Activity: Selective Subsets

Introduction

In this activity you will practice selecting subsets of data from a DataFrame using Pandas. This activity will cover the following topics:

- Creating masks
- Negating masks
- Masks with slicing
- Null value masks

Question 1

Create a `DataFrame` called `df` from the given CSV file `movie_data.csv`, and then create a mask called `before_millennium` to select all movies that were released before 2000.

In [ ]:
```
import pandas as pd

# Your code here

df = pd.read_csv('movie_data.csv')

before millennium = df['Year Released'] > 2000

#before_millennium.head()

df.head()
```

In [ ]:
```
# Question 1 Grading Checks

assert isinstance(df, pd.DataFrame), 'Did you create a DataFrame called df?'
```

Question 2

Using the `before_millennium` mask from Question 1, assign the titles of every movie that was released after 2000 to a `Series` called `newer_titles`.

In [ ]:
```
# Your code here

newer_titles = df[~before_millennium]['Title']

#newer_titles.head()
```

In [ ]:
```
# Question 2 Grading Checks

assert isinstance(newer_titles, pd.Series), 'Did you create a Series called newer_titles?'
```

Question 3

Create a mask to select movies with a `Rating` of `8.9` and a `Box Office ($M)` value higher than `1000.0`. Assign the resulting `Series` to a variable called `popular_pg_movies`.

In [ ]:
```
# Your code here

popular_pg_movies = (df['Rating'] == 8.9) & (df['Box Office ($M)']  > 1000.0)
```

In [ ]:
```
# Question 3 Grading Checks

assert isinstance(popular_pg_movies, pd.Series), 'Did you create a Series called popular_pg_movies?'
```

Question 4

Create a mask to select movies with a null value for `Box Office ($M)` or `Rating`. Assign the resulting `Series` to a variable called `missing_info`.

In [ ]:
```
# Your code here

missing_info = ()
```

In [ ]:
```
# Question 4 Grading Checks

assert isinstance(missing_info, pd.Series), 'Did you create a Series called missing_info?'
```