WDB Mini Challenge Report

Nasdaq listed companies



Rami Tarabishi

12.08.2023

**Table of contents**

[1 The task at hand 3](#_Toc152956871)

[1.1 My ideas: 3](#_Toc152956872)

[1.2 The chosen website 3](#_Toc152956873)

[2 Data science aspects and real-world use cases 3](#_Toc152956874)

[3 Conclusion 4](#_Toc152956875)

# The task at hand

In today’s data heavy world, being able to extract valuable information from the ocean of data that is the internet has become increasingly important. This report covers the idea, methodology and results that I have worked on this semester.

## My ideas

Originally I wanted to combine the modules NPR and WBD in an immersive mini challenge, which in the beginning of the semester would have led me to scrape multilingual websites for machine translation. Coming to the second mini challenge of NPR, I teamed up with some other students and we ended up choosing a different task in the end, which we we’re also considering combining with WDB to scrape hotel information for our chat bot.

Obviously in the end we ended up settling to not combine the modules and I went on to brainstorm more ideas for my solo endeavor. The ideas I had in mind were:

1. Flight pricing data | To model how prices of flights change depending on how early one books.
   1. Chose not to continue with this as the modeling part would have imo been too much for WDB and I didn’t have another module to combine it with and profit off the work.
2. The original idea of hotel information
   1. Chose not to continue with this due to all of the websites I was planning to scrape having either blocked selenium or terms of use that disallowed web scraping.
3. The classic old stock data
   1. I know it isn’t that original, so instead of historic stock prices I made it the companies financials, I was also partially already planning on doing this with another student for WDB but also for its simplicity and most stock/company tracking websites being open to web scraping.

## The chosen website

I originally planned to use [stockanalysis.com](https://stockanalysis.com/stocks/) due to its simplicity and open terms of use, but decided to switch to the [Nasdaq official website](https://www.nasdaq.com/market-activity/stocks/screener) as they had more available stocks and it seemed to be a similar level of “complexity” navigation wise etc.

# Data science aspects and real-world use cases

Scraping financials and market cap data about listed companies presents a wealth of opportunities from both data science and real-world business perspectives. From a data science standpoint, this assignment involved a couple of simple-ish aspects:

1. Data collection
2. Data processing

From a real-world use case perspective, the scraped financial and market cap data can serve several important purposes:

1. Investment strategies
   1. Investors very commonly analyze the financials of listed companies to adapt their investments.
2. Risk assessment
   1. Investors and institutions can leverage financial data to assess the creditworthiness of companies.
3. Sector and industry analysis (I would need to add a couple more industry specific datapoints for this)
   1. By analyzing the financials of companies of a certain sector or industry, one can gain insight into the health of those markets.
4. Economic indicators
   1. Just like with specific sectors, the financials of companies in certain economic zones or countries can contribute to the calculation of economic indicators.

# Conclusion

So, what did I learn from this mini challenge? To start, this was my first forte in web scraping, so that, I gained insight into automation of websites, selenium, and other web scraping tools like puppeteer. Having worked in web development before starting this bachelor, I already had my fair share of knowledge about the structure of websites and creating/using APIs (although none were used in this mc).

And yes I know stock/company data isn’t that exciting or unique of an idea to do, but due to the time constraints and honestly a little lack of motivation to pursue my original flight pricing idea I just took the easier and more accessible path and ran with it. I do genuinely feel like I’ve learned a lot about data gathering and web scraping so maybe in the future I’ll continue with that idea using the knowledge I gained and make something out of it.