Gorilla assessment

1. Use pandas to calculate a Transportation Distribution Charge for 3 Gas meters in the United Kingdom. Save your code in a Jupyter Notebook and upload to a public repo on Github (or any other platform of your choice). All calculations must be done using vectorized operations.

A Transportation Distribution Charge is a charge levied by the Distribution companies for the use of their lower pressure pipelines. The charge covers the cost of physically transporting the gas through the pipeline.
The cost is variable in time and is determined by the Exit Zone (a regional code) and the (estimated) rolling consumption quantity of the meter.

The cost is calculated by finding the correct rates for the meter and calculating the cost in pence per day by multiplying the forecast for the day by the correct rate for that day. All data needed for the calculation can be found in the Excel file.

Calculate the **total cost per meter** by summing the costs per day for the full period of the forecast (2020-10-01 to 2022-09-30) and converting to £ (1p = 0.01£)

Calculate the **total consumption** by summing the forecasted consumption for the full period.

Your result should be a dataframe of the following form with all numerical rounded values up to 2 decimals:

Meter ID	Total Estimated Consumption (kWh)	Total Cost (£)
10626610		
10588707		
1000000603		

Example:

*For meter **10588707**:*
*Exit zone: **EM2**, AQ:* **75123kwh**

Match the meter to a rate in the Rate Table by matching the correct exit zone and date and determining the correct band.

Determine the band by assuring the AQ is between the Min Annual Quantity (included) and Max Annual Quantity (excluded) columns in the Rate Table.

The following rates are found:

Rates determined for meter *10588707 :*

Date	Exit Zone	Min Annual Quantity (kWh)	Max Annual Quantity (kWh)	Rate (p/kWh)
2020-04-01	EM2	73200	732000	0.0228
2020-10-01	EM2	73200	732000	0.0228
2021-04-01	EM2	73200	732000	0.02336014
2021-10-01	EM2	73200	732000	0.02336014
2022-04-01	EM2	73200	732000	0.02435785
2022-10-01	EM2	73200	732000	0.02435785
2023-04-01	EM2	73200	732000	0.02514696
2023-10-01	EM2	73200	732000	0.02514696
2024-04-01	EM2	73200	732000	0.02596247

The rate from 2020-04-01 to 2020-09-30 is 0.0228 p/kWh
The rate from 2021-04-01 to 2021-09-30 is 0.02336014 p/kWh
etc

Calculate the cost per day for each meter by multiplying the forecast for that day (kWh) with the rate for that day (p/kWh) to obtain a cost in p.

Costs calculated for meter *10588707 :*

On 2020-10-01:
Cost: *0.0228* * 126.367711 = 2.8811838108
On 2020-10-02 :
Cost : 0.0228 * 118.322449 = 2.6977518372
etc

---

2. Write a function that generates a list of random meters of any size. Examples of valid Exit Zones can be found in the rate table. You may randomly generate the Annual Quantity.

3. Write a function that generates mock consumption data given a list of meters and a start and end date.

4. Write a function that takes as an input a meter list and a consumption table and returns the Transportation cost table. Benchmark this function using meter lists of different sizes. Try longer periods as well? How does the function scale?

5. What are your observations after benchmarking? Are there any steps in the cost calculation that can be improved? How would you go about improving the performance of such a calculation?