Advanced Python - Module 3 Project - Jake Duffin

Bob's Bookstore Scraper and Analysis

In [2]:
```
import pandas as pd
import requests
from bs4 import BeautifulSoup
```

In [5]:
```
response = requests.get('https://btech-data-analytics.github.io/bridgerland-technic
soup = BeautifulSoup(response, 'html.parser')
```

In [16]:
```
soup.find('table').find_all('tr', class_="book")[0].find_all('td')[1].text
```

Out[16]:
```
'Whiskers of Wisdom: Tales from Feline Philosophers'
```

In [28]:
```
isbn = []
title = []
author = []
language = []
pages = []
topic = []
price = []


for book in soup.find('table').find_all('tr', class_="book"):
  isbn.append(book.find_all('td')[0].text)
  title.append(book.find_all('td')[1].text)
  author.append(book.find_all('td')[2].text)
  language.append(book.find_all('td')[3].text)
  pages.append(book.find_all('td')[4].text)
  topic.append(book.find_all('td')[5].text)
  price.append(book.find_all('td')[6].text.strip('$'))
```

In [30]:
```
df = pd.DataFrame({
    'isbn' : isbn,
    'title' : title,
    'author' : author,
    'language' : language,
    'pages' : pages,
    'topic' : topic,
    'price' : price,

})
```

In [32]:
```
df = df.astype({
    'pages' : 'int64',
    'price' : 'float64'
})
```

```
print(df.info())
df.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15 entries, 0 to 14
Data columns (total 7 columns):
 #   Column    Non-Null Count  Dtype  
---  ------    --------------  -----  
 0   isbn      15 non-null     object 
 1   title     15 non-null     object 
 2   author    15 non-null     object 
 3   language  15 non-null     object 
 4   pages     15 non-null     int64  
 5   topic     15 non-null     object 
 6   price     15 non-null     float64
dtypes: float64(1), int64(1), object(5)
memory usage: 972.0+ bytes
None
```

Out[32]:

	isbn	title	author	language	pages	topic	price
0	978-1234567890	Whiskers of Wisdom: Tales from Feline Philosop...	Penelope Wainwright	English	256	Cats	19.99
1	978-2345678901	Purrfectly Pawesome: A Cat's Life	Jasper Sterling	English	192	Cats	15.99
2	978-3456789012	Cat Tales: Adventures in Whiskerland	Penelope Wainwright	English	320	Cats	21.99
3	978-4567890123	The Enigmatic Paws: Mysteries of Meowville	Maximilian Thorne	English	288	Cats	17.99
4	978-5678901234	Cats in Wonderland	Isadora Harrington	English	224	Cats	16.99

In [35]:

```
# Question 1

df['author'].value_counts()
```

Out[35]:

author	count
Penelope Wainwright	4
Jasper Sterling	2
Benjamin Barkley	2
Maximilian Thorne	1
Celeste Nightshade	1
Isadora Harrington	1
Seraphina Montague	1
Sophie Shepherd	1
Oliver Obedience	1
Ruby Ruffington	1

dtype: int64

In [38]:
```
# Question 2

df['topic'].value_counts()
```

Out[38]:

topic	count
Cats	10
Dogs	5

dtype: int64

In [40]:
```
# Question 3

av_top_price = df.groupby(['topic'], as_index=False).agg({'price' : 'mean'})
av_top_price
```

Out[40]:

	topic	price
0	Cats	17.79
1	Dogs	26.59

In [41]:
```
# Question 4
```

```
av_top_page = df.groupby(['topic'], as_index=False).agg({'pages' : 'mean'})
av_top_page
```

Out[41]:

	topic	pages
0	Cats	238.4
1	Dogs	256.0