COS20028 – Big Data Architecture and Application

Semester 2, 2021 – Application Project

Total Marks: 30 pts

Target Participation: Students who target the Distinction **(D)** and High Distinction **(HD)** levels are required to submit this project.

Description:

In this project, you are required to apply any skills you learned in this unit to solve the application problem. The tool(s) you choose to use is/are limited to those introduced in this unit. The basic goal of the targeted information is given in the description. Aside from that, you are required to add two more goals and sieve out the information toward the answer to the additional goals to get a more comprehensive analysis.

Source Data:

"ratings_2012.txt" and "ratings_2013.txt" can be found in the folder "/home/training/traing_materials/analyst/data".

Given Target:

These two files are the collected records from customer feedbacks along the timeline to the products of the company. The company would like to know whether the customer satisfaction (feedback score) is trending up from 2012 to 2013 or is it actually getting worse.

Schema:

Date and Time	User ID	Product ID	Feedback Score	Comments

Hint and Basic Thoughts:

Aside from checking the trending of the feedback score, you can find out things such as how many highly frequent buyers they have in the dataset, which popular product (having more than n feedbacks) has the highest average score, which product has the worst average score, and many other things that can be used to make decisions in the business model.