PYTHON TASK

I. Data Gathering

a. Go to: https://www.eia.gov/dnav/pet/pet_pnp_inpt_a_epc0_yir_mbbl_m.htm

b. Create a script to download the data

Deliverables:

- Script used to download data
- Raw data (one file in csv or xlsx format)
  - i. The data units should be: Thousand barrels per day
  - ii. Data granularity (period): Monthly

*There are five PADDs or "Petroleum Administration for Defense Districts" in the United States, PADD 1, PADD 2, PADD 3, PADD 4, PADD 5. These are geographical subdivisions or areas.

II. Data Processing

a. From the data downloaded in the previous step, read the time series for "US Refinery and Blender Net Input of Crude Oil" **by PADD**

b. Keep data from January 2016 onwards, delete all other rows

c. Rearrange the data to be displayed in the following columns ['Year', 'Quarter', 'Month', '(PADD 1) Refinery and Blender Net Input of Crude Oil', '(PADD 2) Refinery and Blender Net Input of Crude Oil', '(PADD 3) Refinery and Blender Net Input of Crude Oil', '(PADD 4) Refinery and Blender Net Input of Crude Oil', '(PADD 5) Refinery and Blender Net Input of Crude Oil']

d. For each year/month, sum up the 'Refinery and Blender Net Input of Crude Oil' data for the five PADDs and assign the name 'Total US Refinery Net Input of Crude Oil' to the time series e. Create a new output table that summarizes monthly data by **'Quarter'** for each of the six time series ('PADD 1 Refinery and Blender Net Input of Crude Oil', '(PADD 2)…' '(PADD 3)… ', '(PADD 4)… ', '(PADD 5)…', 'Total US…')

f. Create a new output table that summarizes monthly data by **'Year'** for each of the six (6) time series ('PADD 1 Refinery and Blender Net Input of Crude Oil', '(PADD 2)…' '(PADD 3)… ', '(PADD 4)… ', '(PADD 5)…', 'Total US…')

Deliverable:

- Script or spreadsheet used for this step
- Three (3) output tables (one file in csv or xlsx format with three different tables: one with data by month, another with data summarized by quarter, another with data summarized by year)