Nanyang Technological University

Nanyang Business School

BC2402 – Designing and Developing Databases

Semester 1, 2022

Group Project

The Green World Project

1. INTRODUCTION

Case Background

Pro-environmental behaviors depend firmly on individuals' sustainability knowledge about what sustainability entails. Some individuals pursue sustainability as a narrow concept focusing on responsible consumption, while others might see sustainability as an all-encompassing concept, pursuing pro-environmental practices in every possible way. In some cases, individuals might assume sustainability issues to be a collective problem, hence remaining a-motivated to adopt pro-environment practices. Scientists have repetitively warned that there is a race against time to perhaps preserve what has not been damaged already in the environment and to repair what is left to be restored. Considering the urgency of environmental issues, it is imperative to devise an effective strategy to address these vast differences in sustainability knowledge.

The promotion of sustainability knowledge is increasingly drawing the attention of educators and layperson citizens. Learning about sustainability is essential in changing people's mindsets that endanger our collective survival. Formal ecological discourse focuses on elucidating how the collective responsibilities in environmental problems. For instance, the workplace environmental training program at the Fujitsu group focuses on providing comprehensive environmental education, such as sustainable development goals, organizational zero-emission strategies, and specialized skill training.[1] By contrast, informal ecological discourse often adapts a personalized discourse strategy. To illustrate, in addition to the workplace environmental training program, the Fujitsu group has also adapted

[1] Comprehensive Environmental Training, retrieved at https://www.fujitsu.com/global/about/environment/education/

experiential eco-tour to facilitate employees' personal and direct exposure to actual ecological issues.[2]

2. Dataset

In this project, we utilized two sources of data:

1. Data on Energy by Our World in Data,
   retrieved from https://github.com/owid/energy-data

2. Singapore Energy Consumption
   retrieved from https://www.ema.gov.sg/singapore-energy-statistics/Ch03/index3

1. Data on Energy by Our World in Data

Our complete Energy dataset is a collection of key metrics maintained by Our World in Data. It is updated regularly and includes data on energy consumption (primary energy, per capita, and growth rates), energy mix, electricity mix, and other relevant metrics.

For details on the sources drawn to construct the dataset, refer to https://github.com/owid/energy-data/blob/master/README.md

You are advised to consult the codebook for descriptions of the data, https://github.com/owid/energy-data/blob/master/owid-energy-codebook.csv

2. Singapore Energy Consumption

The dataset provides a holistic understanding of energy consumption in Singapore by focusing on 7 key sections, namely energy supply, energy transformation, energy consumption, energy balances, energy prices, solar, and manpower. In this project, we focus on some important aspects of energy supply (i.e., imports and exports) and energy consumption (i.e., household electricity consumption and household town gas consumption).

For details, refer to https://www.ema.gov.sg/assets/stat_table/SES_Public_2021_tidy.xlsx

2. Project Deliverables

The due date for the group project is 18 November 2022, 23:59 (23:59 hrs NTULearn server time)

[2] Tsushima, one of the most plastic contaminated islands, retrieved at https://www.fujitsu.com/global/about/environment/activities/japan/ecotours/

There are two key deliverables (and one set of optional deliverables), namely

A. 1 x project report
B. 1 x presentation
C. [optional] database implementations (e.g., relational database and nonrelational database)

A. DATABASE IMPLEMENTATIONS

To allow you to focus on query development, you are provided with the mySQL database. Depending on your team's approach to Q15, the mySQL database might need to include additional tables.

Your team is expected to implement the noSQL (MongoDB). It is not necessary to maintain structural consistency between the mySQL database and MongoDB database. You can rework the structure of the MongoDB database (e.g., changing the JSON structures, joining tables) as you deem necessary.

For both mySQL and MongoDB databases, excessive temporary/redundant tables (collections) will be penalized. Please consult with your instructor on implanting additional tables/collections.

A.1 mySQL database implementation

The specific deliverables are:

- Instructions on deploying the mySQL database (i.e., steps to import the .sql package, which contains the schema and records)
- SQL statements (with expected outputs) for queries (i.e., in a sql or text file) in the Appendix. You may include in-line comments to explain your logic/design.

A.2 MongoDB database implementation

The specific deliverables are:

- Instructions on deploying the MongoDB database
- noSQL statements (with expected outputs) for queries (i.e., in a document file) as depicted in the Appendix. You may include in-line comments to explain your logic/design.

B. PROJECT REPORT

The report should contain the following:

1. A cover page that includes a title, and names as well as matric numbers of each team member
2. (non-technical) Discussion on the data-driven insights specific to Q12 to Q15.

C. PRESENTATION

Your team is expected to deliver a video-recorded presentation (which must be made available via YouTube), in which the team is expected to:

1. Present the nonrelational database design (i.e., a brief discussion on the differences in design between relational and nonrelational implementation)
2. Discussion on the data-driven insights specific to Q12 to Q15.

The entire presentation **MUST** be within 20 minutes (video duration beyond 20 minutes will be ignored). Each member is expected to contribute equally to the presentation.

3. SUBMISSION

A submission folder will be made available on NTULearn. Please zip the files and make a single file submission. One member will complete the submission on behalf of the group.

The following files must be submitted to complete this group project:

A. Database implementations (which include database dump, script files, etc.)
B. Project report (in pdf format)
C. YouTube URL (in a text file, please ensure the URL is viewable)
D. Completed Task Allocation form

The submission must be made by 18 November 2022, 23:59. Do note that video processing and YouTube uploading can be computationally intensive and bandwidth demanding. Please ensure ample time for processing and uploading the presentation video.

Appendix – 15 queries

1. How many countries are captured in [owid_energy_data]?
   Note: Be careful! The devil is in the details.

2. Find the earliest and latest year in [owid_energy_data]. What are the countries having a record in <owid_energy_data> every year throughout the entire period (from the earliest year to the latest year)?
   Note: The output must provide evidence that the countries have the same number of records.

3. Specific to Singapore, in which year does <fossil_share_energy> stop being the full source of energy (i.e., <100)? Accordingly, show the new sources of energy.

4. Compute the average <GDP> of each ASEAN country from 2000 to 2021 (inclusive of both years). Display the list of countries based on the descending average GDP value.

5. (Without creating additional tables/collections) For each ASEAN country, from 2000 to 2021 (inclusive of both years), compute the 3-year moving average of <oil_consumption> (e.g., 1st: average oil consumption from 2000 to 2002, 2nd: average oil consumption from 2001 to 2003, etc.). Based on the 3-year moving averages, identify instances of negative changes (e.g., An instance of negative change is detected when 1st 3-yo average = 74.232, 2nd 3-yo average = 70.353). Based on the pair of 3-year averages, compute the corresponding 3-year moving averages in GDP.

6. For each <energy_products> and <sub_products>, display the overall average of <value_ktoe> from [importsofenergyproducts] and [exportsofenergyproducts].

7. For each combination of <energy_products> and <sub_products>, find the yearly difference in <value_ktoe> from [importsofenergyproducts] and [exportsofenergyproducts]. Identify those years where more than 4 instances of export value > import value can be detected.

8. In [householdelectricityconsumption], for each <region>, excluding “overall”, generate the yearly average <kwh_per_acc>.

9. Who are the energy-saving stars? Compute the yearly average of <kwh_per_acc> in each region, excluding “overall”. Generate the moving 2-year average difference (i.e., year 1 average kwh_per_acc for the central region = 1223, year 2 = 1000, the moving 2-year average difference = -223). Display the top 3 regions with the most instances of negative 2-year averages.

10. Are there any seasonal (quarterly) effects on energy consumption? Visualizations are typically required to eyeball the effects. For each region, in each year, compute the quarterly average in <kwh_per_acc>. Exclude “Overall” in <region>.

    Note: 1st quarter = January, February, and March, 2nd quarter = April, May, and June, and so on.

11. Consider [householdtowngasconsumption]. Are there any seasonal (quarterly) effects on town gas consumption? For each <sub_housing_type>, in each year, compute the quarterly average in <avg_mthly_hh_tg_consp_kwh>. Exclude “Overall” in < sub_housing_type>.

12. *Open-ended question* How has Singapore been performing in terms of energy consumption? Find a comparable reference(s) to illustrate changes in energy per capita, energy per GDP, and various types of energy (e.g., solar, gas, and oil) over the years.

    Hint: The formal technique to identify comparable references is “matching” in econometrics (i.e., propensity score matching, see https://en.wikipedia.org/wiki/Propensity_score_matching). For this question, you may consider countries with somewhat comparable GDP and/or population).

13. *Open-ended question* Can renewable energy adequately power continued economic growth?

14. *Open-ended question* Say micro-nuclear reactors (see https://energypost.eu/micro-nuclear-reactors-up-to-20mw-portable-safer/) have become environmentally viable and economically feasible for Singapore. Shall we go nuclear? Why / why not? Substantiate your team’s opinion with the data provided.

15. *Blue-sky question* Despite the increasing awareness of environmental issues, some remain skeptical about climate change being a problem (see https://www.bbc.com/news/science-environment-62225696). Using the data provided in this project and the individual assignment (as well as any other publicly available data, if your team shall desire), build a convincing data narrative to illustrate climate change problems associated with emissions.

    Notes: When additional datasets are considered, your team must provide the formal references/sources to retrieve the original datasets.
    Evaluation will be performed with attention to the coherence of your team's narrative. A coherent data narrative can be achieved using a focused dataset. A rich, diversified dataset can muddle the narrative if the data is not meaningfully integrated.