

BT3017

Feature Engineering for Machine Learning

Course Project

Due date: 14th April 2023 (Friday) 2359 hrs

Semester 2, AY22/23, School of Computing, National University of Singapore

This is a team project. Each team will comprise two students. Register your group at https://docs.google.com/document/d/1lv3morlv8JpRSpEsHtJ53xs-nOZwt4KxiL_nbm5F5PY/edit?usp=sharing

Question 1 **95 marks**

The datafile *mrtJourneys.csv* contains the travel pattern of 100,000 MRT commuters. The header shows the time of arrival at each MRT station. Each row shows the stations visited by one commuter in the entire journey.

You are an urban planner looking at this dataset trying to figure out how to improve the MRT travelling experience.

- (i) List down a few objectives that you think will constitute a good travelling experience, and what kind of information will you need to achieve these objectives.
- (ii) Write python codes to help you analyse the data and extract the information that will be useful for the objectives that you listed in (i).

Files to be submitted

- Project report
 - Page limit: Not more than 3-pages
 - Font: Arial, font-size 12
 - File format: .pdf
- Video file
 - Time limit: Not more than 3 minutes
 - Presentation: Both team members must present
 - File format: .mp4 or .mov
- Python codes

Submission instructions

For this project, you are supposed to submit your project file to CANVAS.

- *Create a folder using the following naming convention:*

Gp#_BT3017

Example: Gp1_BT3017 if you are in Group-1.

- *Put your python code, report (.pdf) and video (.mp4 or .mov) in the folder.*
- *Zip your folder. Name your zip file using the following convention:*

Gp#_BT3017.zip

Note: Do not submit your file in .rar format

- *Submit the zip file in the “Project Submit Here” folder in CANVAS.*

Question 2 **5 marks**

Use ChatGPT to give a pleasant surprise related to this project. You are responsible for the content submitted.

******* IMPORTANT *******

This is a group project with each group comprising 2 persons. You are not supposed to copy the solution from other groups or other sources without acknowledgement.

Both team members must contribute equally to the technical part of this project.

It is ok to learn from internet, but you must acknowledge the source. If you use anything from internet or from others including your peers, you must acknowledge the source. If you fail to acknowledge the source or mislead people into thinking it is your own work, that constitutes plagiarism.

As this is a substantial assignment comprising 30% of the total grade, students who are caught committing plagiarism or any other form of cheating will get a “F” grade for this course and will face other disciplinary actions.