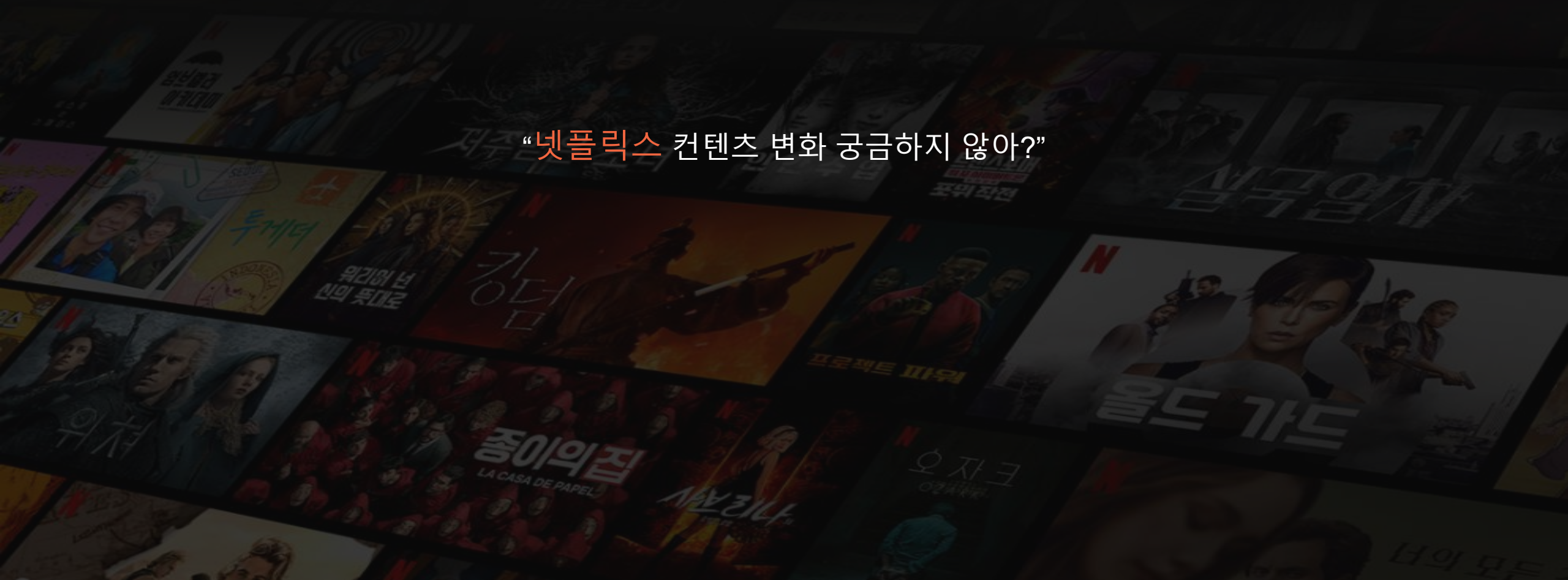


NETFLIX

오리지널 콘텐츠, 영화, 드라마, 예능까지!
넷플릭스 하나로 무제한!

“넷플릭스 콘텐츠 변화 궁금하지 않아?”



Data Analysis

데이터 분석 프로젝트

넷플릭스 전세계 콘텐츠 분석

Project team 김도훈, 박재용, 이하경, 최정근, 홍
성은
Submission Date 2021. 04. 03

First Team Project

목차

01

Project 준비

- 주제 및 목표
- 데이터 설명
- 데이터 상세 설명
- 분석 과정

02

데이터 처리

- 사용 라이브러리
- 결측치 제거

03

데이터분석

- 전체 데이터 분석
- 한국 데이터 분석

04

COVID-19와 상관관계

- 가설 설정
- 가설 1
- 가설 2
- 가설 3

01

Project 준비

- 주제 및 목표
- 데이터 설명
- 데이터 상세 설명
- 분석 과정

주제 및 목표

:Project 주제

주제

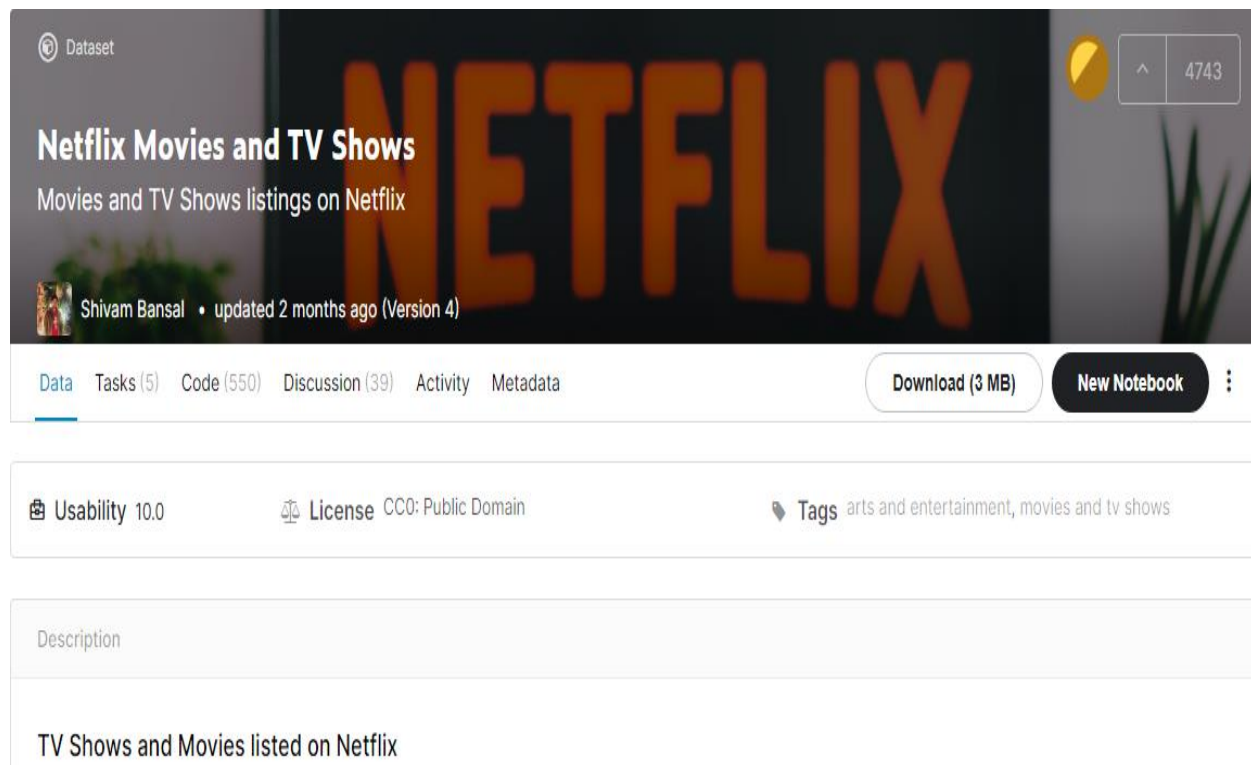
1997년부터 2021년까지 넷플릭스에 등록된 데이터를
활용하여 변화 및 코로나 시대의 콘텐츠 변화 분석

목표

- 1) 넷플릭스 데이터를 활용하여
코로나시기 콘텐츠 등록 상세 현황 분석
- 2)한국의 넷플릭스 콘텐츠 분석

Data 출처

:Project 주제



The screenshot shows the Kaggle dataset page for 'Netflix Movies and TV Shows'. The header features the dataset title, a description, and the creator's name (Shivam Bansal) with an update date. Below this is a navigation bar with tabs for Data, Tasks, Code, Discussion, Activity, and Metadata. A 'Download (3 MB)' button and a 'New Notebook' button are also visible. The 'Usability' is 10.0, the 'License' is CC0: Public Domain, and the 'Tags' are arts and entertainment, movies and tv shows. The 'Description' section is partially visible, stating 'TV Shows and Movies listed on Netflix'.

Dataset

Netflix Movies and TV Shows

Movies and TV Shows listings on Netflix

Shivam Bansal • updated 2 months ago (Version 4)

Data Tasks (5) Code (550) Discussion (39) Activity Metadata

Download (3 MB) New Notebook

Usability 10.0 License CC0: Public Domain Tags arts and entertainment, movies and tv shows

Description

TV Shows and Movies listed on Netflix

Kaggle

Netflix Movies and TV Shows

Netfliex 검색 엔진인 Flixable 에서 수집된 데이터 사용

- 출처 : <https://www.kaggle.com/shivamb/netflix-shows>

Data 설명

:Project 주제

Netflix

Data Set

- 1997년 부터 2021년 까지의 넷플릭스 상세 데이터

- Netflix 검색 엔진인 Flixable 에서 수집된 데이터

- 12개의 열 중 사용하지 않는 3개의 열 드롭 (show_id, director, description 열 드롭)

- 데이터는 총 type / title / cast / country / date_added / release_year / rating / duration / listed_in 으로 총 9개의 항목(열)로 구성

- 데이터 분석 중 필요한 열 추가하면서 사용

- 원본 데이터의 Column

show_id	type	title	director	cast	country	datd_added	release_year	rating	duration	listed_in	description
---------	------	-------	----------	------	---------	------------	--------------	--------	----------	-----------	-------------

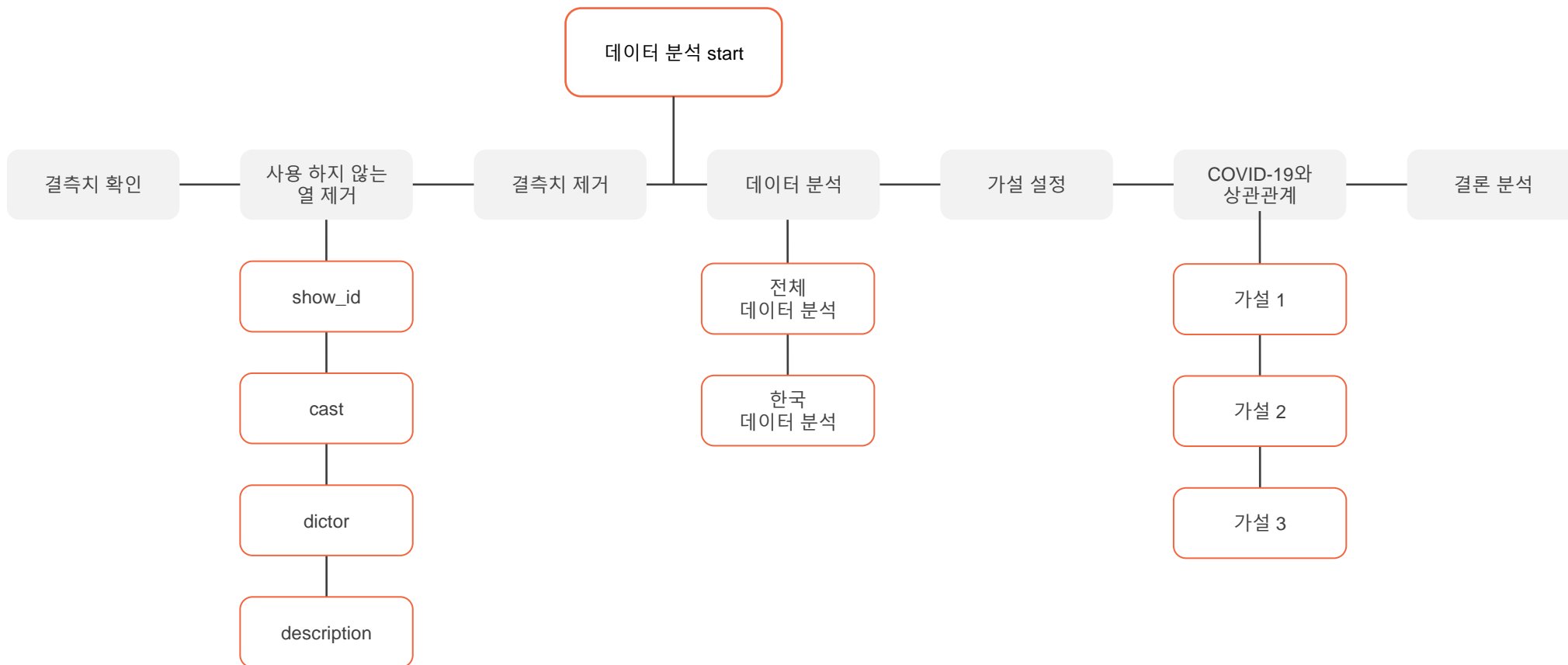
Data 상세 설명

:Project 주제

Column	Explanation	Example Value
type	유형	TV Show
title	제목	Cheese in the Trap
cast	출연자	Park Hae-Jin, Kim Go-eun...
country	나라	South Korea
datd_added	컨텐츠 등록 날짜	01-October-19
release_year	컨텐츠 출시 날짜	2016
rating	시청 연령 (미국 표기법)	TV-14
duration	방영 기간 & 상영 시간	1 Season
listed_in	관련 카테고리 (장르)	International TV Shows, Romantic TV show...

데이터 분석 과정

:데이터 처리



02

데이터 처리

- 사용 라이브러리
- 결측치 처리 과정
- 결측치 제거

사용 라이브러리

:데이터 처리

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import missingno as msno
from matplotlib import font_manager, rc
import matplotlib.font_manager as fm
import plotly.graph_objects as go
import plotly.express as px
import plotly as py
import folium
import datetime
```

```
import os
from plotly.offline import init_notebook_mode, iplot
py.offline.init_notebook_mode(connected = True)
import matplotlib.colors
from collections import Counter

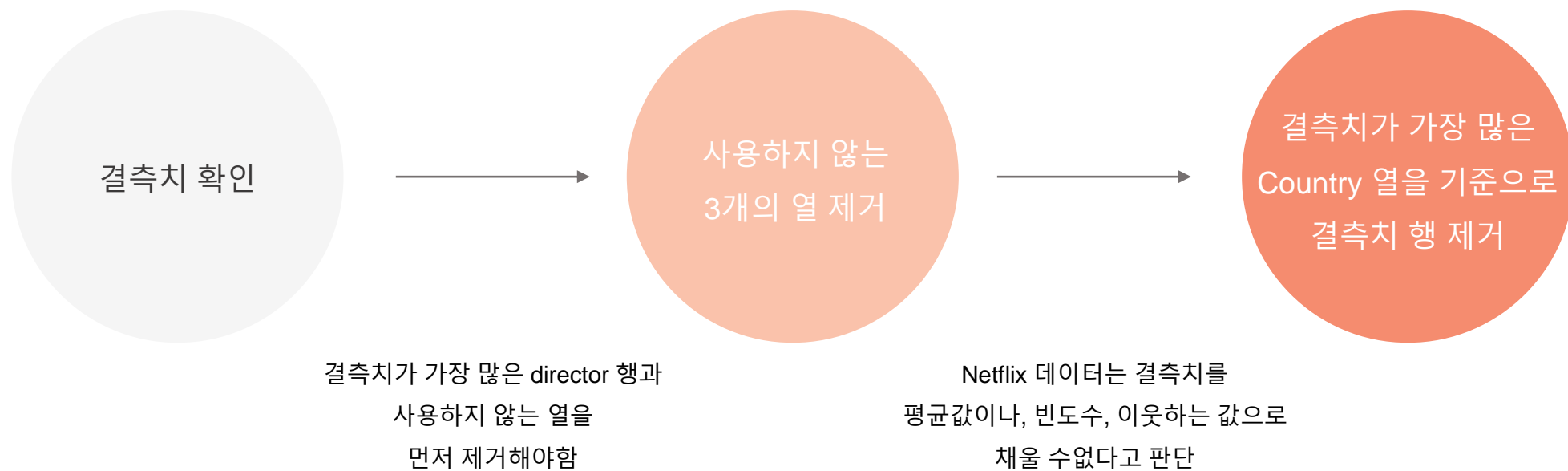
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans, AffinityPropagation
from sklearn.preprocessing import MultiLabelBinarizer

import warnings
warnings.filterwarnings("ignore")
```

```
%matplotlib inline
```

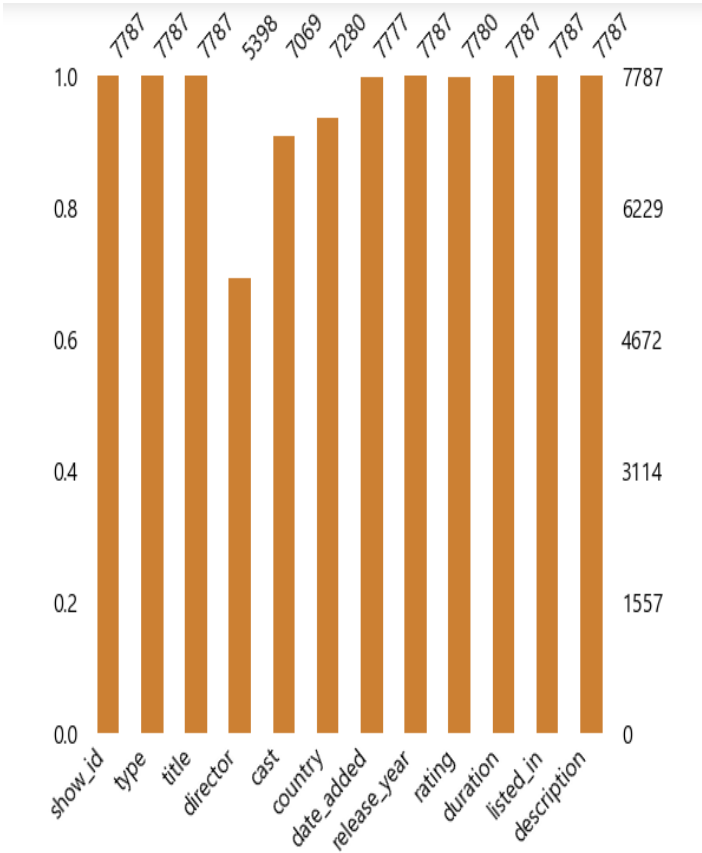
결측치 처리 과정

:데이터 처리

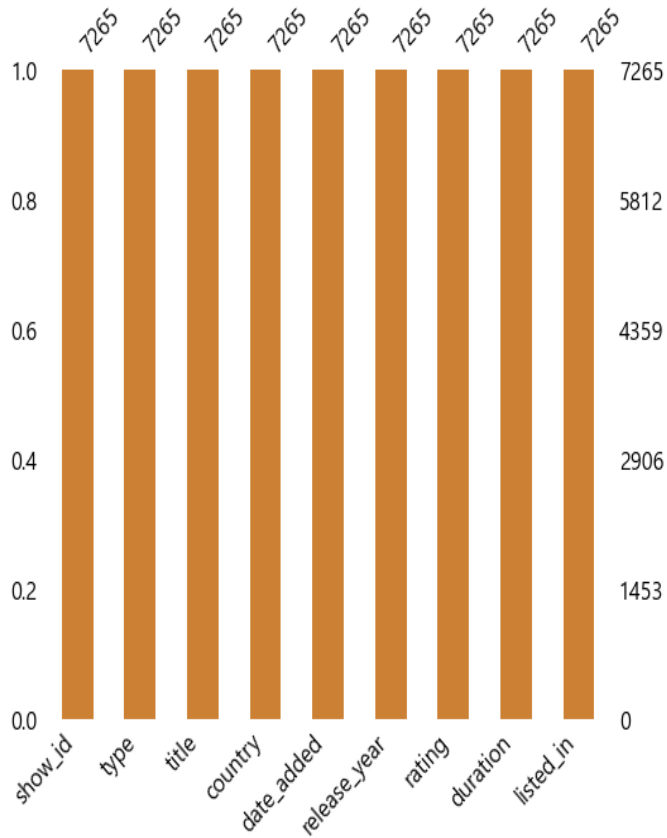


결측치 제거

:데이터 처리



결측치 제거



03

데이터 분석

- 전체 데이터 분석
- 한국 데이터 분석

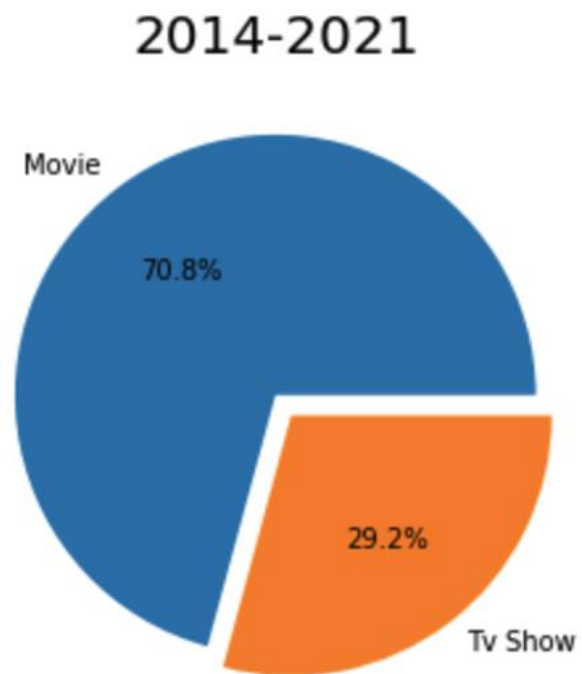
03

데이터 분석

- 전체 데이터 분석
- 한국 데이터 분석

Type

:전체 데이터 분석



유형(TV Show, Movie)별 콘텐츠 비율

- 2014년 ~ 2021년의 TV Show와 Movie의 비율을 나타낸 원 그래프

- Movie 70.8%

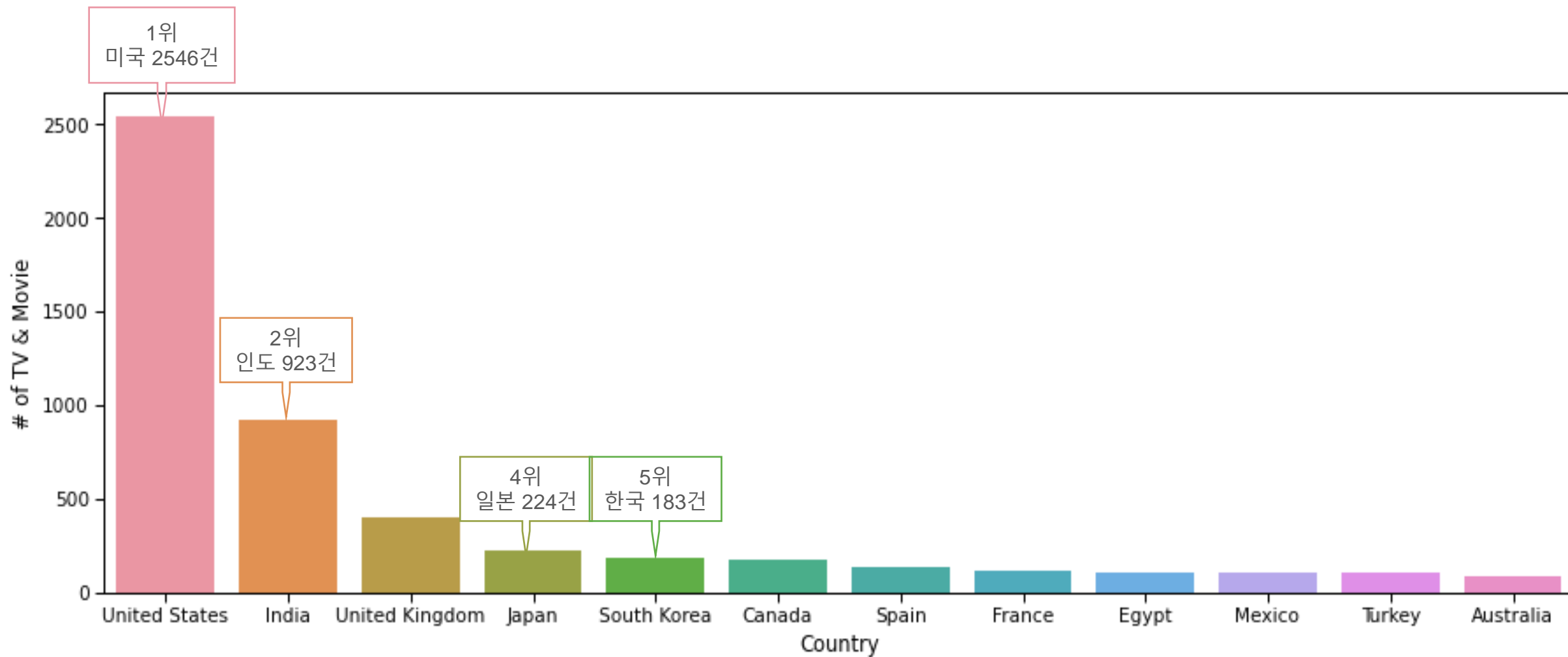
Tv Show 29.2%

- 7 : 3 의 비율

Movie의 비율이 TV Show의 비율에 비해 높게 나타난다.

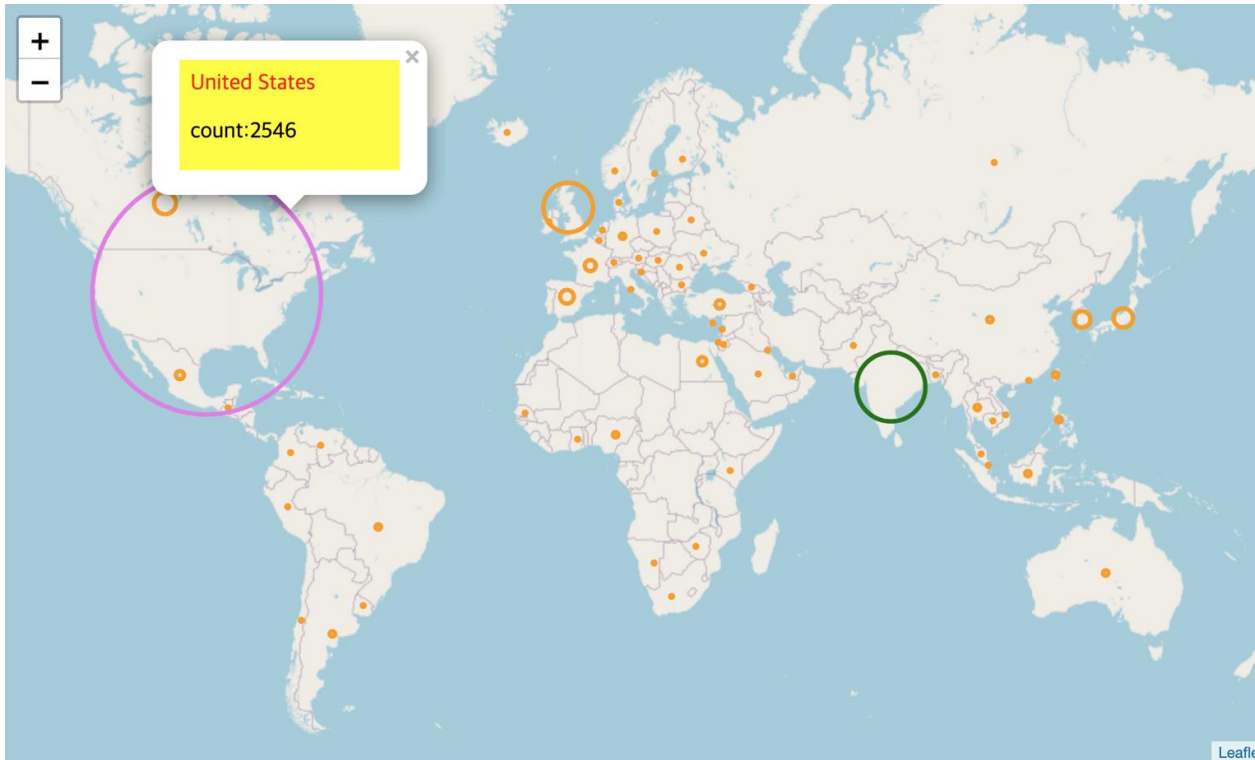
Country

:전체 데이터 분석



Country (지도로 표현)

:전체 데이터 분석

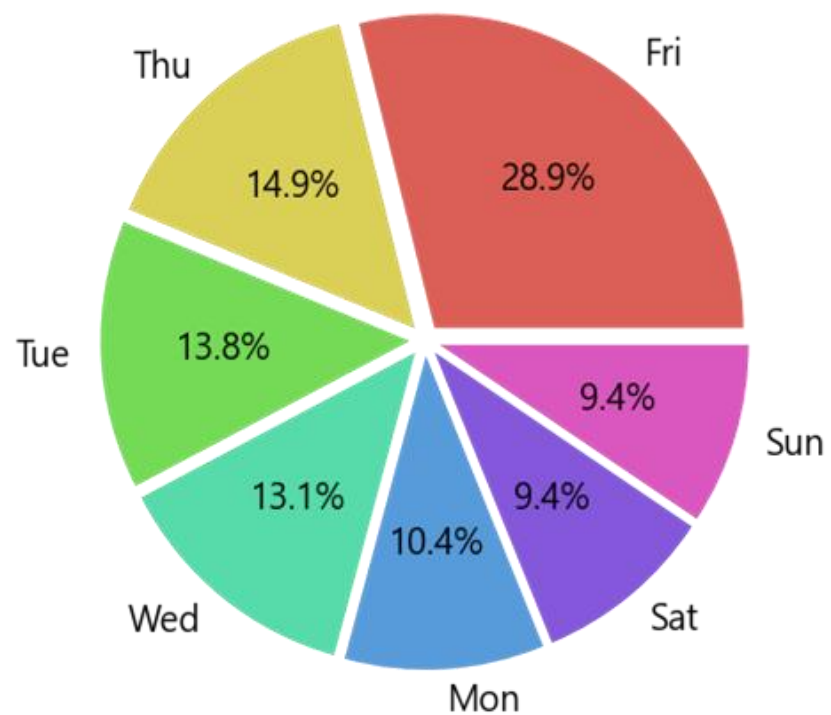


나라별 콘텐츠 수

- 미국이 2500건 이상으로 가장 압도적인 콘텐츠 수를 차지
이어서 인도, 영국의 콘텐츠 생산 비율이 높다.
- 영화 산업이 발달되어 있는 인도, 한국, 일본 등 아시아 권
에서 콘텐츠 등록이 많은 것으로 보아 아시아에서 큰 폭으
로 성장을 한 것으로 분석
- 아프리카권은 인터넷 네트워크 사용의 어려움으로 인해
넷플릭스 등록 콘텐츠가 적은 것으로 분석

Date Added

:전체 데이터 분석



요일별 콘텐츠 출시 비율

- 콘텐츠 등록은 금요일이 28.9%로 가장 많고, 토요일, 일요일이 9.4%로 가장 적다.
- 주말에 사람들의 관심을 끌 수 있도록 금요일 콘텐츠 등록 비율이 주말의 2배 이상이다.

Date Added (crosstab 분석)

:전체 데이터 분석

TV Show 요일 및 월별 출시량

day_week	Fri	Mon	Sat	Sun	Thu	Tue	Wed
month							
1	73	12	7	11	18	18	23
2	52	6	15	7	15	9	13
3	69	5	11	10	19	20	23
4	42	17	19	11	19	15	20
5	56	19	13	6	13	17	8
6	48	17	25	4	14	7	9
7	52	19	28	8	13	10	22
8	56	20	16	3	29	25	7
9	55	5	11	19	24	24	17
10	55	10	21	15	32	27	16
11	78	8	10	19	17	29	23
12	59	20	24	27	37	37	20

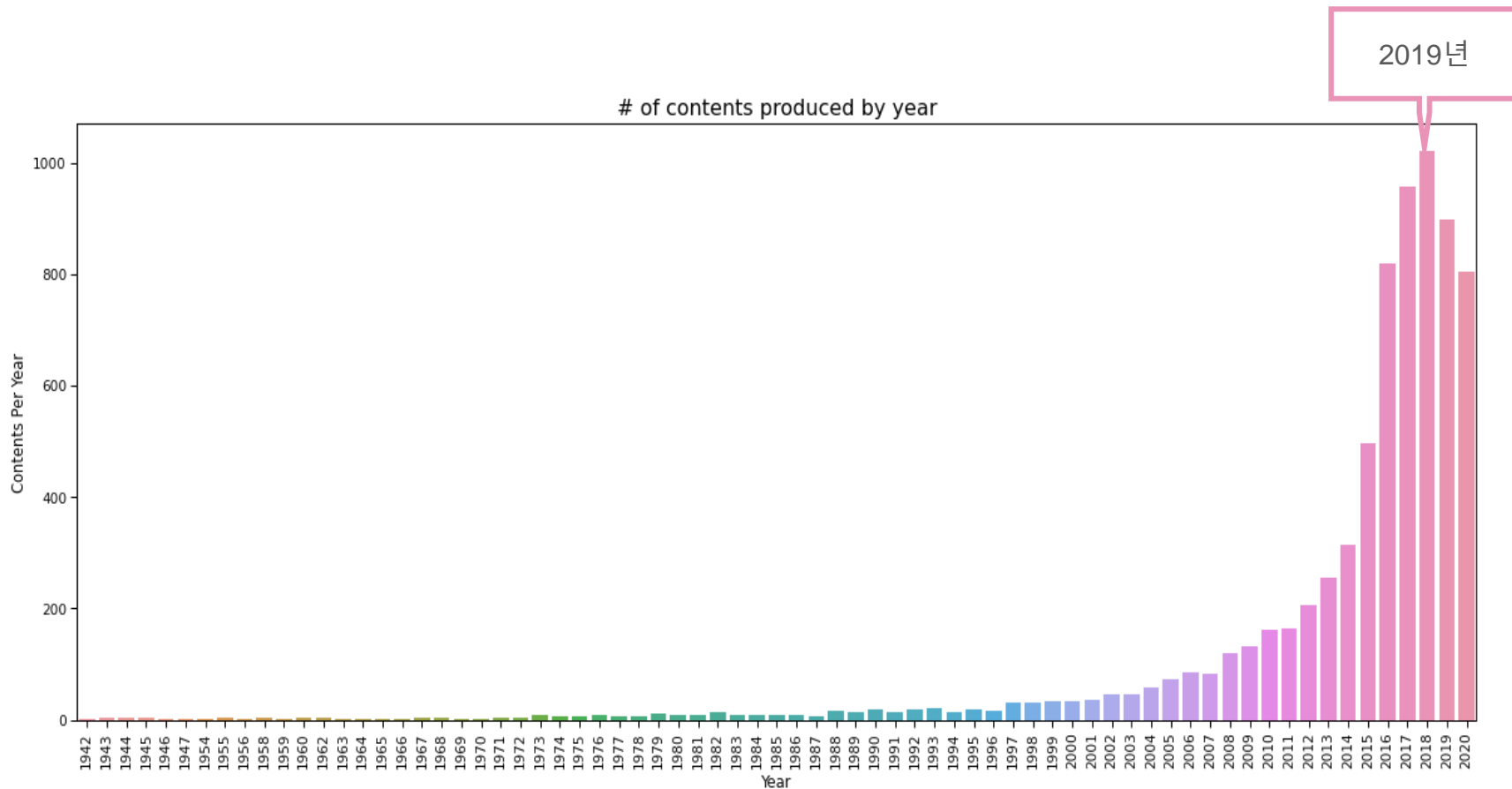
Movie 요일 및 월별 출시량

day_week	Fri	Mon	Sat	Sun	Thu	Tue	Wed	All
month								
1	118	73	25	34	36	80	136	502
2	94	18	25	21	45	28	60	291
3	147	18	20	24	128	42	36	415
4	79	51	49	43	31	42	73	368
5	75	44	22	15	53	69	40	318
6	82	39	50	17	80	38	26	332
7	72	56	38	92	38	29	40	365
8	68	26	41	16	110	49	53	363
9	96	25	51	44	43	55	38	352
10	92	97	50	69	51	102	29	490
11	156	26	27	36	109	42	76	472
12	107	45	54	73	50	130	34	493
All	1186	518	452	484	774	706	641	4761

TV Show와 Movie 에 상관 없이 집에 있는 시간이 늘어나는 연초와 연말, 금요일에 출시가 가장 많다.

Release Year

:전체 데이터 분석



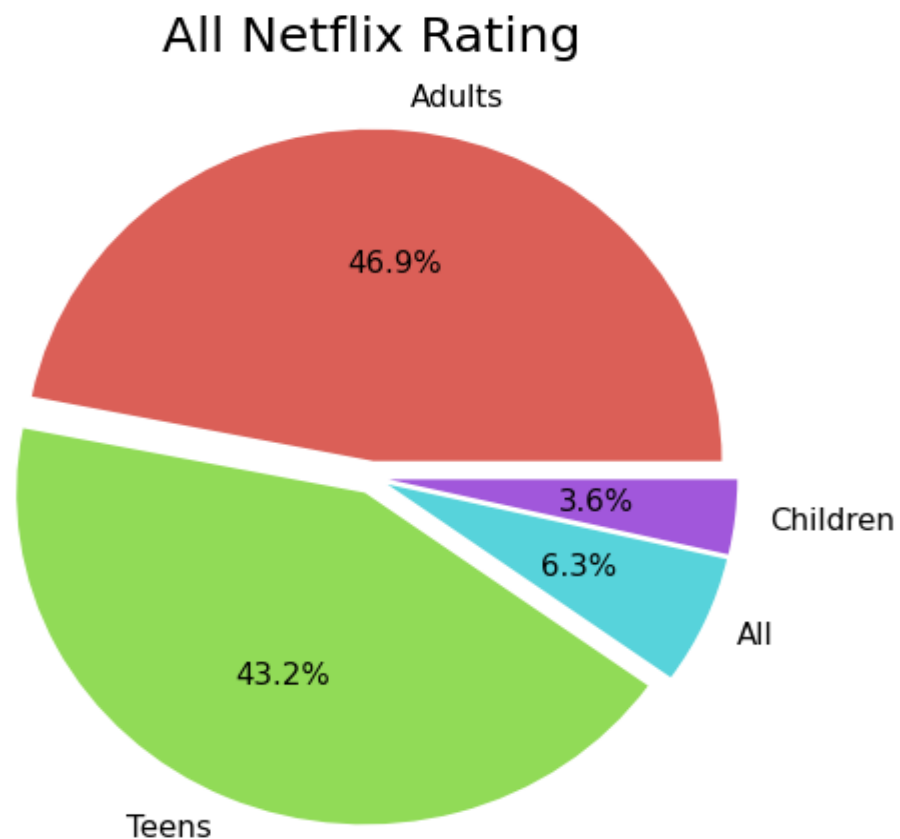
연도별 콘텐츠 출시 수

- 2015년 부터 2018년까지 넷플릭스의 콘텐츠 생산은 기하급수적으로 늘어나는 추세를 보인다.

- 2019년부터 2020년까지 넷플릭스 콘텐츠 생산이 줄어드는 모습을 볼 수 있다.

Rating

:전체 데이터 분석



시청 등급(Rating)별 콘텐츠 비율

- 성인 콘텐츠와 10대 콘텐츠가 압도적인 비율로 높다.
- 성인과 10대 시청 등급 콘텐츠는 총 90.1%
어린 아이들을 위한 콘텐츠는 총 9.9%
- 10 : 1 의 비율

Rating 등급 설명

:전체 데이터 분석

TV-MA	:	17세 미만의 어린이 혹은 청소년한테 부적절한 프로그램
TV-14	:	14세 미만의 어린이 혹은 청소년이 시청하려면 보호자 지도가 권장되는 프로그램
TV-PG	:	어린이가 시청하려면 보호자 지도가 권장되는 프로그램
R	:	일반적 성인 영화. 17세 미만의 경우 부모의 동반 필수
PG-13	:	전체 관람가지만, 부모의 주의가 요구되며 13세 미만에게는 적절하지 않음
TV-Y	:	영유아를 위한 프로그램
TV-Y7	:	7세 이상 어린이를 위한 프로그램
PG	:	전체 관람가지만, 폭력성이 존재하므로 어린이의 경우 보호자의 지도가 요구
TV-G	:	모든 연령이 시청할 수 있는 프로그램. 다만 어린이를 대상으로 하지는 않았다
NR	:	NR은 Not Rated로 등급을 받지 않은 경우
G	:	일반용
TV-Y7-FV	:	7세 이상의 어린이를 위한 프로그램. 매우 경미한 가상의 폭력 묘사가 있음.
UR	:	매겨진 등급을 거부
NC-17	:	수위가 매우 높은 성인 영화. 18세 미만 관람불가

Duration

:전체 데이터 분석



시즌별 콘텐츠 수

- 넷플릭스 콘텐츠는 단기간 시청할 수 있는 콘텐츠가 대부분이다
- 장기 시리즈 프로그램은 거의 존재하지 않는다..
- 시즌제 프로그램이 약소하게 운영되고 있는 것을 알 수 있다.

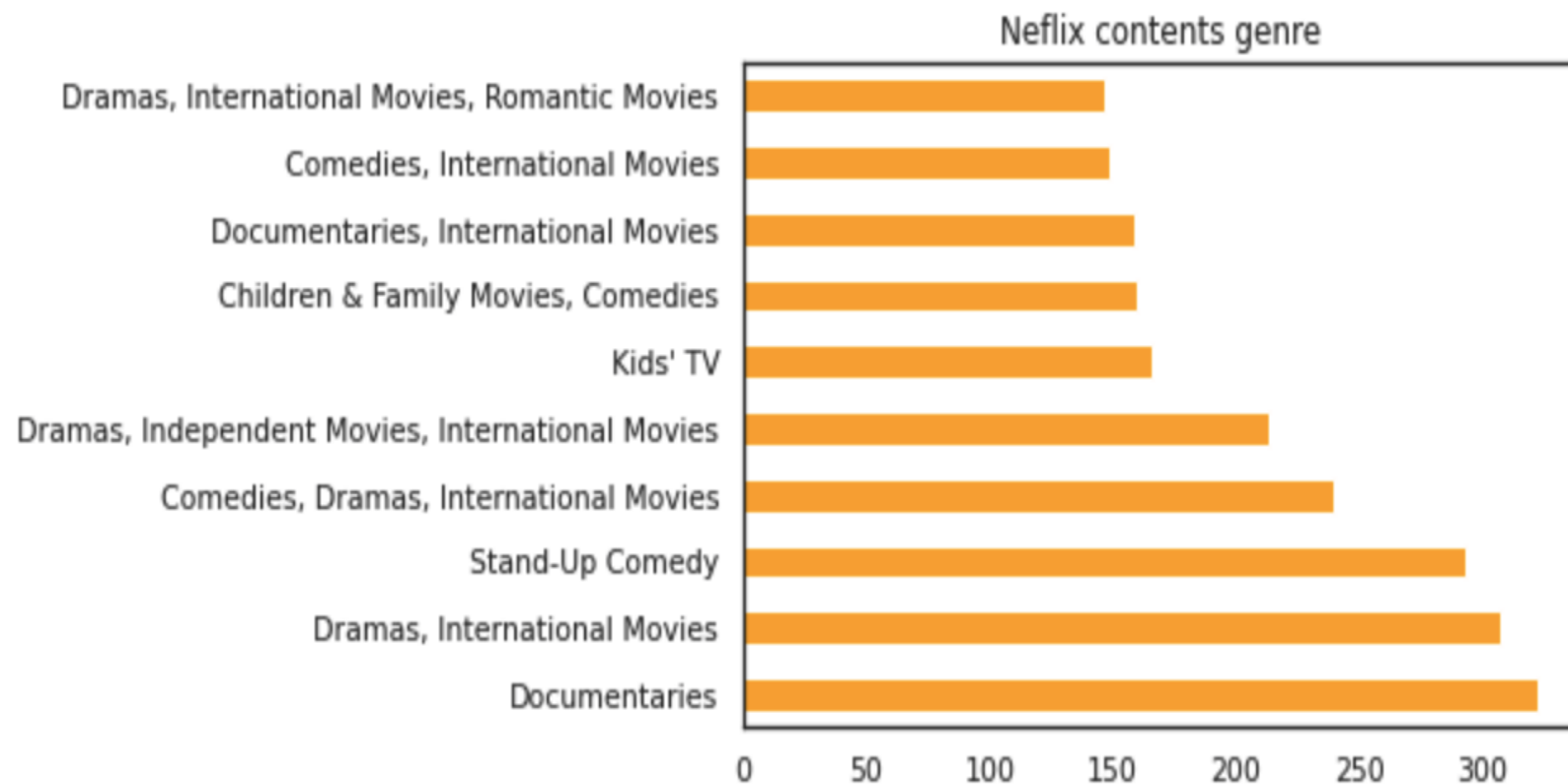
Listed in

:전체 데이터 분석

장르별 콘텐츠 분석

- 2014년 ~ 2020년까지 넷플릭스에서 가장 많이 등록된 장르
: 다큐멘터리 (322개 등록)

- 가족 모두가 볼 수 있는 장르인
Kids TV / Family장르는 6위, 7위에 위치

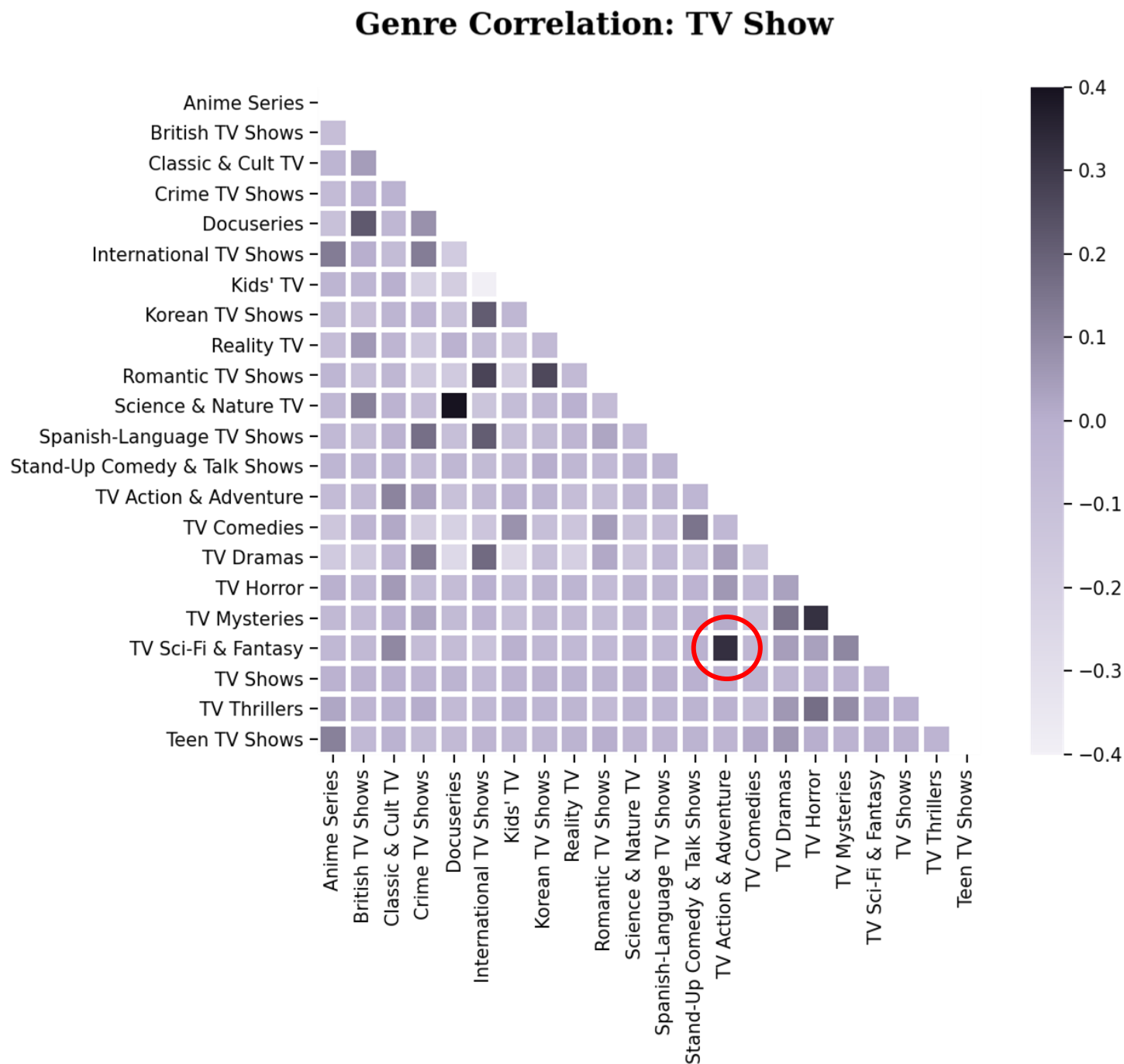


:전체 데이터 분석

- 하나의 콘텐츠 당 1 – 4개로 구성된 관련 장르들을 하나씩 세분화 한 후 분석
- 넷플릭스에 등록된 전세계의 Movie 콘텐츠 중에서
- 드라마 형식의 독립영화와,
드라마 형식의 세계영화의 비율이 가장 높다.

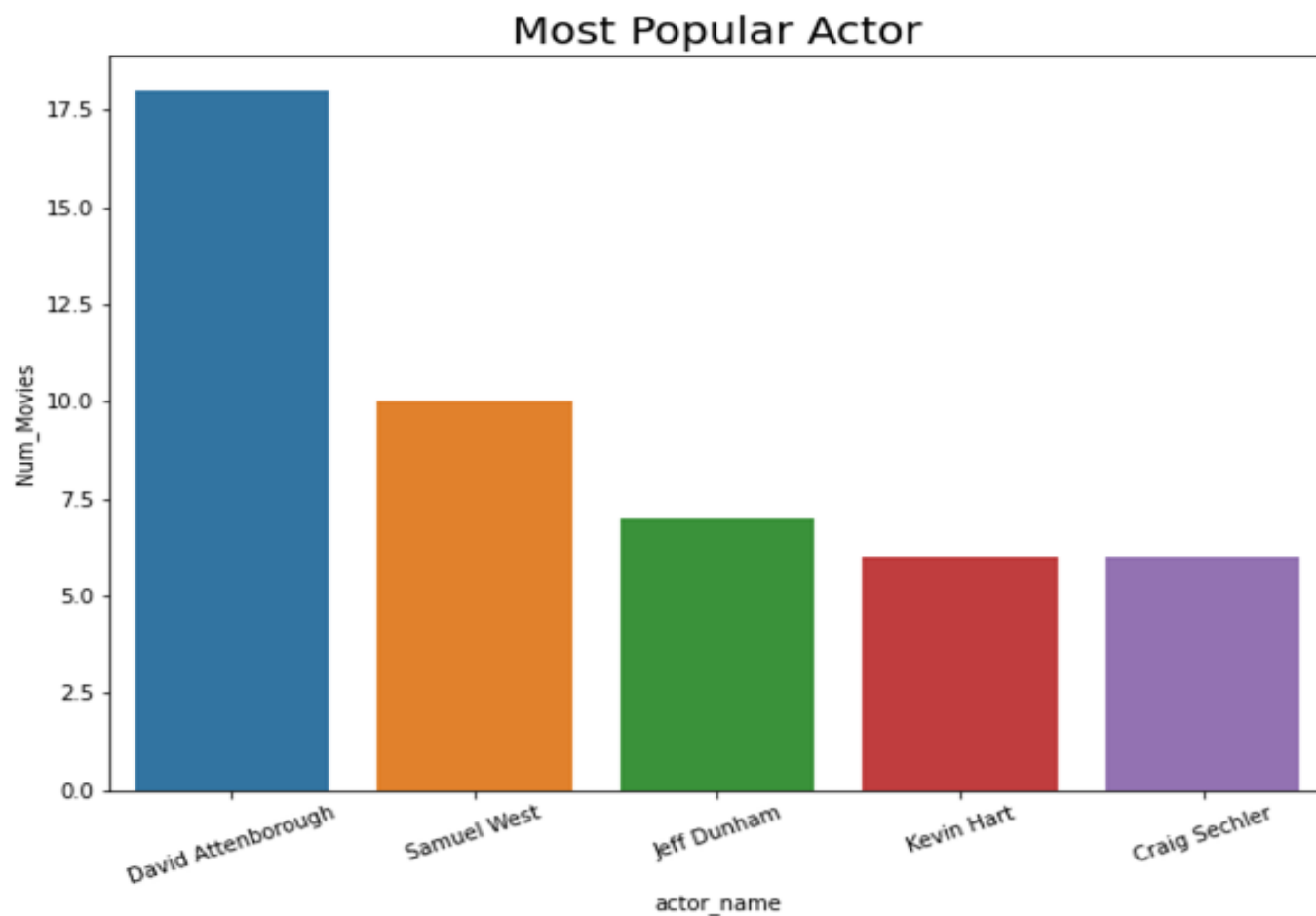
:전체 데이터 분석

- 하나의 콘텐츠 당 1~4개로 구성된 관련 장르들을 하나씩 세분화 한 후 분석
- 넷플릭스에 등록된 전세계의 TV Show 콘텐츠 중에서
- 다큐멘터리 시리즈의 비율이 높게 나타났고, 공상과학과 판타지 콘텐츠를 다룬 액션과 모험 시리즈의 비율이 높게 나타났다.



Cast

:전체 데이터 분석



Question) Netflix 콘텐츠에 출연한 배우 중 우리가 알고 있는 배우가 많이 있을까?

Answer) 가장 많이 출연한 배우 순으로 나열했을 때, 국내 배우 및 국내에서 유명한 해외 배우는 없는 것으로 나타났다

03

데이터 분석

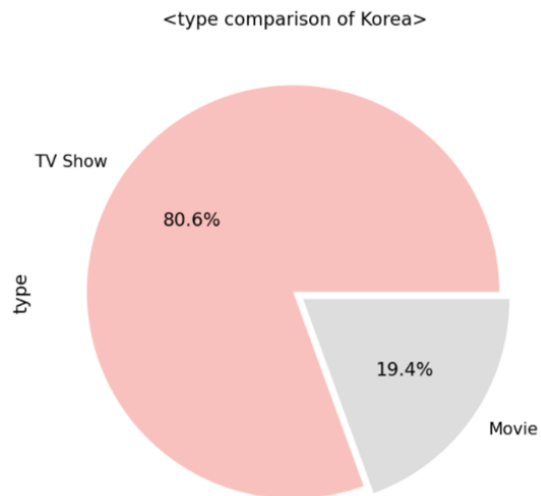
- 전체 데이터 분석
- 한국 데이터 분석

Type 콘텐츠 비교

:한국 데이터 분석

Type 분석

한국의 유형별 콘텐츠 비율

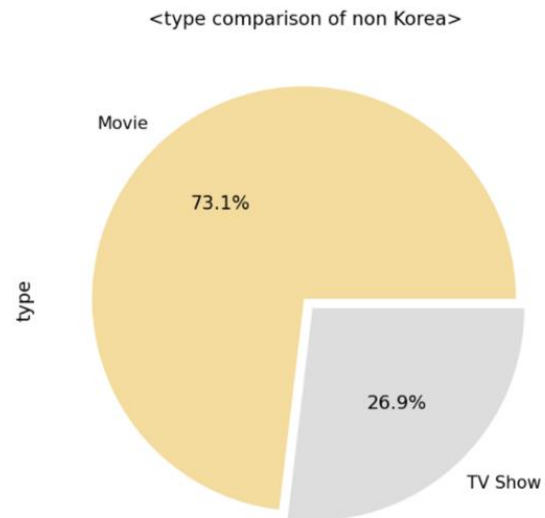


한국

Movie 80.6% , Tv Show 19.4%

약 4 : 1 의 비율

TV Show 의 비율이 Movie 의 비율에 비해 높게 나타난다.



NOT 한국

Movie 73.1% , Tv Show 26.9%

약 4 : 1 의 비율

Movie 의 비율이 TV Show 의 비율에 비해 높게 나타난다.

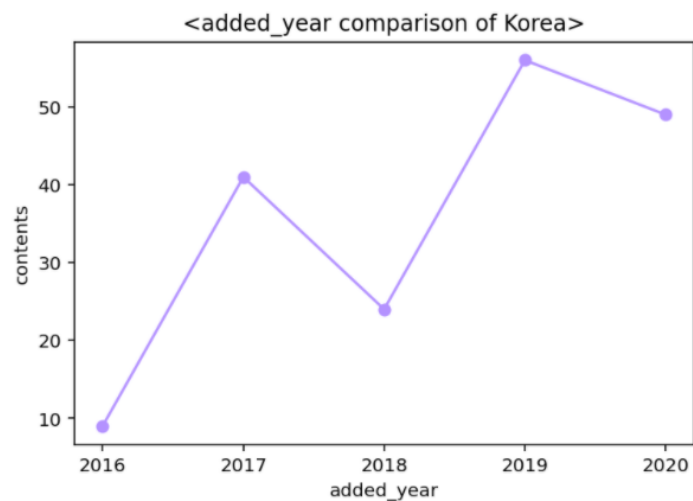
전 세계와 한국은 정 반대의 유형 분포를 보인다.

Date Added 콘텐츠 비교

:한국 데이터 분석

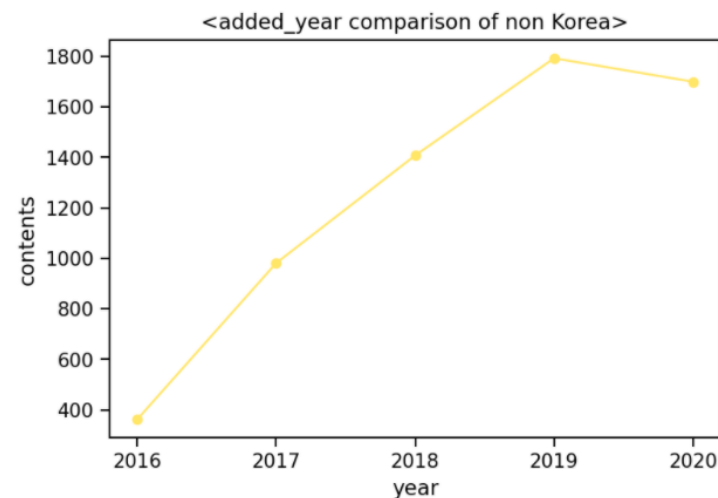
Date Added 분석

한국의 등록 년도별 콘텐츠 비율



한국

2018년 최저, 2019년 최고의 콘텐츠 등록수를 기록
2019년에 비해 2020년에 살짝 감소하지만.
다른 연도와 비교하면 여전히 높은 비율을 차지



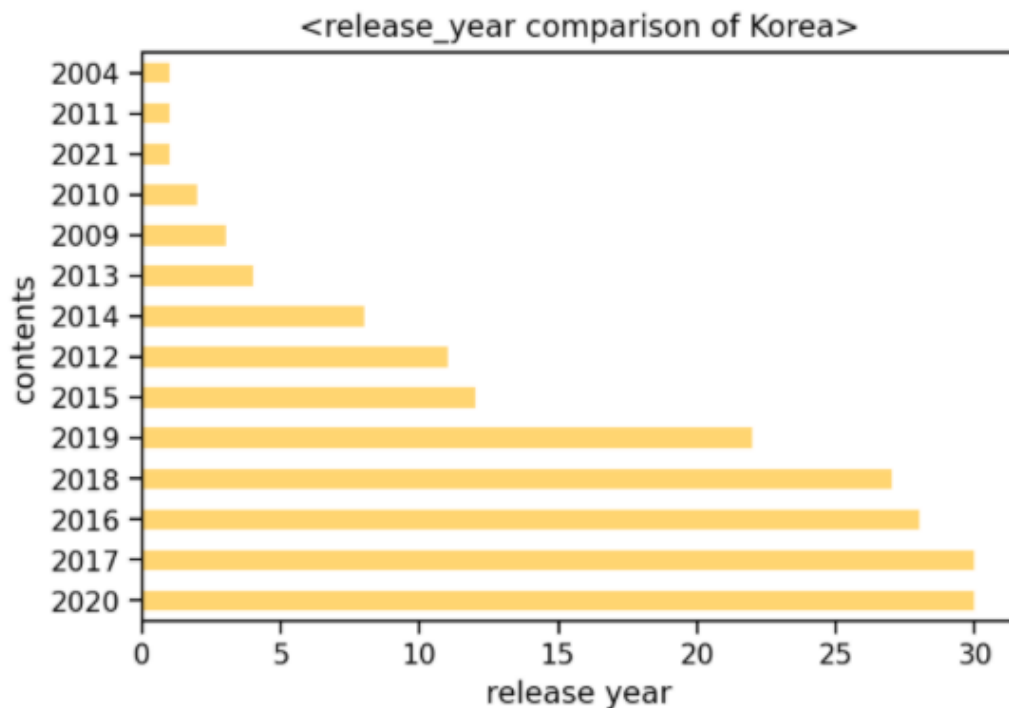
NOT 한국

2016년 이후 꾸준한 증가
2019년 최고의 콘텐츠 등록수를 기록
2019년에 비해 2020년에 살짝 감소하는 형태

한국과 그 밖의 나라 모두 2019년에 절정을 달성한 후, 소폭 감소하는 추세를 보인다.

Release Year 콘텐츠 비교

:한국 데이터 분석



Listed in 분석

출시 년도별 콘텐츠 분석

한국

넷플릭스에 등록된 콘텐츠는
2020년에 출시된 콘텐츠가 가장 많다.

2017년도 2020년과 비슷한 콘텐츠 수를 보이며,
2016년, 2018년 순서로 많다.

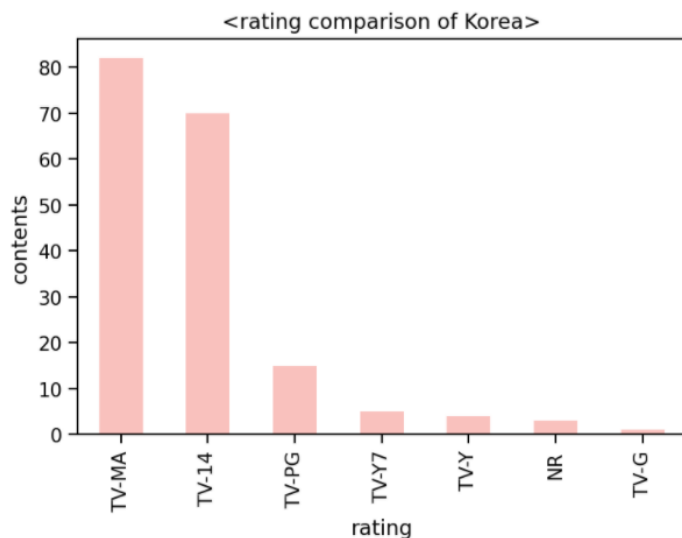
Date_Added 분석에서 2019년에 콘텐츠 등록이 가장 많은 것에 비해,
넷플릭스에 등록된 콘텐츠는 2017년에 가장 많이 출시되었다.

Rating 콘텐츠 비교

:한국 데이터 분석

Type 분석

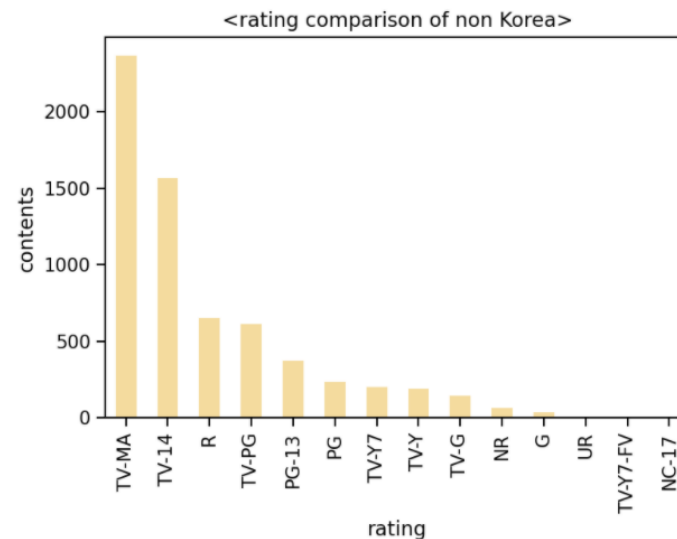
시청 등급별 콘텐츠 비율



한국

17세 미만의 경우 부적절한 프로그램인 TV-MA 콘텐츠가
가장 많다

TV1-14, TV-PG 등 청소년이 볼 수 있는 콘텐츠가 그 다음을 차지



NOT 한국

한국을 제외한 나라 역시 TV-MA 등급 콘텐츠가 가장 많다.

한국 콘텐츠에는 없는 R등급의 비율이 3번째로 높다.

- R : 일반적인 성인영화. 17세 미만의 경우 부모 동반 필수

한국보다 그 밖의 나라에서 선정적이고 폭력적인 콘텐츠의 수가 많다.

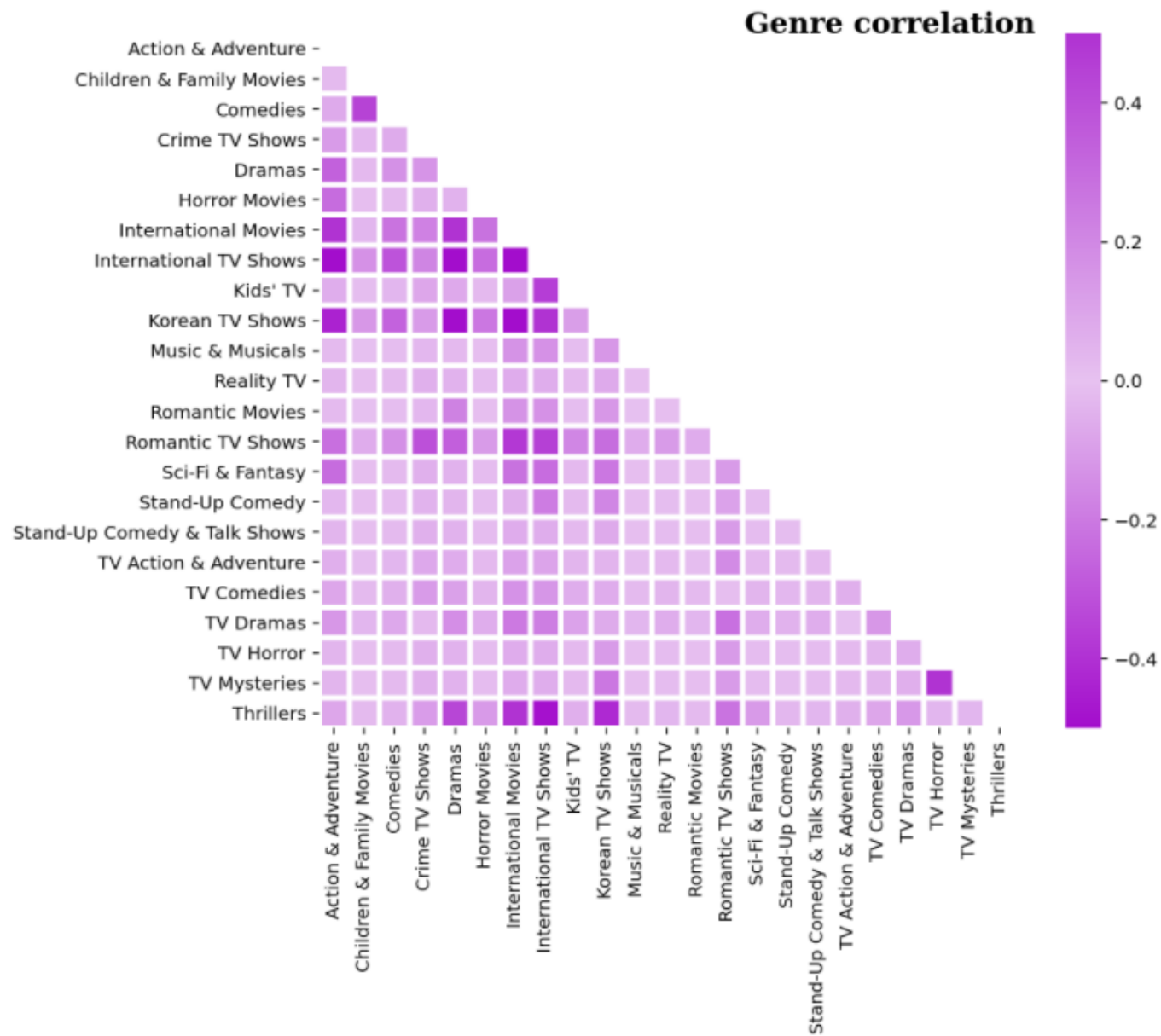
:한국 데이터 분석

장르별 콘텐츠 분석

넷플릭스에 등록된 한국의 콘텐츠는

Korean TV Shows,

비율이 가장 높다.



04

COVID-19와 상관관계

- 가설 설정
- 가설 1
- 가설 2
- 가설 3
- 가설에 대한 결론 분석

04

COVID-19와 상관관계

- 가설 설정
- 가설 1
- 가설 2
- 가설 3
- 가설에 대한 결론 분석

가설 설정

:Project 주제

가설 설정

코로나로 인한 넷플릭스 변화 예상

가설 1



집에 있는 시간이 증가함에 따라 단편영화 보다는 지속성이 있는 시리즈물을 더 많이 보지 않을까?

□ Movie 콘텐츠보다 TV Show 콘텐츠 등록이 증가했을 것으로 예상

가설 2



넷플릭스에 선정적인 콘텐츠가 많다. COVID-19 시대와 상관없이 선정적 콘텐츠는 계속 유지되지 않을까?

□ Rating , Listed in 카테고리 분류 시 성인 기준의 콘텐츠가 많을 것이라고 예상

가설 3



코로나 대유행으로 넷플릭스에 등록하는 콘텐츠의 수가 감소하지 않을까?

□ 코로나 대유행으로 콘텐츠 등록수가 급감할 것으로 예상

04

COVID-19와 상관관계

- 가설 설정

- 가설 1 / 집에 있는 시간이 증가함에 따라 단편 영화 보다는 지속성이 있는 시리즈물을 더 많이 보지 않을까?

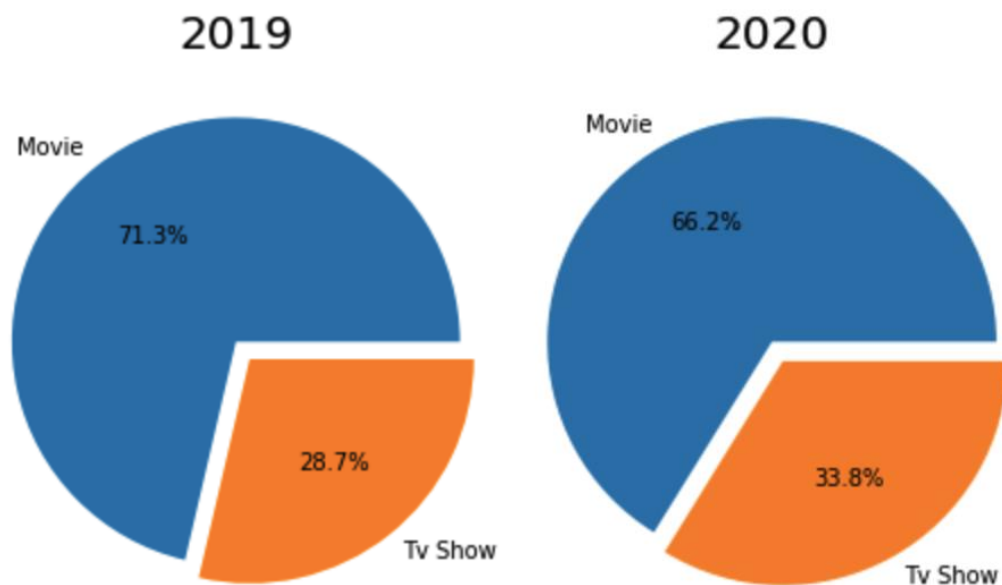
- 가설 2

- 가설 3

- 결론

가설 1

: COVID-19와 상관관계



2019년

- Movie의 비율이 소폭
(0.5%) 증가하며 콘텐츠
점유 우위

2020년

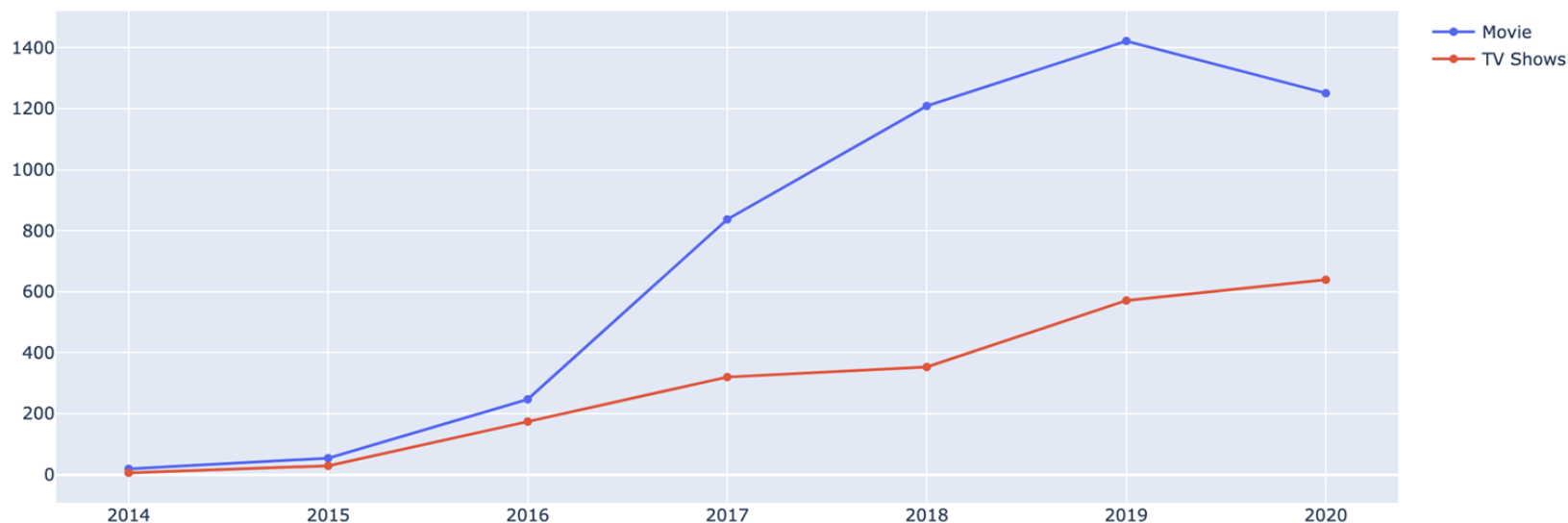
- 2020년도 TV Show의
점유율 33.8%로
2019년 대비 5.1% 증가

- 코로나 이후
TV Show의 점유율 증가

가설 1

: COVID-19와 상관관계

Contents per year



2014년 이후 매년 Movie 콘텐츠가 증가세였지만,
2020년 감소세를 보임

TV Shows 콘텐츠는 지속적으로 점유율 증가



COVID-19로 많은 사람들이 집에 있는 시간이 늘어나 단편영화보다는
지속적으로 볼 수 있는 시리즈물에 대한 관심이 증가하면서,

넷플릭스에서도 이런 추세를 반영하여
Movie의 등록을 소폭 감소, TV Shows의 등록을 소폭 증가시킨 것으로
분석

04

COVID-19와 상관관계

- 가설 설정

- 가설 1

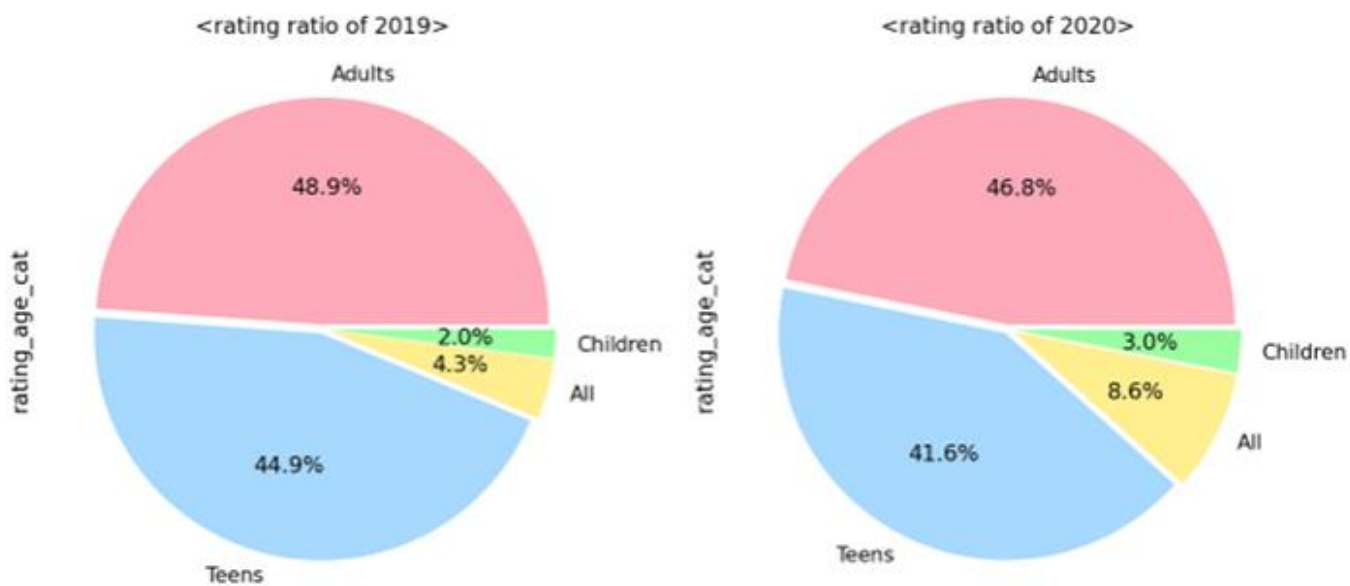
- 가설 2 / 넷플릭스에 선정적인 콘텐츠가 많다. COVID-19 시대에도 선정적 콘텐츠는 계속 유지되지 않을까?

- 가설 3

- 결론

가설 2 (rating)

: COVID-19와 상관관계



2019년

- 성인 이상과 청소년
관람 등급이 월등히 높
음

- 전체 관람가와 7세 이
상 연령 등급은 전체의
10분의 1로 매우 적은
비율

2020년

- 여전히 성인 등급의
비율이 월등히 많음

- 성인 이상, 청소년 관
람 등급 비율이 소폭 감
소

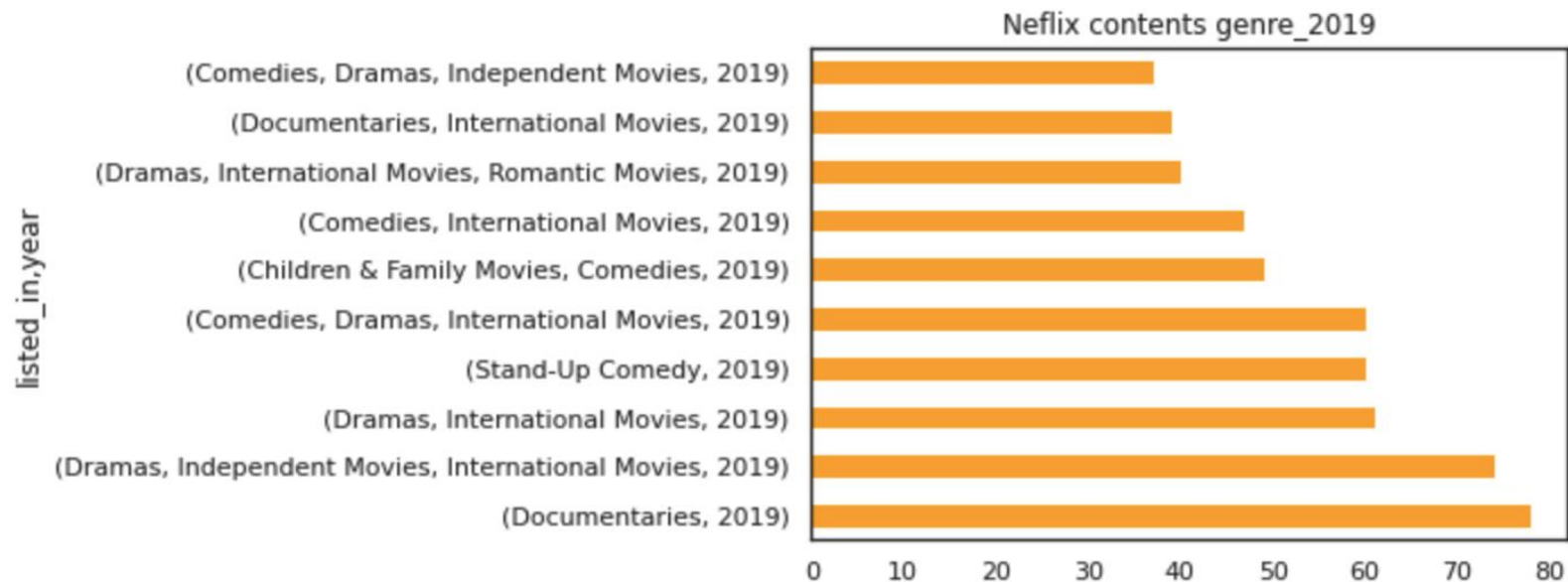
- 전체 관람가와 7세
이상(Children)의 비율
이 소폭 상승

가설 2 (listed_in)

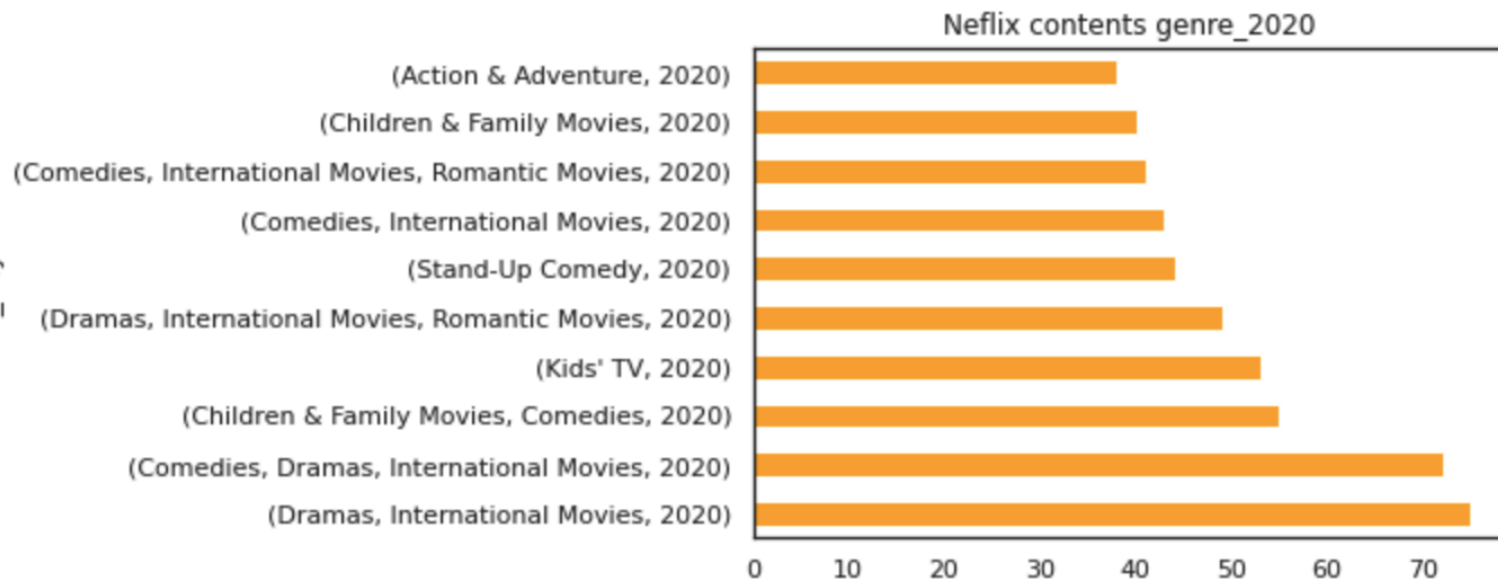
: COVID-19와 상관관계

2019년 장르별 분석

>>



listed_in,year



<<

2020년 장르별 분석

가설 2 (listed_in)

: COVID-19와 상관관계

2019년

- 2019년 넷플릭스에서 가장 많이 등록된 장르: 다큐멘터리 (78개 등록)

- 가족 모두가 볼 수 있는 장르인 Family 장르는 6위, 반면 kids'TV 장르는 상위 10위 안에 포함 안됨

? 2020년 코로나 시대 등록된 장르는 2019년과 어떤 변화가 있을까?

2020년

- 2020년 넷플릭스에서 가장 많이 등록된 장르
: 드라마장르(국제영화부분) 75개 등록

- 가족 모두가 볼 수 있는 장르인 Family 장르는 2위, 2019년 상위 10위 안에 포함이 안 되었던 kids'TV 장르가 4위에 위치하며, 2020년 TV Show 유형 증가와 함께 Children & Family, kids'TV 콘텐츠 등록의 성장이 유의미한 지표로 분석 된다.

- Covid19로 인해 아이들이 밖에 나가지 못함에 따라 아이들이 볼 수 있는 콘텐츠의 필요성이 대두되었을 것으로 예상되며,
넷플릭스에서도 이를 반영해 kids'TV 콘텐츠 등록을 많이 한 것으로 분석 된다.

- kids'TV가 많아진 것으로 보아 선정적인 콘텐츠가 많은 넷플릭스 콘텐츠에도 변화 시작된 것으로 보인다.

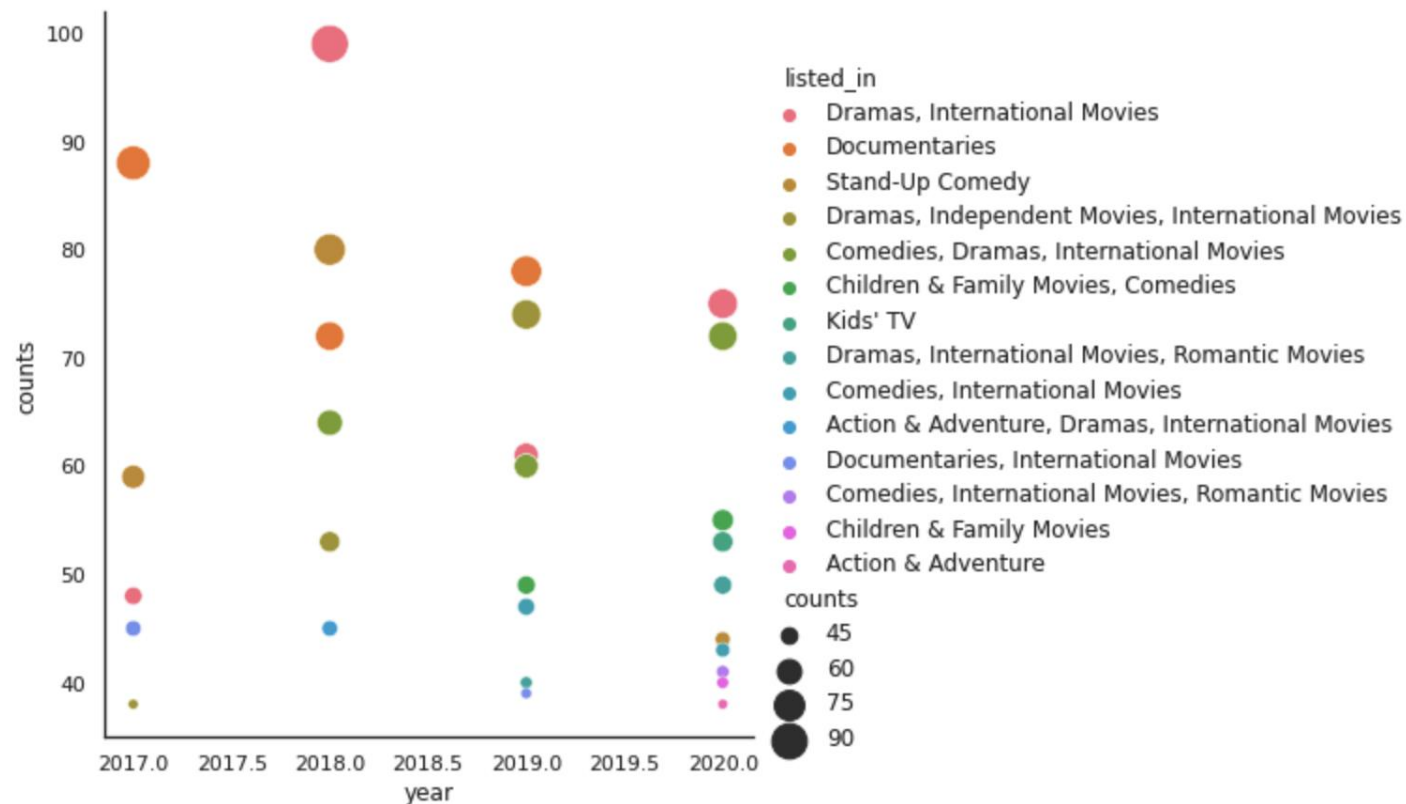
가설 2 (listed_in / relplot)

: COVID-19와 상관관계

가설 2

코로나 시대 장르의 변화가 있을까?

- 가장 많이 등록되어 있는 다큐멘터리 장르의 경우, 2017년과 2019년 등록이 가장 많이 되었으나 2020년에는 상위 10위에 들어가지 못했다.
- 2017년~2019년까지는 5개~8개의 장르가 집중적으로 등록 되었으나, 2020년에는 2017년~2019년에 대비해 콘텐츠가 조금 더 고른 분포로 다양하게 등록된 것으로 분석되었다.



04

COVID-19와 상관관계

- 가설 설정

- 가설 1

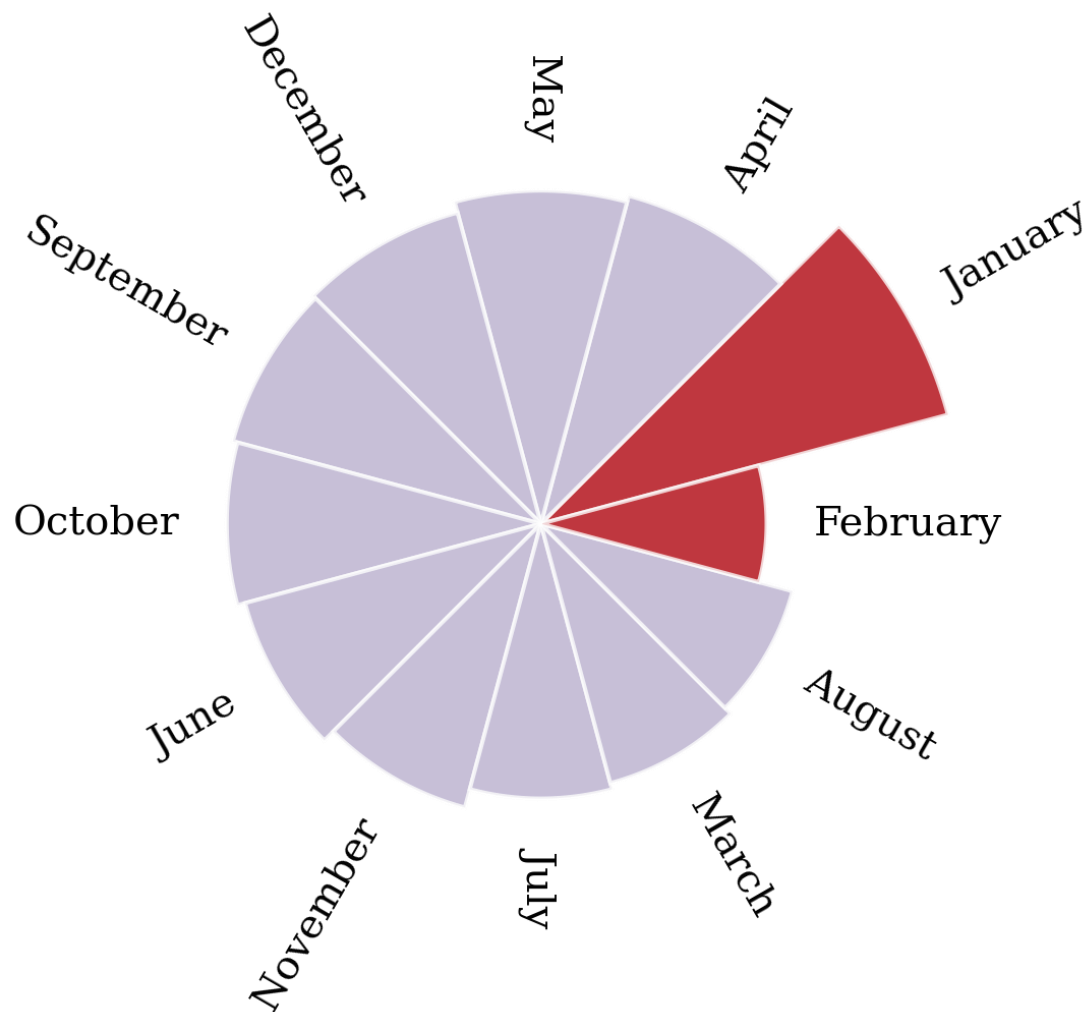
- 가설 2

- 가설 3 / 코로나 대유행으로 넷플릭스에 등록하는 콘텐츠의 수가 감소하지 않을까?

- 가설에 대한 결론 분석

가설 3

: COVID-19와 상관관계



2020년 1월

- 1월 넷플릭스 콘텐츠
등록이 가장 많음

2020년 2월

- 1월에 비해 전세계적으로
코로나의 대확산이
심각해짐에 따라 2월에
콘텐츠 등록 수가 급감

04

COVID-19와 상관관계

- 가설 설정
- 가설 1
- 가설 2
- 가설 3
- 가설에 대한 결론 분석

가설에 대한 결론 분석

: COVID-19와 상관관계

가설 설정

코로나로 인한 넷플릭스 변화 분석

결론 1



TV Shows의 비율이 소폭상승 한 것으로 볼 때
100 % COVID-19 영향이라고 볼 수는 없지만, COVID-19 발생과 넷플릭스 구독이
2억이 넘는 시대인 2020년에 처음으로 유의미한 콘텐츠 등록 변화가 일어난 점에 주목
할 만 함

결론 2



성인 기준의 콘텐츠의 비율이 유지될 것이라고 예상한 것과 달리
COVID-19로 인해 사람들이 집에서 보내는 시간이 증가함에 따라 가족 모두가 시청할
수 있는
전체 관람가와 7세 이상(Children) 비율이 소폭 상승

결론 3



예상대로 코로나 대유행으로 넷플릭스에 등록하는 콘텐츠의 수가 감소하였지만,
코로나 사태로 사람 대 사람의 활동이 줄어들고 개인의 시간이 증가하는 현상에
발맞추어
콘텐츠 등록 수도 회복된 것으로 볼 수 있음

NETFLIX

오리지널 콘텐츠, 영화, 드라마, 예능까지!
넷플릭스 하나로 무제한!

“넷플릭스 콘텐츠 변화 분석 완료!”

