Thought process of my approach to the google AdSense checker

The first thing I thought about is what are the signs that show that there is a google adsense in the site.

How to spot indications and patterns of google AdSense:

Google AdSense has multiple indications that show up on the source code of the site to see the presence of it, I will use these indications for scraping the data that will show up in the 522 sites. Google AdSense ads can be identified through specific HTML tags, JavaScript codes, or network requests. Common identifiers include adsbygoogle.js, googlesyndication.com, and specific div elements with class names related to ads. I will identify them in this project using iframes.

After I know how to spot them, I’m going to make a scraper function that detects and takes that specific data to classify that site as a site with google adsense.

Developing a scraper code that will collect said information:

This code is designed to check for Google AdSense ads on a list of websites using Selenium WebDriver and pandas for handling input and output in CSV format. The script starts by importing the necessary modules and defining the `check\_adsense\_ads` function, which initializes the Edge WebDriver, navigates to a given URL, and waits up to 30 seconds for an iframe containing Google ads to load. If such an iframe is found, it returns True, indicating that AdSense ads are present. The `main` function reads URLs from an input CSV file using pandas, iterates over each URL to check for ads, and writes the results to an output CSV file. The script employs a standard Python idiom (`if \_\_name\_\_ == "\_\_main\_\_":`) to ensure that the `main` function is executed only when the script is run directly, not when it is imported as a module. The use of `WebDriverWait` with `EC.frame\_to\_be\_available\_and\_switch\_to\_it` allows the script to wait for dynamically loaded ads before checking their presence, making it robust for websites where ads load after the main content. This script structure ensures that websites are accurately checked for AdSense ads, and results are systematically logged in a CSV file for further analysis.