Student Data Analysis

A Python program for academic data analysis of students

Student:

Abhilash Reddy Peram (10532966)

Sabitha Maram (10533048)

Srikanth Shilesh Pasam (10387794)

Teacher:

Paul Laird

Course:

Programming for Information Systems (B9IS123)

Table of Contents

[Introduction 2](#_Toc29219118)

[Appendix 1 2](#_Toc29219119)

[Appendix 2 2](#_Toc29219120)

[Bibliography 2](#_Toc29219121)

# Introduction

The Student Data Analysis is a program which is designed to analyze a students or classrooms academic performance. It has been designed with the purpose of addressing the needs of Teach For India. Teach For India is a non-government non-profit organization and is part of Teach For All network. The organizations aim is to tackle the problems of education inequity in the country by providing excellent education to every child. As a part of this organization, qualified and motivated Fellows are placed in low-income private schools or under resourced government schools to drive change. During this time, the fellows collect large volumes of student and school data ranging from academic performance, culture and values of the class, social and economic levels of the parents, teacher resources, training modules, infrastructure of the school, school income and expenditure, external stakeholder investments and much more. All this data is used by the fellows to analyze and create action plans for the improvement of the quality of education received. Currently the organization has 900 fellows, 250 staff and 3000 alumni across 7 cities in India impacting 260 schools and 32000 students directly. As of this day, the organization only uses spreadsheets in order to record and analyze all the student data collected. This task gets increasingly difficult every year as the organization keeps expanding and reaches out to more and more schools and students. The efficiency with which all this data can be analyzed and used is severely limited to the scope and extent of what a spreadsheet can do.

This project aims to tackle a small portion of this issue as a part of this assignment. The program takes in the raw student data from the existing spreadsheets (.xlsx or .csv) runs them through the program to provide an analysis of the student or classrooms academic performance. The current program only focuses on academic analysis within the scope of this CA but plans are made to integrate other aspects like deployment in the Cloud, a backend database system, web interfacing along with a mobile app for GUI in the future.

# Requirements

## Functional Requirements

1. As the School Leader, I need to see a summary of the marks of each subject for each grade so that I can identify the areas of development for the respective grade. (Reporting feature)
2. As a School Leader, I need to see a summary of the marks scored by each student for a subject so that I can identify the areas of development for the respective subject teacher. (Reporting feature)
3. As a Class Teacher, I need to see a summary of the marks scored by each student for all subjects so that I can give feedback to students individually. (Search feature)
4. As a Class Teacher, I would like to be able to add or edit students data into the system. (Entry/update feature)
5. As a Subject Teacher, I want to see a summary of the marks scored by students in my subject so that I can identify which student requires extra classes scheduled.
6. As a Parent, I need to see a summary of the marks scored by my child so that I can keep a track of his/her academic progress. (Search feature)
7. As an external stakeholder, I need to see a summary of the growth of a classroom over 3 years in the form of graph so that I can decide if I should continue funding Teach For India. (Reporting feature)
8. As the City Director of Teach For India, I want access to the Student Data Analysis System to be provided only after proper authentication has taken place because it contains sensitive information of school and students. (Authentication feature)

## Non-Functional Requirements

1. As a fellow who frequently accesses the system, I want the program to load up quickly so that I can finish up my work faster.
2. As the School Relations Manager of Teach For India, I want the graphs to use the Teach For India’s blue color so that it matches the impact growth document shared with the stakeholders.

# Student Data

The organization uses spreadsheets to store their data. For the purpose of the CA, the organization agreed to let us use few of their historic data sets. They have been included in the CA submission under the ‘Source’ folder. From this data sets, eight CSV files have been created which will act as the raw data file for the program demonstration. These CSV files have the student roll, student name, gender, RC (reading comprehension), listening, writing and math columns. The grading system for each subject is different. RC levels range from -0.5 to 5.5, writing starts at 0 and goes all the way to 8, listening is from 0 to 5 and math is percentage based. There is much more data in the source files but for demonstration purposes we will be limiting the scope of the program to only these elements.

# Program Features

## Stage 1 – Login/Register

The program uses CLI as the interaction tool with the user. On executing the program, the user will be asked to either login or register.

* In both the login and register phases the user details are encoded and encrypted for security.
* The password input is taken using the GetPass library function which hides the password entry.
* The encryption is done using the cryptography library Fernet which uses symmetric encryption.
* The Fernet key is stored locally in the project folder under the file name 'Fernet\_key.txt’ for the sake of demonstration.
* The registered user details are encoded into byte code format. This is then encrypted using the Fernet Key created. All this data is stored locally under the file name ‘User\_details.txt’ for the sake of demonstration.

## Stage 2 – Loading files

Once the user logs in successfully they are granted access to the program.

* To access the files on the system the program uses Tkinter library functions.
* The Tkinter provides a dialogue box for the user to select the directory where the files are located.
* Once the user selects the directory, all the files in that directory are listed using the OS library function.
* The user then needs to select a file from that list which needs to be processed.
* After the file is selected, the program loads the CSV file into a DataFrame using the Pandas library functions.

## Stage 3 – Data Analysis

Once the files are loaded, the program can perform data analysis on it depending on what the user wants to do.

* The user has the option to:
  + - * + Calculate the class averages of each grade.
        + Calculate individual subject average for each grade.
        + Search for a student using their full name or part of their name and display their corresponding subject-wise marks.
        + Using the MatPlotLib library function to plot the graph which show the classroom growth for each subject over the span of three years.
        + Plot the graph to show an individual student growth for each subject over the span of three years.
  + The graphs which are generated post the analysis is saved locally in the 'Graphs' folder using a certain naming convention for user access.
  + The user can add new student details or edit existing ones.

# Implementation

The working code is included in Appendix 2 for reference. The implementation of the program is shown using screenshots.

# Individual Contributions

## Srikanth Shilesh Pasam

The three of us as a team contributed towards all the elements of the code. Each aspect has been discussed and brainstormed. This collaboration has helped us to learn from each other. Though inputs were taken from all team members, my major contribution for the CA is as follows:

1. Functional and non-functional requirements
2. User choice logic flow
3. LoadData() class
4. Avg() class
5. SubjectMarks() class
6. StudentMarks() class
7. Auth() class
8. Keys() class
9. Report writing
10. Power point presentation

## Reflection

I come from Electronics and Communication Engineering background with no prior knowledge of the subject. My professional experience has been majorly in Teach for India, a non-government organization within the education sector. So, I have no professional experience working with Python or any other form of programming language either. Because of this I had to put in extra efforts to learn everything form the scratch. This proved extremely difficult during the initial days of the course where I had to learn a new language for each subject in the module. Nonetheless I had to push myself as this was a challenge I willingly signed up for. A major setback I faced was during the first in-class test. I performed poorly and this was my wakeup call to roll up my sleeves. I took it upon myself to study and improve my command over Python as much as possible. The time I got during reading week proved very useful for this. I signed up for courses on Udemy and spent major of my time practicing python and trying to perfect the program I failed to do during the in-class test. Through a lot of hard work and perseverance I slowly but steadily started getting more comfortable and fluent with Python. I was able to complete the program and submit it nine days early on Moodle. This built a good sense of accomplishment within myself and motivated me to push further and harder.

Most of us in class had difficulty during the initial days with the subject due to various reasons and many a times we have let him down. But professor Paul has been extremely patient with us. He taught us from the scratch, helped us practice, shared resources for self-study and even followed up with our progress on Boot Camp. All this constant push and effort is what kept me motivated to try harder. Being a teacher myself I can personally relate to the amount of efforts that go into balancing time crunch in covering the syllabus and making sure the students are able to keep up with the pace. This only made me push myself and put in that extra hour every night. I am especially grateful to professor Paul for introducing me to GitHub. Coming from non tech background I had absolutely no awareness about it. But since then I have been very active on GitHub. I have even used it for all my other modules. In fact, I even based part of my CA in Network and System Administration module on GitHub using the Continuous Integration and Continuous Deployment concept.

This project idea has an association with me on a very personal level. I have been part of Teach For India for many years. My deep desire to impact change in the society by being an advocate for ed-equity is what made me choose this line of work. During my time in the organization I have noticed that one particular area of development here is the requirement of a proper data analysis system. Even in this current age of technology, the organizations vision was severely hampered by the lack of this system. This realization is also one of the major reasons for me to choose this particular course. I intend to do my final year thesis project around developing a system that can address these issues for Teach For India and donate the finished application. To this end I am confident that the module –Programming for Information Systems, the programming language – Python and the professor – Mr. Paul Laird has been a tremendous help in helping me get closer to achieving my goals.

## Sabitha Maram

It was really good experience working in group where we can share ideas and knowledge of different people and put them together. It was really a tough and challenging experience for me to learn python programming as I am not into programming in my previous Education.

For this Assignment we discussed and decided to create Student Data Analysis. The key features in this work are User Authentication, Calculating Class averages and student averages, Visualizing student and class average in the form of pictorial representation in the form of graphs and Student data entry. The area where i worked individually is Graph plots and tried to do as much I learned. Srikanth in my group helped me to go through this to the end.

## Abhilash Reddy Peram

This is very good learning experience with a new group of people who are our fellow members in class. Sitting all together and sharing knowledge on whole new subject with each other is very useful. We learned python it through Udemy classes and worked on the assignment whenever we got time while completing the other assignment challenges. And even data camp. Initially it’s a little bit slow to learn and catch up with each other eventually we got along with the assignment.

We have gone through the requirements and figured out what to be done with the data files we have. So, we chose to deal with student data files based on their academic performance. The actual data sets are gathered from the existing data base of a school by our group member Srikanth because has worked in teaching profession for a while in the past. We tried all the possible ways to meet the given requirement criteria in the module. After getting all the data files we all sorted out our work with each other. I have taken the authentication and data entry for the data files. Working on authentication is quite fun and interesting thing to learn like giving access to the users so that only they can have the access with the decryption side. And data entry is like updating the new student data if any student profile gets missed out. While I got struck with some problem in using keys and encryption my group members helped me with the problems and we worked together if any problem raises for any of the team members.

Finally, while coming to the module learning its very good experience to interact with the lecturer. This is the part I like the most as we came from the places where we have little bit interaction with the lecturer but, here MR. Paul keep on persist us to speak up and ask for the doubts and even ask us to share our ideas with him. We are very much great full for having MR. Paul as our module lecturer.

# Appendix 1

The GitHub link for our project is:

<https://github.com/sabithamaram/PIS_B9IS123_CA2>

We have not realized until it’s too late that all our individual commits were being made to the Master folder directly where instead we needed to fork the Master directory, commit to the forked directory and later on merge it to the Master. Hence, we all have the same GitHub link to share though each of our individual commits are visible in the ‘Insights’ section of the GitHub folder shared above.

# Appendix 2