ICT1002 Programming Fundamentals

Python Team Project Specification

2018 Trimester 1

Objective

The overall objective of this project is to develop one government procurement analysis system to analyze the government procurement every year. The government sectors in Singapore make a large number of procurements via Gebiz. The normal process of the procurement is to submit the tender to the system for quotation and then reward to the vendor with the lowest quotation. The data contains much information for us to understand the important questions such as how the government spend their money, who are the main service providers that always got the tender and so on. To do so, your team is tasked to develop one Python program to analyze the procurement history and provide useful information to users.

The objective of this project is to：

1. Apply the Python programming knowledge learnt in class to solve one real problem.
2. Master the software development circle via a big project.
3. Understand how to develop one big project with teammates using python.

The Detail Requirements

You are given two different datasets: Dataset 1. Government procurement historical records from year 2015 to 2017 in the "government-procurement-via-gebiz.csv". Dataset 2. A list of registered contractors in the "listing-of-registered-contractors.csv".

In dataset1, each row indicates the tender number, agency name of issuing the tender, tender description, award date, tender detail status, supplier name who are awarded and the awarded amount. In dataset2, each row indicates the registered contractors with company name, registration number, grade, expiry date of the contract term, building number/street name/unit no/building name/postal code/telephone number of the company.

The specific requirements and the features that your program shall support are provided as follows.

- Function 1 (5 Marks): As the starting UI, it should allow users to choose two datasets that they want to process. One way to do this is to allow users input the path of the CSV file.
- Function 2 (5 Marks): Read the CSV file and export each government agency's procurement into one text file in the disk. For instance, if there are 10 government agencies, the output of this function will be ten ".txt" files under one folder each of which contains all the procurement information.
- Function 3 (5 Marks): List down the total amount of procurement for each government sectors and allow users to order those sectors either by ascending order or descending order.
- Function 4 (10 Marks): List down the awarded vendors which are the registered contractors.
- Function 5 (10 Marks): List down the amount of the procurement award to registered contractors and non-registered contractors, and top 5 contractors who were awarded by the most procurement.
- Function 6 (25 Marks): Identify one way to categorize the government agency and see how much money the government spend on each category. These categories can be like IHL (institute of higher learning), transportation agency etc.

- Open Functions (45 Marks): Based on the data given, you may provide some other features in the system to identify some insight from the data. Examples like: analyzing the tender description and see what are the main categories of the money spent for each agency (HINT: to do so, some natural language processing techniques can be used such as topic modeling); Analyzing the trends of the money spent; who are the big player (company) that always awarded the tender. PS. to make your system interesting, you can also integrate other dataset in your system.

You need to design one user-friendly UI such that your program is easy to use. As one example, the program should be able to allow users to choose which function to perform and display the result nicely. You can build any innovative UI. Note that it is not compulsory to build a GUI. However, building one GUI will be considered as one advanced features of your program.

Please keep in mind that, in a big data world, the datasets you are handling can be large. Hence, you may want to optimize your code to be memory/algorithm efficient and effective.

Task Allocation

Each student will contribute **at least** 20 marks of the project.

Extra credit

Extra credits (Up to 20 Marks) will be given for innovative features and ideas of how the dataset can be used to help the business.

Timeline and Deliverables

Sunday 14 Oct 2018 11:30pm	Final Submission	Submit your group's xSITE Dropbox ('Project 1 Final Submission') a zip file containing your final report, all source code files and one 2-3 minutes' video to show your system. Your final report should contain the following: • Description of each team member's contribution. NOTE that every member must contribute in coding. • Summarize the technologies applied in the project include the data structures and main APIs. • Description of the main features, especially for the open functions. • The screen capture for each function. You may include the screen captures to show your program's user interface. • Description of the unique features in your project. Point out any innovative and unique feature that your program has. • Limitations of your program. You will have to list all your limitations of your program and interesting future works about how the dataset can be better utilized.

		In the video, please show all the main features of your system. One Zip file containing: report pdf, source code and all related implementation files. ***One submission per team only***
TBD	Project presentation	10 minutes presentation per group

Late Submission

A penalty of 20% per day for each deliverable will be imposed for late submission unless extension has been granted prior to the submission date. Request for extension will be granted on a case-by-case basis. Any work submitted more than 4 days after the submission date will not be accepted and no mark will be awarded.

Plagiarism

SIT's policy on copying does not allow you to copy software as well as your assessment solutions from another person. It is not acceptable to copy other person's work. It is the students' responsibility to guarantee that their assessment solutions are their own work. Meanwhile, you must also ensure that others don't obtain access to your work. Where such plagiarism is detected, both of the assessments involved will receive ZERO mark.

Assessment criteria

Your assignment will be assessed according to the criteria listed in the mark scheme in Table 1.

Table 1 Group Assessment

Criteria	Weight
System functionality and program quality: Each requirement/function will be evaluated based on the following criteria: • Code clarity: easy to read and understand with comments and good modularity • Code completeness and correctness: implement all required features correctly and efficiently. • Code reliability: good data validation, no bugs or errors	40
Report • Your report should include all required information • Your report should be well organized and ease-of-read without errors	20%
Presentation/demonstration • Conduct a clear presentation and demonstration	30%
Video • Your video should cover the main features of your system	10%

The grade for each individual member will be weighted based on the contribution and peer evaluation.