Максименко София ИБМ3-34Б

Домашнее задание

Код на python

import pandas as pd

from bs4 import BeautifulSoup

import requests

import seaborn as sns

import matplotlib.pyplot as plt

url = 'https://webscraper.io/test-sites/e-commerce/allinone/computers/laptops'

response = requests.get(url) #проверяем работу requests

soup = BeautifulSoup(response.text, 'lxml') #выводим исходный код страницы

print(soup)

print(response)

'''

print(soup.header.p.string)

prices = soup.find('h4', class\_ = 'price float-end card-title pull-right')

print(prices.text)

names = soup.find('a', class\_= 'title')

print(names.text)

'''

#выбираем нужные атрибуты (название, цена, отзывы, описание товара)

names = soup.find\_all('a', class\_= 'title')

prices = soup.find\_all('h4', class\_ = 'price float-end card-title pull-right')

reviews = soup.find\_all('p', class\_ = 'review-count float-end')

descriptions = soup.find\_all('p', class\_ = 'description card-text')

#создаем списки и заносим в них данные

names\_list = []

for i in names:

    name = i.text

    names\_list.append(name)

prices\_list = []

for i in prices:

    price = i.text

    prices\_list.append(price)

reviews\_list = []

for i in reviews:

    review = i.text

    reviews\_list.append(review)

descriptions\_list = []

for i in descriptions:

    description = i.text

    descriptions\_list.append(description)

#print(names\_list, prices\_list, reviews\_list, descriptions\_list, sep = '\n' )

#создаем и сохраняем фрейм данных в таблице

data = pd.DataFrame({'Наименование товара': names\_list,

 'Цена': prices\_list,

 'Количество отзывов': reviews\_list,

 'Описание': descriptions\_list})

data.to\_csv('Ассортимент\_ноутбуков.csv', index=False, encoding='utf-8')

print('Данные успешно сохранены в файл Ассортимент\_ноутбуков.csv')

#Разведочный анализ данных

#Диаграмма рассеяния

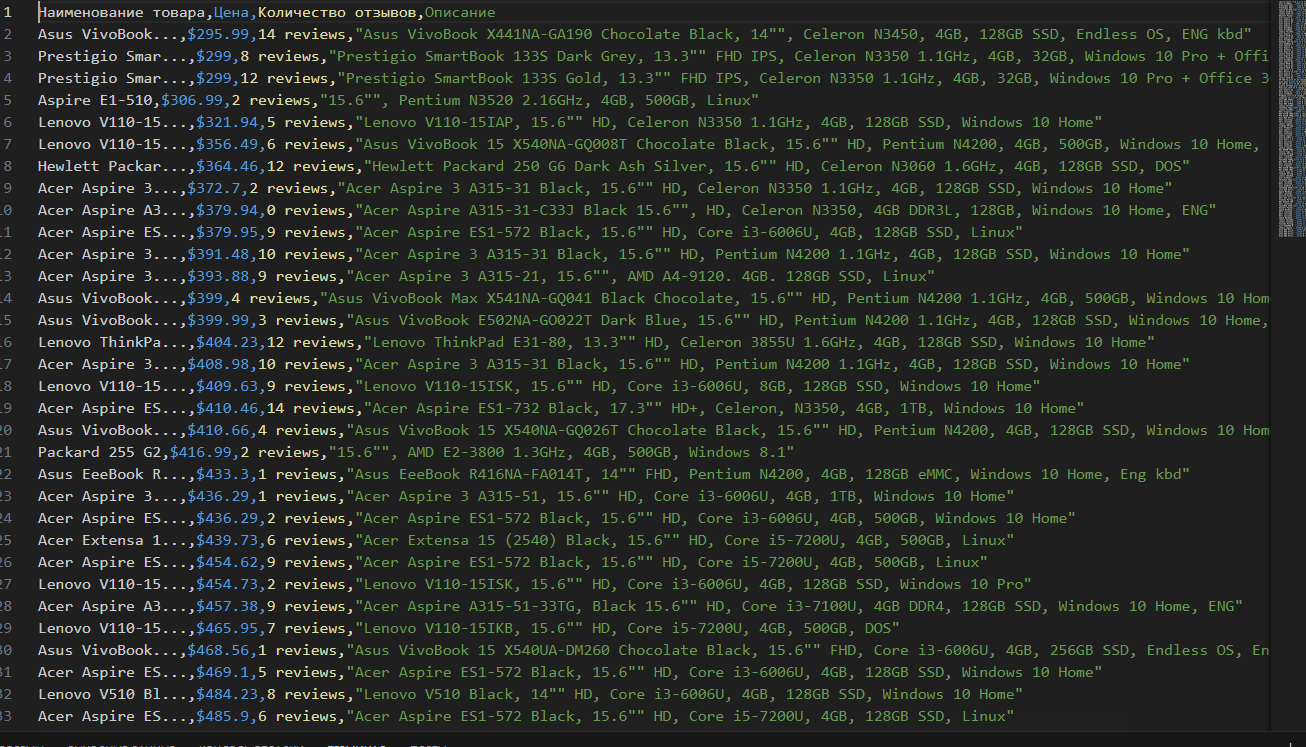
data = pd.read\_csv('Ассортимент\_ноутбуков.csv', sep=",")

fig, ax = plt.subplots(figsize=(10,10))

sns.scatterplot(ax=ax, x='Цена', y='Количество отзывов', data=data)

plt.show()

Создание файла Ассортимент\_ноутбуков.csv



Код для диаграмм

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

#sns.set(style="ticks")

#Разведочный анализ данных

data = pd.read\_csv('Ассортимент\_ноутбуков.csv', sep=",")

#Диаграмма рассеяния

fig, ax = plt.subplots(figsize=(10,10))

sns.scatterplot(ax=ax, x='Цена', y='Количество отзывов', data=data)

plt.show()

#Гистограмма

fig, ax = plt.subplots(figsize=(10,10))

sns.displot(data['Цена'])

plt.show()

#Ящик с усами

sns.boxplot(x='Цена', y='Количество отзывов', data=data)

plt.show()

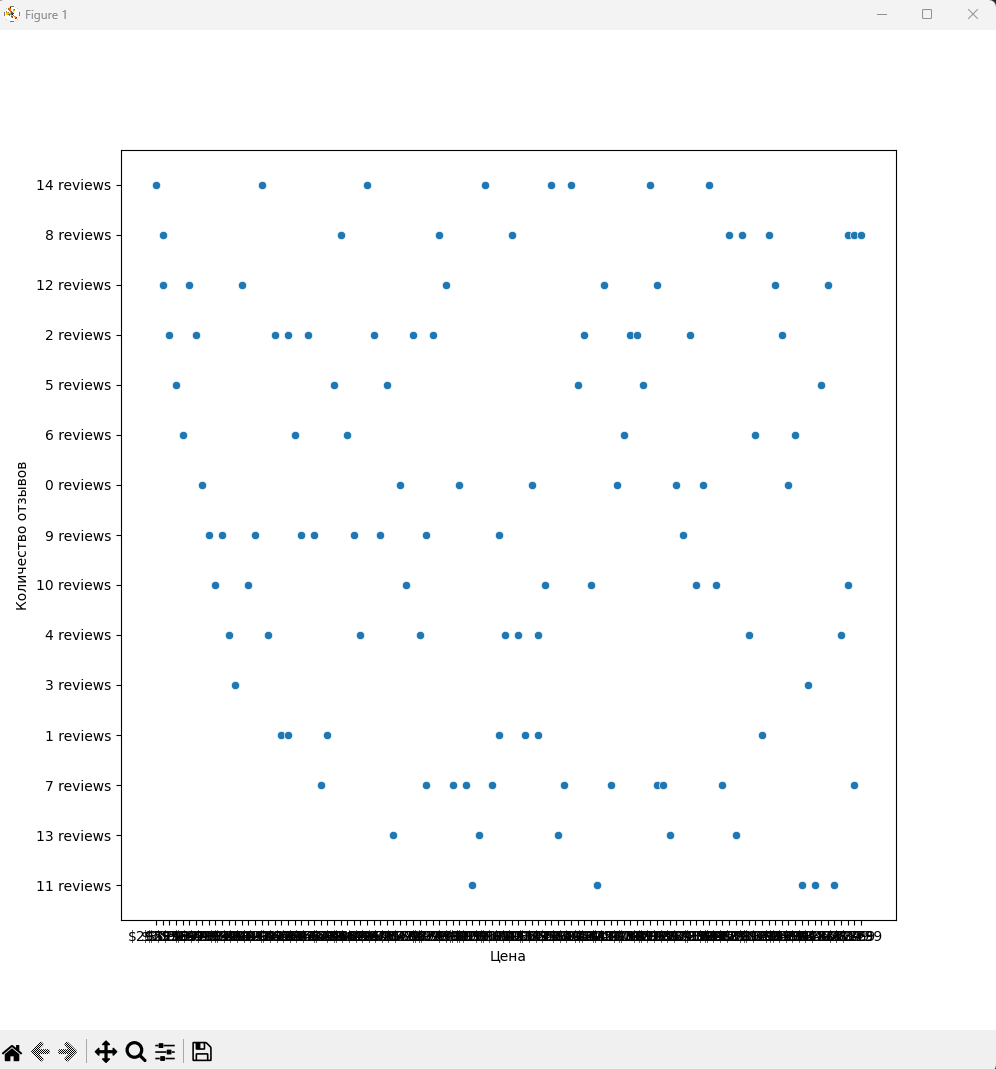
#Violin plot

sns.violinplot(x='Цена', y='Количество отзывов', data=data)

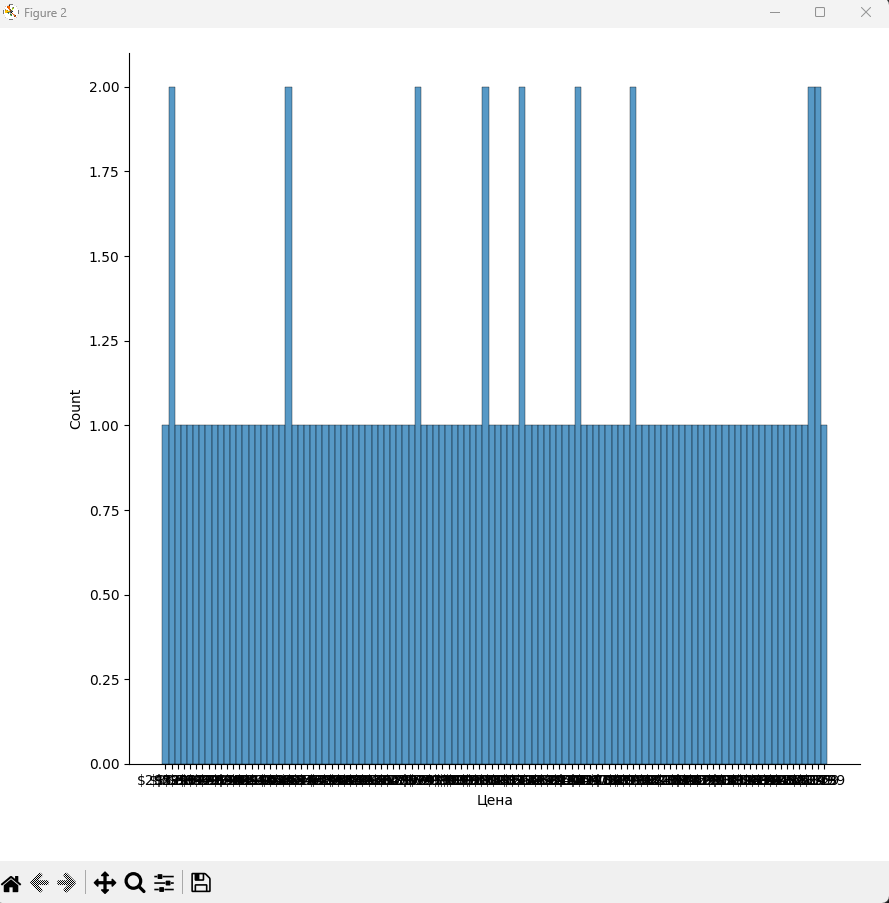
plt.show()

Разведочный анализ данных

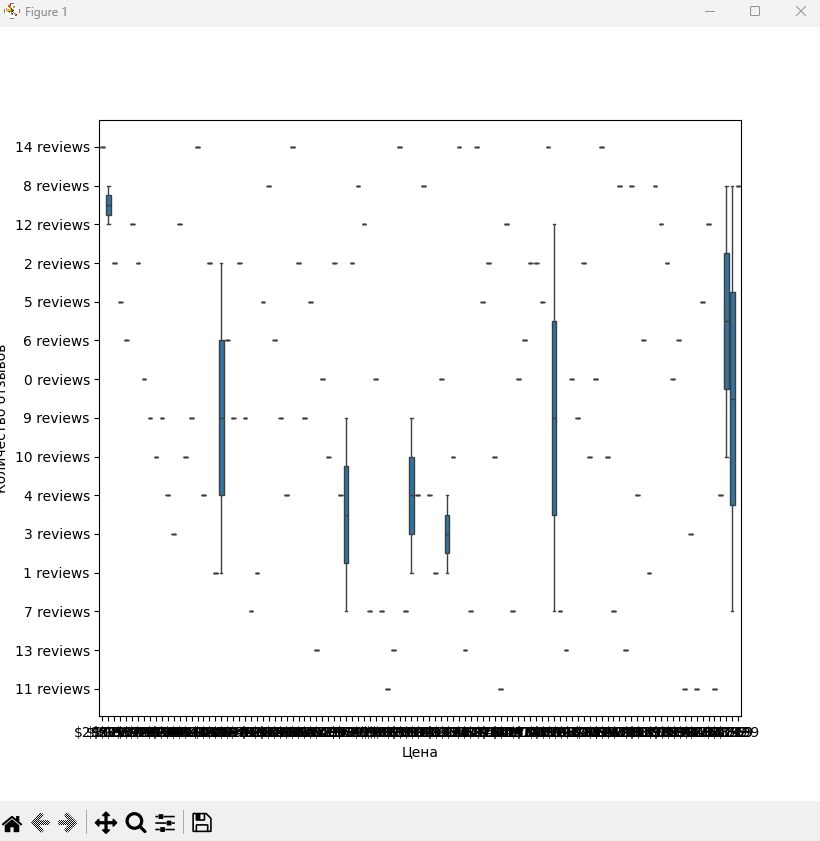
Диаграмма рассеивания



Гистограмма



Ящик с усами



Violin plot

