**LAPORAN PRAKTIKUM**

**PYTHON**

**WEB SCRAPPING**

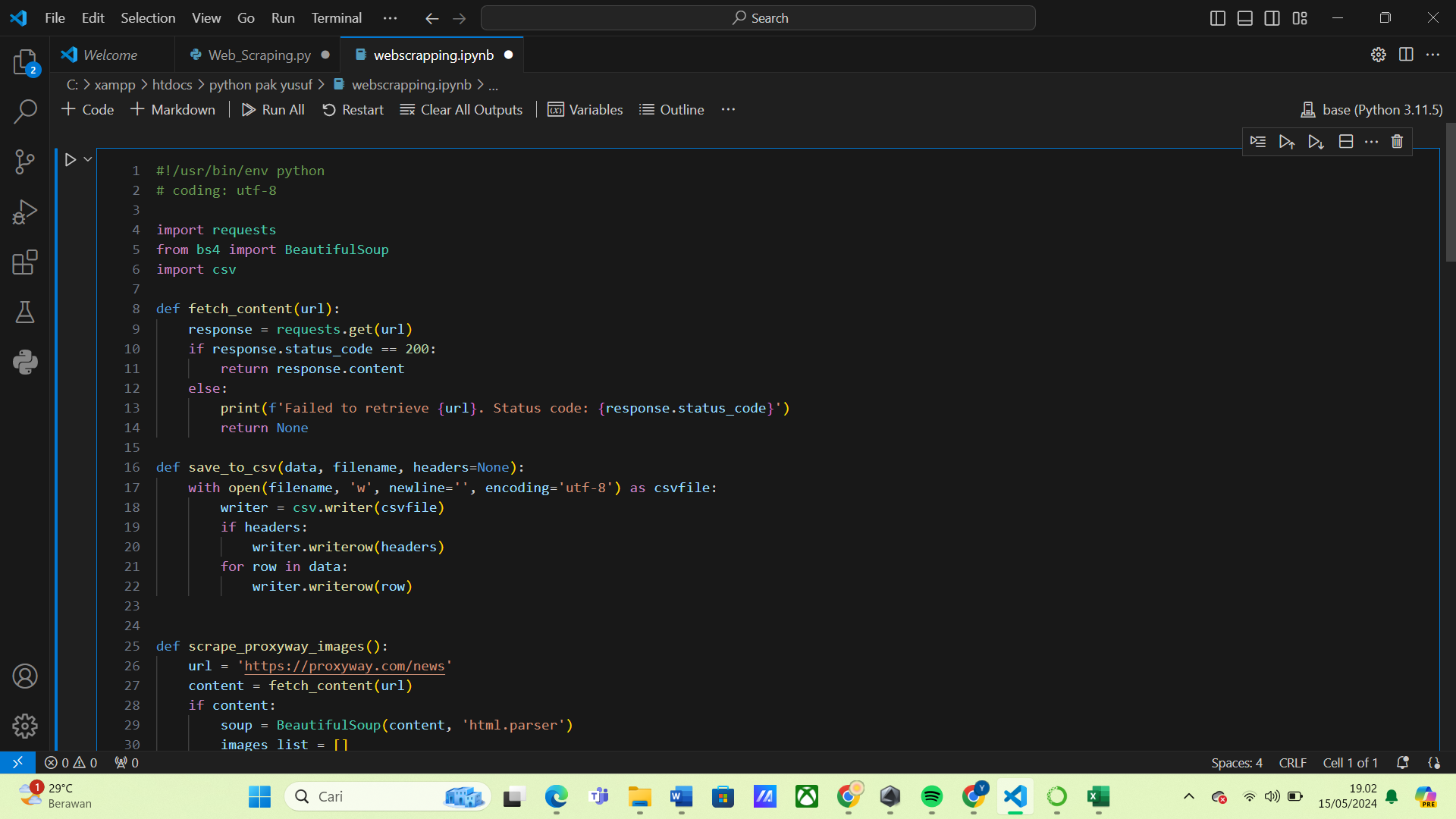
****

**Disusun oleh :**

Yunita Kartika Putri (V3923019)

**Dosen :** Yusuf Fadila Rachman. S.Kom., M.Kom

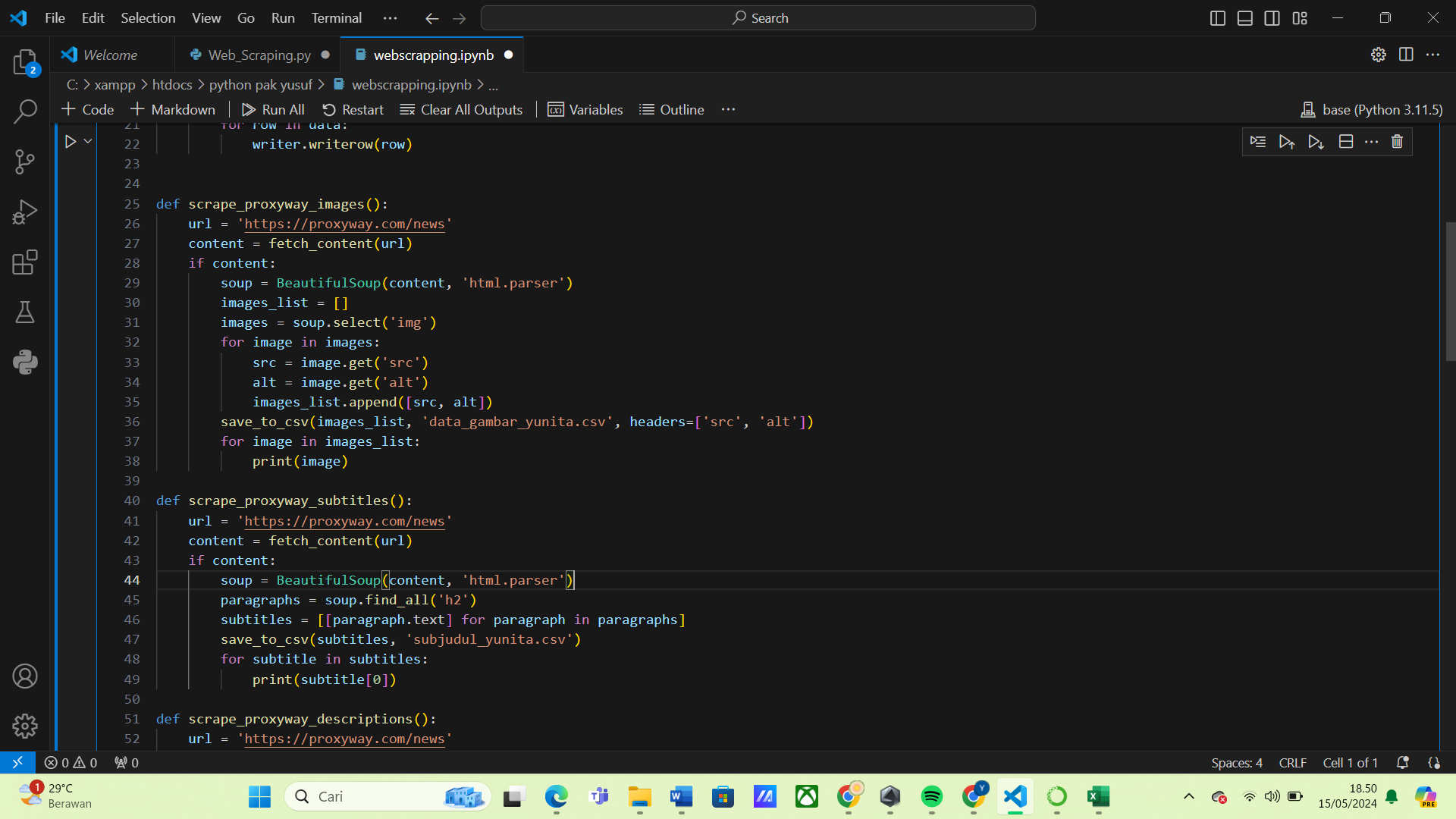
**PS D-III TEKNIK INFORMATIKA SEKOLAH VOKASI UNIVERSITAS SEBELAS MARET 2024**



Fungsi `fetch\_content` bertujuan untuk mengambil konten dari URL yang diberikan menggunakan pustaka `requests`. Jika respons dari URL memiliki status kode 200, yang menunjukkan bahwa permintaan berhasil, fungsi ini mengembalikan konten dari respons tersebut. Jika respons memiliki status kode yang berbeda, fungsi ini mencetak pesan kesalahan yang mencantumkan URL dan status kode respons.

Fungsi `save\_to\_csv` digunakan untuk menyimpan data ke dalam file CSV. Fungsi ini menerima tiga parameter: data yang akan disimpan, nama file untuk disimpan, dan opsional, headers untuk kolom-kolom dalam file CSV. Fungsi ini membuka file CSV dengan mode 'w' untuk menulis, kemudian menulis header jika diberikan, dan selanjutnya menulis setiap baris data ke dalam file CSV tersebut menggunakan objek penulis dari pustaka `csv`.

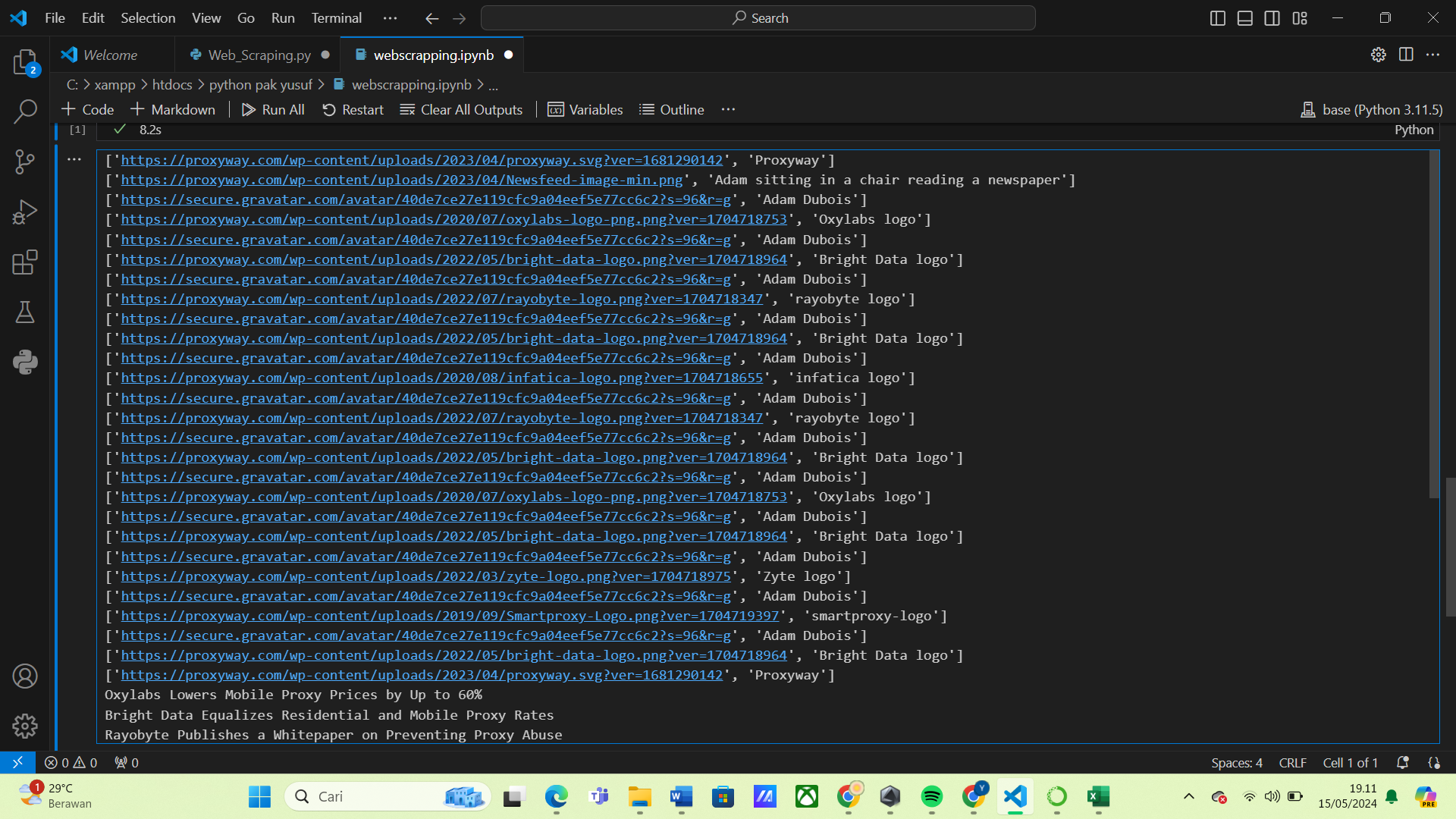
1. **Gambar**
2. **Sourcode**



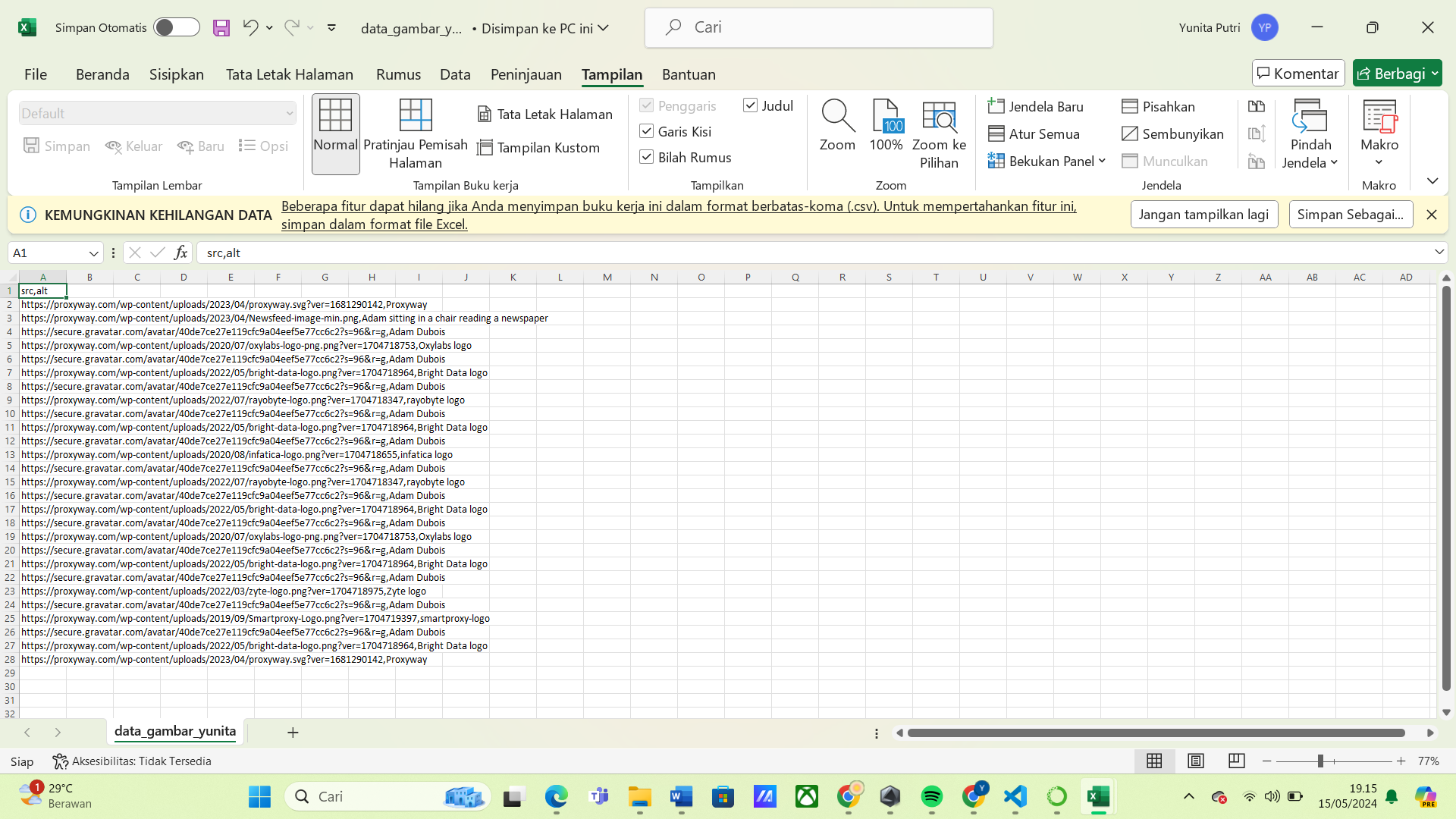
Fungsi `scrape\_proxyway\_images` adalah sebuah fungsi yang dirancang untuk mengumpulkan gambar dari halaman berita situs "Proxyway". Fungsi ini dimulai dengan mendefinisikan URL tujuan sebagai 'https://proxyway.com/news' dan kemudian menggunakan fungsi `fetch\_content` untuk mengambil konten dari halaman tersebut. Jika pengambilan konten berhasil, konten tersebut diuraikan menggunakan `BeautifulSoup`, sebuah perpustakaan Python yang digunakan untuk mengurai dokumen HTML dan XML. Dengan `BeautifulSoup`, fungsi ini mencari semua elemen gambar (`<img>`) dalam halaman tersebut.

Setelah menemukan semua elemen gambar, fungsi ini membuat sebuah daftar yang menyimpan atribut `src` (yang menunjukkan sumber gambar) dan `alt` (yang berisi teks alternatif gambar) dari setiap gambar. Daftar ini kemudian disimpan ke dalam sebuah file CSV bernama 'data\_gambar\_yunita.csv' dengan kolom `src` dan `alt` untuk setiap gambar. Sebagai langkah terakhir, fungsi ini mencetak setiap entri dalam daftar gambar ke layar, yang memudahkan pengguna untuk melihat gambar-gambar yang telah diambil beserta deskripsi teks alternatifnya.

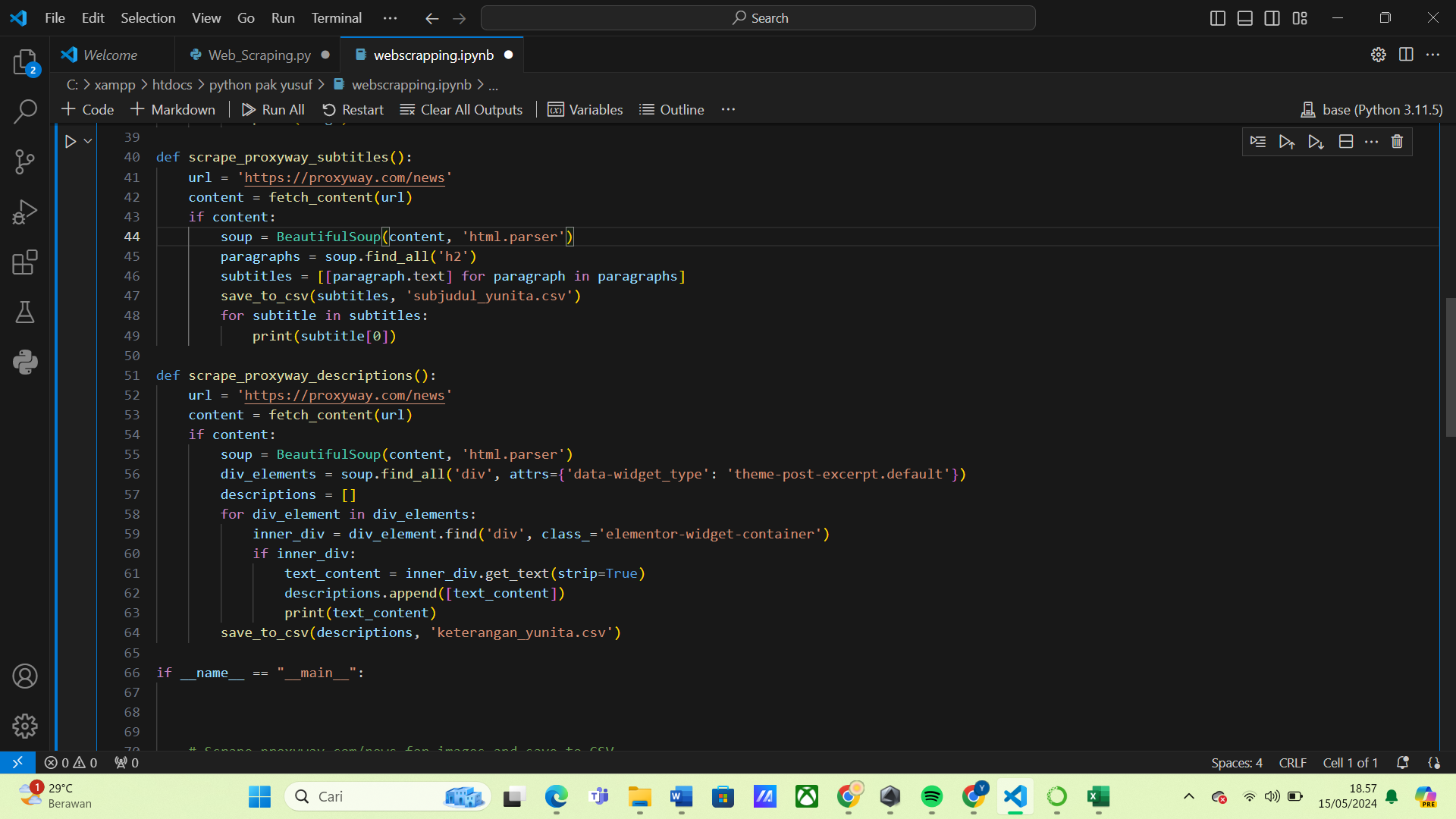
1. **Output**



1. **CSV**



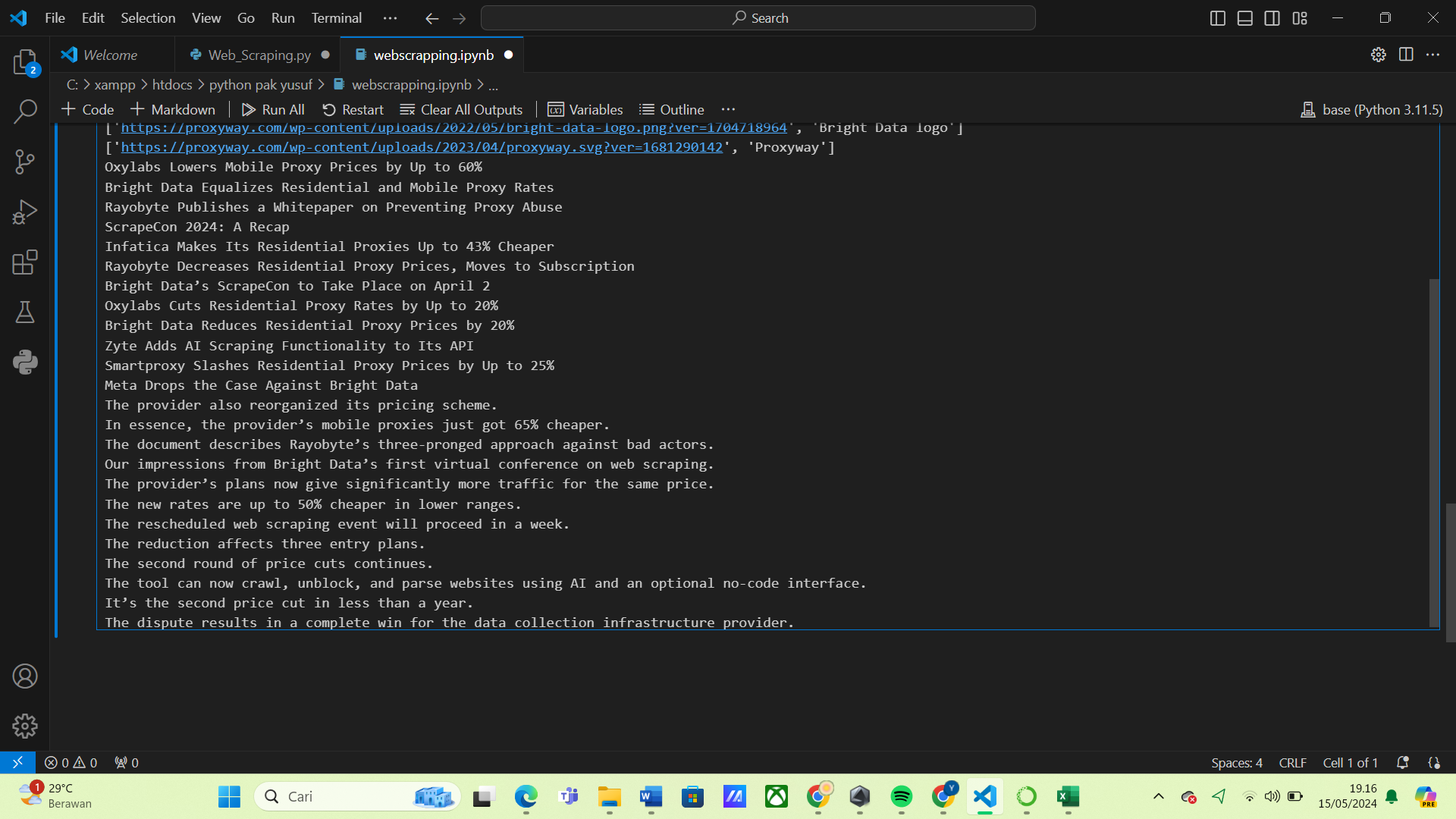
1. **Subjudul**
2. **Sourcode**



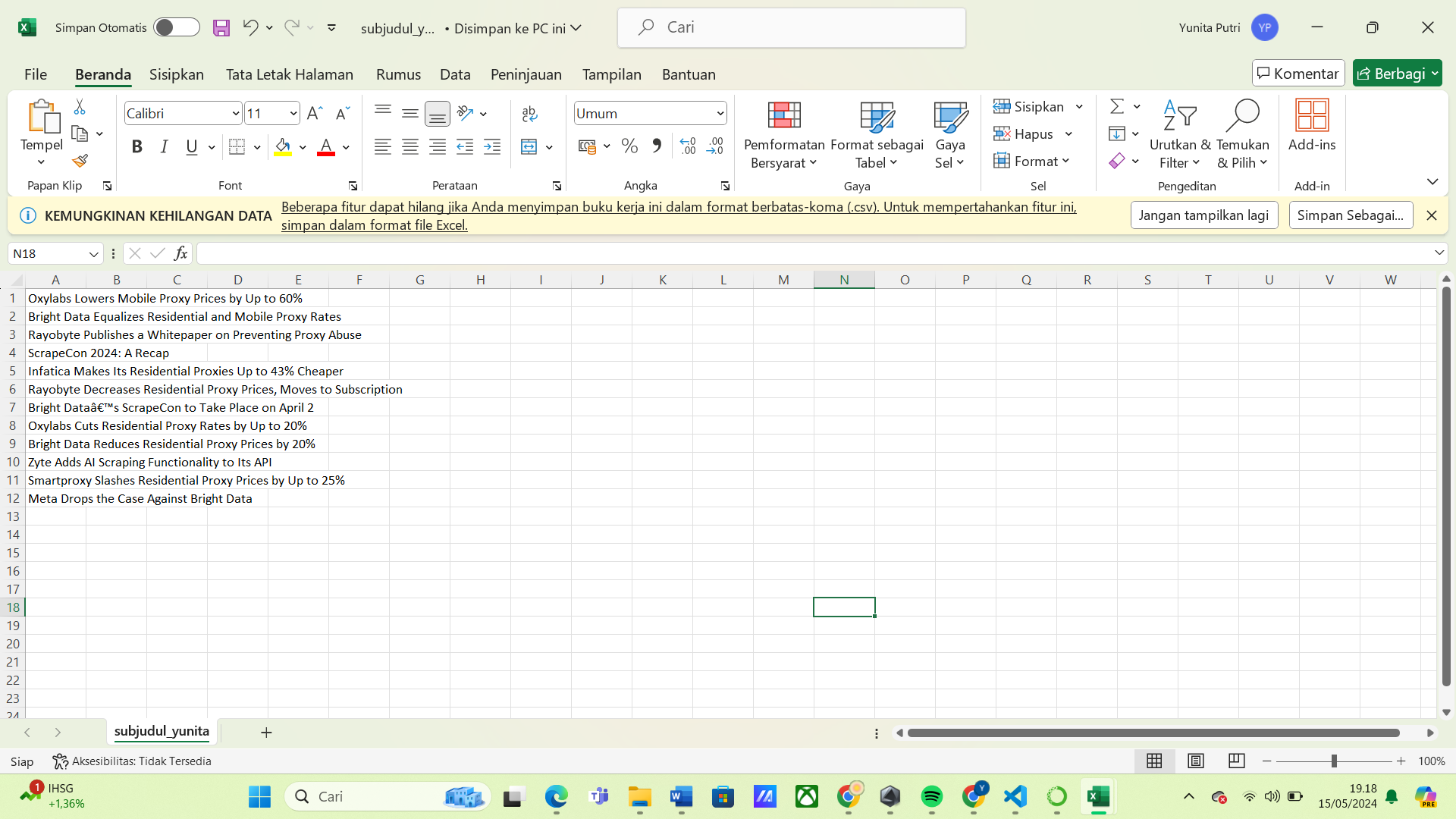
Fungsi `scrape\_proxyway\_subtitles` bertujuan untuk mengekstrak subjudul dari halaman berita situs "Proxyway". Mirip dengan fungsi sebelumnya, fungsi ini dimulai dengan menetapkan URL tujuan sebagai 'https://proxyway.com/news' dan mengambil kontennya menggunakan fungsi `fetch\_content`. Setelah berhasil mendapatkan konten, konten tersebut diurai menggunakan `BeautifulSoup` untuk mempermudah pencarian elemen HTML.

Dalam hal ini, fungsi mencari semua elemen `<h2>`, yang sering digunakan untuk menandai subjudul dalam dokumen HTML. Subjudul-subjudul ini kemudian disimpan dalam sebuah daftar menggunakan pemrosesan daftar, dan daftar tersebut disimpan ke dalam file CSV bernama 'subjudul\_yunita.csv'. Selanjutnya, setiap subjudul dalam daftar diprint ke layar untuk ditampilkan kepada pengguna. Dengan demikian, fungsi ini memungkinkan pengguna untuk melihat daftar subjudul dari halaman berita "Proxyway".

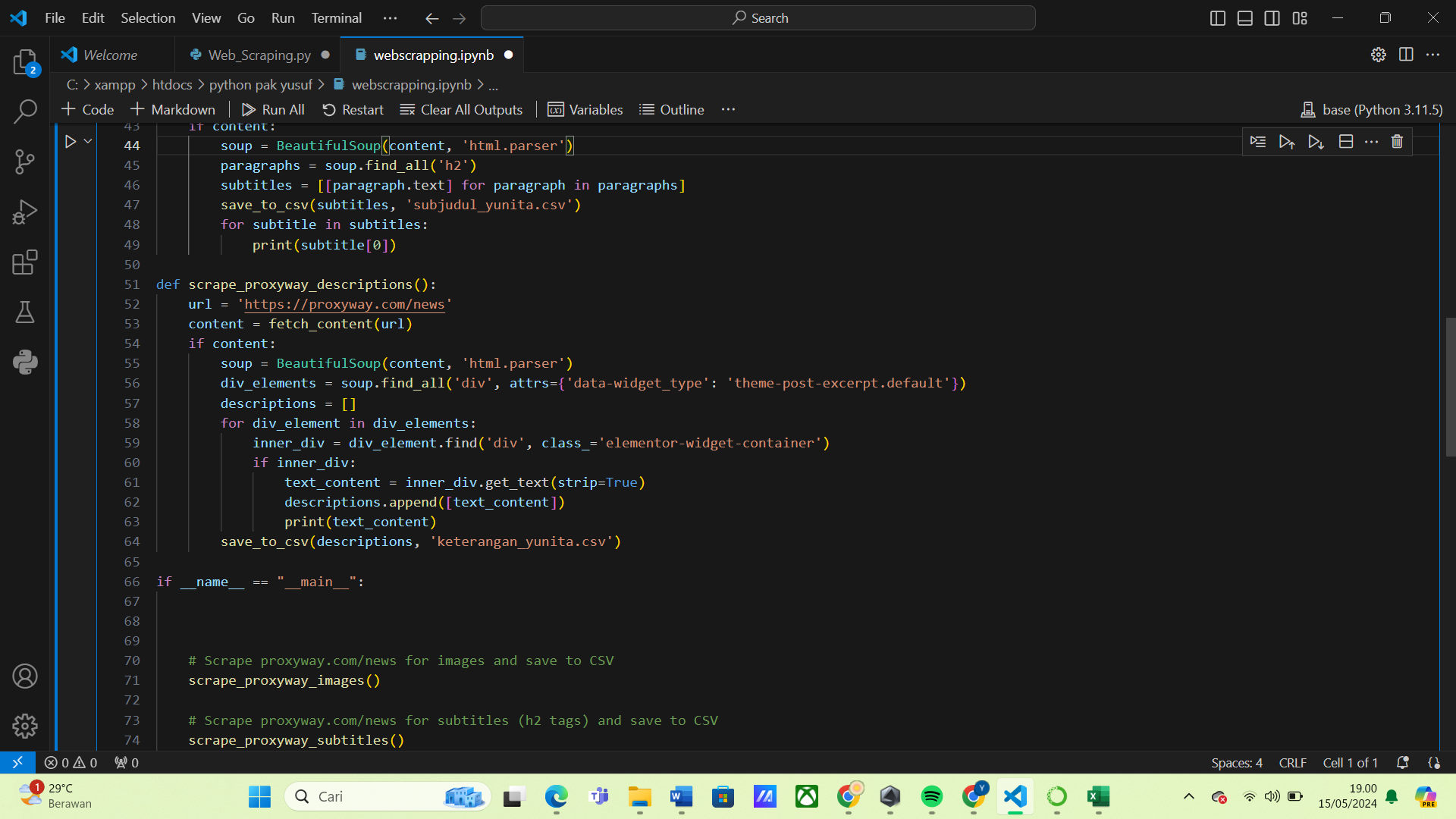
1. **Output**



1. **CSV**



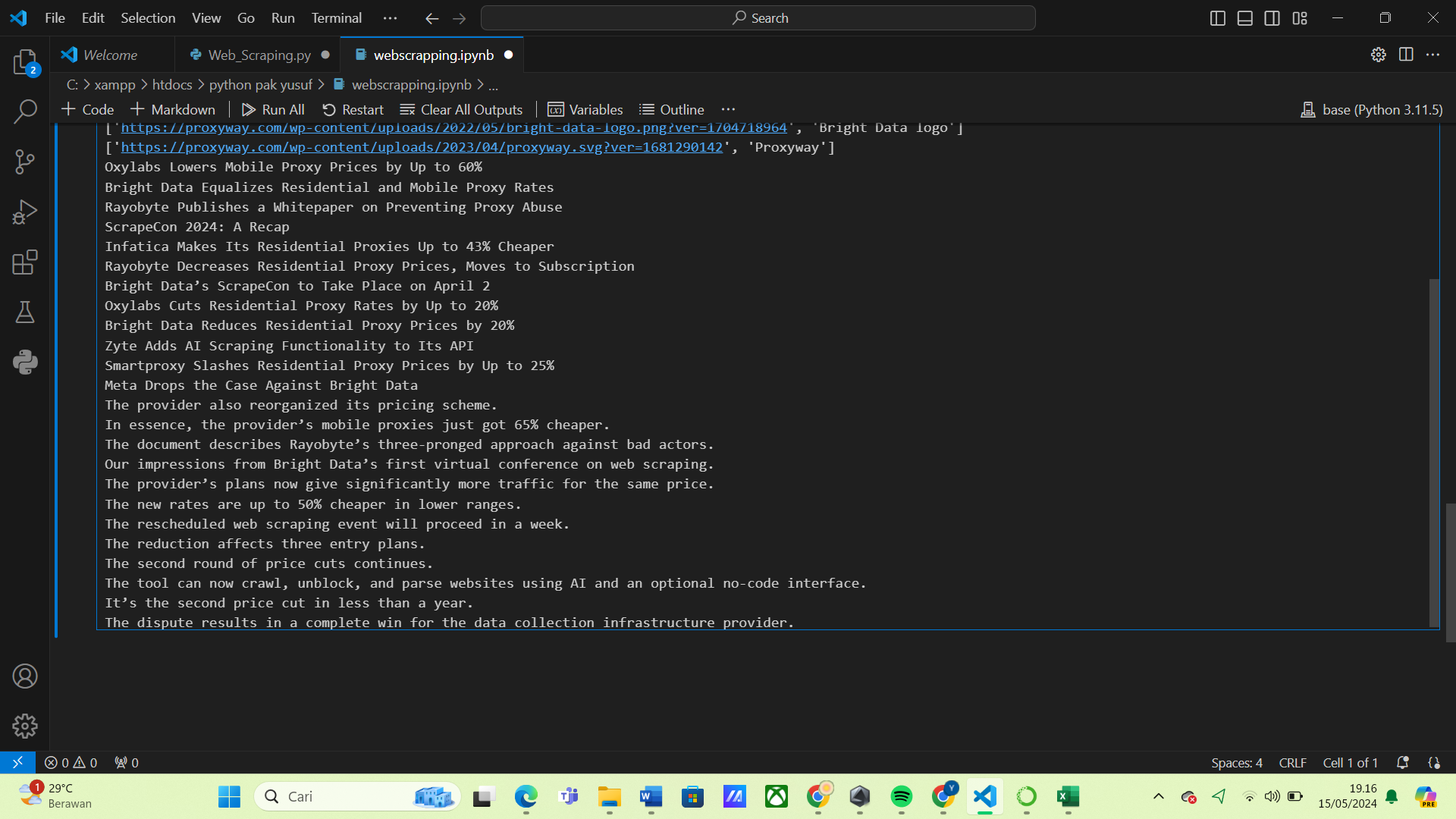
1. **Deskripsi**
2. **Sourcode**



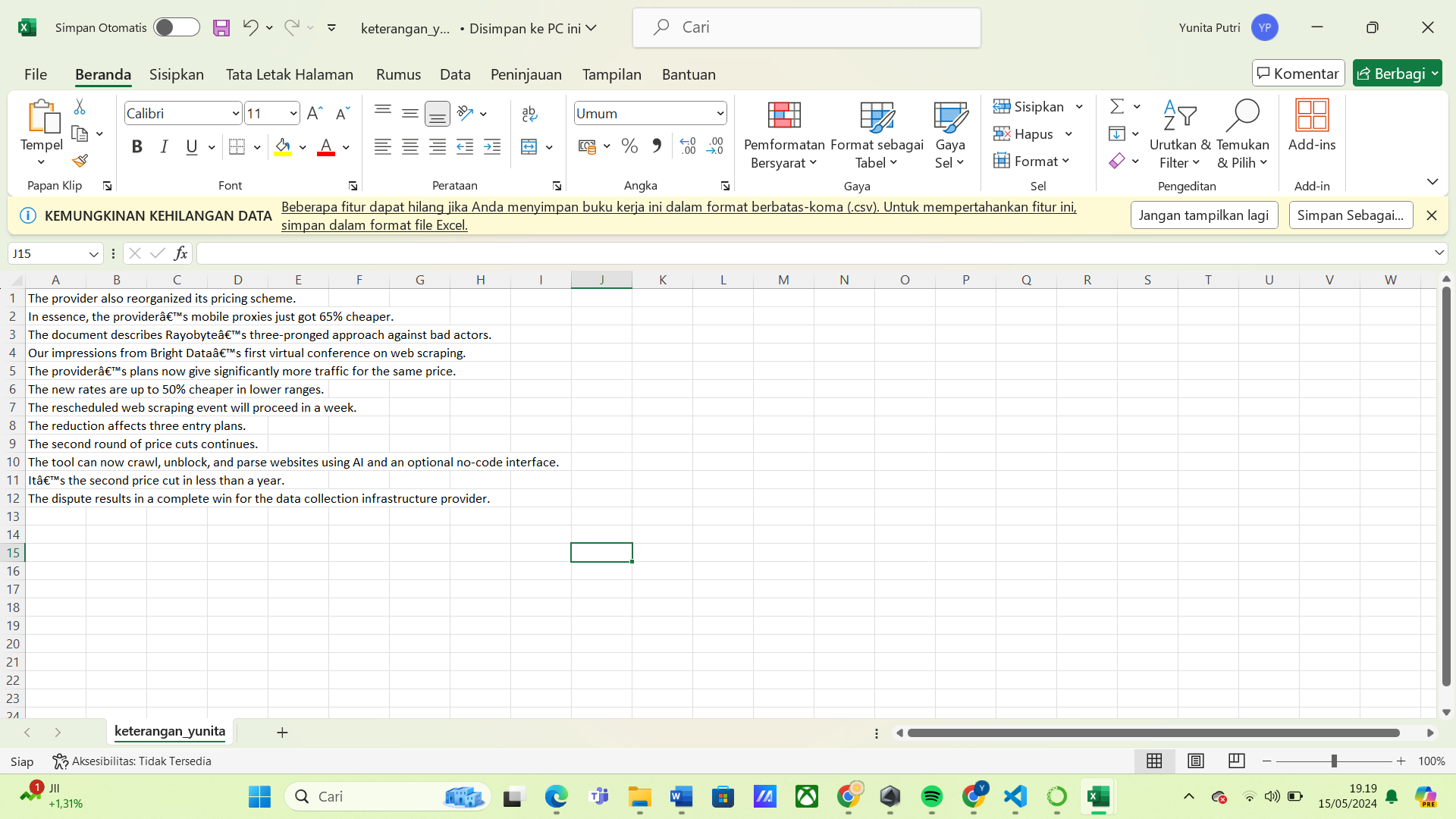
Fungsi `scrape\_proxyway\_descriptions` bertujuan untuk mengekstrak deskripsi atau keterangan dari halaman berita situs "Proxyway". Prosesnya dimulai dengan menetapkan URL target sebagai 'https://proxyway.com/news' dan mengambil kontennya menggunakan fungsi `fetch\_content`. Setelah mendapatkan konten dengan sukses, konten tersebut diurai menggunakan `BeautifulSoup` untuk memudahkan pencarian elemen HTML.

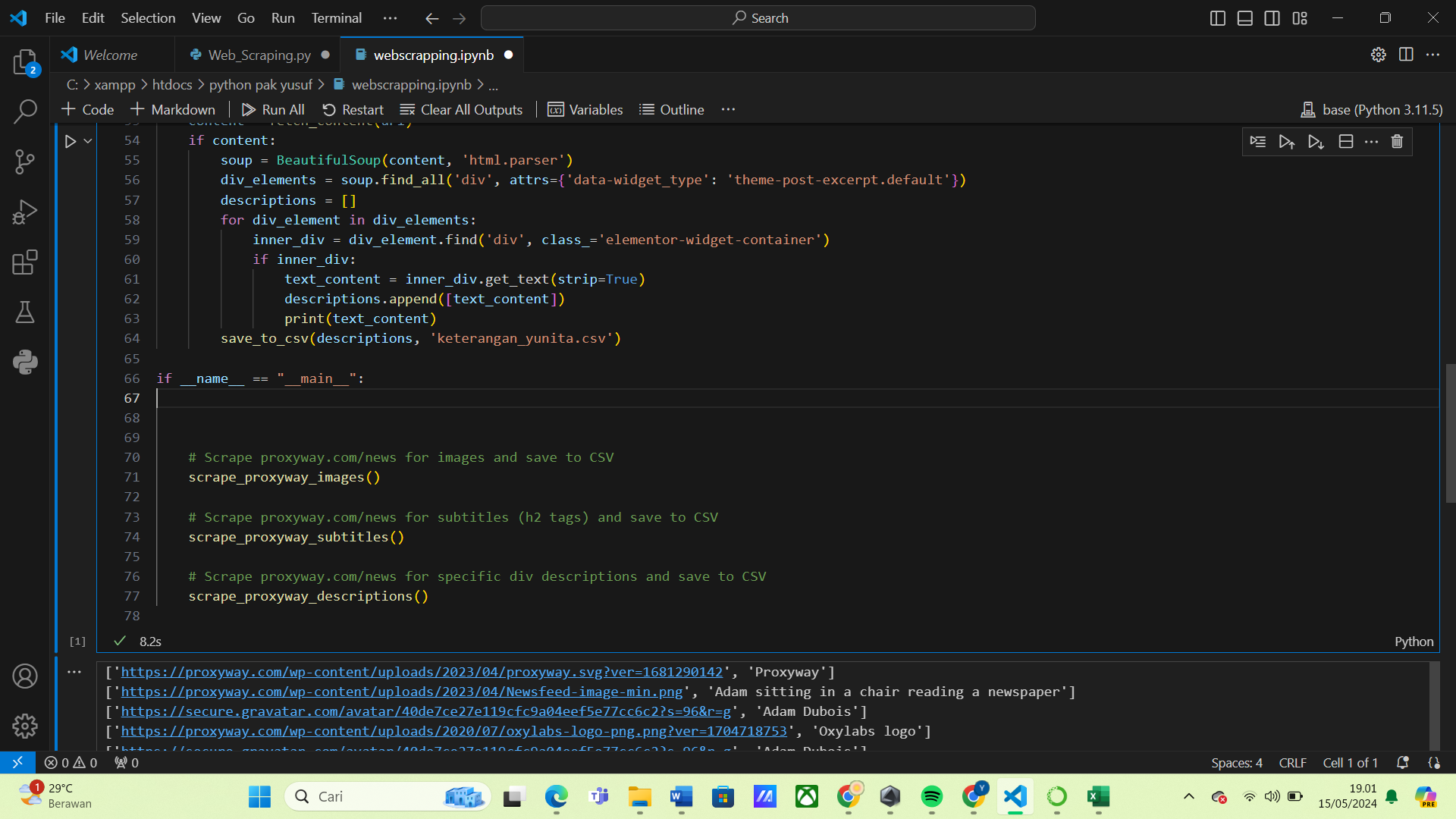
Fungsi ini kemudian mencari semua elemen `<div>` yang memiliki atribut `data-widget\_type` yang sama dengan 'theme-post-excerpt.default', yang umumnya digunakan untuk menampung deskripsi atau keterangan sebuah posting. Setiap deskripsi yang ditemukan diambil dan dimasukkan ke dalam sebuah daftar. Kemudian, deskripsi-deskripsi tersebut diprint ke layar untuk ditampilkan kepada pengguna. Selanjutnya, daftar deskripsi disimpan ke dalam sebuah file CSV bernama 'keterangan\_yunita.csv' menggunakan fungsi `save\_to\_csv`. Dengan demikian, pengguna dapat dengan mudah melihat deskripsi atau keterangan yang terkandung dalam halaman berita "Proxyway".

1. **Output**



1. **CSV**





Pada blok kode yang disediakan, `if \_\_name\_\_ == "\_\_main\_\_":` digunakan untuk mengeksekusi tiga fungsi secara berurutan ketika file ini dijalankan sebagai program utama. Pertama, fungsi `scrape\_proxyway\_images()` dipanggil untuk mengambil gambar dari halaman berita situs "Proxyway" dan menyimpannya ke dalam file CSV. Kemudian, fungsi `scrape\_proxyway\_subtitles()` dipanggil untuk mengekstrak subjudul (yang ditandai dengan tag `<h2>`) dari halaman berita yang sama dan menyimpannya ke dalam file CSV. Terakhir, fungsi `scrape\_proxyway\_descriptions()` dipanggil untuk mengambil deskripsi atau keterangan khusus dari halaman berita tersebut dan juga menyimpannya dalam format CSV. Dengan cara ini, ketiga proses ini dapat dijalankan secara otomatis saat file dieksekusi, memudahkan penggunaan dan pemeliharaan.