A yellow sign with black letters

Description automatically generated

C14 xander talent

IMDB MOVIE DATA ANALYSIS POWERBI PROJECT

Najib soomro

## CONTEXT (Brief)

As a film production analyst, the following business intelligence report addresses the key information the stakeholder is seeking on behalf of a Major Movie Production Studios (MMPS Ltd.) Corporate Executive Director who is interested in better comprehending cinematic trends and industrial insights into the global film industry such as global box office performances, movie ratings and popular movie genres.

The proceeding investigative data analysis methodology complements the data pipelines work conducted the previous week by further analysing the next 10 years across the second decade of the 21st century. Focusing on movie characteristics and film production using the Internet Movie Database (IMDb), I intend to provide insights into how different film production companies perform in terms of their budget and box office figures respectively to understand who key players are in this industry. Furthermore, I also intend to delve into movie analytics and explore what criteria render a particular film successful including movie ratings, genre popularity and box office earnings. A unique aspect not considered previously was the talent behind the curtains. As such, this exploratory data analysis considers who the top directors are based on global box office figures, who are the most critically acclaimed screenwriters based on average movie ratings as well as discovering who are the most in demand actors and actresses throughout this decade. This report concludes with understanding the relationship between movie runtimes and genres. In order to determine celebrity popularity, computing average movie votes across all movies for an actor/actress is favourable to host sought after cast members in a competitive market. Lastly, the frequency of movie productions across the globe is examined to understand profitability and statistically significant margins of success by country throughput and film budgets.

## DATA CLEANING & TRANSFORMATION

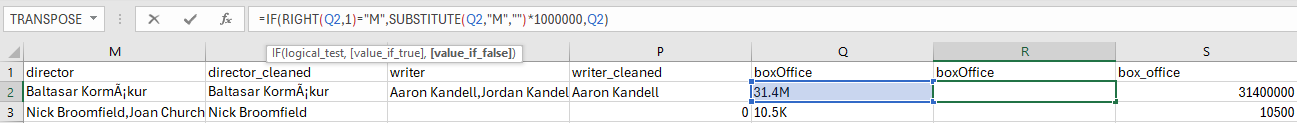
Several data cleaning and transformation steps were executed to ensure only relevant and appropriate data was obtained and within the appropriate data time range for further analysis (i.e. 2010 – 2020). This entailed ensuring uniformity amongst movie languages, movie certificate ratings and

A screenshot of a computer

Description automatically generated

## TRANSFORMING NUMBER FORMATS

As the box office figures in the original dataset contained abbreviated alphanumeric entries, converting this data into a whole integer was paramount prior to loading and transforming the data into PowerBI to perform DAX queries and data filtration processes.

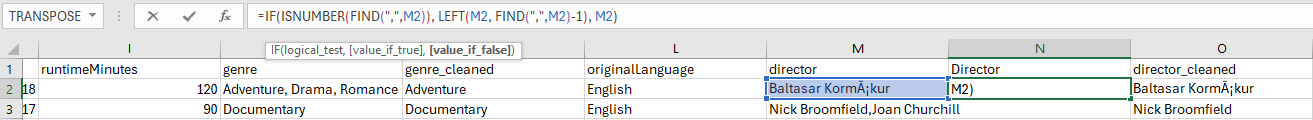


A screenshot of a data sheet

Description automatically generated

### Extracting Text Data Using LEFT and FIND functions with Delimiters

As both writers and directors’ columns contained multiple entries each separated by a comma, this delimiter was specified to extract the positional value of the first entry and extract only that value accordingly.



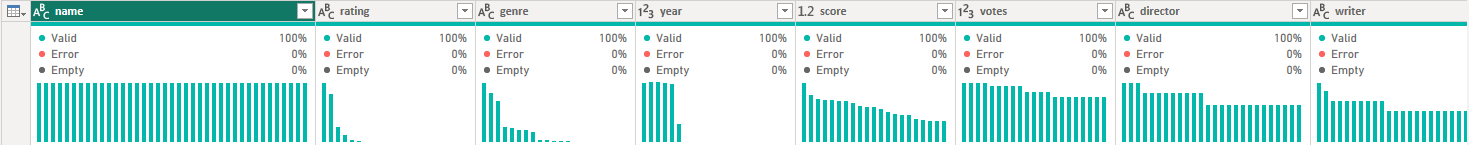
A table of names

Description automatically generated

The column quality, distribution and profile were routinely observed after executing the data cleaning and validation steps previously, ensuring no errors were present before proceeding to data analysis. These steps involved filtering each column to yield any and all ‘null’ or ‘0’ values and populating any such entries with the appropriate and relevant corresponding available data. Furthermore, any irrelevant columns not needed for analysis were made redundant.

A screenshot of a computer

Description automatically generated



After transforming the data in the IMDB dataset, the pending changes are actioned by selecting ‘Close & Apply’. This loads the cleaned data from PowerQuery into Report View wherein visualisation tools and measures can be performed on the data to compare and contrast the columns against one another and yield meaningful insights for data-driven decision-making.

A screenshot of a computer

Description automatically generated

A screenshot of a survey

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated