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 aims 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 are saved locally in the 'Graphs' folder using a certain naming convention for user access.

# Implementation

# Appendix 1

# Appendix 2

# Bibliography