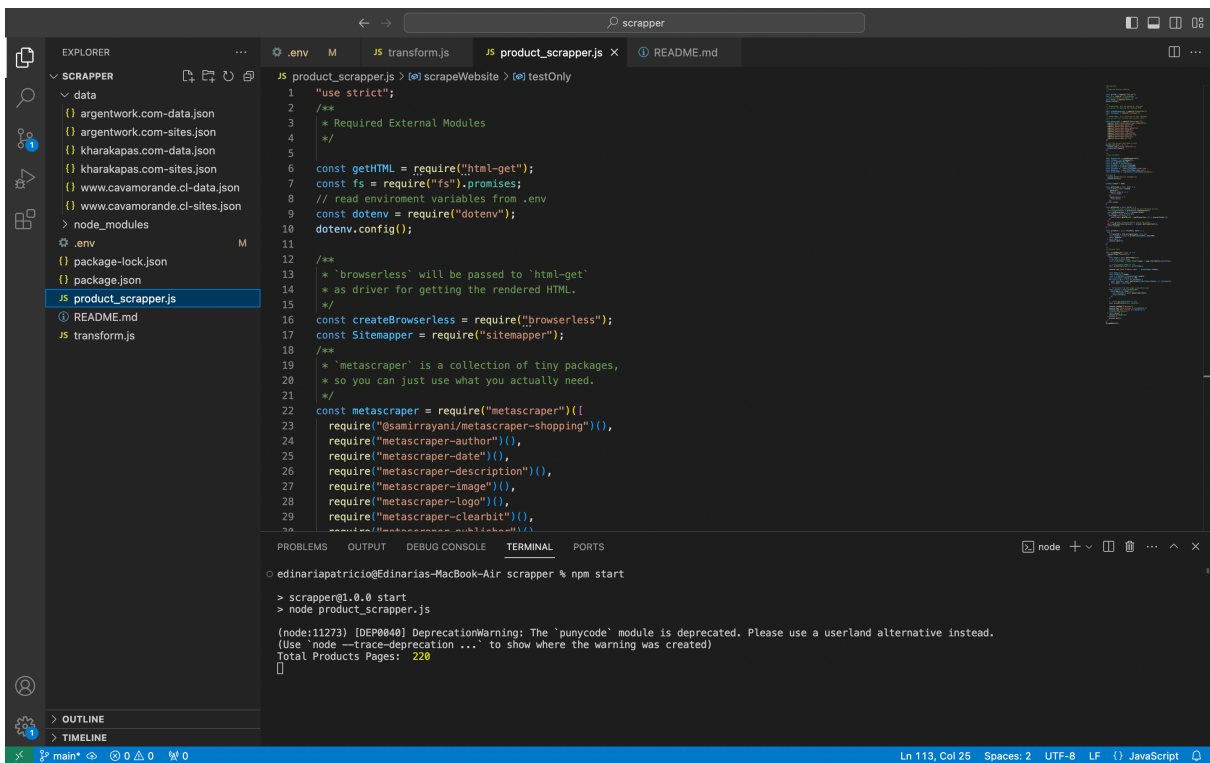


1. Running project

The website tested in the screenshot was URL=<https://www.cavamorande.cl/sitemap.xml>



The screenshot shows the VS Code editor with the file explorer on the left, the editor window in the center, and the terminal at the bottom. The file explorer shows a project named 'SCRAPPER' with a 'data' folder containing several JSON files. The editor window shows the 'product_scrapper.js' file, which is a JavaScript script for scraping a website. The terminal shows the command 'npm start' being executed, which runs the 'product_scrapper.js' file. The output of the script is displayed in the terminal, showing the total number of products (220) and the paths where the data was written.

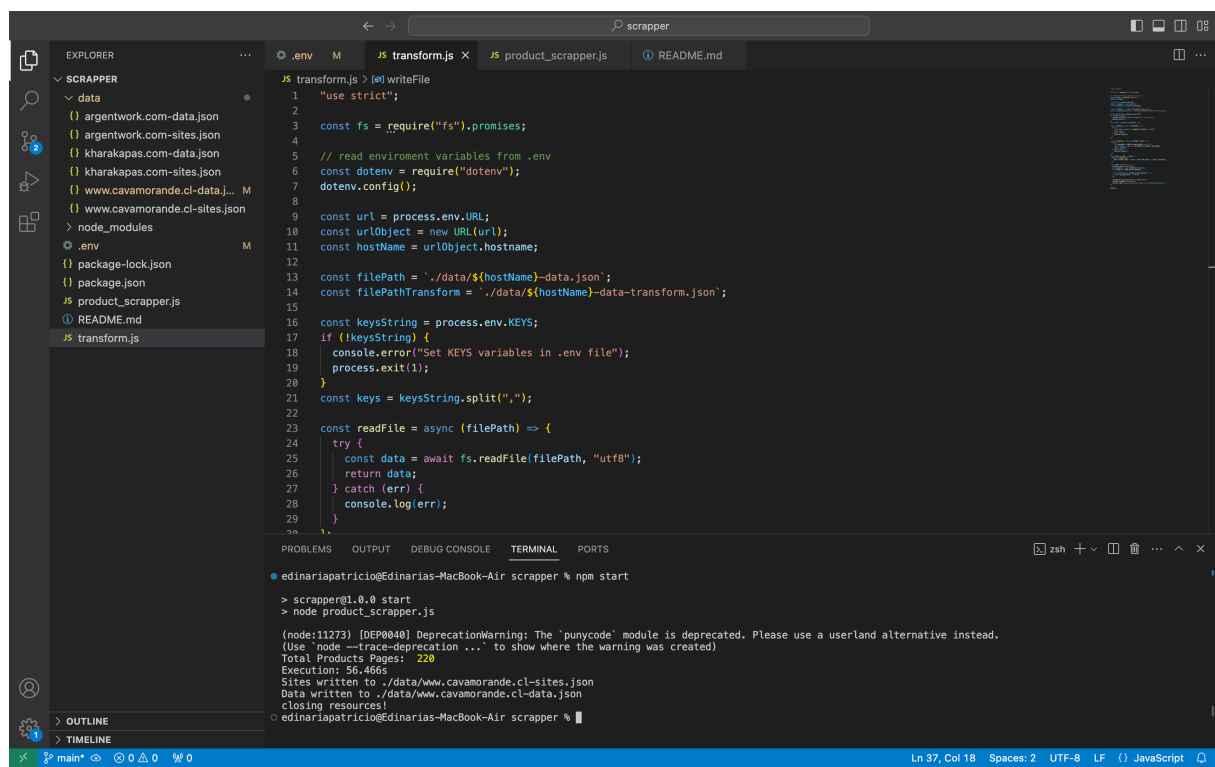
```
1 "use strict";
2 /**
3  * Required External Modules
4  */
5
6 const getHTML = require("html-get");
7 const fs = require("fs").promises;
8 // read environment variables from .env
9 const dotenv = require("dotenv");
10 dotenv.config();
11
12 /**
13  * 'browserless' will be passed to 'html-get'
14  * as driver for getting the rendered HTML.
15  */
16 const createBrowserless = require("browserless");
17 const SitemapParser = require("sitemap-parser");
18 /**
19  * 'metascraper' is a collection of tiny packages,
20  * so you can just use what you actually need.
21  */
22 const metascraper = require("metascraper")({
23   require("@samirrayani/metascraper-shopping")(),
24   require("metascraper-author")(),
25   require("metascraper-date")(),
26   require("metascraper-description")(),
27   require("metascraper-image")(),
28   require("metascraper-logo")(),
29   require("metascraper-clearbit")(),
30 });
31
32 const scraper = createBrowserless({
33   driver: "chromium",
34   headless: true,
35   timeout: 30000,
36 });
37
38 const scraperOptions = {
39   url: "https://www.cavamorande.cl/sitemap.xml",
40   sitemapParser: SitemapParser,
41   metascraper: metascraper,
42 };
43
44 scraper
45   .get(scraperOptions)
46   .then((html) => {
47     const parser = new SitemapParser(html);
48     const products = parser.parse();
49     console.log("Total Products Pages:", products.length);
50     const data = {
51       products: products,
52     };
53     fs.writeFile("./data/www.cavamorande.cl-sites.json", JSON.stringify(data), "utf8");
54   })
55   .catch((err) => console.log(err));
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
o edinariapatricio@Edinarias-MacBook-Air scrapper % npm start
> scrapper@1.0.0 start
> node product_scrapper.js

(node:11273) [DEP0040] DeprecationWarning: The 'punycode' module is deprecated. Please use a userland alternative instead.
(Use 'node --trace-deprecation ...' to show where the warning was created)
Total Products Pages: 220
[]
```

In the terminal I wrote npm start and all files were stored in sites.json.



The screenshot shows the VS Code editor with the file explorer on the left, the editor window in the center, and the terminal at the bottom. The file explorer shows a project named 'SCRAPPER' with a 'data' folder containing several JSON files. The editor window shows the 'transform.js' file, which is a JavaScript script for transforming the scraped data. The terminal shows the command 'npm start' being executed, which runs the 'transform.js' file. The output of the script is displayed in the terminal, showing the total number of products (220) and the paths where the data was written.

```
1 "use strict";
2
3 const fs = require("fs").promises;
4
5 // read environment variables from .env
6 const dotenv = require("dotenv");
7 dotenv.config();
8
9 const url = process.env.URL;
10 const urlObject = new URL(url);
11 const hostName = urlObject.hostname;
12
13 const filePath = `./data/${hostName}-data.json`;
14 const filePathTransform = `./data/${hostName}-data-transform.json`;
15
16 const keysString = process.env.KEYS;
17 if (!keysString) {
18   console.error("Set KEYS variables in .env file");
19   process.exit(1);
20 }
21 const keys = keysString.split(",");
22
23 const readFile = async (filePath) => {
24   try {
25     const data = await fs.readFile(filePath, "utf8");
26     return data;
27   } catch (err) {
28     console.log(err);
29   }
30 }
31
32 const transformData = async (data) => {
33   const transformedData = {};
34   for (const key of keys) {
35     transformedData[key] = data[key];
36   }
37   return transformedData;
38 }
39
40 const transform = async () => {
41   const data = await readFile(filePath);
42   const transformedData = await transformData(JSON.parse(data));
43   fs.writeFile(filePathTransform, JSON.stringify(transformedData), "utf8");
44 }
45
46 transform().catch((err) => console.log(err));
```

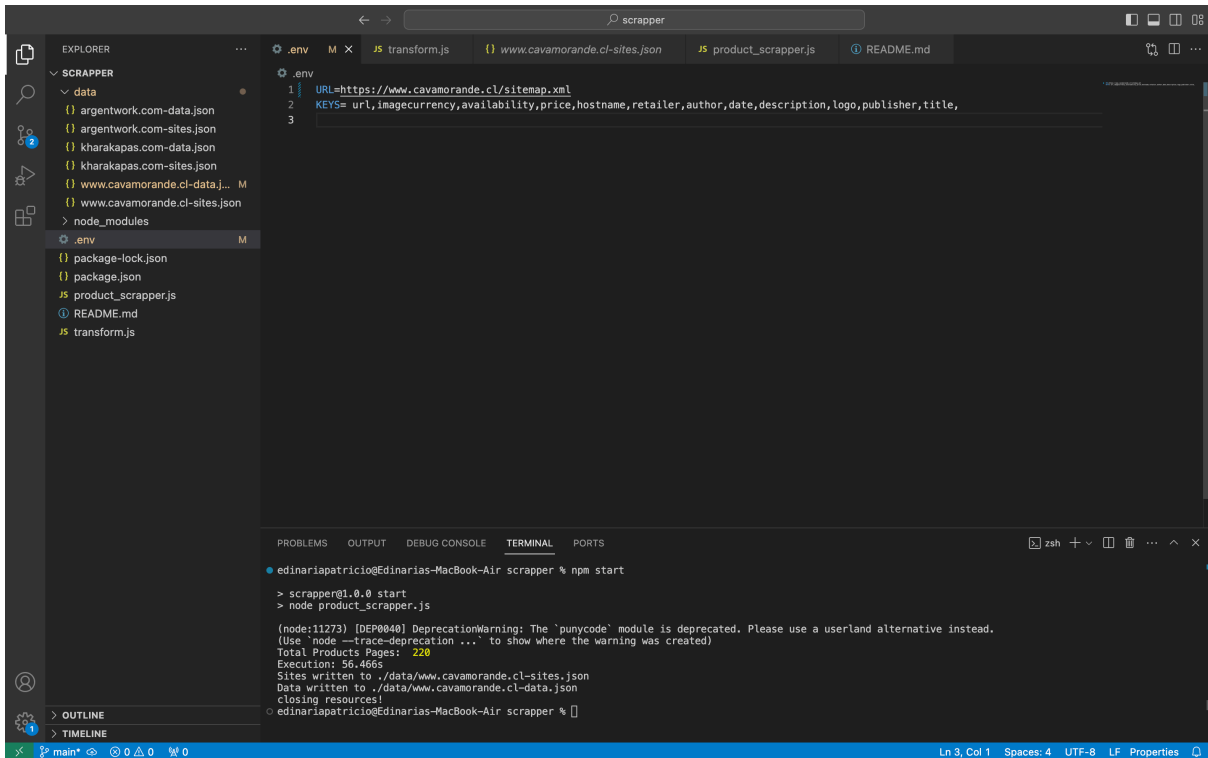
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
o edinariapatricio@Edinarias-MacBook-Air scrapper % npm start
> scrapper@1.0.0 start
> node product_scrapper.js

(node:11273) [DEP0040] DeprecationWarning: The 'punycode' module is deprecated. Please use a userland alternative instead.
(Use 'node --trace-deprecation ...' to show where the warning was created)
Execution: 56.466s
Sites written to ./data/www.cavamorande.cl-sites.json
Data written to ./data/www.cavamorande.cl-data.json
closing resources!
o edinariapatricio@Edinarias-MacBook-Air scrapper %
```

After that, all sites were stored in sites.json and the data of the products containing the keys url/image/currency /availability /price /hostname/retailer/author /date/description,/logo/publisher/title/ were written to the data.json files.

The .env file contains the URL and the keys that will be scrapped after running the code.

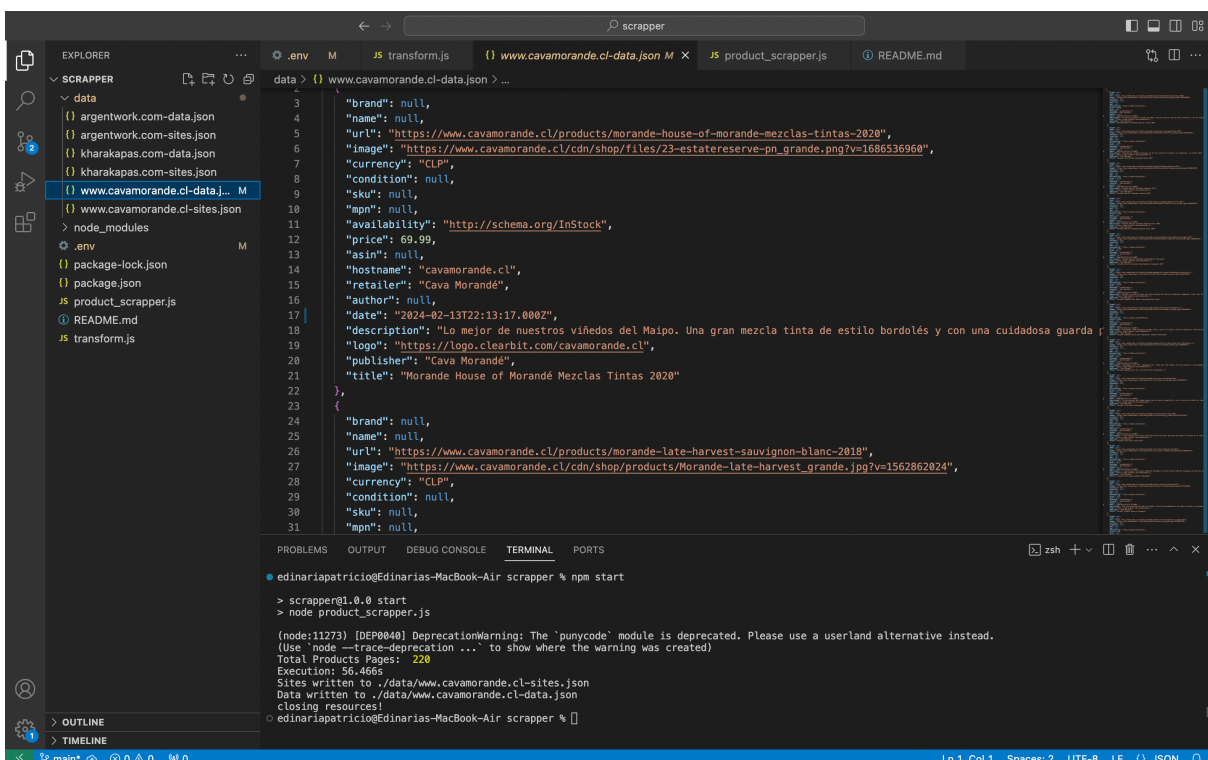


The screenshot shows the VS Code interface with the Explorer panel on the left displaying the project structure. The file explorer shows a folder named 'SCRAPPER' containing a 'data' folder with several .json files, and a '.env' file. The main editor displays the content of the '.env' file, which contains the following code:

```
1 URL=https://www.cavamorande.cl/sitemap.xml
2 KEYS= url,imagecurrency,availability,price,hostname,retailer,author,date,description,logo,publisher,title,
3
```

The terminal panel at the bottom shows the output of running 'npm start' and 'node product_scrapper.js'. The output includes a deprecation warning, the total number of products (220), the execution time (56.466s), and the locations where data was written (./data/www.cavamorande.cl-sites.json and ./data/www.cavamorande.cl-data.json).

The data.json file store all information about the products data.



The screenshot shows the VS Code interface with the Explorer panel on the left displaying the project structure. The file explorer shows a folder named 'SCRAPPER' containing a 'data' folder with several .json files, and a '.env' file. The main editor displays the content of the 'data/www.cavamorande.cl-data.json' file, which contains the following JSON data:

```
{
  "brand": null,
  "name": null,
  "url": "https://www.cavamorande.cl/products/morande-house-of-morande-mezclas-tintas-2020",
  "image": "https://www.cavamorande.cl/cdn/shop/files/23-estatereserve-pn_grande.png?v=1686536960",
  "currency": "CLP",
  "condition": null,
  "sku": null,
  "mpn": null,
  "availability": "http://schema.org/InStock",
  "price": 69.99,
  "asin": null,
  "hostname": "cavamorande.cl",
  "retailer": "Cava Morandé",
  "author": null,
  "date": "2024-02-13T22:13:17.000Z",
  "description": "Lo mejor de nuestros viñedos del Maipo. Una gran mezcla tinta de estilo bordelés y con una cuidadosa guarda",
  "logo": "https://logo.clearbit.com/cavamorande.cl",
  "publisher": "Cava Morandé",
  "title": "Morande House of Morandé Mezclas Tintas 2020"
},
{
  "brand": null,
  "name": null,
  "url": "https://www.cavamorande.cl/products/morande-late-harvest-sauvignon-blanc-2018",
  "image": "https://www.cavamorande.cl/cdn/shop/products/Morande-late-harvest_grande.jpg?v=1562862024",
  "currency": "CLP",
  "condition": null,
  "sku": null,
  "mpn": null,
  "availability": "http://schema.org/InStock",
  "price": 69.99,
  "asin": null,
  "hostname": "cavamorande.cl",
  "retailer": "Cava Morandé",
  "author": null,
  "date": "2024-02-13T22:13:17.000Z",
  "description": "Lo mejor de nuestros viñedos del Maipo. Una gran mezcla tinta de estilo bordelés y con una cuidadosa guarda",
  "logo": "https://logo.clearbit.com/cavamorande.cl",
  "publisher": "Cava Morandé",
  "title": "Morande Late Harvest Sauvignon Blanc 2018"
}
```

The terminal panel at the bottom shows the output of running 'npm start' and 'node product_scrapper.js'. The output includes a deprecation warning, the total number of products (220), the execution time (56.466s), and the locations where data was written (./data/www.cavamorande.cl-sites.json and ./data/www.cavamorande.cl-data.json).