# TEAM - 6

**TEAM LEAD**

**PADMAJA S P    ( CODING)**

SUTHIKSHA S     (DOCUMENTATION )

KAVISH SHARMA   (PRESENTATION)

SANDHIYA S  (CODING)

RAGHUNANDHAN K  (CODING)

SHALINI K  (DOCUMENTATION)

# TITLE OF PROJECT

# AMAZON SCRAPER

# PROJECT INDRODUCTION

**CONTENT:**

**Project Name : Amazon Product Scraper**

*Objective* :

To extract product details (Product Name, Price, Availability, Rating, and Reviews) from Amazon product pages.

*Purpose* :

Automate data collection for market research, price comparisons, and competitor analysis.

*Outcome* :

Save the extracted data in an Excel file for easy access and analysis.

# CODE IMPLEMENTATION

Include the **main parts** of the code (use snippets)

**1.HTTP Request** using request

**2.HTML Parsing** using beautifulsoup

**3.Data Extraction**: Product Name, Price, Availability, Rating, Reviews

**4.Export to Excel** using pandas

# TECH USED IN PROJECT

**CONTENT:**
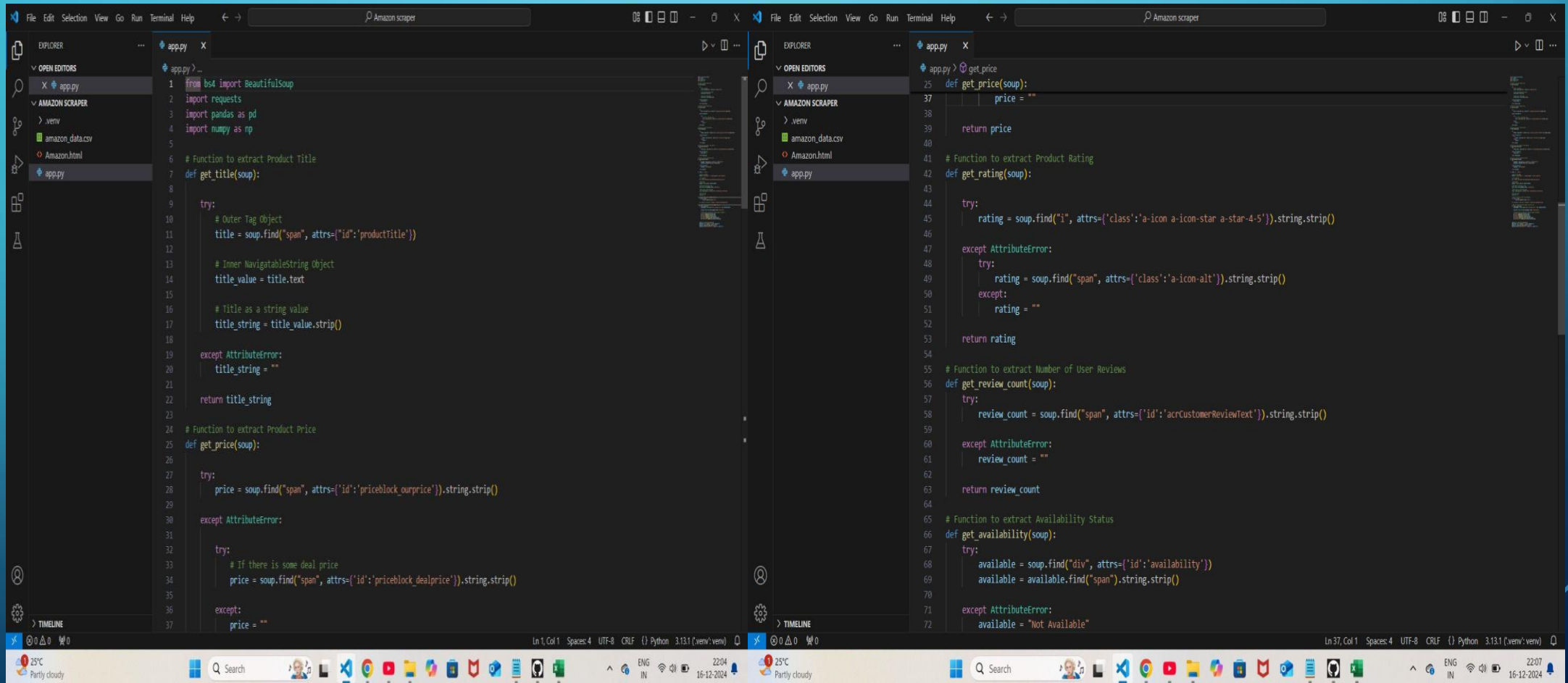
*Programming Language :* **Python**

*Libraries and Tools:*

*Requests  : To fetch web page content.*
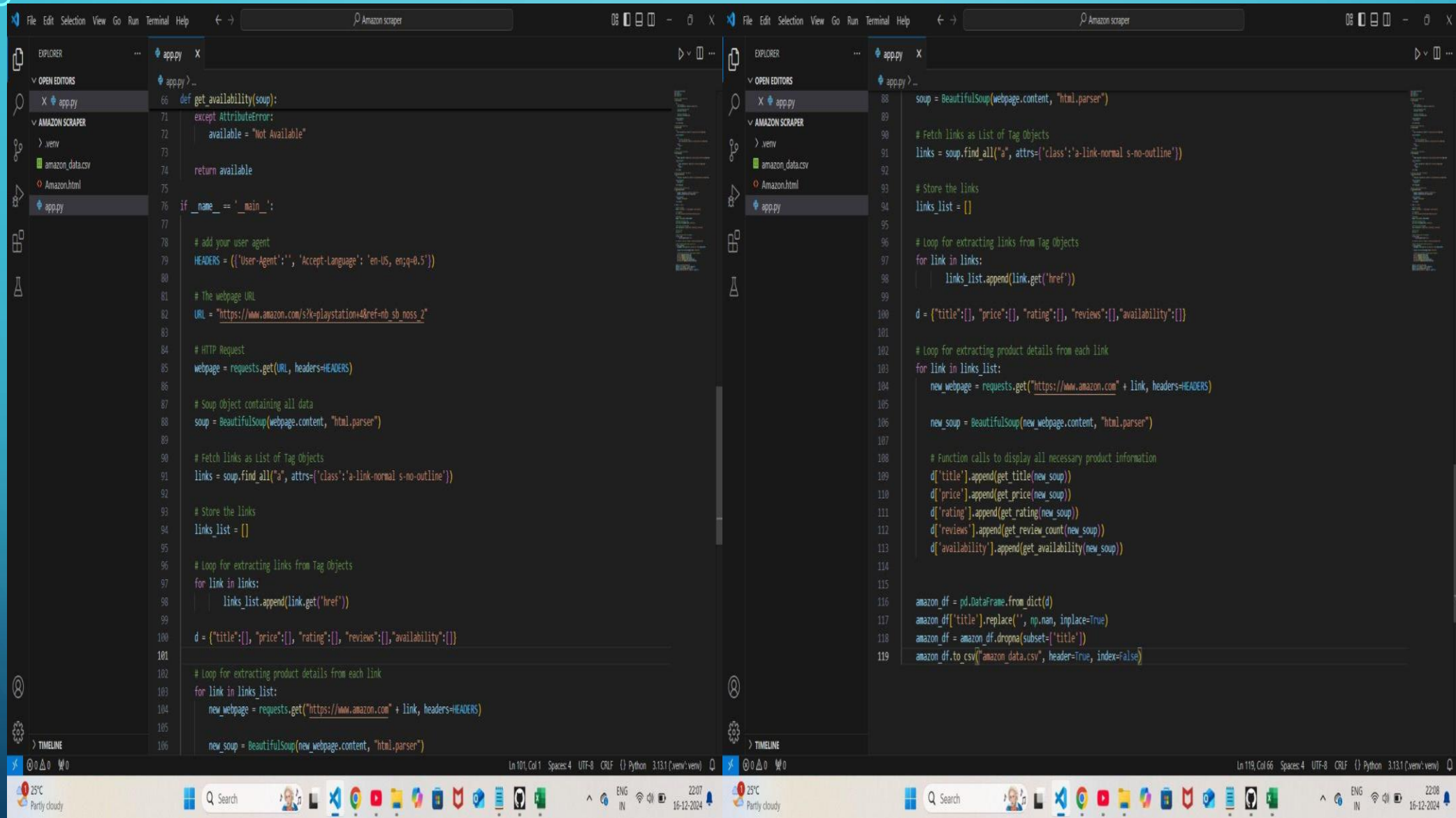
*BeautifulSoup  : To parse and extract data from HTML.*

*Pandas  : To store and save data into an Excel file.*

*Excel  : For organizing and visualizing the output.*

# CODECRAFT

```python
def get_availability(soup):
    except AttributeError:
        available = "Not Available"

    return available

if __name__ == '__main__':

    # add your user agent
    HEADERS = ({'User-Agent':'', 'Accept-Language': 'en-US, en;q=0.5'})

    # The webpage URL
    URL = "https://www.amazon.com/s?k=playstation+4&ref=nb_sb_noss_2"

    # HTTP Request
    webpage = requests.get(URL, headers=HEADERS)

    # Soup Object containing all data
    soup = BeautifulSoup(webpage.content, "html.parser")

    # Fetch links as List of Tag Objects
    links = soup.find_all("a", attrs={'class':'a-link-normal s-no-outline'})

    # Store the links
    links_list = []

    # Loop for extracting links from Tag Objects
    for link in links:
        links_list.append(link.get('href'))

    d = {"title":[], "price":[], "rating":[], "reviews":[],"availability":[]}

    # Loop for extracting product details from each link
    for link in links_list:
        new_webpage = requests.get("https://www.amazon.com" + link, headers=HEADERS)

        new_soup = BeautifulSoup(new_webpage.content, "html.parser")

        # Function calls to display all necessary product information
        d['title'].append(get_title(new_soup))
        d['price'].append(get_price(new_soup))
        d['rating'].append(get_rating(new_soup))
        d['reviews'].append(get_review_count(new_soup))
        d['availability'].append(get_availability(new_soup))


    amazon_df = pd.DataFrame.from_dict(d)
    amazon_df['title'].replace('', np.nan, inplace=True)
    amazon_df = amazon_df.dropna(subset=['title'])
    amazon_df.to_csv("amazon_data.csv", header=True, index=False)
```
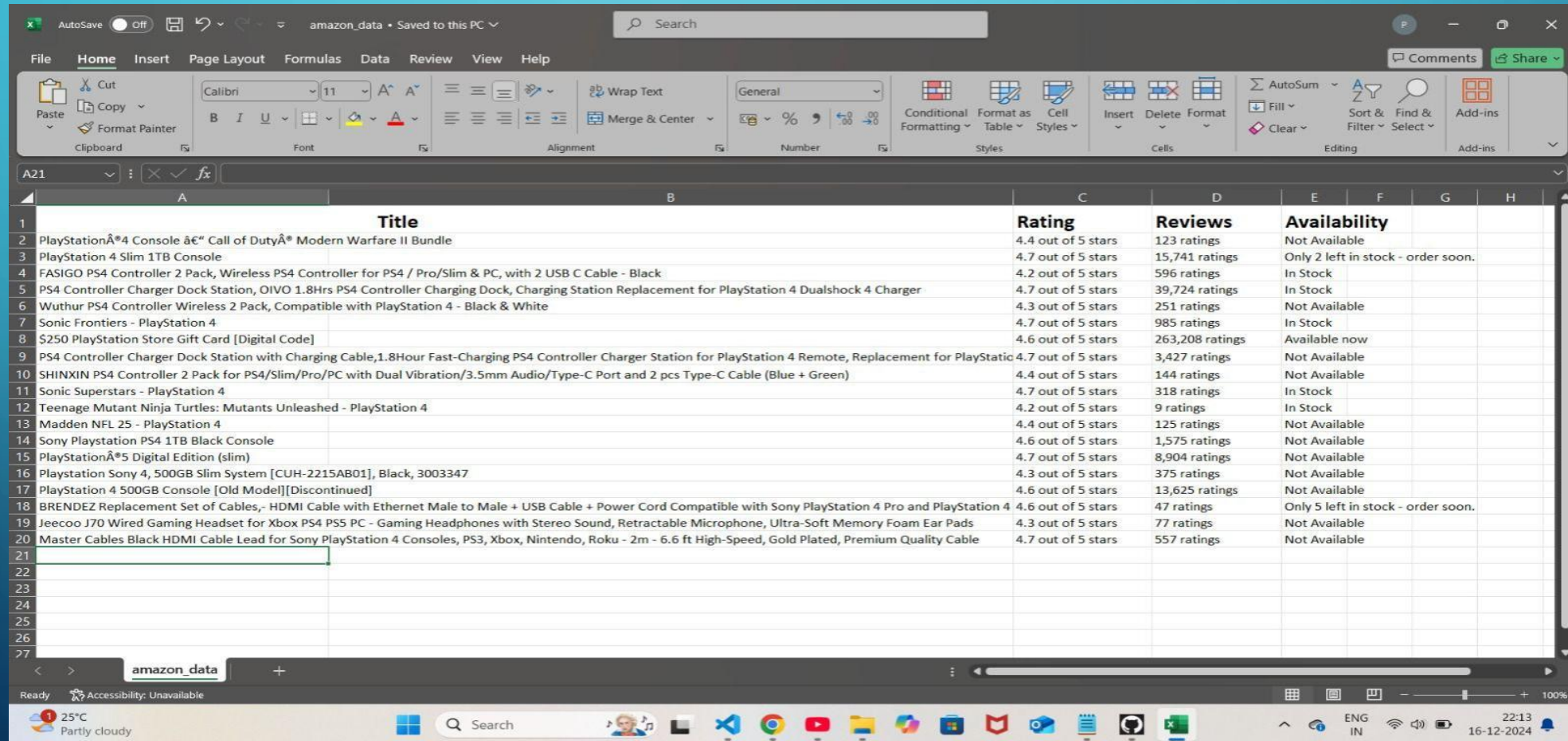
# RESULT

## IMPLEMENTED IN EXCEL SHEET

# REFERENCES

**Content**

Libraries and Tools:

- Requests Documentation
- BeautifulSoup Documentation
- Pandas Documentation

**Amazon Product Page:**

 Amazon Product Link

THANK YOU