In [1]:

```python
import pandas as pd
budgets = pd.read_csv("tn.movie_budgets.csv")
budgets
```

Out[1]:

	id	release_date	movie	production_budget	domestic_gross	worldwide_gross
0	1	Dec 18, 2009	Avatar	$425,000,000	$760,507,625	$2,776,345,279
1	2	May 20, 2011	Pirates of the Caribbean: On Stranger Tides	$410,600,000	$241,063,875	$1,045,663,875
2	3	Jun 7, 2019	Dark Phoenix	$350,000,000	$42,762,350	$149,762,350
3	4	May 1, 2015	Avengers: Age of Ultron	$330,600,000	$459,005,868	$1,403,013,963
4	5	Dec 15, 2017	Star Wars Ep. VIII: The Last Jedi	$317,000,000	$620,181,382	$1,316,721,747
...	...	...	...	...	...	...
5777	78	Dec 31, 2018	Red 11	$7,000	$0	$0
5778	79	Apr 2, 1999	Following	$6,000	$48,482	$240,495
5779	80	Jul 13, 2005	Return to the Land of Wonders	$5,000	$1,338	$1,338
5780	81	Sep 29, 2015	A Plague So Pleasant	$1,400	$0	$0
5781	82	Aug 5, 2005	My Date With Drew	$1,100	$181,041	$181,041

5782 rows × 6 columns

In [2]:

```python
# Check for null values in a specific column
null_values = budgets['movie'].isnull().sum()

# This will give you the count of null values in the specified column
print("Number of null values in 'movie':", null_values)
```

```
Number of null values in 'movie': 0
```

In [3]:

```python
# Check for null values
null_values = budgets['release_date'].isnull().sum()

# This will give the count of null values
print("Number of null values in 'release_date':", null_values)
```

```
Number of null values in 'release_date': 0
```

In [4]:

```python
# Check for null values
null_values = budgets['worldwide_gross'].isnull().sum()

# This will give the count of null values
print("Number of null values in 'worldwide_gross':", null_values)
```

```
Number of null values in 'worldwide_gross': 0
```

In [5]:

```python
# Check for $0 values in the 'domestic_gross' column
zero_domestic_gross = budgets[budgets['domestic_gross'] == '$0']

# Display the rows where 'domestic_gross' is $0
print("Rows with $0 in 'domestic_gross':")
print(zero_domestic_gross)
```

```
Rows with $0 in 'domestic_gross':
      id  release_date                                           movie  \
194   95  Dec 31, 2020                                        Moonfall
479   80  Dec 13, 2017                                          Bright
480   81  Dec 31, 2019                                Army of the Dead
535   36  Feb 21, 2020                                Call of the Wild
617   18  Dec 31, 2012  AstÃ©rix et ObÃ©lix: Au service de Sa MajestÃ©
...   ..           ...                                             ...
5761  62  Dec 31, 2014                            Stories of Our Lives
5764  65  Dec 31, 2007                                     Tin Can Man
5771  72  May 19, 2015                                Family Motocross
5777  78  Dec 31, 2018                                          Red 11
5780  81  Sep 29, 2015                            A Plague So Pleasant

     production_budget domestic_gross worldwide_gross
194       $150,000,000             $0              $0
479        $90,000,000             $0              $0
480        $90,000,000             $0              $0
535        $82,000,000             $0              $0
617        $77,600,000             $0     $60,680,125
...                ...            ...             ...
5761           $15,000             $0              $0
5764           $12,000             $0              $0
5771           $10,000             $0              $0
5777            $7,000             $0              $0
5780            $1,400             $0              $0

[548 rows x 6 columns]
```

In [6]:

```python
# Check for $0 values in the 'worldwide_gross' column
zero_worldwide_gross = budgets[budgets['worldwide_gross'] == '$0']

# Display the rows where 'worldwide_gross' is $0
print("Rows with $0 in 'worldwide_gross':")
print(zero_worldwide_gross)
```

```
Rows with $0 in 'worldwide_gross':
      id  release_date                 movie production_budget domestic_gross
\
194   95  Dec 31, 2020              Moonfall      $150,000,000             $0
479   80  Dec 13, 2017                Bright       $90,000,000             $0
480   81  Dec 31, 2019      Army of the Dead       $90,000,000             $0
535   36  Feb 21, 2020      Call of the Wild       $82,000,000             $0
670   71  Aug 30, 2019             PLAYMOBIL       $75,000,000             $0
...   ..           ...                   ...               ...            ...
```

```
5761  62  Dec 31, 2014  Stories of Our Lives          $15,000              $0
5764  65  Dec 31, 2007           Tin Can Man          $12,000              $0
5771  72  May 19, 2015      Family Motocross          $10,000              $0
5777  78  Dec 31, 2018                Red 11           $7,000              $0
5780  81  Sep 29, 2015  A Plague So Pleasant           $1,400              $0

     worldwide_gross
194               $0
479               $0
480               $0
535               $0
670               $0
...              ...
5761              $0
5764              $0
5771              $0
5777              $0
5780              $0

[367 rows x 6 columns]
```

In [7]:

```python
# Filter out rows where 'worldwide_gross' is not equal to '$0'
budgets_cleaned = budgets[budgets['worldwide_gross'] != '$0']
budgets_cleaned.to_csv('updated_data.csv', index=False)

# Now, 'budgets_without_zero_worldwide_gross' contains the DataFrame with non
```

In [8]:

```python
import pandas as pd

# Read the CSV file into a DataFrame named 'budgets_cleaned2'
budgets_cleaned2 = pd.read_csv('updated_data.csv')

# Convert 'release_date' column to datetime format
budgets_cleaned2['release_date'] = pd.to_datetime(budgets_cleaned2['release_d

# Filter out rows where 'release_date' is greater than or equal to January 1,
filtered_budgets = budgets_cleaned2.loc[budgets_cleaned2['release_date'] > '1
```

In [ ]:

In [9]:

```python
# Remove duplicates from the 'primary_title' column
zero_duplicates = filtered_budgets.drop_duplicates(subset=['movie'])
```

In [10]:

```python
zero_duplicates
```

Out[10]:

	id	release_date	movie	production_budget	domestic_gross	worldwide_gross
0	1	2009-12-18	Avatar	$425,000,000	$760,507,625	$2,776,345,279
1	2	2011-05-20	Pirates of the Caribbean: On Stranger Tides	$410,600,000	$241,063,875	$1,045,663,875
2	3	2019-06-07	Dark Phoenix	$350,000,000	$42,762,350	$149,762,350

	id	release_date	movie	production_budget	domestic_gross	worldwide_gross
3	4	2015-05-01	Avengers: Age of Ultron	$330,600,000	$459,005,868	$1,403,013,963
4	5	2017-12-15	Star Wars Ep. VIII: The Last Jedi	$317,000,000	$620,181,382	$1,316,721,747
...	...	...	...	...	...	...
5410	76	2006-05-26	Cavite	$7,000	$70,071	$71,644
5411	77	2004-12-31	The Mongol King	$7,000	$900	$900
5412	79	1999-04-02	Following	$6,000	$48,482	$240,495
5413	80	2005-07-13	Return to the Land of Wonders	$5,000	$1,338	$1,338
5414	82	2005-08-05	My Date	$1,100	$181,041	$181,041

In [11]:

```
zero_duplicates.to_csv('budgets_cleaned.csv', index=False)
```

In [ ]:

```
```

In [ ]:

```
```