MOVIE GENRE PREDICTION AND RECOMMENDER SYSTEM

DSI18 CAPSTONE PROJECT BY RICHA MARWAH

AGENDA

01. MISSION PROJECT

Background and Problem Statement

02. DATA SPOTLIGHT

Data Cleaning and Exploratory Analysis

03. MODELING MADNESS

Modeling and Evaluation

04. THE FINALE

Conclusion

MISSION PROJECT

Background and Problem Statement

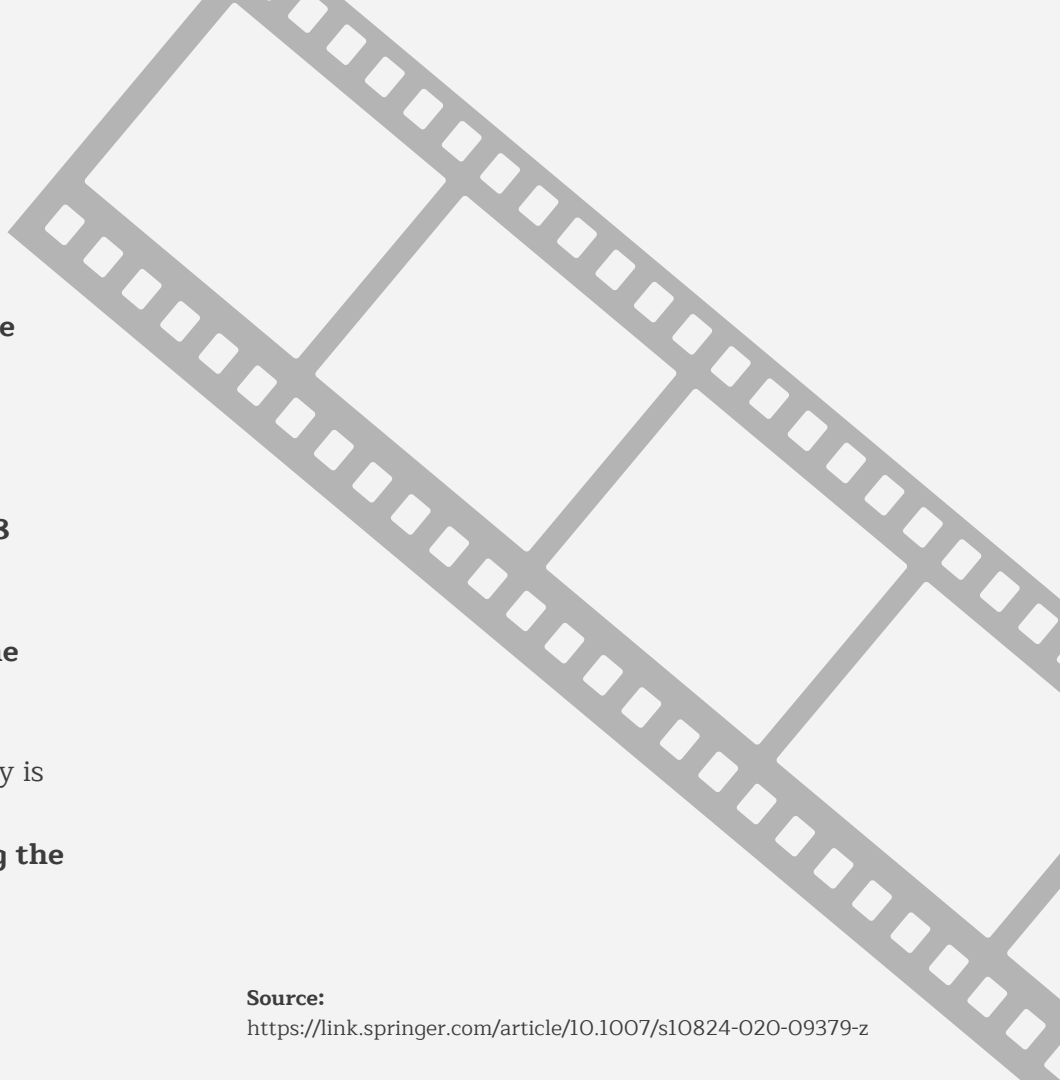
BACKGROUND

There has been a **shift from theatrical to home entertainment**, led by the growth of online streaming services

According to an industry report*, the home entertainment market has grown from **USD 15.8 billion in 2014** to **USD 42.6 billion in 2018**

This change has been further **accelerated by the Covid-19 pandemic**

The competition in the online streaming industry is intensifying, with **greater investments from established players and new players entering the market**

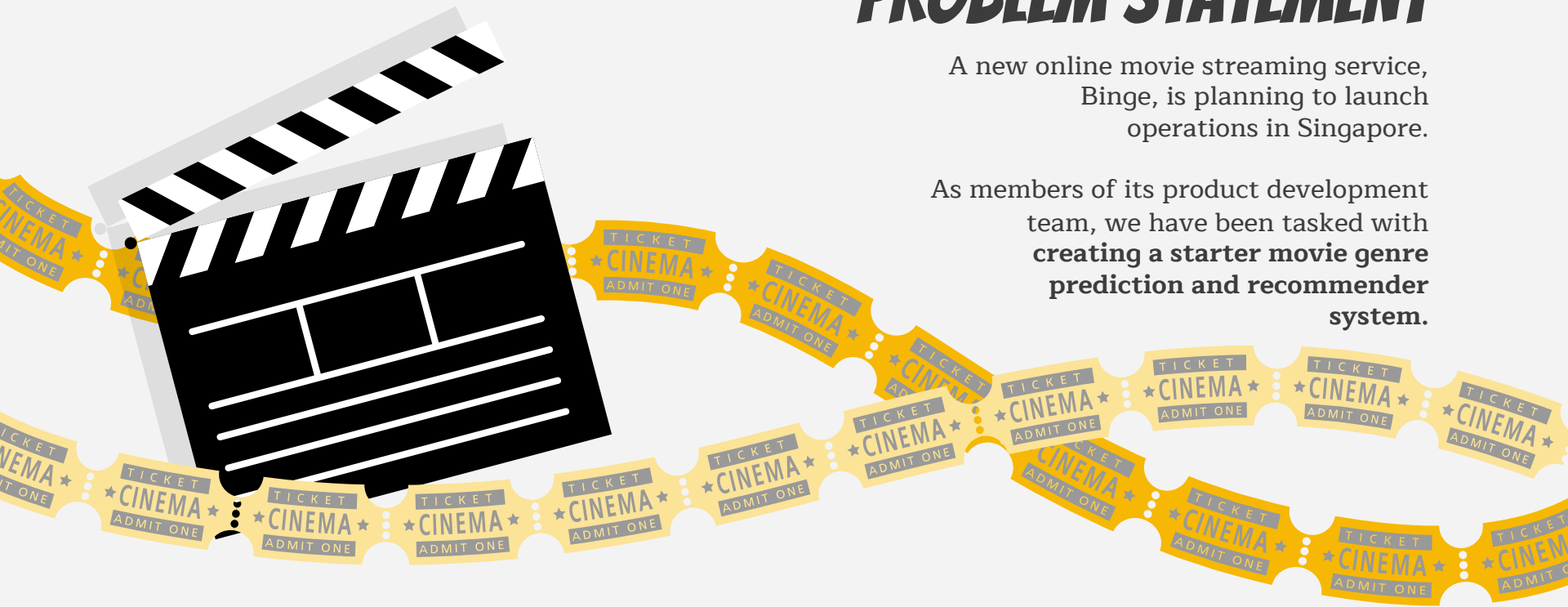


Source:
https://link.springer.com/article/10.1007/s10824-020-09379-z

PROBLEM STATEMENT

A new online movie streaming service, Binge, is planning to launch operations in Singapore.

As members of its product development team, we have been tasked with **creating a starter movie genre prediction and recommender system.**



DATA SPOTLIGHT

Data Cleaning and Exploratory Analysis

DATA SET

IMDB Title ID

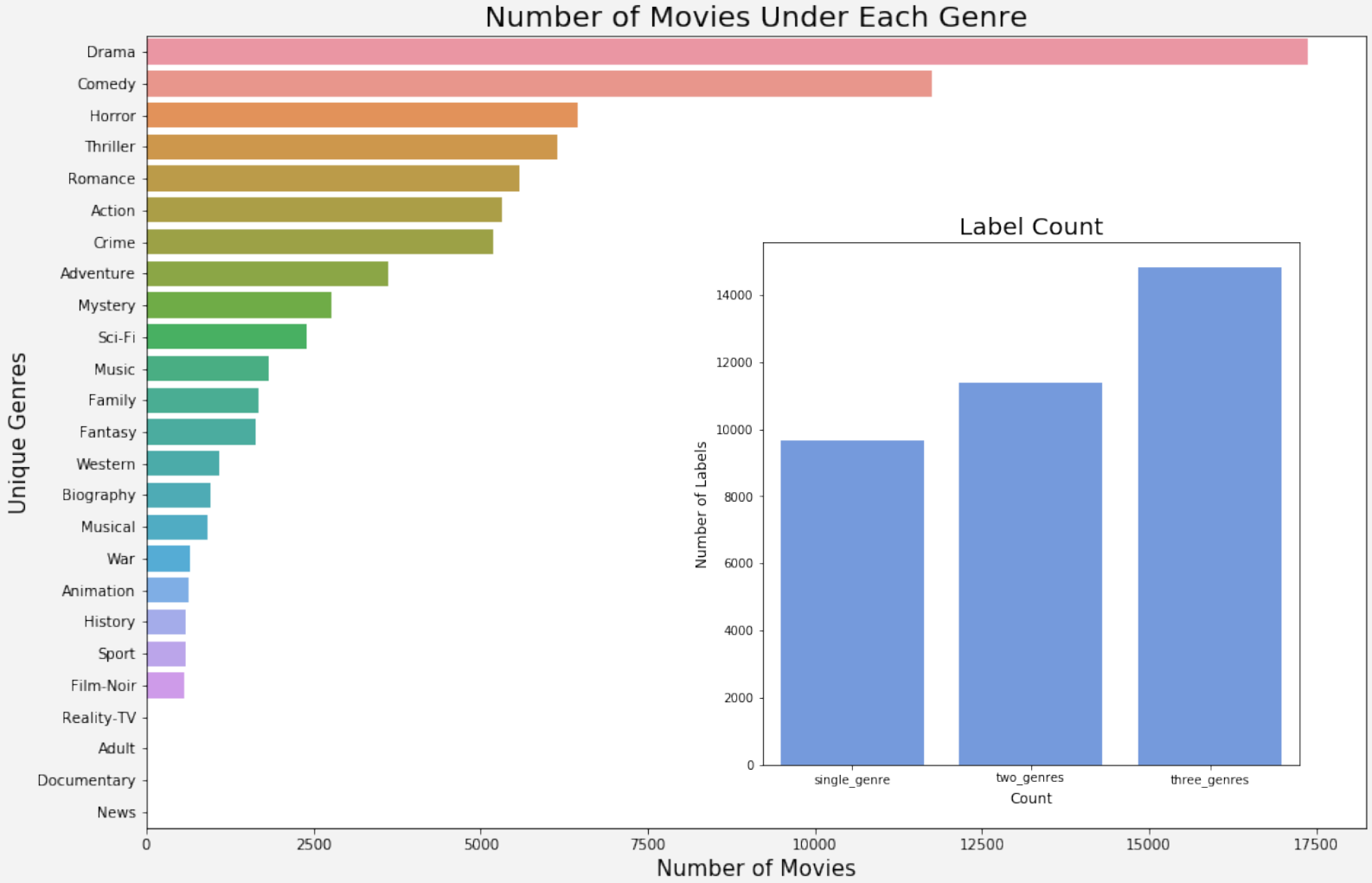
Title/ Original Title

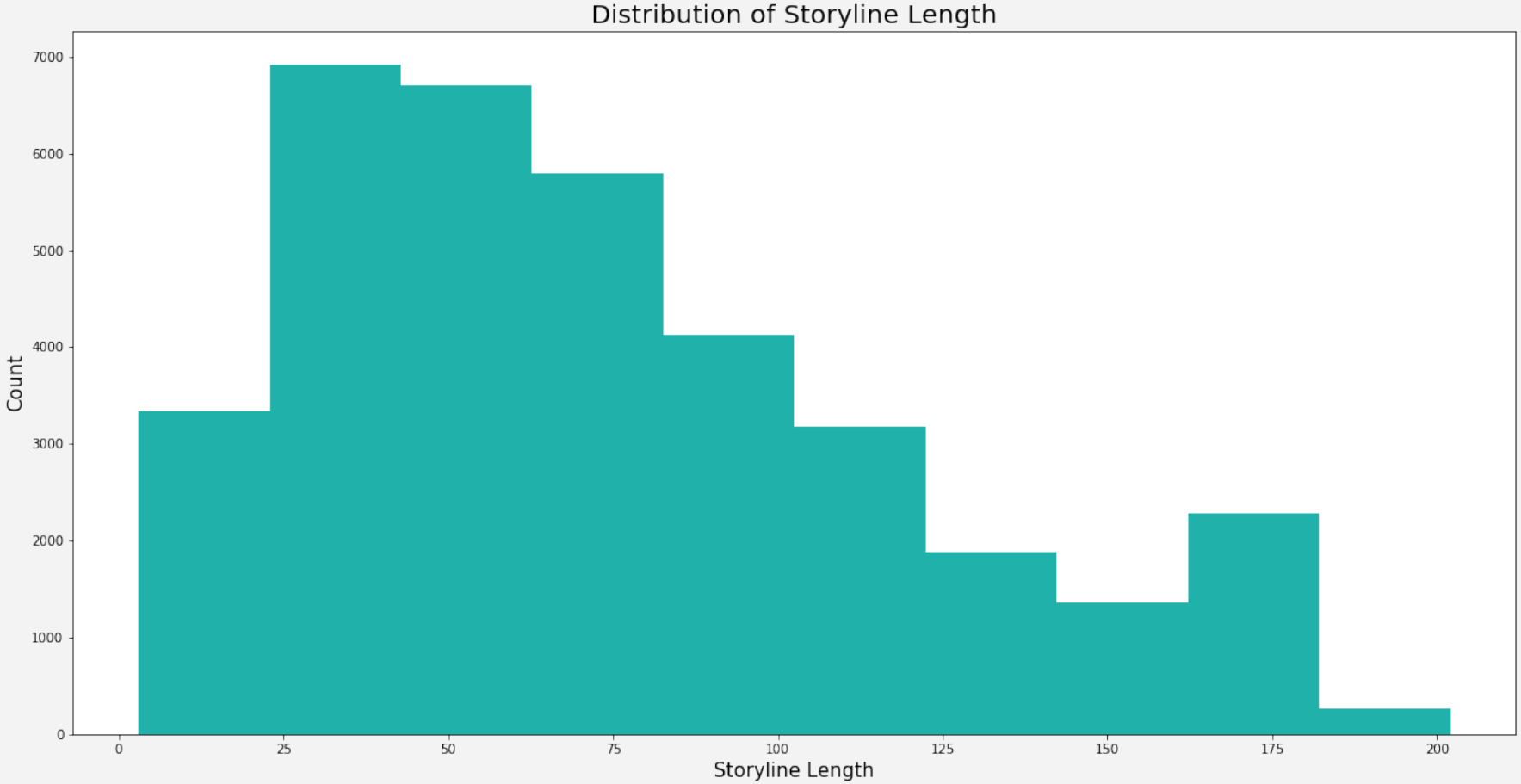
Genre

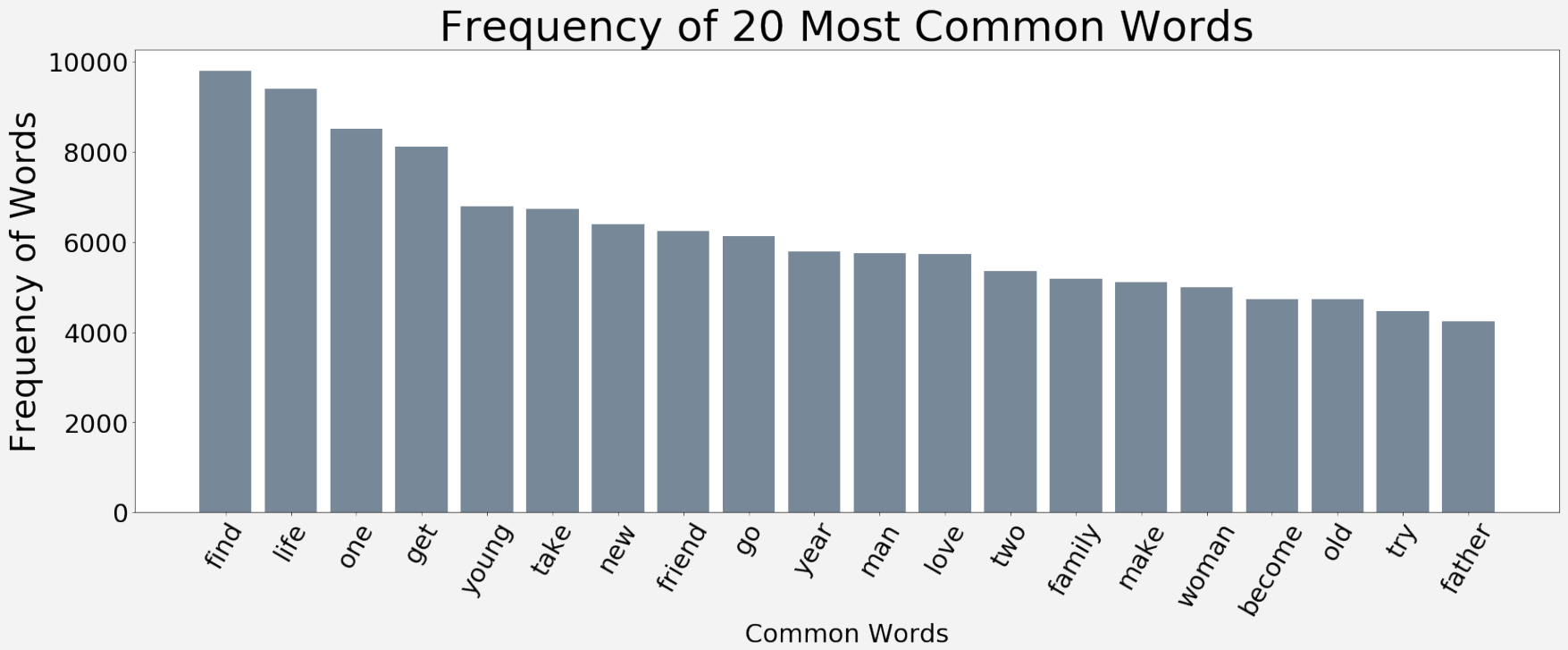
Description/ Storyline

	imdb_title_id	title	original_title	year	genre	description
0	tt0002101	Cleopatra	Cleopatra	1912	Drama, History	The fabled queen of Egypt's affair with Roman ...
1	tt0002199	From the Manger to the Cross; or, Jesus of Naz...	From the Manger to the Cross; or, Jesus of Naz...	1912	Biography, Drama	An account of the life of Jesus Christ, based ...
2	tt0002461	Richard III	Richard III	1912	Drama	Richard of Gloucester uses manipulation and mu...
3	tt0003167	Amore di madre	Home, Sweet Home	1914	Drama	John Howard Payne at his most miserable point ...
4	tt0003471	Traffic in Souls	Traffic in Souls	1913	Crime, Drama	A woman, with the aid of her police officer sw...

	imdb_title_id	title	original_title	year	genre	description	storyline
0	tt0002101	Cleopatra	Cleopatra	1912	Drama, History	The fabled queen of Egypt's affair with Roman ...	When she discovers that a slave named Phar...
1	tt0002199	From the Manger to the Cross; or, Jesus of Naz...	From the Manger to the Cross; or, Jesus of Naz...	1912	Biography, Drama	An account of the life of Jesus Christ, based ...	An account of the life of Jesus Christ, ba...
2	tt0002461	Richard III	Richard III	1912	Drama	Richard of Gloucester uses manipulation and mu...	Shakespeare's tragedy of the hump-backed D...
3	tt0003167	Amore di madre	Home, Sweet Home	1914	Drama	John Howard Payne at his most miserable point ...	Over the protests of his mother and sweeth...
4	tt0003471	Traffic in Souls	Traffic in Souls	1913	Crime, Drama	A woman, with the aid of her police officer sw...	An early social commentary on the New York...

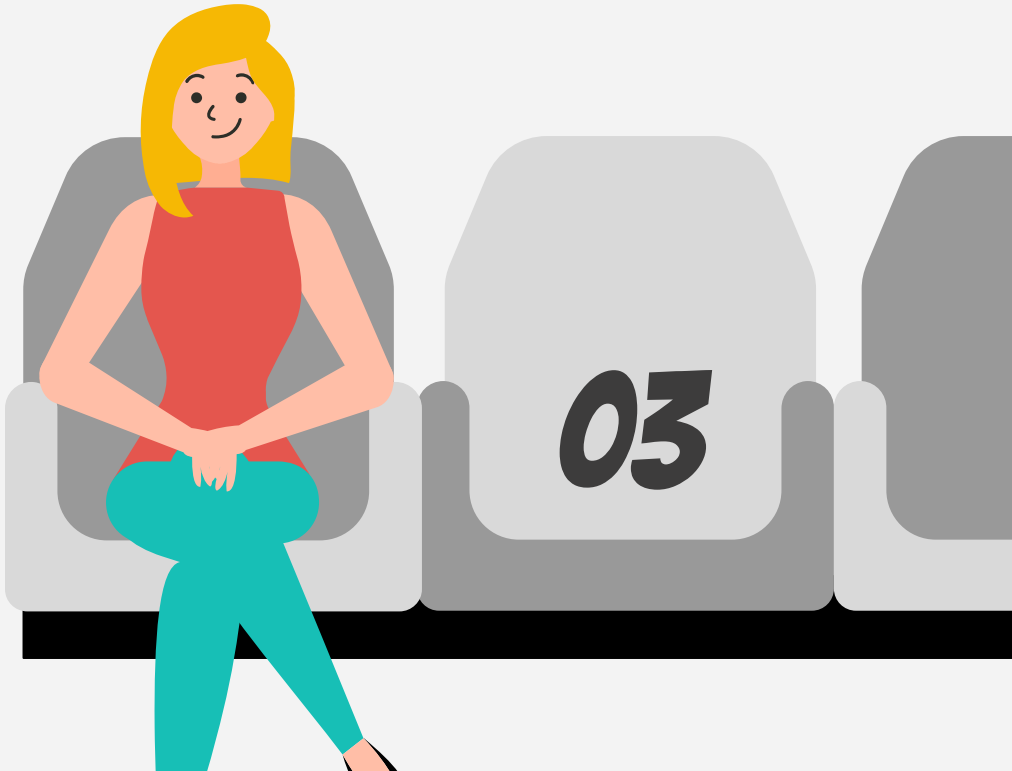

Number of Movies Under Each Genre
Drama
Comedy
Horror
Thriller
Romance
Action
Crime
Adventure
Mystery
Sci-Fi
Music
Family
Fantasy
Western
Biography
Musical
War
Animation
History
Sport
Film-Noir
Reality-TV
Adult
Documentary
News
Unique Genres
0
2500
5000
7500
10000
12500
15000
17500
Number of Movies
Label Count
14000
12000
10000
8000
6000
4000
2000
0
Number of Labels
single_genre
two_genres
three_genres
Count


Distribution of Storyline Length
Count
Storyline Length
0
1000
2000
3000
4000
5000
6000
7000
0
25
50
75
100
125
150
175
200


Frequency of 20 Most Common Words
Frequency of Words
10000
8000
6000
4000
2000
0
find
life
one
get
young
take
new
friend
go
year
man
love
two
family
make
woman
become
old
try
father
Common Words

MODELING MADNESS

Modeling and Evaluation



APPROACH

Multi-label Classification

Using **OneVsRest approach**, with union of all genres that are predicted for a movie to be taken as the multi-label output

MODELS - GENRE PREDICTION

K-Nearest Neighbors

Logistic Regression

Multinomial Naive Bayes

PRODUCTION MODEL

MULTINOMIAL NAIVE BAYES WITH TFIDF VECTORIZER

MODEL PERFORMANCE

Genre	Train	Test
Drama	0.81	0.71
Comedy	0.84	0.78
Horror	0.91	0.89
Thriller	0.87	0.83
Romance	0.89	0.86
Action	0.89	0.87
Crime	0.90	0.87
Adventure	0.94	0.91

Genre	Train	Test
Mystery	0.94	0.92
Sci-Fi	0.95	0.94
Music	0.97	0.95
Family	0.97	0.96
Fantasy	0.97	0.95
Western	0.99	0.97
Others	0.91	0.87

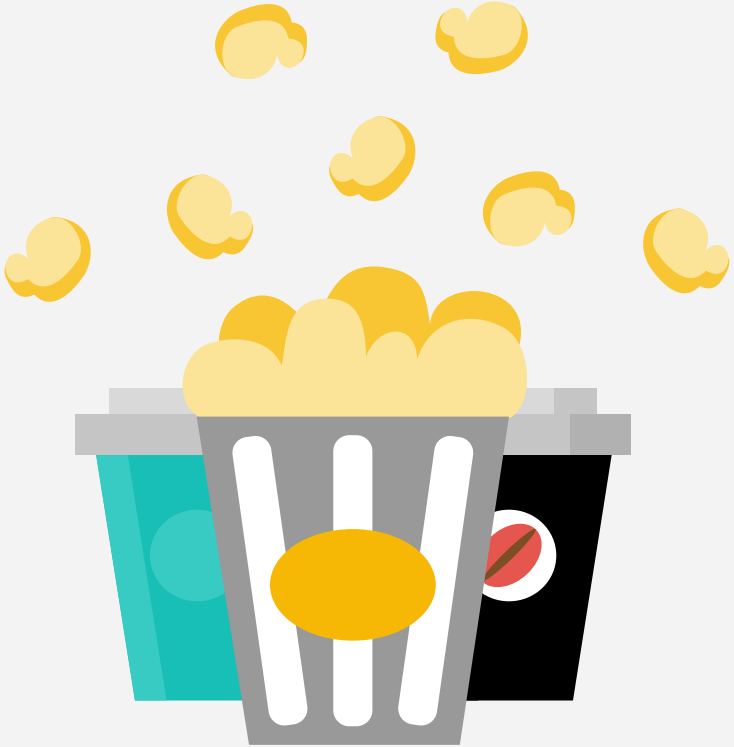
EVALUATION

Hamming Loss

0.11

Accuracy Score

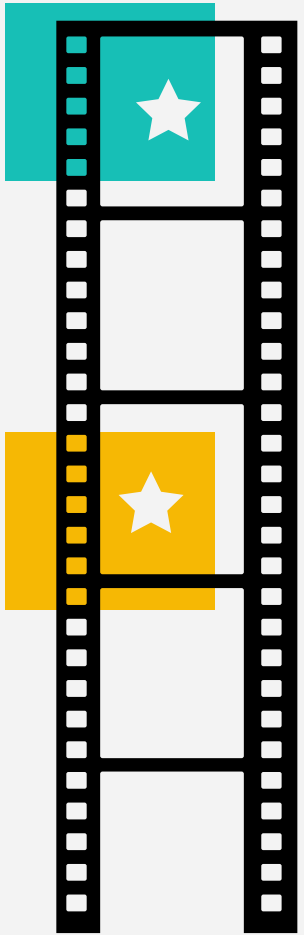
0.14



MISCLASSIFICATION

Despite the Falling Snow
(no label)

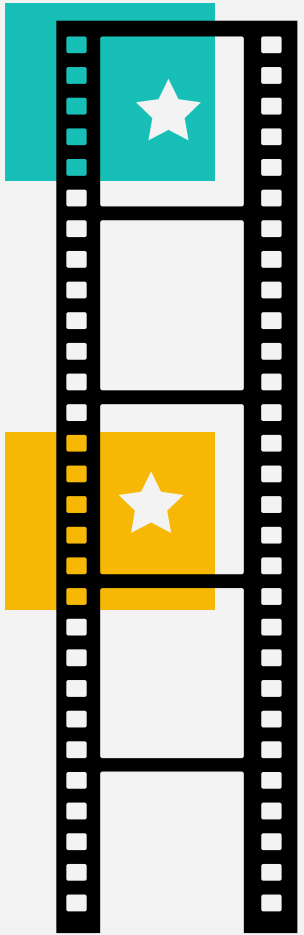
'moscow communist katya secretly spy american cold war arm race land big assignment steal secret rise government star alexander last thing expect fall love reconcile passion belief communism mean make great sacrifice life sacrifice alexander uncovers thirty year later'



MISCLASSIFICATION

Dickie Roberts: Former Child Star
(correctly predicted comedy
but also labeled drama)

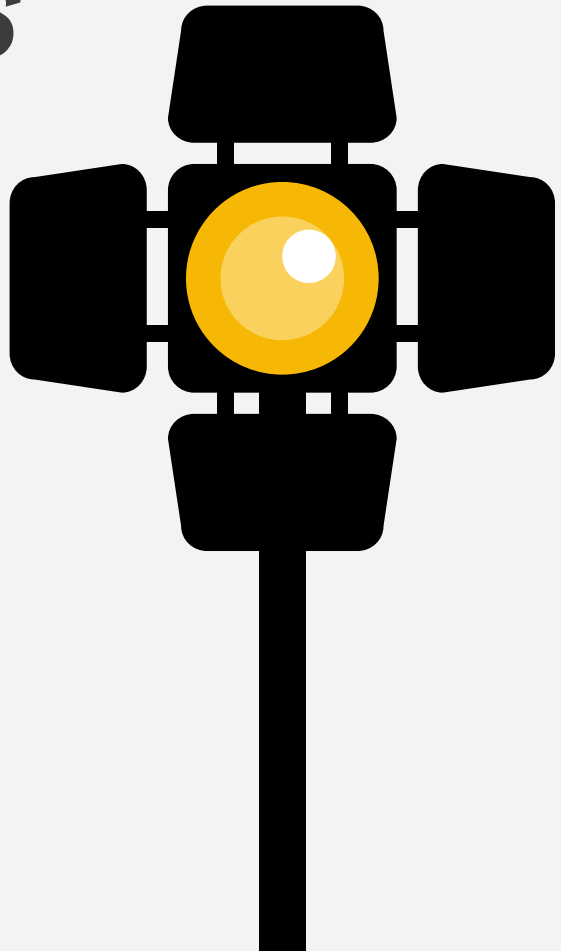
'top world popular child actor tv hollywood punchline everything wrong people famous child dickie roberts want find one gig restore honor everyone love learn rob reiner make ambitious new movie destine sweep oscar dickie first line audition walk learn certainly look part act yet owe unusual childhood research role dickie embark bizarre scheme live suburban family see average american child life put son glove dickie discovers great part true family whether get part attempt method act certainly change life forever'



LIMITATIONS AND NEXT STEPS

In using the OneVsRest classifier for training our model, we ignore any correlation between genres in movies with multiple labels

In addition, labels are predicted based purely on the unique genres found within the IMDB dataset. Further iterations of the model maybe required to train for new genres.

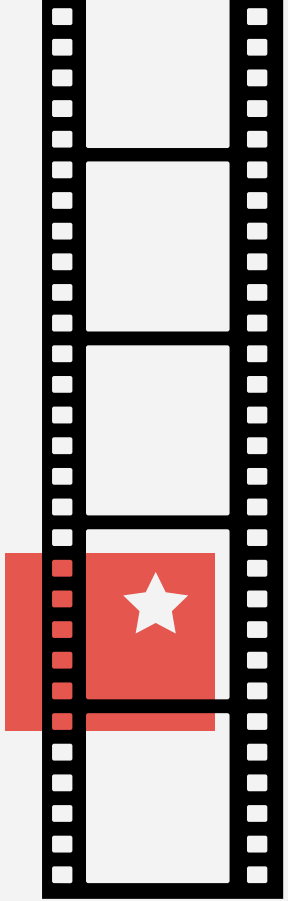


MODELS - RECOMMENDER SYSTEM

Storyline-based Recommender

Genre and Description-based Recommender

EVALUATION



Storyline - Shrek

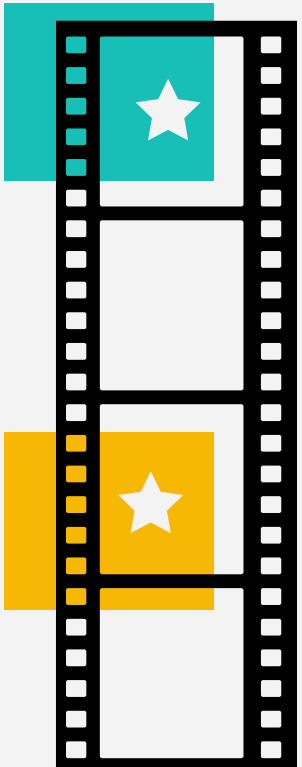
Shrek 2

Shrek Retold

Shrek Forever After

Shrek the Third

Shrek the Musical



Description and Genre - Shrek

Shrek Retold

One More Time

The Lorax

Snow White and the Seven Dwarfs

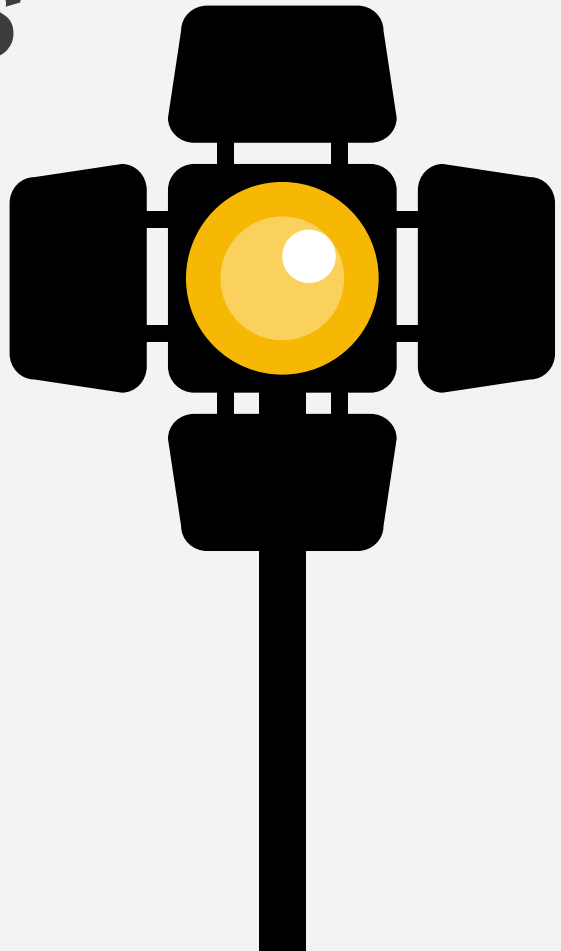
Fairy Tales

LIMITATIONS AND NEXT STEPS

The models developed are based on the **exhaustive list of English movies** within the IMDB dataset

The systems are purely **content-based** as the Binge platform is currently under development

Binge can look into developing collaborative recommender systems that can provide recommendations based on user profiles



THE FINALE

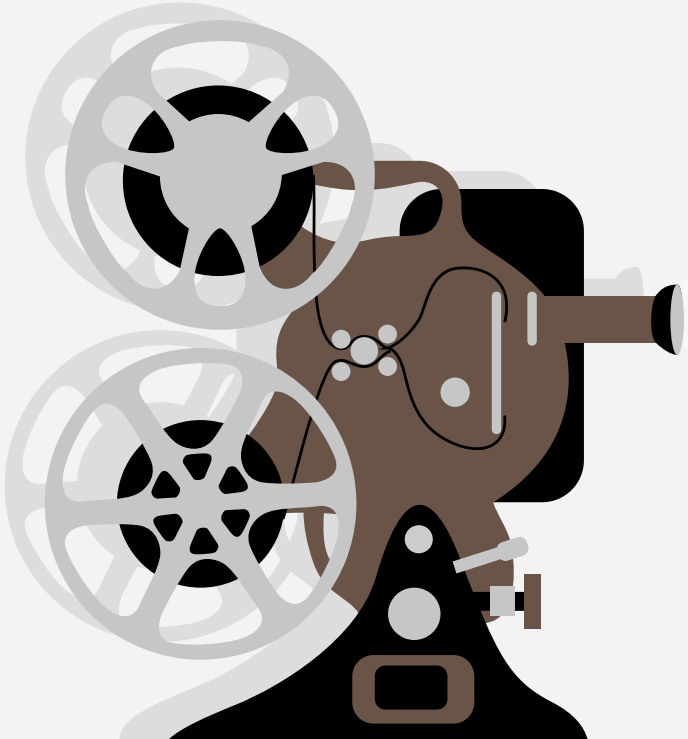
Conclusion

CONCLUSION

New content is being generated everyday

The genre prediction and recommender systems provide a good starting point for Binge to **optimize internal operations and engage users**

Once Binge rolls out its services, it can collect **user feedback and undertake A/B testing** to improve its content classification and recommendation capabilities



THANKS

Do you have any questions?

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik.