

SeSac AI개발자 과정 - 미니 프로젝트

영화 개봉 전략을 위한 인기도 및 시기 분석.

함현지

목차

1 추진 배경 및 목표 설정

2 데이터셋 설명

3 분석 결과 및 인사이트 도출

4 인사이트 확장 가능성

The Movies Dataset

1. 추진 배경 및 목표 설정

1. 추진 배경

영화 산업은 시즌별로 특정 장르가 더 인기를 끌 수 있습니다.

예를 들어, 여름에는 블록버스터 액션 영화가, 겨울에는 가족과 함께 볼 수 있는 휴가 시즌 영화가 더 많은 관객을 끌어들이 수 있습니다.

따라서 영화 개봉 시기를 결정할 때, 장르별로 유리한 개봉 월을 파악하는 것은 중요합니다.

또한, 개봉 시기에 따른 유저 인기도를 분석하면, 어느 시기에 어떤 장르가 더 많은 인기를 끌었는지 이해할 수 있습니다. 이를 통해 영화 배급사와 제작사들은 마케팅 전략을 더욱 효과적으로 수립할 수 있습니다.

2. 목표 설정

이 프로젝트의 목표는 영화 장르별로 개봉 월과 유저 인기도 월 분포를 분석하여, 특정 장르가 어느 달에 유행하는지 파악하는 것입니다.


이를 통해 예측 모델을 개발하고, 다음과 같은 질문에 대한 답을 찾고자 합니다


- 특정 장르가 특히 인기를 끄는 달은 언제인가?
- 영화 개봉 시기를 어떻게 최적화할 수 있는가?



The Movies Dataset

2. 데이터셋 설명

2. 데이터셋 설명

키워드		Keywords		
	영화ID	movieID	Domain	INT
	키워드ID	keywords_id	Domain	DECIMAL
	키워드이름	keywords_name	Domain	CHAR(255)

링크		Links		
	영화ID	movieID	Domain	INT
	IMDb ID	imdbID	Domain	DECIMAL
	TMDb ID	tmdbID	Domain	DECIMAL

평점		Ratings		
	사용자ID	userId	Domain	VARCHAR
	영화ID	movieID	Domain	INT
	평점	rating	Domain	INT
	타임스탬프	timestamp	Domain	Type

영화메타데이터		Movies_metadata		
	영화ID	id	Domain	INT
	성인영화여부	adult	Domain	BOOL
	영화가 속한 컬렉션 정보	belongs_to_collection:	Domain	Type
	제작 예산	budget	Domain	DECIMAL
	영화장르	genres	Domain	CHAR(255)
	영화공식홈페이지	homepage	Domain	Type
	IMDb ID	imdb_id	Domain	DECIMAL
	원래 언어	original_language	Domain	CHAR(255)
	원래 제목	original_title	Domain	CHAR(255)
	영화 개요	overview	Domain	CHAR(255)
	인기도	popularity	Domain	Type
	포스터 이미지 경로	poster_path	Domain	Type
	제작사	production_companies	Domain	Type
	제작 국가	production_countries	Domain	Type
	개봉일	release_date	Domain	Type
	수익	revenue	Domain	Type
	상영시간	runtime	Domain	Type
	사용된 언어	spoken_languages	Domain	Type
	개봉상태	status	Domain	Type
	영화태그라인	tagline	Domain	Type
	제목	title	Domain	Type
	비디오여부	video	Domain	Type
	평균 평점	vote_average	Domain	Type
	평점수	vote_count	Domain	Type

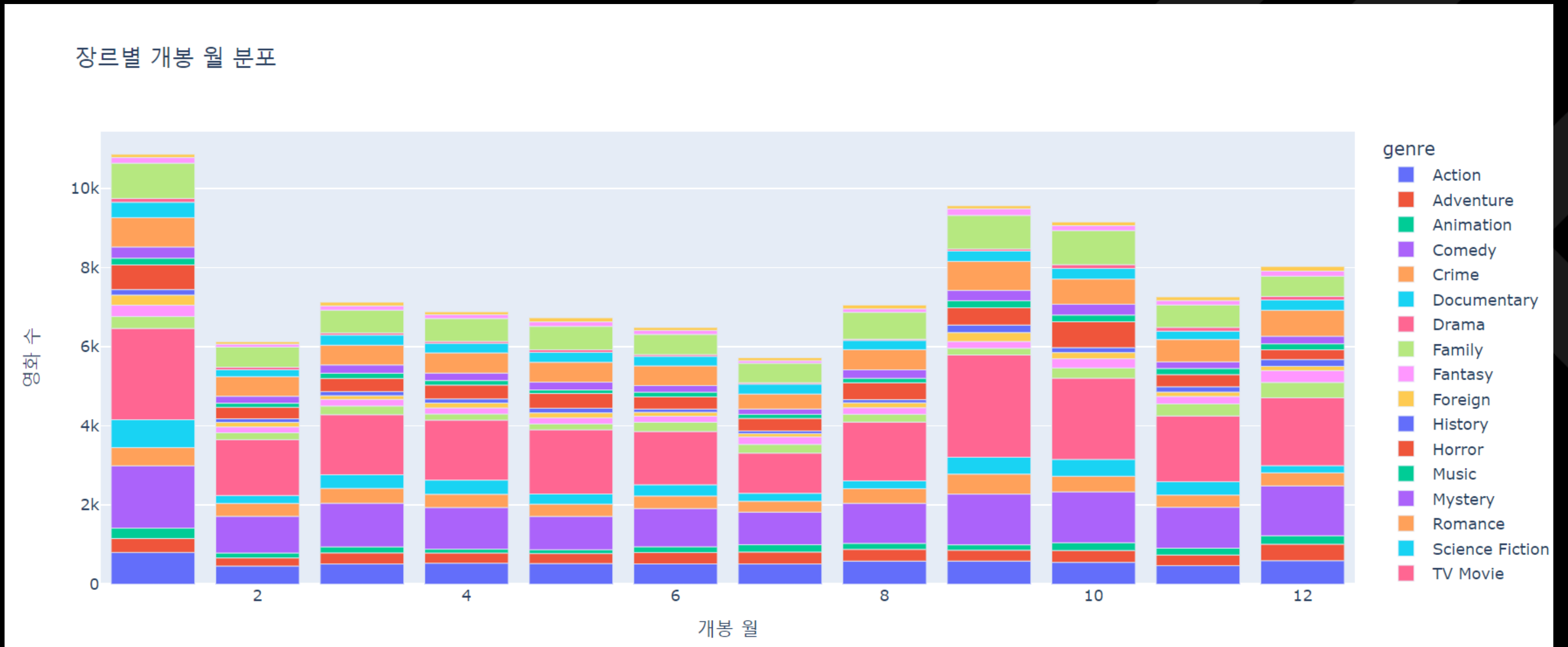
2. 데이터셋 설명

영화메타데이터		Movies_metadatas	
 영화ID	id	Domain	INT
성인영화여부	adult	Domain	BOOL
영화가 속한 컬렉션 정보	belongs_to_collection:	Domain	Type
제작 예산	budget	Domain	DECIMAL
영화장르	genres	Domain	CHAR(255)
영화공식홈페이지	homepage	Domain	Type
IMDb ID	imdb_id	Domain	DECIMAL
원래 언어	original_language	Domain	CHAR(255)
원래 제목	original_title	Domain	CHAR(255)
영화 개요	overview	Domain	CHAR(255)
인기도	popularity	Domain	Type
포스터 이미지 경로	poster_path	Domain	Type
제작사	production_companies	Domain	Type
제작 국가	production_countries	Domain	Type
개봉일	release_date	Domain	Type
수익	revenue	Domain	Type
상영시간	runtime	Domain	Type
사용된 언어	spoken_languages	Domain	Type
개봉상태	status	Domain	Type
영화태그라인	tagline	Domain	Type
제목	title	Domain	Type
비디오여부	video	Domain	Type
평균평점	vote_average	Domain	Type
평점수	vote_count	Domain	Type

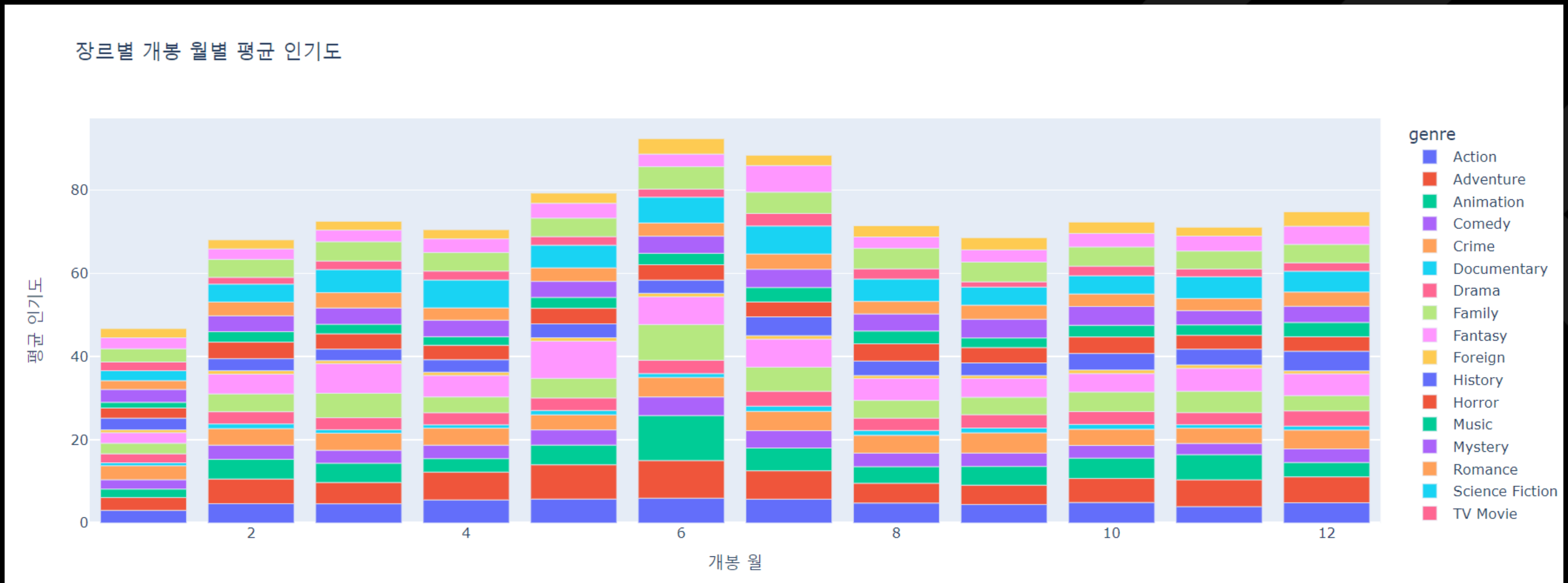
The Movies Dataset

3. 분석 결과 및 인사이트 도출

1. 분석 결과 (장르별 개봉 월 분포)



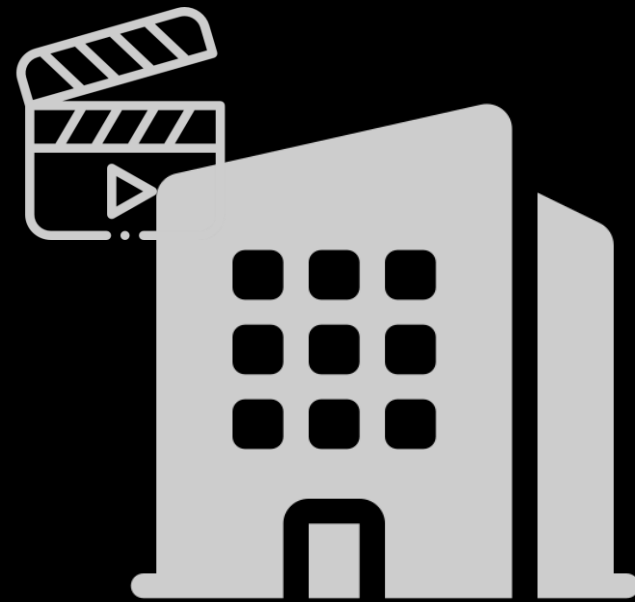
2. 분석 결과 (장르별 개봉 월별 평균 인기도)



The Movies Dataset

3-1. 페르소나 설정 및 데이터 활용

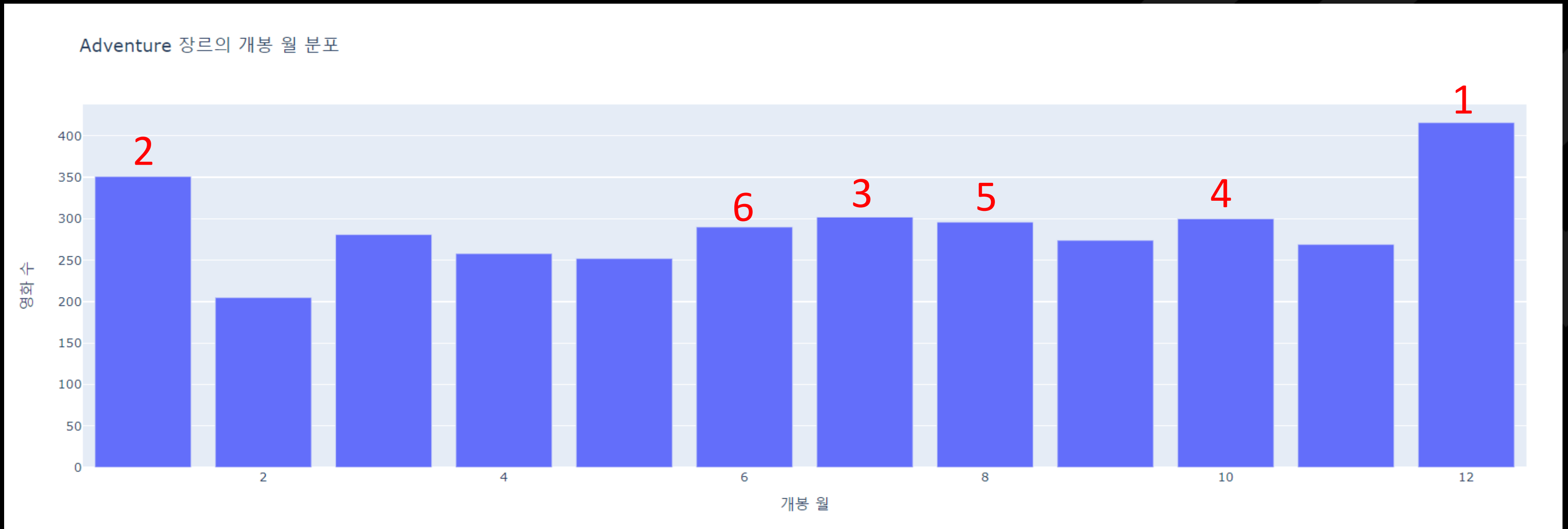
3. 페르소나 설정



제작한 영화 장르 : Adventure

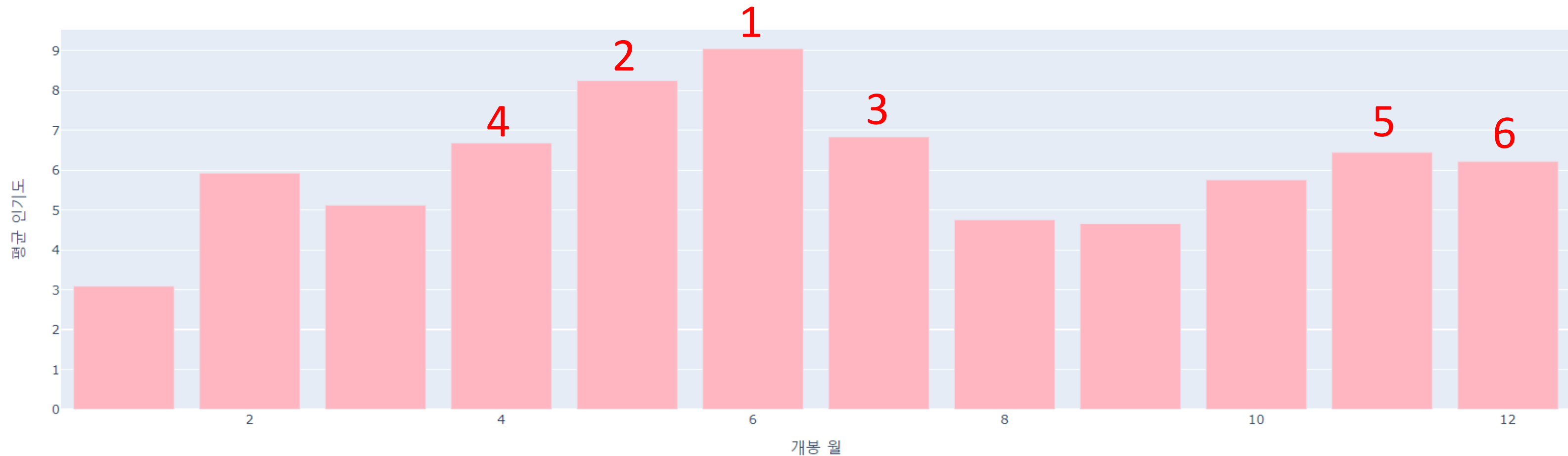
어느 달에 개봉하는 것이
유리할까?

1. 분석 결과 (장르별 개봉 월 분포)



2. 분석 결과 (장르별 개봉 월별 평균 인기도)

Adventure 장르의 월별 평균 인기도



3. 각 결과 순위 합산

월 분포 순위	월별 평균 인기도 순위
12월	6월
1월	5월
7월	7월
10월	4월
8월	11월
6월	12월

‘어드벤처’ 장르의 영화를 개봉하려고 할 경우,
같은 장르의 영화가 적게 개봉하며, 인기도의 평균 순위가 높은 ‘5월, 6월’이 적합합니다.

The Movies Dataset

4. 인사이트 확장 가능성

The Movies Dataset

1. 인사이트 확장 가능성

다른 데이터셋과 함께 분석한다면 다른 분석도 가능합니다.

연도별 twitter 인기 키워드 dataset

연도로별로 인기가 많았던 키워드와
현재 인기가 많은 키워드를 도출

The movies dataset의 연도별 인기 있던
태그라인을 도출해낸 후
재개봉 할 영화를 도출해 낼 수 있습니다.

2. 추가 내용

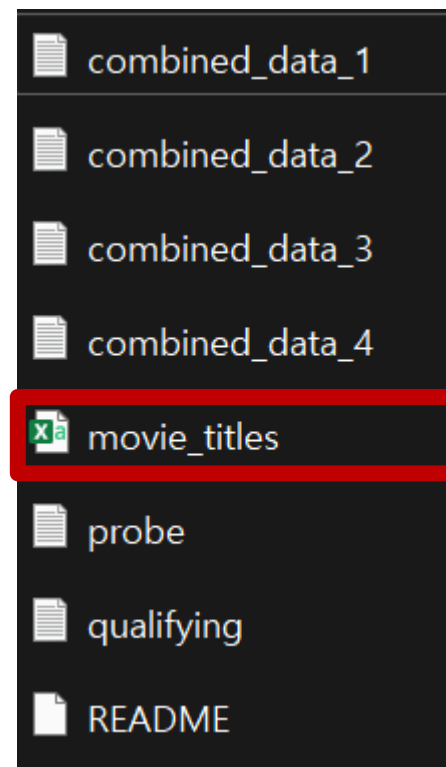
머신러닝과 딥러닝을 배운 이후, 분석한 내용을 이용하여 새로운 결과를 만들어 낼 수 있습니다.

Netflix dataset

Netflix는 영화에 대한 사용자 평가를 예측하는 Netflix Prize 공개 경연 대회를 개최했습니다. 그 경연 대회에서 사용된 데이터 세트입니다.

2. 추가 내용

- Netflix dataset 설명



	영화 id	출시 년도	영화제목		
	A	B	C	D	E
1	1	2003	Dinosaur Planet		
2	2	2004	Isle of Man TT 2004 Review		
3	3	1997	Character		
4	4	1994	Paula Abdul's Get Up & Dance		
5	5	2004	The Rise and Fall of ECW		

2. 추가 내용

- Netflix dataset 설명

1:	1: (영화의 ID)		
1488844,3,2005-09-06	고객 ID	평점	날짜
822109,5,2005-05-13	1488844	3	2005-09-06
885013,4,2005-10-19	822109	5	2005-05-13
30878,4,2005-12-26	885013	4	2005-10-19
823519,3,2004-05-03	30878	4	2005-12-26
893988,3,2005-11-17			
124105,4,2004-08-05			
1248029,3,2004-04-22			

combined_data_1

	User	Rating	Date
0	1:	NaN	NaN
1	1488844	3.0	2005-09-06
2	822109	5.0	2005-05-13
3	885013	4.0	2005-10-19
4	30878	4.0	2005-12-26
...
24058258	2591364	2.0	2005-02-16
24058259	1791000	2.0	2005-02-10
24058260	512536	5.0	2005-07-27
24058261	988963	3.0	2005-12-20
24058262	1704416	3.0	2004-06-02

df_raw

(영화의 id)

(user_id)

2. 추가 내용

- Netflix dataset 설명

	User	Rating	Date	Movie
1	1488844	3.0	2005-09-06	1
2	822109	5.0	2005-05-13	1
3	885013	4.0	2005-10-19	1
4	30878	4.0	2005-12-26	1
5	823519	3.0	2004-05-03	1
...
24058258	2591364	2.0	2005-02-16	4499
24058259	1791000	2.0	2005-02-10	4499
24058260	512536	5.0	2005-07-27	4499
24058261	988963	3.0	2005-12-20	4499
24058262	1704416	3.0	2004-06-02	4499

넷플릭스 데이터셋의 영화 title과
The Movies dataset의 title을 결합하여
넷플릭스 사용자에게 시기에 맞는 영화를
추천하는 시스템을 만들 수 있습니다.

The Movies Dataset

Thank you.

