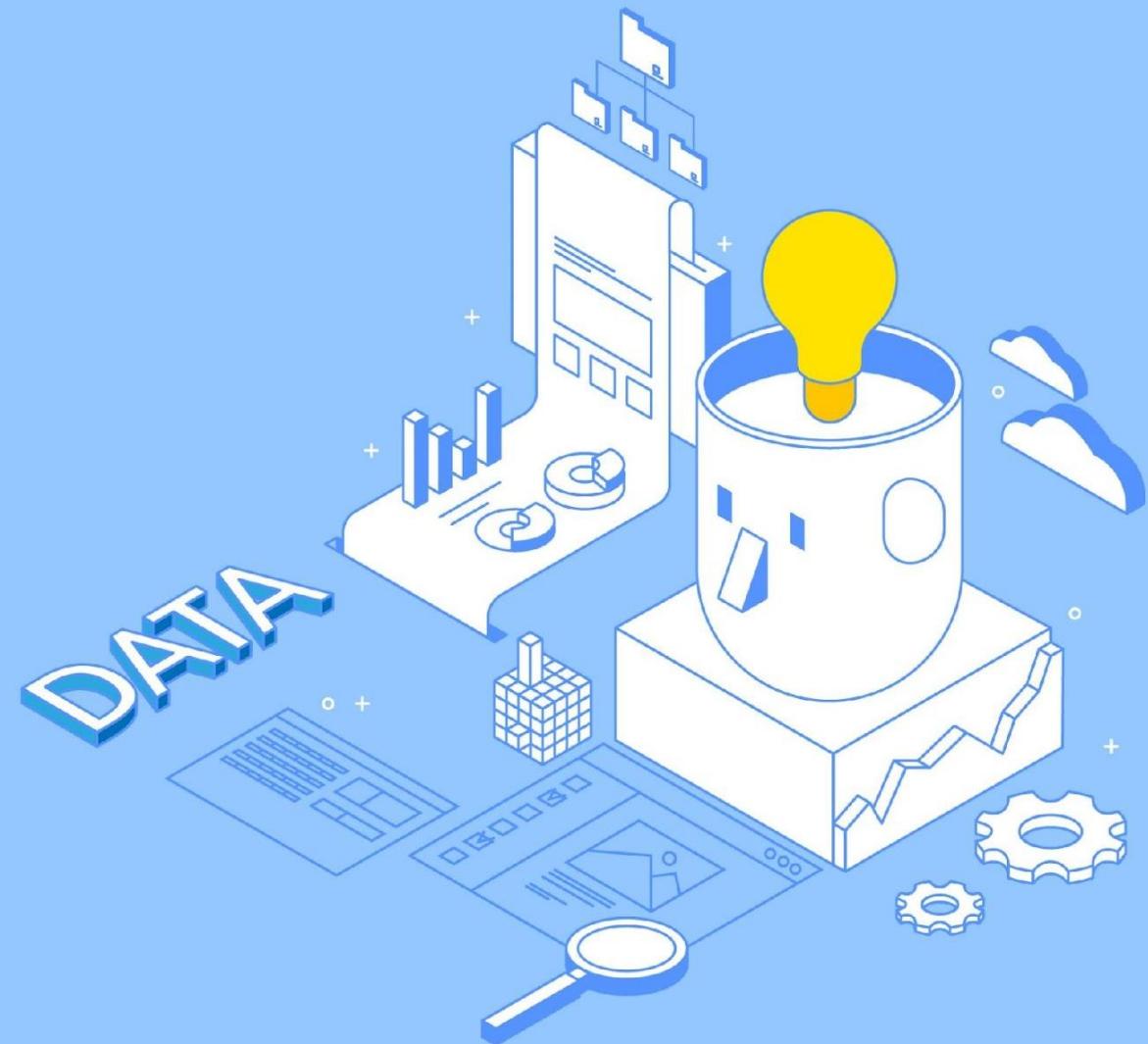




2021 - 2 Big Data

TV shows & Movies on OTT Platforms



Contents



I Introduction

II Data overview

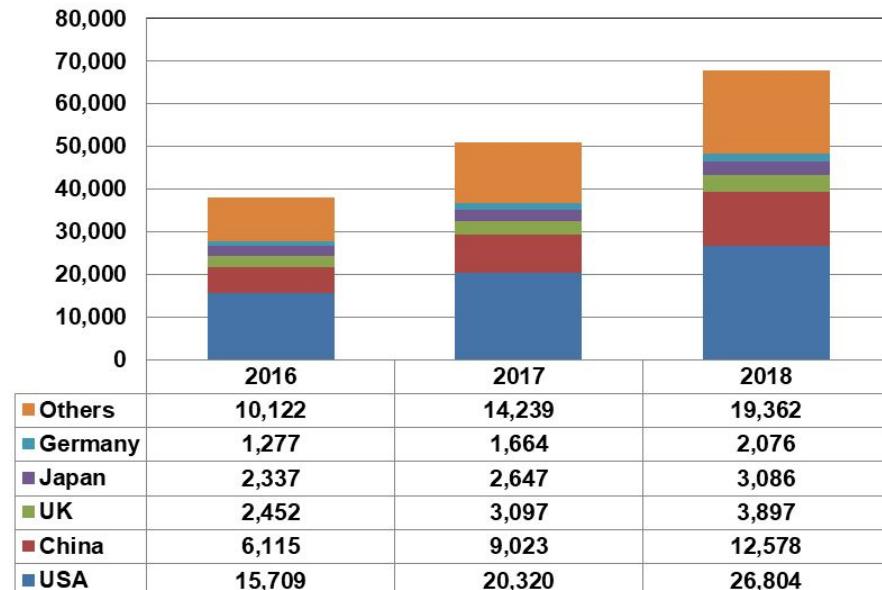
III Methodology & Analysis Plan

Introduction



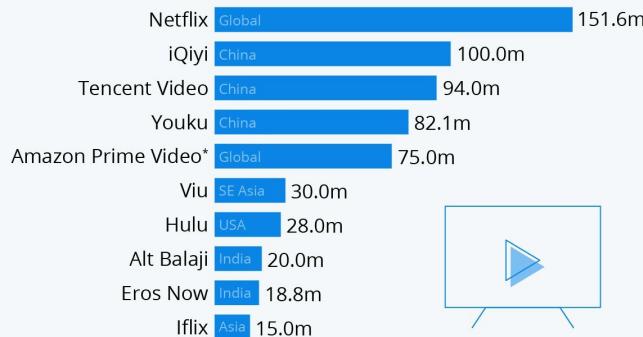
Huge Increase of using OTT & SVOD services

Global OTT TV & video revenue data by top countries (\$ million)



The World's Most Popular Video Streaming Services

Video streaming services with the most subscribers as of November 2019



* Estimate

Source: FIPP Global Digital Subscription Snapshot



statista



NETFLIX

Disney+

amazon
Prime

TVING

wavve

Introduction



→ Recommendation System



Data Overview



→ Datasets from Kaggle

Dataset

Tv Shows on OTT Platforms

A collection of tv shows found on these streaming platforms

NETFLIX

Javagar M • updated a year ago (Version 1)

29

Dataset

Movies on OTT platforms

A collection of movies found on Netflix, Prime Video, Hulu and Disney+

NETFLIX

Javagar M • updated a year ago (Version 1)

18

Data Overview



Datasets from Kaggle

Title	# Year	Age	# IMDb
Title	The year in which the tv show was produced	Target age group	IMDb rating
5564 unique values		[null] 44% 16+ 18% Other (1147) 38%	
Ozark	2017	18+	8.4
Attack on Titan	2013	16+	8.8
Narcos	2015	18+	8.8
Fullmetal Alchemist: Brotherhood	2009	18+	9.1
Community	2009	7+	8.5
Mindhunter	2017	18+	8.6
Parks and Recreation	2009	16+	8.6

Title	Year	# Age	# IMDb
Title - Movie Title (Name) of the movie	Year- year of Released	Age- who are eligible to view the movie	IMDb -IMDB rating towards this movies
16744 unique values	16744 total values	16744 total values	
Inception	2010	13+	8.8
The Matrix	1999	18+	8.7
Avengers: Infinity War	2018	13+	8.5
Back to the Future	1985	7+	8.5
The Good, the Bad and the Ugly	1966	18+	8.8
Spider-Man: Into the Spider-Verse	2018	7+	8.4
The Pianist	2002	18+	8.5
Django Unchained	2012	18+	8.4

Total 17 columns

12 common columns

- Index
- ID
- Title
- Year(제작년도)
- Age(대상 연령)
- IMDb(평점)
- Rotten Tomatoes
- Netflix(Netflix에 존재 여부)
- Hulu(Hulu에 존재 여부)
- Prime Video(Prime Video에 존재 여부)
- Disney+(Disney+에 존재 여부)
- Type(Movie or TV series)

(5점 만점에서 3.5 이상의 평점을 준 관객 비율)

5 columns more in Movie Dataset

- Directors
- Genres
- Country
- Language
- Runtime

5610 rows(TV shows) /
16744 rows(Movie)



66

1. Analyze and Visualize Data distributions and for Data pre-processing

- python-based
- using pandas, seaborn, matplotlib library

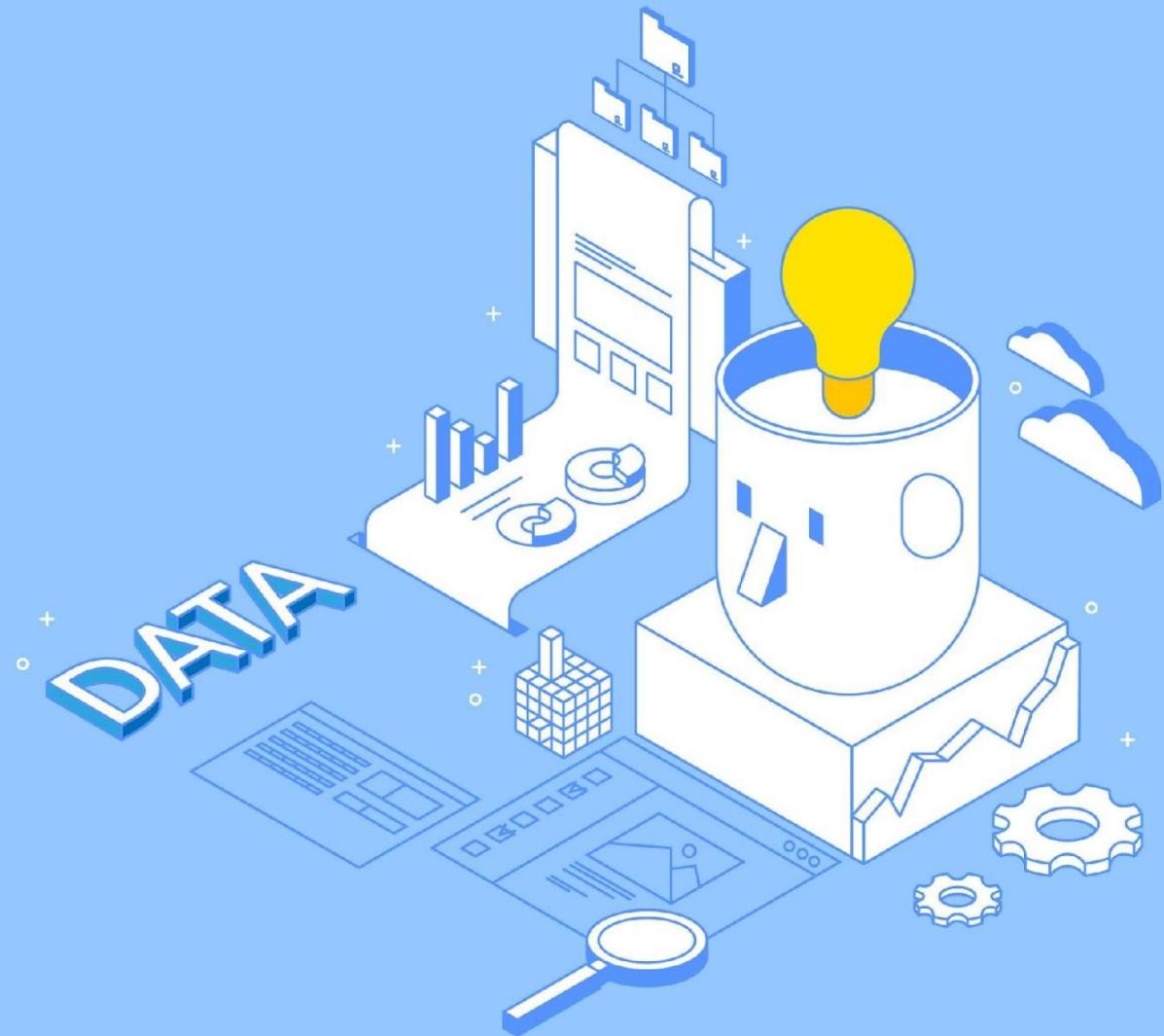
2. Feature Engineering

3. Modeling

- sklearn, pandas, tensorflow, pytorch

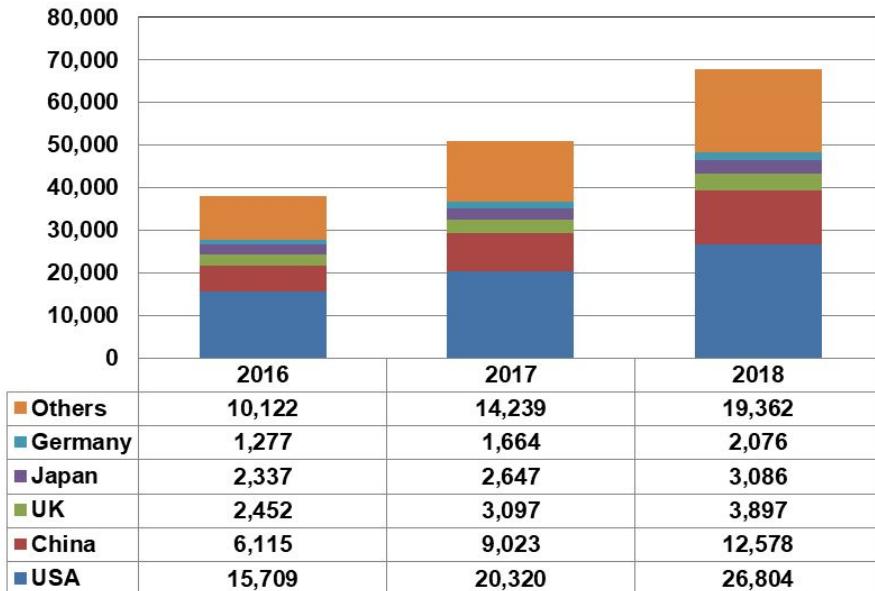


**Thank you
for
listening!**



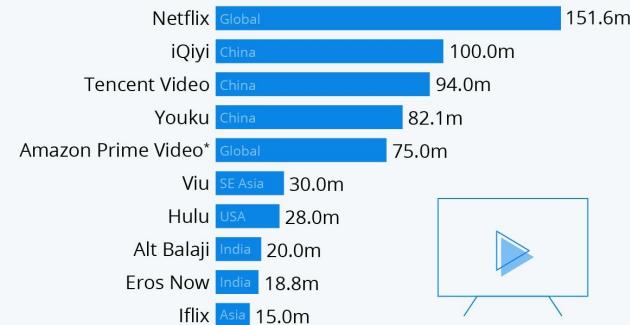
Huge Increase of using OTT & SVOD services

Global OTT TV & video revenue data by top countries (\$ million)



The World's Most Popular Video Streaming Services

Video streaming services with the most subscribers as of November 2019



* Estimate

Source: FIPP Global Digital Subscription Snapshot

statista





Data description

Dataset

TV shows on Netflix, Prime Video, Hulu and Disney+

A collection of tv shows found on these streaming platforms

Ruchi Bhatia • updated 3 months ago (Version 3)

Data Tasks Code (41) Discussion (3) Activity Metadata

Download (313 kB) New Notebook :

Usability 9.7 License CCO: Public Domain

Tags arts and entertainment, online communities, movies and tv shows

Dataset

Movies on Netflix, Prime Video, Hulu and Disney+

A collection of movies found on these streaming platforms

Ruchi Bhatia • updated 3 months ago (Version 3)

Data Tasks (4) Code (69) Discussion (10) Activity Metadata

Download (1 MB) New Notebook :

Usability 10.0 License CCO: Public Domain

Tags arts and entertainment, online communities, movies and tv shows

17 columns

12 공통

- **Index**
- **ID**
- **Title**
- **Year(제작년도)**
- **Age(대상 연령)**
- **IMDb(평점)**
- **Rotten Tomatoes**

(5점 만점에서 3.5 이상의 평점을 준 관객
비율)

- **Netflix**(Netflix에 존재 여부)
- **Hulu**(Hulu에 존재 여부)
- **Prime Video**(Prime Video에 존재
여부)
- **Disney+**(Disney+에 존재 여부)
- **Type**(Movie or TV series)

5 Movies만

- **Directors**
- **Genres**
- **Country**
- **Language**
- **Runtime**

Analysis plan

1. 데이터 분포, 전처리를 위한 분석, 시각화

- pandas
- 시각화툴(seaborn(sns), matplotlib 등)

2. 시각화 후 피처 엔지니어링

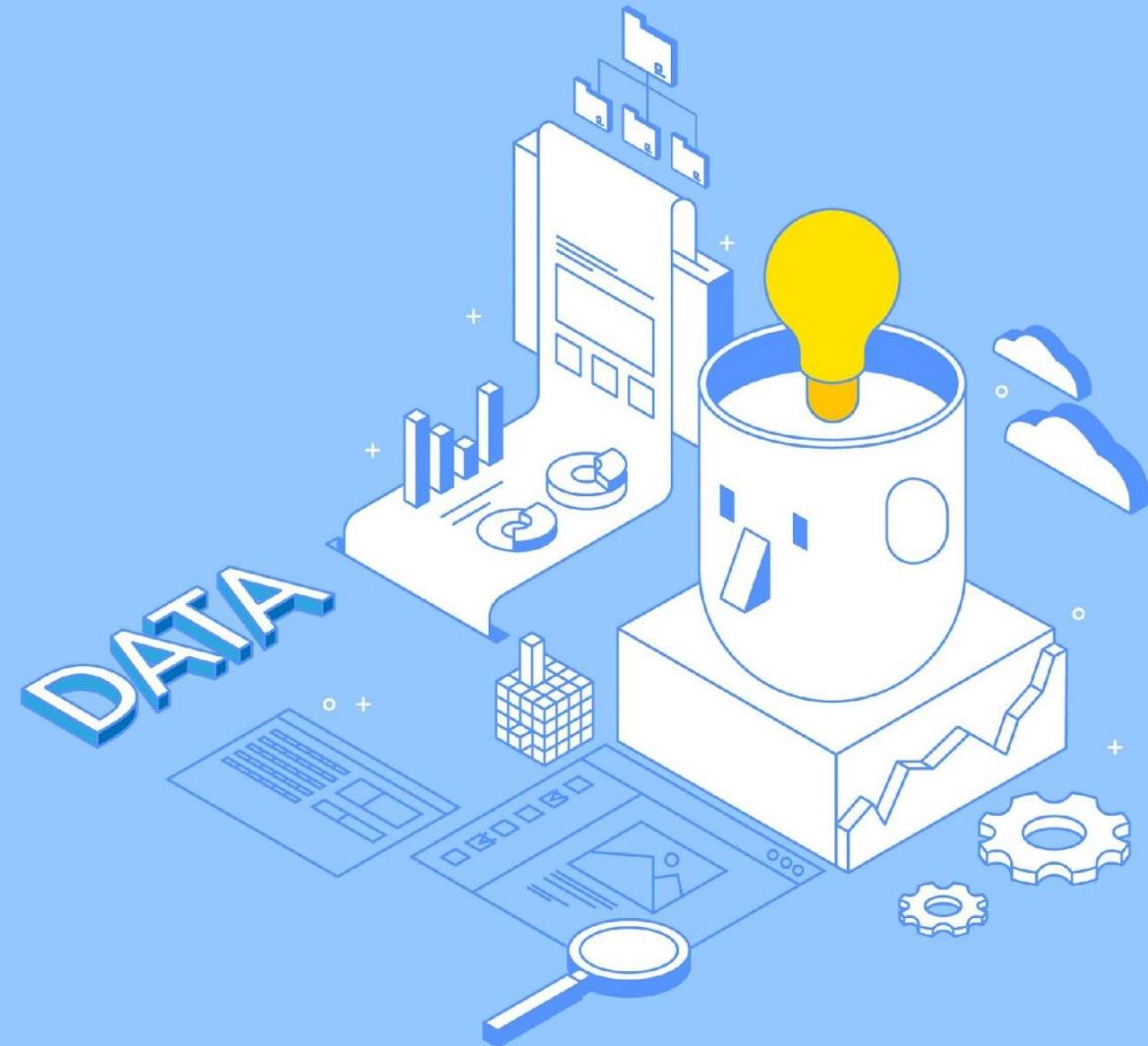
3. 분석진행

- sklearn, pandas, tensorflow, pytorch



2021 - 2 Big Data

Movies on OTT Platforms & MBTI



Contents



I Analyzing Previous Works

II Data Overview

III Proposed Methodology

Analyzing Previous Works



Recommendation System

56

- **Demographic Filtering** ex) based on movie popularity
 - >Sort the scores and recommend the best rated movie
 - > Not sensitive to a particular user
- **Content Based Filtering** ex) genre, director, description, actors
 - >Most likely to be similar
 - >scikit-learn, TfIdfVectorizer, TF-IDF matrix/CountVectorizer()
 - >euclidean, Pearson, cosine similarity scores/linear_kernel()
- **Collaborative Filtering** ex) based on persons matching
 - User based filtering** ex) similar users have liked
 - >pearson correlation/cosine similarity
 - Item Based Collaborative Filtering** ex) similarity with the items
 - >pearson correlation/cosine similarity
 - Surprise, SVD, Root Mean Square Error

Data Overview



Dataset

Movies on OTT platforms

A collection of movies found on Netflix, Prime Video, Hulu and Disney+

NETFLIX

Javagar M • updated a year ago (Version 1)

18



+ Additional Dataset

Dataset

Movie Character MBTI Dataset

Movie Character's Personality Dataset

Subin An • updated 9 months ago (Version 1)



Total 6 columns

- stat(좋아요)
- mbti
- enneagram(성격진단테스트)
- role
- movie
- img_url(캐릭터이미지)

Proposed Methodology



1. What kind of movie character would you be?

- based on movie character's MBTI
- using cosine similarity from sklearn
- sort by imdb + rotten tomatoes scores

2. Top 3 MBTI types of movie characters on OTT platform

- Which is the most common MBTI type of movie character on each OTT platform?



**Thank you
for
listening!**

