

Programming Concepts and Practice

Academic Year: 2021/2022, Semester 2

Assignment 1

Module Leader: Dr. Bayode Ogunleye		Level: 7
Module Name: Programming Concepts and Practice		Module Code: 55-706555
Assignment Title: PCP Assignment 1		
Individual	Weighting: 60%	Magnitude: <i>wordcount: 2500</i>
Submission date/time: 28-March-2022, 14:59	Blackboard submission: Yes Turnitin submission: Yes	Format: source code, digital media, report.
Planned feedback date: 18-04-2022	Mode of feedback: Written and verbal feedback	In-module retrieval available: Yes

Module Learning Outcomes

- LO1: select appropriate programming techniques and data structures to develop effective software implementations of relatively complex systems using an appropriate programming language
- LO2 apply relevant program design strategies to the implementation of software applications using that programming language.
- LO3: design and implement well-engineered, domain specific software using that programming language

Assessment Brief

1. Introduction

In the electronic world, fraud is quite prevalent. Fraudulent transactions can be found in insurance, Internet auctions, benefit claims, and telecommunication services. When you engage in any online based transactions, there is a probability that you are dealing with a fraudster. Thus, the ability to identify when any entity you are engaging with online is not playing by the rules is very important.

Considering the enormous challenges posed by online transactions, you have been hired as a data scientist by a bank that issues credit cards. One of your job roles is to develop a system that uses purchasing transaction data of the bank customers to detect fraudulent behaviours very quickly, possibly in real-time in such a way that appropriate mechanisms for protecting the bank's clients can be activated.

For this purpose, you are provided with datasets consisting of fraudulent and non-fraudulent purchasing transactions. Specifically, you will design and implement the system by analysing the problem of fraud detection systems, designing, and implementing a solution based on the concepts and principles taught in PCP module.

2. Assignment Key Tasks

The following tasks are to be performed in the assignment:

a) Loading transaction datasets

In this task, you will develop a suitable module for retrieving data from the provided datasets. The data to be retrieved is a nested dictionary containing user transactions, i.e. users and transactions they have reformed (including the location and amount of money spent in the transaction). You are provided with sample datasets and information regarding these datasets is in section 3.

b) Computing user transaction statistics

In this task, you will design and implement 6 functions that use the data retrieved in task(a) to compute basic statistics such maximum, minimum, etc. transactions of any user. Details of those functions are provided in section 4.

c) Computing distance between any two transactions of any user

In this task, you will design a function that computes the **distance between any two given transactions** of any user. Another function should be implemented for computing the **distance of transactions of any two users**.

3. Datasets

In the folder (transactions), you will find three datasets. You are required to extract data from relevant dataset(s), namely the *transaction.txt*, *fraud-description.txt*, and *description.txt*.

Three datasets are provided. The first dataset (description.txt) contains the description of genuine transactions. The second dataset (fraud-description.txt) contains the description of fraudulent transactions. The last dataset (transaction.txt) contains the actual transactions (10,000). The transaction dataset contains the description of the transactions, the amount of the transactions and the locations of the transactions. The locations are not real location to avoid tracing any transaction to anyone. The locations are just Euclidean coordinates in x, and y points. The following are the attributes of the transaction dataset, also see Figure 1 below.

1. The user id
2. The transaction id
3. The description of the transaction
4. The amount of the transaction
5. The x coordinate of each transaction
6. The y coordinate of each transaction
7. A Boolean label that represents whether the transaction is fraudulent or not.

Figure 1: sample data from the transaction.txt

```
1:1:EXPEDIA TRAVEL:63.29:856.0:717.0:false
1:2:BENTLEY COLLEGE:58.38:995.0:952.0:false
1:3:JERRY'S FAMOUS DELI:49.05:909.0:750.0:false
1:4:MARSHALLS:71.93:710.0:978.0:false
1:5:AMERCN SCTY FOR INFO S:55.36:886.0:958.0:false
1:6:MGH YAWKEY PRKNG GARAG:42.26:772.0:918.0:false
1:7:MACY'S EAST:66.85:967.0:822.0:false
1:8:HOMEGOODS:68.07:877.0:816.0:false
1:9:H & M:12.71:710.0:925.0:false
1:10:HYATT HOTELS MIAMI F&B:46.11:746.0:973.0:false
1:11:BORDERS BOOKS:33.97:730.0:759.0:false
1:12:EXPEDIA TRAVEL:57.62:776.0:868.0:false
1:13:L'ESPLANADE:56.96:921.0:849.0:false
1:14:FAST LANE:47.6:813.0:713.0:false
1:15:FAST LANE:31.5:942.0:797.0:false
1:16:CVS PHARMACY:67.02:882.0:965.0:false
1:17:VERIZON WEB:53.74:989.0:833.0:false
1:18:PEETS COFFEE/TEA:50.94:703.0:965.0:false
1:19:BLACK DIAMOND COFFEE:80.62:942.0:772.0:false
1:20:BORDERS BOOKS:55.68:782.0:853.0:false
1:21:EXPRESS:50.43:860.0:849.0:false
1:22:REDBOX DVD:36.87:981.0:832.0:false
```

Note that you may not need the three datasets for this assignment. The most important dataset is the *transaction.txt*

4. Implementations

As discussed in the assignment key tasks (section 2), you are expected to design and develop 3 modules namely *load_dataset_module*, *user_statistics_module* and *test_module*. Please see requirement of the modules below.

- I. You will implement a *load_dataset_module* with a function that retrieves the above attributes and returns a dictionary. You must use python file objects. **Please do not use any other libraries for this task.** If you do, you will lose the marks for this task.

- II. You will implement a ***user_statistics_module***. This module will contain the following functions:
- a) A function that returns the maximum and minimum amount of transactions of any user.
 - b) A function that returns the location centroid of any user based on their transaction locations.
 - c) A function that computes the distance of any transaction location from the centroid, based on the computed centroid in 2(b).
 - d) A function that computes the standard deviation of transactions of any user.
 - e) A function that computes the variance of any given user's transactions.
 - f) A function that determines whether a transaction is fraudulent or not. It should provide details of such transactions.
- III. Implement the main module (***test_module***) containing the function that implements the user interface through which users can query and interact with all functions in 2.

NOTE: There will be loss of marks for not implementing these 3 modules. Please do not use any other libraries for this task. If you do, you will lose the marks for this task.

5. Pay attention to the following requirements

- a) This assignment is an individual piece of work, and your submission must be in the form of modules or Jupyter Notebook file. We should be able to open and run your modules on a standard campus computer.
- b) You will submit a report. The report should provide analysis of the problem being solved, justification for your design decisions and pseudocodes for the functions (for example, the statistics functions) you have developed. It should explain the relationships between the modules. A good report should provide **evidence of critical analysis of the implemented system**. Even if your application does not work correctly, you should still submit the report explaining what you have done, what works and what has not worked. If you have any questions regarding the report, please consult with your tutors.
- c) You are required to submit a video demonstration of your implementation. This should not be longer than 15 minutes.
- d) You have been provided with 3 indicative files representing the modules to be implemented. Also included is the folder containing the datasets for your implementation.

- e) Any evidence of collusion/plagiarism will be penalised! If there is some doubt about the authenticity of a particular piece of work, then the person submitting it will be expected to defend such work, including reasons for the programming decisions taken. You must document with references any use of libraries or existing code in your report.
- f) Appropriate use of variable names for clearer understating of your code is important.
- g) Adequate commenting of your code for easier understanding during demonstration/grading is also important.

6. Report

You will write a report of your implementation (Ms word or PDF file), summarising your implementation processes, decisions and justifications, and detailed instructions on how to execute your application. Your report should also contain other details of your program e.g., design/pseudocode/algorithms for your functions/system. In addition, your report should include a reflection section of your experience while executing this project. The reflection should detail what went well or not and lesson learnt. What would you do differently if you have another chance to execute this project again?

7. Submission Process

- Your assignment should be submitted electronically through the module's Blackboard site as a **single ZIP file** that contains **all source code, video demonstration (no longer than 15 minutes) and a report, not more than 5 pages.**
- In addition, you are required to submit a copy of your REPORT only to Turnitin submission point
- If your video is longer than 15 minutes, we will stop watching it during grading at exactly 15th minute.
- Check your upload to ensure you have submitted the correct files successfully as any issues will not be considered after the deadline.
- **Note that late submission will attract penalty. The penalty is capping of your mark to 50%**
- **You must also check your report's similarity score using Turnitin on the Blackboard before final submission. Please do not submit any report with similarity score higher than 20%. Otherwise, you will be penalized for plagiarism or collusion.**

8. Assessment Criteria

This assignment assesses the module's learning outcomes (LO) in the following way:

- This assignment focuses on designing and implementing python a module for loading and extracting data from real-world datasets and a module for computing statistics, using appropriate data structures, string manipulation, iteration, selection, etc. (LO1) and programming design strategies (LO2).
- The coursework will be assessed against the Learning Outcomes (LOs) using a set of assessment criteria. This set of assessment criteria allows assessing how successful a student has met the LOs. In order to ensure consistent application of the relevant criteria, the assessment criteria are summarised in the following assessment matrix and grid. This is an indicator of how the marks will scale across each category of the learning outcomes it covers.

Note that the University's common grading descriptor in Table 3 will be used to determine marks within each area shown in Table 3. The marking scheme embeds the concept of extended work by rewarding only the highest marks to those who demonstrate evidence of independent investigation, learning, and thought. Thus, to achieve top grades, you will need to go beyond the materials presented in lectures and labs and undertake some of your own research (i.e. read and discuss related materials).

Table 1: Assessment Matrix

Assignment	Assessment Criteria	Marks	Learning Outcomes		
			LO1	LO2	LO3
Assignment 1	Basic Understanding programming concepts (/10)	10%	X	X	
	Application of programming processes to solving problems (/15)	15%	X		X
	Implementation of software solution using appropriate programming concepts(/25)	25%	X	X	X
	Quality and usefulness of the submitted mini report (/10)	10%	X	X	

Table 2: Assessment Marking Grid

Fail (<50%)	Pass (50-59)	Merit (60-69)	Distinction (70% +)
Basic Understanding of programming concepts (/10)			
No evidence of understanding of problem being solved. No solution requirements.	Good evidence of understanding of the problem, its definition and analysis. Good understanding of important solution requirements, may lack some clarity, missing requirements, etc.	Very good demonstration of knowledge and understanding of the problem. Very good understanding of the system requirements and its analysis.	Excellent and professional demonstration of understanding of the problem being solved. Excellent details of solution requirements and analysis of the programming problem.
Application of programming processes to solving problems (/15)			
No application of software development processes.	Good use of appropriate software development processes. Minor issues, such as lack of use of some software development processes such as missing pseudocodes or architectural figures/flowcharts	Very good use of appropriate software development processes. Use of pseudocodes, algorithms, flowcharts, use of figures e.g. high-level architectural diagram showing key aspects of your solution, etc.	Excellent and professional use of the program development processes.

Implementation of solution using relevant programming concepts (/25)			
Lack/poor use of programming concepts.	Good understanding and use of appropriate programming concepts such as selection, repetition, sequences, functions, and parameters, etc. Use of appropriate data types, data structures, etc.	Very good understanding and use of appropriate programming concepts such as selection, repetition, sequences, functions and parameters, exception handling, etc. Software does not crash when being run. Entering wrong input does not crash the program but gracefully handles such exceptions, etc.	Excellent, professional level use of programming concepts in the implementation, robust application delivered.
Quality and usefulness of mini-report and video demonstration (/10)			
Report lacking good structure, no personal reflection, no description of the deliverable or explanation and justification of decisions. Poor use of language. No/minimal walkthrough of implementation through video demonstration	Good structure, some evidence of personal reflection on what went well or not. Good justification for design and implementation decisions. Good use of language. Good walkthrough of implementation through video demonstration	Very good structure, evidence of personal reflection on what went well or not. Very good justification for design and implementation decisions. Good use of language. Very good walkthrough of implementation through video demonstration	Excellent structure, excellent personal reflection on what went well or not. Good justification for design and implementation decisions. Excellent use of language. Excellent walkthrough of implementation through video demonstration. Evidence of innovation in the deliverable, e.g. excellent user interaction through UI, etc.

Table 3: Level 7 RUBRIC for Grading

Class	Mark range	CG %	General Characteristics
DISTINCTION (Excellent)	93 - 100	96	Exceptional breadth and depth of knowledge and understanding evidenced by own independent insight and critical awareness of relevant literature and concepts at the forefront of the discipline; evidence of extensive and appropriate independent inquiry operating with advanced concepts, methods and techniques to solve problems in unfamiliar contexts; Cogent arguments and explanations are consistently provided using a range of media demonstrating an ability to communicate effectively in a variety of formats using a sophisticated level of the English language in an eloquent and professional manner to both technical and non-technical audiences; a sustained academic approach to all aspects of the tasks is evidenced; academic work extends boundaries of the disciplines and is beyond expectation of the level and may achieve or be very close to publishable or commercial standard.
	85 - 92	89	Excellent knowledge and understanding evidenced by some clear independent insight and critical awareness of relevant concepts some of which are at the forefront of the discipline; evidence of appropriate independent inquiry operating with core concepts, methods and techniques to solve complex problems in mostly familiar contexts; Arguments and explanations are provided that is well-supported by the literature and in some cases uses a range of media demonstrating an ability to communicate effectively in a limited number of formats using own style that is suited to both technical and non-technical audiences; a sustained academic approach to most aspects of the tasks is evidenced; one or more aspects of the academic work is beyond the prescribed range and evidences a competent understanding of all of the relevant taught content.
	78 - 84	81	
	70 - 77	74	
MERIT (Good)	67 - 69	68	Very good knowledge and understanding is evidenced as the student is typically able to independently relate taught facts/concepts together some of which are at the forefront of the discipline; evidence of some competent independent inquiry operating with core concepts, methods and techniques to solve familiar problems; Arguments and explanations are provided that are typically supported by the literature and in some cases may challenge some received wisdoms; competently uses all taught media and communication methods to communicate effectively in a familiar settings; an academically rigorous approach applied to some aspects of the tasks is evidenced; some beyond the prescribed range, may rely on set sources to advance work/direct
	64 - 66	65	
	60 - 63	62	

			arguments; demonstrates autonomy in approach to learning.
PASS (Satisfactory)	57 - 59	58	Good knowledge and understanding of the area of study balanced towards the descriptive rather than critical or analytical and mostly confined to concepts that are not at the forefront of the discipline; evidence of some independent reading and research to advance work and inform arguments and approaches; Arguments and explanations are limited in range and depth although some are adequately supported by the literature albeit descriptively rather than critically; competently uses at least one taught media and communication method to communicate appropriately in familiar settings; although the approach applied to some aspects of the tasks may lack academic rigour, there are some clear areas of competence within the prescribed range. Relies on set sources to advance work/direct arguments and communicated in a way which shows clarity, but structure may not always be coherent.
	54 - 56	55	
	50 - 53	52	
FAIL (Insufficient)	40 - 49	45	Knowledge and understanding is marginally insufficient as the student is typically only able to deal with terminology, basic facts and concepts; Adequate knowledge of concepts within the prescribed range but fails to add meaningful detail or make sufficient links between concepts and facts to adequately solve problems posed by the assessment; some ability to independently select and evaluate reading/research however there is a strong reliance on set sources and to provide descriptive and unsubstantiated arguments/methods; communication/presentation is competent in places and at a threshold level as it fails to demonstrate clarity and focus; inability to adequately define problems and make reasoned judgements; The general approach to tasks lacks rigor and where there is competence and rigor, it is not sustained.
	30 - 39	35	
	20 - 29	25	Knowledge and understanding is highly insufficient as the student is unable to evidence any meaningful understanding of two or more taught concepts or methods; very limited evidence of reading and research to advance work; inadequate technical and practical skills as the student is unable to use and apply such skills to address problems or make judgements; limited or lack of understanding of the boundaries of the discipline and does not question received wisdom; approach to learning lacks autonomy and approach to tasks is not sustained; inability to communicate coherently.
	10-19	15	
	1-9	5	
ZERO	0	0	Work of no merit OR absent, work not submitted, penalty in some misconduct cases.

*All work must be your own. If evidence of collusion/copying is found, then **such collusion will be penalised, severely if appropriate!** If there is some doubt about the authenticity of a particular piece of work, then the person submitting it will be expected to give a detailed explanation of such work, including reasons for the programming decisions taken.*

You must submit a **demonstration video (not more than 15 minutes)** demonstrating your running prototype system and explaining how it meets the assessment criteria. Please look at the marking criteria and prepare your demonstration accordingly. In case you face any challenge in video recording, please run your system and provide screencast explaining how your system satisfy the assessment criteria,

WARNING

All work must be yours. If evidence of collusion/copying/plagiarism is found, then such collusion will be penalised, severely if appropriate! If there is some doubt about the authenticity of a particular piece of work, then the person submitting it will be expected to give a detailed explanation of such work, including reasons for the programming decisions taken.