

# Movie\_First\_View

June 21, 2024

## 0.0.1 Import Librairies

```
[2]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

### Installer les packages

```
[ ]: #utilisation de !pip install package_name pour installer si celui ci n'est pas
    ↪ encore installed
!pip install matplotlib

!pip install pandas

!pip install numpy

#etc
```

## 0.0.2 Connect to Dataset

```
[4]: df = pd.read_csv("C:/Users/JEHDAI/Desktop/Python Practice/Project/Movies First_
    ↪ View/Dataset/movie.csv")
```

# 1 Nettoyage Complet des datas (Data Preprocessing and Manipulation)

### Structure des datas

```
[5]: #structure type datas
df.info()

#columns
df.columns

#rows, columns
df.shape
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4916 entries, 0 to 4915
Data columns (total 28 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   color                                4897 non-null   object
1   director_name                       4814 non-null   object
2   num_critic_for_reviews              4867 non-null   float64
3   duration                            4901 non-null   float64
4   director_facebook_likes             4814 non-null   float64
5   actor_3_facebook_likes             4893 non-null   float64
6   actor_2_name                       4903 non-null   object
7   actor_1_facebook_likes             4909 non-null   float64
8   gross                              4054 non-null   float64
9   genres                              4916 non-null   object
10  actor_1_name                       4909 non-null   object
11  movie_title                        4916 non-null   object
12  num_voted_users                   4916 non-null   int64
13  cast_total_facebook_likes         4916 non-null   int64
14  actor_3_name                     4893 non-null   object
15  facenumber_in_poster             4903 non-null   float64
16  plot_keywords                    4764 non-null   object
17  movie_imdb_link                  4916 non-null   object
18  num_user_for_reviews             4895 non-null   float64
19  language                        4904 non-null   object
20  country                         4911 non-null   object
21  content_rating                   4616 non-null   object
22  budget                          4432 non-null   float64
23  title_year                      4810 non-null   float64
24  actor_2_facebook_likes          4903 non-null   float64
25  imdb_score                      4916 non-null   float64
26  aspect_ratio                    4590 non-null   float64
27  movie_facebook_likes            4916 non-null   int64
dtypes: float64(13), int64(3), object(12)
memory usage: 1.1+ MB

```

[5]: (4916, 28)

### Valeurs manquantes

```

[6]: # check missing values
df.isnull().sum()

```

```

[6]: color                                19
     director_name                      102
     num_critic_for_reviews              49
     duration                            15
     director_facebook_likes            102

```

actor_3_facebook_likes	23
actor_2_name	13
actor_1_facebook_likes	7
gross	862
genres	0
actor_1_name	7
movie_title	0
num_voted_users	0
cast_total_facebook_likes	0
actor_3_name	23
facenumber_in_poster	13
plot_keywords	152
movie_imdb_link	0
num_user_for_reviews	21
language	12
country	5
content_rating	300
budget	484
title_year	106
actor_2_facebook_likes	13
imdb_score	0
aspect_ratio	326
movie_facebook_likes	0
dtype:	int64

[5]: *#Supprimer les valeurs (rows des valeurs nulles)*

```
df.dropna(inplace = True)

#check
df.isnull().sum()
```

color	0
director_name	0
num_critic_for_reviews	0
duration	0
director_facebook_likes	0
actor_3_facebook_likes	0
actor_2_name	0
actor_1_facebook_likes	0
gross	0
genres	0
actor_1_name	0
movie_title	0
num_voted_users	0
cast_total_facebook_likes	0
actor_3_name	0

```

facenumber_in_poster      0
plot_keywords              0
movie_imdb_link           0
num_user_for_reviews      0
language                  0
country                   0
content_rating            0
budget                    0
title_year                0
actor_2_facebook_likes    0
imdb_score                0
aspect_ratio              0
movie_facebook_likes      0
dtype: int64

```

**Remplacement des valeurs manquantes par la moyenne des valeurs de la colonne**  
`movies['score'].fillna(movies['score'].mean(), inplace = True)`

```

[9]: #look at datas
df.head(10)

df.tail(10)

```

```

[9]:
      color      director_name  num_critic_for_reviews  duration \
4872    Color      Kevin Jordan                21.0      90.0
4882  Black and White      Kevin Smith               136.0     102.0
4885    Color      Neil LaBute                80.0      97.0
4888  Black and White  Richard Linklater                61.0     100.0
4898    Color      John Waters                73.0     108.0
4899    Color  Olivier Assayas                81.0     110.0
4900    Color      Jafar Panahi                64.0      90.0
4906    Color      Shane Carruth               143.0      77.0
4908    Color  Robert Rodriguez                56.0      81.0
4915    Color      Jon Gunn                  43.0      90.0

      director_facebook_likes  actor_3_facebook_likes      actor_2_name \
4872                    4.0                113.0      Christa Miller
4882                    0.0                216.0    Brian O'Halloran
4885                   119.0                  7.0        Matt Malloy
4888                    0.0                  0.0    Richard Linklater
4898                    0.0               105.0        Mink Stole
4899                   107.0                 45.0    Béatrice Dalle
4900                   397.0                  0.0    Nargess Mamizadeh
4906                   291.0                  8.0    David Sullivan
4908                    0.0                  6.0    Peter Marquardt
4915                   16.0                 16.0    Brian Herzlinger

```

	actor_1_facebook_likes	gross	genres \
4872	20000.0	277233.0	Comedy Romance
4882	898.0	3151130.0	Comedy
4885	136.0	2856622.0	Comedy Drama
4888	5.0	1227508.0	Comedy Drama
4898	462.0	180483.0	Comedy Crime Horror
4899	576.0	136007.0	Drama Music Romance
4900	5.0	673780.0	Drama
4906	291.0	424760.0	Drama Sci-Fi Thriller
4908	121.0	2040920.0	Action Crime Drama Romance Thriller
4915	86.0	85222.0	Documentary

	... num_user_for_reviews	language	country	content_rating	budget \
4872	...	26.0	English	USA	R 40000.0
4882	...	615.0	English	USA	R 230000.0
4885	...	197.0	English	Canada	R 25000.0
4888	...	80.0	English	USA	R 23000.0
4898	...	183.0	English	USA	NC-17 10000.0
4899	...	39.0	French	France	R 4500.0
4900	...	26.0	Persian	Iran	Not Rated 10000.0
4906	...	371.0	English	USA	PG-13 7000.0
4908	...	130.0	Spanish	USA	R 7000.0
4915	...	84.0	English	USA	PG 1100.0

	title_year	actor_2_facebook_likes	imdb_score	aspect_ratio \
4872	1999.0	467.0	7.6	1.85
4882	1994.0	657.0	7.8	1.37
4885	1997.0	108.0	7.3	1.85
4888	1991.0	0.0	7.1	1.37
4898	1972.0	143.0	6.1	1.37
4899	2004.0	133.0	6.9	2.35
4900	2000.0	0.0	7.5	1.85
4906	2004.0	45.0	7.0	1.85
4908	1992.0	20.0	6.9	1.37
4915	2004.0	23.0	6.6	1.85

	movie_facebook_likes
4872	0
4882	0
4885	489
4888	2000
4898	0
4899	171
4900	697
4906	19000
4908	0
4915	456

```
[10 rows x 28 columns]
```

### Etat des colonnes et variables

```
[6]: df.columns
```

```
[6]: Index(['color', 'director_name', 'num_critic_for_reviews', 'duration',  
        'director_facebook_likes', 'actor_3_facebook_likes', 'actor_2_name',  
        'actor_1_facebook_likes', 'gross', 'genres', 'actor_1_name',  
        'movie_title', 'num_voted_users', 'cast_total_facebook_likes',  
        'actor_3_name', 'facenumber_in_poster', 'plot_keywords',  
        'movie_imdb_link', 'num_user_for_reviews', 'language', 'country',  
        'content_rating', 'budget', 'title_year', 'actor_2_facebook_likes',  
        'imdb_score', 'aspect_ratio', 'movie_facebook_likes'],  
        dtype='object')
```

#### 1.0.1 Changer le type des colonnes

1. Changer le type de la variable en int `df['score'] = df['score'].astype('int')`

2. Changer le type de la variable en date `df['date_production'] = pd.to_datetime(df['date_production'])`

3. Changer le type de la variable en string `df['org_id'] = df['org_id'].astype('str')`

```
[ ]:
```

```
[11]: #Vvoir la repartition des films selonm la variable "language"
```

```
df.language.value_counts(normalize = True) * 100
```

```
#ou df['language'].value_counts(normalize = True) *100
```

```
[11]: English      95.731874  
      French      0.930233  
      Spanish     0.629275  
      Mandarin    0.383037  
      German      0.273598  
      Japanese    0.273598  
      Cantonese   0.191518  
      Italian     0.191518  
      Portuguese  0.136799  
      Hindi       0.136799  
      Norwegian  0.109439  
      Korean      0.109439  
      Danish      0.082079  
      Thai        0.082079  
      Persian     0.082079
```

Dutch	0.082079
Indonesian	0.054720
Aboriginal	0.054720
Dari	0.054720
Mongolian	0.027360
None	0.027360
Romanian	0.027360
Vietnamese	0.027360
Arabic	0.027360
Hebrew	0.027360
Zulu	0.027360
Czech	0.027360
Russian	0.027360
Filipino	0.027360
Maya	0.027360
Kazakh	0.027360
Hungarian	0.027360
Aramaic	0.027360
Bosnian	0.027360

Name: language, dtype: float64

```
[7]: #choisir les variables importantes pour notre analyse (colonnes importantes)
movies = df[['director_name', 'duration', 'gross', 'genres', 'movie_title',
            'plot_keywords', 'budget', 'title_year', 'language', 'imdb_score']]

#look on our new dataset
movies.head()
```

```
[7]:
```

	director_name	duration	gross	genres \
0	James Cameron	178.0	760505847.0	Action Adventure Fantasy Sci-Fi
1	Gore Verbinski	169.0	309404152.0	Action Adventure Fantasy
2	Sam Mendes	148.0	200074175.0	Action Adventure Thriller
3	Christopher Nolan	164.0	448130642.0	Action Thriller
5	Andrew Stanton	132.0	73058679.0	Action Adventure Sci-Fi

	movie_title \
0	Avatar
1	Pirates of the Caribbean: At World's End
2	Spectre
3	The Dark Knight Rises
5	John Carter

	plot_keywords	budget	title_year \
0	avatar future marine native paraplegic	237000000.0	2009.0
1	goddess marriage ceremony marriage proposal pi...	300000000.0	2007.0
2	bomb espionage sequel spy terrorist	245000000.0	2015.0
3	deception imprisonment lawlessness police offi...	250000000.0	2012.0

```
5 alien|american civil war|male nipple|mars|prin... 263700000.0 2012.0
```

```
language imdb_score
0 English 7.9
1 English 7.1
2 English 6.8
3 English 8.5
5 English 6.6
```

```
[8]: #Arrangement des variables ou colonnes, renommer certaines colonnes
movies.rename(columns = {'title_year':'year', 'imdb_score':'score',
↳ 'movie_title':'title'}, inplace = True)

#check
movies.head()
```

```
[8]:      director_name  duration      gross      genres \
0      James Cameron    178.0  760505847.0 Action|Adventure|Fantasy|Sci-Fi
1      Gore Verbinski    169.0  309404152.0      Action|Adventure|Fantasy
2      Sam Mendes      148.0  200074175.0      Action|Adventure|Thriller
3 Christopher Nolan    164.0  448130642.0      Action|Thriller
5      Andrew Stanton    132.0   73058679.0      Action|Adventure|Sci-Fi
```

```
      title \
0      Avatar
1 Pirates of the Caribbean: At World's End
2      Spectre
3      The Dark Knight Rises
5      John Carter
```

```
      plot_keywords      budget      year \
0      avatar|future|marine|native|paraplegic  237000000.0  2009.0
1 goddess|marriage ceremony|marriage proposal|pi...  300000000.0  2007.0
2      bomb|espionage|sequel|spy|terrorist  245000000.0  2015.0
3 deception|imprisonment|lawlessness|police offi...  250000000.0  2012.0
5 alien|american civil war|male nipple|mars|prin...  263700000.0  2012.0
```

```
language score
0 English 7.9
1 English 7.1
2 English 6.8
3 English 8.5
5 English 6.6
```

```
[9]: #Supprimer une autre colonne pas trop importante
movies.drop(['plot_keywords'], axis = 1, inplace = True)
```



```
#check
movies.head()
```

```
[9]:
```

	director_name	duration	gross	genres \
0	James Cameron	178.0	760505847.0	Action Adventure Fantasy Sci-Fi
1	Gore Verbinski	169.0	309404152.0	Action Adventure Fantasy
2	Sam Mendes	148.0	200074175.0	Action Adventure Thriller
3	Christopher Nolan	164.0	448130642.0	Action Thriller
5	Andrew Stanton	132.0	73058679.0	Action Adventure Sci-Fi

	title	budget	year	language \
0	Avatar	237000000.0	2009.0	English
1	Pirates of the Caribbean: At World's End	300000000.0	2007.0	English
2	Spectre	245000000.0	2015.0	English
3	The Dark Knight Rises	250000000.0	2012.0	English
5	John Carter	263700000.0	2012.0	English

	score
0	7.9
1	7.1
2	6.8
3	8.5
5	6.6

```
[18]: #Segmentation des données
#decouper les films du producteur Steven Spiel

is_Steven = movies[movies['director_name'] == 'Steven Spielberg']

is_Steven
```

```
[18]:
```

	director_name	duration	gross \
54	Steven Spielberg	122.0	317011114.0
177	Steven Spielberg	117.0	52792307.0
187	Steven Spielberg	116.0	234277056.0
192	Steven Spielberg	107.0	77564037.0
281	Steven Spielberg	145.0	132014112.0
369	Steven Spielberg	146.0	78616689.0
517	Steven Spielberg	129.0	229074524.0
530	Steven Spielberg	128.0	77032279.0
539	Steven Spielberg	163.0	47379090.0
583	Steven Spielberg	142.0	119654900.0
597	Steven Spielberg	146.0	79883359.0
641	Steven Spielberg	169.0	216119491.0
646	Steven Spielberg	150.0	182204440.0
689	Steven Spielberg	127.0	356784000.0
902	Steven Spielberg	141.0	164435221.0

1030	Steven Spielberg	127.0	197171806.0
1213	Steven Spielberg	142.0	72306065.0
1237	Steven Spielberg	155.0	44175394.0
1733	Steven Spielberg	118.0	179870271.0
1857	Steven Spielberg	185.0	96067179.0
2128	Steven Spielberg	115.0	242374454.0
2130	Steven Spielberg	135.0	128300000.0
2575	Steven Spielberg	154.0	94175854.0
2870	Steven Spielberg	130.0	260000000.0
3028	Steven Spielberg	120.0	434949459.0

	genres \
54	Action Adventure Fantasy
177	Adventure Family Fantasy
187	Adventure Sci-Fi Thriller
192	Action Adventure Family Mystery
281	Action Mystery Sci-Fi Thriller
369	Adventure Drama Sci-Fi
517	Action Adventure Sci-Fi
530	Comedy Drama
539	Drama History Thriller
583	Adventure Comedy Family Fantasy
597	Drama War
641	Action Drama War
646	Biography Drama History War
689	Adventure Sci-Fi Thriller
902	Biography Crime Drama
1030	Action Adventure Fantasy
1213	Drama History Thriller
1237	Drama History
1733	Action Adventure
1857	Biography Drama History
2128	Action Adventure
2130	Drama Sci-Fi
2575	Drama
2870	Adventure Drama Thriller
3028	Family Sci-Fi

	title	budget	year \
54	Indiana Jones and the Kingdom of the Crystal S...	185000000.0	2008.0
177	The BFG	140000000.0	2016.0
187	War of the Worlds	132000000.0	2005.0
192	The Adventures of Tintin	135000000.0	2011.0
281	Minority Report	102000000.0	2002.0
369	A.I. Artificial Intelligence	100000000.0	2001.0
517	The Lost World: Jurassic Park	73000000.0	1997.0
530	The Terminal	60000000.0	2004.0

539	Munich	70000000.0	2005.0
583	Hook	70000000.0	1991.0
597	War Horse	66000000.0	2011.0
641	Saving Private Ryan	70000000.0	1998.0
646	Lincoln	65000000.0	2012.0
689	Jurassic Park	63000000.0	1993.0
902	Catch Me If You Can	52000000.0	2002.0
1030	Indiana Jones and the Last Crusade	48000000.0	1989.0
1213	Bridge of Spies	40000000.0	2015.0
1237	Amistad	36000000.0	1997.0
1733	Indiana Jones and the Temple of Doom	28000000.0	1984.0
1857	Schindler's List	22000000.0	1993.0
2128	Raiders of the Lost Ark	18000000.0	1981.0
2130	Close Encounters of the Third Kind	19400870.0	1977.0
2575	The Color Purple	15000000.0	1985.0
2870	Jaws	8000000.0	1975.0
3028	E.T. the Extra-Terrestrial	10500000.0	1982.0

	language	score
54	English	6.2
177	English	6.8
187	English	6.5
192	English	7.4
281	English	7.7
369	English	7.1
517	English	6.5
530	English	7.3
539	English	7.6
583	English	6.7
597	English	7.2
641	English	8.6
646	English	7.4
689	English	8.1
902	English	8.0
1030	English	8.3
1213	English	7.6
1237	English	7.2
1733	English	7.6
1857	English	8.9
2128	English	8.5
2130	English	7.7
2575	English	7.8
2870	English	8.0
3028	English	7.9

[21]: *#Prendre juste les films french et plus de 60 minutes...*

```
french_movie = movies[(movies['language'] == 'French') & (movies['duration'] > 60)]
```

```
#check
```

```
french_movie
```

```
[21]:
```

	director_name	duration	gross	\
498	Jacques Perrin	104.0	19406406.0	
880	Jean-Pierre Jeunet	133.0	6167817.0	
1166	Jean-Pierre Jeunet	105.0	1260917.0	
1287	Jean-Pierre Jeunet	122.0	33201661.0	
1989	Guillaume Canet	134.0	183662.0	
2050	Anne Fontaine	111.0	6109075.0	
2095	Christian Carion	116.0	1050445.0	
2101	Laurent Tirard	120.0	634277.0	
2295	Jean-Paul Rappeneau	114.0	2353728.0	
2562	Gérard Krawczyk	94.0	81525.0	
2695	Pierre Morel	84.0	1197786.0	
2718	Gilles Paquet-Brenner	111.0	7691700.0	
2830	Olivier Assayas	120.0	4857376.0	
2833	Patrice Leconte	112.0	3058380.0	
2872	Stéphane Aubier	80.0	71442.0	
3122	François Girard	130.0	9473382.0	
3207	Michael Haneke	127.0	225377.0	
3338	François Ozon	111.0	3076425.0	
3393	Vincent Paronnaud	89.0	4443403.0	
3455	Jean-Marie Poiré	107.0	700000.0	
3483	Denis Villeneuve	139.0	6857096.0	
3584	Cédric Klapisch	111.0	3895664.0	
3608	Christophe Barratier	97.0	3629758.0	
3613	Aki Kaurismäki	93.0	611709.0	
3688	Denys Arcand	112.0	3432342.0	
3701	Xavier Beauvois	122.0	3950029.0	
3766	Michel Leclerc	100.0	513836.0	
3942	Luc Jacquet	80.0	77413017.0	
4065	Alexandre Aja	89.0	3645438.0	
4172	Sylvain Chomet	80.0	7002255.0	
4256	Pascal Arnold	95.0	18435.0	
4303	François Truffaut	123.0	26893.0	
4713	Cédric Klapisch	91.0	1652472.0	
4899	Olivier Assayas	110.0	136007.0	

	genres	title	\
498	Documentary Drama	Oceans	
880	Drama Mystery Romance War	A Very Long Engagement	
1166	Action Comedy Crime	Micmacs	
1287	Comedy Romance	Amélie	

1989	Comedy Drama	Little White Lies
2050	Biography Drama	Coco Before Chanel
2095	Drama History Music Romance War	Joyeux Noel
2101	Comedy History	Molière
2295	Comedy Drama Mystery Romance Thriller War	Bon voyage
2562	Action Comedy Crime Drama Thriller	Wasabi
2695	Action Crime Thriller	District B13
2718	Drama War	Sarah's Key
2830	Comedy Drama Romance	Paris, je t'aime
2833	Drama Romance	The Widow of Saint-Pierre
2872	Animation Comedy Crime Drama Family	Ernest & Celestine
3122	Drama Music Mystery Romance	The Red Violin
3207	Drama Romance	Amour
3338	Comedy Crime Musical Romance	8 Women
3393	Animation Biography Drama War	Persepolis
3455	Comedy Fantasy Sci-Fi	Les visiteurs
3483	Drama Mystery War	Incendies
3584	Comedy Drama	L'auberge espagnole
3608	Drama Music	The Chorus
3613	Comedy Drama	Le Havre
3688	Comedy Crime Drama Mystery Romance	The Barbarian Invasions
3701	Drama	Of Gods and Men
3766	Comedy Drama Romance	The Names of Love
3942	Documentary	March of the Penguins
4065	Horror	High Tension
4172	Animation Comedy Drama	The Triplets of Belleville
4256	Drama	One to Another
4303	Crime Drama Romance	Mississippi Mermaid
4713	Comedy Romance	When the Cat's Away
4899	Drama Music Romance	Clean

	budget	year	language	score
498	40000000.0	2009.0	French	7.8
880	47000000.0	2004.0	French	7.7
1166	27000000.0	2009.0	French	7.2
1287	77000000.0	2001.0	French	8.4
1989	25000000.0	2010.0	French	7.1
2050	19430000.0	2009.0	French	6.7
2095	22000000.0	2005.0	French	7.8
2101	16000000.0	2007.0	French	7.3
2295	20000000.0	2003.0	French	6.9
2562	15300000.0	2001.0	French	6.6
2695	12000000.0	2004.0	French	7.2
2718	10000000.0	2010.0	French	7.5
2830	13000000.0	2006.0	French	7.3
2833	100000000.0	2000.0	French	7.3
2872	9600000.0	2012.0	French	7.9

3122	10000000.0	1998.0	French	7.7
3207	8900000.0	2012.0	French	7.9
3338	8000000.0	2002.0	French	7.1
3393	7300000.0	2007.0	French	8.0
3455	50000000.0	1993.0	French	7.0
3483	6800000.0	2010.0	French	8.2
3584	5300000.0	2002.0	French	7.3
3608	5500000.0	2004.0	French	7.9
3613	3850000.0	2011.0	French	7.2
3688	6000000.0	2003.0	French	7.7
3701	4000000.0	2010.0	French	7.2
3766	4600000.0	2010.0	French	7.2
3942	8000000.0	2005.0	French	7.6
4065	2200000.0	2003.0	French	6.8
4172	9500000.0	2003.0	French	7.8
4256	1400000.0	2006.0	French	5.8
4303	1600000.0	1969.0	French	7.2
4713	300000.0	1996.0	French	6.9
4899	4500.0	2004.0	French	6.9

### Gestion des doublons

```
[22]: #Gestion des doublons
      #identifier les doublons en value TRUE

      movies.duplicated()
```

```
[22]: 0      False
      1      False
      2      False
      3      False
      5      False
      ...
      4899   False
      4900   False
      4906   False
      4908   False
      4915   False
      Length: 3655, dtype: bool
```

```
[10]: #suppression des doublons tout en readaptant les index
      movies.drop_duplicates().reset_index(drop = True)
```

```
[10]:      director_name  duration  gross \
0      James Cameron    178.0  760505847.0
1      Gore Verbinski    169.0  309404152.0
2          Sam Mendes    148.0  200074175.0
3  Christopher Nolan    164.0  448130642.0
```

4	Andrew Stanton	132.0	73058679.0
...	...	...	...
3650	Olivier Assayas	110.0	136007.0
3651	Jafar Panahi	90.0	673780.0
3652	Shane Carruth	77.0	424760.0
3653	Robert Rodriguez	81.0	2040920.0
3654	Jon Gunn	90.0	85222.0

	genres \
0	Action Adventure Fantasy Sci-Fi
1	Action Adventure Fantasy
2	Action Adventure Thriller
3	Action Thriller
4	Action Adventure Sci-Fi
...	...
3650	Drama Music Romance
3651	Drama
3652	Drama Sci-Fi Thriller
3653	Action Crime Drama Romance Thriller
3654	Documentary

	title	budget	year	language \
0	Avatar	237000000.0	2009.0	English
1	Pirates of the Caribbean: At World's End	300000000.0	2007.0	English
2	Spectre	245000000.0	2015.0	English
3	The Dark Knight Rises	250000000.0	2012.0	English
4	John Carter	263700000.0	2012.0	English
...	...	...	...	...
3650	Clean	4500.0	2004.0	French
3651	The Circle	10000.0	2000.0	Persian
3652	Primer	7000.0	2004.0	English
3653	El Mariachi	7000.0	1992.0	Spanish
3654	My Date with Drew	1100.0	2004.0	English

	score
0	7.9
1	7.1
2	6.8
3	8.5
4	6.6
...	...
3650	6.9
3651	7.5
3652	7.0
3653	6.9
3654	6.6

[3655 rows x 9 columns]

```
[24]: #recheck
movies.duplicated()
```

```
[24]: 0      False
      1      False
      2      False
      3      False
      5      False
      ...
      4899   False
      4900   False
      4906   False
      4908   False
      4915   False
      Length: 3655, dtype: bool
```

### Remplacement des valeurs dans les colonnes

```
[11]: #Remplacer une valeur dans une colonne

movies['language'].replace(to_replace = 'French', value = 'Paris', inplace =_
↪True)
```

```
[40]: #check

movies.head(100)

movies['language'].unique()
```

```
[40]: array(['English', 'Mandarin', 'Aboriginal', 'Spanish', 'Paris',
        'Filipino', 'Maya', 'Kazakh', 'Cantonese', 'Japanese', 'Aramaic',
        'Italian', 'Dutch', 'Dari', 'German', 'Mongolian', 'Thai',
        'Bosnian', 'Korean', 'Hungarian', 'Hindi', 'Danish', 'Portuguese',
        'Norwegian', 'Czech', 'Russian', 'None', 'Zulu', 'Hebrew',
        'Arabic', 'Vietnamese', 'Indonesian', 'Romanian', 'Persian'],
      dtype=object)
```

```
[12]: #Remplacer les valeurs English par Eng
movies['language'].replace(to_replace = 'English', value = 'Eng', inplace =_
↪True)
```

```
[42]: #check my new datas

movies.head(50)
```



[42]:	director_name	duration	gross	\
0	James Cameron	178.0	760505847.0	
1	Gore Verbinski	169.0	309404152.0	
2	Sam Mendes	148.0	200074175.0	
3	Christopher Nolan	164.0	448130642.0	
5	Andrew Stanton	132.0	73058679.0	
6	Sam Raimi	156.0	336530303.0	
7	Nathan Greno	100.0	200807262.0	
8	Joss Whedon	141.0	458991599.0	
9	David Yates	153.0	301956980.0	
10	Zack Snyder	183.0	330249062.0	
11	Bryan Singer	169.0	200069408.0	
12	Marc Forster	106.0	168368427.0	
13	Gore Verbinski	151.0	423032628.0	
14	Gore Verbinski	150.0	89289910.0	
15	Zack Snyder	143.0	291021565.0	
16	Andrew Adamson	150.0	141614023.0	
17	Joss Whedon	173.0	623279547.0	
18	Rob Marshall	136.0	241063875.0	
19	Barry Sonnenfeld	106.0	179020854.0	
20	Peter Jackson	164.0	255108370.0	
21	Marc Webb	153.0	262030663.0	
22	Ridley Scott	156.0	105219735.0	
23	Peter Jackson	186.0	258355354.0	
24	Chris Weitz	113.0	70083519.0	
25	Peter Jackson	201.0	218051260.0	
26	James Cameron	194.0	658672302.0	
27	Anthony Russo	147.0	407197282.0	
28	Peter Berg	131.0	65173160.0	
29	Colin Trevorrow	124.0	652177271.0	
30	Sam Mendes	143.0	304360277.0	
31	Sam Raimi	135.0	373377893.0	
32	Shane Black	195.0	408992272.0	
33	Tim Burton	108.0	334185206.0	
34	Brett Ratner	104.0	234360014.0	
35	Dan Scanlon	104.0	268488329.0	
36	Michael Bay	150.0	402076689.0	
37	Michael Bay	165.0	245428137.0	
38	Sam Raimi	130.0	234903076.0	
39	Marc Webb	142.0	202853933.0	
40	Joseph Kosinski	125.0	172051787.0	
41	John Lasseter	106.0	191450875.0	
42	Martin Campbell	123.0	116593191.0	
43	Lee Unkrich	103.0	414984497.0	
44	McG	118.0	125320003.0	
45	James Wan	140.0	350034110.0	
46	Marc Forster	123.0	202351611.0	

47	Bryan Singer	149.0	233914986.0
48	J.J. Abrams	132.0	228756232.0
49	Bryan Singer	114.0	65171860.0
50	Baz Luhrmann	143.0	144812796.0

	genres \
0	Action Adventure Fantasy Sci-Fi
1	Action Adventure Fantasy
2	Action Adventure Thriller
3	Action Thriller
5	Action Adventure Sci-Fi
6	Action Adventure Romance
7	Adventure Animation Comedy Family Fantasy Musi...
8	Action Adventure Sci-Fi
9	Adventure Family Fantasy Mystery
10	Action Adventure Sci-Fi
11	Action Adventure Sci-Fi
12	Action Adventure
13	Action Adventure Fantasy
14	Action Adventure Western
15	Action Adventure Fantasy Sci-Fi
16	Action Adventure Family Fantasy
17	Action Adventure Sci-Fi
18	Action Adventure Fantasy
19	Action Adventure Comedy Family Fantasy Sci-Fi
20	Adventure Fantasy
21	Action Adventure Fantasy
22	Action Adventure Drama History
23	Adventure Fantasy
24	Adventure Family Fantasy
25	Action Adventure Drama Romance
26	Drama Romance
27	Action Adventure Sci-Fi
28	Action Adventure Sci-Fi Thriller
29	Action Adventure Sci-Fi Thriller
30	Action Adventure Thriller
31	Action Adventure Fantasy Romance
32	Action Adventure Sci-Fi
33	Adventure Family Fantasy
34	Action Adventure Fantasy Sci-Fi Thriller
35	Adventure Animation Comedy Family Fantasy
36	Action Adventure Sci-Fi
37	Action Adventure Sci-Fi
38	Adventure Family Fantasy
39	Action Adventure Fantasy Sci-Fi
40	Action Adventure Sci-Fi
41	Adventure Animation Comedy Family Sport

42 Action|Adventure|Sci-Fi  
 43 Adventure|Animation|Comedy|Family|Fantasy  
 44 Action|Adventure|Sci-Fi  
 45 Action|Crime|Thriller  
 46 Action|Adventure|Horror|Sci-Fi|Thriller  
 47 Action|Adventure|Fantasy|Sci-Fi|Thriller  
 48 Action|Adventure|Sci-Fi  
 49 Adventure|Fantasy  
 50 Drama|Romance

	title	budget	year	language	\
0	Avatar	237000000.0	2009.0	Eng	
1	Pirates of the Caribbean: At World's End	300000000.0	2007.0	Eng	
2	Spectre	245000000.0	2015.0	Eng	
3	The Dark Knight Rises	250000000.0	2012.0	Eng	
5	John Carter	263700000.0	2012.0	Eng	
6	Spider-Man 3	258000000.0	2007.0	Eng	
7	Tangled	260000000.0	2010.0	Eng	
8	Avengers: Age of Ultron	250000000.0	2015.0	Eng	
9	Harry Potter and the Half-Blood Prince	250000000.0	2009.0	Eng	
10	Batman v Superman: Dawn of Justice	250000000.0	2016.0	Eng	
11	Superman Returns	209000000.0	2006.0	Eng	
12	Quantum of Solace	200000000.0	2008.0	Eng	
13	Pirates of the Caribbean: Dead Man's Chest	225000000.0	2006.0	Eng	
14	The Lone Ranger	215000000.0	2013.0	Eng	
15	Man of Steel	225000000.0	2013.0	Eng	
16	The Chronicles of Narnia: Prince Caspian	225000000.0	2008.0	Eng	
17	The Avengers	220000000.0	2012.0	Eng	
18	Pirates of the Caribbean: On Stranger Tides	250000000.0	2011.0	Eng	
19	Men in Black 3	225000000.0	2012.0	Eng	
20	The Hobbit: The Battle of the Five Armies	250000000.0	2014.0	Eng	
21	The Amazing Spider-Man	230000000.0	2012.0	Eng	
22	Robin Hood	200000000.0	2010.0	Eng	
23	The Hobbit: The Desolation of Smaug	225000000.0	2013.0	Eng	
24	The Golden Compass	180000000.0	2007.0	Eng	
25	King Kong	207000000.0	2005.0	Eng	
26	Titanic	200000000.0	1997.0	Eng	
27	Captain America: Civil War	250000000.0	2016.0	Eng	
28	Battleship	209000000.0	2012.0	Eng	
29	Jurassic World	150000000.0	2015.0	Eng	
30	Skyfall	200000000.0	2012.0	Eng	
31	Spider-Man 2	200000000.0	2004.0	Eng	
32	Iron Man 3	200000000.0	2013.0	Eng	
33	Alice in Wonderland	200000000.0	2010.0	Eng	
34	X-Men: The Last Stand	210000000.0	2006.0	Eng	
35	Monsters University	200000000.0	2013.0	Eng	
36	Transformers: Revenge of the Fallen	200000000.0	2009.0	Eng	

37	Transformers: Age of Extinction	210000000.0	2014.0	Eng
38	Oz the Great and Powerful	215000000.0	2013.0	Eng
39	The Amazing Spider-Man 2	200000000.0	2014.0	Eng
40	TRON: Legacy	170000000.0	2010.0	Eng
41	Cars 2	200000000.0	2011.0	Eng
42	Green Lantern	200000000.0	2011.0	Eng
43	Toy Story 3	200000000.0	2010.0	Eng
44	Terminator Salvation	200000000.0	2009.0	Eng
45	Furious 7	190000000.0	2015.0	Eng
46	World War Z	190000000.0	2013.0	Eng
47	X-Men: Days of Future Past	200000000.0	2014.0	Eng
48	Star Trek Into Darkness	190000000.0	2013.0	Eng
49	Jack the Giant Slayer	195000000.0	2013.0	Eng
50	The Great Gatsby	105000000.0	2013.0	Eng

	score
0	7.9
1	7.1
2	6.8
3	8.5
5	6.6
6	6.2
7	7.8
8	7.5
9	7.5
10	6.9
11	6.1
12	6.7
13	7.3
14	6.5
15	7.2
16	6.6
17	8.1
18	6.7
19	6.8
20	7.5
21	7.0
22	6.7
23	7.9
24	6.1
25	7.2
26	7.7
27	8.2
28	5.9
29	7.0
30	7.8
31	7.3

```

32    7.2
33    6.5
34    6.8
35    7.3
36    6.0
37    5.7
38    6.4
39    6.7
40    6.8
41    6.3
42    5.6
43    8.3
44    6.6
45    7.2
46    7.0
47    8.0
48    7.8
49    6.3
50    7.3

```

### Traitement des textes de colonnes

```

[13]: #mettre en majuscule les valeurs d'une colonne
      movies['director_name'] = movies['director_name'].str.upper()

      #check
      movies.head(50)

```

```

[13]:      director_name  duration      gross \
0      JAMES CAMERON      178.0  760505847.0
1      GORE VERBINSKI      169.0  309404152.0
2          SAM MENDES      148.0  200074175.0
3  CHRISTOPHER NOLAN      164.0  448130642.0
5      ANDREW STANTON      132.0   73058679.0
6          SAM RAIMI      156.0  336530303.0
7      NATHAN GRENO      100.0  200807262.0
8      JOSS WHEDON      141.0  458991599.0
9      DAVID YATES      153.0  301956980.0
10      ZACK SNYDER      183.0  330249062.0
11      BRYAN SINGER      169.0  200069408.0
12      MARC FORSTER      106.0  168368427.0
13      GORE VERBINSKI      151.0  423032628.0
14      GORE VERBINSKI      150.0   89289910.0
15      ZACK SNYDER      143.0  291021565.0
16      ANDREW ADAMSON      150.0  141614023.0
17      JOSS WHEDON      173.0  623279547.0
18      ROB MARSHALL      136.0  241063875.0
19      BARRY SONNENFELD      106.0  179020854.0

```

20	PETER JACKSON	164.0	255108370.0
21	MARC WEBB	153.0	262030663.0
22	RIDLEY SCOTT	156.0	105219735.0
23	PETER JACKSON	186.0	258355354.0
24	CHRIS WEITZ	113.0	70083519.0
25	PETER JACKSON	201.0	218051260.0
26	JAMES CAMERON	194.0	658672302.0
27	ANTHONY RUSSO	147.0	407197282.0
28	PETER BERG	131.0	65173160.0
29	COLIN TREVORROW	124.0	652177271.0
30	SAM MENDES	143.0	304360277.0
31	SAM RAIMI	135.0	373377893.0
32	SHANE BLACK	195.0	408992272.0
33	TIM BURTON	108.0	334185206.0
34	BRETT RATNER	104.0	234360014.0
35	DAN SCANLON	104.0	268488329.0
36	MICHAEL BAY	150.0	402076689.0
37	MICHAEL BAY	165.0	245428137.0
38	SAM RAIMI	130.0	234903076.0
39	MARC WEBB	142.0	202853933.0
40	JOSEPH KOSINSKI	125.0	172051787.0
41	JOHN LASSETER	106.0	191450875.0
42	MARTIN CAMPBELL	123.0	116593191.0
43	LEE UNKRICH	103.0	414984497.0
44	MCG	118.0	125320003.0
45	JAMES WAN	140.0	350034110.0
46	MARC FORSTER	123.0	202351611.0
47	BRYAN SINGER	149.0	233914986.0
48	J.J. ABRAMS	132.0	228756232.0
49	BRYAN SINGER	114.0	65171860.0
50	BAZ LUHRMANN	143.0	144812796.0

	genres \
0	Action Adventure Fantasy Sci-Fi
1	Action Adventure Fantasy
2	Action Adventure Thriller
3	Action Thriller
5	Action Adventure Sci-Fi
6	Action Adventure Romance
7	Adventure Animation Comedy Family Fantasy Musi...
8	Action Adventure Sci-Fi
9	Adventure Family Fantasy Mystery
10	Action Adventure Sci-Fi
11	Action Adventure Sci-Fi
12	Action Adventure
13	Action Adventure Fantasy
14	Action Adventure Western

15	Action Adventure Fantasy Sci-Fi
16	Action Adventure Family Fantasy
17	Action Adventure Sci-Fi
18	Action Adventure Fantasy
19	Action Adventure Comedy Family Fantasy Sci-Fi
20	Adventure Fantasy
21	Action Adventure Fantasy
22	Action Adventure Drama History
23	Adventure Fantasy
24	Adventure Family Fantasy
25	Action Adventure Drama Romance
26	Drama Romance
27	Action Adventure Sci-Fi
28	Action Adventure Sci-Fi Thriller
29	Action Adventure Sci-Fi Thriller
30	Action Adventure Thriller
31	Action Adventure Fantasy Romance
32	Action Adventure Sci-Fi
33	Adventure Family Fantasy
34	Action Adventure Fantasy Sci-Fi Thriller
35	Adventure Animation Comedy Family Fantasy
36	Action Adventure Sci-Fi
37	Action Adventure Sci-Fi
38	Adventure Family Fantasy
39	Action Adventure Fantasy Sci-Fi
40	Action Adventure Sci-Fi
41	Adventure Animation Comedy Family Sport
42	Action Adventure Sci-Fi
43	Adventure Animation Comedy Family Fantasy
44	Action Adventure Sci-Fi
45	Action Crime Thriller
46	Action Adventure Horror Sci-Fi Thriller
47	Action Adventure Fantasy Sci-Fi Thriller
48	Action Adventure Sci-Fi
49	Adventure Fantasy
50	Drama Romance

	title	budget	year	language	\
0	Avatar	237000000.0	2009.0	Eng	
1	Pirates of the Caribbean: At World's End	300000000.0	2007.0	Eng	
2	Spectre	245000000.0	2015.0	Eng	
3	The Dark Knight Rises	250000000.0	2012.0	Eng	
5	John Carter	263700000.0	2012.0	Eng	
6	Spider-Man 3	258000000.0	2007.0	Eng	
7	Tangled	260000000.0	2010.0	Eng	
8	Avengers: Age of Ultron	250000000.0	2015.0	Eng	
9	Harry Potter and the Half-Blood Prince	250000000.0	2009.0	Eng	

10	Batman v Superman: Dawn of Justice	250000000.0	2016.0	Eng
11	Superman Returns	209000000.0	2006.0	Eng
12	Quantum of Solace	200000000.0	2008.0	Eng
13	Pirates of the Caribbean: Dead Man's Chest	225000000.0	2006.0	Eng
14	The Lone Ranger	215000000.0	2013.0	Eng
15	Man of Steel	225000000.0	2013.0	Eng
16	The Chronicles of Narnia: Prince Caspian	225000000.0	2008.0	Eng
17	The Avengers	220000000.0	2012.0	Eng
18	Pirates of the Caribbean: On Stranger Tides	250000000.0	2011.0	Eng
19	Men in Black 3	225000000.0	2012.0	Eng
20	The Hobbit: The Battle of the Five Armies	250000000.0	2014.0	Eng
21	The Amazing Spider-Man	230000000.0	2012.0	Eng
22	Robin Hood	200000000.0	2010.0	Eng
23	The Hobbit: The Desolation of Smaug	225000000.0	2013.0	Eng
24	The Golden Compass	180000000.0	2007.0	Eng
25	King Kong	207000000.0	2005.0	Eng
26	Titanic	200000000.0	1997.0	Eng
27	Captain America: Civil War	250000000.0	2016.0	Eng
28	Battleship	209000000.0	2012.0	Eng
29	Jurassic World	150000000.0	2015.0	Eng
30	Skyfall	200000000.0	2012.0	Eng
31	Spider-Man 2	200000000.0	2004.0	Eng
32	Iron Man 3	200000000.0	2013.0	Eng
33	Alice in Wonderland	200000000.0	2010.0	Eng
34	X-Men: The Last Stand	210000000.0	2006.0	Eng
35	Monsters University	200000000.0	2013.0	Eng
36	Transformers: Revenge of the Fallen	200000000.0	2009.0	Eng
37	Transformers: Age of Extinction	210000000.0	2014.0	Eng
38	Oz the Great and Powerful	215000000.0	2013.0	Eng
39	The Amazing Spider-Man 2	200000000.0	2014.0	Eng
40	TRON: Legacy	170000000.0	2010.0	Eng
41	Cars 2	200000000.0	2011.0	Eng
42	Green Lantern	200000000.0	2011.0	Eng
43	Toy Story 3	200000000.0	2010.0	Eng
44	Terminator Salvation	200000000.0	2009.0	Eng
45	Furious 7	190000000.0	2015.0	Eng
46	World War Z	190000000.0	2013.0	Eng
47	X-Men: Days of Future Past	200000000.0	2014.0	Eng
48	Star Trek Into Darkness	190000000.0	2013.0	Eng
49	Jack the Giant Slayer	195000000.0	2013.0	Eng
50	The Great Gatsby	105000000.0	2013.0	Eng

score

0	7.9
1	7.1
2	6.8
3	8.5



5	6.6
6	6.2
7	7.8
8	7.5
9	7.5
10	6.9
11	6.1
12	6.7
13	7.3
14	6.5
15	7.2
16	6.6
17	8.1
18	6.7
19	6.8
20	7.5
21	7.0
22	6.7
23	7.9
24	6.1
25	7.2
26	7.7
27	8.2
28	5.9
29	7.0
30	7.8
31	7.3
32	7.2
33	6.5
34	6.8
35	7.3
36	6.0
37	5.7
38	6.4
39	6.7
40	6.8
41	6.3
42	5.6
43	8.3
44	6.6
45	7.2
46	7.0
47	8.0
48	7.8
49	6.3
50	7.3

```
[14]: #mettre en miniscules les valeurs d'une colonne
movies['director_name'] = movies['director_name'].str.lower()

#check
movies.head(50)
```

```
[14]:
```

	director_name	duration	gross	\
0	james cameron	178.0	760505847.0	
1	gore verbinski	169.0	309404152.0	
2	sam mendes	148.0	200074175.0	
3	christopher nolan	164.0	448130642.0	
5	andrew stanton	132.0	73058679.0	
6	sam raimi	156.0	336530303.0	
7	nathan greno	100.0	200807262.0	
8	joss whedon	141.0	458991599.0	
9	david yates	153.0	301956980.0	
10	zack snyder	183.0	330249062.0	
11	bryan singer	169.0	200069408.0	
12	marc forster	106.0	168368427.0	
13	gore verbinski	151.0	423032628.0	
14	gore verbinski	150.0	89289910.0	
15	zack snyder	143.0	291021565.0	
16	andrew adamson	150.0	141614023.0	
17	joss whedon	173.0	623279547.0	
18	rob marshall	136.0	241063875.0	
19	barry sonnenfeld	106.0	179020854.0	
20	peter jackson	164.0	255108370.0	
21	marc webb	153.0	262030663.0	
22	ridley scott	156.0	105219735.0	
23	peter jackson	186.0	258355354.0	
24	chris weitz	113.0	70083519.0	
25	peter jackson	201.0	218051260.0	
26	james cameron	194.0	658672302.0	
27	anthony russo	147.0	407197282.0	
28	peter berg	131.0	65173160.0	
29	colin trevorrow	124.0	652177271.0	
30	sam mendes	143.0	304360277.0	
31	sam raimi	135.0	373377893.0	
32	shane black	195.0	408992272.0	
33	tim burton	108.0	334185206.0	
34	brett ratner	104.0	234360014.0	
35	dan scanlon	104.0	268488329.0	
36	michael bay	150.0	402076689.0	
37	michael bay	165.0	245428137.0	
38	sam raimi	130.0	234903076.0	
39	marc webb	142.0	202853933.0	
40	joseph kosinski	125.0	172051787.0	

41	john lasseter	106.0	191450875.0
42	martin campbell	123.0	116593191.0
43	lee unkrich	103.0	414984497.0
44	mcg	118.0	125320003.0
45	james wan	140.0	350034110.0
46	marc forster	123.0	202351611.0
47	bryan singer	149.0	233914986.0
48	j.j. abrams	132.0	228756232.0
49	bryan singer	114.0	65171860.0
50	baz luhrmann	143.0	144812796.0

	genres \
0	Action Adventure Fantasy Sci-Fi
1	Action Adventure Fantasy
2	Action Adventure Thriller
3	Action Thriller
5	Action Adventure Sci-Fi
6	Action Adventure Romance
7	Adventure Animation Comedy Family Fantasy Musi...
8	Action Adventure Sci-Fi
9	Adventure Family Fantasy Mystery
10	Action Adventure Sci-Fi
11	Action Adventure Sci-Fi
12	Action Adventure
13	Action Adventure Fantasy
14	Action Adventure Western
15	Action Adventure Fantasy Sci-Fi
16	Action Adventure Family Fantasy
17	Action Adventure Sci-Fi
18	Action Adventure Fantasy
19	Action Adventure Comedy Family Fantasy Sci-Fi
20	Adventure Fantasy
21	Action Adventure Fantasy
22	Action Adventure Drama History
23	Adventure Fantasy
24	Adventure Family Fantasy
25	Action Adventure Drama Romance
26	Drama Romance
27	Action Adventure Sci-Fi
28	Action Adventure Sci-Fi Thriller
29	Action Adventure Sci-Fi Thriller
30	Action Adventure Thriller
31	Action Adventure Fantasy Romance
32	Action Adventure Sci-Fi
33	Adventure Family Fantasy
34	Action Adventure Fantasy Sci-Fi Thriller
35	Adventure Animation Comedy Family Fantasy

36	Action Adventure Sci-Fi
37	Action Adventure Sci-Fi
38	Adventure Family Fantasy
39	Action Adventure Fantasy Sci-Fi
40	Action Adventure Sci-Fi
41	Adventure Animation Comedy Family Sport
42	Action Adventure Sci-Fi
43	Adventure Animation Comedy Family Fantasy
44	Action Adventure Sci-Fi
45	Action Crime Thriller
46	Action Adventure Horror Sci-Fi Thriller
47	Action Adventure Fantasy Sci-Fi Thriller
48	Action Adventure Sci-Fi
49	Adventure Fantasy
50	Drama Romance

	title	budget	year	language	\
0	Avatar	237000000.0	2009.0	Eng	
1	Pirates of the Caribbean: At World's End	300000000.0	2007.0	Eng	
2	Spectre	245000000.0	2015.0	Eng	
3	The Dark Knight Rises	250000000.0	2012.0	Eng	
5	John Carter	263700000.0	2012.0	Eng	
6	Spider-Man 3	258000000.0	2007.0	Eng	
7	Tangled	260000000.0	2010.0	Eng	
8	Avengers: Age of Ultron	250000000.0	2015.0	Eng	
9	Harry Potter and the Half-Blood Prince	250000000.0	2009.0	Eng	
10	Batman v Superman: Dawn of Justice	250000000.0	2016.0	Eng	
11	Superman Returns	209000000.0	2006.0	Eng	
12	Quantum of Solace	200000000.0	2008.0	Eng	
13	Pirates of the Caribbean: Dead Man's Chest	225000000.0	2006.0	Eng	
14	The Lone Ranger	215000000.0	2013.0	Eng	
15	Man of Steel	225000000.0	2013.0	Eng	
16	The Chronicles of Narnia: Prince Caspian	225000000.0	2008.0	Eng	
17	The Avengers	220000000.0	2012.0	Eng	
18	Pirates of the Caribbean: On Stranger Tides	250000000.0	2011.0	Eng	
19	Men in Black 3	225000000.0	2012.0	Eng	
20	The Hobbit: The Battle of the Five Armies	250000000.0	2014.0	Eng	
21	The Amazing Spider-Man	230000000.0	2012.0	Eng	
22	Robin Hood	200000000.0	2010.0	Eng	
23	The Hobbit: The Desolation of Smaug	225000000.0	2013.0	Eng	
24	The Golden Compass	180000000.0	2007.0	Eng	
25	King Kong	207000000.0	2005.0	Eng	
26	Titanic	200000000.0	1997.0	Eng	
27	Captain America: Civil War	250000000.0	2016.0	Eng	
28	Battleship	209000000.0	2012.0	Eng	
29	Jurassic World	150000000.0	2015.0	Eng	
30	Skyfall	200000000.0	2012.0	Eng	

31	Spider-Man 2	200000000.0	2004.0	Eng
32	Iron Man 3	200000000.0	2013.0	Eng
33	Alice in Wonderland	200000000.0	2010.0	Eng
34	X-Men: The Last Stand	210000000.0	2006.0	Eng
35	Monsters University	200000000.0	2013.0	Eng
36	Transformers: Revenge of the Fallen	200000000.0	2009.0	Eng
37	Transformers: Age of Extinction	210000000.0	2014.0	Eng
38	Oz the Great and Powerful	215000000.0	2013.0	Eng
39	The Amazing Spider-Man 2	200000000.0	2014.0	Eng
40	TRON: Legacy	170000000.0	2010.0	Eng
41	Cars 2	200000000.0	2011.0	Eng
42	Green Lantern	200000000.0	2011.0	Eng
43	Toy Story 3	200000000.0	2010.0	Eng
44	Terminator Salvation	200000000.0	2009.0	Eng
45	Furious 7	190000000.0	2015.0	Eng
46	World War Z	190000000.0	2013.0	Eng
47	X-Men: Days of Future Past	200000000.0	2014.0	Eng
48	Star Trek Into Darkness	190000000.0	2013.0	Eng
49	Jack the Giant Slayer	195000000.0	2013.0	Eng
50	The Great Gatsby	105000000.0	2013.0	Eng

	score
0	7.9
1	7.1
2	6.8
3	8.5
5	6.6
6	6.2
7	7.8
8	7.5
9	7.5
10	6.9
11	6.1
12	6.7
13	7.3
14	6.5
15	7.2
16	6.6
17	8.1
18	6.7
19	6.8
20	7.5
21	7.0
22	6.7
23	7.9
24	6.1
25	7.2

```

26    7.7
27    8.2
28    5.9
29    7.0
30    7.8
31    7.3
32    7.2
33    6.5
34    6.8
35    7.3
36    6.0
37    5.7
38    6.4
39    6.7
40    6.8
41    6.3
42    5.6
43    8.3
44    6.6
45    7.2
46    7.0
47    8.0
48    7.8
49    6.3
50    7.3

```

```

[15]: #decouper des valeurs sur base d'une condition
      #decouper et creer une colonne surname en separant les deux noms grace a
      ↪ l'espace et recuperant la premiere partie (0), 1 est la deuxieme partie du
      ↪ decoupage
      movies['surname'] = movies['director_name'].str.split(" ").str.get(0)

      #check
      movies.head(50)

```

```

[15]:
      director_name  duration  gross \
0      james cameron    178.0  760505847.0
1      gore verbinski    169.0  309404152.0
2          sam mendes    148.0  200074175.0
3 christopher nolan    164.0  448130642.0
5      andrew stanton    132.0   73058679.0
6          sam raimi    156.0  336530303.0
7      nathan greno    100.0  200807262.0
8      joss whedon    141.0  458991599.0
9      david yates    153.0  301956980.0
10      zack snyder    183.0  330249062.0
11      bryan singer    169.0  200069408.0

```

12	marc forster	106.0	168368427.0
13	gore verbinski	151.0	423032628.0
14	gore verbinski	150.0	89289910.0
15	zack snyder	143.0	291021565.0
16	andrew adamson	150.0	141614023.0
17	joss whedon	173.0	623279547.0
18	rob marshall	136.0	241063875.0
19	barry sonnenfeld	106.0	179020854.0
20	peter jackson	164.0	255108370.0
21	marc webb	153.0	262030663.0
22	ridley scott	156.0	105219735.0
23	peter jackson	186.0	258355354.0
24	chris weitz	113.0	70083519.0
25	peter jackson	201.0	218051260.0
26	james cameron	194.0	658672302.0
27	anthony russo	147.0	407197282.0
28	peter berg	131.0	65173160.0
29	colin trevorrow	124.0	652177271.0
30	sam mendes	143.0	304360277.0
31	sam raimi	135.0	373377893.0
32	shane black	195.0	408992272.0
33	tim burton	108.0	334185206.0
34	brett ratner	104.0	234360014.0
35	dan scanlon	104.0	268488329.0
36	michael bay	150.0	402076689.0
37	michael bay	165.0	245428137.0
38	sam raimi	130.0	234903076.0
39	marc webb	142.0	202853933.0
40	joseph kosinski	125.0	172051787.0
41	john lasseter	106.0	191450875.0
42	martin campbell	123.0	116593191.0
43	lee unkrich	103.0	414984497.0
44	mcg	118.0	125320003.0
45	james wan	140.0	350034110.0
46	marc forster	123.0	202351611.0
47	bryan singer	149.0	233914986.0
48	j.j. abrams	132.0	228756232.0
49	bryan singer	114.0	65171860.0
50	baz luhrmann	143.0	144812796.0

	genres \
0	Action Adventure Fantasy Sci-Fi
1	Action Adventure Fantasy
2	Action Adventure Thriller
3	Action Thriller
5	Action Adventure Sci-Fi
6	Action Adventure Romance

7 Adventure|Animation|Comedy|Family|Fantasy|Musi...  
 8 Action|Adventure|Sci-Fi  
 9 Adventure|Family|Fantasy|Mystery  
 10 Action|Adventure|Sci-Fi  
 11 Action|Adventure|Sci-Fi  
 12 Action|Adventure  
 13 Action|Adventure|Fantasy  
 14 Action|Adventure|Western  
 15 Action|Adventure|Fantasy|Sci-Fi  
 16 Action|Adventure|Family|Fantasy  
 17 Action|Adventure|Sci-Fi  
 18 Action|Adventure|Fantasy  
 19 Action|Adventure|Comedy|Family|Fantasy|Sci-Fi  
 20 Adventure|Fantasy  
 21 Action|Adventure|Fantasy  
 22 Action|Adventure|Drama|History  
 23 Adventure|Fantasy  
 24 Adventure|Family|Fantasy  
 25 Action|Adventure|Drama|Romance  
 26 Drama|Romance  
 27 Action|Adventure|Sci-Fi  
 28 Action|Adventure|Sci-Fi|Thriller  
 29 Action|Adventure|Sci-Fi|Thriller  
 30 Action|Adventure|Thriller  
 31 Action|Adventure|Fantasy|Romance  
 32 Action|Adventure|Sci-Fi  
 33 Adventure|Family|Fantasy  
 34 Action|Adventure|Fantasy|Sci-Fi|Thriller  
 35 Adventure|Animation|Comedy|Family|Fantasy  
 36 Action|Adventure|Sci-Fi  
 37 Action|Adventure|Sci-Fi  
 38 Adventure|Family|Fantasy  
 39 Action|Adventure|Fantasy|Sci-Fi  
 40 Action|Adventure|Sci-Fi  
 41 Adventure|Animation|Comedy|Family|Sport  
 42 Action|Adventure|Sci-Fi  
 43 Adventure|Animation|Comedy|Family|Fantasy  
 44 Action|Adventure|Sci-Fi  
 45 Action|Crime|Thriller  
 46 Action|Adventure|Horror|Sci-Fi|Thriller  
 47 Action|Adventure|Fantasy|Sci-Fi|Thriller  
 48 Action|Adventure|Sci-Fi  
 49 Adventure|Fantasy  
 50 Drama|Romance

	title	budget	year	language	\
0	Avatar	237000000.0	2009.0	Eng	



1	Pirates of the Caribbean: At World's End	300000000.0	2007.0	Eng
2	Spectre	245000000.0	2015.0	Eng
3	The Dark Knight Rises	250000000.0	2012.0	Eng
5	John Carter	263700000.0	2012.0	Eng
6	Spider-Man 3	258000000.0	2007.0	Eng
7	Tangled	260000000.0	2010.0	Eng
8	Avengers: Age of Ultron	250000000.0	2015.0	Eng
9	Harry Potter and the Half-Blood Prince	250000000.0	2009.0	Eng
10	Batman v Superman: Dawn of Justice	250000000.0	2016.0	Eng
11	Superman Returns	209000000.0	2006.0	Eng
12	Quantum of Solace	200000000.0	2008.0	Eng
13	Pirates of the Caribbean: Dead Man's Chest	225000000.0	2006.0	Eng
14	The Lone Ranger	215000000.0	2013.0	Eng
15	Man of Steel	225000000.0	2013.0	Eng
16	The Chronicles of Narnia: Prince Caspian	225000000.0	2008.0	Eng
17	The Avengers	220000000.0	2012.0	Eng
18	Pirates of the Caribbean: On Stranger Tides	250000000.0	2011.0	Eng
19	Men in Black 3	225000000.0	2012.0	Eng
20	The Hobbit: The Battle of the Five Armies	250000000.0	2014.0	Eng
21	The Amazing Spider-Man	230000000.0	2012.0	Eng
22	Robin Hood	200000000.0	2010.0	Eng
23	The Hobbit: The Desolation of Smaug	225000000.0	2013.0	Eng
24	The Golden Compass	180000000.0	2007.0	Eng
25	King Kong	207000000.0	2005.0	Eng
26	Titanic	200000000.0	1997.0	Eng
27	Captain America: Civil War	250000000.0	2016.0	Eng
28	Battleship	209000000.0	2012.0	Eng
29	Jurassic World	150000000.0	2015.0	Eng
30	Skyfall	200000000.0	2012.0	Eng
31	Spider-Man 2	200000000.0	2004.0	Eng
32	Iron Man 3	200000000.0	2013.0	Eng
33	Alice in Wonderland	200000000.0	2010.0	Eng
34	X-Men: The Last Stand	210000000.0	2006.0	Eng
35	Monsters University	200000000.0	2013.0	Eng
36	Transformers: Revenge of the Fallen	200000000.0	2009.0	Eng
37	Transformers: Age of Extinction	210000000.0	2014.0	Eng
38	Oz the Great and Powerful	215000000.0	2013.0	Eng
39	The Amazing Spider-Man 2	200000000.0	2014.0	Eng
40	TRON: Legacy	170000000.0	2010.0	Eng
41	Cars 2	200000000.0	2011.0	Eng
42	Green Lantern	200000000.0	2011.0	Eng
43	Toy Story 3	200000000.0	2010.0	Eng
44	Terminator Salvation	200000000.0	2009.0	Eng
45	Furious 7	190000000.0	2015.0	Eng
46	World War Z	190000000.0	2013.0	Eng
47	X-Men: Days of Future Past	200000000.0	2014.0	Eng
48	Star Trek Into Darkness	190000000.0	2013.0	Eng

49	Jack the Giant Slayer	195000000.0	2013.0	Eng
50	The Great Gatsby	105000000.0	2013.0	Eng

	score	surname
0	7.9	james
1	7.1	gore
2	6.8	sam
3	8.5	christopher
5	6.6	andrew
6	6.2	sam
7	7.8	nathan
8	7.5	joss
9	7.5	david
10	6.9	zack
11	6.1	bryan
12	6.7	marc
13	7.3	gore
14	6.5	gore
15	7.2	zack
16	6.6	andrew
17	8.1	joss
18	6.7	rob
19	6.8	barry
20	7.5	peter
21	7.0	marc
22	6.7	ridley
23	7.9	peter
24	6.1	chris
25	7.2	peter
26	7.7	james
27	8.2	anthony
28	5.9	peter
29	7.0	colin
30	7.8	sam
31	7.3	sam
32	7.2	shane
33	6.5	tim
34	6.8	brett
35	7.3	dan
36	6.0	michael
37	5.7	michael
38	6.4	sam
39	6.7	marc
40	6.8	joseph
41	6.3	john
42	5.6	martin
43	8.3	lee

```

44    6.6      mcg
45    7.2      james
46    7.0      marc
47    8.0      bryan
48    7.8      j.j.
49    6.3      bryan
50    7.3      baz

```

```

[16]: #decouper et prendre la partie 1
movies['name'] = movies['director_name'].str.split(" ").str.get(1)

#check
movies.head(50)

```

```

[16]:      director_name  duration    gross \
0      james cameron    178.0  760505847.0
1      gore verbinski    169.0  309404152.0
2          sam mendes    148.0  200074175.0
3 christopher nolan    164.0  448130642.0
5      andrew stanton    132.0   73058679.0
6          sam raimi    156.0  336530303.0
7      nathan greno    100.0  200807262.0
8      joss whedon    141.0  458991599.0
9      david yates    153.0  301956980.0
10      zack snyder    183.0  330249062.0
11      bryan singer    169.0  200069408.0
12      marc forster    106.0  168368427.0
13      gore verbinski    151.0  423032628.0
14      gore verbinski    150.0   89289910.0
15      zack snyder    143.0  291021565.0
16      andrew adamson    150.0  141614023.0
17      joss whedon    173.0  623279547.0
18      rob marshall    136.0  241063875.0
19      barry sonnenfeld    106.0  179020854.0
20      peter jackson    164.0  255108370.0
21      marc webb    153.0  262030663.0
22      ridley scott    156.0  105219735.0
23      peter jackson    186.0  258355354.0
24      chris weitz    113.0   70083519.0
25      peter jackson    201.0  218051260.0
26      james cameron    194.0  658672302.0
27      anthony russo    147.0  407197282.0
28      peter berg    131.0   65173160.0
29      colin trevorrow    124.0  652177271.0
30          sam mendes    143.0  304360277.0
31          sam raimi    135.0  373377893.0
32      shane black    195.0  408992272.0

```

33	tim burton	108.0	334185206.0
34	brett ratner	104.0	234360014.0
35	dan scanlon	104.0	268488329.0
36	michael bay	150.0	402076689.0
37	michael bay	165.0	245428137.0
38	sam raimi	130.0	234903076.0
39	marc webb	142.0	202853933.0
40	joseph kosinski	125.0	172051787.0
41	john lasseter	106.0	191450875.0
42	martin campbell	123.0	116593191.0
43	lee unkrich	103.0	414984497.0
44	mcg	118.0	125320003.0
45	james wan	140.0	350034110.0
46	marc forster	123.0	202351611.0
47	bryan singer	149.0	233914986.0
48	j.j. abrams	132.0	228756232.0
49	bryan singer	114.0	65171860.0
50	baz luhrmann	143.0	144812796.0

	genres \
0	Action Adventure Fantasy Sci-Fi
1	Action Adventure Fantasy
2	Action Adventure Thriller
3	Action Thriller
5	Action Adventure Sci-Fi
6	Action Adventure Romance
7	Adventure Animation Comedy Family Fantasy Musi...
8	Action Adventure Sci-Fi
9	Adventure Family Fantasy Mystery
10	Action Adventure Sci-Fi
11	Action Adventure Sci-Fi
12	Action Adventure
13	Action Adventure Fantasy
14	Action Adventure Western
15	Action Adventure Fantasy Sci-Fi
16	Action Adventure Family Fantasy
17	Action Adventure Sci-Fi
18	Action Adventure Fantasy
19	Action Adventure Comedy Family Fantasy Sci-Fi
20	Adventure Fantasy
21	Action Adventure Fantasy
22	Action Adventure Drama History
23	Adventure Fantasy
24	Adventure Family Fantasy
25	Action Adventure Drama Romance
26	Drama Romance
27	Action Adventure Sci-Fi

28	Action Adventure Sci-Fi Thriller
29	Action Adventure Sci-Fi Thriller
30	Action Adventure Thriller
31	Action Adventure Fantasy Romance
32	Action Adventure Sci-Fi
33	Adventure Family Fantasy
34	Action Adventure Fantasy Sci-Fi Thriller
35	Adventure Animation Comedy Family Fantasy
36	Action Adventure Sci-Fi
37	Action Adventure Sci-Fi
38	Adventure Family Fantasy
39	Action Adventure Fantasy Sci-Fi
40	Action Adventure Sci-Fi
41	Adventure Animation Comedy Family Sport
42	Action Adventure Sci-Fi
43	Adventure Animation Comedy Family Fantasy
44	Action Adventure Sci-Fi
45	Action Crime Thriller
46	Action Adventure Horror Sci-Fi Thriller
47	Action Adventure Fantasy Sci-Fi Thriller
48	Action Adventure Sci-Fi
49	Adventure Fantasy
50	Drama Romance

	title	budget	year	language	\
0	Avatar	237000000.0	2009.0	Eng	
1	Pirates of the Caribbean: At World's End	300000000.0	2007.0	Eng	
2	Spectre	245000000.0	2015.0	Eng	
3	The Dark Knight Rises	250000000.0	2012.0	Eng	
5	John Carter	263700000.0	2012.0	Eng	
6	Spider-Man 3	258000000.0	2007.0	Eng	
7	Tangled	260000000.0	2010.0	Eng	
8	Avengers: Age of Ultron	250000000.0	2015.0	Eng	
9	Harry Potter and the Half-Blood Prince	250000000.0	2009.0	Eng	
10	Batman v Superman: Dawn of Justice	250000000.0	2016.0	Eng	
11	Superman Returns	209000000.0	2006.0	Eng	
12	Quantum of Solace	200000000.0	2008.0	Eng	
13	Pirates of the Caribbean: Dead Man's Chest	225000000.0	2006.0	Eng	
14	The Lone Ranger	215000000.0	2013.0	Eng	
15	Man of Steel	225000000.0	2013.0	Eng	
16	The Chronicles of Narnia: Prince Caspian	225000000.0	2008.0	Eng	
17	The Avengers	220000000.0	2012.0	Eng	
18	Pirates of the Caribbean: On Stranger Tides	250000000.0	2011.0	Eng	
19	Men in Black 3	225000000.0	2012.0	Eng	
20	The Hobbit: The Battle of the Five Armies	250000000.0	2014.0	Eng	
21	The Amazing Spider-Man	230000000.0	2012.0	Eng	
22	Robin Hood	200000000.0	2010.0	Eng	

23	The Hobbit: The Desolation of Smaug	225000000.0	2013.0	Eng
24	The Golden Compass	180000000.0	2007.0	Eng
25	King Kong	207000000.0	2005.0	Eng
26	Titanic	200000000.0	1997.0	Eng
27	Captain America: Civil War	250000000.0	2016.0	Eng
28	Battleship	209000000.0	2012.0	Eng
29	Jurassic World	150000000.0	2015.0	Eng
30	Skyfall	200000000.0	2012.0	Eng
31	Spider-Man 2	200000000.0	2004.0	Eng
32	Iron Man 3	200000000.0	2013.0	Eng
33	Alice in Wonderland	200000000.0	2010.0	Eng
34	X-Men: The Last Stand	210000000.0	2006.0	Eng
35	Monsters University	200000000.0	2013.0	Eng
36	Transformers: Revenge of the Fallen	200000000.0	2009.0	Eng
37	Transformers: Age of Extinction	210000000.0	2014.0	Eng
38	Oz the Great and Powerful	215000000.0	2013.0	Eng
39	The Amazing Spider-Man 2	200000000.0	2014.0	Eng
40	TRON: Legacy	170000000.0	2010.0	Eng
41	Cars 2	200000000.0	2011.0	Eng
42	Green Lantern	200000000.0	2011.0	Eng
43	Toy Story 3	200000000.0	2010.0	Eng
44	Terminator Salvation	200000000.0	2009.0	Eng
45	Furious 7	190000000.0	2015.0	Eng
46	World War Z	190000000.0	2013.0	Eng
47	X-Men: Days of Future Past	200000000.0	2014.0	Eng
48	Star Trek Into Darkness	190000000.0	2013.0	Eng
49	Jack the Giant Slayer	195000000.0	2013.0	Eng
50	The Great Gatsby	105000000.0	2013.0	Eng

	score	surname	name
0	7.9	james	cameron
1	7.1	gore	verbinski
2	6.8	sam	mendes
3	8.5	christopher	nolan
5	6.6	andrew	stanton
6	6.2	sam	raimi
7	7.8	nathan	greno
8	7.5	joss	whedon
9	7.5	david	yates
10	6.9	zack	snyder
11	6.1	bryan	singer
12	6.7	marc	forster
13	7.3	gore	verbinski
14	6.5	gore	verbinski
15	7.2	zack	snyder
16	6.6	andrew	adamson
17	8.1	joss	whedon

18	6.7	rob	marshall
19	6.8	barry	sonnenfeld
20	7.5	peter	jackson
21	7.0	marc	webb
22	6.7	ridley	scott
23	7.9	peter	jackson
24	6.1	chris	weitz
25	7.2	peter	jackson
26	7.7	james	cameron
27	8.2	anthony	russo
28	5.9	peter	berg
29	7.0	colin	trevorrow
30	7.8	sam	mendes
31	7.3	sam	raimi
32	7.2	shane	black
33	6.5	tim	burton
34	6.8	brett	ratner
35	7.3	dan	scanlon
36	6.0	michael	bay
37	5.7	michael	bay
38	6.4	sam	raimi
39	6.7	marc	webb
40	6.8	joseph	kosinski
41	6.3	john	lasseter
42	5.6	martin	campbell
43	8.3	lee	unkrich
44	6.6	mcg	NaN
45	7.2	james	wan
46	7.0	marc	forster
47	8.0	bryan	singer
48	7.8	j.j.	abrams
49	6.3	bryan	singer
50	7.3	baz	luhrmann

[52]: *#Extraire les datas contenant le nom du director name ZACK SNYDER*

```
movies_contain_zack = movies[movies['director_name'].str.contains("ZACK")]

#check
movies_contain_zack.head(50)
```

[52]:	director_name	duration	gross	\
10	ZACK SNYDER	183.0	330249062.0	
15	ZACK SNYDER	143.0	291021565.0	
164	ZACK SNYDER	215.0	107503316.0	
308	ZACK SNYDER	101.0	55673333.0	
558	ZACK SNYDER	128.0	36381716.0	

701	ZACK SNYDER	117.0	210592590.0
1535	JUSTIN ZACKHAM	89.0	21784432.0
1743	ZACK SNYDER	110.0	58885635.0

	genres \
10	Action Adventure Sci-Fi
15	Action Adventure Fantasy Sci-Fi
164	Action Drama Mystery Sci-Fi
308	Action Adventure Animation Family Fantasy
558	Action Fantasy
701	Action Drama Fantasy War
1535	Comedy
1743	Action Horror Thriller

	title	budget	year \
10	Batman v Superman: Dawn of Justice	250000000.0	2016.0
15	Man of Steel	225000000.0	2013.0
164	Watchmen	130000000.0	2009.0
308	Legend of the Guardians: The Owls of Ga'Hooole	80000000.0	2010.0
558	Sucker Punch	82000000.0	2011.0
701	300	65000000.0	2006.0
1535	The Big Wedding	35000000.0	2013.0
1743	Dawn of the Dead	26000000.0	2004.0

	language	score	surname	surnname	name
10	Eng	6.9	ZACK	[ZACK, SNYDER]	SNYDER
15	Eng	7.2	ZACK	[ZACK, SNYDER]	SNYDER
164	Eng	7.7	ZACK	[ZACK, SNYDER]	SNYDER
308	Eng	7.0	ZACK	[ZACK, SNYDER]	SNYDER
558	Eng	6.1	ZACK	[ZACK, SNYDER]	SNYDER
701	Eng	7.7	ZACK	[ZACK, SNYDER]	SNYDER
1535	Eng	5.6	JUSTIN	[JUSTIN, ZACKHAM]	ZACKHAM
1743	Eng	7.4	ZACK	[ZACK, SNYDER]	SNYDER

```
[53]: #voir le nombre de caracteres des valeurs d'une table
movies_contain_zack['director_name'].str.len() #ceci affiche les index
```

```
[53]: 10      11
      15      11
      164     11
      308     11
      558     11
      701     11
      1535    14
      1743    11
      Name: director_name, dtype: int64
```



```
[55]: #afficher les valeurs y correspondantes
movies_contain_zack.loc[movies_contain_zack['director_name'].str.len().
↳idxmax(), 'director_name']
```

```
[55]: 'JUSTIN ZACKHAM'
```

```
[17]: #Remplacement des valeurs particulieres

movies['year'].replace({'2009.0':'2009', '2010.0':'2010'})
#check
movies.head(50)

#changer le type de year
movies.info()

movies['year'] = movies['year'].astype('int')

#check
movies.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 3655 entries, 0 to 4915
Data columns (total 11 columns):
#   Column          Non-Null Count  Dtype
---  -
0   director_name    3655 non-null   object
1   duration          3655 non-null   float64
2   gross            3655 non-null   float64
3   genres           3655 non-null   object
4   title            3655 non-null   object
5   budget           3655 non-null   float64
6   year             3655 non-null   float64
7   language         3655 non-null   object
8   score            3655 non-null   float64
9   surname          3655 non-null   object
10  name              3645 non-null   object
dtypes: float64(5), object(6)
memory usage: 342.7+ KB
<class 'pandas.core.frame.DataFrame'>
Int64Index: 3655 entries, 0 to 4915
Data columns (total 11 columns):
#   Column          Non-Null Count  Dtype
---  -
0   director_name    3655 non-null   object
1   duration          3655 non-null   float64
2   gross            3655 non-null   float64
3   genres           3655 non-null   object
```

```

4  title          3655 non-null  object
5  budget         3655 non-null  float64
6  year           3655 non-null  int32
7  language       3655 non-null  object
8  score          3655 non-null  float64
9  surname        3655 non-null  object
10 name           3645 non-null  object
dtypes: float64(4), int32(1), object(6)
memory usage: 328.4+ KB

```

```
[60]: #check my new data
```

```
movies.head(50)
```

```
[60]:
```

	director_name	duration	gross	\
0	JAMES CAMERON	178.0	760505847.0	
1	GORE VERBINSKI	169.0	309404152.0	
2	SAM MENDES	148.0	200074175.0	
3	CHRISTOPHER NOLAN	164.0	448130642.0	
5	ANDREW STANTON	132.0	73058679.0	
6	SAM RAIMI	156.0	336530303.0	
7	NATHAN GRENO	100.0	200807262.0	
8	JOSS WHEDON	141.0	458991599.0	
9	DAVID YATES	153.0	301956980.0	
10	ZACK SNYDER	183.0	330249062.0	
11	BRYAN SINGER	169.0	200069408.0	
12	MARC FORSTER	106.0	168368427.0	
13	GORE VERBINSKI	151.0	423032628.0	
14	GORE VERBINSKI	150.0	89289910.0	
15	ZACK SNYDER	143.0	291021565.0	
16	ANDREW ADAMSON	150.0	141614023.0	
17	JOSS WHEDON	173.0	623279547.0	
18	ROB MARSHALL	136.0	241063875.0	
19	BARRY SONNENFELD	106.0	179020854.0	
20	PETER JACKSON	164.0	255108370.0	
21	MARC WEBB	153.0	262030663.0	
22	RIDLEY SCOTT	156.0	105219735.0	
23	PETER JACKSON	186.0	258355354.0	
24	CHRIS WEITZ	113.0	70083519.0	
25	PETER JACKSON	201.0	218051260.0	
26	JAMES CAMERON	194.0	658672302.0	
27	ANTHONY RUSSO	147.0	407197282.0	
28	PETER BERG	131.0	65173160.0	
29	COLIN TREVORROW	124.0	652177271.0	
30	SAM MENDES	143.0	304360277.0	
31	SAM RAIMI	135.0	373377893.0	
32	SHANE BLACK	195.0	408992272.0	

33	TIM BURTON	108.0	334185206.0
34	BRETT RATNER	104.0	234360014.0
35	DAN SCANLON	104.0	268488329.0
36	MICHAEL BAY	150.0	402076689.0
37	MICHAEL BAY	165.0	245428137.0
38	SAM RAIMI	130.0	234903076.0
39	MARC WEBB	142.0	202853933.0
40	JOSEPH KOSINSKI	125.0	172051787.0
41	JOHN LASSETER	106.0	191450875.0
42	MARTIN CAMPBELL	123.0	116593191.0
43	LEE UNKRICH	103.0	414984497.0
44	MCG	118.0	125320003.0
45	JAMES WAN	140.0	350034110.0
46	MARC FORSTER	123.0	202351611.0
47	BRYAN SINGER	149.0	233914986.0
48	J.J. ABRAMS	132.0	228756232.0
49	BRYAN SINGER	114.0	65171860.0
50	BAZ LUHRMANN	143.0	144812796.0

	genres \
0	Action Adventure Fantasy Sci-Fi
1	Action Adventure Fantasy
2	Action Adventure Thriller
3	Action Thriller
5	Action Adventure Sci-Fi
6	Action Adventure Romance
7	Adventure Animation Comedy Family Fantasy Musi...
8	Action Adventure Sci-Fi
9	Adventure Family Fantasy Mystery
10	Action Adventure Sci-Fi
11	Action Adventure Sci-Fi
12	Action Adventure
13	Action Adventure Fantasy
14	Action Adventure Western
15	Action Adventure Fantasy Sci-Fi
16	Action Adventure Family Fantasy
17	Action Adventure Sci-Fi
18	Action Adventure Fantasy
19	Action Adventure Comedy Family Fantasy Sci-Fi
20	Adventure Fantasy
21	Action Adventure Fantasy
22	Action Adventure Drama History
23	Adventure Fantasy
24	Adventure Family Fantasy
25	Action Adventure Drama Romance
26	Drama Romance
27	Action Adventure Sci-Fi

28	Action Adventure Sci-Fi Thriller
29	Action Adventure Sci-Fi Thriller
30	Action Adventure Thriller
31	Action Adventure Fantasy Romance
32	Action Adventure Sci-Fi
33	Adventure Family Fantasy
34	Action Adventure Fantasy Sci-Fi Thriller
35	Adventure Animation Comedy Family Fantasy
36	Action Adventure Sci-Fi
37	Action Adventure Sci-Fi
38	Adventure Family Fantasy
39	Action Adventure Fantasy Sci-Fi
40	Action Adventure Sci-Fi
41	Adventure Animation Comedy Family Sport
42	Action Adventure Sci-Fi
43	Adventure Animation Comedy Family Fantasy
44	Action Adventure Sci-Fi
45	Action Crime Thriller
46	Action Adventure Horror Sci-Fi Thriller
47	Action Adventure Fantasy Sci-Fi Thriller
48	Action Adventure Sci-Fi
49	Adventure Fantasy
50	Drama Romance

	title	budget	year	language	\
0	Avatar	237000000.0	2009	Eng	
1	Pirates of the Caribbean: At World's End	300000000.0	2007	Eng	
2	Spectre	245000000.0	2015	Eng	
3	The Dark Knight Rises	250000000.0	2012	Eng	
5	John Carter	263700000.0	2012	Eng	
6	Spider-Man 3	258000000.0	2007	Eng	
7	Tangled	260000000.0	2010	Eng	
8	Avengers: Age of Ultron	250000000.0	2015	Eng	
9	Harry Potter and the Half-Blood Prince	250000000.0	2009	Eng	
10	Batman v Superman: Dawn of Justice	250000000.0	2016	Eng	
11	Superman Returns	209000000.0	2006	Eng	
12	Quantum of Solace	200000000.0	2008	Eng	
13	Pirates of the Caribbean: Dead Man's Chest	225000000.0	2006	Eng	
14	The Lone Ranger	215000000.0	2013	Eng	
15	Man of Steel	225000000.0	2013	Eng	
16	The Chronicles of Narnia: Prince Caspian	225000000.0	2008	Eng	
17	The Avengers	220000000.0	2012	Eng	
18	Pirates of the Caribbean: On Stranger Tides	250000000.0	2011	Eng	
19	Men in Black 3	225000000.0	2012	Eng	
20	The Hobbit: The Battle of the Five Armies	250000000.0	2014	Eng	
21	The Amazing Spider-Man	230000000.0	2012	Eng	
22	Robin Hood	200000000.0	2010	Eng	

23	The Hobbit: The Desolation of Smaug	225000000.0	2013	Eng
24	The Golden Compass	180000000.0	2007	Eng
25	King Kong	207000000.0	2005	Eng
26	Titanic	200000000.0	1997	Eng
27	Captain America: Civil War	250000000.0	2016	Eng
28	Battleship	209000000.0	2012	Eng
29	Jurassic World	150000000.0	2015	Eng
30	Skyfall	200000000.0	2012	Eng
31	Spider-Man 2	200000000.0	2004	Eng
32	Iron Man 3	200000000.0	2013	Eng
33	Alice in Wonderland	200000000.0	2010	Eng
34	X-Men: The Last Stand	210000000.0	2006	Eng
35	Monsters University	200000000.0	2013	Eng
36	Transformers: Revenge of the Fallen	200000000.0	2009	Eng
37	Transformers: Age of Extinction	210000000.0	2014	Eng
38	Oz the Great and Powerful	215000000.0	2013	Eng
39	The Amazing Spider-Man 2	200000000.0	2014	Eng
40	TRON: Legacy	170000000.0	2010	Eng
41	Cars 2	200000000.0	2011	Eng
42	Green Lantern	200000000.0	2011	Eng
43	Toy Story 3	200000000.0	2010	Eng
44	Terminator Salvation	200000000.0	2009	Eng
45	Furious 7	190000000.0	2015	Eng
46	World War Z	190000000.0	2013	Eng
47	X-Men: Days of Future Past	200000000.0	2014	Eng
48	Star Trek Into Darkness	190000000.0	2013	Eng
49	Jack the Giant Slayer	195000000.0	2013	Eng
50	The Great Gatsby	105000000.0	2013	Eng

	score	surname	surnname	name
0	7.9	JAMES	[JAMES, CAMERON]	CAMERON
1	7.1	GORE	[GORE, VERBINSKI]	VERBINSKI
2	6.8	SAM	[SAM, MENDES]	MENDES
3	8.5	CHRISTOPHER	[CHRISTOPHER, NOLAN]	NOLAN
5	6.6	ANDREW	[ANDREW, STANTON]	STANTON
6	6.2	SAM	[SAM, RAIMI]	RAIMI
7	7.8	NATHAN	[NATHAN, GRENO]	GRENO
8	7.5	JOSS	[JOSS, WHEDON]	WHEDON
9	7.5	DAVID	[DAVID, YATES]	YATES
10	6.9	ZACK	[ZACK, SNYDER]	SNYDER
11	6.1	BRYAN	[BRYAN, SINGER]	SINGER
12	6.7	MARC	[MARC, FORSTER]	FORSTER
13	7.3	GORE	[GORE, VERBINSKI]	VERBINSKI
14	6.5	GORE	[GORE, VERBINSKI]	VERBINSKI
15	7.2	ZACK	[ZACK, SNYDER]	SNYDER
16	6.6	ANDREW	[ANDREW, ADAMSON]	ADAMSON
17	8.1	JOSS	[JOSS, WHEDON]	WHEDON

18	6.7	ROB	[ROB, MARSHALL]	MARSHALL
19	6.8	BARRY	[BARRY, SONNENFELD]	SONNENFELD
20	7.5	PETER	[PETER, JACKSON]	JACKSON
21	7.0	MARC	[MARC, WEBB]	WEBB
22	6.7	RIDLEY	[RIDLEY, SCOTT]	SCOTT
23	7.9	PETER	[PETER, JACKSON]	JACKSON
24	6.1	CHRIS	[CHRIS, WEITZ]	WEITZ
25	7.2	PETER	[PETER, JACKSON]	JACKSON
26	7.7	JAMES	[JAMES, CAMERON]	CAMERON
27	8.2	ANTHONY	[ANTHONY, RUSSO]	RUSSO
28	5.9	PETER	[PETER, BERG]	BERG
29	7.0	COLIN	[COLIN, TREVORROW]	TREVORROW
30	7.8	SAM	[SAM, MENDES]	MENDES
31	7.3	SAM	[SAM, RAIMI]	RAIMI
32	7.2	SHANE	[SHANE, BLACK]	BLACK
33	6.5	TIM	[TIM, BURTON]	BURTON
34	6.8	BRETT	[BRETT, RATNER]	RATNER
35	7.3	DAN	[DAN, SCANLON]	SCANLON
36	6.0	MICHAEL	[MICHAEL, BAY]	BAY
37	5.7	MICHAEL	[MICHAEL, BAY]	BAY
38	6.4	SAM	[SAM, RAIMI]	RAIMI
39	6.7	MARC	[MARC, WEBB]	WEBB
40	6.8	JOSEPH	[JOSEPH, KOSINSKI]	KOSINSKI
41	6.3	JOHN	[JOHN, LASSETER]	LASSETER
42	5.6	MARTIN	[MARTIN, CAMPBELL]	CAMPBELL
43	8.3	LEE	[LEE, UNKRICH]	UNKRICH
44	6.6	MCG	[MCG]	NaN
45	7.2	JAMES	[JAMES, WAN]	WAN
46	7.0	MARC	[MARC, FORSTER]	FORSTER
47	8.0	BRYAN	[BRYAN, SINGER]	SINGER
48	7.8	J.J.	[J.J., ABRAMS]	ABRAMS
49	6.3	BRYAN	[BRYAN, SINGER]	SINGER
50	7.3	BAZ	[BAZ, LUHRMANN]	LUHRMANN

[62]: *#Changer les colonnes en Maj ou Minisucles*

```
movies = movies.rename(columns = str.upper)

#check
movies.head(50)
```

[62]:

	DIRECTOR_NAME	DURATION	GROSS	\
0	JAMES CAMERON	178.0	760505847.0	
1	GORE VERBINSKI	169.0	309404152.0	
2	SAM MENDES	148.0	200074175.0	
3	CHRISTOPHER NOLAN	164.0	448130642.0	
5	ANDREW STANTON	132.0	73058679.0	

6	SAM RAIMI	156.0	336530303.0
7	NATHAN GRENO	100.0	200807262.0
8	JOSS WHEDON	141.0	458991599.0
9	DAVID YATES	153.0	301956980.0
10	ZACK SNYDER	183.0	330249062.0
11	BRYAN SINGER	169.0	200069408.0
12	MARC FORSTER	106.0	168368427.0
13	GORE VERBINSKI	151.0	423032628.0
14	GORE VERBINSKI	150.0	89289910.0
15	ZACK SNYDER	143.0	291021565.0
16	ANDREW ADAMSON	150.0	141614023.0
17	JOSS WHEDON	173.0	623279547.0
18	ROB MARSHALL	136.0	241063875.0
19	BARRY SONNENFELD	106.0	179020854.0
20	PETER JACKSON	164.0	255108370.0
21	MARC WEBB	153.0	262030663.0
22	RIDLEY SCOTT	156.0	105219735.0
23	PETER JACKSON	186.0	258355354.0
24	CHRIS WEITZ	113.0	70083519.0
25	PETER JACKSON	201.0	218051260.0
26	JAMES CAMERON	194.0	658672302.0
27	ANTHONY RUSSO	147.0	407197282.0
28	PETER BERG	131.0	65173160.0
29	COLIN TREVORROW	124.0	652177271.0
30	SAM MENDES	143.0	304360277.0
31	SAM RAIMI	135.0	373377893.0
32	SHANE BLACK	195.0	408992272.0
33	TIM BURTON	108.0	334185206.0
34	BRETT RATNER	104.0	234360014.0
35	DAN SCANLON	104.0	268488329.0
36	MICHAEL BAY	150.0	402076689.0
37	MICHAEL BAY	165.0	245428137.0
38	SAM RAIMI	130.0	234903076.0
39	MARC WEBB	142.0	202853933.0
40	JOSEPH KOSINSKI	125.0	172051787.0
41	JOHN LASSETER	106.0	191450875.0
42	MARTIN CAMPBELL	123.0	116593191.0
43	LEE UNKRICH	103.0	414984497.0
44	MCG	118.0	125320003.0
45	JAMES WAN	140.0	350034110.0
46	MARC FORSTER	123.0	202351611.0
47	BRYAN SINGER	149.0	233914986.0
48	J.J. ABRAMS	132.0	228756232.0
49	BRYAN SINGER	114.0	65171860.0
50	BAZ LUHRMANN	143.0	144812796.0

GENRES \

0	Action Adventure Fantasy Sci-Fi
1	Action Adventure Fantasy
2	Action Adventure Thriller
3	Action Thriller
5	Action Adventure Sci-Fi
6	Action Adventure Romance
7	Adventure Animation Comedy Family Fantasy Musi...
8	Action Adventure Sci-Fi
9	Adventure Family Fantasy Mystery
10	Action Adventure Sci-Fi
11	Action Adventure Sci-Fi
12	Action Adventure
13	Action Adventure Fantasy
14	Action Adventure Western
15	Action Adventure Fantasy Sci-Fi
16	Action Adventure Family Fantasy
17	Action Adventure Sci-Fi
18	Action Adventure Fantasy
19	Action Adventure Comedy Family Fantasy Sci-Fi
20	Adventure Fantasy
21	Action Adventure Fantasy
22	Action Adventure Drama History
23	Adventure Fantasy
24	Adventure Family Fantasy
25	Action Adventure Drama Romance
26	Drama Romance
27	Action Adventure Sci-Fi
28	Action Adventure Sci-Fi Thriller
29	Action Adventure Sci-Fi Thriller
30	Action Adventure Thriller
31	Action Adventure Fantasy Romance
32	Action Adventure Sci-Fi
33	Adventure Family Fantasy
34	Action Adventure Fantasy Sci-Fi Thriller
35	Adventure Animation Comedy Family Fantasy
36	Action Adventure Sci-Fi
37	Action Adventure Sci-Fi
38	Adventure Family Fantasy
39	Action Adventure Fantasy Sci-Fi
40	Action Adventure Sci-Fi
41	Adventure Animation Comedy Family Sport
42	Action Adventure Sci-Fi
43	Adventure Animation Comedy Family Fantasy
44	Action Adventure Sci-Fi
45	Action Crime Thriller
46	Action Adventure Horror Sci-Fi Thriller
47	Action Adventure Fantasy Sci-Fi Thriller



48 Action|Adventure|Sci-Fi  
 49 Adventure|Fantasy  
 50 Drama|Romance

	TITLE	BUDGET	YEAR	LANGUAGE	\
0	Avatar	237000000.0	2009	Eng	
1	Pirates of the Caribbean: At World's End	300000000.0	2007	Eng	
2	Spectre	245000000.0	2015	Eng	
3	The Dark Knight Rises	250000000.0	2012	Eng	
5	John Carter	263700000.0	2012	Eng	
6	Spider-Man 3	258000000.0	2007	Eng	
7	Tangled	260000000.0	2010	Eng	
8	Avengers: Age of Ultron	250000000.0	2015	Eng	
9	Harry Potter and the Half-Blood Prince	250000000.0	2009	Eng	
10	Batman v Superman: Dawn of Justice	250000000.0	2016	Eng	
11	Superman Returns	209000000.0	2006	Eng	
12	Quantum of Solace	200000000.0	2008	Eng	
13	Pirates of the Caribbean: Dead Man's Chest	225000000.0	2006	Eng	
14	The Lone Ranger	215000000.0	2013	Eng	
15	Man of Steel	225000000.0	2013	Eng	
16	The Chronicles of Narnia: Prince Caspian	225000000.0	2008	Eng	
17	The Avengers	220000000.0	2012	Eng	
18	Pirates of the Caribbean: On Stranger Tides	250000000.0	2011	Eng	
19	Men in Black 3	225000000.0	2012	Eng	
20	The Hobbit: The Battle of the Five Armies	250000000.0	2014	Eng	
21	The Amazing Spider-Man	230000000.0	2012	Eng	
22	Robin Hood	200000000.0	2010	Eng	
23	The Hobbit: The Desolation of Smaug	225000000.0	2013	Eng	
24	The Golden Compass	180000000.0	2007	Eng	
25	King Kong	207000000.0	2005	Eng	
26	Titanic	200000000.0	1997	Eng	
27	Captain America: Civil War	250000000.0	2016	Eng	
28	Battleship	209000000.0	2012	Eng	
29	Jurassic World	150000000.0	2015	Eng	
30	Skyfall	200000000.0	2012	Eng	
31	Spider-Man 2	200000000.0	2004	Eng	
32	Iron Man 3	200000000.0	2013	Eng	
33	Alice in Wonderland	200000000.0	2010	Eng	
34	X-Men: The Last Stand	210000000.0	2006	Eng	
35	Monsters University	200000000.0	2013	Eng	
36	Transformers: Revenge of the Fallen	200000000.0	2009	Eng	
37	Transformers: Age of Extinction	210000000.0	2014	Eng	
38	Oz the Great and Powerful	215000000.0	2013	Eng	
39	The Amazing Spider-Man 2	200000000.0	2014	Eng	
40	TRON: Legacy	170000000.0	2010	Eng	
41	Cars 2	200000000.0	2011	Eng	
42	Green Lantern	200000000.0	2011	Eng	

43		Toy Story 3	200000000.0	2010	Eng
44		Terminator Salvation	200000000.0	2009	Eng
45		Furious 7	190000000.0	2015	Eng
46		World War Z	190000000.0	2013	Eng
47		X-Men: Days of Future Past	200000000.0	2014	Eng
48		Star Trek Into Darkness	190000000.0	2013	Eng
49		Jack the Giant Slayer	195000000.0	2013	Eng
50		The Great Gatsby	105000000.0	2013	Eng

	SCORE	SURNAME	SURNMAME	NAME
0	7.9	JAMES	[JAMES, CAMERON]	CAMERON
1	7.1	GORE	[GORE, VERBINSKI]	VERBINSKI
2	6.8	SAM	[SAM, MENDES]	MENDES
3	8.5	CHRISTOPHER	[CHRISTOPHER, NOLAN]	NOLAN
5	6.6	ANDREW	[ANDREW, STANTON]	STANTON
6	6.2	SAM	[SAM, RAIMI]	RAIMI
7	7.8	NATHAN	[NATHAN, GRENO]	GRENO
8	7.5	JOSS	[JOSS, WHEDON]	WHEDON
9	7.5	DAVID	[DAVID, YATES]	YATES
10	6.9	ZACK	[ZACK, SNYDER]	SNYDER
11	6.1	BRYAN	[BRYAN, SINGER]	SINGER
12	6.7	MARC	[MARC, FORSTER]	FORSTER
13	7.3	GORE	[GORE, VERBINSKI]	VERBINSKI
14	6.5	GORE	[GORE, VERBINSKI]	VERBINSKI
15	7.2	ZACK	[ZACK, SNYDER]	SNYDER
16	6.6	ANDREW	[ANDREW, ADAMSON]	ADAMSON
17	8.1	JOSS	[JOSS, WHEDON]	WHEDON
18	6.7	ROB	[ROB, MARSHALL]	MARSHALL
19	6.8	BARRY	[BARRY, SONNENFELD]	SONNENFELD
20	7.5	PETER	[PETER, JACKSON]	JACKSON
21	7.0	MARC	[MARC, WEBB]	WEBB
22	6.7	RIDLEY	[RIDLEY, SCOTT]	SCOTT
23	7.9	PETER	[PETER, JACKSON]	JACKSON
24	6.1	CHRIS	[CHRIS, WEITZ]	WEITZ
25	7.2	PETER	[PETER, JACKSON]	JACKSON
26	7.7	JAMES	[JAMES, CAMERON]	CAMERON
27	8.2	ANTHONY	[ANTHONY, RUSSO]	RUSSO
28	5.9	PETER	[PETER, BERG]	BERG
29	7.0	COLIN	[COLIN, TREVORROW]	TREVORROW
30	7.8	SAM	[SAM, MENDES]	MENDES
31	7.3	SAM	[SAM, RAIMI]	RAIMI
32	7.2	SHANE	[SHANE, BLACK]	BLACK
33	6.5	TIM	[TIM, BURTON]	BURTON
34	6.8	BRETT	[BRETT, RATNER]	RATNER
35	7.3	DAN	[DAN, SCANLON]	SCANLON
36	6.0	MICHAEL	[MICHAEL, BAY]	BAY
37	5.7	MICHAEL	[MICHAEL, BAY]	BAY

38	6.4	SAM	[SAM, RAIMI]	RAIMI
39	6.7	MARC	[MARC, WEBB]	WEBB
40	6.8	JOSEPH	[JOSEPH, KOSINSKI]	KOSINSKI
41	6.3	JOHN	[JOHN, LASSETER]	LASSETER
42	5.6	MARTIN	[MARTIN, CAMPBELL]	CAMPBELL
43	8.3	LEE	[LEE, UNKRICH]	UNKRICH
44	6.6	MCG	[MCG]	NaN
45	7.2	JAMES	[JAMES, WAN]	WAN
46	7.0	MARC	[MARC, FORSTER]	FORSTER
47	8.0	BRYAN	[BRYAN, SINGER]	SINGER
48	7.8	J.J.	[J.J., ABRAMS]	ABRAMS
49	6.3	BRYAN	[BRYAN, SINGER]	SINGER
50	7.3	BAZ	[BAZ, LUHRMANN]	LUHRMANN

```
[18]: #remettre les noms de colonnes en miniscule
movies = movies.rename(columns = str.lower)

#check
movies.head(50)
```

```
[18]:
```

	director_name	duration	gross \
0	james cameron	178.0	760505847.0
1	gore verbinski	169.0	309404152.0
2	sam mendes	148.0	200074175.0
3	christopher nolan	164.0	448130642.0
5	andrew stanton	132.0	73058679.0
6	sam raimi	156.0	336530303.0
7	nathan greno	100.0	200807262.0
8	joss whedon	141.0	458991599.0
9	david yates	153.0	301956980.0
10	zack snyder	183.0	330249062.0
11	bryan singer	169.0	200069408.0
12	marc forster	106.0	168368427.0
13	gore verbinski	151.0	423032628.0
14	gore verbinski	150.0	89289910.0
15	zack snyder	143.0	291021565.0
16	andrew adamson	150.0	141614023.0
17	joss whedon	173.0	623279547.0
18	rob marshall	136.0	241063875.0
19	barry sonnenfeld	106.0	179020854.0
20	peter jackson	164.0	255108370.0
21	marc webb	153.0	262030663.0
22	ridley scott	156.0	105219735.0
23	peter jackson	186.0	258355354.0
24	chris weitz	113.0	70083519.0
25	peter jackson	201.0	218051260.0
26	james cameron	194.0	658672302.0

27	anthony russo	147.0	407197282.0
28	peter berg	131.0	65173160.0
29	colin trevorrow	124.0	652177271.0
30	sam mendes	143.0	304360277.0
31	sam raimi	135.0	373377893.0
32	shane black	195.0	408992272.0
33	tim burton	108.0	334185206.0
34	brett ratner	104.0	234360014.0
35	dan scanlon	104.0	268488329.0
36	michael bay	150.0	402076689.0
37	michael bay	165.0	245428137.0
38	sam raimi	130.0	234903076.0
39	marc webb	142.0	202853933.0
40	joseph kosinski	125.0	172051787.0
41	john lasserter	106.0	191450875.0
42	martin campbell	123.0	116593191.0
43	lee unkrich	103.0	414984497.0
44	mcg	118.0	125320003.0
45	james wan	140.0	350034110.0
46	marc forster	123.0	202351611.0
47	bryan singer	149.0	233914986.0
48	j.j. abrams	132.0	228756232.0
49	bryan singer	114.0	65171860.0
50	baz luhrmann	143.0	144812796.0

	genres \
0	Action Adventure Fantasy Sci-Fi
1	Action Adventure Fantasy
2	Action Adventure Thriller
3	Action Thriller
5	Action Adventure Sci-Fi
6	Action Adventure Romance
7	Adventure Animation Comedy Family Fantasy Musi...
8	Action Adventure Sci-Fi
9	Adventure Family Fantasy Mystery
10	Action Adventure Sci-Fi
11	Action Adventure Sci-Fi
12	Action Adventure
13	Action Adventure Fantasy
14	Action Adventure Western
15	Action Adventure Fantasy Sci-Fi
16	Action Adventure Family Fantasy
17	Action Adventure Sci-Fi
18	Action Adventure Fantasy
19	Action Adventure Comedy Family Fantasy Sci-Fi
20	Adventure Fantasy
21	Action Adventure Fantasy

22 Action|Adventure|Drama|History  
 23 Adventure|Fantasy  
 24 Adventure|Family|Fantasy  
 25 Action|Adventure|Drama|Romance  
 26 Drama|Romance  
 27 Action|Adventure|Sci-Fi  
 28 Action|Adventure|Sci-Fi|Thriller  
 29 Action|Adventure|Sci-Fi|Thriller  
 30 Action|Adventure|Thriller  
 31 Action|Adventure|Fantasy|Romance  
 32 Action|Adventure|Sci-Fi  
 33 Adventure|Family|Fantasy  
 34 Action|Adventure|Fantasy|Sci-Fi|Thriller  
 35 Adventure|Animation|Comedy|Family|Fantasy  
 36 Action|Adventure|Sci-Fi  
 37 Action|Adventure|Sci-Fi  
 38 Adventure|Family|Fantasy  
 39 Action|Adventure|Fantasy|Sci-Fi  
 40 Action|Adventure|Sci-Fi  
 41 Adventure|Animation|Comedy|Family|Sport  
 42 Action|Adventure|Sci-Fi  
 43 Adventure|Animation|Comedy|Family|Fantasy  
 44 Action|Adventure|Sci-Fi  
 45 Action|Crime|Thriller  
 46 Action|Adventure|Horror|Sci-Fi|Thriller  
 47 Action|Adventure|Fantasy|Sci-Fi|Thriller  
 48 Action|Adventure|Sci-Fi  
 49 Adventure|Fantasy  
 50 Drama|Romance

	title	budget	year	language \
0	Avatar	237000000.0	2009	Eng
1	Pirates of the Caribbean: At World's End	300000000.0	2007	Eng
2	Spectre	245000000.0	2015	Eng
3	The Dark Knight Rises	250000000.0	2012	Eng
5	John Carter	263700000.0	2012	Eng
6	Spider-Man 3	258000000.0	2007	Eng
7	Tangled	260000000.0	2010	Eng
8	Avengers: Age of Ultron	250000000.0	2015	Eng
9	Harry Potter and the Half-Blood Prince	250000000.0	2009	Eng
10	Batman v Superman: Dawn of Justice	250000000.0	2016	Eng
11	Superman Returns	209000000.0	2006	Eng
12	Quantum of Solace	200000000.0	2008	Eng
13	Pirates of the Caribbean: Dead Man's Chest	225000000.0	2006	Eng
14	The Lone Ranger	215000000.0	2013	Eng
15	Man of Steel	225000000.0	2013	Eng
16	The Chronicles of Narnia: Prince Caspian	225000000.0	2008	Eng

17		The Avengers	220000000.0	2012	Eng
18	Pirates of the Caribbean: On Stranger Tides		250000000.0	2011	Eng
19		Men in Black 3	225000000.0	2012	Eng
20	The Hobbit: The Battle of the Five Armies		250000000.0	2014	Eng
21		The Amazing Spider-Man	230000000.0	2012	Eng
22		Robin Hood	200000000.0	2010	Eng
23	The Hobbit: The Desolation of Smaug		225000000.0	2013	Eng
24		The Golden Compass	180000000.0	2007	Eng
25		King Kong	207000000.0	2005	Eng
26		Titanic	200000000.0	1997	Eng
27	Captain America: Civil War		250000000.0	2016	Eng
28		Battleship	209000000.0	2012	Eng
29		Jurassic World	150000000.0	2015	Eng
30		Skyfall	200000000.0	2012	Eng
31		Spider-Man 2	200000000.0	2004	Eng
32		Iron Man 3	200000000.0	2013	Eng
33		Alice in Wonderland	200000000.0	2010	Eng
34	X-Men: The Last Stand		210000000.0	2006	Eng
35		Monsters University	200000000.0	2013	Eng
36	Transformers: Revenge of the Fallen		200000000.0	2009	Eng
37	Transformers: Age of Extinction		210000000.0	2014	Eng
38	Oz the Great and Powerful		215000000.0	2013	Eng
39	The Amazing Spider-Man 2		200000000.0	2014	Eng
40		TRON: Legacy	170000000.0	2010	Eng
41		Cars 2	200000000.0	2011	Eng
42		Green Lantern	200000000.0	2011	Eng
43		Toy Story 3	200000000.0	2010	Eng
44		Terminator Salvation	200000000.0	2009	Eng
45		Furious 7	190000000.0	2015	Eng
46		World War Z	190000000.0	2013	Eng
47	X-Men: Days of Future Past		200000000.0	2014	Eng
48	Star Trek Into Darkness		190000000.0	2013	Eng
49	Jack the Giant Slayer		195000000.0	2013	Eng
50		The Great Gatsby	105000000.0	2013	Eng

	score	surname	name
0	7.9	james	cameron
1	7.1	gore	verbinski
2	6.8	sam	mendes
3	8.5	christopher	nolan
5	6.6	andrew	stanton
6	6.2	sam	raimi
7	7.8	nathan	greno
8	7.5	joss	whedon
9	7.5	david	yates
10	6.9	zack	snyder
11	6.1	bryan	singer

12	6.7	marc	forster
13	7.3	gore	verbinski
14	6.5	gore	verbinski
15	7.2	zack	snyder
16	6.6	andrew	adamson
17	8.1	joss	whedon
18	6.7	rob	marshall
19	6.8	barry	sonnenfeld
20	7.5	peter	jackson
21	7.0	marc	webb
22	6.7	ridley	scott
23	7.9	peter	jackson
24	6.1	chris	weitz
25	7.2	peter	jackson
26	7.7	james	cameron
27	8.2	anthony	russo
28	5.9	peter	berg
29	7.0	colin	trevorrow
30	7.8	sam	mendes
31	7.3	sam	raimi
32	7.2	shane	black
33	6.5	tim	burton
34	6.8	brett	ratner
35	7.3	dan	scanlon
36	6.0	michael	bay
37	5.7	michael	bay
38	6.4	sam	raimi
39	6.7	marc	webb
40	6.8	joseph	kosinski
41	6.3	john	lasseter
42	5.6	martin	campbell
43	8.3	lee	unkrich
44	6.6	mcg	NaN
45	7.2	james	wan
46	7.0	marc	forster
47	8.0	bryan	singer
48	7.8	j.j.	abrams
49	6.3	bryan	singer
50	7.3	baz	luhrmann

### 1.0.2 Creation des Colonnes Customized ou Personnalisées

```
[19]: #creation d'une colonne taux rendu

movies['taux_rendu'] = (movies['budget']/movies['score']) * 100

#check
```

```
movies.head(50)
```

```
[19]:
```

	director_name	duration	gross	\
0	james cameron	178.0	760505847.0	
1	gore verbinski	169.0	309404152.0	
2	sam mendes	148.0	200074175.0	
3	christopher nolan	164.0	448130642.0	
5	andrew stanton	132.0	73058679.0	
6	sam raimi	156.0	336530303.0	
7	nathan greno	100.0	200807262.0	
8	joss whedon	141.0	458991599.0	
9	david yates	153.0	301956980.0	
10	zack snyder	183.0	330249062.0	
11	bryan singer	169.0	200069408.0	
12	marc forster	106.0	168368427.0	
13	gore verbinski	151.0	423032628.0	
14	gore verbinski	150.0	89289910.0	
15	zack snyder	143.0	291021565.0	
16	andrew adamson	150.0	141614023.0	
17	joss whedon	173.0	623279547.0	
18	rob marshall	136.0	241063875.0	
19	barry sonnenfeld	106.0	179020854.0	
20	peter jackson	164.0	255108370.0	
21	marc webb	153.0	262030663.0	
22	ridley scott	156.0	105219735.0	
23	peter jackson	186.0	258355354.0	
24	chris weitz	113.0	70083519.0	
25	peter jackson	201.0	218051260.0	
26	james cameron	194.0	658672302.0	
27	anthony russo	147.0	407197282.0	
28	peter berg	131.0	65173160.0	
29	colin trevorrow	124.0	652177271.0	
30	sam mendes	143.0	304360277.0	
31	sam raimi	135.0	373377893.0	
32	shane black	195.0	408992272.0	
33	tim burton	108.0	334185206.0	
34	brett ratner	104.0	234360014.0	
35	dan scanlon	104.0	268488329.0	
36	michael bay	150.0	402076689.0	
37	michael bay	165.0	245428137.0	
38	sam raimi	130.0	234903076.0	
39	marc webb	142.0	202853933.0	
40	joseph kosinski	125.0	172051787.0	
41	john lasseter	106.0	191450875.0	
42	martin campbell	123.0	116593191.0	
43	lee unkrich	103.0	414984497.0	
44	mcg	118.0	125320003.0	



45	james wan	140.0	350034110.0
46	marc forster	123.0	202351611.0
47	bryan singer	149.0	233914986.0
48	j.j. abrams	132.0	228756232.0
49	bryan singer	114.0	65171860.0
50	baz luhrmann	143.0	144812796.0

	genres \
0	Action Adventure Fantasy Sci-Fi
1	Action Adventure Fantasy
2	Action Adventure Thriller
3	Action Thriller
5	Action Adventure Sci-Fi
6	Action Adventure Romance
7	Adventure Animation Comedy Family Fantasy Musi...
8	Action Adventure Sci-Fi
9	Adventure Family Fantasy Mystery
10	Action Adventure Sci-Fi
11	Action Adventure Sci-Fi
12	Action Adventure
13	Action Adventure Fantasy
14	Action Adventure Western
15	Action Adventure Fantasy Sci-Fi
16	Action Adventure Family Fantasy
17	Action Adventure Sci-Fi
18	Action Adventure Fantasy
19	Action Adventure Comedy Family Fantasy Sci-Fi
20	Adventure Fantasy
21	Action Adventure Fantasy
22	Action Adventure Drama History
23	Adventure Fantasy
24	Adventure Family Fantasy
25	Action Adventure Drama Romance
26	Drama Romance
27	Action Adventure Sci-Fi
28	Action Adventure Sci-Fi Thriller
29	Action Adventure Sci-Fi Thriller
30	Action Adventure Thriller
31	Action Adventure Fantasy Romance
32	Action Adventure Sci-Fi
33	Adventure Family Fantasy
34	Action Adventure Fantasy Sci-Fi Thriller
35	Adventure Animation Comedy Family Fantasy
36	Action Adventure Sci-Fi
37	Action Adventure Sci-Fi
38	Adventure Family Fantasy
39	Action Adventure Fantasy Sci-Fi

40 Action|Adventure|Sci-Fi  
 41 Adventure|Animation|Comedy|Family|Sport  
 42 Action|Adventure|Sci-Fi  
 43 Adventure|Animation|Comedy|Family|Fantasy  
 44 Action|Adventure|Sci-Fi  
 45 Action|Crime|Thriller  
 46 Action|Adventure|Horror|Sci-Fi|Thriller  
 47 Action|Adventure|Fantasy|Sci-Fi|Thriller  
 48 Action|Adventure|Sci-Fi  
 49 Adventure|Fantasy  
 50 Drama|Romance

	title	budget	year	language	\
0	Avatar	237000000.0	2009	Eng	
1	Pirates of the Caribbean: At World's End	300000000.0	2007	Eng	
2	Spectre	245000000.0	2015	Eng	
3	The Dark Knight Rises	250000000.0	2012	Eng	
5	John Carter	263700000.0	2012	Eng	
6	Spider-Man 3	258000000.0	2007	Eng	
7	Tangled	260000000.0	2010	Eng	
8	Avengers: Age of Ultron	250000000.0	2015	Eng	
9	Harry Potter and the Half-Blood Prince	250000000.0	2009	Eng	
10	Batman v Superman: Dawn of Justice	250000000.0	2016	Eng	
11	Superman Returns	209000000.0	2006	Eng	
12	Quantum of Solace	200000000.0	2008	Eng	
13	Pirates of the Caribbean: Dead Man's Chest	225000000.0	2006	Eng	
14	The Lone Ranger	215000000.0	2013	Eng	
15	Man of Steel	225000000.0	2013	Eng	
16	The Chronicles of Narnia: Prince Caspian	225000000.0	2008	Eng	
17	The Avengers	220000000.0	2012	Eng	
18	Pirates of the Caribbean: On Stranger Tides	250000000.0	2011	Eng	
19	Men in Black 3	225000000.0	2012	Eng	
20	The Hobbit: The Battle of the Five Armies	250000000.0	2014	Eng	
21	The Amazing Spider-Man	230000000.0	2012	Eng	
22	Robin Hood	200000000.0	2010	Eng	
23	The Hobbit: The Desolation of Smaug	225000000.0	2013	Eng	
24	The Golden Compass	180000000.0	2007	Eng	
25	King Kong	207000000.0	2005	Eng	
26	Titanic	200000000.0	1997	Eng	
27	Captain America: Civil War	250000000.0	2016	Eng	
28	Battleship	209000000.0	2012	Eng	
29	Jurassic World	150000000.0	2015	Eng	
30	Skyfall	200000000.0	2012	Eng	
31	Spider-Man 2	200000000.0	2004	Eng	
32	Iron Man 3	200000000.0	2013	Eng	
33	Alice in Wonderland	200000000.0	2010	Eng	
34	X-Men: The Last Stand	210000000.0	2006	Eng	

35	Monsters University	200000000.0	2013	Eng
36	Transformers: Revenge of the Fallen	200000000.0	2009	Eng
37	Transformers: Age of Extinction	210000000.0	2014	Eng
38	Oz the Great and Powerful	215000000.0	2013	Eng
39	The Amazing Spider-Man 2	200000000.0	2014	Eng
40	TRON: Legacy	170000000.0	2010	Eng
41	Cars 2	200000000.0	2011	Eng
42	Green Lantern	200000000.0	2011	Eng
43	Toy Story 3	200000000.0	2010	Eng
44	Terminator Salvation	200000000.0	2009	Eng
45	Furious 7	190000000.0	2015	Eng
46	World War Z	190000000.0	2013	Eng
47	X-Men: Days of Future Past	200000000.0	2014	Eng
48	Star Trek Into Darkness	190000000.0	2013	Eng
49	Jack the Giant Slayer	195000000.0	2013	Eng
50	The Great Gatsby	105000000.0	2013	Eng

	score	surname	name	taux_rendu
0	7.9	james	cameron	3.000000e+09
1	7.1	gore	verbinski	4.225352e+09
2	6.8	sam	mendes	3.602941e+09
3	8.5	christopher	nolan	2.941176e+09
5	6.6	andrew	stanton	3.995455e+09
6	6.2	sam	raimi	4.161290e+09
7	7.8	nathan	greno	3.333333e+09
8	7.5	joss	whedon	3.333333e+09
9	7.5	david	yates	3.333333e+09
10	6.9	zack	snyder	3.623188e+09
11	6.1	bryan	singer	3.426230e+09
12	6.7	marc	forster	2.985075e+09
13	7.3	gore	verbinski	3.082192e+09
14	6.5	gore	verbinski	3.307692e+09
15	7.2	zack	snyder	3.125000e+09
16	6.6	andrew	adamson	3.409091e+09
17	8.1	joss	whedon	2.716049e+09
18	6.7	rob	marshall	3.731343e+09
19	6.8	barry	sonnenfeld	3.308824e+09
20	7.5	peter	jackson	3.333333e+09
21	7.0	marc	webb	3.285714e+09
22	6.7	ridley	scott	2.985075e+09
23	7.9	peter	jackson	2.848101e+09
24	6.1	chris	weitz	2.950820e+09
25	7.2	peter	jackson	2.875000e+09
26	7.7	james	cameron	2.597403e+09
27	8.2	anthony	russo	3.048780e+09
28	5.9	peter	berg	3.542373e+09
29	7.0	colin	trevorrow	2.142857e+09

30	7.8	sam	mendes	2.564103e+09
31	7.3	sam	raimi	2.739726e+09
32	7.2	shane	black	2.777778e+09
33	6.5	tim	burton	3.076923e+09
34	6.8	brett	ratner	3.088235e+09
35	7.3	dan	scanlon	2.739726e+09
36	6.0	michael	bay	3.333333e+09
37	5.7	michael	bay	3.684211e+09
38	6.4	sam	raimi	3.359375e+09
39	6.7	marc	webb	2.985075e+09
40	6.8	joseph	kosinski	2.500000e+09
41	6.3	john	lasseter	3.174603e+09
42	5.6	martin	campbell	3.571429e+09
43	8.3	lee	unkrich	2.409639e+09
44	6.6	mcg	NaN	3.030303e+09
45	7.2	james	wan	2.638889e+09
46	7.0	marc	forster	2.714286e+09
47	8.0	bryan	singer	2.500000e+09
48	7.8	j.j.	abrams	2.435897e+09
49	6.3	bryan	singer	3.095238e+09
50	7.3	baz	luhrmann	1.438356e+09

colonne conditionnelle avec np.where()

```
[20]: #sur la colonne year , np.where() fonctionne comme if(cond, val1, val2)
      ↪NotCondition)
      #import de numpy
      import numpy as np

      movies['siecple_simple'] = np.where(movies['year'] == 2009, 'Maintain Siecle',
      ↪'Rejected Siecle')

      #check
      movies.head(50)
```

```
[20]:      director_name  duration  gross \
0      james cameron    178.0  760505847.0
1      gore verbinski    169.0  309404152.0
2      sam mendes      148.0  200074175.0
3  christopher nolan    164.0  448130642.0
5      andrew stanton    132.0   73058679.0
6      sam raimi       156.0  336530303.0
7      nathan greno    100.0  200807262.0
8      joss whedon     141.0  458991599.0
9      david yates     153.0  301956980.0
10     zack snyder     183.0  330249062.0
11     bryan singer     169.0  200069408.0
12     marc forster     106.0  168368427.0
```

13	gore verbinski	151.0	423032628.0
14	gore verbinski	150.0	89289910.0
15	zack snyder	143.0	291021565.0
16	andrew adamson	150.0	141614023.0
17	joss whedon	173.0	623279547.0
18	rob marshall	136.0	241063875.0
19	barry sonnenfeld	106.0	179020854.0
20	peter jackson	164.0	255108370.0
21	marc webb	153.0	262030663.0
22	ridley scott	156.0	105219735.0
23	peter jackson	186.0	258355354.0
24	chris weitz	113.0	70083519.0
25	peter jackson	201.0	218051260.0
26	james cameron	194.0	658672302.0
27	anthony russo	147.0	407197282.0
28	peter berg	131.0	65173160.0
29	colin trevorrow	124.0	652177271.0
30	sam mendes	143.0	304360277.0
31	sam raimi	135.0	373377893.0
32	shane black	195.0	408992272.0
33	tim burton	108.0	334185206.0
34	brett ratner	104.0	234360014.0
35	dan scanlon	104.0	268488329.0
36	michael bay	150.0	402076689.0
37	michael bay	165.0	245428137.0
38	sam raimi	130.0	234903076.0
39	marc webb	142.0	202853933.0
40	joseph kosinski	125.0	172051787.0
41	john lasseter	106.0	191450875.0
42	martin campbell	123.0	116593191.0
43	lee unkrich	103.0	414984497.0
44	mcg	118.0	125320003.0
45	james wan	140.0	350034110.0
46	marc forster	123.0	202351611.0
47	bryan singer	149.0	233914986.0
48	j.j. abrams	132.0	228756232.0
49	bryan singer	114.0	65171860.0
50	baz luhrmann	143.0	144812796.0

	genres \
0	Action Adventure Fantasy Sci-Fi
1	Action Adventure Fantasy
2	Action Adventure Thriller
3	Action Thriller
5	Action Adventure Sci-Fi
6	Action Adventure Romance
7	Adventure Animation Comedy Family Fantasy Musi...

8 Action|Adventure|Sci-Fi  
 9 Adventure|Family|Fantasy|Mystery  
 10 Action|Adventure|Sci-Fi  
 11 Action|Adventure|Sci-Fi  
 12 Action|Adventure  
 13 Action|Adventure|Fantasy  
 14 Action|Adventure|Western  
 15 Action|Adventure|Fantasy|Sci-Fi  
 16 Action|Adventure|Family|Fantasy  
 17 Action|Adventure|Sci-Fi  
 18 Action|Adventure|Fantasy  
 19 Action|Adventure|Comedy|Family|Fantasy|Sci-Fi  
 20 Adventure|Fantasy  
 21 Action|Adventure|Fantasy  
 22 Action|Adventure|Drama|History  
 23 Adventure|Fantasy  
 24 Adventure|Family|Fantasy  
 25 Action|Adventure|Drama|Romance  
 26 Drama|Romance  
 27 Action|Adventure|Sci-Fi  
 28 Action|Adventure|Sci-Fi|Thriller  
 29 Action|Adventure|Sci-Fi|Thriller  
 30 Action|Adventure|Thriller  
 31 Action|Adventure|Fantasy|Romance  
 32 Action|Adventure|Sci-Fi  
 33 Adventure|Family|Fantasy  
 34 Action|Adventure|Fantasy|Sci-Fi|Thriller  
 35 Adventure|Animation|Comedy|Family|Fantasy  
 36 Action|Adventure|Sci-Fi  
 37 Action|Adventure|Sci-Fi  
 38 Adventure|Family|Fantasy  
 39 Action|Adventure|Fantasy|Sci-Fi  
 40 Action|Adventure|Sci-Fi  
 41 Adventure|Animation|Comedy|Family|Sport  
 42 Action|Adventure|Sci-Fi  
 43 Adventure|Animation|Comedy|Family|Fantasy  
 44 Action|Adventure|Sci-Fi  
 45 Action|Crime|Thriller  
 46 Action|Adventure|Horror|Sci-Fi|Thriller  
 47 Action|Adventure|Fantasy|Sci-Fi|Thriller  
 48 Action|Adventure|Sci-Fi  
 49 Adventure|Fantasy  
 50 Drama|Romance

	title	budget	year	language	\
0	Avatar	237000000.0	2009	Eng	
1	Pirates of the Caribbean: At World's End	300000000.0	2007	Eng	

2		Spectre	245000000.0	2015	Eng
3		The Dark Knight Rises	250000000.0	2012	Eng
5		John Carter	263700000.0	2012	Eng
6		Spider-Man 3	258000000.0	2007	Eng
7		Tangled	260000000.0	2010	Eng
8		Avengers: Age of Ultron	250000000.0	2015	Eng
9		Harry Potter and the Half-Blood Prince	250000000.0	2009	Eng
10		Batman v Superman: Dawn of Justice	250000000.0	2016	Eng
11		Superman Returns	209000000.0	2006	Eng
12		Quantum of Solace	200000000.0	2008	Eng
13		Pirates of the Caribbean: Dead Man's Chest	225000000.0	2006	Eng
14		The Lone Ranger	215000000.0	2013	Eng
15		Man of Steel	225000000.0	2013	Eng
16		The Chronicles of Narnia: Prince Caspian	225000000.0	2008	Eng
17		The Avengers	220000000.0	2012	Eng
18		Pirates of the Caribbean: On Stranger Tides	250000000.0	2011	Eng
19		Men in Black 3	225000000.0	2012	Eng
20		The Hobbit: The Battle of the Five Armies	250000000.0	2014	Eng
21		The Amazing Spider-Man	230000000.0	2012	Eng
22		Robin Hood	200000000.0	2010	Eng
23		The Hobbit: The Desolation of Smaug	225000000.0	2013	Eng
24		The Golden Compass	180000000.0	2007	Eng
25		King Kong	207000000.0	2005	Eng
26		Titanic	200000000.0	1997	Eng
27		Captain America: Civil War	250000000.0	2016	Eng
28		Battleship	209000000.0	2012	Eng
29		Jurassic World	150000000.0	2015	Eng
30		Skyfall	200000000.0	2012	Eng
31		Spider-Man 2	200000000.0	2004	Eng
32		Iron Man 3	200000000.0	2013	Eng
33		Alice in Wonderland	200000000.0	2010	Eng
34		X-Men: The Last Stand	210000000.0	2006	Eng
35		Monsters University	200000000.0	2013	Eng
36		Transformers: Revenge of the Fallen	200000000.0	2009	Eng
37		Transformers: Age of Extinction	210000000.0	2014	Eng
38		Oz the Great and Powerful	215000000.0	2013	Eng
39		The Amazing Spider-Man 2	200000000.0	2014	Eng
40		TRON: Legacy	170000000.0	2010	Eng
41		Cars 2	200000000.0	2011	Eng
42		Green Lantern	200000000.0	2011	Eng
43		Toy Story 3	200000000.0	2010	Eng
44		Terminator Salvation	200000000.0	2009	Eng
45		Furious 7	190000000.0	2015	Eng
46		World War Z	190000000.0	2013	Eng
47		X-Men: Days of Future Past	200000000.0	2014	Eng
48		Star Trek Into Darkness	190000000.0	2013	Eng
49		Jack the Giant Slayer	195000000.0	2013	Eng

	score	surname	name	taux_rendu	siecle_simple
0	7.9	james	cameron	3.000000e+09	Maintain Siecle
1	7.1	gore	verbinski	4.225352e+09	Rejected Siecle
2	6.8	sam	mendes	3.602941e+09	Rejected Siecle
3	8.5	christopher	nolan	2.941176e+09	Rejected Siecle
5	6.6	andrew	stanton	3.995455e+09	Rejected Siecle
6	6.2	sam	raimi	4.161290e+09	Rejected Siecle
7	7.8	nathan	greno	3.333333e+09	Rejected Siecle
8	7.5	joss	whedon	3.333333e+09	Rejected Siecle
9	7.5	david	yates	3.333333e+09	Maintain Siecle
10	6.9	zack	snyder	3.623188e+09	Rejected Siecle
11	6.1	bryan	singer	3.426230e+09	Rejected Siecle
12	6.7	marc	forster	2.985075e+09	Rejected Siecle
13	7.3	gore	verbinski	3.082192e+09	Rejected Siecle
14	6.5	gore	verbinski	3.307692e+09	Rejected Siecle
15	7.2	zack	snyder	3.125000e+09	Rejected Siecle
16	6.6	andrew	adamson	3.409091e+09	Rejected Siecle
17	8.1	joss	whedon	2.716049e+09	Rejected Siecle
18	6.7	rob	marshall	3.731343e+09	Rejected Siecle
19	6.8	barry	sonnenfeld	3.308824e+09	Rejected Siecle
20	7.5	peter	jackson	3.333333e+09	Rejected Siecle
21	7.0	marc	webb	3.285714e+09	Rejected Siecle
22	6.7	ridley	scott	2.985075e+09	Rejected Siecle
23	7.9	peter	jackson	2.848101e+09	Rejected Siecle
24	6.1	chris	weitz	2.950820e+09	Rejected Siecle
25	7.2	peter	jackson	2.875000e+09	Rejected Siecle
26	7.7	james	cameron	2.597403e+09	Rejected Siecle
27	8.2	anthony	russo	3.048780e+09	Rejected Siecle
28	5.9	peter	berg	3.542373e+09	Rejected Siecle
29	7.0	colin	trevorrow	2.142857e+09	Rejected Siecle
30	7.8	sam	mendes	2.564103e+09	Rejected Siecle
31	7.3	sam	raimi	2.739726e+09	Rejected Siecle
32	7.2	shane	black	2.777778e+09	Rejected Siecle
33	6.5	tim	burton	3.076923e+09	Rejected Siecle
34	6.8	brett	ratner	3.088235e+09	Rejected Siecle
35	7.3	dan	scanlon	2.739726e+09	Rejected Siecle
36	6.0	michael	bay	3.333333e+09	Maintain Siecle
37	5.7	michael	bay	3.684211e+09	Rejected Siecle
38	6.4	sam	raimi	3.359375e+09	Rejected Siecle
39	6.7	marc	webb	2.985075e+09	Rejected Siecle
40	6.8	joseph	kosinski	2.500000e+09	Rejected Siecle
41	6.3	john	lasseter	3.174603e+09	Rejected Siecle
42	5.6	martin	campbell	3.571429e+09	Rejected Siecle
43	8.3	lee	unkrich	2.409639e+09	Rejected Siecle
44	6.6	mcg	NaN	3.030303e+09	Maintain Siecle



45	7.2	james	wan	2.638889e+09	Rejected Siecle
46	7.0	marc	forster	2.714286e+09	Rejected Siecle
47	8.0	bryan	singer	2.500000e+09	Rejected Siecle
48	7.8	j.j.	abrams	2.435897e+09	Rejected Siecle
49	6.3	bryan	singer	3.095238e+09	Rejected Siecle
50	7.3	baz	luhrmann	1.438356e+09	Rejected Siecle

[71]: *#utilisation de For*

```
movies['sicle_for'] = ['Maintain value' if year == 2009 else 'Rejected value']
↳for year in movies['year']]
#check
movies.head(50)
```

[71]:

	director_name	duration	gross	\
0	JAMES CAMERON	178.0	760505847.0	
1	GORE VERBINSKI	169.0	309404152.0	
2	SAM MENDES	148.0	200074175.0	
3	CHRISTOPHER NOLAN	164.0	448130642.0	
5	ANDREW STANTON	132.0	73058679.0	
6	SAM RAIMI	156.0	336530303.0	
7	NATHAN GRENO	100.0	200807262.0	
8	JOSS WHEDON	141.0	458991599.0	
9	DAVID YATES	153.0	301956980.0	
10	ZACK SNYDER	183.0	330249062.0	
11	BRYAN SINGER	169.0	200069408.0	
12	MARC FORSTER	106.0	168368427.0	
13	GORE VERBINSKI	151.0	423032628.0	
14	GORE VERBINSKI	150.0	89289910.0	
15	ZACK SNYDER	143.0	291021565.0	
16	ANDREW ADAMSON	150.0	141614023.0	
17	JOSS WHEDON	173.0	623279547.0	
18	ROB MARSHALL	136.0	241063875.0	
19	BARRY SONNENFELD	106.0	179020854.0	
20	PETER JACKSON	164.0	255108370.0	
21	MARC WEBB	153.0	262030663.0	
22	RIDLEY SCOTT	156.0	105219735.0	
23	PETER JACKSON	186.0	258355354.0	
24	CHRIS WEITZ	113.0	70083519.0	
25	PETER JACKSON	201.0	218051260.0	
26	JAMES CAMERON	194.0	658672302.0	
27	ANTHONY RUSSO	147.0	407197282.0	
28	PETER BERG	131.0	65173160.0	
29	COLIN TREVORROW	124.0	652177271.0	
30	SAM MENDES	143.0	304360277.0	
31	SAM RAIMI	135.0	373377893.0	
32	SHANE BLACK	195.0	408992272.0	

33	TIM BURTON	108.0	334185206.0
34	BRETT RATNER	104.0	234360014.0
35	DAN SCANLON	104.0	268488329.0
36	MICHAEL BAY	150.0	402076689.0
37	MICHAEL BAY	165.0	245428137.0
38	SAM RAIMI	130.0	234903076.0
39	MARC WEBB	142.0	202853933.0
40	JOSEPH KOSINSKI	125.0	172051787.0
41	JOHN LASSETER	106.0	191450875.0
42	MARTIN CAMPBELL	123.0	116593191.0
43	LEE UNKRICH	103.0	414984497.0
44	MCG	118.0	125320003.0
45	JAMES WAN	140.0	350034110.0
46	MARC FORSTER	123.0	202351611.0
47	BRYAN SINGER	149.0	233914986.0
48	J.J. ABRAMS	132.0	228756232.0
49	BRYAN SINGER	114.0	65171860.0
50	BAZ LUHRMANN	143.0	144812796.0

	genres \
0	Action Adventure Fantasy Sci-Fi
1	Action Adventure Fantasy
2	Action Adventure Thriller
3	Action Thriller
5	Action Adventure Sci-Fi
6	Action Adventure Romance
7	Adventure Animation Comedy Family Fantasy Musi...
8	Action Adventure Sci-Fi
9	Adventure Family Fantasy Mystery
10	Action Adventure Sci-Fi
11	Action Adventure Sci-Fi
12	Action Adventure
13	Action Adventure Fantasy
14	Action Adventure Western
15	Action Adventure Fantasy Sci-Fi
16	Action Adventure Family Fantasy
17	Action Adventure Sci-Fi
18	Action Adventure Fantasy
19	Action Adventure Comedy Family Fantasy Sci-Fi
20	Adventure Fantasy
21	Action Adventure Fantasy
22	Action Adventure Drama History
23	Adventure Fantasy
24	Adventure Family Fantasy
25	Action Adventure Drama Romance
26	Drama Romance
27	Action Adventure Sci-Fi

28	Action Adventure Sci-Fi Thriller
29	Action Adventure Sci-Fi Thriller
30	Action Adventure Thriller
31	Action Adventure Fantasy Romance
32	Action Adventure Sci-Fi
33	Adventure Family Fantasy
34	Action Adventure Fantasy Sci-Fi Thriller
35	Adventure Animation Comedy Family Fantasy
36	Action Adventure Sci-Fi
37	Action Adventure Sci-Fi
38	Adventure Family Fantasy
39	Action Adventure Fantasy Sci-Fi
40	Action Adventure Sci-Fi
41	Adventure Animation Comedy Family Sport
42	Action Adventure Sci-Fi
43	Adventure Animation Comedy Family Fantasy
44	Action Adventure Sci-Fi
45	Action Crime Thriller
46	Action Adventure Horror Sci-Fi Thriller
47	Action Adventure Fantasy Sci-Fi Thriller
48	Action Adventure Sci-Fi
49	Adventure Fantasy
50	Drama Romance

	title	budget	year	language	\
0	Avatar	237000000.0	2009	Eng	
1	Pirates of the Caribbean: At World's End	300000000.0	2007	Eng	
2	Spectre	245000000.0	2015	Eng	
3	The Dark Knight Rises	250000000.0	2012	Eng	
5	John Carter	263700000.0	2012	Eng	
6	Spider-Man 3	258000000.0	2007	Eng	
7	Tangled	260000000.0	2010	Eng	
8	Avengers: Age of Ultron	250000000.0	2015	Eng	
9	Harry Potter and the Half-Blood Prince	250000000.0	2009	Eng	
10	Batman v Superman: Dawn of Justice	250000000.0	2016	Eng	
11	Superman Returns	209000000.0	2006	Eng	
12	Quantum of Solace	200000000.0	2008	Eng	
13	Pirates of the Caribbean: Dead Man's Chest	225000000.0	2006	Eng	
14	The Lone Ranger	215000000.0	2013	Eng	
15	Man of Steel	225000000.0	2013	Eng	
16	The Chronicles of Narnia: Prince Caspian	225000000.0	2008	Eng	
17	The Avengers	220000000.0	2012	Eng	
18	Pirates of the Caribbean: On Stranger Tides	250000000.0	2011	Eng	
19	Men in Black 3	225000000.0	2012	Eng	
20	The Hobbit: The Battle of the Five Armies	250000000.0	2014	Eng	
21	The Amazing Spider-Man	230000000.0	2012	Eng	
22	Robin Hood	200000000.0	2010	Eng	

23	The Hobbit: The Desolation of Smaug	225000000.0	2013	Eng
24	The Golden Compass	180000000.0	2007	Eng
25	King Kong	207000000.0	2005	Eng
26	Titanic	200000000.0	1997	Eng
27	Captain America: Civil War	250000000.0	2016	Eng
28	Battleship	209000000.0	2012	Eng
29	Jurassic World	150000000.0	2015	Eng
30	Skyfall	200000000.0	2012	Eng
31	Spider-Man 2	200000000.0	2004	Eng
32	Iron Man 3	200000000.0	2013	Eng
33	Alice in Wonderland	200000000.0	2010	Eng
34	X-Men: The Last Stand	210000000.0	2006	Eng
35	Monsters University	200000000.0	2013	Eng
36	Transformers: Revenge of the Fallen	200000000.0	2009	Eng
37	Transformers: Age of Extinction	210000000.0	2014	Eng
38	Oz the Great and Powerful	215000000.0	2013	Eng
39	The Amazing Spider-Man 2	200000000.0	2014	Eng
40	TRON: Legacy	170000000.0	2010	Eng
41	Cars 2	200000000.0	2011	Eng
42	Green Lantern	200000000.0	2011	Eng
43	Toy Story 3	200000000.0	2010	Eng
44	Terminator Salvation	200000000.0	2009	Eng
45	Furious 7	190000000.0	2015	Eng
46	World War Z	190000000.0	2013	Eng
47	X-Men: Days of Future Past	200000000.0	2014	Eng
48	Star Trek Into Darkness	190000000.0	2013	Eng
49	Jack the Giant Slayer	195000000.0	2013	Eng
50	The Great Gatsby	105000000.0	2013	Eng

	score	surname	surnname	name	taux_rendu	\
0	7.9	JAMES	[JAMES, CAMERON]	CAMERON	3.000000e+09	
1	7.1	GORE	[GORE, VERBINSKI]	VERBINSKI	4.225352e+09	
2	6.8	SAM	[SAM, MENDES]	MENDES	3.602941e+09	
3	8.5	CHRISTOPHER	[CHRISTOPHER, NOLAN]	NOLAN	2.941176e+09	
5	6.6	ANDREW	[ANDREW, STANTON]	STANTON	3.995455e+09	
6	6.2	SAM	[SAM, RAIMI]	RAIMI	4.161290e+09	
7	7.8	NATHAN	[NATHAN, GRENO]	GRENO	3.333333e+09	
8	7.5	JOSS	[JOSS, WHEDON]	WHEDON	3.333333e+09	
9	7.5	DAVID	[DAVID, YATES]	YATES	3.333333e+09	
10	6.9	ZACK	[ZACK, SNYDER]	SNYDER	3.623188e+09	
11	6.1	BRYAN	[BRYAN, SINGER]	SINGER	3.426230e+09	
12	6.7	MARC	[MARC, FORSTER]	FORSTER	2.985075e+09	
13	7.3	GORE	[GORE, VERBINSKI]	VERBINSKI	3.082192e+09	
14	6.5	GORE	[GORE, VERBINSKI]	VERBINSKI	3.307692e+09	
15	7.2	ZACK	[ZACK, SNYDER]	SNYDER	3.125000e+09	
16	6.6	ANDREW	[ANDREW, ADAMSON]	ADAMSON	3.409091e+09	
17	8.1	JOSS	[JOSS, WHEDON]	WHEDON	2.716049e+09	

18	6.7	ROB	[ROB, MARSHALL]	MARSHALL	3.731343e+09
19	6.8	BARRY	[BARRY, SONNENFELD]	SONNENFELD	3.308824e+09
20	7.5	PETER	[PETER, JACKSON]	JACKSON	3.333333e+09
21	7.0	MARC	[MARC, WEBB]	WEBB	3.285714e+09
22	6.7	RIDLEY	[RIDLEY, SCOTT]	SCOTT	2.985075e+09
23	7.9	PETER	[PETER, JACKSON]	JACKSON	2.848101e+09
24	6.1	CHRIS	[CHRIS, WEITZ]	WEITZ	2.950820e+09
25	7.2	PETER	[PETER, JACKSON]	JACKSON	2.875000e+09
26	7.7	JAMES	[JAMES, CAMERON]	CAMERON	2.597403e+09
27	8.2	ANTHONY	[ANTHONY, RUSSO]	RUSSO	3.048780e+09
28	5.9	PETER	[PETER, BERG]	BERG	3.542373e+09
29	7.0	COLIN	[COLIN, TREVORROW]	TREVORROW	2.142857e+09
30	7.8	SAM	[SAM, MENDES]	MENDES	2.564103e+09
31	7.3	SAM	[SAM, RAIMI]	RAIMI	2.739726e+09
32	7.2	SHANE	[SHANE, BLACK]	BLACK	2.777778e+09
33	6.5	TIM	[TIM, BURTON]	BURTON	3.076923e+09
34	6.8	BRETT	[BRETT, RATNER]	RATNER	3.088235e+09
35	7.3	DAN	[DAN, SCANLON]	SCANLON	2.739726e+09
36	6.0	MICHAEL	[MICHAEL, BAY]	BAY	3.333333e+09
37	5.7	MICHAEL	[MICHAEL, BAY]	BAY	3.684211e+09
38	6.4	SAM	[SAM, RAIMI]	RAIMI	3.359375e+09
39	6.7	MARC	[MARC, WEBB]	WEBB	2.985075e+09
40	6.8	JOSEPH	[JOSEPH, KOSINSKI]	KOSINSKI	2.500000e+09
41	6.3	JOHN	[JOHN, LASSETER]	LASSETER	3.174603e+09
42	5.6	MARTIN	[MARTIN, CAMPBELL]	CAMPBELL	3.571429e+09
43	8.3	LEE	[LEE, UNKRICH]	UNKRICH	2.409639e+09
44	6.6	MCG	[MCG]	NaN	3.030303e+09
45	7.2	JAMES	[JAMES, WAN]	WAN	2.638889e+09
46	7.0	MARC	[MARC, FORSTER]	FORSTER	2.714286e+09
47	8.0	BRYAN	[BRYAN, SINGER]	SINGER	2.500000e+09
48	7.8	J.J.	[J.J., ABRAMS]	ABRAMS	2.435897e+09
49	6.3	BRYAN	[BRYAN, SINGER]	SINGER	3.095238e+09
50	7.3	BAZ	[BAZ, LUHRMANN]	LUHRMANN	1.438356e+09

	siecle_simple	siecle_for
0	Maintain Siecle	Maintain value
1	Rejected Siecle	Rejected value
2	Rejected Siecle	Rejected value
3	Rejected Siecle	Rejected value
5	Rejected Siecle	Rejected value
6	Rejected Siecle	Rejected value
7	Rejected Siecle	Rejected value
8	Rejected Siecle	Rejected value
9	Maintain Siecle	Maintain value
10	Rejected Siecle	Rejected value
11	Rejected Siecle	Rejected value
12	Rejected Siecle	Rejected value

13	Rejected	Siecle	Rejected value
14	Rejected	Siecle	Rejected value
15	Rejected	Siecle	Rejected value
16	Rejected	Siecle	Rejected value
17	Rejected	Siecle	Rejected value
18	Rejected	Siecle	Rejected value
19	Rejected	Siecle	Rejected value
20	Rejected	Siecle	Rejected value
21	Rejected	Siecle	Rejected value
22	Rejected	Siecle	Rejected value
23	Rejected	Siecle	Rejected value
24	Rejected	Siecle	Rejected value
25	Rejected	Siecle	Rejected value
26	Rejected	Siecle	Rejected value
27	Rejected	Siecle	Rejected value
28	Rejected	Siecle	Rejected value
29	Rejected	Siecle	Rejected value
30	Rejected	Siecle	Rejected value
31	Rejected	Siecle	Rejected value
32	Rejected	Siecle	Rejected value
33	Rejected	Siecle	Rejected value
34	Rejected	Siecle	Rejected value
35	Rejected	Siecle	Rejected value
36	Maintain	Siecle	Maintain value
37	Rejected	Siecle	Rejected value
38	Rejected	Siecle	Rejected value
39	Rejected	Siecle	Rejected value
40	Rejected	Siecle	Rejected value
41	Rejected	Siecle	Rejected value
42	Rejected	Siecle	Rejected value
43	Rejected	Siecle	Rejected value
44	Maintain	Siecle	Maintain value
45	Rejected	Siecle	Rejected value
46	Rejected	Siecle	Rejected value
47	Rejected	Siecle	Rejected value
48	Rejected	Siecle	Rejected value
49	Rejected	Siecle	Rejected value
50	Rejected	Siecle	Rejected value

colonne conditionnelle avec np.select()

[21]: *#mettre en place les conditions*

*#check a valeur maximale*  
*#movies['year'].max()*

```
conditions = [
    (movies['year'] >= 2005) & (movies['year'] <=2007),
```

```

(movies['year'] >= 2008) & (movies['year'] <= 2010),
(movies['year'] >= 2011) & (movies['year'] <= 2013),
(movies['year'] >= 2014) & (movies['year'] <= 2016),
(movies['year'] > 2016)

]
#mettre en place les choix correspondants aux conditions

choices = ['Between 2005 and 2007', '2008-2010', '2011-2013', '2014-2016', 'more_
↳than 2016' ]

#mettre en place la nulle colonne
movies['complexe_choix'] = np.select(conditions, choices, 'na')

#check
movies.head(50)

#les conditions peuvent se faire sur plusieurs colonnes, 2 ou plus et on peut_
↳utiliser le ou /

#conditions = [
# (movies['year'] == 2009) & (movies['score'] < 7.5),
# (movies['year'] == 2010) & (movies['score'] > 10)

#]

#choices = ['level1', 'level2']

#movies['new'] = np.select(conditions, choices, 'NA')

```

```

[21]:
      director_name  duration  gross \
0      james cameron    178.0  760505847.0
1      gore verbinski    169.0  309404152.0
2          sam mendes    148.0  200074175.0
3 christopher nolan    164.0  448130642.0
5      andrew stanton    132.0   73058679.0
6          sam raimi    156.0  336530303.0
7      nathan greno    100.0  200807262.0
8       joss whedon    141.0  458991599.0
9       david yates    153.0  301956980.0
10       zack snyder    183.0  330249062.0
11      bryan singer    169.0  200069408.0
12      marc forster    106.0  168368427.0
13      gore verbinski    151.0  423032628.0
14      gore verbinski    150.0   89289910.0
15       zack snyder    143.0  291021565.0

```

16	andrew adamson	150.0	141614023.0
17	joss whedon	173.0	623279547.0
18	rob marshall	136.0	241063875.0
19	barry sonnenfeld	106.0	179020854.0
20	peter jackson	164.0	255108370.0
21	marc webb	153.0	262030663.0
22	ridley scott	156.0	105219735.0
23	peter jackson	186.0	258355354.0
24	chris weitz	113.0	70083519.0
25	peter jackson	201.0	218051260.0
26	james cameron	194.0	658672302.0
27	anthony russo	147.0	407197282.0
28	peter berg	131.0	65173160.0
29	colin trevorrow	124.0	652177271.0
30	sam mendes	143.0	304360277.0
31	sam raimi	135.0	373377893.0
32	shane black	195.0	408992272.0
33	tim burton	108.0	334185206.0
34	brett ratner	104.0	234360014.0
35	dan scanlon	104.0	268488329.0
36	michael bay	150.0	402076689.0
37	michael bay	165.0	245428137.0
38	sam raimi	130.0	234903076.0
39	marc webb	142.0	202853933.0
40	joseph kosinski	125.0	172051787.0
41	john lasseter	106.0	191450875.0
42	martin campbell	123.0	116593191.0
43	lee unkrich	103.0	414984497.0
44	mcg	118.0	125320003.0
45	james wan	140.0	350034110.0
46	marc forster	123.0	202351611.0
47	bryan singer	149.0	233914986.0
48	j.j. abrams	132.0	228756232.0
49	bryan singer	114.0	65171860.0
50	baz luhrmann	143.0	144812796.0

	genres \
0	Action Adventure Fantasy Sci-Fi
1	Action Adventure Fantasy
2	Action Adventure Thriller
3	Action Thriller
5	Action Adventure Sci-Fi
6	Action Adventure Romance
7	Adventure Animation Comedy Family Fantasy Musi...
8	Action Adventure Sci-Fi
9	Adventure Family Fantasy Mystery
10	Action Adventure Sci-Fi



11 Action|Adventure|Sci-Fi  
 12 Action|Adventure  
 13 Action|Adventure|Fantasy  
 14 Action|Adventure|Western  
 15 Action|Adventure|Fantasy|Sci-Fi  
 16 Action|Adventure|Family|Fantasy  
 17 Action|Adventure|Sci-Fi  
 18 Action|Adventure|Fantasy  
 19 Action|Adventure|Comedy|Family|Fantasy|Sci-Fi  
 20 Adventure|Fantasy  
 21 Action|Adventure|Fantasy  
 22 Action|Adventure|Drama|History  
 23 Adventure|Fantasy  
 24 Adventure|Family|Fantasy  
 25 Action|Adventure|Drama|Romance  
 26 Drama|Romance  
 27 Action|Adventure|Sci-Fi  
 28 Action|Adventure|Sci-Fi|Thriller  
 29 Action|Adventure|Sci-Fi|Thriller  
 30 Action|Adventure|Thriller  
 31 Action|Adventure|Fantasy|Romance  
 32 Action|Adventure|Sci-Fi  
 33 Adventure|Family|Fantasy  
 34 Action|Adventure|Fantasy|Sci-Fi|Thriller  
 35 Adventure|Animation|Comedy|Family|Fantasy  
 36 Action|Adventure|Sci-Fi  
 37 Action|Adventure|Sci-Fi  
 38 Adventure|Family|Fantasy  
 39 Action|Adventure|Fantasy|Sci-Fi  
 40 Action|Adventure|Sci-Fi  
 41 Adventure|Animation|Comedy|Family|Sport  
 42 Action|Adventure|Sci-Fi  
 43 Adventure|Animation|Comedy|Family|Fantasy  
 44 Action|Adventure|Sci-Fi  
 45 Action|Crime|Thriller  
 46 Action|Adventure|Horror|Sci-Fi|Thriller  
 47 Action|Adventure|Fantasy|Sci-Fi|Thriller  
 48 Action|Adventure|Sci-Fi  
 49 Adventure|Fantasy  
 50 Drama|Romance

	title	budget	year	language	\
0	Avatar	237000000.0	2009	Eng	
1	Pirates of the Caribbean: At World's End	300000000.0	2007	Eng	
2	Spectre	245000000.0	2015	Eng	
3	The Dark Knight Rises	250000000.0	2012	Eng	
5	John Carter	263700000.0	2012	Eng	

6		Spider-Man 3	258000000.0	2007	Eng
7		Tangled	260000000.0	2010	Eng
8		Avengers: Age of Ultron	250000000.0	2015	Eng
9		Harry Potter and the Half-Blood Prince	250000000.0	2009	Eng
10		Batman v Superman: Dawn of Justice	250000000.0	2016	Eng
11		Superman Returns	209000000.0	2006	Eng
12		Quantum of Solace	200000000.0	2008	Eng
13		Pirates of the Caribbean: Dead Man's Chest	225000000.0	2006	Eng
14		The Lone Ranger	215000000.0	2013	Eng
15		Man of Steel	225000000.0	2013	Eng
16		The Chronicles of Narnia: Prince Caspian	225000000.0	2008	Eng
17		The Avengers	220000000.0	2012	Eng
18		Pirates of the Caribbean: On Stranger Tides	250000000.0	2011	Eng
19		Men in Black 3	225000000.0	2012	Eng
20		The Hobbit: The Battle of the Five Armies	250000000.0	2014	Eng
21		The Amazing Spider-Man	230000000.0	2012	Eng
22		Robin Hood	200000000.0	2010	Eng
23		The Hobbit: The Desolation of Smaug	225000000.0	2013	Eng
24		The Golden Compass	180000000.0	2007	Eng
25		King Kong	207000000.0	2005	Eng
26		Titanic	200000000.0	1997	Eng
27		Captain America: Civil War	250000000.0	2016	Eng
28		Battleship	209000000.0	2012	Eng
29		Jurassic World	150000000.0	2015	Eng
30		Skyfall	200000000.0	2012	Eng
31		Spider-Man 2	200000000.0	2004	Eng
32		Iron Man 3	200000000.0	2013	Eng
33		Alice in Wonderland	200000000.0	2010	Eng
34		X-Men: The Last Stand	210000000.0	2006	Eng
35		Monsters University	200000000.0	2013	Eng
36		Transformers: Revenge of the Fallen	200000000.0	2009	Eng
37		Transformers: Age of Extinction	210000000.0	2014	Eng
38		Oz the Great and Powerful	215000000.0	2013	Eng
39		The Amazing Spider-Man 2	200000000.0	2014	Eng
40		TRON: Legacy	170000000.0	2010	Eng
41		Cars 2	200000000.0	2011	Eng
42		Green Lantern	200000000.0	2011	Eng
43		Toy Story 3	200000000.0	2010	Eng
44		Terminator Salvation	200000000.0	2009	Eng
45		Furious 7	190000000.0	2015	Eng
46		World War Z	190000000.0	2013	Eng
47		X-Men: Days of Future Past	200000000.0	2014	Eng
48		Star Trek Into Darkness	190000000.0	2013	Eng
49		Jack the Giant Slayer	195000000.0	2013	Eng
50		The Great Gatsby	105000000.0	2013	Eng

score      surname      name      taux\_rendu      siecle\_simple \

0	7.9	james	cameron	3.000000e+09	Maintain	Siecle
1	7.1	gore	verbinski	4.225352e+09	Rejected	Siecle
2	6.8	sam	mendes	3.602941e+09	Rejected	Siecle
3	8.5	christopher	nolan	2.941176e+09	Rejected	Siecle
5	6.6	andrew	stanton	3.995455e+09	Rejected	Siecle
6	6.2	sam	raimi	4.161290e+09	Rejected	Siecle
7	7.8	nathan	greno	3.333333e+09	Rejected	Siecle
8	7.5	joss	whedon	3.333333e+09	Rejected	Siecle
9	7.5	david	yates	3.333333e+09	Maintain	Siecle
10	6.9	zack	snyder	3.623188e+09	Rejected	Siecle
11	6.1	bryan	singer	3.426230e+09	Rejected	Siecle
12	6.7	marc	forster	2.985075e+09	Rejected	Siecle
13	7.3	gore	verbinski	3.082192e+09	Rejected	Siecle
14	6.5	gore	verbinski	3.307692e+09	Rejected	Siecle
15	7.2	zack	snyder	3.125000e+09	Rejected	Siecle
16	6.6	andrew	adamson	3.409091e+09	Rejected	Siecle
17	8.1	joss	whedon	2.716049e+09	Rejected	Siecle
18	6.7	rob	marshall	3.731343e+09	Rejected	Siecle
19	6.8	barry	sonnenfeld	3.308824e+09	Rejected	Siecle
20	7.5	peter	jackson	3.333333e+09	Rejected	Siecle
21	7.0	marc	webb	3.285714e+09	Rejected	Siecle
22	6.7	ridley	scott	2.985075e+09	Rejected	Siecle
23	7.9	peter	jackson	2.848101e+09	Rejected	Siecle
24	6.1	chris	weitz	2.950820e+09	Rejected	Siecle
25	7.2	peter	jackson	2.875000e+09	Rejected	Siecle
26	7.7	james	cameron	2.597403e+09	Rejected	Siecle
27	8.2	anthony	russo	3.048780e+09	Rejected	Siecle
28	5.9	peter	berg	3.542373e+09	Rejected	Siecle
29	7.0	colin	trevorrow	2.142857e+09	Rejected	Siecle
30	7.8	sam	mendes	2.564103e+09	Rejected	Siecle
31	7.3	sam	raimi	2.739726e+09	Rejected	Siecle
32	7.2	shane	black	2.777778e+09	Rejected	Siecle
33	6.5	tim	burton	3.076923e+09	Rejected	Siecle
34	6.8	brett	ratner	3.088235e+09	Rejected	Siecle
35	7.3	dan	scanlon	2.739726e+09	Rejected	Siecle
36	6.0	michael	bay	3.333333e+09	Maintain	Siecle
37	5.7	michael	bay	3.684211e+09	Rejected	Siecle
38	6.4	sam	raimi	3.359375e+09	Rejected	Siecle
39	6.7	marc	webb	2.985075e+09	Rejected	Siecle
40	6.8	joseph	kosinski	2.500000e+09	Rejected	Siecle
41	6.3	john	lasseter	3.174603e+09	Rejected	Siecle
42	5.6	martin	campbell	3.571429e+09	Rejected	Siecle
43	8.3	lee	unkrich	2.409639e+09	Rejected	Siecle
44	6.6	mcg	NaN	3.030303e+09	Maintain	Siecle
45	7.2	james	wan	2.638889e+09	Rejected	Siecle
46	7.0	marc	forster	2.714286e+09	Rejected	Siecle
47	8.0	bryan	singer	2.500000e+09	Rejected	Siecle

48	7.8	j.j.	abrams	2.435897e+09	Rejected Siecle
49	6.3	bryan	singer	3.095238e+09	Rejected Siecle
50	7.3	baz	luhrmann	1.438356e+09	Rejected Siecle

	complexe_choix
0	2008-2010
1	Between 2055 and 2007
2	2014-2016
3	2011-2013
5	2011-2013
6	Between 2055 and 2007
7	2008-2010
8	2014-2016
9	2008-2010
10	2014-2016
11	Between 2055 and 2007
12	2008-2010
13	Between 2055 and 2007
14	2011-2013
15	2011-2013
16	2008-2010
17	2011-2013
18	2011-2013
19	2011-2013
20	2014-2016
21	2011-2013
22	2008-2010
23	2011-2013
24	Between 2055 and 2007
25	Between 2055 and 2007
26	na
27	2014-2016
28	2011-2013
29	2014-2016
30	2011-2013
31	na
32	2011-2013
33	2008-2010
34	Between 2055 and 2007
35	2011-2013
36	2008-2010
37	2014-2016
38	2011-2013
39	2014-2016
40	2008-2010
41	2011-2013
42	2011-2013

43	2008-2010
44	2008-2010
45	2014-2016
46	2011-2013
47	2014-2016
48	2011-2013
49	2011-2013
50	2011-2013

### colonne conditionnelle avec assign()

```
[22]: #assign() sur une seule colonne et ça peut se faire sur deux ou trois colonnes,
      ↪ il suffit juste de changer le nom de la colonne
      #definition de la fonction
      def oneColumn(row):
          if row["year"] == 2009:
              return "year_2009"
          elif row["year"] == 2010:
              return "year_2010"
          elif row["year"] == 2013:
              return "year_2013"
          else:
              return "VIP"
      #assigne cette fonction a notre dataset

      movies = movies.assign(NewColonne = movies.apply(oneColumn, axis = 1))

      #check
      movies.head(50)
```

```
[22]:      director_name  duration    gross \
0      james cameron    178.0  760505847.0
1      gore verbinski    169.0  309404152.0
2          sam mendes    148.0  200074175.0
3 christopher nolan    164.0  448130642.0
5      andrew stanton    132.0   73058679.0
6          sam raimi    156.0  336530303.0
7      nathan greno    100.0  200807262.0
8      joss whedon    141.0  458991599.0
9      david yates    153.0  301956980.0
10      zack snyder    183.0  330249062.0
11      bryan singer    169.0  200069408.0
12      marc forster    106.0  168368427.0
13      gore verbinski    151.0  423032628.0
14      gore verbinski    150.0   89289910.0
15      zack snyder    143.0  291021565.0
16      andrew adamson    150.0  141614023.0
```

17	joss whedon	173.0	623279547.0
18	rob marshall	136.0	241063875.0
19	barry sonnenfeld	106.0	179020854.0
20	peter jackson	164.0	255108370.0
21	marc webb	153.0	262030663.0
22	ridley scott	156.0	105219735.0
23	peter jackson	186.0	258355354.0
24	chris weitz	113.0	70083519.0
25	peter jackson	201.0	218051260.0
26	james cameron	194.0	658672302.0
27	anthony russo	147.0	407197282.0
28	peter berg	131.0	65173160.0
29	colin trevorrow	124.0	652177271.0
30	sam mendes	143.0	304360277.0
31	sam raimi	135.0	373377893.0
32	shane black	195.0	408992272.0
33	tim burton	108.0	334185206.0
34	brett ratner	104.0	234360014.0
35	dan scanlon	104.0	268488329.0
36	michael bay	150.0	402076689.0
37	michael bay	165.0	245428137.0
38	sam raimi	130.0	234903076.0
39	marc webb	142.0	202853933.0
40	joseph kosinski	125.0	172051787.0
41	john lasseter	106.0	191450875.0
42	martin campbell	123.0	116593191.0
43	lee unkrich	103.0	414984497.0
44	mcg	118.0	125320003.0
45	james wan	140.0	350034110.0
46	marc forster	123.0	202351611.0
47	bryan singer	149.0	233914986.0
48	j.j. abrams	132.0	228756232.0
49	bryan singer	114.0	65171860.0
50	baz luhrmann	143.0	144812796.0

	genres \
0	Action Adventure Fantasy Sci-Fi
1	Action Adventure Fantasy
2	Action Adventure Thriller
3	Action Thriller
5	Action Adventure Sci-Fi
6	Action Adventure Romance
7	Adventure Animation Comedy Family Fantasy Musi...
8	Action Adventure Sci-Fi
9	Adventure Family Fantasy Mystery
10	Action Adventure Sci-Fi
11	Action Adventure Sci-Fi

12 Action|Adventure  
 13 Action|Adventure|Fantasy  
 14 Action|Adventure|Western  
 15 Action|Adventure|Fantasy|Sci-Fi  
 16 Action|Adventure|Family|Fantasy  
 17 Action|Adventure|Sci-Fi  
 18 Action|Adventure|Fantasy  
 19 Action|Adventure|Comedy|Family|Fantasy|Sci-Fi  
 20 Adventure|Fantasy  
 21 Action|Adventure|Fantasy  
 22 Action|Adventure|Drama|History  
 23 Adventure|Fantasy  
 24 Adventure|Family|Fantasy  
 25 Action|Adventure|Drama|Romance  
 26 Drama|Romance  
 27 Action|Adventure|Sci-Fi  
 28 Action|Adventure|Sci-Fi|Thriller  
 29 Action|Adventure|Sci-Fi|Thriller  
 30 Action|Adventure|Thriller  
 31 Action|Adventure|Fantasy|Romance  
 32 Action|Adventure|Sci-Fi  
 33 Adventure|Family|Fantasy  
 34 Action|Adventure|Fantasy|Sci-Fi|Thriller  
 35 Adventure|Animation|Comedy|Family|Fantasy  
 36 Action|Adventure|Sci-Fi  
 37 Action|Adventure|Sci-Fi  
 38 Adventure|Family|Fantasy  
 39 Action|Adventure|Fantasy|Sci-Fi  
 40 Action|Adventure|Sci-Fi  
 41 Adventure|Animation|Comedy|Family|Sport  
 42 Action|Adventure|Sci-Fi  
 43 Adventure|Animation|Comedy|Family|Fantasy  
 44 Action|Adventure|Sci-Fi  
 45 Action|Crime|Thriller  
 46 Action|Adventure|Horror|Sci-Fi|Thriller  
 47 Action|Adventure|Fantasy|Sci-Fi|Thriller  
 48 Action|Adventure|Sci-Fi  
 49 Adventure|Fantasy  
 50 Drama|Romance

	title	budget	year	language	\
0	Avatar	237000000.0	2009	Eng	
1	Pirates of the Caribbean: At World's End	300000000.0	2007	Eng	
2	Spectre	245000000.0	2015	Eng	
3	The Dark Knight Rises	250000000.0	2012	Eng	
5	John Carter	263700000.0	2012	Eng	
6	Spider-Man 3	258000000.0	2007	Eng	

7		Tangled	260000000.0	2010	Eng
8		Avengers: Age of Ultron	250000000.0	2015	Eng
9		Harry Potter and the Half-Blood Prince	250000000.0	2009	Eng
10		Batman v Superman: Dawn of Justice	250000000.0	2016	Eng
11		Superman Returns	209000000.0	2006	Eng
12		Quantum of Solace	200000000.0	2008	Eng
13		Pirates of the Caribbean: Dead Man's Chest	225000000.0	2006	Eng
14		The Lone Ranger	215000000.0	2013	Eng
15		Man of Steel	225000000.0	2013	Eng
16		The Chronicles of Narnia: Prince Caspian	225000000.0	2008	Eng
17		The Avengers	220000000.0	2012	Eng
18		Pirates of the Caribbean: On Stranger Tides	250000000.0	2011	Eng
19		Men in Black 3	225000000.0	2012	Eng
20		The Hobbit: The Battle of the Five Armies	250000000.0	2014	Eng
21		The Amazing Spider-Man	230000000.0	2012	Eng
22		Robin Hood	200000000.0	2010	Eng
23		The Hobbit: The Desolation of Smaug	225000000.0	2013	Eng
24		The Golden Compass	180000000.0	2007	Eng
25		King Kong	207000000.0	2005	Eng
26		Titanic	200000000.0	1997	Eng
27		Captain America: Civil War	250000000.0	2016	Eng
28		Battleship	209000000.0	2012	Eng
29		Jurassic World	150000000.0	2015	Eng
30		Skyfall	200000000.0	2012	Eng
31		Spider-Man 2	200000000.0	2004	Eng
32		Iron Man 3	200000000.0	2013	Eng
33		Alice in Wonderland	200000000.0	2010	Eng
34		X-Men: The Last Stand	210000000.0	2006	Eng
35		Monsters University	200000000.0	2013	Eng
36		Transformers: Revenge of the Fallen	200000000.0	2009	Eng
37		Transformers: Age of Extinction	210000000.0	2014	Eng
38		Oz the Great and Powerful	215000000.0	2013	Eng
39		The Amazing Spider-Man 2	200000000.0	2014	Eng
40		TRON: Legacy	170000000.0	2010	Eng
41		Cars 2	200000000.0	2011	Eng
42		Green Lantern	200000000.0	2011	Eng
43		Toy Story 3	200000000.0	2010	Eng
44		Terminator Salvation	200000000.0	2009	Eng
45		Furious 7	190000000.0	2015	Eng
46		World War Z	190000000.0	2013	Eng
47		X-Men: Days of Future Past	200000000.0	2014	Eng
48		Star Trek Into Darkness	190000000.0	2013	Eng
49		Jack the Giant Slayer	195000000.0	2013	Eng
50		The Great Gatsby	105000000.0	2013	Eng

	score	surname	name	taux_rendu	siecle_simple \
0	7.9	james	cameron	3.000000e+09	Maintain Siecle



1	7.1	gore	verbinski	4.225352e+09	Rejected	Siecle
2	6.8	sam	mendes	3.602941e+09	Rejected	Siecle
3	8.5	christopher	nolan	2.941176e+09	Rejected	Siecle
5	6.6	andrew	stanton	3.995455e+09	Rejected	Siecle
6	6.2	sam	raimi	4.161290e+09	Rejected	Siecle
7	7.8	nathan	greno	3.333333e+09	Rejected	Siecle
8	7.5	joss	whedon	3.333333e+09	Rejected	Siecle
9	7.5	david	yates	3.333333e+09	Maintain	Siecle
10	6.9	zack	snyder	3.623188e+09	Rejected	Siecle
11	6.1	bryan	singer	3.426230e+09	Rejected	Siecle
12	6.7	marc	forster	2.985075e+09	Rejected	Siecle
13	7.3	gore	verbinski	3.082192e+09	Rejected	Siecle
14	6.5	gore	verbinski	3.307692e+09	Rejected	Siecle
15	7.2	zack	snyder	3.125000e+09	Rejected	Siecle
16	6.6	andrew	adamson	3.409091e+09	Rejected	Siecle
17	8.1	joss	whedon	2.716049e+09	Rejected	Siecle
18	6.7	rob	marshall	3.731343e+09	Rejected	Siecle
19	6.8	barry	sonnenfeld	3.308824e+09	Rejected	Siecle
20	7.5	peter	jackson	3.333333e+09	Rejected	Siecle
21	7.0	marc	webb	3.285714e+09	Rejected	Siecle
22	6.7	ridley	scott	2.985075e+09	Rejected	Siecle
23	7.9	peter	jackson	2.848101e+09	Rejected	Siecle
24	6.1	chris	weitz	2.950820e+09	Rejected	Siecle
25	7.2	peter	jackson	2.875000e+09	Rejected	Siecle
26	7.7	james	cameron	2.597403e+09	Rejected	Siecle
27	8.2	anthony	russo	3.048780e+09	Rejected	Siecle
28	5.9	peter	berg	3.542373e+09	Rejected	Siecle
29	7.0	colin	trevorrow	2.142857e+09	Rejected	Siecle
30	7.8	sam	mendes	2.564103e+09	Rejected	Siecle
31	7.3	sam	raimi	2.739726e+09	Rejected	Siecle
32	7.2	shane	black	2.777778e+09	Rejected	Siecle
33	6.5	tim	burton	3.076923e+09	Rejected	Siecle
34	6.8	brett	ratner	3.088235e+09	Rejected	Siecle
35	7.3	dan	scanlon	2.739726e+09	Rejected	Siecle
36	6.0	michael	bay	3.333333e+09	Maintain	Siecle
37	5.7	michael	bay	3.684211e+09	Rejected	Siecle
38	6.4	sam	raimi	3.359375e+09	Rejected	Siecle
39	6.7	marc	webb	2.985075e+09	Rejected	Siecle
40	6.8	joseph	kosinski	2.500000e+09	Rejected	Siecle
41	6.3	john	lasseter	3.174603e+09	Rejected	Siecle
42	5.6	martin	campbell	3.571429e+09	Rejected	Siecle
43	8.3	lee	unkrich	2.409639e+09	Rejected	Siecle
44	6.6	mcg	NaN	3.030303e+09	Maintain	Siecle
45	7.2	james	wan	2.638889e+09	Rejected	Siecle
46	7.0	marc	forster	2.714286e+09	Rejected	Siecle
47	8.0	bryan	singer	2.500000e+09	Rejected	Siecle
48	7.8	j.j.	abrams	2.435897e+09	Rejected	Siecle

49	6.3	bryan	singer	3.095238e+09	Rejected Siecle
50	7.3	baz	luhrmann	1.438356e+09	Rejected Siecle

		complexe_choix	NewColonne
0		2008-2010	year_2009
1	Between 2055 and 2007		VIP
2		2014-2016	VIP
3		2011-2013	VIP
5		2011-2013	VIP
6	Between 2055 and 2007		VIP
7		2008-2010	year_2010
8		2014-2016	VIP
9		2008-2010	year_2009
10		2014-2016	VIP
11	Between 2055 and 2007		VIP
12		2008-2010	VIP
13	Between 2055 and 2007		VIP
14		2011-2013	year_2013
15		2011-2013	year_2013
16		2008-2010	VIP
17		2011-2013	VIP
18		2011-2013	VIP
19		2011-2013	VIP
20		2014-2016	VIP
21		2011-2013	VIP
22		2008-2010	year_2010
23		2011-2013	year_2013
24	Between 2055 and 2007		VIP
25	Between 2055 and 2007		VIP
26		na	VIP
27		2014-2016	VIP
28		2011-2013	VIP
29		2014-2016	VIP
30		2011-2013	VIP
31		na	VIP
32		2011-2013	year_2013
33		2008-2010	year_2010
34	Between 2055 and 2007		VIP
35		2011-2013	year_2013
36		2008-2010	year_2009
37		2014-2016	VIP
38		2011-2013	year_2013
39		2014-2016	VIP
40		2008-2010	year_2010
41		2011-2013	VIP
42		2011-2013	VIP
43		2008-2010	year_2010

```

44          2008-2010  year_2009
45          2014-2016      VIP
46          2011-2013  year_2013
47          2014-2016      VIP
48          2011-2013  year_2013
49          2011-2013  year_2013
50          2011-2013  year_2013

```

**Utilisation de map, replace et category** `Data['sex'].map({'male':0, 'femelle': 1})` #Ceci prend toutes les valeurs male et femelle de la colonne sex et les remplace par 0 et 1

`Data['sex'].replace(['male', 'female'], [1,0])` `Data['sex'].astype('category').cat.codes` #créer automatiquement des codes pour toutes les valeurs uniques de la colonne sex et assigne les valeurs

## 1.1 Decoupage des datas (Extraction des datas)

**Choix d'une colonne, situation d'une colonne, condition sur les colonnes , etc...**

```

[76]: #mettre en evidence la colonne year
      year_movies = movies ["year"]

      year_movies.head(20)

```

```

[76]: 0      2009
      1      2007
      2      2015
      3      2012
      5      2012
      6      2007
      7      2010
      8      2015
      9      2009
     10      2016
     11      2006
     12      2008
     13      2006
     14      2013
     15      2013
     16      2008
     17      2012
     18      2011
     19      2012
     20      2014
      Name: year, dtype: int32

```

```

[78]: #voir le nombre de lignes
      movies["year"].shape

```

[78]: (3655,)

```
[79]: #Avoir juste le titre et le year
title_year = movies[["title", "year"]]

#check
title_year.head(20)
```

```
[79]:
```

	title	year
0	Avatar	2009
1	Pirates of the Caribbean: At World's End	2007
2	Spectre	2015
3	The Dark Knight Rises	2012
5	John Carter	2012
6	Spider-Man 3	2007
7	Tangled	2010
8	Avengers: Age of Ultron	2015
9	Harry Potter and the Half-Blood Prince	2009
10	Batman v Superman: Dawn of Justice	2016
11	Superman Returns	2006
12	Quantum of Solace	2008
13	Pirates of the Caribbean: Dead Man's Chest	2006
14	The Lone Ranger	2013
15	Man of Steel	2013
16	The Chronicles of Narnia: Prince Caspian	2008
17	The Avengers	2012
18	Pirates of the Caribbean: On Stranger Tides	2011
19	Men in Black 3	2012
20	The Hobbit: The Battle of the Five Armies	2014

```
[80]: #movies after ou apres 2012
movies_after_2012 = movies[movies['year'] > 2012]

movies_after_2012.head(20)
```

```
[80]:
```

	director_name	duration	gross	\
2	SAM MENDES	148.0	200074175.0	
8	JOSS WHEDON	141.0	458991599.0	
10	ZACK SNYDER	183.0	330249062.0	
14	GORE VERBINSKI	150.0	89289910.0	
15	ZACK SNYDER	143.0	291021565.0	
20	PETER JACKSON	164.0	255108370.0	
23	PETER JACKSON	186.0	258355354.0	
27	ANTHONY RUSSO	147.0	407197282.0	
29	COLIN TREVORROW	124.0	652177271.0	
32	SHANE BLACK	195.0	408992272.0	
35	DAN SCANLON	104.0	268488329.0	

37	MICHAEL BAY	165.0	245428137.0
38	SAM RAIMI	130.0	234903076.0
39	MARC WEBB	142.0	202853933.0
45	JAMES WAN	140.0	350034110.0
46	MARC FORSTER	123.0	202351611.0
47	BRYAN SINGER	149.0	233914986.0
48	J.J. ABRAMS	132.0	228756232.0
49	BRYAN SINGER	114.0	65171860.0
50	BAZ LUHRMANN	143.0	144812796.0

	genres \
2	Action Adventure Thriller
8	Action Adventure Sci-Fi
10	Action Adventure Sci-Fi
14	Action Adventure Western
15	Action Adventure Fantasy Sci-Fi
20	Adventure Fantasy
23	Adventure Fantasy
27	Action Adventure Sci-Fi
29	Action Adventure Sci-Fi Thriller
32	Action Adventure Sci-Fi
35	Adventure Animation Comedy Family Fantasy
37	Action Adventure Sci-Fi
38	Adventure Family Fantasy
39	Action Adventure Fantasy Sci-Fi
45	Action Crime Thriller
46	Action Adventure Horror Sci-Fi Thriller
47	Action Adventure Fantasy Sci-Fi Thriller
48	Action Adventure Sci-Fi
49	Adventure Fantasy
50	Drama Romance

	title	budget	year	language \
2	Spectre	245000000.0	2015	Eng
8	Avengers: Age of Ultron	250000000.0	2015	Eng
10	Batman v Superman: Dawn of Justice	250000000.0	2016	Eng
14	The Lone Ranger	215000000.0	2013	Eng
15	Man of Steel	225000000.0	2013	Eng
20	The Hobbit: The Battle of the Five Armies	250000000.0	2014	Eng
23	The Hobbit: The Desolation of Smaug	225000000.0	2013	Eng
27	Captain America: Civil War	250000000.0	2016	Eng
29	Jurassic World	150000000.0	2015	Eng
32	Iron Man 3	200000000.0	2013	Eng
35	Monsters University	200000000.0	2013	Eng
37	Transformers: Age of Extinction	210000000.0	2014	Eng
38	Oz the Great and Powerful	215000000.0	2013	Eng
39	The Amazing Spider-Man 2	200000000.0	2014	Eng

45		Furious 7	190000000.0	2015	Eng
46		World War Z	190000000.0	2013	Eng
47		X-Men: Days of Future Past	200000000.0	2014	Eng
48		Star Trek Into Darkness	190000000.0	2013	Eng
49		Jack the Giant Slayer	195000000.0	2013	Eng
50		The Great Gatsby	105000000.0	2013	Eng

	score	surname	surnname	name	taux_rendu \
2	6.8	SAM	[SAM, MENDES]	MENDES	3.602941e+09
8	7.5	JOSS	[JOSS, WHEDON]	WHEDON	3.333333e+09
10	6.9	ZACK	[ZACK, SNYDER]	SNYDER	3.623188e+09
14	6.5	GORE	[GORE, VERBINSKI]	VERBINSKI	3.307692e+09
15	7.2	ZACK	[ZACK, SNYDER]	SNYDER	3.125000e+09
20	7.5	PETER	[PETER, JACKSON]	JACKSON	3.333333e+09
23	7.9	PETER	[PETER, JACKSON]	JACKSON	2.848101e+09
27	8.2	ANTHONY	[ANTHONY, RUSSO]	RUSSO	3.048780e+09
29	7.0	COLIN	[COLIN, TREVORROW]	TREVORROW	2.142857e+09
32	7.2	SHANE	[SHANE, BLACK]	BLACK	2.777778e+09
35	7.3	DAN	[DAN, SCANLON]	SCANLON	2.739726e+09
37	5.7	MICHAEL	[MICHAEL, BAY]	BAY	3.684211e+09
38	6.4	SAM	[SAM, RAIMI]	RAIMI	3.359375e+09
39	6.7	MARC	[MARC, WEBB]	WEBB	2.985075e+09
45	7.2	JAMES	[JAMES, WAN]	WAN	2.638889e+09
46	7.0	MARC	[MARC, FORSTER]	FORSTER	2.714286e+09
47	8.0	BRYAN	[BRYAN, SINGER]	SINGER	2.500000e+09
48	7.8	J.J.	[J.J., ABRAMS]	ABRAMS	2.435897e+09
49	6.3	BRYAN	[BRYAN, SINGER]	SINGER	3.095238e+09
50	7.3	BAZ	[BAZ, LUHRMANN]	LUHRMANN	1.438356e+09

	siecle_simple	siecle_for	complexe_choix	NewColonne
2	Rejected Siecle	Rejected value	2014-2016	VIP
8	Rejected Siecle	Rejected value	2014-2016	VIP
10	Rejected Siecle	Rejected value	2014-2016	VIP
14	Rejected Siecle	Rejected value	2011-2013	year_2013
15	Rejected Siecle	Rejected value	2011-2013	year_2013
20	Rejected Siecle	Rejected value	2014-2016	VIP
23	Rejected Siecle	Rejected value	2011-2013	year_2013
27	Rejected Siecle	Rejected value	2014-2016	VIP
29	Rejected Siecle	Rejected value	2014-2016	VIP
32	Rejected Siecle	Rejected value	2011-2013	year_2013
35	Rejected Siecle	Rejected value	2011-2013	year_2013
37	Rejected Siecle	Rejected value	2014-2016	VIP
38	Rejected Siecle	Rejected value	2011-2013	year_2013
39	Rejected Siecle	Rejected value	2014-2016	VIP
45	Rejected Siecle	Rejected value	2014-2016	VIP
46	Rejected Siecle	Rejected value	2011-2013	year_2013
47	Rejected Siecle	Rejected value	2014-2016	VIP

```

48 Rejected Siecle Rejected value      2011-2013 year_2013
49 Rejected Siecle Rejected value      2011-2013 year_2013
50 Rejected Siecle Rejected value      2011-2013 year_2013

```

```

[81]: #Movies for 2012 ou 2015 or

movie_2012_or_2015 = movies[(movies['year'] == 2012) | (movies['year'] == 2015)]

movie_2012_or_2015.head(20)

```

```

[81]:      director_name  duration      gross \
2          SAM MENDES    148.0  200074175.0
3  CHRISTOPHER NOLAN    164.0  448130642.0
5    ANDREW STANTON    132.0   73058679.0
8      JOSS WHEDON    141.0  458991599.0
17     JOSS WHEDON    173.0  623279547.0
19   BARRY SONNENFELD    106.0  179020854.0
21         MARC WEBB    153.0  262030663.0
28        PETER BERG    131.0   65173160.0
29   COLIN TREVORROW    124.0  652177271.0
30         SAM MENDES    143.0  304360277.0
45        JAMES WAN    140.0  350034110.0
56    MARK ANDREWS     93.0  237282182.0
62   LANA WACHOWSKI    127.0   47375327.0
78    PETE DOCTER     95.0  356454367.0
81   RUPERT SANDERS    132.0  155111815.0
88        BRAD BIRD    130.0   93417865.0
90    RICH MOORE     101.0  189412677.0
103  FRANCIS LAWRENCE    137.0  281666058.0
109        ALAN TAYLOR    126.0   89732035.0
128   GEORGE MILLER    120.0  153629485.0

      genres \
2  Action|Adventure|Thriller
3          Action|Thriller
5  Action|Adventure|Sci-Fi
8  Action|Adventure|Sci-Fi
17 Action|Adventure|Sci-Fi
19 Action|Adventure|Comedy|Family|Fantasy|Sci-Fi
21 Action|Adventure|Fantasy
28 Action|Adventure|Sci-Fi|Thriller
29 Action|Adventure|Sci-Fi|Thriller
30 Action|Adventure|Thriller
45 Action|Crime|Thriller
56 Adventure|Animation|Comedy|Family|Fantasy
62 Action|Adventure|Sci-Fi
78 Adventure|Animation|Comedy|Drama|Family|Fantasy

```

81 Action|Adventure|Drama|Fantasy  
 88 Action|Adventure|Family|Mystery|Sci-Fi  
 90 Adventure|Animation|Comedy|Family|Sci-Fi  
 103 Adventure|Sci-Fi  
 109 Action|Adventure|Sci-Fi  
 128 Action|Adventure|Sci-Fi|Thriller

	title	budget	year	language	score \
2	Spectre	245000000.0	2015	Eng	6.8
3	The Dark Knight Rises	250000000.0	2012	Eng	8.5
5	John Carter	263700000.0	2012	Eng	6.6
8	Avengers: Age of Ultron	250000000.0	2015	Eng	7.5
17	The Avengers	220000000.0	2012	Eng	8.1
19	Men in Black 3	225000000.0	2012	Eng	6.8
21	The Amazing Spider-Man	230000000.0	2012	Eng	7.0
28	Battleship	209000000.0	2012	Eng	5.9
29	Jurassic World	150000000.0	2015	Eng	7.0
30	Skyfall	200000000.0	2012	Eng	7.8
45	Furious 7	190000000.0	2015	Eng	7.2
56	Brave	185000000.0	2012	Eng	7.2
62	Jupiter Ascending	176000000.0	2015	Eng	5.4
78	Inside Out	175000000.0	2015	Eng	8.3
81	Snow White and the Huntsman	170000000.0	2012	Eng	6.1
88	Tomorrowland	190000000.0	2015	Eng	6.5
90	Wreck-It Ralph	165000000.0	2012	Eng	7.8
103	The Hunger Games: Mockingjay - Part 2	160000000.0	2015	Eng	6.6
109	Terminator Genisys	155000000.0	2015	Eng	6.6
128	Mad Max: Fury Road	150000000.0	2015	Eng	8.1

	surname	surnname	name	taux_rendu \
2	SAM	[SAM, MENDES]	MENDES	3.602941e+09
3	CHRISTOPHER	[CHRISTOPHER, NOLAN]	NOLAN	2.941176e+09
5	ANDREW	[ANDREW, STANTON]	STANTON	3.995455e+09
8	JOSS	[JOSS, WHEDON]	WHEDON	3.333333e+09
17	JOSS	[JOSS, WHEDON]	WHEDON	2.716049e+09
19	BARRY	[BARRY, SONNENFELD]	SONNENFELD	3.308824e+09
21	MARC	[MARC, WEBB]	WEBB	3.285714e+09
28	PETER	[PETER, BERG]	BERG	3.542373e+09
29	COLIN	[COLIN, TREVORROW]	TREVORROW	2.142857e+09
30	SAM	[SAM, MENDES]	MENDES	2.564103e+09
45	JAMES	[JAMES, WAN]	WAN	2.638889e+09
56	MARK	[MARK, ANDREWS]	ANDREWS	2.569444e+09
62	LANA	[LANA, WACHOWSKI]	WACHOWSKI	3.259259e+09
78	PETE	[PETE, DOCTER]	DOCTER	2.108434e+09
81	RUPERT	[RUPERT, SANDERS]	SANDERS	2.786885e+09
88	BRAD	[BRAD, BIRD]	BIRD	2.923077e+09
90	RICH	[RICH, MOORE]	MOORE	2.115385e+09



103	FRANCIS	[FRANCIS, LAWRENCE]	LAWRENCE	2.424242e+09
109	ALAN	[ALAN, TAYLOR]	TAYLOR	2.348485e+09
128	GEORGE	[GEORGE, MILLER]	MILLER	1.851852e+09

	siecle_simple	siecle_for	complexe_choix	NewColonne
2	Rejected Siecle	Rejected value	2014-2016	VIP
3	Rejected Siecle	Rejected value	2011-2013	VIP
5	Rejected Siecle	Rejected value	2011-2013	VIP
8	Rejected Siecle	Rejected value	2014-2016	VIP
17	Rejected Siecle	Rejected value	2011-2013	VIP
19	Rejected Siecle	Rejected value	2011-2013	VIP
21	Rejected Siecle	Rejected value	2011-2013	VIP
28	Rejected Siecle	Rejected value	2011-2013	VIP
29	Rejected Siecle	Rejected value	2014-2016	VIP
30	Rejected Siecle	Rejected value	2011-2013	VIP
45	Rejected Siecle	Rejected value	2014-2016	VIP
56	Rejected Siecle	Rejected value	2011-2013	VIP
62	Rejected Siecle	Rejected value	2014-2016	VIP
78	Rejected Siecle	Rejected value	2014-2016	VIP
81	Rejected Siecle	Rejected value	2011-2013	VIP
88	Rejected Siecle	Rejected value	2014-2016	VIP
90	Rejected Siecle	Rejected value	2011-2013	VIP
103	Rejected Siecle	Rejected value	2014-2016	VIP
109	Rejected Siecle	Rejected value	2014-2016	VIP
128	Rejected Siecle	Rejected value	2014-2016	VIP

## 1.2 Analyses Preliminaires (Idees Globales ou Summaries des datas)

### Moyenne

```
[82]: # Moyenne des scores de tous les films

movies["score"].mean()
```

```
[82]: 6.463365253077975
```

### Mediane

```
[83]: movies["score"].median()
```

```
[83]: 6.6
```

```
[84]: #Moyenne et Meidane sur base de deux variables
movies[["score", "duration"]].mean()

movies[["score", "duration"]].median()
```

```
[84]: score          6.6
      duration    106.0
      dtype: float64
```

## Utilisation de la fonction describe

```
[85]: #voir les analyses descriptives preliminaires des nos datas sur base de toutes
      ↪ les variables quantitatives
      movies.describe()
```

```
[85]:
```

	duration	gross	budget	year	score \
count	3655.000000	3.655000e+03	3.655000e+03	3655.000000	3655.000000
mean	110.168263	5.177582e+07	4.242414e+07	2002.904241	6.463365
std	22.699601	6.923325e+07	1.087764e+08	9.877499	1.057851
min	37.000000	1.620000e+02	2.180000e+02	1927.000000	1.600000
25%	96.000000	7.970599e+06	1.000000e+07	1999.000000	5.900000
50%	106.000000	2.920000e+07	2.500000e+07	2004.000000	6.600000
75%	120.000000	6.646734e+07	5.000000e+07	2010.000000	7.200000
max	330.000000	7.605058e+08	4.200000e+09	2016.000000	9.300000

	taux_rendu
count	3.655000e+03
mean	6.629820e+08
std	1.489094e+09
min	3.027778e+03
25%	1.575188e+08
50%	3.846154e+08
75%	8.189270e+08
max	5.454545e+10

## Utilisation de la fonction describe sur des variables bien definies (score, duration)

```
[87]: movies[["score", "duration"]].describe()
```

```
[87]:
```

	score	duration
count	3655.000000	3655.000000
mean	6.463365	110.168263
std	1.057851	22.699601
min	1.600000	37.000000
25%	5.900000	96.000000
50%	6.600000	106.000000
75%	7.200000	120.000000
max	9.300000	330.000000

### 1.2.1 Utilisation de groupby pour categoriser les analyses

```
[95]: #group by genre et voir les analyses

      #movies.columns

      movies_by_genre = movies.groupby("genres")["score", "duration"].mean()

      movies_by_genre
```

[95]:

	score	duration
genres		
Action	5.916667	111.000000
Action Adventure	6.987500	114.125000
Action Adventure Animation Comedy Crime Family ...	6.200000	95.000000
Action Adventure Animation Comedy Drama Family ...	7.950000	96.000000
Action Adventure Animation Comedy Family	7.160000	92.000000
...	...	...
Mystery Thriller	6.713333	115.600000
Romance Sci-Fi Thriller	7.100000	106.000000
Sci-Fi Thriller	6.628571	108.142857
Thriller	4.800000	95.000000
Western	8.100000	128.500000

[745 rows x 2 columns]

[97]: *#Fixer des conditions seulement sur des variables numeriques*

```
movies.groupby("genres").mean(numeric_only = True)
```

[97]:

	duration	gross \
genres		
Action	111.000000	5.952091e+07
Action Adventure	114.125000	1.471453e+08
Action Adventure Animation Comedy Crime Family ...	95.000000	1.359691e+07
Action Adventure Animation Comedy Drama Family ...	96.000000	1.228235e+08
Action Adventure Animation Comedy Family	92.000000	1.801411e+08
...	...	...
Mystery Thriller	115.600000	5.819329e+07
Romance Sci-Fi Thriller	106.000000	6.245332e+07
Sci-Fi Thriller	108.142857	2.979379e+07
Thriller	95.000000	2.468000e+03
Western	128.500000	2.375000e+07

	budget	year \
genres		
Action	9.216667e+07	2000.833333
Action Adventure	7.337500e+07	1998.125000
Action Adventure Animation Comedy Crime Family ...	7.000000e+07	2001.000000
Action Adventure Animation Comedy Drama Family ...	1.175000e+08	2006.500000
Action Adventure Animation Comedy Family	1.330000e+08	2010.400000
...	...	...
Mystery Thriller	4.393333e+07	2005.733333
Romance Sci-Fi Thriller	5.020000e+07	2011.000000
Sci-Fi Thriller	2.772857e+07	2001.285714
Thriller	3.000000e+05	2014.000000
Western	4.050000e+06	1975.500000

	score	taux_rendu
genres		
Action	5.916667	1.677819e+09
Action Adventure	6.987500	1.060507e+09
Action Adventure Animation Comedy Crime Family ...	6.200000	1.129032e+09
Action Adventure Animation Comedy Drama Family ...	7.950000	1.481804e+09
Action Adventure Animation Comedy Family	7.160000	1.860746e+09
...	...	...
Mystery Thriller	6.713333	6.455420e+08
Romance Sci-Fi Thriller	7.100000	7.070423e+08
Sci-Fi Thriller	6.628571	4.572534e+08
Thriller	4.800000	6.250000e+06
Western	8.100000	5.400185e+07

[745 rows x 6 columns]

[98]: *#Grouper sur deux variables ou plus*

```
movies.groupby(["genres", "year"])["budget"].mean()
```

```
[98]: genres      year
      Action      1992      80000000.0
              1993       3000000.0
              1997     125000000.0
              2003      30000000.0
              2008     300000000.0
              ...
      Sci-Fi|Thriller 2010       1100000.0
                  2015     12000000.0
      Thriller      2014       300000.0
      Western      1966      1200000.0
                  1985      6900000.0
      Name: budget, Length: 2450, dtype: float64
```

### 1.2.2 Utilisation de Count Pour compter les valeurs par categories d'analyse

[99]: *#Compter les valeurs de Genres des films pour en voir la proportion*

```
movies["genres"].value_counts()
```

```
[99]: Comedy|Drama|Romance      145
      Drama                    139
      Comedy                   137
      Comedy|Drama             137
      Comedy|Romance           130
      ...
      Action|Adventure|Crime|Drama|Thriller      1
```

```

Action|Adventure|Comedy|Crime|Mystery|Thriller    1
Action|Adventure|Comedy|Drama|Thriller            1
Action|Family|Fantasy|Musical                    1
Comedy|Crime|Horror                              1
Name: genres, Length: 745, dtype: int64

```

```

[100]: #Normalizer (proportion)
movies["genres"].value_counts(normalize = True)

```

```

[100]: Comedy|Drama|Romance                0.039672
Drama                                     0.038030
Comedy                                   0.037483
Comedy|Drama                            0.037483
Comedy|Romance                          0.035568
...
Action|Adventure|Crime|Drama|Thriller      0.000274
Action|Adventure|Comedy|Crime|Mystery|Thriller 0.000274
Action|Adventure|Comedy|Drama|Thriller      0.000274
Action|Family|Fantasy|Musical              0.000274
Comedy|Crime|Horror                       0.000274
Name: genres, Length: 745, dtype: float64

```

```

[103]: # Normaliser et voir les pourcentages (representation)
movies["genres"].value_counts(normalize = True) * 100

```

```

[103]: Comedy|Drama|Romance                3.967168
Drama                                     3.803010
Comedy                                   3.748290
Comedy|Drama                            3.748290
Comedy|Romance                          3.556772
...
Action|Adventure|Crime|Drama|Thriller      0.027360
Action|Adventure|Comedy|Crime|Mystery|Thriller 0.027360
Action|Adventure|Comedy|Drama|Thriller      0.027360
Action|Family|Fantasy|Musical              0.027360
Comedy|Crime|Horror                       0.027360
Name: genres, Length: 745, dtype: float64

```

```

[104]: #voir le plus grand pourcentage
(movies["genres"].value_counts(normalize = True) * 100).max()

```

```

[104]: 3.9671682626538987

```

### Utilisation de Group by complexe et Sort\_values en Mm temps

```

[31]: #avoir la somme des scores de chq type de siecle simple
movies.groupby(['siecle_simple'], as_index = False)['score'].sum().
    ↪sort_values(by = 'score', ascending = False)

```

```
[31]:      siecle_simple      score
1  Rejected Siecle  22497.7
0  Maintain Siecle   1125.9
```

### 1.2.3 Utilisation de Sort pour trier les datas

```
[105]: #Trier les datas par les plus petits scores jusqu'aux grands scores
movies.sort_values(by = "score").head()
```

```
[105]:      director_name  duration      gross      genres \
2789      JON M. CHU    115.0  73000942.0  Documentary|Music
2240    JASON FRIEDBERG    88.0  14174654.0      Comedy
2266      BOB CLARK    88.0  9109322.0  Comedy|Family|Sci-Fi
3438  DON MICHAEL PAUL    93.0  5694308.0      Comedy|Sport
3282  VONDIE CURTIS-HALL  104.0  4273372.0  Drama|Music|Romance

      title      budget  year language  score \
2789  Justin Bieber: Never Say Never  13000000.0  2011      Eng    1.6
2240      Disaster Movie  25000000.0  2008      Eng    1.9
2266  Superbabies: Baby Geniuses 2  20000000.0  2004      Eng    1.9
3438      Who's Your Caddy?   7000000.0  2007      Eng    2.0
3282      Glitter  22000000.0  2001      Eng    2.1

      surname      surname      name  taux_rendu \
2789      JON      [JON, M., CHU]      M.  8.125000e+08
2240      JASON    [JASON, FRIEDBERG]  FRIEDBERG  1.315789e+09
2266      BOB      [BOB, CLARK]      CLARK  1.052632e+09
3438      DON    [DON, MICHAEL, PAUL]  MICHAEL  3.500000e+08
3282  VONDIE  [VONDIE, CURTIS-HALL]  CURTIS-HALL  1.047619e+09

      siecle_simple      siecle_for      complexe_choix  NewColonne
2789  Rejected Siecle  Rejected value      2011-2013      VIP
2240  Rejected Siecle  Rejected value      2008-2010      VIP
2266  Rejected Siecle  Rejected value      na      VIP
3438  Rejected Siecle  Rejected value  Between 2055 and 2007      VIP
3282  Rejected Siecle  Rejected value      na      VIP
```

```
[106]: #Trier sur base de deux variables et d'une maniere desc
#head pour voir juste les 5 premieres lignes
```

```
movies.sort_values(by = ["score", "year"], ascending = False).head()
```

```
[106]:      director_name  duration      gross \
1920      FRANK DARABONT    142.0  28341469.0
3402  FRANCIS FORD COPPOLA    175.0  134821952.0
66      CHRISTOPHER NOLAN    152.0  533316061.0
2791  FRANCIS FORD COPPOLA    220.0  57300000.0
```

335	PETER JACKSON	192.0	377019252.0
-----	---------------	-------	-------------

	genres \
1920	Crime Drama
3402	Crime Drama
66	Action Crime Drama Thriller
2791	Crime Drama
335	Action Adventure Drama Fantasy

	title	budget	year \
1920	The Shawshank Redemption	25000000.0	1994
3402	The Godfather	6000000.0	1972
66	The Dark Knight	185000000.0	2008
2791	The Godfather: Part II	13000000.0	1974
335	The Lord of the Rings: The Return of the King	94000000.0	2003

	language	score	surname	surnname	name \
1920	Eng	9.3	FRANK	[FRANK, DARABONT]	DARABONT
3402	Eng	9.2	FRANCIS	[FRANCIS, FORD, COPPOLA]	FORD
66	Eng	9.0	CHRISTOPHER	[CHRISTOPHER, NOLAN]	NOLAN
2791	Eng	9.0	FRANCIS	[FRANCIS, FORD, COPPOLA]	FORD
335	Eng	8.9	PETER	[PETER, JACKSON]	JACKSON

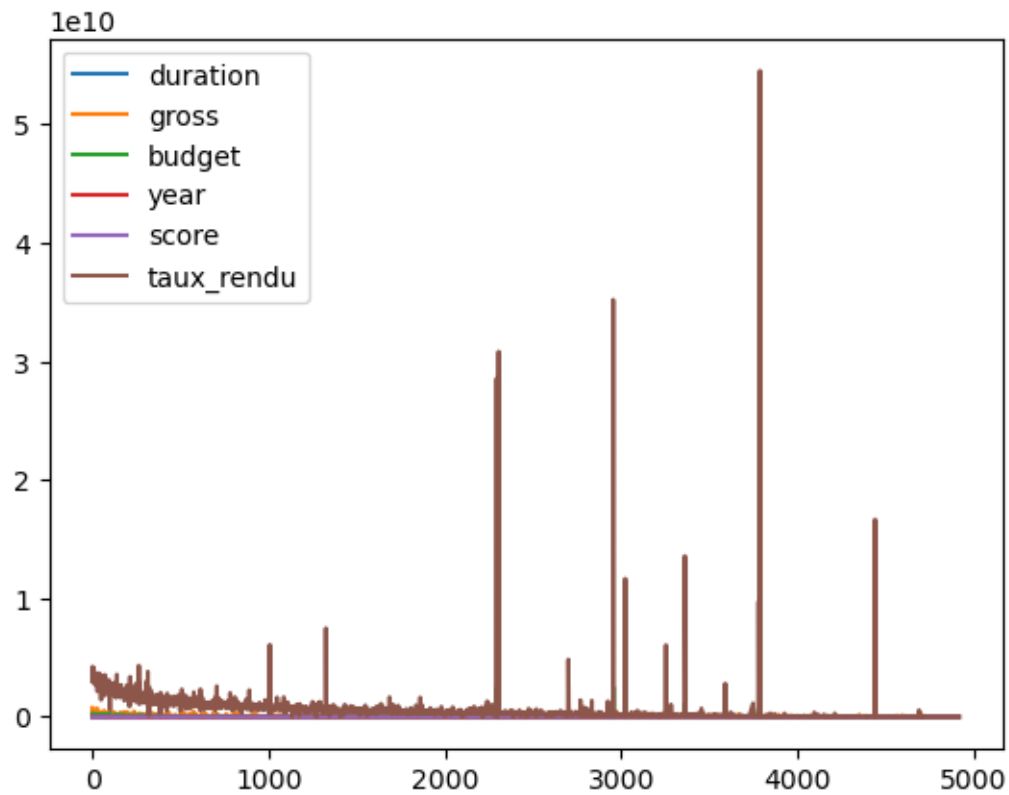
	taux_rendu	siecle_simple	siecle_for	complexe_choix	NewColonne
1920	2.688172e+08	Rejected Siecle	Rejected value	na	VIP
3402	6.521739e+07	Rejected Siecle	Rejected value	na	VIP
66	2.055556e+09	Rejected Siecle	Rejected value	2008-2010	VIP
2791	1.444444e+08	Rejected Siecle	Rejected value	na	VIP
335	1.056180e+09	Rejected Siecle	Rejected value	na	VIP

## 2 Visualisation de certains aspects des datas

### 2.1 1. Utilisation de plot de MATPLOTLIB

```
[107]: #Voir juste la representation des datas en general
movies.plot()

plt.show()
#ce graphique melange juste les valeurs mais tout en tenant compte des valeurs
↳numeriques<<<>>>
```

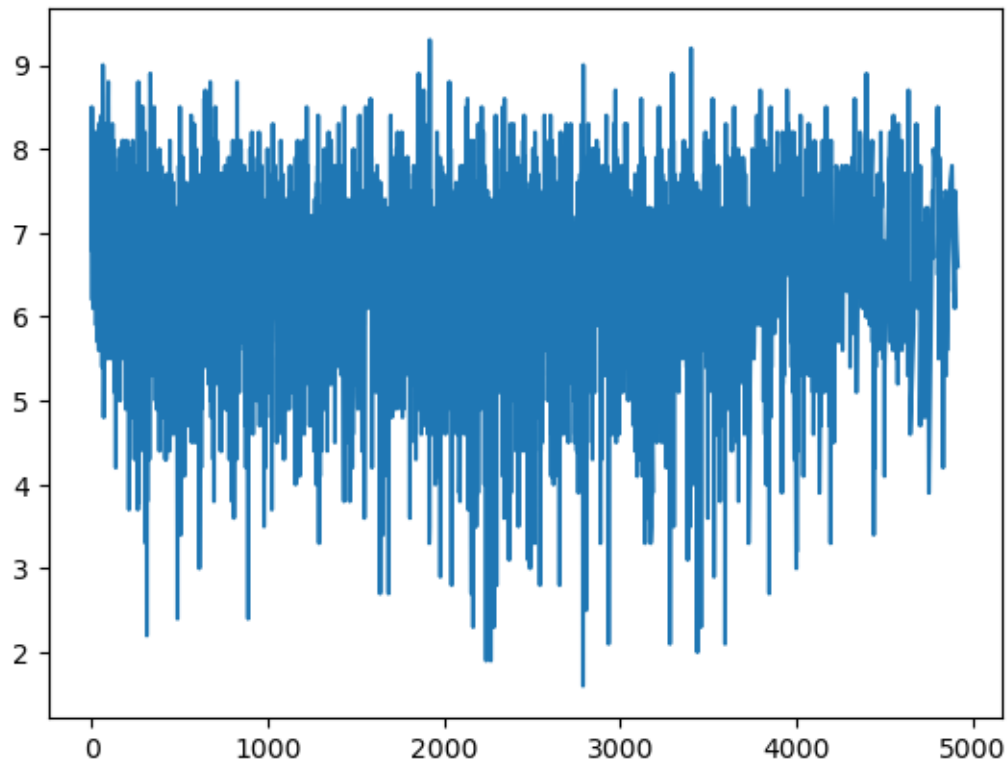


[ ]:

```
[108]: #utilisation de plot() sur une variable bien definie
#exemple de la vraieible score
movies["score"].plot()

plt.show()
```



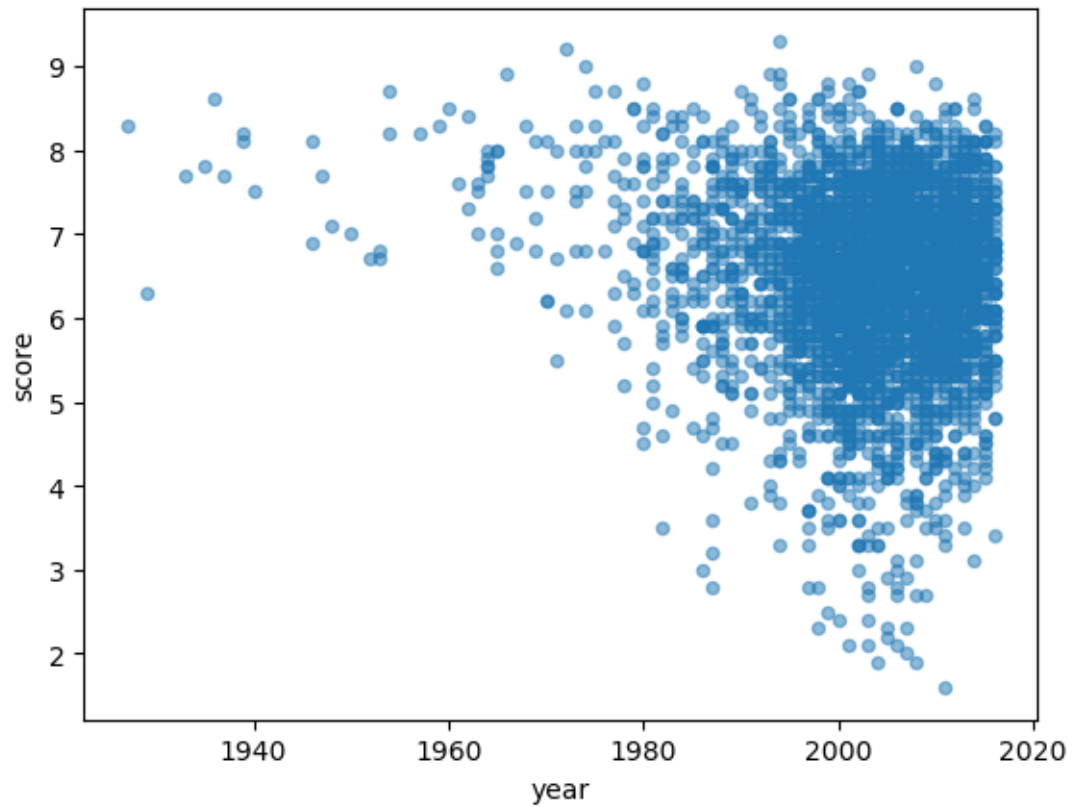


### Scatter Plot (Voir la corrélation entre deux variables)

```
[111]: #voir la relation entre le score et le year
movies.plot.scatter(x = "year", y = "score", alpha = 0.5)

plt.show()

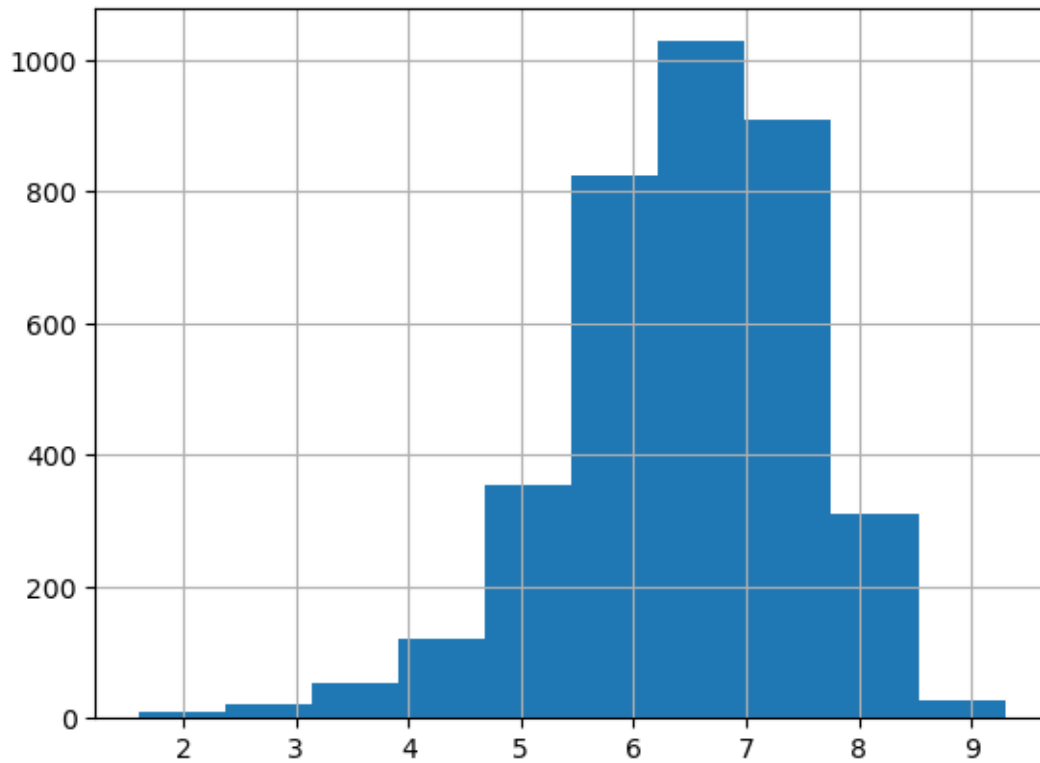
#Sur base de ce graphique nous trouvons qu'il n'y a pas grande relation entre
↳ la variable year et score
```



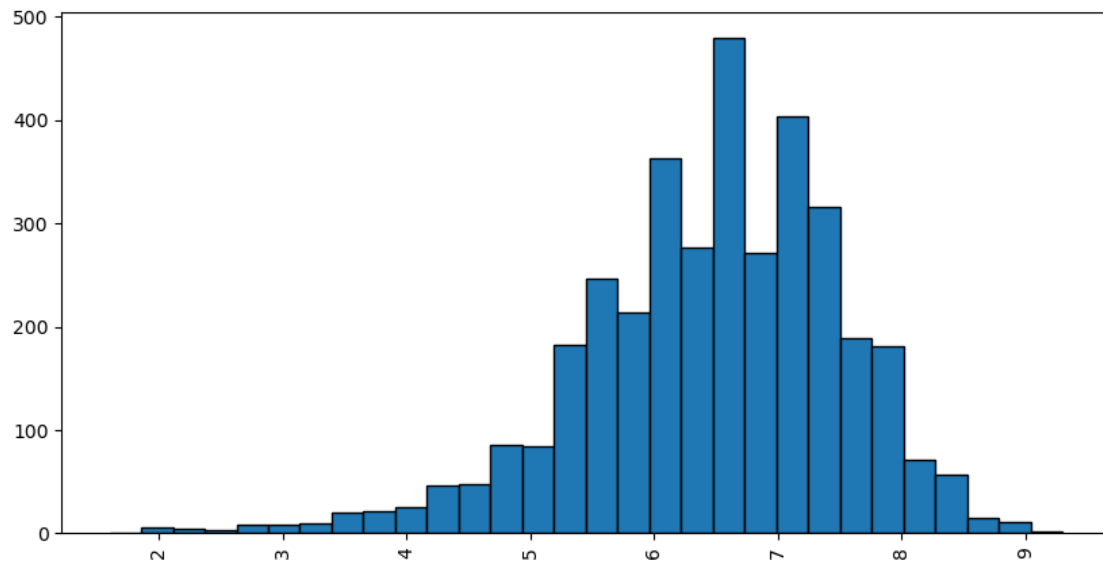
```
[ ]: ##### Utilisation de hist() pour voir la situation de notre score
```

```
[43]: movies['score'].hist()
```

```
[43]: <Axes: >
```



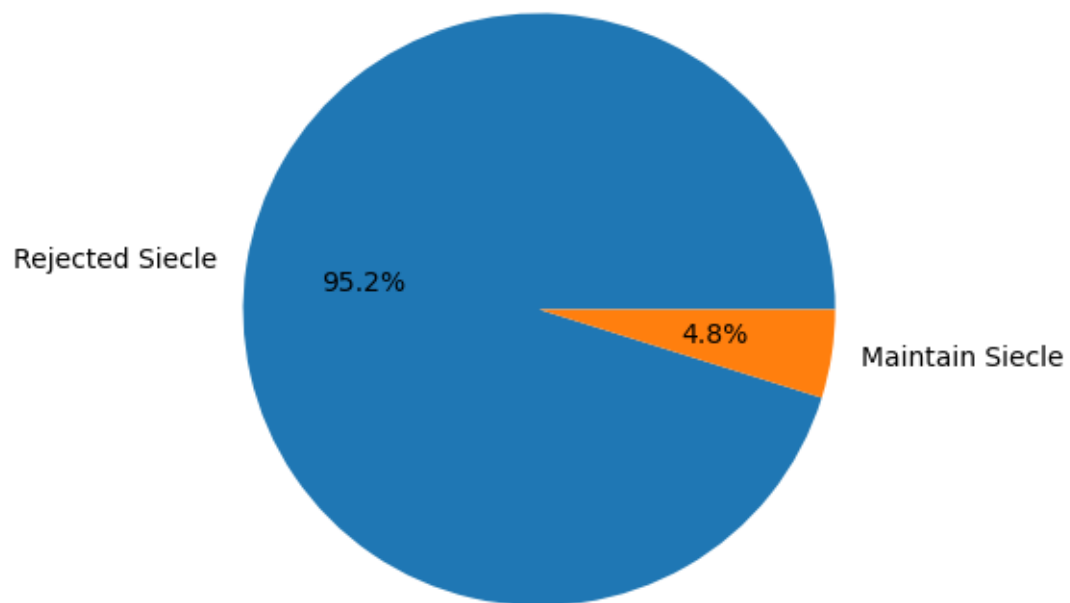
```
[44]: #considerer notre dataframe d'une seule colonne des cores tout simplement  
score_data = movies['score']  
plt.figure(figsize =(10,5))  
plt.hist(score_data, bins = 30, edgecolor = 'Black')  
plt.xticks(rotation=90)  
plt.show()
```



### Pie Plot (Pour repartir la repartition en proportion)

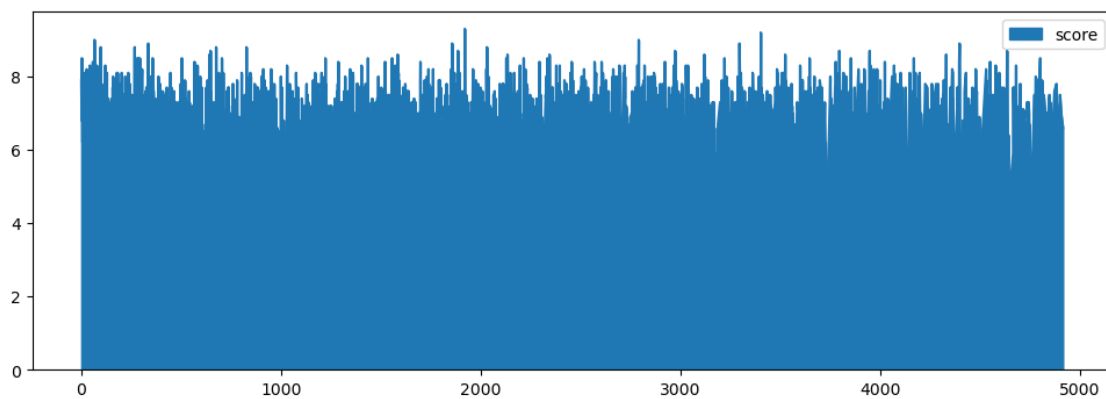
```
[47]: #Créer un sous dataframe pour la colonne/variable à analyser
siecle_simple_datas = movies['siecle_simple'].value_counts()
plt.figure(figsize = (8,5))
plt.pie(siecle_simple_datas, labels = siecle_simple_datas.index, autopct = '%1.
↪1f%%')

plt.show()
```



utilisation de plot, area et subplot

```
[55]: siecle_simple_datas = movies[['score', 'siecle_simple']]  
  
siecle_simple_datas  
  
siecle_simple_datas.plot.area(figsize= (12,4), subplots = True)  
plt.show()
```



```
[ ]:
```

```
[ ]:
```

## 2.2 2. Utilisation de SEABORN

```
[24]: #compter par language
movies['language'].unique() #voir d'abors les valeurs uniques, ceci nous
    ↳ indique que pour cette variable nous avons aussi bcp de valeurs, cherchons
    ↳ une autre varibale plus globale
```

```
[24]: array(['Eng', 'Mandarin', 'Aboriginal', 'Spanish', 'Paris', 'Filipino',
        'Maya', 'Kazakh', 'Cantonese', 'Japanese', 'Aramaic', 'Italian',
        'Dutch', 'Dari', 'German', 'Mongolian', 'Thai', 'Bosnian',
        'Korean', 'Hungarian', 'Hindi', 'Danish', 'Portuguese',
        'Norwegian', 'Czech', 'Russian', 'None', 'Zulu', 'Hebrew',
        'Arabic', 'Vietnamese', 'Indonesian', 'Romanian', 'Persian'],
        dtype=object)
```

```
[25]: #compter par Anne (year)
movies['year'].unique() #elles sont aussi nombreuses, nous pouvons opter pour
    ↳ la valeur siecle_for
```

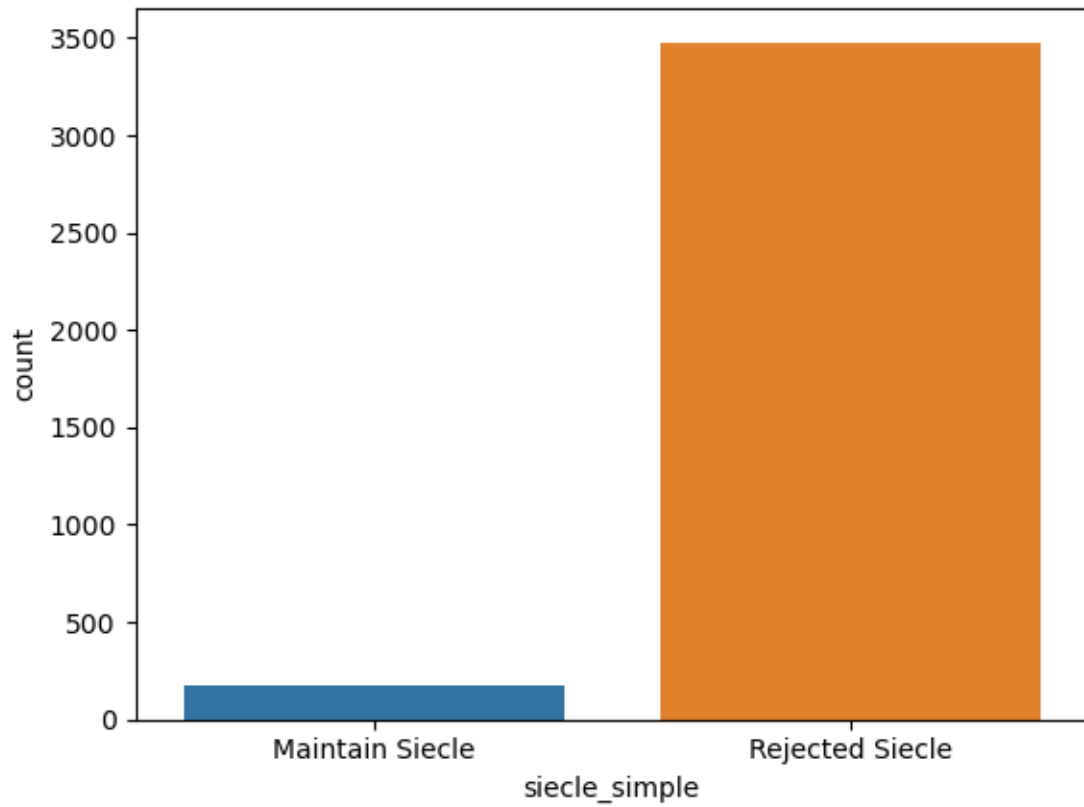
```
[25]: array([2009, 2007, 2015, 2012, 2010, 2016, 2006, 2008, 2013, 2011, 2014,
        2005, 1997, 2004, 1999, 1995, 2003, 2001, 2002, 1998, 2000, 1990,
        1991, 1994, 1996, 1982, 1993, 1992, 1989, 1988, 1978, 1980, 1984,
        1985, 1940, 1963, 1987, 1986, 1973, 1983, 1979, 1981, 1977, 1970,
        1960, 1965, 1964, 1962, 1927, 1974, 1975, 1968, 1967, 1971, 1969,
        1972, 1976, 1961, 1946, 1953, 1952, 1939, 1950, 1948, 1957, 1959,
        1947, 1936, 1937, 1966, 1954, 1935, 1933, 1929])
```

### Utilisation de countplot pour compter les categories

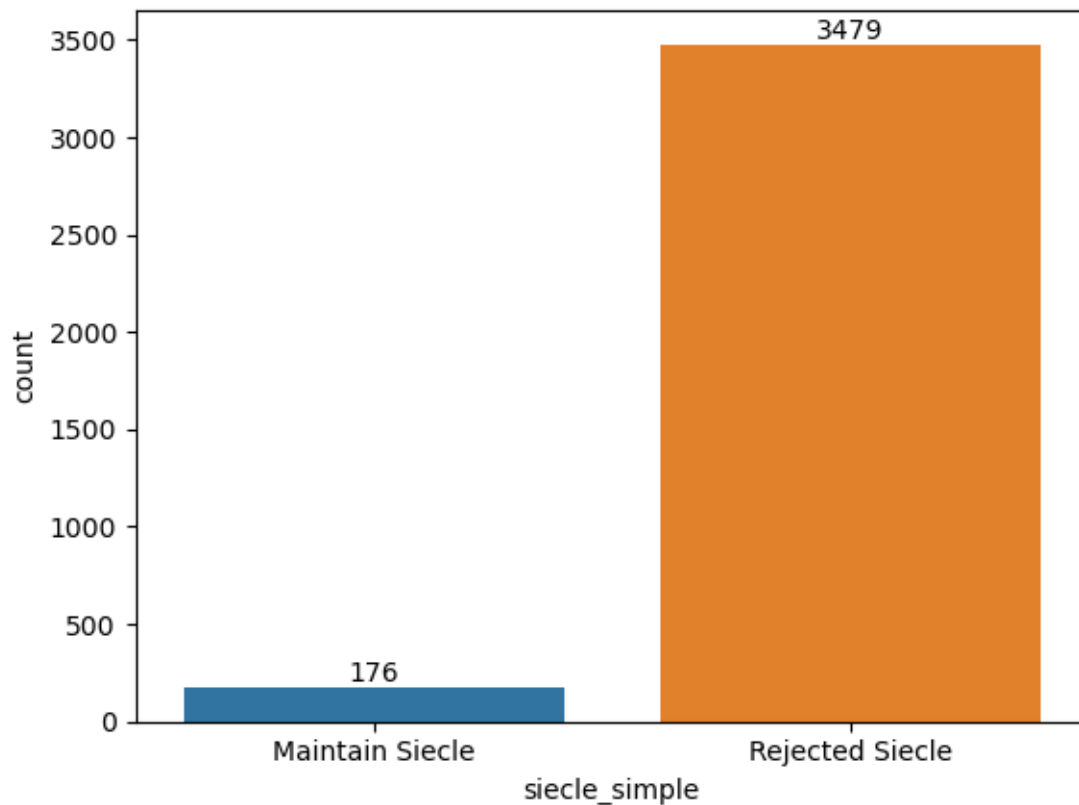
```
[28]: #compter par siecle_simple
#movies.head()
movies['siecle_simple'].unique()
```

```
[28]: array(['Maintain Siecle', 'Rejected Siecle'], dtype=object)
```

```
[29]: #comptons donc les valeurs par la variable qualitative siecle_for
ax = sns.countplot(x = 'siecle_simple', data = movies)
```



```
[30]: #si l'on veut afficher les valeurs sur les bars:  
ax = sns.countplot(x = 'siecle_simple', data = movies)  
  
for bars in ax.containers:  
    ax.bar_label(bars)
```



### Representation des valeurs avec barplot

```
[34]: portion_movies = movies.groupby(['siecle_simple'], as_index = False)['score'].
      ↪sum().sort_values(by = 'score', ascending = False)

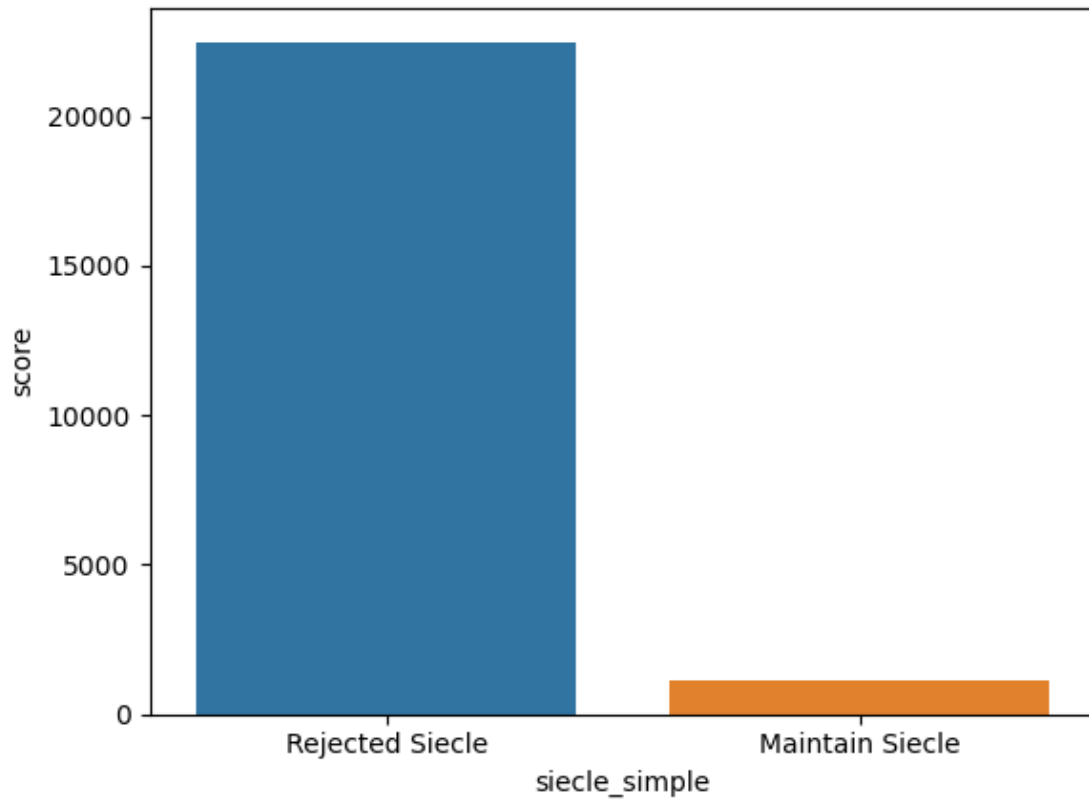
      #voir portion_movies
      portion_movies
```

```
[34]:      siecle_simple    score
      1 Rejected Siecle  22497.7
      0 Maintain Siecle   1125.9
```

```
[35]: #ainsi la sous dataframe contient les variables siecle_simple et score en somme_
      ↪de chaque siecle simple
      sns.barplot(x = 'siecle_simple', y = 'score', data = portion_movies)
```

```
[35]: <Axes: xlabel='siecle_simple', ylabel='score'>
```



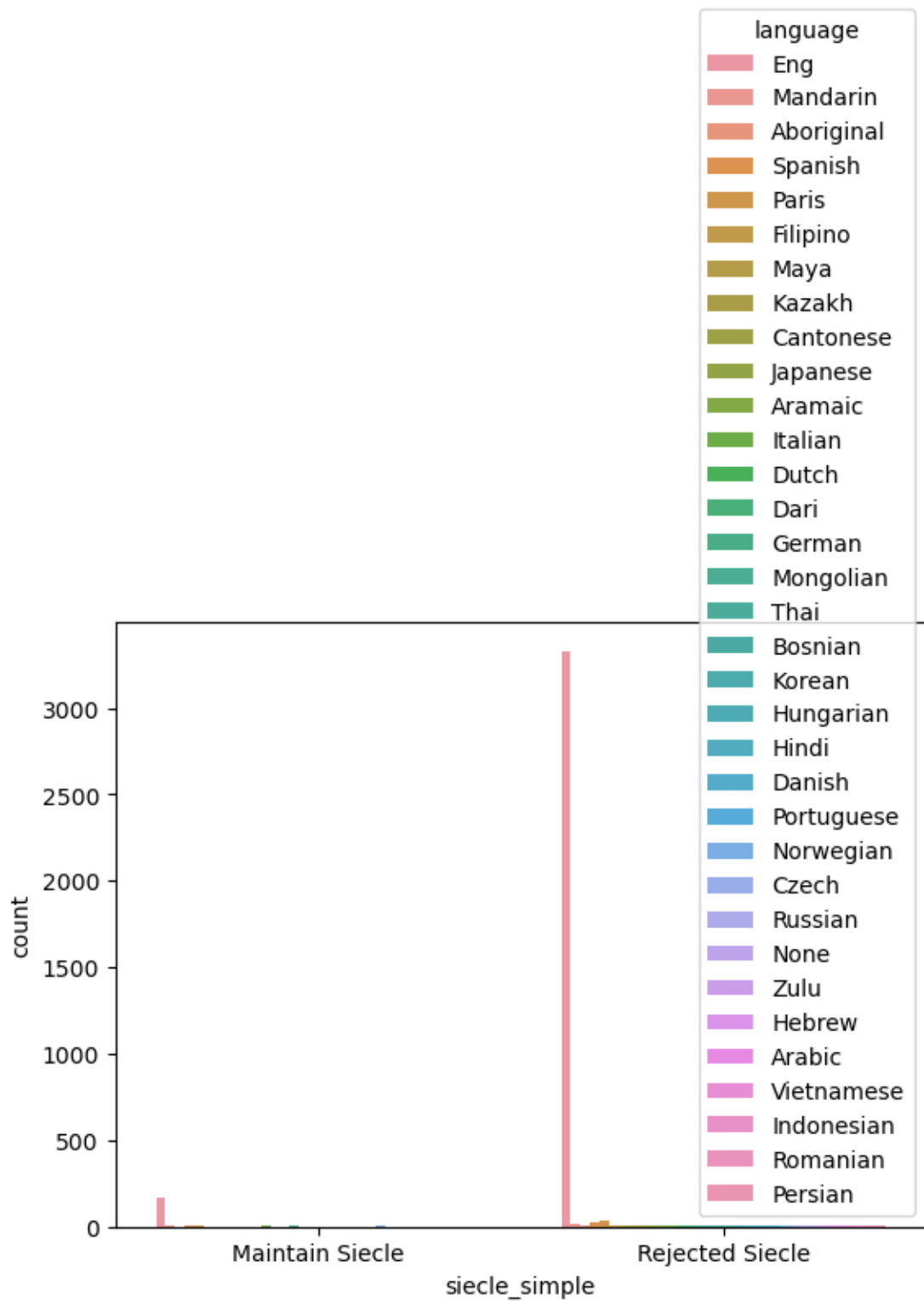


### Utilisation de countplot avec la categorisation hue

```
[36]: #compter les siecple_simple par genres
sns.countplot(data = movies, x = 'siecple_simple', hue = 'language')

#avec ce graphique nous remarquons que nous avons plusieurs datas, raison pour
↳ laquelle ce n'est pas visible
```

```
[36]: <Axes: xlabel='siecple_simple', ylabel='count'>
```



```
[40]: #Decoupons nos donnees sur base de language
```

```
movies_fr_en = movies[(movies['language'] == 'Eng') | (movies['language'] == 'Fr')]
movies_fr_en.head()
```

```
[40]:
```

	director_name	duration	gross	genres
0	james cameron	178.0	760505847.0	Action Adventure Fantasy Sci-Fi
1	gore verbinski	169.0	309404152.0	Action Adventure Fantasy
2	sam mendes	148.0	200074175.0	Action Adventure Thriller
3	christopher nolan	164.0	448130642.0	Action Thriller
5	andrew stanton	132.0	73058679.0	Action Adventure Sci-Fi

	title	budget	year	language
0	Avatar	237000000.0	2009	Eng
1	Pirates of the Caribbean: At World's End	300000000.0	2007	Eng
2	Spectre	245000000.0	2015	Eng
3	The Dark Knight Rises	250000000.0	2012	Eng
5	John Carter	263700000.0	2012	Eng

	score	surname	name	taux_rendu	siecle_simple
0	7.9	james	cameron	3.000000e+09	Maintain Siecle
1	7.1	gore	verbinski	4.225352e+09	Rejected Siecle
2	6.8	sam	mendes	3.602941e+09	Rejected Siecle
3	8.5	christopher	nolan	2.941176e+09	Rejected Siecle
5	6.6	andrew	stanton	3.995455e+09	Rejected Siecle

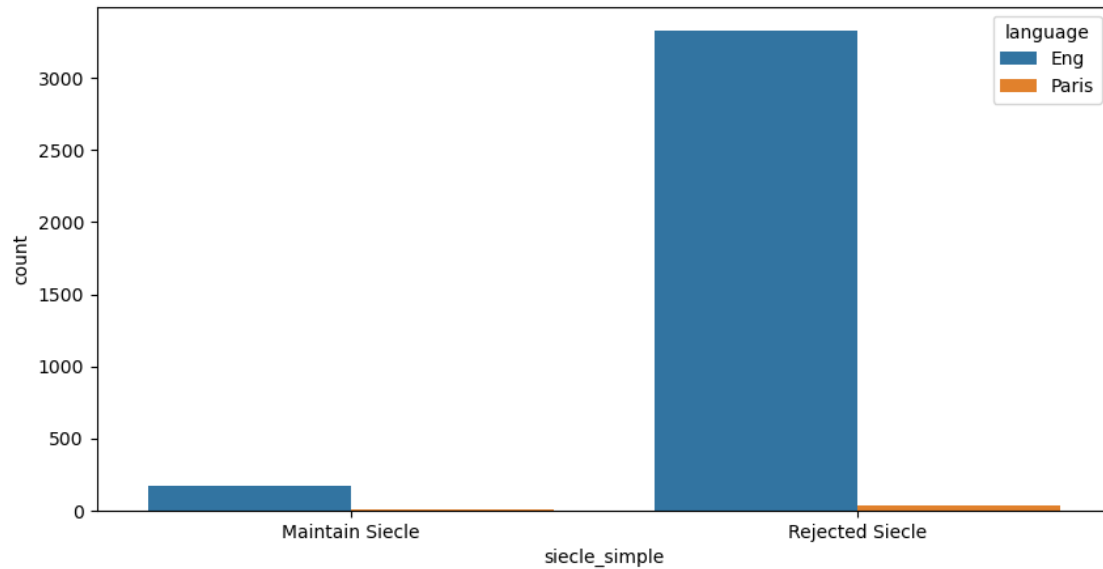
	complexe_choix	NewColonne
0	2008-2010	year_2009
1	Between 2055 and 2007	VIP
2	2014-2016	VIP
3	2011-2013	VIP
5	2011-2013	VIP

```
[42]: #voir la sitaution des siecle_simple par language
```

```
plt.figure(figsize = (10,5))

ax = sns.countplot(data = movies_fr_en, x = 'siecle_simple', hue = 'language')

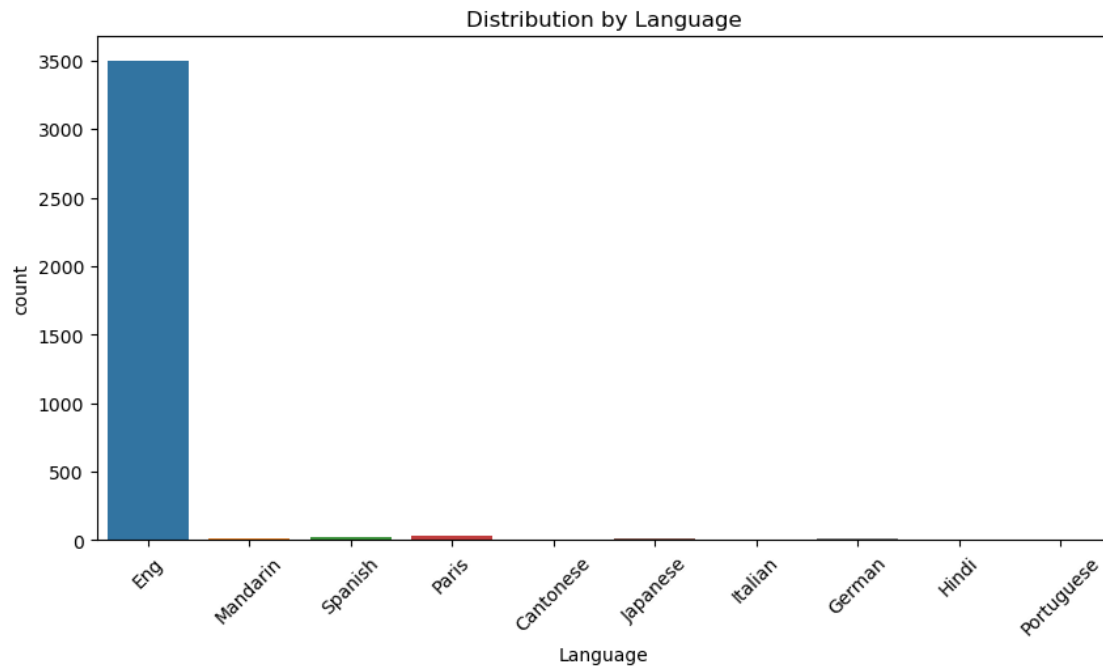
plt.show()
```



```
[50]: #utilisation de countplot et compter les top 10 des valeurs pour des
      ↪ observations nombreuses

top_10_language = movies['language'].value_counts().head(10)

plt.figure(figsize =(10,5))
sns.countplot(data = movies[movies['language'].isin(top_10_language.index)], x=
      ↪ 'language')
plt.xlabel('Language')
plt.ylabel('count')
plt.title('Distribution by Language')
plt.xticks(rotation =45)
#ENREGISTRER LE GRAPHIQUE
plt.savefig("language_distribution.jpg")
plt.show()
```



**2.3** “Un autre Projet de Visualisation avec Pandas et Seaborn est en cours d’implémentation, considérons ce projet comme un projet de data preprocessing, de nettoyage et de manipulation des datas...”

[ ]:

[ ]: