# ICT1002 Programming Fundamentals

Team Assignment 1 Specification

2016 Trimester 1

## Objective

The overall objective of this project is to develop one sales transaction analysis program to analyze all the shopping/food transactions in different restaurant or merchants. Today, most of the merchants use the POS (point of sales) machine to print the receipts for their customers indicating the shopping transactions and prices. Normally, the receipts include important information such as the transaction time, item purchased, the price of each item, merchant information etc. These transactions provide the merchants valuable information for them to analyze the customer behaviors, shopping patterns and so on. Therefore, as a professional software team, you are given a dataset including such transactions and required to develop one Python program to perform different data analytics obtaining all the useful information for the decision makers to understand their business.

The project is designed based on one real-world problem which is also one of the most popular problems that most IT companies are facing. By successfully finishing the project, you will not only get chance to practice what you have learnt in class and beyond, but also obtain the hands on experience in solving one real-world problem. Please note that there 'may' be a potential scope of internship, subject to the successful completion of the project.

## The Detail Requirements

The dataset given is in one CSV file "sample_data.csv". The sample dataset consists the receipts from different merchants. Each merchant includes multiple receipts with various information and one unique receipt_id. The specific requirements are provides as follows.

- Program Input: one CSV file with the receipt information from different merchants. Please note that you will be provided another dataset with much more information to evaluate your program during the project final demonstration and evaluation. However, for simplicity, you can assume the receipts in the final dataset are from the same companies as in the sample dataset with the same format.

The functionalities/features that your program should support are listed below:

- Function 1 (5 Marks): As the starting UI, it should allow users to choose the dataset 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 receipt information into one text file in the disk. For instance, if there are 10-receipt information, the output of this function will be ten ".txt" files under one folder.
- Function 3 (5 Marks): List the total number of receipts of each merchant in the file.
- Function 4 (10 Marks): List the total sales amount for each merchant.
- Function 5 (10 Marks): List all the items sold in each merchant.
- Function 6 (30 Marks): Extract the most important data information from each receipt and store them into another CSV file. The information that you have to extract includes Receipt ID, merchant name, merchant address, shopping date, shopping time, shopping items (separated by ","), the number of items (separated by ","), cost of per item (separated by ","), total cost. Below is one example of new information for receipt id 25007 in the same dataset:

| Receipt ID | Name | Address | Date | Time | Items | Amount | cost | Total |
|---|---|---|---|---|---|---|---|---|
| 25007 | Chin Wan Logic PTE LTD | 10 UBI CRESCENT, #04-07, UBI TECH | 4/7/2016 | 10:40 | Carrot Cake, Chocolate Cake | 1,1 | 2.5,3.5 | 6 |

- Function 7 (20 Marks): Discover all the associated items or user shopping patterns. You may need to use existing association algorithms or propose your own algorithm to detect the correlated items which the customers always purchase together. All of correlated items or shopping patterns can be utilized to help the business perform promotions or recommendations to different customers.
- Function 8 (15 Marks): This is an open question where you can propose your own idea about how you can analyze the dataset to better make sense of the data. In order to improve the merchant's business, you may propose and implement one your self-defined function to better analyze and utilize the data you have.

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 20 Oct 2016<br>11:30pm | Final Submission | Submit your group's xSITE Dropbox ('Assignment 1 Final Submission') a zip file containing your final report, and all source code files.<br>Your final report should contain the following:<br>• Description of each team member's contribution. NOTE that every member must contribute in coding.<br>• Summarize the technologies applied in the project include the data structures and main APIs.<br>• Description of the main methodology/algorithm of the major tasks such as function 5, 6 and 7.<br>• The screen capture for each function. You may include the screen captures to show your program's user interface.<br>• 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.<br>One Zip file containing: report pdf, source code and all related implementation files.<br>***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 (Group Assessment) and Table 2 (Individual assessment). In particular, the assessment includes three different factors: group assessment, peer evaluation and individual assessment. The group assessment takes 70% and individual assessment will take 30%. The group assessment weighted by the peer evaluation marks will be your final group marks. Your final mark will be calculated as follows:

Group_mark*70%*Peer_evaluation + Individual_mark*30%

Table 1 Group Assessment

| Criteria | Weight |
|---|---|
| Each requirement/function will be evaluated based on the following criteria:<br>• Code clarity: easy to read and understand with comments and good modularity<br>• Code completeness and correctness: implement all required features correctly and efficiently.<br>• Code reliability: good data validation, no bugs or errors | 10%<br>40%<br>20% |
| Report<br>• Your report should include all required information<br>• Your report should be well organized and ease-of-read without errors | 20% |
| Presentation/demonstration<br>• Conduct a clear presentation and demonstration | 15% |

Table 2 Individual Assessment

| Criteria |
|---|
| Each individual is required to implement a code portion that is worth of **at least 20 Marks**. For each individual's code, it will be assessed based on the same criteria as the group assessment (see Table 1). |