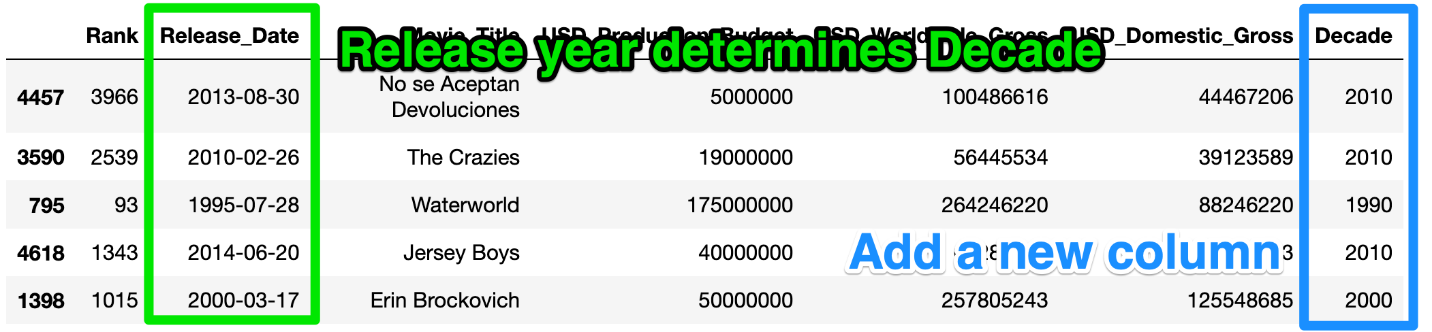
**Floor Division: A Trick to Convert Years to Decades**

In our bubble charts, we've seen how massively the industry has changed over time, especially from the 1970s onwards. This makes me think it makes sense to separate our films out by decade. Here's what I'm after:

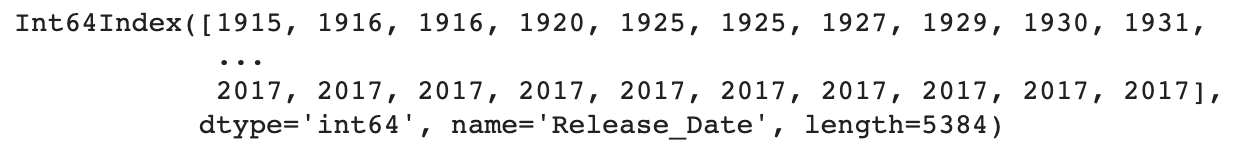


Challenge

Can you create a column in data\_clean that has the decade of the movie release. For example, a film released in 1992 or 1999 should have 1990 in the Decade column.

Here is one approach that you can follow:

1. Create a [DatetimeIndex object](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DatetimeIndex.html) from the Release\_Date column.



1. Grab all the years from the DatetimeIndex object using the .year property.
2. Use floor division // to convert the year data to the decades of the films.
3. Add the decades as a Decade column to the data\_clean DataFrame.

.

.

..

...

..

.

.

**Solution: Using Floor Division to Convert Years to Decades**

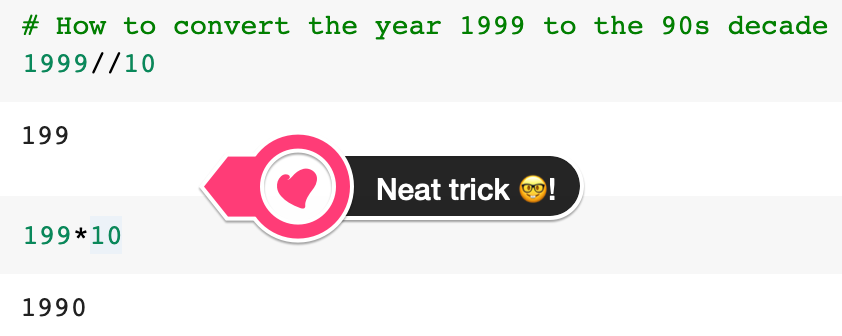
To create a DatetimeIndex, we just call the constructor and provide our release date column as an argument to initialise the DatetimeIndex object. Then we can extract all the years from the DatetimeIndex.

1. dt\_index = pd.DatetimeIndex(data\_clean.Release\_Date)
2. years = dt\_index.year

Now, all we need to do is convert the years to decades. For that, we will use floor division (aka integer division). The difference to regular division is that the result is effectively rounded down.

1. 5.0 / 2
2. # output: 2.5
3. 5.0 // 2
4. # output: 2.0

In our case, we will use the floor division by 10 and then multiplication by 10 to convert the release year to the release decade:



We can do this for all the years and then add the decades back as a column.

1. decades = years//10\*10
2. data\_clean['Decade'] = decades

Challenge

Create two new DataFrames: old\_films and new\_films

* old\_films should include all the films before 1970 (up to and including 1969)
* new\_films should include all the films from 1970 onwards
* How many of our films were released prior to 1970?
* What was the most expensive film made prior to 1970?

.

.

..

...

..

.

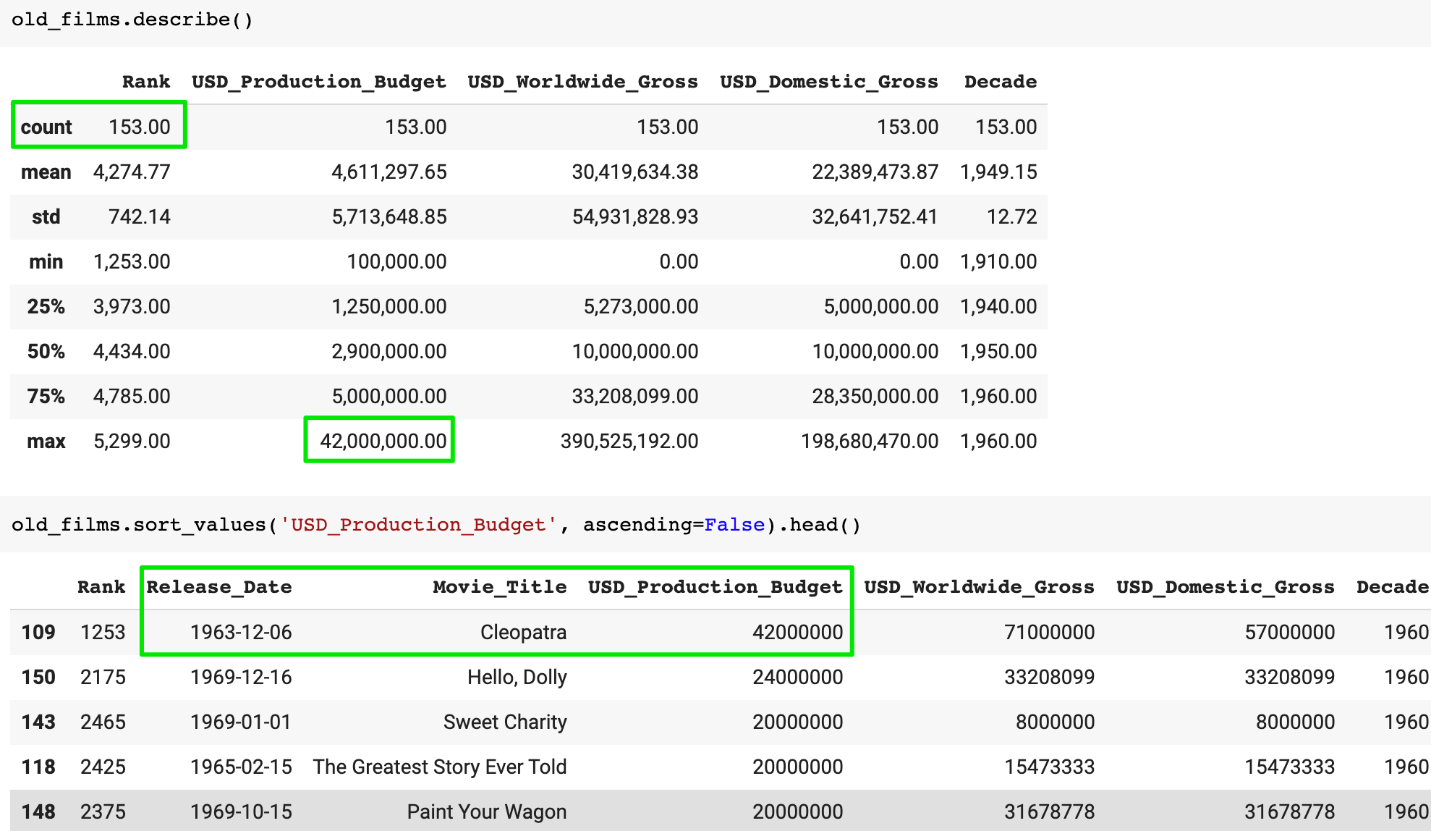
.

**Solution: Separate the films made before and after 1970**

Now that we have our Decades column we can use it to create subsets of our data.

1. old\_films = data\_clean[data\_clean.Decade <= 1960]
2. new\_films = data\_clean[data\_clean.Decade > 1960]

The cut-off for our calculation is 1960 in the Decade column because this will still include 1969. When we inspect our old\_films DataFrame we see that it only includes 153 films. As we saw in the bubble chart, the bulk of films in the dataset have been released in the last 30 years.



The most expensive film prior to 1970 was Cleopatra, with a production budget of $42 million. That's some serious 1960s money, and judging by the trailer, a lot of it went into extravagant costumes, set design, and plenty of extras. Impressive.

