

Mini-Project Research Question 3

Financial services firms spend a lot of time working with financial statements to update their profiles with the U.S. Securities and Exchange Commission. Through SEC, clients, competitors, and potential investors can look at one firm to investigate, analyze, and make a decision based on the firm's historic and current performance. While Cloud's invention has been great to all the companies to store and collaborate files with other team members, downloading the online PDFs with different websites, storing each in different locations, and extracting important information from the documents is a pain. With this repetitive task occurring periodically (monthly, quarterly, semi-annually), an automated process system should be introduced in order to maximize operating time, net revenue, and employees' space to produce more meaningful work. Therefore, I want to contribute to the other scraping projects for financial data by scraping PDF documents, specifically form ADV from the [Investment Adviser Public Disclosure website](#).

Form ADV provides us with the general information about a firm like official address, business hours, phone number, etc. and detailed information like number of employees, types of clients, who the owners are, assets under management (AUM) etc. Without connecting with any analyst from the firm, a person can have enough information to picture the structure and the performance of the company. Therefore, this project will meet the needs of financial analysts or market research assistant by saving their time to go looking, downloading each document, and reading the document to select the valuable information.

Regarding the models for this project, since there is not a way to get access to the previous documents, it is unlikely to perform a temporal analysis with the financial data such as AUM and number of employees and clients resulted from the scraping process. However, we can do firm comparison analysis located in the same state or the same region to explore the financial status of that state or region. Furthermore, we can develop a dashboard with this information to help the analysts or even the board of directors to quickly look at which region is financially hyped, which company is worth to make a visit, etc. In the long term, if we can extract the data from the documents over time, there will be enough data to perform temporal analysis.