

Kongu Engineering College,Erode

Team Name : Tech Nerds

Team Members : Akshara R K – 22ALR001

Madhiuksha S – 22ALR045

Dharaneesh N – 22ADR018

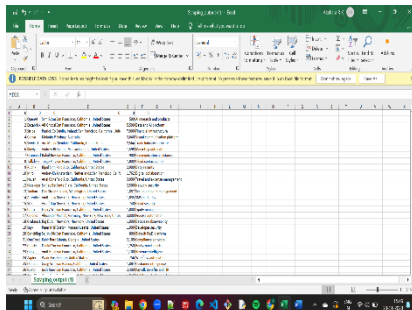
Problem Statement : Market Research

Technology Stack Used : Google Colab,BeautifulSoup,Pandas,Python,Html

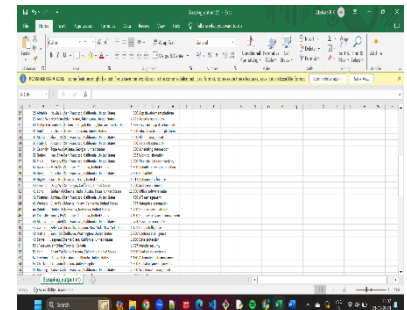
Implementation : Scrapped 100 company names and their details mostly located in USA
using BeautifulSoup tool from the following website

<https://www.forbes.com/lists/cloud100/sh=6636e5517d9c>

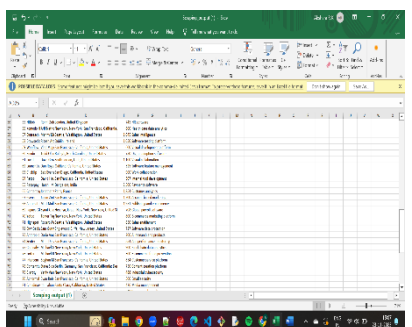
Output Screenshots :



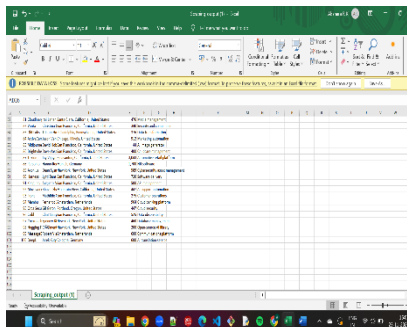
Rank	Company Name	Website
1	Microsoft	Microsoft.com
2	Amazon	Amazon.com
3	Google	Google.com
4	Facebook	Facebook.com
5	Apple	Apple.com



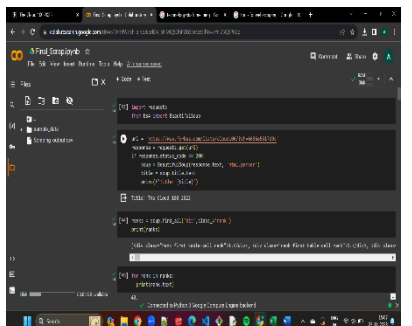
Rank	Company Name	Website
1	Microsoft	Microsoft.com
2	Amazon	Amazon.com
3	Google	Google.com
4	Facebook	Facebook.com
5	Apple	Apple.com



Rank	Company Name	Website
1	Microsoft	Microsoft.com
2	Amazon	Amazon.com
3	Google	Google.com
4	Facebook	Facebook.com
5	Apple	Apple.com



Rank	Company Name	Website
1	Microsoft	Microsoft.com
2	Amazon	Amazon.com
3	Google	Google.com
4	Facebook	Facebook.com
5	Apple	Apple.com



```
[1]: # Importing the required libraries
import requests
from bs4 import BeautifulSoup
import pandas as pd

# Step 1: Get the HTML content from the website
url = "https://www.forbes.com/lists/cloud100/sh=6636e5517d9c"
response = requests.get(url)
html_content = response.text

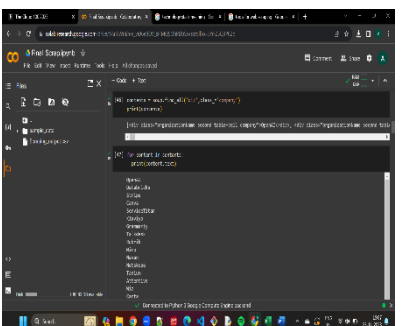
# Step 2: Parse the HTML content using BeautifulSoup
soup = BeautifulSoup(html_content, 'html.parser')

# Step 3: Extract the data from the HTML content
# Find all the company names and their websites
company_names = []
company_websites = []

for company in soup.find_all('li'):
    company_name = company.find('a').text
    company_website = company.find('a').get('href')
    company_names.append(company_name)
    company_websites.append(company_website)

# Step 4: Create a DataFrame from the extracted data
data = pd.DataFrame({'Company Name': company_names, 'Website': company_websites})

# Step 5: Display the first 10 rows of the DataFrame
data.head(10)
```



```
[2]: # Displaying the first 10 rows of the DataFrame
data.head(10)
```

