Nanyang Business School

AN6007 ADVANCE PROGRAMMING

Nov 2024 – Feb 2025

INDIVIDUAL Practical Assignment (20%)

INDIVIDUAL Presentation (20%)

Due: 03 Jan 2025 (Friday) 23:59

Submission instructions:

1. This assignment **must** be done in **python .py files** only. You are allowed to use generative AI to assist you in completing the project but not to start any part of the project. This means that the generative AI is solely used to debug or improve your codings.
2. You should submit your assignment with :

i) a presentation link in the presentation.txt file containing the link and access password if applicable.

ii) a word document containing all logs you have with generative AI, including your questions, data provided to the platform and the response from the platform.

iii) all python .py files containing all algorithms and the relevant testing.

iv) a python file that *import all* ***chosen*** *algorithms* and perform the entire task name ***complete.py***

v) a python file that *import complete.py* and perform the query task name ***query.py***

vi) or more .py files that illustrates your testing and how you generate your analysis

vii) any other supporting files

via the NTU-learn blackboard site in the Seminar Site Assignment folder.

Please ensure that you include your name and admin number in all you .py file(s). You must keep a copy of the final version of your submission and be prepared to provide it on request.

The University treats plagiarism, collusion, theft of other students' work and other forms of dishonesty in assessment seriously. Students shared out their answers or copied answers from others shall receive a zero score. Using AI generated codes without explanation or justification will also be penalized.

Graded result will not be published, only qualitative feedback (if any) will be provided in class after that.

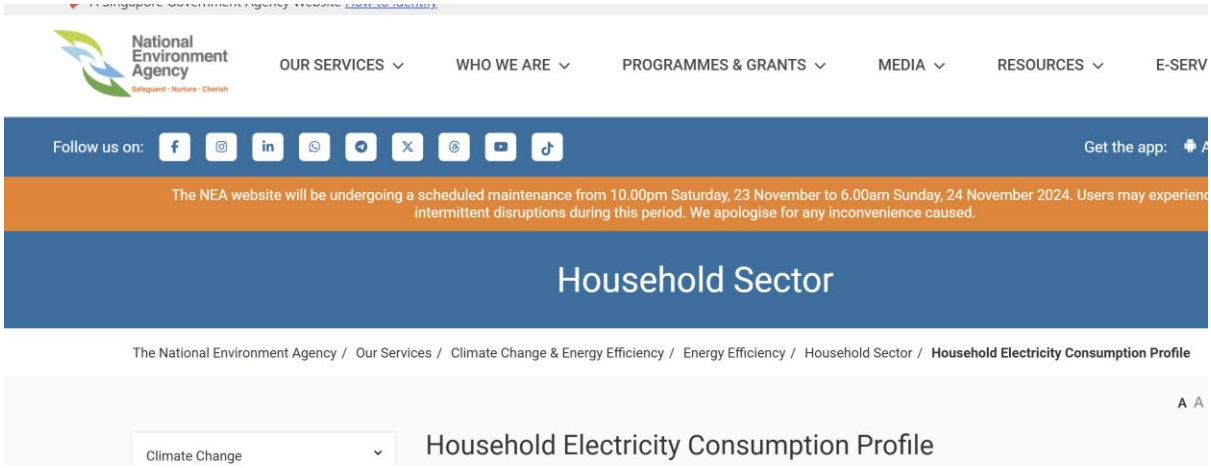
As this is a graded assignment, all debugging shall be done by student himself/herself.

PENALTY FOR LATE SUBMISSION

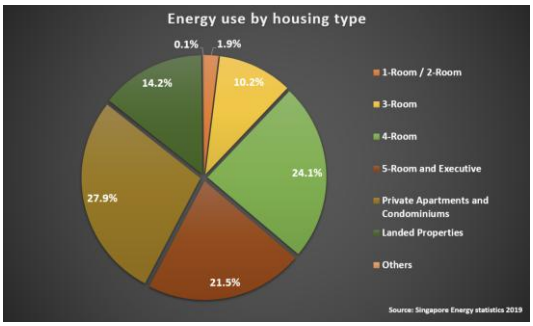
A penalty of 10 marks per day will be applied on the total course marks obtained. No assignments will be accepted on and after 5th day of due date; participants will get "0" marks for this component.

Practical Assignment background knowledge.

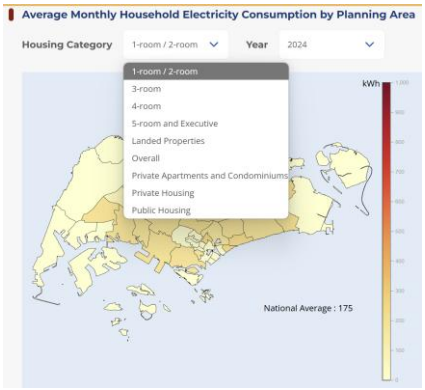
The household sector of NEA (National Environment Agency) of Singapore looks into electricity usage across Singapore.



They will publish on the web site https://www.nea.gov.sg/our-services/climate-change-energy-efficiency/energy-efficiency/household-sector/household-electricity-consumption-profile statistics such as :

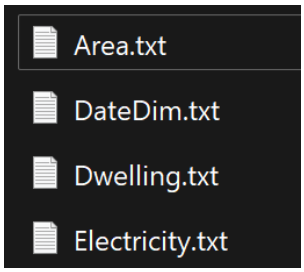


More 'drill-in' analysis is found on https://www.ema.gov.sg/resources/singapore-energy-statistics/chapter3



As this website does not own by NEA, NEA is considering extracting the data from the database and build a prototype for this same purpose with simple tabulations (only at backend, no access from outside world, no web page required).

The IT department has agreed to provide 4 text files to support this initiative:



Area.txt contains AreaID, the actual area and which region is the area belong to, each data element is delimited by ";", a sample as shown here:

```
AreaID;Area;Region
1001;Bishan;Central Region
1002;Sembawang;North Region
1003;Yishun;North Region
1004;Outram;Central Region
1005;Kallang;Central Region
1006;North Region;North Region
1007;Bukit Batok;West Region
```

Dwelling.txt contains the ID and dwelling type delimited by ','

```
TypeID,dwelling_type
1,1-room / 2-room
2,Private Apartments and Condominiums
3,Landed Properties
4,5-room and Executive
5,3-room
6,4-room
```

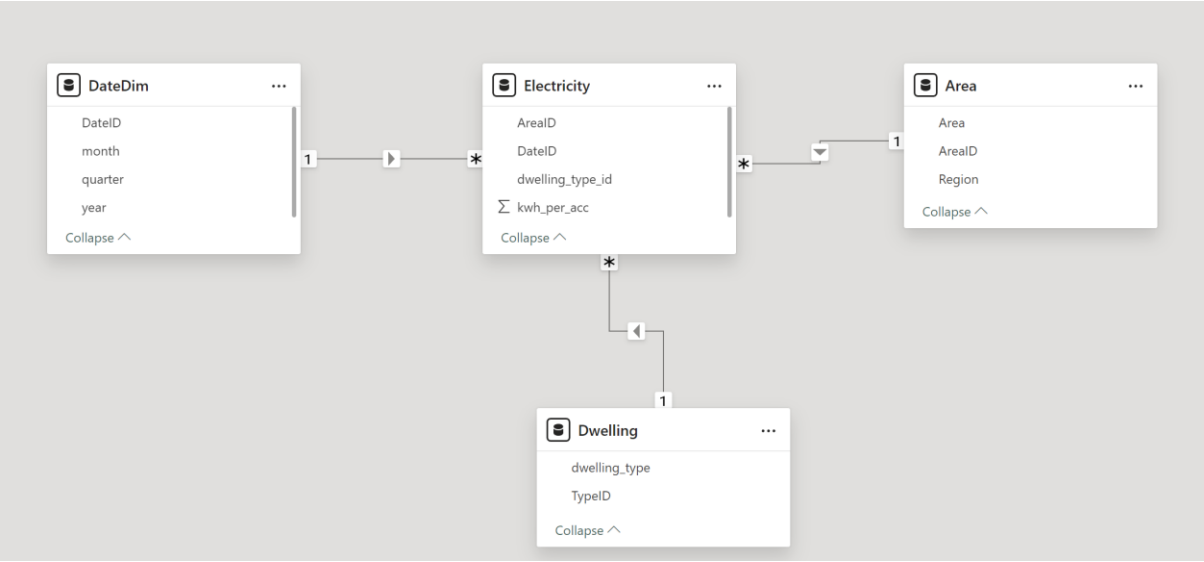
Electricity is containing months of average electricity usage per house hold for a dwelling type in a particular area. Sample as such :

```
DateID;AreaID;dwelling_type_id;kwh_per_acc
54950;1001;1;86.7
54950;1002;1;107.5
54950;1003;1;113.9
54950;1004;1;118.1
54950;1005;1;123.2
54950;1006;1;126.7
54950;1007;1;127
54950;1008;1;127.4
54950;1009;1;131.4
54950;1010;1;131.6
54950;1011;1;137.5
54950;1012;1;138.8
54950;1013;1;142.9
```

```
DateID;year;month;quarter
54950;2010;1;1
19452;2010;2;1
62101;2010;3;1
32574;2010;4;2
12611;2010;5;2
35482;2010;6;2
15862;2010;7;2
89767;2010;8;3
25185;2010;9;3
65896;2010;10;3
80735;2010;11;3
81900;2010;12;4
72867;2011;1;1
62219;2011;2;1
24506;2011;3;1
20409;2011;4;2
```

It is noted that the dateID is the same for the sample shown as this is a unique random number generated for every month, the actual month and year may be obtained from the DateDim.txt in which the records are delimited by ";" :

This, according to the IT department is to ensure that data analytics shall be adhered to the principles of Dimension Modelling with the Star Schema :



Your 2 main tasks for this assignment:

i. Merge all the datasets and produce a denomalised data set in memory with all links in star schema removed :

Region	Area	year	month	dwelling_type	Average of kwh_per_acc
North East Region	Ang Mo Kio	2010	1	1-room / 2-room	144.90
North East Region	Ang Mo Kio	2010	1	Landed Properties	785.20
North East Region	Ang Mo Kio	2010	1	Private Apartments and Condominiums	445.00
North East Region	Ang Mo Kio	2010	2	1-room / 2-room	157.50
North East Region	Ang Mo Kio	2010	2	Landed Properties	896.40
North East Region	Ang Mo Kio	2010	2	Private Apartments and Condominiums	525.40
North East Region	Ang Mo Kio	2010	3	1-room / 2-room	158.10
North East Region	Ang Mo Kio	2010	3	Landed Properties	1,069.90
North East Region	Ang Mo Kio	2010	3	Private Apartments and Condominiums	621.30
North East Region	Ang Mo Kio	2010	4	1-room / 2-room	167.80
North East Region	Ang Mo Kio	2010	4	Landed Properties	985.70
North East Region	Ang Mo Kio	2010	4	Private Apartments and Condominiums	605.70
North East Region	Ang Mo Kio	2010	5	1-room / 2-room	170.40
North East Region	Ang Mo Kio	2010	5	Landed Properties	969.60

ii. Write a simple command line filter program in .py to generate the average household usage of electricity by month for a particular year, region and dwelling type :

Area	1	2	3	4	5	6	7	8	9	10	11	1
Ang Mo Kio	144.90	157.50	158.10	167.80	170.40	169.40	163.70	166.80	156.70	163.70	170.95	178.55
Hougang	147.50	150.00	164.90	161.70	161.10	157.50	144.40	164.00	149.20	155.40	161.95	168.73
North East Region	144.50	156.30	155.90	163.40	165.00	165.60	158.70	163.70	153.80	160.60	167.63	174.99
Punggol				6.40	40.30	66.30	78.20	107.00	98.80	106.20	114.06	122.53
Sengkang	127.40	123.40	106.60	111.10	116.00	163.40	133.90	184.90	156.50	176.10	198.12	222.95
Serangoon	138.80	160.10	145.70	173.10	160.90	183.50	183.60	140.60	180.80	187.70	194.81	202.20

You are required to

a) For task (i)
Use divide and conquer techniques to fulfill the requirements.

For every sub problem, code at least 2 algorithms as functions, perform a Time Complexity analysis to pick the most efficient algorithm. If you have chosen to use dataframe from pandas in 1 of the algorithm, ensure that the other algorithms should not use dataframe again for the same sub problem.

The final dataset generated should be list of objects (a must to test your OOP concepts) where the object is designed for fast access for task (ii)

Combine all the chosen algorithms to form the final solution and the delivery for this section will be ***complete.py*** and all the imports.

b) For task (ii)

The command line program will continue with the data set generated in task(i) by importing the data generation function in task(i).

The program allow user to enter a year, a religion and a dwelling type.

This program should loop until the user enter a '-' for year selection.

The delivery for this section will be ***query.py***

Presentation :

In your presentation, you are **not** required to explain any of your codes.

For task (i),

- explain how you apply divide and conquer method to this problem, what are the algorithms you have proposed
- illustrate how you perform Time complexity analysis.
- discuss how the final selection of algorithm(s) is done
- demo the execution of final.py and show case the dataset required is generated.

For task (ii),

- explain how you design the object and data structure used for this program
- demo how you perform the query with various filtering.
- illustrate the major input validation

Include your learning experience and journey.

If you have used generative AI to assist in completing this assignment, you are required to include why is there a need of use and how have you improved after the use.

The duration of the video should be no more than 5 mins. You are advised to adhere to the time limit strictly.

You are required to show your face in the presentation video and you are advised NOT to read out from scripts while presenting. You may do your presentation with or without powerpoint slides.

Marking Guide : please refer to course outline.