

ОТЧЕТ

По Домашнему заданию

Дисциплина «Парадигмы и конструкции языков программирования»

Студент: Коваленко Е. Ю.

Группа: ИБМ3-34Б

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.

Был проведен разбор данных сайта <https://quotes.toscrape.com/>

Текст программы:

```
import requests

import pandas as pd

from bs4 import BeautifulSoup

url = "https://quotes.toscrape.com/"

response = requests.get(url)

soup = BeautifulSoup(response.text, "html.parser")

quotes_data = []

for quote in soup.find_all("div", class_="quote"):

    text = quote.find("span", class_="text").text

    author = quote.find("small", class_="author").text

    tags = " , ".join(tag.text for tag in quote.find_all("a", class_="tag"))

    quotes_data.append({

        "quote": text,

        "author": author,

        "tags": tags
```

```
}}
```

```
df = pd.DataFrame(quotes_data)
```

```
df.to_csv("quotes.csv", index=False, encoding="utf-8")
```

```
print("Первые строки датасета:")
```

```
print(df.head(), "\n")
```

```
print("Размер датасета:")
```

```
print(df.shape, "\n")
```

```
print("Количество цитат по авторам:")
```

```
print(df["author"].value_counts())
```

Первые строки датасета:

```
          quote ...
tags
0  "The world as we have created it is a process ...
change, deep-thoughts, thinking, world
1  "It is our choices, Harry, that show what we t...
abilities, choices
2  "There are only two ways to live your life. On...
inspirational, life, live, miracle, miracles
3  "The person, be it gentleman or lady, who has ...
aliteracy, books, classic, humor
4  "Imperfection is beauty, madness is genius and...
be-yourself, inspirational
```

[5 rows x 3 columns]

Размер датасета:

(10, 3)

Количество цитат по авторам:

```
Albert Einstein      3
J.K. Rowling         1
Jane Austen          1
Marilyn Monroe       1
André Gide           1
Thomas A. Edison     1
Eleanor Roosevelt   1
Steve Martin         1
Name: author, dtype: int64
```

