INT217 PROJECT REPORT

(Project Semester August-December 2020)

IMBD DASHBOARD

Submitted by

Shreedhar Niskawade

Registration No: 11808727

Programme and Section: CSE(Hons) KM074

Course Code INT217

Under the Guidance of

Komal Arora: 17783

Discipline of CSE/IT

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

DECLARATION

I, Shreedhar Niskawade student of Computer Science and Engineering under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 01-12-2020 Signature

Registration No. 11808727 **Shreedhar Niskawade**

CERTIFICATE

This is to certify that Shreedhar Niskawade Bearing Registration no. 11808727 has completed INT217 project titled, "IMBD DASHBOARD" under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Signature and Name of the Supervisor Ms. Komal Arora
Designation of the Supervisor Assistant Professor
School of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab.

ACKNOWLEDGEMENT

I would like to express my special thanks of gratitude to my teacher Ms. Komal Arora who gave me the golden opportunity to do this wonderful project on the topic superstore dashboard, which also helped me in doing a lot of Research and I came to know about so many new things I am really thankful to them.

Secondly, I would also like to thank my parents and friends who helped me a lot in finalizing this project within the limited time frame.

TABLE OF CONTENTS

INTRODUCTION

The extraction and cleansing of data is the critical step for any organization to extract useless analysis from the data to provide for the future measures and goals.

This dataset gives the deep structural analysis about movies records maintained by IMBD around the globe.
The dataset was pre-assembled on the Kaggle for open analysis, but it still the dataset was not normalized.

Content

This data provides the Movies detail from year 1902 to 2020 around the world.

- Data were measured every day between 1902 and 2020.
- Data were measured for all movies releasing countries associated and credited by IMBD.
- This dataset was cleaned and transformed using tableau prep.

SOURCE OF DATASET

Source Website: Kaggle.com
Dataset Name: MoviesOnStreamingPlatforms_updated.csv
Dataset Owner: https://www.kaggle.com/ruchi798
Dataset Link: https://www.kaggle.com/ruchi798/movies-on-netflix-prime-video-hulu-and-disney
Dataset Type: Arts and Entertainment

OBJECTIVES

Scope of the Analysis for the Movies dataset has been made in consideration with ratings, platform available, runtime, language, genre and PG rating.

Objective 1: Top 10 Popular Movies

- **Scope:** Movies, Release year, ratings, language, platform, country
- **Domain:** 1902 - 2020

Objective 2: Quick stats and Map

- **Scope:** Best Platform, Best Genre, Total Runtime, Total Releases, Most Releases, Average IMBD Rating
- **Domain:** 1902-2020

Objective 3: Top 5 PG Movies

- **Scope:** Movies and PG rating
- **Domain:** 1902-2020

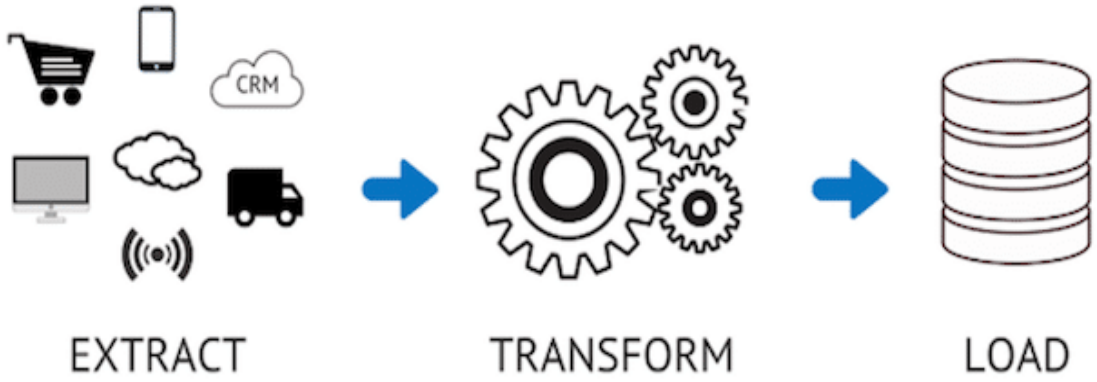
Objective 4: Radar Analysis of IMBD ratings

- **Scope:** IMBD ratings
- **Domain:** Year

Objective 5: Language vs IMBD and Release Count

- **Scope:** Sales, Profit, Quantity, Discount, Region
- **Domain:** Language

ETL PROCESS



Software: Tableau prep

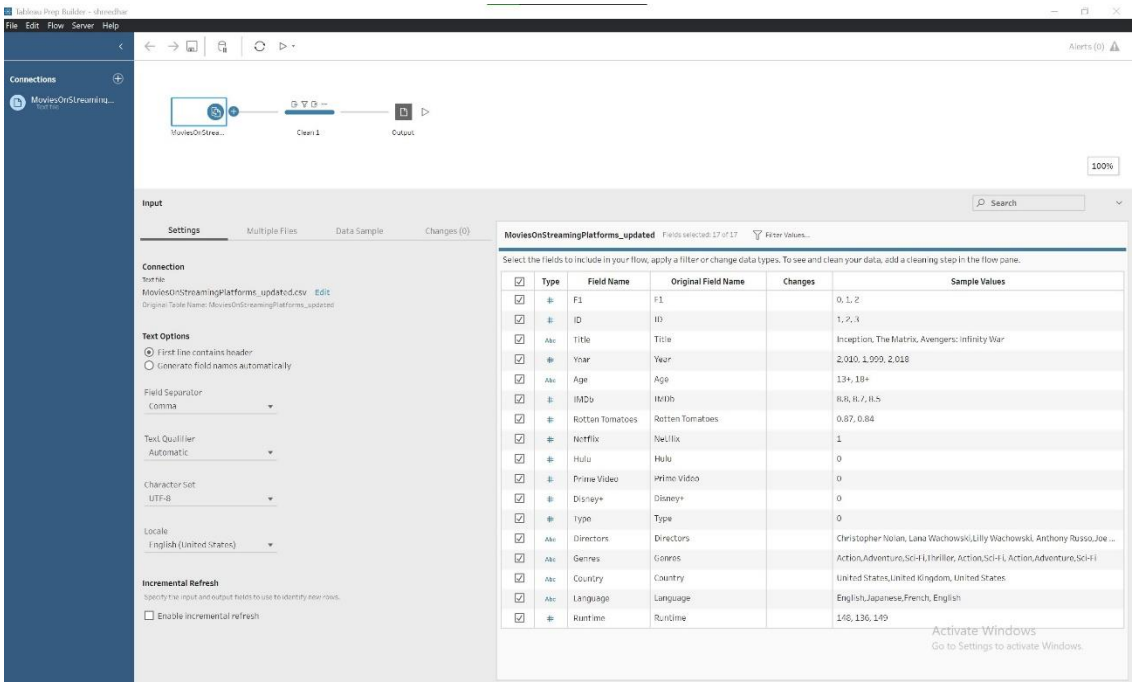
Dataset Metadata:

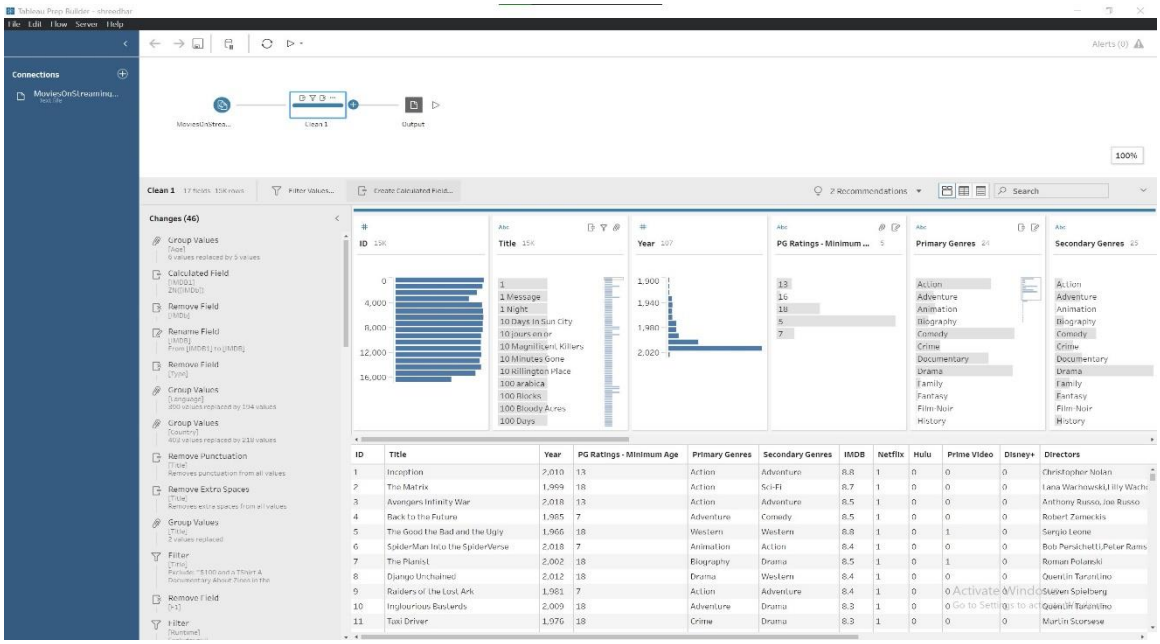
- **Type:** CSV, Medium
- **Size:** 2 MB
- **Rows/Column:** 15k/17

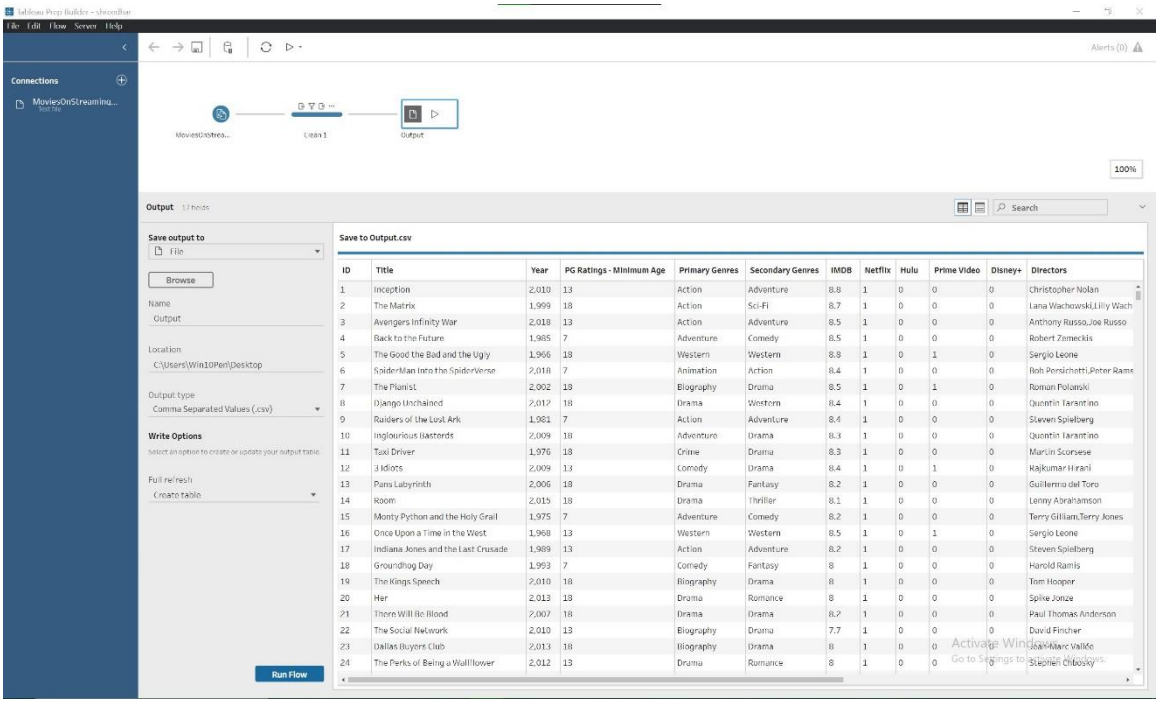
Process:

- Filtering, Grouping. Remove, Renaming, Excluding
- Clean – Punctuation, Commas, Extra Spaces
- Null – Zn(), IFNULL()

Steps:







Flow:

QUICK OVERVIEW

OBJECTIVE 1

General Description: Provide the Quick stats of the IMBD dataset

Specific Requirements:

- Map Chart (Internet Required)
- Slicer – For controlling the Year period

Functions and Formulas:

- Inter-workbook reference: =Sheet_name!Cell_address

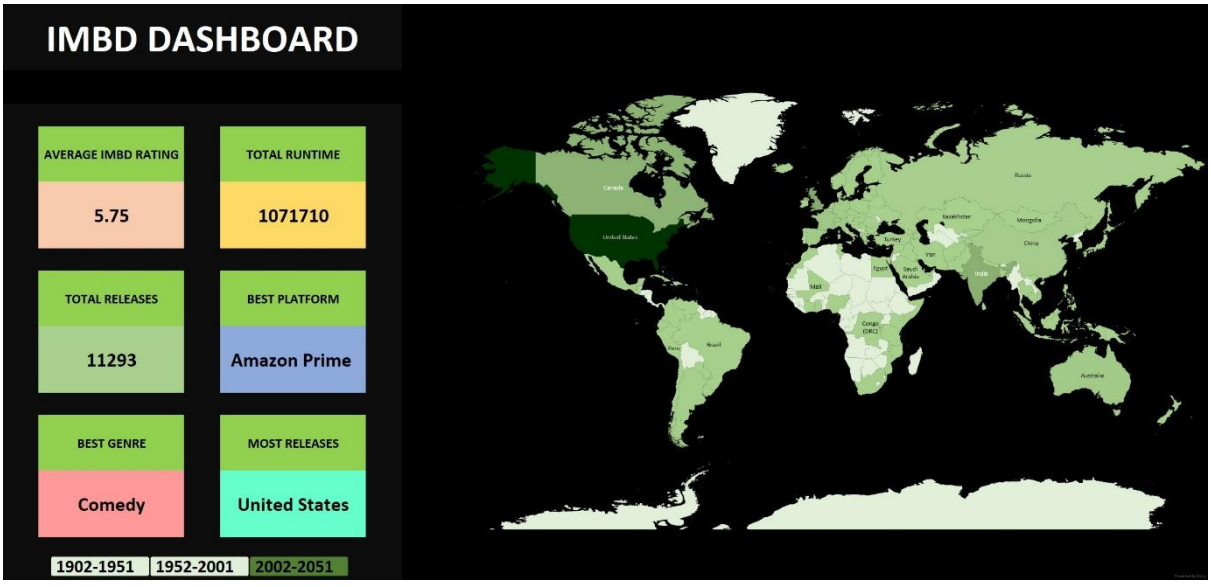
Analysis result:

Total Runtime	Total Release	Average of IMDB	Netflix	Hulu	Amazon Prime	Disney+
1071710	11293	5.75	2938	754	7785	294

Row Labels	Count of Title
Comedy	2394

Row Labels	Count of Title
United States	6225
United Kingdom	947
India	832
Canada	731
France	265
Australia	183
Germany	181
Spain	164
Italy	136
China	119
South Korea	118
Japan	104
Turkey	65

Visualization:



OBJECTIVE 2

Top 10 Popular Movies

General Description: Provide details for the Top 10 popular movies.

Specific Requirements:

- Slicer - Movies.

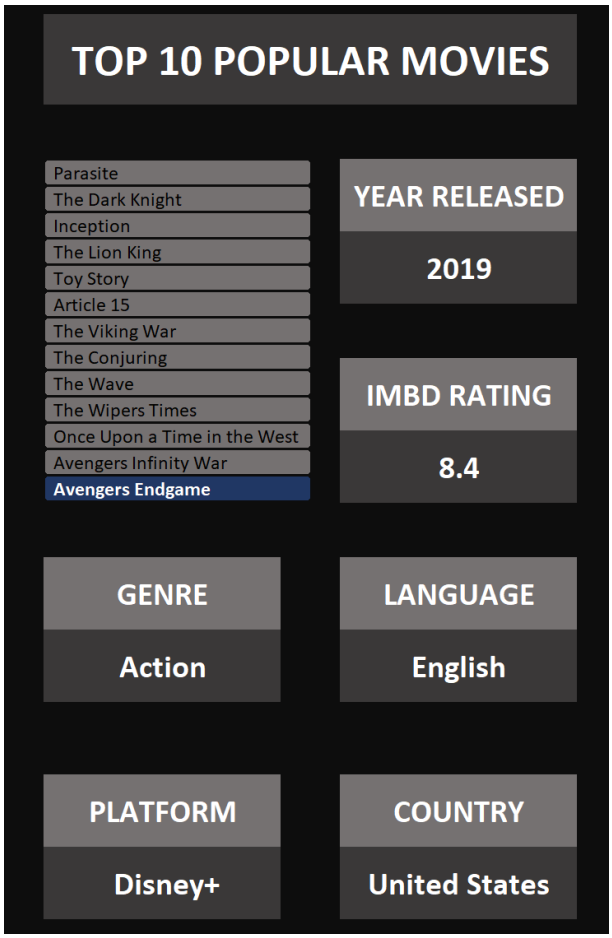
Functions and Formulas:

- Filtering – Movies

Analysis Results:

Title	Primary Genres	Primary Language	Major Release	Max of Year	IMDB	Netflix	Hulu	Prime Video	Disney+
Avengers Endgame	Action	English	United States	2019	8.4	0	0	0	1

Visualization:



OBJECTIVE 3

Top 5 PG Movies

General Description: Provide Top 5 PG rated movies in Category and Platform Available on.

Specific Requirements:

- Slicer – PG ratings

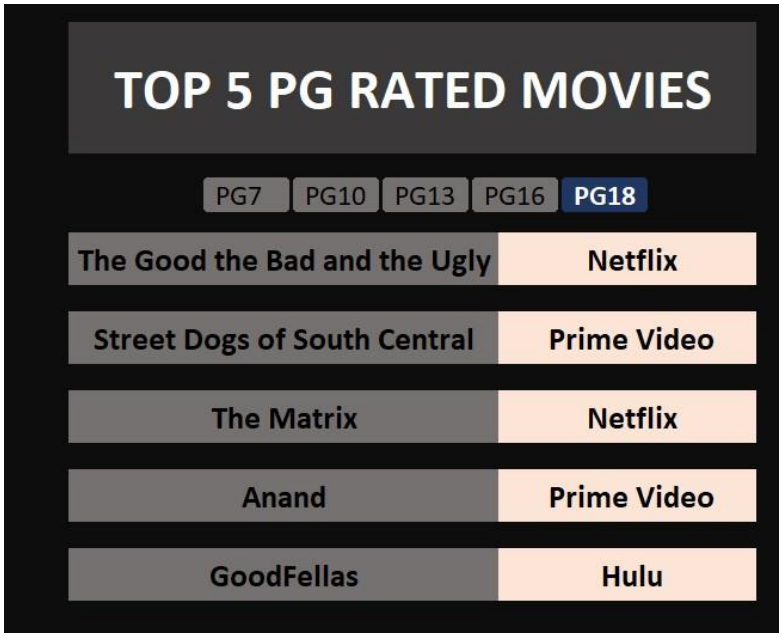
Functions and Formulas:

- Filtering by Max IMBD ratings and platform, title.

Analysis Results:

PG Ratings - Minimum Age	Title	Max of IMDB	Netflix	Hulu	Prime Video	Disney+
PG18	The Good the Bad and the Ugly	8.8	1	0	1	0
	Street Dogs of South Central	8.8	0	0	1	0
	The Matrix	8.7	1	0	0	0
	Anand	8.7	0	0	1	0
	GoodFellas	8.7	0	1	0	0
	The Green Mile	8.6	0	1	0	0
	Louis CK Shameless	8.6	0	0	1	0
	Parasite	8.6	0	1	1	0
	George Carlin You Are All Diseased	8.6	0	0	1	0
	Gol Maal	8.6	1	0	1	0

Visualization:



OBJECTIVE 4

Radar Analysis of IMBD ratings

General Description: Provide the overall reach value for the 15-year group span for IMBD ratings.

Specific Requirements:

- Radar chart

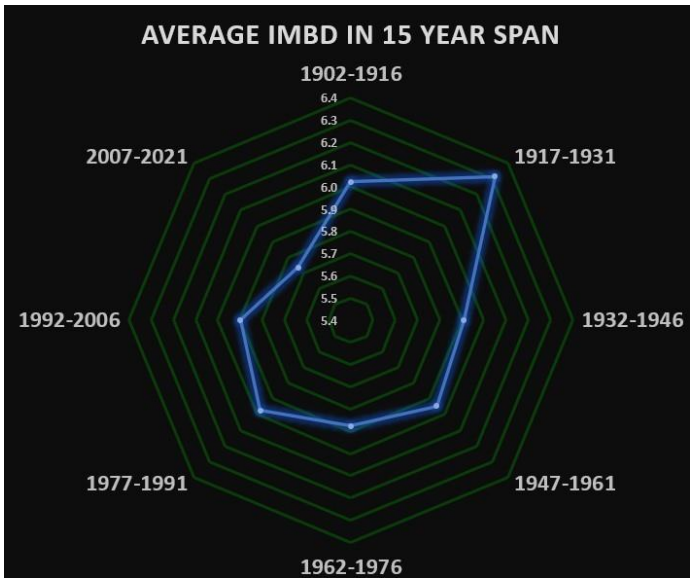
Functions and Formulas:

- Filtering IMBD ratings based on Year
- Date grouping by year

Analysis Results:

Row Labels	Average of IMDB
1902-1916	6.0
1917-1931	6.3
1932-1946	5.9
1947-1961	5.9
1962-1976	5.9
1977-1991	6.0
1992-2006	5.9
2007-2021	5.7

Visualization:



OBJECTIVE 5

Language vs IMBD and Release Count

General Description: Provide comparative analysis between language, IMBD ratings and Release count.

Specific Requirements:

- Custom chart - Primary axis: Release Count :: Secondary axis: IMBD ratings

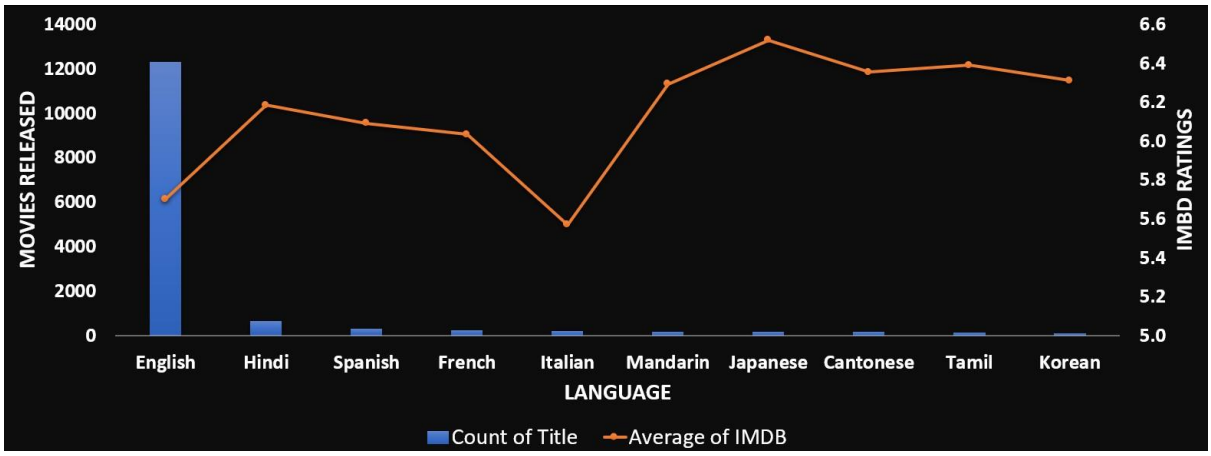
Functions and Formulas:

- Filtering Based on Language

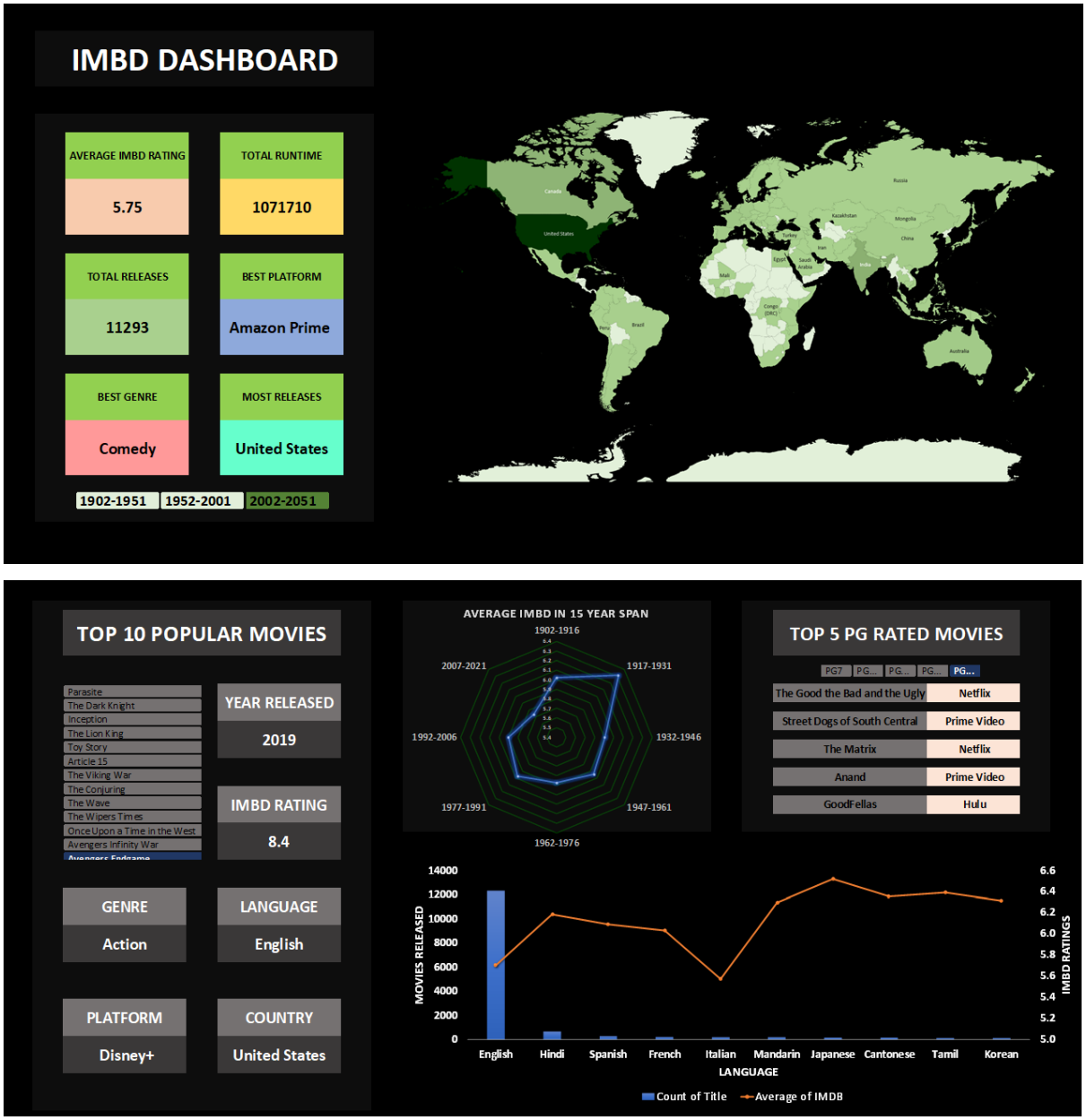
Analysis Results:

Language	Count of Title	Average of IMDB
English	12311	5.7
Hindi	653	6.2
Spanish	290	6.1
French	225	6.0
Italian	185	5.6
Mandarin	176	6.3
Japanese	154	6.5
Cantonese	152	6.4
Tamil	122	6.4
Korean	109	6.3

Visualization:



DASHBOARD VISUALIZATION



REFERENCE

- Exceljet.net
- Stackoverflow.com
- Kaggle.com
- Easy-Excel.com

BIBLIOGRAPHY

- Microsoft help
- Excel Community board
- Tableau reference Dashboard