Webscraping Job #2

**Task I**

1. On CapIQ Pro, navigate to Natural Gas > Index of Customers > Pipeline Contract Details
2. You should be able to filter the data by (a) date by quarter and (b) pipeline(s)
3. For the period Q1 2018 – Q4 2021 (16 quarters), please download data for the 11 pipelines specified in the attached csv file (under “pipeline\_sys”)
4. Please stitch the downloaded excel files together

*Recommendations:*

1. Since you can filter by multiple pipelines at a time, there should be 16 total downloads. Accordingly, it may be more time efficient to download data manually instead of writing a webscraper.

**Task II** *(hold off on this until later)*

1. On CapIQ Pro, navigate to Natural Gas > Tariff Summary
2. You should be able to filter the data by (a) pipeline, (b) date by day, (c) rate schedule, (d) rate type
3. From the download in Task I, you will have gotten the appropriate (c) rate schedule to specify for any given (a) pipeline
4. Please leave all options for (d) rate type selected (this should be the default selection)
5. Please either download the resulting table or scrape directly from the website (the second option should be viable, since the table is quite small) for each day in 2018-2021
6. We will need to figure out a way to format/stitch together the result data later

*Recommendations:*

1. Unlike Part I, since you can only filter by pipeline and by day, you should write a webscraper for this task (you’ll have at least 11 pipelines \* 4 years \* 365.25 days = 16,071 downloads and more if there is more than one rate schedule specified for a given pipeline)

**Task III** *(hold off on this until later)*

1. On CapIQ Pro, navigate to Natural Gas > Rate Calculator
2. You should be able to filter the data by (a) pipeline, (b) date by day, (c) rate schedule, (d) receipt zone, (e) receipt point, (f) delivery zone, (g) delivery point
3. From the attached csv file, for any given (a) pipeline, you should have the appropriate (g) delivery point(s) to specify (“point\_name”)
4. From the download in Task I, you will have gotten the appropriate (c) rate schedule to specify for any given (a) pipeline
5. Please leave all options for (d) receipt zone and (f) delivery zone selected (this should be the default selection)
6. Please leave (e) receipt point to whatever the default selection is **(this step may change, as we may have to iterate through all receipt points if there is variation in the resulting table—I don’t think there is, but let me know otherwise as you play around with this)**
7. Please either download the resulting table or scrape directly from the website (the second option should be viable, since the table is quite small) for each day in 2018-2021
8. We will need to figure out a way to format/stitch together the result data later

*Recommendations:*

1. Same recommendations apply here as in Task II