

EXPERIMENT 1

Load Data in Pandas

```
import pandas as pd

df = pd.read_csv("/content/Top_10000_Movies.csv", on_bad_lines="skip")
df.head()
```

	id	original_language	original_title	popularity	release_date	vote_average	vote_count	genre	overview	revenue	runtime	tagline
0	580489.0	en	Venom: Let There Be Carnage	5401.308	2021-09-30	6.8	1736.0	['Science Fiction', 'Action', 'Adventure']	After finding a host body in investigative rep...	424000000.0	97.0	NaN
1	524434.0	en	Eternals	3365.535	2021-11-03	7.1	622.0	['Action', 'Adventure', 'Science Fiction', 'Fa...']	The Eternals are a team of ancient aliens who ...	165000000.0	157.0	In the beginning...
2	438631.0	en	Dune	2911.423	2021-09-15	8.0	3632.0	['Action', 'Adventure', 'Science Fiction']	Paul Atreides, a brilliant and gifted young ma...	331116356.0	155.0	Beyond fear, destiny awaits.
3	796499.0	en	Army of Thieves	2552.437	2021-10-27	6.9	555.0	['Action', 'Crime', 'Thriller']	A mysterious woman recruits bank teller Ludwig...	0.0	127.0	Before Vegas, one locksmith became a legend.
4	550988.0	en	Free Guy	1850.470	2021-08-11	7.8	3493.0	['Comedy', 'Action', 'Adventure', 'Science Fic...']	A bank teller called Guy realizes he is a back...	331096766.0	115.0	Life's too short to be a background character.

Description of the Dataset

```
df.info()

df.describe()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 10014 entries, 0 to 9999
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0    id                    10002 non-null  float64
1    original_language     10002 non-null  object
2    original_title        10001 non-null  object
3    popularity            10000 non-null  float64
4    release_date          9962 non-null   object
5    vote_average          10000 non-null  float64
6    vote_count            10000 non-null  float64
7    genre                 10000 non-null  object
8    overview              9900 non-null   object
9    revenue               9998 non-null   float64
10   runtime               9989 non-null   float64
11   tagline               7079 non-null   object
dtypes: float64(6), object(6)
memory usage: 1.2+ MB
```

	id	popularity	vote_average	vote_count	revenue	runtime
count	10002.000000	10000.000000	10000.000000	10000.000000	9.998000e+03	9989.000000
mean	250003.082683	34.516871	6.29875	1315.084900	5.737536e+07	98.792772
std	261732.329571	100.693958	1.43426	2501.899103	1.480897e+08	28.771525
min	0.000000	6.269000	0.00000	0.000000	0.000000e+00	0.000000
25%	11864.500000	11.908000	5.90000	118.000000	0.000000e+00	89.000000

✓ Connected to Python 3 Google Compute Engine backend

Drop columns that aren't useful.

Drop Columns

df = df.drop(["popularity", "runtime"], axis=1)

df.head()

	id	original_language	original_title	release_date	vote_average	vote_count	genre	overview	revenue	tagline
0	580489.0	en	Venom: Let There Be Carnage	2021-09-30	6.8	1736.0	['Science Fiction', 'Action', 'Adventure']	After finding a host body in investigative rep...	424000000.0	NaN
1	524434.0	en	Eternals	2021-11-03	7.1	622.0	['Action', 'Adventure', 'Science Fiction', 'Fa...	The Eternals are a team of ancient aliens who ...	165000000.0	In the beginning...
2	438631.0	en	Dune	2021-09-15	8.0	3632.0	['Action', 'Adventure', 'Science Fiction']	Paul Atreides, a brilliant and gifted young ma...	331116356.0	Beyond fear, destiny awaits.
3	796499.0	en	Army of Thieves	2021-10-27	6.9	555.0	['Action', 'Crime', 'Thriller']	A mysterious woman recruits bank teller Ludwig...	0.0	Before Vegas, one locksmith became a legend.
4	550988.0	en	Free Guy	2021-08-11	7.8	3493.0	['Comedy', 'Action', 'Adventure', 'Science Fic...	A bank teller called Guy realizes he is a back...	331096766.0	Life's too short to be a background character.

Standardization and Normalization of columns

Standardization

[16]

from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()

df[['vote_average_std', 'vote_count_std', 'revenue_std']] = scaler.fit_transform(df[['vote_average', 'vote_count', 'revenue']])

df.head()

	id	original_title	release_date	vote_average	vote_count	overview	revenue	tagline	original_language_be	original_language_bn	...	genre_['western', 'horror']	genre_['west', 'myst', 'thrill', 'dr...
0	580489.0	Venom: Let There Be Carnage	2021-09-30	6.8	1736.0	After finding a host body in investigative rep...	424000000.0	Based on a true story.	False	False	...	False	
1	524434.0	Eternals	2021-11-03	7.1	622.0	The Eternals are a team of ancient aliens who ...	165000000.0	In the beginning...	False	False	...	False	
2	438631.0	Dune	2021-09-15	8.0	3632.0	Paul Atreides, a brilliant and gifted young ma...	331116356.0	Beyond fear, destiny awaits.	False	False	...	False	

Normalization

[17]

from sklearn.preprocessing import MinMaxScaler

min_max_scaler = MinMaxScaler()

df[['vote_average_norm', 'vote_count_norm', 'revenue_norm']] = min_max_scaler.fit_transform(df[['vote_average', 'vote_count', 'revenue']])

df.head()

	id	original_title	release_date	vote_average	vote_count	overview	revenue	tagline	original_language_be	original_language_bn	...	genre_['western', 'TV Movie']	genre_['west', 'Thrill
0	580489.0	Venom: Let There Be Carnage	2021-09-30	6.8	1736.0	After finding a host body in investigative rep...	424000000.0	Based on a true story.	False	False	...	False	
1	524434.0	Eternals	2021-11-03	7.1	622.0	The Eternals are a team of ancient aliens who ...	165000000.0	In the beginning...	False	False	...	False	
2	438631.0	Dune	2021-09-15	8.0	3632.0	Paul Atreides, a brilliant and gifted young ma...	331116356.0	Beyond fear, destiny awaits.	False	False	...	False	