



Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ «ИНЖЕНЕРНЫЙ БИЗНЕС И МЕНЕДЖМЕНТ»
КАФЕДРА «ПРОМЫШЛЕННАЯ ЛОГИСТИКА»

ДОМАШНЕЕ ЗАДАНИЕ
по дисциплине «Парадигмы и конструкции
языков программирования»

Студент

ИБМ3-34Б

А.И. Якубович

(Группа)

(Подпись, дата)

(И.О. Фамилия)

Преподаватель

Ю.Е. Гапанюк

(Подпись, дата)

(И.О. Фамилия)

Задание:

1. Осуществите разбор данных сайта с использованием библиотеки BeautifulSoup.
2. Сформируйте датасет табличного типа с использованием Python и сохраните датасет в формате CSV.
3. Проведите разведочный анализ данных для сформированного датасета.

Код программы:

```
import requests
from bs4 import BeautifulSoup
import csv
from datetime import datetime
import pandas as pd

url = 'http://books.toscrape.com/'
response = requests.get(url)
if response.status_code == 200:
    print("Страница успешно загружена!")
else:
    print(f"Ошибка! Код: {response.status_code}")
    exit()

soup = BeautifulSoup(response.text, 'lxml')
books = soup.find_all('article', class_='product_pod')
books_data = []
for book in books:
    title = book.h3.a['title']
    price = book.find('p', class_='price_color').text
    rating_class = book.find('p', class_='star-rating')['class']
    rating_map = {'One': 1, 'Two': 2, 'Three': 3, 'Four': 4, 'Five': 5}
    rating = rating_map.get(rating_class[1], 'N/A')
    link = url + book.h3.a['href']
    books_data.append({
        'Title': title,
        'Price': price,
        'Rating': rating,
        'Link': link
    })
df = pd.DataFrame(books_data)
print(df.head())
df.to_csv('C:\Users\Anna\OneDrive\Desktop\Учёба Бауманка\2 курс\Пяп\дз\books_dataset.csv', index=False,
encoding='utf-8')
print("Датасет успешно сохранен в файл 'books_dataset.csv'")

# 3. ПРОСТОЙ АНАЛИЗ ДАННЫХ
print("\n==== ПРОСТОЙ АНАЛИЗ ===")
rating_names = {1: '★', 2: '★★', 3: '★★★', 4: '★★★★', 5: '★★★★★'}
print("\nКоличество книг по рейтингам:")
rating_counts = df['Rating'].value_counts().sort_index()
for rating, count in rating_counts.items():
    stars = rating_names.get(rating, str(rating))
    print(f"{stars}: {count} книг")
print("\n--- АНАЛИЗ ЦЕН ---")
print(f"Всего книг: {len(df)}")
# Преобразуем Price из строк в числа (убираем £)
df['Price'] = df['Price'].str.replace('£', '').astype(float)
```

```

# Теперь можно считать
print(f"Общая стоимость: {df['Price'].sum():.2f}")
print(f"Средняя цена: {df['Price'].mean():.2f}")
print(f"Самая дорогая: {df['Price'].max():.2f}")
print(f"Самая дешевая: {df['Price'].min():.2f}")
print("\n==== АНАЛИЗ ЗАВЕРШЕН ===")

```

Результат работы программы:

```

Страница успешно загружена!
          Title    Price  Rating
0      A Light in the Attic  £51.77      3
1      Tipping the Velvet  £53.74      1
2      Submission  £50.10      1
3      Sharp Objects  £47.82      4
4 Sapiens: A Brief History of Humankind  £54.23      5
Датасет успешно сохранен в файл 'books_dataset.csv'

==== ПРОСТОЙ АНАЛИЗ ===

Количество книг по рейтингам:
★: 6 книг
★★: 3 книги
★★★: 3 книги
★★★★: 4 книги
★★★★★: 4 книги

--- АНАЛИЗ ЦЕН ---
Всего книг: 20
Общая стоимость: £760.97
Средняя цена: £38.05
Самая дорогая: £57.25
Самая дешевая: £13.99

==== АНАЛИЗ ЗАВЕРШЕН ===

```