

# IMDB 영화 데이터셋

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import koreanize_matplotlib
```

```
In [13]: df = pd.read_csv("imdb.csv")
df.head()
```

Out[13]:

	Rank	Movie_name	Year	Certificate	Runtime_in_min	Genre	Metascore	Gross_in_\$_M	Rating_from_10
0	1	The Shawshank Redemption	1994	R	142	Drama	81.0	28.34	9.3
1	2	The Godfather	1972	R	175	Crime, Drama	100.0	134.97	9.2
2	3	The Dark Knight	2008	PG-13	152	Action, Crime, Drama	84.0	534.86	9.0
3	4	The Lord of the Rings: The Return of the King	2003	PG-13	201	Action, Adventure, Drama	94.0	377.85	9.0
4	5	Schindler's List	1993	R	195	Biography, Drama, History	94.0	96.90	9.0

- (1) 결측값이 없는지 확인을 하시오. 없으면 대체 데이터를 만드세요.
- (2) 액션 장르와 관련된 데이터셋을 추출하고, Metascore를 활용하여 Top 10 리스트를 만들어보세요.
- (3) 'Movie\_name'과 'Genre' 컬럼만을 선택하시오.
- (4) Top 5 판타지 영화를 출력하시오. (Rating\_from\_10)
- (5) 영화 평점을 3개의 등급으로 나누고, 각 등급에 속하는 영화의 개수를 확인하세요.
- (6) 각 장르(Genre)별 영화의 평균 평점('Rating\_from\_10')을 계산하고, 그 결과를 평균 평점이 높은 순서로 정렬하여 상위 5개 장르를 출력하세요.
- (7) 다음은 'Rating\_from\_10'을 활용하여 히스토그램을 그리시오.
- (8) 각 연도(Year)별 영화의 평균 런타임('Runtime\_in\_min')과 최대 평점('Rating\_from\_10')을 계산하여 출력하세요.