ALLOCINÉ

Système de recommandation sur AlloCiné

STRUCTURE DU DOCUMENT

Table des matières

1 Webscraping

Films, Séries, Notes
Stats

2 Data Analysis

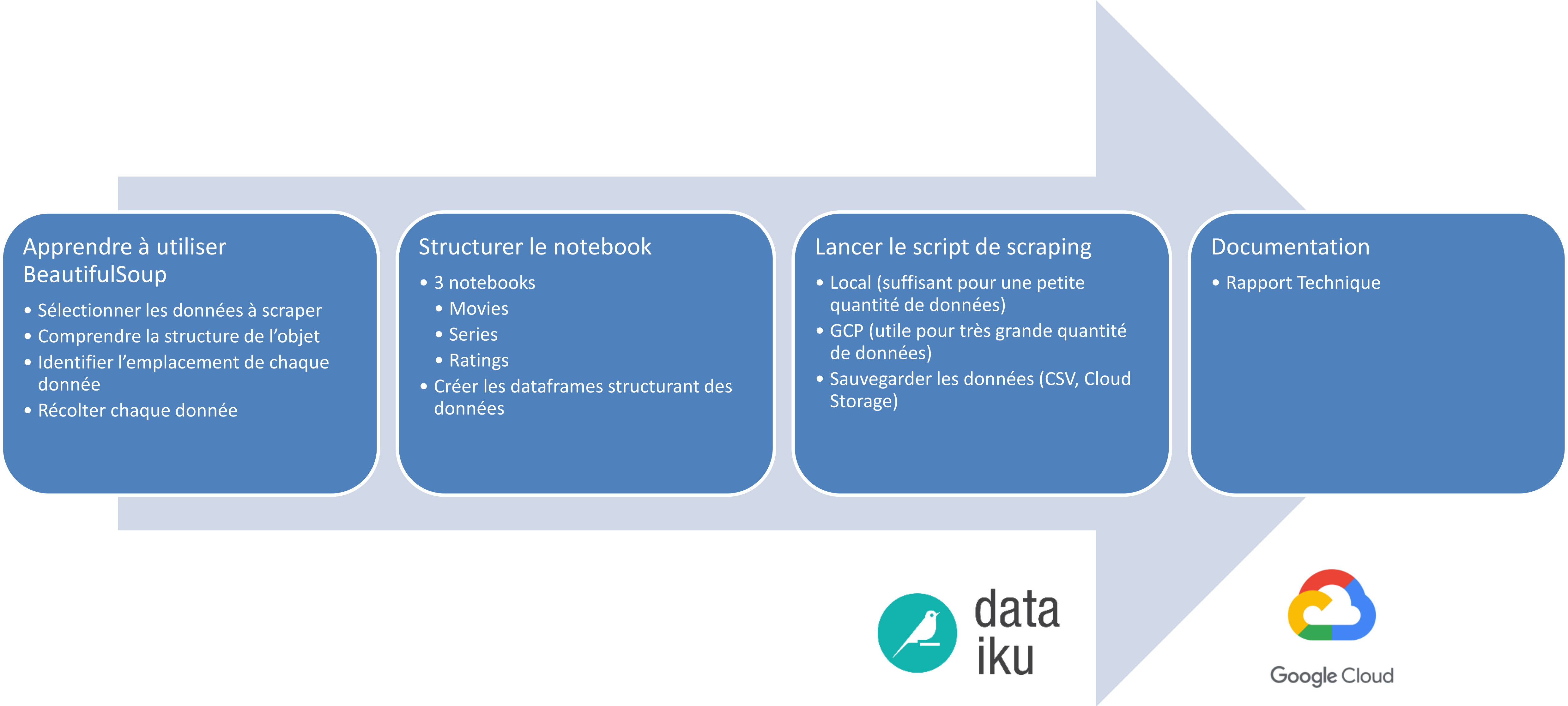
Data cleaning
Analyse Exploratoire N°1
Analyse Exploratoire N°2

3 Moteur de recommandation

Content-based
Collaborative-Filtering (CF)

Webscraping - Films, Séries, Notes

Récolte des données depuis le site d'AlloCiné



Apprendre à utiliser BeautifulSoup

- Sélectionner les données à scraper
- Comprendre la structure de l'objet
- Identifier l'emplacement de chaque donnée
- Récolter chaque donnée

Structurer le notebook

- 3 notebooks
  - Movies
  - Series
  - Ratings
- Créer les dataframes structurant des données

Lancer le script de scraping

- Local (suffisant pour une petite quantité de données)
- GCP (utile pour très grande quantité de données)
- Sauvegarder les données (CSV, Cloud Storage)

Documentation

- Rapport Technique

data iku

Google Cloud

Webscraping - Stats - Tirage 100 pages

Récolte des données depuis le site d'AlloCiné

Objectifs:

100 pages

1500 films

1500 séries

Résultats:

1314 films	Runtime: 1h26
1417 séries	Runtime: 1h38
105 711 notes spectateurs films	Runtime: 6h12
21 582 notes presse films	Runtime: ≃2,8h
60 031 notes spectateurs séries	Runtime: ≃2,8h
4516 notes presse séries	Runtime: ≃2,8h

Webscraping - Stats - Tirage 600 pages

Récolte des données depuis le site d'AlloCiné

Objectifs:

- 600 pages
- 9000 films
- 9000 séries

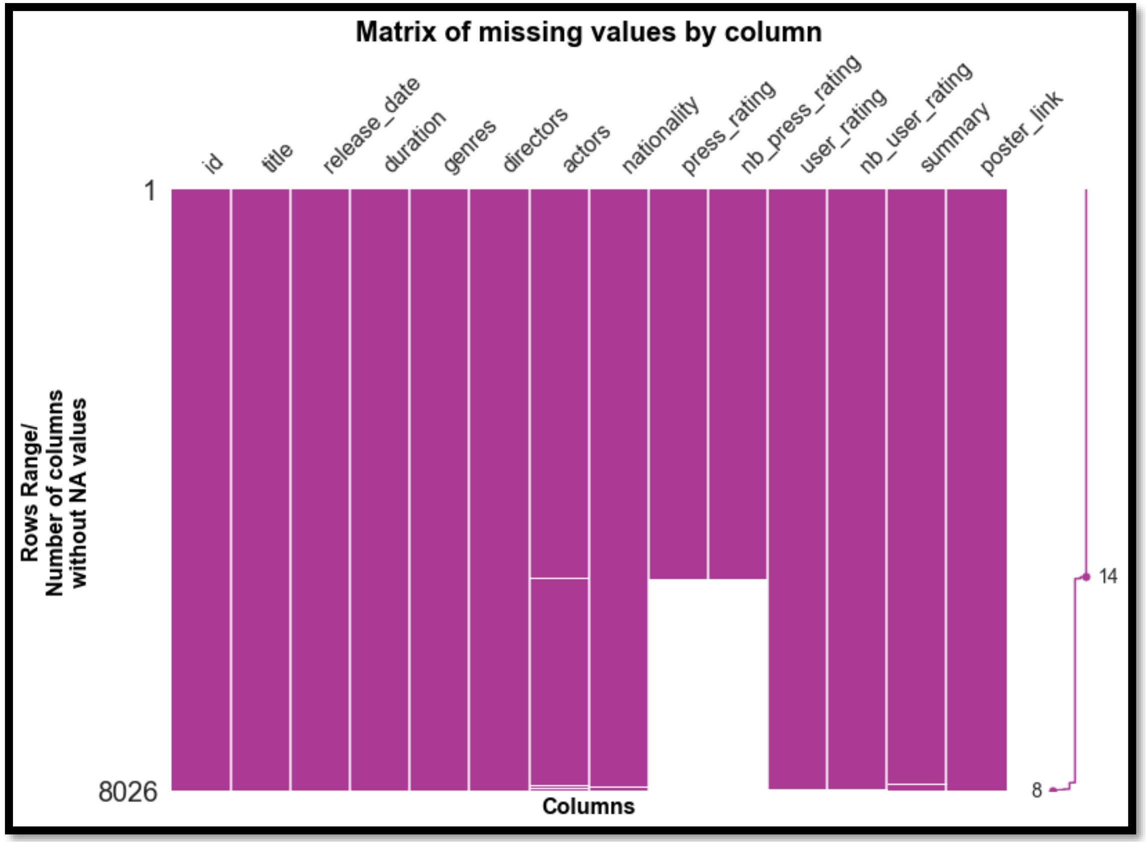
Résultats:

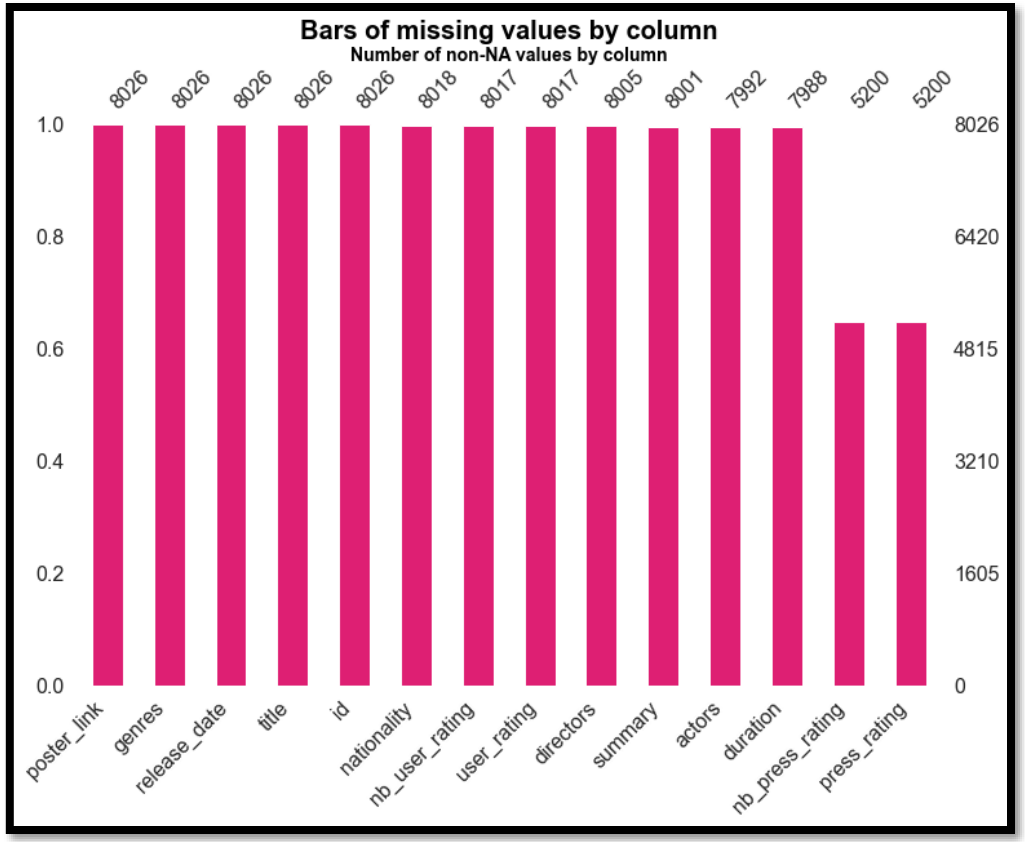
8026 films	Runtime:
8126 séries	Runtime: 9h ±1h
330 413 notes spectateurs films	Runtime: > 24h
90040 notes presse films	Runtime: 8h ±1h
69 946 notes spectateurs séries	Runtime: > 24h
6732 notes presse séries	Runtime: 8h ±1h

Data Analysis - Data cleaning

Analyses des données

MOVIES (BRUT)

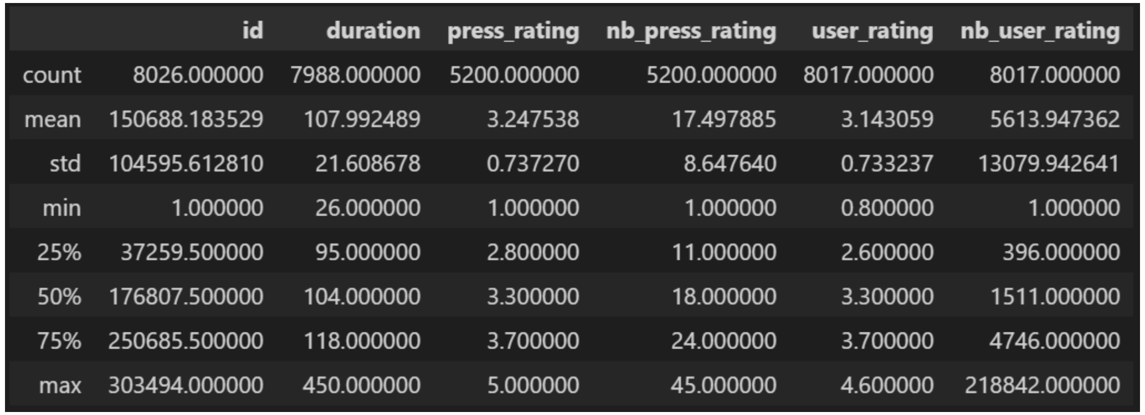




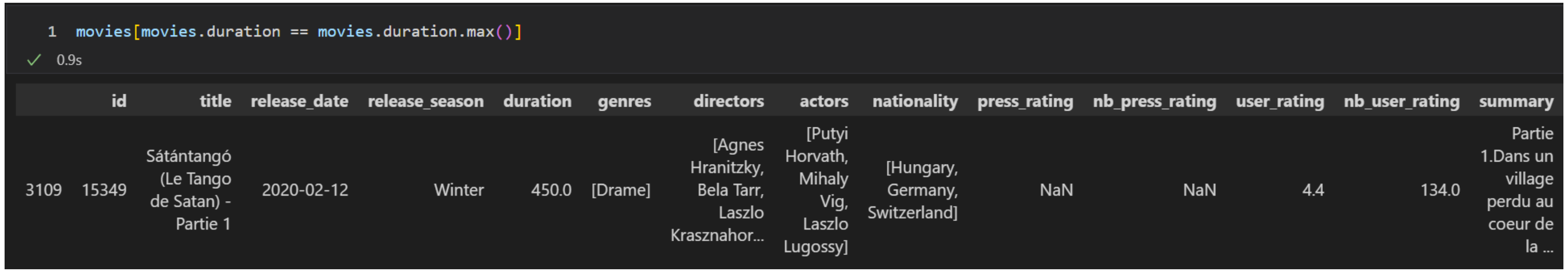
Rows: 8026
Columns: 14

Data Analysis - Data cleaning

MOVIES (BRUT)



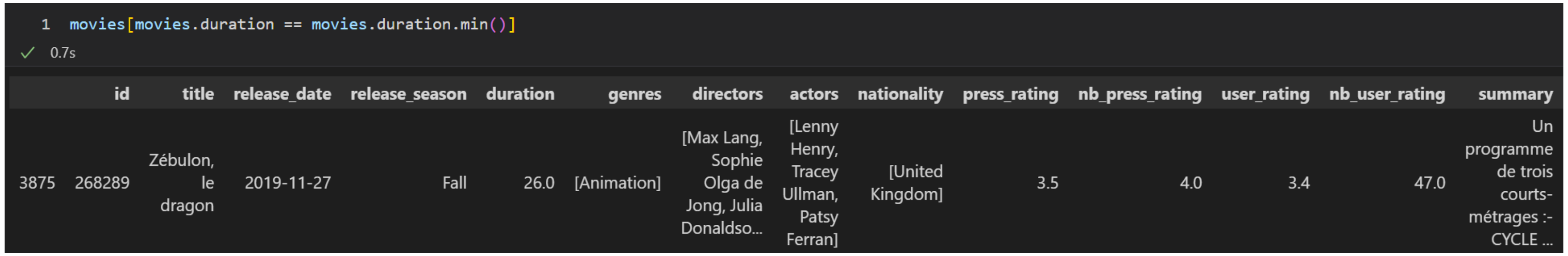
	id	duration	press_rating	nb_press_rating	user_rating	nb_user_rating
count	8026.000000	7988.000000	5200.000000	5200.000000	8017.000000	8017.000000
mean	150688.183529	107.992489	3.247538	17.497885	3.143059	5613.947362
std	104595.612810	21.608678	0.737270	8.647640	0.733237	13079.942641
min	1.000000	26.000000	1.000000	1.000000	0.800000	1.000000
25%	37259.500000	95.000000	2.800000	11.000000	2.600000	396.000000
50%	176807.500000	104.000000	3.300000	18.000000	3.300000	1511.000000
75%	250685.500000	118.000000	3.700000	24.000000	3.700000	4746.000000
max	303494.000000	450.000000	5.000000	45.000000	4.600000	218842.000000



```
movies[movies.duration == movies.duration.max()]
```

✓ 0.9s

	id	title	release_date	release_season	duration	genres	directors	actors	nationality	press_rating	nb_press_rating	user_rating	nb_user_rating	summary
3109	15349	Sátántangó (Le Tango de Satan) - Partie 1	2020-02-12	Winter	450.0	[Drame]	[Agnes Hranitzky, Bela Tarr, Laszlo Krasznahor...	[Putyi Horvath, Mihaly Vig, Laszlo Lugossy]	[Hungary, Germany, Switzerland]	NaN	NaN	4.4	134.0	Partie 1.Dans un village perdu au coeur de la ...



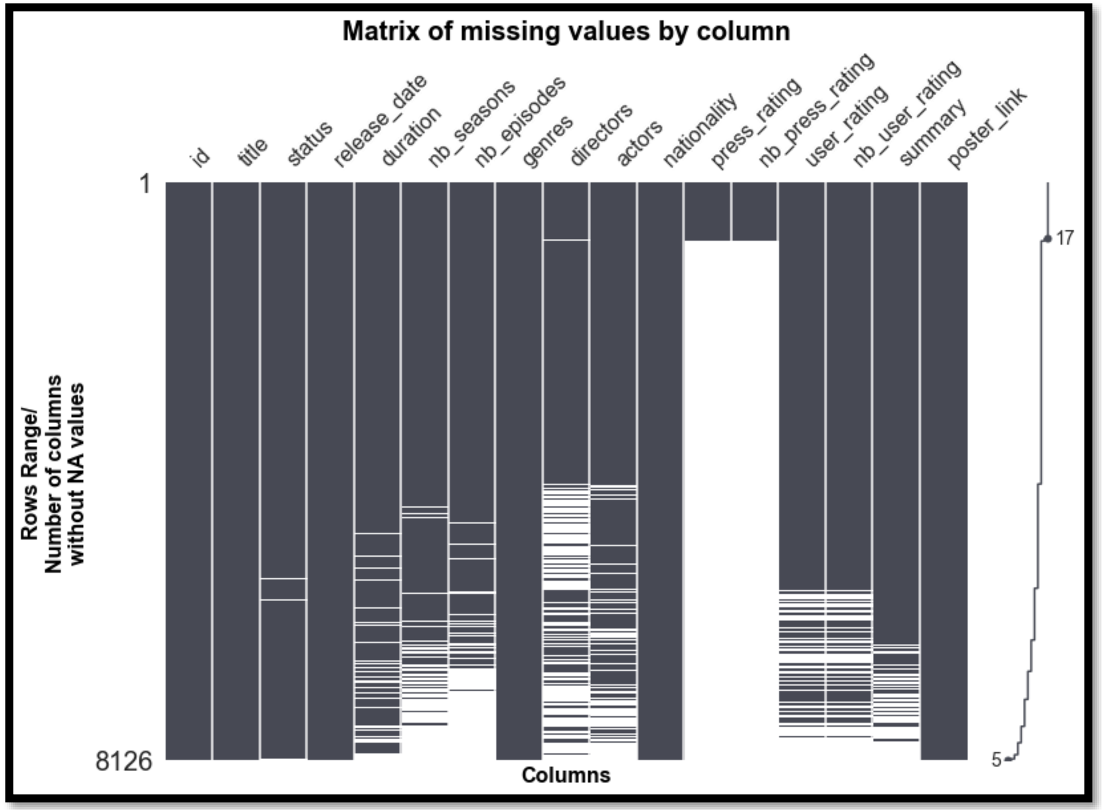
```
movies[movies.duration == movies.duration.min()]
```

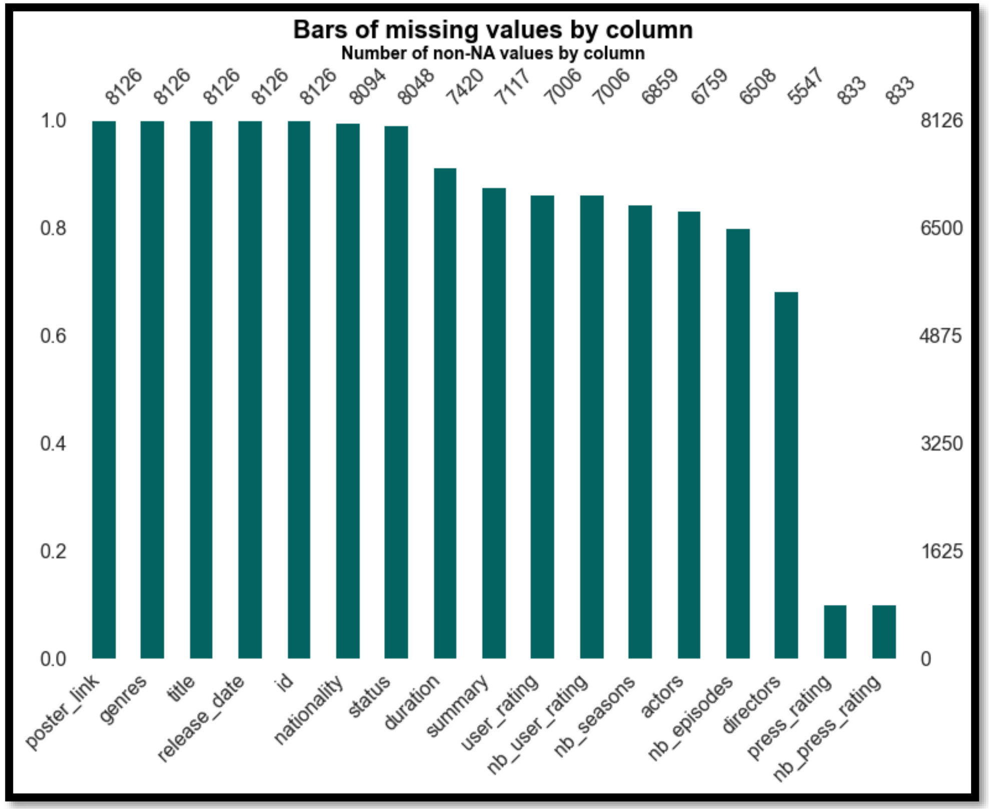
✓ 0.7s

	id	title	release_date	release_season	duration	genres	directors	actors	nationality	press_rating	nb_press_rating	user_rating	nb_user_rating	summary
3875	268289	Zébulon, le dragon	2019-11-27	Fall	26.0	[Animation]	[Max Lang, Sophie Olga de Jong, Julia Donaldso...	[Lenny Henry, Tracey Ullman, Patsy Ferran]	[United Kingdom]	3.5	4.0	3.4	47.0	Un programme de trois courts-métrages :- CYCLE ...

Data Analysis - Data cleaning

SERIES (BRUT)



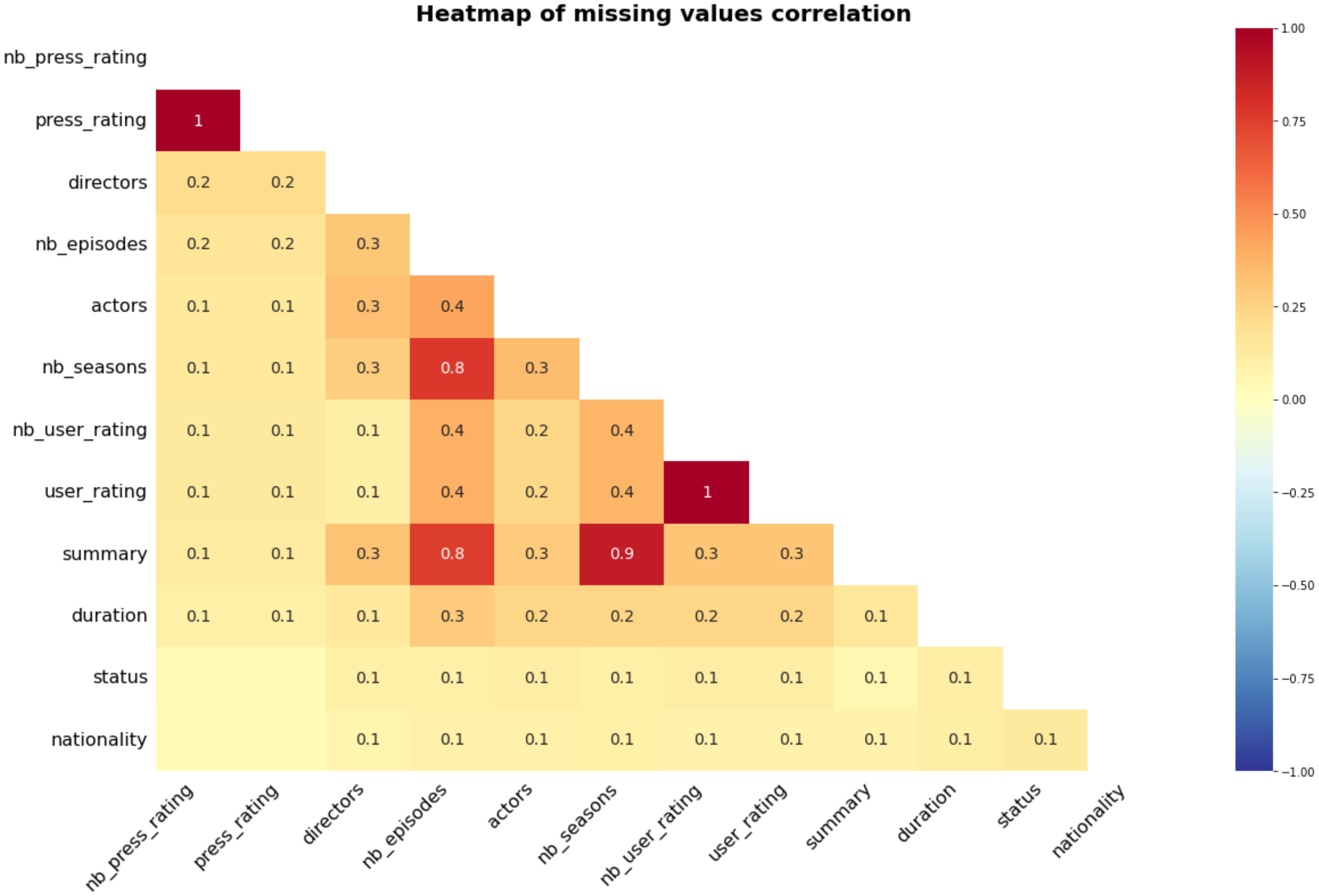


Rows: 8126
Columns: 17

Data Analysis - Data cleaning

Analyses des données

SERIES (BRUT)

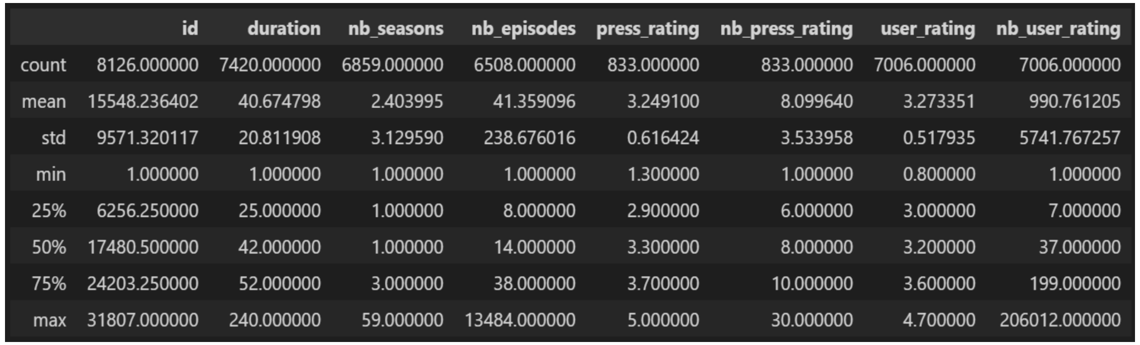


	nb_press_rating	press_rating	directors	nb_episodes	actors	nb_seasons	nb_user_rating	user_rating	summary	duration	status	nationality
nb_press_rating												
press_rating	1											
directors	0.2	0.2										
nb_episodes	0.2	0.2	0.3									
actors	0.1	0.1	0.3	0.4								
nb_seasons	0.1	0.1	0.3	0.8	0.3							
nb_user_rating	0.1	0.1	0.1	0.4	0.2	0.4						
user_rating	0.1	0.1	0.1	0.4	0.2	0.4	1					
summary	0.1	0.1	0.3	0.8	0.3	0.9	0.3	0.3				
duration	0.1	0.1	0.1	0.3	0.2	0.2	0.2	0.2	0.1			
status			0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1		
nationality			0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	

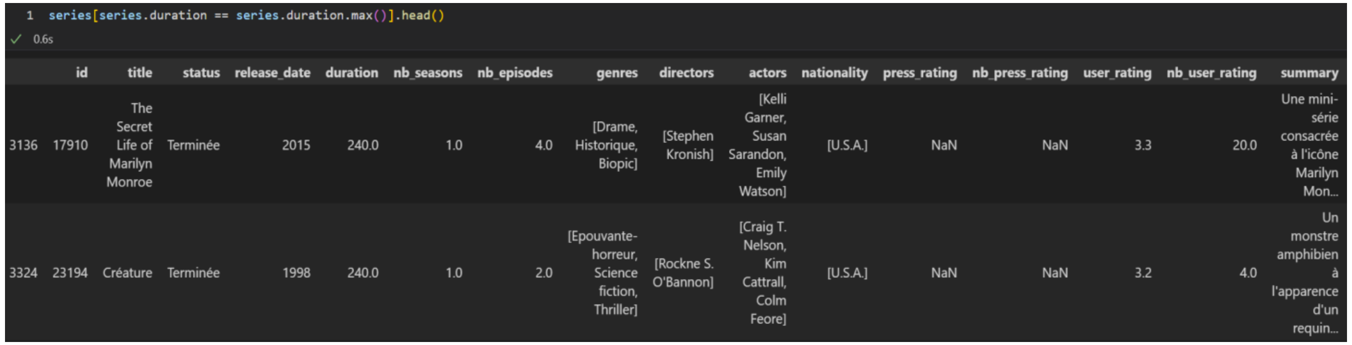
Data Analysis - Data cleaning

Analyses des données

SERIES (BRUT)



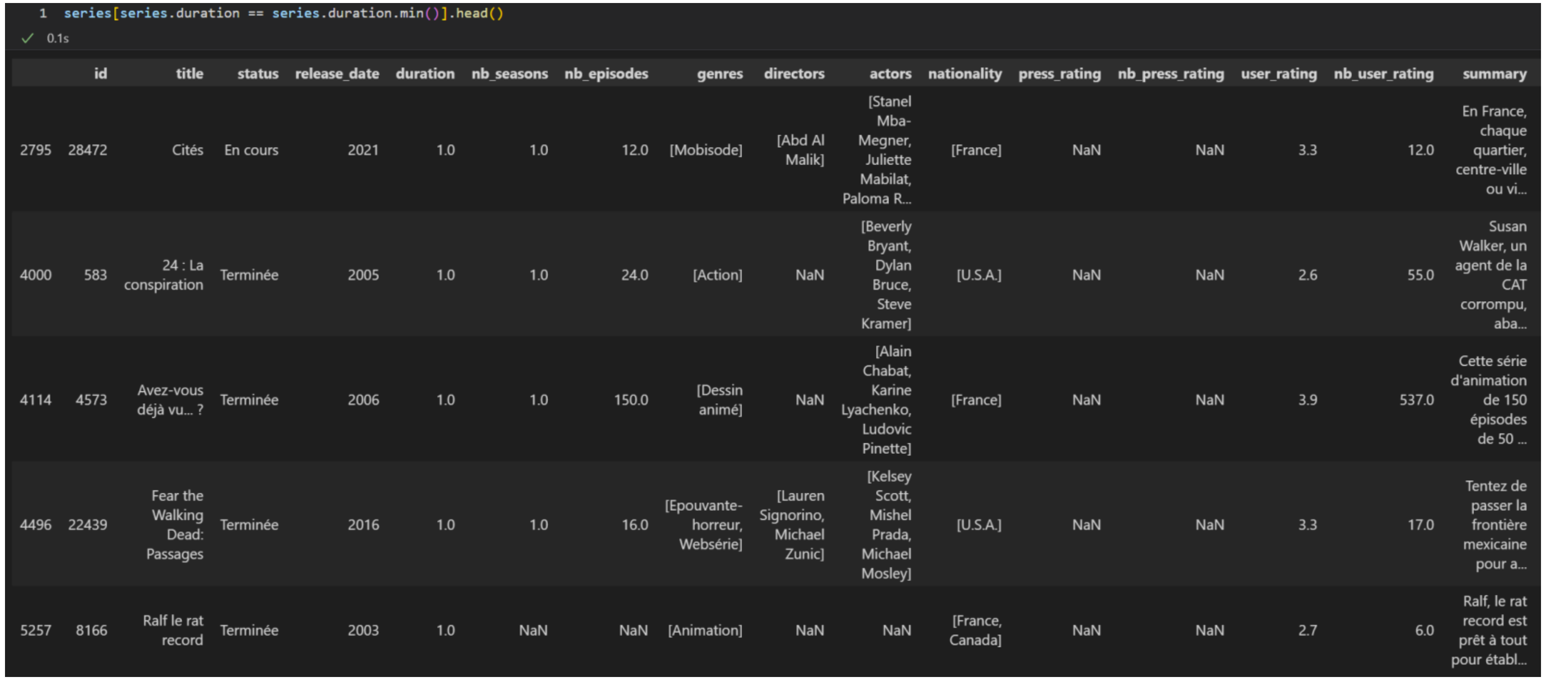
	id	duration	nb_seasons	nb_episodes	press_rating	nb_press_rating	user_rating	nb_user_rating
count	8126.000000	7420.000000	6859.000000	6508.000000	833.000000	833.000000	7006.000000	7006.000000
mean	15548.236402	40.674798	2.403995	41.359096	3.249100	8.099640	3.273351	990.761205
std	9571.320117	20.811908	3.129590	238.676016	0.616424	3.533958	0.517935	5741.767257
min	1.000000	1.000000	1.000000	1.000000	1.300000	1.000000	0.800000	1.000000
25%	6256.250000	25.000000	1.000000	8.000000	2.900000	6.000000	3.000000	7.000000
50%	17480.500000	42.000000	1.000000	14.000000	3.300000	8.000000	3.200000	37.000000
75%	24203.250000	52.000000	3.000000	38.000000	3.700000	10.000000	3.600000	199.000000
max	31807.000000	240.000000	59.000000	13484.000000	5.000000	30.000000	4.700000	206012.000000



```
series[series.duration == series.duration.max()].head()
```

✓ 0.6s

	id	title	status	release_date	duration	nb_seasons	nb_episodes	genres	directors	actors	nationality	press_rating	nb_press_rating	user_rating	nb_user_rating	summary
3136	17910	The Secret Life of Marilyn Monroe	Terminée	2015	240.0	1.0	4.0	[Drame, Historique, Biopic]	[Stephen Kronish]	[Kelli Garner, Susan Sarandon, Emily Watson]	[U.S.A.]	NaN	NaN	3.3	20.0	Une mini-série consacrée à l'icône Marilyn Mon...
3324	23194	Créature	Terminée	1998	240.0	1.0	2.0	[Epouvante-horreur, Science fiction, Thriller]	[Rockne S. O'Bannon]	[Craig T. Nelson, Kim Cattrall, Colm Feore]	[U.S.A.]	NaN	NaN	3.2	4.0	Un monstre amphibien à l'apparence d'un requin...



```
series[series.duration == series.duration.min()].head()
```

✓ 0.1s

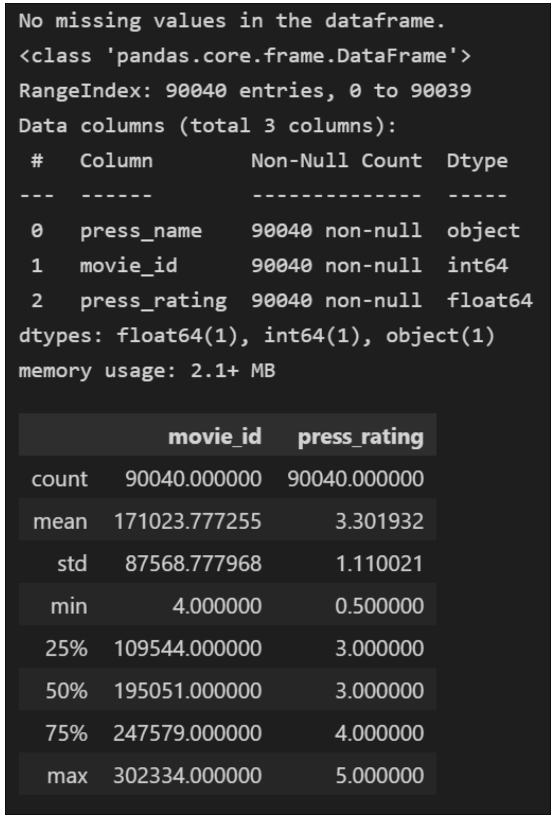
	id	title	status	release_date	duration	nb_seasons	nb_episodes	genres	directors	actors	nationality	press_rating	nb_press_rating	user_rating	nb_user_rating	summary
2795	28472	Cités	En cours	2021	1.0	1.0	12.0	[Mobisode]	[Abd Al Malik]	[Stanel Mba-Megner, Juliette Mabilat, Paloma R...	[France]	NaN	NaN	3.3	12.0	En France, chaque quartier, centre-ville ou vi...
4000	583	24 : La conspiration	Terminée	2005	1.0	1.0	24.0	[Action]	NaN	[Beverly Bryant, Dylan Bruce, Steve Kramer]	[U.S.A.]	NaN	NaN	2.6	55.0	Susan Walker, un agent de la CAT corrompu, aba...
4114	4573	Avez-vous déjà vu... ?	Terminée	2006	1.0	1.0	150.0	[Dessin animé]	NaN	[Alain Chabat, Karine Lyachenko, Ludovic Pinette]	[France]	NaN	NaN	3.9	537.0	Cette série d'animation de 150 épisodes de 50 ...
4496	22439	Fear the Walking Dead: Passages	Terminée	2016	1.0	1.0	16.0	[Epouvante-horreur, Websérie]	[Lauren Signorino, Michael Zunic]	[Kelsey Scott, Mishel Prada, Michael Mosley]	[U.S.A.]	NaN	NaN	3.3	17.0	Tentez de passer la frontière mexicaine pour a...
5257	8166	Ralf le rat record	Terminée	2003	1.0	NaN	NaN	[Animation]	NaN	NaN	[France, Canada]	NaN	NaN	2.7	6.0	Ralf, le rat record est prêt à tout pour établ...

Data Analysis - Data cleaning

Analyses des données

Ratings (BRUT)

Movies Press Ratings

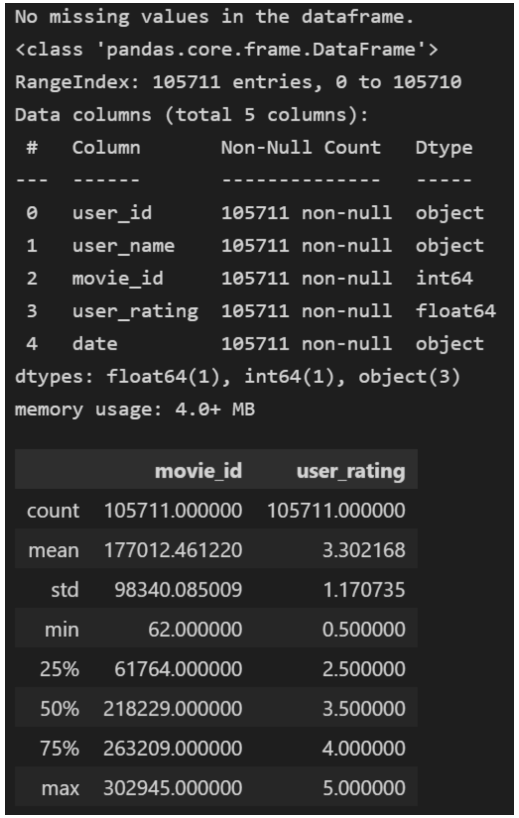


```
No missing values in the dataframe.
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 90040 entries, 0 to 90039
Data columns (total 3 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   press_name    90040 non-null  object
 1   movie_id      90040 non-null  int64
 2   press_rating  90040 non-null  float64
dtypes: float64(1), int64(1), object(1)
memory usage: 2.1+ MB
```

	movie_id	press_rating
count	90040.000000	90040.000000
mean	171023.777255	3.301932
std	87568.777968	1.110021
min	4.000000	0.500000
25%	109544.000000	3.000000
50%	195051.000000	3.000000
75%	247579.000000	4.000000
max	302334.000000	5.000000

Rows: 90040
Columns: 3

Movies User Ratings*

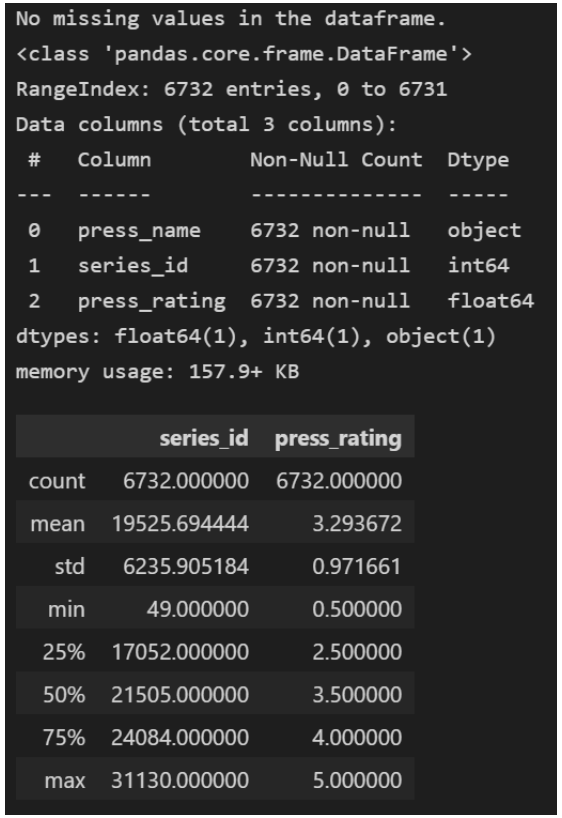


```
No missing values in the dataframe.
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 105711 entries, 0 to 105710
Data columns (total 5 columns):
 #   Column       Non-Null Count   Dtype
---  ------       --------------   -----
 0   user_id      105711 non-null  object
 1   user_name    105711 non-null  object
 2   movie_id     105711 non-null  int64
 3   user_rating  105711 non-null  float64
 4   date         105711 non-null  object
dtypes: float64(1), int64(1), object(3)
memory usage: 4.0+ MB
```

	movie_id	user_rating
count	105711.000000	105711.000000
mean	177012.461220	3.302168
std	98340.085009	1.170735
min	62.000000	0.500000
25%	61764.000000	2.500000
50%	218229.000000	3.500000
75%	263209.000000	4.000000
max	302945.000000	5.000000

Rows: 105711
Columns: 5

Series Press Ratings

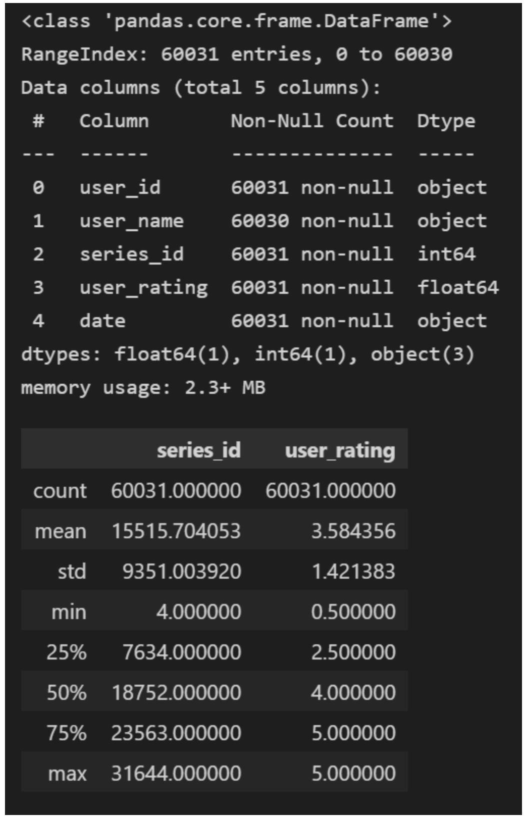


```
No missing values in the dataframe.
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6732 entries, 0 to 6731
Data columns (total 3 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   press_name    6732 non-null   object
 1   series_id     6732 non-null   int64
 2   press_rating  6732 non-null   float64
dtypes: float64(1), int64(1), object(1)
memory usage: 157.9+ KB
```

	series_id	press_rating
count	6732.000000	6732.000000
mean	19525.694444	3.293672
std	6235.905184	0.971661
min	49.000000	0.500000
25%	17052.000000	2.500000
50%	21505.000000	3.500000
75%	24084.000000	4.000000
max	31130.000000	5.000000

Rows: 6732
Columns: 3

Series User Ratings*



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 60031 entries, 0 to 60030
Data columns (total 5 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   user_id      60031 non-null  object
 1   user_name    60030 non-null  object
 2   series_id    60031 non-null  int64
 3   user_rating  60031 non-null  float64
 4   date         60031 non-null  object
dtypes: float64(1), int64(1), object(3)
memory usage: 2.3+ MB
```

	series_id	user_rating
count	60031.000000	60031.000000
mean	15515.704053	3.584356
std	9351.003920	1.421383
min	4.000000	0.500000
25%	7634.000000	2.500000
50%	18752.000000	4.000000
75%	23563.000000	5.000000
max	31644.000000	5.000000

Rows: 60031
Columns: 5

* Tirage de 100 pages

Data Analysis - Data cleaning

Processus de Cleaning

Drop duplicates

Drop genre « Divers » in series (trop peu de valeurs)

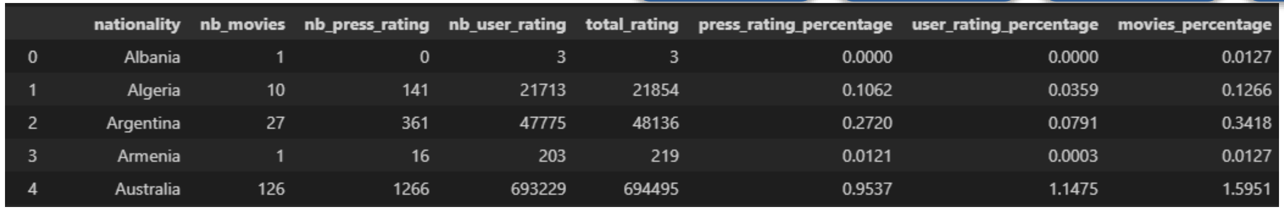
Drop series sans année de sortie

Création de 4 tables supplémentaires

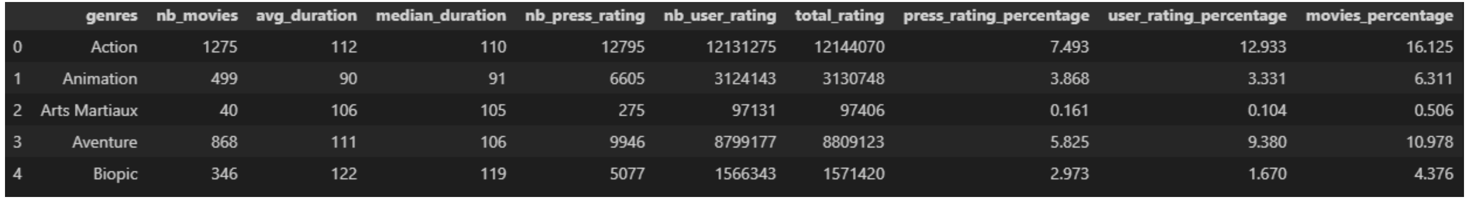
- m_genres
- s_genres
- m_nationality
- s_nationality

Amélioration format donnée et cohérence

- Impute missing duration
- Drop ratings manquants
- Series
  - Drop seuil de NA
  - Date de sortie en année
- Movies
  - Ajout saison de sortie



	nationality	nb_movies	nb_press_rating	nb_user_rating	total_rating	press_rating_percentage	user_rating_percentage	movies_percentage
0	Albania	1	0	3	3	0.0000	0.0000	0.0127
1	Algeria	10	141	21713	21854	0.1062	0.0359	0.1266
2	Argentina	27	361	47775	48136	0.2720	0.0791	0.3418
3	Armenia	1	16	203	219	0.0121	0.0003	0.0127
4	Australia	126	1266	693229	694495	0.9537	1.1475	1.5951



	genres	nb_movies	avg_duration	median_duration	nb_press_rating	nb_user_rating	total_rating	press_rating_percentage	user_rating_percentage	movies_percentage
0	Action	1275	112	110	12795	12131275	12144070	7.493	12.933	16.125
1	Animation	499	90	91	6605	3124143	3130748	3.868	3.331	6.311
2	Arts Martiaux	40	106	105	275	97131	97406	0.161	0.104	0.506
3	Aventure	868	111	106	9946	8799177	8809123	5.825	9.380	10.978
4	Biopic	346	122	119	5077	1566343	1571420	2.973	1.670	4.376

Data Analysis - Data cleaning

Processus de Cleaning

Drop duplicates

Drop genre « Divers » in series (trop peu de valeurs)

Drop series sans année de sortie

Création de 4 tables supplémentaires

- m_genres
- s_genres
- m_nationality
- s_nationality

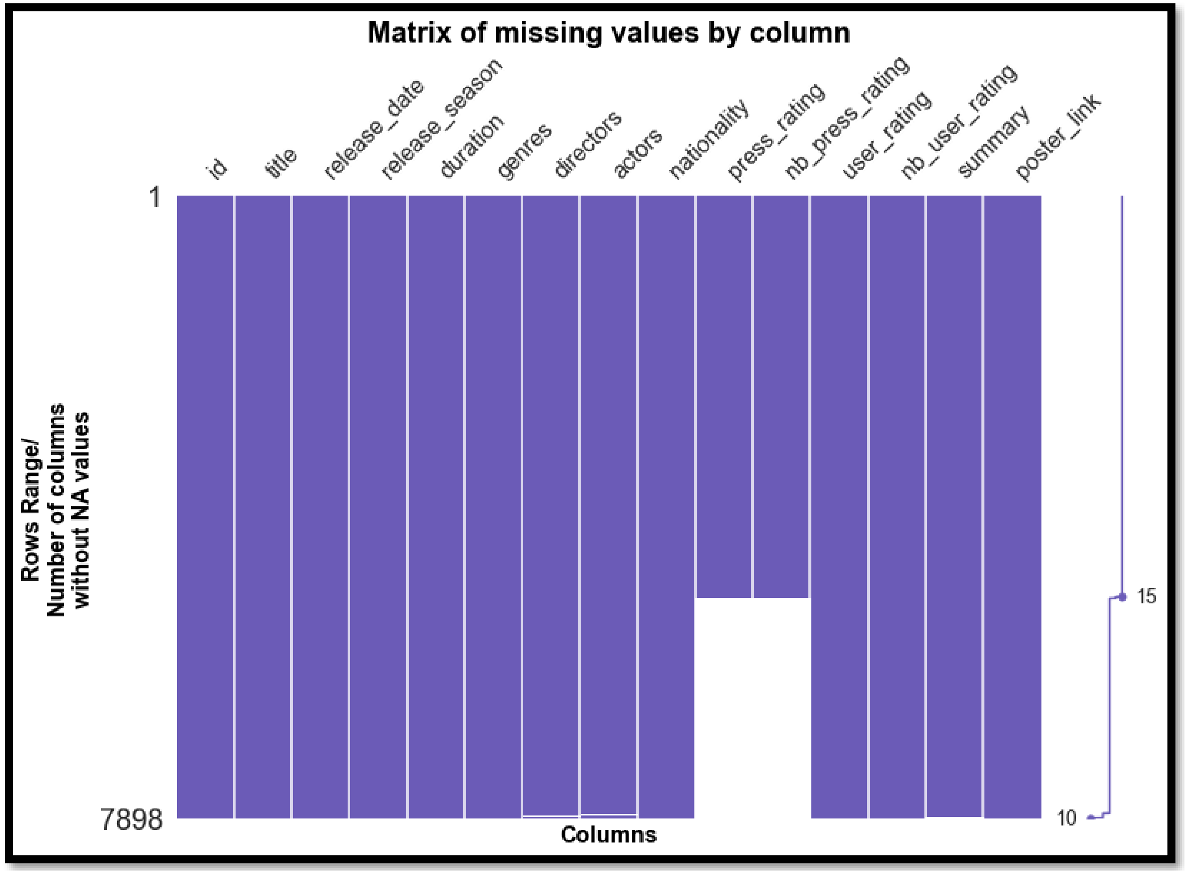
Amélioration format donnée et cohérence

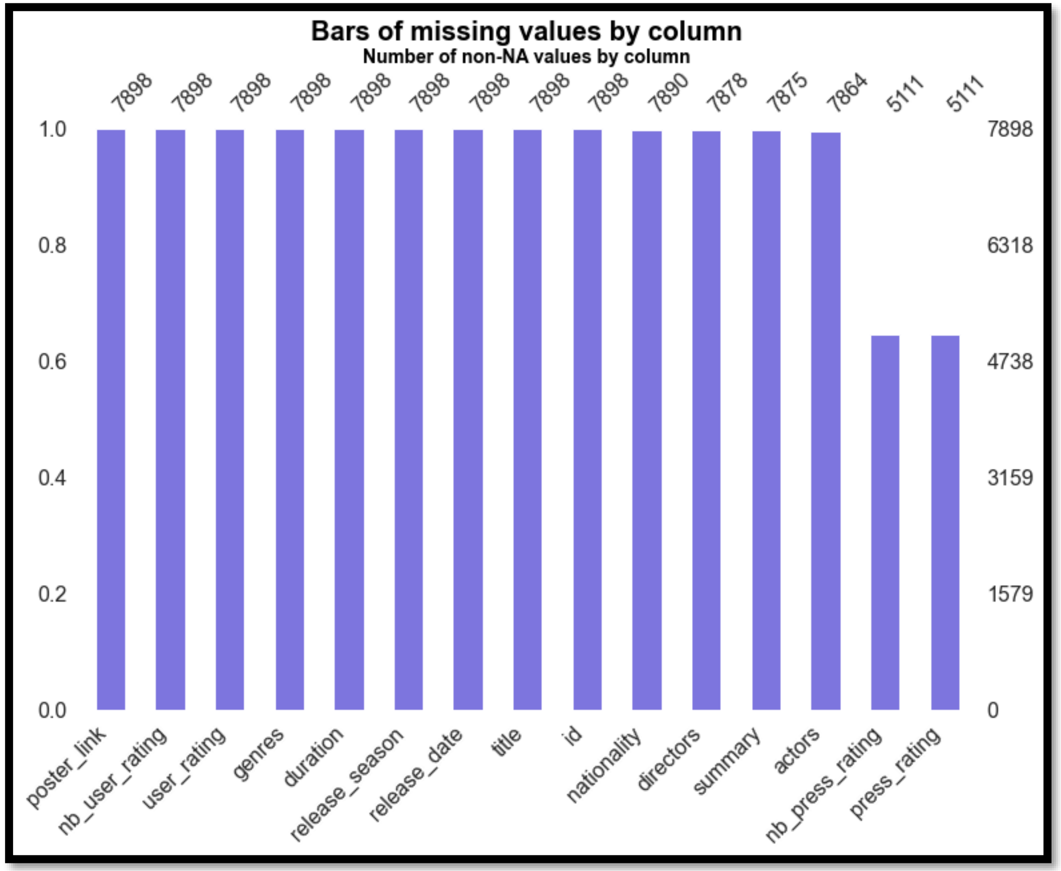
- Impute missing duration
- Drop ratings manquants
- Series
  - Drop seuil de NA
  - Date de sortie en année
- Movies
  - Ajout saison de sortie

Data Analysis - Data cleaning

Analyses des données

MOVIES (CLEAN)

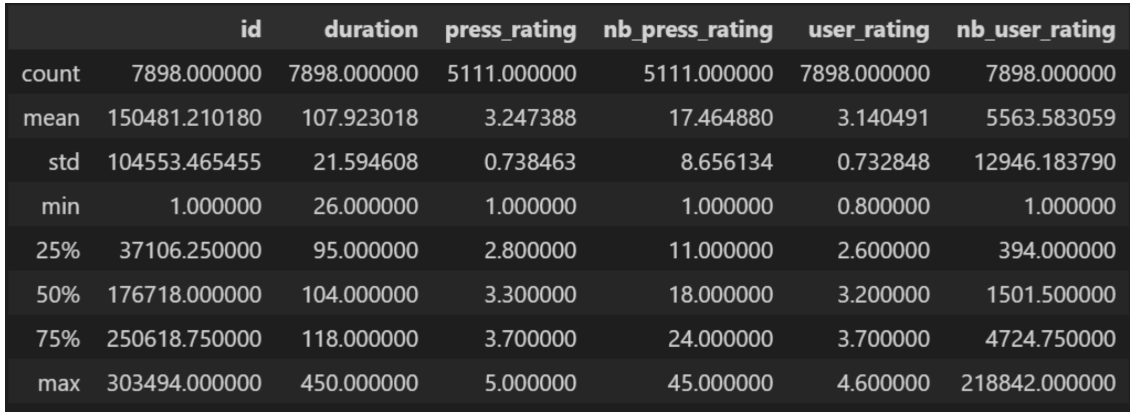




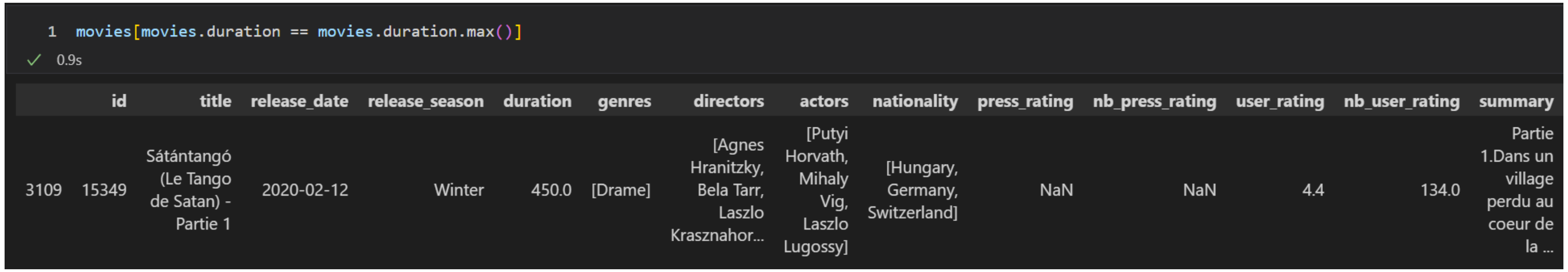
Rows: 7898
Columns: 15
→ -128 rows; +1 column

Data Analysis - Data cleaning

MOVIES (CLEAN)



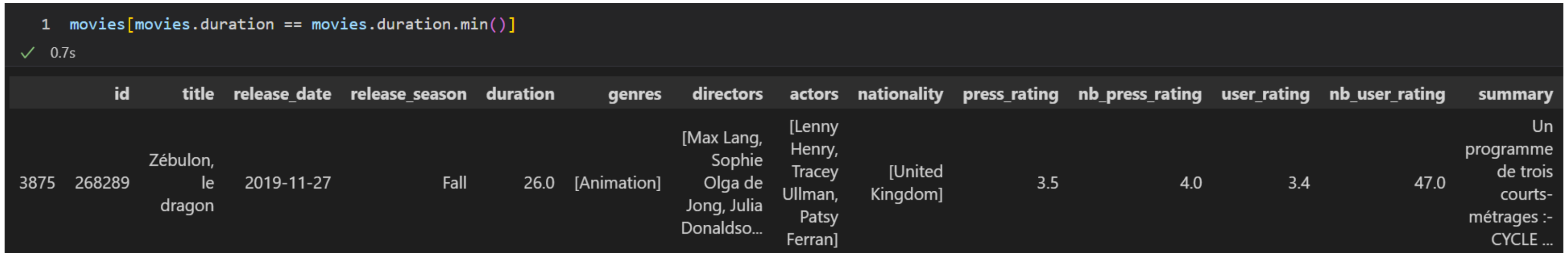
	id	duration	press_rating	nb_press_rating	user_rating	nb_user_rating
count	7898.000000	7898.000000	5111.000000	5111.000000	7898.000000	7898.000000
mean	150481.210180	107.923018	3.247388	17.464880	3.140491	5563.583059
std	104553.465455	21.594608	0.738463	8.656134	0.732848	12946.183790
min	1.000000	26.000000	1.000000	1.000000	0.800000	1.000000
25%	37106.250000	95.000000	2.800000	11.000000	2.600000	394.000000
50%	176718.000000	104.000000	3.300000	18.000000	3.200000	1501.500000
75%	250618.750000	118.000000	3.700000	24.000000	3.700000	4724.750000
max	303494.000000	450.000000	5.000000	45.000000	4.600000	218842.000000



```
movies[movies.duration == movies.duration.max()]
```

✓ 0.9s

	id	title	release_date	release_season	duration	genres	directors	actors	nationality	press_rating	nb_press_rating	user_rating	nb_user_rating	summary
3109	15349	Sátántangó (Le Tango de Satan) - Partie 1	2020-02-12	Winter	450.0	[Drame]	[Agnes Hranitzky, Bela Tarr, Laszlo Krasznahor...	[Putyi Horvath, Mihaly Vig, Laszlo Lugossy]	[Hungary, Germany, Switzerland]	NaN	NaN	4.4	134.0	Partie 1.Dans un village perdu au coeur de la ...



```
movies[movies.duration == movies.duration.min()]
```

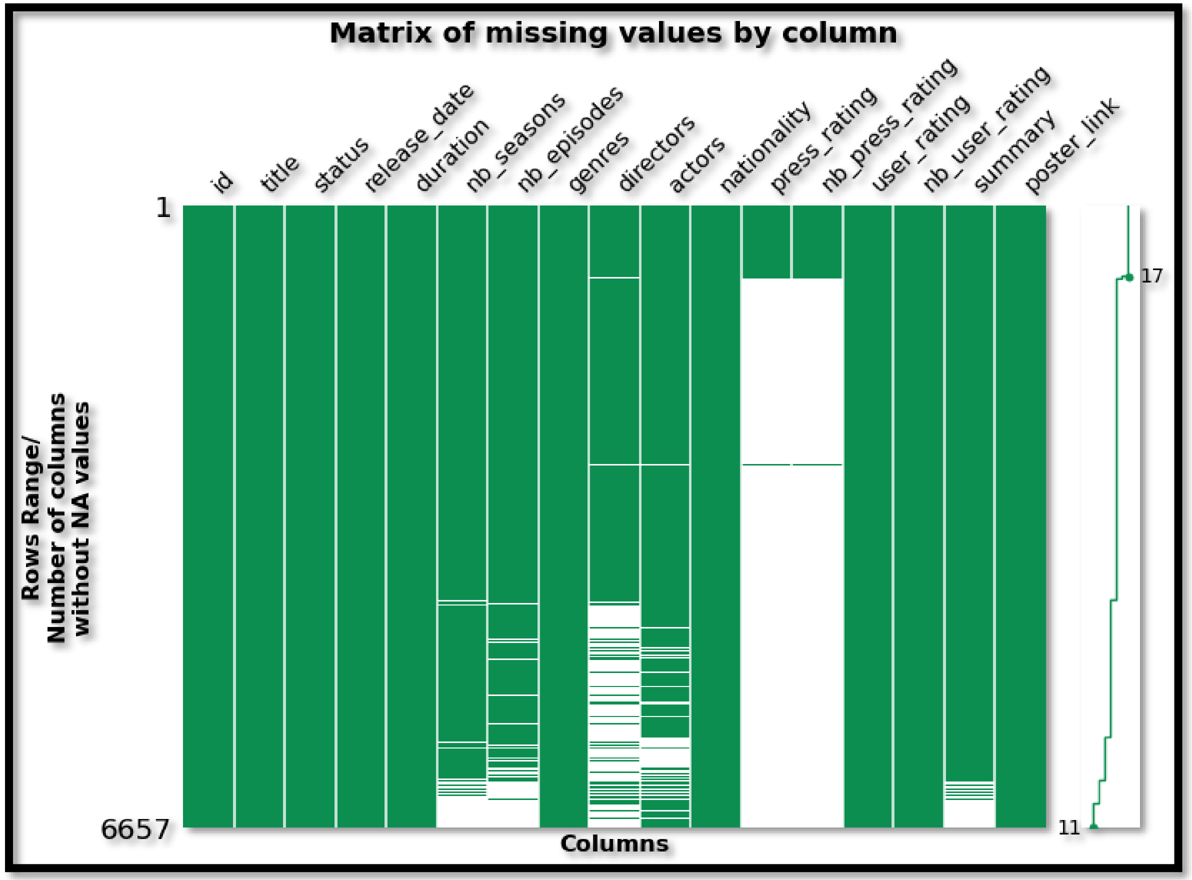
✓ 0.7s

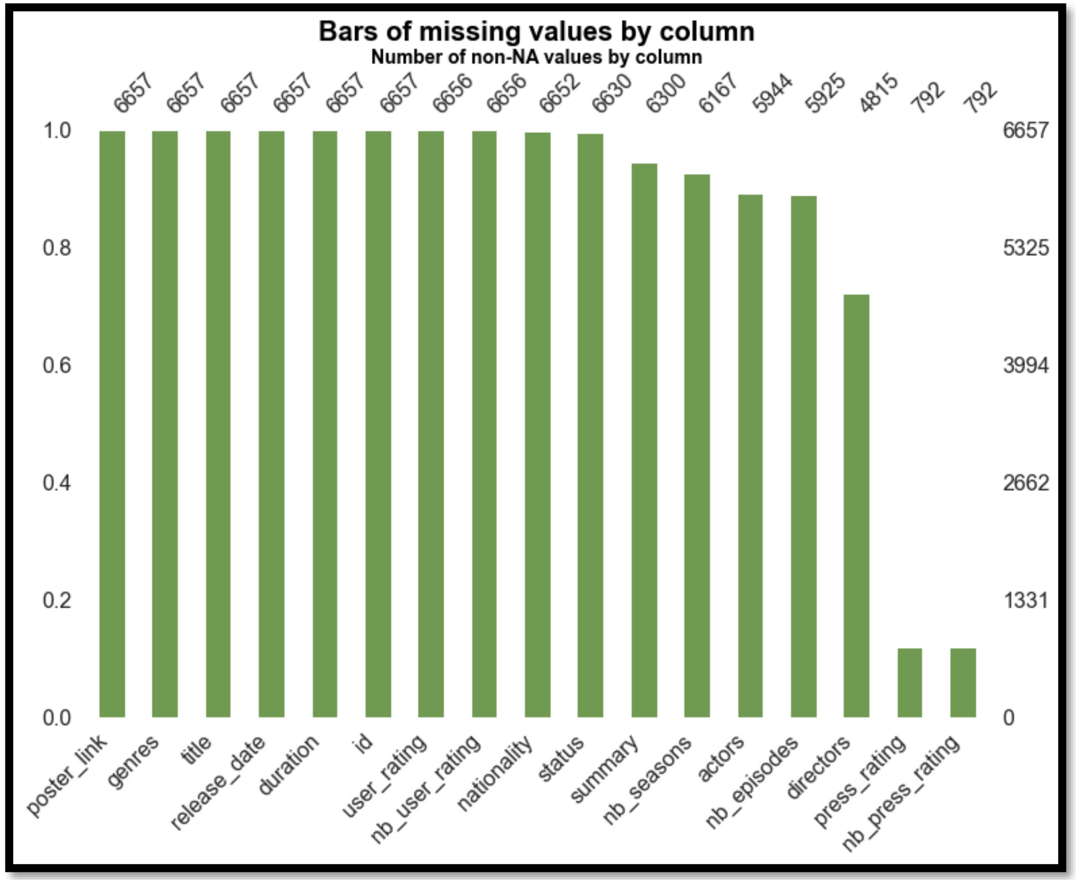
	id	title	release_date	release_season	duration	genres	directors	actors	nationality	press_rating	nb_press_rating	user_rating	nb_user_rating	summary
3875	268289	Zébulon, le dragon	2019-11-27	Fall	26.0	[Animation]	[Max Lang, Sophie Olga de Jong, Julia Donaldso...	[Lenny Henry, Tracey Ullman, Patsy Ferran]	[United Kingdom]	3.5	4.0	3.4	47.0	Un programme de trois courts-métrages :- CYCLE ...

Data Analysis - Data cleaning

Analyses des données

SERIES (CLEAN)



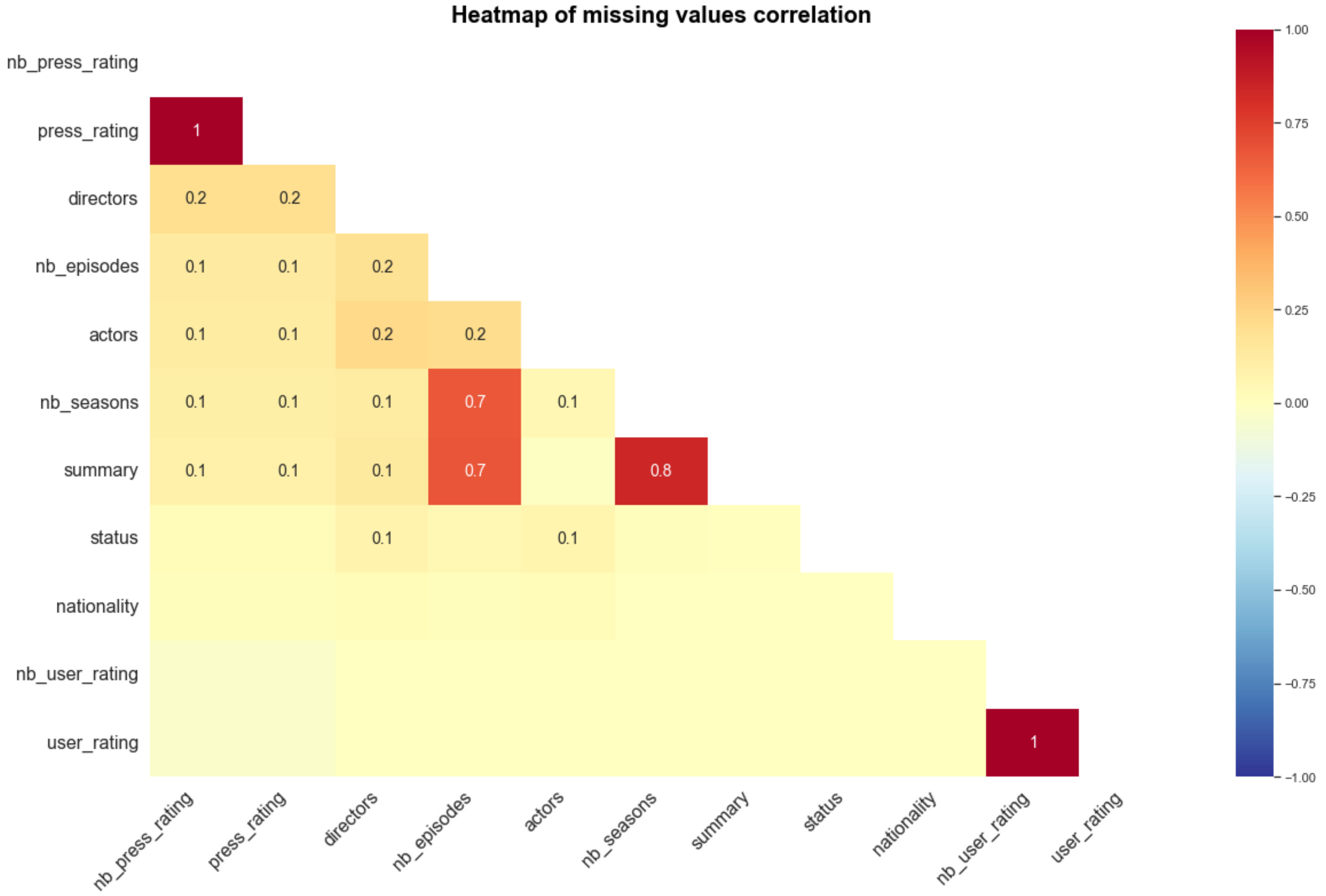


Rows: 6657
Columns: 17
→ -1469 rows

Data Analysis - Data cleaning

Analyses des données

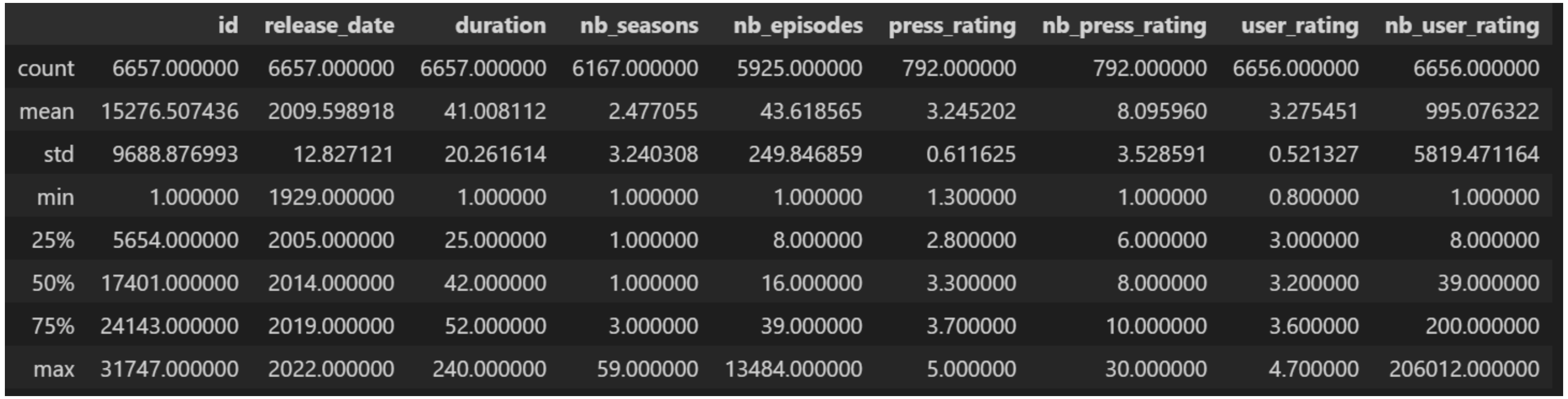
SERIES (CLEAN)



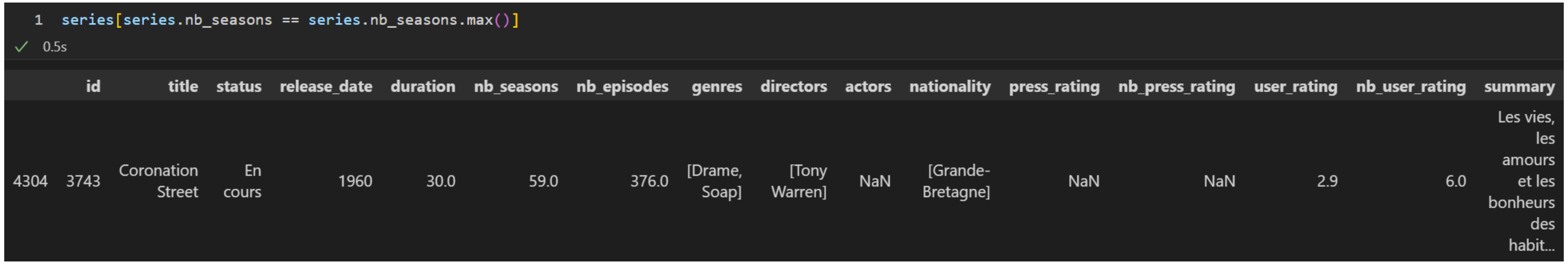
Data Analysis - Data cleaning

Analyses des données

SERIES (CLEAN)



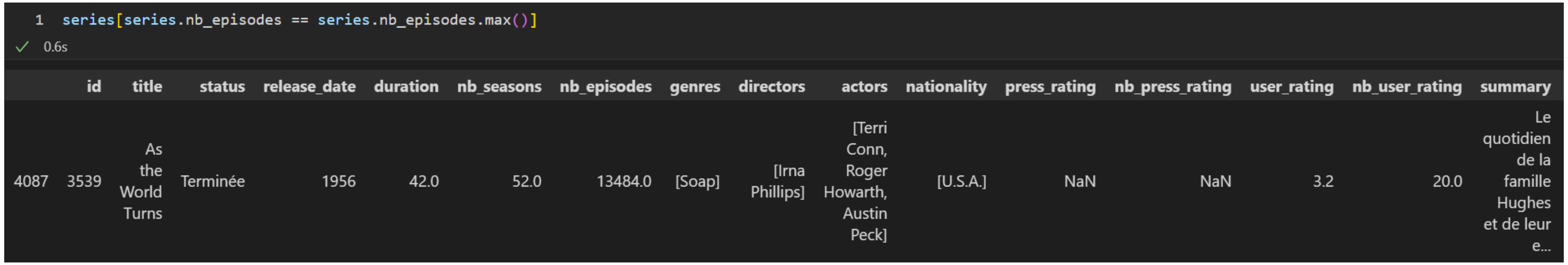
	id	release_date	duration	nb_seasons	nb_episodes	press_rating	nb_press_rating	user_rating	nb_user_rating
count	6657.000000	6657.000000	6657.000000	6167.000000	5925.000000	792.000000	792.000000	6656.000000	6656.000000
mean	15276.507436	2009.598918	41.008112	2.477055	43.618565	3.245202	8.095960	3.275451	995.076322
std	9688.876993	12.827121	20.261614	3.240308	249.846859	0.611625	3.528591	0.521327	5819.471164
min	1.000000	1929.000000	1.000000	1.000000	1.000000	1.300000	1.000000	0.800000	1.000000
25%	5654.000000	2005.000000	25.000000	1.000000	8.000000	2.800000	6.000000	3.000000	8.000000
50%	17401.000000	2014.000000	42.000000	1.000000	16.000000	3.300000	8.000000	3.200000	39.000000
75%	24143.000000	2019.000000	52.000000	3.000000	39.000000	3.700000	10.000000	3.600000	200.000000
max	31747.000000	2022.000000	240.000000	59.000000	13484.000000	5.000000	30.000000	4.700000	206012.000000



```
series[series.nb_seasons == series.nb_seasons.max()]
```

✓ 0.5s

	id	title	status	release_date	duration	nb_seasons	nb_episodes	genres	directors	actors	nationality	press_rating	nb_press_rating	user_rating	nb_user_rating	summary
4304	3743	Coronation Street	En cours	1960	30.0	59.0	376.0	[Drame, Soap]	[Tony Warren]	NaN	[Grande-Bretagne]	NaN	NaN	2.9	6.0	Les vies, les amours et les bonheurs des habit...



```
series[series.nb_episodes == series.nb_episodes.max()]
```

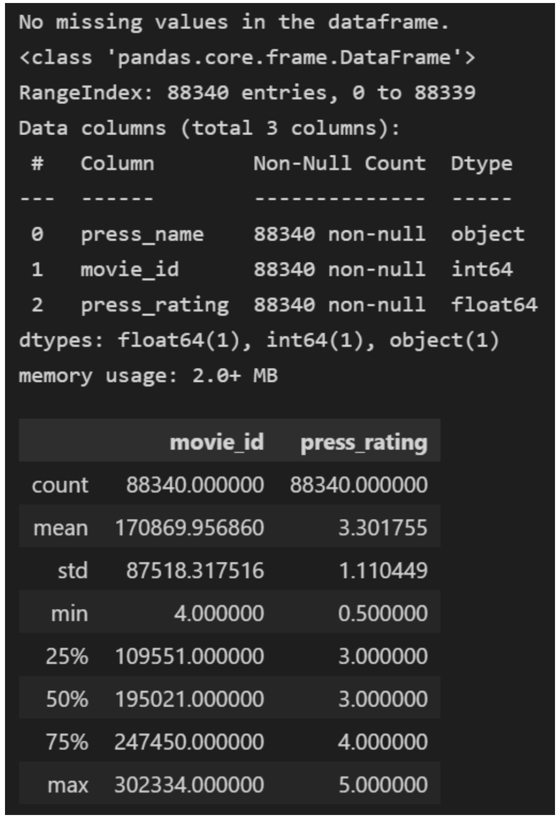
✓ 0.6s

	id	title	status	release_date	duration	nb_seasons	nb_episodes	genres	directors	actors	nationality	press_rating	nb_press_rating	user_rating	nb_user_rating	summary
4087	3539	As the World Turns	Terminée	1956	42.0	52.0	13484.0	[Soap]	[Irna Phillips]	[Terri Conn, Roger Howarth, Austin Peck]	[U.S.A.]	NaN	NaN	3.2	20.0	Le quotidien de la famille Hughes et de leur e...

Data Analysis - Data cleaning

Ratings (CLEAN)

Movies Press Ratings

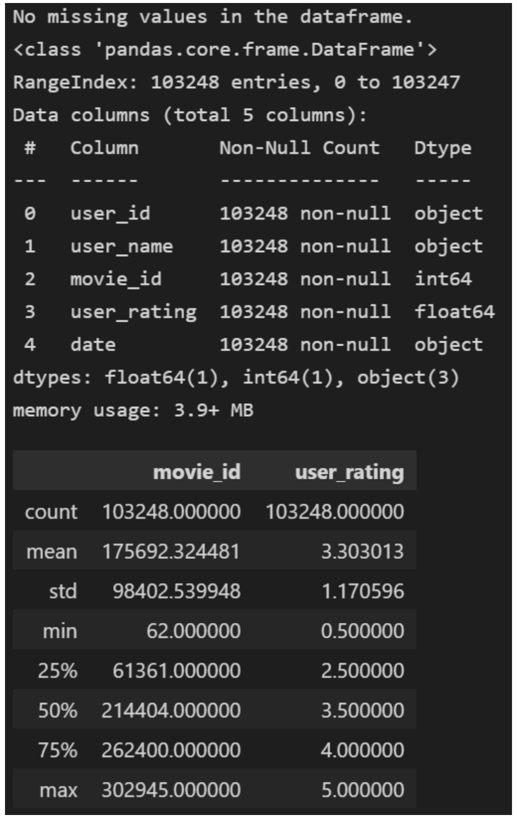


```
No missing values in the dataframe.
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 88340 entries, 0 to 88339
Data columns (total 3 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   press_name    88340 non-null  object
 1   movie_id      88340 non-null  int64
 2   press_rating  88340 non-null  float64
dtypes: float64(1), int64(1), object(1)
memory usage: 2.0+ MB
```

	movie_id	press_rating
count	88340.000000	88340.000000
mean	170869.956860	3.301755
std	87518.317516	1.110449
min	4.000000	0.500000
25%	109551.000000	3.000000
50%	195021.000000	3.000000
75%	247450.000000	4.000000
max	302334.000000	5.000000

Rows: 88340
Columns: 3
→ -1700 rows

Movies User Ratings*

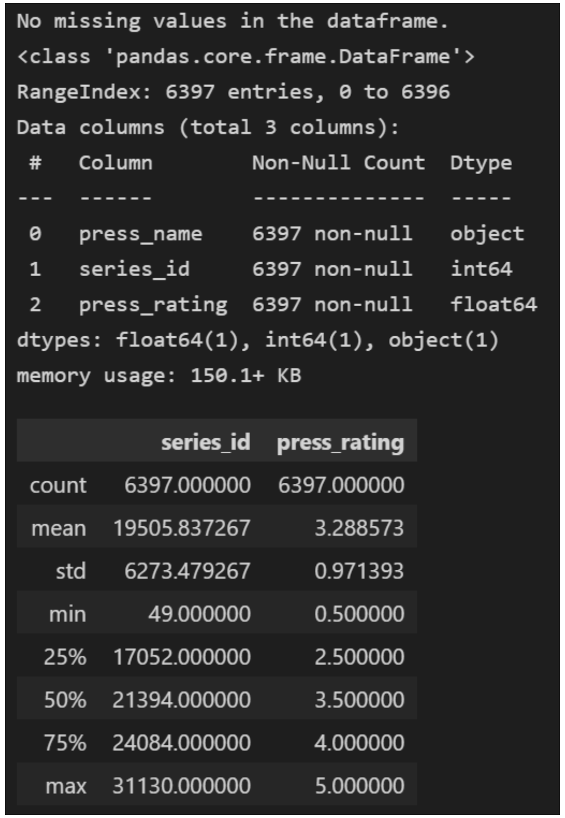


```
No missing values in the dataframe.
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 103248 entries, 0 to 103247
Data columns (total 5 columns):
 #   Column       Non-Null Count   Dtype
---  ------       --------------   -----
 0   user_id      103248 non-null  object
 1   user_name    103248 non-null  object
 2   movie_id     103248 non-null  int64
 3   user_rating  103248 non-null  float64
 4   date         103248 non-null  object
dtypes: float64(1), int64(1), object(3)
memory usage: 3.9+ MB
```

	movie_id	user_rating
count	103248.000000	103248.000000
mean	175692.324481	3.303013
std	98402.539948	1.170596
min	62.000000	0.500000
25%	61361.000000	2.500000
50%	214404.000000	3.500000
75%	262400.000000	4.000000
max	302945.000000	5.000000

Rows: 103248
Columns: 5
→ -2463 rows

Series Press Ratings

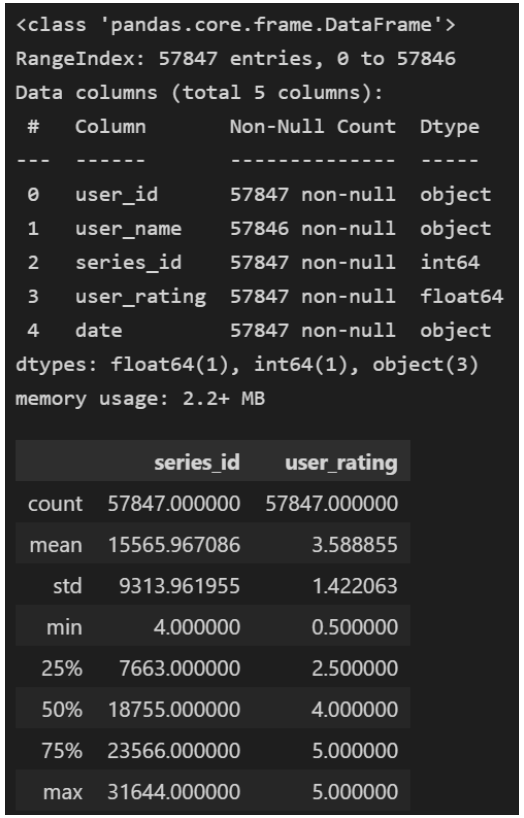


```
No missing values in the dataframe.
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6397 entries, 0 to 6396
Data columns (total 3 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   press_name    6397 non-null   object
 1   series_id     6397 non-null   int64
 2   press_rating  6397 non-null   float64
dtypes: float64(1), int64(1), object(1)
memory usage: 150.1+ KB
```

	series_id	press_rating
count	6397.000000	6397.000000
mean	19505.837267	3.288573
std	6273.479267	0.971393
min	49.000000	0.500000
25%	17052.000000	2.500000
50%	21394.000000	3.500000
75%	24084.000000	4.000000
max	31130.000000	5.000000

Rows: 6397
Columns: 3
→ -6732 rows

Series User Ratings*



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 57847 entries, 0 to 57846
Data columns (total 5 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   user_id      57847 non-null  object
 1   user_name    57846 non-null  object
 2   series_id    57847 non-null  int64
 3   user_rating  57847 non-null  float64
 4   date         57847 non-null  object
dtypes: float64(1), int64(1), object(3)
memory usage: 2.2+ MB
```

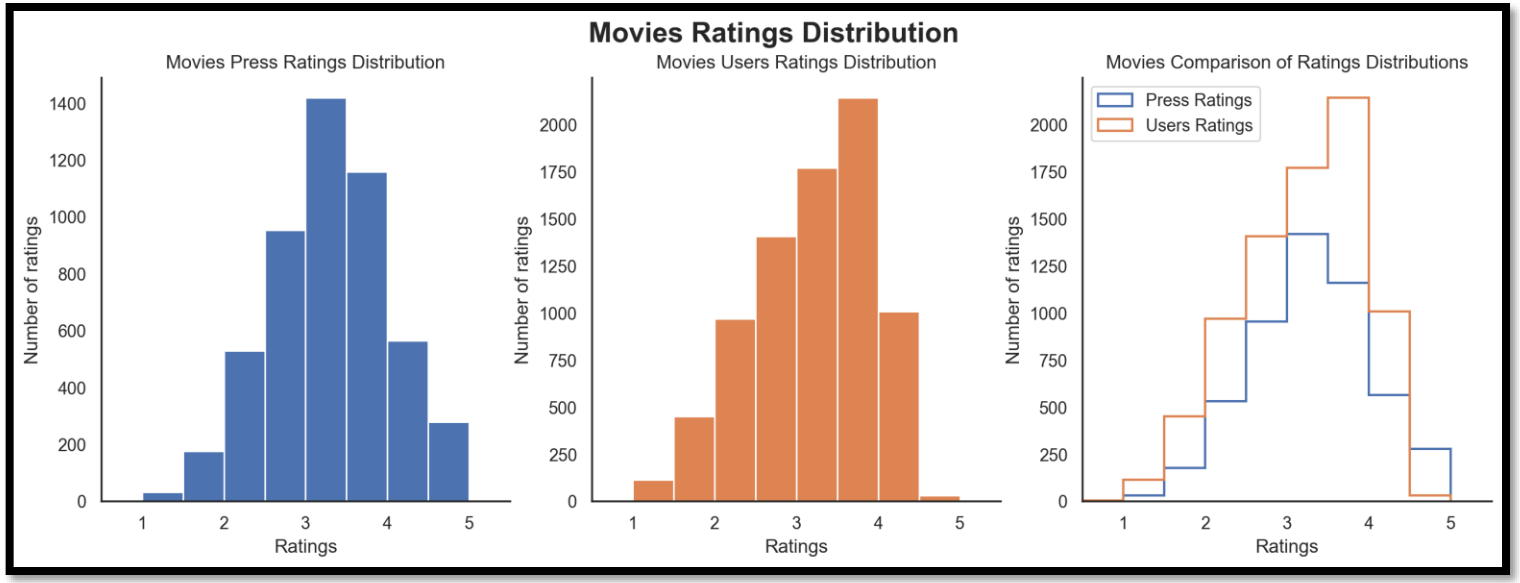
	series_id	user_rating
count	57847.000000	57847.000000
mean	15565.967086	3.588855
std	9313.961955	1.422063
min	4.000000	0.500000
25%	7663.000000	2.500000
50%	18755.000000	4.000000
75%	23566.000000	5.000000
max	31644.000000	5.000000

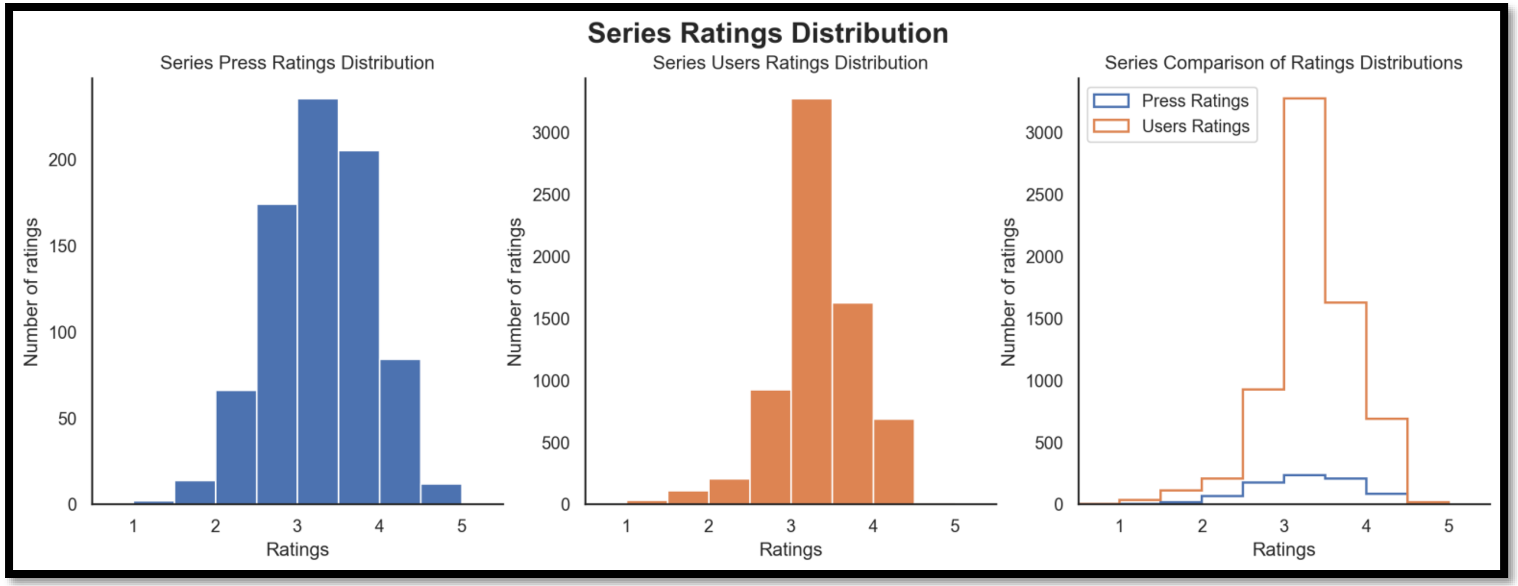
Rows: 57847
Columns: 5
→ -2184 rows

* Tirage de 100 pages

Data Analysis – Analyse Exploratoire 1

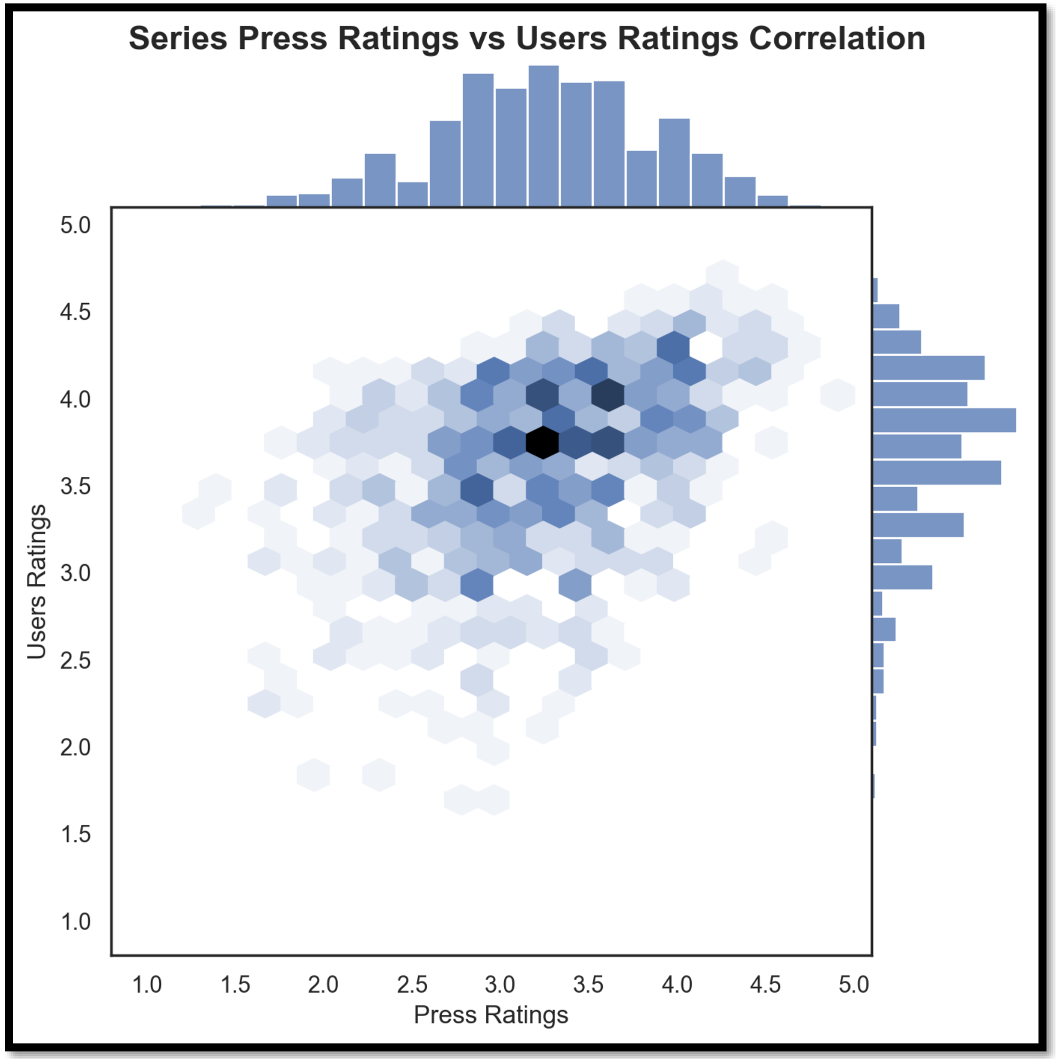
Analyses des données

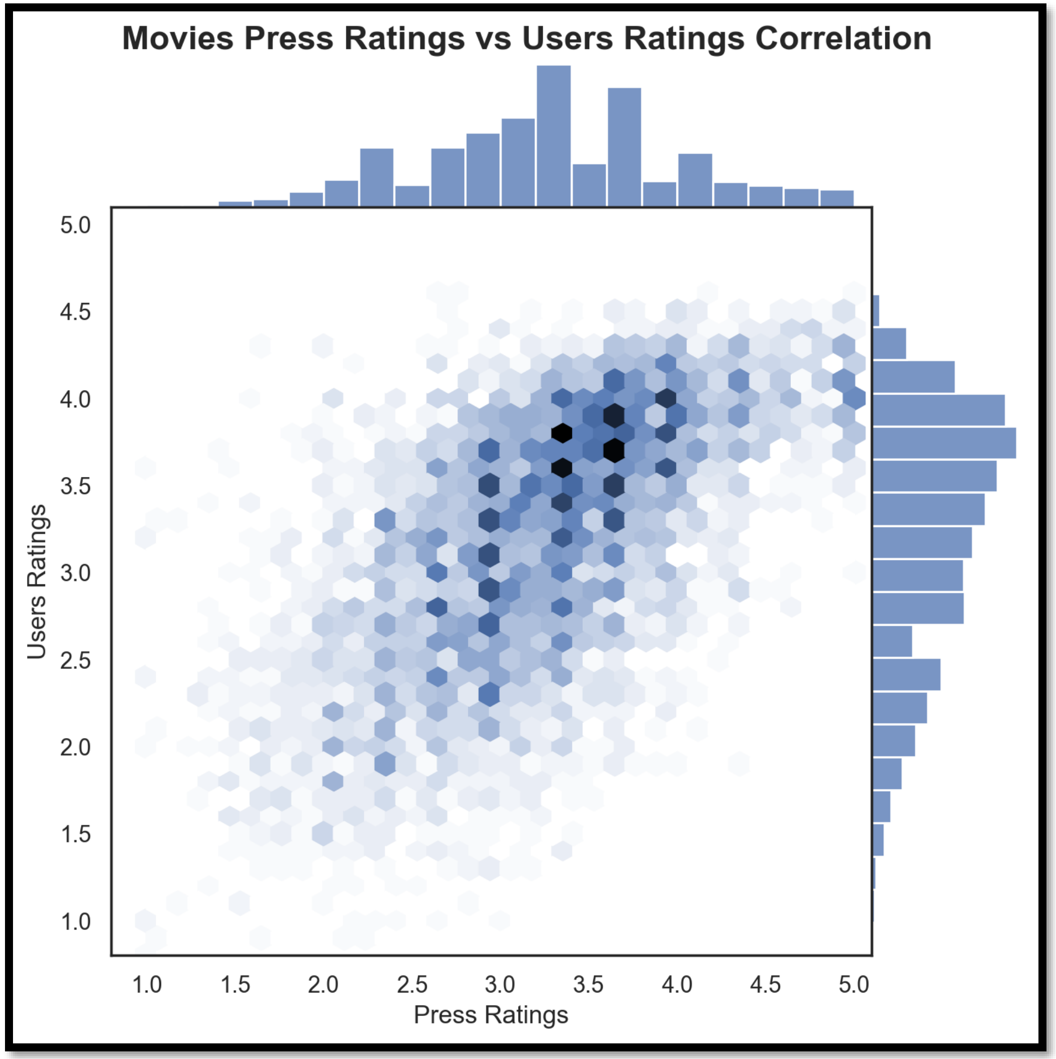




Data Analysis - Analyse Exploratoire 1

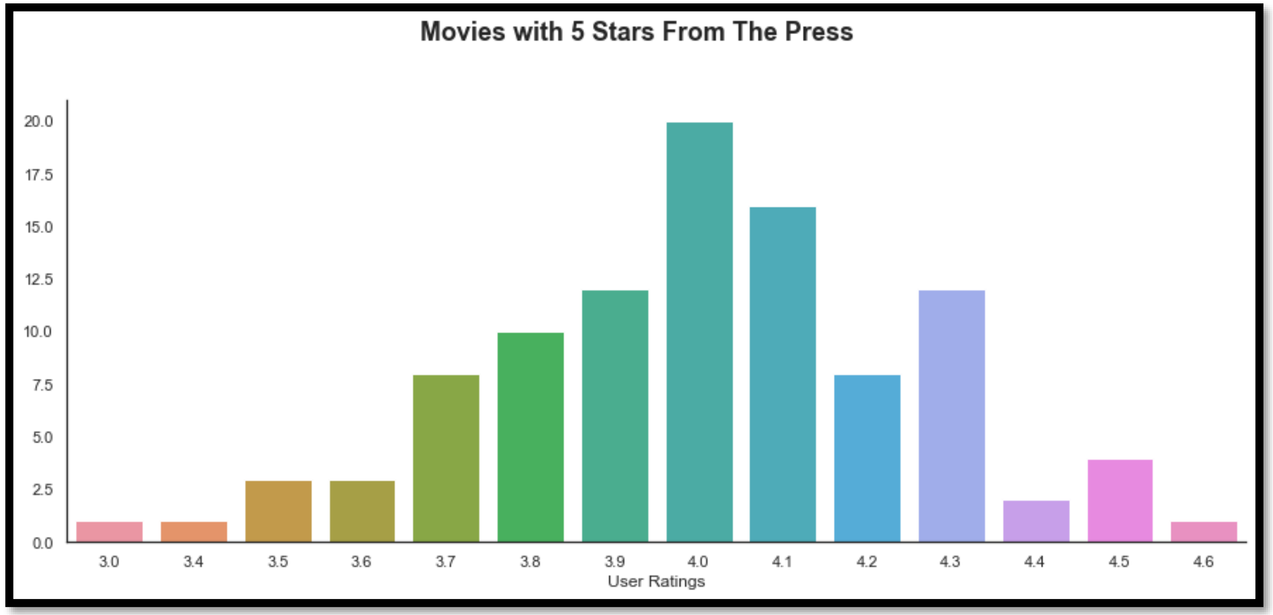
Analyses des données

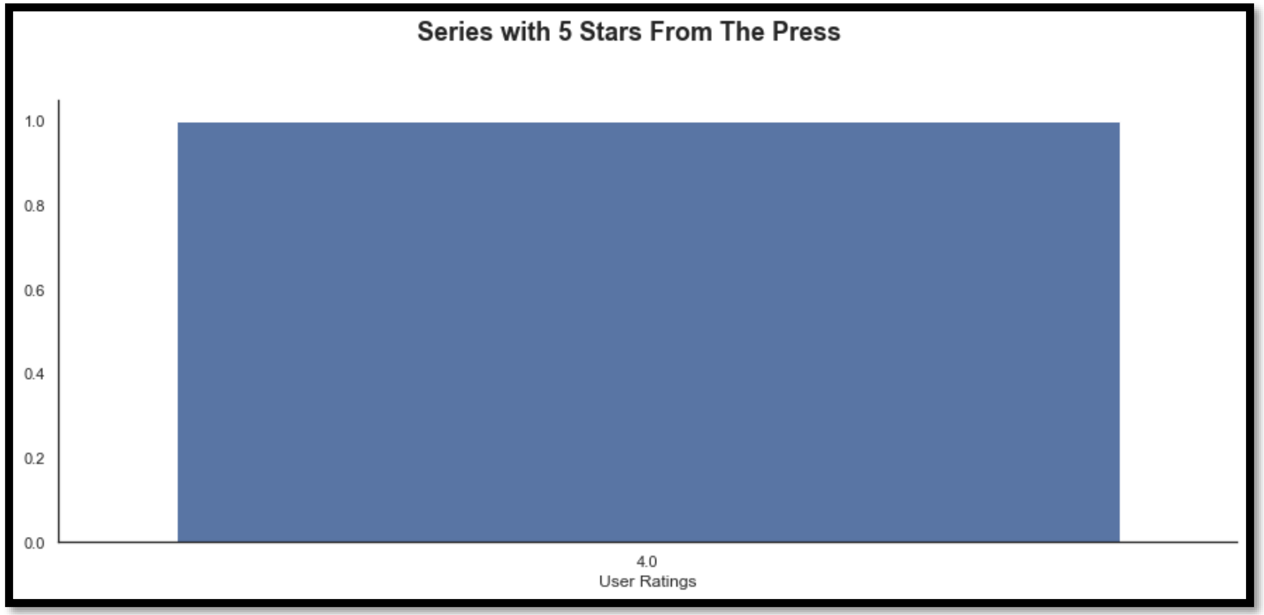


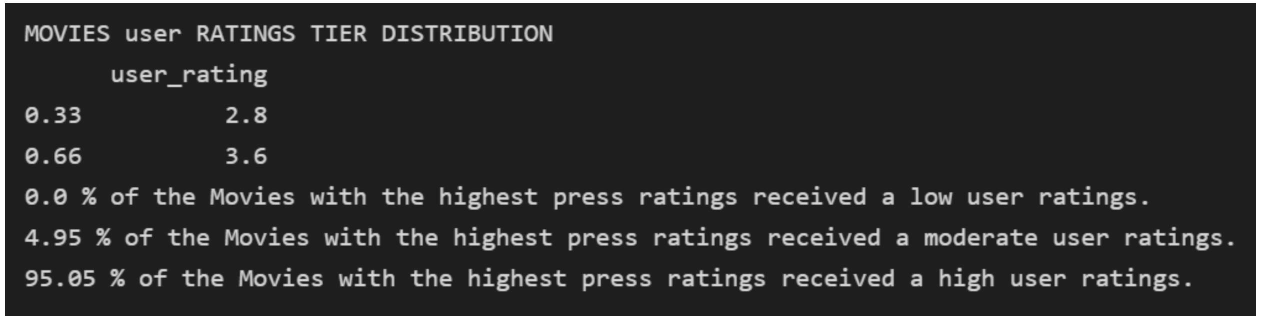


Data Analysis - Analyse Exploratoire 1

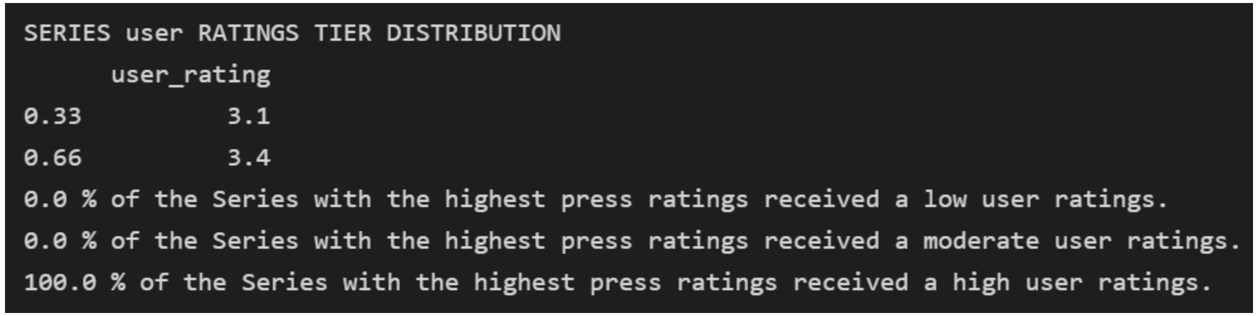
Compare users to press ratings


Movies with 5 Stars From The Press


Series with 5 Stars From The Press



```
MOVIES user RATINGS TIER DISTRIBUTION
      user_rating
0.33          2.8
0.66          3.6
0.0 % of the Movies with the highest press ratings received a low user ratings.
4.95 % of the Movies with the highest press ratings received a moderate user ratings.
95.05 % of the Movies with the highest press ratings received a high user ratings.
```

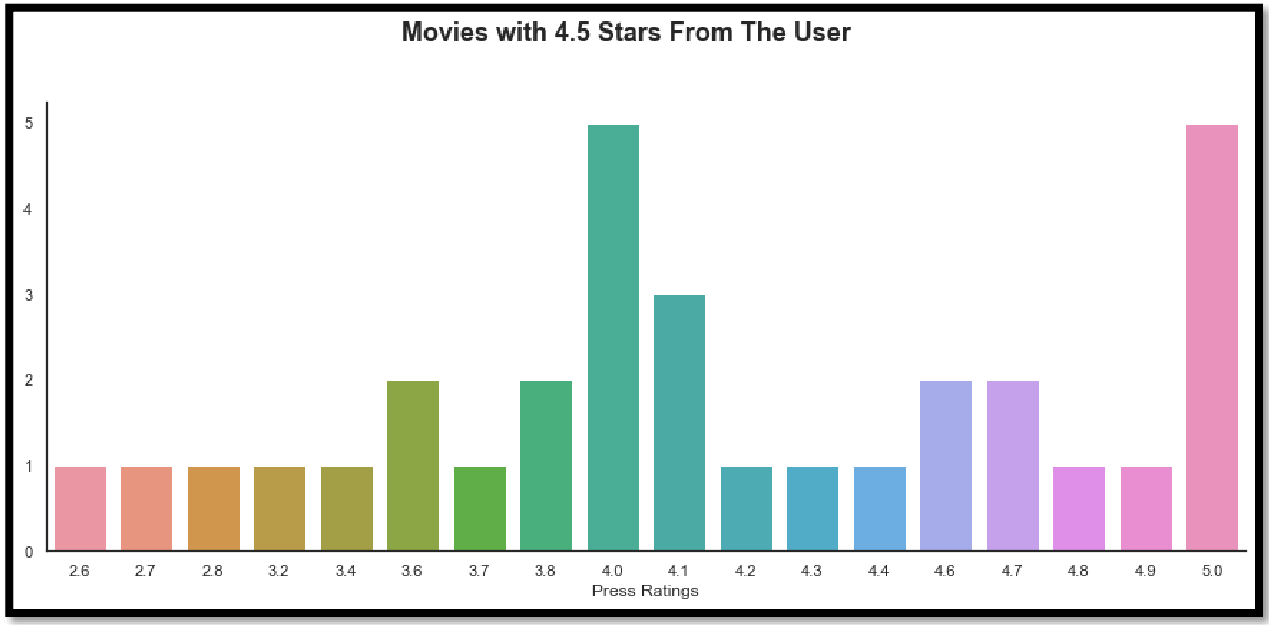


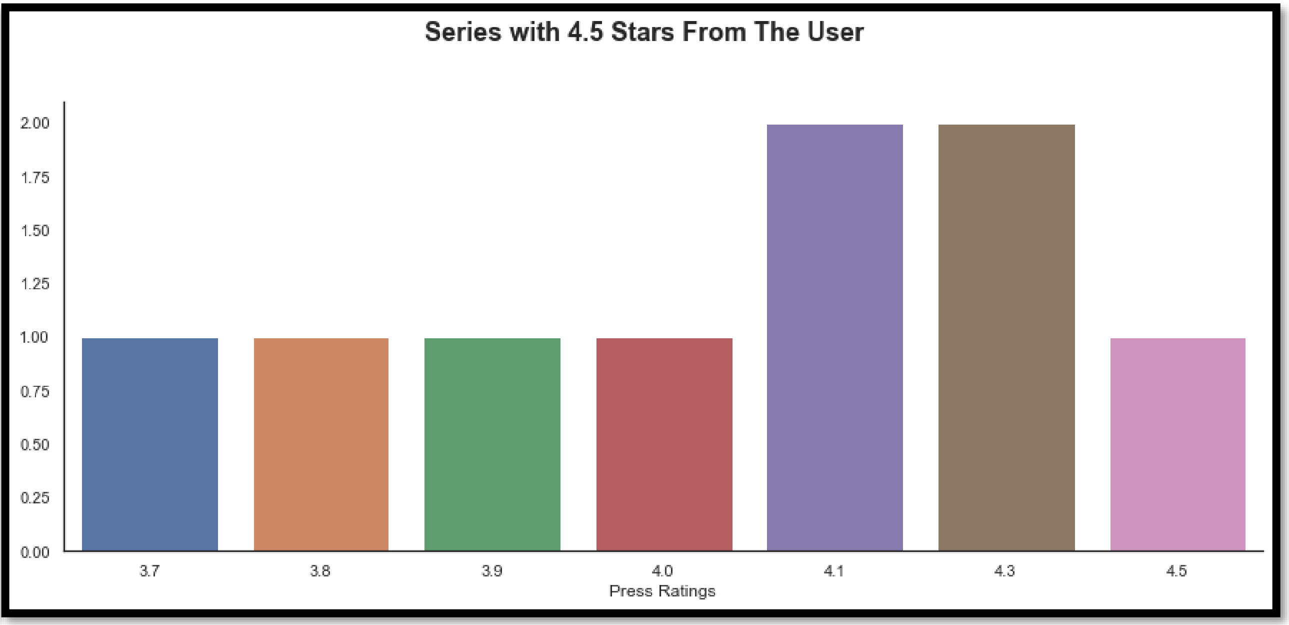
```
SERIES user RATINGS TIER DISTRIBUTION
      user_rating
0.33          3.1
0.66          3.4
0.0 % of the Series with the highest press ratings received a low user ratings.
0.0 % of the Series with the highest press ratings received a moderate user ratings.
100.0 % of the Series with the highest press ratings received a high user ratings.
```

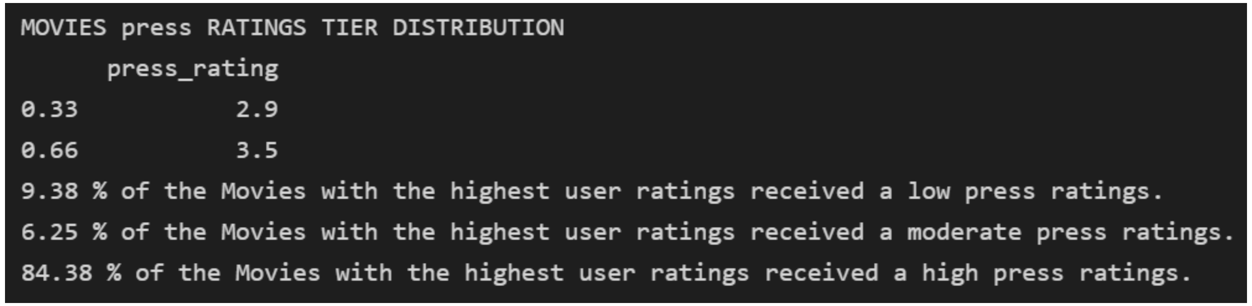
Data Analysis - Analyse Exploratoire 1

Analyses des données

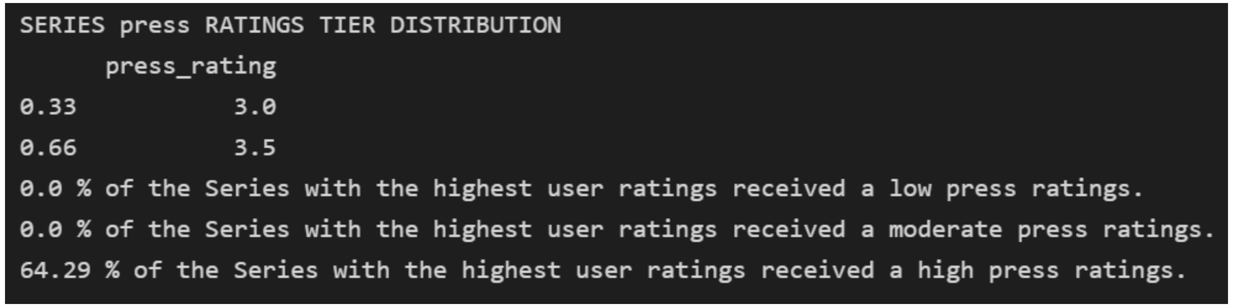
Compare press to user ratings







```
MOVIES press RATINGS TIER DISTRIBUTION
      press_rating
0.33           2.9
0.66           3.5
9.38 % of the Movies with the highest user ratings received a low press ratings.
6.25 % of the Movies with the highest user ratings received a moderate press ratings.
84.38 % of the Movies with the highest user ratings received a high press ratings.
```



```
SERIES press RATINGS TIER DISTRIBUTION
      press_rating
0.33           3.0
0.66           3.5
0.0 % of the Series with the highest user ratings received a low press ratings.
0.0 % of the Series with the highest user ratings received a moderate press ratings.
64.29 % of the Series with the highest user ratings received a high press ratings.
```

DATA, DIGITAL & TECHNOLOGY