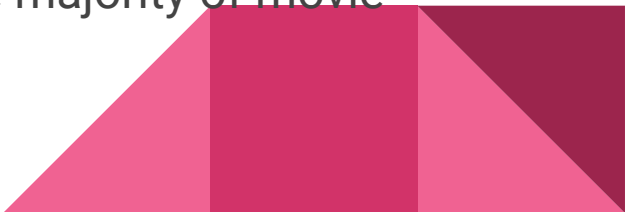# MOVIES ANALYSIS PROJECT

By George Mbugua

PROBLEM STATEMENT

To explore the types of films doing the best on the box office.

My hypothesis will be:

- The international market will be more profitable than the domestic.
- The English language is by far the most available language for movies
- There are consistent times of the month every year which yields lower profits when movies are released on, termed as 'Dump Months' and also there is a preferred length of time for movies which the audience can make time.
- There are henre types which are preferred by the majority of movie watchers.

DATA UNDERRSTANDING

The data was collected from github. There are 6 datasets provided by:

- Internet Movie Database (IMDB)
- The Movie Database (TMDB)
- Box Office Mojo (BOM)
- Rotten Tomatoes website (RT)
- The Numbers (TN)

DESCRIBING THE QUESTION

As a hired Data Scientist, I am to find the types of films which are currently doing the best and translate them into actionable insights.

DEFINING THE METRIC OF SUCCESS

Using the visualisations to show that:

- There exists the worst months to release films
- There are preferred genres.
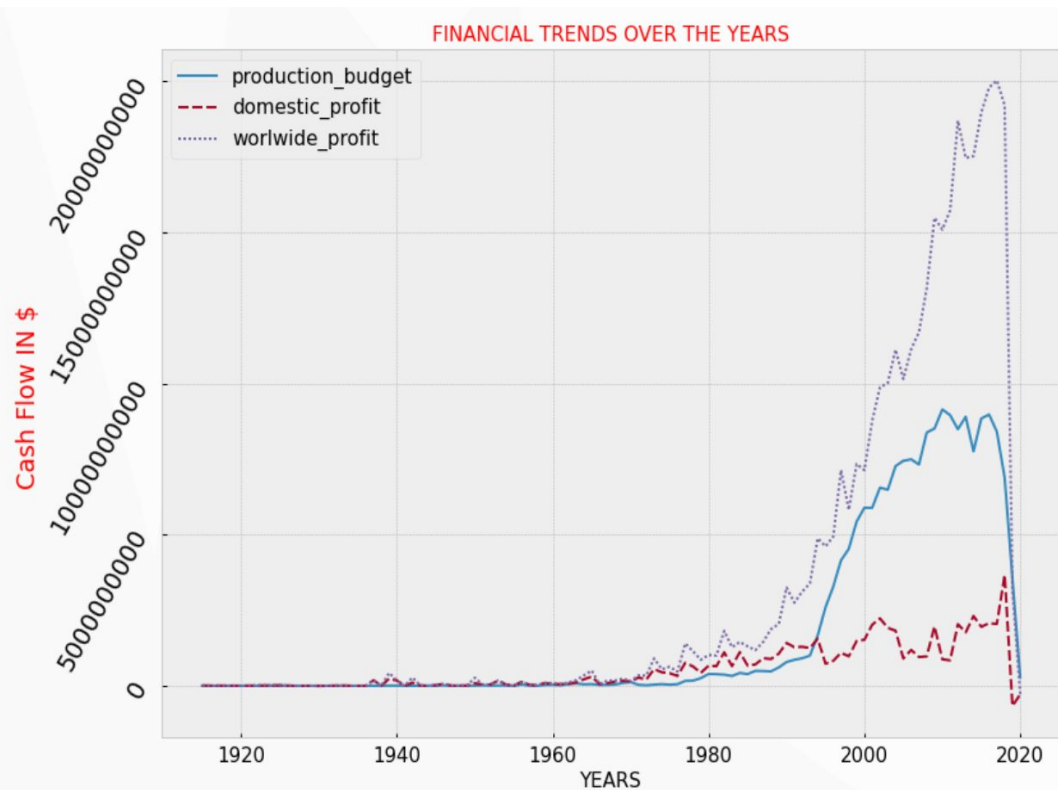- More money is to be made internationally.

DATA PREPARATION

I will select the following tables to work with:

- tn.movie_budgets.csv.gz database.
- Tmdb.movies.csv database
- Imdb.sql database

DATA VISUALIZATION

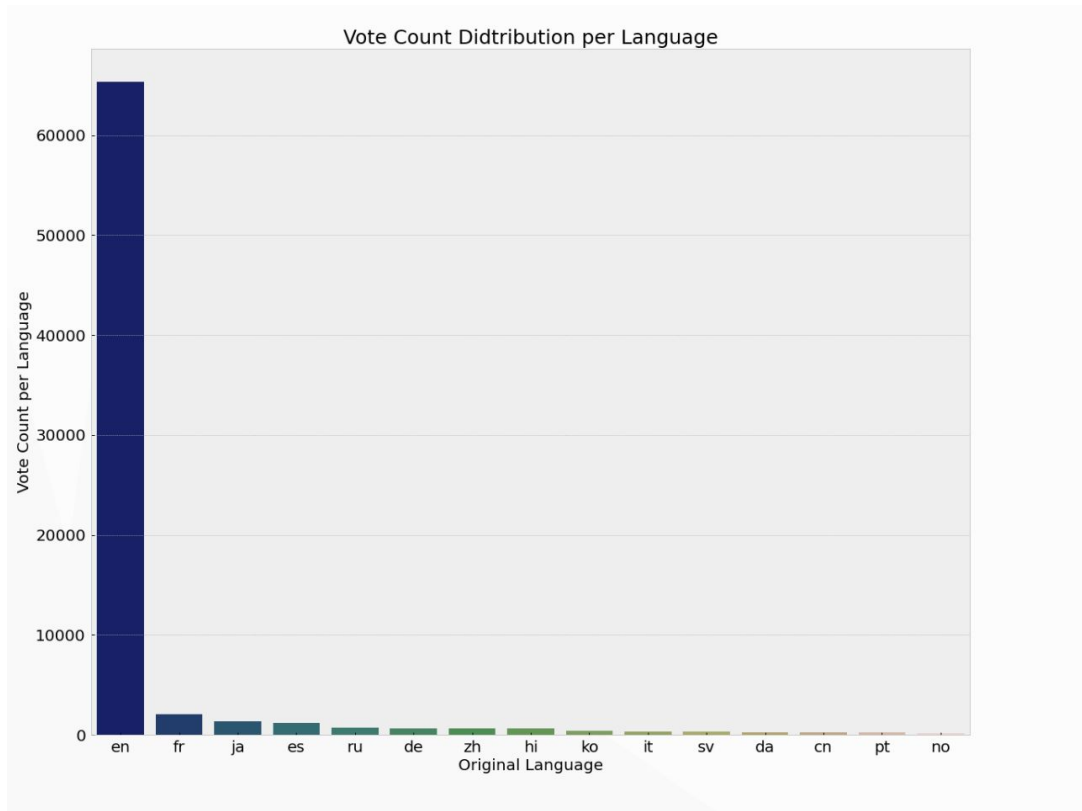FINANCIAL TRENDS OVER THE YEARS

DATA VISUALIZATION

The line plot follows the trend of cashflows based on the included movies. Since 1920's to 2020. It was grouped by year of release and summed to make the Data Frame "Budget_yearly".

It is clear that the industry has been growing steadily and shot up between 1990 and 2000, especially the worldwide_profit, basically the international audience.

The worldwide_profit is clearly more attractive than the domestic profit.
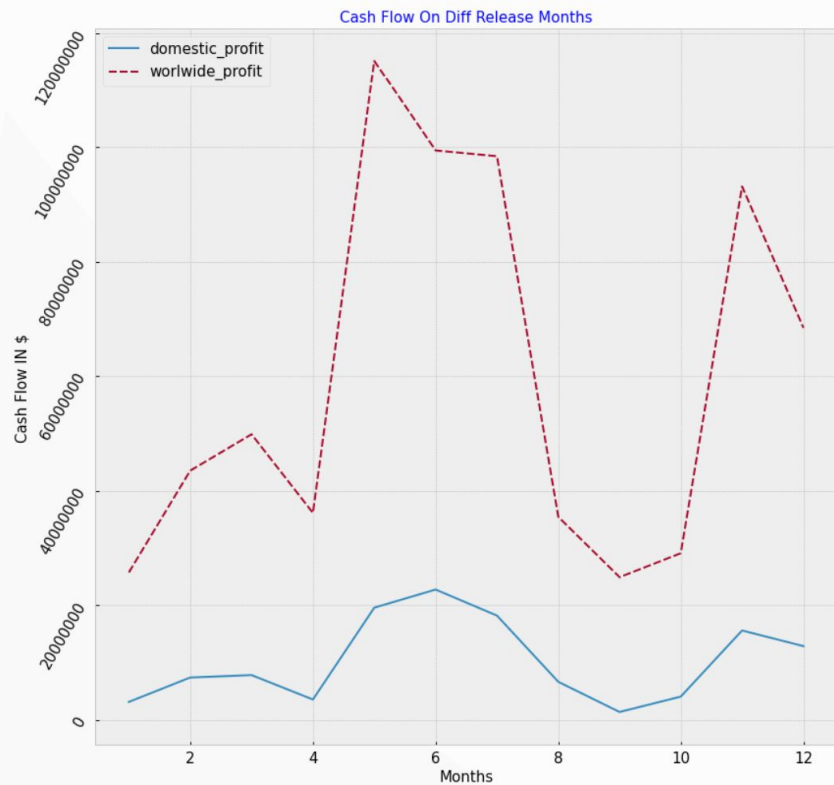
DATA VISUALISATIONS

DATA VISUALISATIONS

It is evident that English is the dominant language.
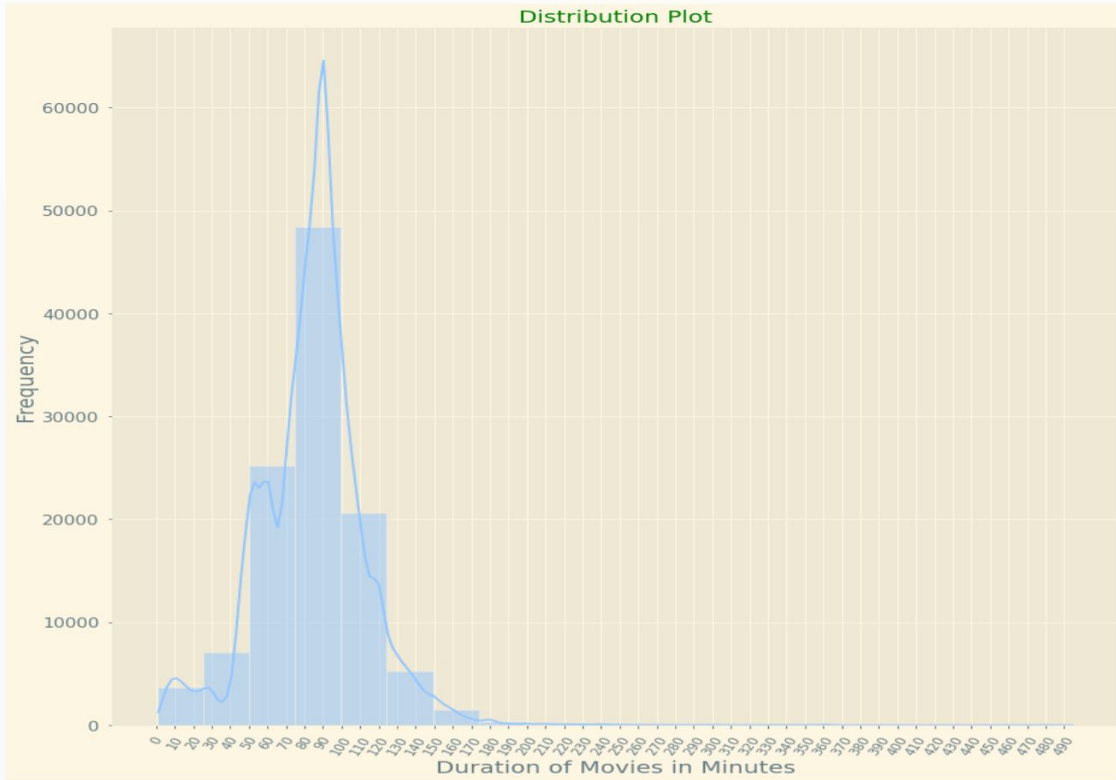
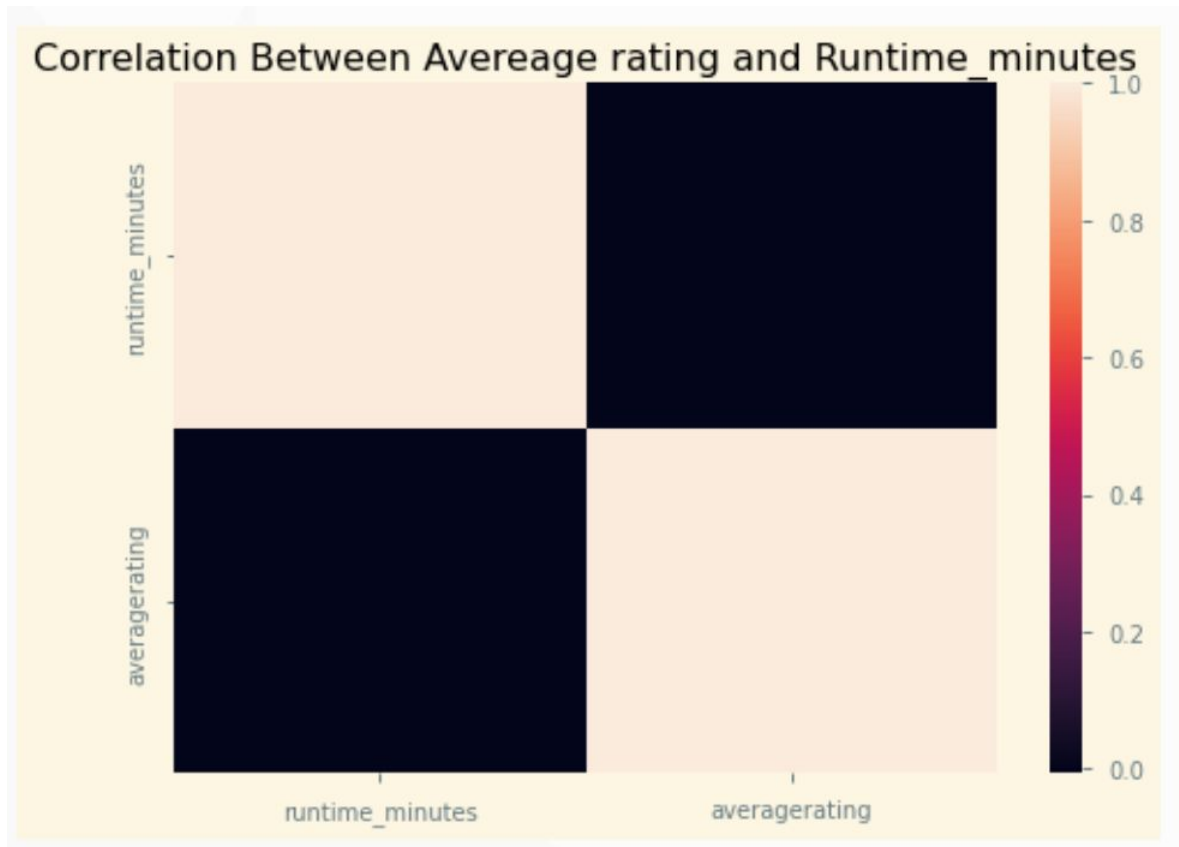# DATA VISUALISATIONS

DATA VISUALISATIONS

There are consistent times of the month every year which yields lower profits when movies are released on, termed as 'Dump Months' and also shows a goldilocks period with high profits.
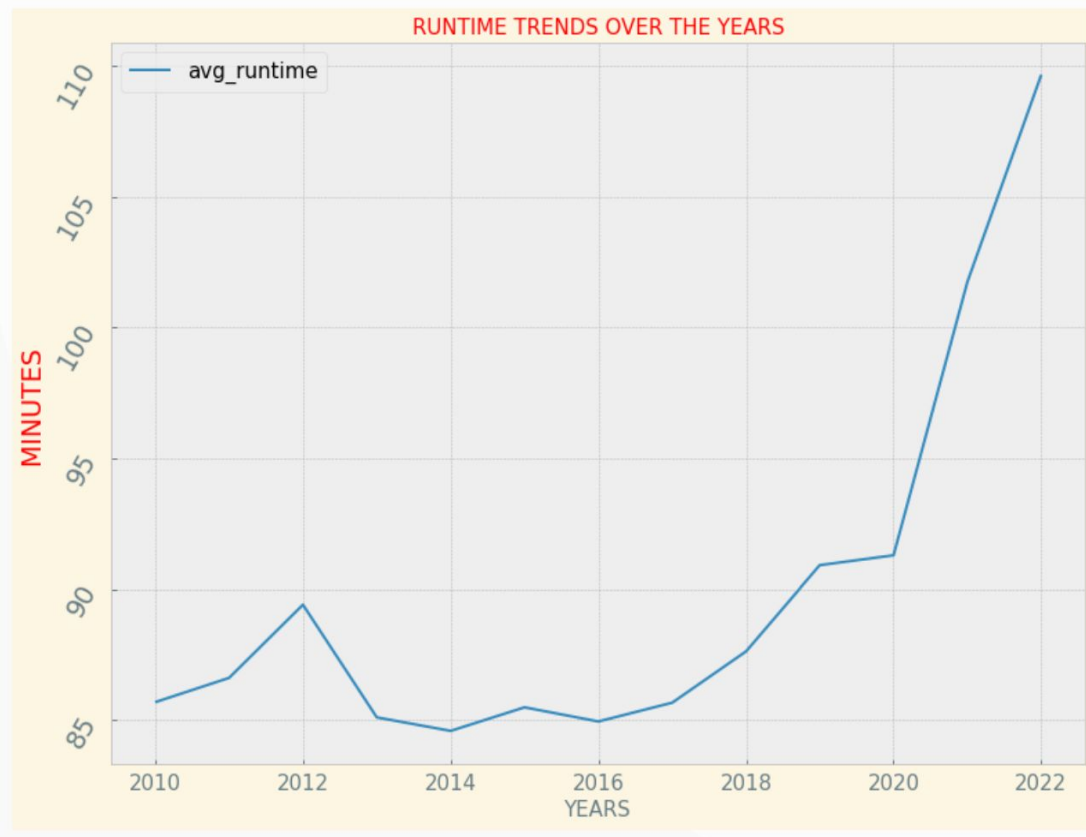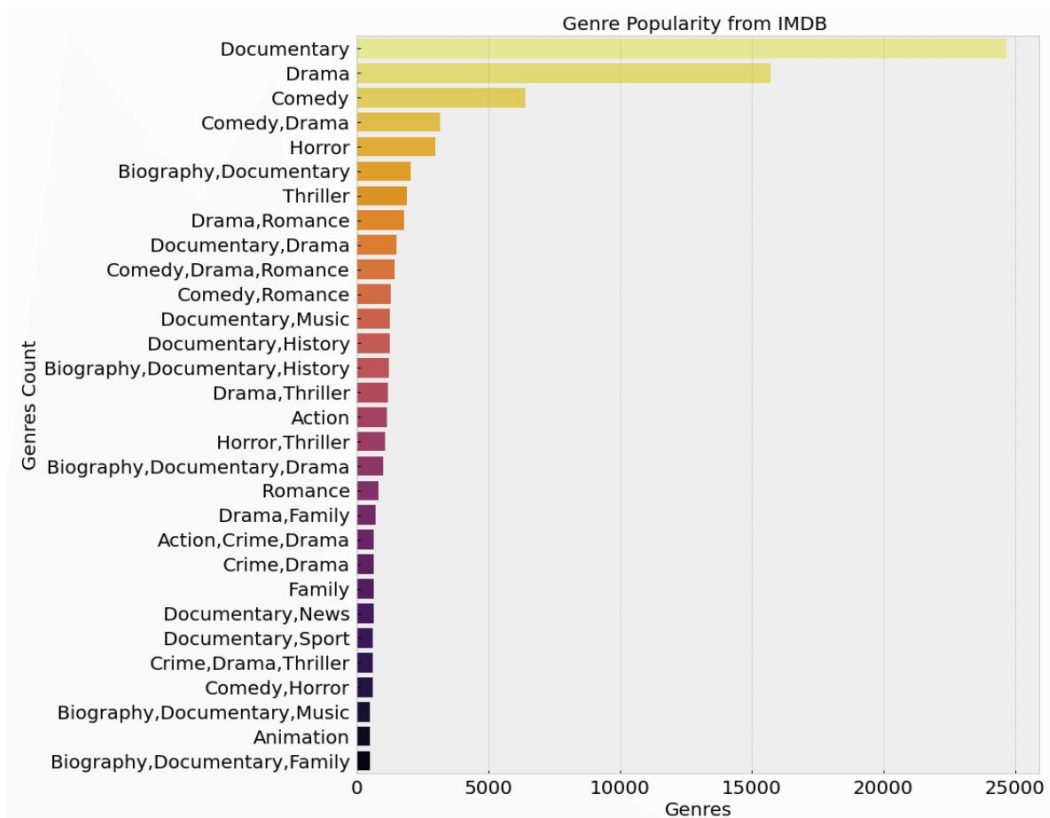
DATA VISUALISATIONS

DATA VISUALISATIONS

Correlation Between Avereage rating and Runtime_minutes

DATA VISUALISATIONS

RUNTIME TRENDS OVER THE YEARS

DATA VISUALISATIONS

DATA VISUALISATIONS

Genres Popularity from AVG ratings from IMDB