SNS COLLEGE OF TECHNOLOGY

Coimbatore-35
An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A++' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

Redesigning Common Mind & Business Towards Excellence

Build an Entrepreneurial Mindset Through Our Design Thinking FrameWork

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

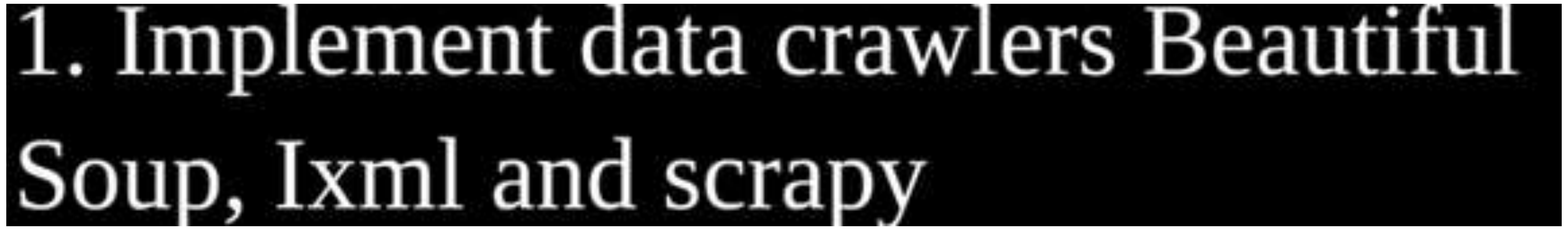
23AMB201 - MACHINE LEARNING

II YEAR IV SEM

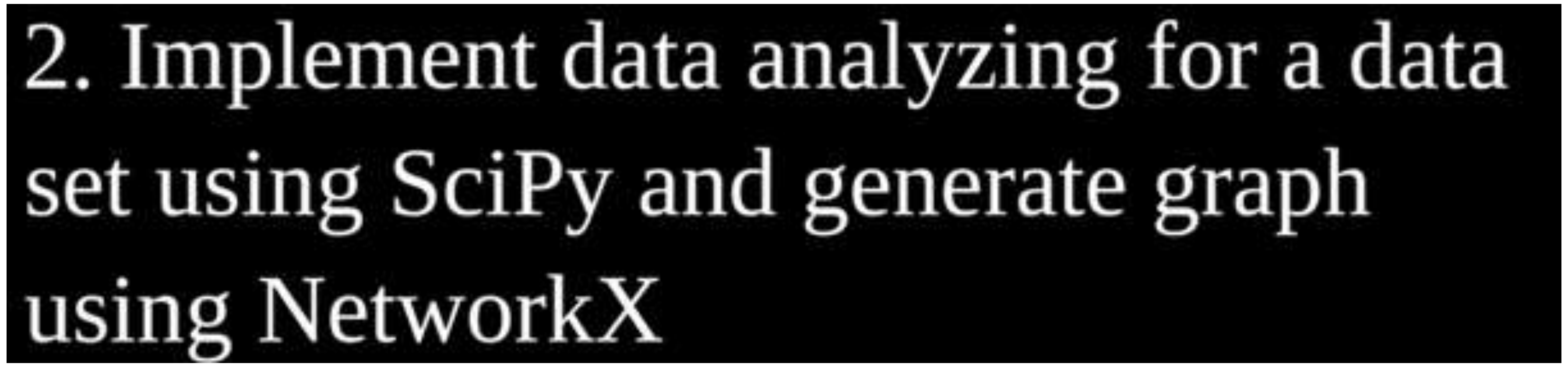
UNIT I – INTRODUCTION

Exercise 1 & 2

Programs



1. Implement data crawlers Beautiful Soup, Ixml and scrapy

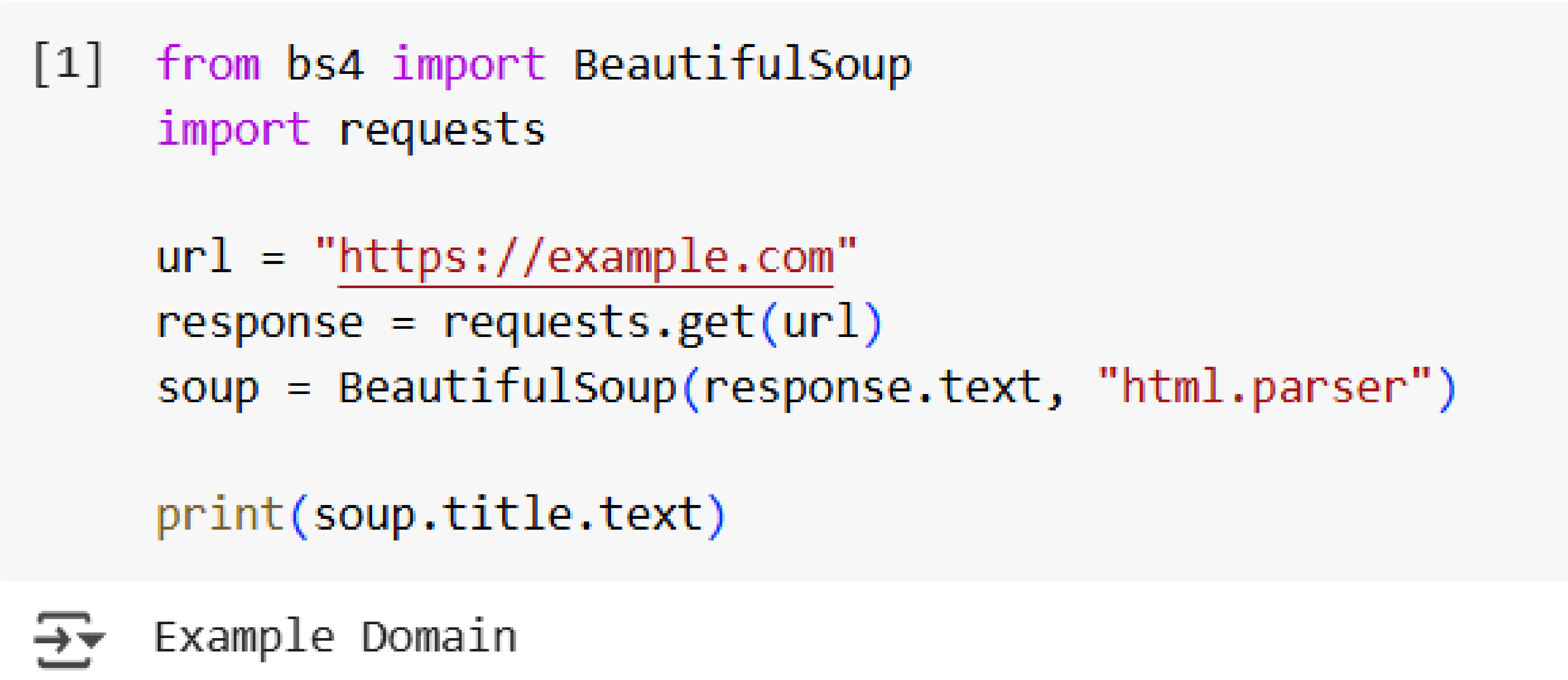


2. Implement data analyzing for a data set using SciPy and generate graph using NetworkX

1. BeautifulSoup, lxml and scrapy –Data Crawlers

```
pip install requests
pip install html5lib
pip install bs4
```

```
import requests
import requests
URL = "https://www.geeksforgeeks.org"
r = requests.get(URL)
print(r.content)
```



```
[1] from bs4 import BeautifulSoup
    import requests

    url = "https://example.com"
    response = requests.get(url)
    soup = BeautifulSoup(response.text, "html.parser")

    print(soup.title.text)
```

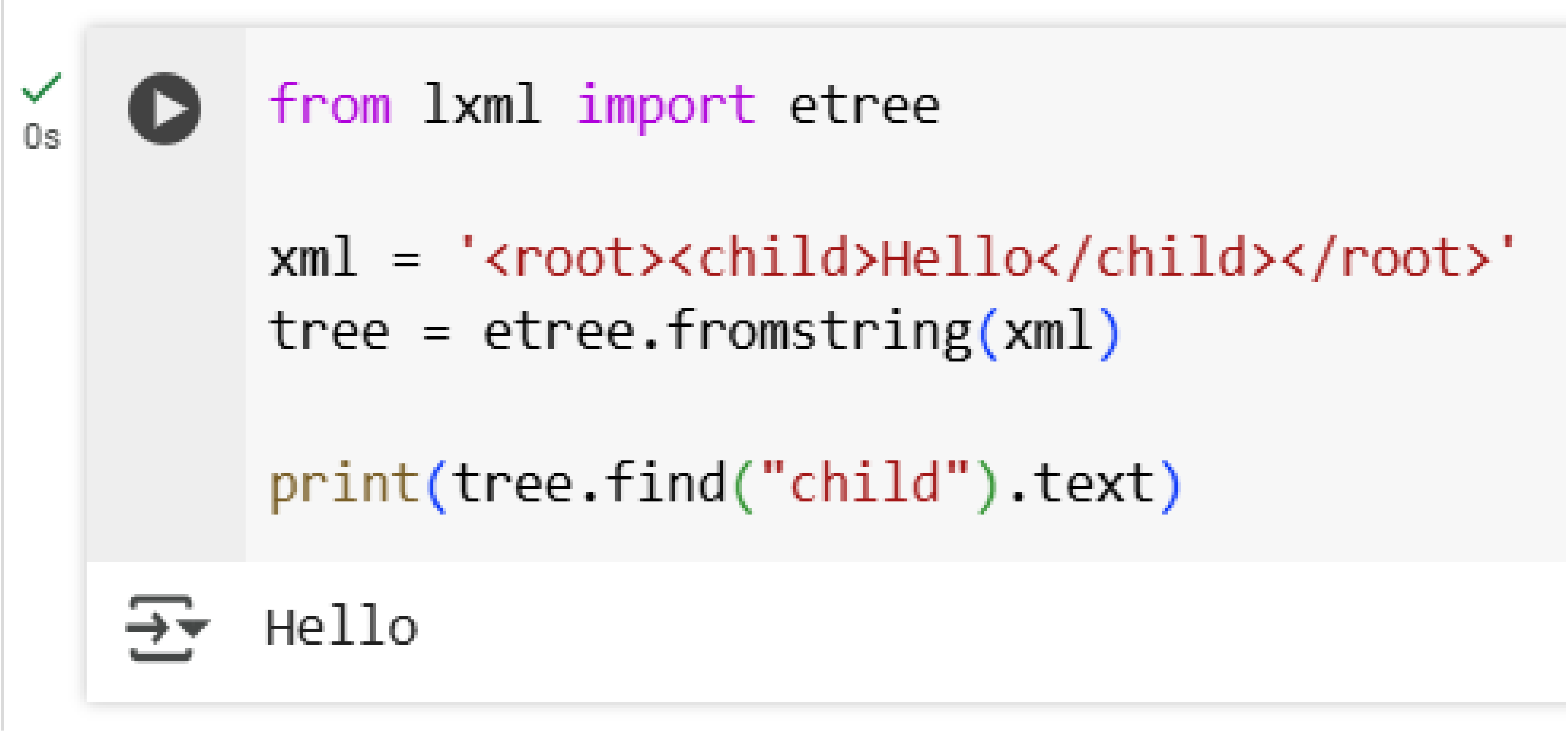
```
Example Domain
```

```
import requests
from bs4 import BeautifulSoup

URL = "https://www.geeksforgeeks.org"
r = requests.get(URL)

soup = BeautifulSoup(r.content)
print(soup.prettify())
h2_tags = soup.find_all("h2")
for tag in h2_tags:
    print(tag.text)
```

1. BeautifulSoup, lxml and scrapy –Data Crawlers



```
from lxml import etree

xml = '<root><child>Hello</child></root>'
tree = etree.fromstring(xml)

print(tree.find("child").text)
```

```
Hello
```

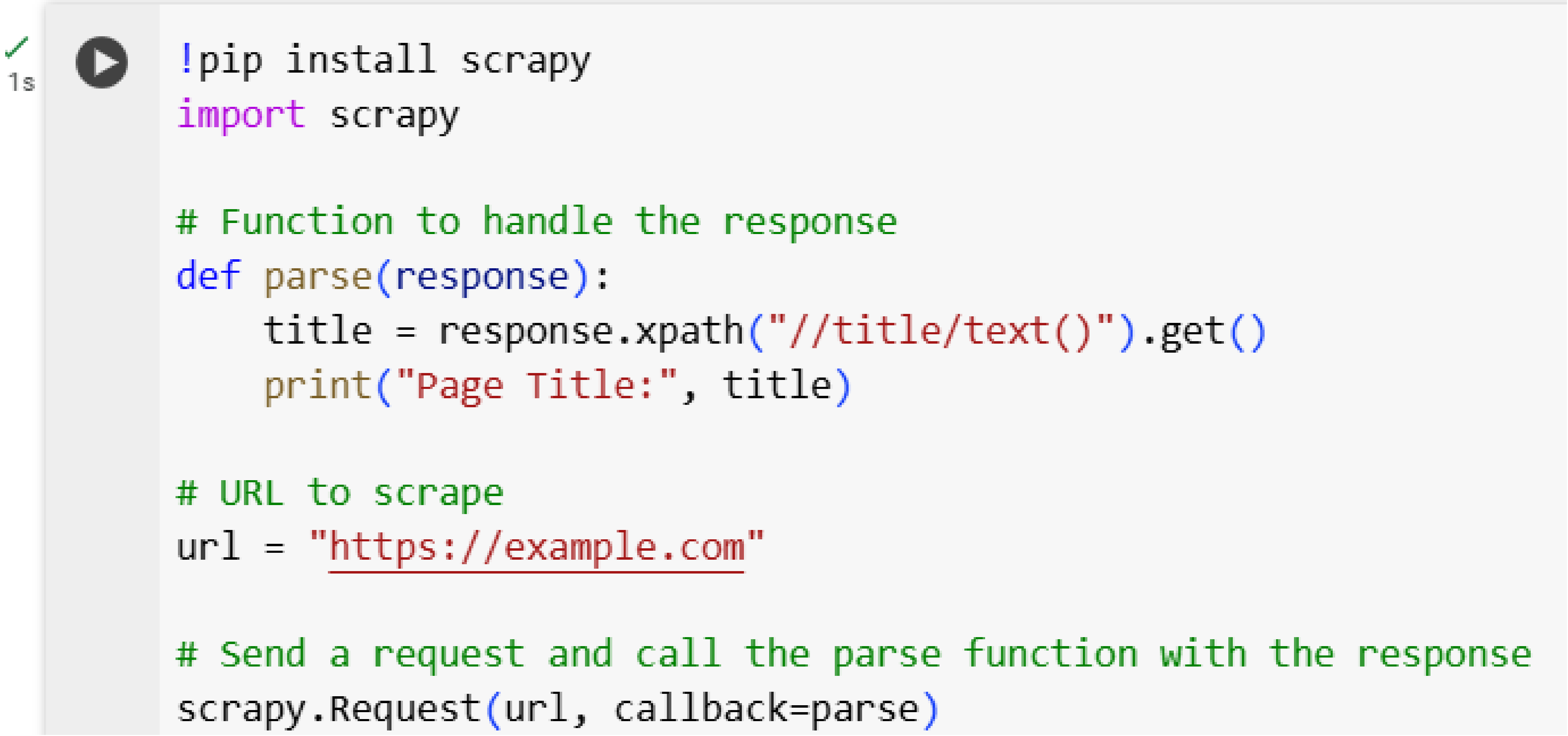
```
import requests
from lxml import html

# Send a GET request to the website
url = "https://example.com"
response = requests.get(url)

# Parse the HTML content using lxml
tree = html.fromstring(response.content)

# Example: Extract all <h2> elements using XPath
h2_elements = tree.xpath('//h2/text()')
for h2 in h2_elements:
    print(h2)
```

1. BeautifulSoup, lxml and scrapy –Data Crawlers



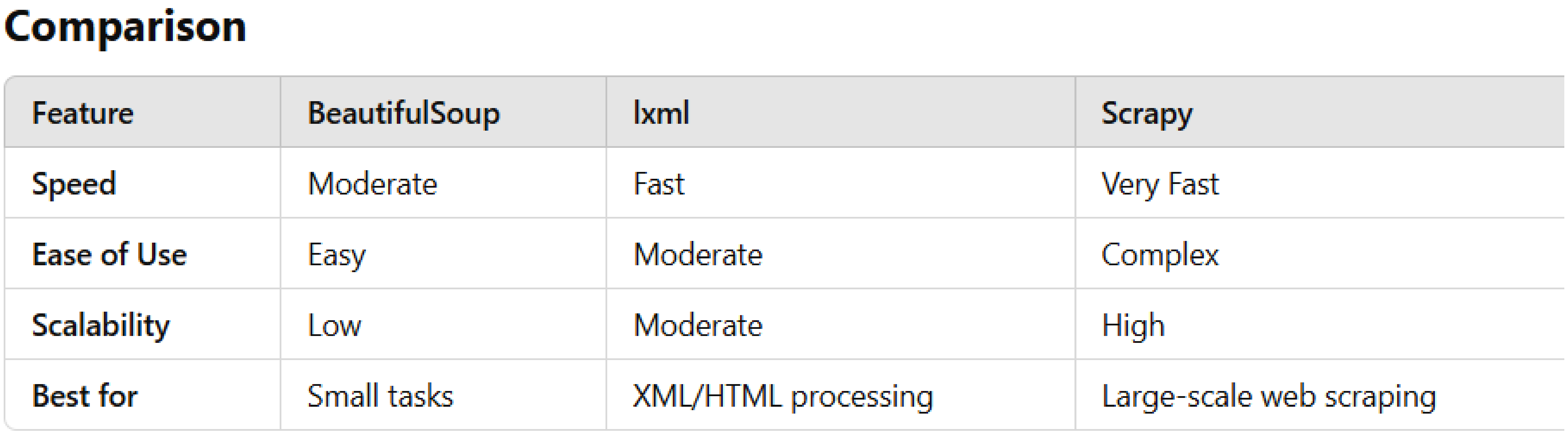
```
!pip install scrapy
import scrapy

# Function to handle the response
def parse(response):
    title = response.xpath("//title/text()").get()
    print("Page Title:", title)

# URL to scrape
url = "https://example.com"

# Send a request and call the parse function with the response
scrapy.Request(url, callback=parse)
```

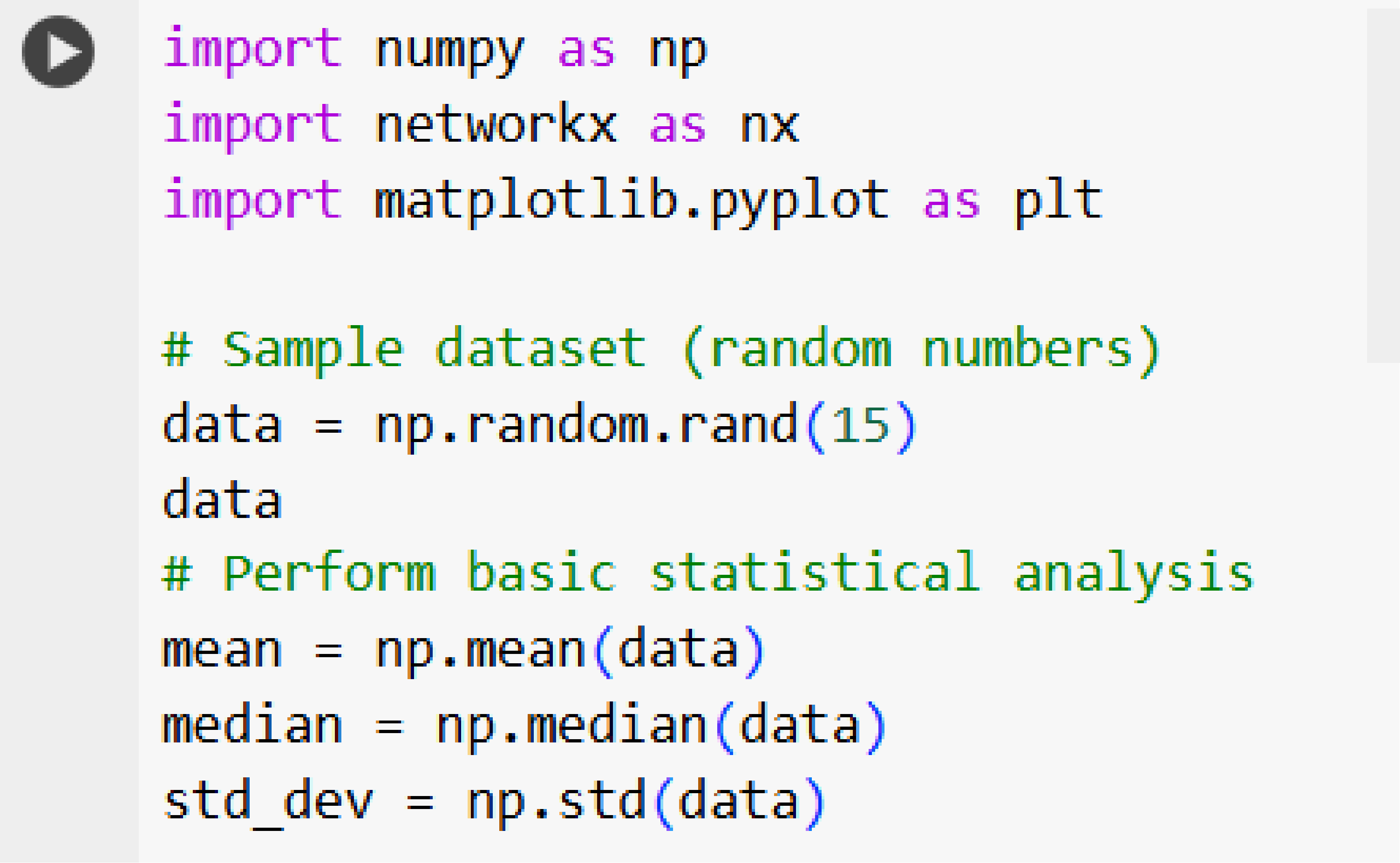
BeautifulSoup, lxml and scrapy –Data Crawlers



Comparison

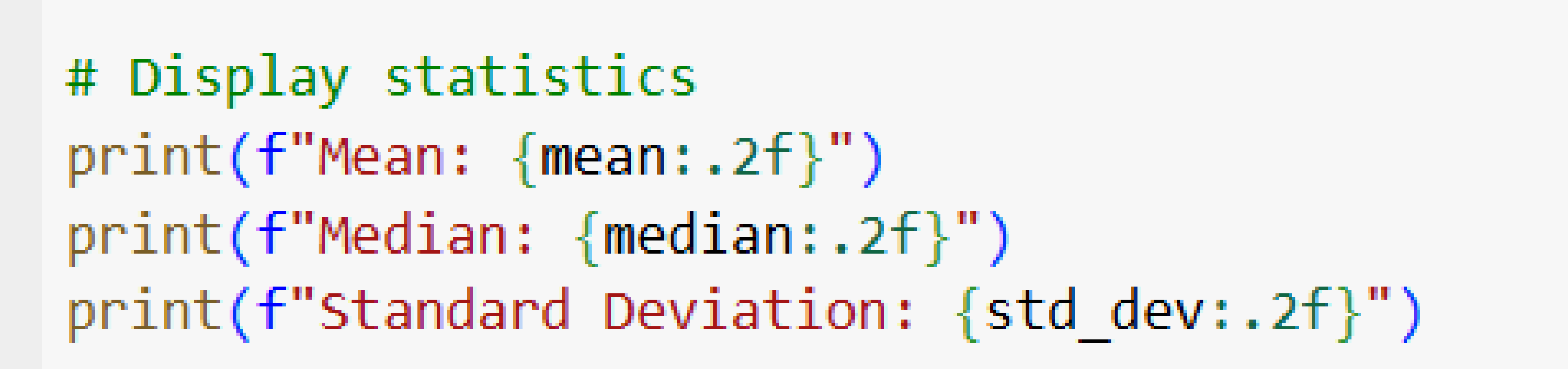
Feature	BeautifulSoup	lxml	Scrapy
Speed	Moderate	Fast	Very Fast
Ease of Use	Easy	Moderate	Complex
Scalability	Low	Moderate	High
Best for	Small tasks	XML/HTML processing	Large-scale web scraping

2. Implement data analyzing for a data set using SciPy and generate graph using NetworkX

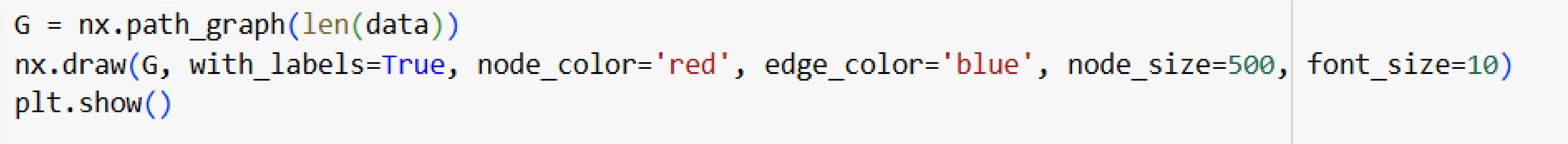


```
import numpy as np
import networkx as nx
import matplotlib.pyplot as plt

# Sample dataset (random numbers)
data = np.random.rand(15)
data
# Perform basic statistical analysis
mean = np.mean(data)
median = np.median(data)
std_dev = np.std(data)
```

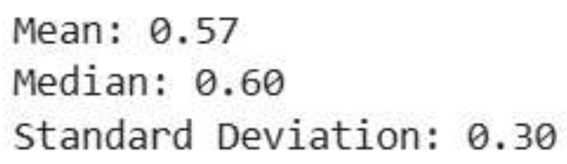


```
# Display statistics
print(f"Mean: {mean:.2f}")
print(f"Median: {median:.2f}")
print(f"Standard Deviation: {std_dev:.2f}")
```

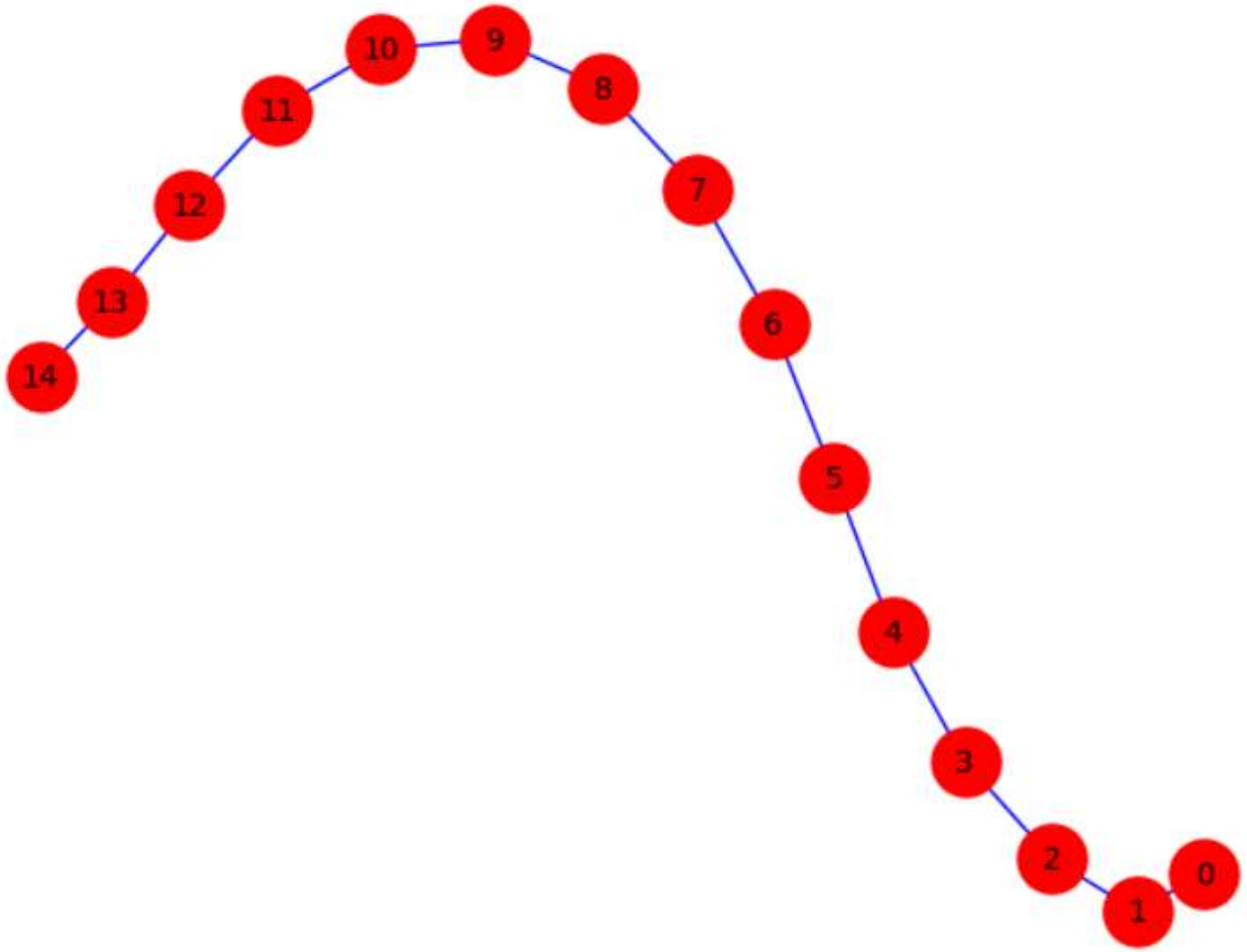


```
G = nx.path_graph(len(data))
nx.draw(G, with_labels=True, node_color='red', edge_color='blue', node_size=500, font_size=10)
plt.show()
```

2. Implement data analyzing for a data set using SciPy and generate graph using NetworkX



```
Mean: 0.57
Median: 0.60
Standard Deviation: 0.30
```



Thank You