from the dataset. These are the columns that were moved: 'keywords', 'original language', 'original title', 'overview', 'production companies', 'status', and 'tagline'. These columns were removed using the *drop()* function.
* There were duplicates that had to be removed and this was done using the *drop\_duplicates()* function.

**MISSING DATA**

There were rows with missing and NaN values. These were removed by replacing zeros with numpy NaN and then using the *drop\_na()* function to remove them.

**DATA STORIES AND VISUALISATIONS**

# Top 5 of the most expensive and cheapest movies



Top 5 most expensive movies

1. Pirates of the Caribbean: On Stranger Tides

2. Pirates of the Caribbean:At World's End

3. Avengers: Age of Uitron

4. Superman Returns

5. John Carter

Top 5 cheapest movies

1. The Prophecy

2. Split Second

3. Brain Nue Dae

4. A Farewell to Arms

5. Modern Times

**NB: The cheapest movies cannot be seen on the graph above because their budgets are generally less than 1% of the budgets of the most expensive movies.**

**Top 5 of the most profitable movies**

**A graph with different colored bars

Description automatically generated**

**Top 5 most profitable movies**

**1. Avatar**

**2. Titanic**

**3. Jurassic World**

**4. Furious 7**

**5. The Avengers**

**Top 5 least profitable movies**

1. **The Adventures of Pluto Nash**
2. **The 13th Warrior**
3. **Mars Needs Moms**
4. **The Alamo**
5. **the Lone Ranger**

**The Most Popular Movies**

**A graph with different colored bars

Description automatically generated with medium confidence**

**Top 5 most profitable movies**

**1. Minions**

**2. Interstellar**

**3. Deadpool**

**4. Guardians of the Galaxy**

**5. Mad Max: Fury Road**

**The Most Successful Genres**

**As it can be seen in the graph below, the drama genre has the highest number of movies and the fantasy genre has the lowest number of movies.**

**A green and yellow bar graph

Description automatically generated**

**Different Interesting Visualisations with a Data Story.**

**A group of blue and orange dotted graphs

Description automatically generated**

**The first graph shows that the movie's budget and revenue have a correlation. The venue of movies grows with the growth in the budget.**

**The second graph shows a slower positive correlation between the movie’s revenue and runtime. The longer the run time of a movie, the higher the revenue.**

**The third graph shows a very low correlation between the release date of a movie and its revenue. The latest movies do not necessarily generate a higher revenue compared to the old ones.**

**The fourth graph shows a strong positive correlation between the popularity of a movie and its revenue. The higher the movie's popularity score, the higher the revenue.**

**THIS REPORT WAS WRITTEN BY : MORONGWA THOBEJANE**

