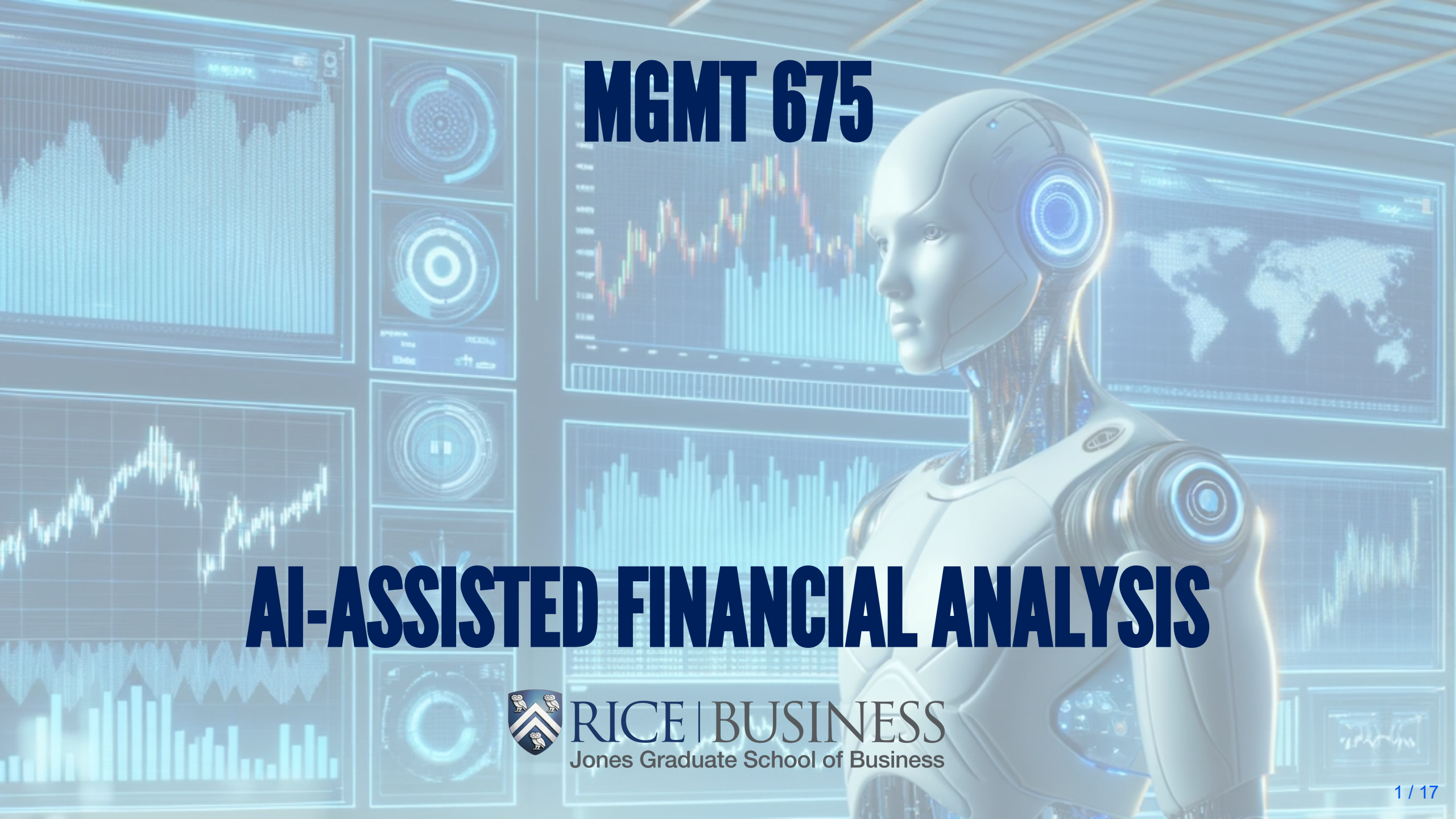

MGMT 675

AI-ASSISTED FINANCIAL ANALYSIS

RICE | BUSINESS
Jones Graduate School of Business

WEB SCRAPING

TOPICS

- Download files from links
- APIs
- Download html tables

DOWNLOAD FILES FROM LINKS

FINRA SHORT INTEREST DATA

- Overview
- Files to Download
- Go to Files to Download and download and view a file.

USE JULIUS

- Give Julius the “files to download” url https://www.finra.org/finra-data/browse-catalog/equity-short-interest/files.
- Ask Julius to read the source code and find the links to .txt files.
  - If you want to read the source code yourself, right-click on the page in your browser and select “View Source” from the dropdown menu.
- Ask Julius to get the first file and print it to the screen.

TXT TO CSV TO EXCEL

- The file uses the pipe symbol “|” to separate items in rows.
- Ask Julius to convert the pipe symbols to commas and to save it as a csv file.
- Download the csv file and double-click on it. It will open in Excel. If you want, you can then use “Save As” and save it as .xlsx.
- You could also ask Julius to read the csv file and then save it as Excel.

CREATE A LOOP

- Ask Julius to read all of the txt files, convert the pipe symbols to commas, and save them as csv files.

GET FILES FOR OTHER MONTHS/YEARS

- Can ask Julius to change the date in form YYYYMMDD in the txt file url to a different date.
- Can loop over dates (all dates in 2024, all dates in 2023, …)

TRY/EXCEPT BLOCK

- Tell Julius to loop over all dates in March 2024, change the url, get the file, convert pipe symbols to commas, and save the file.
- We will need to put the “get file - convert - save” in a try/except block to avoid the loop crashing on non-market days.
- Tell Julius to use a try/except block and to create a list of the dates for which the code exceeded and a list of the dates for which the code failed.

USING DROPDOWN MENUS

- Tell Julius to find the line in the source code with the word “March” and to print the lines surrounding it.
  - To do this yourself, you can use CTRL-F in the source code and search for March.
- The code creates the dropdown menu for months.
- If we can get python to select March, then python could read the March dates. This is not necessary here but could be useful in other cases.

SELENIUM

- Ask Julius to use selenium to select March from the dropdown menu and then find all the links to txt files.
- This may not work. I ran into an error using the built-in Firefox browser.
- Chrome works better, but Julius doesn’t have Chrome. Google colab doesn’t either. But, you can ask Julius to generate code for Chrome and run it on your own machine.

API'S

- Many sites provide APIs that allow you to get data by sending a url crafted in the right way. Frequently, there is a python library to simplify the communication.
- Examples:
  - Yahoo Finance (pip install yfinance)
  - FRED (pip install fred-api)
  - Nasdaq Data Link (pip install Nasdaq-Data-Link)
  - Energy Information Administration (pip install eiapy or eia-python)
  - Yahoo Fantasy Sports (pip install yahoo-fantasy-api)

FRED EXAMPLE

- Ask Julius to communicate via http with FRED and get daily crude oil prices (you can use my API key 3bfa76a79de0dea5d8d19dbf193e6333).
- Ask Julius to show the url it created.
- You can read the code for fred-api at github.

HTML TABLES

- Find the list of S&P 500 companies at Wikipedia.
- Right click and view source code. Do CTRL-F to search for MMM.
- The table shown in the browser is generated from the html code surrounding MMM.
  - tr = table row
  - td = table data
- Ask Julius to read the page and extract the table.