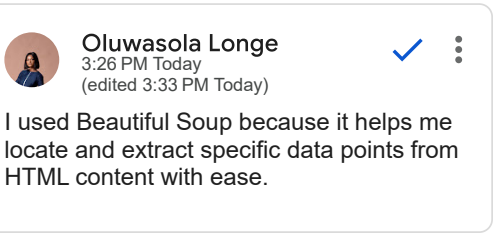
Apple Stocks Web Scrapping

For this project, I scraped data from the Wikipedia website to gather enough data for the stock market analysis project I am working on and I exported it to a CSV file and then later exported it to excel.




```
from bs4 import BeautifulSoup
import requests


url = "https://en.wikipedia.org/wiki/List_of_Apple_products"
page = requests.get(url)
soup = BeautifulSoup(page.text, 'html')


print(soup)
```

```
<!DOCTYPE html>
<html class="client-nojs vector-feature-language-in-heac
<head>
<meta charset="utf-8"/>
<title>List of Apple products - Wikipedia</title>
<script>(function(){var className="client-js vector-feat
"wgMonthNames":["","January","February","March","April",
"Lists of products","Apple Inc. lists","History of Apple
"wgWikibaseItemId":"Q7805629","wgCheckUserClientHintsHea
"jquery.tablesorter","jquery.makeCollapsible","mediawiki
<script>(RLQ=window.RLQ||[]).push(function(){mw.loader.i
}];});});</script>
<link href="/w/load.php?lang=en&amp;modules=ext.cite.sty
<script async="" src="/w/load.php?lang=en&amp;modules=st
<meta content="" name="ResourceLoaderDynamicStyles"/>
<link href="/w/load.php?lang=en&amp;modules=site.styles&
<meta content="MediaWiki 1.44.0-wmf.18" name="generator"
<meta content="origin" name="referrer"/>
<meta content="origin-when-cross-origin" name="referrer"
<meta content="max-image-preview:standard" name="robots"
<meta content="telephone=no" name="format-detection"/>
<meta content="width=1120" name="viewport"/>
<meta content="List of Apple products - Wikipedia" prope
<meta content="website" property="og:type"/>
<link href="//upload.wikimedia.org" rel="preconnect"/>
<link href="//en.m.wikipedia.org/wiki/List_of_Apple_proc
<link href="/w/index.php?title=List_of_Apple_products&am
<link href="/static/apple-touch/wikipedia.png" rel="appl
<link href="/static/favicon/wikipedia.ico" rel="icon"/>
<link href="/w/rest.php/v1/search" rel="search" title="W
<link href="//en.wikipedia.org/w/api.php?action=rsd" rel
<link href="https://en.wikipedia.org/wiki/List_of_Apple_
<link href="https://creativecommons.org/licenses/by-sa/4
<link href="/w/index.php?title=Special:RecentChanges&amp
<link href="//meta.wikimedia.org" rel="dns-prefetch"/>
<link href="login.wikimedia.org" rel="dns-prefetch"/>
</head>
<body class="skin--responsive skin-vector skin-vector-se
<div class="vector-header-container">
<header class="vector-header mw-header">
<div class="vector-header-start">
<nav aria-label="Site" class="vector-main-menu-landmark"
<div class="vector-dropdown vector-main-menu-dropdown ve
<input aria-haspopup="true" aria-label="Main menu" class
<label aria-hidden="true" class="vector-dropdown-label c
<span class="vector-dropdown-label-text">Main menu</spar
</label>
```

```
<div class="vector-dropdown-content">
<div class="vector-unpinned-container" id="vector-main-m
<div class="vector-main-menu vector-pinnable-element" id
<div class="vector-pinnable-header vector-main-menu-pinn
<div class="vector-pinnable-header-label">Main menu</div
<button class="vector-pinnable-header-toggle-button vect
<button class="vector-pinnable-header-toggle-button vect
</div>
<div class="vector-menu mw-portlet mw-portlet-navigation
```

```
soup.find('table')
```

```
<table class="wikitable">
<tbody><tr>
<th>Released
</th>
<th>Model
</th>
<th>Family
</th>
<th>Discontinued
</th></tr>
<tr bgcolor="#FFFF79">
<td bgcolor="#FFF">April 11, 1976
</td>
<td><a href="/wiki/Apple_I" title="Apple I">Apple I</a>
</td>
<td><a href="/wiki/Apple_I" title="Apple I">Apple I</a>
</td>
<td>September 30, 1977
</td></tr>
<tr bgcolor="#FFFF79">
<td bgcolor="#FFF">June 1, 1977
</td>
<td><a href="/wiki/Apple_II" title="Apple II">Apple
II</a>
</td>
<td><a href="/wiki/Apple_II" title="Apple II">Apple
II</a>
</td>
<td>May 1, 1979
</td></tr>
<tr bgcolor="#D8D8F2">
<td bgcolor="#FFF">June 1, 1978
</td>
<td><a href="/wiki/Disk_II" title="Disk II">Disk II</a>
</td>
<td><a href="/wiki/List_of_Apple_drives" title="List of
Apple drives">Drives</a>
</td>
<td>May 1, 1984
</td></tr>
<tr bgcolor="#FFFF79">
<td bgcolor="#FFF" rowspan="6">June 1, 1979
</td>
<td><a href="/wiki/Apple_II_Plus" title="Apple II
Plus">Apple II Plus</a>
</td>
<td rowspan="4"><a href="/wiki/Apple_II" title="Apple
II">Apple II</a>
</td>
<td>December 1, 1982
</td></tr>
<tr bgcolor="#FFFF79">
<td><a
href="/wiki/Apple_II_Plus#Apple_II_Europlus_and_J-Plus"
title="Apple II Plus">Apple II EuroPlus</a>
</td>
```

```
<td>December 1, 1982
</td></tr>
```

```
soup.find_all('table')[1:44]
```

```
[<table class="wikitable">
 <tbody><tr>
 <th>Released
 </th>
 <th>Model
 </th>
 <th>Family
 </th>
 <th>Discontinued
 </th></tr>
 <tr bgcolor="#FFFF79">
 <td bgcolor="#FFF" rowspan="6">November 1, 1980
 </td>
 <td><a href="/wiki/Apple_III" title="Apple III">Apple
III</a>
 </td>
 <td><a href="/wiki/Apple_III" title="Apple III">Apple
III</a>
 </td>
 <td>December 1, 1981
 </td></tr>
 <tr bgcolor="#D8D8F2">
 <td><a href="/wiki/Novation_CAT" title="Novation
CAT">Modem IIB</a> (Novation CAT)
 </td>
 <td><a href="/wiki/Apple_USB_Modem#History"
title="Apple USB Modem">Modems</a>
 </td>
 <td>December 1, 1981
 </td></tr>
 <tr bgcolor="#D8D8F2">
 <td><a href="/wiki/Centronics"
title="Centronics">Printer IIA</a> (Centronics 779)
 </td>
 <td><a href="/wiki/Category:Apple_Inc._printers"
title="Category:Apple Inc. printers">Printers</a>
 </td>
 <td>December 1, 1981
 </td></tr>
 <tr bgcolor="#D8D8F2">
 <td><a href="/wiki/Apple_Monitor_III" title="Apple
Monitor III">Monitor III</a>
 </td>
 <td><a href="/wiki/Apple_displays" title="Apple
displays">Displays</a>
 </td>
 <td>December 1, 1981
 </td></tr>
 <tr bgcolor="#D8D8F2">
 <td><a href="/wiki/Apple_displays" title="Apple
displays">Monitor II</a> (various third party)
 </td>
 <td><a href="/wiki/Apple_displays" title="Apple
displays">Displays</a>
 </td>
 <td>December 1, 1981
 </td></tr>
```

```
soup.find('table',class_ = 'wikitable sortable')
```

```
<table class="wikitable sortable">
<tbody><tr>
<th>Released
```

```
</th>
<th>Model
</th>
<th>Family
</th>
<th>Discontinued
</th></tr>
<tr bgcolor="#D8D8F2">
<td bgcolor="#FFF">March 31, 2010
</td>
<td><a href="/wiki/AirPort_Time_Capsule" title="AirPort
Time Capsule">Time Capsule</a> (3rd gen)
</td>
<td><a href="/wiki/AirPort"
title="AirPort">AirPort</a>, <a
href="/wiki/List_of_Apple_drives" title="List of Apple
drives">drives</a>
</td>
<td>June 21, 2011
</td></tr>
<tr bgcolor="#CCFF99">
<td bgcolor="#FFF">April 3, 2010
</td>
<td><a href="/wiki/IPad_(1st_generation)" title="IPad
(1st generation)">iPad</a> (Wi-Fi)
</td>
<td><a href="/wiki/IPad" title="IPad">iPad</a>
</td>
<td>March 3, 2011
</td></tr>
<tr bgcolor="#95CEFE">
<td bgcolor="#FFF">April 13, 2010
</td>
<td><a href="/wiki/MacBook_Pro" title="MacBook
Pro">MacBook Pro</a> (Mid 2010)
</td>
<td><a href="/wiki/MacBook_Pro" title="MacBook
Pro">MacBook Pro</a>
</td>
<td>February 24, 2011
</td></tr>
<tr bgcolor="#CCFF99">
<td bgcolor="#FFF">April 30, 2010
</td>
<td><a href="/wiki/IPad_(1st_generation)" title="IPad
(1st generation)">iPad</a> (Wi-Fi + 3G)
</td>
<td><a href="/wiki/IPad" title="IPad">iPad</a>
</td>
<td>March 2, 2011
</td></tr>
<tr bgcolor="#95CEFE">
<td bgcolor="#FFF">May 18, 2010
</td>
<td><a href="/wiki/MacBook" title="MacBook">MacBook</a>
```

```
table = soup.find_all('table')[1:44] [1]
print(table)
```

Oluwasola Longe
3:45 PM Today

Trying to get information on the number of tables

```
      <td><a href="/wiki/Compact_Macintosh" title="Compact Mac
      </td>
      <td>September 10, 1984
      </td></tr>
      <tr bgcolor="#D8D8F2">
      <td><a href="/wiki/Macintosh_External_Disk_Drive#400K" t
      </td>
      <td><a href="/wiki/List_of_Apple_drives" title="List of
      </td>
      <td>January 1, 1986
      </td></tr>
      <tr bgcolor="#D8D8F2">
      <td>Apple Modem 300, 1984
      </td>
      <td><a href="/wiki/Apple_USB_Modem#History" title="Apple
      </td>
      <td><span style="white-space:nowrap;font-size:85%;">[<i>
      </td></tr>
      <tr bgcolor="#D8D8F2">
      <td>Apple Modem 1200, 1984
      </td>
      <td><a href="/wiki/Apple_USB_Modem#History" title="Apple
      </td>
      <td><span style="white-space:nowrap;font-size:85%;">[<i>
      </td></tr>
      <tr bgcolor="#FFFF79">
      <td bgcolor="#FFF" rowspan="4">April 1, 1984
      </td>
      <td><a href="/wiki/Apple_IIc" title="Apple IIc">Apple II
      </td>
      <td><a href="/wiki/Apple_II" title="Apple II">Apple II</
      </td>
      <td>August 1, 1988
      </td></tr>
      <tr bgcolor="#D8D8F2">
      <td><a class="mw-redirect" href="/wiki/Apple_Scribe_Prin
      </td>
      <td><a href="/wiki/Category:Apple_Inc._printers" title="
      </td>
      <td>December 1, 1985
      </td></tr>
      <tr bgcolor="#D8D8F2">
      <td><a class="mw-redirect" href="/wiki/Apple_Mouse#Apple
      </td>
      <td><a href="/wiki/Apple_pointing_devices" title="Apple
      </td>
```

```
table.find_all('th')[1:44]
```

```
[<th>Model
 </th>,
 <th>Family
 </th>,
 <th>Discontinued
 </th>]
```

```
world_titles = table.find_all('th')
```

```
world_titles
```

```
[<th>Released
 </th>,
 <th>Model
 </th>,
 <th>Family
 </th>,
```

```
        <th>Discontinued
        </th>]


world_table_titles = [titles.text.strip() for titles in world_t
print(world_table_titles)
```

```
['Released', 'Model', 'Family', 'Discontinued']
```

Oluwasola Longe
3:48 PM Today

I imported pandas to extract the headers in the data frame

```
import pandas as pd


df = pd.DataFrame(columns = world_table_titles)
df
```

Released	Model	Family	Discontinued

```
table.find_all('tr')[1:44]
```

```
 <td><a href="/wiki/Macintosh_128K#Models"
title="Macintosh 128K">Macintosh 128K (revised)</a>
 </td>
 <td>Compact
 </td>
 <td>October 1, 1985
 </td></tr>,
 <tr bgcolor="#D8D8F2">
 <td bgcolor="#FFF">December 1984
 </td>
 <td><a href="/wiki/Apple_displays" title="Apple
```

```
column_data = table.find_all('tr')[1:44]
```

```
for row in column_data:
    print(row.find_all('td'))
```

```
[<td><a href="/wiki/Macintosh_External_Disk_Drive#400K"
</td>, <td><a href="/wiki/List_of_Apple_drives" title="L
</td>, <td>January 1, 1986
</td>]
[<td>Apple Modem 300, 1984
</td>, <td><a href="/wiki/Apple_USB_Modem#History" title
</td>, <td><span style="white-space:nowrap;font-size:85%
</td>]
[<td>Apple Modem 1200, 1984
</td>, <td><a href="/wiki/Apple_USB_Modem#History" title
</td>, <td><span style="white-space:nowrap;font-size:85%
</td>]
[<td bgcolor="#FFF" rowspan="4">April 1, 1984
</td>, <td><a href="/wiki/Apple_IIc" title="Apple IIc">A
</td>, <td><a href="/wiki/Apple_II" title="Apple II">App
</td>, <td>August 1, 1988
</td>]
[<td><a class="mw-redirect" href="/wiki/Apple_Scribe_Pri
</td>, <td><a href="/wiki/Category:Apple_Inc._printers"
</td>, <td>December 1, 1985
</td>]
[<td><a class="mw-redirect" href="/wiki/Apple_Mouse#Appl
</td>, <td><a href="/wiki/Apple_pointing_devices" title=
</td>, <td>December 1, 1985
</td>]
[<td><a class="mw-redirect" href="/wiki/Disk_IIc" title=
</td>, <td><a href="/wiki/List_of_Apple_drives" title="L
</td>, <td><span style="white-space:nowrap;font-size:85%
</td>]
[<td bgcolor="#FFF">May 1, 1984
</td>, <td><a href="/wiki/Disk_II#DuoDisk" title="Disk I
</td>, <td><a href="/wiki/List_of_Apple_drives" title="L
</td>, <td><span style="white-space:nowrap;font-size:85%
</td>]
[<td bgcolor="#FFF" rowspan="2">June, 1984
</td>, <td><a href="/wiki/Apple_410_Color_Plotter" title
</td>, <td><a href="/wiki/Category:Apple_Inc._printers"
</td>, <td><span style="white-space:nowrap;font-size:85%
</td>]
[<td><a class="mw-redirect" href="/wiki/Apple_ImageWrite
```

```
</td>]
[<td bgcolor="#FFF">December 1984
</td>, <td><a href="/wiki/Apple_displays" title="Apple d
</td>, <td><a href="/wiki/Apple_displays" title="Apple d
</td>, <td><span style="white-space:nowrap;font-size:85%
</td>]
```

```
for row in column_data[1:]:
    row_data = row.find_all('td')
    individual_row_data = [data.text.strip() for data in row_d
    print(individual_row_data)
```

```
['January 24, 1984', 'Macintosh (128K)', 'Compact', 'Septem
['Macintosh External Disk Drive (400K)', 'Drives', 'January
['Apple Modem 300, 1984', 'Modems', '[?]']
['Apple Modem 1200, 1984', 'Modems', '[?]']
['April 1, 1984', 'Apple IIc', 'Apple II', 'August 1, 1988'
['Apple Scribe Printer', 'Printers', 'December 1, 1985']
['Apple Mouse IIc', 'Pointing devices', 'December 1, 1985']
['Disk IIc', 'Drives', '[?]']
['May 1, 1984', 'DuoDisk', 'Drives', '[?]']
['June, 1984', 'Apple 410 Color Plotter', 'Printers', '[?]'
['Apple ImageWriter Wide Carriage', 'Printers', '[?]']
['September 10, 1984', 'Macintosh 512K', 'Compact', 'April
['Macintosh 128K (revised)', 'Compact', 'October 1, 1985']
['December 1984', 'AppleColor 100', 'Displays', '[?]']
```

Oluwasola Longe
3:51 PM Today

I extracted both headers and rows. I added an index [1:] to skip the header row

```
headers = [th.text.strip() for th in table.find_all("th")]

rows = []
for tr in table.find_all("tr")[1:]:
    cols = [td.text.strip() for td in tr.find_all("td")]
    rows.append(cols)

df = pd.DataFrame(rows, columns=headers)

df
```

	Released	Model	Family	Discontinued
0	February 28, 2025	iPhone 16e	iPhone	current
1	March 12, 2025	iPad (11th generation)	iPad	current
2	iPad Air 11-in. (M3)	iPad	current	None
3	iPad Air 13-in. (M3)	iPad	current	None
4	MacBook Air 13-in. (M4)	MacBook Air	current	None

Next steps: Generate code with df | View recommended plots | New int

Oluwasola Longe
3:54 PM Today

Concatenating all DataFrames into one table

```
df_list = []

for table in soup.find_all('table')[1:44]:
    headers = [th.text.strip() for th in table.find_all("th")]

    rows = []
```

```
    for tr in table.find_all("tr")[1:]:  # Skip the header row
        cols = [td.text.strip() for td in tr.find_all("td")]
        rows.append(cols)

    df = pd.DataFrame(rows, columns=headers)

    df_list.append(df)

final_df = pd.concat(df_list, ignore_index=True)

final_df
```

	Released	Model	Family	Discontinued
0	November 1, 1980	Apple III	Apple III	December 1, 1981
1	Modem IIB (Novation CAT)	Modems	December 1, 1981	None
2	Printer IIA (Centronics 779)	Printers	December 1, 1981	None
3	Monitor III	Displays	December 1, 1981	None
4	Monitor II (various third party)	Displays	December 1, 1981	None
...	...	...	...	...
695	iPad Air 11-in. (M3)	iPad	current	None

Next steps: Generate code with `final_df` View recommended plots New interactive sheet

```
df_list = pd.read_html(url)

len(df_list)
```

```
50
```

```
df_list[1]
```

	Released	Model	Family	Discontinued
0	November 1, 1980	Apple III	Apple III	December 1, 1981
1	November 1, 1980	Modem IIB (Novation CAT)	Modems	December 1, 1981
2	November 1, 1980	Printer IIA (Centronics 779)	Printers	December 1, 1981
3	November 1, 1980	Monitor III	Displays	December 1, 1981
4	November 1, 1980	Monitor II (various third party)	Displays	December 1, 1981
5	November 1, 1980	Disk III	Drives	May 1, 1984
6	September 1, 1981	Apple ProFile	Drives	September 1, 1986
7	December 1, 1981	Apple III Revised[1]	Apple III	December 1, 1983

```
final_df = pd.concat(df_list[1:45], ignore_index=True)

final_df
```

	Released	Model	Family	Discontinued	Timeline of Apple Inc. products vte
0	November 1, 1980	Apple III	Apple III	December 1, 1981	NaN
1	November 1, 1980	Modem IIB (Novation CAT)	Modems	December 1, 1981	NaN
2	November 1, 1980	Printer IIA (Centronics 779)	Printers	December 1, 1981	NaN
3	November 1, 1980	Monitor III	Displays	December 1, 1981	NaN
4	November 1, 1980	Monitor II (various third party)	Displays	December 1, 1981	NaN
...	...	...	...	...	...
696	March 12, 2025	iPad Air 13-in. (M3)	iPad	current	NaN
697	March 12, 2025	MacBook Air 13-in. (M4)	MacBook Air	current	NaN
	March 12	MacBook	MacBook		

Oluwasola Longe
3:57 PM Today
Exported to CSV file

Oluwasola Longe
3:57 PM Today
Exported to Excel file

Next steps: Generate code with final_df | View recommended plots | New interactive sheet

```
final_df.to_csv(r"C:\Users\USEER\Downloads\List_of_Apple_Stock"


final_df.to_excel("Apple_Events_1980_2024.xlsx", index=False)
print()
```