

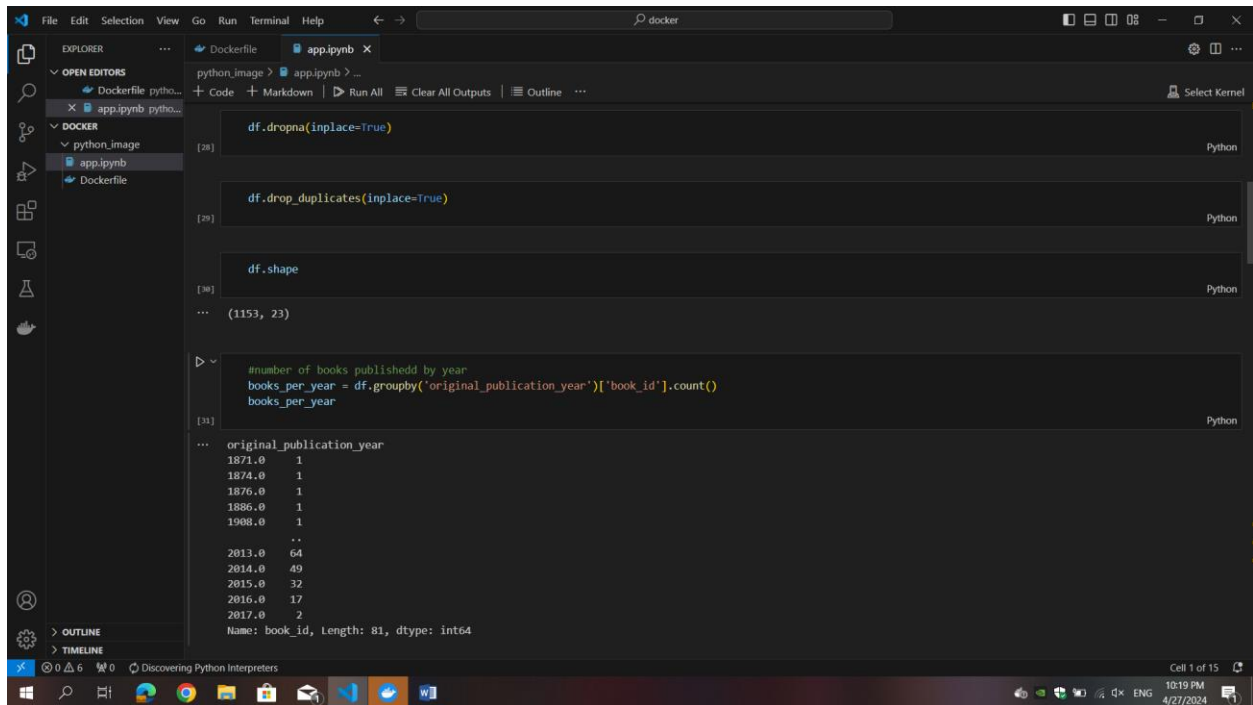
Assignment 2  
Carine Emad Sanad  
20221440878

Side Note:

Code is correct but it doesn't run this is the best I could made

I first uploaded dataset, then I created notebook

I cleaned, analyzed data



The screenshot shows a Jupyter Notebook with the following code cells:

```
[28] df.dropna(inplace=True)
```

```
[29] df.drop_duplicates(inplace=True)
```

```
[30] df.shape
```

```
[31] (1153, 23)
```

```
[31] #number of books publishedd by year
books_per_year = df.groupby('original_publication_year')['book_id'].count()
books_per_year
```

The output of the last cell is a series showing the count of books published per year:

original_publication_year	count
1871.0	1
1874.0	1
1876.0	1
1886.0	1
1908.0	1
...	...
2013.0	64
2014.0	49
2015.0	32
2016.0	17
2017.0	2

The notebook also shows the variable `books_per_year` with its data type: `Name: book_id, Length: 81, dtype: int64`.

Then I created Dockerfile

