```
from google.colab import drive

drive.mount('/content/gdrive')
```

```
Mounted at /content/gdrive
```

```
import pandas as pd
df = pd.read_csv('/content/gdrive/My Drive/pandas practice/books.csv')
print(df.shape)
print(df.info())
print(df.iloc[0:5,1:4])
```

```
(5000, 10)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5000 entries, 0 to 4999
Data columns (total 10 columns):
 #   Column              Non-Null Count  Dtype  
---  ------              --------------  -----  
 0   title               5000 non-null   object 
 1   author              5000 non-null   object 
 2   average_rating      5000 non-null   float64
 3   isbn13              5000 non-null   int64  
 4   language_code       4923 non-null   object 
 5   num_pages           4967 non-null   float64
 6   ratings_count       5000 non-null   int64  
 7   text_reviews_count  5000 non-null   int64  
 8   publisher           5000 non-null   object 
 9   year                5000 non-null   int64  
dtypes: float64(2), int64(4), object(4)
memory usage: 390.8+ KB
None
                        author  average_rating         isbn13
0                 Victor Hugo            4.00  9780345472427
1  Lidia Matticchio Bastianich            4.14  9780767914222
2             James Scott Bell            4.00  9780310243878
3         David Foster Wallace            3.86  9780349111889
4          Laurell K. Hamilton            4.13  9780425209066
```

```
import pandas as pd
df = pd.read_csv('/content/gdrive/My Drive/pandas practice/books.csv')
books_gt_4 = df[df['average_rating'] > 4]
count_books_gt_4 = len(books_gt_4)

print(count_books_gt_4)
```

```
2115
```

```
mean_rating_eng = df.groupby('language_code')['average_rating'].mean()['eng']

print(mean_rating_eng)
```

```
3.931142857142857
```

```
import pandas as pd
df = pd.read_csv('/content/gdrive/My Drive/pandas practice/books.csv')
print(df.loc[[1,4],['language_code', 'num_pages', 'ratings_count']])
```

```
  language_code  num_pages  ratings_count
1         en-GB      288.0             48
4         en-US      368.0          59524
```

```
median_pages = df['num_pages'].median()
print(f"Median value of num_pages: {median_pages}")
max_pages = df['num_pages'].max()
print(f"Maximum value of num_pages: {max_pages}")
mode_pages = df['num_pages'].mode()[0]  # Assuming there's a single mode
print(f"Mode of num_pages: {mode_pages}")
mean_pages = df['num_pages'].mean()
print(f"Mean value of num_pages: {mean_pages}")
```

```
Median value of num_pages: 304.0
Maximum value of num_pages: 3342.0
Mode of num_pages: 288.0
Mean value of num_pages: 341.0940205355345
```

```
import pandas as pd
df = pd.read_csv('/content/gdrive/My Drive/pandas practice/books.csv')

# Check the column names in your dataframe
print(df.columns)

# Use 'year' as the column name
year_counts = df['year'].value_counts()
print(year_counts.get(2010, 0))  # Print 0 if 2010 is not a key

# Split author names (assuming they are comma-separated)
df['authors_list'] = df['author'].str.split(' & ') # 'author' was the correct column name

# Filter for DAW publications and average rating > 4
daw_gt_4_authors = df[
    (df['publisher'] == 'DAW') & (df['average_rating'] > 4)
]['authors_list'].explode()

# Remove duplicates and sort alphabetically (optional)
unique_authors = daw_gt_4_authors.unique()
unique_authors.sort() # Use .sort() for NumPy arrays

# Print result (assuming there are no empty author names in the list)
print(', '.join(unique_authors))
```

```
Index(['title', 'author', 'average_rating', 'isbn13', 'language_code',
       'num_pages', 'ratings_count', 'text_reviews_count', 'publisher',
       'year'],
      dtype='object')
15
Mercedes Lackey, Sharon Green
```

```
long_books_after_2010 = df[(df['num_pages'] > 500) & (df['year'] > 2010)]
count_books = long_books_after_2010.shape[0]

print(f"Number of books with more than 500 pages published after 2010: {count_books}")

agatha_books = df[df['author'].str.contains("Agatha Christie")]
count_books = agatha_books.shape[0]

print(f"Number of books by Agatha Christie: {count_books}")
```

```
Number of books with more than 500 pages published after 2010: 5
Number of books by Agatha Christie: 21
```

```
#Filter for Spanish books published after 2000
spanish_after_2000 = df[(df['language_code'] == 'spa') & (df['year'] > 2000)]
count_books = spanish_after_2000.shape[0]

# Print result
print(f"Number of Spanish books published after 2000: {count_books}")
```

```
Number of Spanish books published after 2000: 84
```

Start coding or generate with AI.