**SLIDE 1**

Hello Everyone! My name is Rudy Martinez, and today I will be presenting my Web Scraping Project – Python Automated Financial & Sentiment Analysis.

I’m passionate about personal finance, and this includes a variety of subtopics – investing in the stock market, real estate investing, and other core fundamental behaviors that lead to long term success.

For this project, I chose to focus on one of these exciting topics – investing in the stock market. This activity can be daunting, and often, as an investor, we look for certain metrics to inform our decisions. This is where my project question came into play: **Will a company in the S&P500 that has characteristics of financial strength also be associated with positive public sentiment, and can this analysis be automated at the click of a button?**

I’ll be diving into the details of the project momentarily, but at a high level, I executed on my question by:

1. Scraping S&P 500 companies from **Wikipedia**
2. Scraping the **Yahoo Finance** site for key statistics, financial statements and stock price history
3. Creating a multi-layered stock screen to determine the company's financial strength (fundamental analysis via Piotroski F-Score)
4. Scraping **finviz** news article headings
5. Performing sentiment analysis on the news article headings
6. Analyzing the data and interpreting the results

**SLIDE 2 (CODE)**

The first step was to acquire Wikipedia page content and data for S&P500 companies using the requests and BeautifulSoup packages. By using a table-id located in the HTML, I was able to pinpoint the table of companies and store it in a variable. Using pandas, I created a data frame with these companies and exported the results to CSV.

With this CSV saved off, I used pandas to read the file, create a company\_list variable, and randomly select a company using the random package. This is the company that was used for the upcoming requests and analysis.

With the selected company, I used the requests and BeautifulSoup packages on Yahoo Finance to acquire the company’s financial statement page content and pinpoint the location of the financial data using div and class identifiers. I then deployed a for loop and while loop to create statement headers and scrape each row of the financials. With pandas, I reformatted some of the data to be numerical, replaced null values with 0, and created a data frame that was exported to CSV for the three financial statements.

I again used the same packages and process to scrape Yahoo Finance for 5-years of stock price history and key statistics for the company. This completed my initial data gathering efforts.

**SLIDE 3 (PRESENTATION)**

Before diving into the automated analysis that runs off of these financials, I first want to share the financial metrics that I implemented within the Python script.

To gauge the financial strength of the selected company, I chose to calculate it’s Piotroski F-Score which is a financial indicator designed with nine criteria based on profitability, leverage and liquidity, and operating efficiency.

As you can see here, the criteria fit nicely with conditional statements focused on improvements of the metrics that are on screen. Each fundamental is given a binary outcome of 0 or 1 based the condition.

Overall, an aggregated score that is higher is better and indicates that the company has the characteristic of financial strength.

**Python Code**

Within the script, I read in each of the financial statements using pandas. I initialized a count for each of the scores. Then, I created three separate functions that applied the conditional Piotroski F-Score logic to the financial statements and added the binary outcomes to the score count.

At the end, I called these functions, stored the results into a data frame, and exported the results to a CSV.

**SLIDE 4**

I now had quantified the company’s financial strength. However, I wanted to be able to measure public sentiment of that same company. To do this, I utilized the Natural Language Toolkit (NLTK) VADER Submodule. This tool is a pre-trained model that uses rule-based values tuned to sentiments to evaluate text.

The VADER tool can evaluate negations, contractions, punctuation, slang, and acronyms, it can score comments within a collection of text, and it can produce a compound score that signifies negative, neutral, or positive sentiment.

**Python Code**

The first step was to scrape news article headings from Finviz using the urllib.request package. I used an id tag to pinpoint the headings in the HTML and created a for loop with conditionals to loop through and grab the stock name, date, time, and title of the news article headings.

I then used pandas and the nltk vader package to run a sentiment analyzer on the headings and produce a data frame with the headings and their associated compound scores. This was then exported to CSV.

**SLIDE 5**

After all data was gathered and processed, I developed the following visualizations. We can see a high-level financial overview of the first randomly selected company, Tesla.

Based on the results of the financial analysis, we can see that this company has a high financial strength score of 7. It’s had strong price appreciation over the last three years, and it shows overall company growth based on increases in Return on Assets, Total Equity, Revenue, Expenses, and relative maintenance of Total Debt.

**SLIDE 6**

Based on the results of the sentiment analysis on the headings here, we can see that TSLA had a negative compound score for the majority of the time period that was analyzed.

**SLIDE 7**

For our second company, Electronic Arts or EA, we can see that this company has a mid-range financial strength score of 5. It’s had strong price appreciation over the last five years, and it shows overall company growth based on increases in Return on Assets, Total Equity, Revenue, Expenses, and relative maintenance of Total Debt.

Based on the results of the sentiment analysis on the headings here, we can see that EA had a positive compound score for the majority of the time period that was analyzed.

**SLIDE 8**

Restating my question: **Will a company in the S&P500 that has characteristics of financial strength also be associated with positive public sentiment, and can this analysis be automated at the click of a button?**

For one, this analysis can definitely be automated. Our Piotroski results for this project show Tesla to be a financially strong company with a sentiment score that indicates that negative sentiment outweighs the positive in its article headings.

On the other hand, the results show Electronic Arts to have less financial strength with a sentiment score that indicates that a positive sentiment outweighs the negative in its article headings.

All in all, there may not be a direct association between financial strength results and sentiment results; however, this process is a great way to screen companies in the initial stages of investment searching.

Before concluding my presentation, I will run the Python script to show how everything comes together.

**End**

With that, thank you for your time!