

A NETFLIX
ORIGINAL



—
MOVIE RECOMMENDATION ENGINE



TEAM ENIGMA

THE TEAM: SET OF WORKAHOLICS

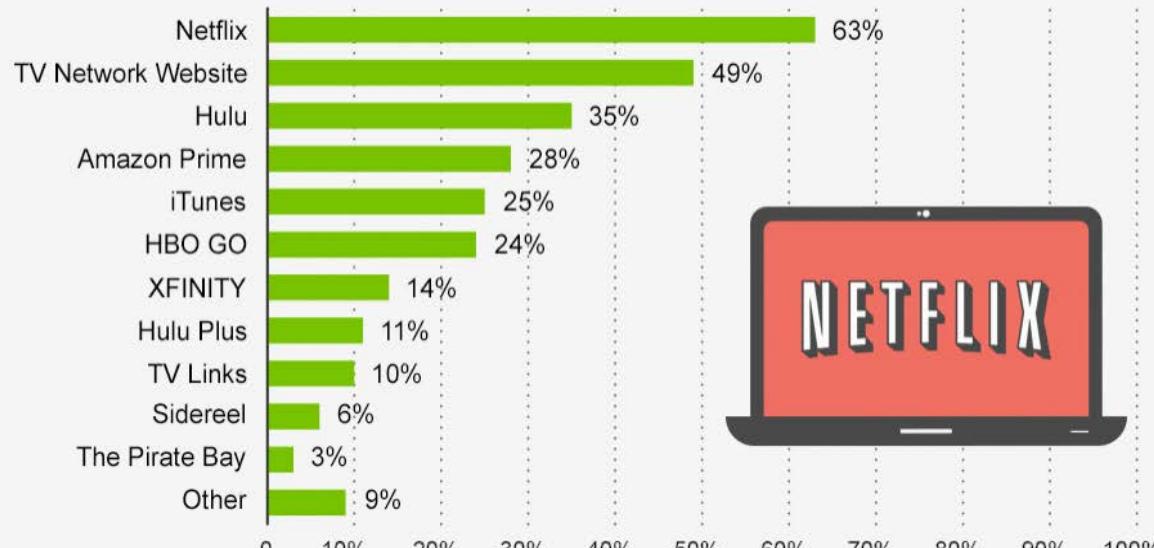


SKIP INTRO

BACKGROUND

Americans Turn to Netflix to Watch TV Content Online

% of respondents who use the following services to access TV content online



NETFLIX

125m
members
in over
190 countries

140m
hours of TV shows
and movies per day

5,500
employees
(2017)

\$11.69b
revenue
(2017)

Headquarter
Los Gatos,
California

30 employees,
925 films on
DVD available
for online
rental

A short history of Netflix

Flat rate for
members

1999

1 billion
films sent
Online streaming
introduced

2007

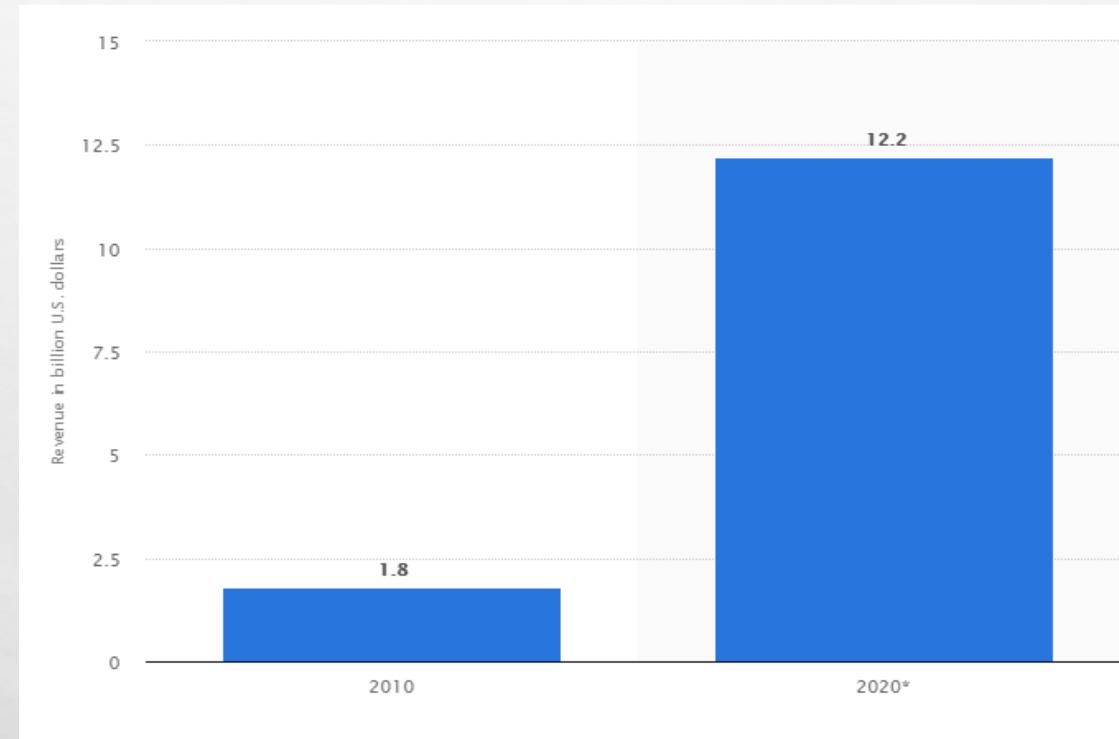
\$1 billion
revenue for the
first time

2014

SKIP INTRO

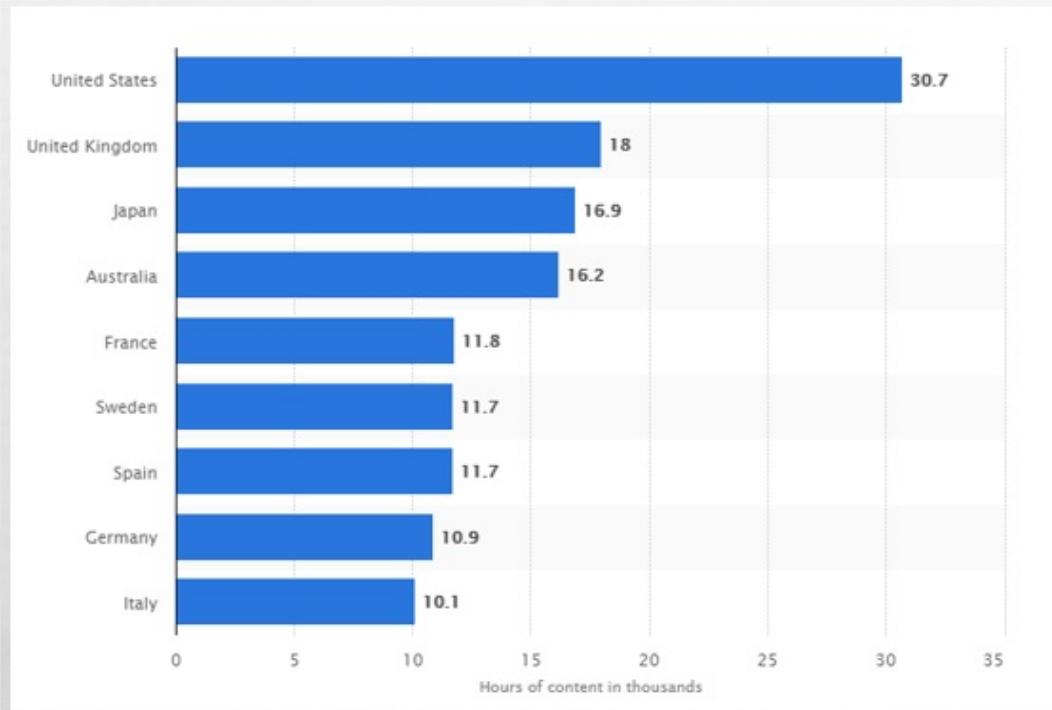
ABOUT THE STREAMING INDUSTRY

Revenue generated
in billion \$ in 2010
and 2020

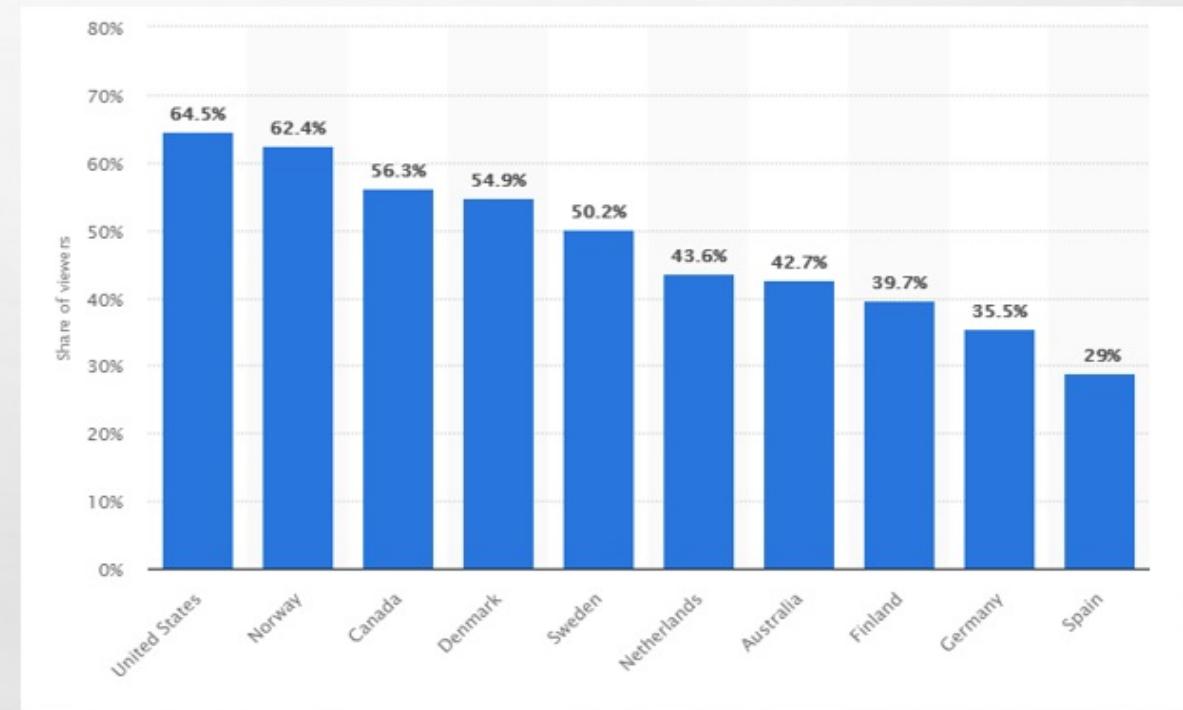


SKIP INTRO

NETFLIX: GAME CHANGER FOR YEARS



Hours of content consumed by countries



Share of viewers by countries

SKIP INTRO

COMPETITION IS GETTING FIERCE

hulu

PlayStation Vue | 

amazon instant video

YouTube

HBO NOW

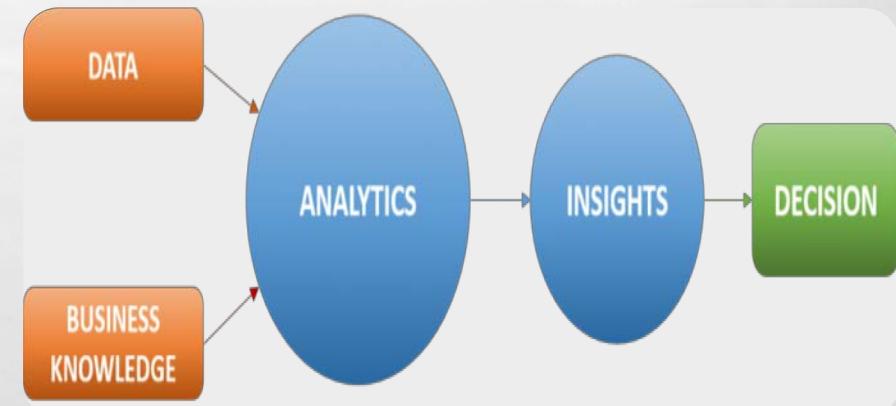
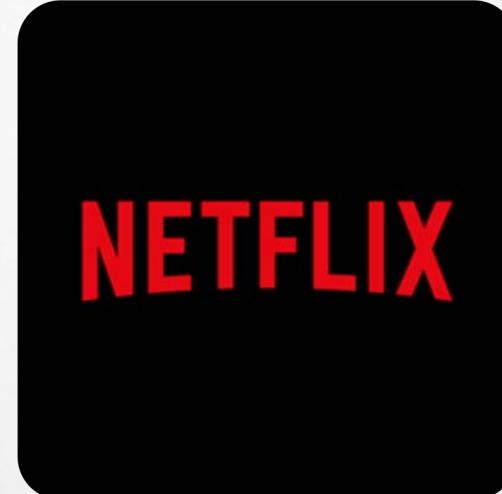


SKIP INTRO

OBJECTIVE

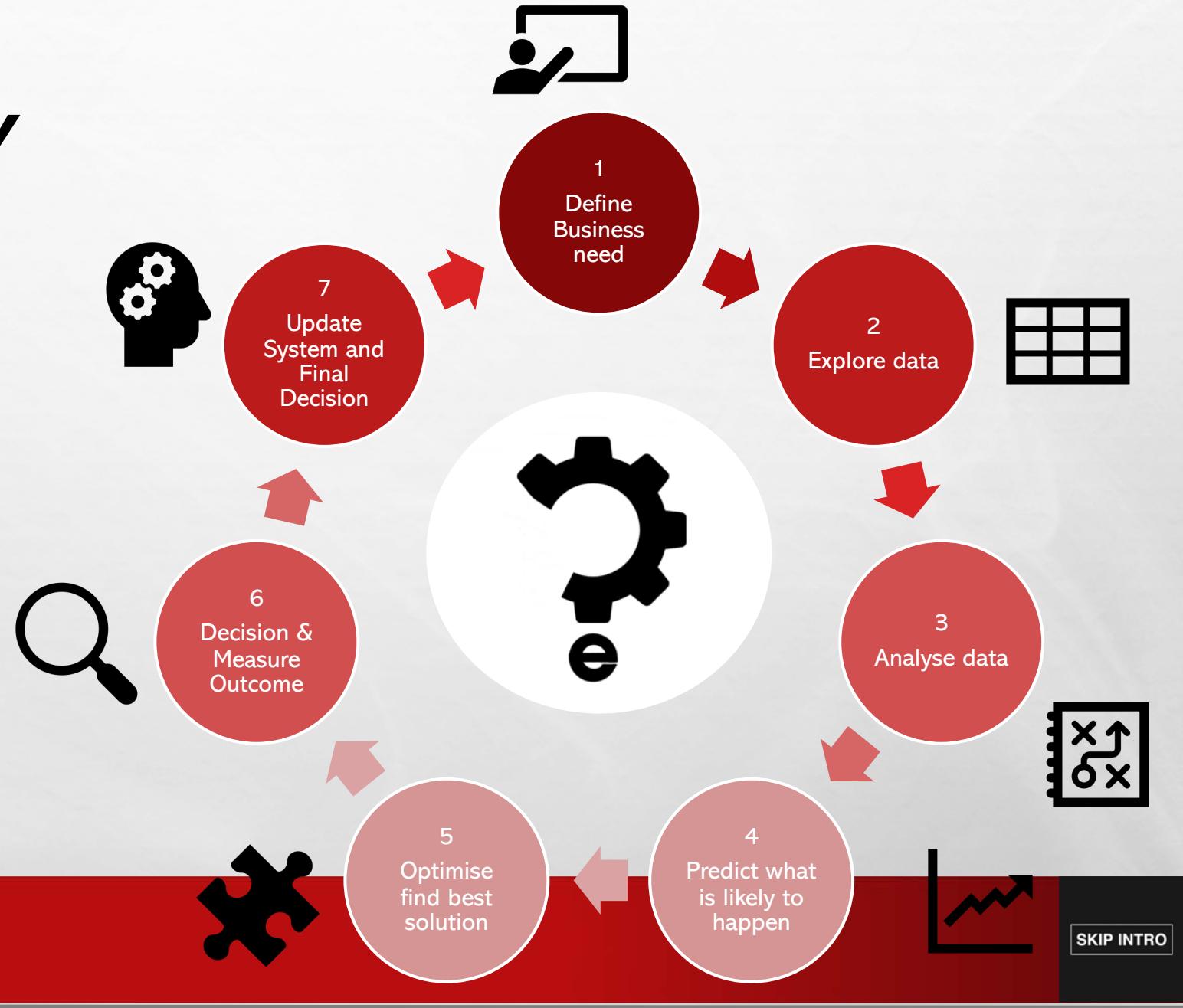
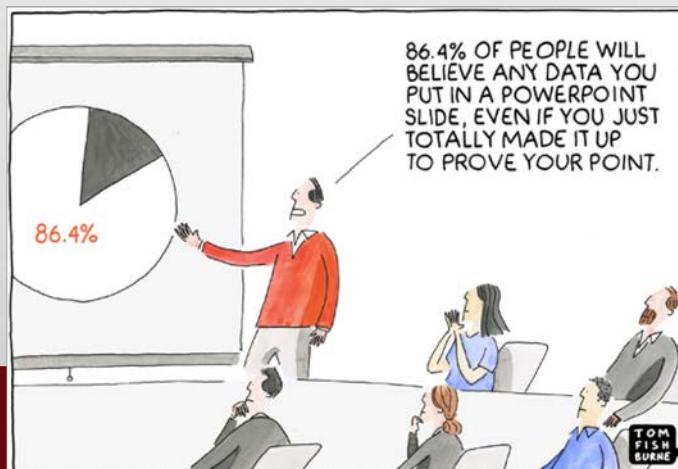


- Develop a Movie Recommendation Engine for NETFLIX.
- The Engine takes an input and must fetch 5 recommendations.
- Business Needs: Popularity + Content based Engine.



THE METHODOLOGY

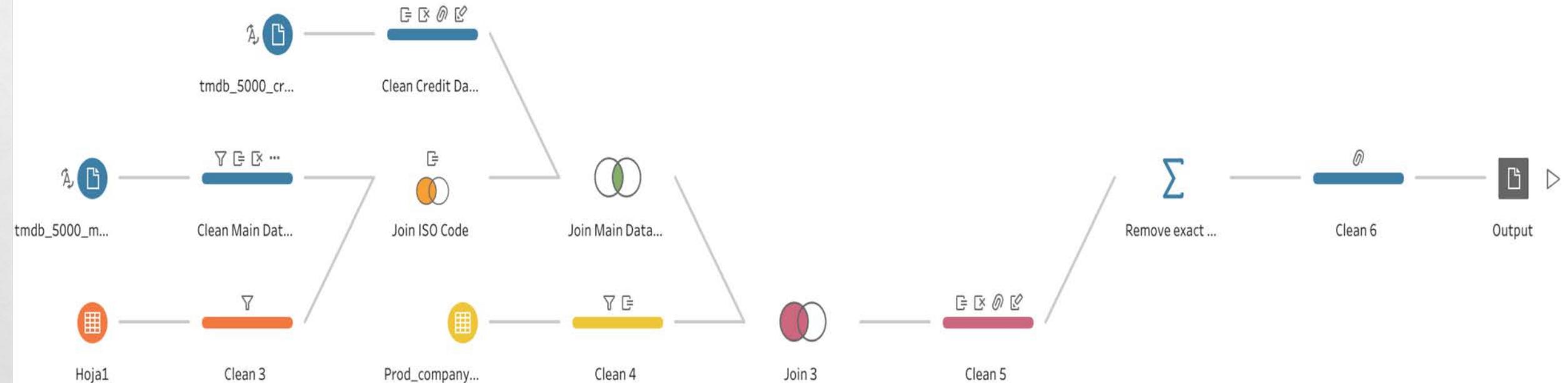
DATA DRIVEN APPROACH



DEFINING THE GAMEPLAN

	Task Name	Day	Percent Complete
Business Problem			
B1	Understanding the business context	22-Oct-18	100%
B2	Identification of key points	22-Oct-18	100%
B3	Understanding industry/sector	22-Oct-18	100%
Data Prep			
DP1	Understanding the data	23-Oct-18	100%
DP2	Cleaning data	23-Oct-18	100%
DP3	Preparing final data set	23-Oct-18	100%
Data Analytics			
DA1	Understanding modelling process	24-Oct-18	100%
DA2	Programming	25-Oct-18	100%
Testing/Iteration			
T1	Testing results	26-Oct-18	100%
T2	Fine tuning results	27-Oct-18	100%
T3	Review session	28-Oct-18	100%
Presentation			
P1	Defining presentation parameters	29-Oct-18	100%
P2	Structuring presentation and speakers	5-Nov-18	100%
P3	Presentation practice	6-Nov-18	100%

DATA CLEANING

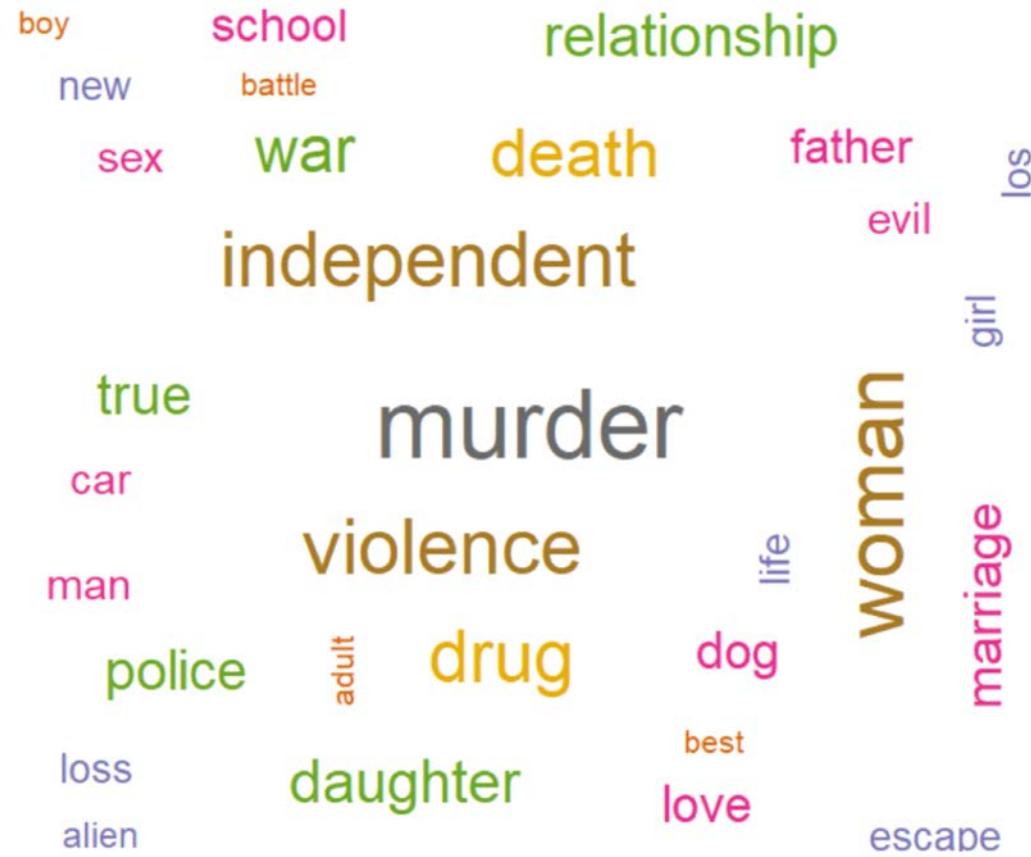
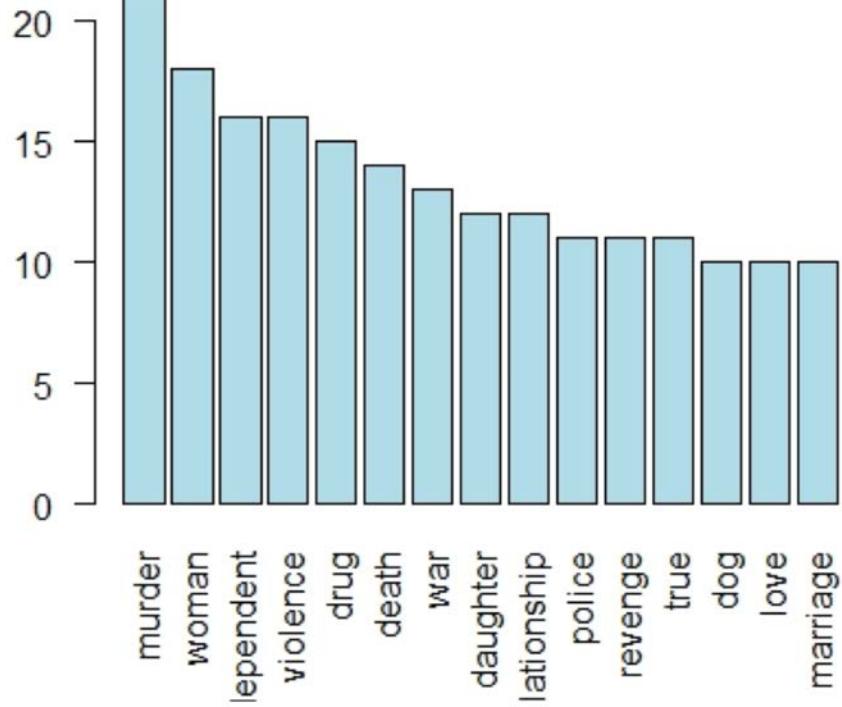


SKIP INTRO

Keyword Analysis

Most frequent words

word frequencies

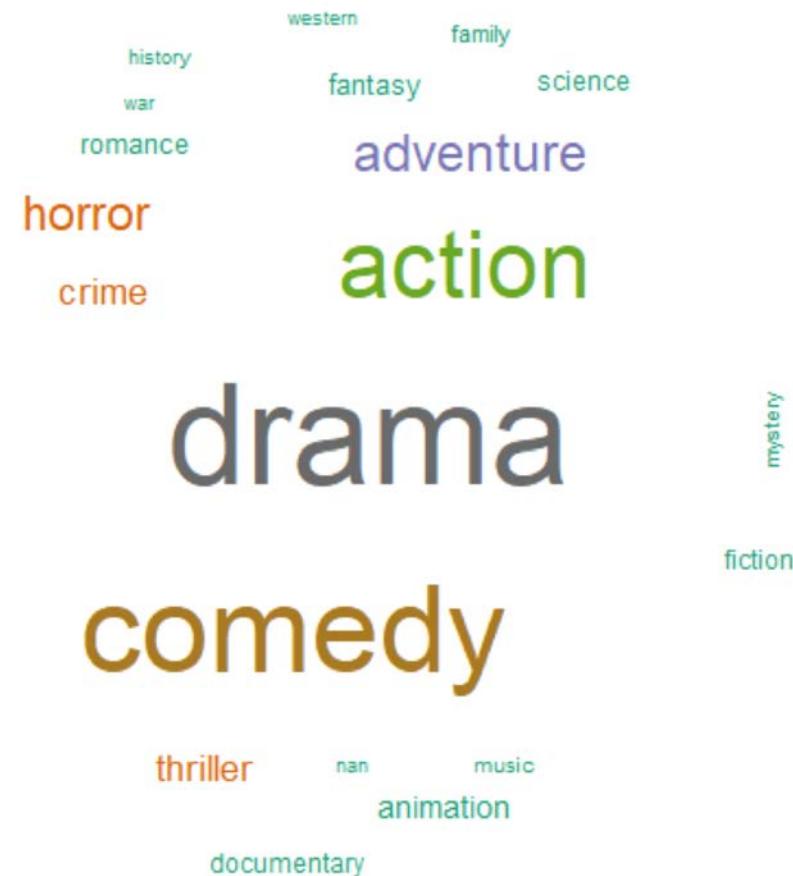
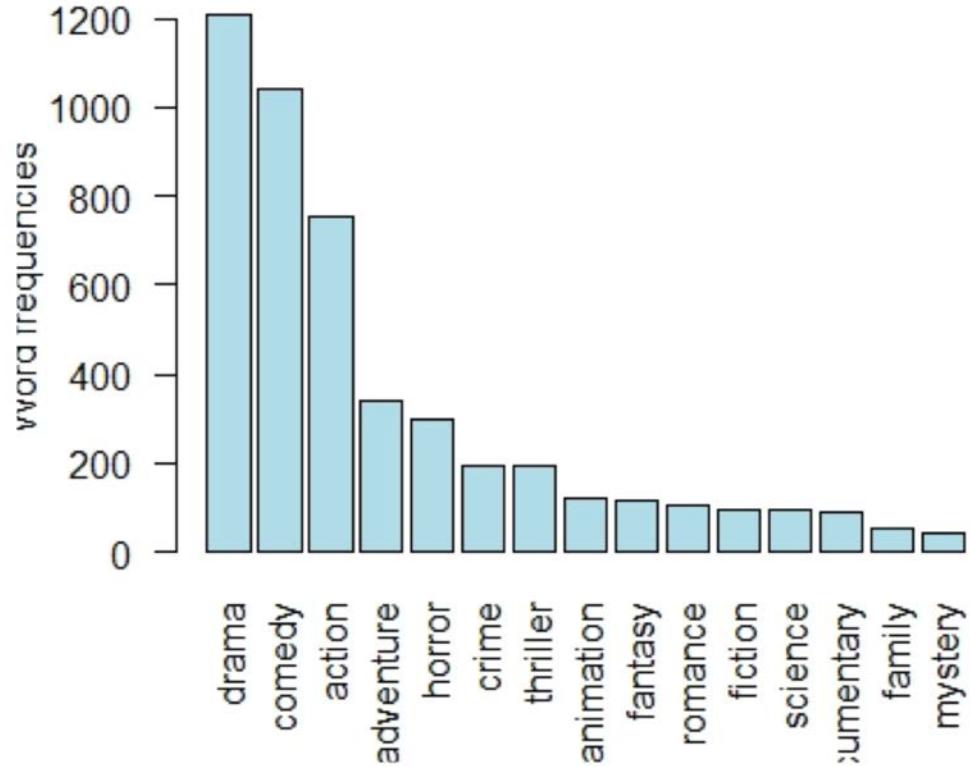


DATA EXPLORATION/ANALYSIS

SKIP INTRO

Genre

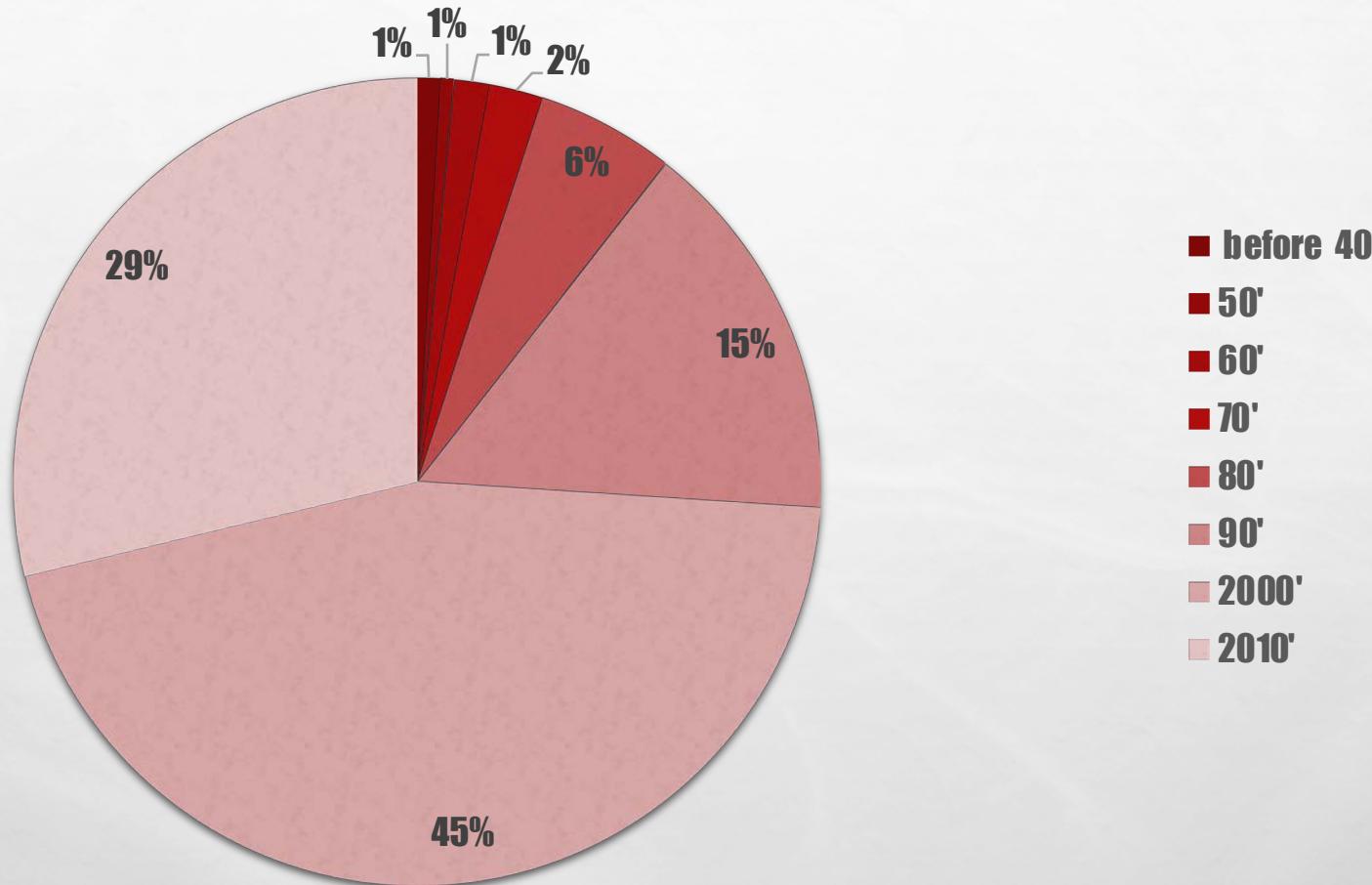
Most frequent words



DATA EXPLORATION/ANALYSIS

SKIP INTRO

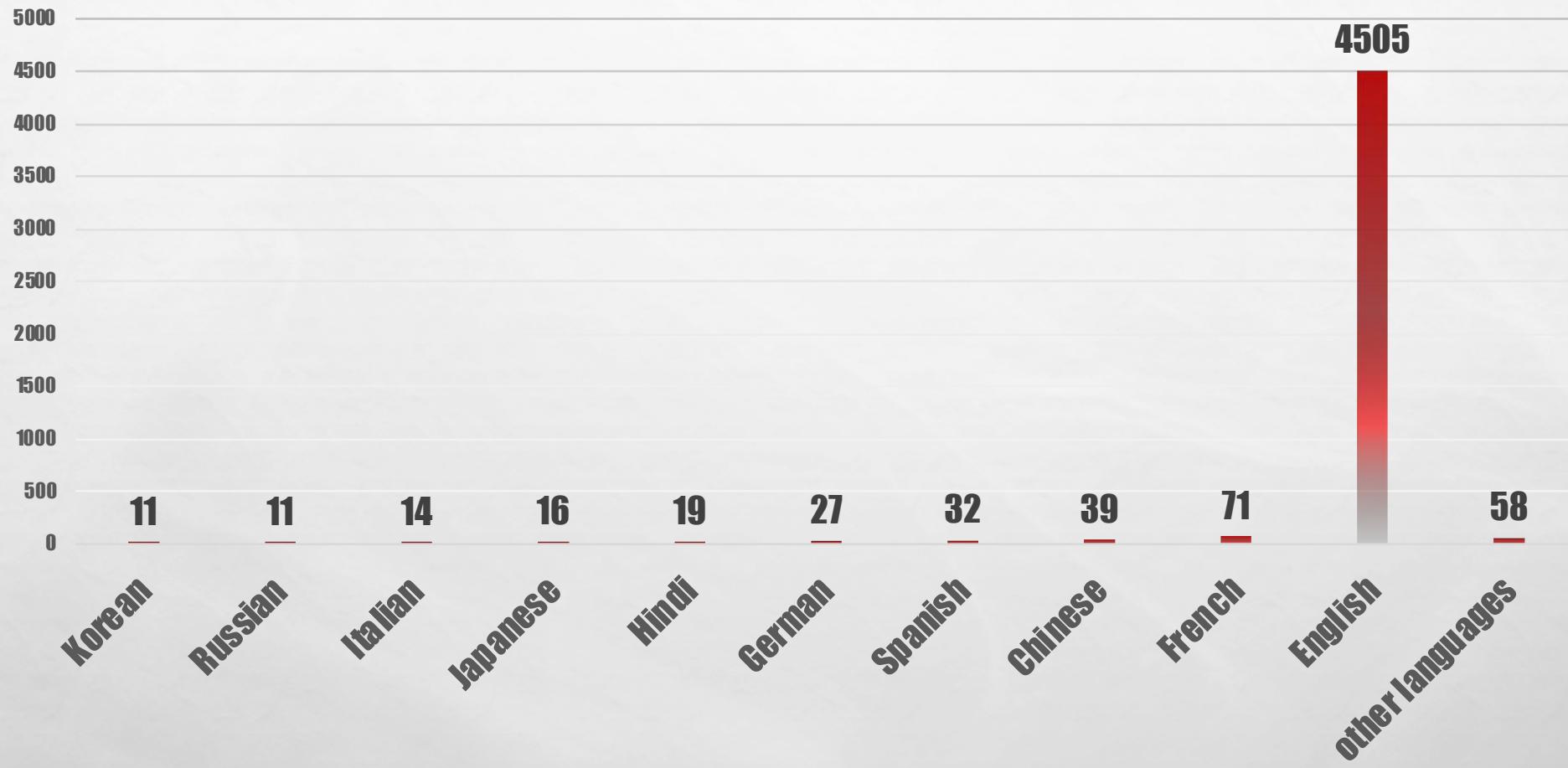
% of films per decade



DATA EXPLORATION/ANALYSIS

SKIP INTRO

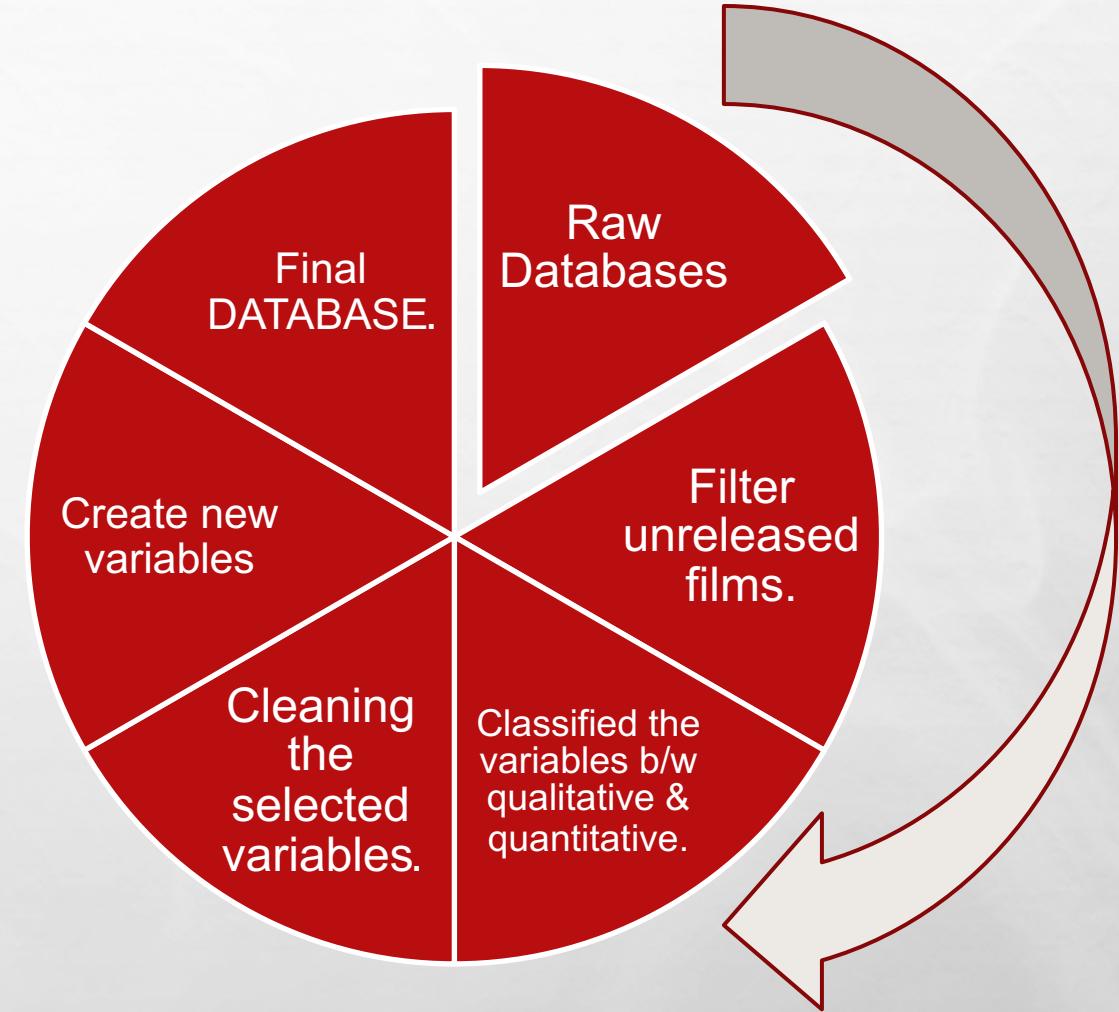
% OF FILMS' ORIGINAL LANGUAGE



DATA EXPLORATION/ANALYSIS

SKIP INTRO

FINAL DATABASE, CHOOSING THE VARIABLES



STEP 5: DATA CLEANING CRITERIA: A DECISION MAKING APPROACH

VARIABLE	QUALITATIVE DATA									
	Title	Director	Original Language	Cast	Production Countries	Keywords	Genres	Independent	Spoken Languages	
Business Value	Key identifier	Preference for films	Geographical Influence	Preference for films	Geographical Influence	Content of the film.	Content of the Film.	Production Preference	Language Preference	
Criteria	Unchanged	Only crew member considered	Unchanged	Five main characters.	Up to 5 countries considered	Up to 9 keywords considered	Up to 5 genres considered	New variable based on budget and production company frequency.	Up to 5	

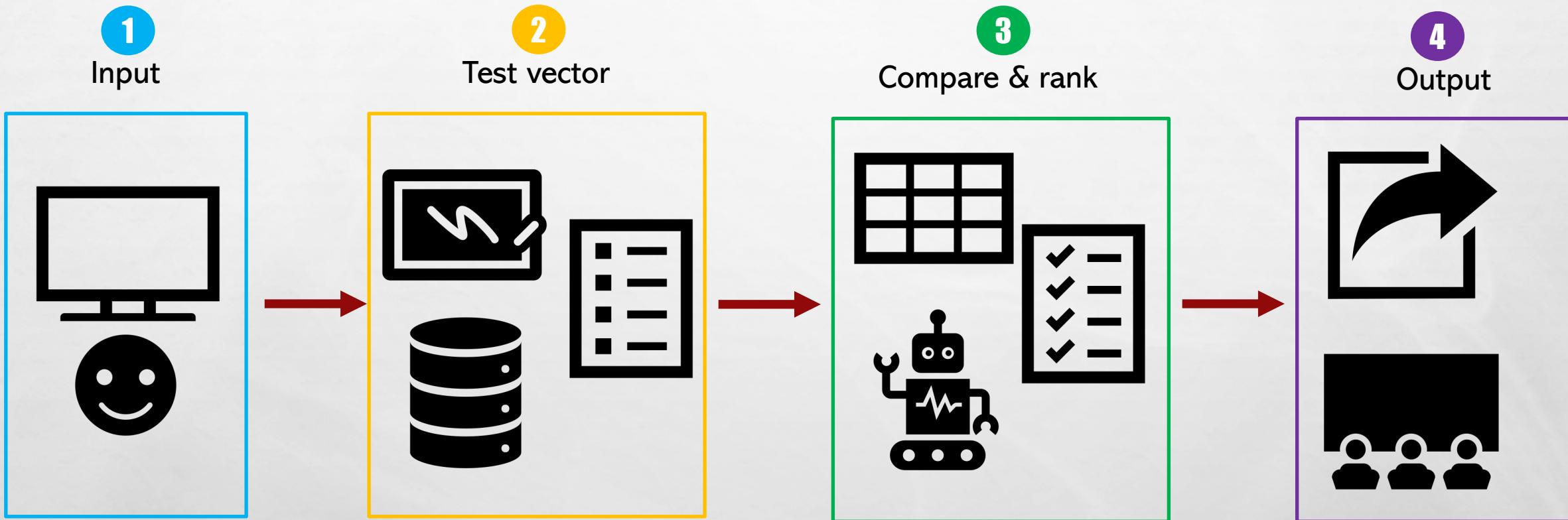
SKIP INTRO

QUANTITATIVE DATA			
VARIABLE	YEAR	VOTE	POPULARITY
Business Value	Preference for new FILMS	Viewers' feedback	Viewers' feedback.
Criteria	Release Year/1000 (1000 parameters based on our scoring scales.)	Adjusted the vote average to the number of votes using IMDB rating formula. (TRUE BAYESIAN ESTIMATE)	Log (Popularity Var)

DATA CLEANING CRITERIA: A DECISION MAKING APPROACH



HOW DOES OUR ENGINE WORK?



SKIP INTRO

1

Input



RAUL

2

Test vector



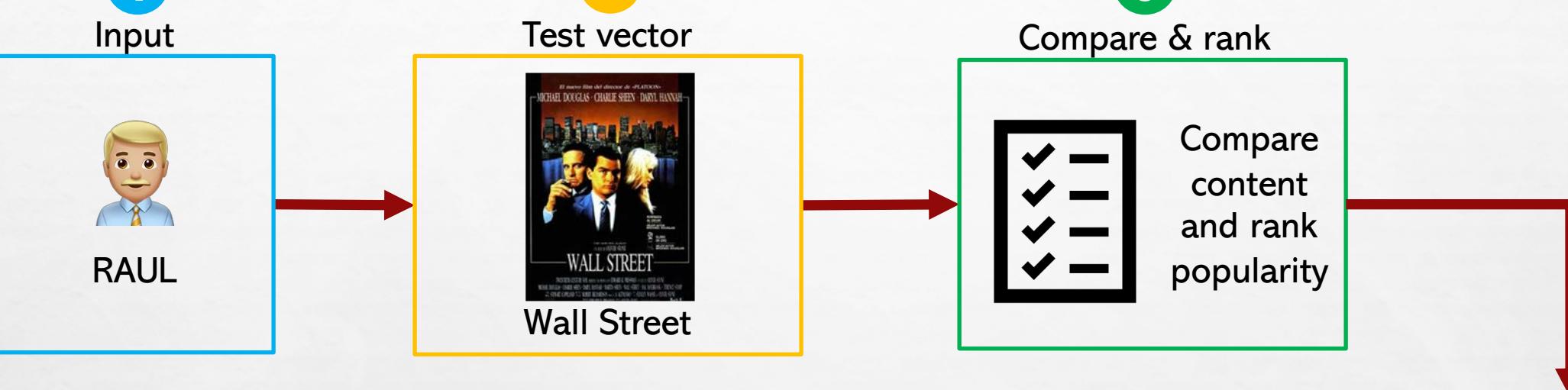
Wall Street

3

Compare & rank

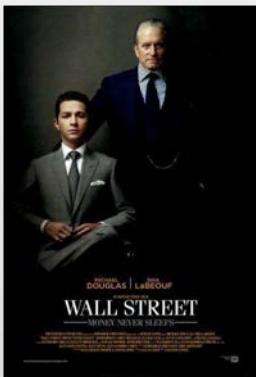


Compare content
and rank popularity



4

Output



Wall Street:
Money Never Sleeps



The Big Short



The Wolf of
Wall Street



The God
Father



Traffic

IN PRACTICE...

SKIP INTRO

1

Input



Ana

2

Test vector



Pulp Fiction

3

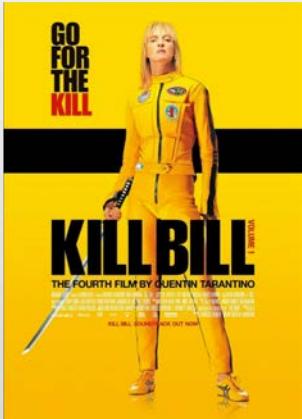
Compare & rank



Compare
content
and rank
popularity

4

Output



Kill Bill: Vol I



Kill Bill: Vol II



Jackie Brown



Se7en



The Dark Knight

IN PRACTICE...

SKIP INTRO

1

Test vector

```
#creating selection
n<-672

for(x in 1:nrow(movie_raw_data))
{
  if(n %in% movie_raw_data[x,1])
  {
    vtest <- as.data.frame(movie_raw_data[x,])
  }
  else

    x+1
}

class(vtest)
vtest<- as.matrix(vtest)
```



2

Compare & rank: Qualitative variables

```
#define variable test vector
vtest_variable<-as.vector(vtest[,30:38])

#sort alphabetically variable test vector
vtest_variable<-sort(vtest_variable)

#define variable database
variable_database<-as.matrix(movie_raw_data[,30:38])

#Remove row/col names
dimnames(variable_database)=NULL

#Loop to alphabetically arrange variable Database by row
s <- variable_database
for(i in 1:NROW(variable_database)) {
  s[i,] <- sort(s[i,],na.last = TRUE)
}
#s
#Loop to compare test vector to variable database
for(j in 1:nrow(variable_database)){

  variable_database[j,] <- as.character(s[j,] %in% vtest_variable)
  j=j+1
}

#Convert T/F to binary
variable_database=ifelse(variable_database=="TRUE",1,0)

#Add results and create a vector
variable_results<-as.vector(rowSums(variable_database))
```

HOW DO WE GET OUR RESULTS

SKIP INTRO

3

Compare & rank: Qualitative variables, cont.

```
comparisonmatrix <- matrix(NA, ncol=12, nrow=4794) #create empty matrix

for(x in 1:nrow(movie_raw_data)){
  comparisonmatrix[x,] <- as.vector(colSums(movie_raw_data[x,1:12]==vtest))
  x=x+1
} #loop to compare each vector to the movie database and then paste it as a row in the resultmatrix
```

5

Output

```
for (i in 1:length(positionvector)){
  ww=positionvector[i] # taking the row number
  recommendation<- rbind(recommendation,movie_raw_data1[ww,1:2])
  #appending all the results in new dataframe
}
print(recommendation) # to print the dataframe
```

4

Compare & rank: Quantitative variables & ranking

```
popularity_vector<-as.vector(movie_raw_data$Adjusted_popularity)
year_vector<-as.vector(movie_raw_data$Adjusted_year)
vote_vector<-as.vector(movie_raw_data$Vote_true_bayesian_rating)

resultmatrix<-cbind(comparisonmatrix,popularity_vector,year_vector,vote_vector,keyword_results,actor_results,genre_results,prod_country_results)

#Add every row and sort the results
|
rankvector<-as.vector(rowSums(resultmatrix,na.rm=TRUE))
resultvector<-as.vector(tail(sort.int(rankvector, partial=length(rankvector)-4,decreasing=FALSE), 6))
positionvector<- which(rankvector %in% resultvector)
```

HOW DO WE GET OUR RESULTS

[SKIP INTRO](#)

SUMMARY

1. Evermore competitive industry, keep innovating
2. What info we have? What info I need?
3. Lets build an engine with the data!
4. You're gonna like our recommendations!
5. Maybe not... Let me try again
6. Sucess! (Keep consumer engaged with our platform)
7. Kevin, Serge, it's up to you now!

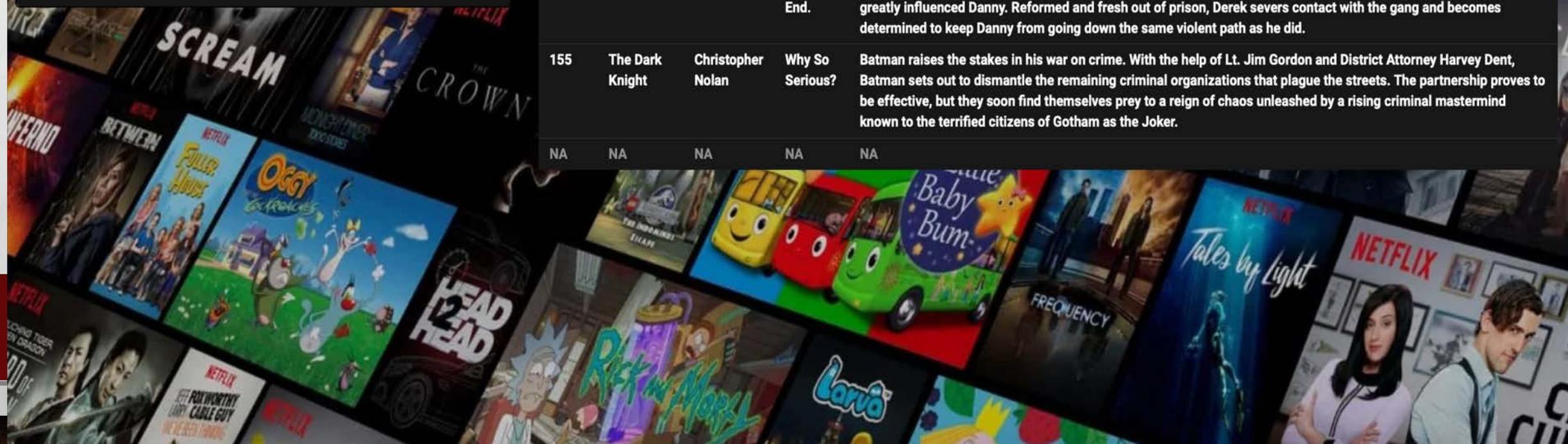


Please Enter Movie ID



Select Movie

91076



Recommendations For you

id	title	Director	tagline	overview
544	There's Something About Mary	Bobby Farrelly	Love Is In The Hair.	Having never fully recovered from a prom date that became a total disaster, a man finally gets a chance to reunite with his old prom date, only to run up against other suitors including the sleazy detective he hired to find her.
75638	Red Lights	Rodrigo Cortu00e9s	How much do you want to believe?	Two investigators of paranormal hoaxes, the veteran Dr. Margaret Matheson and her young assistant, Tom Buckley, study the most varied metaphysical phenomena with the aim of proving their fraudulent origins. Simon Silver, a legendary blind psychic, reappears after an enigmatic absence of 30 years to become the greatest international challenge to both orthodox science and professional sceptics. Tom starts to develop an intense obsession with Silver, whose magnetism becomes stronger with each new manifestation of inexplicable events. As Tom gets closer to Silver, tension mounts, and his worldview is threatened to its core.
73	American History X	Tony Kaye	Some Legacies Must End.	Derek Vineyard is paroled after serving 3 years in prison for killing two thugs who tried to break into/steal his truck. Through his brother, Danny Vineyard's narration, we learn that before going to prison, Derek was a skinhead and the leader of a violent white supremacist gang that committed acts of racial crime throughout L.A. and his actions greatly influenced Danny. Reformed and fresh out of prison, Derek severs contact with the gang and becomes determined to keep Danny from going down the same violent path as he did.
155	The Dark Knight	Christopher Nolan	Why So Serious?	Batman raises the stakes in his war on crime. With the help of Lt. Jim Gordon and District Attorney Harvey Dent, Batman sets out to dismantle the remaining criminal organizations that plague the streets. The partnership proves to be effective, but they soon find themselves prey to a reign of chaos unleashed by a rising criminal mastermind known to the terrified citizens of Gotham as the Joker.



SKIP INTRO

That's all Folks.



"Any questions?"

ANNEX 1: IMBD FORMULA

$$\text{Weighted IMDB rating} = \left(\frac{v}{(v + m)} \right) (R) + \left(\frac{m}{(v + m)} \right) (C)$$

where:

- R = average for the movie (mean) = (Rating)
- v = number of votes for the movie = (votes)
- m = minimum votes required to be listed in the Top 250 (currently 25000)
- C = the mean vote across the whole report (currently 7.0)