# PS5 - ECON 5253

Hannah Bermudez

February 2025

## 1 Scrape Without API

### 1.1 Research Relevance

I extracted data from a Wikipedia article that provides a list of significant fraud cases. I am interested in behavioral research surrounding fraud and aspects of the audit process, including auditor characteristics. Examining the underlying patterns behind fraud cases could help me identify general tendencies among companies, auditors, and individuals, which could then be applied to smaller populations for more targeted analysis.

In this instance, this data could be beneficial to me in the future for manually collecting and analyzing SEC reports, allowing me to cross-reference findings with the compiled list to confirm any trends in behavior or patterns in fraudulent activity.

### 1.2 Potential Use Cases

This data can be useful for:

- Analyzing patterns in major financial fraud cases.
- Identifying behavioral trends in fraud across different industries.

### 1.3 Online Resources Used

I did not use any online resources/tutorials.

## 2 Scrape With API

### 2.1 Exploring Insights into SEC EDGAR API

The SEC EDGAR API offers structured data from public companies, enabling users to retrieve company filings, financial reports, and other relevant information. I successfully accessed recent filings for Wells Fargo and extracted balance sheet data for Apple using two different API endpoints.I didn't necessarily learn anything interesting; I mostly pulled raw financial data to explore how the API works and to test its functionality. However, I can see how this would be useful for analyzing specific aspects of a company's financial performance and other key characteristics.

### 2.2 Packages Used

- tidyverse
- rvest
- httr
- jsonlite
- tibble