



# Box Office Breakdown: Analyzing Movie Industry Trends

Sathvick Sudarsan, Vineeth Ajith John, Haotian Qian & Zhuoer Liu



## INTRODUCTION

Our project delves into an extensive examination of the TMDb movie dataset, a robust repository known as The Movie Database, housing a wealth of film-related information including titles, ratings, release dates, revenue, genres, and beyond.

## DATA RETRIEVAL

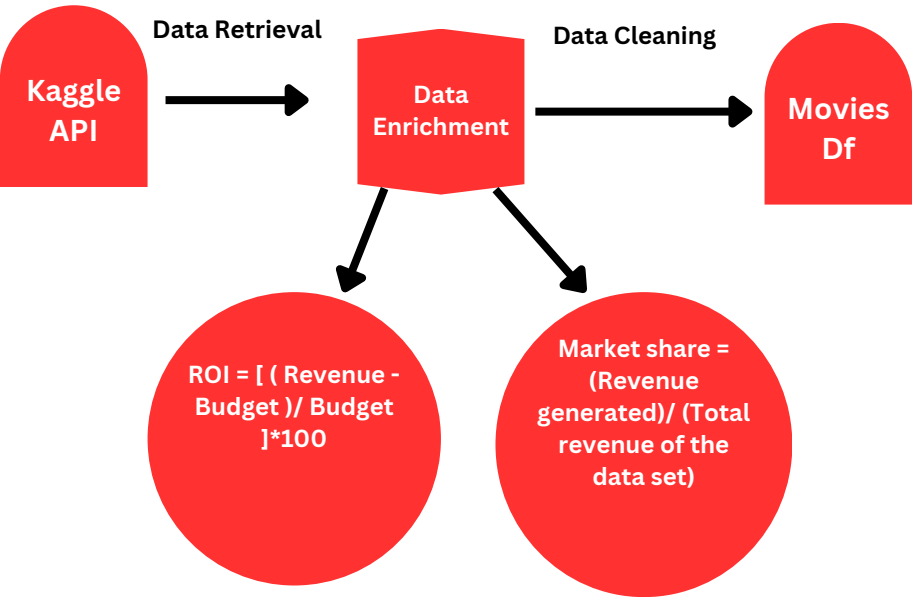
To procure the data, we utilized the Kaggle API to extract information directly from the website.



## DATA CLEANING

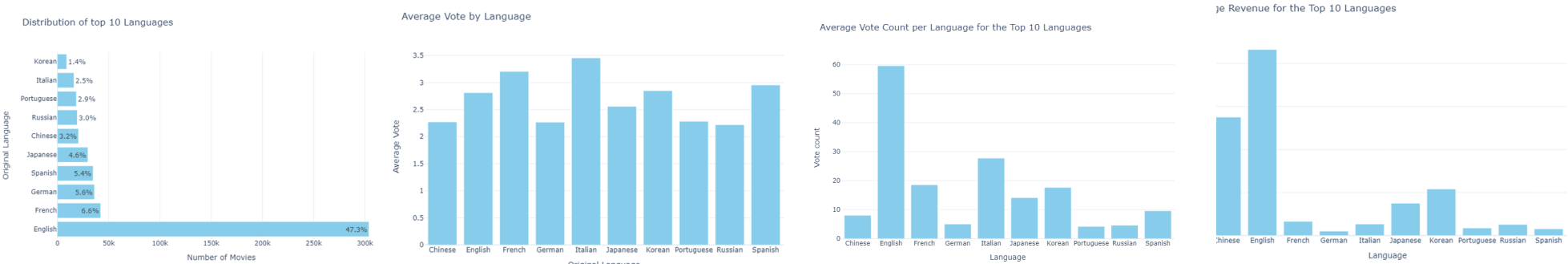
- Converted release\_date to date time format.
- Trimmed the leading and trailing spaces in title and original title
- Remove rows which don't have title.
- Removed columns with more than 50% null values
- Handled the missing values in original\_language and spoken\_languages.
- Handled different values in genres using one hot encoding.

## DATA PROCESSING

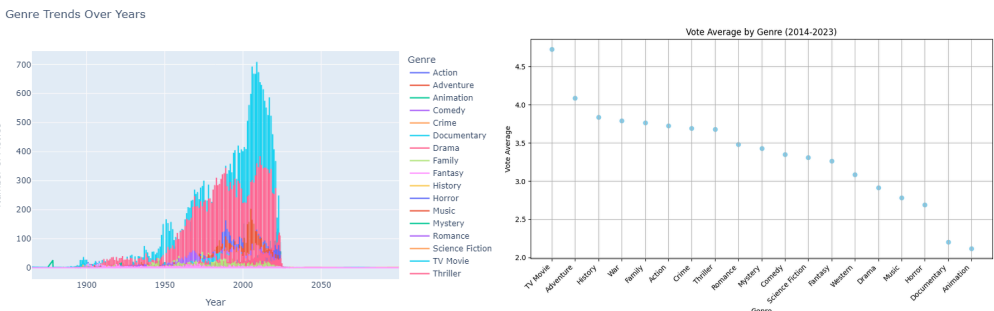


## VISUALIZATIONS

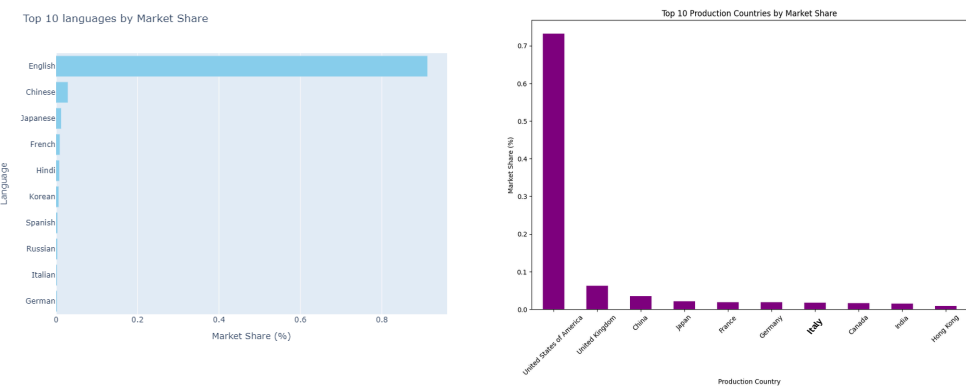
### Language Analysis



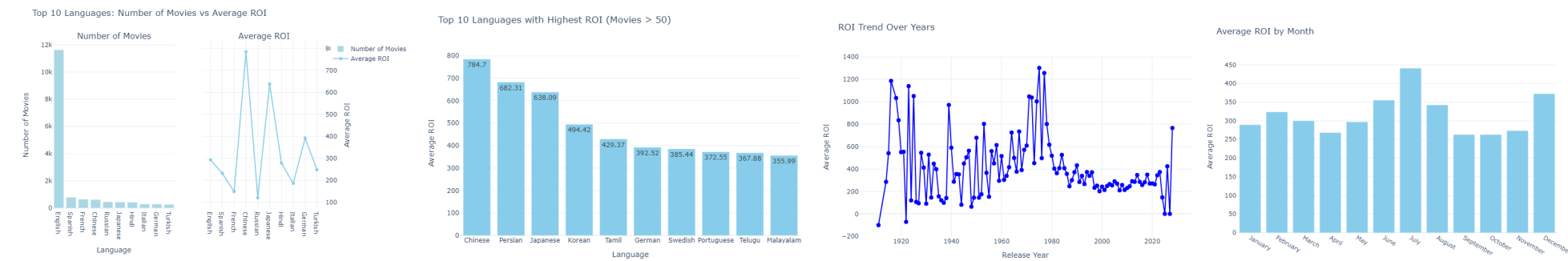
### Genre Analysis



### Market Share Analysis



### ROI Analysis



## CONCLUSION

- The top 10 languages collectively account for 82.5% of the total movies within the dataset.
- The top 10 languages, constituting approximately 1.5% of the overall market, are predominantly represented by the USA, UK, and China as the primary producing nations.
- Drama, with over 200,000 entries, emerges as the most prevalent genre, closely followed by documentary and comedy.
- Analysis of trend charts reveals that drama, documentary, and comedy maintain their dominance as the three most frequent categories over the past decade, while adventure, action, and thriller genres emerge as the most popular during this period.

## COMMIT HISTORY

